

**COMPREHENSIVE PHYLOGENETIC ANALYSES OF ORCHIDACEAE
USING NUCLEAR GENES AND EVOLUTIONARY INSIGHTS INTO
EPIPHYTISM**

Guojin Zhang¹, Yi Hu¹, Ming-Zhong Huang², Wei-Chang Huang³, Ding-Kun Liu²,
Diyang Zhang², Haihua Hu⁴, Jason L. Downing⁵, Zhong-Jian Liu^{2,†}, and Hong
Ma^{1,*}

SUPPORTING INFORMATION

Figure S1

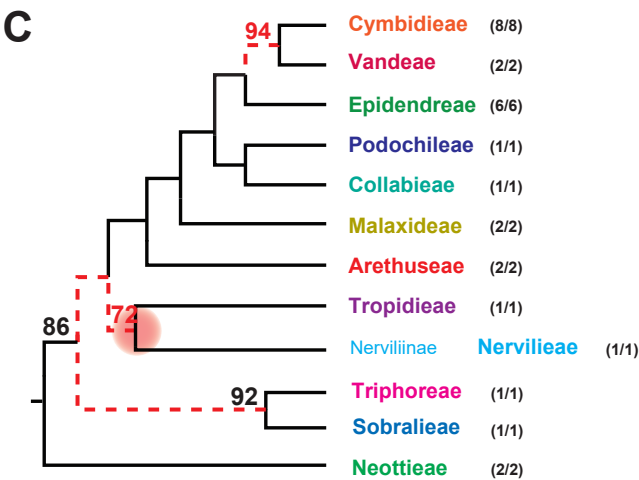
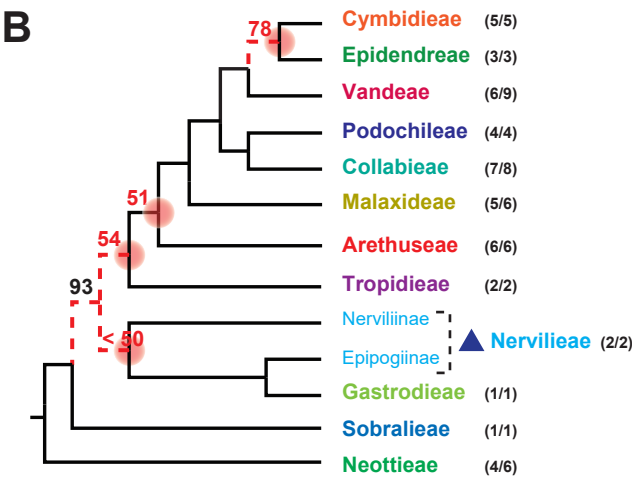
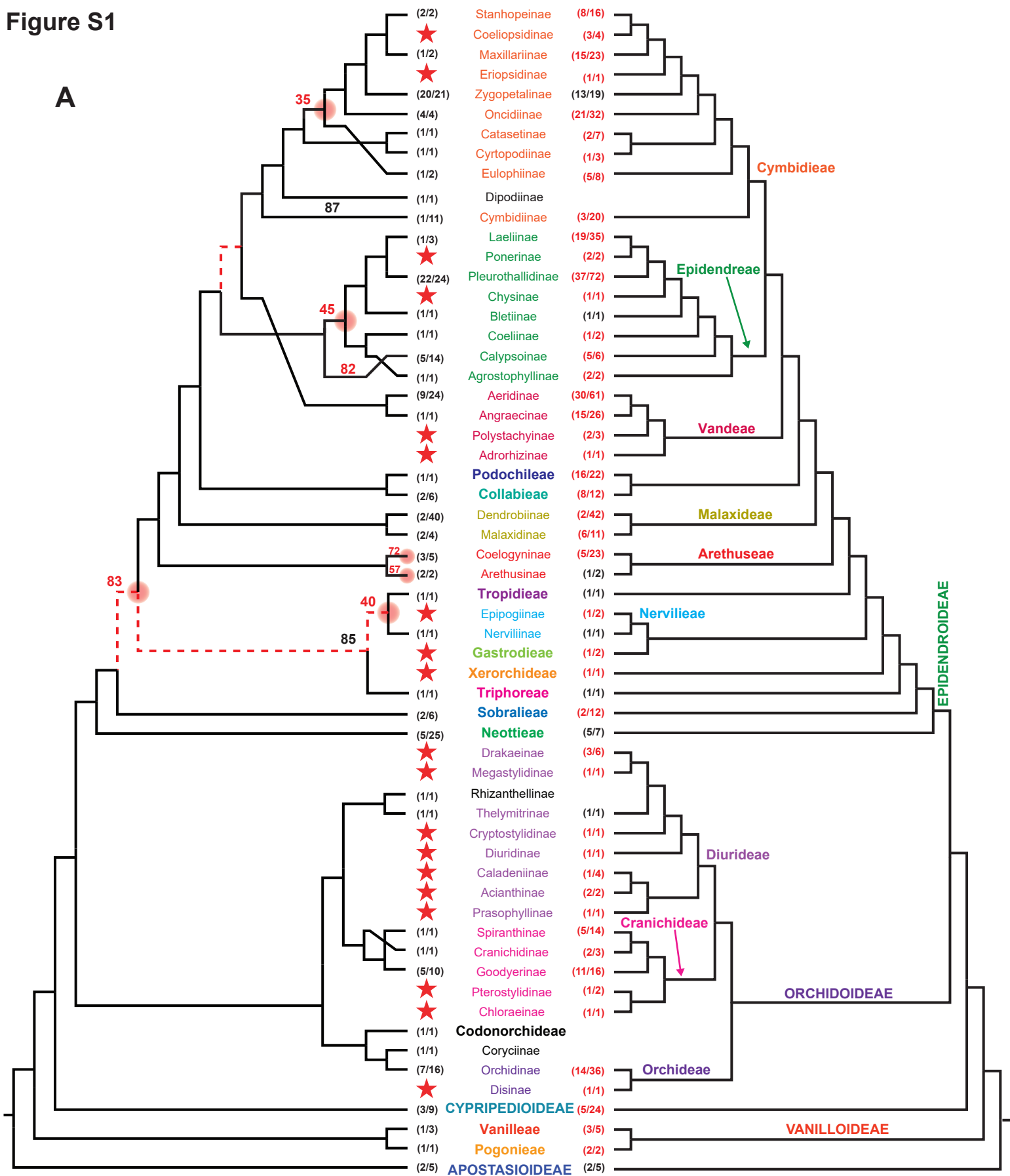


Figure S1. Comparisons of topologies and the sampling of subtribes and higher ranks of Orchidaceae between previous phylogenetic trees using plastomic datasets and the summary nuclear phylogenetic tree in this study.

(A) A comparison of subtribal topologies of Orchidaceae between the maximal likelihood (ML) phylogenetic tree reconstructed using 78 plastid genes from 264 species (left) (Serna-Sánchez et al., 2021) and the summary tree of this study (right). The names of subtribes, tribes, and subfamilies are shown between two trees or on the corresponding branches. Taxa that were not sampled in the previous study are indicated using red stars. The names of one tribe and three subtribes that were not sampled in this study are in black. The number of sampled genus and species of each clade is provided in the parentheses (genus/species) between the branch tip and the taxon name; the red numbers indicate those of sampled taxa here that were greater than those in previous studies. (B) A summary cladogram of Epidendroideae tribes based on the ML tree in a previous study reconstructed using 76 plastid genes from 74 species (Y.X. Li et al., 2019). The non-monophyletic taxon is marked with a blue triangle. (C) A summary cladogram of Epidendroideae tribes based on the ML tree reconstructed using 75 plastid genes from 39 species (Givnish et al., 2015). In all panels, BS values of important nodes are presented on corresponding branches; bootstrap support (BS) values <85 are marked in red, and the corresponding nodes are highlighted with pink circles. Red dashed lines indicate the branches that are inconsistent among the three previous plastid phylogenetic trees.

