Points of View

Syst. Biol. 55(1):138–146, 2006

ISSN: 1063-5157 print / 1076-836X online
DOI: 10.1080/10635150500431155

Are Unequal Clade Priors Problematic for Bayesian Phylogenetics?

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

[1]*Museum of Vertebrate Zoology and Department of Integrative Biology, University of California, Berkeley, California 94720-3160, USA; E-mail: brandley@berkeley.edu (M.C.B)*
[2]*Center for Population Biology, Section of Evolution and Ecology, University of California, Davis, California 95616-5270, USA*

Although Bayesian phylogenetic methodologies were first developed in the 1960s (Felsenstein, 1968, 2004), the approach remained relatively obscure until the initial release of the software application MrBayes (Huelsenbeck and Ronquist, 2001). Since that time, the popularity of Bayesian phylogenetics has increased tremendously, and it now must be considered a primary method of analysis on par with maximum likelihood, parsimony, and distance methods. The popularity of Bayesian analysis can be attributed to computational efficiencies that allow for explicit model-based analyses of large data sets in real time with simultaneous estimation of nodal support in the form of posterior probability values. Despite the initial enthusiasm generated by the availability of a fast likelihood-based approach, Bayesian phylogenetic analysis remains somewhat controversial. Much of the controversy is focused on two related issues: (1) the relationship between posterior probability values and nonparametric bootstrap proportions with the nagging suspicion that posterior probabilities are too liberal (e.g., Suzuki et al., 2002), and (2) the influence of prior probabilities, especially so-called flat or uninformative priors, on resulting Bayesian posteriors (Felsenstein, 2004; Zwickl and Holder, 2004; Pickett and Randle, 2005). Although there has been a spate of simulation studies published during the past 2 years, most (Alfaro et al., 2003; Cummings et al. 2003; Douady et al., 2003; Erixon et al., 2003; Huelsenbeck and Rannala, 2004; Wilcox et al., 2002) have focused on the relationship between posterior probabilities and bootstrap proportions. The relative impact of priors on posteriors has only recently received the detailed study that is required to determine if current Bayesian implementations are appropriate and, if not, how they might be corrected (e.g., Zwickl and Holder, 2004; Lewis et al., 2005).

Bayesian phylogenetic analysis requires the designation of prior probabilities for each parameter in the analysis including those for alternative tree topologies, branch lengths, and the nucleotide substitution model. In each case, we usually have little a priori information that would allow us to select an appropriate informative prior distribution, thus researchers generally attempt to accommodate their ignorance by applying uninformative priors. Because the posterior probability is proportional to the product of the prior probability and the likelihood, a truly uninformative prior should allow the likelihood function to drive the outcome of the analysis (Huelsenbeck et al., 2002; Lewis, 2001a; Zwickl and Holder, 2004). Unfortunately, the designation of truly uninformative priors is notoriously difficult (see Kass and Wasserman, 1996; Zwickl and Holder, 2004), and advocates proceed with the hope that the likelihood will overwhelm inappropriately informative priors when they cannot be avoided. The viability of Bayesian phylogenetics may depend on inferences being robust to these unavoidably informative priors.

In a recent article, Pickett and Randle (2005; hereafter referred to as "PR" for the sake of brevity) provide one of the first investigations of the relationship between prior and posterior probabilities for Bayesian phylogenetic analysis when applying inappropriately informative priors (see also Zwickl and Holder, 2004). They correctly recognized that the designation of uninformative priors on the *tree topology* does not result in uninformative *clade* priors (we note that the prior probability distribution of clades can be viewed either as the joint distribution over all splits, or as the marginal prior distribution for each individual split. Here we are concerned with the former interpretation). This point was clearly illustrated by PR with a simple example—if one considers a fully bifurcating five-taxon tree, there are 15 reconstructions linking each possible pair of taxa and only 9 reconstructions linking any combination of three taxa. Thus, with rooted trees, the prior probability of larger and smaller clades will be greater than those on clades of intermediate size. All else being equal, the posterior probabilities of smaller and larger clades should be inflated relative to those of clades of intermediate size. PR presented two examples of this phenomenon by analyzing both empirical DNA and contrived data sets. We first focus on the contrived data because we believe these are the only results in the PR study that clearly indicate that informative

clade priors could bias the outcome of Bayesian phylogenetic analysis, and our consequent concern that some in the phylogenetics community may misinterpret the results of these analyses. Following our reassessment of the contrived data, we comment on their analysis of the empirical data sets.

Pickett and Randle's contrived data analysis demonstrated the effect of unequal clade priors by contriving 15- to 20-taxon data sets with single, unambiguous synapomorphies supporting each clade (along with two parsimony-uninformative characters) and analyzing these data using MrBayes, a program whose default settings use a uniform tree topology prior. Each contrived matrix was composed of an equal number of Gs and Cs, with no As or Ts included (Fig. 1a). In line with their expectations, the largest and smallest clades had higher posterior probabilities than did clades of intermediate size despite the same amount of data (one synapomorphy) supporting each clade. These results were obtained regardless of whether the Bayesian analysis was conducted using the "no common mechanism" model (Tuffley and Steel, 1997; hereafter NCM) or the F81 model (Felsenstein, 1981)—although PR believed they had applied the Jukes-Cantor model (see below). Nonparametric bootstrap and jackknife analyses under the parsimony criterion exhibited no such bias.

(a)

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

(b)

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

FIGURE 1. (a) The 15-taxon data matrix used by Pickett and Randle (2005) in their contrived data analysis; (b) data matrix with the same number of apomorphies as the above matrix, but with four different nucleotides instead of two.

