Syst. Biol. 58(2):184–198, 2009
Copyright © Society of Systematic Biologists
DOI:10.1093/sysbio/syp019
Advance Access publication on June 29, 2009

Homoplasy and Clade Support

MATTHEW C. BRANDLEY[1,3,*,†], DAN L. WARREN[2,†], ADAM D. LEACHÉ[1,4], AND JIMMY A. MCGUIRE[1]

[1]*Museum of Vertebrate Zoology and Department of Integrative Biology, University of California, Berkeley, CA 94720-3160, USA;*
[2]*Center for Population Biology, Section of Evolution and Ecology, University of California, Davis, CA 95616-5270, USA;*
[3]*Current address: Department of Ecology and Evolutionary Biology, Yale University, New Haven, CT 06520, USA;*
[4]*Current address: Genome Center and Section of Evolution and Ecology, University of California, Davis, CA 95616, USA;*
[*]*Correspondence to be sent to: Department of Ecology and Evolutionary Biology, Yale University, New Haven, CT 06520, USA; E-mail: matthew.brandley@yale.edu.*

[†]Matthew C. Brandley and Dan L. Warren have contributed equally to this work.

Abstract.—Distinguishing phylogenetic signal from homoplasy (shared similarities among taxa that do not arise by common ancestry) is an implicit goal of any phylogenetic study. Large amounts of homoplasy can interfere with accurate tree inference, and it is expected that common measures of clade support, including bootstrap proportions and Bayesian posterior probabilities, should also be impacted to some degree by homoplasy. Through data simulation and analysis of 38 empirical data sets, we show that high amounts of homoplasy will affect all measures of clade support in a manner that is dependent on clade size. More specifically, the smallest taxon bipartitions in an unrooted tree topology will receive higher support relative to clades of intermediate sizes, even when all clades are supported by the same amount of data. We determine that the ultimate causes of this effect are the inclusion of random trees (due to homoplasy) during bootstrap resampling and Markov chain Monte Carlo (MCMC) topology searching and the higher relative proportion of small taxon bipartitions (i.e., 2 or 3 taxa) to larger sized bipartitions. However, the use of explicit model-based methods, especially Bayesian MCMC methods, effectively overcomes this clade size effect even when very small amounts of phylogenetic signal are present. We develop a post hoc statistic, the clade disparity index (CDI), to measure both the relative magnitude of the clade size effect and its statistical significance. In analyses of both simulated and empirical data, CDI values indicate that Bayesian MCMC analyses are substantially more likely to estimate clade support values that are uncorrelated with clade size than are maximum parsimony and maximum likelihood bootstrap analyses and thus less affected by homoplasy. These results may be especially relevant to "deep" phylogenetic problems, such as reconstructing the tree of life, as they represent the largest possible extremes of time and evolutionary rates, 2 factors that cause homoplasy. [Bayesian posterior probability; bootstrap; clade size; homoplasy; prior probability.]

Homoplasy is present in virtually all phylogenetic analyses, and the most commonly used analytical methods seek to either minimize it (e.g., maximum parsimony [MP]) or mitigate its effects through the use of explicit evolutionary models (e.g., maximum likelihood [ML] and Bayesian approaches). For DNA sequence data, it is well known that high rates of molecular evolution coupled with long periods of time will increase the probability that 2 or more taxa will share the same character states due to a process other than common ancestry (Felsenstein 1978, 2004). Large amounts of homoplasy will severely disrupt the accuracy of phylogenetic methods (Huelsenbeck 1995; Swofford et al. 2001), and techniques for overcoming homoplasy include increased taxon sampling (Hillis 1996; Pollock et al. 2002; Zwickl and Hillis 2002), the use of more parameter-rich models of molecular evolution (Yang 1993, 1994; Tavaré 1986), and data partitioning strategies (e.g., Castoe et al. 2004; Brandley et al. 2005; Nylander et al. 2004). Given that it is now conventional that phylogenetic hypotheses must also be accompanied by an assessment of clade support, a detailed exploration of the effects of homoplasy on the most common methods for assessing clade support, including resampling methods (e.g., nonparametric bootstrap [Felsenstein 1985; "bootstrap" hereafter] and jackknife [Lanyon 1985]), and Bayesian posterior probabilities (PPs; Yang and Rannala 1997; Larget and Simon 1999; Mau et al. 1999) is warranted.

The effect of homoplasy on measures of clade support may appear intuitive; if the background noise introduced by homoplasy is high, then the "true" phylogenetic signal (regardless of whether it infers the unknowable true phylogenetic history) will be obscured. Under these conditions, phylogenetic analysis will infer numerous erroneous taxon bipartitions in addition to the "true" tree, reducing support values. However, homoplasy can influence clade support estimates in ways that are not immediately obvious (Pickett and Randle 2005; Brandley et al. 2006; Randle and Pickett 2006). Pickett and Randle (2005) showed an association across empirical studies between the size of a clade (and thus, its prior probability) and its estimated support (bootstrap and jackknife proportions and Bayesian PPs). More specifically, in an unrooted tree, small bipartitions may have inflated support values relative to moderate-sized bipartitions even when all clades are supported by the same amount of data (we will refer to this as the "clade size effect" hereafter). Pickett and Randle (2005) suggested that homoplasy could be responsible for inducing the clade size effect. Although their discussion did not explicitly rule out the possibility that Bayesian methods were subject to the effects of homoplasy, they concluded that the cause of the clade size effect in Bayesian methods is the inclusion of nonuniform clade priors. Brandley et al. (2006) demonstrated that the relationship between clade size and support in Bayesian analyses was not due to the inclusion of clade priors in calculating Bayesian PPs (contra Pickett and Randle 2005; Randle and Pickett 2006) and postulated that a single, yet to be determined factor was the underlying cause of the clade size

effect in both Bayesian and resampling methods. It remains unclear if homoplasy is the source of the clade size effect for any analytical approach as the methods used to study this clade size effect phenomenon have not explicitly accounted for the presence of homoplasy. Furthermore, if homoplasy is driving the clade size effect, it is unclear if different analytical methods are more or less susceptible to this source of bias.

In this study, we conduct data simulations to identify the conditions under which homoplasy will disproportionally affect measures of clade support based on clade size alone and is thus an underlying factor that can induce the clade size effect for bootstrap proportions and Bayesian PPs. Our simulation results demonstrate that the clade size effect is much stronger in MP bootstrap analyses and that explicit model-based methods of analysis (e.g., ML and Bayesian analysis) can reduce the influence of clade size–dependent support. In addition, we analyze 38 empirical data sets, including most of those originally studied by Pickett and Randle (2005), and demonstrate that Bayesian inference is more likely to yield clade support estimates that are not associated with clade size compared with MP and ML bootstrap values.

These results have practical significance for the goal of estimating support for "deep" phylogenies, such as the tree of life (TOL). Homoplasy is dependent on both time and rate. As the TOL encompasses 3.5 billion years of change and a multitude of unique lineages evolving at different rates, it circumscribes the largest biologically possible parameter space for homoplasy. In order to eventually assess our confidence in this phylogeny, or any phylogenetic tree, we must understand how our measures of confidence respond to homoplasy.

Materials and Methods

Analyses of Simulated Data

Creating homoplasy data sets.—To examine the effects of homoplasy on commonly used measures of clade support, we simulated data using a Bayesian framework that incorporates an explicit phylogenetic history (albeit random) and an explicit model of nucleotide substitution (general time reversible [GTR] + Γ). Traditional approaches for simulating data (i.e., simulating multiple data sets on a fixed tree with fixed parameters) violate the basic assumptions of Bayesian analysis, namely, that Bayesian analysis treats all parameters as random variables (Huelsenbeck and Rannala 2004). Instead, the Bayesian simulation regimen draws all parameters of the model (e.g., topology, branch lengths, nucleotide substitution rates, and alpha parameter for the gamma distribution) from a prior distribution. An alternate method of simulating random data is to draw nucleotide characters randomly while assuming no underlying tree structure. From a theoretical perspective, our approach differs in that we assume an explicit common mechanism responsible for the evolution of each character (a phylogenetic history), although the overall effect of either method of producing homoplasious characters may be the same. Under this Bayesian framework, we simulated 200 individual data sets of 560 characters using the following methodology.

For each character:

1. Randomly draw a tree from a uniform distribution of unrooted 15-taxon trees.
2. Draw a set of branch lengths from an exponential distribution ($f(x;\lambda) = \lambda e^{-\lambda x}$, where $x \geq 0$) with $\lambda = 5$.
3. Draw a gamma shape parameter (α) from an exponential distribution with $\lambda = 2$.
4. Draw substitution rate parameters from 6 separate exponential distributions with $\lambda = 1$ corresponding to the rates r_{AC}, r_{AG}, r_{AT}, r_{CG}, r_{CT}, and r_{GT}.
5. Simulate a single nucleotide character with the GTR + Γ model and parameters from Steps 3 and 4 and the tree from Steps 1 and 2 using Seq-Gen v1.3.2 (Rambaut and Grassly 1997).
6. Repeat Steps 1–5 until a data set of 560 characters is created.

The resulting data are a collection of characters simulated on conflicting phylogenies. We chose a data set size of 560 characters because it is a multiple of 14 characters (the number of characters used in our contrived data set) and is small enough that the response to phylogenetic signal is apparent and measurable (something that would be difficult with a data set much larger or much smaller). We acknowledge that this homoplasy simulation methodology assumes that homoplasy is essentially random noise and excludes the potential for convergent evolution due to selection. However, we feel that this is unproblematic because an implicit assumption of most phylogenetic analyses is that the characters included in the analysis are not under directional selection.

Replacing homoplasy with phylogenetic signal.—To evaluate the effects of homoplasy on bootstrap proportions and Bayesian PPs, we conducted phylogenetic analyses (MP and ML bootstrap and Bayesian) for homoplasy data sets that contained progressively larger ratios of contrived data supporting a single topology (Fig. 1a) to homoplasious data. Essentially, replacing increasing quantities of phylogenetic "noise" with phylogenetic "signal" (Fig. 1b) in the data sets allowed us to evaluate the effect of homoplasy on estimates of clade support. Six separate "analysis sets" were constructed with increasing amounts of contrived data (0, 14, 28, 42, 56, and 70 nucleotides), while always maintaining a total data set size of 560 characters. Note that the addition of 14 bp of contrived data adds 1 uncontradicted synapomorphy per clade (as well as a single total autapomorphy and a single symplesiomorphy; Fig. 1), the 28-bp analysis set adds 2 synapomorphies per clade, etc.

