Louisiana State University
LSU Scholarly Repository

Faculty Publications | Department of Biological Sciences

2-1-2007

Assessing systematic error in the inference of seed plant phylogeny

J. Gordon Burleigh

Sarah Mathews

Follow this and additional works at: https://repository.lsu.edu/biosci_pubs

Recommended Citation
Burleigh, J., & Mathews, S. (2007). Assessing systematic error in the inference of seed plant phylogeny. *International Journal of Plant Sciences, 168* (2), 125-135. https://doi.org/10.1086/509588

This Article is brought to you for free and open access by the Department of Biological Sciences at LSU Scholarly Repository. It has been accepted for inclusion in Faculty Publications by an authorized administrator of LSU Scholarly Repository. For more information, please contact ir@lsu.edu.

Int. J. Plant Sci. 168(2):125–135. 2007.
© 2007 by The University of Chicago. All rights reserved.
1058-5893/2007/16802-0002$15.00

ASSESSING SYSTEMATIC ERROR IN THE INFERENCE OF SEED PLANT PHYLOGENY

J. Gordon Burleigh[1] and Sarah Mathews

Section of Evolution and Ecology, University of California, Davis, California 95616, U.S.A.; and Arnold Arboretum of Harvard University, Cambridge, Massachusetts 02138, U.S.A.

We used parametric bootstrapping to assess the performance of maximum parsimony and maximum likelihood phylogenetic analyses of a 12-locus seed plant data set. Evidence of biases in maximum parsimony analyses of single-locus data sets may explain some of the locus-specific variation among DNA-based hypotheses of seed plant phylogeny. In particular, there is strong evidence of bias in maximum parsimony analyses, especially of plastid loci, that favors placing Gnetales sister to other seed plants. We concatenated simulated single-locus data sets to examine biases in analyses of a 12-locus data set in which each locus is simulated with different substitution parameters and branch lengths. Maximum parsimony analyses of the simulated 12-locus data set also show evidence of biases in favor of recovering trees with Gnetales sister to other seed plants and against recovering anthophyte, gnepine, and gnetifer trees. These biases are most evident in analyses that include the fastest-evolving characters. In the maximum likelihood analyses of the simulated 12-locus data sets, there is evidence of a bias against recovering the anthophyte hypothesis. Otherwise, there is little evidence that the heterogeneous branch lengths and substitution processes among loci influence the results from maximum likelihood phylogenetic analyses.

Keywords: seed plants, Gnetales, parametric bootstrapping, multigene analysis, systematic error, phylogeny.

Online enhancements: data files.

Introduction

Phylogenetic analyses of molecular data sets with large numbers of characters per taxon may not be as susceptible to random, or stochastic, error as analyses of data sets with fewer characters (Rokas et al. 2003). However, analyses of such large molecular data sets still are vulnerable to systematic errors or errors caused by failings of the phylogenetic method rather than a lack of data (Phillips et al. 2004; Soltis et al. 2004; Stefanovic et al. 2004). In fact, adding data can exacerbate the symptoms of systematic error by increasing support for erroneous relationships. Thus, it is critical to examine phylogenetic analyses for evidence of systematic error, even when the data sets are large and support for phylogenetic relationships is high. We explore the role of systematic error in phylogenetic analyses of a 12-locus data set from seed plants.

The evolutionary relationships among seed plant lineages remain disputed despite numerous phylogenetic analyses of increasingly large data sets (Bowe et al. 2000; Chaw et al. 2000; Magallón and Sanderson 2002; Rydin et al. 2002; Soltis et al. 2002; Rai et al. 2003; Burleigh and Mathews 2004, 2007). Many of the analyses reveal support for conflicting phylogenetic hypotheses. For example, maximum likelihood (ML) and maximum parsimony (MP) analyses often support different topologies (Magallón and Sanderson 2002; Soltis et al. 2002; Burleigh and Mathews 2004, 2007). Similarly, analyses of different loci, and even analyses of different partitions of the same locus, often give conflicting results (e.g., Chaw et al. 2000; Frohlich and Parker 2000; Sanderson et al. 2000; Magallón and Sanderson 2002; Rydin et al. 2002; Soltis et al. 2002; Burleigh and Mathews 2004, 2007). Although the variety of strongly supported phylogenetic results suggest that systematic error affects at least some of the seed plant phylogenetic analyses, it is difficult to diagnose systematic error from the results of the phylogenetic analyses alone.

Sanderson et al. (2000) used parametric bootstrapping to show evidence of systematic error in MP analyses of *psaA* and *psbB* from seed plants. Their results indicated a bias favoring the placement of Gnetales as sister to other seed plants in MP analyses of *psbB* third codon position sites and also revealed a bias against recovery of the anthophyte hypothesis from MP analyses of both loci (Sanderson et al. 2000). We use parametric bootstrapping to further examine the role of error in both ML and MP analyses of 11 individual loci and in a 12-locus concatenated data set previously used in the seed plant analyses of Burleigh and Mathews (2007). Specifically, we explore the extent to which systematic error contributes to the differences observed in MP and ML analyses, to locus-specific variation, and to the differences found in analyses of partitioned data sets. We also investigate the effect of heterogeneous processes of evolution on MP and ML analyses to see how well they perform in analyses of multilocus data set in which the patterns of evolution vary among the loci that have been sampled to infer seed plant phylogeny.

Methods

Data Sets and Previous Phylogenetic Analyses

The parametric bootstrapping experiments are based on the 12-locus seed plant data set used by Burleigh and Mathews

[1] Author for correspondence; e-mail jgb12@duke.edu.

Manuscript received July 2006; revised manuscript received September 2006.

(2007). This data set contains sequences from 12 loci and 25 exemplar genera (23 seed plants and two outgroups). The genera comprise representatives of angiosperms (*Arabidopsis*, *Chloranthus*, *Drimys*, *Glycine*, *Illicium*, *Magnolia*, *Nicotiana*, *Nymphaea*, *Oryza*, *Piper*, *Pisum*, *Trochodendron*, and *Zea*), cycads (*Cycas* and *Zamia*), Gnetales (*Ephedra*, *Gnetum*, and *Welwitschia*), Pinaceae (*Abies* and *Pinus*), non-Pinaceae conifers (*Araucaria* and *Juniperus*), and *Ginkgo*. *Lycopodium* and *Angiopteris* were included as outgroups. Four of the loci are from the nuclear genome (18S rDNA, 26S rDNA, *PHYP/B*, and *PHYN/A*), five loci are from the plastid genome (*atpB*, *matK*, *psaA*, *psbB*, and *rbcL*), and three loci are from the mitochondrial genome (*atpA*, *coxI*, and *mtSSU*). Burleigh and Mathews (2007) describe details regarding the assembly and alignment of the data set, and the complete accession table and sequence alignment are available in a zip archive in the online edition of the *International Journal of Plant Sciences*, in both Excel and tab-delimited ASCII files; the archive also includes a Nexus file. In the analyses of Burleigh and Mathews (2007), sites in the 12-locus alignment also were partitioned into one of four rate classes using a previously described method (Burleigh and Mathews 2004). Rate class 1 (RC1) sites are the most slowly evolving sites, and in fact, all RC1 sites were constant. Rate class 2 (RC2) and 3 (RC3) sites are estimated to be evolving at intermediate rates, and these were combined into a single partition for simulation (the RC23 sites). The rate class 4 (RC4) sites are estimated to be evolving at the fastest rates.