PR's results suggest that unequal clade priors could be problematic for Bayesian phylogenetic analysis. Indeed, PR interpret their results to have rather dire consequences for the current interpretation of Bayesian posterior probabilities. For example, they conclude that "a return to optimality per se is warranted," and suggest that calculation of Bayes factors might allow unequal clade priors to be accounted for in such a way as to prevent researchers from having to "abandon Bayesian support values altogether." Given these rather strong statements, we decided to investigate this problem further. Our objective was to confirm PR's results and then to determine whether Bayesian posteriors are strongly influenced by the priors when analyzing more realistic data sets (i.e., those with more than one informative character per node and with all four nucleotides present in the data). Along the way, we discovered that the relationship inferred by PR is real, but can only be detected under very specific circumstances unlikely to be met with in analyses of empirical data (throwing into question PR's interpretation of their results obtained in the reanalysis of published empirical studies; see below). Indeed, had PR either used the JC model in their analyses as they intended or constructed their contrived data sets from all four nucleotides (instead of just Gs and Cs), the fairly substantial percentage differences in posterior probabilities detected in their analyses would not have been observed.

CONTRIVED DATA ANALYSES

Reanalysis of the PR Contrived Data Set

In our initial attempt to replicate PR's results, we discovered that their JC model Bayesian analyses were in fact run using the F81 model. This occurred because they did not modify the MrBayes 3.0b4 default settings, thus allowing unequal base frequencies in their analyses (the JC and F81 models differ only in that JC imposes equal base frequencies). We confirmed this by analyzing their contrived data set (Fig. 1a) under both models; our JC results disagreed with PR's findings, whereas our F81 results are essentially identical (see below). However, given that the contrived data sets were composed only of Gs and Cs, it is in fact more appropriate to analyze the contrived data using the F81 model; thus the PR results, though incorrectly labeled, are the relevant ones for evaluation of the question at hand. We performed all of our Bayesian analyses of the contrived data under the two models most relevant to PR's conclusions: NCM and

F81. We also present results from analyses implementing the JC model in order to illustrate that PR's results are in fact based on analyses under the F81 model, as well as how PR would not have obtained such substantial differences in clade posterior probabilities had they analyzed their data using the JC model.

For each model, we conducted 100 Bayesian analyses initiated from different random starting trees. Each analysis consisted of 10^6 generations sampled every 100 generations, four chains with default heating values (T = 0.2), and the following priors: topologies = uniform, branch lengths = unconstrained:exponential (10.0), and state frequencies = dirichlet (except in analyses in which state frequencies were fixed to particular values). Stationarity was assessed using the program Converge (Warren et al., 2003), which tracks the cumulative posterior probabilities of each clade. Stationarity was assumed when the clade posterior probabilities no longer changed over time. As it would be impractical to assess convergence for all of the many analyses in this paper, we instead examined 5 to 10 analyses per model. All of these analyses reached stationarity before 200,000 generations; thus, we calculated posterior probabilities from the last 800,000 generations of each analysis. We then calculated a grand mean of the posterior probability for each clade from the individual posterior probabilities estimated from each set of 100 analyses. We use the tree in Figure 2 as a reference for all subsequent discussion of individual clade posterior probabilities. To determine if estimated posterior probabilities were correlated with a clade's prior probability, we used the same statistical test as PR, the two-dimensional Kolmogorov-Smirnov test (using the program EZ2DK-S available from J. Garvey at http://www.science.siu.edu/zoology/garvey/2dks.html).

In Table 1, we present the results of our reanalyses of the PR contrived data set. The NCM results are consistent with those of PR, with the largest and smallest clades receiving substantially greater posterior probabilities than internal clades ($P = 0.040$). Analyses based on the F81 model (the model actually employed by PR) also suggest that the posteriors on the smallest and largest clades are greater than those of clades of intermediate size ($P = 0.001$). Interestingly, analyses applying the JC model find only a 0.6% maximum difference between the highest and lowest nodal posterior probabilities, although the correlation between the prior and posterior probabilities is again significant ($P = 0.021$; Fig. 3). At first glance, it would appear that the JC model is performing better than the F81 model in these analyses despite the fact that the F81 model utilizes estimated base frequencies that are much closer to reality (A = 0, C = 0.50, G = 0.50, T = 0) than does the JC model (equal base frequencies). Why might this be the case?

To further explore this unexpected result, we first tested for base composition effects by constructing and analyzing a data set with the same number of apomorphies as the data set used by PR, but replacing some of the Gs and Cs with As and Ts (Fig. 1b). Following this modification, the JC and F81 results were almost identical (Table 1), which is unsurprising given that the base frequencies in this modified version of the contrived data set are more similar, thus better matching the expectations of the JC model. The discrepancy between the

