







# Asterid Phylogenomics/Phylotranscriptomics Uncover Morphological Evolutionary Histories and Support Phylogenetic Placement for Numerous Whole-Genome Duplications

Caifei Zhang <sup>†,1</sup> Taikui Zhang <sup>†,1</sup> Federico Luebert <sup>†,2,3</sup> Yezi Xiang,<sup>1</sup> Chien-Hsun Huang,<sup>1</sup> Yi Hu,<sup>4</sup> Mathew Rees,<sup>†,5</sup> Michael W. Frohlich,<sup>5</sup> Ji Qi <sup>\*,1</sup> Maximilian Weigend <sup>\*,2</sup> and Hong Ma <sup>\*,4</sup>

<sup>1</sup>Ministry of Education Key Laboratory of Biodiversity Science and Ecological Engineering, School of Life Sciences, Fudan University, Shanghai, China

<sup>2</sup>Nees Institute for Biodiversity of Plants, University of Bonn, Bonn, Germany

<sup>3</sup>Department of Silviculture and Nature Conservation, University of Chile, Santiago, Chile

<sup>4</sup>Department of Biology, The Eberly College of Science, and The Huck Institutes of the Life Sciences, The Pennsylvania State University, University Park, PA

<sup>5</sup>Royal Botanic Gardens, Kew, Richmond, Surrey, United Kingdom

<sup>†</sup>These authors contributed equally to this work.

<sup>\*</sup>Present address: Royal Horticultural Society, Wisley, Woking, Surrey, United Kingdom

**\*Corresponding authors:** E-mails: qij@fudan.edu.cn; mweigend@uni-bonn.de; hxm16@psu.edu.

**Associate editor:** Naruya Saitou

## Abstract

Asterids are one of the most successful angiosperm lineages, exhibiting extensive morphological diversity and including a number of important crops. Despite their biological prominence and value to humans, the deep asterid phylogeny has not been fully resolved, and the evolutionary landscape underlying their radiation remains unknown. To resolve the asterid phylogeny, we sequenced 213 transcriptomes/genomes and combined them with other data sets, representing all accepted orders and nearly all families of asterids. We show fully supported monophyly of asterids, Berberidopsidales as sister to asterids, monophyly of all orders except Icaciniales, Aquifoliales, and Bruniales, and monophyly of all families except Icacinaceae and Ehretiaceae. Novel taxon placements benefited from the expanded sampling with living collections from botanical gardens, resolving hitherto uncertain relationships. The remaining ambiguous placements here are likely due to limited sampling and could be addressed in the future with relevant additional taxa. Using our well-resolved phylogeny as reference, divergence time estimates support an Aptian (Early Cretaceous) origin of asterids and the origin of all orders before the Cretaceous–Paleogene boundary. Ancestral state reconstruction at the family level suggests that the asterid ancestor was a woody terrestrial plant with simple leaves, bisexual, and actinomorphic flowers with free petals and free anthers, a superior ovary with a style, and drupaceous fruits. Whole-genome duplication (WGD) analyses provide strong evidence for 33 WGDs in asterids and one in Berberidopsidales, including four suprafamilial and seven familial/subfamilial WGDs. Our results advance the understanding of asterid phylogeny and provide numerous novel evolutionary insights into their diversification and morphological evolution.

**Key words:** asterids, phylogeny, adaptive radiation, morphological diversity, WGD, polyploidy.

## Introduction

The asterids (*Asteridae*), with ~100,000 species, include nearly one quarter of the extant angiosperm species and are thus the largest subgroup in eudicots. The bulk of the species belongs to the former subclass Asteridae and includes many economically important crops (Magallón and Castillo 2009). Asterids are subdivided into 110 families in 17 orders, including the orders Apiales (e.g., carrot, celery, and ginseng), Asterales (sunflower, lettuce, and artichoke), Cornales (dogwood), Ericales (tea, kiwifruit, and blueberry), Gentianales (coffee),

Lamiales (mint, basil, sesame, and olive), and Solanales (tomato, potato, pepper, tobacco, and sweet potato). Several asterid orders, for example, Asterales, Lamiales, Gentianales, Solanales, Apiales, and Boraginales, show some of the highest diversification rates in angiosperms (Magallón and Sanderson 2001; Soltis et al. 2019). Asterids are also highly diversified in ecology and morphology, from annual herbs to large trees, and from regular, autotrophic to parasitic and carnivorous plants, with habitats ranging from terrestrial to aquatic, from tropical rainforests to hyperarid deserts, and from warm lowlands to cold high-elevation environments.

© The Author(s) 2020. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact [journals.permissions@oup.com](mailto:journals.permissions@oup.com)

Open Access

This great diversity might be explained in part by morphological and biochemical traits that appear to unify this group and possibly confer evolutionary advantages, including potential synapomorphies such as iridoid compounds, unitegmic ovules, and cellular endosperm (Stull et al. 2018). Asterids have been characterized by a fused corolla tube (sympetaly), which in turn is fused with filaments (stapet), stamens in a single whorl (haplostemony), ovules with a simple integument, and the presence of endothelium and endosperm haustoria (Endress 2011a, 2011b), although some of these traits are not unique to asterids and/or not universally present across the group. In addition, asterid reproductive structures are highly diverse and some characters might have evolved multiple times independently, such as sympetaly (Stull et al. 2018), zygomorphic flowers (in 41 families across several orders) (Reyes et al. 2016), parasitism (Barkman et al. 2007; Schneider et al. 2018), and carnivory (Ellison and Gotelli 2009).

The diversity and economic and ecological importance of the asterids have made this group a focus of molecular phylogenetics (Olmstead et al. 1992; 1993). The monophyly of the asterids has been supported using plastid sequences (e.g., Olmstead et al. 1992; Moore et al. 2010; Li, Yi, et al. 2019) and also nuclear data from small numbers of taxa (e.g., Zhang et al. 2012; Zeng et al. 2017). A few studies have recently addressed phylogenetic relationships across asterids with good representation of orders and families (Albach et al. 2001; Bremer et al. 2002; Wikström et al. 2015; Stull et al. 2018, 2020). Among the well-resolved relationships of the asterid orders are 1) monophyly of core asterids consisting of all orders besides Cornales and Ericales, 2) separation of core asterid orders into two large clades lamiids and campanulids, 3) Gentianales and Boraginales as sister clades within lamiids, and 4) Dipsacales and Paracryphiales as sister clades within campanulids (APG IV 2016; Leebens-Mack et al. 2019; Li, Yi, et al. 2019). In addition, more specialized studies have focused on lamiids (Refulio-Rodriguez and Olmstead 2014; Stull et al. 2015) and campanulids (Winkworth et al. 2008; Tank and Donoghue 2010), sampling eight and seven of the 17 orders, respectively.

However, some relationships within lamiids and campanulids remain unresolved or contradictory, such as the branching sequence of early diverging orders in lamiids (Icacinales, Garryales, and Metteniusales) (e.g., Refulio-Rodriguez and Olmstead 2014; Stull et al. 2015; Li, Yi, et al. 2019). The relationships among the other five lamiids orders in the well-defined clade (Lamianae: Boraginales, Gentianales, Lamiales, Solanales, and Vahliales) are also inconsistent. In the campanulids, Aquifoliales is consistently retrieved as sister to six other orders (Apiales, Asterales, Bruniales, Dipsacales, Escalloniales, and Paracryphiales) and Apiales are sister to a clade of Dipsacales + Paracryphiales, but relationships among the remaining nodes are unresolved or only receive low to moderate supports. Broadly sampled nuclear phylogenies could potentially resolve these issues.

Recent advances in transcriptome sequencing are revolutionizing the understanding of plant molecular systematics by identifying single-copy nuclear genes to resolve deep-level phylogeny (Zhang et al. 2012; Wickett et al. 2014; Zeng

et al. 2017). Nuclear genes are inherited biparentally and show higher substitution rates than plastid genes (Birky 2001; Springer et al. 2001; Zhang et al. 2012; Davis et al. 2014; Lu et al. 2018). Single-copy nuclear genes thus provide an alternative line of evidence for resolving deep relationships and may resolve incongruences between phylogenies (Zeng et al. 2014, 2017; Zhao et al. 2016; Li et al. 2017). They have been successfully used to resolve relationships at the family (Asteraceae, Brassicaceae, and Rosaceae; Huang, Sun, et al. 2016; Huang, Zhang, et al. 2016; Xiang et al. 2017), order (Caryophyllales; Yang et al. 2015), and higher levels (rosids, eudicots, angiosperm, seed plants, ferns, and land plants; Wickett et al. 2014; Zeng et al. 2014, 2017; Zhao et al. 2016; Li et al. 2017; Jin et al. 2018). Recent nuclear phylogenomics of eudicots and across green plants (the 1KP study) placed a clade of Ericales and Cornales as sister to the core asterids (Zeng et al. 2017; Leebens-Mack et al. 2019).

Widely accepted whole-genome duplications (WGDs) include ones shared by all extant angiosperms and by core eudicots ( $\gamma$ ), and ones in the early histories of several families, such as Brassicaceae, Fabaceae, Poaceae, Asteraceae, and Rosaceae (Tang et al. 2010; Edger et al. 2015; Huang, Zhang, et al. 2016; Xiang et al. 2017; Qiao et al. 2019). WGDs contribute to genome structure variation and organismal complexity and have been hypothesized to be a major mechanism supporting key functional innovations and organismal diversity (Schranz et al. 2012; Soltis PS and Soltis DE 2016; Van de Peer et al. 2017). Several WGDs have been reported for some asterid lineages, including those in Solanaceae, Asteraceae, in kiwifruit and carrot, all following the gamma event shared by core eudicots (Barker et al. 2008; Potato Genome Sequencing Consortium 2011; Tomato Genome Consortium 2012; Huang et al. 2013; Kim et al. 2014; Hoshino et al. 2016; Huang, Zhang, et al. 2016; Iorizzo et al. 2016; Badouin et al. 2017; Reyes-Chin-Wo et al. 2017; Xu et al. 2017; Ren et al. 2018; Leebens-Mack et al. 2019; Qiao et al. 2019). Nevertheless, detection of additional WGDs in asterids and placement of previously reported events in a phylogenetic context would enable us to investigate, on a broader basis, possible links between WGDs and morphological evolution associated with high species richness.

We generated 213 new transcriptome and/or genome data sets from 210 species and combined these with other public and collaborative data sets, for a total of 365 asterid species, representing all established orders and nearly all families of asterids to resolve the deep phylogenetic relationships of asterids and to address some related evolutionary questions. The dramatically expanded sampling in asterids here compared with previous studies was facilitated in large part by extensive living collections in botanical gardens, including a wide range of taxa from across the globe (with 99 samples from the Bonn University Botanic Gardens alone, including taxa in Columelliaceae, Paracryphiaceae, Sphenocleaceae, and Loasaceae). Five sets of low-copy nuclear genes were identified from these transcriptomic/genomic data sets and used for phylogenetic analyses using a coalescent approach; in addition, the smallest gene set was also used for phylogenetic analyses as a concatenated supermatrix. Our analyses support

a highly resolved asterid phylogeny, which provides a framework for additional analyses to estimate divergence times, to reconstruct ancestral morphological characters, and to detect evidence for WGDs.

## Results and Discussion

### Transcriptome and Genome Sequencing of Representative Taxa and Identification of Low-Copy Nuclear Genes

To resolve asterid relationships at the order and family levels, a total of 365 species of the asterids representing all 17 orders and 102 of 110 families were sampled (supplementary table S1, Supplementary Material online; see also supplementary note, Supplementary Material online), with 39 species from other orders for relationships of asterids with other major clades of eudicots (supplementary table S1, Supplementary Material online). Among these, 208 transcriptomic and 5 genomic data sets were newly generated, with a total of 1,482.0 Gb raw data (supplementary table S2, Supplementary Material online). In addition, we also included transcriptomic data sets for 46 species we generated previously (Zeng et al. 2014, 2017; Huang, Zhang, et al. 2016; Ren et al. 2018) and other data sets (from Phytozome v12, 1KP [Leebens-Mack et al. 2019], NCBI and other databases) (supplementary table S1, Supplementary Material online). From trimmed reads, an average of 58,815 nonredundant unigenes per species were assembled with an average N50 length (1,221.9 bp; supplementary table S2, Supplementary Material online), similar to that (1,128.9 bp) in recently published de novo assembled transcriptomes for phylogenomics (Xiang et al. 2017).

Previously, <200 low-copy genes were found to contain sufficient phylogenetic signal to resolve the relationships among members of a family or order, and even for deeper angiosperm lineages (Zeng et al. 2014, 2017; Huang, Sun, et al. 2016; Huang, Zhang, et al. 2016; Xiang et al. 2017; Leebens-Mack et al. 2019). We selected low-copy putative orthologs with consideration for sufficient taxon coverage and sequence lengths, as well as aiming at elimination of hidden paralogs due to gene duplication (GD) followed by loss (see Materials and Methods and supplementary table S3, Supplementary Material online) for phylogenetic analyses. An initial set of 1,769 genes was iteratively filtered for those that were recovered from at least 80% of the taxa and had, respectively, overall gene alignment length of 600, 800, and 1,000 bp, successively generating data sets of 1,041, 784, and 565 genes. Finally, a 387-gene set was identified from the 565-gene set to reduce phylogenetic noise and to provide a relatively small gene set for phylogenetic analysis using the concatenation approach and molecular clock analysis.

### Well-Resolved Relationships among Major Lineages of Asterids

The five data sets containing, respectively, 1,769, 1,041, 784, 565, and 387 genes were used for coalescent analyses of asterids phylogeny (supplementary figs. S1–S5, Supplementary Material online) (a family-level summary phylogeny in shown

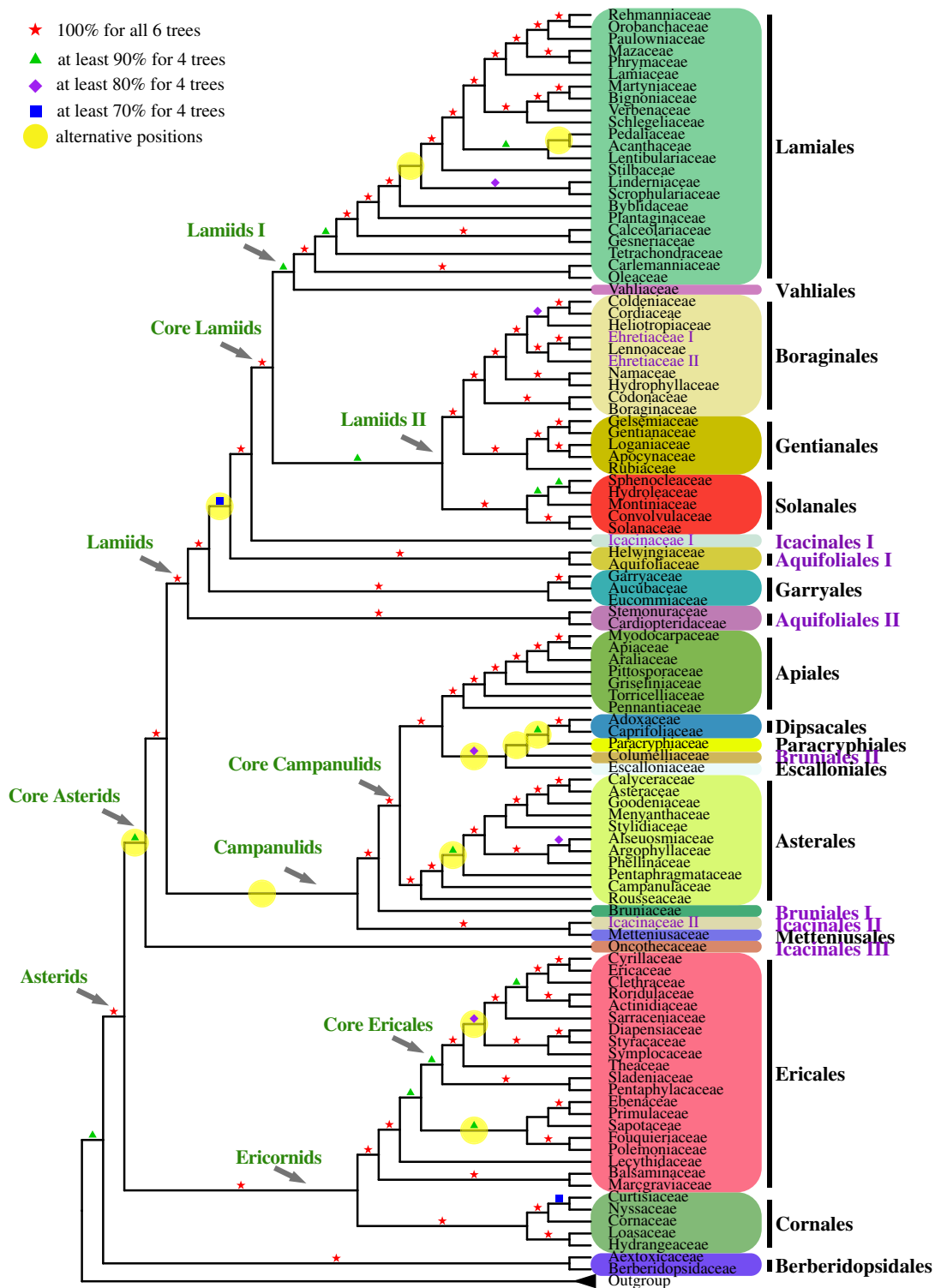
in fig. 1), with additional maximum-likelihood (ML) inference using a concatenated data set of 387 genes (supplementary figs. S6 and S7, Supplementary Material online). In all analyses, the monophyly of asterids is maximally supported and Berberidopsidales is highly supported as the sister to asterids (fig. 1; see supplementary note, Supplementary Material online, for information about relationships of asterids with other eudicot lineages).

