Syst. Biol. 63(3):368–382, 2014
© The Author(s) 2014. Published by Oxford University Press, on behalf of the Society of Systematic Biologists. All rights reserved.
For Permissions, please email: journals.permissions@oup.com
DOI:10.1093/sysbio/syt108
Advance Access publication January 3, 2014

Another Look at the Root of the Angiosperms Reveals a Familiar Tale

BRYAN T. DREW[1,2,*], BRAD R. RUHFEL[1,2,3], STEPHEN A. SMITH[4], MICHAEL J. MOORE[5], BARBARA G. BRIGGS[6], MATTHEW A. GITZENDANNER[1], PAMELA S. SOLTIS[2], AND DOUGLAS E. SOLTIS[1,2]

[1]*Department of Biology, University of Florida, Gainesville, FL 32611, USA;* [2]*Florida Museum of Natural History, University of Florida, Gainesville, FL 32611, USA;* [3]*Department of Biological Sciences, Eastern Kentucky University, Richmond, KY 40475, USA;* [4]*Department of Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, MI 48105, USA;* [5]*Department of Biology, Oberlin College, Oberlin, OH 44074, USA; and* [6]*National Herbarium of New South Wales, Botanic Gardens Trust, Sydney, NSW 2000, Australia*

**Correspondence to be sent to: Department of Biology, University of Florida, Gainesville, FL 32611-8525, USA; E-mail: bdrewfb@yahoo.com*

Received 27 May 2013; reviews returned 19 July 2013; accepted 16 December 2013
Associate Editor: Vincent Savolainen

Abstract.—Since the advent of molecular phylogenetics more than 25 years ago, a major goal of plant systematists has been to discern the root of the angiosperms. Although most studies indicate that *Amborella trichopoda* is sister to all remaining extant flowering plants, support for this position has varied with respect to both the sequence data sets and analyses employed. Recently, Goremykin et al. (2013) questioned the "*Amborella*-sister hypothesis" using a "noise-reduction" approach and reported a topology with *Amborella* + Nymphaeales (water lilies) sister to all remaining angiosperms. Through a series of analyses of both plastid genomes and mitochondrial genes, we continue to find mostly strong support for the *Amborella*-sister hypothesis and offer a rebuttal of Goremykin et al. (2013). The major tenet of Goremykin et al. is that the *Amborella*-sister position is determined by noisy data—that is, characters with high rates of change and lacking true phylogenetic signal. To investigate the signal in these noisy data further, we analyzed the discarded characters from their noise-reduced alignments. We recovered a tree identical to that of the currently accepted angiosperm framework, including the position of *Amborella* as sister to all other angiosperms, as well as all other major clades. Thus, the signal in the "noisy" data is consistent with that of our complete data sets—arguing against the use of their noise-reduction approach. We also determined that one of the alignments presented by Goremykin et al. yields results at odds with their central claim—their data set actually supports *Amborella* as sister to all other angiosperms, as do larger plastid data sets we present here that possess more complete taxon sampling both within the monocots and for angiosperms in general. Previous unpartitioned, multilocus analyses of mitochondrial DNA (mtDNA) data have provided the strongest support for *Amborella* + Nymphaeales as sister to other angiosperms. However, our analysis of third codon positions from mtDNA sequence data also supports the *Amborella*-sister hypothesis. Finally, we challenge the conclusion of Goremykin et al. that the first flowering plants were aquatic and herbaceous, reasserting that even if *Amborella* + water lilies, or water lilies alone, are sister to the rest of the angiosperms, the earliest angiosperms were not necessarily aquatic and/or herbaceous. [Angiosperms; *Amborella*; Nymphaeales; plastid genome; water lilies.]

Researchers have long sought to discern the root of flowering plants and, with it, the patterns of change in floral and vegetative characters. Prior to molecular systematics, it was generally accepted that taxa in the traditionally recognized subclass Magnoliidae (Cronquist 1981, 1988; Thorne 1992; Takhtajan 1997) represented ancestral (i.e., 'primitive') flowering plants. However, with the advent of molecular systematics over 25 years ago, definitively answering the question of which extant angiosperm lineage(s) are sister to the remainder finally became tractable. For example, molecular analyses have shown that the Magnoliidae of Cronquist and Takhtajan are polyphyletic. Moreover, molecular evidence, as well as DNA + morphology (Doyle and Endress 2000), strongly supports a placement of Amborellaceae, Nymphaeales, and Austrobaileyales as early-diverging groups of angiosperms, although the relative branching order of Amborellaceae and Nymphaeales (water lilies) has proven contentious. Most large-scale molecular phylogenetic studies (e.g., Parkinson et al. 1999; Qiu et al. 1999; Soltis et al. 1999, 2000; Zanis et al. 2002; Borsch et al. 2003; Hilu et al. 2003; Stefanović et al. 2004; Leebens-Mack et al. 2005; Qiu et al. 2005; Jansen et al. 2007; Moore et al. 2007; Moore et al. 2010; Graham and Iles 2009; Lee et al. 2011; Moore et al. 2011; Soltis et al. 2011; Zhang et al. 2012) from the past 15 years have found that *Amborella* alone is sister to all other angiosperms (the "*Amborella*-sister hypothesis"). However, internal support for this placement varies among studies, and some analyses support *Amborella* + Nymphaeales as sister to all other extant angiosperms (e.g., Barkman et al. 2000; Soltis et al. 2000; Goremykin et al. 2009b; Finet et al. 2010; Qiu et al. 2010; Wodniok et al. 2011; Goremykin et al. 2013).

In the early to mid-1990s, molecular phylogenetic studies using plastid DNA sequences (primarily the slowly evolving *rbcL* gene) were ambiguous with regard to the angiosperm root (Chase et al. 1993; Qiu et al. 1993). However, as additional plastid gene regions were added, plastid analyses generally supported *Amborella* alone as sister to all other extant angiosperms, followed by Nymphaeales and then Austrobaileyales (Borsch et al. 2003; Hilu et al. 2003; Soltis et al. 2011). These relationships have received strong support in analyses employing complete plastid genomes when sampling of angiosperms was adequate (Leebens-Mack et al. 2005; Cai et al. 2006; Jansen et al. 2007; Moore et al. 2007, 2010).

In contrast to the widespread use of mitochondrial DNA (mtDNA) in animals, the use of mtDNA in plant phylogenetics has been stymied by various issues such as complex genome structure (Palmer and Herbon 1988;

Downloaded from https://academic.oup.com/sysbio/article/63/3/368/1648791 by guest on 19 April 2024

Andre et al. 1992), low nucleotide sequence variability (Palmer and Herbon 1988; Palmer 1992), and horizontal gene transfer (Bergthorsson et al. 2003; Brown 2003; Bergthorsson et al. 2004; Richardson and Palmer 2007; Keeling and Palmer 2008; Goremykin et al. 2009a; Renner and Bellot 2012; Xi et al. 2013). Despite these hindrances, mtDNA sequence data have been used in several deeper-scale phylogenetic studies, mostly in combination with plastid DNA and nuclear ribosomal DNA (nrDNA; Qiu et al. 1999; Barkman et al. 2000; Zanis et al. 2002; Qiu et al. 2005; Soltis et al. 2011), but also alone (Qiu et al. 2010). When mtDNA data have been used in concert with sequences from other genomic compartments, the combined data have usually supported the *Amborella*-sister hypothesis (e.g., Qiu et al. 1999; Zanis et al. 2002; Qiu et al. 2005; Soltis et al. 2011). In contrast, the recent study of Qiu et al. (2010), which included sequence data from 4 mtDNA genes for 380 seed plants, estimated (using total evidence without gene or codon partitioning) that *Amborella* + Nymphaeales form a clade that is sister to all other angiosperms.

Nuclear DNA sequences have not been widely used in large-scale phylogenetic studies of angiosperms to date, and until recently have largely been confined to the 18S and 26S nrDNA regions (Soltis et al. 1997, 1999; Qiu et al. 1999; Barkman et al. 2000; Soltis et al. 2000; Zanis et al. 2002; Qiu et al. 2005; Soltis et al. 2011) and phytochrome genes (Mathews and Donoghue 1999). Nearly all of these analyses found *Amborella* or *Amborella* + Nymphaeales as the sister to all other angiosperms. Recently, however, several large-scale phylogenetic studies have used low copy nuclear genes (Finet et al. 2010; Lee et al. 2011; Morton 2011; Wodniok et al. 2011; Zhang et al. 2012), with some supporting the *Amborella*-sister hypothesis (Lee et al. 2011, ML BS >65%; Zhang et al. 2012, ML BS = 83%, posterior probability [PP] = 0.99) and others recovering *Amborella* + Nymphaeales as sister to all other extant angiosperms, although with BS support less than 50% in one case (Finet et al. 2010; ML BS = 49%) and with only 7 angiosperms sampled in the other (Wodniok et al. 2011).

Although most research has found either *Amborella* alone or *Amborella* + Nymphaeales as sister to all other angiosperms, a few DNA studies have suggested alternative rootings such as a topology that moderately supported *Ceratophyllum* as sister to all other angiosperms (e.g., Chase et al. 1993 [*rbcL*]; Savolainen et al. 2000 [*rbcL* + *atpB*]; Morton 2011 [*xdh*]), but these appear to be the result of spurious rooting based on *rbcL* (e.g., Chase et al. 1993), or low taxon density (Morton 2011). Additionally, some researchers have also proposed that monocots may be sister to all other angiosperms (e.g., Burger 1981; Goremykin et al. 2003). Studies that have combined data from different organellar genomes and that have employed broad taxonomic sampling (e.g., Qiu et al. 1999; Zanis et al. 2002; Qiu et al. 2005; Soltis et al. 2011) have generally concluded that *Amborella* alone is sister to all other extant angiosperms, with the notable exception of Barkman et al. (2000), who employed the noise-reducing program RASA (relative apparent synapomorphy analysis; Lyons-Weiler et al. 1996) and found support for *Amborella* + Nymphaeales as sister to other angiosperms. However, the results of Barkman et al. (2000) varied depending on the method of analysis (e.g., parsimony vs. likelihood).

As a consensus emerged regarding the position of *Amborella* as sister to other extant angiosperms, researchers (e.g., Doyle and Endress 2000; Feild et al. 2000) sought to identify morphological features to corroborate the finding. A suite of morphological characters, including floral merosity, floral organization, phyllotaxy, perianth differentiation, stamen and carpel morphology, nodal anatomy, presence of vessel elements, and embryo sac formation, has been analyzed in light of the molecular angiosperm phylogeny, but no feature or combination of features unambiguously discriminates between *Amborella* alone vs. *Amborella* + Nymphaeales as sister to all other angiosperms (Baily and Swamy 1948; Carlquist 1987; Doyle and Endress 2000; Feild et al. 2000; Herendeen and Miller 2000; Carlquist and Schneider 2001, 2002; Soltis et al. 2005; Doyle 2008; reviewed in Endress and Doyle 2009; Doyle 2012). In fact, many of these features consistently support Amborellaceae, Nymphaeales, and Austrobaileyales as sisters to all other angiosperms, but without providing clear insights into the branching order. Although some researchers favor a woody, rather than herbaceous, origin for angiosperms because all extant gymnosperms (the sister group of angiosperms) and nearly all early lineages of angiosperms are woody, the ancestral habit of the angiosperms remains equivocal in rigorous character-state analyses (e.g., Soltis et al. 2005, 2008a; Doyle 2012), regardless of whether *Amborella* alone or *Amborella* + Nymphaeales is placed as sister to all other living flowering plants. As a result, two hypotheses are currently advanced for the ancestral angiosperms based on these reconstructions—early angiosperms may have been either understory shrubs ("dark and disturbed"; Feild et al. 2004; Coiffard et al. 2007; Soltis et al. 2008a) or aquatic ("wet and wild"; Coiffard et al. 2007; Soltis et al. 2008a).