Burleigh and Mathews (2007) performed MP and ML analyses on all single-locus data sets as well as the combined 12-locus data set. Additionally, the MP and ML analyses were performed on just the RC23 sites and just the RC4 sites of each locus but not on RC1 sites since there is no variation at these sites. Heuristic MP searches were conducted with PAUP*, version 4.0b10 (Swofford 2002), using 10,000 random taxon addition replicates with tree-bisection-reconnection (TBR) branch swapping. All characters were unordered, and gaps were treated as missing data. The MP nonparametric bootstrap analyses (Felsenstein 1985) consisting of 1000 replicates, each with 100 random taxon addition replicates and TBR branch swapping, were performed on each data set. The bootstrap scores are based on sampling a single tree per bootstrap replicate, the first tree saved. All ML analyses used the HKY model (Hasegawa et al. 1985) that allows separate rates for transitions and transversions and uses empirical base frequencies. We chose the HKY model because it is relatively simple and thus computationally tractable yet it incorporates the major substitution bias found in all loci, the unequal rates of transitions and transversions. The ML analyses of data sets with all rate classes also incorporated gamma (Γ) distributed rate variation among sites (Yang 1994), and ML analyses of data sets consisting of only RC23 or RC4 sites assumed equal rates among sites. All ML searches used substitution parameter values estimated from a neighbor-joining (NJ; Saitou and Nei 1987) topology. Each ML bootstrap replicate used TBR branch swapping starting from the NJ tree with a time limit of 6 h.

Assessment of Error Rates with Parametric Bootstrapping

Parametric bootstrapping was used to examine how well MP and ML methods would recover five different seed plant hypotheses inferred in analyses of molecular and/or morphological data (see review in Magallón and Sanderson 2002). The anthophyte (AN) hypothesis places Gnetales in a clade with angiosperms and the extinct Bennettitales and *Pentoxylon* (Doyle 1998*a*). Trees of extant taxa are consistent with the AN hypothesis only if they place angiosperms and Gnetales as sister taxa. The gnepine (GP) hypothesis places Gnetales as sister to Pinaceae, and the gnetifer (GF) hypothesis places Gnetales as sister to all conifers. The Gnetales sister (GS) hypothesis places Gnetales as sister to all other seed plants, and the Gnetales sister gymnosperms (GSG) hypothesis places Gnetales as sister to the other gymnosperms.

In parametric bootstrapping, DNA sequence alignments are simulated based on a specified topology, branch lengths, and substitution parameters. Parametric bootstrapping techniques were used to create pseudoreplicate data sets for each of the 12 single-locus data sets using a topology for each of the five seed plant hypothesis previously listed. To find optimal topologies representing each seed plant hypothesis, we first performed constrained MP searches for each hypothesis using the entire 12-locus data set. In preliminary analyses, we did not find major differences in the topologies of MP and ML constraint topologies. If multiple MP trees were found in the constraint searches, we used the first saved tree. For each locus, we then estimated the branch lengths and substitution parameters for each of the five constraint trees using all sites assuming the general time reversible (GTR) Γ substitution model (Tavaré 1986; Yang 1994). If the locus did not contain all genera from the concatenated data set, these taxa were pruned from the trees. For each locus, the constraint trees and estimated parameters were used to simulate 100 replicate data sets using HYPHY batch files (Kosakovsky Pond et al. 2005). We created 100 replicate 12-locus data sets by concatenating individual locus simulation data sets. Thus, the 12-locus simulation data sets account for different substitution parameters, branch lengths, and even taxon sampling from each locus. To obtain the simulated RC23 and RC4 data sets for each locus, we estimated the optimal rate class assignment for all sites in each concatenated 12-locus data set using the method described by Burleigh and Mathews (2004).

Both MP and ML were used to analyze each of the simulated data sets of all sites, of only the RC23 sites, and of only the RC4 sites from 11 of the 12 loci (outgroup sequences were not available for 26S rDNA when these analyses were begun) as well as the 12-locus concatenated data sets. The MP and ML analyses of the simulated data sets used the same heuristic strategies as those used in the analyses of the original empirical data sets.

Error rates can be estimated through parametric bootstrapping by comparing the topology used to simulate data sets with the topologies inferred from the simulated data sets (Huelsenbeck et al. 1996). The parametric bootstrap estimates the probability of inferring a topology, given that the simulated topology is true. For example, in the seed plant simulations, parametric bootstrapping allows us to estimate the probability of inferring a GS hypothesis, given that the GP hypothesis is true. We would denote this P[infer GS|GP true]. With a complete set of parametric bootstrapping results, it is then possible to use Bayes's rule to calculate the probability that a topology is true, given the topology we inferred. This simply summarizes how the results of the

parametric bootstrapping affect our confidence in the topology that was inferred from the original empirical analysis. So if we inferred a GS hypothesis from our original analyses, we can use the results of the parametric bootstrapping to calculate P[GP true|infer GS] or the posterior probability that the GP is true, given that we inferred GS. Furthermore, we can incorporate the nonparametric bootstrap support to scale the posterior probability based on the percentage of times that a hypothesis was inferred from bootstrap data sets. In our example, if the GS hypothesis received 80% nonparametric bootstrap support and the GP hypothesis received 20% support, the probability that the GP hypothesis is true is P[GP true|80% infer GS; 20% infer GP] $= 0.80 \times P$[GP true|infer GS] $+ 0.20 \times P$[GP true|infer GP]. We call this posterior probability that is scaled by the bootstrap scores the adjusted bootstrap score.

To use Bayes's rule, we first assume that the prior probabilities of each of five hypotheses of seed plant phylogeny are equal (P[AN] $= P$[GP] $= P$[GF] $= P$[GB] $= P$[GBG] $= 0.20$) and that other hypotheses have a prior probability of 0. In other words, in the absence of data, we assume that each of the five seed plant hypotheses is equally likely to be correct. In our example using Bayes's rule P[GP true|infer GS] $= (P$[infer GS|GP true] $\times$ P[GP])/P[data], where P[data] represents the sum of the probabilities of inferring the specified hypothesis, given that each possible hypothesis is true (in this case, P[infer GS|GP true] $\times P$[GP]+ P[infer GS|GF true] $\times$ P[GF] + P[infer GS|AN true] $\times$ P[AN] + P [infer GS|GS true] $\times$ P[GS] + P[infer GS|GSG true] $\times$ P[GSG]). Since the prior probabilities of the five seed plant hypotheses are equal, they cancel out of the calculation of the Bayes's rule formula. Phylogenetic analyses of some single-locus data sets may support trees that are not compatible with any of the five seed plant hypotheses we examine. These trees represent many very unlikely phylogenetic hypotheses, for example, trees in which the angiosperms are not monophyletic. Although there appear to be cases of horizontal transfer of loci among distantly related seed plant clades (Won et al. 2003) and mistakes in GenBank are possible, there is little evidence to attribute the anomalous trees to these causes. In single-locus data sets, the anomalous trees are mostly recovered in the analyses of the small RC23 or RC4 data sets, and they are less frequently recovered in larger combined data sets. These trees also are rarely recovered in phylogenetic analyses of multilocus data sets. Thus, they likely reflect random error and anomalies in the phylogenetic analyses of small data sets rather than horizontal transfer or viable hypotheses of seed plant relationships. We chose to simulate data only on those trees that have been supported in published analyses of molecular or morphological data. We place all of the trees that conflict with all five seed plant hypotheses into a category called "other" trees. The prior probability of an "other" hypothesis is 0%, and therefore the adjusted bootstrap score for an "other" hypothesis will always be 0%.

We detail the full calculation of the adjusted bootstrap with two examples. In the first example, the results of the parametric bootstrap in table 1 show little evidence of bias; the analyses of the simulated data sets recover the hypothesis used to simulate the data 90% of the time. The nonparametric bootstrap support in the original data set is 95% for the GP hypothesis and 5% for an "other" hypothesis (table 1). Therefore, the adjusted bootstrap value that the GP hypothesis is true will be $0.95 \times P$[GP true|infer GP] $+ 0.05 \times P$[GP true|infer other]. We can calculate the P[GP|infer GP] with Bayes's rule as shown (note we have canceled out the prior probabilities since they are equal):

$$P[\text{GP true}|\text{infer GP}] = P[\text{infer GP}|\text{GP true}]/(P[\text{infer GP}|\text{GP true}] + P[\text{infer GP}|\text{GF true}] + P[\text{infer GP}|\text{AN true}] + P[\text{infer GP}|\text{GS true}] + P[\text{infer GP}|\text{GSG true}]).$$

$$P[\text{GP true}|\text{infer GP}] = 0.90/(0.90 + 0.00 + 0.00 + 0.00 + 0.00) = 1.$$

Now we can calculate P[GP|infer other] with the following formula:

$$P[\text{GP true}|\text{infer other}] = P[\text{infer other}|\text{GP true}]/(P[\text{infer other}|\text{GP true}] + P[\text{infer other}|\text{GF true}] + P[\text{infer other}|\text{AN true}] + P[\text{infer other}|\text{GS true}] + P[\text{infer other}|\text{GSG true}]).$$

$$P[\text{GP true}|\text{infer other}] = 0.10/(0.10 + 0.10 + 0.10 + 0.10 + 0.10) = 0.20.$$

So, in table 1, the adjusted bootstrap value that the GP hypothesis is true $= (0.95 \times 1) + (0.05 \times 0.20) = 0.96$. We can use the same method to determine that the adjusted bootstrap values for each of the other four seed plant hypotheses is 0.01 (table 1).

