Nested phylogenetic conflicts, combinability, and deep phylogenomics in plants

Stephen A. Smith[1], Nathanael Walker-Hale[1,2], Joseph F. Walker[1], Joseph W. Brown[3]

Author Affiliation: [1]Department of Ecology and Evolution Biology, University of Michigan, [2]Department of Plant Sciences, University of Cambridge, [3]Department of Animal and Plant Sciences, University of Sheffield

Keywords: phylogenomics, phylogenetics, angiosperms, gene-tree conflict, genomics, coalescent

Abstract

Studies have demonstrated that pervasive gene tree conflict underlies several important phylogenetic relationships where different species tree methods produce conflicting results. Here, we present a means of dissecting the phylogenetic signal for alternative resolutions within a dataset in order to resolve recalcitrant relationships and, importantly, identify relationships the dataset is unable to resolve. These procedures extend upon methods for isolating conflict and concordance involving specific candidate relationships, and can be used to identify systematic error and disambiguate sources of conflict among species tree inference methods. We demonstrate these procedures on a large phylogenomic plant dataset. Our results support the placement of *Amborella* as sister to the remaining extant angiosperms, the monophyly of extant gymnosperms, and that Gnetales are sister to pines. Several other contentious relationships, including the resolution of relationships within both the bryophytes and the eudicots, remain uncertain given the low number of supporting gene trees. To address whether concatenation of filtered genes amplified phylogenetic signal for particular relationships, we implemented a combinatorial heuristic to test combinability of genes. We found that nested conflicts limited the ability of data filtering methods to fully ameliorate conflicting signal amongst gene trees. These analyses confirmed that the underlying conflicting signal does not support broad concatenation of genes. Our approach provides a means of dissecting a specific dataset to address deep phylogenetic relationships while highlighting the limitations of the dataset.

Introduction

Over the last few years, we have come to understand that phylogenetic conflict is common and presents several analytical challenges. Researchers have amassed large genomic and transcriptomic datasets meant to resolve fundamental phylogenetic relationships in plants (Wickett et al. 2014), animals (Jarvis et al. 2014; Dunn et al. 2008; Simion et al. 2017; Whelan et al. 2017), fungi (Shen et al. 2016), and bacteria (Ahrenfeldt et al. 2017). While the goals of these data collection efforts have been to increase the overall phylogenetic support, analyses have demonstrated that different datasets and analytical approaches often reconstruct strongly-supported but conflicting relationships (Feuda et al. 2017; Walker et al. 2018; Shen, Hittinger, and Rokas 2017). Underlying these discordant results are strongly conflicting gene trees (Smith et al. 2015). In some cases, one or two "outlier" genes with large likelihood differences between alternative relationships can drive results (Shen, Hittinger, and Rokas 2017; Brown and Thomson 2016; Walker, Brown, and Smith 2018). Detailed gene tree analysis

of phylogenomic datasets is essential to identifying and analyzing overall gene tree conflict and outlier genes.

Phylogenomic datasets are often analyzed as concatenated supermatrices or with coalescent gene-tree / species tree methods. Supermatrix methods were, in part, developed to amplify the strongest phylogenetic signal. However, it has long been understood that the "total evidence" paradigm (Kluge 1989), where the true history will 'win out' if enough data are collected, is untenable. Genes with real and conflicting histories are expected within datasets due to biological processes like hybridization and incomplete lineage sorting (ILS) (Maddison 1997) in addition to outlying genes and sites as mentioned above (Shen, Hittinger, and Rokas 2017; Brown and Thomson 2016; Walker, Brown, and Smith 2018). "Species tree" inference accommodates for gene tree conflict due to ILS (Edwards, Liu, and Pearl 2007; Liu et al. 2009; Edwards 2009; Edwards et al. 2016) and is often conducted alongside concatenated supermatrix analyses. Differences in the results from these two approaches are often explained by the differences in assumptions each makes. The concatenated supermatrix allows for mixed molecular models and gene-specific branch lengths but assumes a single underlying tree topology common to all genes. This procedure is known to perform poorly in the presence of extensive ILS. Coalescent approaches, depending on the implementation, may assume that all conflict is the result of ILS (but see Boussau et al. (2013) and Ané et al. (2006)), that all genes evolved under selective neutrality and constant effective population size, that all genes contain enough information to properly resolve nodes, and that gene trees are estimated accurately (Springer and Gatesy 2016).

While supermatrix and coalescent methods perform well in many scenarios, when unresolved nodes or discordance between species trees remain after large data collection efforts, researchers can further examine the processes leading to conflict or further dissect the phylogenetic signal within datasets. For example, Bayesian methods have been developed that incorporate processes in addition to ILS that lead to gene tree discordance (Ané et al. 2006; Boussau et al. 2013). However, these methods are often computationally intractable for current genomic datasets and may not handle systematic error well. Recently, network methods that scale to large datasets have been developed (Wen et al. 2018,@snaq), but these do not allow for dissecting signal within datasets. Filtering approaches where subsets of genes are analyzed based on model similarity or the relationships displayed by the genes (Chen, Liang, and Zhang 2015; Shen et al. 2016; Smith, Brown, and Walker 2018), help to enable computational tractability and distill signal. For example, Chen, Liang, and Zhang (2015) filtered for question-specific genes in the phylogeny of jawed vertebrates using two methods: one where only gene trees capable of supporting one of three resolutions for a given relationship were included in the analysis, and another where only gene trees which agreed with a widely-accepted control locus were retained for the analysis. Researchers have also examined alternative phylogenetic hypotheses in order to isolate the supporting signal (Shen, Hittinger, and Rokas 2017; Brown and Thomson 2016; Walker, Brown, and Smith 2018).

In plants, several large data collection efforts aimed at resolving difficult nodes have found extensive conflicts (Smith et al. 2015; Walker et al. 2018, 2017; Wickett et al. 2014). Resolution of these clades is not only important for systematics, but crucial to an evolutionary understanding of key biological questions. For example, the relationships among the lineages of bryophytes (i.e., hornworts, liverworts, and mosses) remain unclear despite extensive data collection efforts (Wickett et al. 2014; Puttick et al. 2018). One of the most heavily debated lineages in plant phylogenetics is the monotypic *Amborella*, the conflicting placement of which alters our understanding of early flowering plant evolution. *Amborella* has been inferred as sister to Nymphaeales, as sister to all angiosperms, or as sister to the remaining Angiosperms excluding Nymphaeales (Xi et al. 2014). The resolution of *Amborella*, along with other

contentious relationships across land plants, would provide greater confidence in our understanding of the evolution of early reproductive ecology, the evolution of floral development, and the life history of early land plants (Feild et al. 2004; Sauquet et al. 2017).

We conducted a detailed analysis of nested phylogenomic conflict and signal across a phylogenomic dataset in hopes of presenting a computationally tractable and practical way to examine contentious relationships. We extended methods for examining phylogenetic alternatives and present an approach that can be widely applied to empirical datasets to determine the support, or lack thereof, for phylogenetic hypotheses. We applied these methods to a large plant genomic dataset (Wickett et al. 2014). We identified systematic error, nested conflicting relationships, support for alternative resolutions, and we present a practical means to test the topological combinability of subsets of genes based on a combinatorial heuristic and information criteria statistics. By taking this broad information-centric approach, we hope to shed more light on the evolution of plants and present a tractable approach for dissecting signal with broad applicability for phylogenomic datasets across the Tree of Life.

Materials and Methods

Datasets

We analyzed the Wickett et al. (2014) dataset of transcriptomes and genomes covering plants available from http://mirrors.iplantcollaborative.org/onekp_pilot. There were several different filtering methods and approaches used in the original manuscript and, based on conversations with the corresponding author, we analyzed the filtered nucleotide dataset with third codon positions removed. These sites were removed because of problems with excessive variation and GC content that caused problems with the placement of the lycophytes (Wickett et al. 2014). This dataset consisted of 852 aligned genes. We did not conduct any other filtering or alteration of these data before conducting the analyses performed as part of this study.

Phylogenetic analyses

We inferred gene trees for each of the 852 genes using IQ-TREE (v. 1.6.3; Nguyen et al. 2014). We used the GTR+G model of evolution and calculated maximum likelihood trees along with SH-aLRT values (Guindon et al. 2010). For all constrained analyses, we conducted additional maximum likelihood analyses with the same model of evolution but constrained on the relationship of interest, although the rest of the tree topology was free to vary.

Conflict analyses

We conducted several different conflict analyses. First, we identified the congruent and conflicting branches between the maximum likelihood gene trees (ignoring branches that had less than 80% SH-aLRT (Guindon et al. 2010)0, and the maximum likelihood species tree from the original publication (Fig. 2; Wickett et al. 2014). These analyses were conducted using the program `bp` available from `https://github.com/FePhyFoFum/gophy`. We placed these conflicting and supporting statistics in a temporal context by calculating the divergence times of each split based on the TimeTree of Life (Hedges, Dudley, and Kumar 2006; Hedges et al. 2015). By examining the dominant conflicting

alternatives, we established which constraints to implement and compare for further analyses. Because the gene regions contain partially overlapping taxa, automated discovery of all conflicting relationships concurrently can be challenging. To overcome these challenges, we examine each constraint individually.

To determine the difference in the log-likelihood (lnL) values among conflicting resolutions, we conducted the constrained phylogenetic analyses (with parameters described in the *Phylogenetic analyses* section above) and compared the lnL values of the alternative resolutions. We then examined those results that had a difference in the lnL of greater than 2, considering this difference as statistically significant (Edwards 1984). For each gene, we noted the relationship with the highest log-likelihood and summed the difference of that and the second best relationship (DlnL) across all genes.

