Strange bayes indeed: uniform topological priors imply non-uniform clade priors

Kurt M. Pickett[a], Christopher P. Randle[b],[*],[1]

[a] Division of Invertebrate Zoology, American Museum of Natural History, 79th Street at Central Park West, New York, NY 10024, United States
[b] Department of Ecology and Evolutionary Biology and Natural History Museum, University of Kansas, 1200 Sunnyside Ave.
Lawrence, KS 66045, United States

Abstract

While Bayesian analysis has become common in phylogenetics, the effects of topological prior probabilities on tree inference have not been investigated. In Bayesian analyses, the prior probability of topologies is almost always considered equal for all possible trees, and clade support is calculated from the majority rule consensus of the approximated posterior distribution of topologies. These uniform priors on tree topologies imply non-uniform prior probabilities of clades, which are dependent on the number of taxa in a clade as well as the number of taxa in the analysis. As such, uniform topological priors do not model ignorance with respect to clades. Here, we demonstrate that Bayesian clade support, bootstrap support, and jackknife support from 17 empirical studies are significantly and positively correlated with non-uniform clade priors resulting from uniform topological priors. Further, we demonstrate that this effect disappears for bootstrap and jackknife when data sets are free from character conflict, but remains pronounced for Bayesian clade supports, regardless of tree shape. Finally, we propose the use of a Bayes factor to account for the fact that uniform topological priors do not model ignorance with respect to clade probability.
© 2004 Elsevier Inc. All rights reserved.

Keywords: Bayesian phylogenetics; Clade support; Prior probability; Jackknife; Bootstrap

1. Introduction

Over the past few years, a method of complex problem-solving known as Markov Chain Monte Carlo (MCMC) has been gaining popularity among phylogeneticists (see reviews in Holder and Lewis, 2003; Huelsenbeck et al., 2002; Lewis, 2001). MCMC itself is not new, dating from Metropolis et al. (1953), and its Bayesian character—its ability to sample a posterior distribution—is well established (Tierney, 1994). But the implementation of the MCMC algorithm as an application in phylogenetics is fairly new, originating with the doctoral dissertation work of Li (1996) and Mau (1996). Others have discussed Bayesian interpretations of phylogenetic problems (Farris, 1973; Harper, 1979; Wheeler, 1991), but these did not involve MCMC, and so we do not treat those interpretations here.

Considering only explicitly statistical methods of phylogenetics, a Bayesian approach is, in some ways, more appealing than the likelihood approach. As is well known, the likelihood of a hypothesis (here, the tree) given the data is proportional to the probability of the data given the hypothesis (see Edwards, 1992, p. 9). Phylogeneticists—whose primary investigation usually relates to tree selection—are more concerned with the probability of the tree, conditional on the model and the data (rather than the probability of

the data), and this is what Bayes' formula provides. But to accomplish this inversion (see Farris, 1973), information regarding the prior probabilities of the trees is needed. Many advocates of Bayesian phylogenetics have commented on the importance of this prior assessment of tree probabilities (Huelsenbeck et al., 2001, 2002; Lewis, 2001) and have mentioned that prior selection can be problematic (Holder and Lewis, 2003; Lewis, 2001).

Because of these issues, some questions remain open. What constitutes a reasonable topological prior probability, and how does one arrive at such a distribution? The estimation of prior probabilities is difficult when little is known about the phylogeny of a group of organisms beforehand, which is most often, or arguably *always*, the case for systematic studies.

Proponents of the new Bayesian approach to phylogeny have advocated the use of uniform topological priors to reflect ignorance (see review in Huelsenbeck et al., 2002; Lewis, 2001). When nothing is known regarding the relationships of taxa prior to analysis, all tree hypotheses are considered to be equally probable. This may be valid given the "principle of insufficient reason" (LaPlace 1820, as cited in Kass and Wasserman, 1996). For example, Farris (1973, p. 251) argued that, in a Bayesian framework, "… $P\{E\}$, the probability of evolutionary hypothesis [tree] E not conditional upon any data, may be treated as if equal for all E." While it may seem paradoxical on the one hand to claim the superiority of a method due to its ability to incorporate prior knowledge, and, on the other, to claim that ignorance should be modeled, this concern is not unique to Bayesian *phylogenetics*, and forms the kernel of the schism between the "empirical" and "subjective" schools of Bayesian statistics. This debate is beyond the scope of the present study.

Other than modeling ignorance, justifications for the use of uniform topological priors are that the likelihood function will overwhelm any information in the topological priors anyway (see review in Lewis, 2001), and that the topological prior information is unimportant to the Metropolis-Hastings algorithm (Hastings, 1970; Metropolis et al., 1953) because it is identical in both the numerator and denominator, when all topologies are considered equally probable a priori. The conditions under which the former will be true in phylogenetic analyses have not been established. The latter, however, is necessarily true when tree topology is the hypothesis being evaluated. Because of this property of the Metropolis-Hastings algorithm, and because every tree is given an equal probability a priori, uniform priors are said to model ignorance effectively. However, it has long been established that no prior can be devised that models ignorance for all hypotheses simultaneously (Franklin, 2001; Kass and Wasserman, 1996; Walley, 1996). This

applies to phylogenetics, when the hypothesis being evaluated is not the entire topology, but the presence of individual clades.

