

INVITED SPECIAL ARTICLE

For the Special Issue: Using and Navigating the Plant Tree of Life

# Plastid phylogenomic analysis of green plants: A billion years of evolutionary history

Matthew A. Gitzendanner<sup>1,3,8</sup> , Pamela S. Soltis<sup>2,3</sup>, Gane K.-S. Wong<sup>4,5,6</sup> , Brad R. Ruhfel<sup>7</sup>, and Douglas E. Soltis<sup>1,2,3</sup>

Manuscript received 8 September 2017; revision accepted 13 November 2017.

<sup>1</sup> Department of Biology, University of Florida, Gainesville, FL 32611, USA

<sup>2</sup> Florida Museum of Natural History, University of Florida, Gainesville, FL 32611, USA

<sup>3</sup> Genetics Institute, University of Florida, Gainesville, FL 32610, USA

<sup>4</sup> Department of Biological Sciences, University of Alberta, Edmonton AB, T6G 2E9, Canada

<sup>5</sup> Department of Medicine, University of Alberta, Edmonton AB, T6G 2E1, Canada

<sup>6</sup> BGI-Shenzhen, Beishan Industrial Zone, Yantian District, Shenzhen, 518083, China

<sup>7</sup> Department of Biological Sciences, Eastern Kentucky University, Richmond, KY 40475, USA

<sup>8</sup> Author for correspondence (e-mail: magitz@ufl.edu)

**Citation:** 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(3): 291–301.

doi:10.1002/ajb2.1048

**PREMISE OF THE STUDY:** For the past one billion years, green plants (*Viridiplantae*) have dominated global ecosystems, yet many key branches in their evolutionary history remain poorly resolved. Using the largest analysis of *Viridiplantae* based on plastid genome sequences to date, we examined the phylogeny and implications for morphological evolution at key nodes.

**METHODS:** We analyzed amino acid sequences from protein-coding genes from complete (or nearly complete) plastomes for 1879 taxa, including representatives across all major clades of *Viridiplantae*. Much of the data used was derived from transcriptomes from the One Thousand Plants Project (1KP); other data were taken from GenBank.

**KEY RESULTS:** Our results largely agree with previous plastid-based analyses. Noteworthy results include (1) the position of Zygnematomyceae as sister to land plants (*Embryophyta*), (2) a bryophyte clade (hornworts, mosses + liverworts), (3) *Equisetum* + Psilotaceae as sister to *Marattiales* + leptosporangiate ferns, (4) cycads + *Ginkgo* as sister to the remaining extant gymnosperms, within which *Gnetophyta* are placed within conifers as sister to non-Pinaceae (Gne-Cup hypothesis), and (5) *Amborella*, followed by water lilies (Nymphaeales), as successive sisters to all other extant angiosperms. Within angiosperms, there is support for *Mesangiospermae*, a clade that comprises magnoliids, Chloranthales, monocots, *Ceratophyllum*, and eudicots. The placements of *Ceratophyllum* and Dilleniaceae remain problematic. Within *Pentapetalae*, two major clades (superasterids and superrosids) are recovered.

**CONCLUSIONS:** This plastid data set provides an important resource for elucidating morphological evolution, dating divergence times in *Viridiplantae*, comparisons with emerging nuclear phylogenies, and analyses of molecular evolutionary patterns and dynamics of the plastid genome.

**KEY WORDS** chloroplast phylogeny; green plants; plant evolution; morphological innovations; plastid phylogenomics; transcriptomes; *Viridiplantae*.

*Viridiplantae* (green plants), comprising perhaps 500,000 species (Govaerts, 2003; Guiry, 2012; Judd et al., 2016), dominate terrestrial and many aquatic environments. Species of *Viridiplantae* fall into two major clades: *Chlorophyta* and *Streptophyta*. *Chlorophyta* (chlorophytes) include approximately 4300 species of

marine, freshwater, terrestrial, and symbiotic “algae”. The remaining green plants, *Streptophyta* (streptophytes), include freshwater “algae” (Mesostigmatophyceae, Chlorokybophyceae, Klebsormidiphyceae, Charophyceae, Coleochaetophyceae, and Zygnematomyceae); hornworts, mosses, and liverworts (together “bryophytes”;

comprising 16,000 species); *Lycopodiophyta* (lycophytes; 1200 species); *Monilophyta* (monilophytes; ferns; 10,000 species); and *Spermatophyta* (seed plants; gymnosperms and angiosperms; with the angiosperms having perhaps 350,000–400,000 species); *Lycopodiophyta*, *Monilophyta*, and *Spermatophyta* collectively make up *Tracheophyta* (tracheophytes; vascular plants). (Note that throughout this paper, italicized names of larger groups correspond to clade names for which phylogenetic definitions have been provided, e.g., Cantino et al. [2007] and Podani [2015]; non-italicized names reflect traditional Linnaean taxonomy.)

Despite the huge ecological and economic importance of *Viridiplantae*, many fundamental relationships remain unclear. Relationships within the *Viridiplantae* branch of the tree of life have been difficult to reconstruct for a combination of reasons, including the enormous age of the clade, significant extinction of major lineages, as well as extensive heterogeneity in the rates of molecular evolution (Soltis et al., 2002; Smith and Donoghue, 2008; Smith, 2009; Zhong et al., 2011; Rothfels et al., 2012; Ruhfel et al., 2014). Nonetheless, enormous progress has been made in recent years in clarifying our understanding of the evolution of *Viridiplantae* based on both molecular and paleobotanical research (reviewed by Ruhfel et al., 2014; Wickett et al., 2014; Chen et al., 2016; Hall and McCourt, 2016; Soltis et al., 2017).

Early phylogenetic analyses of *Viridiplantae* relied heavily on 18S–26S rDNA sequences (Chaw et al., 1997; Soltis et al., 1999b; Nickrent et al., 2000), with occasional use of mtDNA genes (Duff and Nickrent, 1999; Qiu et al., 1999, 2007), but by far most analyses have focused on plastid genes (e.g., Qiu and Palmer, 1999; Magallón and Sanderson, 2002; Qiu, 2008; Smith et al., 2009; Ruhfel et al., 2014). Importantly, regardless of genes or genome, most analyses have also been severely limited in terms of taxonomic sampling (but see Källersjö et al., 1998; Smith et al., 2009; Zanne et al., 2014 for increasingly large numbers of species but limited numbers of plastid genes)—broad comprehensive, phylogenomic analyses have not been conducted. The largest plastid phylogenomic study to date sampled across *Viridiplantae*, but the 360 species included in that study represent <0.1% of *Viridiplantae* species diversity (Ruhfel et al., 2014). The nuclear phylogenomic study of Wickett et al. (2014) sampled 852 genes but for only 103 species. Broad mitochondrial phylogenies remain problematic in *Viridiplantae* in part because of frequent horizontal gene transfer.

As a result of several decades of phylogenetic study, many of the major clades of *Viridiplantae* are now well defined. Nonetheless, major questions remain regarding the relationships among these major clades of green life. For example, the closest relatives of *Embryophyta* remain enigmatic (Nickrent et al., 2000; Karol et al., 2001; Qiu et al., 2006; Lemieux et al., 2007, 2014; Turmel et al., 2006, 2007). Similarly, the relationships among the three bryophyte lineages (mosses, liverworts, and hornworts) have varied greatly among analyses, with all possible relationships recovered (e.g., Nickrent et al., 2000; Shaw and Renzaglia, 2004; Renzaglia et al., 2007; Ruhfel et al., 2014; Wickett et al., 2014). Relationships among seed plant lineages are also contentious (reviewed by Soltis et al., 2017; Bower et al., 2000; Chaw et al., 2000; Mathews, 2009; Finet et al., 2012; Zhong et al., 2010; Ruhfel et al., 2014).

