



## OPEN ACCESS

## EDITED BY

Lisa Pokorny,  
Botanical Institute of Barcelona,  
Spanish National Research Council  
(CSIC), Spain

## REVIEWED BY

Alexandre Rizzo Zuntini,  
Royal Botanic Gardens, Kew,  
United Kingdom  
Zhi-Zhong Li,  
Wuhan Botanical Garden, Chinese  
Academy of Sciences (CAS), China  
Felix Forest,  
Royal Botanic Gardens, Kew,  
United Kingdom

## \*CORRESPONDENCE

Claude W. dePamphilis  
cwd3@psu.edu  
James Leebens-Mack  
jleebensmack@uga.edu

## †PRESENT ADDRESS

Prakash Raj Timilsena,  
School of Plant and Environmental  
Sciences (SPES), Virginia Polytechnic  
Institute and State University,  
Blacksburg, VA, United States

## SPECIALTY SECTION

This article was submitted to  
Plant Systematics and Evolution,  
a section of the journal  
Frontiers in Plant Science

RECEIVED 15 February 2022

ACCEPTED 29 September 2022

PUBLISHED 22 November 2022

## CITATION

Timilsena PR, Wafula EK, Barrett CF,  
Ayyampalayam S, McNeal JR,  
Rentsch JD, McKain MR, Heyduk K,  
Harkess A, Villegente M, Conran JG,  
Illing N, Fogliani B, Ané C, Pires JC,  
Davis JI, Zomlefer WB, Stevenson DW,  
Graham SW, Givnish TJ,  
Leebens-Mack J and dePamphilis CW  
(2022) Phylogenomic resolution of  
order- and family-level monocot  
relationships using 602 single-copy  
nuclear genes and 1375 BUSCO genes.  
*Front. Plant Sci.* 13:876779.  
doi: 10.3389/fpls.2022.876779

# Phylogenomic resolution of order- and family-level monocot relationships using 602 single-copy nuclear genes and 1375 BUSCO genes

Prakash Raj Timilsena<sup>1†</sup>, Eric K. Wafula<sup>1</sup>, Craig F. Barrett<sup>2</sup>, Saravanaraj Ayyampalayam<sup>3,4</sup>, Joel R. McNeal<sup>5</sup>, Jeremy D. Rentsch<sup>6</sup>, Michael R. McKain<sup>7</sup>, Karolina Heyduk<sup>8</sup>, Alex Harkess<sup>9</sup>, Matthieu Villegente<sup>10</sup>, John G. Conran<sup>11</sup>, Nicola Illing<sup>12</sup>, Bruno Fogliani<sup>10</sup>, Cécile Ané<sup>13,14</sup>, J. Chris Pires<sup>15</sup>, Jerrold I. Davis<sup>16</sup>, Wendy B. Zomlefer<sup>4</sup>, Dennis W. Stevenson<sup>17</sup>, Sean W. Graham<sup>18</sup>, Thomas J. Givnish<sup>13</sup>, James Leebens-Mack<sup>4\*</sup> and Claude W. dePamphilis<sup>1\*</sup>

<sup>1</sup>Department of Biology and Huck Institutes of the Life Sciences, Pennsylvania State University, University Park, PA, United States, <sup>2</sup>Department of Biology, West Virginia University, Morgantown, WV, United States, <sup>3</sup>Georgia Advanced Computing Resource Center, University of Georgia, Athens, GA, United States, <sup>4</sup>Department of Plant Biology, University of Georgia, Athens, GA, United States, <sup>5</sup>Department of Ecology, Evolution, and Organismal Biology, Biology Kennesaw State University, Kennesaw, GA, United States, <sup>6</sup>Department of Biology, Francis Marion University, Florence, SC, United States, <sup>7</sup>Department of Biological Sciences, University of Alabama, Tuscaloosa, AL, United States, <sup>8</sup>School of Life Sciences, University of Hawai'i at Mānoa, Honolulu, HI, United States, <sup>9</sup>HudsonAlpha Institute for Biotechnology, Huntsville, AL, United States, <sup>10</sup>Institut des Sciences Exactes et Appliquées (ISEA), University of New Caledonia, Noumea, New Caledonia, <sup>11</sup>Australian Centre for Evolutionary Biology and Biodiversity & Sprigg Geobiology Centre, School of Biological Sciences, University of Adelaide, Adelaide, SA, Australia, <sup>12</sup>Department of Molecular and Cell Biology, University of Cape Town, Cape Town, South Africa, <sup>13</sup>Department of Botany, University of Wisconsin-Madison, Madison, WI, United States, <sup>14</sup>Department of Statistics, University of Wisconsin-Madison, Madison, WI, United States, <sup>15</sup>Division of Biological Sciences and Bond Life Sciences Center, University of Missouri, Columbia, MO, United States, <sup>16</sup>School of Integrative Plant Sciences and L.H. Bailey Hortorium, Cornell University, Ithaca, NY, United States, <sup>17</sup>Department of Botany, University of British Columbia, Vancouver, BC, Canada, <sup>18</sup>New York Botanical Garden, New York, NY, United States

We assess relationships among 192 species in all 12 monocot orders and 72 of 77 families, using 602 conserved single-copy (CSC) genes and 1375 benchmarking single-copy ortholog (BUSCO) genes extracted from genomic and transcriptomic datasets. Phylogenomic inferences based on these data, using both coalescent-based and supermatrix analyses, are largely congruent with the most comprehensive plastome-based analysis, and nuclear-gene phylogenomic analyses with less comprehensive taxon sampling. The strongest discordance between the plastome and nuclear gene analyses is the monophyly of a clade comprising Asparagales and Liliales in our nuclear gene analyses, versus the placement of Asparagales and Liliales as successive sister clades to the commelinids in the plastome tree. Within orders, around six of 72 families shifted

positions relative to the recent plastome analysis, but four of these involve poorly supported inferred relationships in the plastome-based tree. In Poales, the nuclear data place a clade comprising Ecdeiocoleaceae+Joinvilleaceae as sister to the grasses (Poaceae); Typhaceae, (rather than Bromeliaceae) are resolved as sister to all other Poales. In Commelinales, nuclear data place Philydraceae sister to all other families rather than to a clade comprising Haemodoraceae+Pontederiaceae as seen in the plastome tree. In Liliales, nuclear data place Liliaceae sister to Smilacaceae, and Melanthiaceae are placed sister to all other Liliales except Campynemataceae. Finally, in Alismatales, nuclear data strongly place Tofieldiaceae, rather than Araceae, as sister to all the other families, providing an alternative resolution of what has been the most problematic node to resolve using plastid data, outside of those involving achlorophyllous mycoheterotrophs. As seen in numerous prior studies, the placement of orders Acorales and Alismatales as successive sister lineages to all other extant monocots. Only 21.2% of BUSCO genes were demonstrably single-copy, yet phylogenomic inferences based on BUSCO and CSC genes did not differ, and overall functional annotations of the two sets were very similar. Our analyses also reveal significant gene tree-species tree discordance despite high support values, as expected given incomplete lineage sorting (ILS) related to rapid diversification. Our study advances understanding of monocot relationships and the robustness of phylogenetic inferences based on large numbers of nuclear single-copy genes that can be obtained from transcriptomes and genomes.

#### KEYWORDS

phylogenomics, phylotranscriptomics, monocots, conserved single-copy genes, BUSCO, concordance analysis

## Introduction

The monocots are a large monophyletic group of angiosperms, comprising 12 orders, 77 families and about 60,000–85,000 species (Bremer et al., 2009; Chase and Reveal, 2009; Lughadha et al., 2016; Givnish et al., 2018). They underpin some of the most productive ecosystems, including grasslands (e.g., prairies and steppes) and many aquatic habitats (e.g., seagrass meadows) (Waycott et al., 2009). Human civilization depends on cereal crops such as rice, oats and wheat (e.g., Mabblerley, 2008). In addition to cereals and grains, major berry crops (e.g., plantain/banana), forage/fodder species (grasses), and various stem and “root” crops (e.g., sugar cane, onion, yam tubers), collectively provide core food sources for billions of humans. Some individual species have been put to versatile uses: the coconut (*Cocos nucifera*), for example, has fruits, stems, and leaves that are important sources of food, beverages, timber, and fiber. Other monocot crops provide a rich variety of spices (e.g., vanilla, cardamom), herbs (e.g., lemongrass), and beverages (e.g., plant-based milks, beer, and many grain- and sugar-based spirits). In addition, monocots provide biofuel, feedstock (e.g., palm oil, maize, sugarcane, switchgrass), timber (bamboo, *Pandanus*), and other material for housing, thatching and lawns (multiple grass species).

Monocots are also important sources of pharmaceuticals and essential oils, and they provide many attractive ornamental species, including large numbers of bulbous and cormous herbs—such as crocuses, irises, lilies, onions, and trilliums—as well as the extraordinarily diverse orchids. Some monocots are also used in culturally important ceremonies (e.g., sweetgrass use by North American indigenous peoples). Thus, monocots are arguably the most economically and socially important group of green plants (Viridiplantae).

Monocots are estimated to have originated 136–140 million years ago (Mya) (Magallon et al., 2015; Smith and Brown, 2018) and comprise about one-fourth of angiosperm species. Over the intervening time, they have evolved great diversity in ecology and growth form, including: tiny free-floating duckweeds; seagrasses; grassy, often fire-resistant herbs with parallel leaf venation; broad-leaved and gigantic herbs with net venation; resurrection plants; shrubs; vines; tall, highly lignified tree-like plants without true secondary vascular growth; tropical epiphytes; non-green mycoheterotrophs that parasitize fungi and often lurk in dense shade; and at least five species of carnivorous plants (Dahlgren et al., 1985; Kress, 1990; Givnish et al., 2005; Kress and Specht, 2005; Givnish et al., 2010; Merckx

et al., 2010; Merckx et al., 2013; Givnish et al., 2016; Lam et al., 2016; Givnish et al., 2018; Lin et al., 2021). Plants within a single order may show enormous morphological diversity, as illustrated by Asparagales, Liliales, and Pandanales (The Angiosperm Phylogeny Group et al., (APG) IV, 2016; Dahlgren et al., 1985; Kubitzki et al., 1998). Monocots also exhibit substantial diversity in the size and shape of their reproductive organs, including species with the smallest flowers (*Wolffia*), the most massive unbranched (*Amorphophallus*) and branched (*Corypha*) inflorescences, and the smallest, dustlike seeds (Orchidaceae, some less than one-millionth of a gram), and the most massive seeds (*Lodoicea*, at 18 kg). Confident resolution of monocot relationships based on multiple robust lines of evidence is a critical goal of evolutionary systematics, and is essential for understanding patterns of morphological, ecological, and geographical diversification (Givnish et al., 2018).

## Phylogeny of monocots

Recent work on relationships among monocot orders and families has been based largely on DNA sequences of plastid-encoded genes or genes extracted from whole plastid genomes (plastomes) (Graham et al., 2006; Givnish et al., 2010; Soltis et al., 2011; Steele et al., 2012; Davis et al., 2014; Govindarajulu et al., 2015; Barrett et al., 2016a; Givnish et al., 2016; Lam et al., 2018; Givnish et al., 2018). Even with whole plastome sequences, some key uncertainties regarding familial and ordinal relationships, and some plastome-based inferences conflict with those based on phylogenomics analyses of nuclear gene sequences (Sass and Specht, 2010; Zeng et al., 2014; McKain et al., 2016; Sass et al., 2016; One Thousand Plant Transcriptomes (OTPT) Initiative, 2019; Baker et al., 2022). Plastome genes are inherited as a single locus (Doyle, 2022) and plastome tree-species tree discordance may be a consequence of incomplete lineage sorting, hybridization/introgression, or misspecification of substitution models (e.g., Linder and Rieseberg, 2004; Willyard et al., 2009; Sessa et al., 2012; Davis et al., 2014; Garcia et al., 2014; Davis and Xi, 2015; Vargas et al., 2017).

In this study, we use nuclear gene sequences to resolve phylogenetic relationships among all orders and almost all families of monocots. We identify 602 nuclear genes that are conserved in single-copy form across a 12-genome dataset including nine monocots and three non-monocot outgroups. We also assessed the robustness of some inferences based on the 602 conserved single-copy (CSC) genes by comparing species trees estimated using the CSC gene set and the 1,375 Benchmarking Universal Single Copy Ortholog (BUSCO) gene set (Simão et al., 2015; Waterhouse et al., 2017). BUSCO genes are typically used for genome and transcriptome quality assessments, and increasingly extracted from genome and transcriptome data

for phylogenomic analyses in plants (Simão et al., 2015; Waterhouse et al., 2018; Manni et al., 2021; Zhao et al., 2021). Lastly, we use concordance factor analysis to more deeply explore branches that have been contentious in previous studies or that disagree with relationships based on analyses of genes extracted from complete plastomes.