We will show that when uniform topological priors are stipulated, clade probabilities are not equal a priori. Specifically, the number of taxa in a clade, given the number of taxa in the entire analysis, affects the prior probability of that clade in a predictable way. Because of this, if the hypothesis being investigated is monophyly (i.e., the probability of the clade), uniform topological priors do not model ignorance, an undesirable property of a prior distribution when little is known a priori. While we do not argue that clade priors must be uniform, the clade priors that result from uniform topological priors are difficult to justify as reasonable in any case. Under these conditions, the claim that uniform topological priors do not influence results in a Bayesian framework is false.

2. Uniform topological priors and clade priors

Considering a pool of fully bifurcating, equiprobable, rooted trees for n taxa, the probability of a given clade of T taxa is equivalent to the probability of randomly choosing a tree containing that clade, or, considered another way, the sum of the (equal) probabilities of all trees containing that clade. Here, the probability of a clade is obtained by multiplying the number of rearrangements of that clade by the number of rearrangements of all taxa not in that clade, divided by the number of possible rooted trees for n taxa (see Eq. 1). This means that if all trees are considered equally probable, the probability of a clade is dependent on the number of taxa it contains, T, and the number of total taxa in the analysis, n.

Because a monophyletic group of T taxa is rooted, the number of rearrangements is the same as the number of rooted trees for T taxa (Felsenstein, 1978; see review in Swofford et al., 1996). This value is multiplied by the number of possible $n - T$ rearrangements (that do not compromise the monophyly of T). The denominator is simply the number of possible labeled trees for n taxa (as in Felsenstein, 1978). Therefore,

$$\frac{\left[\prod_{i=2}^{T} 2i - 3\right]\left[\prod_{i=T+1}^{n} 2i - 2T - 1\right]}{\prod_{i=2}^{n} 2i - 3}. \tag{1}$$

Eq. (1) calculates the probability of monophyly for T taxa, given that all possible rooted topologies are equally probable (see Formula 12 of Brown (1994) for a similar, independently derived formula. However, Brown's formula results in somewhat different values than those obtained here [see reported values therein]).

To demonstrate this more intuitively, consider a set of $n = 5$ taxa, A–E, for which there are 105 bifurcating, rooted trees. Considering the monophyly of $T = 3$ taxa,

A, B, and C, there are 9 of 105 trees that contain this clade; this comprises 8.6% of all possible trees. However, if considering a clade of $T = 2$ taxa, A and B, there are nearly twice as many trees containing this clade, 15 of 105 trees, or 14.3%. The probability of selecting a tree containing a clade of A and B exclusively is a priori 1.7 times that of selecting a tree containing a clade of exclusively A, B, and C.

As Fig. 1 shows, very large clades and very small clades will always be more probable a priori than middle-sized clades (if n is constant); as taxa are added, this disparity becomes more dramatic. To demonstrate this for an analysis of typical size, a clade of two taxa in a tree of 50 taxa has a prior probability of approximately 1%, while a clade containing 25 taxa has a prior probability of 2.53×10^{-15}, or about one out of 400 trillion.

In this study, we examine the relationship between clade priors under the assumption of uniform topological priors and reported Bayesian clade support in published articles. We hypothesized that, due to the extremely non-uniform nature of clade priors, the reported Bayesian clade support will be influenced in a non-uniform way. Specifically, we predicted that the clade priors would be correlated with the Bayesian clade support values. In contrast, we predicted that bootstrap

and jackknife support values deriving from both parsimony and maximum likelihood analyses will not correlate with clade priors, because uniform topological priors cannot be specified in the tree search algorithms.

3. Empirical data and clade priors

3.1. Analysis

We gathered clade support values from seventeen studies in which Bayesian clade support was reported with at least one other support measure, either bootstrap or jackknife (Table 1). We included only those values for which Bayesian support was generated from

Table 1
Empirical analyses of natural data calculating both Bayesian support values and bootstrap or jackknife values

Study	Number of clade support values provided		
	Bayesian support	Bootstrap	Jackknife
Berendzen and Dimmick (2002)	44	43[a]	
Cox and Hedderson (2003)	22	30[b]	
Danforth et al. (2003)	47	31	
Delsuc et al. (2002)	46	46	
Inoue et al. (2002)	25		25
Jordan et al. (2003)	74	74	
Kiefer et al. (2002)	15	30[c]	
Kornhall et al. (2001)	56		48
Leache and Reeder (2002)	75	154[d]	
Mateos et al. (2002)	18	55[e]	
Nicholson (2002)	37	18	
Reed et al. (2002)	47	91[f]	
Rokas et al. (2003)	51	43	
Rydin and Källersjö (2002)[g]	70		54
Simmons et al. (2004)	95		97
Voris et al. (2002)	10	7	
Wilcox et al. (2002)	20	20	
Total	752	642	224

Values for clades not appearing in all methods were included, as such resolution (or lack thereof) may be dependent on prior influence. These values were used to measure correlation of clade prior probabilities (assuming equiprobability of tree topologies, as calculated in Eq. (1)) and clade support, using the two-dimensional Kolmogorov–Smirnov test.