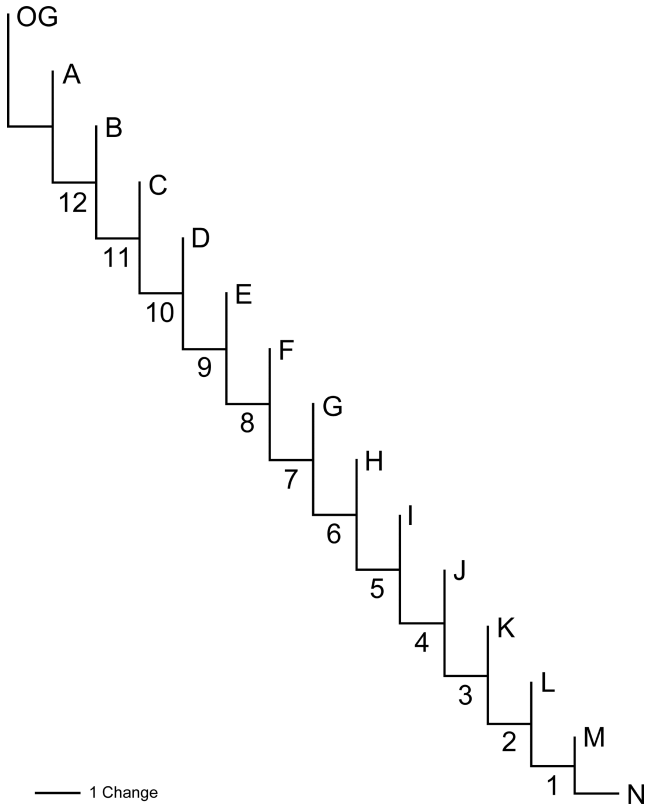

FIGURE 2. Reference phylogram for all analyses. Numbers above clades represent labels used in Table 1. Relative branch lengths represent those from a phylogenetic analysis of one of Pickett and Randle's (2005) contrived data sets (Fig. 1a).

TABLE 1. Results of the Bayesian analyses of Picket and Randle's (2005) contrived data set containing only Gs and Cs (Fig. 1a) versus a data set containing all four nucleotides (Fig. 1b). Grand means are calculated from posterior probabilities (sampled at stationarity) from each of the 100 individual analyses, per model. Results of the two-dimensional Kolmogorov-Smirnov test are shown at the bottom of the table; P values ≤ 0.05 are shown in bold. Clade numbers refer to Figure 2. Partition size refers to the number of taxa in each group at that particular split and is the same for all tables.

		Gs and Cs only			Equal base frequencies		
Clade	Partition size	NCM	JC	F81	NCM	JC	F81
1	2, 13	0.569	0.896	0.750	0.568	0.896	0.895
2	3, 12	0.485	0.892	0.725	0.485	0.892	0.890
3	4, 11	0.483	0.891	0.718	0.483	0.891	0.888
4	5, 10	0.488	0.891	0.715	0.488	0.891	0.888
5	6, 9	0.491	0.891	0.715	0.490	0.891	0.888
6	7, 8	0.491	0.891	0.714	0.491	0.891	0.889
7	8, 7	0.491	0.891	0.713	0.491	0.891	0.889
8	9, 6	0.490	0.891	0.715	0.489	0.891	0.887
9	10, 5	0.487	0.891	0.716	0.487	0.891	0.888
10	11, 4	0.483	0.891	0.720	0.483	0.892	0.889
11	12, 3	0.485	0.893	0.726	0.486	0.894	0.890
12	13, 2	0.569	0.896	0.751	0.568	0.897	0.895
		$P = 0.040$	**0.021**	**0.001**	**0.042**	**0.020**	**0.042**

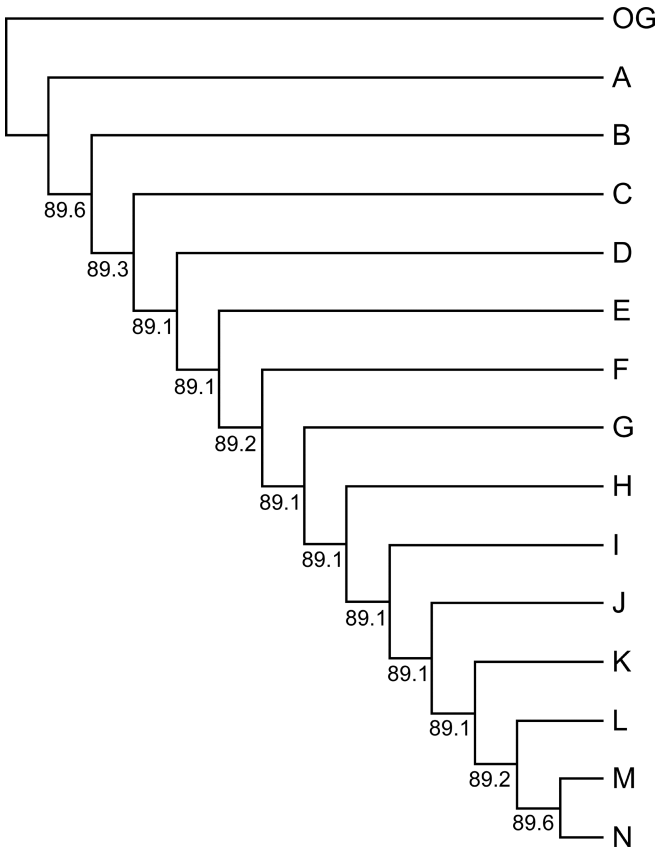

FIGURE 3. Results of the Bayesian analysis of Pickett and Randle's (2005) contrived data set (Fig. 1a) using the JC model. Numbers below clades represent grand means of estimated posterior probabilities from all 100 separate analyses.

TABLE 2. Results of Bayesian analyses of Picket and Randle's (2005) contrived data set composed only of G and C nucleotides (Fig. 1a) assuming three different sets of nucleotide frequencies in the model. Grand means are calculated from posterior probabilities (sampled at stationarity) from each of the 100 individual analyses, per model.