Table 2 shows a case in which parametric bootstrapping results indicate a strong bias in favor of recovering the GP hypothesis. The GP hypothesis is recovered in analyses of 90% of the simulated data sets no matter which seed plant hypothesis is used to simulate the data. In this case, the adjusted bootstrap value for the GP hypothesis is still $0.95 \times P$[GP true|infer GP] $+ 0.05 \times P$[GP true|infer other], and we can calculate P[GP|infer GP] as follows:

$$P[\text{GP true}|\text{infer GP}] = P[\text{infer GP}|\text{GP true}]/(P[\text{infer GP}|\text{GP true}] + P[\text{infer GP}|\text{GF true}] + P[\text{infer GP}|\text{AN true}] + P[\text{infer GP}|\text{GS true}] + P[\text{infer GP}|\text{GSG true}]).$$

$$P[\text{GP true}|\text{infer GP}] = 0.90/(0.90 + 0.90 + 0.90 + 0.90 + 0.90) = 0.20.$$

We calculate P[GP|infer other] as

$$P[\text{GP true}|\text{infer other}] = P[\text{infer other}|\text{GP true}]/(P[\text{infer other}|\text{GP true}] + P[\text{infer other}|\text{GF true}] + P[\text{infer other }|\text{AN true}] + P[\text{infer other}|\text{GS true}] + P[\text{infer other}|\text{GSG true}]).$$

$$\text{P[GP true|infer other]} = 0.10/(0.10 + 0.10 + 0.10 + 0.10 + 0.10) = 0.20.$$

So, in table 2, the adjusted bootstrap for the GP hypothesis = $(0.95 \times 0.20) + (0.05 \times 0.20) = 0.20$. Similar calculations show that the adjusted bootstrap values for each of the other four seed plant hypotheses is also 0.20 (table 2). In this example, we cannot distinguish between the seed plant hypotheses because the GP hypothesis is equally likely to be inferred no matter which seed plant hypothesis is true (table 2). The adjusted bootstrap score accounts for the strong observed bias and indicates that the nonparametric bootstrap score should not be trusted. The adjusted bootstrap values for each hypothesis should be interpreted with respect to the prior probability of 0.20. In this example, the adjusted bootstrap value for each hypothesis is the same as the prior probabilities, indicating that the data cannot help to distinguish among seed plant hypotheses, even though the GP hypothesis receives strong bootstrap support (table 2).

Results

A table with the complete results of the parametric bootstrapping experiments from all single-locus analyses (the number of times each seed plant hypothesis was recovered from each set of simulations) is available as supplemental data. The simulations show evidence of different errors or biases in single-locus analyses. Adjusted bootstrap values for single-locus MP or ML analyses rarely indicate strong support for any seed plant hypothesis (tables 3, 4). In the single-locus MP analyses, there is evidence of a bias favoring the GS hypothesis in the plastid loci (table 3). In nine of 10 cases in which the GS hypothesis receives greater than 50% MP bootstrap support, the adjusted bootstrap is below 50% (table 3). There also is evidence of a bias favoring the GSG hypothesis in MP analyses of all sites and RC23 sites of *PHYN/A*, as determined by the low adjusted bootstrap scores compared with the nonparametric bootstrap scores (table 3). The adjusted bootstrap support for the GP hypothesis generally is lower than MP bootstrap support, though the difference generally is not as extreme as bias toward the GS hypothesis in the plastid loci (table 3). In four of the six cases in which the single-locus MP bootstrap values are above 50%, the adjusted bootstrap also is above 50%. The three hypotheses to receive greater than 50% adjusted bootstrap support in the single-locus MP analyses are the GP hypothesis (three loci, five partitions), the GF hypothesis (three loci, three partitions), and the GS hypothesis (one partition; table 3). Error is less evident in the single-locus ML than in MP analyses (tables 3, 4). In 13 of 16 cases, whenever ML bootstrap support for a hypothesis is greater than 50%, the corresponding adjusted bootstrap support is also more than 50% (table 4). Most of the support from the single-locus ML analyses is for the GP hypothesis. In 10 of 11 cases in which the ML support for the GP hypothesis exceeds 50%, the adjusted bootstrap also is greater than 50%, and in one case (*psbB* RC23), the adjusted bootstrap for the GP hypothesis is greater than 50% even when the nonparametric bootstrap support is not (table 4). In three cases for the GF hypothesis and one case for the GS hypothesis, the adjusted bootstrap from single-locus ML analyses exceeds 50% (table 4). Thus, in the single-locus analyses, bias is more evident in MP than ML analyses, and the strongest apparent biases favor recovering the GS hypothesis.

When the 12-locus data sets are analyzed, there is more evidence of error in MP than ML analyses (tables 5, 6). In the MP analyses, evidence of error is most apparent in the analyses of the all sites data sets, in which there is evidence of bias against recovering the AN and GF hypotheses, and the RC4 data sets, in which there is strong evidence of bias against recovering the AN, GP, and GF hypotheses (table 5). Bias is less evident in the 12-locus MP analyses of the RC23 data sets, in which only the AN hypothesis appears difficult to recover when it is the true hypothesis and the adjusted bootstrap scores are similar to the bootstrap scores (table 5). In all of the MP analyses of 12-locus data sets, the GS and GSG trees are always recovered when they represent the true tree (table 5). In the ML 12-locus simulations, there is no evidence of biases affecting the recovery of the GP, GF, GS, or GSG hypotheses (table 6). All of these hypotheses are recovered in at least 92% of the ML analyses in which they are used to simulate the data (table 6). There is evidence of bias

Table 1

Calculating the Adjusted Bootstrap Values from Parametric Bootstrap Results with No Evidence of Bias

	GP	GF	AN	GS	GSG	BS	ABS
GP	0.90	0.00	0.00	0.00	0.00	0.95	0.96
GF	0.00	0.90	0.00	0.00	0.00	0.00	0.01
AN	0.00	0.00	0.90	0.00	0.00	0.00	0.01
GS	0.00	0.00	0.00	0.90	0.00	0.00	0.01
GSG	0.00	0.00	0.00	0.00	0.90	0.00	0.01
Other	0.10	0.10	0.10	0.10	0.10	0.05	

Note. Five seed plant hypotheses were examined: GP = gnepine; GF = gnetifer; AN = anthophyte; GS = Gnetales sister to seed plants; GSG = Gnetales sister to gymnosperms. Columns GP–GSG represent the seed plant hypothesis that was simulated; the rows represent the percentage of times each seed plant hypothesis was recovered from the simulated data sets. The row "Other" represents the percentage of times a topology that is not consistent with any of the five listed seed plant hypotheses was recovered. For example, when the GP hypothesis was simulated, 90% of the time a GP hypothesis was recovered from the analysis of the simulated data sets, and 10% of the time none of the five hypotheses was recovered. BS is the nonparametric bootstrap score obtained from the empirical data set, and ABS is the adjusted bootstrap.