[a] 23 unweighted parsimony bootstrap + 20 weighted parsimony bootstrap.

[b] 14 parsimony bootstrap + 16 likelihood bootstrap.

[c] 15 likelihood bootstrap + 15 parsimony bootstrap.

[d] 51 unweighted parsimony bootstrap + 55 weighted parsimony bootstrap + 48 likelihood bootstrap.

[e] 30 parsimony bootstrap + 25 likelihood bootstrap.

[f] 29 unweighted parsimony bootstrap + 31 weighted parsimony bootstrap + 31 likelihood bootstrap.

[g] Multiple data sets were analyzed representing different taxon sampling. Only data derived from analysis of the same data sets were utilized; Bayesian support and jackknife supports derived from matrix 38:4 in Figs. 1A and C; Bayesian support and jackknife supports derived from matrix 38:2 in Figs. 1B and D.
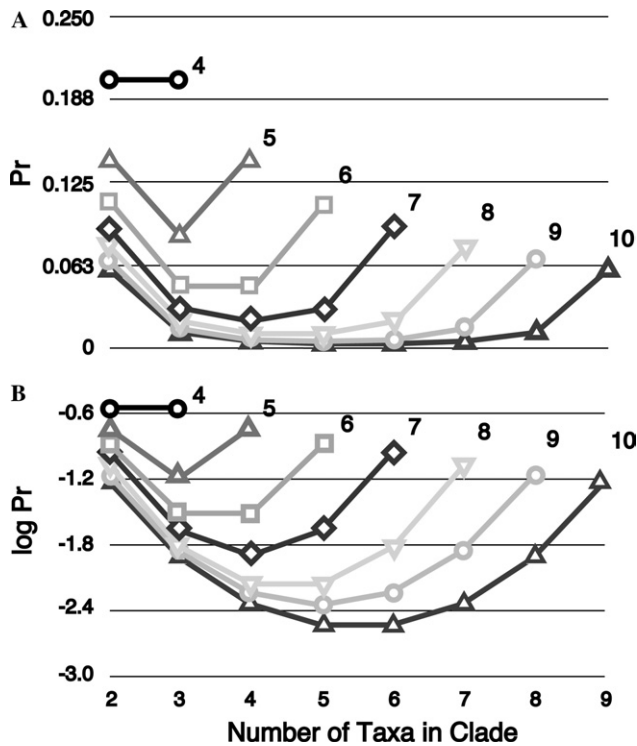

Fig. 1. (A) Clade prior probabilities calculated using Eq. (1), which assumes that all trees are equally probable. The prior probability of a clade (P_r) is dependent on both the number of taxa in the clade and the number of taxa in the tree (numbers above each plotted curve). For each curve, the prior probability of small and large clades is greater than that of intermediate clades. (B) Log transformation of the same.

identical data. A total of 752 Bayesian clade support values, 642 bootstrap support values and 224 jackknife support values were compared with clade priors calculated using Eq. (1). Linear regression is inappropriate for comparing the correlation of clade priors with support values, because, in our data, residual variance decreases as prior probability increases (Fig. 2). This violation of homoscedasticity makes regression, and other non-parametric tests that are sensitive to non-linear transforms, inappropriate. However, the two-dimensional Kolmogorov–Smirnov test makes no distribution-based assumptions and detects correlation between two distributions even when the controlling variable's influence on the response variable differs over the range of values

(Garvey et al., 1998). To implement the two-dimensional Kolmogorov–Smirnov test, we used EZ2DK-S, which compares observed data to the distribution that results from 5000 random permutations of the observed data (distributed by J. Garvey at http://www.science.siu.edu/zoology/garvey/2dks.html).

3.2. Results

All measures of support were significantly correlated with clade prior probabilities calculated under the assumption of uniform topological priors (Bayesian support: $P = 0.0048$; bootstrap: $P = 0.0058$; jackknife: $P = 0.0020$). This indicates that, for all support measures on trees of a given size, the largest clades and the smallest clades are supported most strongly, whereas medium sized clades receive lower support (see Fig. 1). We expected that this might be the case for Bayesian clade support values, because the method involved the specification of non-uniform clade priors (i.e., uniform topological priors). We did not expect such a relationship between the bootstrap or the jackknife and clade priors, because no tree priors, uniform or otherwise, are stipulated in either of the optimality criteria (parsimony and maximum likelihood) or the support calculations (bootstrap and jackknife). Indeed, we included these comparisons essentially as controls. However, the finding that both the bootstrap and jackknife, in addition to Bayesian clade support values, are significantly correlated with clade priors begs the question: Why?