We also examined nested conflicts. In particular, for the genes identified as supporting the dominant relationship of the eudicot lineages, we examined the distribution of conflict. We then examined those genes that supported both the eudicot lineages and the relationship of *Amborella* as sister to the rest of angiosperms. Finally, of those genes, we determined which supported the alternative gymnosperm relationships. We conducted each of these nested analyses using the same methods as described above.

Combinability test

We describe a simple but fast procedure for testing the combinability within a dataset based on gene tree similarity and information criteria (Fig. 1). A typical concatenated phylogenetic analysis assumes that the entire alignment used to calculate the tree was generated with the same underlying topology. When that is not the case, the likelihood of the tree using the entire alignment will be lower than the when considering the gene regions separately. It follows that those genes that should be combined (i.e., concordant histories) will have more similar gene trees than those that should be considered separately (i.e., conflicting histories). To determine similarity between gene trees, we calculated the pairwise weighted Robinson-Foulds (RFW) distance (Robinson and Foulds 1981). We then constructed a graph where genes are nodes and edges are the weights between gene trees based on RFW. Then, beginning with the strongest edge, we tested for the combinability between the two connecting nodes. If they were combinable, based on the information criteria discussed below, we merged the nodes, along with the connecting edges for each.

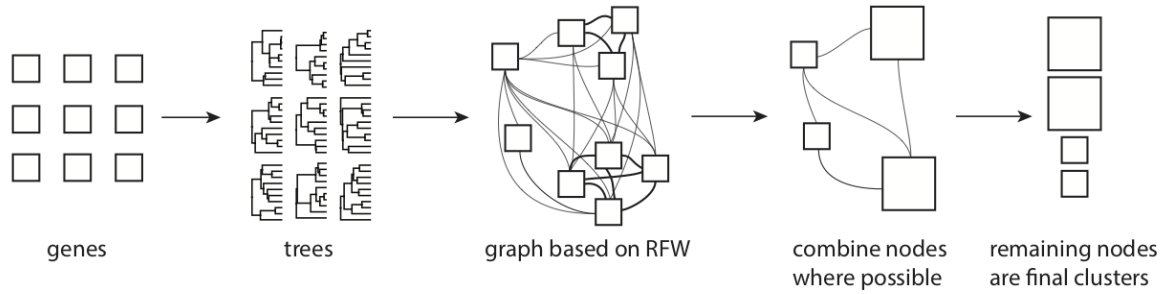

Figure 1: Procedure described in the methods section. Gene trees are constructed for genes and weighted Robinson-Foulds distances are calculated between gene trees. A graph is constructed with genes as nodes and edge weights from the weighted distances. The strongest edges are then tested for combinability and combined if possible. The final nodes in the graph are the final clusters (i.e., clusters that cannot be justifiably combined).

Non-nested likelihood-based analyses that have different numbers of parameters cannot be compared

directly. Instead, in the likelihood framework, information criteria are commonly used to accommodate and penalize for the increase in the number of parameters to prevent overfitting. The Akaike Information Criterion (AIC; Akaike 1973), the AIC with the correction for dataset size (AICc; see Burnham and Anderson 2003), and the Bayesian Information Criterion (BIC; Schwarz 1978) may all be used to compare likelihood scores that are produced from different numbers of parameters. Each of these criteria has different assumptions and different potential utility. Here, we examine the differences in considering AICc and BIC.

The number of parameters for a single gene in a phylogenetic analysis include those for the molecular model (e.g., $GTR = 8$, 5 for substitution rates (the 6 rates are expressed relative to one arbitrary rate that is fixed as 1.0) and 3 for stationary nucleotide frequencies, with an additional 1 when including gamma-distributed rate variation) and the branch lengths of the unrooted phylogenetic tree ($2n - 3$). There are several ways by which multiple genes may be combined. For example, often molecular models are allowed to vary between these genes, or partitions. It is possible to test whether the genes should share models and programs exist to conduct such tests (e.g. PartitionFinder: Lanfear et al. (2016)). If models vary between gene regions, then for a x gene dataset, the number of molecular model parameters y would be $x \times y$. The parameterization of branch lengths has several options: shared ($2n - 3$), exactly proportional ('scaled'; $(2n - 3) + (x - 1)$), and independent ($(2n - 3) \times x$). Here, we considered the molecular models to be independent between gene regions and tested both scaled and independent branch lengths.

With these considerations, the tree comparison calculation proceeded as follows: for each gene, calculate the information criterion of the ML gene tree. Next, sum the information criterion statistic for the set of genes being tested. Further, concatenate the genes and calculate the information criterion for the ML tree. The genes may have different model parameters or branch lengths (shared, scaled, or independent), but they share the same topology. Lastly, compare the values of the information criterion for the summed gene trees and the concatenated genes. If the concatenated genes have a lower value of the information criterion than the summed gene trees, accept the combined genes and continue to the next comparison. If genes are already a member of a merged set, then compare the new gene to the merged set. Given this procedure, our algorithm is a greedy clustering method. Our approach is somewhat similar to the GARD method for detection of recombination breakpoints (Kosakovsky Pond et al. 2006a, 2006b). Here, the 'breakpoints' are the ends of the gene partitions, and we allow full maximum likelihood inference of the topologies of each partition, as well as selection of different branch length models and information criteria. Furthermore, instead of a genetic algorithm, we use tree distances to select which pairs to test. These methods are implemented in an open source python package, phyckle, available at https://github.com/FePhyFoFum/phyckle.

Simulations

We verified the performance of our combinatorial method using a variety of simulations across tree depths, branch length heterogeneity, topological variation, and model variation. Each simulation is described below. In general, we attempted to simplify the simulations in order to isolate the specific element being tested in order to better describe the expected behavior. While alignments were simulated under differing models, all clustering tests were conducted using GTR+G as this is typical of empirical analyses. For all simulations below, trees were simulated using `pxbdsim` from the phyx package (Brown, Walker, and Smith 2017) and alignments were generated using INDELible (Fletcher and Yang 2009).

Comparing information criteria and branch length models– In order to determine the efficacy of different information criteria as well as different branch length models, we conducted several simulation analyses. For each simulation, we generated a tree from a pure birth model with 25 tips and then three gene regions under JC model of evolution with 1000 sites each. This analysis was conducted with 100 replicates. While the JC model of evolution is, perhaps, overly simplistic, we aimed to isolate the factors that caused genes to be considered separate or combined. We test more complex models below. Tree heights were tested for 0.05, 0.25, 0.75, and 1.25. We also conducted tests where branches could vary between gene regions. For each gene region, the species tree branch lengths were perturbed randomly with a sliding window of 0.01, 0.05, and 0.1, so $U(x-w, x+w)$. We examined scaled and independent branch length models with both BIC and AICc.

Examining the impact of branch differences–The above tests examined variation between simulated genes involving branch length heterogeneity and model complexity, but all had the same underlying topology. We also examined the impact of having different underlying topologies between gene regions. To do this, we simulated a pure birth tree of 25 tips and a tree depth of 0.5 and simulated two gene regions under this model. Then for one additional gene region, we chose one node randomly and swapped nearest neighbors and then simulated gene regions. This resulted in three gene regions with two different underlying topologies. The difference in the underlying topologies varied from one swapping move to five swapping moves. All genes trees also had branch lengths perturbed with branch length heterogeneity of 0.01 as described above.

Examining the impact of different models on different genes– In order to examine whether different models may cause the gene regions to be considered separate we conducted similar simulations to those described above but with distinct substitution models applied to individual gene regions. Two gene regions were simulated for each of three substitution models (i.e., six gene regions total), each with 1000 bases and the same underlying pure birth topology of 25 taxa and tree depth of 0.5. Branch length heterogeneity varied from 0.01, 0.05, and 0.1. The first two gene regions were evolved under JC, the second set of two gene regions under HKY with $\kappa = 2.5$, proportion of invariable sites $= 0.25$, $\Gamma = 0.5$, number of Γ categories$= 10$, and state frequencies of $0.2, 0.3, 0.1, 0.4$ for A, C, G, and T, respectively, and the third set of two gene region under HKY with $\kappa = 1.5$, proportion of invariable sites $= 0.25$, $\Gamma = 0.5$, number of Γ categories$= 10$, and state frequencies of $0.1, 0.4, 0.3, 0.2$. Two gene regions were simulated for each model in order to verify that those two continued to be clustered together regardless of how the separate models clustered. This test was not intended to be comprehensive as variation in molecular models in relation to information criteria has already been thoroughly explored (e.g., Lanfear et al. 2016; Seo and Thorne 2018). Instead, we aimed to better understand the conditions under which variation in molecular model would result in consideration as completely separate analyses.

Examining the impact of missing taxa–Because genes often do not have completely overlapping taxa, we conducted simulations where some taxa may be missing from each gene region. For these simulations, 25 taxon pure birth trees were generated and three gene regions of 1000 bases each were simulated. Then from one to three tips were randomly removed from one gene. We also conducted simulations where from one to three tips were randomly removed from each of the three genes. Random taxa were removed from each gene and so some genes would have the same taxa removed and others would not. All genes trees also had branch lengths perturbed with branch length heterogeneity of 0.01 as described above.