Angiosperms contain both ancient clades and recent radiations, and there is heterogeneity of evolutionary rates both among genes and among lineages (Bousquet et al. 1992; Qiu et al. 2000; Soltis et al. 2002). Some sites, when viewed across all angiosperms, are highly variable and others are nearly constant (Chase and Albert 1998; Olmstead et al. 1998; Soltis and Soltis 1998). Various methods have been developed to try to accommodate this heterogeneity and these highly variable sites (e.g., Barkman et al. 2000; Rokas et al. 2003; Delsuc et al. 2005; Parks et al. 2012; Rajan 2013). Recently, the effects of iteratively removing "saturated" nucleotide positions ("noise reduction") in plastid DNA were investigated by Goremykin et al. (2009b, 2013). They used a novel method (described in detail in Goremykin et al. (2010, 2013)) to sort their alignment according to variability and then excluded the least conserved (most variable) characters before analyses. While insufficient taxon sampling has been shown to adversely affect results (e.g., Hillis 1996;

Downloaded from https://academic.oup.com/sysbio/article/63/3/368/1648791 by guest on 19 April 2024

Zwickl and Hillis 2002; Soltis and Soltis 2004; Soltis et al. 2004; Stefanović et al. 2004; Heath et al. 2008), and increased sampling has been advocated as the solution, there is no consensus on how to handle variable sites. Although the effects of iteratively removing variable plastid DNA sites have been explored (e.g., Delsuc et al. 2005; Philippe et al. 2005; Regier and Zwick 2011; Rajan 2013), no general agreement has been reached as to methodology, or perhaps more importantly, how much and which data to remove. Furthermore, the impact on phylogenetic inference from sparse taxon sampling combined with iterative reduction of the most variable sites is even less clear.

Goremykin et al. (2013) reported that removing the most variable sites (2000 out of 40,553) in a data set of 31 taxa yields *Amborella* + Nymphaeales as sister to all other angiosperms (PP = 1.00). Moreover, they also claim that phylogenetic relationships throughout the tree are affected by the number of highly variable sites they remove (Goremykin et al. 2009b, 2013). That is, as characters from the "noisy" end of the alignment are iteratively removed, even otherwise well-supported monocot and eudicot relationships collapse. Their approach conflates character variability with loss of phylogenetic signal. We argue instead that the relationship among character variability, homoplasy and its distribution, and taxon sampling determines whether or not highly variable characters carry phylogenetic signal. Thus, the utility of characters for phylogenetic analysis cannot be determined a priori on the basis of character variability. It is especially noteworthy that although the explicit goal of Goremykin et al. (2009b, 2013) was to "reduce noise" in the data, they used nonvascular plants (Goremykin et al. 2009b) and Gnetales (Goremykin et al. 2013), a gymnosperm clade characterized by very long branches, as outgroups. The inclusion of these groups seemingly adds little, if anything, toward resolving the angiosperm root, but increases the potential for variation across the data set, possibly resulting in "noisy" characters across the matrix.

Determining the root of extant angiosperms is important not only because it will drive how we think about angiosperm evolution as a whole, but also because it will orient specific character-state reconstructions and thus permit inferences of ancestral states. In this article, we analyze a suite of data sets, including plastid DNA, mtDNA, and nuclear gene regions, to address the root of flowering plants. We re-examine some previously published studies (Qiu et al. 2010; Soltis et al. 2011; Goremykin et al. 2013) and also present several new data sets, including the largest complete plastid genome data set yet assembled for angiosperms. We then use these data sets to address the following questions: (i) Which angiosperm lineage is sister to all others? (ii) Do the noise-reduced data alignments presented by Goremykin et al. (2013) convincingly show that Nymphaeales, either alone or with *Amborella*, are sister to the remaining angiosperms? (iii) What are the implications of these phylogenetic analyses for a terrestrial versus aquatic origin of angiosperms?

Materials and Methods

Genomic Sequencing, Taxon Sampling, and Sequence Alignment

The goals of this study were to ascertain the most likely root of flowering plants using plastid and mitochondrial DNA sequence data, and also to assess how taxon and character sampling affect this inference. We therefore designed a series of analyses and taxon sampling schemes to examine the effects that different codon partition analyses, taxon sampling, outgroup selection, and gene sampling would have on the inference of the angiosperm root. We also included, as did Goremykin et al. (2013), a recently recognized member of Nymphaeales, *Trithuria* (Hydatellaceae), in most of our analyses. *Trithuria,* formerly placed in Centrolepidaceae, one of 16 families of the monocot clade Poales, was shown to fall instead in Nymphaeales (Saarela et al. 2007). With the exception of Goremykin et al. (2013), *Trithuria* has not previously been included in genomic-scale plastid DNA phylogenetic studies. We obtained fresh plant material of *Trithuria filamentosa* (B.G. Briggs 9859; GenBank# KF696682) for whole plastome sequencing. Purified plastid DNA was isolated using sucrose gradient ultracentrifugation and amplified via rolling circle amplification (RCA) following the protocols in Moore et al. (2006, 2007). The RCA product was sequenced at the Interdisciplinary Center for Biotechnology Research (University of Florida) using the Genome Sequencer 20 System (GS 20; 454 Life Sciences, Branford, CT, USA), as outlined in Moore et al. (2006, 2007). Gaps between the contigs derived from 454 sequence assembly were bridged by designing custom primers near the ends of the GS 20 contigs for PCR and conventional capillary-based sequencing. The completed plastid genome was annotated using DOGMA (Wyman et al. 2004) and subsequently aligned to the other sequences using MAFFT v. 6.859 (Katoh et al. 2002). The entire alignment was inspected in Mesquite (Maddison and Maddison 2011) to ensure that all nucleotide positions were in reading frame.

In total, coding plastome sequence data from 235 seed plants with complete or nearly complete plastid genome sequences were downloaded from GenBank. Only one exemplar per genus was downloaded; when multiple accessions from a genus were available, we chose the taxon with the most complete sequence data. Our sampling included virtually all angiosperm orders sensu APG III (2009) and up to 19 gymnosperm genera as outgroups. To investigate the effect of long branches and accompanying alignment uncertainty typically associated with Gnetales, these taxa were not included in some of our analyses. The first plastid DNA data set consisted of sequences for 235 taxa that were downloaded from GenBank. This alignment of

Downloaded from https://academic.oup.com/sysbio/article/63/3/368/1648791 by guest on 19 April 2024

235 accessions (216 angiosperms and 19 gymnosperm outgroups) contained 78 plastid DNA genes and 58,218 characters. A second alignment of 236 taxa includes our previously unreported *Trithuria filamentosa* plastid genome with the above (59,944 aligned characters). Four additional plastid DNA alignments are as above except that we: (i) removed Gnetales accessions (78 genes, 233 taxa, 58,935 characters), (ii) removed Gnetales accessions and excluded *rps16* and all *ndh* genes, which have been lost in many gymnosperms (Braukmann et al. 2009) (66 genes, 233 taxa, 48,222 characters), (iii) removed the gymnosperms that are missing *ndh* and *rps16* genes (78 genes, 222 taxa, 58,860 characters), and (iv) excluded all characters except the *ndh* genes (11 genes, 177 taxa, 10,479 characters). For the latter alignment, we deleted all taxa that had more than 10% missing data, and the latter two alignments included only the five gymnosperms from our data set that have not lost *ndh* genes (*Cycas taitungensis, Ginkgo biloba, Cephalotaxus wilsoniana, Taiwania cryptomerioides*, and *Cryptomeria japonica*). Because *ndh* genes were not included in the analyses of Goremykin et al. (2013), we considered it important to investigate what effect, if any, their inclusion or exclusion had on the topology.

We attempted to run the Goremykin et al. (2010, 2013) NoiseReductor scripts several times on our full data set, as well as a reduced data set of 25 taxa and 56,838 characters. However, the Goremykin et al. (2010, 2013) scripts were never able to run to completion, requiring more than 16 GB of RAM and long-run times, despite some attempts to modify the scripts to improve performance. In the end, we were unable to run Goremykin's (2010, 2013) Perl scripts to completion on either the full or reduced data set, and we therefore substituted a new Perl script (Miao et al., unpublished data) that sorts characters according to the sum of pairwise distances (least variable sites at the beginning of the alignment, most variable sites at the end; see Supplementary Files S15 and S17 at http://datadryad.org, doi:10.5061/dryad.68n85). This method is similar to that used by Goremykin et al. (2010, 2013) to obtain their sorted alignments. We ran the Miao et al. script to sort two of our alignments: (i) the 236-taxon, 78-gene alignment and (ii) the 222-taxon, 78-gene alignment.

We also analyzed three plastid DNA alignments from Goremykin et al. (2013). First, we created an alignment using the most variable 2000 characters from the "observed variability" (OV) sorted alignment (S3 from Goremykin et al. (2013); in Dryad). This corresponded to the final 2000 characters from their 40,553-character sorted alignment. Next, we used their 31,674-character alignment that maintained the reading frame (S4 from Goremykin et al. (2013); from "A MUSCLE alignment of translated nucleotide sequences from 56 individual Fasta files"; in Dryad). This in-frame alignment was produced using a different approach than their "noise-reduction" method. Although no trees were shown from this alignment in their paper, Goremykin et al. (2013) stated: "similar analytical results were obtained for both alignments" Third, we reanalyzed their 25,246-character alignment composed of first and second codon positions from the in-frame alignment (S2 from Goremykin et al. (2013); in Dryad).

Furthermore, we reanalyzed the published mtDNA alignment of Qiu et al. (2010). Because the alignment used in their analysis was not in frame, we modified their published alignment so the data could also be partitioned by codon position. These alignment adjustments consisted of simple manual modifications such as moving characters at the beginning/end of indels and ensuring all character blocks and gaps occurred in multiples of three to maintain the reading frame. Our final mtDNA alignment had 356 taxa and 7752 characters. Each gene (*atp1*, *matR*, *nad5*, *rps3*) was also analyzed individually to see if they exhibited different phylogenetic signal. We also analyzed an alignment that excluded Gnetales to investigate whether the long-branch and alignment uncertainty associated with Gnetales would affect the root of the angiosperm phylogeny. This Gnetales-absent alignment had 354 taxa and 7701 characters.