	Nucleotide frequencies		
Clade	$\pi_{A,T} = 0.01$, $\pi_{C,G} = 0.49$	$\pi_{A,T,C,G} = 0.25$	$\pi_{A,T} = 0.49$, $\pi_{C,G} = 0.01$
1	0.743	0.896	0.996
2	0.718	0.892	0.995
3	0.710	0.891	0.996
4	0.707	0.891	0.996
5	0.707	0.891	0.996
6	0.706	0.891	0.996
7	0.706	0.891	0.996
8	0.705	0.891	0.996
9	0.707	0.891	0.996
10	0.712	0.891	0.996
11	0.720	0.893	0.996
12	0.745	0.896	0.996

original JC and F81 results may be explained by the manner in which base frequency information is accommodated in the calculation of the likelihood. In analyses invoking the JC model (with base frequencies fixed at 25% for each nucleotide), the likelihood calculation is based on the assumption that every substitution has equal information content. If the empirical base frequencies are skewed, this information is not accommodated in the likelihood calculation and has no influence on the likelihood score, inferred branch lengths, or posterior probability values. If the F81 model is invoked, the base frequencies are an important parameter in the model and skewed empirical frequencies *will* influence the likelihood score, inferred branch lengths, and posterior probability values (Table 2). When particular nucleotides are expected to be abundant in the data set (such as the Gs and Cs in the PR contrived data sets), then substitutions between these states are expected to be relatively frequent and therefore less informative than substitutions between rare character states. Thus, F81 analyses of the PR contrived data result in reduced posterior probability values relative to those obtained under JC (Table 1). However, if the base frequencies are fixed in the model so that Gs and Cs are expected to be rare (e.g., A = 0.49, C = 0.01, G = 0.01, T = 0.49), then G $\leftrightarrow$ C transformations are assumed to be unlikely and therefore more informative. Under these conditions, the posterior probability values approach 1.0 at each node (see Table 2). Because the analyses employing the F81 model utilize the correct base frequencies and thus model the data more appropriately, we interpret the F81 results to provide a more accurate assessment of branch support. These results, though of limited relevance to empirical studies (which are unlikely to involve data comprised of only two bases, or to analyze such data using the JC model), may nevertheless have important implications for analyses of simulated data beyond those of Pickett and Randle (2005). For example, Goloboff and Pol (2005) recently criticized Bayesian methods, in part, by analyzing contrived data sets comprised of only two bases (in their case, As and Gs).

The Effect of Adding Characters

We hypothesized that the influence of the priors was due to the low phylogenetic signal in PR's contrived data set. As the posterior probability is proportional to the prior multiplied by the likelihood function, the influence of the priors will be greater when there is little phylogenetic signal in the data. In contrast, if the likelihood function has more data, then it effectively overcomes the priors, thus leading to the desirable situation in which the data determine the outcome of the analysis. We tested this hypothesis by doubling and tripling the number of characters in the contrived data set analyses (Fig. 1a), effectively doubling and tripling the number of apomorphies. We analyzed these data with the NCM and F81 models using the same analytical conditions as above. With the two-apomorphy analyses, the effect of the priors on posterior probabilities is still evident with the NCM analysis (Table 3), but it is not statistically significant ($P = 0.066$). The effect of the priors on the F81 analysis also was greatly diminished when compared with the single-apomorphy analyses, also becoming nonsignificant ($P = 0.161$). The effect of the unequal clade priors decreases further with the inclusion

TABLE 3. Results of the Bayesian analyses of Picket and Randle's (2005) contrived data set (Fig. 1a) with the number of apomorphies doubled and tripled. Grand means are calculated from posterior probabilities (sampled at stationarity) from each of the 100 individual analyses, per model. Results of the two-dimensional Kolmogorov-Smirnov test are shown at the bottom of the table; P values ≤ 0.05 are shown in bold. Clade numbers refer to Figure 2.

	Two apomorphies			Three apomorphies		
Clade	NCM	JC	F81	NCM	JC	F81
1	0.873	0.995	0.958	0.969	1.000	0.994
2	0.857	0.995	0.956	0.967	1.000	0.994
3	0.857	0.995	0.956	0.967	1.000	0.994
4	0.858	0.995	0.956	0.967	1.000	0.994
5	0.858	0.995	0.956	0.967	1.000	0.994
6	0.858	0.995	0.955	0.967	1.000	0.994
7	0.858	0.995	0.956	0.967	1.000	0.994
8	0.858	0.995	0.956	0.967	1.000	0.994
9	0.858	0.995	0.956	0.967	1.000	0.994
10	0.857	0.995	0.956	0.967	1.000	0.994
11	0.857	0.994	0.956	0.967	1.000	0.994
12	0.873	0.995	0.957	0.968	1.000	0.994
	$P = 0.066$	0.670	0.161	1.000	1.000	1.000

of three apomorphies (Table 3). We also note that it is not even necessary to add apomorphies in order to reduce the influence of the biased priors. If the data set is subsidized with uninformative nucleotide positions (thus resulting in shorter branch length estimates for the terminal branches), then the inferred phylogenetic information content of the original 14 apomorphies increases and the bias in the posterior probabilities diminishes (results not shown). Given that most empirical data sets are unlikely to be comprised of 15 nonhomoplastic yet variable base positions with no intervening invariant sites, the contrived data sets are clearly a poor proxy for the sorts of data typically encountered by systematists.