Examining the potential for snowballing–Based on initial observations, we hypothesized that the use of particular combinations of model and information criteria may lead to genes being erroneously

combined because of the size of the cluster they were compared to, i.e. that clusters would snowball in size. We assessed this possibility by simulating 1000 base-pair alignments under a JC model of evolution on a 25-taxon pure birth tree with a tree depth of 0.5, and comparing these alignments to another alignment simulated on a tree three NNI-moves away. In each iteration, we increased the number of alignments simulated on the same tree. Thus iteration one compared one gene on one tree and another on a tree three NNI moves away, while iteration two compared two genes simulated on one tree with another on a tree three NNI moves away, and so on. Each comparison was repeated 100 times for linked (proportionally scaled) and unlinked (independent) branch lengths and analyzed with both AICc and BIC.

Empirical Demonstration

For demonstration purposes, we did not conduct exhaustive testing of combinability of the entire Wickett et al. (2014) dataset. Instead, we conducted these tests on two gene sets that supported the eudicot relationship. First, we tested the set of genes that supported the eudicot relationship in the ML tree that did not have a branch length longer than 2.5 and did not have outgroup taxa falling in the ingroup. Long branch lengths (e.g., >2.5 substitutions per site) suggest multiple substitutions at *each* site and therefore little to no remaining phylogenetic information (e.g., systematic error or extremely rapid rates of evolution). Second, we tested the set of genes that did not only support the relationship in the ML tree but also displayed the relationship in the ML gene tree with SH-aLRT support higher than 80 and with no outlying branch lengths or outgroup taxa falling in the ingroup. These control methods echo the classes of filtering evoked in Chen, Liang, and Zhang (2015), that of non-specific data filtering (branch length, support values) and 'node-control' (outgroup relationships, eudicot relationships).

Clustering analyses were conducted using IQ-TREE with AICc and the `-spp` option for scaled branch lengths partitions, as simulations demonstrated that it split the most accurately based on conflicting topologies (see *Results*).

We compared the results of our analyses to the PartitionFinder 'greedy' algorithm implemented in IQ-TREE using the option `-m MERGE`, specifying the GTR+G model and assessing partitions with the edge-linked proportional model with `-spp`. We compared the individual gene trees of each merged partition in IQ-TREE with `-spp` and `-m GTR+G` and for comparison assessed the optimal partitioning scheme on the full data similarly with `-spp` and `-m GTR+G`. In addition we compared the results of treating the clusters from the combination procedure as an optimal partitioning scheme, using `-spp` and `-m GTR+G`. In each case AICc was used for a direct comparison to the results of our method.

Results

Conflict analyses

We compared gene trees (Fig. 1) based on the concatenated maximum likelihood (ML) analysis from Wickett et al. (2014) and found that both gene tree conflict and support varied through time with support increasing toward the present (Fig. 2). We aimed to resolve contentious relationships, with a focus on those that have either been debated in the literature or been considered important in resolving key evolutionary questions, to the best of the ability of the underlying data (Table 1).

The massive scale of genomic datasets can cause substantial noise that is often difficult to identify when taking the dataset as a whole. When analyzing specific genes, we found that several conflicting relationships were the result of systematic error in the underlying data. In order to minimize the impact of systematic error on the estimation of relationships, we excluded obvious errors where possible. For example, we found 258 of 852 gene trees contained non-land plant taxa that fell within the land plants. While these errors may not impact the estimation of relationships within eudicots, they will impact the estimation of relationships at the origin of land plants. Therefore, we excluded gene trees for which there was not previously well established monophyly of the focal taxa (i.e., involving the relationship of interest). We also identified 68 gene trees that possessed very long estimated branch lengths ($>$ 2.5 expected substitutions per site). We conservatively considered these to contain potential errors in homology (Yang and Smith 2014). While these genes demonstrate patterns associated with systematic error, they also likely contain information for several relationships. However, some error may be the result of misidentified orthology that will mislead estimation of phylogenetic relationships, even if this error may not impact all relationships inferred by the gene. Therefore, to minimize sources of systematic error, we took a conservative approach and excluded these genes from additional analyses.

We explored both numbers of gene trees and differences in log-likelihoods for several key relationships. In some cases both number of gene trees and differences in log-likelihood support the same resolution, as was the case for the monophyly of Gymnosperms. However, other relationships are more equivocal or contradictory. For example, Gnetales and conifers as sisters ("Gnetifers") is supported by more genes, but Gnetales and Pines as sisters ("Gnepine") is supported by differences in log-likelihood (Table 1).

*Table 1. Comparison of the number of genes and the difference in the likelihood (DlnL) with relationships ordered based on support. * indicates relationships present in the ML tree.*

Major clade	Resolutions	Genes	Genes ($>$ 2lnL)	DlnL	DlnL $>$ 2
Bryophytes	Hornworts sister*	110	83	677.6	654.1
	Liverworts sister	56	41	294.1	280.8
	Mosses+liverworts	81	40	228.9	190.2
	All monophyly	81	37	185.3	148.5
Gymnosperms	monophyly*	288	264	7259.0	7233.8
	Gnetum sister	45	31	229.8	216.0
	Cycas sister	39	18	120.3	105.2
Gymno relat.	Gnepine*	107	85	1017.2	994.4
	conifers	93	79	800.0	787.2
	Gnetifers	134	55	288.1	217.8
	Gnetales sister	76	40	211.2	176.3
Amborella	*Amborella* sister*	184	152	1501.1	1470.0
	Amborella+*Nuphar*	118	75	564.2	526.3
	Nuphar sister	111	62	392.2	345.2
Eudicots	Magnoliids+eudicots*	114	98	1223.4	1204.3
	Monocots+eudicots	66	49	541.5	526.5
	Monocots+magnoliids	90	58	453.3	425.5

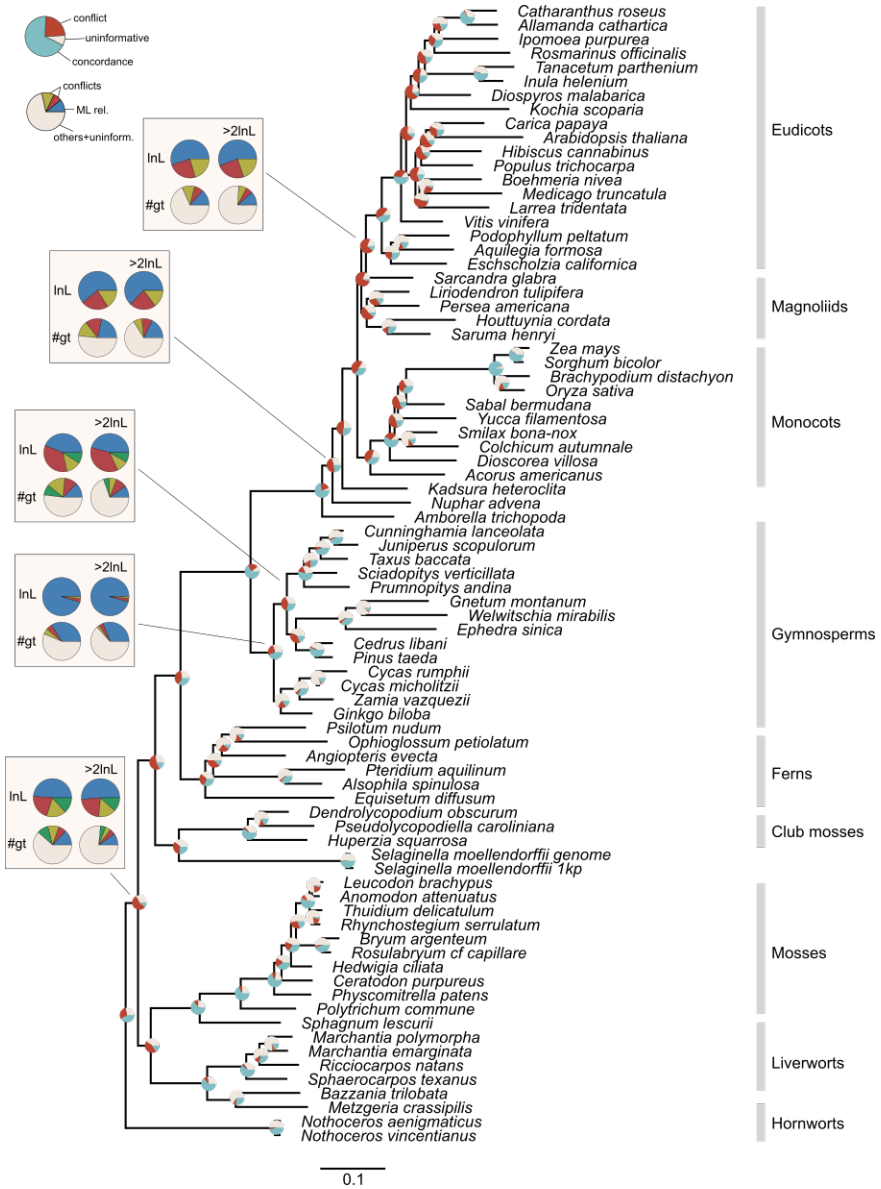

Figure 2: Phylogeny of land plants with pie charts at nodes illustrating conflict, concordance, and informativeness of the gene tree set without any filtering. Inset boxes show summed differences in log likelihoods (top row) and the number of gene trees (bottom row) that support the relationship shown in the tree and the dominant conflicting relationships. Right pie charts in the inset box show results when only differences greater than 2 log likelihoods are considered. See also Table 1.