Here we provide the largest phylogenomic analysis of the *Viridiplantae* clade yet conducted, based on 1879 taxa with representatives across all major subclades of *Viridiplantae* using complete (or nearly complete) plastid protein-coding sequences. This comprehensive phylogenetic framework for *Viridiplantae* reflects a

wealth of new sequence data. We compare our results to recent plastid and nuclear phylogenomic analyses of *Viridiplantae* based on far fewer species (Ruhfel et al., 2014; Wickett et al., 2014), highlight differences in nuclear and plastid topologies, discuss implications for morphological innovation, and illustrate areas for future research. We hope that the broad phylogenetic framework we provide for *Viridiplantae* will be of enormous value to evolutionary biologists, ecologists, molecular biologists, and genomicists.

## MATERIALS AND METHODS

The One Thousand Plants Project (1KP project; onekp.com) used various Illumina sequencing technologies to generate approximately 2 Gbp of sequence from each of approximately 1400 RNA-seq libraries built in most cases from vegetative samples (leaves for vascular plants or leaves and stems for tracheophytes with microphyllous leaves) collected from taxonomically diverse *Viridiplantae*. These data were filtered and assembled with SOAP-denovo-Trans (Wickett et al., 2014), and these assemblies were the starting point of the present study.

To obtain the plastid gene sequences from the SOAP-denovo-Trans transcriptome assemblies, we developed a new pipeline. We first used BLASTX (v. 2.2.29, Altschul et al., 1990) searches against each of the 78 protein-coding genes over 100 bp long typical of plastid genomes from the data set of Ruhfel et al. (2014). BLAST hits having e-values  $<10 \times e^{-25}$  were used to extract the matching gene regions from scaffolds. Any fragment that had BLAST hits to multiple plastid genes was discarded. Because the plastid data were found to contain many transcripts that did not correspond to the correct taxon (likely due to either misassembly or contamination), we performed a BLASTN search, initially against the Ruhfel et al. (2014) data set. Samples were divided into 13 broad groups (“algae”, nonseed land plants, gymnosperms, basal angiosperms, monocots, basal eudicots, Saxifragales, campanulids, Caryophyllales, Cornales and Ericales, malvids, fabids, and lamiids), and if the best BLAST hit was to a sample outside these groups, the fragment was discarded. To take advantage of the increased coverage provided by the 1KP data added in the initial round, a second round of BLASTN searches and filtering was conducted, using the Ruhfel et al. (2014) data, plus the new fragments that had passed the initial filter above. This approach was used to increase sampling in the BLAST database across the phylogenetic breadth and reduce the chances of missing true matches where the initial data set may have had limited samples to match in some under-represented groups. All fragments that passed this second BLAST filter were retained. As many samples had multiple fragments for each gene, these were aligned to the Ruhfel et al. (2014) data for the corresponding gene using the –add option of Mafft (v. 7.125, Katoh and Standley, 2013). A single sequence for each taxon was obtained by joining non-overlapping fragments with gap characters and computing the consensus of overlapping portions of fragments. Fragments with greater than 5% sequence divergence in regions of overlap were discarded.

Data from all complete plastid genomes in GenBank as of 27 December 2014 were downloaded and parsed into gene files based on their annotations. Our data set thus consisted of the original Ruhfel et al. (2014) samples (360), additional genomes in GenBank (380), and 1KP samples (1139) (Appendix S1, see the Supplemental Data with this article). We used 52 red and brown algal samples to root the tree, leaving 1827 representatives of *Viridiplantae*.

Ruhfel et al. (2014), as well as initial analyses of our own data, found that shifts in GC content over the tree result in unconventional placement of some groups (e.g., *Lycopodiophyta*) in analyses of DNA sequences (see also Smith, 2009); thus, we restricted our analyses to an amino acid data set and did not conduct analyses removing third positions, or RY-coding as was done in Ruhfel et al. (2014). All gene sequences for all samples were therefore translated to amino acids using transeq (EMBOSS v. 6.5.7.0, Rice et al., 2000).

The amino acid data sets for each gene were aligned using MAFFT. Each gene alignment was then processed with TrimAL (v. v1.4.rev15, Capella-Gutiérrez et al., 2009) using the gappyout option to remove poorly aligned positions. The individual gene alignments were concatenated, and samples with a concatenated total of fewer than 1000 amino acids were removed from the data set.

The resulting data set was analyzed in EXaML (v. 3.0.14, Kozlov et al., 2015). For the maximum likelihood (ML) analysis, 10 random and 10 parsimony starting trees were used. The per-site rate category model (PSR) was used. One hundred bootstrap (BS) data sets were analyzed; each bootstrap replicate used a single, randomly selected starting tree from the 10 random trees used for the ML search.

## RESULTS

### Data set properties

We recovered 78 protein-coding genes across the data set (Table 1). The concatenated alignment for the 78 genes from 1879 taxa consisted of 18,328 amino acid positions. The trimmed data set had approximately 30% missing data. In some cases, entire genes were missing from many samples (e.g., *petN* was recovered from only 650 taxa, as compared to *psbA*, recovered from 1829 taxa; see Table 1). Figure 1 summarizes data set completeness across the genes and taxa.

Missing genes may be due to lack of expression at the time of sampling or true loss from the plastomes of some species. Gene recovery was low in the outgroups, possibly due to the lack of a close reference genome for the BLAST searches among these taxa. Within *Chlorophyta*, the *ndh* genes are missing in both transcriptome and genome sequences, consistent with previous reports (Martín and Sabater, 2010; Wicke et al., 2011). Similarly, in the gymnosperms, Pinaceae and *Gnetophyta* lack the *ndh* genes. Poales have lost *accD* (Wicke et al., 2011). The eurosids also show a loss of *infA* in both genome sequences and transcriptomes, a result consistent with Millen et al. (2001). In contrast, many genes not recovered from transcriptomes are found in related genome sequences; for example, recovery of *psbF* through *psbZ* genes was low from transcriptomes, although these genes are generally present in genome sequences.

### The plastid topology

The topology we obtained is summarized in Fig. 2, with the full tree with species names provided in Appendices S2 (cladogram with BS values) and S3 (phylogram). *Viridiplantae* are recovered as monophyletic (BS = 99%), within which *Chlorophyta* and *Streptophyta* (each with BS = 97%) are sisters (BS = 54%), with a small clade (BS = 100%) of *Mesostigma*, *Chlorokybus*, and the enigmatic *Spirotaenia* (Gontcharov and Melkonian, 2004) sister to this larger clade. The glaucophytes are sister to *Viridiplantae* (BS = 99%).

Within *Chlorophyta*, Prasinophyceae, Trebouxiophyceae, and Ulvophyceae are each nonmonophyletic, and the relationship of

*Chlorophyceae* (monophyletic with BS = 99%) to these lineages is uncertain.

Within *Streptophyta*, Klebsormidiophyceae, along with *Entransia* and *Interfilum*, form a clade (BS = 100%) that is sister to the remaining streptophytes, which form a grade of Charales, Coleochaetales (BS = 100%), and a clade of Zygnematophyceae (BS = 96%) and embryophytes (BS = 97%); this clade of Zygnematophyceae and embryophytes has BS support of 93%.

Within embryophytes, bryophytes are strongly supported as monophyletic (BS = 95%), with hornworts (BS = 100%) sister to a clade (BS = 93%) of liverworts and mosses, each with BS = 100%. The monophyly of *Tracheophyta* is poorly supported (BS = 54%); within *Tracheophyta*, a well-supported (BS = 100%) *Lycopodiophyta* is sister to a weakly supported (BS = 60%) *Euphyllophyta* clade comprising *Monilophyta* (ferns; BS = 100%) + *Spermatophyta* (BS = 97%).

