RESEARCH ARTICLE

Short Title: Stull et al.—Nuclear phylogenomics of *Asteridae*

Nuclear phylogenomic analyses of asterids conflict with plastome trees and support novel relationships among major lineages

Gregory W. Stull[1,2,7], Pamela S. Soltis[3,4], Douglas E. Soltis[3,4,5], Matthew A. Gitzendanner[5], and Stephen A. Smith[6]

Manuscript received 26 January 2019; revision accepted 25 February 2020.

[1] Kunming Institute of Botany, Chinese Academy of Sciences, Kunming, Yunnan 650204, China

[2] Department of Botany, Smithsonian Institution, Washington, D.C. 20013, USA

[3] Florida Museum of Natural History, University of Florida, Gainesville, Florida 32611, USA

[4] Biodiversity Institute, University of Florida, Gainesville, Florida 32611, USA

[5] Department of Biology, University of Florida, Gainesville, Florida 32611, USA

[6] Department of Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, Michigan 48109, USA

[7] Author for correspondence (e-mail: gwstull@gmail.com)

Citation: Stull, G. W., P. S. Soltis, D. E. Soltis, M. A. Gitzendanner, and S. A. Smith. 2020. Nuclear phylogenomic analyses of asterids conflict with plastome trees and support novel relationships among major lineages. *American Journal of Botany* 107(5): XXXX.

DOI: XXXX

PREMISE: Discordance between nuclear and organellar phylogenies (cytonuclear discordance) is a well-documented phenomenon at shallow evolutionary levels but has been poorly investigated at deep levels of plant phylogeny. Determining the extent of cytonuclear discordance across major plant lineages is essential not only for elucidating evolutionary processes, but also for evaluating the currently used framework of plant phylogeny, which is largely based on the plastid genome.

METHODS: We present a phylogenomic examination of a major angiosperm clade (*Asteridae*) based on sequence data from the nuclear, plastid, and mitochondrial genomes as a means of evaluating currently accepted relationships inferred from the plastome and exploring potential sources of genomic conflict in this group.

RESULTS: We recovered at least five instances of well-supported cytonuclear discordance concerning the placements of major asterid lineages (i.e., Ericales, Oncothecaceae, Aquifoliales, *Cassinopsis*, and Icacinaceae). We attribute this conflict to a combination of incomplete lineage sorting and hybridization, the latter supported in part by previously inferred whole-genome duplications.

CONCLUSIONS: Our results challenge several long-standing hypotheses of asterid relationships and have implications for morphological character evolution and for the importance of ancient whole-genome duplications in early asterid evolution. These findings also highlight the value of reevaluating broad-scale angiosperm and green-plant phylogeny with nuclear genomic data.

KEY WORDS: Aquifoliales; asterids; cytonuclear discordance; Ericales; incomplete lineage sorting; phylogenomics; whole-genome duplication.

Discordance between nuclear and organellar phylogenies (cytonuclear discordance) can arise from several evolutionary processes, including hybridization, incomplete lineage sorting (ILS), horizontal gene transfer, and gene duplication and loss (Doyle, 1992; Maddison, 1997; Galtier and Daubin, 2008; Soltis and Soltis, 2009; Smith et al., 2015). In plants, hybridization in particular—often accompanied by polyploidy, or whole-genome duplication (WGD)—has been emphasized as an important evolutionary process (e.g., Stebbins, 1950; Soltis and Soltis, 2009) and a major source of observed cytonuclear discordance due to chloroplast capture (e.g., Rieseberg and Soltis, 1991; Soltis and Kuzoff, 1995; Linder and Rieseberg, 2004). While such instances of discordance are typically observed at shallow phylogenetic levels (e.g., Huang et al., 2014), recent studies have shown that ancient hybridization events can produce lasting signatures of cytonuclear discordance (Folk et al., 2017; García et al., 2017; Morales-Briones et al., 2018), even among major lineages of angiosperms (e.g., members of the rosid COM clade; Sun et al.,

2015). We might expect ILS, as well, to result in persistent cytonuclear discordance in some cases, but this possibility has been poorly explored across the phylogeny of green plants.

Our limited understanding of deep cytonuclear discordance stems, in part, from the historical reliance on the chloroplast genome for plant phylogenetic studies. For example, with the exception of a few nuclear and mitochondrial regions (e.g., Qui et al., 1999, 2010; Soltis et al., 2000) and the recently published 1KP phylogeny (One Thousand Plant Transcriptomes Initiative, 2019), our current framework of angiosperm phylogeny (e.g., APG IV, 2016) has primarily been informed by analyses of plastid genes and genomes (e.g., Jansen et al. 2007; Moore et al., 2007, 2010, 2011; Soltis et al., 2011; Ruhfel et al., 2014; Gitzendanner et al., 2018; Li et al., 2019). While the number of plant phylogenetic studies employing nuclear genes is certainly increasing (e.g., Wickett et al., 2014; Yang et al., 2015; Zeng et al., 2017; Couvreur et al., 2019; Johnson et al., 2019), few have explicitly investigated instances of deep cytonuclear discordance between comparable plastid and nuclear genomic data sets using tools for dissecting genomic conflict (e.g., Smith et al., 2015). The evolutionary history of angiosperms includes numerous cases of rapid radiation, evident from both the fossil record (e.g., Friis et al., 2011) and phylogenetic diversification analyses (e.g., Magallón and Castillo, 2009; Magallón et al., 2015; Tank et al., 2015; Landis et al., 2018; Smith and Brown, 2018). As a consequence, we might expect genomic signatures of ILS and introgression to be relatively common at deep levels across flowering plant phylogeny (Soltis et al., 2019), but this has been largely unexplored.

Of course, the erosion of phylogenetic signal for deep nodes—due to saturation, for example—can render the resolution of deep relationships and associated evolutionary processes a considerable challenge (King and Rokas, 2017; Smith et al., 2019), especially in cases of rapid radiation (Whitfield and Lockhart, 2007; Parks et al., 2017). For some radiations, complete and confident resolution of all relationships and the identification of discrete evolutionary processes may be impossible, at least with current methods (Morales-Briones et al., 2019). Nevertheless, these types of investigations are necessary for a more nuanced understanding of the evolutionary history of angiosperms and of plants more broadly.

An important corollary of the points above is that the widely used framework of angiosperm phylogeny, based largely on the plastid genome (e.g., APG IV, 2016; Soltis et al., 2018a, b), may not reflect the true species tree and/or may be incomplete as a result of evolutionary processes leading to cytonuclear discordance. More broad-scale phylogenomic

studies of angiosperms and green plants in general, explicitly comparing results from both the nuclear and plastid genomes, are clearly needed to evaluate the plastid-based phylogenetic framework and to elucidate the extent of cytonuclear discordance at deep phylogenetic levels. Here, we present a phylogenomic investigation of the major angiosperm clade *Asteridae* (Cantino et al., 2007), based on sequence data from the nuclear, plastid, and mitochondrial genomes, as a means of evaluating relationships inferred from the plastome (e.g., APG IV, 2016) and exploring potential sources of phylogenomic conflict in this group.

The asterid clade (*Asteridae*), with ~80,000 species or ~25% of flowering plant diversity, represents one of the largest angiosperm radiations (Soltis et al., 2018a). This clade has been the subject of numerous phylogenetic studies and, to a large extent, most major relationships appear well resolved (e.g., Olmstead et al., 1993, 2000; Albach et al., 2001a, b; Bremer et al., 2002; Tank and Donoghue, 2010; Stull et al., 2015, 2018). Cornales and Ericales are consistently recovered as successively sister to *Gentianidae*, which in turn comprises two major clades, *Lamiidae* and *Campanulidae*, each with ~40,000 species (Soltis et al., 2018a). Within campanulids, Aquifoliales have been consistently recovered as sister to the remainder of the clade (Judd and Olmstead, 2004). Within lamiids, several species-poor lineages (Icacinales, Metteniusales, Garryales) form a grade of successive sisters to the large core lamiid clade (= *Lamianae*; see Stull et al., 2015), which comprises four species-rich orders (Boraginales, Gentianales, Lamiales, Solanales) and the phylogenetically isolated genus *Vahlia* (Refulio-Rodriguez and Olmstead, 2014). This framework, however, is almost entirely based on data derived from the plastid genome. Significantly, several recent studies of angiosperm phylogeny based on the nuclear genome (Zeng et al., 2017; One Thousand Plant Transcriptomes Initiative, 2019) recovered notably different placements for several major asterid lineages than those reported on the basis of plastid sequences. A more comprehensive nuclear genomic perspective of asterid phylogeny is therefore urgently needed to evaluate the reliability of broadly accepted plastid-based relationships.

To examine asterid relationships in a phylogenomic context, we assembled a large data set for 129 species comprising numerous loci from the nuclear (307) and plastid (77) genomes, using a combination of publicly available data (1KP: http://www.onekp.com; GenBank: https://www.ncbi.nlm.nih.gov/genbank/) and newly generated sequences from target enrichment (hyb-seq). A data set of 35 mitochondrial genes across 150 species (representing the same major

lineages, but different species in many cases) was also assembled from GenBank to provide a more comprehensive perspective on inter-genomic conflict. We used both species-tree and concatenation methods to infer phylogenetic relationships from each genomic data set, as well as methods for examining conflict and concordance among gene trees from the nuclear genome (Smith et al., 2015) and for inferring ancient reticulation events (e.g., Than et al., 2008; Folk et al., 2017; Wen et al., 2018). The results of these analyses provide important new insights on major relationships and patterns of morphological character evolution in *Asteridae* and establish a critical framework for understanding genomic conflict in light of previously inferred ancient WGDs. More broadly, our results underscore the importance of explicitly investigating phylogenomic discordance in major plant clades, even those whose relationships are thought to be largely settled.

<h1>MATERIALS AND METHODS

<h2>Sampling and sequencing

We sampled 129 species (Appendix S1) for both the nuclear and plastid data sets, which ultimately included 307 and 77 loci, respectively. All major asterid groups were represented, with an emphasis on the lineages forming the basal lamiid nodes (Stull et al., 2015), as well as numerous outgroups spanning non-asterid eudicots.