Among 17 asterid orders, one is monotypic and 13 are monophyletic, whereas representatives of the orders Icacinales, Aquifoliales, and Bruniales were found in two or three lineages. All families are found to be monophyletic except Icacinae and Ehretiaceae, and all 70 genera with two or three species sampled are monophyletic (figs. 1–5 and supplementary figs. S1–S8, Supplementary Material online). The plastid gene-based phylogeny found monophyly of Icacinales (APG IV 2016), but we retrieved members of two Icacinales families in three nonsister clades Icacinales I, Icacinales II, and Icacinales III (fig. 1). The phylogenetic relationships among asterid families are shown in figures 2–5, with detailed descriptions of the relationships among orders, families, and even some genera within larger families are provided in the supplementary notes, Supplementary Material online. Below, we mainly focused on the comparison between the asterids ordinal and familial phylogeny here versus what has been shown in previous phylogenetic studies (supplementary fig. S9, Supplementary Material online) (e.g., the recent asterids phylogenies using 410 nuclear genes from 217 species, 162 genera, 70 families, and 13 orders [Leebens-Mack et al. 2019], and using 80 plastid genes from 719 species, 375 genera, 85 families, and 17 orders [Li, Yi, et al. 2019]).

### Ordinal Relationships in Asterids

Our analyses reconstruct a robust (100% bootstrap [BS] for all trees) relationship for basal asterids, with Cornales and Ericales forming a maximally supported clade (referred to here as Ericornids) as sister to core asterids (fig. 1 and supplementary fig. S7, Supplementary Material online). This is congruent with the topologies found in previous studies using nuclear genes and a smaller taxon sampling (e.g., Morton 2011; Zhang et al. 2012; Maia et al. 2014; Zeng et al. 2014, 2017; Leebens-Mack et al. 2019). Conversely, plastome data retrieved Cornales and Ericales as successive sisters to all remaining asterids with high support (e.g., Moore et al. 2010; Gitzendanner et al. 2018; Li, Yi, et al. 2019). The topology of the nuclear phylogenies seems to be supported by morphological congruence between Ericales and Cornales (e.g., Huber 1963; Dahlgren 1975, 1983). We suggest that the distinct plastid topology might be due to the fixation of an older plastid genome sequence in the most recent common ancestors of Cornales.

Core asterids (=Gentianidae; fig. 1) represent all taxa other than Cornales and Ericales and form a highly supported clade in both plastid and nuclear phylogenetic studies, including the present one. Plastid phylogenies find two major subclades, lamiids and campanulids, and these are also retrieved here using nuclear sequences (except for the family



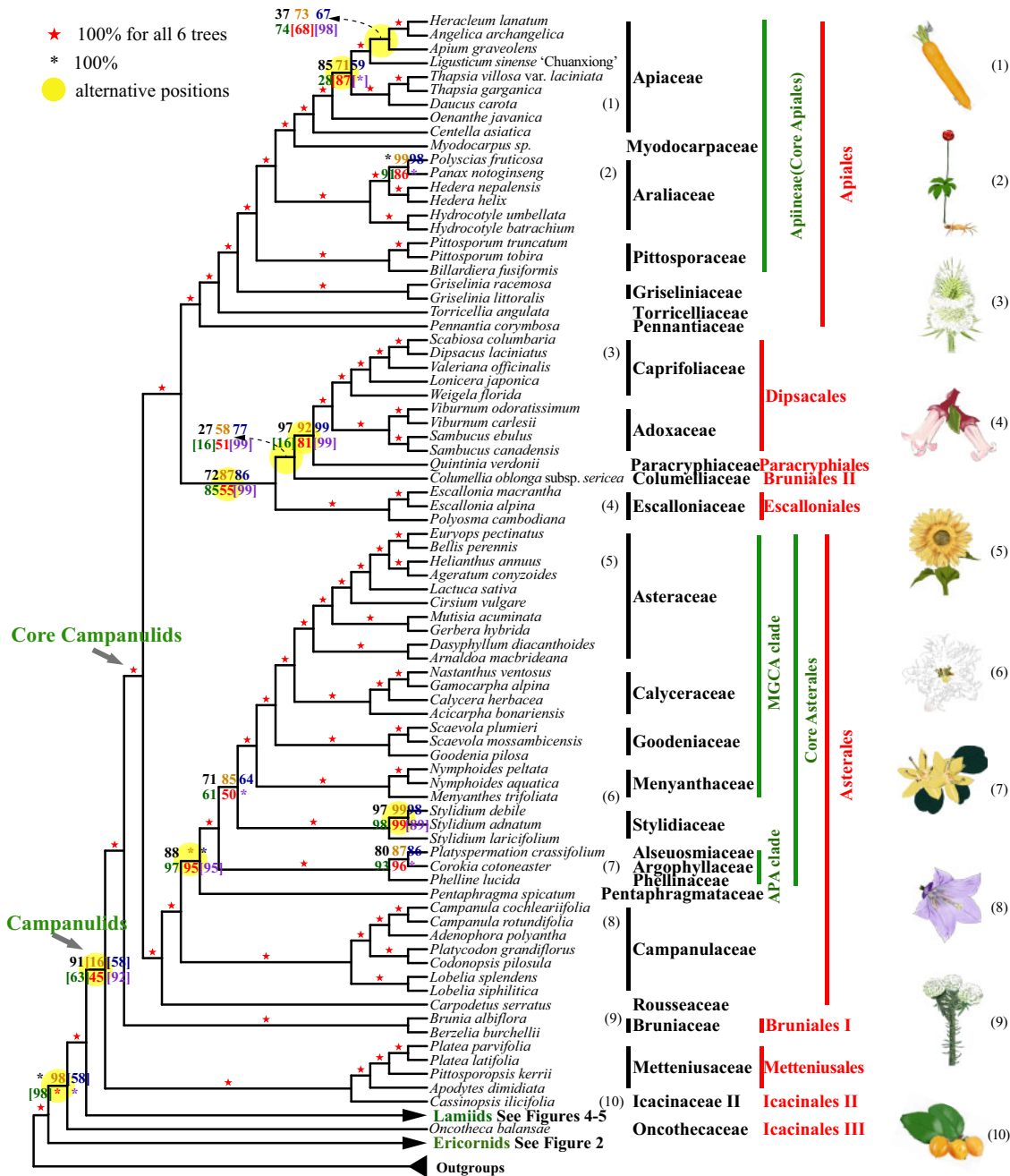
Downloaded from https://academic.oup.com/mbe/article/37/1/3/1881587/0293 by guest on 21 August 2022

**Fig. 1.** A summary tree showing phylogenetic relationships among asterid families. Red pentagrams represent support of 100% in all six trees. For other branches, green triangles, purple rhombuses, and blue squares represent support of  $\geq 90\%$ ,  $\geq 80\%$ , and  $\geq 70\%$  in at least four trees, respectively. Yellow circles indicate nodes with alternative topologies. Family names are given at the terminals, with order names to the far right. Names representing paraphyletic or polyphyletic groups are given in purple. (See figures 2–5 for detailed phylogenetic relationships among terminal taxa in various groups.)

Oncothecaceae, Icaciniales III, fig. 1), but the detailed relationships among orders and families found here are markedly different from previous hypotheses (see below). The new

placement of Oncothecaceae here (see detailed discussion in [supplementary note, Supplementary Material](#) online) supports the recognition of the order Oncothecales. This finding

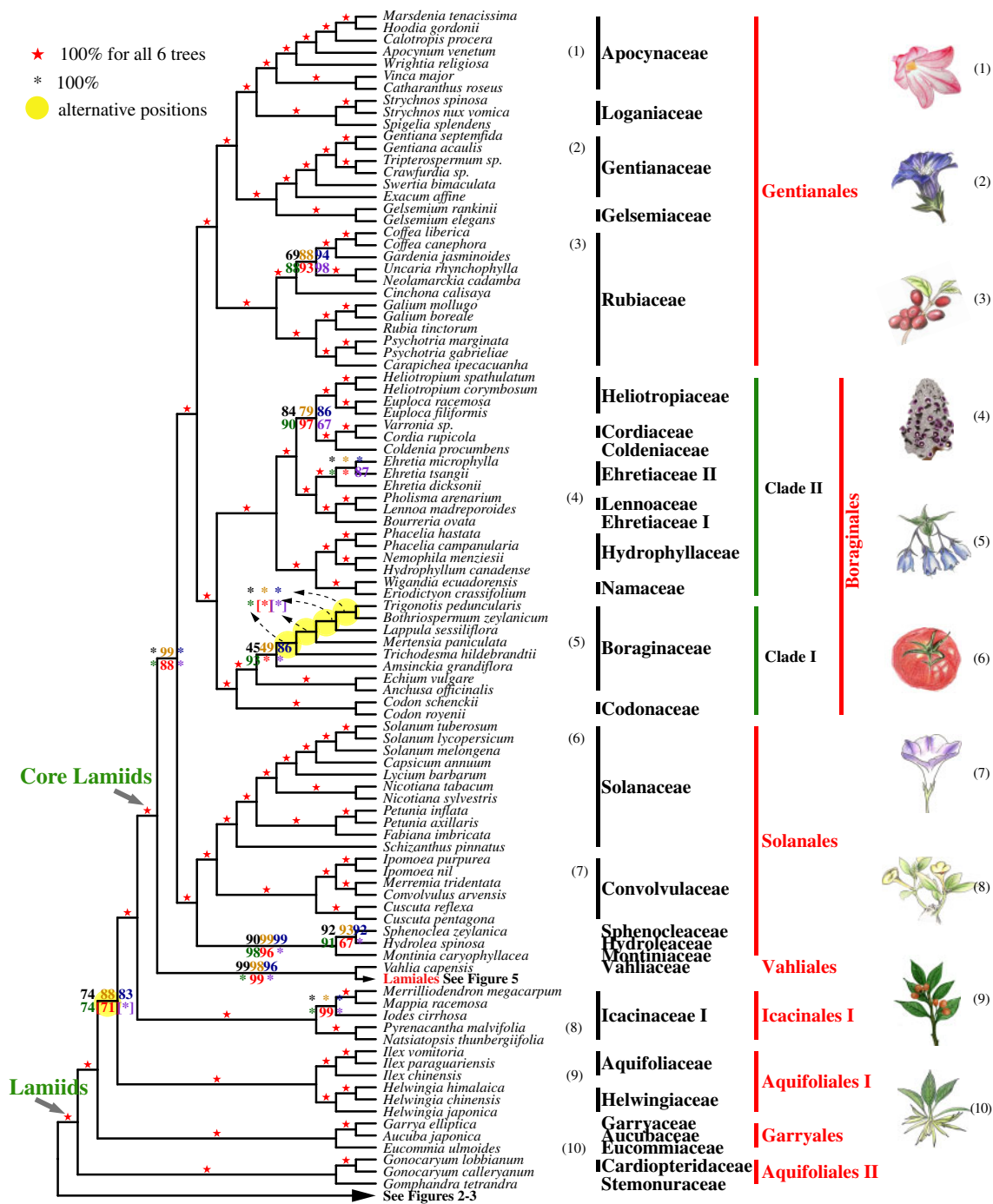




**FIG. 3.** A summary of the coalescent and concatenation trees of campanulids. Black, orange, blue, green, and red digits on the internode represent the support values from five coalescence analyses of 1,769-, 1,041-, 784-, 565-, and 387-gene sets, respectively, and purple numbers represent that from a concatenation analysis using the 387-gene supermatrix. A number within square brackets indicates support level for an alternative position in the corresponding tree. Numbers next to the species names correspond to the numbered plant drawings at the far right.

(Bruniales II), which are placed separately in the campanulid phylogeny (fig. 1). Bruniaceae (Bruniales I) are retrieved with maximum support (in all topologies) as sister to core campanulids, which are divided into Asterales (100% BS support in all trees) and a maximally supported clade including the remaining orders (Escalloniales, Bruniales II, Paracryphiales, Dipsacales, and Apiales; figs. 1 and 3). Columelliaceae (Bruniales II) is embedded in a clade with uncertain relationships among the orders Dipsacales, Paracryphiales, and Escalloniales (fig. 1); this clade has relatively high support values and is sister to Apiales. Similarly, the recent 1KP

phylogenomic study also retrieved Asterales and Apiales as successive sisters to Dipsacales + Escalloniales although Paracryphiales were not sampled (supplementary fig. S9, Supplementary Material online) (Leebens-Mack et al. 2019). In an earlier study, Dipsacales and Paracryphiales were sisters, but Escalloniales were sister to Asterales (Soltis et al. 2011). Lamiids are another very diverse asterid group of nine orders (fig. 1). The early branching lineages of lamiids are represented here by clades (orders) Aquifoliales I (here with Aquifoliaceae and Helwingiaceae), Garryales (Garryaceae, Aucubaceae, and Eucommiaceae), Aquifoliales II

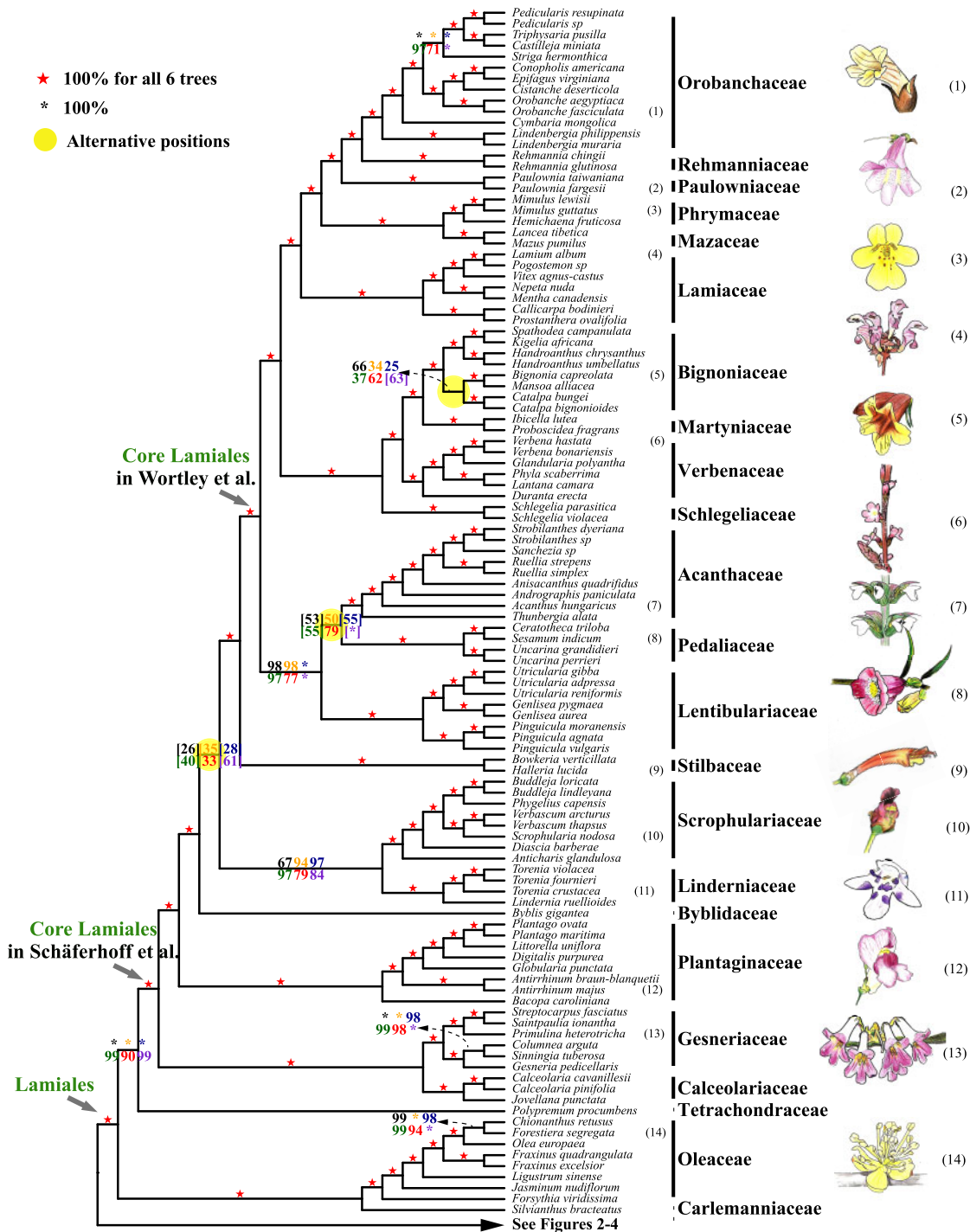


Downloaded from https://academic.oup.com/mbe/article/37/1/1/181881587/0293 by guest on 21 August 2022

**FIG. 4.** A summary of the coalescent and concatenation trees of lamiids. Black, orange, blue, green, and red digits on the internode represent the support values from five coalescence analyses of 1,769-, 1,041-, 784-, 565-, and 387-gene sets, respectively, and purple numbers represent that from a concatenation analysis using the 387-gene supermatrix. A number within square brackets indicates support level for an alternative position in the corresponding tree. Numbers next to the species names correspond to the numbered plant drawings at the far right.