Finally, we assessed per-site variation between alternative topologies by examining data from the 17-gene (nrDNA, plastid DNA, mtDNA) alignment from Soltis et al. (2011), both with mtDNA gene regions present and after their removal. Goremykin et al. (2013) questioned the results reported by Soltis et al. (2011), stating they "were unable to confirm this [*Amborella* sister to remaining angiosperms] finding" instead inferring "a phylogenetic tree wherein a clade comprising *Amborella*, *Trithuria* and Nymphaeales received 94% non-parametric bootstrap support"

Phylogenetic Analyses

Maximum likelihood (ML) analyses were conducted using a parallel version (RAxML-PTHREADS-SSE3) of RAxML v.7.3.0 (Stamatakis et al. 2005, 2008) as implemented on the high-performance computing cluster at the University of Florida. To assess the effects that different phylogenetic methods might have on our results, ML and Bayesian analysis were used to analyze the concatenated DNA alignment from Goremykin et al. (2013) that maintains the reading frame (Supplementary Data S4 from Goremykin et al. 2013). Bayesian analysis was performed in MrBayes v.3.1.2 (Huelsenbeck and Ronquist 2001) as implemented on the Cyberinfrastructure for Phylogenetic Research (CIPRES) cluster (http://www.phylo.org/; last accessed January 12, 2014). All analyses were run for 5 million generations with the covarion-like model option turned on and GTR + Γ as the model of evolution. The first 25% of trees were discarded as burnin, and gaps were treated as missing data in all the analyses. Convergence and mixing of two independent runs were assessed using Tracer 1.5 (Suchard et al. 2001; Rambaut and Drummond 2009). After removing burnin, effective sample size (ESS) values of the combined sump files were ~2000, and

Downloaded from https://academic.oup.com/sysbio/article/63/3/368/1648791 by guest on 19 April 2024

visual observation of the associated trace files indicated that the separate runs had converged. To assess whether our runs achieved convergence, we checked that the standard deviation of split frequencies fell below 0.01.

Alignments were analyzed using several partitioning schemes. For the six plastid DNA alignments assembled for this study, as well as the mtDNA alignment from Qiu et al. (2010), the data were partitioned and analyzed by codon position (first and second positions only [third positions excluded], third positions only [first and second positions excluded], all three codon positions included), gene, and codon position (different substitution rates were allowed for first, second, and third positions) + gene. For the reanalysis of the Goremykin et al. (2013) reduced-variability, in-frame data set, we analyzed the alignment by codon position (first and second positions only, third positions only, all three positions combined), but were unable to partition the alignment by gene because the boundaries of the gene regions are unknown. For the alignment featuring the 2000 most variable characters as specified by the noisereduction program of Goremykin et al. (2013; supplementary alignment S3), we conducted a single ML analysis that included all characters. No modifications were performed on the Goremykin et al. (2013) alignments downloaded from Dryad.

To assess the number of deleted variable characters necessary to achieve an *Amborella* + Nymphaeales topology, we iteratively removed the most variable characters from our two sorted alignments in 0.5% increments. This amounted to a removal of 295 characters per iteration in our 236-taxon alignment and 294 characters per iteration in the 222-taxon alignment. After removing the variable characters, we analyzed the reduced alignments in RAxML as described above. Furthermore, to investigate how our sorting script compared with the sorting scheme of Goremykin et al. (2013), we used our sorting script on the unsorted 40,553-character alignment from Goremykin et al. (2013) and followed their general approach of iteratively removing the most variable 250 characters. We then analyzed the reduced alignments in RAxML and compared the results from our sorting script to the results they obtained using their noise-reduction protocol. Files containing all of our original alignments and corresponding character partitions used for this study have been deposited at datadryad.org.

Another way to examine where the differences lie between the competing hypotheses of *Amborella* sister to Nymphaeales versus *Amborella* sister to all other angiosperms is to examine the differences in site likelihoods between trees constrained to support the different hypotheses (Evans et al. 2010; Smith et al. 2011). This analysis allows for the examination of patterns in the support between the two trees. With this comparison, the expectation is that the sum of the differences will equal the difference in total *ln* likelihood. To conduct these analyses, we constructed two constraint trees based on the alignment of Soltis et al. (2011): one with *Amborella* sister to the rest of the angiosperms and one with *Amborella* sister to Nymphaeales. We ran at least 10 ML analyses using these two constraint trees with RAxML v. 7.2.8 using the GTR + Γ model (used for Figure 6c in Goremykin et al. 2013) of molecular evolution. For each constraint, we analyzed the data set with and without mtDNA data. Using the ML trees, we calculated the site likelihoods and the differences between the trees with *Amborella* sister to all other angiosperms and *Amborella* sister to Nymphaeales.

Results

Most of the alignments used in this study contain ca. 200-350 taxa, and thus the resultant trees are difficult to view on a single page. Because the purpose of this study is to evaluate only a small portion of the larger trees (at the root of the angiosperms), the pertinent results from this study have been summarized in Table 1, and the relevant sections of the trees are summarized in Figures 1–4. The only results presented in this article not shown in Table 1 are from the 2000-character plastid DNA alignment from Goremykin et al. (2013), which was not analyzed in partitions. The individual trees are available as Supplementary Data at datadryad.org.

Plastid DNA Analyses

For all original plastid DNA alignments and partitions analyzed here, *Amborella* is sister to a clade of all other angiosperms in the ML trees (Fig. 1, Table 1). The unpartitioned analyses using all three codon positions and only third codon positions without partitioning by gene all found 100% bootstrap support for *Amborella* sister to all other angiosperms (i.e., all angiosperms except *Amborella* formed a clade with 100% BS) as did all the analyses of these positions partitioned by gene. Five of our six analyses that included just first and second codon positions supported *Amborella* as the sole sister to all other extant angiosperms (bootstrap values for all other angiosperms ranged from 66% to 100%). The 66-gene, 233-taxon alignment that excluded the *ndh* and *rps16* genes was the lone outlier among our plastid DNA data partitions in that the bootstrap analysis of first and second codon positions weakly supported (BS=57%) a clade of *Amborella* + Nymphaeales as sister to the remaining angiosperms. As previously mentioned, however, the single best ML tree shows *Amborella* sister to all other angiosperms (Fig. 1d). For all data alignments other than first and second codon positions, *Amborella* alone was sister to all remaining angiosperms, which formed a clade with 100% bootstrap support.

Our analysis of the in-frame plastid DNA alignment from Goremykin et al. (2013; S4 in Dryad) produced similar results to those noted above for the analyses of our own plastid DNA data sets. The ML trees and the majority-rule Bayesian consensus trees that we obtained from their data each showed *Amborella* alone as sister to all other angiosperms in all partitioned analyses

Downloaded from https://academic.oup.com/sysbio/article/63/3/368/1648791 by guest on 19 April 2024

TABLE 1. Summary of *Amborella* placement relative to Nymphaeales and the rest of angiosperms

Alignment	Three codon positions	First and second codon positions	Third codon positions	Partitioned by gene	Partitioned by gene and codon
cpDNA-235 taxa, (no *Trithuria*), 78 genes 58,218 chars	*Ambo.* sister; 100% BS for rest of angio	*Ambo.* sister; 84% BS for rest of angio	*Ambo.* sister; 100% BS for rest of angio	*Ambo.* sister; 100% BS for rest of angio	*Ambo.* sister; 100% BS for rest of angio
cpDNA-236 taxa, (w/*Trithuria*), 78 genes, 58,950 chars	*Ambo.* sister; 100% BS for rest of angio	*Ambo.* sister; 66% BS for rest of angio	*Ambo.* sister; 100% BS for rest of angio	*Ambo.* sister; 100% BS for rest of angio	*Ambo.* sister; 100% BS for rest of angio
cpDNA-233 taxa, 78 genes, 58,935 chars	*Ambo.* sister; 100% BS for rest of angio	*Ambo.* sister; 76% BS for rest of angio	*Ambo.* sister; 100% BS for rest of angio	*Ambo.* sister; 100% BS for rest of angio	*Ambo.* sister; 100% BS for rest of angio
cpDNA-233 taxa, 66 genes, 48,222 chars	*Ambo.* sister; 100% BS for rest of angio	*Ambo.* + Nymphaeales 57% BS; **	*Ambo.* sister; 100% BS for rest of angio	*Ambo.* sister; 100% BS for rest of angio	*Ambo.* sister; 100% BS for rest of angio
cpDNA-222 taxa, 78 genes, 58,860 chars	*Ambo.* sister; 100% BS for rest of angio	*Ambo.* sister; 100% BS for rest of angio	*Ambo.* sister; 100% BS for rest of angio	*Ambo.* sister; 100% BS for rest of angio	*Ambo.* sister; 100% BS for rest of angio
cpDNA-ndh genes, 177 taxa; 10,479 chars	*Ambo.* sister; 100% BS for rest of angio	*Ambo.* sister; 76% BS for rest of angio	*Ambo.* sister; 100% BS for rest of angio	*Ambo.* sister; 100% BS for rest of angio	*Ambo.* sister; 100% BS for rest of angio
Goremykin et al. (2013) align. S4: ML; 31,674 chars	*Ambo.* sister; 97% BS for rest of angio	*Ambo.* sister; 53% BS for rest of angio; **	*Ambo.* sister; 98% BS for rest of angio	NA	NA
Goremykin et al. (2013) align. S4: BI; 31,674 chars	*Ambo.* sister; 1.00 PP for rest of angio	*Ambo.* sister; 0.92 PP for rest of angio	*Ambo.* sister; 1.00 PP for rest of angio	NA	NA
mtDNA-356 taxa, 4 genes, 7752 chars	*Ambo.* + Nymphaeales 71% BS	*Ambo.* + Nymphaeales 89% BS	*Ambo.* sister; 66% BS for other angio	*Ambo.* + Nymphaeales 78% BS	*Ambo.* + Nymphaeales 75% BS
mtDNA-no Gnetales, 4 genes, 7752 chars	*Ambo.* + Nymphaeales 68% BS	*Ambo.* + Nymphaeales 79% BS	*Ambo.* sister; 78% BS for other angio	*Ambo.* + Nymphaeales 61% BS	*Ambo.* + Nymphaeales 72% BS

**Indicates *Amborella* is sister to angiosperms in best ML tree.

(Figs. 2 and 3). Other than the discrepancy regarding the placement of *Amborella*, the trees we obtained were nearly identical in topology to the noise-reduced tree shown by Goremykin et al. (2013). The results from our analyses of the Goremykin et al. (2013) alignment varied from those obtained using our new plastid DNA data set and alignments in two major ways: (i) Support was lower in our analyses for the clade of angiosperms sister to *Amborella*, but still very high; ML BS = 97%, 98% and PP = 1.00 in analyses including all three codon positions and third codon positions, respectively, and (ii) *Amborella* was sister to all other angiosperms in all of our bootstrap analyses as well (albeit with only 53% bootstrap support for the data set of first and second codon positions). The analysis of the 2000 most variable plastid DNA characters identified by the Goremykin et al. (2013) noise-reduction protocol yielded a tree that was congruent with all other plastid DNA trees in our study (Fig. 4). *Amborella* was recovered as sister to all other angiosperms in the ML tree, with the clade of remaining angiosperms having 75% bootstrap support. The reanalysis of the 25,246-character alignment of only first and second codon positions yielded results similar to those found in Goremykin et al. (2013): the ML tree showed Nymphaeales as sister to all other angiosperms, with *Amborella* and *Illicium* as successive sisters to all remaining angiosperms. This relationship had less than 50% bootstrap support, however.