4. Homoplasy-free data and clade priors

4.1. Analysis

Although others have addressed why different tree shapes (and hence different clades) may have differential support (Huelsenbeck and Kirkpatrick, 1996), and why clades will be differentially supported merely by virtue of size (Sanderson and Wojciechowski, 2000), a general explanation of these observations remains elusive. Sanderson and Wojciechowski (2000) noted a monotonic decrease in bootstrap support of the *Neoastragalus* (Fabaceae) clade as taxon sampling increased, under both parsimony and neighbor-joining algorithms. However, in their study, increasing clade size was accompanied with an increasing number of terminals in the whole tree. As we have shown in Fig. 1, clade priors are dependent not only on the number of taxa in the clade of interest, but also on the number of taxa in the tree. Upon examining their data, we found that clade prior probabilities for *Neoastragalus* (as calculated from the various topologies presented by Sanderson and Wojciechowski, 2000) also decreased monotonically with increasing clade size, simply because the number of taxa
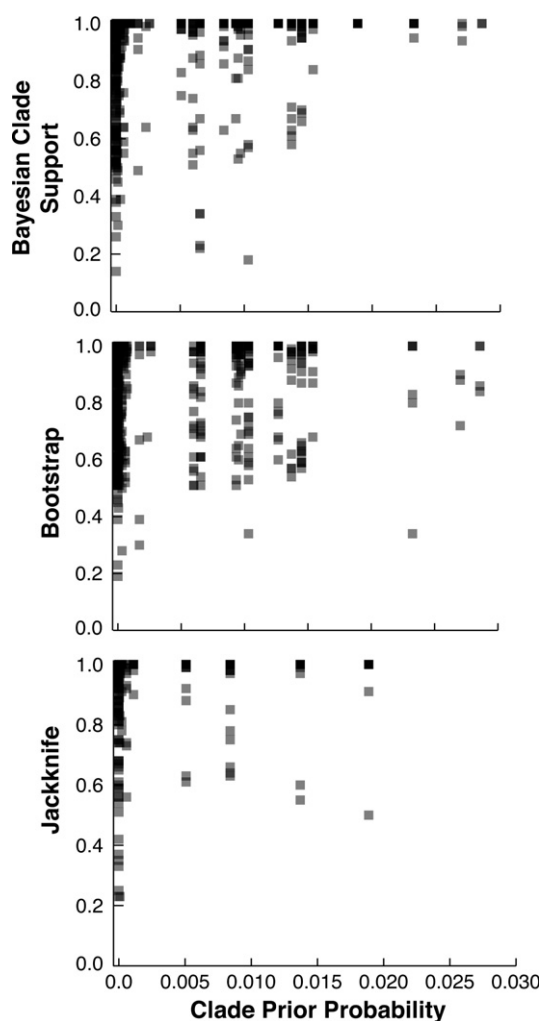

Fig. 2. Distribution of measures of nodal support (Bayesian clade support, bootstrap, and jackknife) plotted against prior clade probability, as calculated in Eq. (1). Values range from 2.75×10^{-6}–0.037. Intensity of symbols relates to the number of superimposed data points. Support data were gathered from empirical studies of natural data (summarized in Table 1). All studies utilized provided both Bayesian and at least one other type of support calculated from the same matrix of sequence data. These distributions show that variance in support decreased with increasing clade prior probability for all three measures of support.

in each analysis was increased. Each matrix was composed of approximately half *Neoastragalus* species (5/ 13, 8/19, 11/25, 15/33, 25/53, 35/73, and 79/140); clade priors for these nearly-equal ratios of T and n decrease logarithmically (Fig. 1B).

Huelsenbeck and Kirkpatrick (1996) speculated that bias in tree shape may be related to randomness of data. While that discussion related specifically to random branching tree models, they proposed that the bias may result from noise at the tips of a tree, a consequence of rate heterogeneity. An explanation of our own results and the phenomenon observed by Sanderson and Wojciechowski, which follows, is similar to this. However, our explanation does not depend on the *distribution* of noise on a tree, but *is* related to the noisy, imperfect nature of real data. Multiple hits at nucleotide sites (such that synapomorphies are overwritten) result in phylogenetic noise (sensu Wenzel and Siddall, 1999). As noise in data increases, the probability of selecting a noisy character by re-sampling (whether bootstrap or jackknife) will increase. Random characters will necessarily support random topologies. So, when re-sampled, the degree to which a matrix is composed of noisy characters will relate directly to the degree of *support* for random trees appearing across re-sampling replicates. As noise increases, group membership frequencies will approach the probabilities of clades described by Eq. (1). However, data in which every clade in a tree is supported by a single uncontradicted synapomorphy do not have any character noise, and so would not be expected to produce this relationship. Therefore, we predict that bootstrap and jackknife values will not correlate with clade priors when they are inferred from homoplasy-free data. If, however, the correlation with Bayesian clade support values persists, influence due to clade priors is the only explanation.