(Cardiopteridaceae and Stemonuraceae), and Icacinales I (Icacinaceae I). Published plastid gene phylogenies supported the monophyly of Aquifoliales as the first diverging clade of campanulids (e.g., Bremer et al. 2002; Soltis et al. 2011; Li, Yi, et al. 2019), and Aquifoliales I and Aquifoliales II each as strongly supported clades in a sister relationship (e.g., Li, Yi, et al. 2019). However, earlier studies using concatenated data sets of multiple nuclear genes supported the sister

relationship of Aquifoliales I and Garryales (74–98% BS, Zhang et al. 2012; Zeng et al. 2017; Leebens-Mack et al. 2019; Stull et al. 2020), consistent with our results based on a 387-gene sets with coalescent and supermatrix methods (71% and 100% BS, respectively, see supplementary figs. 5 and 6, Supplementary Material online). Our results differ from previous ones in placing Aquifoliales I and Aquifoliales II in nonsister lineages (figs. 1 and 4; see supplementary note



**FIG. 5.** Summary of the coalescent and concatenation trees of Lamiales. Black, orange, blue, green, and red digits on the internode represent the support values from five coalescence analyses of 1,769-, 1,041-, 784-, 565-, and 387-gene sets, respectively, and purple numbers represent that from a concatenation analysis using the 387-gene supermatrix. Numbers next to the species names correspond to the numbered plant drawings at the far right.

and [fig. S9, Supplementary Material](#) online, for more information).

Core lamiids fall into two sister clades, lamiids I (Lamiales + Vahliales) and lamiids II ([Boraginales + Gentianales] + Solanales) ([figs. 1 and 4](#)), both receiving 98% or higher BS values in five of the six trees ([supplementary fig. S8, Supplementary Material](#) online); in addition, the monophyly of each of the four orders with multiple species and the

sisterhood of Gentianales and Boraginales received maximal supports in all six trees. The 1KP phylogeny included members of four orders Lamiales, Boraginales, Gentianales, and Solanales and their relationships are consistent with this study ([supplementary fig. S9, Supplementary Material](#) online) ([Leebens-Mack et al. 2019](#)). Previous nuclear phylogenies found low support values for Solanales sister to Lamiales or Boraginales, based on a small sampling ([Zhang et al. 2012](#);



Zeng et al. 2017). Some recent studies using plastid genes or plastomes found a sister relationship Gentianales + Boraginales (50–87% BS), whereas Solanales was found as sister to either Vahliales or to Lamiales (<80% BS, Stull et al. 2018; Li, Yi, et al. 2019), but Refulio-Rodriguez and Olmstead (2014) found a different topology between the five orders with <80% BS using nine plastid regions and the mitochondrial rps3 region.

### Familial Relationships in Asterids

We focus here on the familial relationships within relatively large and phylogenetically diverse orders, including Ericales, Asterales, Boraginales, and Lamiales; additional discussions for all families are provided in the [supplementary note, Supplementary Material](#) online.

Ericales are a large order in basal asterids (the Ericornids), with 22 families, 346 genera, and 11,545 species (APG IV 2016) and their relationships have remained uncertain probably due to their ancient and rapid radiation (Schönenberger 2009; Soltis et al. 2011; Rose et al. 2018). Our robust topologies with sampling for 20 families provide at least 90% BS support at most nodes along the Ericales backbone in four or more trees (fig. 2). This is largely consistent with previous publications ([supplementary fig. S9, Supplementary Material](#) online) (Soltis et al. 2011; Leebens-Mack et al. 2019), but we are able to provide well-supported placements for previously problematic families such as Theaceae and Pentaphylacaceae (cf., Rose et al. 2018). Previous studies have recognized several suprafamilial clades (for a recent reference, see Rose et al., 2018), which are indicated in green on the right in figure 2 (see [supplementary note, Supplementary Material](#) online, for details on relationships of suprafamilial clades). The sampling here is expanded compared with the 1KP study (Leebens-Mack et al. 2019) and includes six additional families (in order of divergence from other Ericales): Lecythidaceae, Sladeniaceae, Theaceae, Sarraceniaceae, Clethraceae, and Cyrillaceae (fig. 2 and [supplementary fig. S9, Supplementary Material](#) online). Lecythidaceae is maximally supported here in all six trees as the second lineage to diverge from other Ericales families, after the separation of a maximally supported clade of Balsaminaceae + Marcgraviaceae. Sladeniaceae + Pentaphylacaceae form a maximally supported clade that separates from others after the divergence of a highly supported clade of five families (from Fouquieriaceae to Primulaceae, [supplementary fig. S9, Supplementary Material](#) online). Theaceae (including the tea plant, the most commercially important species in Ericales) occupies the next lineage to diverge, followed by a clade of three families with the same topology as in previous studies based on plastid (Li, Yi, et al. 2019) or nuclear genes (Leebens-Mack et al. 2019). Plastid analysis placed Theaceae as sister (67% BS) to the three-family clade (Li, Yi, et al. 2019), which is not confirmed here. Sarraceniaceae is maximally supported as sister to a clade of five families, two of which are Clethraceae, and Cyrillaceae, placed here as successive sisters to Ericaceae (fig. 2 and [supplementary fig. S9, Supplementary Material](#) online).

Within core campanulids, Asterales represent a particularly important lineage, with many species of agricultural and/or pharmaceutical importance. They are also extraordinarily species rich (26,870 species), possibly due to several rounds of WGDs (Barker et al. 2008, 2016; APG IV 2016; Huang, Zhang, et al. 2016; Badouin et al. 2017; Ren et al. 2018; Leebens-Mack et al. 2019). We sampled 35 species, representing all 11 Asterales families (1.66% [29/1,743] of the genera) (APG IV 2016), and retrieved robust relationships of the 11 families in each of the well-supported topologies (figs. 1 and 3). The interfamilial phylogeny here is consistent with the 1KP study (Leebens-Mack et al. 2019) for lineages sampled in both studies, with the only exception of Styliidiaceae ([supplementary fig. S9, Supplementary Material](#) online). However, we include three additional families (namely Rouseaceae, Pentaphragmataceae, and Calyceraceae; [supplementary fig. S9, Supplementary Material](#) online). Among three early-divergent families, Rouseaceae was placed as the basalmost lineage (100% BS in all six trees; figs. 1 and 3), in agreement with results from some previous analyses (Soltis et al. 2007; Zanne et al. 2014). Campanulaceae and Pentaphragmataceae are the next two divergent lineages (for further discussions of relationships for basal lineages, see [supplementary note, Supplementary Material](#) online). The remaining eight Asterales families (core Asterales) form a maximally supported clade, in agreement with previous publications (Lundberg and Bremer 2003; Lundberg 2009). Seven of the families in core Asterales comprise two suprafamilial clades, the MGCA clade (100% BS, all analyses) with Menyanthaceae, Goodeniaceae, Calyceraceae, and Asteraceae and the APA clade (100% BS, all analyses) with Argophyllaceae, Phellinaceae, and Alseuosmiaceae (fig. 3). Our results indicate that the APA clade and Styliidiaceae are consecutive sisters to the MGCA clade (100% BS for APA clade, 50–85% for Styliidiaceae in five coalescent trees, and 100% BS for Styliidiaceae in the concatenation tree; fig. 3 and [supplementary fig. S8, Supplementary Material](#) online). This is consistent with some previous studies (Tank and Donoghue 2010; Soltis et al. 2011); however, recent studies using plastome data or nuclear genes highly supported the APA and MGCA clades as sisters (Leebens-Mack et al. 2019; Li, Yi, et al. 2019) and other studies proposed different relationships with low to moderate supports (e.g., Bremer et al. 2002; Stull et al. 2018).

Within core lamiids, we sampled 29 Boraginales species covering 9 of 11 families (no samples for Hoplestigmataceae and Wellstediaceae) and 15.33% of the genera and found all families to be monophyletic except Ehretiaceae (figs. 1 and 4). The sampling here includes four additional families, Codonaceae, Namaceae, Cordiaceae, and Coldeniaceae, compared with the 1KP study. Our analyses indicate that Boraginales contain two suprafamilial clades: clade I consisting of Codonaceae + Boraginaceae with 100% BS in all six trees (Boraginales I, Weigend et al. 2014), and clade II, including the remaining seven Boraginales families with maximal BS support (fig. 4) (Boraginales II, Weigend et al. 2014). Within Boraginales II, Hydrophyllaceae + Namaceae form a clade (100% BS) that is sister to the other five families of Boraginales II. Among the latter five families, a clade with

Lennoaceae nested within Ehretiaceae (see [supplementary note, Supplementary Material](#) online) is sister to Heliotropiaceae + (Coldeniaceae + Cordiaceae). Among the five shared families, the topology is consistent with that of the 1KP study ([supplementary fig. S9, Supplementary Material](#) online) (Leebens-Mack et al. 2019) but slightly different from the phylogenies based on plastid markers from Weigend et al. (2014), Refulio-Rodriguez and Olmstead (2014), Stull et al. (2015), and Luebert et al. (2016), which supported Hydrophyllaceae and Namaceae as successive sisters to the remaining families. Conversely, recently published plastid phylogenetics and morphological character analyses support the hypothesis that Heliotropiaceae are sister to a well-supported clade consisting of two clades, one including Cordiaceae, Hoplestigmataceae, and Coldeniaceae and the other including Ehretiaceae and Lennoaceae, collectively form a super clade with species bearing multilayered endocarp and four or fewer ovules (Refulio-Rodriguez and Olmstead 2014; Weigend et al. 2014; Luebert et al. 2016). The relationships among Boraginales families are also supported by previous fruit morphological and phylogenetic analyses (Weigend et al. 2014; Luebert et al. 2016).

Lamiales are one of the largest orders of asterids and are particularly diverse, with 24 families comprising 23,810 species, many of them of economic or horticultural importance, such as sesame, olive, and mint (Wortley et al. 2005; APG IV 2016). We sampled 110 species and 7.93% (84/1,059) of the genera representing all families except the Plocospermataceae, which has been placed as the first diverging lineage of Lamiales (APG IV 2016). Compared with the recent 1KP study, we include five additional families (starting from the earliest diverging): Carlemanniaceae, Linderniaceae, Stilbaceae, Martyniaceae, and Mazaceae; the phylogenetic relationships among the 18 shared families among these phylogenomic studies are largely consistent ([supplementary fig. S9, Supplementary Material](#) online) (Leebens-Mack et al. 2019), except for Tetracondraceae, Schlegeliaceae, Scrophulariaceae, and Byblidaceae ([supplementary fig. S9, Supplementary Material](#) online). Among the families sampled here, Carlemanniaceae and Oleaceae form a maximally supported clade as sister to the other Lamiales, with Tetracondraceae occupying the next divergent branch, congruent with the plastome analysis, whereas in the 1KP phylogeny (without Carlemanniaceae), Oleaceae and Tetracondraceae form a strongly supported early branch ([supplementary fig. S9, Supplementary Material](#) online).

A large clade of 20 Lamiales families received 100% BS in all analyses here ([fig. 5](#)) and 100% jackknife values support with the *matK* gene previously (Hilu et al. 2003) and was proposed as core Lamiales by Schäferhoff et al. (2010). However, within core Lamiales, some interfamilial relationships were uncertain and no synapomorphy was recognized apart from the shared chemical trait of the presence of cornoside respectively absence of iridoids (Schäferhoff et al. 2010). The two early diverging clades in core Lamiales (Schäferhoff et al. 2010) are Calceolariaceae + Gesneriaceae (maximal support) and Plantaginaceae, with the remaining families forming a

maximally supported clade ([fig. 5](#)). Our results agree with recent nuclear gene and plastome analyses (Leebens-Mack et al. 2019; Li, Yi, et al. 2019). The placement of Byblidaceae remains unresolved – it is either retrieved as sister to the remaining lineages in the coalescent tree of 387 genes (with Linderniaceae + Scrophulariaceae being the next branch; [supplementary fig. S5, Supplementary Material](#) online) or in other positions with low supports in other analyses (26–61% BS, [supplementary fig. S8, Supplementary Material](#) online). In the 1KP study, Byblidaceae and Scrophulariaceae together occupy the next early-divergent lineage (Linderniaceae not sampled), whereas in the plastome study, Scrophulariaceae diverge first, followed by Byblidaceae and Linderniaceae (57 BS). Future analyses should probably expand sampling to resolve relationships of Byblidaceae and Linderniaceae + Scrophulariaceae to other core Lamiales families. Stilbaceae are found as sister to a highly supported clade (100% BS across all analyses), also named core Lamiales by Wortley et al. (2005). According to the Wortley et al. (2005) definition, core Lamiales comprise 13 families, including Orobanchaceae, Rehmanniaceae, Paulowniaceae, Mazaceae, Phrymaceae, Lamiaceae, Martyniaceae, Bignoniaceae, Verbenaceae, Schlegeliaceae, Acanthaceae, Lentibulariaceae, and Pedaliaceae. Topology retrieved here is in agreement with the 1KP study for the 11 shared families ([supplementary fig. S9, Supplementary Material](#) online) and is maximally supported for all 13 families included here ([fig. 5](#)). Two families not included in the 1KP study are firmly placed in the present study: Martyniaceae as sister to Bignoniaceae and Mazaceae sister to Phrymaceae ([fig. 5](#)). In our subsequent analyses, especially WGD analyses, we employed this definition of core Lamiales with 13 families (Wortley et al. 2005).

Overall, the present study has resolved several previously enigmatic relationships in asterids. The improved resolution here is a benefit of the expanded sampling with the inclusion of additional families compared with other recent studies. At the same time, the ambiguous nodes remaining in our phylogenies are often associated with limited sampling, for example, the uncertain placement of Byblidaceae (one of eight species sampled), Pentaphragmataceae (1 of ca. 25 species sampled), Styliidiaceae (2 of >240 species sampled). Thus, an expanded sampling in these clades will likely resolve such ambiguities in asterid phylogeny.

### Molecular Clock Estimate of an Aptian Origin of Crown Asterids

We used molecular clock analysis to estimate the origin and divergence times of asterids. We calibrate the asterid phylogeny with a total of 25 unambiguous fossils, representing the oldest known occurrence of relevant asterid and eudicot lineages ([supplementary table S5, Supplementary Material](#) online; see [supplementary note, Supplementary Material](#) online, for additional information on fossil calibration; [supplementary fig. S11, Supplementary Material](#) online). Fossil-derived minimum ages for 24 internal nodes among asterids were set in r8s, including one intrafamilial clade, 17 families, 2 suprafamilial clades, and 4 orders.