Figure S2

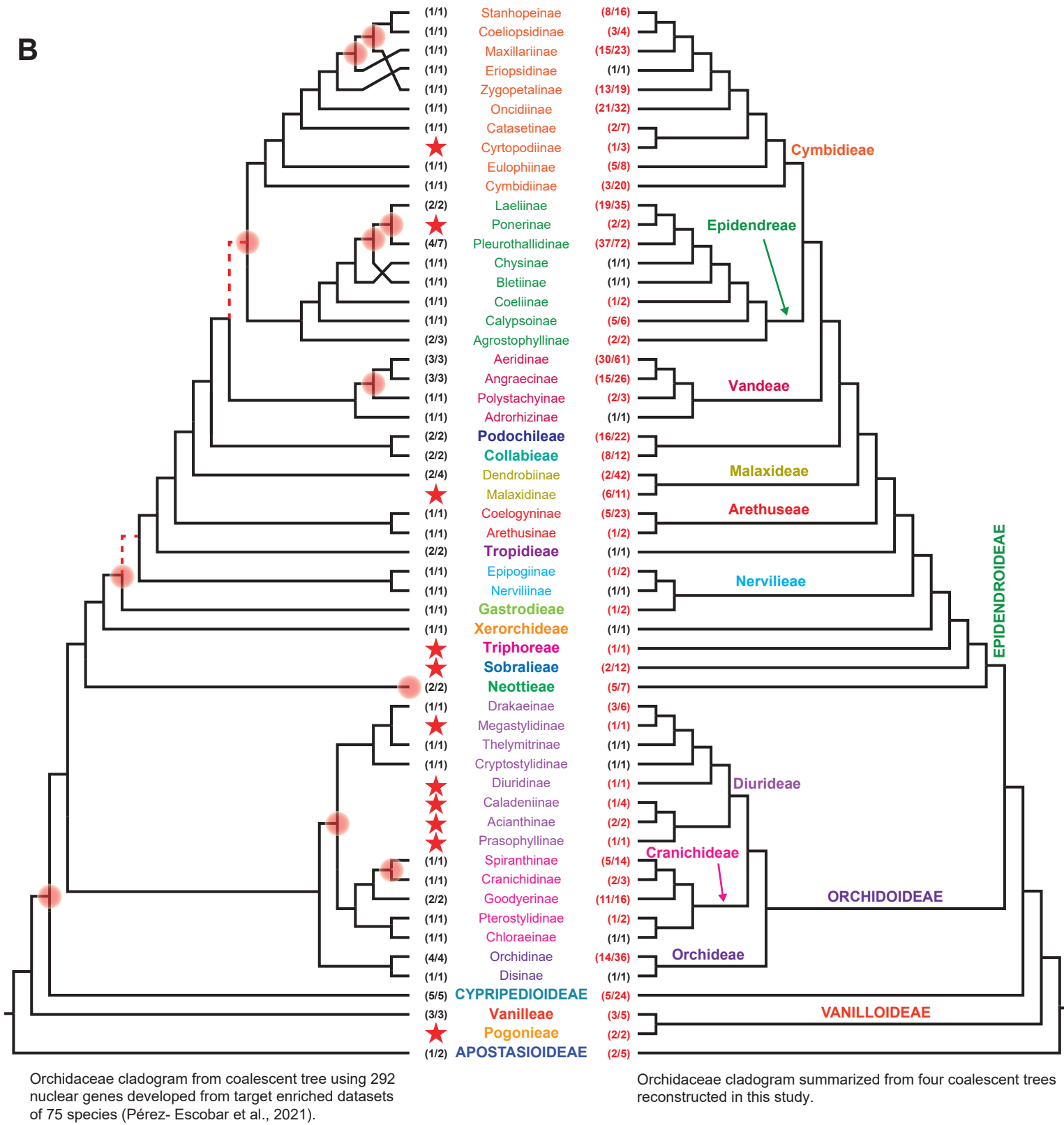
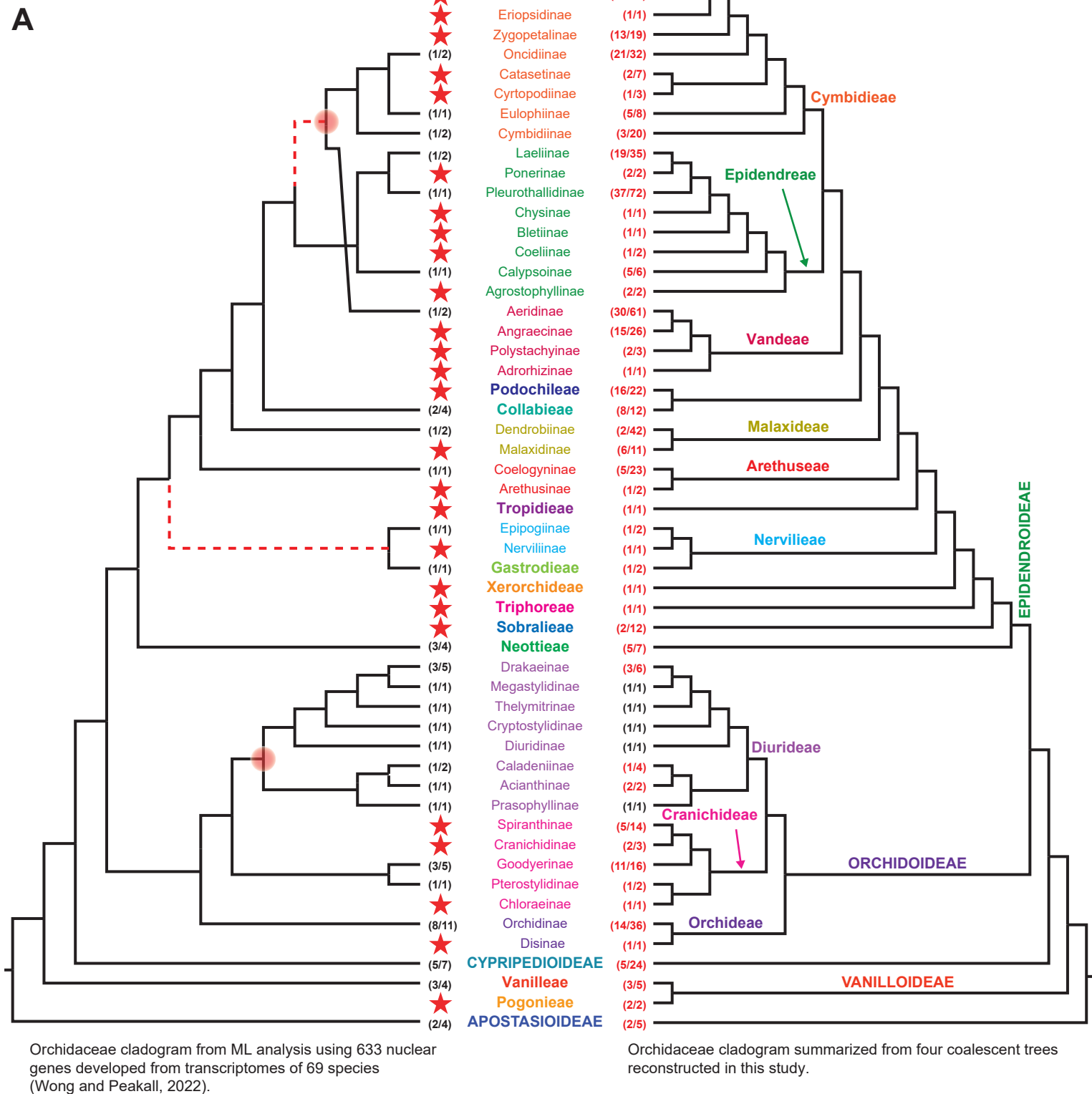