To perform this test, we constructed binary data matrices, all of which give rise to cladograms in which each clade is supported by a single uncontradicted synapomorphy. Simulated data cannot accomplish this test, because the entire cladogram must be free of homoplasy. To eliminate the possibility that the results are biased based on tree shape, five of the contrived matrices yielded pectinate trees, and five of the matrices yielded balanced trees. For both types, the matrices contained 20, 19, 18, 17, 16, and 15 taxa. For balanced trees, starting with the 20-taxon-case, one taxon was alternately pruned from each of the two largest clades, such that every other taxon removal restored symmetry to the cladogram. The support values from these differently sized cladograms were combined to increase the amount of nodal-support data generated (to afford power) and to gather a more complete range of clade priors. For example, a clade of 2 taxa on a tree of 20 taxa has a different clade prior (0.027027027) than a clade of two on a tree of 19 taxa (0.028571429), and so on (see above). By

staggering the taxon number, a more complete distribution of clade priors was obtained.

Each contrived matrix was subjected to Bayesian, parsimony bootstrap and parsimony jackknife analyses. For the Bayesian analyses, MrBayes 3.0B4 was used to implement the "no common mechanism" (NCM; "lset parsmodel = yes") model of Tuffley and Steel (1997). This was done for two reasons: (1) this model is much more computationally efficient than typical maximum average likelihood models that stipulate homogeneity of the underlying process, and (2) by using NCM, the optimality criterion that generated the trees themselves (i.e., parsimony) was identical for all three support measures. Of these two, the most important, of course, was ensuring that differences in optimality criteria are eliminated, and the differences among the three support measures are due solely to the methods by which the support values are calculated.

For Bayesian analyses, all the default settings were retained (see MrBayes3.0B4 for details; Huelsenbeck and Ronquist, 2001). In each case, stationarity was obtained by the 5000th generation, so the first 5000 trees were discarded as burn-in. Both bootstrap and jackknife analyses were implemented in PAUP* 4.0 beta10 (Swofford, 2002), issuing 10,000 replicates per matrix, with one random addition sequence per replicate, and TBR swapping on two trees held (these settings derive from Freudenstein et al., 2004). Additional settings for jackknife analyses included specifying 37% deletion probability and selecting the "emulate jac" option (As discussed in Freudenstein et al. (2004), these options in combination result in the deletion probability of e^{-1} set fourth in Farris et al. (1996)).

4.2. Results

Bayesian support values resulting from the contrived, homoplasy-free data matrices showed a significant, positive correlation with their clade priors, whether the clades being examined were from balanced trees (trees = 5, nodes = 87, $P = 0.0002$), pectinate trees (trees = 5, nodes = 87, $P = 0.0002$) or both groups combined (trees = 10, nodes = 174, $P = 0.0002$). However, the bootstrap support values from the homoplasy-free data did not correlate significantly with clade priors in any of these cases (balanced: trees = 5, nodes = 87, $P = 0.268$; pectinate: trees = 5, nodes = 87, $P = 0.319$; combined: $P = 0.100$). Similarly, the jackknife support values showed no relationship with clade priors regardless of tree shape or taxon number (balanced: trees = 5, nodes = 87, $P = 0.538$; pectinate: trees = 5, nodes = 87, $P = 0.446$; combined: $P = 0.525$). As expected from theory (Farris et al., 1996), jackknife values were almost all 63%, varying from 62 to 64%, regardless of clade size (see Fig. 3A). Bootstrap values were slightly higher than
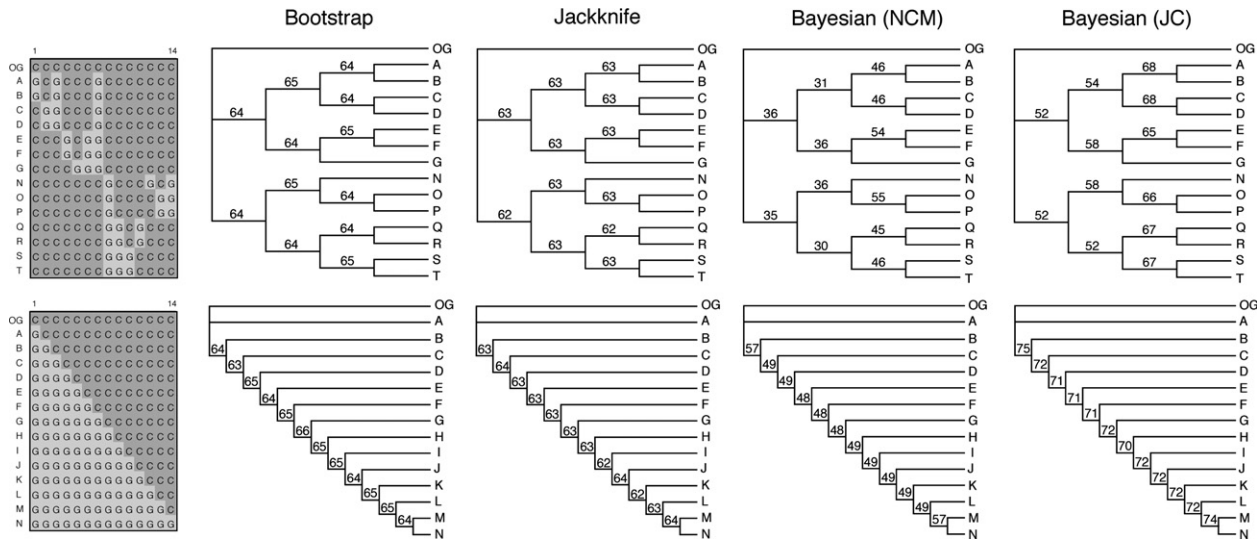
Fig. 3. Contrived matrices and their phylogenetic results, for the 15-taxon case. Balanced (top row) and pectinate (bottom row) matrices were analyzed using the parsimony criterion; in the case of the Bayesian analysis (second column), this was accomplished by implementing the "no common mechanism" model of Tuffley and Steel (1997). Every node in the analyses is supported by a single uncontradicted synapomorphy. Bootstrap and Jackknife support values are essentially consistent regardless of clade size, showing no relationship with calculated clade priors. Bayesian clade support values, on the other hand, show a significant, positive relationship with clade priors (two-dimensional Kolmogorov–Smirnov test: $P < 0.0002$; see text for further details). This influence is present under Bayesian analysis employing Jukes-Cantor. This bias in Bayesian analyses is consistent with bias introduced from the stipulation of uniform topological priors alone.