Our molecular clock estimations date the origin of the crown core eudicots to a median age of 125.64 Ma, close to the divergence time (123.75 Ma; [fig. 6](#)) between stem asterids and stem rosids in [Magallón et al. \(2015\)](#). The divergence time of stem asterids is dated to 124.92 Ma and that of crown asterids to 121.41 Ma (predating the asterid fossil record by 32.11 Ma) and suggesting an Aptian origin (Early Cretaceous, [fig. 6](#) and [supplementary table S5, Supplementary Material online](#)). This is in agreement with the previously proposed Early Cretaceous origin ([Magallón et al. 2015](#)) but is slightly older than that in [Li, Yi, et al. \(2019\)](#). The divergence times of asterid lineages ([fig. 6](#); see details in [supplementary table S5, Supplementary Material online](#)) support the idea that the crown groups of core asterids, lamiids, campanulids, core lamiids, and core campanulids originated in the Early Cretaceous ([fig. 6](#) and [supplementary figs. S11–S14 and table S5, Supplementary Material online](#)), with a small difference in the median ages of the origins. Divergences along the asterids backbone took place over a shorter period of time than divergence between the orders in the major asterids clades ([fig. 6](#)), suggesting rapid radiation of major asterids lineages, similar to what has been reported for rosids ([Magallón et al. 2015](#)). The radiation of backbone lineages also coincides with a dramatic climate shift (green bands in [fig. 6](#)), indicating a possible link between climate change and macroevolution ([Lyson et al. 2019](#)).

Ordinal-level lineages all appear to have originated in the Cretaceous and thus predate mass extinctions at the Cretaceous–Paleogene (K–Pg) boundary ([fig. 6](#)) as previously proposed by [Li, Yi, et al. \(2019\)](#) and [Barba-Montoya et al. \(2018\)](#). Again, these results are parallel to what has been found for rosids ([Magallón et al. 2015](#); [Barba-Montoya et al. 2018](#); [Li, Yi, et al. 2019](#)). At the family level, we find that 79 of 102 (77.5%) of the asterid families originated before the K–Pg mass extinction and an additional 21 (20.6%) before the Eocene–Oligocene extinction, indicating that the ancestors of many orders and families in asterids had survived these periods of mass extinctions. These divergences at the level of orders and families of the asterids in the Cretaceous (see divergences among other eudicot orders in [Li, Yi, et al. \[2019\]](#)) concomitantly paralleled the development of an increasingly wet climate and massive volcanic activity (see detailed discussion about links between climate change, geological events, and plant evolution in [Piombino \[2016\]](#)), leading to vicariance events and likely related to angiosperm-wide radiations ([Chaboureaud et al. 2014](#); [Piombino 2016](#); [Li, Yi, et al. 2019](#)).

### Ancestral State Reconstruction for Morphological Characters

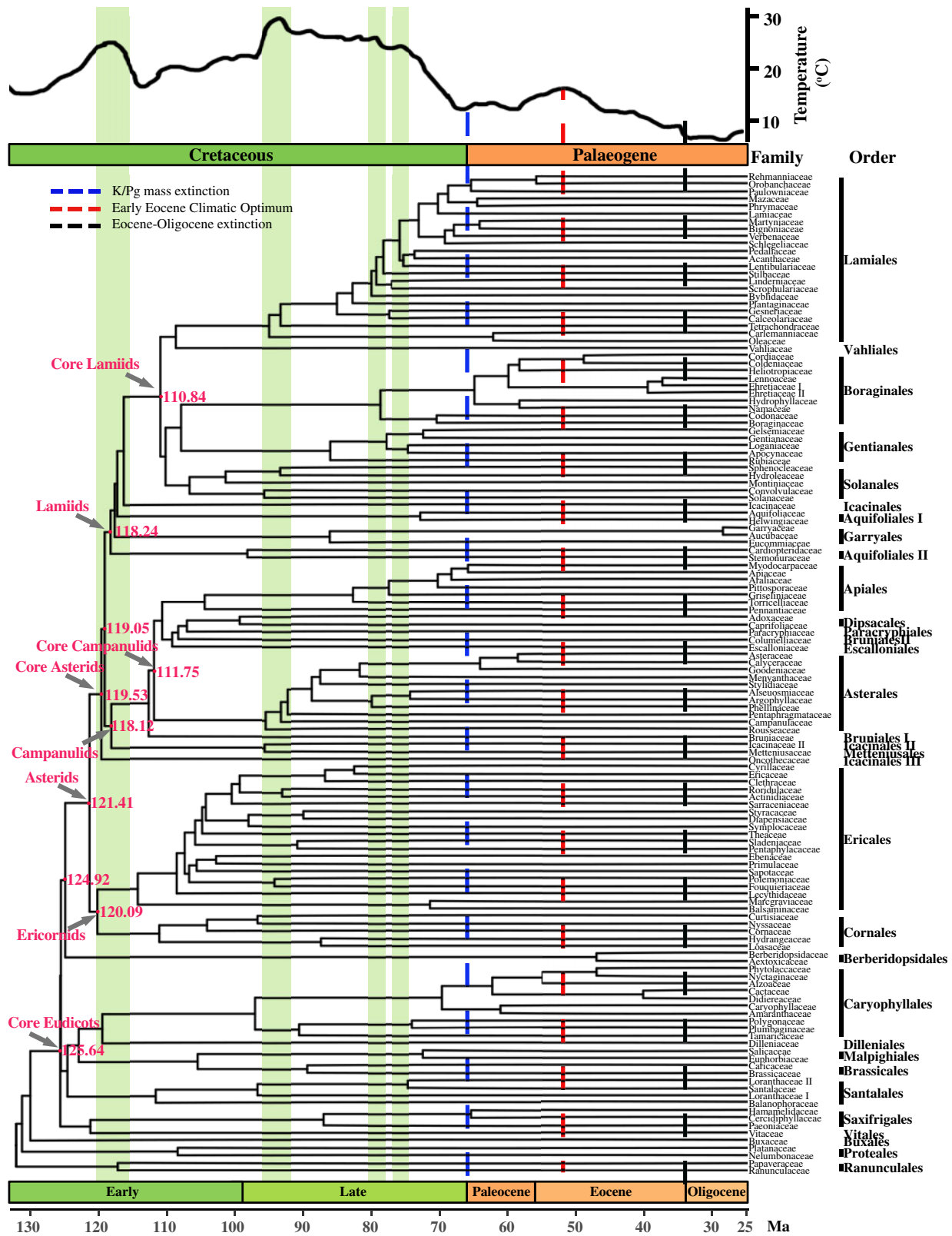
We performed ancestral state reconstruction of 12 characters at the family level to investigate the morphological evolution of asterids ([fig. 7](#) and [supplementary figs. S15–S24 and tables S6 and S7, Supplementary Material online](#)). Our results suggest that the asterids ancestor was a woody land plant with simple leaves, actinomorphic and bisexual flowers with free petals and free anthers, a superior ovary with a style, and drupaceous fruits ([fig. 7](#) and [supplementary figs. S15–S24, Supplementary Material online](#)). This combination of

characters is present in eight asterid families in the orders Escalloniales (Escalloniaceae), Ericales (Cyrillaceae, Ericaceae, Lecythidaceae, and Primulaceae), Icaciniales (Icacinaeae), and Aquifoliales II (Cardiopteridaceae and Stemonuraceae). Given the relatively early divergence of Ericornids, the asterids ancestor may have been similar to what is today found in this clade. In spite of the topological differences among phylogenies, the results of the morphological reconstructions agree with those based on plastid phylogenies ([Stull et al. 2018](#)) and previously proposed ancestral floral character states of the asterids ([Sauquet et al. 2017](#)).

Most asterids families are woody plants, but herbaceous growth forms evolved independently in all major orders, especially in multiple families of Asterales and Boraginales ([supplementary fig. S15, Supplementary Material online](#)), with transitions probably early in their evolutionary histories. Predominantly herbaceous families are also present in Cornales, Ericales, Solanales, Vahliales, and Lamiales. Marsh and water plants originated 14 times independently ([supplementary fig. S16, Supplementary Material online](#)) within families of Ericales, Asterales, Apiales, Gentianales, and especially in Lamiales. All those families have herbaceous representatives, but only Menyanthaceae are exclusively marsh and water plants.

Leaves are mostly simple, with 11 intrafamilial transitions toward compound leaves and only two families where compound leaves are predominant, namely Araliaceae and Bignoniaceae ([supplementary fig. S17, Supplementary Material online](#)). Although vegetative morphology is relatively homogeneous at family level in asterids ([Smith and Donoghue 2008](#)), several transitions are reported in this study. Since climatic niche evolution seems to be influenced by life history traits ([Smith and Beaulieu 2009](#)), and these traits have long been linked to ecological factors, homoplasy in vegetative morphology is to be expected, even at the intrafamilial level. For example, in Heliotropiaceae, [Luebert et al. \(2011\)](#) reported high levels of variability within genera with regard to growth form and leaf morphology.

Most asterids families have bisexual flowers, but dioecy (with male and female plants) evolved several times, especially in the early evolution of Garryales and Aquifoliales I ([supplementary figs. S18 and S19, Supplementary Material online](#)). Dioecy also occurs in Ebenaceae, Phellinaceae, Torricelliaceae, Griselinaceae, and Montiniaceae, all of which are predominately woody plants, as previously noted (e.g., [Bawa 1980](#)). In addition, sexual systems tend to be highly variable in certain asterid families ([Renner and Ricklefs 1995](#)), especially Nyssaceae, Sapotaceae, Asteraceae, Caprifoliaceae, and Cardiopteridaceae, with probably several transitions from and to dioecy ([Renner 2014](#)). Flower symmetry shows at least 13 transitions from actinomorphic to zygomorphic perianth ([supplementary fig. S20, Supplementary Material online](#)), especially during the early evolution of Lamiales, an order with predominantly zygomorphic families. Lamiales also appears to be the only order with the reversal from zygomorphic to actinomorphic perianth, at least seven times independently: Byblidaceae being the only exclusively actinomorphic family of the order. Flower

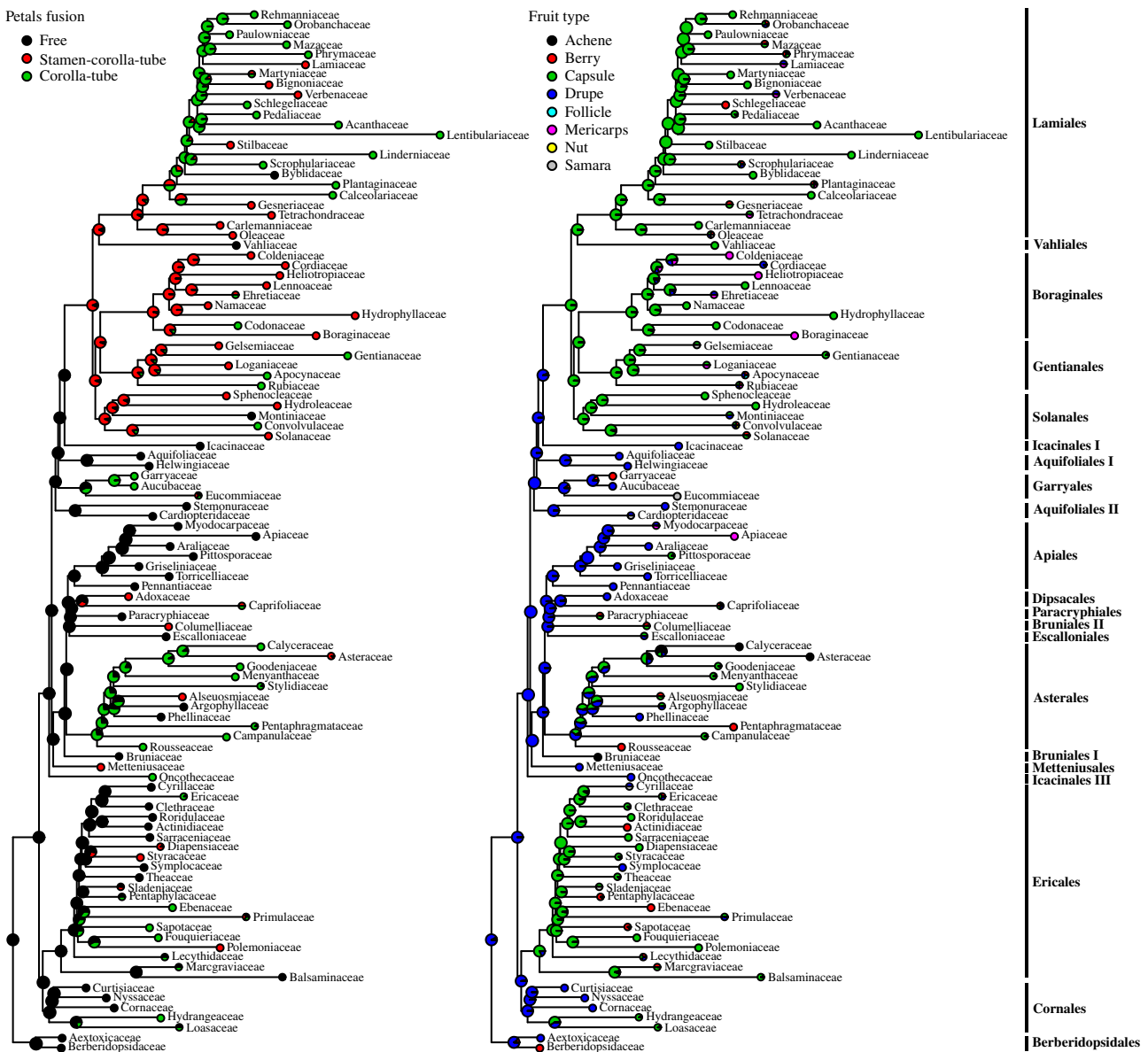


Downloaded from https://academic.oup.com/mbe/article/37/1/1/3188158/70293 by guest on 21 August 2022

**FIG. 6.** A dated phylogeny of family relationships in asterids. Red numbers next to nodes indicate the median age of divergence time. Top panel shows a graph of global temperature change (data compiled from [Veizer et al. \[2000\]](#) and [Zachos et al. \[2008\]](#)). Green bars indicate the periods with global temperature excursions in the Cretaceous.

symmetry likely diverged even more often at the intrafamilial level ([Reyes et al. 2016](#)), but this is beyond the scope of the present study.

The evolution of petal fusion is rather complex, with cho-ripetaly being the likely ancestral condition in asterids and sympetaly evolving several times independently from



**FIG. 7.** Ancestral character state reconstructions mapped at the family level, on the asterid tree. Petal fusion is shown on the left tree and fruit type on the right tree. Asterid orders are indicated at far right. The asterid phylogeny is from [supplementary figure S7, Supplementary Material](#) online.

choripetaly (fig. 7). There are also several reversals to chori-petaly in Asterales, Solanales, Vahliales, and Lamiales. Stamen–corolla–tubes (i.e., filaments fused with the corolla tube), evolved several times from chori-petalous ancestors or groups with corolla–tubes only. There are seven reversals in the lamiids from stamen–corolla–tubes to corolla–tubes with free filaments, especially in the early evolution of lamiids. Sympetaly has traditionally been considered as one of the diagnostic characters of asterids (e.g., Endress 1996, 2011b) but can be shown to be homoplasious in the present and previous studies (e.g., Sauquet et al. 2017; Stull et al. 2018), with both early and late sympetaly present in the asterids (Endress 1996).

Anthers are free in most families and some type of anther fusion likely evolved 12 times independently ([supplementary fig. S21, Supplementary Material](#) online). Pairwise-united anthers are sometimes present in Solanales and Lamiales.

Anther tubes evolved in Ericales, Asterales, and Gentianales, whereas anther caps evolved exclusively in Heliotropiaceae in Boraginales. This variation is expected, since anther fusion is intimately linked to flower biology (e.g., Endress 2009; Ren and Tang 2010) and thus arguably under very strong selective pressure.

A superior ovary is the most common and likely ancestral condition in asterids ([supplementary fig. S22, Supplementary Material](#) online), with inferior and semisuperior ovaries common in early branches such as Cornales, campanulids and Garryales. Inferior ovaries also evolved in individual families in the orders Ericales, Aquifoliales I, Solanales, Gentianales, Vahliales, and Lamiales. At the suprafamilial level, evolution of inferior ovaries from superior ones appears to be the major trend, but secondarily superior ovaries are also found (Endress 2011b). Representatives with free styloids evolved from ancestors with a united style in most orders of asterids

(supplementary fig. S23, Supplementary Material online), namely Cornales, Ericales, Icaciniales III, Bruniales I, Asterales, and Dipsacales and in the early evolution of Apiales, Garryales, Solanales, Gentianales, Boraginales, and Vahliales. Evolution of the style has rarely been addressed in evolutionary studies and no review is known to us.