Figure S2. Comparisons of topologies and sampling of subtribes and higher ranks of Orchidaceae between previous nuclear phylogenetic trees and the summary nuclear phylogenetic tree in this study. (A) A comparison of subtribal topologies of Orchidaceae between the ML phylogenetic tree (left) reconstructed using 633 nuclear genes from transcriptomes of 69 orchids (Wong and Peakall, 2022) and the summary tree of this study (right). (B) A comparison of subtribal topologies of Orchidaceae between the coalescent tree (left) reconstructed using 292 nuclear genes from targeted enriched datasets of 75 species (Pérez- Escobar et al., 2021) and the summary tree of this study (right). Pink circles indicate the nodes that have <80 bootstrap support (BS) value in Wong and Peakall (2022) (panel A) and less than 50% quartet support in Pérez- Escobar et al. (2021) (panel B). In both panels, taxa that were not sampled in previous studies are marked using red stars. Red dashed lines indicate the branches that are inconsistent among the two previous nuclear phylogenetic trees. The numbers of sampled genus and species of each clade are provided in the parentheses (genus/species) between the branch tip and the taxon name; the red numbers indicate those of genera and/or species that were greater in this study than previous studies.

phylogenomic studies (Givnish et al. 2015; Li et al. 2019; Kim et al. 2020; Eserman et al. 2021; Perez-Escobar et al. 2021; Serna-Sánchez et al. 2021; Wong and Peakall 2022).

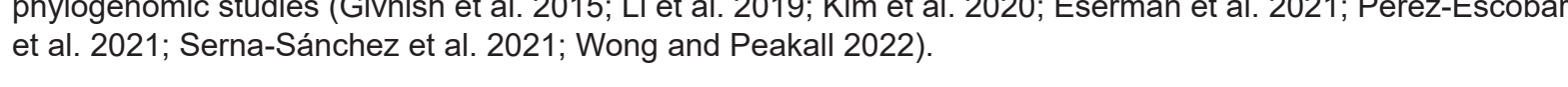


Figure S4

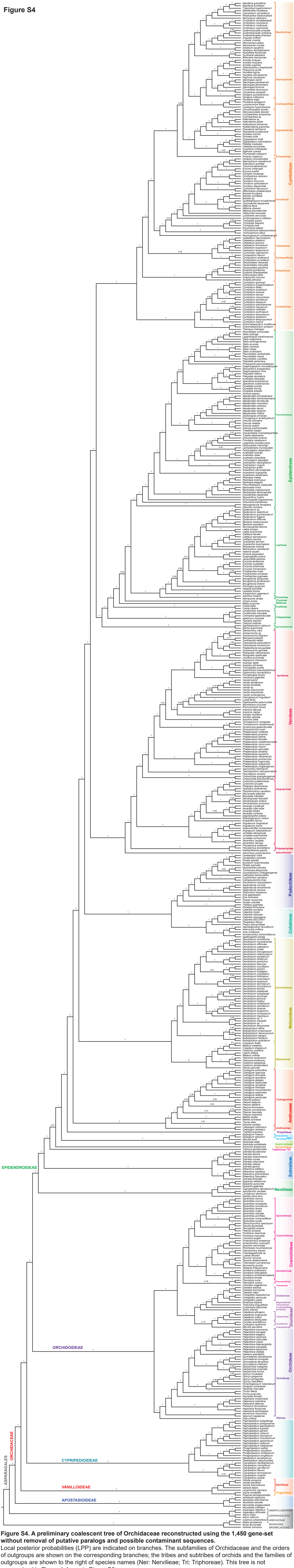
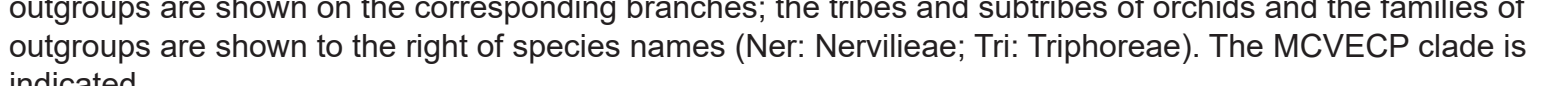


Figure S4. A preliminary coalescent tree of Orchidaceae reconstructed using the 1,450 gene-set without removal of putative paralogs and possible contaminant sequences. Local posterior probabilities (LPP) are indicated on branches. The subfamilies of Orchidaceae and the orders of outgroups are shown on the corresponding branches; the tribes and subtribes of orchids and the families of outgroups are shown to the right of species names (Ner: Nervilieae; Tri: Triphoreae). This tree is not

Figure S5. A coalescent tree of Orchidaceae reconstructed using the 1,195 gene-set.



Multi-locus bootstrapping support values (MBS) (in red) and local posterior probabilities (LPP) (in black) are indicated above and below branches, respectively. The subfamilies of Orchidaceae and the orders of outgroups are shown on the corresponding branches; the tribes and subtribes of orchids and the families of outgroups are shown to the right of species names (Nari: Nariaceae; Tri: Tripterocarpaceae). The MGV5CB clade is



Multi-locus bootstrapping support values (MBS) (in red) and local posterior probabilities (LPP) (in black) are indicated above and below branches, respectively. The subfamilies of Orchidaceae and the orders of outgroups are shown on the corresponding branches; the tribes and subtribes of orchids and the families of outgroups are shown to the right of species names (Ner: Nerveleae; Tri: Triphoreae). The MCVECP clade is

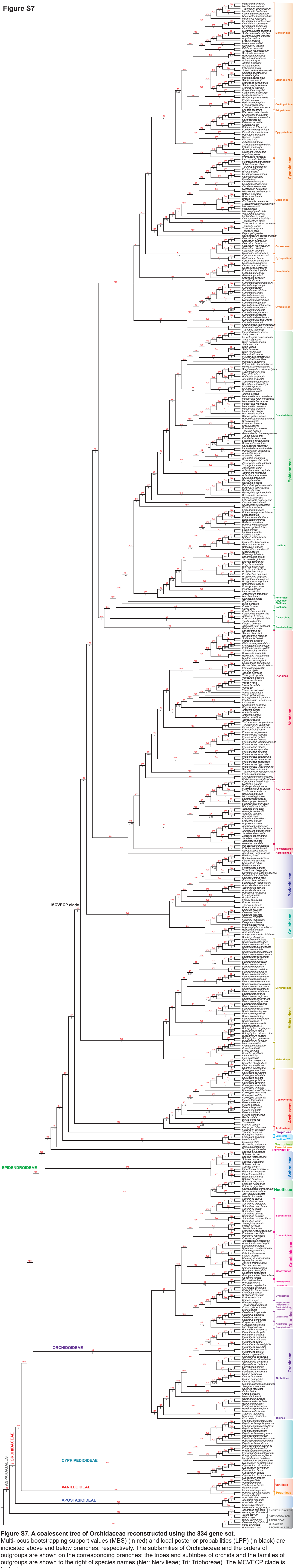


Figure S8. A coalescent tree of Orchidaceae reconstructed using the 639 gene-set. Multi-locus bootstrapping support values (MBS) (in red) and local posterior probabilities (LPP) (in black) are indicated above and below branches, respectively. The subfamilies of Orchidaceae and the orders of outgroups are shown on the corresponding branches; the tribes and subtribes of orchids and the families of outgroups are shown to the right of species names (Nary, Nargilaceae; Tri, Tripharaceae). The MCV/CPD clade is