jackknife values, varying from 63 to 65% regardless of clade size (see Fig. 3A), which is not outside theoretical expectation (Efron et al., 1996). But unlike the support from re-sampling methods, Bayesian clade support values did vary according to clade size; specifically, for pectinate trees the largest ($T = n - 2$) and smallest ($T = 2$) clades were *always* supported by 57%, regardless of the number of the taxa in the tree. Support for intermediate clades on Bayesian trees varied from 48 to 49% (see Fig. 3A). Similarly, for balanced trees, the smallest clades ($T = 2$) were supported by 45–46% when the clade was sister to a clade of two taxa; but when the smallest clades were sister to a single taxon, the support varied from 54 to 55%. This bias was never present for bootstrap or jackknife values. Because there are no $n - 2$ sized clades on balanced trees, we cannot report those values. However, intermediate-sized clades on balanced Bayesian trees consistently showed the lowest support, varying from 29 to 36%.

To address the suggestion that our results may merely be the result of the interface of the Bayesian method and what is, compared to typical ML, a peculiar model (NCM), we have conducted some additional analyses. The Bayesian analyses were conducted again, this time using the homogeneous ML model chosen according to the hierarchic likelihood ratio test, as implemented in ModelTest 3.5 (Posada and Crandall, 1998). Bayesian support values resulting from analysis under that model, Jukes-Cantor, also showed significant correlation with clade priors (balanced: trees = 5, nodes = 87, $P = 0.0002$; pectinate: trees = 5, nodes = 87,

$P = 0.0026$; combined: $P = 0.0002$), just as when NCM was used.

These results support our hypotheses 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. However, it should be noted that despite the measurable influence of clade priors on Bayesian clade support values, the correct trees were always inferred by Bayesian analysis, but some two-taxon clades received higher support (see Fig. 3; clades A + B) than did others (see Fig. 3; clades E + F). Also, the pectinate trees under NCM seem to be less influenced by the priors (always giving 57% for small and large clades, but virtually flat values for intermediate sized clades) than the pectinate Jukes-Cantor support values, which follow the clade prior distribution more closely (see Fig. 3). These data suggest a complex interaction between the likelihood function, clade priors and data.

5. Discussion

The new Bayesian approach to phylogenetics offers at least two potential speed advantages over Maximum Likelihood methods. The first is that MCMC searches much more of the parameter space than do ML heuristics, and if the chain is run sufficiently long, the poster-

ior distribution of trees results. Another time-saving advantage of MCMC is that it computes support during tree search. As is well known, traditional re-sampling support measures, like the bootstrap, are extremely time-consuming, especially under ML. Today's standards of taxon sampling and sequence length may prohibit thorough bootstrapping. MCMC has been advocated as a method that allows for the substitution of clade frequency across the distribution of trees for laborious bootstrapping (Larget and Simon, 1999). However, this second time-saving step only makes sense if the clade priors adequately model ignorance, or if they incorporate prior knowledge. We have shown here that, when all trees are treated as equally probable, a priori, Bayesian clade support distributions are affected by priors that fail to model ignorance accurately, and the distribution of clade priors thus implied is not a reasonable statement of a priori knowledge. Few, if any, systematists believe a priori that the probability of monophyly has anything to do with the number of taxa hypothesized to be monophyletic. Certainly, the prior assertion that small clades and large clades are more probable than mid-sized clades lacks biological relevance. As such, a return to optimality per se is warranted.

If the goal of Bayesian phylogenetics is to calculate the posterior probability of trees, the tree(s) of most interest from that distribution should be the most probable one(s). By choosing the most probable tree(s) and using that as the optimality criterion, the Maximum A posteriori Probability (MAP) tree would be approximated; this was the initial suggestion of Rannala and Yang (1996).