mtDNA Analyses

Our analyses of the mtDNA alignment of Qiu et al. (2010) generally yielded results similar to their published total-evidence study. A weakly to moderately supported clade (71%–89% bootstrap values) consisting of *Amborella* + Nymphaeales was sister to a clade containing all remaining angiosperms (Table 1). However, Qiu et al. (2010) did not examine data partitions. Although two partitioning schemes also supported *Amborella* + Nymphaeales (first and second codon positions, all three positions), the analysis that included only third codon positions found *Amborella* alone to be sister to all other angiosperms, which formed a weakly supported (BS = 66%) clade. Analyses of single mtDNA gene alignments yielded trees that were generally poorly resolved, especially within angiosperms. In the *atp1* analysis, the ML tree found Nymphaeales (BS = 76%) as sister to all angiosperms, but branches were very short throughout the tree, and this relationship did not receive BS > 50%; *Amborella* formed a clade with Austrobaileyales (BS = 70%). The *matR* analysis found *Amborella* + Nymphaeales (BS = 87%) as

Downloaded from https://academic.oup.com/sysbio/article/63/3/368/1648791 by guest on 19 April 2024

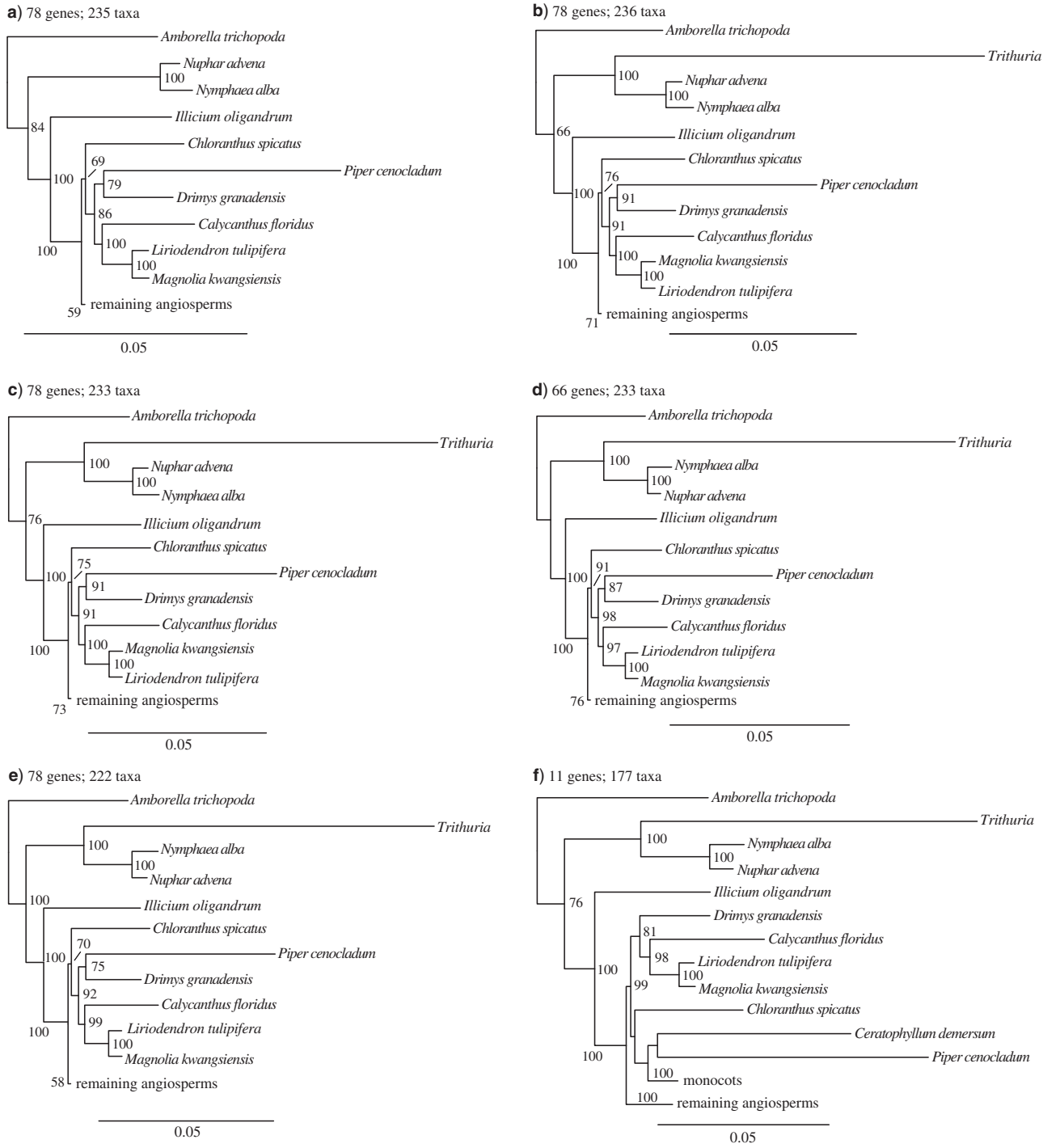

FIGURE 1. Best ML trees showing relationships of basal angiosperms. All alignments involve first and second codon positions only. a) 78-gene, 235-taxon plastid DNA alignment not including *Trithuria*. b) 78-gene, 236-taxon alignment including *Trithuria*. c) 78-gene, 233-taxon alignment that excludes Gnetales. d) 66-gene, 233-taxon alignment that excludes *ndh* and *rps16* genes. e) 78-gene, 222-taxon alignment that excludes gymnosperms that have lost *ndh* genes. f) 177-taxon alignment including only *ndh* genes. ML bootstrap values indicated near nodes (complete trees available at datadryad.org).

sister to the remaining angiosperms (BS = 100%). The *nad5* tree recovered a clade consisting of *Amborella* + Nymphaeales (BS = 83%) that was in turn sister to all other angiosperms (BS = 80%). The *rps3* analysis found *Amborella* as sister to the remaining angiosperms, which received BS support of 64%.

Removing Gnetales from the mtDNA analyses reduced support for *Amborella* + Nymphaeales with all

Downloaded from https://academic.oup.com/sysbio/article/63/3/368/1648791 by guest on 19 April 2024

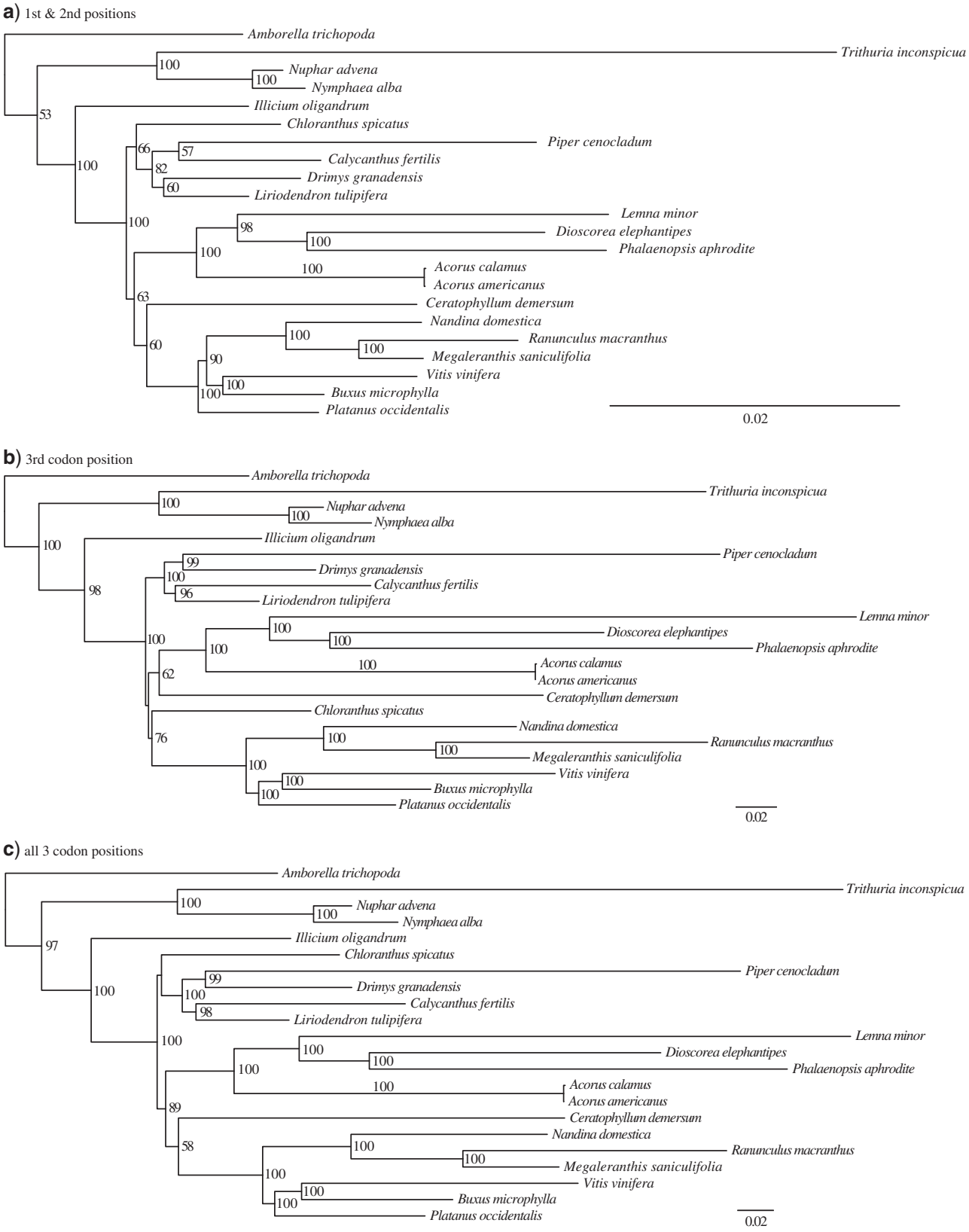

FIGURE 2. Phylograms resulting from ML analyses of Goremykin et al. (2013; S4 in Dryad) in-frame alignment. a) first and second codon positions. b) third codon positions. c) all three codon positions.