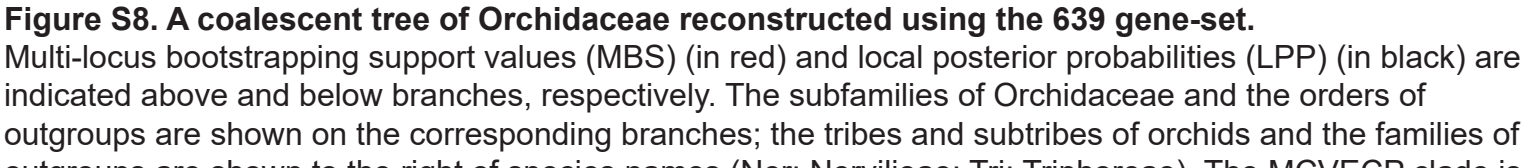


Figure S
A

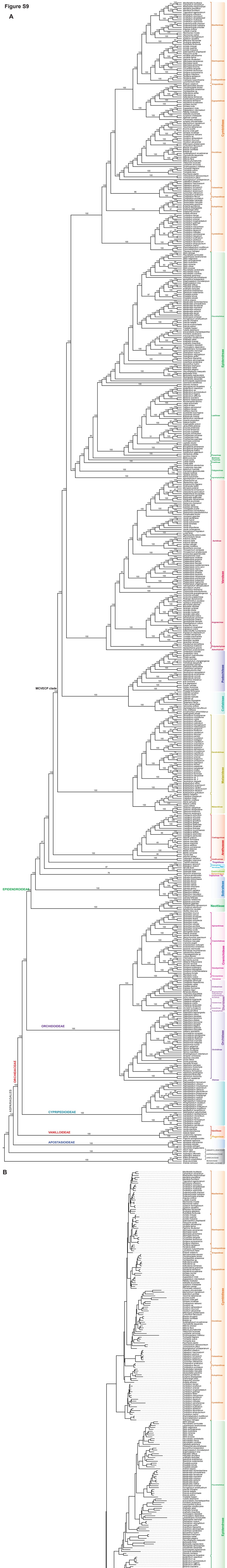
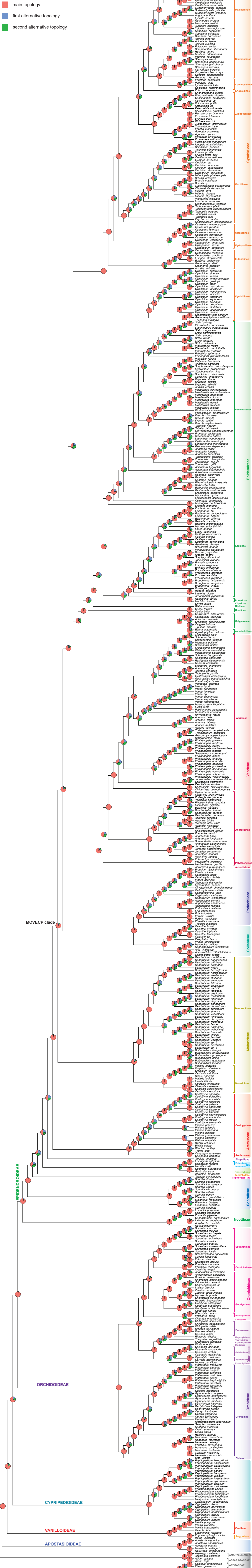


Figure S9. A supermatrix tree of Orchidaceae reconstructed using a concatenated dataset composed of

6



between the coalescent tree and gene trees of the 1,195 gene-set.

subfamilies or Orchidoideae and the orders or outgroups are shown on the corresponding branches, the tribes and subtribes of orchids and the families of outgroups are shown to the right of species names. The MCVECP

Figure S11

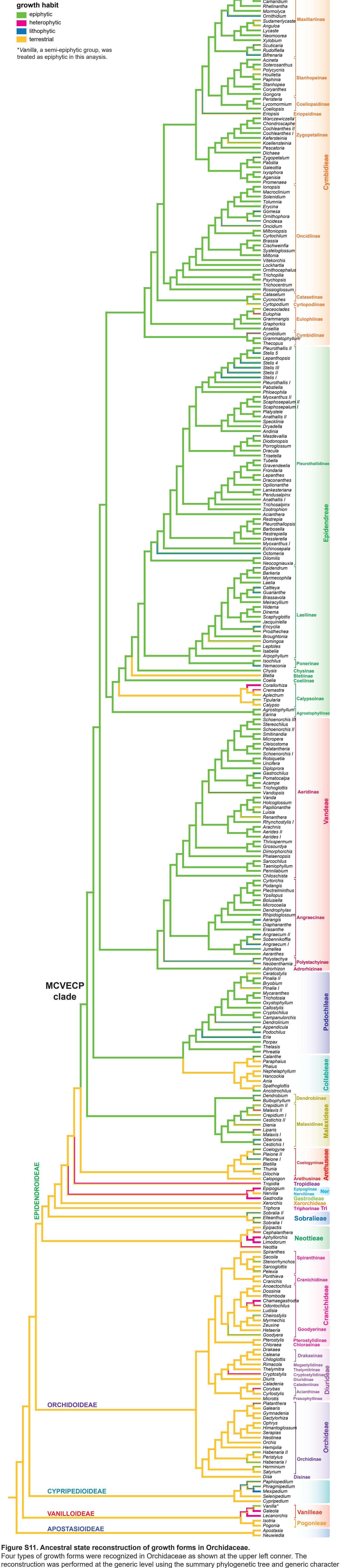
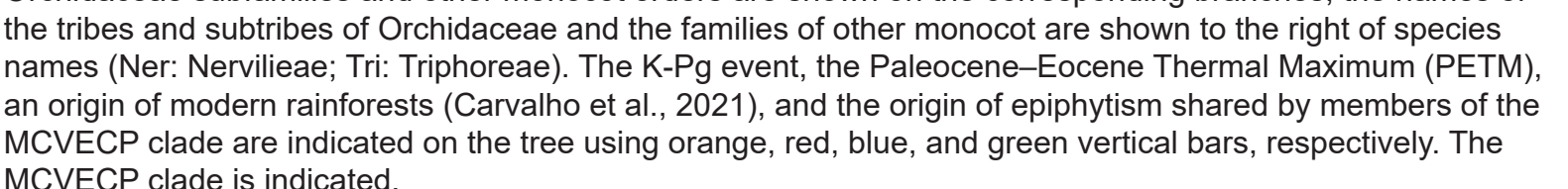


Figure S11. Ancestral state reconstruction of growth forms in Orchidaceae.