Fleshy, indehiscent drupes likely represent the ancestral fruit type (fig. 7), with two principal transitions to capsular (and dehiscent) fruits, once in Ericales and one at the base of core lamiids. The ancestral fruit type in Asterales is equivocal: Capsules may represent the plesiomorphic condition in Asterales or they may have arisen several times independently. Another transition to capsules took place within Cornales, toward the origin of the clade Loasaceae + Hydrangeaceae. Berries originated from drupes at least twice independently, in Asterales and Garryales. Reversals from capsular fruits to indehiscent fruits (drupes, mericarps, and berries) took place several times in Ericales and in all major orders of core Lamiids. Reconstruction in Asterales is equivocal; berries may have arisen from drupaceous ancestors. Our reconstruction contradicts Beaulieu and Donoghue (2013), who inferred capsule as the ancestral condition of campanulids (as opposed to drupes in the present study). Differences may be due to taxon sampling, to reconstruction methods, and/or to phylogenetic topology. Fruit evolution has long been studied in the evolutionary literature, especially in relation to the evolution of the ovary, from which fruits develop (reviewed by Endress [2011b]). Several processes might be involved in the evolutionary transitions between fruit types. They include changes in ovary position, number and survival of ovules, number and fusion of carpels, bulging of ovaries, and/or placentation. Additionally, phytochemistry and structural properties (such as the occurrence of mineralized trichomes in Boraginaceae and Loasaceae + Hydrangeaceae) may determine the degrees of freedom in fruit evolution.

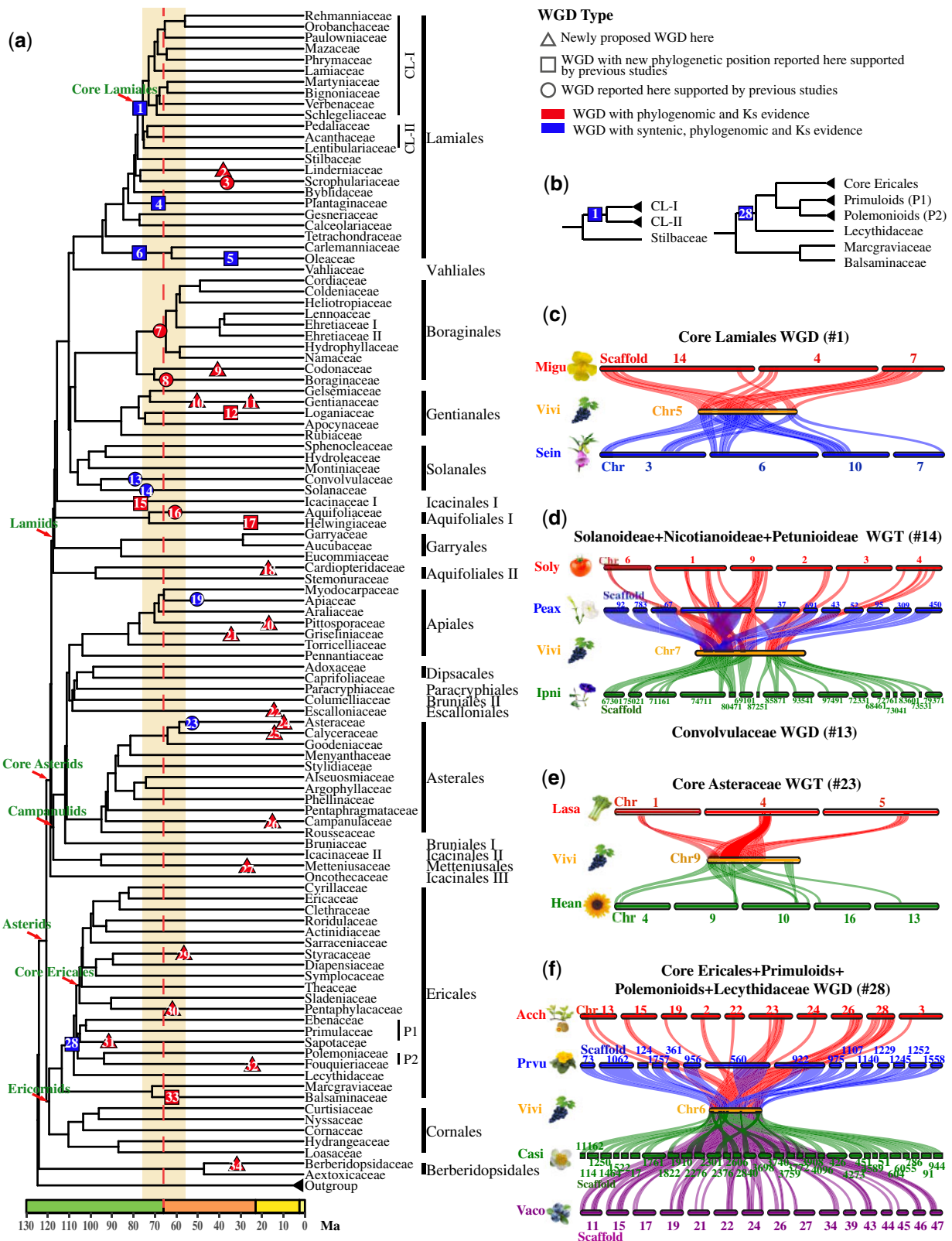
### Phylogenomic Analyses Uncover Strong Evidence for Numerous WGDs in Asterids

WGDs and the resulting new gene copies are thought to support key angiosperm functional innovations and coevolution with animals, contributing to angiosperm diversification (Schranz et al. 2012; Soltis PS and Soltis DE 2016; Sauquet and Magallón 2018). Numerous WGD and whole-genome triplication (WGT) events in plants are supported by chromosomal synteny, dating of peaks of synonymous substitution rates ( $K_s$ ) between paralogs, and phylogenetic placement of many GDs onto a species phylogeny (Cui et al. 2006; Barker et al. 2008, 2016; Shi et al. 2010; Jiao et al. 2011; Li et al. 2015; Huang, Zhang, et al. 2016; Ruprecht et al. 2017; Xiang et al. 2017; Julca et al. 2018; Ren et al. 2018; Wang et al. 2018; Yuan et al. 2018; Leebens-Mack et al. 2019; Qiao et al. 2019; Zwaenepoel and Van de Peer 2019). We integrated the phylogenomic methods,  $K_s$  evidence, and genome syntenic analyses and identified 33 asterid WGDs/WGTs (#1–33 in fig. 8), one WGD in Berberidopsidales, and numerous additional large-scale GD bursts in asterids (see supplementary figs. S25–S45 and tables S8–S13, Supplementary Material online). These events are consistent with reported WGDs/WGTs

from analyses of genome sequences (at least 14 WGD events [including lineage-specific WGDs] and four WGT events [including the  $\gamma$  WGT shared by asterids and other core eudicots]) (Tomato Genome Consortium 2012; Hellsten et al. 2013; Ibarra-Laclette et al. 2013; Denoeud et al. 2014; Kim et al. 2014, 2018; Wang et al. 2014; Bombarely et al. 2016; Hoshino et al. 2016; Iorizzo et al. 2016; Badouin et al. 2017; Reyes-Chin-Wo et al. 2017; Sollars et al. 2017; Unver et al. 2017; Xia et al. 2017; Xu et al. 2017; Zhang et al. 2017; Wuyun et al. 2018; Li, Zhang, et al. 2019; Pu et al. 2020; Tang et al. 2020). Our results also agree with the WGDs in some asterid families (e.g., Actinidiaceae, Asteraceae, Boraginaceae, Oleaceae, and Solanaceae) identified by previous phylogenomic analyses (Barker et al. 2008, 2016; Shi et al. 2010; Huang, Zhang, et al. 2016; Xu et al. 2017; Julca et al. 2018; Ren et al. 2018; Wang et al. 2018), and with most of the comparable WGDs in the 1KP study (Leebens-Mack et al. 2019) (supplementary table S11, Supplementary Material online; see below for a comparison).

The GD bursts supporting WGDs here were detected on branches with different lengths, which could result in errors in GD detection (Li et al. 2018). To test this, we applied the MAPS pipeline as used in the 1KP study (Leebens-Mack et al. 2019) to test the WGDs shared by some large groups and the newly proposed WGDs (see supplementary note and fig. S33, Supplementary Material online). The MAPS analyses here retrieved consistent results supporting several WGDs shared by multiple species, such as the core Lamiales WGD (#1 in fig. 8; supplementary figs. S25 and S33, Supplementary Material online), Boraginales II (#7 in fig. 8; supplementary figs. S27 and S33, Supplementary Material online), Boraginaceae (#8 in fig. 8; supplementary figs. S27 and S33, Supplementary Material online), and a clade with core Ericales + Primuloids + Polemonioids + Lecythidaceae WGD (#28 in fig. 8; supplementary figs. S31 and S33, Supplementary Material online), and also identified a new Sapotaceae WGD (#30 in fig. 8; supplementary figs. S31 and S33, Supplementary Material online).

To obtain further evidence for the WGDs proposed from phylogenomic analyses here and to reduce possible effects of recent small-scale duplication,  $K_s$  analyses were performed following the methods in the 1KP study (Leebens-Mack et al. 2019) for paralogs (including syntenic gene pairs) from representative taxa sharing one of the 33 asterid WGDs (#1–33) and one (#34) in Berberidopsidales (supplementary fig. S34, Supplementary Material online). Estimation of overall  $K_s$  distribution significantly different from null  $K_s$  simulation, which assumes a constant rate over time for independent event of gene birth and death (Cui et al. 2006), has been used to detect the effect of recent small-scale duplications (e.g., Barker et al. 2008; Leebens-Mack et al. 2019). For this purpose,  $K_s$  values of syntenic gene pairs were analyzed for species with public genome sequences, and  $K_s$  distribution of paralogs for the species with transcriptome data sets was compared with  $K_s$  from null simulation (see detailed methods in supplementary note, Supplementary Material online, and a summary of  $P$  value in supplementary fig. S34, Supplementary Material online), further supporting the 34 events (#1–34) as



proposed WGDs. In addition, the age distribution estimated from  $K_s$  values of all GDs (nodes) in a gene phylogeny (e.g., Vanneste et al. 2013; Leebens-Mack et al. 2019) and/or from  $K_s$  of retained paralogs after WGDs shared by two or more species can provide a highly sensitive approach to support WGDs, as discussed previously (see a detailed discussion in Li and Barker [2020]).

Recently, the 1KP study reported 41 WGDs (including 18 previously reported and 23 newly proposed ones) in asterids (Leebens-Mack et al. 2019). Among the 41 WGDs, nine events (including eight previously reported and one new ones; referred to as Type-I WGDs) were identified by both GD clusters with MAPS tests and dated  $K_s$  peaks; ten events (including five previously reported and five new ones; referred to as Type-II WGDs) were each supported by  $K_s$  peak evidence from at least two shared species; and 22 events (including five previously reported and 17 new ones; referred to as Type-III WGDs) were each supported by  $K_s$  peak evidence from a single species (Leebens-Mack et al. 2019).

An examination of taxon sampling related to the 1KP WGDs (Leebens-Mack et al. 2019) indicates that 27 of them correspond to groups with sampling of two or more species in our study, sufficient for potential detection of WGD(s) according to the criteria here. Among these 27 WGDs, our results support 23 (supplementary table S11, Supplementary Material online). Specifically, our results agree with eight Type-I WGDs in 1KP (Leebens-Mack et al. 2019), including four proposed WGDs (#13, 14, 16, and 23 in fig. 8), two candidate WGDs and two GD bursts (supplementary figs. S27–S30, S32, S34, and S35, Supplementary Material online). Among the ten Type-II WGDs reported in the 1KP study (Leebens-Mack et al. 2019), seven events are supported by five proposed WGDs here (#1, 3, 5, 6, and 19, fig. 8) and two additional GD bursts (supplementary table S11, Supplementary Material online) (supplementary figs. S25, S26, S28, S30, and S34, Supplementary Material online). Also, eight of the 22 Type-III 1KP WGDs (Leebens-Mack et al. 2019) are for groups with two or more species in this study and are supported by results here, with four proposed WGDs (#12, 15, 17, and 33, fig. 8), one candidate WGD and three GD burst (supplementary table S11, Supplementary Material online). Thus, the difference in WGD detection between the 1KP study and this one is largely due to differences in sampling, as 14 of the 1KP WGDs do not have sufficient sampling here for the criterion of at least two species sharing an event. Also, compared with the 1KP study and other previous reports, we present 17 newly proposed WGDs, ten of which are shared by taxa unique to our sampling (see detailed comparison in supplementary table S12, Supplementary Material online). Furthermore, our well-resolved phylogeny and extensive sampling here allowed refinement of the phylogenetic placement of nine (#1, 4–6, 12, 15, 17, 28, and 33) previously reported WGDs (supplementary table S12, Supplementary Material online).

Among the 34 WGDs we detected (fig. 8), six are within Ericornids (#28–33 in fig. 8; including the newly identified WGDs for *Sinojackia-Pterostyrax* [#29; Styracaceae], Sapotaceae [#31], *Fouquieria* [#32; Fouquieriaceae], and

*Eurya* [#30; Pentaphragaceae]) and a new event within Berberidopsidales, sister of asterids (#34 in fig. 8; see also supplementary figs. S31 and S32 and table S12, Supplementary Material online). Among the four new WGDs in Ericornids, the Sapotaceae WGD with  $K_s$  and phylogenomic evidence had highly significant support ( $P$  value =  $8.37e-28$ ) from the MAPS test (supplementary figs. S31, S33, and S34 and tables S10 and S12, Supplementary Material online). In addition, a WGD shared by three *Impatiens* species (#33 in fig. 8; supplementary fig. S31, Supplementary Material online) was also detected by the 1KP study (Leebens-Mack et al. 2019).

Previous analysis of the kiwifruit (*Actinidia chinensis*) genome identified a WGD called Ad- $\beta$  (Shi et al. 2010; Huang et al. 2013), which was recently mapped to the core Ericales (from Ericaceae to Sladeniaceae) with the MAPS pipeline (Leebens-Mack et al. 2019). Our analyses further revised the position of Ad- $\beta$  at the common ancestor (#28, fig. 8; supplementary table S12, Supplementary Material online) of Lecythidaceae, two suprafamilial clades Polemonioids (Fouquieriaceae and Polemoniaceae) and Primuloids (Sapotaceae, Ebenaceae, and Primulaceae), in addition to core Ericales, with an estimate age of  $\sim 108.59$  Ma. This conclusion is supported by multiple lines of evidence: 1) strong phylogenomic signal (748 GDs with 257 GDs in (AB)(AB) retention type; supplementary fig. S31, Supplementary Material online); 2) a percentage (19.07%) of subtree duplication significantly above the null simulation from the MAPS test (supplementary fig. S33, Supplementary Material online); 3)  $K_s$  peak value (0.7) of syntenic genes from Ad- $\beta$  WGD event higher than that (0.68) of orthologs between *A. chinensis* and *Napoleonea vogelii* and lower than that (1.42) of orthologs between *A. chinensis* and *Impatiens notolophora* (supplementary fig. S34, Supplementary Material online); 4) phylogenetic placement of GDs represented by paralogs in syntenic blocks in the kiwifruit and tea genomes at the MRCA of core Ericales, Primuloids, Polemonioids, and Lecythidaceae (supplementary fig. S45, Supplementary Material online); 5) mapping of kiwifruit duplicates in 36.51% of gene trees exhibiting the (AB)(AB) retention type to syntenic genomic blocks (supplementary table S8, Supplementary Material online), providing further syntenic support for the Ad- $\beta$  WGD event at the common ancestor of Core Ericales, Primuloids, Polemonioids, and Lecythidaceae (fig. 8 and supplementary figs. S44 and S45, Supplementary Material online; for additional discussions about WGDs in Ericornids, see supplementary note, Supplementary Material online).

Our results support 18 WGDs in Lamiids (#1–18 in fig. 8), including five newly proposed WGDs (*Torenia-Lindernia* [#2 Linderniaceae], *Codon* [#9 Codonaceae], Gentianaceae [#10 and 11], and *Gonocaryum* [#18 Cardiopteridaceae]) with phylogenomic and  $K_s$  evidence (see also supplementary figs. S25–S28 and S32–S34 and table S12, Supplementary Material online). Among the 18 WGDs, six WGDs (#1, 4–7, and 13) and one WGT (#14) previously revealed by genomic synteny (Tomato Genome Consortium 2012; Hellsten et al. 2013; Ibarra-Laclette et al. 2013; Bombarely et al. 2016; Hoshino et al. 2016; Unver et al. 2017; Li, Zhang, et al. 2019; Tang



et al. 2020) are also strongly supported by  $K_s$  and phylogenomic evidence here (see [supplementary figs. S25–S28 and S34, Supplementary Material](#) online); moreover, among these seven WGDs/WGT with genomic support, the phylogenetic positions of four are revised here, respectively, to be core Lamiales WGD (#1 in [fig. 8; supplementary fig. S37, Supplementary Material](#) online), *Antirrhinum*-specific WGD (#4 in Plantaginaceae, [fig. 8; supplementary fig. S36, Supplementary Material](#) online), a WGD in Oleaceae (#5 in [fig. 8; supplementary figs. S34, S38, and S39, Supplementary Material](#) online), and a WGD shared by Carlemanniaceae and Oleaceae (#6 in [fig. 8; see details in supplementary note and table S12, Supplementary Material](#) online).