Phylogenetic analyses.—For all data sets, MP bootstraps were performed using PAUP* v4b10 (Swofford 2002). Each analysis consisted of 500 pseudoreplicates, 100 random stepwise addition replicates per pseudoreplicate, and tree bisection and reconnection (TBR) branch

a)

OG	A	T	A	T	A	T	A	C	G	C	G	C	G	C
A	T	T	A	T	A	T	A	C	G	C	G	C	G	C
B	T	A	A	T	A	T	A	C	G	C	G	C	G	C
C	T	A	T	T	A	T	A	C	G	C	G	C	G	C
D	T	A	T	A	A	T	A	C	G	C	G	C	G	C
E	T	A	T	A	T	T	A	C	G	C	G	C	G	C
F	T	A	T	A	T	A	A	C	G	C	G	C	G	C
G	T	A	T	A	T	A	T	C	G	C	G	C	G	C
H	T	A	T	A	T	A	T	G	G	C	G	C	G	C
I	T	A	T	A	T	A	T	G	C	C	G	C	G	C
J	T	A	T	A	T	A	T	G	C	G	G	C	G	C
K	T	A	T	A	T	A	T	G	C	G	C	C	G	C
L	T	A	T	A	T	A	T	G	C	G	C	G	G	C
M	T	A	T	A	T	A	T	G	C	G	C	G	C	C
N	T	A	T	A	T	A	T	G	C	G	C	G	C	G

b)

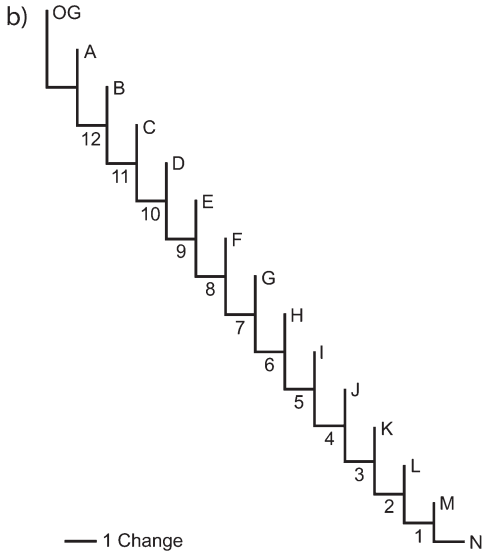

FIGURE 1. a) The 15-taxon contrived data matrix used to replace homoplastic data in the simulation analyses and b) reference phylogram for all analyses. Numbers above clades represent labels used in Supplementary Appendices 1–3. Relative branch lengths represent those from a phylogenetic analysis of the contrived data set.

swapping. ML bootstraps were performed using RAxML v7.0.4 (Stamatakis 2006), and each analysis consisted of 200 pseudoreplicates, using the GTR + Γ model (GTRMIX with 4 rate categories) with all model parameters estimated from the data. All Bayesian analyses were performed using MrBayes v3.1.2 (Ronquist and Huelsenbeck 2003). Each Bayesian analysis consisted of 2 million generations, using 4 chains, sampled every 1000 generations, and used the default priors (substitution rates, Dirichlet [1, 1, 1, 1, 1, 1]; base frequencies, Dirichlet [1, 1, 1, 1]; gamma shape parameter, uniform [0, 200]; topologies, uniform; branch lengths, unconstrained exponential [$\lambda = 10$]). To determine convergence, cumulative posterior probability plots were constructed for several analyses using the "cumulative" function in Are We There Yet? (AWTY; Nylander et al. 2008). These plots indicated that excluding the first 1 million generations as burn-in was sufficient, and thus PPs for individual clades (see Sukumaran and Linkem 2008) for all analyses were estimated from trees sampled after this point (1000 trees). Bootstrap proportions and PPs for each of the 6 analysis sets were summarized by calculating the grand mean of the values for each of the contrived data clades in Figure 1b (i.e., the true phylogenetic signal) for all data sets analyzed (Brandley et al. 2006).

Modulating the strength of the homoplasy.—To test how clade support values respond to situations where the "noisy" data more strongly prefer fewer, incorrect random topologies, we repeated our Bayesian simulations making fewer draws from the prior distribution of trees. Thus, instead of evolving a single character on each of 560 random trees, we evolved 2 characters on each of 280 random trees and 16 characters on 35 trees. Although this approach generated data that supported conflicting topologies, we expected that the influence of having fewer underlying trees would render the reconstruction of the true (contrived data) tree more difficult. These alternative homoplasy data sets were mixed with contrived data and analyzed using the methods described above.

Note that our simulation strategy assumes that the phylogenetic "signal" supports a fully pectinate tree (Fig. 1a). We chose this tree because it exhibits a uniform range of possible clade sizes. To test the possibility that this tree shape biases our results, we conducted additional simulations using the above methodology (reducing the number of simulations from 200 to 100 and only simulating one character per random tree) but using contrived data that support a completely symmetric, 16-taxon unrooted tree.

Analyses of Empirical Data Sets

We tested 38 empirical data sets for an association between the size of a clade (i.e., its prior probability based on its size) and its MP and ML bootstrap proportion and Bayesian PPs to assess the existence and magnitude of the clade size effect in real data. A similar analysis was presented by Pickett and Randle (2005) based on 17 empirical data sets; however, our analytical regimen differs in several important ways. First, we conducted separate analyses for each empirical data set instead of pooling clade support values across all 17 studies. In addition, we evaluated MP and ML bootstrap results separately. These distinctions are critical because there are reasons to suspect that pooling samples from across studies and across various methods of analysis may bias the results. For instance, some studies provide bootstrap results from multiple analyses of the same data, and pooling these values together and then combining them with other empirical studies (as done by Pickett and Randle 2005) effectively weight the results of one empirical data set over another. Some analyses employed character weighting schemes, and different weighting schemes may affect bootstrap proportions (e.g., Milinkovitch et al. 1996). Other more cryptic problems can arise if a data set has an underlying bias that can affect clade support, including different gene trees and taxon-sampling artifacts. Most importantly, equating parsimony and

likelihood bootstraps may not be justified given the known biases of parsimony toward long-branch attraction (Felsenstein 1978; Huelsenbeck 1995; Swofford et al. 2001; see Results). Second, many of the empirical studies examined by Pickett and Randle (2005) did not report clade support values less than 50%. Some of these clades were small (i.e., had high clade prior probabilities), yet had low clade support, and thus not including them excludes evidence contrary to the hypothesis that small clade size is associated with high clade support. Finally, for a variety of reasons (discussed below) we develop a new ad hoc statistic to evaluate the relationship between clade prior probabilities and their estimated clade support that differs from the 2-dimensional Kolmogorov–Smirnov (2DKS) test employed by both Pickett and Randle (2005) and Brandley et al. (2006).

To provide a more level playing field from which we could then evaluate the association between the prior probability of a clade and MP and ML bootstrap proportions and Bayesian PPs, we reanalyzed 38 empirical data sets using similar analytical conditions. We included most of the data sets used by Pickett and Randle (2005), as well as additional data sets that differed in number of taxa, type of organism, number of characters, and genes (nuclear, mitochondrial, protein coding, RNA, etc.; Table 3). We used NEXUS files (with the authors' DNA alignment) obtained directly from the authors or from TreeBASE (Sanderson et al. 1994). We excluded characters if indicated in the file. Equally weighted MP bootstrap, ML bootstrap, and Bayesian analyses were performed. MP bootstrap analyses were conducted with PAUP* 4b10 using 500 pseudoreplicates, 20 random addition replicates per pseudoreplicate, saving a maximum of 100 trees per replicate (chuckscore =1, nchuck = 100), and TBR branch swapping. ML bootstrap analyses were conduced using RAxML 2.2.0 (Stamatakis 2006). Each analysis consisted of 200 pseudoreplicates using the GTR + Γ model (GTRMIX with 4 rate categories) with all model parameters estimated from the data. Bayesian analyses were performed using MrBayes v3.1.2 using default priors (see above), the GTR + Γ model, initially run for 10^7 generations, and convergence was assessed using AWTY. Due to lack of convergence, some data sets were reanalyzed with 2×10^7 generations, and because of extremely slow convergence, the data sets of Ruiz-Trillo et al. (1999) and Wu et al. (2001) were reanalyzed using the parallel version of MrBayes v3.1.1 (Altekar et al. 2004) and were run for 6×10^7 and 5×10^7 generations, respectively, with the number of MCMC chains increased to 8. All Bayesian analyses of empirical data were performed twice, and provided that both analyses converged on a similar PP distribution (assessed using the "compare" option in AWTY), post-"burn-in" samples were pooled to calculate individual clade PPs.

Determining the relationship between clade size and estimated clade support.—Because a clade's prior probability is directly related to its size (Pickett and Randle 2005), to determine if clade size alone influences clade support, our strategy was to determine whether clade support was significantly associated with a clade's prior probability. Although MP and ML bootstrap analyses do not include prior probabilities in their calculations (unlike Bayesian analyses), it is nonetheless justified to use clade priors as a proxy for clade size in these analyses because the trees inferred by each bootstrap pseudoreplicate may be affected by the same underlying factor that induces the size-based bias in clade priors, that is, the distribution of clade sizes on randomly drawn trees (Pickett and Randle 2005; see Discussion).

To calculate clade prior probabilities, we used the following formula:

$$\frac{\prod_{i=2}^{T}(2i-3)\prod_{i=2}^{n-T}(2i-3)}{\prod_{i=2}^{n-1}(2i-3)}, \tag{1}$$

where n is the number of species in the tree and T is the number of species in the clade in question. Note that this differs from the equation used by Pickett and Randle (2005) in that our formula assumes unrooted rather than rooted phylogenies. As all our analyses (and indeed, most commonly used phylogenetic algorithms) infer unrooted phylogenies, it is the more appropriate formula to use in this setting. This is also the same formula we used, but neglected to mention, in our previous paper (Brandley et al. 2006).