Within *Monilophyta*, a well-supported clade (BS = 100%) of *Equisetum* + Psilotaceae is sister to *Marattiales* + leptosporangiate ferns (Polypodiidae sensu the Pteridophyte Phylogeny Group: PPG I, 2016) (BS = 100%). Within *Spermatophyta*, extant gymnosperms (*Acrogymnospermae*; BS = 100%) are sister to *Angiospermae* (BS = 97%). Within gymnosperms, *Cycadophyta* (cycads; BS = 100%) and *Ginkgo* form a clade (BS = 98%) that is sister to a clade in which *Gnetophyta* are nested within conifers (BS = 100%). We find strong evidence for the “Gne-Cup” hypothesis, with *Gnetophyta* forming a well-supported clade (BS = 98%) with *Cupressophyta* sensu Cantino et al. (2007), the latter comprising a clade (BS = 100%) of Araucariaceae (BS = 100%), Podocarpaceae (BS = 100%), *Sciadopitys*, Cephalotaxaceae (BS = 100%), Taxaceae (BS = 100%), and Cupressaceae (BS = 100%). This clade of *Gnetophyta* + *Cupressophyta* is sister to Pinaceae (BS = 100%).

The monophyly of angiosperms receives 97% BS support, with *Amborella* sister to all other extant angiosperms (BS = 80%); *Nymphaeales* (BS = 100%) and *Austrobaileyales* (BS = 100%) are successive sisters (BS = 80%) to all other angiosperms (*Mesangiospermae*; BS = 97%), comprising *Magnoliophyta* (magnoliids), Chloranthales, monocots, *Ceratophyllum*, and eudicots. Magnoliids (BS = 95%) and Chloranthales (BS = 100%) form a weakly supported clade (BS = 61%) that is sister to a clade (BS = 56%) of monocots (BS = 100%) and eudicots (BS = 97%) + *Ceratophyllum* (41%).

The placement of *Ceratophyllum* remains problematic, with only 41% bootstrap support for its placement as sister to the eudicots. Ranunculales are sister to the remaining eudicots (BS = 97%), followed by a grade of Proteales (BS = 100%), *Meliosma*, and Trochodendrales + Buxales (forming a clade with BS = 52%). Gunnerales are sister to *Pentapetalae* (97%), and *Dillenia* is sister to the remaining eudicots (BS = 97%), which comprise a clade of superrosids (BS = 99%) and superasterids (BS = 48%).

Within the superrosids, Vitales are sister to the remainder (BS = 61%), which comprise Saxifragales (BS = 100%) sister to the core rosids/eurosids (BS = 99%); the core rosids comprise the malvid and fabid clades, although each is weakly supported with only BS = 61% and 62%, respectively. The COM clade (Celastrales, Oxalidales, Malpighiales), long a problematic region of angiosperm phylogeny (e.g., Wurdack and Davis, 2009; Sun et al., 2015), is strongly supported as a clade (BS = 90%) and placed within the fabids as sister (BS = 74%) to Zygophyllales. Within the COM clade, there is weak support (BS = 59%) for (Celastrales + Oxalidales), which are in turn sister to Malpighiales (BS = 100%).

**TABLE 1.** List of the 78 protein-coding plastid genes used in the analysis and alignment properties. AA = amino acids

Gene	Number of taxa with sequences	Maximum ungapped length (AA)	Aligned length, pre-trimAL (AA)	Aligned length, post-trimAL (AA)	Average pairwise %-identity (AA)
<i>accD</i>	1216	1017	4376	235	79.3
<i>atpA</i>	1616	556	629	488	89.7
<i>atpB</i>	1589	577	731	498	92.1
<i>atpE</i>	1651	149	185	128	74.4
<i>atpF</i>	1597	230	395	180	70.6
<i>atpH</i>	1536	130	147	79	96.1
<i>atpI</i>	1621	313	448	246	88.0
<i>ccsA</i>	1476	447	1112	316	66.1
<i>cemA</i>	1514	582	1207	229	67.0
<i>clpP</i>	1538	556	1097	195	71.8
<i>infA</i>	1090	116	159	61	75.6
<i>matK</i>	1257	546	1057	364	64.9
<i>ndhA</i>	1443	385	649	362	79.7
<i>ndhB</i>	1304	545	786	518	88.0
<i>ndhC</i>	1389	135	170	119	86.4
<i>ndhD</i>	1403	531	681	495	80.7
<i>ndhE</i>	1301	111	129	99	84.1
<i>ndhF</i>	1327	791	1186	778	68.0
<i>ndhG</i>	1349	222	271	174	75.3
<i>ndhH</i>	1459	506	668	388	87.4
<i>ndhI</i>	1401	189	233	179	86.4
<i>ndhJ</i>	1447	174	228	155	85.8
<i>ndhK</i>	1403	318	520	285	83.7
<i>petA</i>	1711	394	548	320	84.1
<i>petB</i>	1751	235	264	213	95.5
<i>petD</i>	1718	224	307	155	93.3
<i>petG</i>	717	50	55	36	90.5
<i>petL</i>	671	86	119	30	79.4
<i>petN</i>	650	237	238	28	94.9
<i>psaA</i>	1750	769	944	745	92.7
<i>psaB</i>	1746	748	863	727	92.7
<i>psaC</i>	1717	103	110	74	95.4
<i>psaI</i>	669	226	275	33	78.3
<i>psaJ</i>	731	77	89	40	82.7
<i>psbA</i>	1837	359	417	350	96.5
<i>psbB</i>	1789	510	645	505	91.7
<i>psbC</i>	1787	487	570	449	94.6
<i>psbD</i>	1823	373	425	351	96.1
<i>psbE</i>	1748	490	497	81	90.6
<i>psbF</i>	718	49	56	38	94.5
<i>psbH</i>	1577	94	120	71	81.1
<i>psbI</i>	697	230	254	34	95.3
<i>psbJ</i>	717	50	62	40	87.8
<i>psbK</i>	733	91	117	42	83.4
<i>psbL</i>	679	43	43	37	93.9
<i>psbM</i>	709	97	135	33	87.2
<i>psbN</i>	851	97	122	42	87.7
<i>psbT</i>	717	508	508	32	93.0
<i>psbZ</i>	1196	99	123	60	82.5
<i>rbcl</i>	1774	490	548	474	92.2
<i>rpl2</i>	1713	314	491	263	80.1
<i>rpl14</i>	1732	147	195	121	81.6
<i>rpl16</i>	1716	171	260	130	81.4
<i>rpl20</i>	1618	157	290	112	66.8
<i>rpl22</i>	1407	194	335	184	66.9
<i>rpl23</i>	1432	112	201	89	73.3
<i>rpl32</i>	662	88	169	41	70.9
<i>rpl33</i>	1479	76	112	65	76.4
<i>rpl36</i>	729	196	196	35	84.8
<i>rpoA</i>	1565	858	2417	325	67.5
<i>rpoB</i>	1345	2143	4523	1069	72.3

(continues)



**TABLE 1.** (Continued)

Gene	Number of taxa with sequences	Maximum ungapped length (AA)	Aligned length, pre-trimAL (AA)	Aligned length, post-trimAL (AA)	Average pairwise %-identity (AA)
<i>rpoC1</i>	1351	2348	5920	682	72.9
<i>rpoC2</i>	1264	3629	9252	1355	63.3
<i>rps2</i>	1658	951	1131	235	75.9
<i>rps3</i>	1690	745	1392	202	69.6
<i>rps4</i>	1637	257	538	200	76.5
<i>rps7</i>	1666	279	439	154	79.4
<i>rps8</i>	1650	177	339	129	72.7
<i>rps11</i>	1686	161	258	135	73.9
<i>rps12</i>	1511	165	324	103	87.8
<i>rps14</i>	1664	116	134	98	78.8
<i>rps15</i>	1400	116	168	39	77.2
<i>rps16</i>	1151	99	123	77	80.8
<i>rps18</i>	1566	251	596	98	76.0
<i>rps19</i>	1699	215	371	91	76.0
<i>ycf2</i>	657	2481	6703	335	76.9
<i>ycf3</i>	1614	201	330	167	85.7
<i>ycf4</i>	1,570	209	400	183	75.1

Within the superasterid clade, Santalales (BS = 100%) are sister to the rest of the clade (BS = 63%), with Berberidopsidales sister to a clade (BS = 94%) of Caryophyllales (BS = 100%) and asterids (BS = 98%). The asterids comprise a clade of [Cornales (BS = 100%) sister to a clade of Ericales (BS = 100%) + (campanulids (BS = 98%) + lamiids (BS = 91%))].