Among the remaining 11 WGDs in Lamiids, which are supported with evidence from phylogenomic and  $K_s$  analyses (but without available genomic sequences) (#2, 3, 8–12, and 15–18 in [fig. 8](#)), three WGDs have newly revised phylogenetic positions from phylotranscriptomics analyses here and are shared, respectively, by taxa in Icacinae I (#15, [fig. 8](#)), *Helwingia* (#17 in Helwingiaceae, [fig. 8](#)), and *Strychnos* (#12 in Loganiaceae, [fig. 8](#)) (see [supplementary figs. S26–S28, S32, and S34, Supplementary Material](#) online). In addition, these three WGDs were also supported each by  $K_s$  evidence of a single species in the 1KP study ([Leebens-Mack et al. 2019](#)).

In Campanulids, our phylogenomic and  $K_s$  analyses support nine WGDs, including seven newly proposed events (*Pittosporum* [#20 in Pittosporaceae], *Griselinia* [#21 in Griselinaceae], *Escallonia* [#22 in Escalloniaceae], *Dasyphyllum-Arnaldoa* [#24 in Asteraceae], *Nastanthus-Gamocarpha-Calycera* [#25 in Calyceraceae], *Campanula* [#26 in Campanulaceae], and *Platea* [#27 in Metteniusaceae]; [fig. 8 and supplementary figs. S29, S30, S33, and S34 and table S12, Supplementary Material](#) online]. In Apiaceae, previous analysis of the carrot (*Daucus carota*) genome supports two lineage-specific WGDs, referred to as Dc- $\alpha$  and Dc- $\beta$  ([Iorizzo et al. 2016](#)). Our phylogenetic,  $K_s$ , and syntenic results (#19 in [fig. 8; supplementary figs. S30, S34, and S42, Supplementary Material](#) online) placed the Dc- $\alpha$  WGD to the ancestor of the subfamily Apioideae (see details in [supplementary note, Supplementary Material](#) online), in agreement with the 1KP study. In Asteraceae, the largest family in asterids, our analyses identified 215 GDs to core Asteraceae (#23 in [fig. 8; see details in supplementary note, figs. S29 and S43, and table S12, Supplementary Material](#) online). These results are consistent with a previously proposed WGD that is shared by several subfamilies of core Asteraceae reported in multiple studies ([Barker et al. 2008, 2016; Huang, Zhang, et al. 2016](#)) and a WGT event shared by genomes of sunflower (*Helianthus annuus*, Asteroideae), lettuce (*Lactuca sativa*, Cichorioideae), and artichoke (*Cynara scolymus*, Carduoideae) ([Badouin et al. 2017; Reyes-Chin-Wo et al. 2017](#)).

### Phylogenomics of WGDs in Asterids Support Contribution to Diversity

Our phylogenomics here provide strongly supported resolution of asterids phylogeny at the order and family levels, with molecular clock estimates of the origins and divergence times

of asterids and their major lineages. These results combined with the detection and phylogenetic placement of numerous WGDs, many of which are supported by previous analyses, allow for possible linkage of WGDs to changes in diversification. Among the asterids WGDs detected here, eight were placed near the K–Pg boundary ( $\pm 10$  My) ([fig. 8 and supplementary table S10, Supplementary Material](#) online), including one suprafamilial WGD (#7 in Boraginales), one familial WGD (#8 at Boraginaeae), one subfamilial WGD (#14 within Solanaceae), and five generic WGD (#4 within Plantaginaceae, #16 within Aquifoliaceae, #29 within Styracaceae, #30 within Pentaphragaceae, and #33 within Balsaminaceae). Recently, it was suggested that many WGDs are associated with upshifts in diversification rates in angiosperms and younger WGDs tend to be followed by an increase in diversification more than older WGDs ([Landis et al. 2018](#)).

Sometimes, ancient WGDs are followed by “lag-times” before increased diversification ([Schranz et al. 2012; Landis et al. 2018](#)). Indeed, after some of the WGDs reported here, there were taxon radiations in the corresponding lineages (e.g., core Lamiales, Boraginales II, core Asteraceae, and Apioideae). It was recently reported that a substantial acceleration of diversification rates (rate = 0.171, [Mandel et al. 2019](#)) occurred in Asteraceae after the origin of Carduoideae and before the divergence into Cichorioideae and Asteroideae, ultimately resulting in nearly 95% of the extant diversity in Asteraceae. This diversification rate upshift occurred after a time lag following the ancient core Asteraceae WGT (#23, [fig. 8](#)), consistent with the idea that WGDs might have contributed to the survival of some lineages through periods of severe environmental stress ([Van de Peer et al. 2017](#)), allowing their subsequent radiation to occupy new niches. Therefore, our results provide new evidence for the evolutionary hypothesis linking polyploidy to survival in periods of environmental and ecological crisis ([Van de Peer et al. 2017](#)). Long-term survival of lineages may depend on finding a novel adaptive plateau; this condition might be achieved only after a time interval and only a few of the surviving lineages might then diversify to occupy available niches.

Our results also showed a putative link between the evolutionary histories of morphological characters and the change of temperature ([figs. 6 and 7](#)). The transitions from woody to herbaceous ([supplementary fig. S15, Supplementary Material](#) online), from bisexual to unisexual ([supplementary fig. S18, Supplementary Material](#) online), from hermaphroditic to other sexual system ([supplementary fig. S19, Supplementary Material](#) online), and from indehiscent fruit to dehiscent fruit ([supplementary fig. S24, Supplementary Material](#) online) in lamiids, and the transition of ovary positions in campanulids occurred coincidentally with the change of high temperature (the first green band from the left in [fig. 6](#)). During the second period of obvious temperature change (the second green band in [fig. 6](#)), a coincident transition also occurred in Lamiales from actinomorphic to zygomorphic ([supplementary fig. S20, Supplementary Material](#) online). The coincidence between the temperature change and the transition of these characters of key

significances to radiation prior to K–Pg boundary might suggest that improved ecological conditions had a positive contribution to morphological character evolution.

Flower and fruit play a crucial role in the adaptive radiation of angiosperms. Duplicates retained after ancient WGDs include genes with important functions for developmental processes, with a strong tendency for such genes to be retained in long-term evolution (Van de Peer et al. 2017). Indeed, the retained duplicates from the WGD associated with core Lamiales can be implicated in the regulation of flower development. Nine genes with the (AB)(AB) retention type after WGD and with synteny support from both the sesame and monkey flower genomes are homologs of regulators of reproductive development, including pollen development and pollen tube growth (supplementary table S14, Supplementary Material online). Pollen tube growth is a key process for double-fertilization, which is a defining feature of angiosperms and is believed to have contributed to their evolutionary success (Lopes et al. 2019). Specifically, duplicates retained after the core Lamiales WGD are homologous to the *ALMT* gene encoding a transducer of GABA signaling in preovular guidance (Ramesh et al. 2015) and the *MIK1* gene encoding a receptor for perception of the female attractant LURE1 in ovular guidance (Wang et al. 2016). These and other duplicates from the WGDs might have impacted the evolution of relevant lineages by enhancing key regulatory networks and thus contributing to evolutionary diversification.

Moreover, paralogs from the WGT shared by Solanoideae, Nicotianoideae, and Petunioideae (the single-nucleotide polymorphism WGT) have also been linked to the evolution of fleshy fruits in tomato (Tomato Genome Consortium 2012) and potentially in other berry-bearing species in Solanoideae, although not in Nicotianoideae nor Petunioideae. Fruits protect developing seeds and facilitate their dispersal, contributing to angiosperm diversification (reviewed by Seymour et al. [2013]). Analyses of genes retained from the single-nucleotide polymorphism WGT for overrepresented Gene Ontology categories revealed genes related to crucial developmental programs, especially for ethylene response and fruit ripening (supplementary fig. S46, Supplementary Material online), as almost two-thirds of genes related to the ethylene signal pathway were retained after the WGT (supplementary figs. S47–S51, Supplementary Material online). It is plausible that the genes from the WGT enhanced the ripening process of the berry type fleshy fruit of tomato and other Solanoideae members and contributed to the evolutionary success of these species by optimizing fruit quality for animal-facilitated seed dispersal.

## Materials and Methods

### Sequencing and Assembly of Unigenes and Selection of Low-Copy Nuclear Genes

Total RNA was extracted from fresh leaves and/or floral buds, silica-dried leaves, or seeds using TRIzol reagent (Invitrogen, Carlsbad, CA). Total DNA was extracted using the CTAB method from dried leaf or other samples. Paired-end cDNA or genomic DNA libraries were generated and sequenced on a

HiSeq 3000 platform (Illumina, San Diego, CA). Raw sequence reads were trimmed by using Trimmomatic v0.38 (Bolger et al. 2014) and were de novo assembled into contigs by using Trinity v2.4.0 (Grabherr et al. 2011) and SOAPdenovo v2.04-r240 (Luo et al. 2012). The longest isoforms of predicted coding sequences of transcripts were identified by CD-HIT v4.6 (Li and Godzik 2006). Additional information is provided in the supplementary note, Supplementary Material online. Previously a set of 4,180 low-copy nuclear genes was identified from nine representative angiosperm species using HaMStR (Ebersberger et al. 2009) and further filtered (see supplementary note, Supplementary Material online) to five sets of nuclear genes with 1,769, 1,041, 784, 565, and 387 putative ortholog groups (supplementary table S3, Supplementary Material online). The alignments of the ortholog groups in various gene sets were generated using MUSCLE v3.8.31 (Edgar 2004) with default setting and then manually adjusted by using AliView v1.19 (Larsson 2014). The raw sequence reads have been deposited in GenBank under a BioProject with accession number PRJNA636634.

### Phylogenetic Analyses

We utilized a coalescent approach to infer asterids phylogeny using data sets of 1,769, 1,041, 784, 565, and 387 genes, respectively (see species trees in supplementary figs. S1–S5, Supplementary Material online). Protein sequences were aligned by MUSCLE v3.8.31 (Edgar 2004) and used for constructing gene trees using a ML method as implemented in RAxML v7.0.4 (Stamatakis 2014). BS values were estimated with 100 replicates using the GTRCAT model. The best ML trees were used for a coalescent tree inference using Accurate Species Tree Algorithm (ASTRAL v4.4.4) (Zhang et al. 2018). Then, using the summarized phylogeny from the above mentioned five trees, we computed branch lengths using IQ-TREE v1.6.10 (Nguyen et al. 2015) with the 387 gene data set, generating a final ML tree (supplementary fig. S7, Supplementary Material online) for subsequent analyses. All trees were edited and drawn using ggtree v1.14.6 (Yu et al. 2017).

### Fossil Data and Molecular Dating

We inferred the evolutionary timescale of asterids using the penalized likelihood approach implemented in r8s v1.8.1 (Sanderson 2003). We conservatively calibrated the crown eudicots with a fixed range (132–125 Ma). Other fossil calibration information can be found in supplementary table S4, Supplementary Material online. An optimal smoothing value of 0.32 was selected from the cross-validation with a range of smoothing parameters from 0.01 to 100,000 (cvstart = −2, cvinc = 0.5, and cvnum = 15). The mean and 95% confidence time intervals of each node were summarized onto a final consensus time tree using the program TreeAnnotator v1.7.5.

### Ancestral Character Reconstruction

We reconstructed the ancestral states of 12 characters (supplementary table S6, Supplementary Material online) at family level. These characters were selected based on four criteria: 1) they are variable within the asterids, 2) they have been used for diagnosing taxa, 3) the information can be obtained from

traditional taxonomic works, and 4) their coding is straightforward at family level. Information was obtained from several sources, mostly comprehensive family revisions or detailed family descriptions (Kubitzki 1990–2018; Stevens 2001; Takhtajan 2009; Simpson 2010), as well as specific works for certain families (supplementary table S7, Supplementary Material online). Ancestral character state reconstruction was carried out based on the ML tree of 387-gene data set and using a stochastic character mapping (SIMMAP; Bollback 2006) as implemented in the R-package phytools v.0.6-60 (Revell 2012). Equal prior probabilities of character states were assigned to each family, based on the information from the taxonomical descriptions, when different character states were reported for a single family and equally represented across its species. Character states that were reported as rare in the taxonomical descriptions were assigned a prior probability of 0.1 or an equal fraction of 0.1 if several states were reported as rare. Stochastic character mapping was conducted with the function `make.simmap` of package `phytools` with 1,000 simulations, setting the prior distribution of the root node of the tree ( $\pi$ ) to be estimated and using a Bayesian Markov chain Monte Carlo to obtain 1,000 values for the transition matrix  $Q$  from the posterior distribution.

### WGD Identification

Putative WGDs shared by at least two asterids species were detected using a phylogenomic tool, `tree2gd` v2.4 (the custom software is available from <https://sourceforge.net/projects/tree2gd/>; last accessed October 30, 2019). To reduce computing cost, asterids species were divided into eight overlapping groups (supplementary figs. S25–S32, Supplementary Material online) according to the phylogeny here, with consideration for the quality of transcriptomes and genomes. Additional WGD identification information using phylogenomic, MAPS,  $K_s$ , and syntenic analyses is provided in the supplementary note, Supplementary Material online, with a summary provided as follows. Gene families were identified by all-against-all BlastP followed by clustering with MCL. Multiple sequence alignments were performed by using MUSCLE v3.8.31 (Edgar 2004) and used for constructing ML trees using RAxML (Stamatakis 2014), with BS values estimated from 100 replicates using the GTRCAT model. The gene family trees were compared with the species tree to detect the presence and positions of GD events using a method used in a number of recent studies (e.g., Jiao et al. 2011; Ren et al. 2018). A node associated with a cluster of GDs (number > 200) was retained for further evaluation by considering the GD number and fraction of GDs with the (AB)(AB) or other types (see supplementary note, Supplementary Material online, for additional information). The WGD candidates shared by taxa without genome sequenced were further estimated by the  $K_s$  distribution using the approach in the 1KP study (Leebens-Mack et al. 2019; Li and Barker 2020). Some WGD candidates are also evaluated by using evidence from colinearity (synteny) analyses between genomic regions if genome sequence is available for at least one species in the studied lineage. To test the influence of branch length and gene birth and death rate on WGD

identification, several interfamilial WGDs and newly proposed ones were further tested by using the MAPS pipeline (Li et al. 2015; Godden et al. 2019; Leebens-Mack et al. 2019). To elucidate the functional enrichment of the retained duplicates from polyploidization, Gene Ontology and KEGG analyses were performed.

### Supplementary Material

Supplementary data are available at *Molecular Biology and Evolution* online.

### Acknowledgments

We thank Dr Yaping Chen, Dr Holly Forbes, Dr Xuejun Ge, Dr Jian Huang, Dr Qingfeng Wang, Dr Jun Xiang, Dr Tingshuang Yi, Dr Junwen Zhai, Dr Liangsheng Zhang, Dr Qiang Zhang, Dr Peter Brownless, the Bonn University Botanic Gardens, the Botanic Garden in Berlin, the Missouri Botanical Garden, the New York Botanical Garden, the Royal Botanic Garden Edinburgh, the Royal Botanic Gardens, Kew, the South China Botanical Garden, and the UC Berkeley Botanical Garden for plant materials; Prof. Claude W. dePamphilis for the transcriptome data set of *Pholisma arenarium*; Weicheng Yuan, Jiayi Tang, Beining Zhou, Huaqi Liu, Shaowen Lve, Yaning Shi, Xin, and Qing Qin for the plant drawings in figure 2; Chao Han for the plant drawings in figures 3 and 5; and Dr Xinhui Zhang the plant drawings in figure 4. This work was supported by funds from the National Natural Science Foundation of China (31770244 and 31970224), China Postdoctoral Science Foundation (2019M661344), and the Eberly College of Science and the Huck Institutes of the Life Sciences at the Pennsylvania State University, the Ministry of Education Key Laboratory of Biodiversity Science and Ecological Engineering, and State Key Laboratory of Genetic Engineering at Fudan University.