SIMULATED DATA ANALYSES

It is clear from our analyses of contrived, homplasy-free data that few data are required to overcome the influence of biased clade priors. However, as the contrived data clearly are not representative of empirical DNA sequence data, we attempted to further evaluate the impact of biased clade priors on more realistic, yet properly controlled, simulated DNA sequence data. To this end, we employed Seq-Gen v.1.2.6 (Rambout and Grassly, 1997) to simulate DNA data using the JC model on the 15-taxon tree shown in Figure 2. This tree topology corresponds to that inferred from PR's 15-taxon contrived data set; internal branch lengths and the taxa OG and N terminals were set to 1/14 or 0.071 substitutions per site and all other terminal branches were set to zero. We simulated 14, 28, 42, 56, 70, 84, and 98 character data sets 1000 times each and analyzed these data with MrBayes 3.0b4 using both the NCM and JC models. Each simulated data matrix was analyzed one time using the methods described above. To compare posterior probabilities, we calculated grand means of the posterior probabilities of each clade across all 1000 analyses conducted for each "character number" data set. The results of these analyses are presented in Table 4 and Figure 4. We predicted that the effect of the priors should diminish with increasing data because of the relatively larger influence of the likelihood function on the inferred posteriors. This is indeed what we observed. We found that 14-character data sets were strongly influenced by the unequal clade priors when applying both NCM and JC models ($P = 0.001$ and 0.005, respectively). The influence of the priors dissipated rapidly with increasing data and effectively disappeared for JC analyses of 28-character data sets ($P = 0.067$). Analyses applying the NCM model required more data (42 characters) before the relationship between clade prior and posterior probabilities was no longer significant ($P = 0.275$). We note that the two-dimensional Kolmogorov-Smirnov test appears to be insensitive to cases where the posterior probabilities of the largest and smallest clade are higher than the intermediate values, but intermediate values are similar (pers. obs.). Thus, it may be more informative to compare the differences between the maximum and minimum values (Table 4), which also decrease dramatically with the larger data sets. In either case (JC or NCM), the simulated data sets were much smaller than those typically employed in phylogenetic studies suggesting that the influence of unequal clade priors on posterior probabilities should not be cause for concern in the vast majority of empirical studies.

We were intrigued by our finding that unequal clade priors influence NCM analyses to a greater degree than is the case for the JC analyses. This may be related to the many more parameters estimated by the NCM model, which is equal to the number of sites multiplied by the number of internal branches (Tuffley and Steel, 1997). With extremely overdispersed models and little phylogenetic signal, the likelihood surface is expected to be relatively flat (Holder and Lewis, 2004), thus diminishing the contribution of the likelihood function relative to the priors. Despite being commonly referred to as a "parsimony model," NCM is anything but parsimonious. If researchers seek a model in which to perform phylogenetic analyses of combined DNA and morphological data sets, we recommend using the Mk model (Lewis, 2001b) as it assumes far fewer parameters than the NCM model.

BIASED SUPPORT VALUES IN EMPIRICAL STUDIES

Pickett and Randle demonstrated a significant correlation between clade prior and posterior probabilities in a selection of 17 published phylogenetic studies. That is, larger and smaller clades received significantly higher posterior probabilities than did clades of intermediate size (as in their analyses of the contrived data sets). However, PR observed the same pattern in jackknife and bootstrap proportions obtained in analyses of these same data, despite that these methods were included in the PR study as controls with the expectation that they would not generate biased support values. PR hypothesized that this unexpected finding indicated two things: (1) that noise inherent in empirical data leads to biased bootstrap and jackknife support values because these methods are

TABLE 4. Results from the simulated DNA data analyses. Grand means and standard deviations are calculated from posterior probabilities (sampled at stationarity) estimated for each of the 1000 individual data sets. The highest and lowest posterior probabilities are shown in bold. The results of the two-dimensional Kolmogorov-Smirnov test (P values ≤ 0.05 shown in bold) and the difference between the highest and lowest estimated posterior probability are shown at the bottom of the table.

	14 characters				28 characters				42 characters				56 characters				70 characters				84 characters				98 characters			
	NCM		JC		NCM		JC		NCM		JC		NCM		JC		NCM		JC		NCM		JC		NCM		JC	
Clade	Mean	σ	Mean	σ	Mean	σ	Mean	σ	Mean	σ	Mean	σ	Mean	σ	Mean	σ	Mean	σ	Mean	σ	Mean	σ	Mean	σ	Mean	σ	Mean	σ
1	**0.396**	0.246	**0.593**	0.338	**0.651**	0.256	**0.823**	0.269	**0.802**	0.220	0.920	0.190	**0.895**	0.161	**0.968**	0.121	0.930	0.133	0.980	0.088	**0.965**	0.096	0.992	0.056	0.973	0.081	0.995	0.045
2	0.302	0.244	0.530	0.355	0.584	0.283	0.787	0.295	0.767	0.245	0.902	0.213	0.869	0.181	0.963	0.121	**0.934**	0.124	**0.986**	0.071	0.957	0.102	0.992	0.053	0.974	0.083	0.995	0.052
3	0.285	0.232	0.536	0.358	0.584	0.282	0.805	0.285	0.779	0.235	**0.921**	0.188	0.874	0.180	**0.967**	0.111	0.924	0.139	0.980	0.099	0.958	0.100	0.992	0.053	0.973	0.085	0.994	0.053
4	0.259	0.231	0.527	0.359	0.545	0.278	0.783	0.291	0.762	0.234	0.918	0.181	0.868	0.191	0.957	0.138	0.927	0.135	0.983	0.080	0.958	0.095	**0.995**	0.033	0.975	0.084	0.995	0.056
5	0.243	0.227	0.497	0.350	0.559	0.283	0.798	0.284	0.770	0.235	**0.921**	0.186	0.863	0.191	0.956	0.141	0.931	0.129	**0.985**	0.078	0.950	0.119	0.989	0.068	0.969	0.093	0.993	0.059
6	0.236	0.243	0.485	0.355	0.555	0.280	0.790	0.290	0.752	0.242	0.911	0.197	0.862	0.186	0.961	0.120	0.928	0.135	0.983	0.091	0.957	0.105	0.991	0.064	**0.976**	0.075	**0.996**	0.046
7	0.240	0.241	0.504	0.359	0.564	0.283	0.798	0.289	0.761	0.247	0.910	0.201	0.874	0.182	0.966	0.117	0.920	0.146	0.980	0.087	0.958	0.104	0.990	0.063	0.971	0.092	0.993	0.057
8	0.240	0.231	0.485	0.356	0.559	0.285	0.786	0.294	0.768	0.242	0.917	0.192	0.874	0.191	0.959	0.139	0.923	0.139	**0.985**	0.066	0.959	0.096	**0.993**	0.052	**0.976**	0.083	**0.996**	0.040
9	0.254	0.232	0.517	0.353	0.575	0.280	0.805	0.282	0.754	0.245	0.905	0.204	0.860	0.199	0.954	0.142	0.923	0.151	0.977	0.107	0.957	0.101	0.992	0.053	**0.977**	0.073	**0.996**	0.044
10	0.272	0.235	0.522	0.355	0.575	0.280	0.798	0.285	0.756	0.240	0.912	0.191	0.858	0.204	0.949	0.154	0.924	0.142	0.979	0.098	0.957	0.104	0.991	0.062	0.974	0.080	**0.996**	0.043
11	0.322	0.253	0.543	0.351	0.594	0.274	0.808	0.279	0.766	0.245	0.904	0.205	0.867	0.188	0.959	0.136	0.933	0.130	**0.985**	0.077	0.948	0.121	0.988	0.069	0.973	0.083	0.995	0.049
12	**0.401**	0.239	**0.599**	0.337	**0.653**	0.259	**0.817**	0.277	**0.799**	0.221	0.917	0.194	**0.885**	0.171	0.964	0.125	**0.936**	0.126	**0.985**	0.071	**0.962**	0.100	0.990	0.070	0.975	0.081	0.994	0.061
	***P*=0.001**		**0.005**		**0.001**		0.067		0.275		0.586		0.532		0.251		0.488		0.862		0.959		0.926		0.364		0.535	
Max − min =	0.164		0.114		0.107		0.041		0.050		0.018		0.037		0.019		0.016		0.009		0.016		0.006		0.008		0.003	