The nuclear data set was constructed using a combination of publicly available transcriptome data (100 species, obtained from the 1KP project: Matasci et al., 2014) and newly generated sequences for key asterid lineages (namely, Icacinaceae, Metteniusaceae) obtained through target enrichment and high-throughput sequencing (29 species, 104 nuclear loci; Appendix S1). The transcriptomes had been assembled previously (Matasci et al., 2014) using SOAPdenovo-Trans (Xie et al., 2014). We used hyb-seq (rather than transcriptome sequencing) to complete our sampling because most of the additional species needed are tropical and practically inaccessible for fresh material, necessitating an approach, such as hyb-seq, that is amenable to herbarium specimens.

New hyb-seq data were generated as follows. DNAs were extracted using the CTAB method (Doyle and Doyle, 1987; Cullings, 1992), and genomic libraries were built following Stull et al. (2013) or by RAPiD Genomics (http://rapid-genomics.com/home; see Appendix S1). Nuclear loci for probe design and target enrichment were selected using MarkerMiner version

1.0 (Chamala et al., 2015), based on analysis of five input transcriptomes (*Aucuba japonica*, *Eucommia ulmoides*, *Iodes vitiginea*, *Oncotheca balansae*, *Pyrenacantha malvifolia*) related to the taxa intended for enrichment and sequencing; these transcriptomes were also obtained from the 1KP project (http://www.onekp.com/samples/list.php). We then performed target enrichment of the 104 selected nuclear loci using a custom-designed MYbaits kit from Arbor Biosciences (Ann Arbor, MI, USA), following MYbaits Manual version 3 (https://arborbiosci.com/wp-content/uploads/2017/10/MYbaits-manual-v3.pdf). The custom MYbaits kit included 120mer baits with flexible (45 bp) tiling density (~2.67×) designed across each included sequence (i.e., the five species listed above) for each locus (104), totaling 10,001 baits. After enrichment, the samples were sequenced on the Illumina NextSeq 500 or MiSeq platform (2 × 150 bp) at the University of Florida Interdisciplinary Center for Biotechnology Research (Appendix S1).

The plastid data set (Appendix S1) was similarly assembled using publicly available data and newly generated sequence data. Most samples (100) represented the same 1KP accessions as above, but the data were obtained directly from Gitzendanner et al. (2018), who reconstructed green-plant relationships using 78 protein-coding plastid genes. The remaining 29 samples were obtained from Stull et al. (2015) or newly sequenced for this study (Appendix S1). One gene (*petG*) was ultimately excluded due to its low representation throughout our sampling, resulting in 77 genes in the plastid data set.

Although our primary goal was to compare phylogenetic results between the nuclear and plastid genomes based on a comparable sampling of species, we also generated a mitochondrial data set using publicly available sequences from GenBank (https://www.ncbi.nlm.nih.gov/genbank/), reasoning that the mitochondrial tree may provide useful additional evidence for interpreting instances of nuclear-plastid incongruence. However, given the data available on GenBank, it was not possible to sample the same species from the nuclear/plastid data sets for mitochondrial sequences. We therefore attempted to represent the same major lineages (e.g., Cornales, Ericales, Icacinaceae, Aquifoliales, Garryales) to the greatest extent possible, while maximizing genic representation for the species included. Our mitochondrial data set included 150 species and 35 genes (Appendix S2). Most species were represented by only two or three mitochondrial genes, but the more "complete" species (i.e., 40 species with >20 genes) are reasonably dispersed phylogenetically, helping anchor the placements of the more poorly sampled species.

<h2>**New sequence assembly and alignment**

The newly generated plastid and nuclear sequence data were assembled as follows. Remnant adapter sequences were removed from the demultiplexed reads using Cutadapt version 1.5 (Martin, 2011); low-quality nucleotides were removed from the reads using Sickle version 1.33 (Joshi, 2011). HybPiper version 1.0 (Johnson et al., 2016) was then used to assemble the 104 nuclear and 77 plastid loci. Although HybPiper includes an optional step to assemble (partial to complete) intron sequences flanking the exons, we opted to assemble exons only, given that the transcriptomic references used for probe design and assembly lack introns. The assembled sequences were then combined with their corresponding gene regions from publicly available data sets for subsequent alignment. In the case of the nuclear data set, however, the transcriptome data were first subjected to a preliminary ortholog identification step before being combined with the corresponding hyb-seq data. Because the hyb-seq loci were selected from the curated gene set implemented in MarkerMiner, which was also used to obtain orthologs from the transcriptome sequences (outlined below), this readily facilitated the combinability of these data sets. Once combined, the nuclear data set was then subjected to additional filtering steps to further ensure orthology (outlined below).

Each gene region (for each of the genomic data sets: nuclear, plastid, and mitochondrial) was aligned individually using MAFFT version 7.220 (Katoh and Standley, 2013). Following alignment, the program "pxclsq" in Phyx (Brown et al., 2017) was used to remove columns from each alignment with >50% missing data. The final alignments are available from the Dryad Digital Repository: https://doi.org/10.5061/dryad.h70rxwdfq (Stull et al., 2020).

<h2>**Ortholog identification**

We used MarkerMiner to identify putative single-copy genes across the assembled transcriptomes. MarkerMiner conducts a reciprocal BLAST search across the included transcriptomes to identify putative orthologs, which are then filtered for single-copy loci based on a curated set of 2809 genes that have been shown to be mostly single-copy across angiosperms (De Smet et al., 2013). This approach resulted in 2119 putative single-copy genes, which were subjected to several filtering steps before the final phylogenetic analyses. However, prior to filtering, we combined the sequences obtained from target enrichment with their

corresponding genes in the transcriptome gene set (because MarkerMiner was used for loci selection in the hyb-seq data set, these two data sets were readily combined based on their reference gene IDs). We then filtered the nuclear loci such that only genes including at least 30 species were retained. The remaining 604 genes were then aligned individually, as described above, and used to generate gene trees in RAxML version 8.2.12, using the GTRGAMMA molecular model and including a search for the best-scoring maximum likelihood (ML) tree as well as 200 bootstrap replicates (Stamatakis, 2014). The resulting gene trees were used for two subsequent filtering steps to remove genes with potential paralogy issues: (1) gene trees with polyphyly of well-established major clades (i.e., *Asteridae* and *Rosidae*, *sensu* Soltis et al., 2018a) were excluded, as were (2) gene trees with extreme root-to-tip variation or tree lengths (calculated using the Phyx program "pxstr"; Brown et al., 2017). Following these filtering steps, 307 loci/gene trees remained for subsequent phylogenetic analyses.

<h2>**Phylogenetic analyses**

We inferred phylogenetic trees for each genomic region (nuclear, plastid, mitochondrial) separately. For the plastid and mitochondrial genomes, we concatenated each gene set (77 loci for the plastome, 35 loci for the mitochondrial genome) using the phyx program "pxcat" (Brown et al., 2017) for subsequent phylogenetic analysis in RAxML, using the GTRGAMMA molecular model partitioned by gene region and including a search for the best-scoring ML tree and 200 bootstrap replicates. Because organellar genomic regions are typically analyzed in concatenation (i.e., treated as a single locus), we followed that approach here. However, recent studies have shown that plastid genomes are not free of gene-tree conflict (e.g., Walker et al. 2019), suggesting that concatenation approaches may be inappropriate for plastomes, at least in some cases (Gonçalves et al., 2019). In light of this, multispecies coalescent (MSC) methods have been suggested as an alternative framework for plastome analyses (Gonçalves et al., 2019), but the appropriateness of this type of approach for plastid data, both theoretically and practically, has not been thoroughly considered. In addition to potential model violations (e.g., linkage, selection; Edwards et al., 2016), gene-tree estimation error—which is likely to be considerable with plastid genes, given that many of them are short and largely uninformative (Walker et al., 2019)—could potentially undermine analyses of the plastome, given that many species tree methods are sensitive to such errors (Chou et al., 2015).

With these caveats in mind, we also analyzed our plastome data set with ASTRAL version 5.6.3 (Zhang et al., 2018), a "species tree" method that infers the maximum quartet support species tree (MQSST), given a set of input gene trees. While ASTRAL does not explicitly model the MSC, it nevertheless has been shown to be statistically consistent in species tree inference from gene trees generated by the MSC process (Mirarab et al., 2014). We reasoned that using both of these approaches (concatenation and species tree) might help highlight weakly supported (or inaccurate) plastid relationships that may not actually represent true conflicts with the nuclear genome, even if both methods, in a sense, constitute model violations (Edwards et al., 2016). Plastid gene trees for the ASTRAL analysis were inferred using RAxML with the GTRGAMMA model and 200 bootstrap replicates. Low-supported branches (<10% bootstrap support) in the gene trees were collapsed prior to analysis in ASTRAL, as this has been shown to improve species tree inference (Zhang et al., 2017). The ASTRAL analysis, based on 77 plastid loci, was implemented with the default settings, resulting in an estimated species tree topology with branch lengths and local posterior probabilities (LPPs) as branch support values.

For the nuclear genome, as above, we inferred phylogenies using both concatenation and species tree methods. For the concatenation analysis, we used "pxcat" to concatenate the alignments of the 307 genes that passed all filtering steps (described above), and this combined set of genes was then analyzed in RAxML, using the same parameters and model partitioning scheme noted above. ASTRAL was used to infer an MQSST, with the 307 filtered gene trees from above as the input. As above, low-supported branches (<10% bootstrap support) were collapsed prior to the analysis, which employed the default settings.

<h2>**Conflict analyses**

We used PhyParts (Smith et al., 2015) to examine patterns of gene-tree concordance and conflict within the nuclear genome and to reveal subsets of the nuclear genome supporting alternative relationships in the plastid topology. We performed these analyses because the characterization of patterns of conflict is an important step for identifying areas of a phylogeny that deserve more detailed attention and, potentially, additional analyses (e.g., for investigation of ancient reticulation). Operationally, PhyParts maps a given set of gene trees on an input species tree to determine the number of gene trees that are concordant, conflicting, or uninformative with respect to each node in the species tree. Among the conflicting genes, it also determines the

number of gene trees supporting a dominant alternative relationship. We performed two PhyParts analyses, each with a bootstrap support (BS) threshold of 70, meaning that gene-tree branches/nodes with <70% BS were considered uninformative. This BS value has long been considered a baseline for strong support (Hillis and Bull, 1993), although this notion has since been rightfully challenged (e.g., Soltis and Soltis, 2003). Nevertheless, it is a useful, albeit somewhat arbitrary, value for filtering out poorly supported (and possibly spurious) branches, thus alleviating noise in the results of the conflict analysis (Smith et al., 2015).