### Author Contributions

H.M., M.W., and J.Q. designed research; C.Z., T.Z., and F.L. contributed equally to this work. C.Z., F.L., Y.X., C.-H.H., Y.H., M.R., M.W.F., M.W., and H.M. contributed to the taxon sampling; C.Z. performed DNA/RNA data processing and phylogenetic analyses; F.L. conducted ancestral character reconstruction; T.Z. performed the molecular dating and WGD analyses, with assistance from J.Q. and H.M., and prepared figures and supplementary files; T.Z., C.Z., F.L., and H.M. wrote the draft manuscript; H.M., F.L., M.W., and M.W.F. revised the manuscript.

### References

- Albach DC, Soltis PS, Soltis DE, Olmstead RG. 2001. Phylogenetic analysis of asterids based on sequences of four genes. *Ann Mo Bot Gard*. 88(2):163–212.
- APG IV. 2016. An update of the angiosperm phylogeny group classification for the orders and families of flowering plants: APG IV. *Bot J Linn Soc*. 181:1–20.
- Badouin H, Gouzy J, Grassa CJ, Murat F, Staton SE, Cottret L, Lelandais-Brière C, Owens GL, Carrère S, Mayjonade B, et al. 2017. The sunflower genome provides insights into oil metabolism, flowering and Asterid evolution. *Nature* 546(7656):148–152.

- Barba-Montoya J, dos Reis M, Schneider H, Donoghue PCJ, Yang Z. 2018. Constraining uncertainty in the timescale of angiosperm evolution and the veracity of a Cretaceous Terrestrial Revolution. *New Phytol.* 218(2):819–834.
- Barker MS, Kane NC, Matvienko M, Kozik A, Michelmore RW, Knapp SJ, Rieseberg LH. 2008. Multiple paleopolyploidizations during the evolution of the Compositae reveal parallel patterns of duplicate gene retention after millions of years. *Mol Biol Evol.* 25(11):2445–2455.
- Barker MS, Li Z, Kidder TI, Reardon CR, Lai Z, Oliveira LO, Scascitelli M, Rieseberg LH. 2016. Most Compositae (Asteraceae) are descendants of a paleohexaploid and all share a paleotetraploid ancestor with the Calyceraceae. *Am J Bot.* 103(7):1203–1211.
- Barkman TJ, McNeal JR, Lim S-H, Coat G, Croom HB, Young ND, dePamphilis CW. 2007. Mitochondrial DNA suggests at least 11 origins of parasitism in angiosperms and reveals genomic chimerism in parasitic plants. *BMC Evol Biol.* 7(1):248.
- Bawa KS. 1980. Evolution of dioecy in flowering plants. *Annu Rev Ecol Syst.* 11(1):15–39.
- Beaulieu JM, Donoghue MJ. 2013. Fruit evolution and diversification in campanulid Angiosperms. *Evolution* 67(11):3132–3144.
- Birky CW. 2001. The inheritance of genes in mitochondria and chloroplasts: laws, mechanisms, and models. *Annu Rev Genet.* 35:26–33.
- Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30(15):2114–2120.
- Bollback JP. 2006. SIMMAP: stochastic character mapping of discrete traits on phylogenies. *BMC Bioinf.* 7(1):88.
- Bombarely A, Moser M, Amrad A, Bao M, Bapaume L, Barry CS, Bliker M, Boersma MR, Borghi L, Bruggmann R, et al. 2016. Insight into the evolution of the Solanaceae from the parental genomes of *Petunia hybrida*. *Nat Plants* 2(6):16074.
- Bremer B, Bremer K, Heidari N, Erixon P, Olmstead RG, Anderberg AA, Källersjö M, Barkhordarian E. 2002. Phylogenetics of asterids based on 3 coding and 3 non-coding chloroplast DNA markers and the utility of non-coding DNA at higher taxonomic levels. *Mol Phylogen Evol.* 24(2):274–301.
- Chaboureau A-C, Sepulchre P, Donnadiou Y, Franc A. 2014. Tectonic-driven climate change and the diversification of angiosperms. *Proc Natl Acad Sci U S A.* 111(39):14066–14070.
- Cui L, Wall PK, Leebens-Mack JH, Lindsay BC, Soltis DE, Doyle JJ, Soltis PS, Carlson JE, Arumuganathan K, Barakat A, et al. 2006. Widespread genome duplications throughout the history of flowering plants. *Genome Res.* 16(6):738–749.
- Dahlgren R. 1983. General aspects of angiosperm evolution and macro-systematics. *Nord J Bot.* 3(1):119–149.
- Dahlgren R. 1975. System of classification of the angiosperms to be used to demonstrate the distribution of characters. *Bot Notiser.* 128:119–147.
- Davis CC, Xi Z, Mathews S. 2014. Plastid phylogenomics and green plant phylogeny: almost full circle but not quite there. *BMC Biol.* 12(1):11.
- Denoed F, Carretero-Paulet L, Dereeper A, Droc G, Guyot R, Pietrella M, Zheng C, Alberti A, Anthony F, Aprea G, et al. 2014. The coffee genome provides insight into the convergent evolution of caffeine biosynthesis. *Science* 345(6201):1181–1184.
- Ebersberger I, Strauss S, von Haeseler A. 2009. HaMStR: profile hidden Markov model based search for orthologs in ESTs. *BMC Evol Biol.* 9(1):157.
- Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32(5):1792–1797.
- Edger PP, Heidel-Fischer HM, Bekaert M, Rota J, Glöckner G, Platts AE, Heckel DG, Der JP, Wafula EK, Tang M, et al. 2015. The butterfly plant arms-race escalated by gene and genome duplications. *Proc Natl Acad Sci U S A.* 112(27):8362–8366.
- Ellison AM, Gotelli NJ. 2009. Energetics and the evolution of carnivorous plants—Darwin's 'most wonderful plants in the world'. *J Exp Bot.* 60(1):19–42.
- Endress PK. 1996. Homoplasmy in angiosperm flowers. In: Sanderson MJ, Hufford L, editors. Homoplasmy. San Diego (CA): Academic Press. p. 303–325.
- Endress PK. 2009. Diversity and evolutionary trends in angiosperm anthers. In: D'Arcy WG, Keating RC, editors. The anther: form, function and phylogeny. Cambridge: Cambridge University Press. p. 92–110.
- Endress PK. 2011a. Angiosperm ovules: diversity, development, evolution. *Ann Bot.* 107(9):1465–1489.
- Endress PK. 2011b. Evolutionary diversification of the flowers in angiosperms. *Am J Bot.* 98(3):370–396.
- Gitzendanner MA, Soltis PS, Wong GKS, Ruhfel BR, Soltis DE. 2018. Plastid phylogenomic analysis of green plants: a billion years of evolutionary history. *Am J Bot.* 105(3):291–301.
- Godden GT, Kinser TJ, Soltis PS, Soltis DE. 2019. Phylotranscriptomic analyses reveal asymmetrical gene duplication dynamics and signatures of ancient polyploidy in mints. *Genome Biol Evol.* 11(12):3393–3408.
- Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, et al. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol.* 29(7):644–652.
- Hellsten U, Wright KM, Jenkins J, Shu S, Yuan Y, Wessler SR, Schmutz J, Willis JH, Rokhsar DS. 2013. Fine-scale variation in meiotic recombination in *Mimulus* inferred from population shotgun sequencing. *Proc Natl Acad Sci U S A.* 110(48):19478–19482.
- Hilu KW, Borsch T, Müller K, Soltis DE, Soltis PS, Savolainen V, Chase MW, Powell MP, Alice LA, Evans R, et al. 2003. Angiosperm phylogeny based on *matK* sequence information. *Am J Bot.* 90(12):1758–1776.
- Hoshino A, Jayakumar V, Nitasaka E, Toyoda A, Noguchi H, Itoh T, Shin-I T, Minakuchi Y, Koda Y, Nagano AJ, et al. 2016. Genome sequence and analysis of the Japanese morning glory *Ipomoea nil*. *Nat Commun.* 7(1):13295.
- Huang C-H, Sun R, Hu Y, Zeng L, Zhang N, Cai L, Zhang Q, Koch MA, Al-Shehbaz I, Edger PP, et al. 2016. Resolution of Brassicaceae phylogeny using nuclear genes uncovers nested radiations and supports convergent morphological evolution. *Mol Biol Evol.* 33(2):394–412.
- Huang C-H, Zhang C, Liu M, Hu Y, Gao T, Qi J, Ma H. 2016. Multiple polyploidization events across Asteraceae with two nested events in the early history revealed by nuclear phylogenomics. *Mol Biol Evol.* 33(11):2820–2835.
- Huang S, Ding J, Deng D, Tang W, Sun H, Liu D, Zhang L, Niu X, Zhang X, Meng M, et al. 2013. Draft genome of the kiwifruit *Actinidia chinensis*. *Nat Commun.* 4(1):2640.
- Huber H. 1963. Die Verwandtschaftsverhältnisse der Rosifloren. *Mitt. Bot. Staatssamml. München* 5:1–48.
- Ibarra-Laclette E, Lyons E, Hernandez-Guzman G, Perez-Torres CA, Carretero-Paulet L, Chang T-H, Lan T, Welch AJ, Juarez MJA, Simpson J, et al. 2013. Architecture and evolution of a minute plant genome. *Nature* 498(7452):94–98.
- Iorizzo M, Ellison S, Senalik D, Zeng P, Satapoomin P, Huang J, Bowman M, Iovene M, Sanseverino W, Cavagnaro P, et al. 2016. A high-quality carrot genome assembly provides new insights into carotenoid accumulation and asterid genome evolution. *Nat Genet.* 48(6):657–666.
- Jiao Y, Wickert NJ, Ayyampalayam S, Chanderbali AS, Landherr L, Ralph PE, Tomsho LP, Hu Y, Liang H, Soltis PS, et al. 2011. Ancestral polyploidy in seed plants and angiosperms. *Nature* 473(7345):97–100.
- Jin D, Wei H-J, Shang H, Shen H, Shu J-P, Liu L, Zhang R, Zhou X-L, Gu Y-F, Yan Y-H, et al. 2018. Large-scale phylogenomic analysis resolves a backbone phylogeny in ferns. *GigaScience* 7(2):gix116.
- Julca I, Marcet-Houben M, Vargas P, Gabaldón T. 2018. Phylogenomics of the olive tree (*Olea europaea*) reveals the relative contribution of ancient allo- and autopolyploidization events. *BMC Biol.* 16(1):15.
- Kim N-H, Jayakodi M, Lee S-C, Choi B-S, Jang W, Lee J, Kim HH, Waminal NE, Lakshmanan M, van Nguyen B, et al. 2018. Genome and

- evolution of the shade-requiring medicinal herb *Panax ginseng*. *Plant Biotechnol J*. 16(11):1904–1917.
- Kim S, Park M, Yeom S-I, Kim Y-M, Lee JM, Lee H-A, Seo E, Choi J, Cheong K, Kim K-T, et al. 2014. Genome sequence of the hot pepper provides insights into the evolution of pungency in *Capsicum* species. *Nat Genet*. 46(3):270–278.
- Kubitzki K. 1990–2018. The families and genera of vascular plants. Berlin (Germany): Springer.
- Landis JB, Soltis DE, Li Z, Marx HE, Barker MS, Tank DC, Soltis PS. 2018. Impact of whole-genome duplication events on diversification rates in angiosperms. *Am J Bot*. 105(3):348–363.
- Larsson A. 2014. AliView: a fast and lightweight alignment viewer and editor for large datasets. *Bioinformatics* 30(22):3276–3278.
- Leebens-Mack JH, Barker MS, Carpenter EJ, Deyholos MK, Gitzendanner MA, Graham SW, Grosse I, Li Z, Melkonian M, Mirarab S, et al. 2019. One thousand plant transcriptomes and the phylogenomics of green plants. *Nature* 574(7780):679–685.
- Li H-T, Yi T-S, Gao L-M, Ma P-F, Zhang T, Yang J-B, Gitzendanner MA, Fritsch PW, Cai J, Luo Y, et al. 2019. Origin of angiosperms and the puzzle of the Jurassic gap. *Nat Plants* 5(5):461–470.
- Li M, Zhang D, Gao Q, Luo Y, Zhang H, Ma B, Chen C, Whibley A, Zhang Y, Cao Y, et al. 2019. Genome structure and evolution of *Antirrhinum majus* L. *Nat Plants* 5(2):174–183.
- Li W, Godzik A. 2006. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* 22(13):1658–1659.
- Li Z, Baniaga AE, Sessa EB, Scascitelli M, Graham SW, Rieseberg LH, Barker MS. 2015. Early genome duplications in conifers and other seed plants. *Sci Adv*. 1(10):e1501084.
- Li Z, Barker MS. 2020. Inferring putative ancient whole-genome duplications in the 1000 Plants (1KP) initiative: access to gene family phylogenies and age distributions. *GigaScience* 9(2):1–11.
- Li Z, De La Torre AR, Sterck L, Cánovas FM, Avila C, Merino I, Cabezas JA, Cervera MT, Ingvarsson PK, Van de Peer Y. 2017. Single-copy genes as molecular markers for phylogenomic studies in seed plants. *Genome Biol Evol*. 9(5):1130–1147.
- Li Z, Tiley GP, Galuska SR, Reardon CR, Kidder TI, Rundell RJ, Barker MS. 2018. Multiple large-scale gene and genome duplications during the evolution of hexapods. *Proc Natl Acad Sci U S A*. 115(18):4713–4718.
- Lopes AL, Moreira D, Ferreira MJ, Pereira AM, Coimbra S. 2019. Insights into secrets along the pollen tube pathway in need to be discovered. *J Exp Bot*. 70(11):2979–2992.
- Lu L-M, Mao L-F, Yang T, Ye J-F, Liu B, Li H-L, Sun M, Miller JT, Mathews S, Hu H-H, et al. 2018. Evolutionary history of the angiosperm flora of China. *Nature* 554(7691):234–238.
- Luebert F, Brokamp G, Wen J, Weigend M, Hilger HH. 2011. Phylogenetic relationships and morphological diversity in neotropical *Heliotropium* (Heliotropiaceae). *Taxon* 60(3):663–680.
- Luebert F, Cecchi L, Frohlich MW, Gottschling M, Williams CM, Hasenstab-Lehman KE, Hilger HH, Miller JS, Mittelbach M, Nazaire M, et al. 2016. Familial classification of the Boraginales. *Taxon* 65(3):502–522.
- Lundberg J. 2009. Asteraceae and relationships within Asterales. In: Funk VA, Sussana A, Stuessy TF, Bayer RJ, editors. Systematics, evolution, and biogeography of Compositae. Vienna (Austria): International Association for Plant Taxonomy, University of Vienna. p. 157–169.
- Lundberg J, Bremer K. 2003. A phylogenetic study of the order Asterales using one morphological and three molecular data sets. *Int J Plant Sci*. 164(4):553–578.
- Luo R, Liu B, Xie Y, Li Z, Huang W, Yuan J, He G, Chen Y, Pan Q, Liu Y, et al. 2012. SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *GigaScience* 1(1):6.
- Lyson TR, Miller IM, Bercovici AD, Weissenburger K, Fuentes AJ, Clyde WC, Hagadorn JW, Butrim MJ, Johnson KR, Fleming RF, et al. 2019. Exceptional continental record of biotic recovery after the Cretaceous–Paleogene mass extinction. *Science* 366(6468):977–983.
- Magallón S, Castillo A. 2009. Angiosperm diversification through time. *Am J Bot*. 96(1):349–365.
- Magallón S, Gómez-Acevedo S, Sánchez-Reyes LL, Hernández-Hernández T. 2015. A metacalibrated time-tree documents the early rise of flowering plant phylogenetic diversity. *New Phytol*. 207(2):437–453.
- Magallón S, Sanderson MJ. 2001. Absolute diversification rates in angiosperm clades. *Evolution* 55(9):1762–1780.
- Maia VH, Gitzendanner MA, Soltis PS, Wong G-S, Soltis DE. 2014. Angiosperm phylogeny based on 18S/26S rDNA sequence data: constructing a large data set using next-generation sequence data. *Int J Plant Sci*. 175(6):613–650.
- Mandel JR, Dikow RB, Siniscalchi CM, Thapa R, Watson LE, Funk VA. 2019. A fully resolved backbone phylogeny reveals numerous dispersals and explosive diversifications throughout the history of Asteraceae. *Proc Natl Acad Sci U S A*. 116(28):14083–14088.
- Moore MJ, Soltis PS, Bell CD, Burleigh JG, Soltis DE. 2010. Phylogenetic analysis of 83 plastid genes further resolves the early diversification of eudicots. *Proc Natl Acad Sci U S A*. 107(10):4623–4628.
- Morton CM. 2011. Newly sequenced nuclear gene (*Xdh*) for inferring angiosperm phylogeny. *Ann Mo Bot Gard*. 98(1):63–89.
- Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol*. 32(1):268–274.
- Olmstead RG, Bremer B, Scott KM, Palmer JD. 1993. A parsimony analysis of the Asteridae sensu lato based on rbcL sequences. *Ann Mo Bot Gard*. 80(3):700–722.
- Olmstead RG, Michaels HJ, Scott KM, Palmer JD. 1992. Monophyly of the Asteridae and identification of their major lineages inferred from DNA sequences of rbcL. *Ann Mo Bot Gard*. 79(2):249–265.
- Piombino A. 2016. The heavy links between geological events and vascular plants evolution: a brief outline. *Int J Evol Biol*. 2016:1–9.
- Potato Genome Sequencing Consortium. 2011. Genome sequence and analysis of the tuber crop potato. *Nature* 475(7355):189–195.
- Pu X, Li Z, Tian Y, Gao R, Hao L, Hu Y, He C, Sun W, Xu M, Peters RJ, et al. 2020. The honeysuckle genome provides insight into the molecular mechanism of carotenoid metabolism underlying dynamic flower coloration. *New Phytol*. 227(3):930–943.
- Qiao X, Li Q, Yin H, Qi K, Li L, Wang R, Zhang S, Paterson AH. 2019. Gene duplication and evolution in recurring polyploidization–diploidization cycles in plants. *Genome Biol*. 20(1):38.
- Ramesh SA, Tyerman SD, Xu B, Bose J, Kaur S, Conn V, Domingos P, Ullah S, Wege S, Shabala S, et al. 2015. GABA signalling modulates plant growth by directly regulating the activity of plant-specific anion transporters. *Nat Commun*. 6(1):7879.
- Refulio-Rodríguez NF, Olmstead RG. 2014. Phylogeny of Lamiidae. *Am J Bot*. 101(2):287–299.
- Ren MX, Tang JY. 2010. Anther fusion enhances pollen removal in *Campsis grandiflora*, a hermaphroditic flower with didynamous stamens. *Int J Plant Sci*. 171(3):275–282.
- Ren R, Wang H, Guo C, Zhang N, Zeng L, Chen Y, Ma H, Qi J. 2018. Widespread whole genome duplications contribute to genome complexity and species diversity in angiosperms. *Mol Plant* 11(3):414–428.
- Renner SS. 2014. The relative and absolute frequencies of angiosperm sexual systems: dioecy, monoecy, gynodioecy, and an updated online database. *Am J Bot*. 101(10):1588–1596.
- Renner SS, Ricklefs RE. 1995. Dioecy and its correlates in the Flowering Plants. *Am J Bot*. 82(5):596–606.
- Revell LJ. 2012. phytools: an R package for phylogenetic comparative biology (and other things). *Methods Ecol Evol*. 3(2):217–223.
- Reyes E, Sauquet H, Nadot S. 2016. Perianth symmetry changed at least 199 times in angiosperm evolution. *Taxon* 65(5):945–964.
- Reyes-Chin-Wo S, Wang Z, Yang X, Kozik A, Arikat S, Song C, Xia L, Froenicke L, Lavelle DO, Truco M-J, et al. 2017. Genome assembly with in vitro proximity ligation data and whole-genome triplication in lettuce. *Nat Commun*. 8(1):14953.
- Rose JP, Kleist TJ, Löfstrand SD, Drew BT, Schönenberger J, Sytsma KJ. 2018. Phylogeny, historical biogeography, and diversification of angiosperm order Ericales suggest ancient Neotropical and East Asian connections. *Mol Phylog Evol*. 122:59–79.