To determine whether estimated clade support is correlated with that clade's size (i.e., prior probability), both Pickett and Randle (2005) and Brandley et al. (2006) previously employed the 2DKS test (Garvey et al. 1998). As an alternative metric, Brandley et al. (2006) also provided the difference in maximum and minimum clade support for the tree as they noted that the 2DKS test seemed to be insensitive to the "magnitude" of the clade size effect. This and subsequent use of the 2DKS test led us to develop a more informative statistic that allows us to evaluate the existence of the clade size effect (i.e., the influence on clade support due solely to clade size) and also evaluate its magnitude relative to the mean support for clades on a given tree. We call this statistic the clade disparity index (CDI), and it is represented by the following formula:

$$\text{Clade disparity index} = 1 - \frac{\sum_{h=1}^{k}\left|\left(\frac{s_h}{\sum_{i=1}^{k}s_i} - \frac{p_h}{\sum_{j=1}^{k}p_j}\right)\right|}{\sum_{h=1}^{k}\left|\left(\frac{1}{k} - \frac{p_h}{\sum_{j=1}^{k}p_j}\right)\right|}, \tag{2}$$

where k = number of nodes in the tree (number of taxa – 3), s_h = support for node h, and p_h = prior probability of node h. This metric simply scales the priors and the support values to have the same mean ($1/k$) and sums the deviation between each support value and its respective prior. The denominator scales this sum by the departure of the priors from a uniform distribution with the same mean. A positive value is consistent with a positive relationship between a clade's size and estimated clade support (e.g., smaller clades with higher prior

probabilities also have higher clade support relative to larger clades; the classic "clade size effect"). A negative value is consistent with a negative relationship between a clade's size and estimated clade support (i.e., smaller clades with higher prior probabilities have lower support relative to larger clades). A value of 0 is consistent with clade support values that are uniformly distributed and are uncorrelated with clade size. The upper bound of the CDI is 1, which occurs when all clade support values are equal to the prior probabilities multiplied by the same constant. We have not yet developed a formula for the lower bound, which depends on the precise distribution of the priors. Figure 2 shows 5 example CDI values for a 15-taxon tree. We emphasize that the CDI is most valuable as a comparative tool to assess different clade support analyses of the same data set. Although a value closer to 0 indicates little relationship between overall clade size and support, much like a likelihood score, the CDI value has no absolute meaning in isolation and cannot be used to precisely compare analyses among different data sets, although the qualitative comparisons of these values may be somewhat useful. For example, although one cannot claim that a data set with a CDI of 0.3 shows evidence of 3 times the clade size effect of a data set with a CDI = 0.1, one can nonetheless infer that the clade size effect is much less pronounced in the latter data set.

We tested for significance of the CDI using a randomization test. For the simulation analyses, actual support values (MP bootstrap, ML bootstrap, and Bayesian PPs; Fig. 3 and Supplementary Appendices 1–3 [available from http://sysbio.oxfordjournals.org/]) for each of the different homoplasy treatments were randomly assigned to nodes and the CDI was calculated. This process was repeated 10 000 times, producing a distribution of expected CDI values under the null hypothesis of no relationship between clade size and support values. This process was also used for the 38 empirical data sets. For each data set, the CDI for actual, nonrandomized data was compared with the null distribution and a P value was calculated as the percent of the null distribution with values the same as or greater than the observed CDI. We determined the significance of the P value using sequential Bonferroni correction (Holm 1979; Rice 1989) for 3 comparisons. Significant support for the alternative hypothesis indicates a significant association between clade size and support value. Because trees of different sizes were used in these empirical analyses and because the statistical significance of CDI values depends on the number of nodes in a tree and the number of species subtending each node, similar CDI values may be statistically significant on one tree but not another.

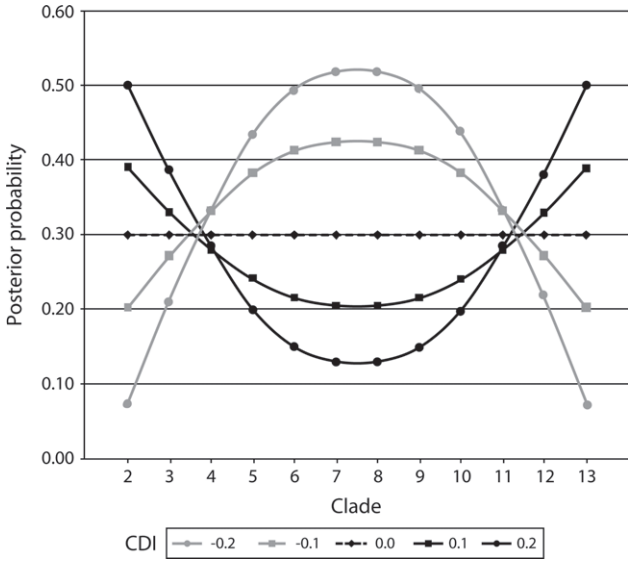

FIGURE 2. Example CDIs (equation 2) for 5 different distributions of PPs for a 15-taxon tree. A positive CDI value indicates a positive relationship between a clade's size and estimated clade support (e.g., smaller clades with higher prior probabilities also have higher clade support relative to larger clades; the classic "clade size effect"). A negative value indicates a negative relationship between a clade's size and estimated clade support (i.e., smaller clades with higher prior probabilities have lower support relative to larger clades). A CDI value of 0 indicates that there is no relationship between a clade's size and support. Clade numbers (x axis) refer to the tree in Figure 1b.

The strength of the CDI calculation is that it can be used to assess 1) whether clade size and support values are associated (the P value) and 2) the magnitude of the effect (the CDI value). It is critically important that both the P values of the randomization test and the raw CDI values are taken into account when interpreting the results of the CDI statistic in this paper. For example, the contrived example in Table 1 shows a significant relationship between a clade's prior probability and support. Although this pattern of clade support is significant ($P = 0.0085$), the magnitude of the difference between support for the smallest and largest clades is only on the scale of 0.001, values that most phylogeneticists would interpret as trivial. This is reflected in the CDI that is very close to zero (6.14×10^{-5}).

Evaluating the effects of homoplasy in empirical data.—To evaluate which phylogenetic methods are least prone to the clade size effect in empirical data, we ranked the CDIs of the MP bootstrap, ML bootstrap, and Bayesian analyses for each empirical data set based on their deviation from a CDI = 0 (i.e., when there exists no relationship between clade size and support). We then employed a nonparametric rank-based test (Friedman 1937) to make pairwise comparisons of the ranks of different measures of clade support. For example, if one analysis had a CDI = 0.10 and the other 0.20, the first analysis would receive a rank of "1" and the other "2". If one clade support method is less affected by the clade size effect compared with another method, its CDI will be consistently closer to 0 (i.e., have a mean rank closer to 1.0) for a significant number of empirical data sets compared with the other methods. Sequential Bonferroni corrections for 3 comparisons were made during assessment of significance. We did not use the Friedman test for the simulation data due to the small sample size ($n = 5$).

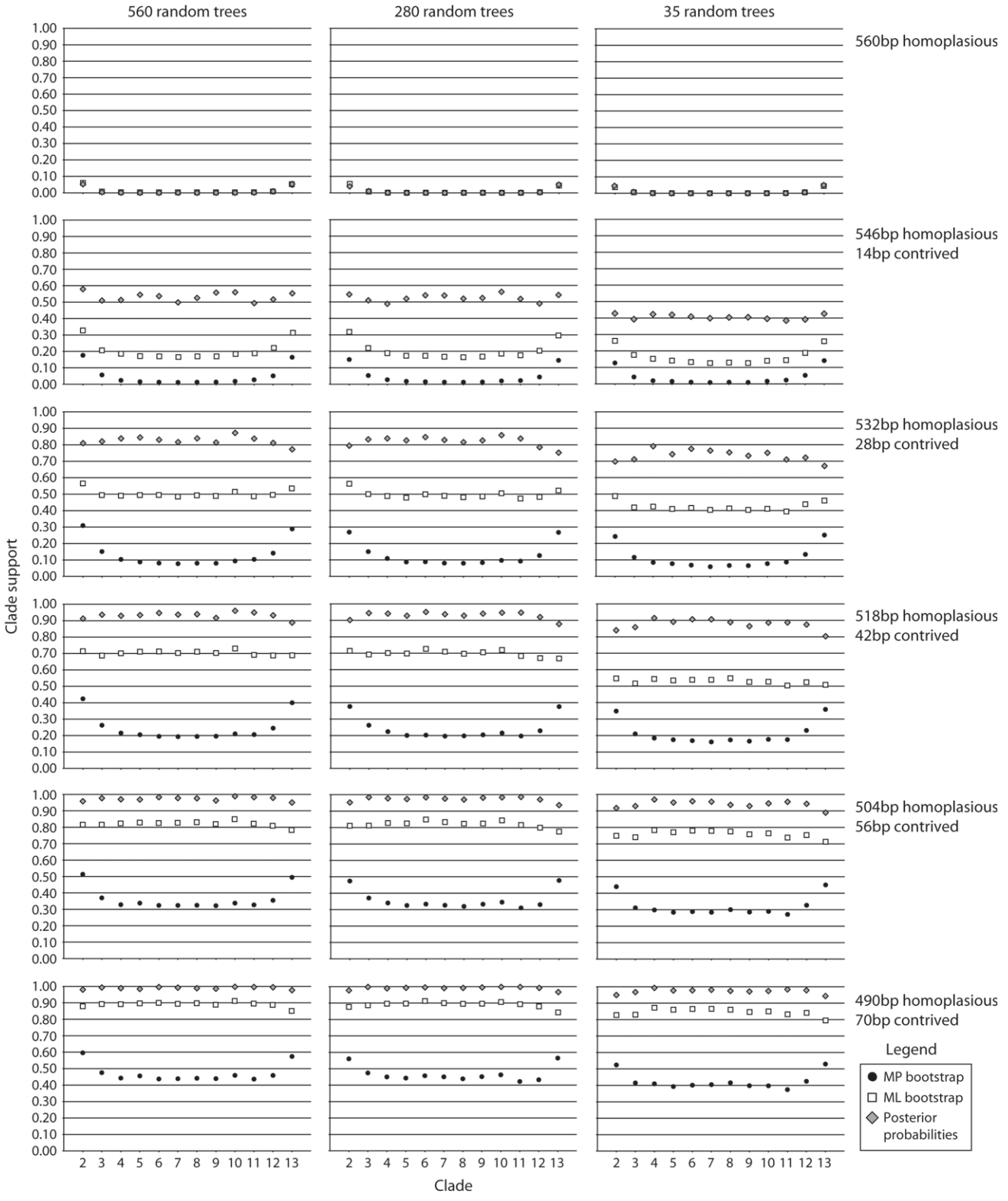

FIGURE 3. MP and ML bootstrap proportions and Bayesian PPs of data sets containing variable amounts of homoplasy. Homoplasy for each data set was modeled by evolving 1, 2, or 16 characters per randomly drawn, unrooted tree (560 characters, 560 trees). These homoplasious data were replaced with increasing amounts of contrived data supporting a single topology (Fig. 1). Values are grand means of 200 separate analyses for the MP bootstraps and Bayesian analyses and 100 analyses for the ML bootstraps.