For the first analysis, we mapped the 307 nuclear gene trees onto the species-tree phylogeny (which, as discussed below, is largely consistent with the concatenated nuclear tree in major relationships) to characterize gene-tree conflict against the prevailing phylogenetic signal from the nuclear genome. In the second analysis, we mapped the 307 nuclear gene trees onto the inferred plastome topology to determine if subsets of the nuclear genome support the plastome-based topology in instances of plastid-nuclear conflict. We were unable to map the nuclear gene trees onto the mitochondrial phylogeny using PhyParts because of numerous differences in species sampling; instead, we visually inspected the mitochondrial tree to locate well-supported major relationships in conflict with the nuclear and plastid topologies. The output of the PhyParts analyses was visualized using a Python script by M. Johnson (https://github.com/mossmatters/phyloscripts/tree/master/phypartspiecharts). We also visualized differences between the nuclear species tree and the plastid phylogeny using the "cophylo" function in Phytools (Revell, 2012); this function takes two input phylogenies with identical sampling and makes a plot juxtaposing them, with lines connecting each species to its match in the alternative phylogeny, thus revealing differences in placement.

<h2>**Coalescent simulations, gene-tree dating, and network analyses**

One major instance of incongruence between the nuclear and plastid genomes (concerning the placement of Ericales) was revealed in the conflict analyses and was explored further to determine whether the observed conflict was likely due to ancient chloroplast capture or ILS, although in many cases these are very difficult to tease apart (Pease and Hahn, 2015). Most of the informative nuclear genes (i.e., >70% BS for the relevant node) place Ericales sister to Cornales (61 genes), while a subset of the nuclear genome (26 genes) and the plastid genome place Ericales sister to *Gentianidae*.

To distinguish between chloroplast capture and ILS, following several studies (Folk et al., 2017; García et al., 2017; Morales-Briones et al., 2018), we simulated 1000 organellar trees under the coalescent using the program DendroPy (Sukumaran and Holder, 2010), with the inferred species tree acting as the guide tree. The guide tree (which was inferred from the nuclear data set) was scaled by a factor of four to approximate the branch lengths expected from organellar inheritance; this was done because the effective population size of the plastome is generally expected to be one-fourth that of the nuclear genome given the assumptions of equal sex ratios, haploidy (homoplasmic), and uniparental inheritance (McCauley, 1994). This approach allowed us to evaluate the plausibility of plastid ILS with respect to this particular species tree. The rationale for this approach is that, if the branch subtending the species-tree relationship (in this case, Cornales + Ericales) is very short, we would expect to see alternative relationships (e.g., Ericales + *Gentianidae*) with an appreciable frequency among the simulated trees due to ILS. However, if this branch is sufficiently long, we should not observe ILS from the simulations, suggesting that the actual observed plastid topology (Ericales + *Gentianidae*) is due not to ILS, but rather to a later reticulation event (resulting in chloroplast capture). The clade frequencies of the simulated gene trees (with respect to the main alternative topologies: Cornales + Ericales vs. Ericales + *Gentianidae*) were calculated using PhyParts and visualized using the same Python script noted above.

To further explore this issue, we conducted dating analyses of the plastome phylogeny and the individual nuclear gene trees supporting the alternative placements of Ericales. We reasoned that, in the case of ILS, the timing of the Ericales + *Gentianidae* divergences (in the plastome tree and the gene trees supporting this relationship) should precede those of Ericales + Cornales (in the main/concordant gene set), reflecting that the gene trees supporting Ericales + *Gentianidae* are a consequence of incomplete sorting prior to the Ericales + Cornales split. Conversely, in the case of later chloroplast capture, the timing of the Ericales + *Gentianidae* divergences should occur after the Ericales + Cornales divergences.

We used treePL to date the plastome phylogeny and each individual gene tree, and for each we ran cross-validation tests prior to the final analysis to determine the optimal smoothing parameters (see configuration files for more details, available from the Dryad Digital Repository: https://doi.org/10.5061/dryad.h70rxwdfq [Stull et al., 2020]). Given the difficulty of consistently calibrating individual gene trees with differences in both relationships and included species, we

employed only two calibrations: (1) one for crown eudicots, with the minimum age set at 125.0 Ma (corresponding to the first appearance of tricolpate pollen; Doyle et al., 1977; Hughes and McDougall, 1990) and the maximum age set at 135.0 Ma (corresponding approximately to the oldest reliable fossil evidence of angiosperms; Trevisan, 1988; Brenner, 1996; Friis et al., 2011; Coiro et al., 2019); and (2) another for crown *Asteridae*, with the minimum age set at 115.75 Ma (based on the penalized likelihood minimum estimate from 100 ML bootstrap trees in Magallón et al., 2015) and the maximum age set at 125.0 Ma (the first fossil appearance of eudicots, the broader clade including *Asteridae*). This more minimal calibration scheme allowed us to consistently calibrate nearly all gene trees, while weighting the significance of the eudicot calibration, which is generally considered one of the better angiosperm calibration points in that it likely captures eudicots relatively close to their actual geologic origin (Friis et al., 2011).

Finally, we inferred species networks using PhyloNet version 3.8.0 (Than et al., 2008; Wen et al., 2018; Cao et al., 2019), which models both ILS and gene flow and permits the inference of reticulate nodes (hybridization events). For these analyses, we used the set of 307 nuclear gene trees, with the sampling reduced to a computationally tractable size (i.e., <30 taxa; Than et al., 2008; Wen et al., 2018). Specifically, we retained all species of Cornales and Ericales present in each gene tree, but reduced the outgroups to a single species and *Gentianidae* to three species (*Oncotheca balansae* and one species each from the lamiid and campanulid clades). The reduced gene trees are available on Dryad (https://doi.org/10.5061/dryad.h70rxwdfq). This sampling allowed us to specifically explore the processes responsible for the conflicting placements of Ericales. Species networks (with three, two, and one maximum reticulations) were inferred under maximum pseudo-likelihood, with branches including <70% BS collapsed, and five optimal networks returned for each analysis. We additionally inferred a strictly bifurcating species tree as a means of testing whether the best network (i.e., the one with the highest log probability) or a strictly bifurcating tree better fits the gene trees. The command CalGTProb was used to compute the likelihood scores of the best network and the bifurcating phylogeny (inferred using maximum pseudo-likelihood) given the set of gene trees, and these were then compared using a likelihood-ratio test.

<h1>RESULTS

<h2>Nuclear relationships

The nuclear phylogenomic analyses, using both concatenation and species-tree methods, recovered largely congruent relationships among major lineages of *Asteridae* (Figs. 1 and 2). Cornales and Ericales were recovered as sister with maximal support (Figs. 1 and 2), and together these were sister to the remaining asterids (= *Gentianidae*). *Oncotheca* (of the monogeneric Oncothecaceae) was recovered with maximal support as sister to the remainder of *Gentianidae*, which in turn was divided into two major clades corresponding roughly to the campanulids and lamiids (*sensu* APG IV, 2016). However, Aquifoliales (*Ilex* spp. and *Helwingia*) were recovered within the lamiid clade with maximal support, positioned either as sister to Garryales (concatenated phylogeny; Fig. 1 and Appendix S4) or along the backbone (species tree; Fig 2 and Appendix S5); Aquifoliales have been conventionally considered a member of the campanulid clade, sister to all other campanulid orders (e.g., Tank and Donoghue, 2010). Icacinaceae *sensu* Stull et al. (2015) were recovered as monophyletic with the exception of *Cassinopsis*, which was placed sister to Metteniusaceae with strong support (Figs. 1 and 2). Icacinaceae (minus *Cassinopsis*) were placed sister to the core lamiids (LPP = 1.0, BS = 100). Within the core lamiids, both analyses recovered Lamiales and Solanales as successively sister to Boraginales + Gentianales. Metteniusaceae + *Cassinopsis* were placed sister to all remaining lamiids (Figs. 1 and 2). Within the campanulids, relationships differed slightly between the concatenation and species-tree analyses. However, two major clades were recovered, one with Pennantiaceae, *Escallonia* (Escalloniales), and members of Dipsacales (*Viburnum* and *Lonicera*), and the other with Apiales and Asterales (Figs. 1 and 2).

<h2>**Phylogenomic conflict**

Regarding major asterid relationships, the results from the nuclear analyses (Figs. 1 and 2) show multiple instances (at least five) of well-supported conflict with those recovered from the plastid genome (Fig. 3 and Appendices S6, S7, S8), the latter of which are largely consistent with previous plastid and large-scale analyses of asterids (e.g., Stull et al., 2015). Slight differences in some relationships were observed between the concatenated and ASTRAL analyses of the plastid genes (e.g., the positions of *Oncotheca* and Garryales; Appendix S8), as described below, but these differences are largely confined to areas of poor support, and results from both plastid analyses differ consistently in important ways from those of the nuclear genome.

Results based on the mitochondrial genome, on the other hand, show areas of agreement and conflict with both the nuclear and plastid genomes, but internal support for major relationships in the mitochondrial tree is generally low (Appendix S9). The plastome tree, for example, places Ericales sister to *Gentianidae* with maximal support (Fig. 3 and Appendix S8), while both the nuclear (Figs. 1 and 2) and mitochondrial (Appendix S9) trees recovered Cornales and Ericales as sister (mitochondrial support is very low [BS = 16], while the nuclear support is maximal). However, in the nuclear phylogeny, the branch subtending Cornales + Ericales exhibits considerable well-supported gene tree conflict (Fig. 2), with 61 gene trees supporting this topology while 32 gene trees support conflicting/alternative resolutions; of these conflicting gene trees, 26 support the plastid relationship, as revealed by the conflict analysis with the nuclear genes mapped on the plastid topology (Fig. 3). This nuclear-plastid conflict, with a subset of the nuclear genome supporting the plastid relationship, is suggestive of a possible ancient chloroplast capture event.