The topology for monocots generally agrees with recent analyses (with the exception of the placement of Petrosaviales; see below) (Chase et al., 2006; Graham et al., 2006; Givnish et al., 2010; Soltis et al., 2011; Hertweck et al., 2015). We found that Acoraceae (Acorales), followed by Alismatales, are each well supported (BS = 100%) and are subsequent sisters to all other monocots, which form a well-supported clade (BS = 100%). Within this remaining large clade of monocots, the major subclades are typically well supported: *Commelinidae* (Arecales, Commelinales, Poales, and Zingiberales), Dioscoreales, Petrosaviales, Pandanales, Liliales, and Asparagales. After Acoraceae and Alismatales, we recover Pandanales as sister to all remaining monocots, a placement that is weakly supported (remaining monocots have BS = 76%). Following Pandanales, Liliales and then Asparagales are sister to the remaining monocots. We then recover Petrosaviales as sister to the weakly supported (BS = 62%) remaining monocots consisting of Arecales + *Commelinidae* (BS = 83%). This placement of Petrosaviales differs from previous analyses in which following Acorales and Alismatales, Petrosaviales have been sister to all other monocots (Chase et al., 2006; Graham et al., 2006; see APG, 2016). Relationships obtained here within the larger clades of monocots are in general agreement with other recent analyses (see Chase et al., 2006; Graham et al., 2006; Givnish et al., 2010; Soltis et al., 2011; Hertweck et al., 2015). For example, within *Commelinidae*, we recovered Arecales as sister to Poales + (Commelinales + Zingiberales).

## DISCUSSION

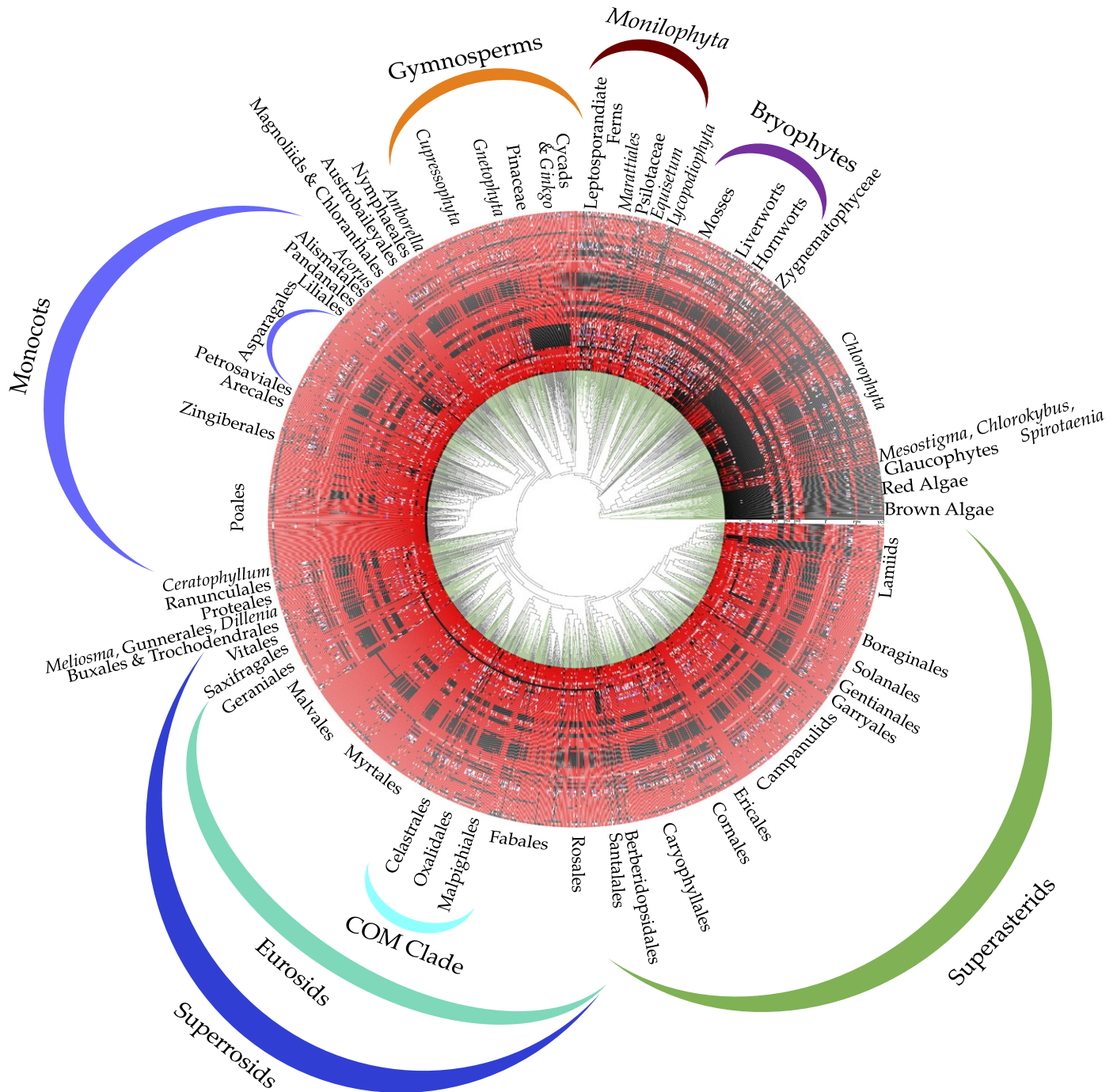
### Low bootstrap support for some nodes

Several historically well-supported nodes in the *Viridiplantae* phylogeny received relatively weak support in our analyses (e.g.,

monophyly of vascular plants [BS = 54%]; monophyly of superasterids [BS = 48%]). While this could certainly be a sign of low support in this data set for these relationships, we believe most cases are due to the limited analyses we were able to conduct on each of the bootstrap data sets (a single analysis from one random starting tree). For the vascular plants, for example, about 40% of topologies found a large group of monilophytes sister to the remaining land plants (including the bryophytes). We reanalyzed six of these bootstrap data sets using a different random starting tree and found that all six found a tree consistent with the monophyletic vascular plant topology presented in Fig. 2. Thus, it appears that this reduced effort in analyzing bootstrap data sets generated trees that may have been stuck on local optima or suboptimal parts of tree space; these suboptimal trees varied in topology, yielding lower support for the monophyly of certain clades (see Mort et al., 2000). However, given that each bootstrap analysis uses about 300 CPU hours, using the same analytical rigor used for the maximum likelihood tree search (20 starting trees) would use over 600,000 CPU hours for the bootstrap analysis. Such an analysis was beyond the scope of our current study, although as data sets continue to grow, we encourage further research into the effects of limited analysis of bootstrap datasets (Mort et al., 2000) and new methods of inferring support for relationships (see Smith et al., 2018). It should also be acknowledged that it is possible that reduced rigor in analyses of bootstrap data sets may lead to higher bootstrap values, and support values should therefore be interpreted with caution.