However, researchers may not want to abandon Bayesian support values altogether. To determine how much the evidence has improved prior confidence in clade hypotheses. A Bayes factor can be calculated as in Huelsenbeck and Imennov (2002, p. 161):

$$\text{Bayes factor} = \frac{\text{posterior}/(1 - \text{posterior})}{\text{prior}/(1 - \text{prior})}.$$

Consider an analysis of ten terminals. A clade of two taxa has a prior probability of 0.0584 and a clade of five taxa has a prior probability of 0.0029. If both clades receive a Bayesian support value of 0.70, the Bayes factors are 24.0 and 808.1, respectively. In this case we should have 808.1/24.0 or approximately 34 times more confidence in the clade of five than in the clade of two. As a caveat, clade priors are rather close to zero, even for small data sets, resulting in generally high Bayes factors. For a given tree, Bayes factors could be ranked, or otherwise normalized to indicate relative support among clades of different sizes. Such normalization will not necessarily render the results the same as traditional Maximum Likelihood, as the tree upon which they will be

reported could be the 50% majority-rule consensus of the posterior distribution of trees, and not the MAP tree.

At a minimum, researchers should be aware that Bayesian clade support values are only comparable for clades of equal size on a given tree (or clades of equal size across trees of equal size). Indeed, this warning applies to the bootstrap and jackknife with equal force. However, as we have shown, this caveat applies to Bayes even when data are free of character conflict. This is because Bayesian clade values are uniquely afflicted by an influence that derives from the stipulation of uniform tree priors (unequal clade priors).

This sheds light on the disparity between data and theory, which suggests bootstrap support and Bayesian support should be comparable (Larget and Simon, 1999). As in Efron et al. (1996, p. 13431), bootstrap support can be thought of as Bayesian support "if we begin with an 'uninformative' prior density…" However, the prior density on clades cannot be uninformative when tree priors are. This may explain why a growing body of evidence suggests no general correspondence between re-sampling support and Bayesian support (Alfaro et al., 2003; Cummings et al., 2003; Douady et al., 2003; Erixon et al., 2003; Simmons et al., 2004; Suzuki et al., 2002; Wilcox et al., 2002).

Acknowledgments

References

Alfaro, M.E., Zoller, S., Lutzoni, F., 2003. Bayes or bootstrap? A simulation study comparing the performance of Bayesian Markov Chain Monte Carlo sampling and bootstrapping in assessing phylogenetic confidence. Mol. Biol. Evol. 20, 255–266.

Berendzen, P.B., Dimmick, W.W., 2002. Phylogenetic relationships of Pleuronectiformes based on molecular evidence. Copeia 3, 642–652.

Brown, J.K.M., 1994. Probabilities of evolutionary trees. Syst. Biol. 43, 78–91.

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.

Cummings, M.P., Handley, S.A., Myers, D.S., Reed, D.L., Rokas, A., Winka, K., 2003. Comparing bootstrap and posterior probability values in the four-taxon case. Syst. Biol. 52, 477–487.

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.

Douady, C.J., Delsuc, F., Boucher, Y., Doolittle, W.F., Douzery, E.J.P., 2003. Comparison of Bayesian and maximum likelihood bootstrap measures of phylogenetic reliability. Mol. Biol. Evol. 20, 248–254.

Edwards, A.W.F., 1992. Likelihood. Johns Hopkins University Press, Baltimore, MD.

Efron, B., Halloran, E., Holmes, S., 1996. Bootstrap confidence levels for phylogenetic trees. Proc. Natl. Acad. Sci. USA 93, 13429–13434.

Erixon, P., Svennblad, B., Britton, T., Oxelman, B., 2003. Reliability of Bayesian posterior probabilities and bootstrap frequencies in phylogenetics. Syst. Biol. 52, 665–673.

Farris, J.S., 1973. A probability model for inferring evolutionary trees. Syst. Zool. 22, 250–256.

Farris, J.S., Albert, V.A., Källersjö, M., Lipscomb, D., Kluge, A., 1996. Parsimony jackknifing outperforms neighbor-joining. Cladistics 12, 99–124.

Felsenstein, J., 1978. The number of evolutionary trees. Syst. Zool. 27, 27–33.

Franklin, J., 2001. Resurrecting logical probability. Erkenntnis 55, 277–305.

Freudenstein, J.V., van den Berg, C., Chase, M.W., 2004. An expanded plastid DNA phylogeny of Orchidaceae and analysis of jackknife branch support strategy. Am. J. Bot. 91, 149–157.

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.

Harper, C.W., 1979. A Bayesian probability view of phylogenetic systematics. Syst. Zool. 28, 547–553.

Hastings, W.K., 1970. Monte Carlo sampling methods using Markov chains and their applications. Biometrika 57, 97–109.

Holder, M., Lewis, P.O., 2003. Phylogeny estimation: traditional and Bayesian approaches. Nat. Rev. Genet. 4, 275–284.