The plastid analyses (Fig. 3) place Oncothecaceae in the lamiids either sister to Icacinaceae (consistent with Stull et al., 2015) with strong support (BS = 94, concatenated analysis; Fig. 3) or unresolved within a clade including Garryales and the core lamiids (ASTRAL analysis; Appendix S8). The nuclear analyses place this family sister to the rest of *Gentianidae* (Fig. 1); no nuclear gene trees (with >70% BS) support the plastid topologies (Fig. 3). Although the mitochondrial data are consistent with plastid data in placing Oncothecaceae in the lamiids, none of the branches subtending this placement are well supported in the mitochondrial tree (Appendix S5). Additional instances of nuclear-plastid conflict, in which no or few nuclear genes support the plastid relationships, include the placements of Aquifoliales, Icacinaceae, and Metteniusaceae. Notably, the plastid trees place Aquifoliales in the campanulid clade, sister to the remaining orders with strong support (Fig. 3 and Appendix S8), while the nuclear genome places Aquifoliales with maximal support in the lamiids (Figs. 1 and 2); although only six nuclear gene trees strongly support Aquifoliales sister to Garryales, the nuclear gene trees in general are overwhelmingly concordant in placing Aquifoliales in the lamiid clade (Fig. 2). The mitochondrial genome also places Aquifoliales in the lamiids, albeit with weak support (BS = 28).

The nuclear placement of Icacinaceae (minus *Cassinopsis*) sister to the core lamiids, although well supported in the concatenation and species-tree analyses, shows notable gene-tree

conflict, with 29 gene trees in support of this relationship and 36 in conflict (Fig. 2). Within Icacinaceae, generic relationships are largely consistent with previous plastid analyses (e.g., Stull et al., 2015). The strongly supported (LPP = 1.0, BS = 98) placement of *Cassinopsis* sister to Metteniusaceae in the nuclear tree (Figs. 1 and 2) is at odds with the plastid placement of this genus as sister to the rest of Icacinaceae (BS = 91; Fig. 3 and Appendix S8). In the nuclear trees, the position of this clade as a whole (*Cassinopsis* + Metteniusaceae) is well to poorly supported as sister to the remainder of *Lamiidae* (LPP = 0.93, BS = 33; Fig. 1). The mitochondrial phylogeny shows notably different placements for Icacinaceae (sister to the remaining *Gentianidae*) and Garryales (sister to campanulids), but in both instances with weak support (Appendix S5); additionally, only two species of Icacinaceae (*Icacina mannii* and *Mappianthus iodoides*), and none of Metteniusaceae, were available in GenBank for inclusion in the mitochondrial analyses.

Within the core lamiids, there is extensive gene-tree conflict in the nuclear genome concerning the relationships among Boraginales, Gentianales, Lamiales, and Solanales (Fig. 2), as well as conflict among relationships inferred using the nuclear, plastid, and mitochondrial data sets. However, support for core lamiid relationships in the organellar phylogenies (Fig. 3 and Appendices S3, S5 S8) was generally low, in contrast to the more strongly supported nuclear relationships (Fig. 1).

<h2>Coalescent simulations, gene-tree dating, and network analyses

When organellar phylogenies were simulated on the scaled species tree under the coalescent, considerable gene-tree conflict was recovered at many of the nodes in the species tree conflicting with the actual plastid topology, suggesting that ILS is a likely explanation of at least some of the observed conflict (Fig. 4). In particular, for Cornales + Ericales, 759 of the simulated gene trees (out of 1000) were concordant with this relationship while 241 were in conflict, with roughly half of these conflicting genes (124) concordant with the plastid topology (Fig. 4). This suggests that the rapidity of the speciation events separating the lineages *Gentianidae* + (Cornales + Ericales) created an opportunity for extensive ILS. If the branches separating the speciation events were longer, little or no conflict should be observed in the simulated organellar trees (Folk et al., 2017; Morales-Briones et al., 2018). Similarly, conflict among the simulated organellar

trees is also appreciable along the backbone of *Lamiidae* (Fig. 4), where multiple conflicting placements were observed (Figs. 1 and 2), suggesting that ILS is likely also at play here.

The plastome and nuclear gene-tree dating analyses (Fig. 5 and Appendix S10) showed overlapping age ranges for the alternative topologies, making it difficult to distinguish between the alternative scenarios (ILS vs. hybridization) presented above. The divergences of Ericales + *Gentianidae* in the plastid tree (114.5 Ma) and in nuclear gene trees supporting this topology (median age = 112.5 Ma, range: 108.8–114.2) mostly preceded the divergences of Cornales + Ericales in the main nuclear gene set (median age = 109.8 Ma, range: 84.491–114.5), but with notable overlap.

From the PhyloNet analyses, the inferred network with the highest log probability (−7185.1706) included a single reticulation, with the lineage Ericales descending from a reticulate node involving Cornales and *Gentianidae* (Fig. 5). This model (a network with one reticulation) was found to be significantly better ($P = 0.0001$) than a simpler alternative (a strictly bifurcating tree) based on the likelihood-ratio test. The CalGTProb command was used to compute the likelihood scores for these alternatives (−1070.2706, −1062.7161) given the set of gene trees; these scores were used for the likelihood-ratio test.

<h1>DISCUSSION

<h2>Novel insights into asterid phylogeny

Our results provide important new insights into relationships among major asterid lineages, highlighting multiple instances of well-supported conflict between the nuclear and plastid genomes (cytonuclear discordance). Some of these points of conflict pertain to areas of asterid phylogeny that have long been problematic—for example, the relationships among basal lamiid lineages (Stull et al., 2015) and the core lamiid orders (Refulio-Rodriguez and Olmstead, 2014). Others, however—for example, the positions of Aquifoliales and Ericales—had seemed firmly settled, based on well-supported placements in phylogenies inferred from plastid or primarily plastid data (Soltis et al., 2018a). These findings have important implications beyond relationships in *Asteridae*. Our results underscore the importance of reevaluating broad-scale angiosperm phylogeny, as well as green-plant phylogeny in general, with more extensive evidence from the nuclear genome, because many other "accepted" relationships inferred from the plastome may be incorrect or incompletely understood due to evolutionary processes such as

ILS or organellar capture via hybridization. Unfortunately, our results from the mitochondrial analyses were largely unresolved/poorly supported, suggesting that the slowly evolving mitochondrial genome may generally hold little promise for examining instances of deep cytonuclear discordance. Nevertheless, our results show that conflict analysis of the nuclear and plastid genomes alone is useful for characterizing deep cytonuclear discordance, which can be explored further using simulations, gene-tree dating, network analyses, or other approaches (e.g., Folk et al., 2017; García et al., 2017; Morales-Briones et al., 2018) to help elucidate the underlying biological causes.

Our analyses showed that perhaps both ancient reticulation and ILS were at play in the initial radiation of *Asteridae*. The network analyses inferred Ericales to be a reticulate lineage formed by hybridization between Cornales and *Gentianidae*. However, the organellar simulations indicated that, with the rapidity of speciation events at the base of *Asteridae*, extensive ILS should have been common as well. Furthermore, results from the gene-tree dating analyses were largely overlapping and therefore plausibly consistent with either scenario. Teasing apart ILS from hybridization in ancient radiations is a challenging task (Morales-Briones et al., 2019), in part because rapid radiations create ample opportunity for both of these evolutionary processes (Fontaine et al., 2015; Pease and Hahn, 2015), and additionally because phylogenetic signal is often eroded over long periods of evolutionary time (Smith et al., 2019), reducing our ability to infer any type of evolutionary process. However, it is noteworthy that the observed conflict coincides with inferred WGDs for *Asteridae* and for Ericales + Cornales (Landis et al., 2018); this provides additional potential evidence for reticulation in early asterid diversification. We elaborate on the potential significance of these WGD events below.

The processes responsible for the different placements of Aquifoliales are less clear, as the considerable phylogenetic distance between the alternative positions makes it more difficult to evaluate alternative scenarios. Nevertheless, the nuclear genome overwhelmingly supports the placement of this order in the lamiid clade, whereas our plastid analyses as well as earlier plastid results (e.g., Moore et al., 2010, 2011; Tank and Donoghue, 2010) place the order with strong support as sister to all other campanulids. This nuclear-based placement is actually more consistent with morphological data and traditional classifications, given that Aquifoliales (and specifically Cardiopteridaceae and Stemonuraceae) include numerous genera of what was formerly Icacinaceae *s.l.* (Kårehed, 2001).

<h2>Implications for character evolution

Aquifoliales, Icacinaceae, and Metteniusaceae (which also include genera of Icacinaceae *s.l.*) are all generally woody, evergreen plants, mostly of tropical habitats, with alternate, exstipulate leaves (although minute stipules are present in Aquifoliaceae), relatively inconspicuous flowers with free to slightly fused petals, superior ovaries, and drupaceous fruits (Stull et al., 2018). Many of these features, however, are likely ancestral for *Gentianidae* or *Asteridae* as a whole (Stull et al., 2018) and therefore do not necessarily offer morphological support for the placement of Aquifoliales in the lamiids, although they also do not contradict it. Nevertheless, the position of Aquifoliales in lamiids, as opposed to campanulids, might help clarify the reconstructions of several morphological characters (e.g., ovary position, fruit type) in the latter clade. For example, inferior ovaries likely would be reconstructed as ancestral for campanulids (see Stull et al., 2018), as would capsular fruits (see Beaulieu and Donoghue, 2013).