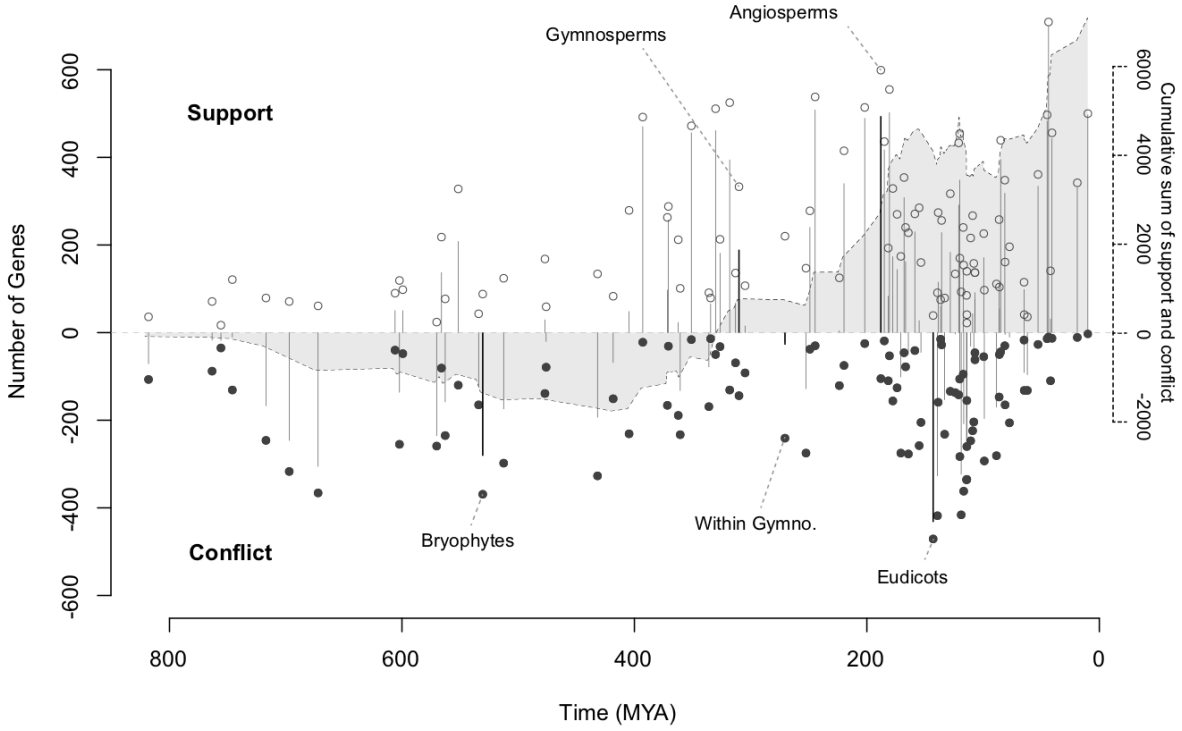

Figure 3: Examination of support and conflict in relation to time across all nodes with node ages taken from TimeTree (Hedges, Dudley, and Kumar 2006; Hedges et al. 2015). The differences between support and conflict are noted with vertical lines. The cumulative sum of support and conflict through time is noted in solid grey. Focal nodes from Fig. 2 are identified.

Nested analyses

Given the variation in support and conflict through time (Fig. 3), many genes that contain signal for a particular relationship may disagree with the resolution at other nodes. To examine these patterns of nested conflict, we examined the genes that support the resolution of the eudicot relationships (Fig. 4). In a set of 127 genes which supported the eudicot relationships recovered in the original ML analysis, 98 survived filtering for outgroup placement, branch length, and support with a statistically significant difference in lnL (> 2; Edwards 1984). 63 of these genes supported the monophyly of gymnosperms, and among those 63 only 25 supported a sister relationship between pines and *Gnetum*.

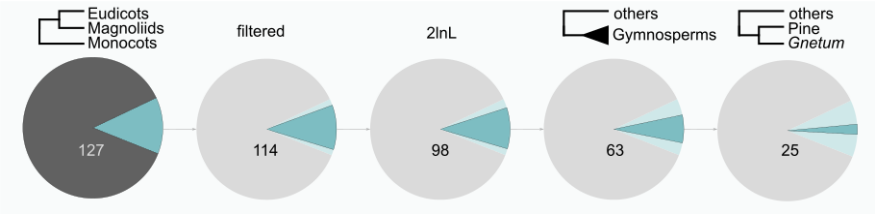

Figure 4: Nested patterns of support with genes associated with the resolution of eudicots. From left to right are shown the genes that support eudicots as sister to magnoliids (far left), those genes filtered as not having any outgroup errors or long branch lengths, those genes that support the resolution by at least 2lnL, those genes that support monophyletic gymnosperms, and finally those genes that support the Gnepine relationship.

Simulations of combinability

The procedure described here consists of two components: the information criterion for testing model complexity and the hill-climbing greedy clustering algorithm. First we conducted analyses to compare the performance of the difference information criteria measures (Fig. 5). In our tests, BIC with scaled branch lengths performed the best overall while AICc with scaled branch lengths performed well when branch length heterogeneity was low but poorly when branch length heterogeneity was medium to high. AICc with independent branch lengths tended to overfit when tree depths were higher but was more consistent across a range of branch length heterogeneity than any other information criterion. BIC with independent branch lengths (not shown) failed to recover any clusters and therefore was not considered further. High branch length heterogeneity generally resulted in overfitting. Because of the propensity of AICc with independent branch lengths to erroneously split clusters with both increasing tree depth and low levels of branch length heterogeneity, we did not consider it further.

Phylogenomic datasets often have only partially overlapping taxa sets for each gene, therefore we tested the influence of this in two ways (Fig. 5B). First, we randomly removed from one to three taxa for a single gene. These results demonstrate that the procedure will tend to overfit as the number of missing taxa increases. AICc with scaled branch lengths was highly sensitive to missing taxa, with between 33% and 87% overfitting for missing taxa in one gene and only one replicate correctly recovering one cluster for the highest amount of missing taxa in all genes. BIC with scaled branch lengths was less sensitive to missing taxa, with between 4% and 12% overfitting for missing taxa in one gene, and up to 52% overfitting for missing taxa in all genes.

The results above all had the same underlying species tree topology for each gene simulated. In order to determine not only whether the procedure overfitted models, we also examined the ability for the

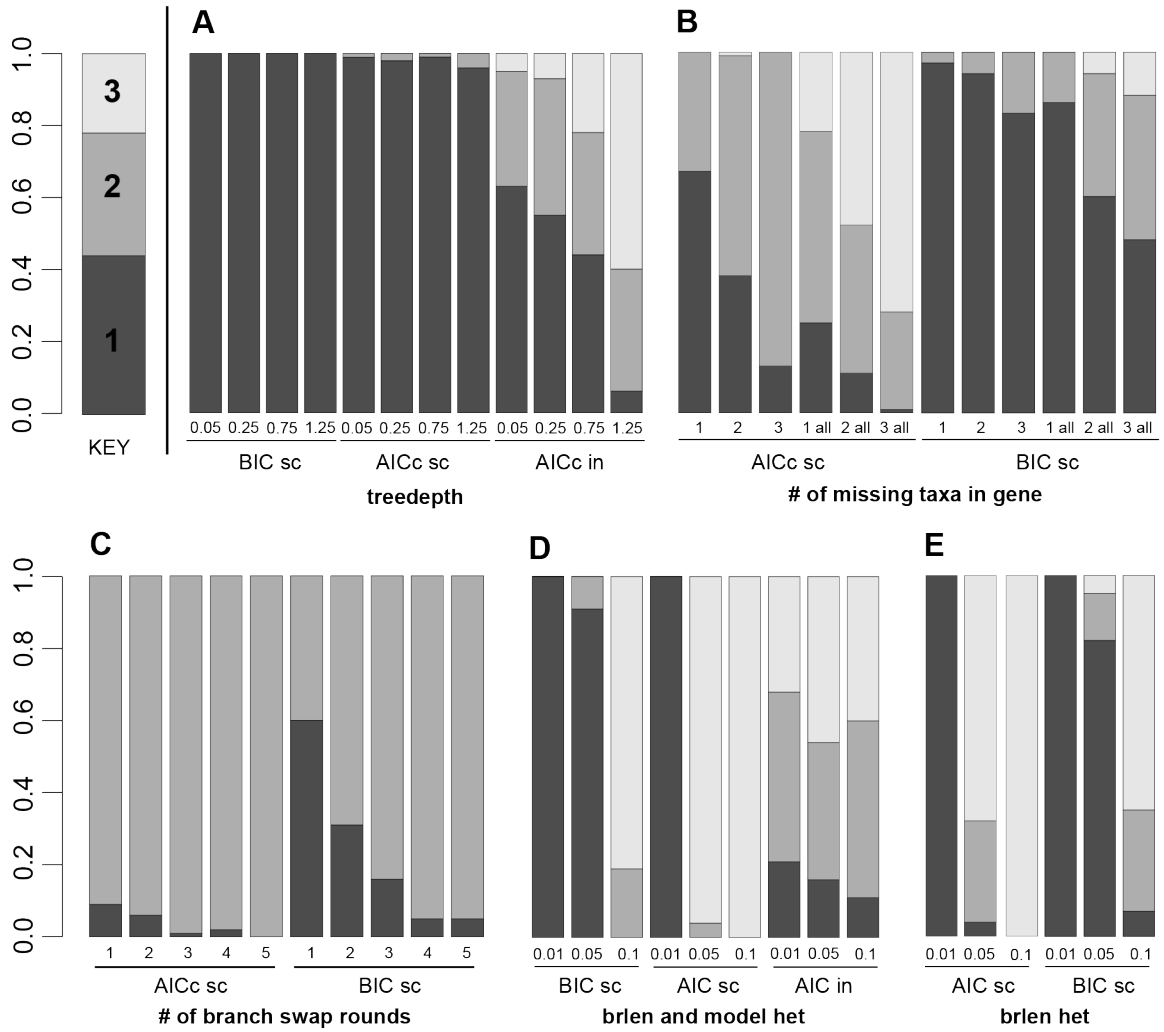

Figure 5: Simulations of clustering behaviour for the information criteria-based clustering under different models and data perturbation. 'sc' indicates that branch lengths are scaled (proportional) between gene regions, while 'in' indicates independent branch lengths. **A** performance of varied tree depths for three gene regions simulated on the same topology. Ideally all would recover one cluster. **B** performance for decreasing taxon overlap, with three gene regions simulated on the same topology but with one gene missing 1-3 taxa or with all genes missing 1-3 taxa. Ideally all would recover one cluster. **C** ability to detect topological differences amongst three gene regions, with two simulated on one topology and one simulated on a topology 1-5 NNI moves away. Ideally all would recover two clusters. **D** performance of varied branch length heterogeneity for three gene regions simulated on the same topology using the same model. Ideally all would recover one cluster. **E** performance for three gene regions simulated under different models and increasing branch length heterogeneity on the same topology. Ideally all would recover one cluster.