### Comparison to other plastid trees

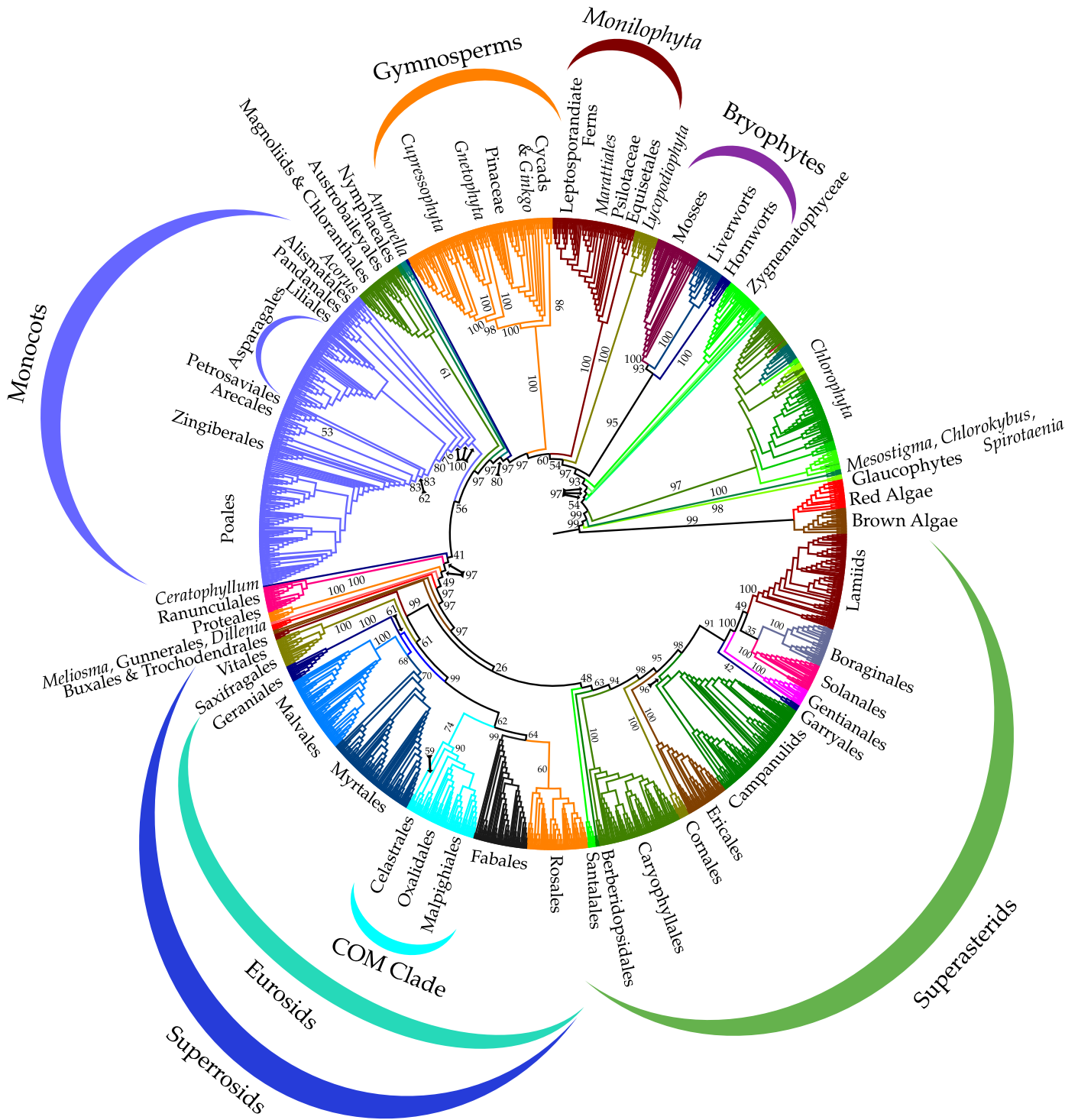
Our results are generally congruent with the previous plastid genome analysis of Ruhfel et al. (2014); here we focus primarily on the amino acid analysis of that study. However, the current study includes five times as many taxa (1879) and provides much better coverage of all lineages of *Viridiplantae*. There are differences, however, between our results and those of Ruhfel et al. (2014). We recovered a well-supported (BS = 95%) bryophyte clade, but Ruhfel et al. (2014) found various relationships among hornworts, liverworts, and mosses, with the amino acid analysis supporting mosses and liverworts together (BS = 99%) as found here, but as



**FIGURE 1.** Heatmap showing presence of sequence data for each gene and taxon across the phylogeny. Branches on the phylogeny are colored green to indicate 1KP transcriptome data and black to indicate data from NCBI or Ruhfel et al. (2014; mostly NCBI and genomic data). Genes are arranged alphabetically, as in Table 1 with *accD* on the inside and *ycf4* on the outside of the arcs; one or more letters are placed where the letter of the gene name changes, e.g., “i” indicates the position of *infA*, “*psb*” indicates where *psa* genes end and *psb* genes begin, etc. Heatmap colors: black, no sequence data for gene; blue, 0–25%; white, 25–50%; pink, 50–75%; red, 75–100% of full-length gene.

sister to a clade of the single hornwort sample plus the remaining *Embryophyta* (BS = 53%). Indeed, none of the several different analyses of Ruhfel et al. (2014) recovered the bryophytes as monophyletic. Ruhfel et al. (2014) recovered *Ceratophyllum* and Piperales as sister to the monocots, although with low support (BS = 56%); we recovered Piperales as sister to Laurales (BS = 95%) and

*Ceratophyllum* as sister to the eudicots, but with very low support (BS = 41%). The placement of Piperales in Ruhfel et al. (2014) disagrees with recent analyses and likely reflects limited taxon sampling in the magnoliids. *Ceratophyllum* has often been placed as sister to eudicots (e.g., Soltis et al., 2011; Sun et al. 2016; summarized by APG, 2016), but its position as sister to monocots has also



**FIGURE 2.** Summary of the plastid phylogenomic tree based on analysis of 1827 *Viridiplantae* taxa and 52 outgroups using 78 protein-coding genes. Bootstrap values are those highlighted in the text. Appendices S2 and S3 present the full topology with species names, with all bootstrap values given in Appendix S2. Clades are colored consistently across figures.

been reported repeatedly (e.g., Zanis et al., 2002). The placement here of Dilleniaceae as sister to all core eudicots except *Gunnera* (i.e., to superrosids + superasterids) also conflicts with its position in Ruhfel et al. (2014), where it was sister to the superasterids (BS = 70%). The placement of Dilleniaceae has been problematic

in all phylogenetic analyses of angiosperms (e.g., Soltis et al., 1999a, 2011).

In general, the increased taxon sampling in our analysis relative to Ruhfel et al. (2014) did not change the topology greatly or add bootstrap support to the nodes that were previously problematic.



These problematic nodes may represent hard polytomies that result from rapid radiations (as proposed for mesangiosperms; Moore et al., 2010). Comparison of the plastid tree with a robust nuclear phylogeny may help reveal the causes of these remaining areas of poor resolution and support.

### Discordance with other studies

In general, there is strong agreement between the plastid tree we report here and nuclear topologies based on 852 nuclear genes for 103 species reported by Wickett et al. (2014). However, several areas of discordance deserve discussion. We stress that most differences are weakly supported in either the plastid or nuclear trees (or both) and also conflict between concatenated vs. gene-tree-based nuclear trees (Wickett et al., 2014). Furthermore, some of the more noteworthy differences are represented by short branches in the ASTRAL nuclear tree. These branch lengths are proportional to coalescent time, and we infer that these problematic relationships reflect “non-tree-like” events, including ILS and hybridization. Perhaps the most noteworthy differences between plastid and nuclear topologies involve the relationships among bryophytes, among gymnosperms, and of magnoliids to either monocots or eudicots. These differences could be due to differences in taxon sampling between Wickett et al. (2014) and the current study; it will be important to rule out this issue through a comparable analysis of plastid and nuclear genes from the same large set of species (Green Plant Consortium, unpublished manuscript).