The placement (and slightly revised circumscription, see below) of Icacinaceae based on the nuclear data set also has several broader implications for character evolution. Previously, unilacunar nodes and simple perforation plates were considered likely synapomorphies for the subclade of Icacinaceae excluding *Cassinopsis* (Stull et al., 2015). However, these morphological features also characterize the core lamiids, to which Icacinaceae (minus *Cassinopsis*) appear to be sister; therefore, these features may be synapomorphic for this broader clade instead. The possible placement of *Oncotheca* as sister to the rest of *Gentianidae* would also be consequential for understanding broader patterns of character evolution in *Gentianidae* and *Asteridae*, although this phylogenetic position for *Oncotheca* deserves further scrutiny. It is noteworthy that *Oncotheca* has five carpels while *Gentianidae* more broadly are characterized by only two or three; perhaps carpel reduction in *Gentianidae* occurred after the divergence of *Oncotheca*. However, carpel number is ambiguous in Icacinaceae, Metteniusaceae, and Aquifoliales, where pseudomonomery seems to predominate (Engler, 1872; Baillon, 1874; Sleumer, 1942; González and Rudall, 2010; Tobe, 2012; Kong et al., 2014), and the genus *Emmotum* (Metteniusaceae) is notable for having a pseudotrimerous gynoecium of five carpels (Endress and Rapini, 2014). Thus, more detailed developmental studies of asterid gynoecia, in addition to further phylogenetic investigation, will be necessary to resolve these questions.

<h2>Phylogeny and circumscription of Icacinaceae and Metteniusaceae

The placement of *Cassinopsis* in Icacinaceae was questioned in earlier phylogenetic studies (e.g., Kårehed, 2001), although analyses of the plastid genome (Stull et al., 2015; present study) have supported its inclusion in Icacinaceae as sister to the rest of the family. Morphologically, however, it is isolated from the other members of Icacinaceae—for example, it differs from other genera of Icacinaceae in having scalariform perforation plates, multilacunar nodes, opposite leaves, and occasional axillary thorns (Potgieter and Duno, 2016). Several of these features (scalariform perforation plates and multilacunar nodes) are shared with Metteniusaceae, although not necessarily synapomorphic. The nuclear analyses presented here suggest that *Cassinopsis* is indeed phylogenetically isolated from Icacinaceae and more closely related to Metteniusaceae, and should perhaps be included within a slightly expanded circumscription of the latter family. For Icacinaceae, the absence of *Cassinopsis* thus renders the family more morphologically coherent (Kårehed, 2001; Stull et al., 2015). Excluding *Cassinopsis*, the circumscription of Icacinaceae would thus include 23 genera: all those listed in Stull et al. (2015) plus *Vadensea*, a genus recently established to accommodate the continental African species of *Desmostachys* (Jongkind and Lachenaud, 2019), which was previously shown to be polyphyletic (Byng et al., 2014; Stull et al., 2015).

<h2>Ancient hybridization and whole-genome duplication in *Asteridae*

Recent analyses of transcriptome and genome data (e.g., Cannon et al., 2014; Landis et al., 2018; Smith et al., 2018; Yang et al., 2018; Larson et al., 2019; One Thousand Plant Transcriptomes Initiative, 2019) have greatly expanded our understanding of the extent of paleopolyploidy (or WGD) in the evolutionary history of green plants, and especially of angiosperms. These studies indicate that polyploidy played an important role in the early evolution of *Asteridae*, with WGD events inferred in the common ancestor of *Asteridae* as a whole and in the common ancestor of Cornales + Ericales. There also appears to be a WGD event in the early diversification of Ericales (Larson et al., 2019), as well as in numerous other, more nested areas of asterid phylogeny (Landis et al., 2018; One Thousand Plant Transcriptomes Initiative, 2019; see also Barker et al., 2008). Notably, the most prominent genomic conflict we observed in our analyses seems to coincide with these deep asterid WGD events. Namely, the earliest divergences in asterid phylogeny—those of *Gentianidae* + (Cornales + Ericales)—exhibit well-supported

conflict both between the nuclear and plastid genomes and within the nuclear genome. This suggests that these deep WGD events were significantly involved in the initial radiation of asterids. While the relationship between WGD and diversification remains unclear, or perhaps idiosyncratic (e.g., Vamos and Dickinson, 2006; Tank et al. 2015; Kellogg, 2016; Landis et al., 2018), the potential for WGD to contribute directly to evolutionary novelty—in terms of biochemistry, physiology, morphology, or ecological niche—has long been recognized (e.g., Stebbins, 1950; Roose and Gottleib, 1976; Levin, 1983; Edger et al., 2015; Soltis and Soltis, 2016; Van de Peer et al., 2017; Baniaga et al., 2020). It is therefore possible that these WGD events played a role in the evolution of major asterid innovations, such as iridoid production; integument, nucellus, and endosperm type; and sympetaly (Stull et al., 2018). This hypothesis deserves further study.

The methods used to detect WGDs include examination of K_s plots—that is, the distribution of synonymous substitutions per site (K_s) between paralogs in a genome (Lynch and Conery, 2000; Cui et al., 2006; One Thousand Plant Transcriptomes Initiative, 2019); a spike of many paralog pairs with a similar level of divergence in the plot suggests a simultaneous point of origin stemming from a WGD event. Because autopolyploidy results in extra alleles (not duplicate genes) across homologous chromosomes with polysomic inheritance, we would not necessarily expect K_s plots to capture autopolyploidy events, or at least not as distinctly as they should capture allopolyploid events. Consequently, WGDs inferred, at least in part, using K_s plots are perhaps more likely to be the result of allopolyploidy. Given these considerations—as well as the patterns of genomic conflict observed here (Figs. 2 and 3) and the results of our phylogenetic network analyses (Fig. 5)—the WGD event detected in the common ancestor of Ericales + Cornales (e.g., Landis et al., 2018; Larson et al., 2019; One Thousand Plant Transcriptomes Initiative, 2019) might therefore have resulted from a reticulation event involving the ancestors of Cornales and *Gentianidae*. This possibility deserves further scrutiny, as identifiability of sources of conflict is generally low (Morales-Briones et al., 2019), and other types of data (e.g., complete genomes) would be useful to fully resolve this question.

In contrast to most major asterid groups, Cornales and Ericales both have an extensive Cretaceous fossil record tracing roughly to the Turonian-Coniacian boundary, ~90 mya (e.g., Crepet et al., 2013, 2018; Atkinson, 2018; Atkinson et al., 2019). These fossil reports, along with various molecular dating analyses (e.g., Wikström et al., 2001; Bremer et al., 2004; Bell et al.,

2010; Magallón et al., 2015), indicate that Cornales and Ericales were actively diversifying in the early part of the Upper Cretaceous—perhaps spurred, at least in part, by these WGD events (Landis et al., 2018). Backbone relationships in both Cornales (e.g., Xiang et al., 2002; 2011; Atkinson, 2018; Fu et al., 2019) and Ericales (e.g., Schönenberger et al., 2005; Rose et al., 2018; Larson et al., 2019) have been notoriously difficulty to resolve, which likely reflects the rapid (and ancient) diversification of these lineages. Ericales also seem to share an additional WGD in their early evolutionary history, shortly after divergence of the crown node (Landis et al., 2018; Larson et al., 2019; One Thousand Plant Transcriptomes Initiative, 2019). Collectively, these lines of evidence reveal a complex early evolutionary history for asterid angiosperms.

<h1>FUTURE DIRECTIONS

While this study provides important initial insight into asterid phylogeny based on the nuclear genome, additional sampling is necessary to further clarify some relationships. For example, greater sampling of Aquifoliales, Metteniusaceae (including *Cassinopsis*), and Icacinaceae for transcriptome or target-capture sequencing would more firmly resolve major lamiid relationships and perhaps help elucidate the biological processes underlying this complex radiation. In addition to the deepest asterid WGD events, described above, there appear to be multiple WGD events early in the diversification of both lamiids and campanulids (Landis et al., 2018). However, among the basal lamiids especially, poor transcriptomic sampling has made it difficult to pinpoint the precise locations of several WGD events. For example, *Pyrenacantha malvifolia* (Icacinaceae) shows signatures of a WGD in its history (Landis et al., 2018), but in the absence of transcriptomes from close relatives, it is difficult to say if a WGD event is limited to this species or occurred deeper in the evolutionary history of Icacinaceae. A similar situation is evident for *Ilex paraguariensis* (Landis et al., 2018). Increased sampling is essential for accurately determining the number and locations of WGDs within a clade (Yang et al., 2018). This issue is demonstrated by comparing analyses of dense (Yang et al., 2018) and sparse (Landis et al., 2018) samplings of the clade Caryophyllales; the former provides a much more nuanced picture of the extent of WGD in this clade.

Future nuclear phylogenomic studies of asterids will continue to illuminate the complex evolutionary history of this important clade. We anticipate that similar reinvestigations of other major groups—not only angiosperms, but other lineages of green life—will reveal other

unexpected nuclear-cytoplasmic discordances that will ultimately enrich our understanding of evolutionary history. Although the generation of new sequence data will remain important, we are quickly approaching a time when some of the most significant new insights about the Tree of Life and the complexities of genome history will likely result from careful reanalysis of existing data.

<h1>ACKNOWLEDGMENTS

We acknowledge support from a National Science Foundation (NSF) Doctoral Dissertation Improvement Grant (DEB-1310805 to P.S.S. and G.W.S.) and an NSF Postdoctoral Fellowship (DBI-1612032 to G.W.S.). We thank J. Freudenstein and two anonymous reviewers for helpful comments that improved the manuscript.

<h1>AUTHOR CONTRIBUTIONS

G.W.S., S.A.S., P.S.S., and D.E.S. conceived the study. G.W.S. and M.A.G. compiled the data set. G.W.S. performed analyses. G.W.S. wrote the manuscript with help from S.A.S., P.S.S., D.E.S., and M.A.G.

<h1>>DATA AVAILIBILITIY

The raw Illumina data generated for this study are available through the Sequence Read Archive (SRA accession PRJNA602434). The DNA alignments, configuration files, and major results are available from the Dryad Digital Repository: https://doi.org/10.5061/dryad.h70rxwdfq (Stull et al., 2020).

<h1>SUPPORTING INFORMATION

Appendix S1. Taxon sampling in the nuclear and plastid data sets.

Appendix S2. Taxon sampling in the mitochondrial data set.

Appendix S3. GenBank accession numbers.

Appendix S4. Best tree from the concatenated maximum likelihood RAxML analysis of the 307 nuclear genes, including outgroup species, branch lengths, and bootstrap support at each node. This tree was used to make Figure 1.

Appendix S5. Species tree, inferred by ASTRAL from the 307 nuclear genes, including outgroup species, branch lengths, and local posterior probability values at each node. This tree was used for the first PhyParts analysis and to make Figure 2.