TABLE 1. An example of a 15-taxon pectinate tree (e.g., Fig. 1b), where overall clade support is extremely high and the values differ by only a very small value (0.001). There is a significant association between the clade size/prior probabilities and clade support (P value), but the magnitude of this effect is extremely small (CDI $\sim$ 0)

Prior probability	Clade support
0.04	0.9999
0.00522	0.9998
0.00124	0.9997
0.000458	0.9996
0.000242	0.9995
0.000178	0.9994
0.000178	0.9994
0.000242	0.9995
0.000458	0.9996
0.00124	0.9997
0.00522	0.9998
0.04	0.9999
	CDI = 6.14×10^{-5}
	$P = 0.0085$

Because the CDI measures the clade size effect, we used it to determine if there is a relationship between this effect and levels of homoplasy in the empirical data sets. To do this, we calculated Farris' distortion coefficient (d) (Farris 1973), which measures the "fraction of possible homoplasy" (Farris 1989), and plotted these values against the CDI for each empirical data set. We determined the distortion coefficient by calculating the retention index and subtracting this value from 1 (Farris 1989). We calculated the retention index in PAUP* using an ML tree (estimated by RAxML under the same analytical conditions above) instead of the MP tree because the goal of parsimony is to minimize homoplasy, and calculating the retention index from the ML tree will more accurately reveal the amount of homoplasy. Given that CDI values are not precisely comparable (see above), in addition to the limitations of the distortion coefficient (and homoplasy statistics in general; see Archie 1989; Naylor and Kraus 1995), we acknowledge that this is a somewhat crude way to measure the relationship between homoplasy and the clade size effect in empirical data. However, we point out that the homoplasy will likely be underestimated given that this methodology is dependent solely on tree topology (rather than branch lengths and explicit estimation of character change). Thus, we do not subject the results of this analysis to statistical evaluation and instead restrict ourselves to a qualitative interpretation of the results.

RESULTS

Simulation Analyses

The grand means of all 200 MP and ML bootstrap and Bayesian analyses for each manipulation of contrived data and homoplasy creation regimen are provided in Supplementary Appendices 1–3 and Figure 3. The results of the CDI analyses are presented in Table 2 and Figure 4. Not surprisingly, the general clade support for the true, contrived data tree is lower when homoplasy simulations are restricted to fewer underlying trees (thus providing stronger support for a smaller set of random trees). However, the magnitude of the clade size effect is small and is very similar among all 3 homoplasy creation regimens. Thus, unless noted, discussions of the results will be generalized across all 3 regimens.

As expected, in the analyses of homoplasious data that contain no contrived phylogenetic signal, the Bayesian PPs are significantly associated with the prior probabilities and the CDI is extremely high (0.77–0.94; Table 2 and Fig. 4). The 2 smallest taxon bipartitions in an unrooted tree show the highest bootstrap proportions, and clade support is lower for clades of intermediate sizes. However, given that resampling methods such as bootstrapping do not incorporate prior probability information, it may be somewhat surprising that the bootstrap analyses display the same behavior in the absence of phylogenetic signal and have CDIs that are actually slightly higher (Table 2) than the Bayesian analyses (i.e., the clade size effect is stronger) in 2 of the 3 homoplasy simulation regimens.

The addition of phylogenetic signal reduces the disparity of support between the different-sized clades in the likelihood-based analyses (ML and Bayesian), but the effect remains severe for MP (Table 2 and Figs. 3 and 4). With the replacement of 14 bp of homoplastic data with contrived data (one uncontradicted synapomorphy per clade), the CDIs of the Bayesian analyses drop to almost 0 (Fig. 4), although the values are still significant in 1 of the 3 homoplasy simulation regimens (Table 2). More interestingly, this occurs even well before the threshold for individual clade statistical significance $\geq$0.95 (Fig. 3). With the same amount of contrived data, the CDIs of the ML bootstrap also decrease dramatically (from $\sim$0.9 to $\sim$0.13, but are still significant; Table 2 and Fig. 4) but is higher than that for the Bayesian analyses. Although the CDIs of MP bootstrap analyses also decrease when the relative amount of homoplasy is reduced, the improvement is much smaller than that for the ML bootstrap and Bayesian analyses (from $\sim$0.9 to $\sim$0.6).

In the Bayesian analyses of 28 bp of contrived data (and 532 bp of homoplasious data), the disparity of clade support based on clade size disappears (CDI $\sim$ 0, randomization tests not significant; Table 2 and Figs. 3 and 4), and overall clade support is moderate to high but still not above the commonly used thresholds for "strong" support ($\geq$0.95 posterior probability). The further addition of contrived data only results in more support for the clades in the true underlying tree (Figs. 1b and 3). The CDIs of the ML bootstrap analyses are also quite low ($\sim$0.03), but still significant. However, the CDIs become nonsignificant with the addition of 42 bp or more of contrived data. On the other hand, the CDIs of the MP bootstrap analyses remain comparatively high (CDI $\sim$ 0.3, P = 0.018–0.049), and overall support for the contrived data tree is low (Table 2 and Fig. 4).

The performance of the parsimony bootstrap in terms of inferring the correct clades with strong support, as well as overcoming clade size effects, is poor compared with ML and Bayesian methods. Even with the addition of 70 bp of contrived data, the MP bootstrap

TABLE 2. CDI results of MP and ML bootstrap and Bayesian analyses (PP) of simulated data, where homoplasy was modeled by evolving 1, 2, or 16 characters per random tree (for a total of 560 characters) and was increasingly replaced with contrived data (Fig. 1a) that supported a single tree (Fig. 1b). Clade support values used in the CDI calculation (equation 2) were the grand means of 200 separate analyses for the MP and ML bootstraps and Bayesian analyses (Supplementary Appendices 1–3). Clade priors were calculated using equation (1). Statistical significance was assessed using a randomization test (see text) and significant *P* values (after sequential Bonferroni correction) are in bold

Homoplasy simulation	Contrived characters	Homoplasious characters	MP		ML		PP	
			CDI	*P*	CDI	*P*	CDI	*P*
One homoplasious character per tree	0	560	0.9025	**0.0001**	0.9272	**0.0001**	0.9422	**0.0001**
	14	546	0.6166	**0.0002**	0.1364	**0.0115**	0.0157	**0.0204**
	28	532	0.3055	**0.0049**	0.0230	**0.0155**	−0.0103	0.9856
	42	518	0.1662	**0.0157**	−0.0007	0.5154	−0.0083	0.9885
	56	504	0.0954	**0.0075**	−0.0064	0.9847	−0.0046	0.9935
	70	490	0.0594	**0.0155**	−0.0070	0.9999	−0.0027	0.9955
	84	476	0.0382	**0.0118**				
	98	462	0.0241	**0.0062**				
	112	448	0.0162	**0.0149**				
	126	434	0.0104	**0.0103**				
	140	420	0.0066	**0.0030**				
Two homoplasious characters per tree	0	560	0.8988	**0.0001**	0.8960	**0.0001**	0.8504	**0.0005**
	14	546	0.5709	**0.0004**	0.1263	**0.0107**	0.0090	0.0795
	28	532	0.2704	**0.0018**	0.0223	**0.0129**	−0.0141	0.9855
	42	518	0.1392	**0.0133**	−0.0027	0.7201	−0.0107	0.9917
	56	504	0.0816	**0.0152**	−0.0080	0.9861	−0.0071	0.9940
	70	490	0.0500	**0.0154**	−0.0083	0.9940	−0.0046	0.9943
	84	476	0.0318	**0.0155**				
	98	462	0.0211	**0.0175**				
	112	448	0.0129	**0.0175**				
	126	434	0.0081	**0.0138**				
	140	420	0.0056	**0.0054**				
Sixteen homoplasious characters per tree	0	560	0.9153	**0.0001**	0.9286	**0.0001**	0.7698	**0.0016**
	14	546	0.5900	**0.0002**	0.1424	**0.0081**	0.0125	**0.0084**
	28	532	0.3026	**0.0030**	0.0293	**0.0111**	−0.0170	0.9934
	42	518	0.1668	**0.0148**	−0.0010	0.5697	−0.0154	0.9924
	56	504	0.0970	**0.0147**	−0.0089	0.9695	−0.0096	0.9903
	70	490	0.0597	**0.0158**	−0.0099	0.9966	−0.0066	0.9964
	84	476	0.0357	**0.0108**				
	98	462	0.0219	**0.0148**				
	112	448	0.0126	**0.0163**				
	126	434	0.0067	**0.0326**				
	140	420	0.0017	0.2599				

proportions remain below 0.60 (when all the Bayesian PPs and most ML bootstrap proportions are >0.90; Fig. 2) and the CDI values are significant (Table 2 and Fig. 4). To determine how much data are necessary to overcome persistent clade size–dependent support, we performed additional MP bootstrap analyses replacing 84, 98, 112, 126, and 140 bp of homoplasious data with contrived data using the same analysis protocol as before. Although the CDIs approach 0 as more contrived data are added (and thus, the magnitude of the clade size effect becomes very small; Table 2 and Fig. 4), in 2 of the homoplasy simulation regimens, the CDI is nonetheless significant even with the replacement of 140 bp of homoplasious data with contrived data (10 uncontradicted synapomorphies per clade).

The results of the simulation analyses including contrived data supporting a 16-taxon symmetric tree are very similar to those of the pectinate tree (not shown). The clade size effect is present in MP and ML bootstraps and Bayesian PPs, with 2-taxon clades having higher support than the 4-taxon clades and the single 8-taxon bipartition. As phylogenetic signal is added, overall support for the clades rises, but the effects remain severe in the parsimony analyses even with the addition of 5 uncontradicted synapomorphies per clade. The clade size effect is quickly lost in the analyses using explicit models. Thus, discussion will be limited to the results from the analyses using the contrived data in Figure 1.