procedure to correctly break up gene regions when underlying topologies differed (Fig. 5C). As the simulations were conducted with two topologies differing from one to five NNIs, we expected the procedure to identify two clusters. We found that AICc with scaled branch lengths was much more sensitive to topological differences, with a highest error of 9% of replicates, and perfect recovery at five NNIs. BIC with scaled branch lengths tended to underfit, with error rates up to 60%, and producing two clusters in 5% of replicates even at five NNIs.

While isolating the behavior of the information criteria in relation to tree depth and branch length heterogeneity is helpful, it is likely that most datasets will have variation in substitution models between genes as well (Fig. 5E). We found that the BIC with scaled branch lengths was mostly robust to model variation except in the presence of large branch length heterogeneity (i.e., 10% of total tree height). AICc with scaled branch lengths was prone to overfitting based on model discrepancies, particularly with increasing branch length heterogeneity, correctly recovering one cluster in all replicates with branch length heterogeneity of 0.01, but incorrectly recovering three clusters in all replicates with branch length heterogeneity of 0.01. The discrepancy between the branch length heterogeneity of 0.1 in this analysis and the one above reflect that there were six genes simulated in this case with two for each model versus three gene regions as above.

Initial observations from some empirical data suggested the potential for clusters to snowball in size. We therefore simulated increasing numbers of genes on the same topology and tested clustering them against a single gene simulated on a topology three NNI moves away. For an proportional branch length model, two clusters were obtained in all replicates regardless of the number of genes in the cluster, for both AICc and BIC. For an independent branch length model, two clusters were also obtained in all replicates for AICc and BIC (not shown).

Empirical combinability of genes

We greedily tested the combinability of genes sets based on Robinson-Foulds distances to examine whether genes can be justifiably concatenated despite heterogeneity in information content throughout the phylogeny. We refer to our method as the COMBination of datasets (COMB) method. Because our approach bears conceptual similarity to algorithms used to estimate the optimal partitioning schemes (e.g. PartitionFinder, Lanfear et al. 2012, 2016), we compared combinable subsets to those recommended by the implementation of the PartitionFinder algorithm in IQ-TREE (Kalyaanamoorthy et al. 2017, referred to as MERGE here). Since an exhaustive search of the entire dataset is intractable, we examined the combinability of those genes that support the eudicot lineages to be sister to the magnoliid lineages (Fig. 2). We conducted analyses of two sets of genes: those that support the relationship with greater than 2 lnL versus alternative relationships (98 genes; 'CombinedSet'), and those that display the relationship in the ML gene tree and have SH-aLRT support greater than 80 (44 genes; 'MLSet'). These two sets were chosen because the first set was already examined as part of this study and the second is a typical cutoff used in standard systematics analyses (Guindon et al. 2010).

No method or gene set supported the concatenation of all genes that supported the focal eudicot relationship (see Table 2). The COMB method on the 'CombinedSet' supported concatenation of only two sets: one of three genes and one of two. The MERGE method supported merging partitions of 46 genes out of 98 (see Table 2 for more details). MERGE supported partition merging for a much greater number of genes than COMB supported combination. The COMB and MERGE results did not contain any identical concatenated sets. We constructed phylogenies of each concatenated set and

compared the inferred topologies (Table 2). Despite filtering on the magnoliids as sister to eudicots relationship, not all concatenated sets recovered this relationship with greater than 80 SH-aLRT. In one case, a merged partition supported a contradictory relationship to the filtered one.

Table 2. Comparison of partitioned subsets between combining strategies

Algorithm	Gene set	Genes	Sets	Partitioned Topology	Subset Relationships
MERGE	combined	98	20 (2x4, 2x3, 16x2)	magnoliids+eudicots (100)	magnoliids+eudicots (40%)
	ML	44	9 (1x4, 2x3, 6x2)	magnoliids+eudicots (100)	magnoliids+eudicots (67%)
					monocots+eudicots (11%)
COMB	combined	98	2 (1x3, 1x2)	magnoliids+eudicots (100)	magnoliids+eudicots(100%)
	ML	44	5 (1x3, 4x2)	magnoliids+eudicots (100)	magnoliids+eudicots (25%)

Brackets following a partitioned topology give the SH-aLRT score for that branch, while percentages following a subset relationship give the proportion of individual partition gene trees supporting the specified relationship with ≥ 80 SH-aLRT

Discussion

Conflict analysis

Several contentious relationships show strong contrast between the number of genes supporting the relationship, the number of genes *strongly* supporting the relationship (>2 lnL), the lnL supporting the relationship, and the lnL of genes that strongly support the relationship. Our analyses demonstrate that the differences in the number of gene trees supporting relationships and the difference in the summed likelihoods can provide insight into the cause for discordance between concatenated ML analyses and coalescent analyses. For example, the relationship involving Gnetales and the conifers as sister (Gnetifers) was recovered in coalescent-based analysis and is supported by more genes. However, the sum of the differences in the log-likelihoods of alternative resolutions support the Gnepine relationship (i.e., Gnetales sister to Pinales), the relationship found in the ML supermatrix analyses. Other relationships, including the placement of *Amborella* (Table 1), unequivocally support *Amborella* as sister to the rest of the angiosperms. For some relationships, gene support was equivocal (e.g. for relationships in eudicots and Bryophytes), but differences in *strongly* supporting genes and in summed lnL differences showed a clear preference.

Nested analysis

Filtering genes by the specific relationship they display provides an opportunity to examine nested conflicts (i.e., subsets of genes that do not conflict in one relationship may conflict in another). Furthermore, if conflict was reduced as a result of filtering, concatenation may be more tenable on such a filtered datasets. However, our nested conflict analyses demonstrated significant conflict and variation in the support for different relationships (Fig. 4) and that filtering genes based on specific relationships did not reduce conflict in other parts of the tree. While filtering genes may provide some means for lessening some systematic errors (Brown and Thomson 2016), or reducing some conflict (the question-specific 'node-control' approach of Chen, Liang, and Zhang (2015)) our analyses suggest that it will not likely solve general problems regarding conflicting genes.

A test for the combinability of genes

It is perhaps naïve to expect a single gene to have high support throughout a large part of the Tree of Life (see Penny et al. (1990); MUTOG: the 'Myth of a Universal Tree from One Gene'). For this reason, some researchers have thus argued that concatenating genes effectively combines data informative at various scales and so provides the necessary information to better resolve deep and shallow nodes (e.g., Mirarab, Bayzid, et al. 2014). Despite the potential benefits of concatenation (i.e., amplifying weak phylogenetic signal), the underlying model of evolution for a concatenated analysis assumes topological concordance among gene tree histories. Extensive gene conflicts should often violate these assumptions. Filtering genes could be one means of reducing conflict, though our filtered analyses demonstrated that conflict remained in other parts of the tree. However, this conflict may have been weak enough to still support concatenation. Whether genes should be combined for a concatenated analysis has been discussed at length (Huelsenbeck, Bull, and Cunningham 1996; Leigh et al. 2008; Seo and Thorne 2018; Theobald 2010; Walker, Brown, and Smith 2018) and Bayesian methods have recently been developed to address some of these issues (Neupane et al. 2018). However, due to the large scale of genomic datasets, Bayesian methods are often computationally intractable.

We developed a heuristic to test if genes should be combined based on information criteria, and validated its performance through simulation. Our approach bears some similarity to methods which test the combinability of partition models in concatenated analyses (Lanfear et al. 2012, 2016), but additionally considers topological heterogeneity between gene regions, rather than evaluating them on a fixed topology (Neupane et al. 2018; Seo and Thorne 2018). In some cases the two approaches are expected to perform similarly. For example, if two genes have identical topologies, then our results and the results of PartitionFinder should be identical. One key difference lies in the interpretation of the results. If two genes are not merged in a PartitionFinder-like analysis, they are still included in the same concatenation analysis, albeit in different partitions. However, if two genes are not clustered by our approach, we argue that they should not be concatenated at all.