Appendix S6. Tree from the concatenated maximum likelihood RAxML analysis of the 77 plastid genes, including outgroup species, branch lengths, and bootstrap support at each node. This tree was used for the second PhyParts analysis and to make Figure 3.

Appendix S7. Plot juxtaposing the nuclear species tree and the plastid tree, generated using the "cophylo" function in Phytools, to highlight major topological differences between the trees. Note that the degree of conflict concerning the placement of *Dillenia* is exaggerated because of how the trees are rotated; in both trees (nuclear and chloroplast), this genus has a deep, isolated, and poorly resolved/supported position.

Appendix S8. Species tree, inferred by ASTRAL from the 77 plastid genes, including branch lengths and local posterior probability values at each node.

Appendix S9. Tree from the concatenated maximum likelihood RAxML analysis of the 35 mitochondrial genes, including branch lengths and bootstrap support at each node.

Appendix S10. Summary results from the gene-tree dating analyses implemented in treePL.

<h1>LITERATURE CITED

Albach, D. C., P. S. Soltis, and D. E. Soltis. 2001a. Patterns of embryological and biochemical evolution in the asterids. *Systematic Botany* 26: 242–262.

Albach, D. C., P. S. Soltis, D. E. Soltis, and R. G. Olmstead. 2001b. Phylogenetic analysis of asterids based on sequences of four genes. *Annals of the Missouri Botanical Garden* 88: 162–212.

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

Atkinson, B. A. 2018. The critical role of fossils in inferring deep-node phylogenetic relationships and macroevolutionary patterns in Cornales. *American Journal of Botany* 105: 1401–1411.

Baillon, H. 1874. Deuxième étude sur les Mappiées (suite). *Adansonia* 11: 187–203.

Baniaga, A. E., H. E. Marx, N. Arrigo, and M. S. Barker. 2020. Polyploidy plants have faster rates of multivariate niche differentiation than their diploid relatives. *Ecology Letters* 23: 68–78.

Barker, M. S., N. C. Kane, M. Matvienko, A. Kozik, R. M. Michelmore, S. J. Knapp, and L. H. Rieseberg. 2008. Multiple paleopolyploidizations during the evolution of the Compositae reveal parallel patterns of duplicate gene retention after millions of years. *Molecular Biology and Evolution* 25: 2445–2455.

Beaulieu, J. M., and M. J. Donoghue. 2013. Fruit evolution and diversification in campanulid angiosperms. *Evolution* 67: 3132–3144.

Bell, C. D., D. E. Soltis, and P. S. Soltis. 2010. The age and diversification of the angiosperms re-revisited. *American Journal of Botany* 97: 1296–1303.

Bouchaert, R. R., and J. Heled. 2014. DensiTree 2: Seeing trees through the forest. bioRxiv doi: https://doi.org/10.1101/012401.

Bremer, B., K. Bremer, N. Heidari, P. Erixon, R. G. Olmstead, A. A. Anderberg, M. Källersjö, and E. Barkhordarian. 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. *Molecular Phylogenetics and Evolution* 24: 274–301.

Brenner, G. J. 1996. Evidence for the earliest stage of angiosperm pollen evolution: a paleoequatorial section from Israel. *In*: D. W. Taylor, and L. J. Hickey [eds.], Flowering plant origin, evolution and phylogeny, 91–115. Chapman and Hall, New York, NY, USA.

Brown, J. W., J. F. Walker, and S. A. Smith. 2017. Phyx: phylogenetic tools for unix. *Bioinformatics* 33: 1886–1888.

Cannon, S. B., M. R. McKain, A. Harkess, M. N. Nelson, S. Dash, M. K. Deyholos, Y. Peng, et al. 2014. Multiple polyploidy events in the early radiation of nodulating and nonnodulating legumes. *Molecular Biology and Evolution* 32: 193–210.

Cantino, P. D., J. A. Doyle, S. W. Graham, W. S. Judd, R. G. Olmstead, D. E. Soltis, P. S. Soltis, and M. J. Donoghue. 2007. Towards a phylogenetic nomenclature of Tracheophyta. *Taxon* 56: 1E–44E.

Cao, Z., X. Liu, H. A. Ogilvie, Z. Yan, and L. Nakhleh. 2019. Practical aspects of phylogenetic network analysis using PhyloNet. *bioRxiv* doi: https://doi.org/10.1101/746362.

Chamala, S., N. García, G. T. Godden, V. Krishakumar, I. E. Jordon-Thaden, R. De Smet, W. B. Barbazuk, et al. 2015. MarkerMiner 1.0: A new application for phylogenetic marker development using angiosperm transcriptomes. *Applications in Plant Sciences* 3: 1400115.

Chou, J., A. Gupta, S. Yaduvanshi, R. Davidson, M. Nute, S. Mirarab, and T. Warnow. 2015. A comparative study of SVDquartets and other coalescent-based species tree estimation methods. *BMC Genomics* 16: S2.

Coiro, M., J. A. Doyle, and J. Hilton. 2019. How deep is the conflict between molecular and fossil evidence on the age of angiosperms? *New Phytologist* 223: 83–99.

Couvreur, T. L. P., A. J. Helmstetter, E. J. M. Koenen, K. Bethune, R. D. Brandão, S. A. Little, H. Sauquet, and R. H. J. Erkens. 2019. Phylogenomics of the major tropical plant family Annonaceae using target enrichment of nuclear genes. *Frontiers in Plant Science* 9: 1941.

Cui, L., P. K. Wall, J. H. Leebens-Mack, B. G. Lindsay, D. E. Soltis, J. J. Doyle, P. S. Soltis, et al. 2006. Widespread genome duplications throughout the history of flowering plants. *Genome Research* 16: 738–749.

De Smet, R., K. L. Adams, K. Vanderpoele, M. C. E. Van Montagu, S. Maere, and Y. Van de Peer. 2013. Convergent gene loss following gene and genome duplications creates single-copy families in flowering plants. *Proceedings of the National Academy of Sciences, USA* 110: 2898–2903.

Doyle, J. A., P. Biens, A. Doerenkamp, and S. Jardiné. 1977. Angiosperm pollen from the pre-Albian Lower Cretaceous of Equatorial Africa. *Bulletin Centres Recherches Exploration-Production Elf-Aquitaine* 1: 451–473.

Doyle, J. J. 1992. Gene trees and species trees: Molecular systematics as one-character taxonomy. *Systematic Biology* 17: 144–163.

Edger, P. P., H. M. Heidel-Fischer, M. Bekaert, J. Rota, G. Glöckner, A. E. Platts, D. G. Heckel, et al. 2015. The butterfly plant arms-race escalated by gene and genome duplications. *Proceedings of the National Academy of Sciences, USA* 112: 8362–8366.

Edwards, S. V., Z. Xi, A. Janke, B. C. Faircloth, J. E. McCormack, T. C. Glenn, B. Zhong, et al. 2016. Implementing and testing the multispecies coalescent model: A valuable paradigm for phylogenomics. *Molecular Phylogenetics and Evolution* 94: 447–462.

Endress, P. K., and A. R. Rapini. 2014. Flora structure of *Emmotum* (Icacinaceae *sensu stricto* or Emmotaceae), a phylogenetically isolated genus of lamiids with a unique pseudotrimerous gynoecium, bitegmic ovules and monossporangiate thecae. *Annals of Botany* 114: 945–959.

Engler, A. 1872. Icacineae. *In* C. F. P. Martius, and A. W. Eichler [eds.], Flora Brasiliensis, vol. 12, pars 2., 40–62. Fleischer, Leipzig, Germany.

Folk, R. A., J. R. Mandel, and J. V. Freudenstein. 2017. Ancestral Gene Flow and Parallel Organellar Genome Capture Result in Extreme Phylogenomic Discord in a Lineage of Angiosperms. *Systematic Biology* 66: 320–337.

Fontaine, M. C., J. B. Pease, A. Steele, R. M. Waterhouse, D. E. Neafsey, I. V. Sharakhov, X. Jiang, et al. 2015. Extensive introgression in a malaria vector species complex revealed by phylogenomics. *Science* 347: 1258524.

Fu, C.-N., Z.-Q. Mo, J.-B. Yang, X.-J. Ge, D.-Z. Li, Q.-Y. Xiang, L.-M. Gao. 2019. Plastid phylogenomics and biogeographic analysis support a trans-Tethyan origin and rapid early radiation of Cornales in the Mid-Cretaceous. *Molecular Phylogenetics and Evolution* 140: 106601.

Galtier, N., and V. Daubin. 2008. Dealing with incongruence in phylogenomic analyses. *Philosophical Transactions of the Royal Society B: Biological Sciences* 363: 4023–4029.

García, N., R. A. Folk, A. W. Meerow, S. Chamala, M. A. Gitzendanner, R. Souza de Oliveira, D. E. Soltis, and P. E. Soltis. 2017. Deep reticulation and incomplete lineage sorting obscure the diploid phylogeny of rain-lilies and allies (Amaryllidaceae tribe Hippeastreae). *Molecular Phylogenetics and Evolution* 111: 231–247.

Gitzendanner, M. A., P. S. Soltis, G. K. S. Wong, B. R. Ruhfel, and D. E. Soltis. 2018. Plastid phylogenomic analysis of green plants: A billion years of evolutionary history. *American Journal of Botany* 105: 291–301.

Gonçalves, D. J. P., B. B. Simpson, E. M. Ortiz, G. H. Shimizu, and R. K. Jansen. 2019. Incongruence between gene trees and species trees and phylogenetic signal variation in plastid genes. *Molecular Phylogenetics and Evolution* 138: 219–232.

González, F. A., and P. J. Rudall. 2010. Flower and fruit characters in the early-divergent lamiid family Metteniusaceae, with particular reference to the evolution of pseudomonomery. *American Journal of Botany* 97: 191–206.

Hillis, D. M., and J. J. Bull. 1993. An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analysis. *Systematic Biology* 42: 182–192.

Huang, D. I., C. A. Hefer, N. Kolosova, C. J. Douglas, and Q. C. B. Cronk. 2014. Whole plastome sequencing reveals deep plastid divergence and cytonuclear discordance between closely related balsam poplars, *Populus balsamifera* and *P. trichocarpa* (Salicaceae). *New Phytologist* 204: 693–703.