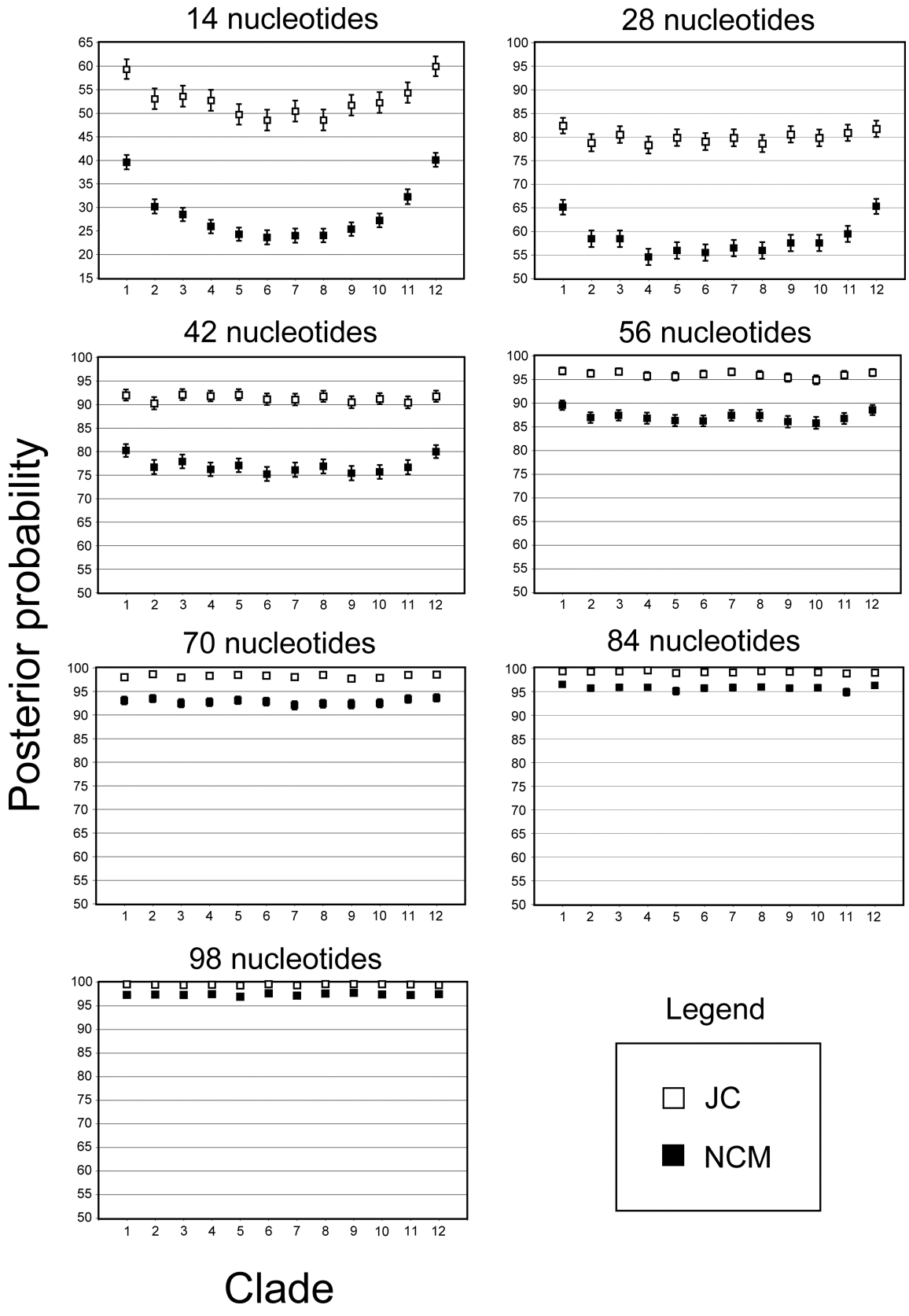

FIGURE 4. Results of the simulated DNA data analyses. Posterior probabilities are calculated as the grand mean of those estimated from each of the 1000 data sets. Error bars represent the 95% confidence interval of the grand mean. Values on the *y*-axis may differ among analyses, but the scale is equal. Flattening of the curve indicates decreased influence of the prior probabilities on the posteriors.