## Materials and methods

### Taxon sampling, data collection, and sequencing

Our sampling included representatives of 72 of 77 recognized families of monocots (APG IV, 2016; the unsampled families are Blandfordiaceae, Corsiaceae, Juncaginaceae, Ripogonaceae, and Ruppiaceae); we analyzed 173 transcriptomes and 25 genomes, for a total of 198 taxa (Table S1). These data include 79 newly sequenced transcriptomes derived from RNA extracted from flash-frozen young leaf material (NCBI BioProjects PRJNA313089, PRJNA752894, SRP009920, PRJNA412930, and PRJNA752837). RNA was extracted following the methods described by Johnson et al. (2012). Illumina Tru-Seq libraries were constructed following the manufacturer's protocols (Illumina, San Diego, CA, USA) and sequenced on Illumina HiSeq or NextSeq 500 platforms (Table S1). Additional transcriptomes and genomes were also obtained from Phytozome (Goodstein et al., 2012), Ensembl Plants (Bolser et al., 2017), NCBI (Agarwala et al., 2017), the One Thousand Plant Transcriptomes Project (1KP) (OTPT Initiative, 2019) and other genome project databases (Tables S1–S3).

### Transcript assembly

Quality assessments of reads and adapter contamination analysis were performed using FastQC (Andrews, 2010) and any adapters were removed with Cutadapt (Martin, 2011). The reads were trimmed from the ends at positions with three consecutive bases with scores less than Q20. After trimming, reads with median quality scores less than Q22 and more than 3 uncalled bases were removed. Any read less than 40 bp in length after filtering was also removed. Cleaned reads were assembled using the Trinity v. 2013-02-25-*de novo* assembler (Haas et al., 2013). They were then aligned back to the Trinity assembly multifasta file using Bowtie (v. 0.12.8) (Langmead, 2010). RSEM v. 1.1.21 (Li and Dewey, 2011) was used to quantify the abundance of different isoforms. The assembly was then filtered to remove isoforms that had less than 1% of FPKM (Fragments Per Kilobase of transcript per Million mapped reads). Assembled transcript sequences for each species were translated using ESTScan v. 2.1 (Iseli et al., 1999), using *Oryza sativa* gene models as the training set.

## Gene-family circumscription and assignment of transcript assemblies to orthogroups

We created a PlantTribes database (Wall et al., 2008) from protein-coding sequences extracted from the annotations to enable the global identification of conserved single copy (CSC) genes across a diverse set of monocot genomes. All protein-coding gene models from nine and three published monocot and non-monocot angiosperm genomes, respectively (Table S2), were clustered using OrthoMCL (Li et al., 2003) to circumscribe orthogroups approximating gene families. OrthoMCL was run with a 1E-5 BLASTP e-value cutoff and an inflation factor of 1.2. The resulting gene family scaffold comprised 24,873 orthogroups of which 602 stringently defined single copy gene families. The 602 CSC orthogroups, with exactly one gene from each of the 12 reference genomes, were used for phylogenomic analyses.

Gene sequences from transcriptome assemblies and additional genomes were assigned to orthogroups using a combination of protein BLAST and Hidden Markov Models (HMMs) using a two-step approach. Transcript assemblies were translated using ESTScan to obtain the corresponding open reading frames (ORFs) and protein translations (Iseli et al., 1999). Hmmscan v. 3.3.2 within the HMMER package (Eddy, 2011) was then used to interrogate translated sequences for each sample with orthogroup HMM profiles. Queries of the 12-genome scaffold protein database were then conducted using BLASTp v. 2.2.26 (Altschul et al., 1990) with a threshold of 1e-5. Orthogroup assignment was based on the hmmscan results, which typically corresponded to the orthogroup that included the best BLAST hit.

Transcript assemblies and genome models assigned to the 602 putatively CSC orthogroups were inspected further. Following methods used by the One Thousand Plant Transcriptome Initiative (Matasci et al. 2014; Wickett et al., 2014; OTPT Initiative, 2019), and implemented through the AssemblyPostProcessor steps in the PlantTribes toolkit (<https://github.com/dePamphilis/PlantTribes>), if multiple transcript assemblies from a single sample were assigned to a CSC orthogroup, they were scaffolded using the banana genome (*Musa*) as a reference. If the transcript sequence overlapped with a sequence similarity of 95% or better, a consensus sequence was retained for downstream analyses. If divergence among multiple transcript assemblies for a sample sorted to a CSC orthogroup was greater than 5%, the sequences for that sample were treated as missing data for downstream analyses of that CSC orthogroup. This scaffolding process could combine splice variants into consensus sequences or treat splice variants as paralogs when they do not align well. Similarly, for the

genomes included beyond the 12 used for orthogroup construction (Table S1, S3), paralogous gene models sorted to a CSC orthogroup were also discarded. All retained transcript assemblies and scaffolds were included in multiple sequence alignments and phylogenetic analyses.

DNA and protein sequences from all taxa were brought together to create fasta files for each CSC orthogroup. Protein sequences were aligned using MAFFT v. 7.4 (Kato and Standley, 2013), trimmed using trimAl (Capella-Gutierrez et al., 2009), and then DNA sequences were forced onto the protein alignments, all using the PlantTribes GeneFamilyAligner tool (Wafula, 2019; <https://github.com/dePamphilis/PlantTribes>). A maximum of 10 alignment iterations was run; for each iteration, sites in the alignments with less than 90% occupancy or sequences with gene length less than 90% of the alignment were removed, and the remaining sequences were realigned.

Species relationships were estimated using the coalescence-based gene tree summary method implemented in ASTRAL III (Zhang et al., 2018) with default settings. Input gene trees were estimated for each of the 602 CSC orthogroup alignments using RAXML v. 8.2, with analyses partitioned by codon position as below, and a GTRGAMMA model of rate variation, with 100 rapid bootstrap replicates. TreeShrink (Mai and Mirarab, 2018) on the “per-species basis” was used to identify and filter out “rogue” taxa (that is, single genes were removed from individual taxa) that exhibited significantly greater than expected variation in placement among gene trees, possibly due to sequence error resulting in out of frame indels and mistranslation, unspliced introns, contamination, or issues with paralogy (Table S4). ASTRAL species trees were estimated from the filtered gene trees using all 602 gene trees and using filtered sets of gene trees with at least 100 or 150 taxa, respectively. Local posterior probabilities were recorded as measures of support for each branch, and the polytomy test in ASTRAL III (Sayyari and Mirarab, 2018) was also applied.

For a supermatrix analysis, CSC orthogroup alignments were concatenated into DNA and protein supermatrices using FASconCAT (Kuck and Meusemann, 2010). Phylogenetic trees were estimated from the concatenated alignment including all 602 single copy gene alignments using RAXML v. 8.2 (Stamatakis, 2014). DNA alignments were partitioned by codon position, where the first and second codon positions were made into one partition, and the third codon position was a second partition. In addition, concatenated trees were also run with gene-based partitioning, where each gene was treated as a separate partition. We used GTRGAMMA for modeling rate variation of the DNA sequences. In addition to super matrix

analyses including all 602 CSC orthogroups, analyses were performed on subsets that retained the 100 and 150 orthogroups with greatest species representation.

We further explored the placement of Asparagales and Liliales, which conflicted with plastid-based studies (see below), using, for computational efficiency, a subsample of 67 species and 1,375 BUSCO genes (Simão et al., 2015). For each taxon, only BUSCO sequences that had a single transcript were used for phylogenomic analysis, leaving missing data in places where multiple sequences were recovered from a single sample. Multiple sequence alignment and tree estimation were performed as described above. Species trees and clade support were estimated from the gene trees using ASTRAL III (Zhang et al., 2018). In order to understand how the CSC genes compared with the BUSCO sets, enrichment clustering was run with DAVID (Huang et al., 2009) for *Arabidopsis* sequences sorted to BUSCO sets and CSC sets separately. The BUSCO sets were also separated into those classified into the same or different orthogroups as the monocot conserved CSC genes.

## Concordance analysis

To further explore patterns of support and conflict for coalescent-based relationships, we calculated both gene and site “concordance factors” in IQ-TREE v. 2.2.0 (gCF and sCF, respectively; Baum, 2007; Minh et al., 2020a; Minh et al., 2020b). Branches may receive 100% bootstrap support or posterior probabilities of 1.0, yet these measures of sampling variance (Felsenstein, 1985) may obscure patterns and potential processes contributing to genealogical discordance. The gCF summarizes the proportion of ‘decisive’ individual gene trees containing a particular branch in the specified reference tree (here, the species tree inferred by ASTRAL). The sCF summarizes the average proportion of sites decisive for a particular branch in the reference tree concordant for that branch, averaged across 1000 subsampled quartets (Minh et al., 2020b). Here, ‘decisive’ denotes that a site is parsimony-informative for a particular quartet, yet decisive sites can be either concordant or discordant with a particular branch, and thus sCF represents the proportion of concordant sites relative to decisive sites. IQ-TREE 2.2.0 takes as input the reference (i.e., ASTRAL) species tree estimate, all gene trees, and all gene alignments, and produces a table with gCF, sCF, and other information for each branch, including ‘discordance factors.’ Discordance factors  $gDF_1$  and  $gDF_2$  summarize the proportion of genes concordant with the nearest-neighbor relationships of a particular branch in the reference tree, while  $gDF_p$  (‘paraphyly’) summarizes all other discordance. Further, we tested the expected pattern under a scenario of incomplete lineage sorting (ILS) using a chi-square test, with the null hypothesis being that the number of genes or sites supporting the two nearest-neighbor relationships for a

node should be roughly equal (represented by P-values for gEF and sEF). gCF and sCF were plotted along with LPP for each branch of the ASTRAL species tree estimate, using ggplot2 v.3.3.5 (Wickham, 2016).

## Results

### Transcriptomes assembly and single copy assignment

We started with a set of 4.1 billion paired-end transcript fragment reads averaging 24 million pairs of raw reads per sample. Following adapter removal and quality trimming, an average of 21.1 million pairs of reads were recovered per transcriptome and used for *de novo* assembly (Table S1). The *de novo* assembly files contained an average of 86,211 contigs + singletons (median = 75,241). These sequences (scaffolded contigs + singletons) had a mean length of 745 bases and N-50 length of 1,161 bases (medians = 715 and 1,098. Bases, respectively). An average of 60,679 (median = 58,712) coding DNA sequences and inferred protein sequences were recovered per transcriptome following translation by ESTScan. This number dropped to 57,126 contig sequences (median = 55,582) after post processing and deduplication using genome tools (Gremme et al., 2013). The mean and median N-50 lengths for these deduplicated sequences was 935 and 959 bases, respectively.

On average, 537 (median = 560) of 602 CSC orthologs were recovered per transcriptome, but after scaffolding, removing taxon-specific duplicated genes, unscaffolded alternative splice variants of unduplicated genes, and short transcripts using the PlantTribes toolkit (<https://github.com/dePamphilis/PlantTribes>), an average of 395 single copy genes per transcriptome (median = 410) were retained. Only 17 transcriptomes retained 301 or fewer single copy genes after the post-processing steps (Figure 1 and Table S1) with *Helmholtzia* retaining the fewest, with just 26 CSC gene assemblies.

### Phylogenetic inferences

The ASTRAL species trees and RAxML supermatrix trees were nearly identical as summarized in Figures 2 and S1. Both analyses yielded strong support across most of the tree. Topologies were identical at the ordinal level and nearly identical within familial levels when different stringencies of filtering (based on completeness), and or different data partitioning schemes were used, and so we focused on the presentation of results on the full nucleotide alignments and with partitions based on codon positions.