The position of *Mesostigma* + *Chlorokybus* has fluctuated in different analyses, with Lemieux et al. (2007) finding this clade either sister to *Streptophyta* or *Streptophyta* + *Chlorophyta*. The nuclear analyses of Wickett et al. (2014) and the Green Plant Consortium (unpublished manuscript) both found *Mesostigma* + *Chlorokybus* sister to *Streptophyta*.

Our analyses of plastid data provide strong support (BS = 95%) for a bryophyte clade with hornworts sister to mosses + liverworts. In contrast, the nuclear tree of Wickett et al. (2014) based on concatenated genes provides 100% bootstrap support for mosses + liverworts, with hornworts sister to the tracheophytes; however, the ASTRAL analyses of nuclear gene trees recovered a bryophyte clade with 97% bootstrap support (Wickett et al., 2014).

Within gymnosperms, our plastid tree and that of Ruhfel et al. (2014) place the enigmatic *Gnetophyta* within conifers as sister to non-Pinaceae (the Gne-Cup hypothesis) with strong support (BS = 98%). In contrast, the concatenated nuclear tree of Wickett et al. (2014) placed *Gnetophyta* sister to Pinaceae with strong support (Gne-pine hypothesis; BS = 100%), while the ASTRAL tree found *Gnetophyta* sister to the conifers (Gnetifer hypothesis; BS = 89%). These topological differences in gymnosperm relationships have been noted previously (e.g., Bowe et al., 2000; Chaw et al., 1997; Lee et al., 2011; reviewed by Soltis et al., 2017).

Both concatenated and ASTRAL nuclear trees place magnoliids + Chloranthales sister to the eudicots with 100% bootstrap support (Wickett et al., 2014), while, in contrast, our analyses place magnoliids + Chloranthales sister to a clade of monocots and eudicots + *Ceratophyllum*. However, bootstrap support for these relationships is weak.

Another noteworthy discrepancy between plastid and nuclear trees is the placement of the COM clade (Celastrales–Oxalidales–Malpighiales) in differing positions within rosids. Previous analyses of nuclear genes (Wickett et al., 2014) have placed the clade in

*Malvidae*, whereas plastid data place the clade within *Fabidae*, as seen here (reviewed by Sun et al., 2015).

### Implications for Embryophyta evolution

Our results have major implications for resolving the sister group of *Embryophyta*. Despite many studies having found either Charophyceae or Coleochaetophyceae as sister to embryophytes (reviewed by Ruhfel et al. 2014; Wickett et al., 2014), recent analyses have converged with Zygnematophyceae as sister to all *Embryophyta* (BS = 93% in this study) (Timme et al., 2012; Ruhfel et al., 2014; Wickett et al., 2014). Thus, both plastid and nuclear genes support this placement. The data indicate that the complex morphological characters found in Zygnematophyceae, such as branching, parental retention of the egg, and plasmodesmata, either originated once in the common ancestor of Charophyceae, Coleochaetophyceae, Zygnematophyceae, and *Embryophyta* and were then lost in most lineages of Zygnematophyceae or these traits originated independently in Charophyceae, Coleochaetophyceae, and *Embryophyta* (reviewed by Judd et al., 2016).

Resolving relationships among the three lineages of bryophytes (hornworts, liverworts, and mosses) has been extremely difficult (reviewed by Ruhfel et al., 2014; Wickett et al., 2014; Judd et al., 2016). As multiple papers have indicated (Qiu et al., 1998; Nickrent et al., 2000; Renzagalia et al., 2000; Nishiyama et al., 2004; Chang and Graham, 2011; reviewed by Ruhfel et al., 2014; Wickett et al., 2014), every possible branching order of relationships among these three lineages has been obtained in analyses using diverse molecular and morphological data. Our result indicating a bryophyte clade mirrors the recent results of Wickett et al. (2014) using a large number of nuclear genes (but fewer taxa); however, as noted, this result for nuclear genes was obtained only from the coalescent analysis and not the concatenated analysis. Importantly, recent analyses of more taxa (but fewer nuclear genes than in Wickett et al. [2014]) also reveals a bryophyte clade (Green Plant Consortium, unpublished manuscript).

Resolving the relationships among the three lineages of bryophytes has major implications for the origin and evolutionary transformation of the heteromorphic alternation of generations typical of all embryophytes. The traditional hypothesis, based on a grade of bryophytes, as often reported in previous studies, suggests that a gametophyte-dominant life cycle is ancestral in embryophytes (Ligrone et al., 2012; Judd et al., 2016) and the sporophyte-dominant life cycle of tracheophytes is derived. However, if bryophytes are indeed a clade, as our results indicate, then the alternation of generations of the common ancestor of all embryophytes cannot be unequivocally reconstructed, with gametophyte-dominant and sporophyte-dominant life cycles both plausible hypotheses of the ancestral state.

### Future of plastid phylogenomics

We may be close to exhausting the power of the plastid genome for resolving deep-level phylogenetic questions of relationships in *Viridiplantae* (cf. Davis et al., 2014). Particularly enigmatic areas of the *Viridiplantae* tree of life first revealed with a few genes remain problematic with the entire protein-coding component of the plastid genome, despite dense taxon sampling in the problematic portions of the tree. Moreover, although some of the uncertain nodes are deep within *Viridiplantae* phylogeny, many others



occur at more shallow levels, suggesting that perhaps different processes—extinction vs. recent radiations, for example—may be responsible at different points in the tree. Notable areas of uncertainty include the placement of Chloranthales and magnoliids, the positions of eudicot lineages, such as Dilleniaceae, and the placement of Ceratophyllaceae. Adding more taxa will continue to fill in the leaves of the tree, but will likely do little to improve support for deeper nodes. Filling missing genes may help with resolution and support, and adding noncoding regions will provide more characters (although such characters will also augment challenges of alignment and varying GC content). Thus, it seems unlikely that plastid genome analysis alone will recover a single, well-supported topology resolving all remaining deep-level questions.

Sequencing of the plastid genome has now become routine, and obtaining complete or nearly complete plastid genomes will soon be as commonplace as the sequencing of a few genes via PCR. Although current work suggests that plastid data will likely not resolve all long standing problems in *Viridiplantae* phylogeny, plastid genome sequencing has a very bright future and broad application as we move to the tips of the *Viridiplantae* tree of life, particularly within clades recognized as orders, families, and genera.

Future analyses should focus on elucidating the underlying causes responsible for the noteworthy discrepancies between plastid and nuclear phylogenetic trees and for regions of poor support. It is our hope that these studies will usher in a new era in molecular systematics that explicitly incorporates phylogenetic uncertainty and conflict in studying and using the tree of life. Despite its longstanding role in plant phylogenetics, the plastid genome provides only one perspective on plant evolutionary history, and a plastid-based tree should not be blindly accepted as the backbone tree for *Viridiplantae*. Incongruent placements between nuclear and plastid trees (e.g., the COM clade) are points of concern for reconstructing phylogeny, but opportunities for understanding the intricacies of plant evolution and should be the focus of intensive study.

## ACKNOWLEDGEMENTS

The authors thank the many researchers across the world who supplied samples and, in many cases, extracted RNA for the 1KP project. Without these samples, this project would not have been possible. These major contributors are recognized as coauthors on the 1KP synthesis paper (Green Plant Consortium, unpublished manuscript). In addition, we thank University of Florida Research Computing for providing access to HiPerGator for the computational analyses conducted in this study. The authors also thank two anonymous reviewers and the editor for their helpful suggestions to improve the manuscript.