Table 2

Calculating the Adjusted Bootstrap Values from Parametric Bootstrap Results with Evidence of Strong Bias

	GP	GF	AN	GS	GSG	BS	ABS
GP	0.90	0.90	0.90	0.90	0.90	0.95	0.20
GF	0.00	0.00	0.00	0.00	0.00	0.00	0.20
AN	0.00	0.00	0.00	0.00	0.00	0.00	0.20
GS	0.00	0.00	0.00	0.00	0.00	0.00	0.20
GSG	0.00	0.00	0.00	0.00	0.00	0.00	0.20
Other	0.10	0.10	0.10	0.10	0.10	0.05	

Note. Five seed plant hypotheses were examined: GP = gnepine; GF = gnetifer; AN = anthophyte; GS = Gnetales sister to seed plants; GSG = Gnetales sister to gymnosperms. Columns GP–GSG represent the seed plant hypothesis that was simulated; the rows represent the percentage of times each seed plant hypothesis was recovered from the simulated data sets. The row "Other" represents the percentage of times a topology that is not consistent with any of the five listed seed plant hypotheses was recovered. For example, when the GP hypothesis was simulated, 90% of the time a GP hypothesis was recovered from the analysis of the simulated data sets, and 10% of the time none of the five hypotheses was recovered. BS is the nonparametric bootstrap score obtained from the empirical data set, and ABS is the adjusted bootstrap. Evidence of bias favors recovering the GP hypothesis.

in 12-locus ML analyses against recovering the AN hypothesis. In the 12-locus ML simulations of all sites and only the RC23 sites, the AN hypothesis is recovered 74% and 82% of the times it is simulated, respectively. There is stronger evidence of biases against the AN hypothesis the 12-locus ML analyses of just the RC4 sites, in which the AN hypothesis is recovered only 26% of the times it is used to simulate the data (table 6). In both the MP and ML analyses of the 12-locus data sets, only the GP hypothesis ever has an adjusted bootstrap score above 72% (99% in MP and ML analyses of RC23 data set; 100% in ML analyses of all sites), and in cases where the GS hypothesis receives notable bootstrap support, its adjusted bootstrap scores are always lower (tables 5, 6).

Discussion

The results of the parametric bootstrapping experiments demonstrate the potential for systematic error in both MP and ML analyses of seed plant data sets. There is extensive evidence of bias favoring the GS hypothesis in MP analyses, especially of faster-evolving sites, and both MP and ML analyses demonstrate a bias against recovering the AN hypothesis (tables 3–6). With the exception of the potential bias against recovering the AN hypothesis, ML analyses of the 12-locus data set show little evidence of error associated with heterogeneous processes of evolution among loci. There is greater evidence of bias in MP analyses of the 12-locus data set. The results emphasize that analyses of concatenated data sets with many characters are susceptible to error and must be interpreted with caution (Phillips et al. 2004; Soltis et al. 2004; Stefanovic et al. 2004).

Biases in Single-Locus Analyses

Biases observed in the MP analyses of single-locus simulations appear to explain some of the variation in the phylogenetic signal among loci and rate class partitions observed previously in phylogenetic analyses of seed plant data (Burleigh and Mathews 2007). For example, the MP bootstrap analysis for 18S rDNA supports the GF hypothesis much more strongly than other loci (Burleigh and Mathews 2007), and there is evidence for a bias favoring recovery of the GF hypothesis when it is not the true (simulated) hypothesis in 18S rDNA data (table 3). The same situation occurs for *PHYN/A* and the GSG hypothesis (Burleigh and Mathews 2007; table 3). Most strikingly, MP analyses of plastid loci strongly support the GS hypothesis compared with the other loci (Burleigh and Mathews 2007), and the simulations show a strong bias favoring the recovery of the GS hypothesis in analyses of plastid loci, even if the GS hypothesis is not true (table 3). In an extreme case, MP analyses of the simulated *matK* data set always recover the GS hypothesis, no matter which phylogenetic hypothesis is used for the simulations (see supplemental data). Bias in MP analyses favoring the GS hypothesis is overall most evident in analyses of the RC4 sites (table 3). This result extends the observations of bias favoring the GS hypothesis in the fast-evolving third codon position of *psaA* from Sanderson et al. (2000). It is not obvious why MP analyses of the plastid loci seem especially susceptible to bias, though it may be due to the relative branch lengths associated with plastid loci. The plastid RC23 sites have more biased empirical nucleotide frequencies (GC content = 32.4%) than the RC4 sites (GC content = 42.6%), and thus, the phylogenetic bias likely is not due to a limited character-state space.

Compared with the MP analyses, there is little evidence of bias in the single-locus ML analyses (table 4). Some ML analyses rarely recover the hypothesis used to simulate the data, but there also is rarely much support for an alternate topology (table 4; supplemental data). Since we grouped every topology that is not consistent with the five chosen seed plant hypotheses in the category "other," high "other BS" values do not necessarily mean that any specific alternate hypothesis is well supported. The results of the ML analyses of single-locus simulated data sets are consistent with a lack of power to resolve the phylogeny and do not necessarily indicate a bias favoring recovery of an erroneous phylogeny. The

Table 3

Nonparametric Bootstrap Values and Adjusted Bootstrap Values from Maximum Parsimony Analyses of 11 Loci for Each of the Five Seed Plant Hypotheses Examined

	AN BS	AN ABS	GP BS	GP ABS	GF BS	GF ABS	GS BS	GS ABS	GSG BS	GSG ABS	Other BS
18S all	0.08	0.09	0.28	0.29	0.58	0.45	0.02	0.08	0.00	0.08	0.04
18S RC23	0.00	0.04	0.00	0.09	0.00	0.02	0.00	0.38	0.00	0.47	1.00
18S RC4	0.04	0.13	0.29	0.33	0.57	0.34	0.01	0.07	0.00	0.14	0.09
PHYP/B all	0.04	0.22	0.02	0.25	0.00	0.16	0.00	0.16	0.01	0.20	0.93
PHYP/B RC23	0.02	0.02	0.47	0.55	0.01	0.40	0.00	0.01	0.01	0.03	0.49
PHYP/B RC4	0.00	0.08	0.00	0.19	0.00	0.37	0.00	0.26	0.00	0.10	1.00
PHYN/A all	0.00	0.17	0.00	0.21	0.00	0.16	0.12	0.32	0.62	0.15	0.26
PHYN/A RC23	0.00	0.34	0.02	0.01	0.02	0.11	0.06	0.08	0.85	0.45	0.05
PHYN/A RC4	0.02	0.20	0.00	0.09	0.00	0.09	0.05	0.37	0.03	0.25	0.91
atpB all	0.02	0.24	0.02	0.08	0.00	0.14	0.71	0.24	0.15	0.30	0.10
atpB RC23	0.27	0.21	0.00	0.20	0.00	0.52	0.01	0.05	0.00	0.01	0.72
atpB RC4	0.02	0.18	0.04	0.19	0.00	0.21	0.53	0.19	0.23	0.23	0.17
matK all	0.00	0.20	0.00	0.20	0.00	0.20	0.97	0.20	0.03	0.20	0.00
matK RC23	0.00	0.29	0.00	0.03	0.00	0.16	0.95	0.31	0.03	0.21	0.02
matK RC4	0.12	0.16	0.00	0.16	0.00	0.18	0.60	0.16	0.11	0.33	0.17
psaA all	0.00	0.23	0.12	0.08	0.00	0.10	0.74	0.42	0.13	0.20	0.00
psaA RC23	0.00	0.19	0.00	0.00	0.00	0.02	0.98	0.62	0.00	0.17	0.02
psaA RC4	0.01	0.17	0.22	0.12	0.00	0.12	0.43	0.18	0.22	0.41	0.13
psbB all	0.09	0.30	0.00	0.07	0.00	0.11	0.79	0.35	0.12	0.17	0.01
psbB RC23	0.00	0.04	0.59	0.45	0.15	0.37	0.22	0.14	0.00	0.05	0.04
psbB RC4	0.10	0.23	0.00	0.19	0.00	0.14	0.71	0.22	0.17	0.22	0.02
rbcL all	0.07	0.17	0.00	0.09	0.00	0.17	0.60	0.44	0.11	0.13	0.22
rbcL RC23	0.15	0.21	0.00	0.00	0.00	0.52	0.32	0.21	0.00	0.07	0.53
rbcL RC4	0.10	0.23	0.00	0.17	0.00	0.14	0.37	0.24	0.11	0.22	0.42
atpA all	0.00	0.01	0.96	0.86	0.00	0.09	0.01	0.01	0.00	0.04	0.03
atpA RC23	0.00	0.03	0.00	0.00	0.00	0.69	0.00	0.00	0.00	0.28	1.00
atpA RC4	0.00	0.04	0.87	0.62	0.00	0.19	0.06	0.06	0.00	0.11	0.06
cox1 all	0.00	0.18	0.00	0.19	0.06	0.14	0.01	0.19	0.01	0.30	0.93
cox1 RC23	0.00	0.13	0.00	0.16	0.44	0.32	0.00	0.18	0.00	0.22	0.56
cox1 RC4	0.00	0.20	0.00	0.20	0.00	0.19	0.00	0.20	0.00	0.20	1.00
mtSSU all	0.00	0.05	1.00	0.62	0.00	0.27	0.00	0.01	0.00	0.05	0.00
mtSSU RC23	0.00	0.06	0.72	0.60	0.00	0.10	0.00	0.07	0.00	0.17	0.28
mtSSU RC4	0.00	0.10	0.99	0.37	0.00	0.28	0.00	0.14	0.00	0.17	0.01