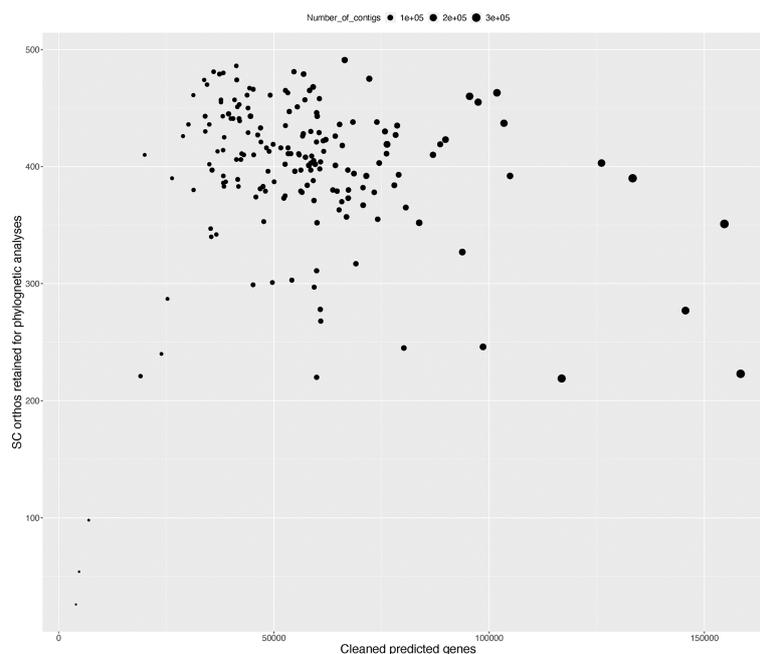


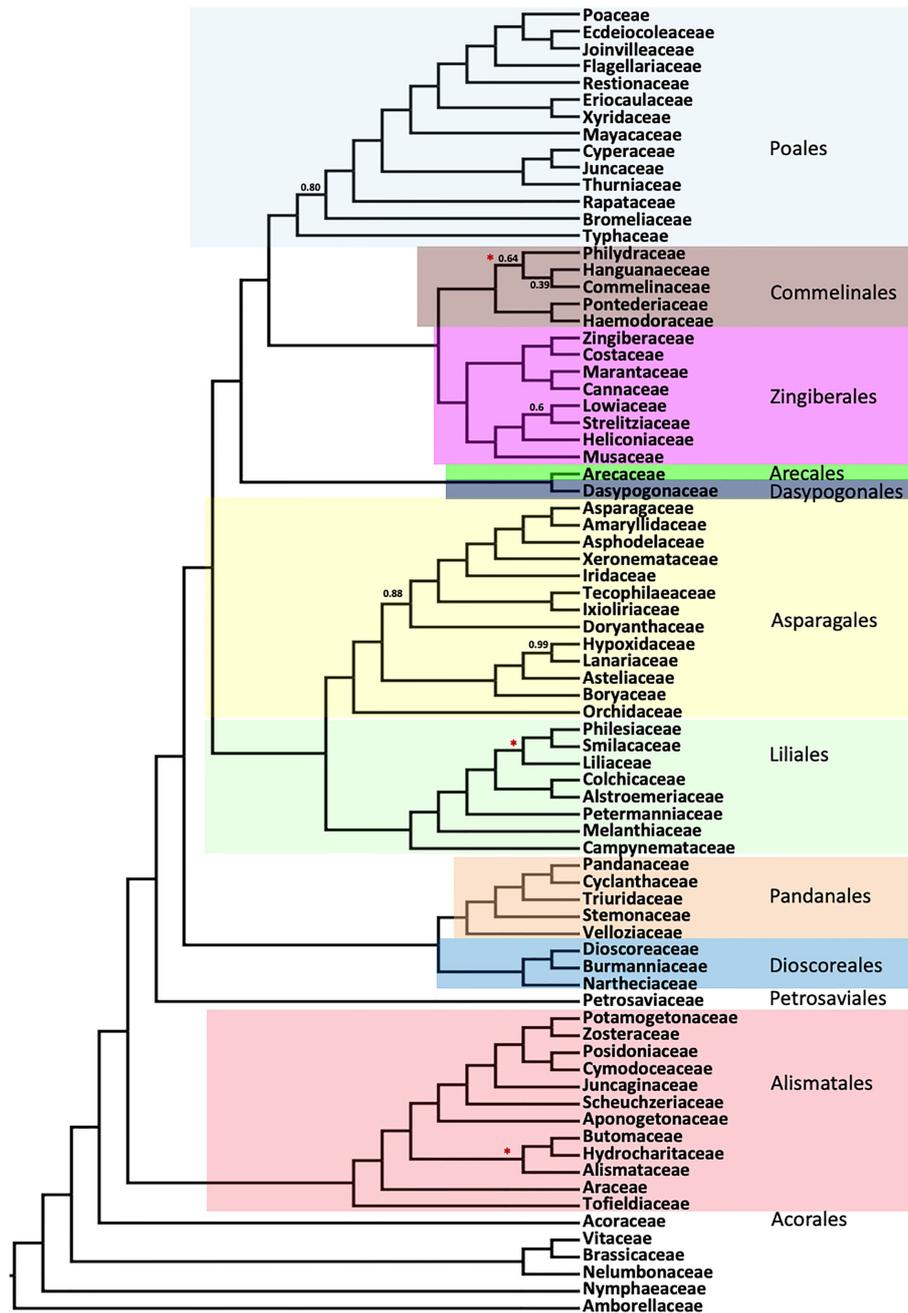
FIGURE 1

Summary of CSC orthogroup representation in transcriptomes used in the phylogenetic analysis. Figure shows the number of CSC genes retained for phylogenetic inference (Y-axis) versus the total number of translated and deduplicated genes after filtering (X-axis). Dot size corresponds to the total number of Trinity assembled contigs for each sample.

Inter-ordinal relationships within the commelinid clade are identical between the coalescent (ASTRAL) and concatenated (RAxML) analyses, with posterior probability (LPP) of 1.0 for the former and 100% bootstrap support (BS) for the latter (Figures 3, S1). Within Poales, the position of *Setaria* differs between the coalescent (Figure 3) and concatenation trees (Figure S1), though with weak support (LPP 0.01) in the former and a strong support (BS 100%) in the latter. Typhaceae are resolved as sister to a clade comprising the remainder of the order Poales with strong support. A clade comprising Commelinales and Zingiberales is sister to Poales in both the ASTRAL and supermatrix RAxML trees. Arecales and Dasygogonales comprise a clade that is sister to the rest of the commelinids. The relationships within Dasygogonales and Arecales were identical between the RAxML and ASTRAL trees.

As seen in previous species tree estimates using nuclear genes (Zeng et al., 2014; McKain et al., 2016; OTPT Initiative, 2019; Baker et al., 2022), Asparagales and Liliales formed a clade in both coalescent (ASTRAL) and concatenated supermatrix (RAxML) trees (Figure 4). In the ASTRAL tree, seven nodes in the Asparagales + Liliales clade had local posterior support values less than 0.9, while all but five nodes were fully supported in the concatenated analysis (Figure 4). There were a few topological differences between the two analyses, often at nodes that received

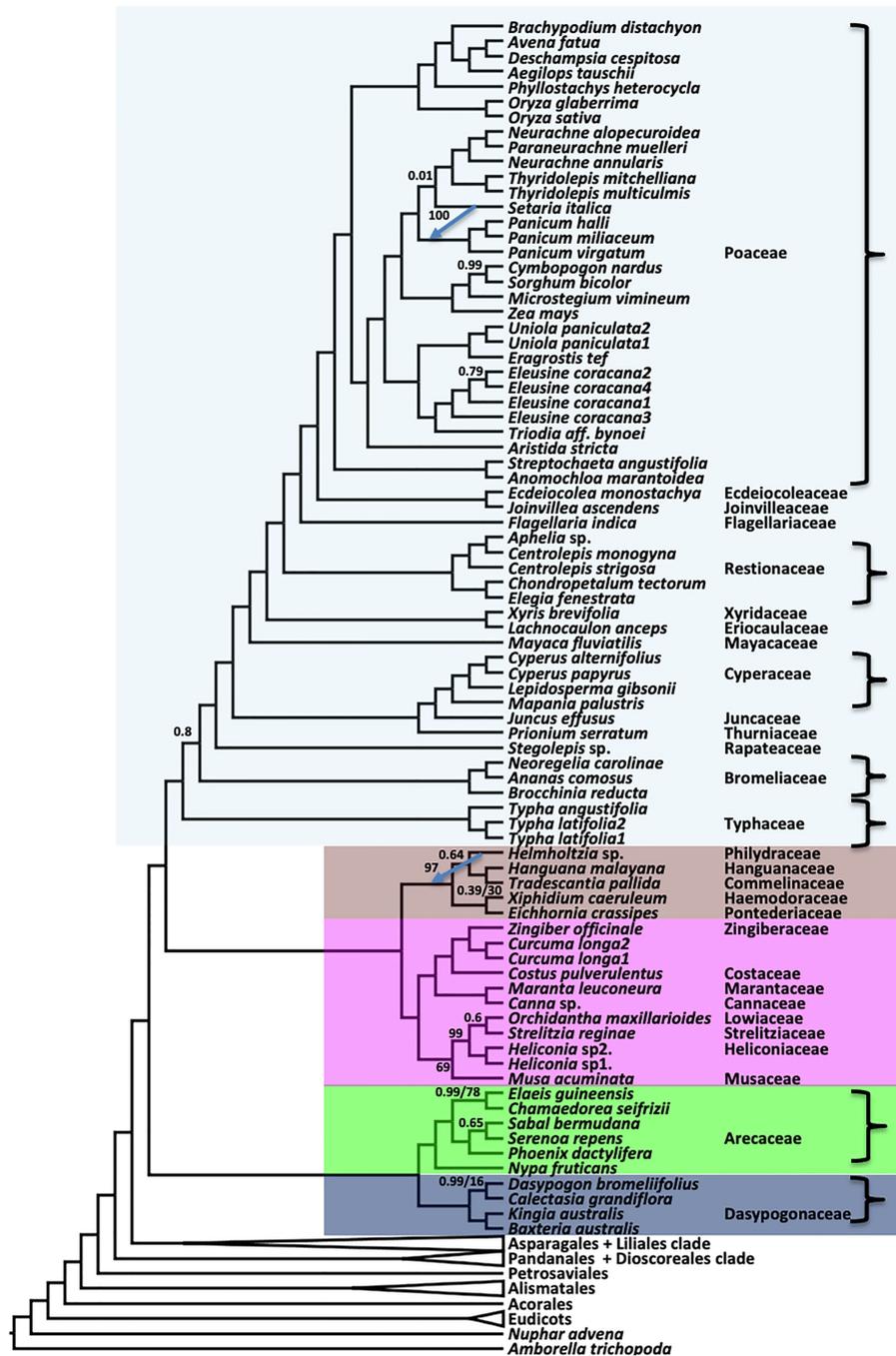
less than full support in one of the trees: (1) *Lomandra* was placed as the sister of the Asparagoideae clade (the latter including *Asparagus* and *Hemiphyllacus*) in the concatenated analysis, whereas it is sister to a larger clade in the coalescent analysis; (2) Within Asparagaceae, the positions of *Peliosanthes minor* and *Aphyllanthes monspeliensis* differ; (3) *Cypripedium* and *Selenipedium* formed a clade in the concatenated analysis, but *Cypripedium* was sister to other slipper orchids (*Phragmipedium*, *Mexipedium*, *Paphiopedilum*, and *Cypripedium*) in the ASTRAL tree; (4) The relationship among the four other orchids *Oncidium*, *Lechochilus*, *Corallorhiza*, and *Masdevallia* is also slightly different, although that relationship has BS of 0% in the concatenated analysis and LPP of 1 in the coalescent tree (5) *Smilax* and *Lilium* were sister taxa in the concatenated analysis, but *Smilax* was sister to a clade comprising *Philesia* and *Lapageria* in the ASTRAL analysis. Both the ASTRAL and concatenated analyses resolved Doryanthaceae as sister to a clade including Ixioliriaceae-Tecophilaceae, Iridaceae, Xeronemataceae, Asphodelaceae, Amaryllidaceae and Asparagaceae, with the latter seven-family clade well supported in the ASTRAL analysis (LPP 0.86) as well as in the concatenated analysis (Figure 4). Campynemataceae were resolved as the sister to the remainder of the Liliales with strong support in both analyses. The recently published analysis of Baker et al. (2022) using the Angiosperm353 bait set (Johnson et al., 2019) resolved



**FIGURE 2**  
 Summary of the order- and family-level clades inferred using ASTRAL coalescent analyses with 602 single copy genes. Numbers indicate the local posterior probability (LPP) for the main topology, for any LPP value less than 1.0. Tree topologies from RAxML concatenated analyses (Figure S1) are identical to this summary, except where a red star (“\*”) indicates a difference between the two trees (specific differences are shown in detailed Figures 3–5).