Four types of growth forms were recognized in Orchidaceae as shown at the upper left conner. The reconstruction was performed at the generic level using the summary phylogenetic tree and generic character information. The classification information is shown on the right side of the tree and follows the color scheme of Figures 1–4.

the tribes and subtribes of Orchidaceae and the families of other monocot are shown to the right of species names (Ner: Nerviileae; Tri: Triphoreae). The K-Pg event, the Paleocene–Eocene Thermal Maximum (PETM), an origin of modern rainforests (Carvalho et al., 2021), and the origin of epiphytism shared by members of the MCVECP clade are indicated on the tree using orange, red, blue, and green vertical bars, respectively. The MCVECP clade is indicated



supermatrix dataset with 299 genes. The estimated 95% HDP intervals are indicated using blue bar of nodes and showed in brackets. The fossil and secondary calibrations used in the estimation are indicated at the corresponding nodes using red diamonds and circles, respectively. The subfamilies of Orchidaceae and the orders of monocots are shown on the corresponding branches; the tribes and subtribes of orchids and the families of monocots are shown to the right of species names (Ner: Nerveleae; Tri: Triphoreae). The MCVECP

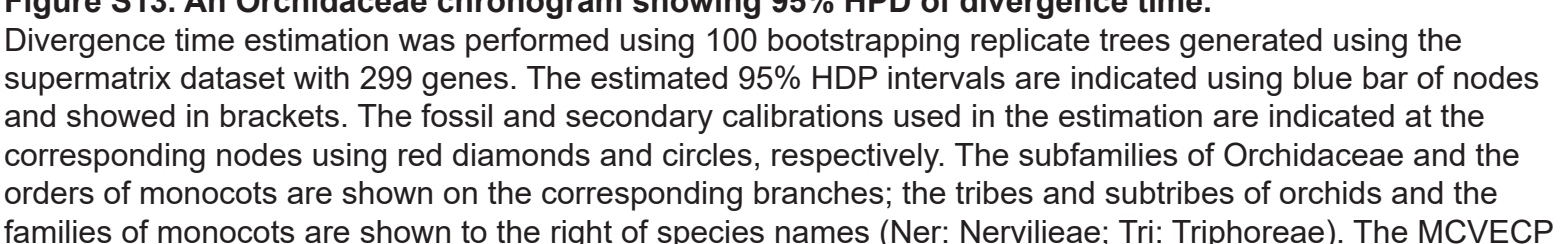


Figure S14

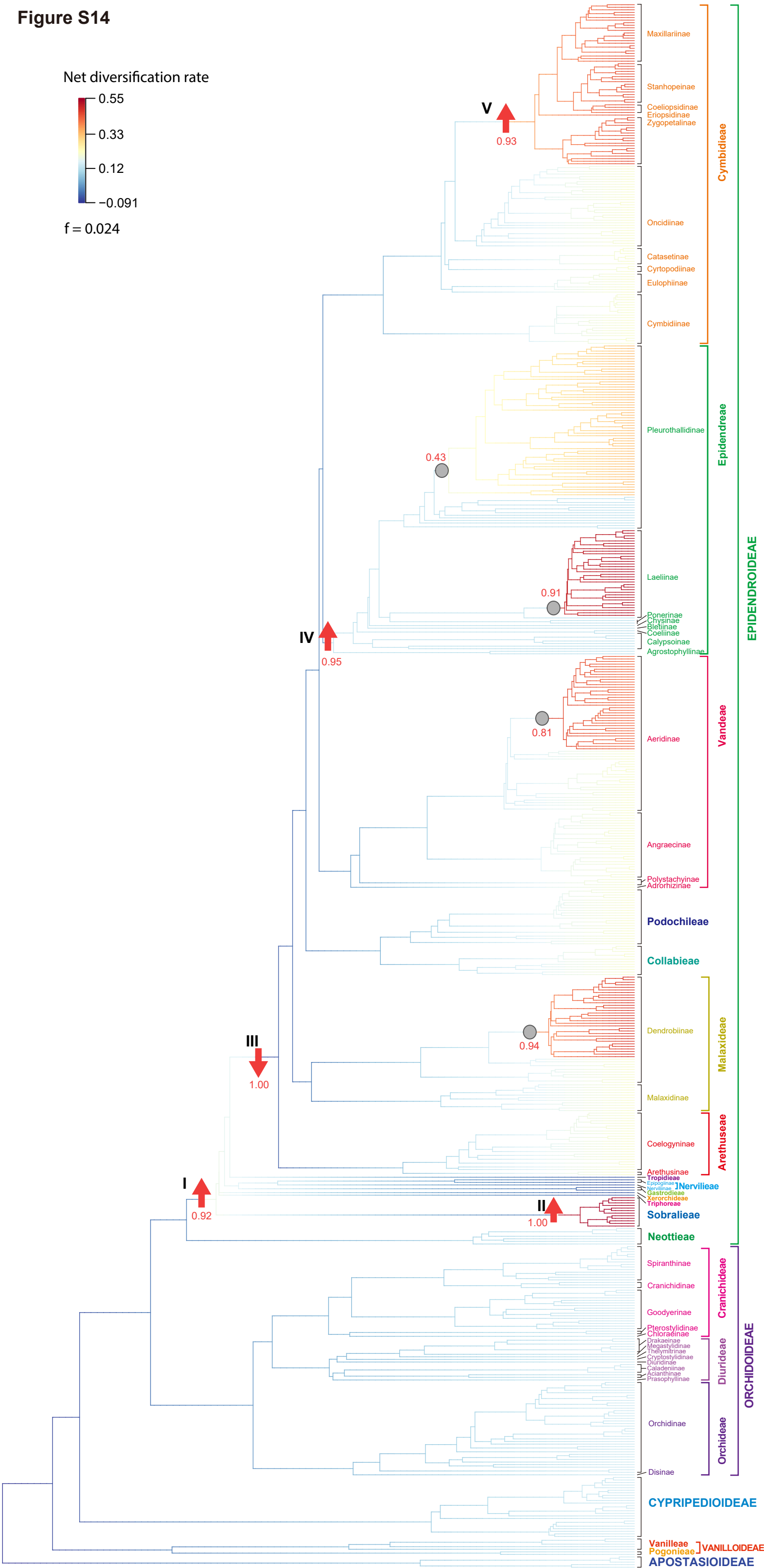


Figure S14. A cladogram of the Orchidaceae with shifts of diversification rate. Shown here is the best configuration (combination of positioned shifts) inferred using BAMM with subtribe-level sampling fractions. The branch colors indicate net diversification rates of corresponding branches as shown in the legend at the upper-left corner. The f value at the upper-left corner indicates the sampled frequency of the best shift configuration in the posterior. The red arrows indicate the five rate shifts (with Roman numerals) detected above the subtribe level, and the four gray circles indicate the rate shifts detected within subtribes. The marginal shift probability of each shift is indicated using numbers in red near the shifts. Names of subfamilies, tribes, and subtribes are shown to the right of the tree, and the color scheme of Figure 6 is followed.

Figure S15

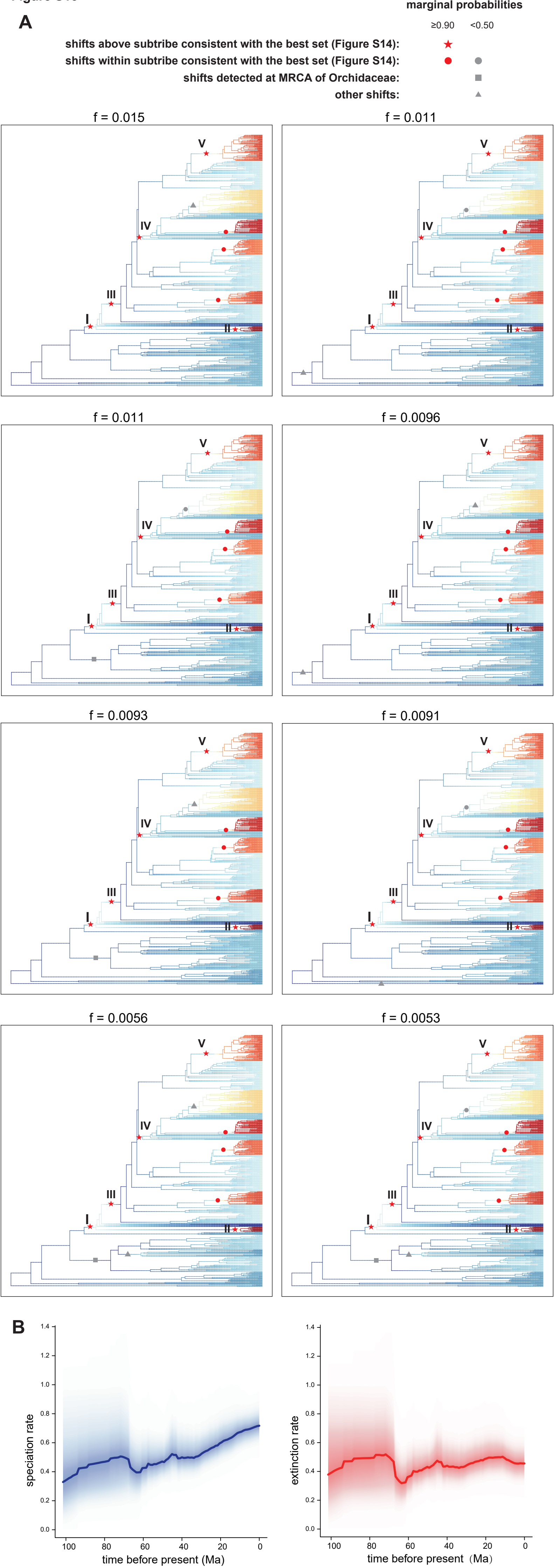


Figure S15. Additional results on shifts of Orchidaceae diversification rates and speciation and extinction rates.