Note. GP = gnepine; GF = gnetifer; AN = anthophyte; GS = Gnetales sister to seed plants; GSG = Gnetales sister to gymnosperms. "Other" specifies the bootstrap support for topologies that are not consistent with any of these hypotheses. Columns list the nonparametric bootstrap (BS) or adjusted bootstrap (ABS) scores from each hypothesis. The rows represent maximum parsimony (MP) analyses of each locus data set. RC23 indicates an analysis that only used the rate class 2 and rate class 3 sites from the specified locus, and RC4 indicates an analysis that only used the rate class 4 sites. Underlined values are >50%.

lack of evidence of bias in single-locus ML simulations is likely due to the similarity of the model used to simulate the data with the model used to analyze the simulated data sets. In cases in which the model used to infer the phylogeny is identical to the true process generating the sequences, ML analyses using reversible nucleotide models will be consistent; that is, they converge toward the correct result as more data are added (Rogers 2001). Furthermore, in simulation, ML analyses often appear to behave in a consistent manner when the nucleotide model used to analyze the data is similar to the model used to simulate the data (Sullivan and Swofford 2001). The GTR model used to simulate the single-locus simulation data sets is a generalized form of the HKY model used to analyze the data. Thus, the ML analyses of the single-locus data sets should have little systematic error associated with model misspecification. The apparent difficulty of recovering the simulated seed plant hypothesis from single-locus simulated data sets (table 4) demonstrates the strength of the random error, which may necessitate combining loci to infer seed plant phylogeny.

Biases in 12-Locus Simulated Data Sets

Combining loci to make longer alignments may reduce sampling error and increase the power of a phylogenetic analysis, but it is unclear whether analyses of concatenated data sets with many characters remain susceptible to the biases observed in the single-locus analyses. On the one hand, combining a large collection of unlinked loci with different patterns of evolution might allow the true phylogenetic signal to emerge above the divergent sets of biases observed in the single-locus analyses (Rokas et al. 2003). On the other hand, in some cases, heterogeneity in the processes of evolution or phylogenetic signal among loci can complicate phylogenetic analyses, especially when the analyses assume a homogeneous process of evolution (Wilgenbush and de Queiroz 2000; Brandley et al. 2005). The analyses

Table 4

Nonparametric Bootstrap Values and Adjusted Bootstrap Values from Maximum Likelihood Analyses of 11 Loci for Each of the Five Seed Plant Hypotheses Examined

	AN BS	AN ABS	GP BS	GP ABS	GF BS	GF ABS	GS BS	GS ABS	GSG BS	GSG ABS	Other BS
18S all	0.04	0.07	0.11	0.12	0.84	0.53	0.00	0.14	0.00	0.14	0.01
18S RC23	0.01	0.01	0.19	0.19	0.77	0.77	0.01	0.02	0.00	0.01	0.02
18S RC4	0.01	0.09	0.19	0.25	0.77	0.45	0.01	0.06	0.00	0.16	0.02
PHYP/B all	0.09	0.13	0.53	0.50	0.05	0.06	0.00	0.06	0.00	0.16	0.33
PHYP/B RC23	0.04	0.04	0.83	0.82	0.03	0.08	0.00	0.05	0.01	0.01	0.09
PHYP/B RC4	0.00	0.03	0.01	0.18	0.00	0.68	0.00	0.08	0.00	0.03	0.99
PHYN/A all	0.00	0.09	0.33	0.35	0.21	0.22	0.15	0.14	0.24	0.21	0.07
PHYN/A RC23	0.03	0.14	0.07	0.07	0.01	0.02	0.26	0.29	0.53	0.48	0.10
PHYN/A RC4	0.01	0.01	0.21	0.44	0.05	0.55	0.00	0.00	0.00	0.00	0.73
atpB all	0.22	0.21	0.05	0.13	0.18	0.22	0.02	0.40	0.01	0.03	0.52
atpB RC23	0.17	0.20	0.00	0.19	0.00	0.29	0.00	0.29	0.00	0.03	0.83
atpB RC4	0.18	0.16	0.32	0.32	0.02	0.30	0.01	0.13	0.03	0.15	0.44
matK all	0.00	0.18	0.26	0.32	0.13	0.19	0.33	0.22	0.04	0.10	0.24
matK RC23	0.00	0.15	0.12	0.19	0.28	0.43	0.09	0.11	0.01	0.13	0.50
matK RC4	0.00	0.21	0.00	0.13	0.00	0.20	0.06	0.29	0.00	0.17	0.94
psaA all	0.00	0.03	0.79	0.65	0.00	0.20	0.16	0.12	0.00	0.02	0.05
psaA RC23	0.00	0.17	0.00	0.00	0.00	0.08	0.77	0.61	0.00	0.13	0.23
psaA RC4	0.00	0.07	0.71	0.51	0.00	0.37	0.08	0.07	0.00	0.07	0.21
psbB all	0.00	0.04	0.58	0.68	0.07	0.19	0.03	0.04	0.00	0.04	0.32
psbB RC23	0.00	0.06	0.48	0.51	0.27	0.29	0.02	0.08	0.00	0.06	0.23
psbB RC4	0.00	0.19	0.13	0.20	0.09	0.23	0.19	0.18	0.00	0.21	0.59
rbcL all	0.02	0.06	0.29	0.32	0.01	0.43	0.14	0.17	0.01	0.04	0.53
rbcL RC23	0.21	0.34	0.00	0.27	0.00	0.30	0.00	0.01	0.00	0.08	0.79
rbcL RC4	0.00	0.09	0.50	0.38	0.04	0.57	0.03	0.07	0.01	0.08	0.42
atpA all	0.00	0.00	1.00	0.70	0.00	0.18	0.00	0.05	0.00	0.10	0.00
atpA RC23	0.00	0.02	0.88	0.74	0.00	0.19	0.00	0.04	0.00	0.03	0.12
atpA RC4	0.00	0.05	1.00	0.58	0.00	0.26	0.00	0.07	0.00	0.09	0.00
cox1 all	0.00	0.18	0.00	0.16	0.18	0.15	0.00	0.15	0.00	0.36	0.82
cox1 RC23	0.00	0.15	0.00	0.18	0.45	0.35	0.00	0.16	0.00	0.15	0.55
cox1 RC4	0.00	0.19	0.00	0.22	0.00	0.17	0.00	0.21	0.00	0.21	1.00
mtSSU all	0.00	0.03	0.90	0.76	0.00	0.19	0.00	0.01	0.00	0.03	0.10
mtSSU RC23	0.00	0.09	0.61	0.54	0.00	0.17	0.00	0.08	0.00	0.12	0.39
mtSSU RC4	0.00	0.18	0.30	0.25	0.00	0.21	0.00	0.17	0.00	0.19	0.70

Note. GP = gnepine; GF = gnetifer; AN = anthophyte; GS = Gnetales sister to seed plants; GSG = Gnetales sister to gymnosperms "other" specifies the bootstrap support for topologies that are not consistent with any of these hypotheses. Columns contain the nonparametric bootstrap (BS) or adjusted bootstrap (ABS) scores from each hypothesis. The rows represent maximum likelihood analyses of each locus data set. RC23 indicates an analysis that only used the rate class 2 and rate class 3 sites from the specified locus, and RC4 indicates an analysis that only used the rate class 4 sites. Underlined values are >50%.

of the 12-locus simulated data sets indicate systematic error in MP and, to a lesser degree, ML analyses, demonstrating that combining heterogeneous loci does not eliminate the potential effects of systematic error on the inference of seed plant phylogeny.