based on character resampling (which does not apply to Bayesian analysis), and (2) that Bayesian support values are biased by unequal clade priors. This hypothesis suggests that the similar patterns observed for bootstrap and jackknife support values on one hand, and Bayesian support values on the other, was coincidental (because PR imply that posteriors are expected to be biased by unequal clade priors regardless of the presence of homoplasy in the data, whereas bootstrap and jackknife proportions are not). PR attempted to test these hypotheses by creating and analyzing their contrived data sets. They reasoned that if they did not detect a correlation between bootstrap and jackknife proportions and clade priors in analyses of homoplasy-free contrived data, then this would be consistent with their hypothesis that the relationship between bootstrap and jackknife support values observed in the empirical data sets must be the result of noisy character data. They further reasoned that, should they detect a correlation between Bayesian posteriors and clade priors in analyses of the contrived data, this would be consistent with their hypothesis that the empirical bias in posterior probabilities is causally linked to unequal clade priors. Although both of these statements may be true, this does not necessarily suggest a compelling causal link between either of the hypothesized mechanisms and patterns observed in the empirical data—we believe this test requires a number of logical leaps that we find questionable.

First, we question PR's fundamental position that unequal clade priors are likely to produce detectably biased posteriors in Bayesian analyses of empirical data. The results obtained in our reanalyses of the contrived data, as well as in our analyses of simulated data, clearly indicate that the effect of nonuniform clade priors can only be detected with extremely sparse data—conditions that are unlikely to pertain to the empirical data sets evaluated by PR.

Second, we are not convinced that different underlying mechanisms are responsible for bootstrap, jackknife, and Bayesian support values that are similarly correlated with clade priors in the empirical studies (in other words, we are not convinced that the similar patterns observed for all three support values are coincidental). Indeed, without compelling evidence to the contrary, we believe that it is highly likely that Bayesian support values reflect the same phylogenetic signal in these data sets as do the bootstrap and jackknife proportions, regardless of whether that signal is biased in some manner by noisy data, tree shape, or other factors. For example, the observed pattern could simply be the result of taxon sampling. Many investigators root their trees by including taxa that clearly reside outside of the ingroup such that support for the large ingroup node tends to be high. The smallest clades in empirical studies often include multiple specimens of the same species, or perhaps small sets of taxa that share a common ancestor relatively recently when compared with other branches on the tree (e.g., sampling two mammals, two reptiles, two amphibians, and two actinopterygian fish in a phylogenetic study of vertebrates). Under these circumstances, we might expect the branch support values for these smaller clades to be greater on average than those associated with clades of intermediate size. Although taxon-sampling might not explain the patterns observed in the 17 studies considered by PR, we believe that this potential explanation (and others) are viable possibilities that must be given due consideration before we will know why the clade priors and support values in the empirical studies appear to be correlated. This is not to say that all possibilities must be considered simultaneously—just that tests must be devised that allow hypotheses to be adequately evaluated. For example, we believe that the contrived data analyses (as well as the analyses of simulated data) provide compelling evidence that unequal clade priors can be rejected as an explanation for the observed empirical relationship.

Should we be Concerned by the Effect of Unequal Clade Priors?

PR presented an example demonstrating that even when all clades are supported by the same amount of data, unequal clade prior probabilities (due to equal topological priors) will influence the posterior probabilities. It is not surprising that the prior distribution may influence the posterior when there are few data; we suspect that most practitioners of Bayesian phylogenetics are more interested in knowing whether biased clade priors are likely to influence inferred posteriors when analyzing their own data sets. Simply put, the likelihood function needs data with which to work. If the phylogenetic signal in the data is extremely low, the priors will mask the influence of the likelihood function, but given enough data, the opposite will occur and the likelihood will drive the outcome.

We show that PR's conclusion that unequal clade priors can influence estimated posterior probabilities is only relevant for exceptionally simple data sets, and thus, not particularly relevant to modern phylogenetic analysis. By simply doubling or tripling the number of apomorphies in their contrived data set, the influence of the priors became negligible. We further established, through simulations, that few characters are needed to overcome the effect of the unequal clade priors, and we therefore believe that the unequal clade priors are not problematic for Bayesian phylogenetics as a whole. Finally, we emphasize that PR's analysis of multiple empirical data sets is no more damning for Bayesian phylogenetics than for bootstrap or jackknife analyses, and may merely be an innocuous artifact of taxon sampling.