Simulations demonstrated that our approach performed well with clustering success decreasing with increasing tree depth and increasing branch length heterogeneity (Fig. 5). Simply put, trees that were more different were easier to separate into clusters. Overfitting increased as taxon overlap was reduced. Based on these results, we find that our method provides a feasible approach to partition data into combinable subsets and to determine the degree of combinability (or lack thereof) of a set of genes. Despite the shortcuts employed, however, it may still involve long computational times or be intractable for some large datasets. Therefore, methods that reduce computational time, for

example the training of machine learning discriminative models for metrics like RFW from data subsets, could be explored. Because of the extensive gene tree conflict within datasets and the improbable nature of supporting combining genes that differ extensively in topology, generally researchers would be better to test subsets of the datasets instead of the entire dataset, reducing computational time and effort extensively. Additionally, the results of our simulations show that different information criteria and branch length models may be applicable in different situations. For example, AICc with scaled branch lengths is likely to produce few clusters when gene tree conflict is extensive, while BIC with scaled branch lengths may produce more. Therefore, researchers wishing to apply our approach should consider the characteristics of the data they are analyzing when making this choice.

Combinability of empirical data

Using our heuristic, we tested combinability of the subset of the genes from Wickett et al. (2014) that supported magnoliids sister to eudicots as inferred in the original ML analysis. We found that only a very small set of genes supported concantenation. Because concatenation is a common means for analyzing large phylogenomic analyses, it may be surprising that our metric does not support widespread concatenation. However, given the extensive underlying gene tree conflict that remains even after filtering for a particular focal relationship (Fig. 4) this should be expected. In particular, simulations demonstrated that our approach using AICc with scaled branch lengths is very sensitive to topological heterogeneity. Therefore, very small numbers of concatenated sets are probably the result of the extensive gene tree conflict that remains even after node-specific filtering. Furthermore, it is notable that even after filtering for gene trees supporting a particular relationships, some concatenated subsets still did not provide strong support for that relationship. While concatenation can be helpful for exploratory inference to identify dominant signal, it is not capable of addressing specific and contentious relationships. We suggest that when exploring specific relationships analyses such as those described above should be used to uncover the most robust phylogenetic hypothesis upon which to base other evolutionary hypotheses.

Implications for plant phylogenetics

The results presented here provide strong support for several relationships that have long been considered contentious, and indicate probable resolutions for others. For example, we found support for *Amborella* being sister to the rest of angiosperms and that gymnosperms are monophyletic. Several relationships (e.g., among the eudicots and relatives as well as the hornworts, liverworts, and mosses) lack enough information to confidently accept any of the alternative resolutions. Rather than being dismayed at this apparent failure, we regard this lack of signal as extremely valuable information, as it informs where future effort should be focused. Though we identified the relationship that was more strongly supported by the data (Table 1), the differences between the alternatives were so slight that the current dataset is likely unable to confidently resolve this debate and conducting additional analyses with expanded taxa and gene regions is warranted.

Among the strongly supported hypotheses, the placement of *Amborella* continues to be a point of major contention within the plant community. *Amborella* is a tropical tree with relatively small flowers, while the Nymphaeales are aquatic plants with relatively large flowers. The resolution of these taxa in relation to the remainder of the flowering plants will inform the life history or early angiosperms (Feild et al. (2004)) as well as the lability of life history and floral traits. Our results suggest *Amborella* is

sister to all other extant angiosperms, and imply that rates of evolution need not be particularly fast in order to understand the morphological differences between a tropical tree (*Amborella*) and water lilies (Nymphaeales). Strong support for the monophyly of gymnosperms implies that the morphological disparity of extant gymnosperm taxa, including the especially diverse Gnetales, emerged post-divergence from the angiosperm lineage. This reinforces analyses of LEAFY homologs, which recover gymnosperm paralogs as monophyletic groups (Sayou et al. 2014), and also lends support to shared characteristics between Gnetales and angiosperms resulting from convergent evolution (Bowe, Coat, and dePamphilis 2000; Hansen et al. 1999).

For contentious relationships only weakly supported here, there are several biological questions that will be answered once these are confidently resolved. The data and analyses presented here suggest that hornworts are sister to all other land plants. This is consistent with some studies (Nickrent et al. 2000; Nishiyama and Kato 1999), but contradicts the results of others (Cox et al. 2014; Karol et al. 2010; Qiu et al. 2006), including some but not all results of a recent re-analysis of this dataset (Puttick et al. 2018). If the position of hornworts presented here holds with additional data, it implies that the absence of stomata in liverworts and some mosses is a derived state resulting from loss of the trait, suggests a single loss of pyrenoids in non-hornwort land plants (but see Villarreal and Renner 2012), and questions some inferences on the characteristics of hornwort sporophytes (Qiu et al. 2006). Among gymnosperms, these data suggest that Gnetales are sister to pines (the "Gnepine" hypothesis; Chaw et al. 2000), further supporting the lability and rapid evolution of morphological disparity within the group. Finally, magnoliids are inferred as sister to the eudicot lineages, which has implications on the origin and divergence times of eudicots and monocots.

Despite the ability of the methods explored here to accomodate the underlying gene tree uncertainty, our results depend on the information available in the underlying dataset. While this dataset is not comprehensive, it *does* represent extensive sequencing of transcriptomes and genomes for the taxa included. We can say, with confidence, what these data support or do not support, but different datasets (e.g., based on different taxa, different homology analyses) may have stronger signal for relationships that are resolved more equivocally here. We recommend analyzing these future datasets with an eye toward hypotheses of specific phylogenetic relationships. Our novel approach provides insight into several of the most contentious relationships across land plants and is broadly applicable among different groups. Approaches that ascertain the support for alternative resolutions should be used to resolve contentious branches across the Tree of Life.

Implications for future phylogenomic studies

A panacea does not currently exist for phylogenomic analyses. Some researchers aim to determine the relative support for contentious relationships. Others want to construct a reasonable, if not ideal, phylogeny for downstream analyses. Others still may be primarily interested in gene trees. Here, we suggest that more detailed analyses of the gene trees will yield more informative results regarding the information within a particular dataset and the ability of the dataset to resolve relationships. Our results also speak to the common analyses conducted on phylogenomic datasets.

The underlying conflict identified by many researchers (Wickett et al. 2014; Puttick et al. 2018) suggests that concatenation, while helpful for identifying the dominant signal, should not be used to address contentious nodes. Our targeted exploration of the combinability of gene regions found that very few genes are optimally modelled by concatenation, even when filtering on those genes that

support a relationship. However, our analyses of combinability leaves many unanswered questions. For example, how should we adequately address the problem of low signal when gene tree conflict is high and concatenation is statistically unsupported? Are genes that are statistically supported to be analyzed together linked? And perhaps, most importantly, when faced with several clusters of combined genes, how does one move forward with inference? Some have suggested feeding the clusters into a coalescent analysis (Mirarab, Bayzid, et al. 2014), however this most likely violates many assumptions of the coalescent. Alternatively, researchers are faced with multiple species trees. Here, we suggest that examining each of the dominant relationships in more detailed helps resolve these conflicts, though additional work is necessary to translate these results to species tree analyses.

The most common alternative to concatenation, coalescent species tree approaches, often accomodate one major source of conflict in gene trees without concatenation, ILS (Mirarab, Reaz, et al. 2014). However, the most sophisticated model-based coalescent approaches are often not computationally tractable for phylogenomic analyses because of the large sizes of the datasets (Ané et al. 2006; Boussau et al. 2013). Instead, most phylogenomic analyses that accommodate ILS use quartet methods (e.g., ASTRAL) that, while fast and effective, do not account for multiple sources of conflict and make several other assumptions that may or may not be reasonable given the dataset (e.g. equal weighting of gene trees regardless of properties of the underlying genes). Some researchers have suggested filtering the data to include only those genes that conflict due to ILS (Knowles et al. 2018; Huang et al. 2017) or that agree with accepted relationships or specific relationships to be tested (Chen, Liang, and Zhang 2015; Doyle et al. 2015; Smith, Brown, and Walker 2018). However, for datasets with a broad scope, several processes may be at play throughout the phylogeny and it may not be possible to filter based on a single underlying process.

While a single species tree may be necessary for some downstream analyses, these obfuscate the biological realities that underlie these data. By uncovering the support and lack thereof, we can determine the limits of our data, identify troublesome phylogenetic relationships that require more attention, and put to rest debates over specific relationships (at least in regard to specific datasets). The approach we adopt here is akin to the 'hypothesis-control' method of Chen, Liang, and Zhang (2015), but instead of relying on the results of typical inference on the filtered subsets, we profile the signal for different resolutions and processes within them. Overall, we suggest that species trees, because of the cacophany of signal and conflict, are not the best units of analysis for resolving specific relationships. Instead, analyses which focus on the support for a particular relationship in isolation, without requiring the data to speak to the full set of relationships in a species tree, should be pursued.

Acknowledgments

This work was supported by funding from NSF DEB 1354048 (J.F.W. and S.A.S.) and NSF AVATOL 1207915 (J.W.B. and S.A.S.). We appreciate comments from Ning Wang, Caroline Parins-Fukuchi, Diego Alvarado Serrano, Greg Stull, Drew Larson, Hector Fox, and Richie Hodel.

Literature Cited

Ahrenfeldt, Johanne, Carina Skaarup, Henrik Hasman, Anders Gorm Pedersen, Frank Møller Aarestrup, and Ole Lund. 2017. "Bacterial Whole Genome-Based Phylogeny: Construction of a New Benchmarking Dataset and Assessment of Some Existing Methods." *BMC Genomics* 18 (1). BioMed Central: 19.

Akaike, Hirotogu. 1973. "Information Theory and an Extension of the Maximum Likelihood Principle." In *Second International Symposium on Information Theory*, edited by Petrov P. N. and Csaki F., 267–81. Akademiai Kiado.