Model misspecification can cause systematic error (e.g., Swofford et al. 2001). Though the data representing each locus in the 12-locus simulated data sets were generated based on different branch lengths and substitution parameters, the ML models used to infer the phylogeny assume homogeneous branch lengths and substitution patterns across all sites. Therefore, unlike in the single-locus simulations, model misspecification may lead to systematic error in the ML analyses of the simulated 12-locus data sets. In fact, under some simulation conditions when different sites have very different sets of branch lengths, MP performs better than ML or Bayesian analyses that assume homogeneous branch lengths across sites (Kolaczkowski and Thornton 2004; Simmons et al. 2006; but see Gadagkar and Kumar 2005; Gaucher and Miyamoto 2005; Philippe et al. 2005; Spencer et al. 2005). However, in the analyses of the 12-locus simulated data sets, not only do the ML analyses appear to outperform MP but also, with the exception of a bias against recovering the AN hypothesis, they generally are robust to the heterogeneous branch lengths and substitution parameters among seed plant loci (tables 5, 6).

When using MP, evidence of bias is strongest in analyses of the fast-evolving RC4 sites, with relatively little evidence of bias in analyses of the RC23 data sets (table 5). This is consistent with the idea that removing the fast-evolving sites will reduce the effects of long-branch attraction (e.g., Brinkman and Philippe 1999; Philippe et al. 2000; Pisani 2004). However, simply eliminating all RC4 or fast-evolving sites may not be the most effective use of the seed plant data. There is evidence of a heterogeneous phylogenetic signal in the RC4 sites among loci (Burleigh and Mathews 2007), and some of these sites may be phylogenetically informative and not subject to biases in MP analyses. Furthermore, increasing taxon

Table 5

Results from the Maximum Parsimony 12-Locus Data Set Simulations

	AN	GP	GF	GS	GSG	BS	ABS
All:							
AN	0.00	0.00	0.00	0.00	0.00	0.00	0.33
GP	0.00	1.00	0.00	0.00	0.00	0.00	0.00
GF	0.00	0.00	0.19	0.00	0.00	0.00	0.14
GS	0.84	0.00	0.00	1.00	0.00	0.66	0.36
GSG	0.16	0.00	0.81	0.00	1.00	0.34	0.17
Other	0.00	0.00	0.00	0.00	0.00	0.00	
RC23:							
AN	0.11	0.00	0.00	0.00	0.00	0.00	0.00
GP	0.00	1.00	0.00	0.00	0.00	1.00	0.99
GF	0.00	0.00	1.00	0.00	0.00	0.00	0.01
GS	0.79	0.00	0.00	1.00	0.00	0.00	0.00
GSG	0.10	0.00	0.00	0.00	1.00	0.00	0.00
Other	0.00	0.00	0.00	0.00	0.00	0.00	
RC4:							
AN	0.00	0.00	0.00	0.00	0.00	0.00	0.01
GP	1.00	0.09	0.00	0.00	0.00	0.00	0.07
GF	0.00	0.00	0.00	0.00	0.00	0.00	0.15
GS	0.00	0.02	0.12	1.00	0.00	0.82	0.72
GSG	0.00	0.89	0.88	0.00	1.00	0.15	0.05
Other	0.00	0.00	0.00	0.00	0.00	0.03	

Note. Data were simulated according to five different seed plant hypotheses (AN = anthophyte; GP = gnepine; GF = gnetifer; GS = Gnetales sister to seed plants; GSG = Gnetales sister to gymnosperms). Columns AN–GSG contain the percentage of simulation replicates in which the seed plant hypothesis listed in each row was inferred from the simulations. The row labeled "Other" lists the percentage of times that a topology not consistent with any of the five hypotheses was inferred. The column titled "BS" contains the nonparametric bootstrap scores obtained from maximum likelihood analysis of the empirical 12-locus data set, and the column titled "ABS" contains the adjusted bootstrap scores for each seed plant hypothesis. The 12-locus maximum parsimony simulations were done using all sites (All), only the RC23 sites, and only the RC4 sites.

sampling or using different approaches to MP analyses, including different weighting schemes for characters or types of substitutions, may ameliorate or possibly eliminate some of the observed biases. Finally, analyses of slowly evolving sites also can be susceptible to error (e.g., Stanger-Hall and Cunningham 1998), and this error may not be detected from our parametric bootstrapping because of a lack of power from small data sets and few substitutions or complex processes of evolution that are not incorporated into the simulations.

In MP analyses of the simulated 12-locus RC23 sites and in ML analyses of all 12-locus data sets, the only notable bias is against recovering the AN hypothesis (tables 5, 6). This bias may help explain the dearth of support for the AN hypothesis from molecular phylogenetic analyses (but see Stefanovic et al. 1998; Rydin et al. 2002). Again, this result extends a similar finding of a bias in MP analyses of *psaA* and *psbB* by Sanderson et al. (2000). It is difficult to reconcile the general bias against recovering the AN hypothesis with the numerous phylogenetic analyses of morphological characters that have favored the AN hypothesis (Parenti 1980; Crane 1985; Doyle and Donoghue 1986; Rothwell and Serbet 1994; Doyle 1996, 1998*b*, 2006; Hilton and Bateman 2006). A recent analysis of a revised morphological data set found that MP analyses still favor the AN hypothesis, but trees that place Gnetales within conifers, though not sister to Pinaceae, are only a single step longer (Doyle 2006). Since the AN hypothesis is often recovered from phylogenetic analyses of morphological data sets, it appears that the morphological data do not suffer from the bias against recovering the AN hypothesis that we observe in the parametric bootstrapping of molecular data. The support for the AN hypothesis in some morphological analyses may be partly due to errors in character coding or mistaken assignments of homology (Donoghue and Doyle 2000; Doyle 2006), and the morphological characters likely have very different patterns of rate variation than the molecular characters. However, the morphological analyses also often incorporate data from important extinct lineages, and the differences in taxon sampling may help explain the differences in results of analyses of morphological and molecular characters. Since molecular data are unavailable for the extinct lineages, taxon sampling for any seed plant analysis using only molecular data is extremely limited.

Implications for the Inference of Seed Plant Phylogeny

Although the results of the simulation experiments raise many questions regarding the accuracy of seed plant phylogenetic analyses, they also provide some insights into seed plant phylogeny. Foremost, the simulation results provide grounds for questioning the legitimacy of the GS hypothesis. Support for the GS hypothesis generally comes from MP analyses of molecular data (see Magallón and Sanderson 2002; Soltis et al. 2002; Burleigh and Mathews 2004), and this study finds much evidence of bias in MP analyses of single-locus data sets and the 12-locus data set that result in recovering the GS hypothesis even when it is untrue. The presence of