## DATA ACCESSIBILITY

The assembled transcriptomes from the 1KP project used for this study are publicly available at [http://www.onekp.com/public\\_data.html](http://www.onekp.com/public_data.html). All raw sequence data have been deposited in the SRA: see [http://www.onekp.com/public\\_read\\_data.html](http://www.onekp.com/public_read_data.html) for list of accession numbers. Scripts used to analyze the data as well as single-gene and concatenated amino acid alignments are available at [https://github.com/magitz/1KP\\_Plastid](https://github.com/magitz/1KP_Plastid).

## SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

## LITERATURE CITED

- Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. 1990. Basic local alignment search tool. *Journal of Molecular Biology* 215: 403–410.
- 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.
- Bowe, L. M., G. Coat, and C. W. dePamphilis. 2000. Phylogeny of seed plants based on all three genomic compartments: Extant gymnosperms are monophyletic and Gnetales' closest relatives are conifers. *Proceedings of the National Academy of Sciences* 97: 4092–4097.
- Canino, 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.
- Capella-Gutiérrez, S., J. M. Silla-Martínez, and T. Gabaldón. 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25: 1972–1973.
- Chang, Y., and S. W. Graham. 2011. Inferring the higher-order phylogeny of mosses (Bryophyta) and relatives using a large, multigene plastid data set. *American Journal of Botany* 98: 839–849.
- Chase, M., M. Fay, D. Devey, O. Maurin, N. Rønsted, T. Davies, Y. Pillon, et al. 2006. Multigene analyses of monocot relationships. *Aliso* 22: 63–75.
- Chaw, S.-M., C. L. Parkinson, Y. Cheng, T. M. Vincent, and J. D. Palmer. 2000. Seed plant phylogeny inferred from all three plant genomes: monophyly of extant gymnosperms and origin of Gnetales from conifers. *Proceedings of the National Academy of Sciences, USA* 97: 4086–4091.
- Chaw, S.-M., A. Zharkikh, H.-M. Sung, T.-C. Lau, and W.-H. Li. 1997. Molecular phylogeny of extant gymnosperms and seed plant evolution: analysis of nuclear 18S rRNA sequences. *Molecular Biology and Evolution* 14: 56–58.
- Chen, Z.-D., T. Yang, L. Lin, L.-M. Lu, H.-L. Li, M. Sun, B. Liu, et al. 2016. Tree of life for the genera of Chinese vascular plants. *Journal of Systematics and Evolution* 54: 277–306.
- Davis, C. C., Z. Xi, and S. Mathews. 2014. Plastid phylogenomics and green plant phylogeny: almost full circle but not quite there. *BMC Biology* 12: 11.
- Duff, R. J., and D. L. Nickrent. 1999. Phylogenetic relationships of land plants using mitochondrial small-subunit rDNA sequences. *American Journal of Botany* 86: 372–386.
- Finet, C., R. E. Timme, C. F. Delwiche, and F. Marlétaz. 2012. Multigene phylogeny of the green lineage reveals the origin and diversification of land plants. *Current Biology* 22: 1456–1457.
- Givnish, T. J., M. Ames, J. R. McNeal, M. R. McKain, P. R. Steele, C. W. dePamphilis, S. W. Graham, et al. 2010. Assembling the tree of the Monocotyledons: plastome sequence phylogeny and evolution of Poales. *Annals of the Missouri Botanical Garden* 97: 584–616.
- Gontcharov, A. A., and M. Melkonian. 2004. Unusual position of the genus *Spirotaenia* (Zygnematophyceae) among streptophytes revealed by SSU rDNA and *rbcL* sequence comparisons. *Phycologia* 43: 105–113.
- Govaerts, R. 2003. How many species of seed plants are there? a response. *Taxon* 52: 583–584.
- Graham, S., J. Zgurski, M. McPherson, D. Cherniawsky, J. Saarela, E. Horne, S. Smith, et al. 2006. Robust inference of monocot deep phylogeny using an expanded multigene plastid data set. *Aliso* 22: 3–21.
- Guiry, M. D. 2012. How many species of algae are there? *Journal of Phycology* 48: 1057–1063.
- Hall, J.D., and R. McCourt. 2016. Zygnematophyta. In J. M. Archibald, A. G. B. Simpson, C. H. Slamovits, L. Margulis, M. Melkonian, D. J. Chapman, and J. O. Corliss [eds.], *Handbook of the protists*, 1–30. Springer International