Downloaded from https://academic.oup.com/sysbio/article/63/3/368/1648791 by guest on 19 April 2024

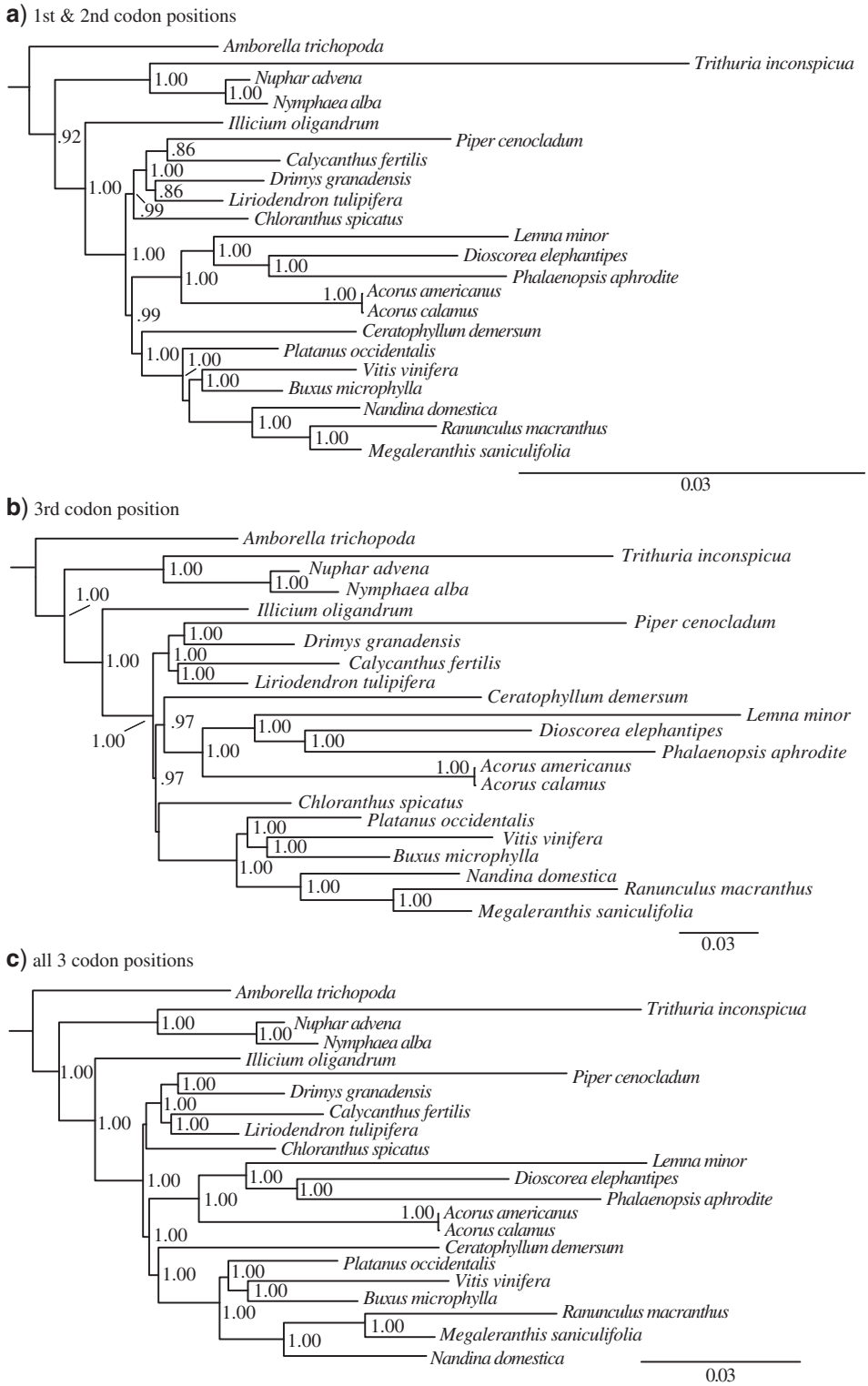

FIGURE 3. Phylograms resulting from Bayesian analyses of Goremykin et al. (2013; S4 in Dryad) in-frame alignment. a) first and second codon positions. b) third codon positions. c) all three codon positions.

Downloaded from https://academic.oup.com/sysbio/article/63/3/368/1648791 by guest on 19 April 2024

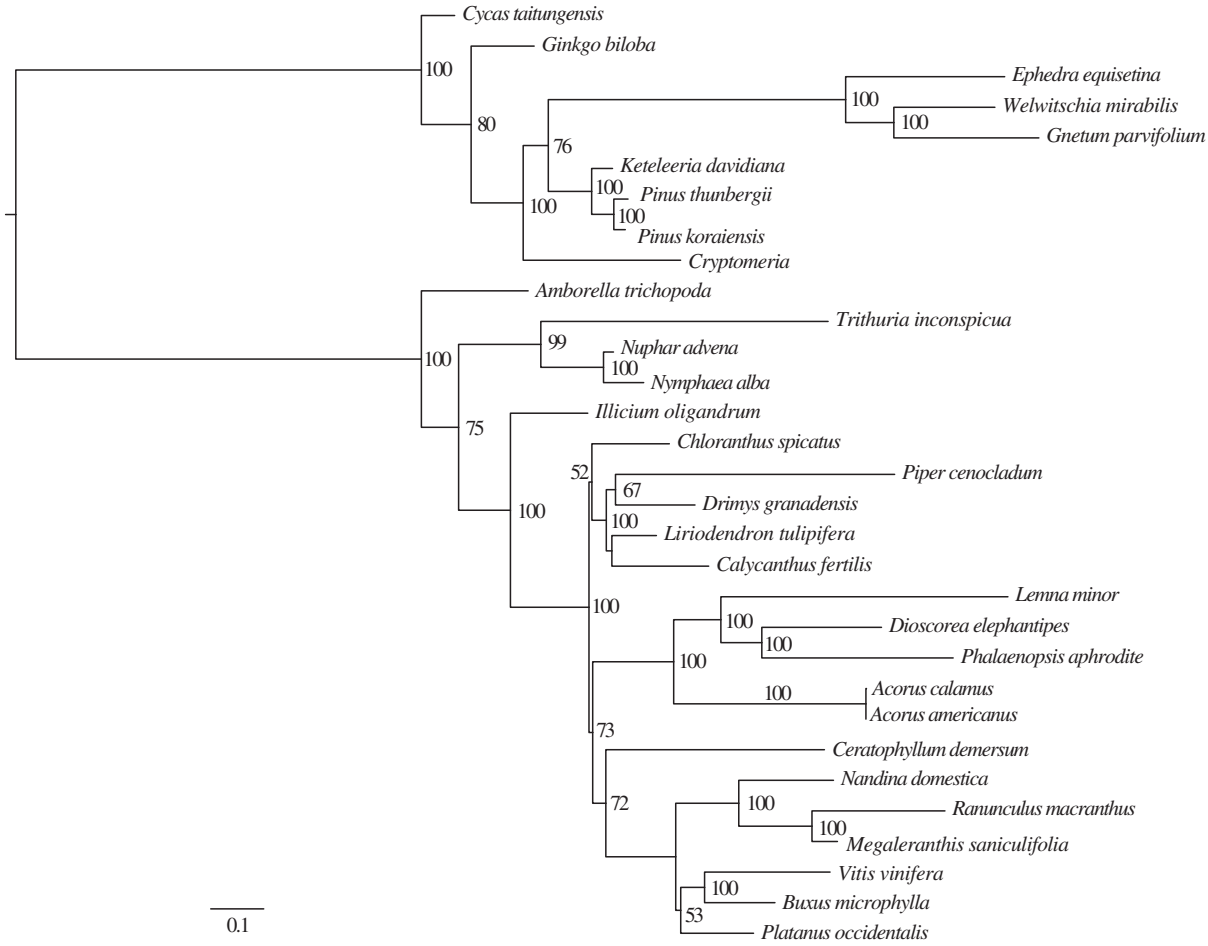

FIGURE 4. Phylogram showing ML analysis of deleted characters from noise-reduced alignment in Goremykin et al. (2013) (final 2000 characters from alignment S3 in Dryad).

codon positions included (BS = 68% vs. BS = 71% with Gnetales), first and second codon positions (BS = 79% vs. BS = 89%), gene-partitioned (BS = 61% vs. BS = 78%), and gene- and codon-partitioned (BS = 72% vs. BS = 75%) alignments and increased support for the topology of *Amborella* sister to all other angiosperms (BS = 78% vs. BS = 66%) in the analysis of the third codon position alignment (Table 1).

Analyses of Variability-Sorted Alignments

Using our variability-sorted, 236-taxon, 58,944-character alignment, it was necessary to remove 22% of the most variable positions (12,968 characters) to recover a topology showing *Amborella* + Nymphaeales as sister to the remaining angiosperms (S16 in Dryad). Using our sorted 222-taxon, 58,860-character alignment, 35% of the most variable positions (20,601 characters) needed to be deleted to achieve an *Amborella* + Nymphaeales clade (S18 in Dryad). In analyses of both data sets, monophyly of recognized clades (e.g., monocots, eudicots) broke down after removal of about 10% (~5900) of the most variable characters, and at the point at which an *Amborella* + Nymphaeales clade was recovered, virtually no structure remained in the backbone of either tree. After applying our variability sorting script to the 40,553-character alignment from Goremykin et al. (2013), we needed to remove 1750 of the most variable characters to find *Amborella* + Nymphaeales as sister to the remaining angiosperms, similar to the findings of Goremykin et al. (2013), who needed to remove 1250 sites to find this topology. This result suggests that the sorting script used here (Miao et al., unpublished data) behaves similarly to the sorting script used by Goremykin et al. (2010, 2013).

Analysis of Per-Site Likelihoods

The per-site likelihood results for analyses including mitochondrial genes found that the topology with *Amborella* sister to all other angiosperms is 19.4 log likelihood units better than a topology with *Amborella* sister to Nymphaeales (Supplementary Fig. S1). The results for analyses without mtDNA sequence data show that *Amborella* as sister to the rest of the angiosperms is favored by 8.4 log likelihood units (Supplementary

Downloaded from https://academic.oup.com/sysbio/article/63/3/368/1648791 by guest on 19 April 2024

TABLE 2. Per-site likelihood variation per gene region, based on ML tree constructed using the 17-gene alignment in Soltis et al. (2011)

Gene region	With mtDNA regions	Without mtDNA regions
18S	−2.678305	−1.371077
26S	22.568016	18.267124
atp1	7.354322	—
atpB	−8.169379	5.765824
matK	7.182720	5.765824
*mat*R	18.149114	—
nad5	−34.125334	—
ndhF	17.404057	9.612408
psbBTNH	−19.613889	9.065335
rbcL	16.653941	−51.84577
rpoC2	−7.595023	10.924675
rps16	1.057146	−4.009307
rps3	6.649084	—
rps4	−5.274964	−2.599938

Note: Values above 0 indicate gene regions that support *Amborella* sister to angiosperms, whereas values below 0 support *Amborella* as sister to the Nymphaeales.

Fig. S2). The significant differences in gene regions for these trees show 26S rDNA, *atpB*, *matK*, *ndhF*, *psbBTNH*, and *rboC2* supporting *Amborella* sister to the rest of the angiosperms, and *rbcL*, *rps16*, and *rps4* supporting the alternative of *Amborella* + Nymphaeales (Table 2). 18S rDNA weakly (not significantly) supports *Amborella* sister to Nymphaeales.

DISCUSSION

The ML trees from each of the 30 analyses from our six original plastid DNA alignments all show *Amborella* alone as sister to all remaining extant angiosperms. In only one instance, in our analysis of first and second codon positions from the 66-gene, 233-taxon alignment that excluded *rps16* and the *ndh* genes, did the *Amborella*-sister position receive less than 50% bootstrap support. Additionally, our analyses of the supplemental plastid DNA data set from Goremykin at al. (2013; S4 at datadryad.org) also found *Amborella* alone as sister to the remaining angiosperms in all three partitioned analyses (first and second codon position [BS = 53%, PP = 0.92], third codon position [BS = 97%, PP = 1.00], all three codon positions [BS = 98%, PP = 1.00]) using both ML and Bayesian approaches. Our analysis of the 2000 most variable characters that were deleted from analysis by Goremykin et al. (2013) also placed *Amborella* sister to all other angiosperms. More noteworthy, however, is that the tree we obtained from the 2000 characters that purportedly provide "incorrect" phylogenetic signal with regard to the angiosperm root gave essentially the same angiosperm tree as shown by all of the other partitioned analyses and is consistent with current hypotheses of angiosperm phylogeny (e.g., Jansen et al. 2007; Moore et al. 2007, 2010; Bremer et al. 2009; Soltis et al. 2011). Moreover, the only differences (within angiosperms) between the tree produced from the 2000 deleted characters and the tree produced by the 38,553 retained characters in the Goremykin et al. (2013) paper are the placement of *Ceratophyllum*, which was not strongly supported in either analysis, and the placement of *Amborella*. Our analyses of first and second codon positions of the Goremykin et al. (2013; alignment S4 on Dryad) data place *Amborella* alone as sister to the remaining angiosperms; this result contradicts reports by Goremykin et al. (2013) that removal of third codon positions (i.e., consideration of only first and second positions) consistently recovers an *Amborella* + Nymphaeales clade.

