

Syst. Biol. 55(1):147–151, 2006
Copyright © Society of Systematic Biologists
ISSN: 1063-5157 print / 1076-836X online
DOI: 10.1080/10635150500481630

Are Nonuniform Clade Priors Important in Bayesian Phylogenetic Analysis? A Response to Brandley et al.

CHRISTOPHER P. RANDLE¹ AND KURT M. PICKETT²

¹*Department of Ecology and Evolutionary Biology and The Natural History Museum and Biodiversity Research Center,
University of Kansas, Lawrence, Kansas 66045, USA*

²*Division of Invertebrate Zoology, American Museum of Natural History, Central Park West at 79th St.
New York, New York 10024, USA; E-mail: kpickett@amnh.org*

The use and design of prior distributions that reflect prior ignorance have long been controversial in statistics. The use of any prior distribution marks the difference between Bayesian and frequentist schools of thought. Disagreements regarding the design of prior distributions to reflect ignorance, and the interpretation of posterior distributions derived from such priors, have resulted in conflicting Bayesian philosophies. In a recent study, Pickett and Randle (2005) examined uniform prior distribution on tree topologies, which has been commonly used as a reflection of prior ignorance. We demonstrated that if all trees are considered to be equally probable a priori, all clades *cannot* be (except when n taxa ≤ 4). In a combined analysis of 17 empirical studies, nonuniform clade priors were correlated with Bayesian clade support. However, resampling supports (bootstrap and jackknife values) were also correlated with clade priors in these empirical studies. Using contrived data, we demonstrated that the correlation between resampling support and nonuniform clade priors is due to character conflict. When data sets lacked conflict, correlation disappeared, which follows from well-known properties of the bootstrap and jackknife (Farris, 1996). On the other hand, Bayesian clade support was correlated with nonuniform clade priors, even when there was no character conflict. Brandley et al. have called these findings into question, and although we consider their arguments flawed in several ways, they have posed a provocative question; are unequal clade priors problematic in Bayesian analysis? We address their concerns with our work, and conclude with a consideration of whether the question they pose has been answered.

EMPIRICAL ANALYSES IN PICKETT AND RANDLE (2005)

Brandley et al.'s criticism of Pickett and Randle (2005) is primarily focused on our use and analysis of contrived data, an argument that requires our analysis of 17 empirical, published studies to be invalid. There would have been little use in our contriving data had not a correlation been found between nonuniform clade priors and *both* Bayesian clade posteriors and resampling support values

using real data. Therefore, we will respond to Brandley et al.'s criticism of the empirical studies first.

In tests of correlation between clade prior probability and clade support, Pickett and Randle (2005) found significant correlations for Bayesian and resampling support (bootstrap and jackknife). It was not surprising that Bayesian support shows such a correlation, as prior assertions are involved in the estimation of posterior probabilities. It was surprising, however, that jackknife and bootstrap values were also correlated with clade priors, because topological priors are never included in their estimation. We hypothesized that the correlation seen for resampling supports was due to homoplasy. We then proposed a test and subjected *all* support measures to that test. The test's results were decisive, and our conclusions were drawn from those results.

Brandley et al. question our findings, saying that "the effect of non-uniform clade priors can only be detected with extremely sparse data—conditions that are unlikely to pertain to the empirical data sets of PR." It is not evident what "sparse" means in reference to data, but we believe that here, Brandley et al. are making the argument that our finding of correlation between priors and Bayesian clade support, and the absence of that finding for resampling support, is only applicable to small contrived data sets. We did not employ small data sets to explain the correlation between priors and posteriors alone, but to test the correlation between priors and *resampling support*. Nonetheless, we believe that, in the empirical analyses, for both measures of support, this correlation is due to a lack of data decisiveness, caused by the presence of character conflict for resampling support, or a likelihood function that is insufficiently decisive to overwhelm *nonequal* prior assumptions for Bayesian clade support. Of course, character conflict may result in data that are indecisive for either Bayesian or resampling support. At a minimum, the following can be logically inferred from the results of our contrived data regarding Bayesian support: conflict alone is not the cause of prior/posterior correlation.

Brandley et al. further dismiss the results of our empirical analysis, writing, "Secondly, 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

Christopher P. Randle and Kurt M. Pickett contributed equally to this article.

studies." We agree that this would be reasonable, absent a test that suggests otherwise. In constructing two hypotheses to explain a single phenomenon, Brandley et al. make the same logical leap that they find dubious in Pickett and Randle (2005); when explaining the correlation of priors in small, contrived data sets, and large empirical ones they hypothesize that the former is due to the effect of non-uniform clade priors and the latter is due to taxon sampling. The critical difference is that our hypothesis (that the correlation seen between priors and resampling support is due to character conflict) was tested and theirs was not.

There can be no doubt that the mechanisms differ between resampling and Bayesian support as the correlation is apparent for Bayesian support and evaporates entirely for jackknife and bootstrap support when analyzing precisely the same (homoplasy-free) data. Notwithstanding, Brandley et al. offer two hypotheses which, in conjunction, are proposed to explain the correlation between clade priors and clade support that we detected in the empirical analyses. First, they reason that the number of outgroup taxa is generally smaller than the number of ingroup taxa; therefore ingroup clades are big. Ingroup clades generally have higher support than other clades, and this will cause large clades to have higher support than intermediate ones. Secondly, they offer that, "the smallest clades in empirical studies often include multiple specimens of the same species, or perhaps small sets of taxa that share an ancestor relatively recently when compared to other branches on a tree."