Empirical Data Reanalyses

For many data sets, the CDIs for each MP, ML, and Bayesian analysis are similar. With few exceptions (see Discussion), CDI values are positive, indicating that smaller clades have, on average, higher support than medium-sized clades. However, 2 notable exceptions are the data sets of Kiefer et al. (2002) and Ilves and Taylor (2007) for which the CDI is markedly negative (and is the lowest in the Bayesian analysis of the latter data set).

When positive CDIs differ substantially among analytical methods, with a single exception (dePamphilis et al. 1997), the Bayesian analysis is much more likely to have a CDI closer to 0.0 than will the MP and ML

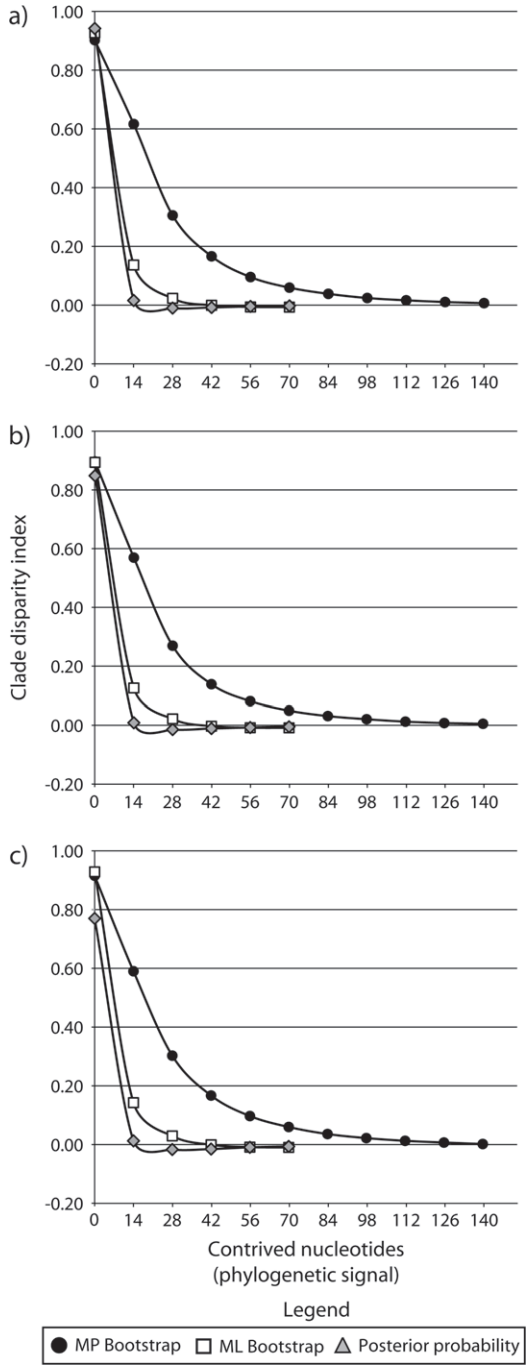

FIGURE 4. CDIs (equation 2) of MP and ML bootstrap and Bayesian analyses (PP). Homoplasy for each data set was modeled by evolving 1 (graph a), 2 (graph b), or 16 (graph c) characters per randomly drawn, unrooted tree. These homoplasious data were replaced with increasing amounts of contrived data (Fig. 1b) supporting a single topology (Fig. 1a) while maintaining a data set of 560 characters. A CDI value of 0 indicates that there is no relationship between a clade's size and support.

bootstrap analyses (Table 3 and Fig. 5). Pairwise Friedman tests (Table 4) also indicate that the Bayesian analyses consistently resulted in support values that showed less evidence of the clade size effect than ML or MP bootstraps.

In multiple empirical data sets, there exists a statistically significant relationship between the prior probability of a clade (based on clade size) and either its MP bootstrap, ML bootstrap, or Bayesian PP (Table 3). Of the 38 data sets, 17 show this relationship in at least one of the clade support analyses; however, the relationship is more prevalent in the MP (11 of 38 data sets) and ML (14 of 38) bootstraps than Bayesian analyses (6 of 38; Table 3).

With few exceptions, if the relationship between clade size and MP bootstrap proportions is significant, it is also significant in the ML bootstrap proportions. In 5 of the 6 cases in which the CDI was significant in the Bayesian analyses, it was also significant in the MP and ML bootstrap analyses. There is only a single example of an empirical data set that implicates Bayesian analyses as having significant clade size–dependent support that does not also affect both MP and ML (dePamphilis et al. 1997) and is thus the only example (of 38) supporting the hypothesis of Pickett and Randle (2005) that PPs are more prone to be biased by clade size.

The plot of CDI against the distortion coefficient (a measure of homoplasy) shows that there is considerable variation in CDI values, even among data sets that show similar levels of homoplasy. If a pattern can be inferred, it is that in the MP and ML analyses, data sets with more homoplasy may have higher CDIs than data sets with less homoplasy (Fig. 6). At the very least, there are no data sets with low homoplasy that have high CDIs. With the exception of 2 data sets with strongly negative CDIs (Kiefer et al. 2003; Ilves and Taylor 2007), the CDI scores for Bayesian analyses remain fairly constant across data sets.

DISCUSSION

Homoplasy Induces the Clade Size Effect

Clade support values estimated by both bootstrap approaches and Bayesian MCMC methods are partially dependent on clade size rather than completely informed by the actual data. These 2 classes of methods have very different assumptions and mechanics, which suggests an underlying factor that is common to both methods. Homoplasy is a likely factor because this effect is induced by high levels of homoplasy and diminishes with the addition of phylogenetic signal.

How would homoplasy induce this effect? The answer centers on the inclusion of random trees in calculating both bootstrap proportions and PPs and, in fact, is closely related to unequal clade prior probabilities. The underlying cause of the association between clade size and bootstrap proportions and Bayesian PPs is the distribution of clades of varying size on randomly chosen trees and the inclusion of effectively random trees in the calculation of support values. This hypothesis was originally proposed by Pickett and Randle (2005) but was only discussed as an explanation for the clade

TABLE 3. CDI results of MP and ML bootstrap and Bayesian (PP) reanalyses of 38 empirical data sets. CDIs were calculated using equation (2) using clade support values from the phylogenetic reanalyses and clade priors calculated from equation (1). Statistical significance was assessed using a randomization test (see text) and significant *P* values (after sequential Bonferroni correction) are in bold

Study	Organism	No. of taxa	No. of characters	MP		ML		PP		
				CDI	*P*	CDI	*P*	CDI	*P*	*d*
Anderson et al. (2003)	Bilateria	63	2943	0.1624	**0.0002**	0.1205	**0.0028**	0.0131	0.1740	0.606
Barns et al. (1996)	Archaea	64	1620	0.1452	**0.0001**	0.1579	**0.0004**	0.0222	0.0527	0.421
Bell and Donoghue (2003)	Morinaceae plants	23	3025	0.0203	0.2387	0.0314	0.1433	0.0028	0.2930	0.307
Berbee and Taylor (2001)	Fungi	52	1531	0.0314	0.2333	0.0304	0.2133	0.0326	0.0576	0.409
Brandley and de Queiroz (2004)	*Anolis* lizards	18	962	0.0068	0.4506	0.0571	0.1158	0.0321	0.1764	0.588
Brandley et al. (2005)	Scincid lizards	60	2654	0.1583	**0.0080**	0.1980	**0.0002**	0.0862	**0.0002**	0.641
Cox and Hedderson (2003)	Bryaceae moss	53	2773	0.0904	0.0375	0.0887	**0.0165**	0.0540	0.1174	0.216
Danforth et al. (2003)	*Lasioglossum* bees	53	3331	0.0675	0.0512	0.0604	0.0430	0.0276	0.0700	0.497
Delsuc et al. (2002)	Xenarthran mammals	50	2789	0.0562	0.0594	0.0374	0.0708	−0.0047	0.6263	0.485
dePamphilis et al. (1997)	Plants	35	614	0.0124	0.1964	0.0747	0.0950	0.0994	**0.0320**	0.366
Des Marais et al. (2003)	Horsetails	22	2342	−0.0109	0.6389	0.0072	0.4344	−0.0080	0.6204	0.223
Dohrmann et al. (2008)	Glass sponges	51	3435	0.0289	0.1474	0.0046	0.4376	−0.0028	0.6173	0.212
Edwards et al. (2005)	*Pereskia* cactus	38	6150	0.0162	0.3425	0.0258	0.1989	−0.0059	0.6238	0.207
Garey et al. (1998)	Rotifera	29	2520	0.0041	0.4750	0.0364	0.2362	0.0216	0.1728	0.525
Ilves and Taylor (2007)	*Hypomesus* fish	15	2732	−0.2693	0.9938	−0.1358	0.9477	−0.2076	0.9790	0.068
Inoue et al. (2003)	Actinopterygian fish	28	11509	0.0430	0.0939	0.0437	0.0362	0.0129	0.1046	0.647
Jordan et al. (2003)	*Megalagrion* damselflies	68	2326	0.0271	0.1659	0.0476	0.0435	−0.0083	0.6550	0.241
Kelch and Baldwin (2002)	*Cirsium* plants	52	1303	0.1991	**0.0076**	0.1870	**0.0059**	0.0335	0.2183	0.341
Kiefer et al. (2002)	*Plecotus* bats	19	1714	−0.0399	0.7124	−0.0700	0.8868	−0.1153	0.9239	0.153
Leaché and Reeder (2002)	*Sceloporus* lizards	78	3688	0.0152	0.3071	0.0118	0.3234	−0.0242	0.8479	0.289
Lewis et al. (1997)	Liverworts	41	1428	0.2260	**0.0002**	0.1935	**0.0008**	0.0917	**0.0034**	0.538
Maddison et al. (2007)	Salticid spiders	33	4849	0.1212	0.0342	0.0459	0.2268	0.0032	0.4586	0.658
McGuire and Kiew (2001)	*Draco* lizards	57	1165	0.0746	**0.0170**	0.0689	**0.0292**	0.0441	**0.0157**	0.446
Nicholson (2002)	*Anolis* lizards	55	945	0.3233	**0.0001**	0.3055	**0.0002**	0.0167	0.0417	0.428
Reed et al. (2002)	Carangid fish	64	1140	0.1143	**0.0070**	0.1115	**0.0016**	0.0352	**0.0297**	0.420
Rokas et al. (2003)	Gallwasps	85	433	0.0987	**0.0028**	0.1184	**0.0005**	0.0003	0.5050	0.346
Ruiz-Trillo et al. (1999)	Bilateria	78	2555	0.2890	**0.0001**	0.2582	**0.0001**	0.0679	**0.0025**	0.635
Rydin and Källersjö (2002)	Seed plants	38	1428	0.0344	0.2331	0.0941	**0.0082**	0.0276	0.0563	0.442
Santos et al. (2002)	Dinoflagellates	20	1006	0.0173	0.4132	0.0251	0.3008	0.0127	0.3779	0.177
Sikes et al. (2008)	*Nicrophorus* beetles	50	2129	0.0373	0.1413	0.0482	0.1139	0.0269	0.2133	0.195
Smith et al. (2005)	Tephridid flies	44	1027	0.1249	0.0364	0.1237	**0.0172**	0.0698	0.0352	0.461
Swain and Taylor (2003)	Water fleas	21	3531	0.0442	0.2531	0.0447	0.2420	−0.0391	0.9888	0.459
Townsend et al. (2004)	Squamate reptiles	47	4621	0.0220	0.2034	−0.0194	0.7916	−0.0162	0.9437	0.577
Voigt and Wostemeyer (2001)	Zygomycete fungi	43	1660	0.0564	0.1139	0.0494	0.1030	0.0206	0.1455	0.425
Voris et al. (2002)	Homalopsine snakes	17	1490	0.2539	0.0278	0.2694	0.0344	0.0709	0.1794	0.609
Weisrock et al. (2006)	Salamandrid salamanders	96	2765	0.2154	**0.0001**	0.0293	0.0617	0.0110	0.1578	0.358
Wilcox et al. (2002)	Snakes	23	1545	0.0753	0.1625	0.0409	0.2458	0.0214	0.1385	0.522
Wu et al. (2001)	*Aleurodiscus* fungi	71	997	0.1346	**0.0055**	0.1843	**0.0002**	0.0478	0.0661	0.265