Ané, Cécile, Bret Larget, David A Baum, Stacey D Smith, and Antonis Rokas. 2006. "Bayesian Estimation of Concordance Among Gene Trees." *Molecular Biology and Evolution* 24 (2). Oxford University Press: 412–26.

Boussau, Bastien, Gergely J Szöllősi, Laurent Duret, Manolo Gouy, Eric Tannier, and Vincent Daubin. 2013. "Genome-Scale Coestimation of Species and Gene Trees." *Genome Research* 23 (2). Cold Spring Harbor Lab: 323–30.

Bowe, L Michelle, Gwénaële Coat, and Claude 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 (8). National Acad Sciences: 4092–7.

Brown, Jeremy M, and Robert C Thomson. 2016. "Bayes Factors Unmask Highly Variable Information Content, Bias, and Extreme Influence in Phylogenomic Analyses." *Systematic Biology* 66 (4). Oxford University Press: 517–30.

Brown, Joseph W, Joseph F Walker, and Stephen A Smith. 2017. "Phyx: Phylogenetic Tools for Unix." *Bioinformatics* 33 (12). Oxford University Press: 1886–8.

Burnham, Kenneth P, and David R Anderson. 2003. *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*. Springer Science & Business Media.

Chaw, Shu-Miaw, Christopher L Parkinson, Yuchang Cheng, Thomas M Vincent, and Jeffrey 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* 97 (8). National Acad Sciences: 4086–91.

Chen, Meng-Yun, Dan Liang, and Peng Zhang. 2015. "Selecting Question-Specific Genes to Reduce Incongruence in Phylogenomics: A Case Study of Jawed Vertebrate Backbone Phylogeny." *Systematic Biology* 64 (6). Oxford University Press: 1104–20.

Cox, Cymon J., Blaise Li, Peter G. Foster, T. Martin Embley, and Peter Civáň. 2014. "Conflicting Phylogenies for Early Land Plants Are Caused by Composition Biases Among Synonymous Substitutions." *Systematic Biology* 63 (2). Oxford University Press: 272–79. http://dx.doi.org/10.1093/sysbio/syt109.

Doyle, Vinson P, Randee E Young, Gavin JP Naylor, and Jeremy M Brown. 2015. "Can We Identify Genes with Increased Phylogenetic Reliability?" *Systematic Biology* 64 (5). Oxford University Press: 824–37.

Dunn, Casey W, Andreas Hejnol, David Q Matus, Kevin Pang, William E Browne, Stephen A Smith, Elaine Seaver, et al. 2008. "Broad Phylogenomic Sampling Improves Resolution of the Animal Tree of Life." *Nature* 452 (7188). Nature Publishing Group: 745.

Edwards, Anthony William Fairbank. 1984. *Likelihood*. CUP Archive.

Edwards, Scott V. 2009. "Is a New and General Theory of Molecular Systematics Emerging?" *Evolution* 63 (1). Wiley Online Library: 1–19.

Edwards, Scott V, Liang Liu, and Dennis K Pearl. 2007. "High-Resolution Species Trees Without Concatenation." *Proceedings of the National Academy of Sciences* 104 (14). National Acad Sciences: 5936–41.

Edwards, Scott V, Zhenxiang Xi, Axel Janke, Brant C Faircloth, John E McCormack, Travis C Glenn, Bojian Zhong, et al. 2016. "Implementing and Testing the Multispecies Coalescent Model: A Valuable Paradigm for Phylogenomics." *Molecular Phylogenetics and Evolution* 94. Elsevier: 447–62.

Feild, Taylor S, Nan Crystal Arens, James A Doyle, Todd E Dawson, and Michael J Donoghue. 2004. "Dark and Disturbed: A New Image of Early Angiosperm Ecology." *Paleobiology* 30 (1). BioOne: 82–107.

Feuda, Roberto, Martin Dohrmann, Walker Pett, Hervé Philippe, Omar Rota-Stabelli, Nicolas Lartillot, Gert Wörheide, and Davide Pisani. 2017. "Improved Modeling of Compositional Heterogeneity Supports Sponges as Sister to All Other Animals." *Current Biology* 27 (24). Elsevier: 3864–70.

Fletcher, William, and Ziheng Yang. 2009. "INDELible: A Flexible Simulator of Biological Sequence Evolution." *Molecular Biology and Evolution* 26 (8). SMBE: 1879–88.

Guindon, Stéphane, Jean-François Dufayard, Vincent Lefort, Maria Anisimova, Wim Hordijk, and Olivier Gascuel. 2010. "New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of Phyml 3.0." *Systematic Biology* 59 (3). Oxford University Press: 307–21.

Hansen, Andrea, Sabine Hansmann, Tahir Samigullin, Andrey Antonov, and William Martin. 1999. "Gnetum and the Angiosperms: Molecular Evidence That Their Shared Morphological Characters Are Convergent, Rather Than Homologous." *Molecular Biology and Evolution* 16 (7). Oxford University Press: 1006–6.

Hedges, S Blair, Joel Dudley, and Sudhir Kumar. 2006. "TimeTree: A Public Knowledge-Base of Divergence Times Among Organisms." *Bioinformatics* 22 (23). Oxford University Press: 2971–2.

Hedges, S. Blair, Julie Marin, Michael Suleski, Madeline Paymer, and Sudhir Kumar. 2015. "Tree of Life Reveals Clock-Like Speciation and Diversification." *Molecular Biology and Evolution* 32 (4): 835–45. https://doi.org/10.1093/molbev/msv037.

Huang, Huateng, Jeet Sukumaran, Stephen A Smith, and LLacey Knowles. 2017. "Cause of Gene Tree Discord? Distinguishing Incomplete Lineage Sorting and Lateral Gene Transfer in Phylogenetics." *PeerJ PrePrints* 5. PeerJ, Inc.: e3489v1.

Huelsenbeck, John P, JJ Bull, and Clifford W Cunningham. 1996. "Combining Data in Phylogenetic Analysis." *Trends in Ecology & Evolution* 11 (4). Elsevier: 152–58.

Jarvis, Erich D, Siavash Mirarab, Andre J Aberer, Bo Li, Peter Houde, Cai Li, Simon YW Ho, et al. 2014. "Whole-Genome Analyses Resolve Early Branches in the Tree of Life of Modern Birds." *Science* 346 (6215). American Association for the Advancement of Science: 1320–31.

Kalyaanamoorthy, Subha, Bui Quang Minh, Thomas KF Wong, Arndt von Haeseler, and Lars S Jermiin. 2017. "ModelFinder: Fast Model Selection for Accurate Phylogenetic Estimates." *Nature Methods* 14 (6). Nature Publishing Group: 587.

Karol, Kenneth G., Kathiravetpillai Arumuganathan, Jeffrey L. Boore, Aaron M. Duffy, Karin DE Everett, John D. Hall, S. Kellon Hansen, et al. 2010. "Complete Plastome Sequences of Equisetum

631 Arvense and Isoetes Flaccida: Implications for Phylogeny and Plastid Genome Evolution of Early Land
632 Plant Lineages." *BMC Evolutionary Biology* 10 (1): 321. https://doi.org/10.1186/1471-2148-10-321.

633 Kluge, Arnold G. 1989. "A Concern for Evidence and a Phylogenetic Hypothesis of Relationships
634 Among Epicrates (Boidae, Serpentes)." *Systematic Biology* 38 (1). Society of Systematic Zoology: 7–25.

635 Knowles, L Lacey, Huateng Huang, Jeet Sukumaran, and Stephen A Smith. 2018. "A Matter of
636 Phylogenetic Scale: Distinguishing Incomplete Lineage Sorting from Lateral Gene Transfer as the Cause
637 of Gene Tree Discord in Recent Versus Deep Diversification Histories." *American Journal of Botany*
638 105 (3). Wiley Online Library: 376–84.

639 Kosakovsky Pond, Sergei L, David Posada, Michael B Gravenor, Christopher H Woelk, and Simon
640 DW Frost. 2006a. "Automated Phylogenetic Detection of Recombination Using a Genetic Algorithm."
641 *Molecular Biology and Evolution* 23 (10). Oxford University Press: 1891–1901.

642 ———. 2006b. "GARD: A Genetic Algorithm for Recombination Detection." *Bioinformatics* 22 (24).
643 Oxford University Press: 3096–8.

644 Lanfear, Robert, Brett Calcott, Simon YW Ho, and Stephane Guindon. 2012. "PartitionFinder:
645 Combined Selection of Partitioning Schemes and Substitution Models for Phylogenetic Analyses."
646 *Molecular Biology and Evolution* 29 (6). Oxford University Press: 1695–1701.

647 Lanfear, Robert, Paul B Frandsen, April M Wright, Tereza Senfeld, and Brett Calcott. 2016. "Partition-
648 Finder 2: New Methods for Selecting Partitioned Models of Evolution for Molecular and Morphological
649 Phylogenetic Analyses." *Molecular Biology and Evolution* 34 (3). Oxford University Press: 772–73.

650 Leigh, Jessica W, Edward Susko, Manuela Baumgartner, and Andrew J Roger. 2008. "Testing
651 Congruence in Phylogenomic Analysis." *Systematic Biology* 57 (1). Taylor & Francis: 104–15.

652 Liu, Liang, Lili Yu, Dennis K Pearl, and Scott V Edwards. 2009. "Estimating Species Phylogenies
653 Using Coalescence Times Among Sequences." *Systematic Biology* 58 (5). Oxford University Press:
654 468–77.