(A) After the best configuration (shown in Figure S14), shown here are the next eight diversification shift configurations of Orchidaceae from the 95% credible set inferred using subtribe-level sampling fractions. The branch colors indicate the net diversification rates of the corresponding branches (warm = high; cool = low). Detected shifts are shown in four types indicated using different symbols, and the marginal shift probabilities are indicated using two colors, as shown in the legend at the top. The five shifts detected in the best configuration (Figure S14) are indicated using the corresponding Roman numerals as those in Figure S14. The f value at the top of each plot indicates the sampled frequency of the corresponding shift configuration in the posterior. (B) The speciation (left) and extinction (right) rate-through-time plots of Orchidaceae inferred with subtribal sampling fractions.

Figure S16

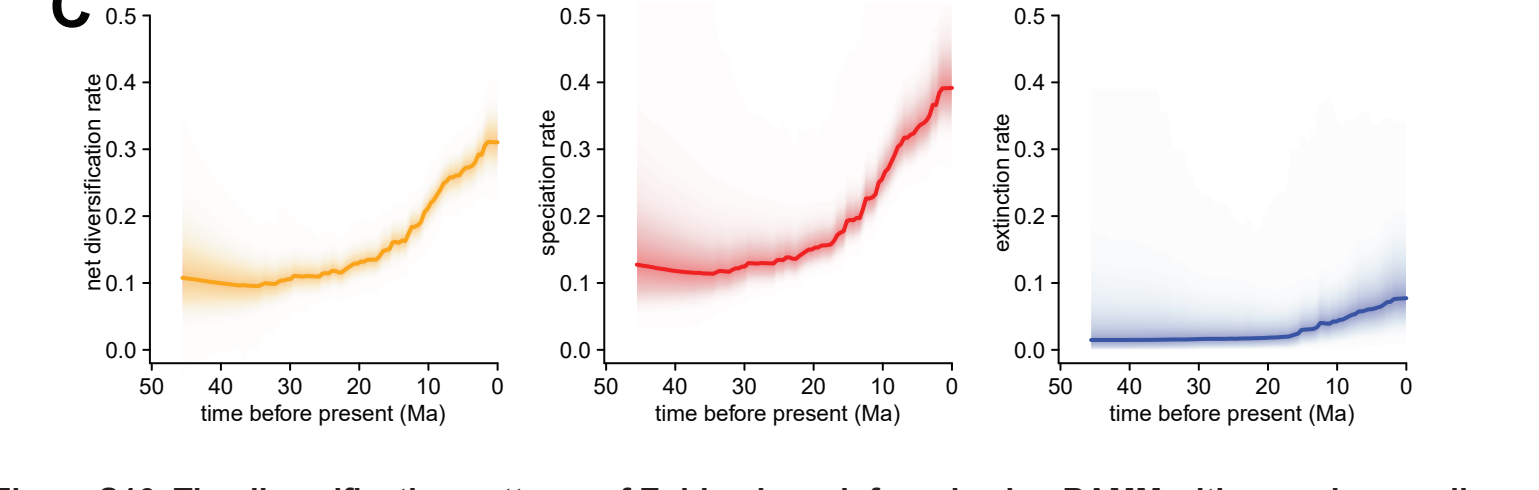
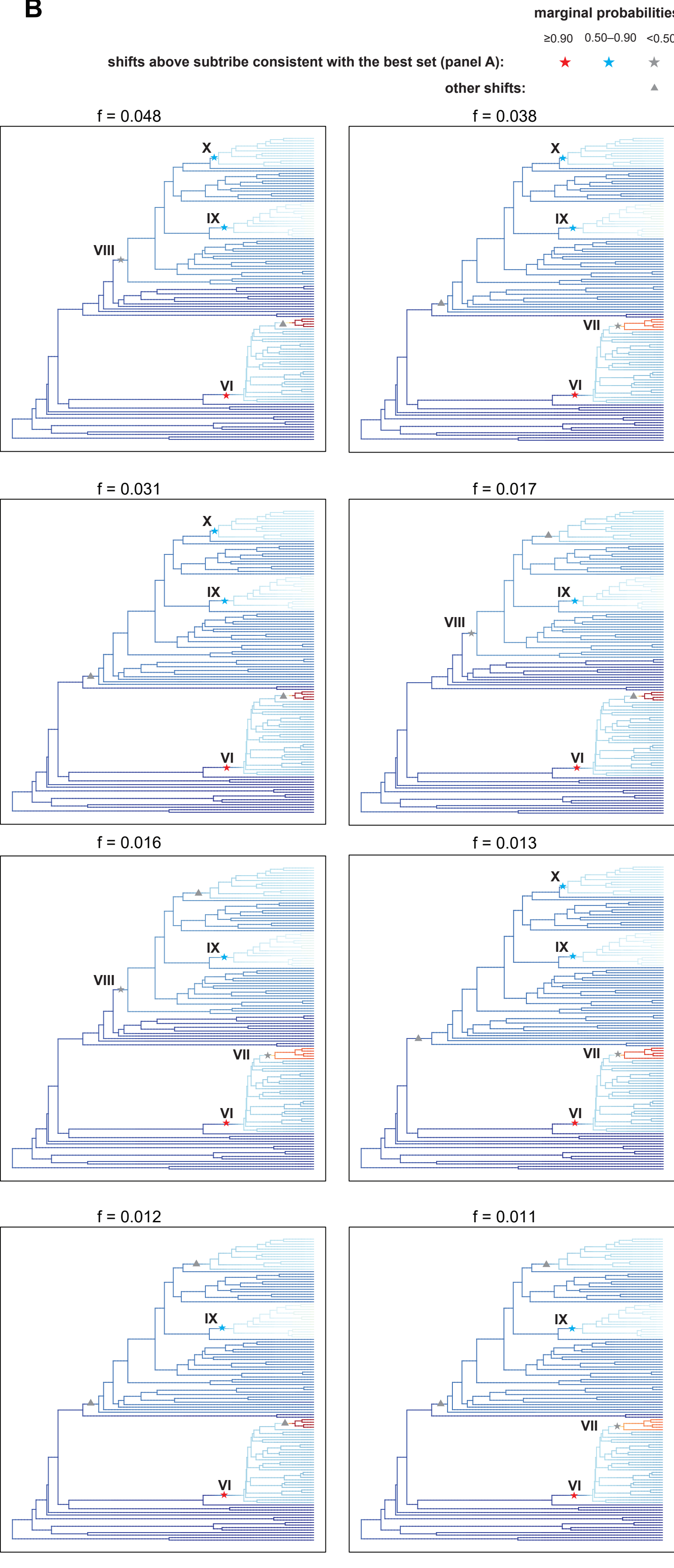
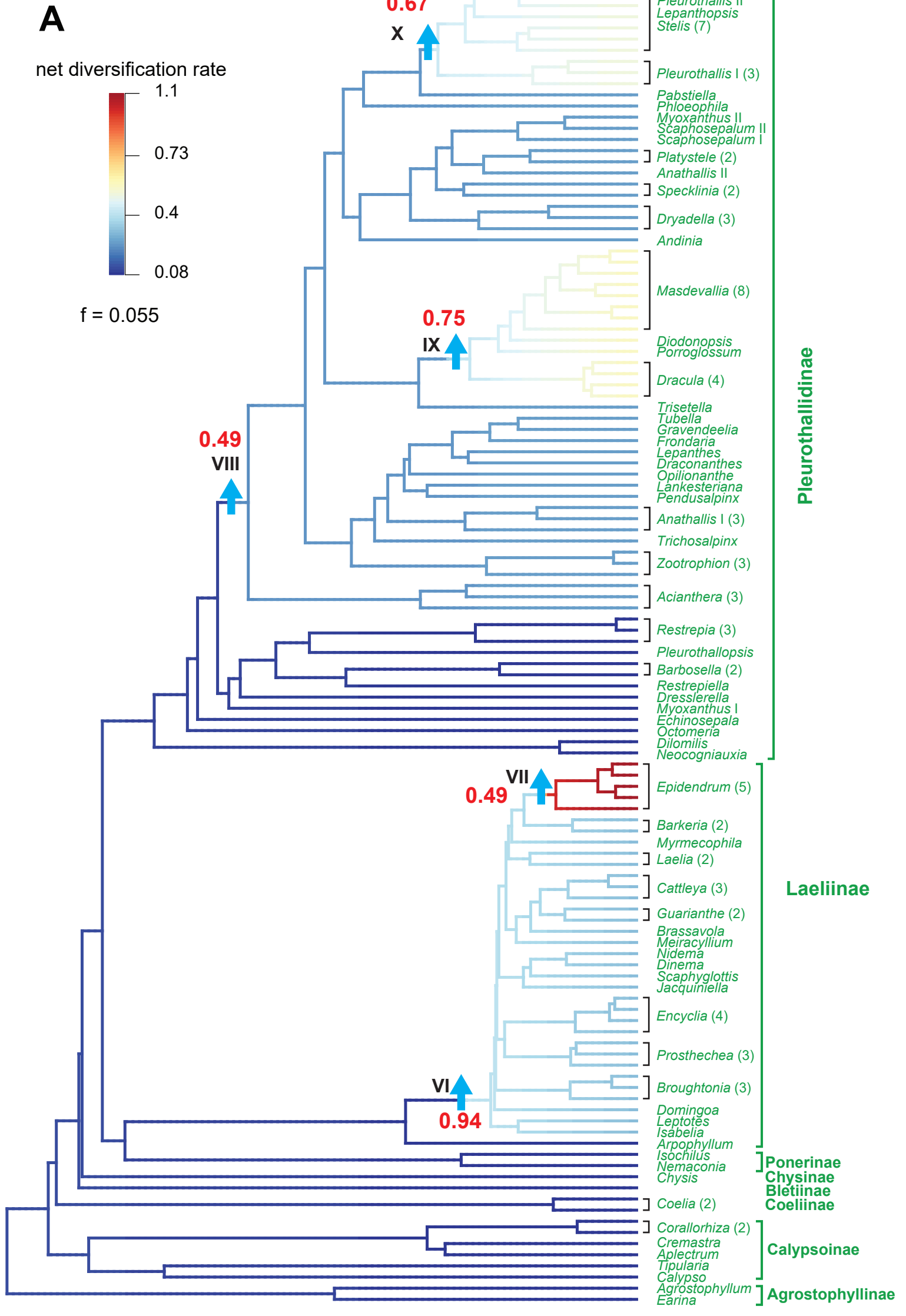