Our reanalysis of the in-frame alignment of Goremykin et al. (2013; S4 in Dryad), an alignment based on a procedure (Lockhart et al. 1996) that produces conservative alignments using a different method than their noise-reduction protocol, surprisingly recovered the *Amborella*-sister topology (Figs. 2 and 3). Because no results were shown from analysis of this alignment in Goremykin et al. (2013), it is unclear why this alignment was presented, given that it does not support their main conclusion. Our results for this alignment are at odds with their central claim: that *Amborella* is not the sister to all other extant angiosperms. Even the ML and Bayesian trees (which had weak BS/moderate PP support; 53% and 0.92, respectively) from analyses of first and second codon positions within this conservative alignment recovered the *Amborella*-sister topology. Our reanalysis of the 25,246-character data set of Goremykin et al. (2013) showed that Nymphaeales alone were sister to all other angiosperms, in agreement with their findings (although bootstrap support was less than 50%).

The analyses we performed using our variability sorting script (Miao et al., unpublished data) indicate that taxon sampling is extremely important in the sorting process. The 222-taxon data set differed from the 236-taxon data set in that the former alignment excluded 14 gymnosperm outgroup taxa (including Gnetales) that lacked *ndh* genes. All analyses of these two data sets conducted prior to sorting recovered *Amborella* as sister to all other angiosperms. In our 222-taxon alignment, an additional ~7044 variable (20,012 vs. 12,968) characters needed to be removed before an *Amborella* + Nymphaeales clade was found. These results show that outgroup choice (and number of outgroup taxa used) can have a dramatic effect on how variable characters are sorted, and that outgroup choice can influence interpretation of results. Furthermore, the fact that we needed to remove more than 20% of the most variable characters in our sorted alignments, versus ~4% for the Goremykin et al. (2013) alignment, indicates that the limited taxon sampling in Goremykin et al. (2013) affected their results.

Although the results of our unpartitioned mtDNA 4-gene analyses largely concurred with Qiu et al. (2010), our four-gene mtDNA analysis of only third codon positions supported *Amborella* as sister to the remaining angiosperms, and our analyses of individual genes yielded three different topologies at the angiosperm root. It is possible that analyses using additional mtDNA sequence data will find results that differ from those based on just four genes, in the same way that early

Downloaded from https://academic.oup.com/sysbio/article/63/3/368/1648791 by guest on 19 April 2024

rbcL studies and *rbcL* + *atpB* analyses (Chase et al. 1993; Qiu et al. 1993; Savolainen et al. 2000) yielded different topologies from those obtained with many more plastid DNA genes (e.g., Moore et al. 2007, 2010, 2011; Soltis et al. 2011). Our per-site likelihood analyses of the Soltis et al. (2011) 17-gene data set is in agreement with most other analyses reported in this article and reaffirms the original findings of Soltis et al. (2011) regarding the placement of *Amborella*. Given the results of our mtDNA analyses in this article, it is noteworthy that the per-site likelihood support for the *Amborella*-sister topology increased with the inclusion of the four mtDNA genes. These results suggest that the likelihood surface regarding the placement of *Amborella*, at least for this data set, is relatively flat. So, with longer searches, slight modifications of the model, and other variation, trees that result in small increases in likelihood may be found. Goremykin et al. (2013) discovered alternative topologies in their analysis of these same data, but possible explanations include a different model specification and nature of the likelihood surface for the data set.

Taxon sampling will greatly affect which characters are to be excluded under virtually any character-exclusion scheme. The fact that Goremykin et al. (2009b) included taxa known to have long branches (*Huperzia*, *Marchantia*, and *Psilotum*) as outgroups in their analyses seems misguided if the goal is to reduce saturation, particularly given that their inclusion does not seem necessary to answer questions about relationships within angiosperms. Moreover, others (Graham and Iles 2009) have cautioned against the use of Gnetales as an outgroup for angiosperms due to the long branch leading to Gnetales, yet three species of Gnetales were included by Goremykin et al. (2013). The use of highly divergent outgroups by Goremykin et al. (2009b, 2013) is noteworthy because their noise-reducing protocol discards characters based (in part) upon distance between the ingroup and outgroup: "The extreme branch length separating the outgroup and ingroup sequences is a property of the 2000 sites at the most varied end of the OV alignment" (Goremykin et al. 2013, p. 54). In addition, prior to even implementing the noise-reduction protocol, outgroup choice affected the Goremykin et al. (2013) alignment: "The resulting 122 alignment files were each manually edited, such that regions of low similarity between the ingroup and outgroup sequences were discarded" (Goremykin et al. 2013, p. 51). It is surprising, given the apparent importance of outgroup distance in the procedures of Goremykin et al. (2009b, 2013), that the authors were not more conservative with their outgroup selection. Furthermore, the noise-reduction approach described by Goremykin et al. (2010, 2013) essentially collapses branches by reducing the number of characters. As they point out, as characters are iteratively removed, *Amborella* forms a clade with Nymphaeales. As more characters are removed, *Illicium* forms a clade with the aforementioned two lineages, and as even more variable characters are removed "disintegration of the monocot and dicot clusters" are observed (Goremykin et al. 2009b), etc. What they show, in essence, is that removing characters leads to reduced phylogenetic signal, and that if enough characters are removed, even (virtually) universally accepted clades break down, a result that should surprise no one. With limited taxon sampling and highly divergent outgroups, as in the data sets of Goremykin et al. (2013), the effect of character removal on the resulting topology can be rapid and dramatic.

Alignment uncertainty and/or nucleotide saturation are important phenomena in phylogenetic analyses. Several studies (Barkman et al. 2000; Goremykin et al. 2009b; Finet et al. 2010; Qiu et al. 2010; Goremykin et al. 2013) have shown that using highly conserved alignments, slowly evolving genes, and/or noise-reduction protocols can yield alternative topologies within angiosperms. However, we argue that most studies, including most analyses presented here, favor an *Amborella*-sister position, even when dismissing third codon positions. Furthermore, regardless of whether *Amborella* alone is the sister to all other extant angiosperms or whether *Amborella* + Nymphaeales form a clade, one cannot infer the habit or habitat of the first angiosperms based on the morphology of extant taxa. Based on formal character-state analyses conducted to date, the ancestral habit and habitat of the first angiosperms remain equivocal, regardless of the placement of *Amborella* and Nymphaeales in the angiosperm tree (e.g., Soltis et al. 2005, 2008a; Doyle 2012). The genome of *Amborella* has recently been fully sequenced (www.amborella.org; *Amborella* Genome Project, in press), and a water lily genome will also undoubtedly be sequenced in the near future. These data will lead to unprecedented insights within flowering plants (Soltis et al. 2008b; *Amborella* Genome Project, in press). The complete genome of *Amborella* will allow future researchers to use this wealth of information not only to study *Amborella*, but also hopefully to analyze further the root of extant angiosperms.

Supplementary Material

Supplementary material, including data files and/or online-only appendices, can be found in the Dryad data repository at http://datadryad.org, doi:10.5061/dryad.68n85.

S1. 78-gene, 236-taxon in-frame plastid DNA alignment.

S2. List of gene and codon partitions for plastid DNA alignment.

S3. 4-gene, 356-taxon in-frame mtDNA alignment.

S4. List of gene and codon partitions for mtDNA alignment.

S5. Trees resulting from 78-gene, 235-taxon, plastid DNA matrix.

S6. Trees resulting from 78-gene, 236-taxon plastid DNA matrix.

S7. Trees resulting from 78-gene, 233-taxon (Gnetales removed) plastid DNA matrix.

Downloaded from https://academic.oup.com/sysbio/article/63/3/368/1648791 by guest on 19 April 2024

S8. Trees resulting from 66-gene, 233-taxon (Gnetales removed) plastid DNA matrix.

S9. Trees resulting from 78-gene, 222-taxon plastid DNA matrix.

S10. Trees resulting from 177-taxon *ndh* matrix.

S11. Trees resulting from Goremykin et al. (2013) alignment S4.

S12. Trees resulting from 4-gene, 356-taxon mtDNA alignment.

S13. Trees resulting from 4-gene, 354-taxon (Gnetales removed) mtDNA alignment.

S14. Trees resulting from analysis of the 2000 excluded characters from Goremykin et al. (2013) noise-reduction protocol.

S15. 236-taxon, 58,944-character alignment sorted according to pairwise distances.

S16. ML tree resulting from 236-taxon sorted alignment with the most variable 22% of characters removed.

S17. 222-taxon, 58,860-character alignment sorted according to pairwise distances.

S18. ML tree resulting from 222-taxon sorted alignment with the most variable 35% of characters removed.

Figure S1. Per-site likelihoods calculated from ML trees based on a 17-gene region alignment from Soltis et al. (2011). Values above 0 are sites that support *Amborella* sister to all other extant angiosperms, and values below 0 support *Amborella* sister to Nymphaeales.

Figure S2. Per-site likelihoods calculated from ML trees based on the 13-gene region alignment (4 mtDNA genes excluded) of Soltis et al. (2011). Values above 0 are sites that support *Amborella* sister to angiosperms, and values below 0 support *Amborella* sister to Nymphaeales.

Funding

This study was made possible by United States National Science Foundation funding for the Open Tree of Life (DEB-12008809) and the *Amborella* Genome Project (IOS-0922742).

Acknowledgments

Thanks to Yin-Long Qiu, who graciously provided us the mtDNA alignment used in this study. J. Gordon Burleigh provided valuable guidance and commentary in the planning and writing of this article and also provided the alignment-sorting script. Vincent Savolainen, Mark Chase, Frank Anderson, and an anonymous reviewer provided constructive comments and suggestions that greatly improved this article. Author Contributions: B.T.D., D.E.S., and P.S.S designed the study. B.G.B. acquired tissue of *Trithuria* in the field, and M.J.M. sequenced and assembled the *Trithuria* plastome. M.A.G. and B.T.D annotated the *Trithuria* genome. B.R.R. assembled the plastid DNA data set. M.A.G. and B.T.D. created sorted character alignments. B.T.D. performed phylogenetic analyses; S.A.S. conducted analyses of per-site variation in likelihood. B.T.D. and D.E.S. wrote the article. B.T.D., D.E.S., P.S.S., S.A.S., and M.J.M. revised the article. The authors declare no competing financial interests.

References

Andre C., Levy A., Walbot V. 1992. Small repeated sequences and the structure of plant mitochondrial genomes. Trends Gen. 8:128–132.

Bailey I.W., Swamy B.G.L. 1948. *Amborella trichopoda* Baill., a new morphological type of vesselless dicotyledon. J. Arnold. Arbor. Harv. Univ. 29:245–254.

Barkman T.J., Chenery G., McNeal J.R., Lyons-Weiler J., Ellisens W.J., Moore G., Wolfe A.D., dePamphilis C.W. 2000. Independent and combined analyses of sequences from all three genomic compartments converge on the root of flowering plant phylogeny. Proc. Natl Acad. Sci. USA 97:13166–13171.