655 Maddison, Wayne P. 1997. "Gene Trees in Species Trees." *Systematic Biology* 46 (3). Oxford University
656 Press: 523–36.

657 Mirarab, Siavash, Md Shamsuzzoha Bayzid, Bastien Boussau, and Tandy Warnow. 2014. "Statistical
658 Binning Enables an Accurate Coalescent-Based Estimation of the Avian Tree." *Science* 346 (6215).
659 American Association for the Advancement of Science: 1250463.

660 Mirarab, Siavash, Rezwana Reaz, Md S Bayzid, Théo Zimmermann, M Shel Swenson, and Tandy
661 Warnow. 2014. "ASTRAL: Genome-Scale Coalescent-Based Species Tree Estimation." *Bioinformatics*
662 30 (17). Oxford University Press: i541–i548.

663 Neupane, Suman, Karolina Fucikova, Louise A Lewis, Lynn Kuo, Ming-Hui Chen, and Paul Lewis.
664 2018. "Assessing Combinability of Phylogenomic Data Using Bayes Factors." *bioRxiv*. Cold Spring
665 Harbor Laboratory, 250969.

666 Nguyen, Lam-Tung, Heiko A Schmidt, Arndt von Haeseler, and Bui Quang Minh. 2014. "IQ-Tree: A
667 Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies." *Molecular*
668 *Biology and Evolution* 32 (1). Oxford University Press: 268–74.

669 Nickrent, Daniel L., Christopher L. Parkinson, Jeffrey D. Palmer, and R. Joel Duff. 2000. "Multigene
670 Phylogeny of Land Plants with Special Reference to Bryophytes and the Earliest Land Plants." *Molecular*

Biology and Evolution 17 (12): 1885–95. https://doi.org/10.1093/oxfordjournals.molbev.a026290.

Nishiyama, T, and M Kato. 1999. "Molecular Phylogenetic Analysis Among Bryophytes and Tracheophytes Based on Combined Data of Plastid Coded Genes and the 18S rRNA Gene." *Molecular Biology and Evolution* 16 (8): 1027–36. https://doi.org/10.1093/oxfordjournals.molbev.a026192.

Penny, D, MD Hendy, EA Zimmer, and RK Hamby. 1990. "Trees from Sequences: Panacea or Pandora's Box?" *Australian Systematic Botany* 3 (1). CSIRO: 21–38.

Puttick, Mark N, Jennifer L Morris, Tom A Williams, Cymon J Cox, Dianne Edwards, Paul Kenrick, Silvia Pressel, et al. 2018. "The Interrelationships of Land Plants and the Nature of the Ancestral Embryophyte." *Current Biology* 28 (5). Elsevier: 733–45.

Qiu, Yin-Long, Libo Li, Bin Wang, Zhiduan Chen, Volker Knoop, Milena Groth-Malonek, Olena Dombrovska, et al. 2006. "The Deepest Divergences in Land Plants Inferred from Phylogenomic Evidence." *Proceedings of the National Academy of Sciences* 103 (42). National Academy of Sciences: 15511–6. https://doi.org/10.1073/pnas.0603335103.

Robinson, David F, and Leslie R Foulds. 1981. "Comparison of Phylogenetic Trees." *Mathematical Biosciences* 53 (1-2). Elsevier: 131–47.

Sauquet, Hervé, Maria von Balthazar, Susana Magallón, James A Doyle, Peter K Endress, Emily J Bailes, Erica Barroso de Morais, et al. 2017. "The Ancestral Flower of Angiosperms and Its Early Diversification." *Nature Communications* 8. Nature Publishing Group: 16047.

Sayou, Camille, Marie Monniaux, Max H Nanao, Edwige Moyroud, Samuel F Brockington, Emmanuel Thévenon, Hicham Chahtane, et al. 2014. "A Promiscuous Intermediate Underlies the Evolution of Leafy Dna Binding Specificity." *Science* 343 (6171). American Association for the Advancement of Science: 645–48.

Schwarz, Gideon. 1978. "Estimating the Dimension of a Model." *The Annals of Statistics* 6 (2). Institute of Mathematical Statistics: 461–64.

Seo, Tae-Kun, and Jeffrey L Thorne. 2018. "Information Criteria for Comparing Partition Schemes." *Systematic Biology* 67 (4): 616–32. https://doi.org/10.1093/sysbio/syx097.

Shen, Xing-Xing, Chris Todd Hittinger, and Antonis Rokas. 2017. "Contentious Relationships in Phylogenomic Studies Can Be Driven by a Handful of Genes." *Nature Ecology & Evolution* 1 (5). Nature Publishing Group: 0126.

Shen, Xing-Xing, Xiaofan Zhou, Jacek Kominek, Cletus P Kurtzman, Chris Todd Hittinger, and Antonis Rokas. 2016. "Reconstructing the Backbone of the Saccharomycotina Yeast Phylogeny Using Genome-Scale Data." *G3: Genes, Genomes, Genetics* 6 (12). G3: Genes, Genomes, Genetics: 3927–39.

Simion, Paul, Hervé Philippe, Denis Baurain, Muriel Jager, Daniel J. Richter, Arnaud Di Franco, Béatrice Roure, et al. 2017. "A Large and Consistent Phylogenomic Dataset Supports Sponges as the Sister Group to All Other Animals." *Current Biology* 27 (7): 958–67. https://doi.org/https://doi.org/10.1016/j.cub.2017.02.031.

Smith, Stephen A, Joseph W Brown, and Joseph F Walker. 2018. "So Many Genes, so Little Time: A Practical Approach to Divergence-Time Estimation in the Genomic Era." *PloS One* 13 (5). Public Library of Science: e0197433.

Smith, Stephen A, Michael J Moore, Joseph W Brown, and Ya Yang. 2015. "Analysis of Phylogenomic Datasets Reveals Conflict, Concordance, and Gene Duplications with Examples from Animals and Plants." *BMC Evolutionary Biology* 15 (1). BioMed Central: 150.

Solís-Lemus, Cécile, Claudia AND Ané. 2016. "Inferring Phylogenetic Networks with Maximum Pseudolikelihood Under Incomplete Lineage Sorting." *PLOS Genetics* 12 (3). Public Library of Science: 1–21. https://doi.org/10.1371/journal.pgen.1005896.

Springer, Mark S, and John Gatesy. 2016. "The Gene Tree Delusion." *Molecular Phylogenetics and Evolution* 94. Elsevier: 1–33.

Theobald, Douglas L. 2010. "A Formal Test of the Theory of Universal Common Ancestry." *Nature* 465 (7295). Nature Publishing Group: 219.

Villarreal, Juan Carlos, and Susanne S Renner. 2012. "Hornwort Pyrenoids, Carbon-Concentrating Structures, Evolved and Were Lost at Least Five Times During the Last 100 Million Years." *Proceedings of the National Academy of Sciences* 109 (46). National Acad Sciences: 18873–8.

Walker, Joseph F, Joseph W Brown, and Stephen A Smith. 2018. "Analyzing Contentious Relationships and Outlier Genes in Phylogenomics." *Systematic Biology* syy043. Oxford University Press. https://doi.org/10.1093/sysbio/syy043.

Walker, Joseph F, Ya Yang, Tao Feng, Alfonso Timoneda, Jessica Mikenas, Vera Hutchison, Caroline Edwards, et al. 2018. "From Cacti to Carnivores: Improved Phylotranscriptomic Sampling and Hierarchical Homology Inference Provide Further Insight into the Evolution of Caryophyllales." *American Journal of Botany* 105 (3). Wiley Online Library: 446–62.

Walker, Joseph F, Ya Yang, Michael J Moore, Jessica Mikenas, Alfonso Timoneda, Samuel F Brockington, and Stephen A Smith. 2017. "Widespread Paleopolyploidy, Gene Tree Conflict, and Recalcitrant Relationships Among the Carnivorous Caryophyllales." *American Journal of Botany* 104 (6). Botanical Soc America: 858–67.

Wen, Dingqiao, Yun Yu, Jiafan Zhu, and Luay Nakhleh. 2018. "Inferring Phylogenetic Networks Using Phylonet." *Systematic Biology* 67 (4): 735–40. https://doi.org/10.1093/sysbio/syy015.

Whelan, Nathan V, Kevin M Kocot, Tatiana P Moroz, Krishanu Mukherjee, Peter Williams, Gustav Paulay, Leonid L Moroz, and Kenneth M Halanych. 2017. "Ctenophore Relationships and Their Placement as the Sister Group to All Other Animals." *Nature Ecology & Evolution* 1 (11). Nature Publishing Group: 1737.

Wickett, Norman J, Siavash Mirarab, Nam Nguyen, Tandy Warnow, Eric Carpenter, Naim Matasci, Saravanaraj Ayyampalayam, et al. 2014. "Phylotranscriptomic Analysis of the Origin and Early Diversification of Land Plants." *Proceedings of the National Academy of Sciences* 111 (45). National Acad Sciences: E4859–E4868.

Xi, Zhenxiang, Liang Liu, Joshua S Rest, and Charles C Davis. 2014. "Coalescent Versus Concatenation Methods and the Placement of Amborella as Sister to Water Lilies." *Systematic Biology* 63 (6). Oxford University Press: 919–32.

Yang, Ya, and Stephen A Smith. 2014. "Orthology Inference in Nonmodel Organisms Using Transcriptomes and Low-Coverage Genomes: Improving Accuracy and Matrix Occupancy for Phylogenomics." *Molecular Biology and Evolution* 31 (11). Oxford University Press: 3081–92.