Table 6

Maximum Likelihood 12-Locus Simulations

	AN	GP	GF	GS	GSG	BS	ABS
All:							
AN	0.74	0.00	0.00	0.00	0.00	0.00	0.00
GP	0.00	1.00	0.00	0.00	0.00	1.00	1.00
GF	0.00	0.00	1.00	0.00	0.00	0.00	0.00
GS	0.19	0.00	0.00	1.00	0.00	0.00	0.00
GSG	0.07	0.00	0.00	0.00	1.00	0.00	0.00
Other	0.00	0.00	0.00	0.00	0.00	0.00	
RC23:							
AN	0.82	0.00	0.00	0.00	0.00	0.00	0.00
GP	0.00	1.00	0.00	0.00	0.00	0.99	0.99
GF	0.00	0.00	1.00	0.00	0.00	0.01	0.01
GS	0.06	0.00	0.00	0.92	0.00	0.00	0.00
GSG	0.04	0.00	0.00	0.00	1.00	0.00	0.00
Other	0.08	0.00	0.00	0.08	0.00	0.00	0.00
RC4:							
AN	0.26	0.00	0.00	0.00	0.01	0.00	0.28
GP	0.00	1.00	0.03	0.00	0.00	0.09	0.09
GF	0.00	0.00	0.95	0.00	0.00	0.02	0.02
GS	0.48	0.00	0.00	0.98	0.07	0.89	0.57
GSG	0.26	0.00	0.00	0.02	0.92	0.00	0.04
Other	0.00	0.00	0.02	0.00	0.00	0.00	

Note. Data were simulated according to five different seed plant hypotheses (AN = anthophyte; GP = gnepine; GF = gnetifer; GS = Gnetales sister to seed plants; GSG = Gnetales sister to gymnosperms). Columns AN–GSG contain the percentage of simulation replicates in which the seed plant hypothesis listed in each row was inferred from the simulations. The row labeled "Other" lists the percentage of times that a topology not consistent with any of the five major seed plant hypotheses was inferred. The column titled "BS" contains the nonparametric bootstrap scores obtained from maximum likelihood analysis of the empirical 12-locus data set, and the column titled "ABS" contains the adjusted bootstrap scores for each seed plant hypothesis. The 12-locus maximum parsimony simulations were done using all sites (All), only the RC23 sites, and only the RC4 sites.

such a bias represents systematic error, but it does not necessarily mean that the favored hypothesis is not true. However, further evidence leads us to doubt the GS hypothesis. First, there appears to be much less evidence of bias in MP analyses of the slower-evolving RC23 sites, which generally support a GP hypothesis. Also, it is difficult to reconcile the GS hypothesis with morphological and stratigraphic data (e.g., Doyle 1998*a*; Donoghue and Doyle 2000; Burleigh and Mathews 2004), which further suggests that the results of the MP analyses of the fast-evolving sites may be erroneous. Thus, it appears likely that observed biases favoring the GS hypothesis are misleading some MP analyses.

Second, though there is evidence of bias against recovering the AN hypothesis, a closer look at the results from parametric bootstrapping suggests that the AN hypothesis is not correct. In other words, simulations based on an AN topology reveal a potential bias against recovering the AN hypothesis, but this bias is not consistent with the results from the MP and ML analyses of the original empirical data set. For example, the MP analyses of the 12-locus simulated data sets suggest that, were the AN hypothesis true, the GP hypothesis would be recovered by the fast-evolving sites, and the GS hypothesis would be recovered by the slowly evolving sites (table 5). Yet we observe the opposite trend in the MP nonparametric bootstrap analyses (table 5; Burleigh and Mathews 2007). Furthermore, the adjusted bootstrap score quantifies how much the results of the parametric bootstrapping should affect our interpretation of the seed plant phylogeny, and the adjusted bootstrap values for the AN hypothesis are never greater than 33% (tables 5, 6).

Taken together, the results of the parametric bootstrap suggest that the GP hypothesis is the best-corroborated seed plant hypothesis. However, the results also illustrate limitations of the parametric bootstrapping approach that may lessen our confidence in the GP hypothesis. Since the simulated data sets are generated using an assumed model of evolution and estimated topology, branch lengths, and substitution parameters, the accuracy and effectiveness of parametric bootstrapping analyses depends on how closely the simulation conditions reflect the true patterns of evolution (Buckley 2002). In contrast, by sampling the original data set with replacement, nonparametric bootstrapping is essentially creating pseudoreplicate data sets by sampling from the actual distribution of the data. The variability observed in the parametric and nonparametric bootstrap values would be similar if the parametric bootstrap were simulated using the true model of evolution and true tree (Felsenstein 2004). In our analyses, it is impossible to reconcile the results of the parametric bootstrapping for any hypothesis with the results of the nonparametric bootstrap analyses. For example, ML analyses of all sites and just the RC23 sites from all 12 loci strongly support the GP hypothesis, and ML analyses of just the RC4 sites supports the GS hypothesis (Burleigh and Mathews 2007; table 6). Yet this pattern of support is not consistent with any seed plant hypothesis in the ML analyses of the simulated data sets (table 6). The same discrepancy

between the nonparametric and parametric bootstrap results occurs in the 12-locus MP analyses (table 5). The difficulty of reconciling parametric and nonparametric bootstrapping results suggests that the simulations do not accurately reflect processes of evolution, whether in their assumptions of the topology, branch lengths, the model of nucleotide evolution, or some combination of these. Perhaps no simulation will ever accurately reflect the true complexities of molecular evolution, and thus, the estimates of systematic bias through parametric bootstrapping may be conservative. In other words, a lack of evidence of bias through parametric bootstrapping should not be interpreted as an absence of bias.

To further understand potential systematic errors in seed plant analyses and assess and perhaps enhance the performance of MP and ML analyses, it will be important to identify the major factors that influence molecular evolution that are not yet incorporated into the simulation models used in this study. For example, there is evidence of covarion evolution, in which the rate of evolution at a site changes through time, in nearly half of the plastid loci (Ané et al. 2005). This is particularly notable because covarion patterns of evolution appear much more frequently in the slowly evolving first and second codon position sites than in the faster-evolving third codon position sites (Ané et al. 2005). Thus, we might expect effects of covarion patterns of evolution to influence the slowly evolving sites, where we observed relatively little evidence of bias in these simulations (tables 3–6). Similarly, codon models appear to provide a better fit to the data than nucleotide models in some plant genes (J. G. Burleigh, unpublished manuscript). In some simulation examples, ML using the HKY Γ model, which was used in the seed plant analyses of Burleigh and Mathews (2004, 2007) as well as in the parametric bootstrapping experiments, performs in an inconsistent manner when sequences are evolving under a codon model of evolution (J. G. Burleigh, unpublished manuscript).

Conclusions

Burleigh and Mathews (2007) showed that the 12-locus seed plant data set is susceptible to high sampling variance associated with locus-specific evolution. This study demonstrates that the 12-locus data set is susceptible to systematic error in both MP and ML analyses and that combining loci does not eliminate evidence of error. Together these studies provide reasons for taking a cautious view of results from published analyses of seed plant data and, more broadly, from analyses of multilocus phylogenetic analyses. Despite very large data sets and numerous seed plant trees with 100% bootstrap values (Sanderson and Magallón 2002; Burleigh and Mathews 2004, 2007), it is far from clear that we have obtained an accurate estimate of seed plant phylogeny. These studies demonstrate the possibility of being easily misled by seemingly unambiguous phylogenetic results. Resolving the phylogeny of seed plants will require more than adding new data; it will require a better understanding of the complexities within data sets along with analytical methods that incorporate these insights.

Acknowledgments

This study benefited from discussions with Cecile Ané and Mike Sanderson, and we are grateful for the comments of Mark Simmons and an anonymous reviewer. Taum Hanlon provided critical computer assistance. This study was funded by National Science Foundation grant 0431154.

Literature Cited

Ané C, JG Burleigh, MM McMahon, MJ Sanderson 2005 Covarion structure in plastid genome evolution: a new statistical test. Mol Biol Evol 22:914–924.

Bowe LM, G Coat, CW de Pamphilis 2000 Phylogeny of seed plants based on all three genomic compartments: extant gymnosperms are monophyletic and Gnetales' closest relatives are conifers. Proc Natl Acad Sci USA 97:4092–4097.

Brandley MC, A Schmitz, TW Reeder 2005 Partitioned Bayesian analyses, partition choice, and the phylogenetic relationships of scincid lizards. Syst Biol 54:373–390.

Brinkmann H, H Philippe 1999 Archaea sister group of bacteria? indications from tree reconstruction artifacts in ancient phylogenies. Mol Biol Evol 16:817–825.