Figure S16. The diversification patterns of Epidendreae inferred using BAMM with generic sampling fractions. (A) The best diversification shift configuration of Epidendreae showing five diversification rate shifts within subtribes (blue arrows, with Roman numerals). The branch colors indicate the net diversification rates of the corresponding branches as shown in the legend at the upper-left corner. The f value at the upper-left corner indicates the sampled frequency of the best shift configuration in the posterior. The marginal shift probability of each shift is indicated using numbers in red near the shifts. The names of subtribes and genera are shown to the right of the tree. If more than one species were sampled from a genus, the number of sampled species is shown in parentheses after the genus name. (B) The first eight diversification shift configurations of Epidendreae after the best one (shown in panel A) from the 95% credible set. The branch colors indicate the net diversification rates of the corresponding branches (warm = high; cool = low). The detected shifts are classified into two types indicated using different symbols, and the marginal shift probabilities are indicated using three colors, as shown in the legend at the top. The five shifts detected in the best configuration (panel A) are indicated using the corresponding Roman numerals as those in panel A. The f value at the top of each plot indicates the sampled frequency of the corresponding shift configuration in the posterior. (C) The net diversification (left), speciation (center), and extinction (right) rate-through-time plots of Epidendreae.

Figure S17

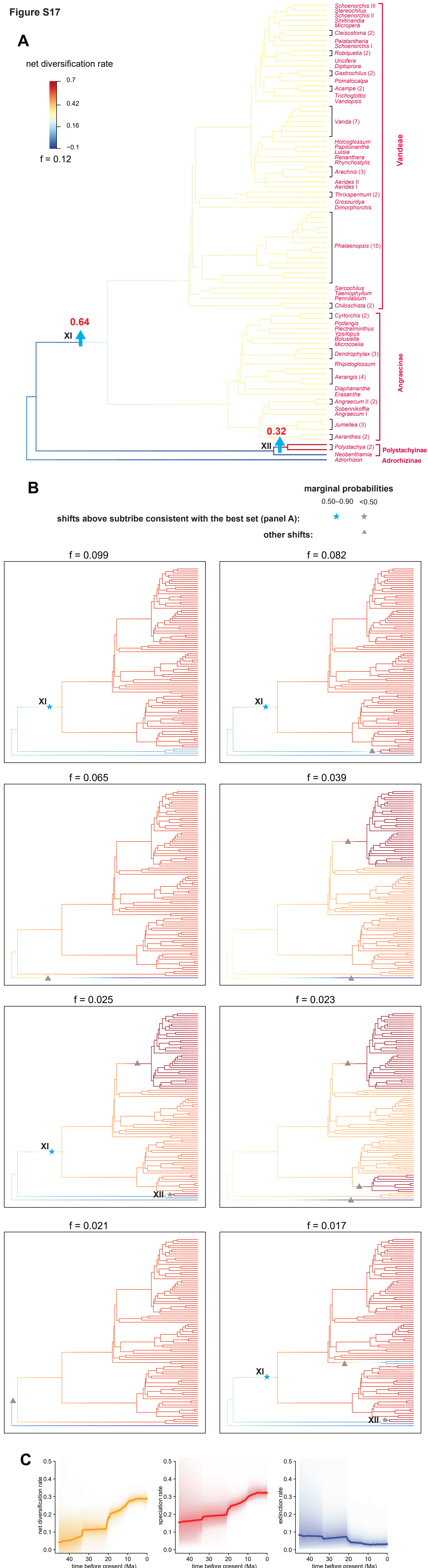
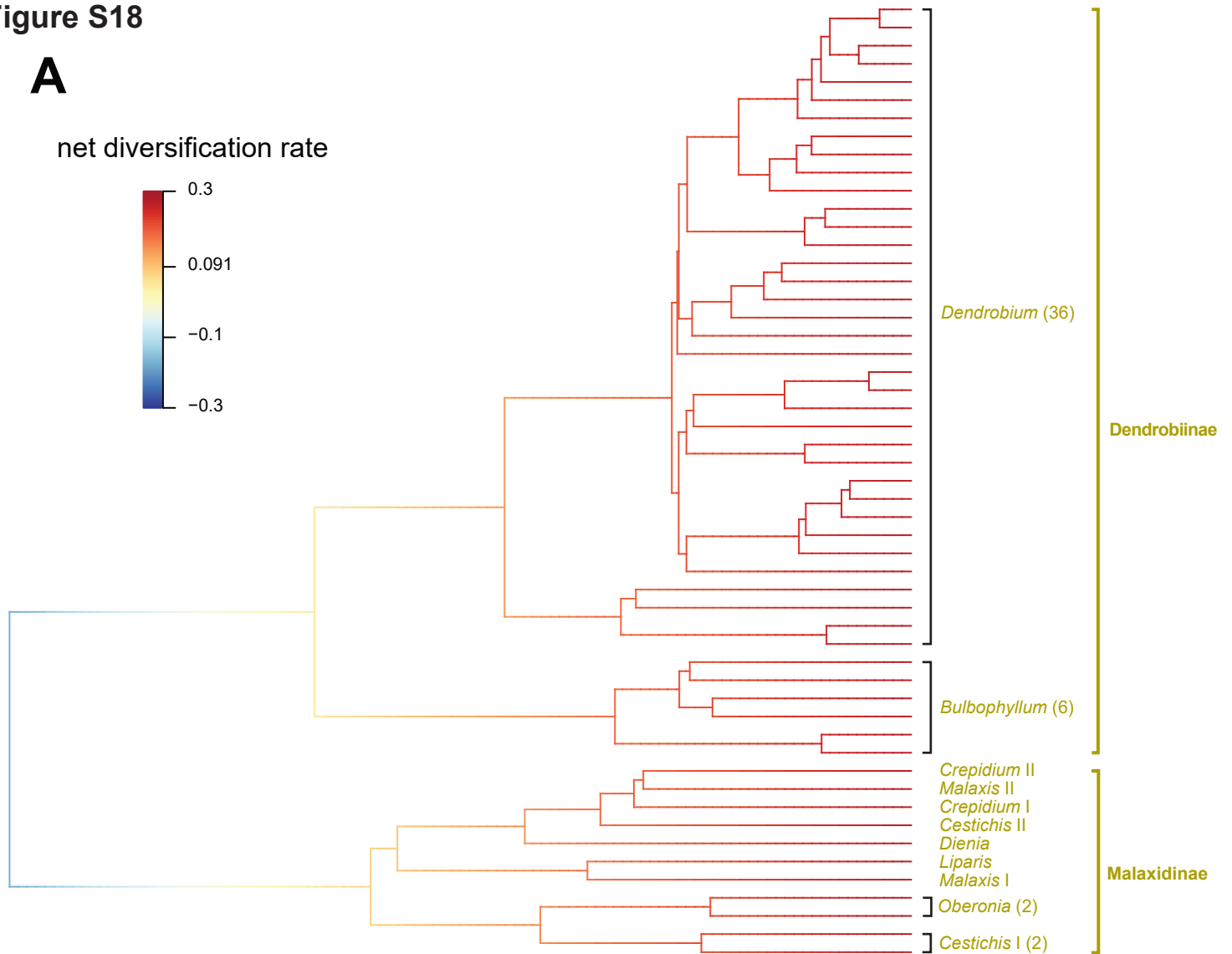


Figure S17. The diversification patterns of Vandeeae inferred using BAMB with generic sampling fractions.