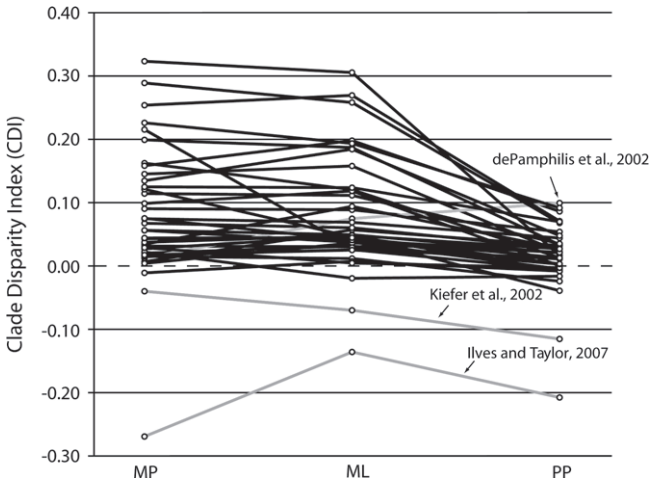

FIGURE 5. CDIs (equation 2) of MP and ML bootstrap and Bayesian analyses of the 38 empirical data sets in Table 3. Each line connects the 3 CDIs calculated from each analysis of each empirical data set. A CDI value of 0 indicates that there is no relationship between a clade's size and support. Data sets with atypical behavior (see Discussion) are indicated by the gray line and arrows.

size effect in resampling methods (but see Randle and Pickett 2006). It should be noted that the implied model of homoplasy underlying this explanation almost certainly underestimates the complexity of homoplasy in real data; for instance, long-branch attraction will result in increased support for clades that conflict with the true tree, but in a nonrandom fashion (Felsenstein 1978; Swofford et al. 2001). Nonetheless, we feel that this model is adequate to demonstrate that homoplasy is at present the best explanation for the clade size effect seen in empirical studies.

Given a data set with no conflicting characters, bootstrap resampling will construct data sets that are 100% consistent with the original data (as will jackknife methods; Pickett and Randle 2005). As these are heuristic optimization methods and there is no character conflict, the trees produced by the resampled data sets will all be mutually consistent and match the true tree, albeit with the possible addition of polytomies. In this example, the possibility of including inconsistent trees is absent and the association between clade priors and support values disappears. Because the goal of Bayesian MCMC analysis is to approximate the posterior distribution of trees, topologies that are inconsistent with the true topology are sampled independent of the occurrence of homoplasy. It is for this reason that the distribution of clade sizes on randomly drawn trees influences PP estimates when homoplasy is absent.

Pickett and Randle (2005) argued that "(1) the correlation between re-sampling support and clade priors is an artifact of noise in natural data and not due to any influence of clade priors per se", and (2) "Bayesian support values are influenced by clade priors, even when the signal from the data is homoplasy-free, exhibiting no noise" (p. 208). Although previous simulation work (Brandley et al. 2006) demonstrated that the effect of the priors disappears with the addition of even a small amount of phylogenetic signal, there could still be legitimate cause for concern if the effect of the priors amplified the effects of homoplasy on size-dependent clade support. Contrary to this, we find that the clade size effect is smallest in Bayesian analysis and that both bootstrap resampling and Bayesian methods will respond similarly in the presence of homoplasy (Table 2 and Fig. 3). We suggest that homoplasy and prior probabilities should not be considered separate explanations for clade size–dependent clade support in resampling and Bayesian methods. The inclusion of random trees is the common factor in all 3 analyses, and the source of random trees in real analyses is mostly homoplasy.

The poor performance of the ML bootstrap analyses of empirical data, despite their superior performance with simulated data, is perplexing. We speculate that Bayesian analyses' use of probabilistic distributions of evolutionary model parameters, rather than parameters that are typically fixed in the beginning of the analysis (as in ML bootstrapping), may play a factor. It is also possible that bootstrapping is more subject to the clade size effect regardless of the optimality criterion and that this pattern is obscured in the simulation study due to the simplicity of our homoplastic data. The model parameters in the ML bootstrap analyses were estimated from an initial neighbor-joining tree and were not fixed to be the same as those used to simulate the data. Nonetheless, both the simulation and the analytical strategy used the GTR + Γ model, and thus it is reasonable to expect that the parameters used in the phylogenetic analysis were at least similar to those used to simulate the data. Thus, the performance difference between the ML bootstrap analyses of the simulated and empirical data may result because the model of molecular evolution used in the analyses of simulated data fit the true process fairly well, whereas the models used in the empirical studies may fit the true process poorly.

Should We Be Concerned?

That homoplasy and clade size can influence the 2 most commonly used methods of clade support (nonparametric bootstrapping and Bayesian PP estimation) is understandably disturbing to the empirical phylogeneticist. Despite the manifestation of this phenomenon in many simulations and more than half of the tested empirical data sets, we nonetheless feel that researchers using likelihood-based methods of analysis should have little reason for concern. First, the simulation analyses presented here reveal that, at least with the ML and Bayesian analyses, the maximum disparity between support values for the small- and medium-sized clades is small (less than 10%). More importantly, the effect, as measured by CDI, becomes negligible with the addition of more data (Table 2 and Fig. 4). Thus, homoplasy alone should not transform "weak" clade support into "strong" clade support (e.g., changing the PP of ≥ 0.50 to ≥ 0.95). In a worst-case scenario, excessive homoplasy may alter a clade's PP above or below the 0.95 significance level by 1% or 2%; however, this may be a

TABLE 4. Results of Friedman tests of pairwise comparisons of the ranks of different measures of clade support with 38 empirical data sets. With each comparison, the CDI for both analyses of each data set were calculated and ranked based on their proximity to 0 (e.g., if one analysis had a CDI = 0.10 and the other, 0.20, the first analysis would receive a rank of "1" and the other, "2"). If one clade support method is less affected by the clade size effect compared with another method, its mean rank will be closer to 1.0 (i.e., the CDI of that method will be consistently closer to 0). Statistical significance was assessed using a randomization test (see text) and significant *P* values (after sequential Bonferroni correction) are in bold

Comparison	*N*	Mean rank MP	Mean rank ML	Mean rank Bayesian	*Q*	*P*
MP versus ML	37	1.500	1.500		0.000	1.000
MP versus Bayesian	37	1.842		1.158	21.362	<**0.001**
ML versus Bayesian	37		1.921	1.079	32.360	<**0.001**

greater criticism of using a strict significance cutoff than of the validity of Bayesian phylogenetics as a whole. Second, the clade size effect is most evident in clades that are already accompanied by low support. This phenomenon is illustrated in the simulation analyses, where increasing phylogenetic signal reduces the disparity of support among clades. Of course, this is also intuitively obvious, as both low overall support and the clade size effect are due to the same phenomenon—the inclusion of randomly resolved trees due to homoplasy.

Another potentially troubling result is that, in 5 empirical data sets (Table 3), the CDI is significant for MP and ML bootstrap and Bayesian analyses. This suggests that the underlying factor causing a significant relationship between clade size and support is so strong that none of the commonly used analyses of clade support can overcome it. Visual inspection of the topologies inferred from these 5 empirical studies reveals relatively weak support for the medium-sized nodes and high support for the small nodes, the exact situation from which a positive CDI value is expected. Given the complexity of empirical data, it is difficult to single out a primary causal factor for this "middle clade crisis." Some possible explanations may include biased taxon sampling and the inclusion of specimens with no sequence divergence (see Brandley et al. 2006). A biological situation that could result in a topology that is prone to the middle clade crisis is a rapid evolutionary radiation (Poe and Chubb 2004). These factors could be acting alone or in addition to homoplasy (a factor explicitly accounted for in the simulated data). Of course, the significant CDIs may simply be an indication of the different amounts of actual support in the data for nodes of different sizes. Regardless, as with the simulated data, the CDIs of the Bayesian analyses, although significant, are almost always much lower than those of the bootstrap analyses. Thus, when faced with a pessimistic situation where no method is immune to factors confounding accurate clade support values, Bayesian analyses suffer the least.