Modifying the Posteriors to Accommodate Bias

If an investigator remains concerned about the effects of priors on a clade, it is possible to adjust the posteriors to accommodate bias. More specifically, for a set of mutually exclusive hypotheses that completely partition the probability space, we can transform a posterior probability obtained under one set of priors into the

posterior probability given a different set of priors. This process is straightforward, and has been used previously with clade priors (Huelsenbeck and Immenov, 2002). As the posterior probability of a clade is proportional to its likelihood multiplied by its prior probability, we have simply to divide the posterior by the original prior and multiply by the new one, and scale by the sum of the posterior probabilities of all hypotheses under the new set of priors.

$$p_2(X \mid data) = \frac{\frac{p_1(X|data)\,p_2(X)}{p_1(X)}}{\sum_{i=1}^{n} \frac{p_1(X_i|data)\,p_2(X_i)}{p_1(X_i)}} \quad (1)$$

Where $p_j(X_i \mid data)$ is the posterior probability of hypothesis X_i given the data and prior probability $p_j(X_i)$. For clade posteriors, this simplifies to:

$$p_2(X \mid data) = \frac{\frac{p_1(X|data)\,p_2(X)}{p_1(X)}}{\frac{p_1(X|data)\,p_2(X)}{p_1(X)} + \frac{p_1(not\,X|data)\,p_2(not\,X)}{p_1(not\,X)}} \quad (2)$$

with X representing the clade in question. Investigators that are concerned about the effect of a particular set of priors on a clade of interest may therefore substitute any other set of priors that they deem appropriate after the fact. The problem of constructing an uninformative joint split prior that does not result in an invalid topology prior remains, so this transformation is primarily useful in examining the effects of priors on a particular clade of interest rather than correcting for concerns about the joint distribution of split priors across all splits. Implicit in this process is the assumption that the posterior probability estimate is reliable, and it is worth noting that convergence in the Markov chain methods used in Bayesian phylogenetics could be affected by altering the prior distribution, so that the posterior obtained by this method may not exactly match the posterior that would result from running a separate MCMC analysis of the same length under the new set of priors.

CONCLUSION

We believe that Bayesian phylogenetics will remain controversial until the systematics community is convinced that so-called uninformative priors do not influence estimated posterior probability values in unexpected ways. Phylogeneticists are just now beginning to explore these issues, and we envision that many studies will be required before the controversy is resolved (one way or the other). However, we feel that the present study provides some cause for optimism in that the unequal clade priors seem to have little influence on posterior clade probabilities for data sets comprised of more than a handful of informative characters. It remains to be seen whether priors on other model parameters will also be so easily overwhelmed by phylogenetically informative data.

ACKNOWLEDGMENTS

We thank Michael Alfaro, Paul Lewis, Christopher Randle, Michael Sanderson, Ainsley Seago, Michael Turelli, and the UC Berkeley phylogenetics discussion group for discussion and comments.

REFERENCES

Alfaro, M. E., S. Zoller, and F. Lutzoni. 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.

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

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

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

Felsentstein, J. 1968. Statistical inference and the estimation of phylogenies. Ph.D. Dissertation, University of Chicago, Chicago.

Felsenstein, J. 1981. Evolutionary trees from DNA sequences: Maximum likelihood approach. J. Mol. Evol. 17:368–376.

Felsenstein, J. 2004. Inferring phylogenies. Sinauer Associates, Sunderland, Massachusetts, U.S.A.

Goloboff, P. A., and D. Pol. 2005. Parsimony and Bayesian phylogenetics. Pages 148–162 *in* Parsimony, phylogeny, and genomics (V. A. Albert, ed.). Oxford University Press, Oxford, United Kingdom.

Huelsenbeck, J. P., and N. S. Imennov. 2002. Geographic origin of human mitochondrial DNA: Accommodating phylogenetic uncertainty and model comparison. Syst. Biol. 51:155–165.

Huelsenbeck, J. P., and B. Rannala. 2004. Frequentist properties of Bayesian posterior probabilities of phylogenetic trees under simple and complex substitution models. Syst. Biol. 53:904–910.

Huelsenbeck, J. P., and F. Ronquist. 2001. MrBayes: Bayesian inference of phylogeny. Bioinformatics 17:754–755.

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

Lemmon, A. R., and E. C. Moriarty. 2004. The importance of proper model assumption in Bayesian phylogenetics. Syst. Biol. 53:265–277.

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

Lewis, P. O. 2001b. A likelihood approach to estimating phylogeny from discrete morphological character data. Syst. Biol. 50:913–925.

Lewis, P. O., M. T. Holder, and K. E. Holsinger. 2005. Polytomies and Bayesian phylogenetic inference. Syst. Biol. 54:241–253.

Pickett, K. M., and C. P. Randle. 2005. Strange bayes indeed: Uniform topological prior simply non-uniform clade priors. Mol. Phylogenet. Evol. 34:203–211.

Rambaut, A., and N. C. Grassly. 1997. Seq-Gen: An application for the Monte Carlo simulation of DNA sequence evolution along phylogenetic trees. Comput. Appl. Biosci.13:235–238.

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

Tuffley, C., and M. Steel. 1997. Links between maximum likelihood and maximum parsimony under a simple model of site substitution. Bull. Math. Biol. 59:581–607.

Warren, D. L., J. Wilgenbusch, and D. L. Swofford. 2003. Converge: A program for implementing MCMC convergence diagnostics. Available from the authors at danwarren@ucdavis.edu.

Wilcox, T. P., D. J. Zwickl, T. A. Heath, and D. M. Hillis. 2002. Phylogenetic relationships of the dwarf boas and a comparison of Bayesian and bootstrap measures of phylogenetic support. Mol. Phylogenet. Evol. 25:361–371.

Zwickl, D. J., and M. T. Holder. 2004. Model parameterization, prior distributions, and the general time-reversible model in Bayesian phylogenetics. Syst. Biol. 53:877–888.

First received 27 January 2005; reviews returned 31 May 2005; final acceptance 15 August 2005
Associate Editor: Paul Lewis