- Publishing, Cham, Switzerland. Available at [http://link.springer.com/referenceworkentry/10.1007/978-3-319-32669-6\\_41-1](http://link.springer.com/referenceworkentry/10.1007/978-3-319-32669-6_41-1).
- Hertweck, K. L., M. S. Kinney, S. A. Stuart, O. Maurin, S. Mathews, M. W. Chase, M. A. Gandolfo, and J. C. Pires. 2015. Phylogenetics, divergence times and diversification from three genomic partitions in monocots. *Botanical Journal of the Linnean Society* 178: 375–393.
- Judd, W. S., C. S. Campbell, E. A. Kellogg, P. F. Stevens, and M. J. Donoghue. 2016. *Plant systematics: a phylogenetic approach*, 4th ed. Sinauer, Sunderland, MA, USA.
- Källersjö, M., J. S. Farris, M. W. Chase, B. Bremer, M. F. Fay, C. J. Humphries, G. Petersen, et al. 1998. Simultaneous parsimony jackknife analysis of 2538 *rbcl* DNA sequences reveals support for major clades of green plants, land plants, seed plants and flowering plants. *Plant Systematics and Evolution* 213: 259–287.
- Karol, K. G., R. M. McCourt, M. T. Cimino, and C. F. Delwiche. 2001. The closest living relatives of land plants. *Science* 294: 2351–2353.
- Katoh, K., and D. M. Standley. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution* 30: 772–780.
- Kozlov, A. M., A. J. Aberer, and A. Stamatakis. 2015. ExaML version 3: a tool for phylogenomic analyses on supercomputers. *Bioinformatics* 31: 2577–2579.
- Lee, E. K., A. Cibrian-Jaramillo, S.-O. Kolokotronis, M. S. Katari, A. Stamatakis, M. Ott, J. C. Chiu, et al. 2011. A functional phylogenomic view of the seed plants. *PLOS Genetics* 7: e1002411.
- Lemieux, C., C. Otis, and M. Turmel. 2007. A clade uniting the green algae *Mesostigma viride* and *Chlorokybus atmophyticus* represents the deepest branch of the Streptophyta in chloroplast genome-based phylogenies. *BMC Biology* 5: 2.
- Lemieux, C., C. Otis, and M. Turmel. 2014. Chloroplast phylogenomic analysis resolves deep-level relationships within the green algal class Trebouxiophyceae. *BMC evolutionary biology* 14: 211.
- Ligrone, R., J. G. Duckett, and K. S. Renzaglia. 2012. Major transitions in the evolution of early land plants: a bryological perspective. *Annals of Botany* 109: 851–871.
- Magallón, S., and M. J. Sanderson. 2002. Relationships among seed plants inferred from highly conserved genes: sorting conflicting phylogenetic signals among ancient lineages. *American Journal of Botany* 89: 1991–2006.
- Martin, M., and B. Sabater. 2010. Plastid *ndh* genes in plant evolution. *Plant Physiology and Biochemistry* 48: 636–645.
- Mathews, S. 2009. Phylogenetic relationships among seed plants: persistent questions and the limits of molecular data. *American Journal of Botany* 96: 228–236.
- Millen, R. S., R. G. Olmstead, K. L. Adams, J. D. Palmer, N. T. Lao, L. Heggie, T. A. Kavanagh, et al. 2001. Many parallel losses of *infA* from chloroplast DNA during angiosperm evolution with multiple independent transfers to the nucleus. *Plant Cell Online* 13: 645–658.
- Mort, M. E., P. S. Soltis, D. E. Soltis, and M. L. Mabry. 2000. A comparison of bootstrap, fast bootstrap, and jackknife values in three empirical data sets. *Systematic Biology* 49: 102–113.
- 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.
- Nickrent, D. L., C. L. Parkinson, J. D. Palmer, and R. J. Duff. 2000. Multigene phylogeny of land plants with special reference to bryophytes and the earliest land plants. *Molecular Biology and Evolution* 17: 1885–1895.
- Nishiyama, T., P. G. Wolf, M. Kugita, R. B. Sinclair, M. Sugita, C. Sugiura, T. Wakasugi, et al. 2004. Chloroplast phylogeny indicates that bryophytes are monophyletic. *Molecular Biology and Evolution* 21: 1813–1819.
- Podani, J. 2015. A növények evolúciója és osztályozása. (Evolution and systematics of plants). Eötvös Kiadó, Budapest, Hungary.
- PPG I. 2016. A community-derived classification for extant lycophytes and ferns. *Journal of Systematics and Evolution* 54: 563–603.
- Qiu, Y.-L. 2008. Phylogeny and evolution of charophytic algae and land plants. *Journal of Systematics and Evolution* 46: 287–306.
- Qiu, Y.-L., Y. Cho, J. C. Cox, and J. D. Palmer. 1998. The gain of three mitochondrial introns identifies liverworts as the earliest land plants. *Nature* 394: 671–674.
- Qiu, Y.-L., J. Lee, F. Bernasconi-Quadroni, D. E. Soltis, P. S. Soltis, M. Zanis, Z. Chen, et al. 1999. The earliest angiosperms: evidence from mitochondrial, plastid and nuclear genomes. *Nature* 402: 404–407.
- Qiu, Y., L. Li, B. Wang, Z. Chen, O. Dombrowska, J. Lee, L. Kent, et al. 2007. A nonflowering land plant phylogeny inferred from nucleotide dequences of seven chloroplast, mitochondrial, and nuclear genes. *International Journal of Plant Sciences, USA* 168: 691–708.
- Qiu, Y.-L., L. Li, B. Wang, Z. Chen, V. Knoop, M. Groth-Malonek, O. Dombrowska, et al. 2006. The deepest divergences in land plants inferred from phylogenomic evidence. *Proceedings of the National Academy of Sciences, USA* 103: 15511–15516.
- Qiu, Y.-L., and J. D. Palmer. 1999. Phylogeny of early land plants: insights from genes and genomes. *Trends in Plant Science* 4: 26–30.
- Renzaglia, K. S., R. J. T. Duff, D. L. Nickrent, and D. J. Garbary. 2000. Vegetative and reproductive innovations of early land plants: implications for a unified phylogeny. *Philosophical Transactions of the Royal Society, B, Biological Sciences* 355: 769–793.
- Renzaglia, K. S., S. Schuette, R. J. Duff, R. Ligrone, A. J. Shaw, B. D. Mishler, and J. G. Duckett. 2007. Bryophyte phylogeny: advancing the molecular and morphological frontiers. *Bryologist* 110: 179–213.
- Rice, P., I. Longden, and A. Bleasby. 2000. EMBOSS: the European molecular biology open software suite. *Trends in Genetics* 16: 276–277.
- Rothfels, C. J., A. Larsson, L.-Y. Kuo, P. Korall, W.-L. Chiou, and K. M. Pryer. 2012. Overcoming deep roots, fast rates, and short internodes to resolve the ancient rapid radiation of eupolypod II ferns. *Systematic Biology* 61: 490–509.
- 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.
- Shaw, J., and K. Renzaglia. 2004. Phylogeny and diversification of bryophytes. *American Journal of Botany* 91: 1557–1581.
- Smith, D. R. 2009. Unparalleled GC content in the plastid DNA of *Selaginella*. *Plant Molecular Biology* 71: 627.
- Smith, S. A., J. M. Beaulieu, and M. J. Donoghue. 2009. Mega-phylogeny approach for comparative biology: an alternative to supertree and supermatrix approaches. *BMC Evolutionary Biology* 9: 37.
- 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., and M. J. Donoghue. 2008. Rates of molecular evolution are linked to life history in flowering plants. *Science* 322: 86–89.
- Soltis, D., P. Soltis, P. Endress, M. Chase, S. Manchester, W. Judd, L. Majure, and E. Mavrodiev. 2017. *Phylogeny and evolution of the Angiosperms: revised and updated edition*. University of Chicago Press, Chicago, IL, USA.
- 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, P. S., D. E. Soltis, and M. W. Chase. 1999a. Angiosperm phylogeny inferred from multiple genes as a tool for comparative biology. *Nature* 402: 402–404.
- Soltis, P. S., D. E. Soltis, V. Savolainen, P. R. Crane, and T. G. Barraclough. 2002. Rate heterogeneity among lineages of tracheophytes: integration of molecular and fossil data and evidence for molecular living fossils. *Proceedings of the National Academy of Sciences, USA* 99: 4430–4435.
- Soltis, P. S., D. E. Soltis, P. G. Wolf, D. L. Nickrent, S.-M. Chaw, and R. L. Chapman. 1999b. The phylogeny of land plants inferred from 18S rDNA sequences: pushing the limits of rDNA signal? *Molecular Biology and Evolution* 16: 1774–1784.
- 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.
- Sun, Y., M. J. Moore, S. Zhang, P. S. Soltis, D. E. Soltis, T. Zhao, A. Meng, et al. 2016. Phylogenomic and structural analyses of 18 complete plastomes across nearly all families of early-diverging eudicots, including an angiosperm-wide analysis of IR gene content evolution. *Molecular Phylogenetics and Evolution* 96: 93–101.

- Timme, R. E., T. R. Bachvaroff, and C. F. Delwiche. 2012. Broad phylogenomic sampling and the sister lineage of land plants. *PLOS ONE* 7: e29696.
- Turmel, M., C. Otis, and C. Lemieux. 2006. The chloroplast genome sequence of *Chara vulgaris* sheds new light into the closest green algal relatives of land plants. *Molecular Biology and Evolution* 23: 1324–1338.
- Turmel, M., J. Pombert, P. Charlebois, C. Otis, and C. Lemieux. 2007. The green algal ancestry of land plants as revealed by the chloroplast genome. *International Journal of Plant Sciences, USA* 168: 679–689.
- Wicke, S., G. M. Schneeweiss, C. W. dePamphilis, K. F. Müller, and D. Quandt. 2011. The evolution of the plastid chromosome in land plants: gene content, gene order, gene function. *Plant Molecular Biology* 76: 273–297.
- 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* 111: E4859–E4868.
- Wurdack, K. J., and C. C. Davis. 2009. Malpighiales phylogenetics: gaining ground on one of the most recalcitrant clades in the angiosperm tree of life. *American Journal of Botany* 96: 1551–1570.
- Zanis, M., D. E. Soltis, P. S. Soltis, S. Matthews, and M. J. Donoghue. 2002. The root of the angiosperms revisited. *Proceedings of the National Academy of Sciences, USA* 99: 6848–6853.
- Zanne, A. E., D. C. Tank, W. K. Cornwell, J. M. Eastman, S. A. Smith, R. G. FitzJohn, D. J. McGlenn, et al. 2014. Three keys to the radiation of angiosperms into freezing environments. *Nature* 506: 89–92.
- Zhong, B., O. Deusch, V. V. Goremykin, D. Penny, P. J. Biggs, R. A. Atherton, S. V. Nikiforova, and P. J. Lockhart. 2011. Systematic error in seed plant phylogenomics. *Genome Biology and Evolution* 3: 1340–1348.
- Zhong, B., T. Yonezawa, Y. Zhong, and M. Hasegawa. 2010. The position of Gnetales among seed plants: overcoming pitfalls of chloroplast phylogenomics. *Molecular Biology and Evolution* 27: 2855–2863.