Bergthorsson U., Adams K.L., Thomason B., Palmer J.D. 2003. Widespread horizontal transfer of mitochondrial genes in flowering plants. Nature 424:197–201.

Bergthorsson U., Richardson A.O., Young G.J., Goertzen L.R., Palmer J.D. 2004. Massive horizontal transfer of mitochondrial genes from diverse land plant donors to the basal angiosperm *Amborella*. Proc. Natl Acad. Sci. USA 101:17747–17752.

Bessey C.E. 1915. The phylogenetic taxonomy of flowering plants. Ann. Missouri Bot. Gard. 2(1/2):109–164.

Borsch T., Hilu K.W., Quandt D., Wilde V., Neinhuis C., Barthlott W. 2003. Noncoding plastid *trnT - trnF* sequences reveal a well resolved phylogeny of basal angiosperms. J. Evol. Biol. 16:558–576.

Bousquet J., Strauss S.H., Doerksen A.H., Price R.A. 1992. Extensive variation in evolutionary rate of *rbcL* gene sequences among seed plants. Proc. Natl Acad. Sci. USA 89:7844–7848.

Braukmann T.W., Kuzmina M., Stefanovicì S. 2009. Loss of all plastid *ndh* genes in Gnetales and conifers: extent and evolutionary significance for the seed plant phylogeny. Curr. Genet. 55:323–337.

Bremer, B., Bremer, K., Chase, M.W., Fay, M.F., Reveal, J.L., Soltis, D.E., Soltis, P.S., Stevens, P. 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.

Brown J.R. 2003. Ancient horizontal gene transfer. Nat. Rev. Gen. 4: 121–132.

Burger W. 1981. Heresy revived: the monocot theory of angiosperm origin. Evol. Theory 5:189–225.

Cai Z., Penaflor C., Kuehl J.V., Leebens-Mack J., Carlson J.E., dePamphilis C.W., Boore J.L., Jansen R.K. 2006. Complete plastid genome sequences of *Drimys*, *Liriodendron*, and *Piper*: implications for the phylogenetic relationships of magnoliids. BMC Evol. Biol. 6:77.

Carlquist S. 1987. Presence of vessels in wood of *Sarcandra* (Chloranthaceae); comments on vessel origins in angiosperms. Am. J. Bot. 74:1765–1771.

Carlquist S.J., Schneider E.L. 2001. Vegetative anatomy of the New Caledonian endemic *Amborella trichopoda*: relationships with the Illiciales and implications for vessel origin. Pac. Sci. 55:305–312.

Carlquist S., Schneider E.L. 2002. The tracheid–vessel element transition in angiosperms involves multiple independent features: cladistic consequences. Am. J. Bot. 89:185–195.

Chase M.W., Albert V.A. 1998. A perspective on the contribution of plastid *rbcL* DNA sequences to angiosperm phylogenetics. In: Molecular systematics of plants II. Springer US, p. 488–507.

Chase M.W., Soltis D.E., Olmstead R.G., Morgan D., Les D.H., Mishler B.D., Duvall M.R., Price R.A., Hills H.G., Qiu Y.L., Kron K.A., Rettig J.H., Conti E., Palmer J.D., Manhart J.R., Sytsma K.J., Michael H.J., Kress W.J., Karol K.A., Clark W.D., Hedrén M., Gaut B.S., Jansen R.K., Kim K.J., Wimpee C.F., Smith J.F., Furnier G.R., Strauss S.H., Xiang Q.Y., Plunkett G.M., Soltis P.S., Swensen S.M., Williams S.E., Gadek P.A., Quinn C.J., Eguiarte L.E., Golenberg E., Learn G.H., Graham S.W., Barrett S.C.H., Dayanandan S., Albert V.A. 1993. Phylogenetics of seed plants: an analysis of nucleotide sequences

Downloaded from https://academic.oup.com/sysbio/article/63/3/368/1648791 by guest on 19 April 2024

from the plastid gene *rbcL*. Ann. Missouri Bot. Gard. 80:528–580.
Coiffard C., Gomez B., Thevenard F. 2007. Early cretaceous angiosperm invasion of western Europe and major environmental changes. Ann. Bot. 100:545–553.
Cronquist A. 1981. An integrated system of classification of flowering plants. New York, USA: Columbia University Press.
Cronquist A. 1988. The evolution and classification of flowering plants. Bronx (NY): New York Botanical Garden.
Delsuc F., Brinkmann H., Philippe H. 2005. Phylogenomics and the reconstruction of the tree of life. Nat. Rev. Gen. 6:361–375.
Doyle J.A. 2008. Integrating molecular phylogenetic and paleobotanical evidence on origin of the flower. Int. J. Plant Sci. 169:816–843.
Doyle J.A. 2012. Molecular and fossil evidence on the origin of angiosperms. Ann. Rev. Earth Plan. Sci. 4:301–326.
Doyle J.A., Endress P.K., 2000. Analysis of basal angiosperms: comparison and combination with molecular data. Int. J. Plant Sci. 161:S121–S153.
Endress P.K., Doyle J.A. 2009. Reconstructing the ancestral angiosperm flower and its initial specializations. Am. J. Bot. 96:22–66.
Evans N.M., Holder M.T., Barbeitos M.S., Okamura B., Cartwright P. 2010. The phylogenetic position of Myxozoa: exploring conflicting signals in phylogenomic and ribosomal data sets. Mol. Biol. Evol. 27:2733–2746.
Feild T.S., Arens N.C., Doyle J.A., Dawson T.E., Donoghue M.J. 2004. Dark and disturbed: a new image of early angiosperm ecology. Paleobiol. 30:82–107.
Feild T.S., Zweiniecki M.A., Brodribb T., Jaffré T., Donoghue M.J., Holbrook N.M. 2000. Structure and function of tracheary elements in *Amborella trichopoda*. Int. J. Plant Sci. 161:705–712.
Finet C., Timme R.E., Delwiche C.F., Marletaz F. 2010. Multigene phylogeny of the green lineage reveals the origin and diversification of land plants. Curr. Biol. 20:2217–2222.
Goremykin V.V., Hirsch-Ernst K.I, Woelfl S., Hellwig F.H. 2003. Analysis of the *Amborella trichopoda* chloroplast genome sequence suggests that *Amborella* is not a basal angiosperm. Mol. Biol. Evol. 20:1499–1505.
Goremykin V.V., Nikiforova S.V., Bininda-Emonds O.R. 2010. Automated removal of noisy data in phylogenomic analyses. J. Mol. Evol. 71:319–331.
Goremykin V.V., Nikiforova S.V., Biggs P.J., Zhong B., Delange P., Martin W., Woetzel S., Atherton R.A., Mclenachan T., Lockhart P.J. 2013. The evolutionary root of flowering plants. Syst. Biol. 62:50–61.
Goremykin V.V., Salamini F., Velasco R., Viola R. 2009a. Mitochondrial DNA of *Vitis vinifera* and the issue of rampant horizontal gene transfer. Mol. Biol. Evol. 26:99–110.
Goremykin V.V., Viola R., Hellwig F.H. 2009b. Removal of noisy characters from chloroplast genome-scale data suggests revision of phylogenetic placements of *Amborella* and *Ceratophyllum*. J. Mol. Evol. 68:197–204.
Graham S.W., Iles W.J.D. 2009. Different gymnosperm outgroups have (mostly) congruent signal regarding the root of flowering plant phylogeny. Am. J. Bot. 96:216–227.
Heath T.A., Hedtke S.M., Hillis D.M. 2008. Taxon sampling and the accuracy of phylogenetic analyses. J. Syst. Evol. 46:239–257.
Herendeen P.S., Miller R.B. 2000. Utility of wood anatomic characters in cladistic analysis. IAWA J. 21:247–276.
Hillis D.M. 1996. Inferring complex phylogenies.Nature 383:130.
Hilu K.W., Borsch T., Müller K., Soltis D.E., Soltis P.S., Savolainen V., Chase M.W., Powell M.P., Alice L.A., Evans R., Sauquet H., Neinhuis C., Slotta T.A.B., Rohwer J.G., Campbell C.S., Chatrou L. W. 2003. Angiosperm phylogeny based on *matK* sequence information. Am. J. Bot. 90:1758–1776.
Jansen R.K., Cai Z., Raubeson L.A., Daniell H., dePamphilis C.W., Leebens-Mack J., Mueller K.F., Guisinger-Bellian M., Haberle R.C., Hansen A.K., Chumley T.W., Lee S.-B., Peery R., McNeal J.R., Kuehl J.V., Boore J.L. 2007. Analysis of 81 genes from 64 plastid genomes resolves relationships in angiosperms and identifies genome-scale evolutionary patterns. Proc. Natl Acad. Sci. USA 104:19369–19374.
Katoh K, Misawa K., Kuma K.Ä., Miyata T. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res.30:3059–3066.
Keeling P.J., Palmer J.D. 2008. Horizontal gene transfer in eukaryotic evolution. Nat. Rev. Gen. 9:605–618.
Lee E.K., Cibrian-Jaramillo A., Kolokotronis S.O., Katari M.S., Stamatakis A., Ott, M., Chiu J.C., Little D.P., Stevenson D.W., McCombie R.W., Martiensen R.A., Coruzzi G. DeSalle R. 2011. A functional phylogenomic view of the seed plants. PLoS Gen. 7(12):e1002411.
Leebens-Mack J., Raubeson L.A., Cui L., Kuehl J.V., Fourcade M.H., Chumley T.W., Boore J.L., Jansen R.K., dePamphilis C.W. 2005. Identifying the basal angiosperm node in chloroplast genome phylogenies: sampling one's way out of the Felsenstein zone. Mol. Biol. Evol. 22:1948–1963.
Lockhart P.J., Larkum A.W.D., Steel M.A., Waddell P.J., Penny D. 1996. Evolution of chlorophyll and bacteriochlorophyll: the problem of invariant sites in sequence analysis. Proc. Natl Acad. Sci. USA 93:1930–1934.
Lyons-Weiler J., Hoelzer G.A., Tausch R.J. 1996. Relative apparent synapomorphy analysis (RASA). I: the statistical measurement of phylogenetic signal. Mol. Biol. Evol. 13:749–757.
Maddison W.P., Maddison D.R. 2011. Mesquite: a modular system for evolutionary analysis, v. 2.75. Available from: http://mesquiteproject.org.
Mathews S., Donoghue M.J. 1999. The root of angiosperm phylogeny inferred from duplicate phytochrome genes. Science 286:947–950.
Moore M.J., Bell C.D., Soltis P.S., Soltis D.E. 2007. Using plastid genomic-scale data to resolve enigmatic relationships among basal angiosperms. Proc. Natl Acad. Sci. USA 104:19363–19368.
Moore M.J., Dhingra A., Soltis P.S., Shaw R., Farmerie W.G., Folta K.M., Soltis D.E. 2006. Rapid and accurate pyrosequencing of angiosperm plastid genomes. BMC Plant Biol. 6:17–30.
Moore M.J., Hassan N., Gitzendanner M.A., Bruenn R.A., Croley M., Vandeventer A., Horn J.W., Dhingra A., Brockington S.F., Latvis M., Ramdial J., Alexandre R., Piedrahita A., Xi Z., Davis C.C., Soltis P.S., Soltis D.E. 2011. Phylogenetic analysis of the plastid inverted repeat for 244 species: Insights into deeper-level angiosperm relationships from a long, slowly evolving sequence region. Int. J. Plant Sci. 172:541–558.
Moore M.J., Soltis P.S., Bell C.D., Burleigh J.G., Soltis D.E. 2010. Phylogenetic analysis of 83 plastid genes further resolves the early diversification of eudicots. Proc. Natl Acad. Sci. USA 107: 4623–4628.
Morton C.M. 2011. Newly sequenced nuclear gene (*Xdh*) for inferring angiosperm phylogeny. Ann. Missouri Bot. Gard. 98:63–89.
Olmstead R.G., Reeves P.A., Yen A.C. 1998. Patterns of sequence evolution and implications for parsimony analysis of chloroplast DNA. In: Soltis D.E., Soltis P.S., Doyle, J.J. editors. Molecular systematics of plants II. Boston: Springer and Kluwer. p. 164–187.
Palmer J.D. 1992. Mitochondrial DNA in plant systematics: applications and limitations. In: Soltis D.E., Soltis P.S., Doyle, J.J. editors. Molecular systematics of plants, 1st edition. New York: Chapman and Hall. p. 36–49.
Palmer J.D., Herbon L.A. 1988. Plant mitochondrial DNA evolved rapidly in structure, but slowly in sequence. J. Mol. Evol. 28:87–97.
Parkinson C.L., Adams K.L., Palmer J.D. 1999. Multigene analyses identify the three earliest lineages of extant flowering plants. Curr. Biol. 9:1485–1488.
Parks M., Cronn R., Liston, A. 2012. Separating the wheat from the chaff: mitigating the effects of noise in a plastome phylogenomic data set from *Pinus* L. (Pinaceae). BMC Evol. Biol. 12(1):100.
Philippe H., Delsuc F., Brinkmann H., Lartillot N. 2005. Phylogenomics. Ann. Rev. Ecol. Evol. Systematics 36:541–562.
Qiu Y.-L., Chase M.W., Les D.H., Parks C.R. 1993. Molecular phylogenetics of the Magnoliidae: cladistic analyses of nucleotide sequences of the plastid gene *rbcL*. Ann. Missouri Bot. Gard. 80:587–606.
Qiu Y.-L., Dombrovska O., Lee J., Li L.B., Whitlock B.A., Bernasconi-Quadroni F., Rest J.S., Davis C.C., Borsch T., Hilu K.W., Renner S.S., Soltis D.E., Soltis P.S., Zanis M.J., Cannone J.J., Gutell R.R., Powell M., Savolainen V., Chatrou L.W., Chase M.W. 2005. Phylogenetic analyses of basal angiosperms based on nine plastid, mitochondrial, and nuclear genes. Int. J. Plant Sci. 166:815–842.
Qiu Y.-L., Lee J., Bernasconi-Quadroni F., Soltis D.E., Soltis P.S., Zanis M., Zimmer E.A., Chen Z., Savolainen V., Chase M.W. 1999. The earliest angiosperms: evidence from mitochondrial, plastid and nuclear genomes. Nature 402:404–407.