Hughes, N. F., and A. B. McDougall. 1990. Barremian-Aptian angiospermid pollen records from southern England. *Review of Palaeobotany and Palynology* 65: 145–51.

Johnson, M. G., L. Pokorny, S. Dodsworth, L. R. Botigue, R. S. Cowan, A. Devault, W. L. Eiserhardt, et al. 2019. A universal probe set for targeted sequencing of 353 nuclear genes from any flowering plant designed using k-medoids clustering. *Systematic Biology* 68: 594–606.

Jongkind, C. H. C., and O. Lachenaud. 2019. *Vadensea* (Icacinaceae), a new genus to accommodate the continental African species of *Desmostachys*. *Phytotaxa* 405: 237–247.

Judd, W. S., and R. G. Olmstead. 2004. A survey of tricolpate (eudicot) phylogenetic relationships. *American Journal of Botany* 91: 1627–1644.

Kellogg, E. A. 2016. Has the connection between polyploidy and diversification actually been tested? *Current Opinion in Plant Biology* 30: 25–32.

King, N., and A. Rokas. 2017. Embracing uncertainty in reconstructing early animal evolution. *Current Biology* 27: R1081–R1088.

Kong, D.-R., M. Schori, S.-G. Lu, L. Li, and H. Peng. 2014. Floral development of *Cardiopteris*, with emphasis on gynoecial structure and ovular morphology. *Journal of Systematics and Evolution* 54: 629–642.

Larson, D. A., J. F. Walker, O. M. Vargas, and S. A. Smith. 2019. A consensus phylogenomic approach highlights paleopolyploid and rapid radiation in the history of Ericales. *bioRxiv* https://doi.org/10.1101/816967.

Levin, D. A. 1983. Polyploidy and novelty in flowering plants. *The American Naturalist* 122: 1–25.

Li, H.-T., T.-S. Yi, L.-M. Gao, P.-F. Ma, T. Zhang, J.-B. Yang, M.-A. Gitzendanner, et al. 2019. Origin of angiosperms and the puzzle of the Jurassic gap. *Nature Plants* 5: 461–470.

Lynch, M., and J. S. Conery. 2000. The evolutionary fate and consequences of duplicate genes. *Science* 290: 1151–1155.

Maddison, W. P. 1997. Gene trees in species trees. *Systematic Biology* 46: 523–536.

Magallón, S., and A. Castillo. 2009. Angiosperm diversification through time. *American Journal of Botany* 96: 349–365.

Magallón, S., S. Gómez-Acevedo, L. L. Sánchez-Reyes, and T. Hernández-Hernández. 2015. A metacalibrated time-tree documents the early rise of flowering plant phylogenetic diversity. *New Phytologist* 207: 437–453.

Matasci, N., L.-H. Hung, Z. Yan, E. J. Carpenter, N. J. Wickett, S. Mirarab, N. Nguyen, et al. 2014. Data access for the 1,000 Plants (1KP) project. *Gigascience* 3: 17.

McCauley, D. E. 1994. Contrasting the distribution of chloroplast DNA and allozyme polymorphism among local populations of *Silene alba*: Implications for studies of gene flow in plants. *Proceedings of the National Academy of Sciences, USA* 91: 8127–8131.

Mirarab, S., R. Reaz, M. S. Bayzid, T. Zimmermann, M. S. Swenson, and T. Warnow. 2014. ASTRAL: genome-scale coalescent-based species tree estimation. *Bioinformatics* 30: i541–i548.

Moore, M. J., C. D. Bell, P. S. Soltis, and D. E. Soltis. 2007. Using plastid genomic-scale data to resolve enigmatic relationships among basal angiosperms. *Proceedings of the National Academy of Sciences, USA* 104: 19363–19368.

Moore, M. J., N. Hassan, M. A. Gitzendanner, R. A. Bruenn, M. Croley, A. Vandeventer, J. W. Horn, et al. 2011. Phylogenetic analysis of the plastid inverted repeat for 244 species: insights into deeper-level angiosperm relationships from a long, slowly evolving sequence region. *International Journal of Plant Sciences* 172: 541–558.

Moore, M. J., P. S. Soltis, C. D. Bell, J. G. Burleigh, and D. E. Soltis. 2010. Phylogenetic analysis of 83 plastid genes further resolves the early diversification of eudicots. *Proceedings of the National Academy of Sciences, USA* 107: 4623–4628.

Morales-Briones, D. F., G. Kadereit, D. T. Tefarikis, M. J. Moore, S. A. Smith, S. F. Brockington, A. Timoneda, et al. 2019. Disentangling sources of gene tree discordance in phylotranscriptomic datasets: a case study from Amaranthaceae s.l. *bioRxiv*: https://doi.org/10.1101/794370.

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

Olmstead, R. G., B. Bremer, K. M. Scott, and J. D. Palmer. 1993. A parsimony analysis of the Asteridae sensu lato based on *rbcL* sequences. *Annals of the Missouri Botanical Garden* 80: 700–722.

Olmstead, R. G., K.-J. Kim, Jansen RK, Wagstaff SJ. 2000. The phylogeny of the *Asteridae* sensu lato based on chloroplast *ndhF* gene sequences. *Molecular Phylogenetics and Evolution* 16: 96–112.

Olmstead, R. G., H. J. Michaels, K. M. Scott, and J. D. Palmer. 1992. Monophyly of the Asteridae and identi cation of their major lineages inferred from DNA sequences of *rbcL*. *Annals of the Missouri Botanical Garden* 79: 249–265.

One Thousand Plant Transcriptomes Initiative. 2019. One thousand plant transcriptomes and phylogenomics of green plants. *Nature* 574: 679–685.

Parks, M. B., N. J. Wickett, and A. J. Alverson. 2017. Signal, uncertainty, and conflict in phylogenomic data for a diverse lineage of microbial eukaryotes (diatoms, Bacillariophyta). *Molecular Biology and Evolution* 35: 80–93.

Pease, J. B., and M. W. Hahn. Detection and polarization of introgression in a five-taxon phylogeny. *Systematic Biology* 64: 651–662.

Potgieter, M. J., and R. Duno. 2016. Icacinaceae. *In*: J. Kadereit, and V. Bittrich [eds.], Flowering plants. Eudicots. The families and genera of vascular plants, vol. 14., 239–256. Springer-Verlag, Berlin, Germany.

Qiu, Y.-L., J. Lee, F. Bernasconi-Quadroni, D. E. Soltis, P. S. Soltis, M. Zanis, E. A. Zimmer, Z. Chen, V. Savolainen, and M. W. Chase. 1999. The earliest angiosperms: evidence from mitochondrial, plastid and nuclear genomes. *Nature* 402: 404–407.

Qiu, Y.-L., L. Li, B. Wang, J.-Y. Xue, T. A. Hendry, R.-Q. Li, J. Brown, et al. 2010. Angiosperm phylogeny inferred from sequences of four mitochondrial genes. *Journal of Systematics and Evolution* 48: 391–425.

Refulio-Rodriguez, N. F., and R. G. Olmstead. 2014. Phylogeny of Lamiidae. *American Journal of Botany* 101: 287–299.

Revell, L. J. 2012. Phytools: An R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution* 3: 217–223.

Rieseberg, L. H., and D. E. Soltis. 1991. Phylogenetic consequences of cytoplasmic gene flow in plants. *Evolutionary Trends in Plants* 5: 65–84.

Roose, M. L., and L. D. Gottlieb. 1976. Genetic and biochemical consequences of polyploidy in *Tragapogon*. *Evolution* 30: 818–830.

Rose, J. P., T. J. Kleist, S. D. Löfstrand, B. T. Drew, J. Schönenberger, and K. J. Sytsma. 2018. Phylogeny, historical biogeography, and diversification of angiosperm order Ericales suggest ancient Neotropical and East Asian connections. *Molecular Phylogenetics and Evolution* 122: 59-79.

Ruhfel, B. R., M. A. Gitzendanner, P. S. Soltis, D. E. Soltis, and J. G. Burleigh. 2014. From algae to angiosperms—inferring the phylogeny of green plants (*Viridiplantae*) from 360 plastid genomes. *BMC Evolutionary Biology* 14: 23.

Schönenberger, J, A. A. Anderberg, and K. J. Sytsma. 2005. Molecular phylogenetics and patterns of floral evolution in the Ericales. *International Journal of Plant Sciences* 166.2: 265–288.

Sleumer, H. 1942. Icacinaceae. *In* A. Engler, and K. Prantl [eds.], Die natürlichen Pflanzenfamilien, 2nd edn, vol. 20b, 322–396. Engelmann, Leipzig, Germany.

Smith, S. A., and J. W. Brown. 2018. Constructing a broadly inclusive seed plant phylogeny. *American Journal of Botany* 105: 302–314.

Smith, S. A., J. W. Brown, Y. Yang, R. Bruenn, C. P. Drummond, S. F. Brockington, J. F. Walker, et al. 2018. Disparity, diversity, and duplications in the Caryophyllales. *New Phytologist* 217: 836–854.

Smith, S. A., M. J. Moore, J. W. Brown, and Y. Yang. 2015. Analysis of phylogenomic datasets reveals conflict, concordance, and gene duplications with examples from animals and plants. *BMC Evolutionary Biology* 15: 150.

Smith, S. A., N. Walker-Hale, J. F. Walker, and J. W. Brown. 2019. Phylogenetic conflicts, combinability, and deep phylogenomics in plants. *Systematic Biology*, in press.

Soltis, D. E., and R. K. Kuzoff. 1995. Discordance between nuclear and chloroplast phylogenies in the *Heuchera* group (Saxifragaceae). *Evolution* 49: 727–742.

Soltis, D. E., M. J. Moore, E. B. Sessa, S. A. Smith, and P. S. Soltis. 2018b. Using and navigating the plant tree of life. *American Journal of Botany* 105: 287–290.

Soltis, D. E., S. A. Smith, N. Cellinese, K. J. Wurdack, D. C. Tank, S. F. Brockington, N. F. Refulio-Rodriguez, et al. 2011. Angiosperm phylogeny: 17 genes, 640 taxa. *American Journal of Botany* 98: 704–730.