Buckley TR 2002 Model misspecification and probabilistic tests of topology: evidence from empirical data sets. Syst Biol 51:509–523.

Burleigh JG, S Mathews 2004 Phylogenetic signal in nucleotide data from seed plants: implications for resolving the seed plant tree of life. Am J Bot 91:1599–1613.

——— 2007 Assessing among-locus variation in the inference of seed plant phylogeny. Int J Plant Sci 168:111–124.

Chaw S-M, CL Parkinson, Y Cheng, TM Vincent, JD Palmer 2000 Seed plant phylogeny inferred from all three plant genomes: monophyly of extant gymnosperms and the origin of Gnetales from conifers. Proc Natl Acad Sci USA 97:4086–4091.

Crane PR 1985 Phylogenetic analysis of seed plants and the origin of angiosperms. Ann Mo Bot Gard 72:716–793.

Donoghue MJ, JA Doyle 2000 Seed plant phylogeny: demise of the anthophyte hypothesis? Curr Biol 10:R106–R109.

Doyle JA 1996 Seed plant phylogeny and the relationships of Gnetales. Int J Plant Sci 157(suppl):S3–S39.

——— 1998*a* Molecules, morphology, fossils, and the relationship of angiosperms and Gnetales. Mol Phylogenet Evol 9:448–462.

——— 1998*b* Phylogeny of vascular plants. Annu Rev Ecol Syst 9:365–392.

——— 2006 Seed ferns and the origin of angiosperms. J Torrey Bot Soc 133:169–209.

Doyle JA, MJ Donoghue 1986 Seed plant phylogeny and the origin of angiosperms: an experimental cladistic approach. Bot Rev 52:321–431.

Felsenstein J 1985 Confidence limits on phylogenies: an approach using the bootstrap. Evolution 38:16–24.

——— 2004 Inferring phylogenies. Sinauer, Sunderland, MA.

Frohlich MW, Parker DS 2000 The mostly male theory of flower evolutionary origins: from genes to fossils. Syst Bot 25:155–170.

Gadagkar SR, S Kumar 2005 Maximum likelihood outperforms maximum parsimony even when evolutionary rates are heterotachous. Mol Biol Evol 22:2139–2141.

Gaucher EA, MM Miyamoto 2005 A call for likelihood phylogenetics even when the process of sequence is heterogeneous. Mol Phylogenet Evol 37:928–931.

Hasegawa M, H Kishino, T Yano 1985 Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. J Mol Evol 22:160–174.

Hilton J, RM Bateman 2006 Pteridosperms are the backbone of seed-plant phylogeny. J Torrey Bot Soc 133:119–168.

Huelsenbeck JP, DM Hillis, R Jones 1996 Parametric bootstrapping in molecular phylogenetics: applications and performance. Pages 19–45 *in* JD Ferraris, SR Palumbi, eds. Molecular zoology: advances, strategies, and protocols. Wiley-Liss, New York.

Kolaczkowski B, JW Thornton 2004 Performance of maximum parsimony and likelihood phylogenetics when evolution is heterogeneous. Nature 431:980–984.

Kosakovsky Pond SL, DW Frost, SV Muse 2005 HyPhy: hypothesis testing using phylogenies. Bioinformatics 21:676–679.

Magallón S, MJ Sanderson 2002 Relationships among seed plants inferred from highly conserved genes: sorting conflicting phylogenetic signals among ancient lineages. Am J Bot 89:1991–2006.

Parenti LR 1980 A phylogenetic analysis of the land plants. Biol J Linn Soc 13:225–242.

Philippe H, P Lopez, H Brinkmann, K Budin, A Germot, J Laurent, D Moreira, M Müller, H Le Guyader 2000 Early-branching or fast evolving eukaryotes? an answer based on slowly evolving positions. Proc Biol Sci 267:1213–1221.

Philippe H, Y Zhou, H Brinkmann, N Rodrigue, F Delsuc 2005 Heterotachy and long-branch attraction in phylogenetics. BMC Evol Biol 5:50.

Phillips MJ, F Delsuc, D Penny 2004 Genome-scale phylogeny and the detection of systematic biases. Mol Biol Evol 21:1455–1458.

Pisani D 2004 Identifying and removing fast-evolving sites using compatibility analysis: an example from the arthropods. Syst Biol 53:978–989.

Rai HS, HE O'Brian, PA Reeves, RG Olmstead, SW Graham 2003 Inference of higher-order relationships in the cycads from a large chloroplast data set. Mol Phylogenet Evol 29:350–359.

Rogers JS 2001 Maximum likelihood estimation of phylogenetic trees is consistent when substitution rates vary according to the invariable sites plus gamma distribution. Syst Biol 50:713–722.

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

Rothwell GR, R Sebert 1994 Lignophyte phylogeny and the evolution of spermatophytes: a numerical cladistic analysis. Syst Bot 19:443–482.

Rydin C, M Källersjö, EM Friis 2002 Seed plant relationships and the systematic position of Gnetales based on nuclear and chloroplast DNA: conflicting data, rooting problems, and the monophyly of conifers. Int J Plant Sci 163:197–214.

Saitou N, M Nei 1987 The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406–425.

Sanderson MJ, MF Wojciechowski, J-M Hu, TS Khan, SG Brady 2000 Error, bias, and long-branch attraction in data for two chloroplast photosystem genes in seed plants. Mol Biol Evol 17:782–797.

Simmons MP, L-B Zhang, CT Webb, A Reeves, JA Miller 2006 The relative performance of Bayesian and parsimony approaches when sampling characters under homogeneous and heterogeneous sets of parameters. Cladistics 22:171–185.

Soltis DE, VA Albert, V Savolainen, K Hilu, Y-L Qiu, MW Chase, JS Farris, et al 2004 Genome-scale data, angiosperm relationships, and "ending incongruence": a cautionary tale in phylogenetics. Trends Plant Sci 9:477–483.

Soltis DE, PS Soltis, MJ Zanis 2002 Phylogeny of seed plants based on evidence from eight genes. Am J Bot 89:1670–1681.

Spencer M, E Susko, AJ Roger 2005 Likelihood, parsimony, and heterogeneous evolution. Mol Biol Evol 22:1161–1164.

Stanger-Hall K, CW Cunningham 1998 Support for a monophyletic lemuriformes: overcoming incongruence between data partitions. Mol Biol Evol 15:1572–1577.

Stefanovic S, M Jager, J Deutsch, J Broutin, M Masselot 1998 Phylogenetic relationships of conifers inferred from partial 28S rRNA gene sequences. Am J Bot 85:688–697.

Stefanovic S, DW Rice, JD Palmer 2004 Long branch attraction, taxon sampling, and the earliest angiosperms: *Amborella* or monocots? BMC Evol Biol 4:35.

Sullivan J, DL Swofford 2001 Should we use model-based methods for phylogenetic inference when we know that assumptions about among-site rate variation and nucleotide substitution pattern are violated? Syst Biol 50:723–729.

Swofford DL 2002 PAUP*: phylogenetic analysis using parsimony (*and other methods), version 4.0b10. Sinauer, Sunderland, MA.

Swofford DL, PJ Waddell, JP Huelsenbeck, PG Foster, PO Lewis, JS Rogers 2001 Bias in phylogenetic estimation and its relevance to the choice between parsimony and likelihood methods. Syst Biol 50: 525–539.

Tavaré S 1986 Some probabilistic and statistical problems on the analysis of DNA sequences. Pages 57–86 *in* RM Miura, ed. Lectures on mathematics in life sciences. American Mathematics Society, Providence, RI.

Wilgenbush J, K de Queiroz 2000 Phylogenetic relationships among the phrynosomatid sand lizards inferred from mitochondrial DNA sequences generated by heterogeneous evolutionary processes. Syst Biol 49:592–612.

Won H, SS Renner 2003 Horizontal gene transfer from flowering plants to *Gnetum*. Proc Natl Acad Sci USA 100:10824–10829.

Yang Z 1994 Maximizing likelihood phylogenetic estimation from DNA sequences with variable rates over sites: approximate methods. J Mol Evol 39:306–314.