Petermanniaceae, Campynemataceae and Melanthiaceae, as successive sister clades to the remainder of the Liliales, but the LPP for the Campynemataceae + remaining Liliales clade was quite low (0.59) there. Our study (Figures 2, 4, S1) and that of Baker et al.

(2022) provide maximum support for the placement of Melanthiaceae as sister to a clade comprising all Liliales families other than Campynemataceae, whereas the plastome analysis (Givnish et al., 2018) placed Melanthiaceae sister to the following

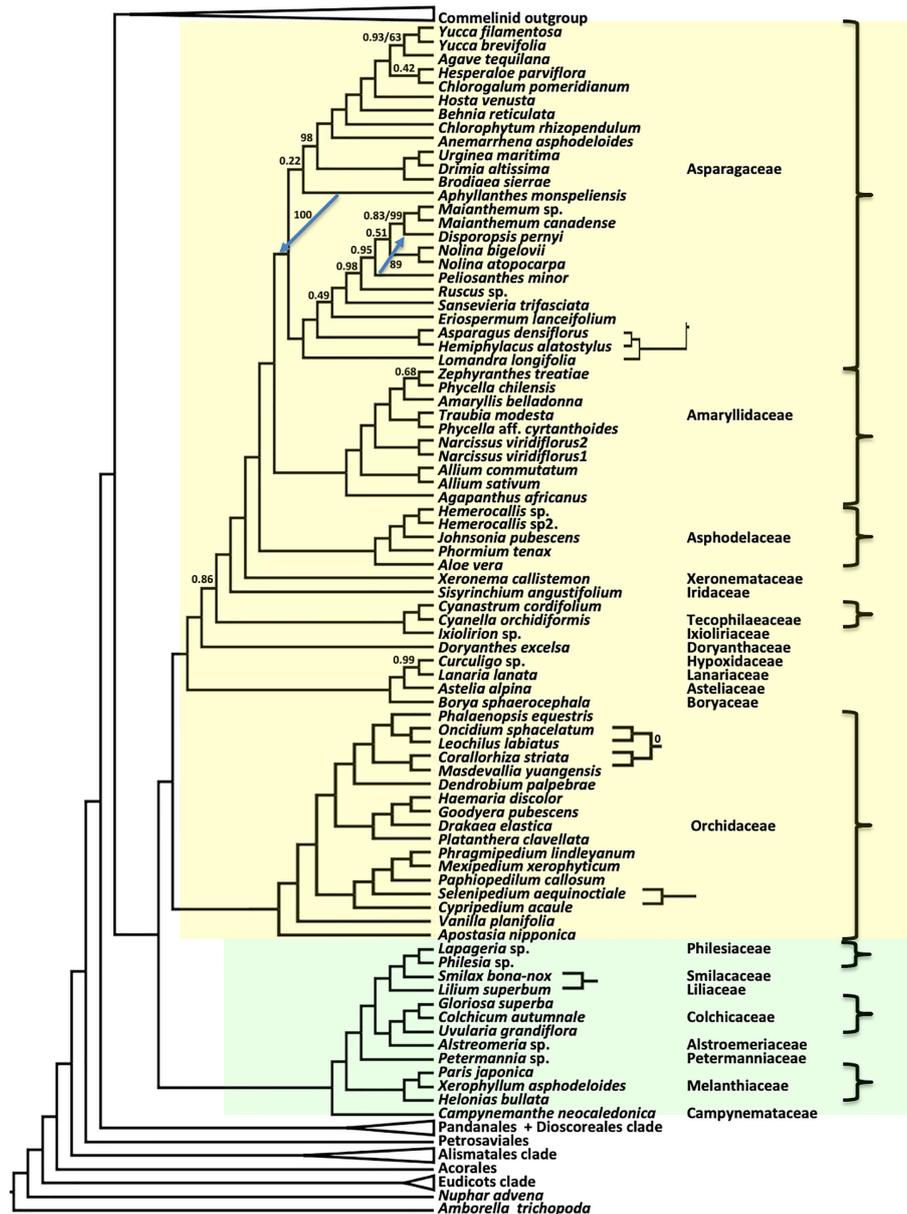


**FIGURE 3**  
 Relationships in the commelinid clade inferred using ASTRAL coalescent analyses with 602 single copy genes. RAxML topology is identical, except for the branches indicated with a blue arrow, where the tip of the arrow shows the position of the indicated branch in RAxML. Ordinal color scheme same as Figure 2. All support values have an LPP=1.0 and BS=100%, unless indicated in the diagram.

clade: (Smilacaceae, (Liliaceae, (Philesiaceae, Ripogonaceae))).

The remaining inter-ordinal and inter-familial relationships were strongly supported (all LPP of = 1.0 and all but two BS of 100%; Figure 5). The order Pandanales had identical topologies

between the ASTRAL (Figure 5) and RAxML trees (Figure S1), and only one weakly supported branch in the RAxML tree (BS of 32%) regarding the placement of Triuridaceae. Similarly, there was no difference between the two trees for Dioscoreales and the



**FIGURE 4**  
 Relationships in the Asparagales (yellow) + Liliales (green) clades inferred using ASTRAL coalescent analyses with 602 CSC genes. Ordinal color scheme same as Figure 2. RAXML topology is identical, except for four small subtrees shown on the right, and two blue arrows showing different branch positions in the RAXML tree. All support values have an LPP=1.0 and BS=100%, unless indicated in the diagram.

placement of Petrosaviales. All analyses placed Tofieldiaceae as sister to a clade comprising all other Alismatales taxa, as seen in the phylogenomic analyses of Ross et al. (2016); Baker et al. (2022) and Chen et al. (2022). Also in agreement with both plastome and nuclear gene phylogenomic analyses, the order Acorales was sister to all other monocot orders.

The BUSCO-based coalescent tree based on 1375 nuclear universal single copy orthologs was consistent with the results from the larger analyses. Analysis of the BUSCO genes resolved

the Asparagales+Liliales clade (Figure 6), and all the ordinal relationships were also identical to results from the 602 single copy gene analyses. All branches except four had local posterior probabilities of 1.0 and only one branch had weak support (LPP=0.79). The polytomy test (Sayyari and Mirarab, 2018) rejected the null hypothesis of polytomy for all but 3 nodes. Also, the positions of all families in all orders were identical to the CSC analyses (Figures 2-4) except within Poales and Asparagales.

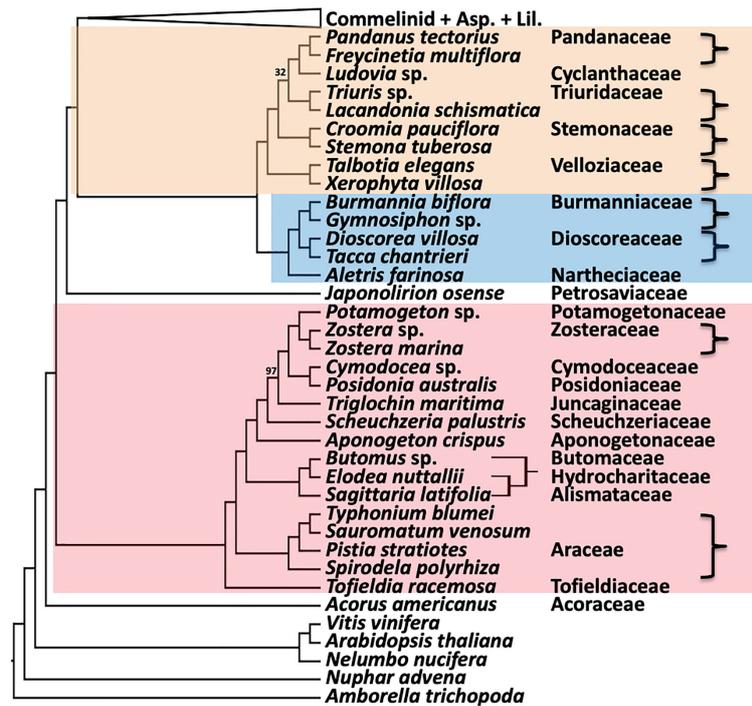


FIGURE 5

Relationships in the remaining five orders inferred using ASTRAL coalescent analyses with 602 CSC genes. Ordinal color scheme same as Figure 2. RAXML topology is identical, except for resolution of relationships among *Butomus*, *Elodea*, and *Sagittaria* (RAXML subtree shown on the right). All support values have an LPP=1.0 and BS=100%, unless indicated in the diagram.

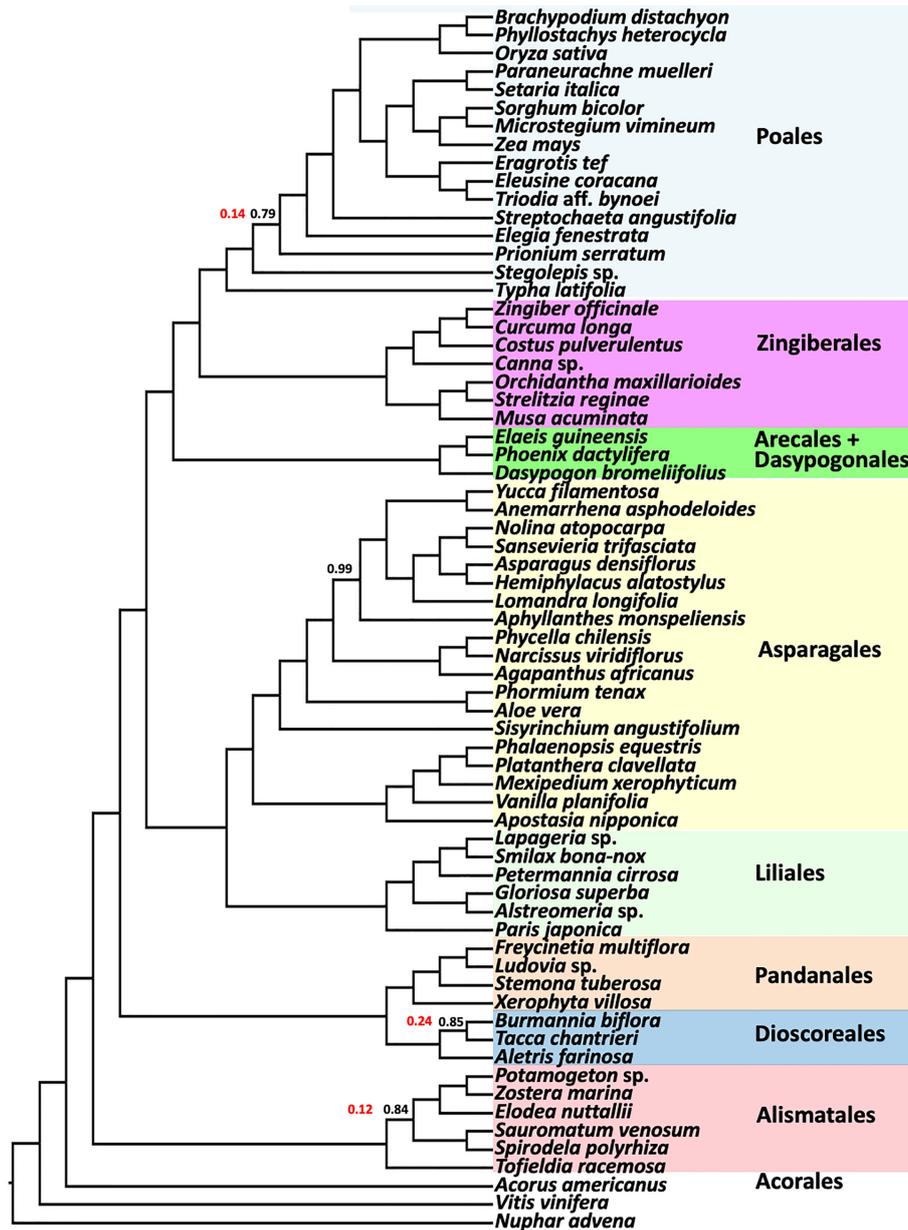
## Comparison between the CSC and BUSCO gene sets

Functional annotation clustering of CSC and BUSCO gene sets showed similarly enriched clusters between the two datasets (Tables S5, S6). The most enriched clusters contained the same uniprot keywords: Transit peptide, Chloroplast, plastid, DNA repair, DNA damage, methyl transferase, Helicase, DNA replication, DNA-binding, TPR repeat etc., indicating the photosynthetic, plastidic, and household nature of these gene sets. CSC and BUSCO gene sets had no significant differences in their enrichment patterns, meaning they were functionally indistinguishable. However, the specific genes present in the two gene sets were quite different. Overlap analysis between CSC and BUSCO showed that, out of the 1373 BUSCOs present in the *Arabidopsis* genome, 291 belonged to the CSC gene set (21.2% of 1373), and were single copy in all the scaffold genomes, while 1082 (78.8% of 1373) were not single copy genes in monocots (Tables S7, S8). The BUSCO gene sets that were not single copy had an overall average of 1.24 copies per genome, with gene numbers ranging from 0.58 to 28.9, implicating lineage-specific loss and retention of BUSCO genes following duplications. In an extreme case, as many as 59 genes were annotated in the genome of a single taxon (*Aegilops tauschii*). Our analyses revealed that

only a small fraction of BUSCO genes are actually single copy in this broad sampling of monocot genomes, while others were highly duplicated gene families. 313 genes that were exactly single copy in each of the 12 scaffold genomes (and therefore included in the CSC set) were not present in the BUSCO gene set (Table S9).