Huelsenbeck, J.P., Imennov, N.S., 2002. Geographic origin of human mitochondrial DNA: accommodating phylogenetic uncertainty and model comparison. Syst. Biol. 51, 152–163.

Huelsenbeck, J.P., Kirkpatrick, M., 1996. Do phylogenetic methods produce trees with biased shapes?. Evolution 50, 1418–1424.

Huelsenbeck, J.P., Larget, B., Miller, R.E., Ronquist, F., 2002. Potential applications and pitfalls of Bayesian inference of phylogeny. Syst. Biol. 51, 673–688.

Huelsenbeck, J.P., Ronquist, F., 2001. MrBayes: Bayesian inference of phylogenetic trees. Biometrics 17, 754–755.

Huelsenbeck, J.P., Ronquist, F., Nielsen, R., Bollback, J.P., 2001. Bayesian inference of phylogeny and its impact on evolutionary biology. Science 294, 2310–2314.

Inoue, J.G., Miya, M., Tsukamoto, K., Nishida, M., 2002. Basal actinopterygian relationships: a mitogenomic perspective on the phylogeny of the "ancient fish". Integr. Comp. Biol. 42, 1249–1249.

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.

Kass, R.E., Wasserman, L., 1996. The selection of prior distributions by formal rules. J. Am. Stat. Assoc. 91, 1343–1370.

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.

Kornhall, P., Heidari, N., Bremer, B., 2001. Selagineae and Manuleeae, two tribes or one? Phylogenetic studies in the Scrophulariaceae. Plant Syst. Evol. 228, 199–218.

Larget, B., Simon, D.L., 1999. Markov chain Monte Carlo algorithms for the Bayesian analysis of phylogenetic trees. Mol. Biol. Evol. 16, 750–759.

Leache, 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, P.O., 2001. Phylogenetic systematics turns over a new leaf. Trends Ecol. Evol. 16, 30–37.

Li, S., 1996. Phylogenetic reconstruction using Markov chain Monte Carlo. Ph.D. Dissertation, The Ohio State University, Columbus.

Mateos, M., Sanjur, O.I., Vrijenhoek, R.C., 2002. Historical biogeography of the livebearing fish genus *Poeciliopsis* (Poeciliidae:Cyprinodontiformes). Evolution 56, 972–984.

Mau, B., 1996. Bayesian phylogenetic inference via Markov chain Monte Carlo. Ph.D. Dissertation, University of Wisconsin-Madison, Madison.

Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H., Teller, E., 1953. Equations of state calculations by fast computing machines. J. Chem. Phys. 21, 1087–1092.

Nicholson, K.E., 2002. Phylogenetic analysis and a test of the current infrageneric classification of Norops (beta Anolis). Herpetol. Monogr. 16, 93–120.

Posada, D., Crandall, K.A., 1998. Modeltest: testing the model of DNA substitution. Bioinformatics 14, 817–818.

Rannala, B., Yang, Z.H., 1996. Probability distribution of molecular evolutionary trees: a new method of phylogenetic inference. J. Mol. Evol. 43, 304–311.

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.

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.

Rydin, C., Källersjö, M., 2002. Taxon sampling and seed plant phylogeny. Cladistics 18, 485–513.

Sanderson, M.J., Wojciechowski, M.F., 2000. Improved bootstrap confidence limits in large-scale phylogenies, with and example from *Neoastragalus* (Leguminosae). Syst. Biol. 49, 671–685.

Simmons, M.P., Pickett, K.M., Miya, M., 2004. How meaningful are Bayesian support values?. Mol. Biol. Evol. 21, 188–199.

Suzuki, Y., Glazko, G.V., Nei, M., 2002. Overcredibility of molecular phylogenies obtained by Bayesian phylogenetics. Proc. Natl. Acad. Sci. USA 99, 16138–16143.

Swofford, D.L., 2002. PAUP*: phylogenetic analysis using parsimony (* and other methods), version 4.0. Sinauer and Associates.

Swofford, D.L., Olsen, G.J., Waddell, P.J., Hillis, D.M., 1996. Phylogenetic inference. In: Hillis, D.M., Moritz, C., Mable, B.K (Eds.), Molecular Systematics. Sinauer Associates, Sunderland, MA, pp. 407–514.

Tierney, L., 1994. Markov chains for exploring posterior distributions. Ann. Stat. 4, 1701–1762.

Tuffley, C., Steel, M., 1997. Links between maximum likelihood and maximum parsimony under a simple model of site substitution. B. Math. Biol. 59, 1–18.

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 4, 906–915.

Walley, P., 1996. Inferences from multinomial data: learning about a bag of marbles. J. Roy. Stat. Soc. B Met. 58, 3–34.

Wenzel, J.W., Siddall, M.E., 1999. Noise. Cladistics 15, 51–64.

Wheeler, W.C., 1991. Congruence among data sets: a Bayesian approach. In: Miyamoto, M.M., Cracraft, J. (Eds.), Phylogenetic Analysis of DNA Sequences. Oxford University Press, Oxford, pp. 334–346.

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.