Only a single empirical data set was found in which the clade size effect was significant in the Bayesian analysis and not in the MP or ML bootstrap analyses (dePamphilis et al. 1997; Table 3). The data sets of Kiefer et al. (2002) and Ilves and Taylor (2007) are also notable because of their strongly negative CDI, and in the latter data set, this is most severe in the Bayesian analysis (Table 3 and Fig. 5). Of course, this would suggest that this data set has generally low support for small-sized clades and high support for medium-sized clades, a situation one may expect when sampling multiple, closely related individuals among multiple species. Indeed, Kiefer et al. (2002) and Ilves and Taylor (2007) sampled multiple individuals of 5 species of *Hypomesus* smelts and 4 species of *Plecotus* bats, respectively, and the resulting phylogenies (not shown, but similar to that in the original papers) reveal very little phylogenetic structure among the individuals in each species. On the other hand, the interrelationships of the species are well supported. Thus, the smallest clades have very low support, whereas the middle-sized clades are well supported. Another contributing factor to the unusually negative CDI may be due to the small size of the data sets; with few nodes the size/support relationship of each clade will greatly influence the CDI statistic.

If any result should concern researchers, it should be the relatively poor performance of MP bootstrapping in both analyses of simulated and empirical data. It is disturbing that even under generous simulation conditions in which each clade is supported by at least 6, 7, or even 8 uncontradicted synapomorphies, the inclusion of homoplasy can nonetheless differentially affect MP bootstrap proportions based on clade size. Our conclusion is that the increased magnitude of the clade size effect seen in parsimony as compared with the other methods is due to parsimony's poor performance in the presence of homoplasy.

CONCLUSIONS

The results of this study have much practical relevance. It is critical that, as we build the TOL, we assess confidence in the tree while simultaneously being confident in these assessments. As the phylogenetics community collects increasing amounts of DNA data for an ever-growing TOL and other "deep" phylogenetic analyses, we will encounter the biological extremes of homoplastic data. We have demonstrated that homoplasy may be a common underlying factor that may bias the results of both resampling and Bayesian analyses, but explicit statistical methods for determining clade confidence (ML bootstrap and Bayesian PPs) are far more robust to homoplasy than MP bootstrap analyses.

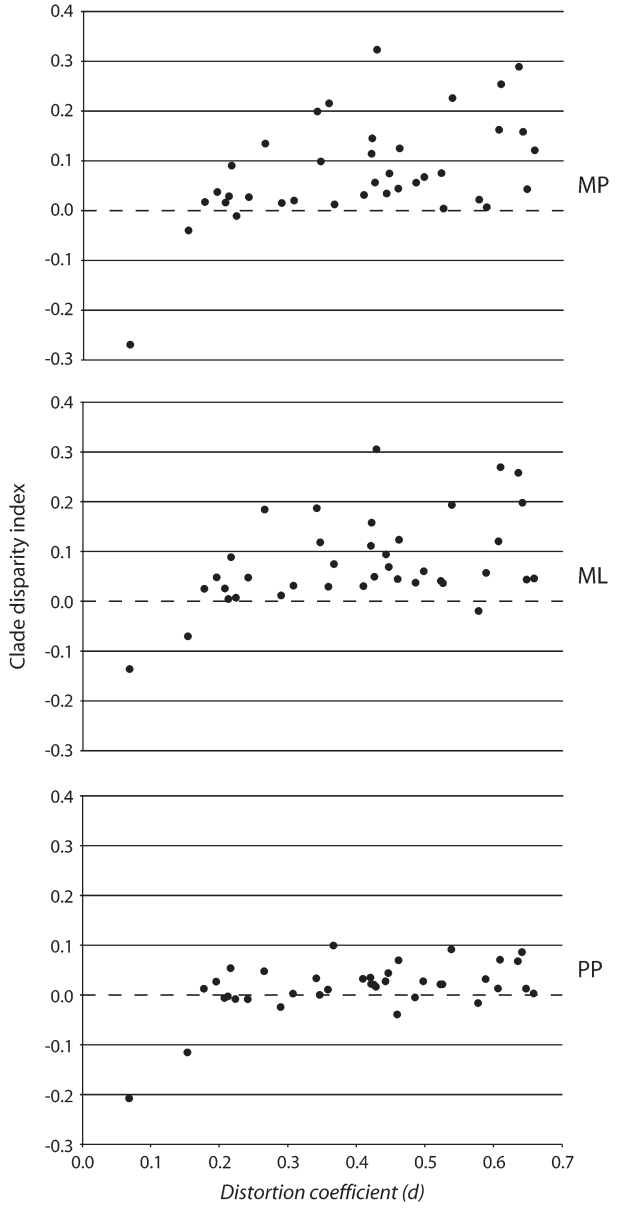

FIGURE 6. CDIs (equation 2) plotted against Farris' distortion coefficients (*d*; a measure of homoplasy) for MP and ML bootstrap and Bayesian analyses of the 38 empirical data sets in Table 3. A CDI value of 0 indicates that there is no relationship between a clade's size and support.

Perhaps more importantly, we reveal that Bayesian methods consistently infer clade support values that are less likely to be associated with clade size in both simulated and empirical data.

SUPPLEMENTARY MATERIAL

Supplementary material can be found at http://sysbio.oxfordjournals.org/.

FUNDING

This project was funded by an National Science Foundation (NSF) Doctoral Dissertation Improvement Grant (DEB 0709885) awarded to M.C.B and NSF grants DEB 0724679 and DEB 0640967 awarded to J.A.M.

ACKNOWLEDGMENTS

We thank the authors of the empirical studies who graciously supplied us with their empirical data sets and K. Pickett and C. Randle for providing data used in their original study. Ideas presented in this paper were much improved by discussion with J. Huelsenbeck, A. Seago, and M. Turelli as well as the comments of C. Ané, J. Sullivan, and 2 anonymous reviewers. We also thank D. Zwickl for bringing to our attention the omission of the clade priors formula in our previous paper and David Hillis for suggesting an explanation for the difference in the performance of ML in simulation and empirical studies.

REFERENCES

Altekar G., Dwarkadas S., Huelsenbeck J.P., Ronquist F. 2004. Parallel Metropolis-coupled Markov chain Monte Carlo for Bayesian phylogenetic inference. Bioinformatics. 20:407–415.

Anderson F.E., Cordoba A.J., Thollesson M. 2003. Bilaterian phylogeny based on analyses of a region of the sodium-potassium ATPase alpha subunit gene. J. Mol. Evol. 58:252–268.

Archie J.W. 1989. Homoplasy excess ratios: new indices for measuring levels of homoplasy in phylogenetics systematics and a critique of the consistency index. Syst. Zool. 38:253–268.

Barns S.M., Delwiche C.F., Palmer J.D., Pace N.R. 1996. Perspectives on archaeal diversity, thermophily and monophyly from environmental rRNA sequences. Proc. Natl. Acad. Sci. USA. 93: 9188–9193.

Bell C.D., Donoghue M.J. 2003. Phylogeny and biogeography of Morinaceae (Dipsacales) based on nuclear and chloroplast DNA sequences. Org. Divers. Evol. 3:227–237.

Berbee M.L., Taylor J.W. 2001. Fungal molecular evolution: gene trees and geologic time. In: McLaughlin D.J., McLaughlin E.G., Lemke P.A., editors. The mycota VIIB systematics and evolution. Berlin: Springer, pp. 229–245.

Brandley M.C., de Queiroz K. 2004. Phylogeny, ecomorphological evolution, and historical biogeography of the *Anolis cristatellus* series. Herpetol. Monogr. 18:90–126.

Brandley M.C., Leache A.D., Warren D.L., McGuire J.A. 2006. Are unequal clade priors problematic for Bayesian phylogenetics? Syst. Biol. 55:138–146.

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

Castoe T.A., Doan T.M., Parkinson C.L. 2004. Data partitions and complex models in Bayesian analysis: the phylogeny of gymnophthalmid lizards. Syst. Biol. 53:448–469.

Cox C.J., Hedderson T.A.J. 2003. Phylogenetic relationships within the moss family Bryaceae based on chloroplast DNA evidence. J. Bryol. 25:31–40.

Danforth B.N., Conway L., Ji S.Q. 2003. Phylogeny of eusocial *Lasioglossum* reveals multiple losses of eusociality within a primitively eusocial clade of bees (Hymenoptera: Halictidae). Syst. Biol. 52:23–36.

Delsuc F., Scally M., Madsen O., Stanhope M.J., de Jong W.W., Catzeflis F.M., Springer M.S., Douzery E.J.P. 2002. Molecular phylogeny of living xenarthrans and the impact of character and taxon sampling on the placental tree rooting. Mol. Biol. Evol. 19:1656–1671.

dePamphilis C.W., Young N.D., Wolfe A.D. 1997. Evolution of plastid gene rps2 in a lineage of hemiparasitic and holoparasitic plants:

many losses of photosynthesis and complex patterns of rate variation. Proc. Natl. Acad. Sci. USA. 94:7367–7372.
Des Marais D.L., Smith A.R., Britton D.M., Pryer K.M. 2003. Phylogenetic relationships and evolution of extant horsetails, *Equisetum*, based on chloroplast DNA sequence data (*rbcL* and *trnL-f*). Int. J. Plant Sci. 164:737–751.
Edwards E.J., Nyffeler R., Donoghue M.J. 2005. Basal cactus phylogeny: *Pereskia* paraphyly and implications for early cactus evolution. Am. J. Bot. 92:1177–1188.
Farris J.S. 1973. On comparing the shapes of taxonomic trees. Syst. Zool. 22:50–54.
Farris J.S. 1989. The retention index and the rescaled consistency index. Cladistics. 5:417–419.
Felsenstein J. 1978. Cases in which parsimony and compatibility methods will be positively misleading. Syst. Zool. 27:401–410.
Felsenstein J. 1985. Confidence limits on phylogenies: an approach using the bootstrap. Evolution. 39:783–791.
Felsenstein J. 2004. Inferring phylogenies. Sunderland (MA): Sinauer Associates.
Friedman M. 1937. The use of ranks to avoid the assumption of normality implicit in the analysis of variance. J. Am. Stat. Assoc. 32:675–701.
Garey J.R., Near T.J., Nonnemacher M.R., Nadler S.A. 1996. Molecular evidence for Acanthocephala as a subtaxon of Rotifera. J. Mol. Evol. 43:287–292.
Garvey J.E., Marschall E.A., Wright R.A. 1998. From star charts to stoneflies: detecting relationships in continuous bivariate data. Ecology. 79:442–447.
Hillis D.M. 1996. Inferring complex phylogenies. Nature. 383:130–131.
Holm S. 1979. A simple sequential rejective multiple test procedure. Scand. J. Stat. 6:65–70.
Huelsenbeck J.P. 1995. Performance of phylogenetic methods in simulation. Syst. Biol. 44:17–48.
Huelsenbeck J.P., Rannala B. 2004. Frequentist properties of Bayesian posterior probabilities of phylogenetic trees under simple and complex substitution models. Syst. Biol. 53:904–910.
Ilves K.L., Taylor E.B. 2007. Evolutionary and biogeographic patterns within the smelt genus *Hypomesus* (Pisces: Osmeridae) in the North Pacific Ocean. J. Biogeogr. 35:48–64.
Inoue J.G., Miya M., Tsukamoto K., Nishida M. 2003. Basal actinopterygian relationships: a mitogenomic perspective on the phylogeny of the ancient fish. Mol. Phylogenet. Evol. 26:110–120.
Jordan S., Simon C., Polhemus D. 2003. Molecular systematics and adaptive radiation of Hawaii's endemic damselfly genus *Megalagrion* (Odonata: Coenagrionidae). Syst. Biol. 52:89–109.
Kelch D.G., Baldwin B.G. 2002. Phylogeny and ecological radiation of New World thistles (*Cirsium*, Cardueae–Compositae) based on ITS and ETS rDNA sequence data. Mol. Ecol. 12:141–151.
Kiefer A., Mayer F., Kosuch J., von Helversen O., Veith A. 2002. Conflicting molecular phylogenies of European long-eared bats (*Plecotus*) can be explained by cryptic diversity. Mol. Phylogenet. Evol. 25:557–566.
Lanyon S.M. 1985. Detecting internal inconsistencies in distance data. Syst. Zool. 34:397–403.
Larget B., Simon D. 1999. Markov chain Monte Carlo algorithms for the Bayesian analysis of phylogenetic trees. Mol. Biol. Evol. 16:750–759.
Leaché A.D., Reeder T.W. 2002. Molecular systematics of the Eastern fence lizard (*Sceloporus undulatus*): a comparison of parsimony, likelihood, and Bayesian approaches. Syst. Biol. 51:44–68.
Lewis L.A., Mishler B.D., Vilgalys R. 1997. Phylogenetic relationships of the liverworts (Hepaticae), a basal embryophyte lineage, inferred from nucleotide sequence data of the chloroplast gene *rbcL*. Mol. Phylogenet. Evol. 7:377–393.
Maddison W.P., Zhang J.X., Bodner M.R. 2007 A basal phylogenetic placement for the salticid spider *Eupoa*, with descriptions of two new species (Araneae: Salticidae). Zootaxa. 1432:23–33.
Mau B., Newton M., Larget B. 1999. Bayesian phylogenetic inference via Markov chain Monte Carlo methods. Biometrics. 55:1–12.
McGuire J.A., Kiew B.-H. 2001. Phylogenetic systematics of Southeast Asian flying lizards (Iguania: Agamidae: *Draco*) as inferred from mitochondrial DNA sequence data. Biol. J. Linn. Soc. 73:203–229.
Milinkovitch M.C., LeDuc R.G., Adachi J., Farnir F., Georges M., Hasegawa M. 1996. Effects of character weighting and species sampling on phylogeny reconstruction: a case study based on DNA sequence data in cetaceans. Genetics. 144:1817–1833.
Naylor G., Kraus F. 1995. The relationship between *s* and *m* and the retention index. Syst. Biol. 44:559–562.
Nicholson K.E. 2002. Phylogenetic analysis and a test of the current infrageneric classification of *Norops* (beta *Anolis*). Herpetol. Monogr. 16:93–120.
Nylander J.A.A., Ronquist F., Huelsenbeck J.P., Nieves-Aldrey J.L. 2004. Bayesian phylogenetic analysis of combined data. Syst. Biol. 53:47–67.
Nylander J.A.A., Wilgenbusch J.C., Warren D.L., Swofford D.L. 2008. AWTY (are we there yet?): a system for graphical exploration of MCMC convergence in Bayesian phylogenetics. Bioinformatics. 24:581–583.
Pickett K.M., Randle C.P. 2005. Strange Bayes indeed: uniform topological priors imply non-uniform clade priors. Mol. Phylogenet. Evol. 34:203–211.
Poe S., Chubb A.L. 2004. Birds in a bush: five genes indicate explosive evolution of avian orders. Evolution. 58:404–415.
Pollock D.D., Zwickl D.J., McGuire J.A., Hills D.M. 2002. Increased taxon sampling is advantageous for phylogenetic inference. Syst. Biol. 51:664–671.
Rambaut A., Grassly N.C. 1997. Seq-Gen: an application for the Monte Carlo simulation of DNA sequence evolution along phylogenetic trees. Comput. Appl. Biosci. 13:235–238.
Randle C.P., Pickett K.M. 2006. Are nonuniform clade priors important in Bayesian phylogenetic analysis? A response to Brandley et al. Syst. Biol. 55:147–151.
Reed D.L., Carpenter K.E., deGravelle M.J. 2002. Molecular systematics of the jacks (Perciformes: Carangidae) based on mitochondrial cytochrome *b* sequences using parsimony, likelihood, and Bayesian approaches. Mol. Phylogenet. Evol. 23:513–524.
Rice W.R. 1989. Analyzing tables of statistical tests. Evolution. 43:223–225.
Rokas A., Melika G., Abe Y., Nieves-Aldrey J.L., Cook J.M., Stone G.N. 2003. Lifecycle closure, lineage sorting, and hybridization revealed in a phylogenetic analysis of European oak gallwasps (Hymenoptera: Cynipidae: Cynipini) using mitochondrial sequence data. Mol. Phylogenet. Evol. 26:36–45.
Ronquist F., Huelsenbeck J.P. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics. 19:1572–1574.
Ruiz-Trillo I., Riutort M., Littlewood D.T.J., Herniou E.A., Bagun J. 1999. Acoel flatworms: earliest extant bilaterian metazoans, not members of the Platyhelminthes. Science. 283:1919–1923.
Rydin C., Källersjö M. 2002. Taxon sampling and seed plant phylogeny. Cladistics. 18:485–513.
Sanderson M.J., Donoghue M.J., Piel W., Eriksson T. 1994. TreeBASE: a prototype database of phylogenetic analyses and an interactive tool for browsing the phylogeny of life. Am. J. Bot. 81:183.
Santos S.R., Taylor D.J., Kinzie R.A. III, Hidaka M., Sakai K., Coffroth M.A. 2002. Molecular phylogeny of symbiotic dinoflagellates inferred from partial chloroplast large submit (23S)-rDNA sequences. Mol. Phylogenet. Evol. 23:97–111.
Sikes D.S., Vamosi S.M., Trumbo S.T., Ricketts M., Venables C. 2008. Molecular systematics and biogeography of *Nicrophorus* in part—the *investigator* species group (Coleoptera: Silphidae) using mixture model MCMC. Mol. Phylogenet. Evol. 48:646–666.
Smith J.J., Jaycox M.A., Smith M., Bush G.L. 2005. Analysis of mitochondrial DNA and morphological characters in the subtribe Carpomyina (Diptera: Tephritidae). Isr. J. Entomol. 35–36:317–340.
Stamatakis A. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics. 22:2688–2690.
Sukumaran J., Linkem C.W. 2008. Choice of topology estimators in Bayesian phylogenetic analysis. Mol. Biol. Evol. 26:1–3.
Swain T.D., Taylor D.J. 2003. Structural rRNA characters support monophyly of raptorial limbs and paraphyly of limb specialization in water fleas. Proc. R. Soc. B. 270:887–896.
Swofford D.L. 2002. PAUP*: phylogenetic analysis using parsimony (*and other methods), version 4. Sunderland (MA): Sinauer Associates.
Swofford D.L., Waddell P.J., Huelsenbeck J.P., Foster P.G., Lewis P.O., Rogers J.S. 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. Lect. Math. Life Sci. 17:57–86.
Townsend T.M., Larson A., Louis E., Macey J.R. 2004. Molecular phylogenetics of Squamata: the position of snakes, amphisbaenians, and dibamids, and the root of the squamate tree. Syst. Biol. 53:1–23.
Voigt K., Wostemeyer J. 2001. Phylogeny and origin of 82 zygomycetes from all 54 genera of the Mucorales and Mortierellales based on combined analysis of actin and translation elongation factor EF-1 alpha genes. Gene. 270:113–120.
Voris H.K., Alfaro M.E., Karns D.R., Starnes G.L., Thompson E., Murphy J.C. 2002. Phylogenetic relationships of the Oriental-Australian rear-fanged water snakes (Colubridae: Homalopsinae) based on mitochondrial DNA sequences. Copeia. 2002: 906–915.
Weisrock D.W., Papenfuss T.J., Macey J.R., Litvinchuk S.N., Polymeni R., Ugurtas I.H., Zhao E., Larson A. 2006. A molecular assessment of phylogenetic relationships and lineage accumulation rates within the family Salamandridae (Amphibia, Caudata). Mol. Phylogenet. Evol. 41:368–383.
Wilcox T.P., Zwickl D.J., Heath T.A., Hillis D.M. 2002. Phylogenetic relationships of the dwarf boas and a comparison of Bayesian and bootstrap measures of phylogenetic support. Mol. Phylogenet. Evol. 25:361–371.
Wu S.-H., Hibbett D.S., Binder M. 2001. Phylogenetic analyses of *Aleurodiscus s.l.* and allied genera. Mycologia. 93:720–731.
Yang Z. 1993. Maximum likelihood estimation of phylogeny from DNA sequences when substitution rates differ over sites. Mol. Biol. Evol. 10:1396–1401.
Yang Z. 1994. Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: approximate methods. J. Mol. Evol. 39:306–314.
Yang Z., Rannala B. 1997. Bayesian phylogenetic inference using DNA sequences: a Markov chain Monte Carlo method. Mol. Biol. Evol. 14:717–724.
Zwickl D.J., Hillis, D.M. 2002. Increased taxon sampling greatly reduces phylogenetic error. Syst. Biol. 51:588–598.

First submitted 4 April 2008; reviews returned 11 June 2008;
final acceptance 13 January 2009
Associate Editor: Cécile Ané