## Measurement of gene and species trees concordance and discordance

Figure 7 shows the relationship between gene concordance factor (gCF), Site concordance factor (sCF), and branch support (LPP, local posterior probability) for all internal branches of the tree inferred with ASTRAL. All branches above a gCF of ~30 and sCF ~50 had LPP of 1.0 (Table S10). However, some branches with LPP = 1.0 had low gCF and sCF values, with the lowest gCF value for a branch with LPP = 1.0 of ~15 (node 293 Table S10, Figure S3) and the lowest sCF values for a branch with LPP = 1.0 at ~27 (nodes 243, 280). Overall, 66.3% of branches had a gCF value >50, meaning more than half of all genes are concordant for that particular branch (Figure 7; Table S10). 87.7% of branches had a sCF value > 33 indicating that there is a predominant signal across sites for most branches. Internal

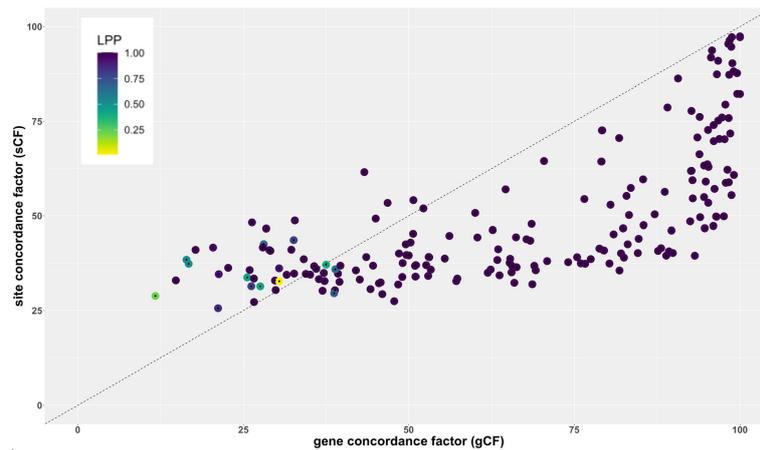


**FIGURE 6**  
 ASTRAL coalescent analysis of a subsample of 67 taxa using 1375 BUSCO genes. Ordinal color scheme same as Figure 2, showing the same relationships. All values have a LPP=1.0, except as indicated in black. Polyotomy test result p-values are shown in red for branches where the null hypothesis of a polyotomy is not rejected. Polyotomy is rejected (p-value of 0) for all other nodes.

branch lengths from the tree inferred by ASTRAL are strongly correlated with gCF and sCF (Pearson’s  $r = 0.88$ ,  $p < 0.0001$ ;  $r = 0.77$ ,  $p < 0.0001$ , respectively). Similarly, internode certainty was strongly correlated to both gene concordance factors and branch lengths (Figures S5-S6).

The branch indicating a sister relationship among Liliales and Asparagales had LPP = 1.0 but low values for gCF (22.7) and sCF (36.3). Gene discordance factors for the two nearest-neighbor

relationships for this branch were also low ( $gDF_1 = 8.2$ ,  $gDF_2 = 12.2$ ), whereas the  $gDF_p$  had a value of 56.9, indicating that over half of all gene trees decisive for this branch were discordant with the ASTRAL species tree estimate and both nearest neighbor relationships (Table S10 and Figures S3-S4). Site discordance factors for the two nearest neighbors at this branch were similar to the sCF for this branch in the ASTRAL tree, with 31.8 and 31.9% of all decisive sites being discordant. A chi-square test,



**FIGURE 7**  
Relationship between site concordance factor (sCF), gene concordance factor (gCF) and LPP in the ASTRAL-based coalescent tree used in the analyses (Figures 2–5).

however, failed to reject the null hypothesis of the pattern expected under incomplete lineage sorting (ILS) for genes and sites (gEF P-value = 0.029, sEF P-value = 0.98), underscoring the impact of rapid diversification and ILS at this branch (Table S10 and Figure S3-S4).

The branch leading to the ‘commelinid’ clade had a relatively high concordance among genes, but relatively even concordance among sites for different topologies, although the null hypothesis under ILS was rejected considering sites concordant with the two nearest-neighbor relationships (gCF = 61.7, gDF<sub>1</sub> = 0.2, gDF<sub>2</sub> = 0, gDFP = 37.94; gEF P-value = 0.3; sCF = 35, sDF<sub>1</sub> = 34.92; sDF<sub>2</sub> = 29.9; sEF P-value < 0.001). Within the commelinid clade, the branch leading to ((Zingiberales, Commelinales), Poales) received relatively low gene and site concordance factors, and the null hypothesis expected under ILS was rejected for both genes and sites (gCF = 26.6, gDF<sub>1</sub> = 18.9, gDF<sub>2</sub> = 5.8, gDFP = 48.7; gEF P-value < 0.0001; sCF = 33.4, sDF<sub>1</sub> = 36.2; sDF<sub>2</sub> = 30.3; sEF P-value < 0.001).

## Discussion

Our transcriptome-based analyses resolve and robustly support both ordinal and family-level relationships across monocot phylogeny. Aside from the strongly-supported resolution of an Asparagales+Liliales clade seen here and in other phylogenomic analyses of nuclear loci (Zeng et al., 2014; McKain et al., 2016; OTPT Initiative, 2019; Baker et al., 2022), our results support large-scale molecular analyses of monocot relationships based on plastome analyses. Notably our results corroborate inferences of Givnish et al. (Givnish et al., 2010; Givnish et al., 2018) and Barrett et al. (Barrett et al., 2013; Barrett

et al., 2016b) with respect to long-standing questions regarding relationships among commelinid monocot orders. Poales is sister to Commelinales+Zingiberales in the so-called herbaceous clade, and Arecales (Arecaceae) are sister to Dasypogonales (Dasypogonaceae) in the so-called woody clade (Givnish et al., 2018). However, while the Givnish et al. (2018) plastome analysis provided 74% bootstrap support for the sister relationship of Arecaceae and Dasypogonaceae, our evidence based on hundreds of nuclear loci strongly support that conclusion, with 1.0 LPP and 100% BS. Givnish et al. (2018) proposed that Dasypogonaceae should be recognized as order Dasypogonales (Givnish et al., 1999; Givnish et al., 2010), rather than being included in Arecales (as proposed by APG IV 2016), because the two families are highly distinctive, share few if any potential morphological synapomorphies other than a “woody” habit (making it very hard to diagnose an order containing both), and diverged earlier (119 Mya) than any other pair of sister families among the monocots.

Of the five families with placements in our nuclear phylogenies that differ from those in the plastome tree (Givnish et al., 2018), three are among those with the weakest levels of support for familial placement based on the plastome data: Tofieldiaceae (35% BS for supporting node in Givnish et al., 2018), Philydraceae (50.6% BS) and Typhaceae (62.6% BS). Each of these weakly supported nodes in the plastome phylogeny is resolved with 1.0 LPP in the current analysis, except for the placement of Philydraceae which has a LPP of 0.6 in ASTRAL tree. The poor resolution for the placement of Philydraceae are not surprising given that we only recovered 26 CSC genes in the small RNA seq dataset for *Helmholtzia*. As expected based on simulation-based experiments for phylogenomic studies (Molloy and Warnow, 2018), removing *Helmholtzia* had no impact on other inferred relationships here (Figure S6). Moreover, a recent

comprehensive phylogenomic analysis of the Commelinales using the Angiosperm353 bait set (Zuntini et al., 2021) also placed Philydraceae sister to the Hanguanaceae+Commelinaceae clade with only slightly higher support in the multispecies coalescent analysis (LPP = 0.74) and good support in the concatenated analysis (96% BS).

Concordance analysis for the placement of Tofieldiaceae as sister to the remainder of Alismatales showed that most genes were concordant (62.3%), and only a few (2.9% + 5.6% (NNI1+NNI2)) were discordant with the estimated topology. As mentioned above, other phylogenomic analyses of nuclear loci also recover strong support for Tofieldiaceae as sister to a clade including the remainder of Alismatales (Baker et al., 2022; Chen et al., 2022). Similarly, the placement of Typhaceae as sister to the remainder of Poales is supported with good gene concordance (77.5%, discordance 0.4% + 0.2%) and previous phylogenomic inference (McKain et al., 2016), although Baker et al. (2022) recovered a Typhaceae+Bromeliaceae clade using the Angiosperm353 bait set.

Interestingly, the placement of Musaceae (represented by *Musa acuminata*) as sister to a clade comprising Heliconiaceae, Lowiaceae and Strelitziaceae (LPP=1.0; BS=69%) is consistent with the plastome tree of Givnish et al. (2018) and phylogenomic analyses of nuclear genes (Carlsen et al., 2018; Baker et al., 2022; but see Sass et al., 2016), but this placement of the Musaceae had low gene concordance (28.4%) and high gene discordance (13% and 22%, for NN1 and NN2 placements, respectively) in the current analysis (Table S10). The concordance/discordance data together with the conflicting placement of Musaceae recovered by Sass et al. (2016) and earlier studies may be a consequence of reticulation in the early diversification of the Zingiberales. Relationships among the eight families of order Zingiberales have also been contentious, with studies recovering different relationships, even when employing large phylogenomic datasets based on plastomes or nuclear data (Kress et al., 2001; Kress and Specht, 2005; Barrett et al., 2014; Sass et al., 2016). Carlsen et al. (2018) did not rule out the possibility of a 'hard polytomy' at the base of Zingiberales, possibly representing a rapid, simultaneous radiation among the major lineages. Although a polytomy is rejected at the base of Zingiberales (Figure 6), quartet analysis finds no evidence to reject the null hypothesis expected under the coalescence model (ILS) for a scenario in which the major lineages of Zingiberales diverged nearly simultaneously (over a short time span).

Within Poales, we find 1.0 LPP and 100% BS and for Echeiocolaceae as sister to Joinvilleaceae, in a clade that is sister to Poaceae (Figure 4). This resolution is consistent with the previous phylotranscriptomic analysis of McKain et al. (2016) and the Angiosperm353 bait capture analysis of Baker et al. (2022), but conflicts with the most complete plastome phylogeny to date (Givnish et al., 2018), which places Echeiocolaceae as sister to Poaceae with 100% BS, and Joinvilleaceae sister to both

with 98% BS. Concordance analysis shows that 85.3% of all gene trees support resolution of the Echeiocolaceae as sister to Joinvilleaceae clade (Table S10, Figure S3-S4).

The commelinid clade is another interesting region of the monocot tree; plastomes provide moderate support for [(Zingiberales, Commelinales), Poales], [Arecales, Dasygogonales]), and nuclear loci provide overall strong support for the same relationships (Givnish et al., 2010; Barrett et al., 2013; Givnish et al., 2018). However, our test failed to reject the null hypothesis expected under a simple coalescence process (ILS) for gene counts, but strongly rejected the null hypothesis for site counts (Table S10). This suggests that while individual genes seem to fit the expectation of ILS, sites across the genome do not, possibly reflecting differences in information content among the CSC gene loci. Taking a closer look at the commelinids, the ILS test strongly rejects the null hypothesis for the clades representing [(Zingiberales, Commelinales), Poales], for both genes and sites (Table S10), whereas these relationships are strongly supported by plastomes alone (e.g. Givnish et al., 2010; Barrett et al., 2013; Givnish et al., 2018). Rejection of the expected pattern of ILS for both genes and sites may suggest an alternative explanation for conflict among these orders, for example due to ancient reticulation, or the effect of whole genome duplication and differential loss of paralogous regions (e.g., the 'sigma' event in Poales vs. the 'gamma' event in Zingiberales; D'Hont et al., 2012; McKain et al., 2016; Li et al., 2021).