Soltis, D. E., P. S. Soltis, M. W. Chase, M. E. Mort, D. C. Albach, M. Zanis, V. Savolainen, et al. 2000. Angiosperm phylogeny inferred from 18S rDNA, *rbcL,* and *atpB* sequences. *Botanical Journal of the Linnean Society* 133: 381–461.

Soltis, D. E., P. S. Soltis, P. Endress, M. Chase, S. R. Manchester, W. Judd, L. Majure, E. Mavrodiev. 2018a. Phylogeny and evolution of the angiosperms. The University of Chicago Press, Chicago, Illinois, USA.

Soltis, P. S., R. A. Folk, and D. E. Soltis. 2019. Darwin review: angiosperm phylogeny and evolutionary radiations. *Proceedings of the Royal Society B: Biological Sciences* 286: 20190099.

Soltis, P. S., and D. E. Soltis. 2003. Applying the bootstrap in phylogeny reconstruction. *Statistical Science* 18: 256–267.

Soltis, P. S., and D. E. Soltis. 2009. The role of hybridization in plant speciation. *Annual Review of Plant Biology* 60: 561–588.

Soltis, P. S., and D. E. Soltis. 2016. Ancient WGD events as drivers of key innovations in angiosperms. *Current Opinion in Plant Biology* 30: 159–165.

Soltis, P. S., D. E. Soltis, and M. W. Chase. 1999. Angiosperm phylogeny inferred from multiple genes as a tool for comparative biology. *Nature* 402: 402–404.

Spoelhof, J. P., P. S. Soltis, and D. E. Soltis. 2017. Pure polyploidy: Closing the gaps in autopolyploid research. *Journal of Systematics and Evolution* 55: 340–352.

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

Stebbins, G. L. 1950. Variation and evolution in plants. Columbia University Press, New York, NY, USA.

Stull, G. W., R. Duno de Stefano, D. E. Soltis, and P. S. Soltis. 2015. Resolving basal lamiid phylogeny and the circumscription of Icacinaceae with a plastome-scale data set. *American Journal of Botany* 102: 1794–1813.

Stull, G. W., M. Schori, D. E. Soltis, and P. S. Soltis. 2018. Character evolution and missing (morphological) data across *Asteridae*. *American Journal of Botany* 105: 470–479.

Stull, G. W., P. S. Soltis, D. E. Soltis, M. A. Gitzendanner, and S. A. Smith. 2020. Data from: Nuclear phylogenomic analyses of asterids conflict with plastome trees and support novel relationships among major lineages. Dryad Digital Repository. doi:10.5061/dryad.h70rxwdfq.

Sukumaran, J., and M. T. Holder. 2010. DendroPy: A Python library for phylogenetic computing. *Bioinformatics* 26: 1569–1571.

Sun, M., D. E. Soltis, P. S. Soltis, X. Zhu, J. G. Burleigh, and Z. Chen. 2015. Deep phylogenetic incongruence in the angiosperm clade *Rosidae*. *Molecular Phylogenetics and Evolution* 83: 156–166.

Tank, D. C., and M. J. Donoghue. 2010. Phylogeny and phylogenetic nomenclature of the Campanulidae based on an expanded sample of the genes and taxa. *Systematic Botany* 35: 425–441.

Tank, D. C., J. M. Eastman, M. W. Pennell, P. S. Soltis, D. E. Soltis, C. E. Hinchliff, J. W. Brown, et al. 2015. Nested radiations and the pulse of angiosperm diversification: increased diversification rates often follow whole genome duplications. *New Phytologist* 207: 454–467.

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

Tobe, H. 2013. Floral morphology and structure of *Phyllonoma* (Phyllonomaceae): systematic and evolutionary implications. *Journal of Plant Research* 126: 709–718.

Trevisan, L. 1988. Angiospermous pollen (monosulcate–trichotomosulcate phase) from the very early Lower Cretaceous of southern Tuscany (Italy): some aspects. 7th International Palynological Congress Abstracts Volume 165. University of Queensland, Brisbane, Australia.

Vamosi, J. C., and T. A. Dickinson. 2006. Polyploidy and diversification: a phylogenetic investigation in Rosaceae. *International Journal of Plant Sciences* 167: 349–358.

Van de Peer, Y., E. Mizrachi, and K. Marachi. 2017. The evolutionary significance of polyploidy. *Nature Review Genetics* 18: 411–424.

Walker, J. W., N. Walker-Hale, O. M. Vargas, D. A. Larson, and G. W. Stull. 2019. Characterizing gene-tree conflict in plastome-inferred phylogenies. *PeerJ*, in press.

Wang, H., M. J. Moore, P. S. Soltis, C. D. Bell, S. F. Brockington, R. Alexandre, C. C. Davis, et al. 2009. Rosid radiation and the rapid rise of angiosperm-dominated forests. *Proceedings of the National Academy of Sciences, USA* 106: 3853–3858.

Wen, D., Y. Yu, J. Zhu, and L. Nakhleh. 2018. Inferring phylogenetic networks using PhyoNet. *Systematic Biology* 67: 735–740.

Whitfield, J. B., and P. J. Lockhart. 2007. Deciphering ancient rapid radiations. *Trends in Ecology and Evolution* 22: 258–265.

Wickett, N. J., S. Mirarab, N. Nguyen, T. Warnow, E. Carpenter, N. Matasci, S. Ayyampalayam, et al. 2014. Phylotranscriptomic analysis of the origin and early diversification of land plants. *Proceedings of the National Academy of Sciences, USA* 111: E4859–E4868.

Xiang, Q.-Y., M. L. Moody, D. E. Soltis, C. Fan, and P. S. Soltis. 2002. Relationships within Cornales and circumscription of Cornaceae—*matK* and *rbcL* sequence data and effects of outgroups and long branches. *Molecular Phylogenetics and Evolution* 24: 35–57.

Xiang, Q.-Y., D. T. Thomas, and Q. P. Xiang. 2011. Resolving and dating the phylogeny of Cornales – Effects of taxon sampling, data partitions, and fossil calibrations. *Molecular Phylogenetics and Evolution* 59: 123–138.

Xie, Y., G. Wu, J. Tang, R. Luo, J. Patterson, S. Liu, W. Huang, G. He, S. Gu, S. Li, et al. 2014. SOAPdenovo-Trans: De novo transcriptome assembly with short RNA-Seq reads. *Bioinformatics* 30: 1660–1666.

Yang, Y., M. J. Moore, S. F. Brockington, J. Mikenas, J. Olivieri, J. F. Walker, and S. A. Smith. 2018. Improved transcriptome sampling pinpoints 26 ancient and more recent polyploidy events in Caryophyllales, including two allopolyploidy events. *New Phytologist* 217: 855–870.

Yang, Y., M. J. Moore, S. F. Brockington, D. E. Soltis, G. K. S. Wong, E. J. Carpenter, Y. Zhang, et al. 2015. Dissecting molecular evolution in the highly diverse plant clade Caryophyllales using transcriptome sequencing. *Molecular Biology and Evolution* 32: 2001–2014.

Zeng, L., N. Zhang, Q. Zhang, P. K. Endress, J. Huang, and H. Ma. 2017. Resolution of deep eudicot phylogeny and their temporal diversification using nuclear genes from transcriptomic and genomic datasets. *New Phytologist* 214: 1338–1354.

Zeng, L., Q. Zhang, R. Sun, H. Kong, N. Zhang, and H. Ma. 2014. Resolution of deep angiosperm phylogeny using conserved nuclear genes and estimates of early divergence times. *Nature Communications* 5: 4956.

Zhang, C., M. Rabiee, E. Sayyari, and S. Mirarab. 2018. ASTRAL-III: Polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinformatics* 19: 153.

Zhang, C., E. Sayyari, and S. Mirarab. 2017. ASTRAL-III: Increased scalability and impacts of contracting low support branches. In: Meidanis J., Nakhleh L. (eds) Comparative Genomics. RECOMB-CG 2017. Lecture Notes in Computer Science, vol 10562. Springer, Cham.

FIGURE 1. Tree from the concatenated maximum likelihood RAxML analysis of the 307 nuclear genes, including relationship support from this analysis (bootstrap values) as well as from the species-tree analysis of the 307 nuclear gene trees (local posterior probability values). All branches received 100% support from both analyses unless otherwise indicated. A relationship in the concatenated phylogeny not present in the species tree is denoted by a hyphen.

FIGURE 2. Species tree topology, inferred by ASTRAL from the 307 nuclear genes, showing patterns of gene-tree concordance and conflict based on the PhyParts analysis. The pie charts at each node show the proportion of genes in concordance (blue), conflict (green = a single dominant alternative; red = all other conflicting trees), and without information (gray). The numbers above and below each branch are the numbers of concordant and conflicting genes at each bipartition, respectively.

FIGURE 3. Plastid topology, from the concatenated ML analysis of the 77 plastid genes, showing subsets of the nuclear genome in concordance and conflict with the plastid relationships, based on the PhyParts analysis of the 307 nuclear gene trees mapped against this tree. The pie charts at each node show the proportion of genes in concordance (blue), conflict (green = a single dominant alternative; red = all other conflicting trees), and without information (gray). The numbers above and below each branch are the numbers of concordant and conflicting genes at each bipartition, respectively.

FIGURE 4. Summary of the organellar phylogenies simulated under the coalescent, obtained by mapping the simulated trees against the guide tree (i.e., the nuclear species tree) using PhyParts. The pie charts show the proportion of simulated trees concordant (blue) and conflicting (green = a single dominant alternative; red = all other conflicting trees) with each node; the numbers above and below each pie chart are the numbers of concordant and conflicting trees, respectively.

FIGURE 5. (A) The optimal phylogenetic network inferred using PhyloNet, depicting Ericales as a reticulate lineage formed by hybridization between Cornales and *Gentianidae*. The inheritance probabilities (IP) are shown for each hybrid branch. (B) The dated nuclear gene trees (with ages estimated using treePL) visualized using DensiTree (Bouckaert and Heled, 2014). The time scale spans from 150 million years ago to the present.