- Ruprecht C, Lohaus R, Vanneste K, Mutwil M, Nikoloski Z, Van de Peer Y, Persson S. 2017. Revisiting ancestral polyploidy in plants. *Sci Adv* 3(7):e1603195.
- Sanderson MJ. 2003. r8s: inferring absolute rates of molecular evolution and divergence times in the absence of a molecular clock. *Bioinformatics* 19(2):301–302.
- Sauquet H, Magallón S. 2018. Key questions and challenges in angiosperm macroevolution. *New Phytol* 219(4):1170–1187.
- Sauquet H, von Balthazar M, Magallón S, Doyle JA, Endress PK, Bailes EJ, Barroso de Morais E, Bull-Hereñu K, Carrive L, Chartier M, et al. 2017. The ancestral flower of angiosperms and its early diversification. *Nat Commun* 8(1):16047.
- Schäferhoff B, Fleischmann A, Fischer E, Albach DC, Borsch T, Heubl G, Kai FM. 2010. Towards resolving Lamiales relationships: insights from rapidly evolving chloroplast sequences. *BMC Evol Biol* 10(1):352.
- Schneider AC, Braukmann T, Banerjee A, Stefanović S. 2018. Convergent plastome evolution and gene loss in holoparasitic Lennoaceae. *Genome Biol Evol* 10(10):2663–2670.
- Schönenberger J. 2009. Comparative floral structure and systematics of Fouquieriaceae and Polemoniaceae (Ericales). *Int J Plant Sci* 170(9):1132–1167.
- Schranz ME, Mohammadin S, Edger PP. 2012. Ancient whole genome duplications, novelty and diversification: the WGD radiation lag-time model. *Curr Opin Plant Biol* 15(2):147–153.
- Seymour GB, Østergaard L, Chapman NH, Knapp S, Martin C. 2013. Fruit development and ripening. *Annu Rev Plant Biol* 64(1):219–241.
- Shi T, Huang H, Barker MS. 2010. Ancient genome duplications during the evolution of kiwifruit (*Actinidia*) and related Ericales. *Ann Bot* 106(3):497–504.
- Simpson MG. 2010. Plant systematics. 2nd ed. Burlington (VT): Academic Press.
- Smith SA, Beaulieu JM. 2009. Life history influences rates of climatic niche evolution in flowering plants. *Proc Biol Sci* 276(1677):4345–4352.
- Smith SA, Donoghue MJ. 2008. Rates of molecular evolution are linked to life history in flowering plants. *Science* 322(5898):86–89.
- Sollars ESA, Harper AL, Kelly LJ, Sambles CM, Ramirez-Gonzalez RH, Swarbreck D, Kaithakottil G, Cooper ED, Uauy C, Havlickova L, et al. 2017. Genome sequence and genetic diversity of European ash trees. *Nature* 541(7636):212–216.
- Soltis DE, Gitzendanner MA, Soltis PS. 2007. A 567-taxon data set for angiosperms: the challenges posed by Bayesian analyses of large data sets. *Int J Plant Sci* 168(2):137–157.
- Soltis DE, Smith SA, Cellinese N, Wurdack KJ, Tank DC, Brockington SF, Refulio-Rodriguez NF, Walker JB, Moore MJ, Carlswald BS, et al. 2011. Angiosperm phylogeny: 17 genes, 640 taxa. *Am J Bot* 98(4):704–730.
- Soltis PS, Folk RA, Soltis DE. 2019. Darwin review: angiosperm phylogeny and evolutionary radiations. *Proc Biol Sci* 286(1899):20190099.
- Soltis PS, Soltis DE. 2016. Ancient WGD events as drivers of key innovations in angiosperms. *Curr Opin Plant Biol* 30:159–165.
- Springer MS, DeBry RW, Douady C, Amrine HM, Madsen O, de Jong WW, Stanhope MJ. 2001. Mitochondrial versus nuclear gene sequences in deep-level mammalian phylogeny reconstruction. *Mol Biol Evol* 18(2):132–143.
- Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30(9):1312–1313.
- Stevens PF. 2001. Angiosperm Phylogeny Website. Version 14, July 2017 [and more or less continuously updated since]. St. Louis: Missouri Botanical Garden. Available from: href="http://www.mobot.org/mobot/research/apweb/welcome.html">http://www.mobot.org/mobot/research/apweb/welcome.html.
- Stull GW, Duno de Stefano R, Soltis DE, Soltis PS. 2015. Resolving basal lamiid phylogeny and the circumscription of Icacinaeaceae with a plastome-scale data set. *Am J Bot* 102(11):1794–1813.
- Stull GW, Schori M, Soltis DE, Soltis PS. 2018. Character evolution and missing (morphological) data across Asteridae. *Am J Bot* 105(3):470–479.
- Stull GW, Soltis PS, Soltis DE, Gitzendanner MA, Smith SA. 2020. Nuclear phylogenomic analyses of asterids conflict with plastome trees and support novel relationships among major lineages. *Am J Bot* 107(5):790–805.
- Takhtajan A. 2009. Flowering plants. 2nd ed. New York: Springer.
- Tang C-Y, Li S, Wang Y-T, Wang X. 2020. Comparative genome/transcriptome analysis probes Boraginales' phylogenetic position, WGDs in Boraginales, and key enzyme genes in the alkannin/shikonin core pathway. *Mol Ecol Resour* 20(1):228–241.
- Tang H, Bowers JE, Wang X, Paterson AH. 2010. Angiosperm genome comparisons reveal early polyploidy in the monocot lineage. *Proc Natl Acad Sci U S A* 107(1):472–477.
- Tank DC, Donoghue MJ. 2010. Phylogeny and phylogenetic nomenclature of the Campanulidae based on an expanded sample of genes and taxa. *Syst Bot* 35(2):425–441.
- Tomato Genome Consortium. 2012. The tomato genome sequence provides insights into fleshy fruit evolution. *Nature* 485(7400):635–641.
- Unver T, Wu Z, Sterck L, Turktas M, Lohaus R, Li Z, Yang M, He L, Deng T, Escalante FJ, et al. 2017. Genome of wild olive and the evolution of oil biosynthesis. *Proc Natl Acad Sci U S A* 114(44):E9413–E9422.
- Van de Peer Y, Mizrachi E, Marchal K. 2017. The evolutionary significance of polyploidy. *Nat Rev Genet* 18(7):411–424.
- Vanneste K, Van de Peer Y, Maere S. 2013. Inference of genome duplications from age distributions revisited. *Mol Biol Evol* 30(1):177–190.
- Veizer J, Godderis Y, François LM. 2000. Evidence for decoupling of atmospheric CO<sub>2</sub> and global climate during the Phanerozoic eon. *Nature* 408(6813):698–701.
- Wang J-P, Yu J-G, Li J, Sun P-C, Wang L, Yuan J-Q, Meng F-B, Sun S-R, Li Y-X, Lei T-Y, et al. 2018. Two likely auto-tetraploidization events shaped kiwifruit genome and contributed to establishment of the Actinidiaceae family. *iScience* 7:230–240.
- Wang L, Yu S, Tong C, Zhao Y, Liu Y, Song C, Zhang Y, Zhang X, Wang Y, Hua W, et al. 2014. Genome sequencing of the high oil crop sesame provides insight into oil biosynthesis. *Genome Biol* 15(2):R39.
- Wang T, Liang L, Xue Y, Jia P-F, Chen W, Zhang M-X, Wang Y-C, Li H-J, Yang W-C. 2016. A receptor heteromer mediates the male perception of female attractants in plants. *Nature* 531(7593):241–244.
- Weigend M, Luebert F, Gottschling M, Couvreur TLP, Hilger HH, Miller JS. 2014. From capsules to nutlets-phylogenetic relationships in the Boraginales. *Cladistics* 30(5):508–518.
- Wickett NJ, Mirarab S, Nguyen N, Warnow T, Carpenter E, Matasci N, Ayyampalayam S, Barker MS, Burleigh JG, Gitzendanner MA, et al. 2014. Phylotranscriptomic analysis of the origin and early diversification of land plants. *Proc Natl Acad Sci U S A* 111(45):E4859–E4868.
- Wikström N, Kainulainen K, Razafimandimbison SG, Smedmark JEE, Bremer B. 2015. A revised time tree of the asterids: establishing a temporal framework for evolutionary studies of the coffee family (Rubiaceae). *PLoS One* 10(5):e0126690.
- Winkworth RC, Lundberg J, Donoghue MJ. 2008. Toward a resolution of Campanulid phylogeny, with special reference to the placement of Dipsacales. *Taxon* 57:53–65.
- Wortley AH, Rudall PJ, Harris DJ, Scotland RW. 2005. How much data are needed to resolve a difficult phylogeny?: case study in Lamiales. *Syst Biol* 54(5):697–709.
- Wuyun T-N, Wang L, Liu H, Wang X, Zhang L, Bennetzen JL, Li T, Yang L, Liu P, Du L, et al. 2018. The hardy rubber tree genome provides insights into the evolution of polyisoprene biosynthesis. *Mol Plant* 11(3):429–442.
- Xia E-H, Zhang H-B, Sheng J, Li K, Zhang Q-J, Kim C, Zhang Y, Liu Y, Zhu T, Li W, et al. 2017. The tea tree genome provides insights into tea flavor and independent evolution of caffeine biosynthesis. *Mol Plant* 10(6):866–877.
- Xiang Y, Huang C-H, Hu Y, Wen J, Li S, Yi T, Chen H, Xiang J, Ma H. 2017. Evolution of Rosaceae fruit types based on nuclear phylogeny in the context of geological times and genome duplication. *Mol Biol Evol* 34(2):262–281.
- Xu S, Brockmüller T, Navarro-Quezada A, Kuhl H, Gase K, Ling Z, Zhou W, Kreitzer C, Stanke M, Tang H, et al. 2017. Wild tobacco genomes reveal the evolution of nicotine biosynthesis. *Proc Natl Acad Sci U S A* 114(23):6133–6138.
- Yang Y, Moore MJ, Brockington SF, Soltis DE, Gane Ka-Shu W, Carpenter EJ, Yong Z, Li C, Zhixiang Y, Yinlong X, et al. 2015. Dissecting

- molecular evolution in the highly diverse plant clade Caryophyllales using transcriptome sequencing. *Mol Biol Evol.* 32(8):2001–2014.
- Yu G, Smith DK, Zhu H, Guan Y, Lam TT-Y. 2017. ggtree: an R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods Ecol Evol.* 8(1):28–36.
- Yuan Z, Fang Y, Zhang T, Fei Z, Han F, Liu C, Liu M, Xiao W, Zhang W, Wu S, et al. 2018. The pomegranate (*Punica granatum* L.) genome provides insights into fruit quality and ovule developmental biology. *Plant Biotechnol J.* 16(7):1363–1374.
- Zachos JC, Dickens GR, Zeebe RE. 2008. An early Cenozoic perspective on greenhouse warming and carbon-cycle dynamics. *Nature.* 451(7176):279–283.
- Zanne AE, Tank DC, Cornwell WK, Eastman JM, Smith SA, FitzJohn RG, McGlinn DJ, O'Meara BC, Moles AT, Reich PB, et al. 2014. Three keys to the radiation of angiosperms into freezing environments. *Nature* 506(7486):89–92.
- Zeng L, Zhang N, Zhang Q, Endress PK, Huang J, Ma H. 2017. Resolution of deep eudicot phylogeny and their temporal diversification using nuclear genes from transcriptomic and genomic datasets. *New Phytol.* 214(3):1338–1354.
- Zeng L, Zhang Q, Sun R, Kong H, Zhang N, Ma H. 2014. Resolution of deep angiosperm phylogeny using conserved nuclear genes and estimates of early divergence times. *Nat Commun.* 5(1):4956.
- Zhang C, Rabiee M, Sayyari E, Mirarab S. 2018. ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinf.* 19(S6):153.
- Zhang D, Li W, Xia E-h, Zhang Q-J, Liu Y, Zhang Y, Tong Y, Zhao Y, Niu Y-C, Xu J-h, et al. 2017. The medicinal herb *Panax notoginseng* genome provides insights into ginsenoside biosynthesis and genome evolution. *Mol Plant* 10(6):903–907.
- Zhang N, Zeng L, Shan H, Ma H. 2012. Highly conserved low-copy nuclear genes as effective markers for phylogenetic analyses in angiosperms. *New Phytol.* 195(4):923–937.
- Zhao L, Li X, Zhang N, Zhang S-D, Yi T-S, Ma H, Guo Z-H, Li D-Z. 2016. Phylogenomic analyses of large-scale nuclear genes provide new insights into the evolutionary relationships within the rosids. *Mol Phylogen Evol.* 105:166–176.
- Zwaenepoel A, Van de Peer Y. 2019. Inference of ancient whole-genome duplications and the evolution of gene duplication and loss rates. *Mol Biol Evol.* 36(7):1384–1404.