Liliales and Asparagales have been recovered as successive sister lineages to the commelinid clade in several analyses of plastid genes and genomes (Chase et al., 2000; Rudall et al., 2000; Chase et al., 2006; Graham et al., 2006; Chase and Reveal, 2009; Givnish et al., 2010; Soltis et al., 2011; Givnish et al., 2018). However, very few known morphological synapomorphies separate the two clades. Dahlgren et al. (1985), segregated the Liliales from other tepaloid monocots based on introrse anthers and tepal nectaries; Asparagales were distinguished from Liliales based on the phytomelan crust covering the seeds, which is absent in the Liliales, but is also absent from most Orchidaceae and certain succulent Asparagales (Bogler and Simpson, 1995; Zomlefer, 1999). Stevens (2017) points to only two frequently reversed traits potentially supporting a clade defined by Liliales, Asparagales, and the commelinid orders: cymose inflorescence branches and protandry. The single potential morphological synapomorphy for a clade formed by Asparagales and the commelinids is more dubious: long styles. Long style is a somewhat subjective character state, and orchids have highly modified, fused columns that are variable in length.

In fact, the lilioid group of monocots is complex and highly diverse, leading to confusion about exact placements (e.g., Cronquist and Takhtadzhian, 1981; Dahlgren et al., 1985; Chase et al., 1995). Both Asparagales and Liliales exhibit diverse growth forms, but similarities in reproductive or vegetative morphology among taxa in both orders have long been noted (Dahlgren et al., 1985; Goldblatt, 1995; Rudall et al., 2000).

Dahlgren et al. (1985) considered the superorder Liliales (including families in Dioscoreales, Asparagales, and Liliales) as monophyletic, but subsequent analyses using plastid genes and genomes rejected this. All analyses using nuclear genome-scale nucleotide and amino acid sequence alignments recover a strongly supported clade comprising Liliales+Asparagales. The plastid genome is inherited as a single linkage group comparable to a genetic locus (Doyle, 2022) and the apparent conflict between nuclear and plastome phylogenomic inferences could potentially be accounted for by rapid divergence and incomplete sorting of ancestral plastome variation. Discordance, presumably due to ILS, is also seen among the nuclear gene trees.

Overall comparison of gCF and sCF values indicate that most genes individually contain low information content (Figure 7), but together contribute to a highly resolved and supported coalescent 'species tree.' The sister relationship of Liliales and Asparagales is strongly supported but differs from relationships based on recent plastome studies, which place Liliales and Asparagales as successive sister lineages to the commelinids (Davis et al., 2004; Graham et al., 2006; Givnish et al., 2010; Barrett et al., 2013; Barrett et al., 2016; Givnish et al., 2018). Analysis of sCF and gCF for the Asparagales +Liliales clade reveals a pattern that is in line with a coalescence process and ILS. Comparisons of quartet frequencies (Figure S5) are also consistent with expectations given a coalescence process with rapid diversification. The quartet frequency for the Asparagales+Liliales clade is 0.45 with similar frequency for the other two alternative resolutions, Asparagales+commelinids (0.29) and Liliales+commelinids (0.26). Therefore, the conflict between the nuclear gene-based species tree and the plastome tree (e.g. Givnish et al., 2018) is easily interpreted as random sampling of ancestral variation as the commelinid and Asparagales+Liliales lineages diverged. A recent mitochondrial genome based phylogenetic study focused on placing mycoheterotrophic lineages recovered Asparagales as sister to most monocots except Acorales and Alismatales (Lin et al., 2022); the authors speculated that this was due to sparse taxon sampling in this part of the tree.

As resolved in most previous molecular phylogenetic analyses, Dioscoreales and Pandanales form a clade and were sisters to the clade comprising commelinids and the Asparagales+Liliales clade. Relationships within Pandanales have also been controversial (Davis et al., 2004; Rudall & Bateman, 2006; Lam et al., 2015; Soto Gomez et al., 2020), perhaps due to increased substitution rates in the mycoheterotrophic Triuridaceae. The positions of Triuridaceae (Triuris, Lacandonia) and Stemonaceae (Croomia, Stemonia) with respect to Pandanaceae (Pandanus, Freycinetia), Cyclanthaceae (Ludovia), and Velloziaceae (Talbotia and Xerophyta) were the same as has been seen in combined analyses of genes encoded in the plastid and mitochondrial genomes (Soto Gomez et al., 2020; Figure 5). Quartet frequencies estimated from gene trees in the ASTRAL analysis (Figure S5) are quite similar for the placement of the Pandanaceae+Cyclanthaceae clade sister to Triuridaceae ( $Q=0.45$ ) or Stemonaceae ( $Q=0.39$ ), and the third alternative, Pandanaceae+Cyclanthaceae sister to a Stemonaceae

+Triuridaceae clade has a significantly lower quartet frequency ( $Q=0.16$ ). The skewed quartet frequencies are not expected given divergence under a coalescence model and may be due to biased gene flow after these three ancestral lineages diverged or possibly heterotachy associated with a shift from autotrophy to mycoheterotrophy in Triuridaceae. Both the ASTAL and RAxML tree estimates are also similar to published plastome trees (Givnish et al., 2018; Lam et al., 2018), which seem to have successfully placed several mycoheterotrophic taxa using plastome data despite multiple gene losses and relaxation of selection on plastid encoded photosynthetic genes (Lam et al., 2018). Mycoheterotrophic monocots have a nucleotide substitution rate for plastid genes that is  $6.9 \pm 4.1$  times that of their green sisters, with *Thismia* plastid sites evolving 364 times faster than its close relative *Tacca* (Givnish et al., 2018). Clearly, the evolution of plastomes has been strongly affected by the shift to mycoheterotrophy, which could interfere with phylogenetic inferences, unless dense taxon sampling is available and large data sets are subjected to careful analysis (Lam et al., 2018). A recent study based on slowly evolving mitochondrial genomes (Lin et al., 2022) also found the relationships among the five Pandanales families found here for the ASTRAL analysis (Figure 5), i.e., (Velloziaceae, (Stemonaceae, (Triuridaceae, (Pandanaceae, Cyclanthaceae))), but with improved support.

The placements of Petrosaviales, Alismatales and Acorales were consistent with previous phylogenomic analyses of both plastid (Lam et al., 2016; Ross et al., 2016; Givnish et al., 2018; Lam et al., 2018) and nuclear genes (Zeng et al., 2014; OTPPT Initiative, 2019). This study includes deeper sampling of Alismatales than previous phylotranscriptomic analyses including Tofieldiaceae (*Tofieldia*). The plastome phylogenies reported in Ross et al. (2016); Givnish et al. (2018), and Lam et al. (2018) generally had poor to moderate support for either Araceae or Tofieldiaceae as sister to the rest of the order (e.g., maximum of 76% support for Tofieldiaceae sister in Ross et al., 2016). In Givnish et al. (2018), the position of Tofieldiaceae had the weakest support of any family in the plastid phylogeny with a bootstrap support of less than 50% for being sister to all alismatids except Araceae. The plastome analysis by Givnish et al. (2018) indicates that the branches involved are very short, and very deep: the inferred stem age of Araceae was 123.96 Mya, and 123.56 Mya for Tofieldiaceae and the clade formed by the remaining 11 Alismatales families. Nonetheless, our analyses return strong support for this placement of Tofieldiaceae as sister to the remaining Alismatales.

Finally, the vast majority of BUSCO genes are not strictly single copy in monocots, suggesting that these genes may return to single copy following duplication more slowly than the strictly single copy gene set, increasing the chance of orthology misspecification with BUSCO genes. Nevertheless, BUSCO trees in this study were largely congruent with those based on CSC genes, and both gene sets have indistinguishable functional biases, suggesting that both are samples of a larger gene set that

can both provide similarly strong evidence for phylogenomic analyses. Key data handling steps for both data sets were the removal of genes from any taxon that had more than a single gene, or were identified as “rogue” taxa based on their unusually long branch lengths, suggesting that these steps alone may minimize orthology misspecification.

## Data availability statement

The data presented in the study are deposited in the NCBI’s SRA repository under BioProject accessions PRJNA313089, PRJNA752894, SRP009920, PRJNA412930, and PRJNA752837; SRA accession IDs for each sample are reported in Table S1.

## Author contributions

PT, CdP, JL-M, TG, DS, CA, JP, JD, WZ, SG, and CB, contributed to conception and design of the study. DS, JL-M, JP, CdP, JM, JR, MM, KH, AH, MV, JC, NI, and BF performed fieldwork, obtained samples, and processed samples for transcriptome analysis. SA, JL-M, PT, EW, and CdP organized the database. PT, CdP, JL-M, EW, and CB performed analyses and created graphics. PT and CdP wrote the first draft of the manuscript. JL-M, CB, CA, TG, and JD wrote sections of the manuscript. All authors contributed to the article and approved the submitted version.

## Acknowledgments

Data generation was performed under the Monocot Tree of Life project (MonAToL), DEB-0829868, at Cold Spring Harbor

## References

- Agarwala, R., Barrett, T., Beck, J., Benson, D. A., Bollin, C., Bolton, E., et al. (2017). Database resources of the national center for biotechnology. *Nucleic Acids Res.* 45, D12–D17. doi: 10.1093/nar/gkw1071
- Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410. doi: 10.1016/S0022-2836(05)80360-2
- Andrews, S. (2010) *FastQC: A quality control tool for high throughput sequence data*. Available at: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/> (Accessed 01/20/2022).
- Baker, W. J., Bailey, P., Barber, V., Barker, A., Bellot, S., Bishop, D., et al. (2022). A comprehensive phylogenomic platform for exploring the angiosperm tree of life. *Syst. Biol.* 71, 301–319. doi: 10.1093/sysbio/syab035
- Barrett, C. F., Specht, C. F., Leebens-Mack, J. H., Stevenson, D., Zomlefer, W. B., Davis, J. I., et al. (2014). Resolving ancient radiations: can complete plastid gene sets elucidate deep relationships among the tropical ginger (Zingiberales)? *Ann. Bot.* 113, 119–133. doi: 10.1093/aob/mct264
- Barrett, C. F., Bacon, C. D., Antonelli, A., Cano, A., and Hofmann, T. (2016a). An introduction to plant phylogenomics with a focus on palms. *Bot. J. Linn. Soc.* 182, 234–255. doi: 10.1111/boj.12399
- Barrett, C. F., Baker, W. J., Comer, J. R., Conran, J. G., Lahmeyer, S. C., Leebens-Mack, J. H., et al. (2016b). Plastid genomes reveal support for deep phylogenetic

Laboratories, University of Georgia, and Penn State University. We thank Sarah Johnson, Riva Bruenn, Nina Hobbhahn, and Peter Linder for tissue samples, and Lisa DeGironimo, Chang Liu, Charlotte Quigley, and Paula Ralph for assisting with RNA isolations. We thank Norman Wickett for early discussions about this work. PT and EW and computer resources for this paper were supported in part by DEB-0829868 and IOS-1238057, and by the Huck Institutes of the Life Sciences and Department of Biology at Penn State University. We also thank three reviewers for their helpful comments.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher’s note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## Supplementary Material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2022.876779/full#supplementary-material>

- relationships and extensive rate variation among palms and other commelinid monocots. *New Phytol.* 209, 855–870. doi: 10.1111/nph.13617
- Barrett, C. F., Davis, J. I., Leebens-Mack, J., Conran, J. G., and Stevenson, D. W. (2013). Plastid genomes and deep relationships among the commelinid monocot angiosperms. *Cladistics* 29, 65–87. doi: 10.1111/j.1096-0031.2012.00418.x
- Baum, D. A. (2007). Concordance trees, concordance factors, and the exploration of reticulate genealogy. *Taxon* 56 (2), 417–426. doi: 10.1002/tax.562013
- Bogler, D. J., and Simpson, B. B. (1995). A chloroplast DNA study of the agavaceae. *Syst. Bot.* 20, 191–205. doi: 10.2307/2419449
- Bolser, D. M., Staines, D. M., Perry, E., and Kersey, P. J. (2017). Ensembl plants: Integrating tools for visualizing, mining, and analyzing plant genomic data. *Plant Genomics Databases: Methods Protoc.* 1533, 1–31. doi: 10.1007/978-1-4939-6658-5\_1
- Bremer, B., Bremer, K., Chase, M. W., Fay, M. F., Reveal, J. L., Soltis, D. E., et al. (2009). An update of the angiosperm phylogeny group classification for the orders and families of flowering plants: APG III. *Bot. J. Linn. Soc.* 161, 105–121. doi: 10.1111/j.1095-8339.2009.00996.x
- Capella-Gutierrez, S., Silla-Martinez, J. M., and Gabaldon, T. (2009). trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25, 1972–1973. doi: 10.1093/bioinformatics/btp348