Downloaded from https://academic.oup.com/sysbio/article/63/3/368/1648791 by guest on 19 April 2024

Qiu Y.-L., Lee J., Bernasconi-Quadroni F., Soltis D.E., Soltis P.S., Zanis M., Zimmer E.A., Chen Z., Savolainen V., Chase M.W. 2000. Phylogeny of basal angiosperms: analyses of five genes from three genomes. Int. J. Plant Sci. 161:S3–S27.

Qiu Y.-L., Wang B., Xue J.-Y., Hendry T.A., Li R.-Q., Brown J. W., Liu Y., Hudson G.T., Chen Z.-D. 2010. Angiosperm phylogeny inferred from sequences of four mitochondrial genes. J. Syst. Evol. 48: 391–425.

Rajan V. 2013. A method of alignment masking for refining the phylogenetic signal of multiple sequence alignments. Mol. Biol. Evol. 30:689–712.

Rambaut A., Drummond A.J. 2009. Tracer, version 1.5. Available from: http://tree .bio.ed.ac.uk/software/tracer/.

Regier J.C., Zwick A. 2011. Sources of signal in 62 protein-coding nuclear genes for higher-level phylogenetics of arthropods. PloS One 6(8):e23408.

Renner S.S., Bellot S. 2012. Horizontal gene transfer in eukaryotes: fungi-to-plant and plant-to-plant transfers of organellar DNA. In: Bock R, Koop B, editors. Genomics of chloroplasts and mitochondria. Dordrecht (The Netherlands): Springer, p. 223–235.

Richardson A.O., Palmer J.D. 2007. Horizontal gene transfer in plants. J. Exp. Bot. 58:1–9.

Rokas A., Williams B.L, King N., Carroll S.B. 2003. Genome-scale approaches to resolving incongruence in molecular phylogenies. Nature 425:798–804.

Saarela J.M., Rai H.S., Doyle J.A., Endress P.K., Mathews S., Marchant A.D., Briggs B.G., Graham S.W. 2007. Hydatellaceae identified as a new branch near the base of the angiosperm phylogenetic tree. Nature 446:5–8.

Savolainen V., Chase M.W., Hoot S.B., Morton C.M., Soltis D.E., Bayer C., Fay M.F., De Bruijn A.Y., Sullivan S., Qiu Y.L. 2000. Phylogenetics of flowering plants based on combined analysis of plastid atpB and *rbcL* gene sequences. Sys. Biol. 49:306–362.

Smith S.A., Wilson N.G., Goetz F.E., Feehery C., Andrade S.C., Rouse G.W., Giribet G., Dunn C.W. 2011. Resolving the evolutionary relationships of molluscs with phylogenomic tools. Nature 480:364–367.

Soltis D.E., Albert V.A., Leebens-Mack J., Palmer J.D., Wing R.A., dePamphilis C.W., Ma H., Carlson J.E., Altman N., Kim S., Wall P.K., Zuccolo A., Soltis P.S. 2008b. The *Amborella* genome: an evolutionary reference for plant biology. Genome Biol. 9:402.

Soltis D.E., Albert V.A., Savolainen V., Hilu K., Qiu Y.L., Chase M.W., Farris J.S., Stefanović S., Rice D.W., Palmer J.D., Soltis P.S. 2004. Genome-scale data, angiosperm relationships, and "ending incongruence": a cautionary tale in phylogenetics. Trends Plant Sci. 9:477–483.

Soltis D.E., Bell C.D., Kim S., Soltis P.S. 2008a. Origin and early evolution of angiosperms. Ann. New York Acad. Sci. 1133:3–25.

Soltis D.E., Soltis P.S. 2004. *Amborella* not a "basal angiosperm"? Not so fast. Am. J. Bot. 91:997–1001.

Soltis P.S., Soltis D.E. 1998. Molecular evolution of 18S rDNA in angiosperms: implications for character weighting in phylogenetic analysis. In: Soltis D.E., Soltis P.S., Doyle, J.J. editors. Molecular systematics of plants II. Boston: Springer and Kluwer. p. 188–210.

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

Soltis D.E., Soltis P.S., Chase M.W., Mort M.E., Albach D.C., Zanis M., Savolainen V., Hahn W.H., Hoot S.B., Fay M.F., Axtell M., Swensen S.M., Prince L.M., Kress W.J., Nixon K.C., Farris J.S. 2000. Angiosperm phylogeny inferred from 18S rDNA, *rbcL*, and *atpB* sequences. Bot. J. Linn. Soc. 133:381–461.

Soltis D.E., Soltis P.S., Endress P.K., Chase M.W. 2005. Phylogeny and evolution of angiosperms. Sunderland (MA): Sinauer.

Soltis D.E., Soltis P.S., Nickrent D.L., Johnson L.A., Hahn W.J., Hoot S.B., Sweere J.A., Kuzoff R.K., Kron K.A., Chase M.W., Swensen S.M., Zimmer E.A., Chaw S.M., Gillespie L.J., Kress W.J., Sytsma K.J. 1997. Angiosperm phylogeny inferred from 18S ribosomal DNA sequences. Ann. Missouri Bot. Gard. 84:1–49.

Soltis P.S., Soltis D.E., Savolainen V., Crane P.R., Barraclough T.G. 2002. Rate heterogeneity among lineages of tracheophytes: integration of molecular and fossil data and evidence for molecular living fossils. Proc. Natl Acad. Sci. USA 99:4430–4435.

Soltis P.S., Soltis D.E., Zanis M.J., Kim S. 2000. Basal lineages of angiosperms: relationships and implications for floral evolution. Int. J. Plant Sci. 161:S97–S107.

Soltis D.E., Smith S.A., Cellinese N., Wurdack K.J., Tank D.C., Brockington S.F., Refulio- Rodriguez N.F., Walker J.B., Moore M.J., Carlsward B.S., Bell C.D., Latvis M., Crawley S., Black C., Diouf D., Xi Z., Rushworth C.A., Gitzendanner M.A., Sytsma K.J., Qiu Y.L., Hilu K.W., Davis C.C., Sanderson M.J., Beaman R.S., Olmstead R.G., Judd W.S., Donoghue M.J., Soltis P.S. 2011. Angiosperm phylogeny: 17 genes, 640 taxa. Am. J. Bot. 98:704–730.

Stamatakis A., Hoover P., Rougemont J. 2008. A rapid bootstrap algorithm for the RAxML web servers. Syst. Biol. 57: 758–771.

Stamatakis A., Ludwig T., Meier H. 2005. RAxML-III: a fast program for maximum likelihood-based inference of large phylogenetic trees. Bioinformatics 21:456–463.

Stefanović S., Rice D.W., Palmer J.D. 2004. Long branch attraction, taxon sampling, and the earliest angiosperms: *Amborella* or monocots? BMC Evol. Biol. 4:35.

Suchard M.A., Weiss R.E., Sinsheimer J.S. 2001. Bayesian selection of continuous-time Markov chain evolutionary models. Molec. Biol. Evol.18:1001–1013.

Takhtajan A.L. 1997. Diversity and classification of flowering plants. New York, USA: Columbia University Press.

Thorne R.F. 1992. Classification and geography of the flowering plants. Bot. Rev. 58:225–327.

Wodniok S., Brinkmann H., Glöckner G., Heidel A.J., Philippe H., Melkonian M., Becker B. 2011. Origin of land plants: do conjugating green algae hold the key? BMC Evol. Biol. 11:104.

Wyman S.K., Jansen R.K., Boore J.L. 2004. Automatic annotation of organellar genomes with DOGMA. Bioinformatics 20: 3252–3255.

Xi Z., Wang Y., Bradley R.K., Sugumaran M., Marx C.J., Rest J.S., Davis C.C. 2013. Massive mitochondrial gene transfer in a parasitic flowering plant clade. PLoS Gen. 9(2):e1003265.

Zanis M.J., Soltis D.E., Soltis P.S., Mathews S., Donoghue M.J. 2002. The root of the angiosperms revisited. Proc. Natl Acad. Sci. USA 99:6848–6853.

Zwickl D.J., Hillis D.M. 2002. Increased taxon sampling greatly reduces phylogenetic error. Syst. Biol. 51:588–598.

Zhang N., Zeng L., Shan H., Hong M. 2012. Highly conserved low-copy nuclear genes as effective markers for phylogenetic analyses in angiosperms. New Phyt. 195:123–137.

Downloaded from https://academic.oup.com/sysbio/article/63/3/368/1648791 by guest on 19 April 2024