(A) The best diversification shift configuration of Vandeeae showing two diversification rate shifts within subtribes (blue arrows, with Roman numerals). The branch colors indicate the net diversification rates of the corresponding branches as shown in the legend at the upper-left corner. The f value at the upper-left corner indicates the sampled frequency of the best shift configuration in the posterior. The marginal shift probability of each shift is indicated using numbers in red near the shifts. The names of subtribes and genera are shown to the right of the tree. If more than one species were sampled from a genus, the number of sampled species is shown in parentheses after the genus name. The upshift shared by *Polystachya* with low marginal shift probability is not shown in Figure 6 and not accounted as a reliable upshift in the discussion. (B) The first eight diversification shift configurations of Vandeeae after the best one (shown in panel A) from the 95% credible set. The branch colors indicate the net diversification rates of the corresponding branches (warm = high; cool = low). The detected shifts are classified into three types indicated using different symbols, and the marginal shift probabilities are indicated using three colors, as shown in the legend at the top. The two shifts detected in the best configuration (panel A) are indicated using the corresponding Roman numerals as those in panel A. The f value at the top of each plot indicates the sampled frequency of the corresponding shift configuration in the posterior. (C) The net diversification (left), speciation (center), and extinction (right) rate-through-time plots of Vandeeae.

Figure S18

A



B

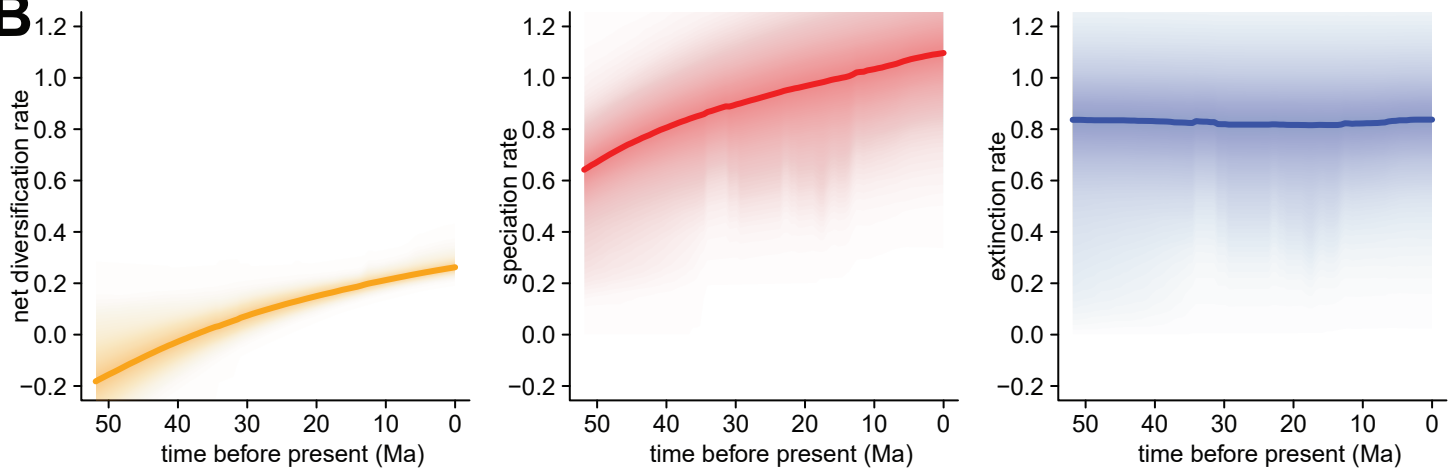


Figure S18. The diversification patterns of Malaxideae inferred using BAMM with generic sampling fractions.

(A) The best diversification shift configuration of Malaxideae showing that no diversification rate shift was detected within subtribes of this tribe. The branch colors indicate the net diversification rates as shown in the legend at the upper-left corner. The names of subtribes and genera are shown to the right of the tree. If more than one species were sampled from a genus, the number of sampled species is shown in parentheses after the genus name. (B) The net diversification (left), speciation (center), and extinction (right) rate-through-time plots of Malaxideae.