- Carlsen, M. M., Fér, T., Schmickl, R., Leong-Skorničková, J., Newman, M., and Kress, W. J. (2018). Resolving the rapid plant radiation of early diverging lineages in the tropical zingiberales: pushing the limits of genomic data. *Mol. Phylogenet. Evol.* 128, 55–68. doi: 10.1016/j.ympev.2018.07.020
- Chase, M. W., Duvall, M. R., Hills, H. G., Conran, J. G., Cox, A. V., Eguiarte, L. E., et al. (1995). "Molecular phylogenetics of Liliaceae", in *Monocotyledons: systematics and evolution* P. J. Rudall, P. J. Cribb, D. F. Cutler and C. J. Humphries eds. (Kew, Richmond, Surrey, UK: Royal Botanic Gardens), 109–137.
- Chase, M. W., and Reveal, J. L. (2009). A phylogenetic classification of the land plants to accompany APG III. *Bot. J. Linn. Soc.* 161, 122–127. doi: 10.1111/j.1095-8339.2009.01002.x
- Chase, M. W., Fay, M. F., Devey, D. S., Maurin, O., Ronsted, N., Davies, T. J., et al. (2006). Multigene analyses of monocot relationships. *Aliso: A Journal of Systematic and Floristic Botany* 22(1), 63–75.
- Chase, M. W., Soltis, D. E., Soltis, P. S., Rudall, P. J., Fay, M. F., Hahn, W. H., et al. (2000). "Higher-level systematics of the monocotyledons: an assessment of current knowledge of a new classification," in *Monocots: systematics and evolution*. Eds. K. L. Wilson and D. A. Morrison (Collingwood, Victoria, Australia: CSIRO), 3–16.
- Chen, L.-Y., Lu, B., Morales-Briones, D. F., Moody, M. L., Liu, F., Hu, G.-W., et al. (2022). Phylogenomic analyses of alismatales shed light into adaptations to aquatic environments. *Mol. Biol. Evol.* 39. doi: 10.1093/molbev/msac079
- Cronquist, A., and Takhtadzian, A. L. (1981). *An integrated system of classification of flowering plants* (New York, USA: Columbia University Press).
- Dahlgren, R. M. T., Clifford, H. T., and Yeo, P. F. (1985). *The families of monocotyledon* (Berlin: Springer-Verlag).
- Davis, J. I., Mcneal, J. R., Barrett, C. F., Chase, M. W., Cohen, J. I., Duvall, M. R., et al. (2013). "Contrasting patterns of support among plastid genes and genomes for major clades of the monocotyledons," in *Early events in monocot evolution. systematics association special volume series*. Eds. P. Wilkin and S. J. Mayo (Cambridge, UK: Cambridge University Press), 315–349.
- Davis, J. I., Stevenson, D. W., Petersen, G., Seberg, O., Campbell, L. M., Freudenstein, J. V., et al. (2004). A phylogeny of the monocots, as inferred from *rbcl* and *atpA* sequence variation, and a comparison of methods for calculating jackknife and bootstrap values. *Systematic Botany* 39 (3), 467–510.
- Davis, C. C., and Xi, Z. (2015). Horizontal gene transfer in parasitic plants. *Cur. Opin. Plant Biol.* 26, 14–19. doi: 10.1016/j.pbi.2015.05.008
- Davis, C. C., Xi, Z. X., and Mathews, S. (2014). Plastid phylogenomics and green plant phylogeny: almost full circle but not quite there. *BMC Biol.* 12, 11. doi: 10.1186/1741-7007-12-11
- D'Hont, A., Denoeuf, F., Aury, J. M., Baurens, F. C., Carreel, F., Garsmeur, O., et al. (2012). The banana (*Musa acuminata*) genome and the evolution of monocotyledonous plants. *Nature* 488 (7410), 213–217. doi: 10.1038/nature11241
- Doyle, J. J. (2022). Defining coalescent genes: theory meets practice in organelle phylogenomics. *Syst. Biol.* 71, 476–489. doi: 10.1093/sysbio/syab053
- Eddy, S. R. (2011). Accelerated profile HMM searches. *PLoS Comput. Biol.* 7, e1002195. doi: 10.1371/journal.pcbi.1002195
- Felsenstein, J. (1985). Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* 39, 783–791. doi: 10.1111/j.1558-5646.1985.tb00420.x
- Garcia, N., Meerow, A. W., Soltis, D. E., and Soltis, P. S. (2014). Testing deep reticulate evolution in amarillidaceae tribe hippeastreae (Asparagales) with ITS and chloroplast sequence data. *Syst. Bot.* 39, 75–89. doi: 10.1600/03636441X678099
- Givnish, T. J., Ames, M., Mcneal, J. R., Mckain, M. R., Steele, P. R., Depamphilis, C. W., et al. (2010). Assembling the tree of the monocotyledons: plastome sequence phylogeny and evolution of poales. *Ann. Missouri Bot. Gard.* 97, 584–616. doi: 10.3417/2010023
- Givnish, T. J., Evans, T. M., Pires, J. C., and Sytma, K. J. (1999). Polyphyly and convergent morphological evolution in commelinales and commelinidae: evidence from *rbcl* sequence data. *Mol. Phylogenet.* 12 (3), 360–385. doi: 10.1006/mpev.1999.0601
- Givnish, T. J., Pires, J. C., Graham, S. W., McPherson, M. A., Prince, L. M., Patterson, T. B., et al. (2005). Repeated evolution of net venation and fleshy fruits among monocots in shaded habitats confirms *a priori* predictions: evidence from an *ndhF* phylogeny. *Proc. Biol. Sci.* 272, 1481–1490. doi: 10.1098/rspb.2005.3067
- Givnish, T. J., Zuluaga, A., Marques, I., Lam, V. K. Y., Gomez, M. S., Iles, W. J. D., et al. (2016). Phylogenomics and historical biogeography of the monocot order liliales: out of Australia and through Antarctica. *Cladistics* 32, 581–605. doi: 10.1111/cla.12153
- Givnish, T. J., Zuluaga, A., Spalink, D., Soto Gomez, M., Lam, V. K. Y., Saarela, J. M., et al. (2018). Monocot plastid phylogenomics, timeline, net rates of species diversification, the power of multi-gene analyses, and a functional model for the origin of monocots. *Am. J. Bot.* 105, 1888–1910. doi: 10.1002/ajb2.1178
- Goldblatt, P. (1995). "The status of r. dahlgren's order liliales and melanthiales," in *Monocotyledons: systematics and evolution*. Eds. P. J. Rudall, P. J. Cribb, D. F. Cutler and C. J. Humphries (Royal Botanic Gardens, Kew, Richmond, Surrey, UK), 181–200.
- Goodstein, D. M., Shu, S. Q., Howson, R., Neupane, R., Hayes, R. D., Fazo, J., et al. (2012). Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Res.* 40, D1178–D1186. doi: 10.1093/nar/gkr944
- Govindarajulu, R., Parks, M., Tennessen, J. A., Liston, A., and Ashman, T. L. (2015). Comparison of nuclear, plastid, and mitochondrial phylogenies and the origin of wild octoploid strawberry species. *Am. J. Bot.* 102, 544–554. doi: 10.3732/ajb.1500026
- Graham, S. W., Zgurski, J. M., McPherson, M. A., Cherniawsky, D. M., Saarela, J. M., Horne, E. F. C., et al. (2006). "Robust inference of monocot deep phylogeny using an expanded multigene plastid data set," in *Monocots: Comparative biology and evolution (excluding poales)*. Eds. J. T. Columbus, E. A. Friar, J. M. Porter, L. M. Prince and M. G. Simpson (Rancho Santa Ana Botanic Garden, Claremont, California, USA), 3–21.
- Gremme, G., Steinbiss, S., and Kurtz, S. (2013). GenomeTools: a comprehensive software library for efficient processing of structured genome annotations. *IEEE/ACM Trans. Comput. Biol. Bioinform.* 10, 645–656.
- Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J., et al. (2013). *De novo* transcript sequence reconstruction from RNA-seq using the trinity platform for reference generation and analysis. *Nat. Protoc.* 8, 1494–1512. doi: 10.1038/nprot.2013.084
- Huang, D. W., Sherman, B. T., and Lempicki, R. A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* 4, 44–57. doi: 10.1038/nprot.2008.211
- Iseli, C., Jongeneel, C. V., and Bucher, P. (1999). ESTScan: a program for detecting, evaluating, and reconstructing potential coding regions in EST sequences. *Proc. - Int. Conf. Intell. Syst. Mol. Biol.* 1999, 138–148.
- Johnson, M. T., Carpenter, E. J., Tian, Z., Bruskiwich, R., Burris, J. N., Carrigan, C. T., et al. (2012). Evaluating methods for isolating total RNA and predicting the success of sequencing phylogenetically diverse plant transcriptomes. *PLoS One* 7 (11), e50226. doi: 10.1371/journal.pone.0050226
- Johnson, M. G., Pokorny, L., Dodsworth, S., Botigué, L. R., Cowan, R. S., Devault, A., et al. (2019). A universal probe set for targeted sequencing of 353 nuclear genes from any flowering plant designed using k-medoids clustering. *Syst. Biol.* 68, 594–606. doi: 10.1093/sysbio/syy086
- Katoh, K., and Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* 30, 772–780. doi: 10.1093/molbev/mst010
- Kress, W. J. (1990). The phylogeny and classification of zingiberales. *Ann. Missouri Bot. Gard.* 77, 698–721. doi: 10.2307/2399669
- Kress, W. J., Prince, L. M., Hahn, W. J., and Zimmer, E. A. (2001). Unraveling the evolutionary radiation of the families of the zingiberales using morphological and molecular evidence. *Syst. Biol.* 50, 926–944. doi: 10.1080/106351501753462885
- Kress, W. J., and Specht, C. D. (2005). Between cancer and Capricorn: phylogeny, evolution and ecology of the primarily tropical zingiberales. *Kongelige Danske Videnskabernes Selskab Biologiske Skrifter* 55, 459–478.
- Kubitzki, K., Rudall, P. J., and Chase, M. C. (1998). "Systematics and evolution," in *Flowering plants - monocotyledons. the families and genera of vascular plants*, vol. 3. Ed. K. Kubitzki (Germany: Springer, Berlin, Heidelberg). doi: 10.1007/978-3-662-03533-7\_3
- Kuck, P., and Meusemann, K. (2010). FASconCAT: convenient handling of data matrices. *Mol. Phylogenet. Evol.* 56, 1115–1118. doi: 10.1016/j.ympev.2010.04.024
- Lam, V. K. Y., Soto Gomez, M., and Graham, S. W. (2015). The highly reduced plastome of mycoheterotrophic Sciaphila (Triuridaceae) is colinear with its green relatives and is under strong purifying selection. *Genome Biol. Evol.* 7 (8), pp.2220–2236. doi: 10.1002/ajb2.1070
- Lam, V. K. Y., Darby, H., Merckx, V., Lim, G., Yukawa, T., Neubig, K. M., et al. (2018). Phylogenomic inference in *extremis*: a case study with mycoheterotroph plastomes. *Am. J. Bot.* 105, 480–494. doi: 10.1002/ajb2.1070
- Lam, V. K. Y., Merckx, V. S. F. T., and Graham, S. W. (2016). A few-gene plastid phylogenetic framework for mycoheterotrophic monocots. *Am. J. Bot.* 103, 692–708. doi: 10.3732/ajb.1500412
- Langmead, B. (2010). Aligning short sequencing reads with bowtie. *Curr. Protoc. Bioinf. Chapter 11 Unit 11*, 17. doi: 10.1002/0471250953.bi1107s32
- Li, B., and Dewey, C. N. (2011). RSEM: accurate transcript quantification from RNA-seq data with or without a reference genome. *BMC Bioinform.* 12, 323. doi: 10.1186/1471-2105-12-323
- Lin, Q., Ane, C., Givnish, T. J., and Graham, S. W. (2021). A new carnivorous plant lineage (*Triantha*) with a unique sticky-influorescence trap. *Proc. Natl. Acad. Sci. U.S.A.* 118, 33. doi: 10.1073/pnas.2022724118
- Lin, Q., Braukmann, T. W. A., Soto Gomez, M., Mayer, J. L. S., Pinheiro, F., Merckx, V. S. F. T., et al. (2022). Mitochondrial genomic data are effective at

- placing mycoheterotrophic lineages in plant phylogeny. *New Phytol.* doi: 10.1111/nph.18335
- Linder, C. R., and Rieseberg, L. H. (2004). Reconstructing patterns of reticulate evolution in plants. *Am. J. Bot.* 91 (10), 1700–1708. doi: 10.3732/ajb.91.10.1700
- Li, L., Stoeckert, C. J. Jr., and Roos, D. S. (2003). OrthoMCL: Identification of ortholog groups for eukaryotic genomes. *Genome Res.* 13, 2178–2189. doi: 10.1101/gr.1224503
- Li, H. L., Wu, L., Dong, Z., Jiang, Y., Jiang, S., Xing, H., et al. (2021). Haplotype-resolved genome of diploid ginger (*Zingiber officinale*) and its unique gingerol biosynthetic pathway. *Hortic. Res.* 8, 189. doi: 10.1038/s41438-021-00627-7
- Lughadha, E. N., Govaerts, R., Belyaeva, I., Black, N., Lindon, H., Allkin, R., et al. (2016). Counting counts: Revised estimates of numbers of accepted species of flowering plants, seed plants, vascular plants and land plants with a review of other recent estimates. *Phytotaxa* 272 (1), 82–88. doi: 10.11646/phytotaxa.272.1.5
- Mabberley, D. J. (2008). *Mabberley's plant-book: A portable dictionary of plants, their classifications and uses* (No. ed. 3) (Cambridge, UK: Cambridge University Press).
- Magallon, S., Gomez-Acevedo, S., Sanchez-Reyes, L. L., and Hernandez-Hernandez, T. (2015). A metacalibrated time-tree documents the early rise of flowering plant phylogenetic diversity. *New Phytol.* 207, 437–453. doi: 10.1111/nph.13264
- Mai, U., and Mirarab, S. (2018). TreeShrink: fast and accurate detection of outlier long branches in collections of phylogenetic trees. *BMC Genomics* 19, 272. doi: 10.1186/s12864-018-4620-2
- Manni, M., Berkeley, M. R., Seppey, M., Simão, F. A., and Zdobnov, E. M. (2021). BUSCO update: Novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. *Mol. Biol. Evol.* 38, 4647–4654. doi: 10.1093/molbev/msab199
- Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. J.* 17, 10–12. doi: 10.14806/ej.17.1.200
- Matasci, N., Hung, L. H., Yan, Z., Carpenter, E. J., Wickett, N. J., Mirarab, S., et al. (2014). Data access for the 1,000 plants (1KP) project. *Gigascience* 3, 17. doi: 10.1186/2047-217X-3-17
- McKain, M. R., Tang, H., Mcneal, J. R., Ayyampalayam, S., Davis, J. I., Depamphilis, C. W., et al. (2016). A phylogenomic assessment of ancient polyploidy and genome evolution across the poales. *Genome Biol. Evol.* 8, 1150–1164. doi: 10.1093/gbe/evw060
- Merckx, V. S. F. T., Mennes, C. B., Peay, K. G., and Geml, J. (2013). “Evolution and diversification,” in *Mycoheterotrophy the biology of plants living on fungi*, vol. 356. Ed. V. S. F. T. Merckx (New York, NY: Springer New York : Imprint: Springer), 377.
- Merckx, V. S. F. T., Stockel, M., Fleischmann, A., Bruns, T. D., and Gebauer, G. (2010). 15N and 13C natural abundance of two mycoheterotrophic and a putative partially mycoheterotrophic species associated with arbuscular mycorrhizal fungi. *New Phytol.* 188, 590–596. doi: 10.1111/j.1469-8137.2010.03365.x
- Minh, B. Q., Hahn, M. W., and Lanfear, R. (2020a). New methods to calculate concordance factors for phylogenomic datasets. *Mol. Biol. Evol.* 37 (9), 2727–2733. doi: 10.1093/molbev/msaa106
- Minh, B. Q., Schmidt, H. A., Schrepf, O. D., Woodhams, M. D., von Haeseler, A., and Lanfear, R. (2020b). IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol.* 37, 1530–1534. doi: 10.1093/molbev/msaa015
- Molloy, E. K., and Warnow, T. (2018). To include or not to include: The impact of gene filtering on species tree estimation methods. *Syst. Biol.* 67, 285–303. doi: 10.1093/sysbio/syx077
- One Thousand Plant Transcriptomes Initiative (2019). One thousand plant transcriptomes and the phylogenomics of green plants. *Nature* 574, 679–685. doi: 10.1038/s41586-019-1693-2
- Ross, T. G., Barrett, C. F., Soto Gomez, M., Lam, V. K. Y., Henriquez, C. L., Les, D. H., et al. (2016). Plastid phylogenomics and molecular evolution of alismatales. *Cladistics* 32, 160–178. doi: 10.1111/cla.12133
- Rudall, P. J., and Bateman, R. M. (2006). Morphological phylogenetic analysis of pandanales: testing contrasting hypotheses of floral evolution. *Systematic Bot.* 31 (2), 223–238. doi: 10.1600/03636440677585766
- Rudall, P. J., Stobart, K. L., Hong, W. P., Conran, J. G., Furness, C. A., Kite, G. C., et al. (2000). “Consider the lilies: Systematics of liliales,” in *Monocots: systematics and evolution*. Eds. K. L. Wilson and D. A. Morrison (Collingwood, Victoria, Australia: CSIRO), 347–359.
- Sass, C., Iles, W. J., Barrett, C. F., Smith, S. Y., and Specht, C. D. (2016). Revisiting the zingiberales: using multiplexed exon capture to resolve ancient and recent phylogenetic splits in a charismatic plant lineage. *PeerJ* 4, e1584. doi: 10.7717/peerj.1584
- Sass, C., and Specht, C. D. (2010). Phylogenetic estimation of the core bromelioids with an emphasis on the genus *Aechmea* (Bromeliaceae). *Mol. Phylogenet. Evol.* 55, 559–571. doi: 10.1016/j.ympev.2010.01.005
- Sayyari, E., and Mirarab, S. (2018). Testing for polytomies in phylogenetic species trees using quartet frequencies. *Genes (Basel)* 9 (3), 132. doi: 10.3390/genes9030132
- Sessa, E. B., Zimmer, E. A., and Givnish, T. J. (2012). Reticulate evolution on a global scale: A nuclear phylogeny for new world *Dryopteris* (Dryopteridaceae). *Mol. Phylogenet. Evol.* 64 (3), 563–581. doi: 10.1016/j.ympev.2012.05.009
- Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., and Zdobnov, E. M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31, 3210–3212. doi: 10.1093/bioinformatics/btv351
- Smith, S. A., and Brown, J. W. (2018). Constructing a broadly inclusive seed plant phylogeny. *Am. J. Bot.* 105, 302–314. doi: 10.1002/ajb2.1019
- Soltis, D. E., Smith, S. A., Cellinese, N., Wurdack, K. J., Tank, D. C., Brockington, S. F., et al. (2011). Angiosperm phylogeny: 17 genes, 640 taxa. *Am. J. Bot.* 98, 704–730. doi: 10.3732/ajb.1000404
- Soto Gomez, M., Lin, Q., da Silva Leal, E., Gallaher, T. J., Scherberich, D., Mennes, C. B., et al. (2020). A bi-organellar phylogenomic study of pandanales: inference of higher-order relationships and unusual rate-variation patterns. *Cladistics* 36 (5), 481–504. doi: 10.1111/cla.12417
- Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinf. (Oxford England)* 30, 1312–1313. doi: 10.1093/bioinformatics/btu033
- Steele, P. R., Hertweck, K. L., Mayfield, D., Mckain, M. R., Leebens-Mack, J., and Pires, J. C. (2012). Quality and quantity of data recovered from massively parallel sequencing: examples in asparagales and poaceae. *Am. J. Bot.* 99, 330–348. doi: 10.3732/ajb.1100491
- Stevens, P. F. (2017). *Angiosperm phylogeny website. version 13. angiosperm phylogeny website. version 14.* Available at: <http://www.mobot.org/MOBOT/research/APweb/>
- The Angiosperm Phylogeny Group, Chase, M. W., Christenhusz, M. J. M., Fay, M. F., Byng, J. W., Judd, W. S., et al. (2016). An update of the angiosperm phylogeny group classification for the orders and families of flowering plants: APG IV. *Botanical J. Lin Soc* 181, 1–20. doi: 10.1111/boj.12385
- Vargas, O. M., Ortiz, E. M., and Simpson, B. B. (2017). Conflicting phylogenomic signals reveal a pattern of reticulate evolution in a recent high-Andean diversification (Asteraceae: Astereae: *Diplostephium*). *New Phytol.* 214 (4), 1736–1750. doi: 10.1111/nph.14530
- Wafula, E. K. (2019). *Computational methods for comparative genomics of non-model species: a case study in the parasitic plant family orobanchaceae* (PhD Dissertation. University Park (PA: The Pennsylvania State University)).
- Wall, P. K., Leebens-Mack, J., Muller, K. F., Field, D., Altman, N. S., and dePamphilis, C. W. (2008). PlantTribes: A gene and gene family resource for comparative genomics in plants. *Nucleic Acids Res.* 36, D970–D976. doi: 10.1093/nar/gkm972
- Waterhouse, R. M., Seppey, M., Simão, F. A., Manni, M., Ioannidis, P., Klioutchnikov, G., et al. (2017). BUSCO applications from quality assessments to gene prediction and phylogenomics. *Mol. Biol. Evol.* 35, 543–548. doi: 10.1093/molbev/msx319
- Waterhouse, R. M., Seppey, M., Simão, F. A., Manni, M., Ioannidis, P., Klioutchnikov, G., et al. (2018). BUSCO applications from quality assessments to gene prediction and phylogenomics *Mol Biol. Evol.* 35(3), 543–548.
- Waycott, M., Duarte, C. M., Carruthers, T. J., Orth, R. J., Dennison, W. C., Olyarnik, S., et al. (2009). Accelerating loss of seagrasses across the globe threatens coastal ecosystems. *Proc. Natl. Acad. Sci. U.S.A.* 106, 12377–12381. doi: 10.1073/pnas.0905620106
- Wickett, N. J., Mirarab, S., Nam, N., Warnow, T., Carpenter, E., Matasci, N., et al. (2014). Phylotranscriptomic analysis of the origin and early diversification of land plants. *Proc. Natl. Acad. Sci. U.S.A.* 111, E4859–E4868. doi: 10.1073/pnas.1329261111
- Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis* (New York, USA: Springer-Verlag New York). Available at: <http://ggplot2.org>, ISBN: 978-3-319-24277-4.
- Willyard, A., Cronn, R., and Liston, A. (2009). Reticulate evolution and incomplete lineage sorting among the ponderosa pines. *Mol. Phylogenet. Evol.* 52 (2), 498–511. doi: 10.1016/j.ympev.2009.02.011
- Zeng, L., Zhang, Q., Sun, R., Kong, H., Zhang, N., and Ma, H. (2014). Resolution of deep angiosperm phylogeny using conserved nuclear genes and estimates of early divergence times. *Nat. Commun.* 5, 4956–4956. doi: 10.1038/ncomms5956
- Zhang, C., Rabiee, M., Sayyari, E., and Mirarab, S. (2018). ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinform.* 19, 153. doi: 10.1186/s12859-018-2129-y
- Zhao, T., Zwaenepoel, A., Xue, J.-Y., Kao, S.-M., Li, Z., Schranz, M. E., et al. (2021). Whole-genome microsynteny-based phylogeny of angiosperms. *Nat. Commun.* 12, 3498. doi: 10.1038/s41467-021-23665-0
- Zomlefer, W. B. (1999). Advances in angiosperm systematics: examples from the liliales and asparagales. *J. Torrey Botanical Soc.* 126, 58–62. doi: 10.2307/2997255
- Zuntini, A. R., Frankel, L. P., Pokorny, L., Forest, F., and Baker, W. J. (2021). A comprehensive phylogenomic study of the monocot order commelinales, with a

new classification of commelinaceae. *Am. J. Bot.* 108, 1066–1086. doi: 10.1002/ajb2.1698

COPYRIGHT

© 2022 Timilsena, Wafula, Barrett, Ayyampalayam, McNeal, Rentsch, McKain, Heyduk, Harkess, Villegente, Conran, Illing, Fogliani, Ané, Pires, Davis, Zomlefer, Stevenson, Graham, Givnish, Leebens-Mack and

dePamphilis. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.