The first hypothesis can be easily tested by removing the ingroup nodes from the correlation analysis. Of the seventeen studies, three do not use traditional outgroup rooting; the ingroup is constrained (Jordan et al., 2003), midpoint rooting is used (Rokas et al., 2003), or a single outgroup terminal defines the root (Simmons et al., 2004). Two studies provide analysis of two different matrices and therefore have posteriors for two ingroup nodes each (Kiefer et al., 2002; Rydin and Källersjö, 2002). Thus only 16 prior/posterior coordinates were removed from a total of 752 pairs, and the reanalysis again results in a significant correlation ($P = 0.005$), thus rejecting this hypothesis.

It is not immediately clear how the second hypothesis could be meaningfully tested; we discuss this below. However, it is worth pointing out now that, regardless of the validity of Brandley et al.'s second hypothesis (which seeks to explain only why small clades have high support), our refutation of their first means that they have no explanation for the entire distributional phenomenon that we reported in Pickett and Randle (2005).

The factors governing taxon selection are multifarious and often unstated. Brandley et al.'s second hypothesis, that researchers often choose exemplars to represent the smallest groups in an analysis based on the prior conviction that they are monophyletic, is a psychological rather than biological/inferential explanation. It may well be true; however, we can think of no experiment that could corroborate or refute it (save interviews or similar psychological assessment). In our experience, it

does not seem to characterize the methods by which researchers choose taxa (for example, we doubt that many researchers spend their funds deliberately duplicating many species in their molecular *phylogenetic* analyses). Further, the hypothesis does not seem to be generally true of the 17 empirical studies that we examined. As to the asserted view that sister taxa should be expected to have high support, although it is reasonable that clades of sister taxa share many more characters by descent than more distantly related taxa, we see no reason to assume that a high proportion of that support need necessarily optimize on the final branch leading to the sister clade. Regardless, abandoning our tested hypothesis in favor of theirs requires a preference for "plausibility" over hypothesis testing. The perception of plausibility cannot be the criterion upon which we discern the merit of hypotheses; prediction and test are those criteria.

Nevertheless, one possible test might involve the sequential pruning of taxa from sister clades. As we show, however, such a test does not result in mutually exclusive predictions between theirs and our hypotheses. If Brandley et al.'s hypothesis is correct, then reanalysis of the data minus the pruned taxa should generate new smallest clades with support values closer to those of the middle-sized clades. In essence, Brandley et al.'s second hypothesis predicts that sequential pruning will result in a slow flattening of the concave distribution of Bayesian support. Unfortunately, this outcome is precisely what our hypothesis predicts. Specifically, the difference in prior probability of smallest and mid-sized clades decreases and the curve becomes more shallow with every terminal removed from the analysis. This *must* at some point mitigate the effect of unequal clade priors (ultimately converging to flat [i.e., no] prior influence in the $n = 4$ case; see Pickett and Randle, 2005; fig. 1B). Therefore, sequential pruning does not, in fact, amount to a test that can distinguish between these two hypotheses.

Both of Brandley et al.'s hypotheses must be true to be explanatory of the entire distribution of posteriors (high for large and small clades and lower for middle-sized clades). At least the first hypothesis is false. Notwithstanding, were they both true, it would then follow that the non-uniform clade prior is *appropriately* (if subjectively) informative, as a reflection of biological data (for the first hypothesis) and as a reflection of the way that systematists choose taxa for analysis (for the second hypothesis). If that view is to be accepted, there is little that could be done experimentally to disentangle their hypotheses from our own.

Brandley et al. further offer that "other" unstated hypothesis "must be given due consideration." That other, unarticulated explanations may be imagined is certainly true. However, even given a stated, credible hypothesis, empiricism demands that it be exposed to refutation before other empirically corroborated hypotheses are abandoned. If scientific hypotheses that have survived refutation attempts can be called into question by the mere suggestion that other as yet unspecified explanations might one day come into view, then science

is in peril indeed. The entire materialist program could be criticized by use of such a view. Evolution itself can be criticized as not being the best explanation of biological diversity, if the only argument one need present is that some other force—whatever that may be—might be at play. Empiricism, however, demands that only articulated hypotheses be considered among possible explanations, and only tested hypotheses can supercede other tested hypotheses. Until such an alternative hypothesis is articulated, is tested, and survives that test, our hypothesis is the only one that has been subjected to experiment and not found wanting.

THE USE OF CONTRIVED DATA SETS IN PICKETT AND RANDLE (2005)

Brandley et al.'s criticism of Pickett and Randle (2005) is largely focused on our use of contrived data. Brandley et al. state: "We first focus on 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. . . ." In the portion of the study to which Brandley et al. refer, data were not contrived to demonstrate the relationship between clade prior and posterior. We hypothesized that for resampling support this relationship may be the result of character conflict, rather than clade prior influence. The important point is that when there is no character conflict, there is no relationship between non-uniform clade probability and resampling support, but the lack of conflict does nothing to lessen the relationship between Bayesian clade support and non-uniform clade priors.

We acknowledge that we erroneously referred to F81 as JC69; we are not too proud to admit that our error here is as Brandley et al. suggest. As regrettable as this error is, it does not alter, in any way, the findings of our study. Indeed, Brandley et al. conclude that the model employed was appropriate, regardless of what we called it. When this model was used by Brandley et al. the results are the same as our own. There is a significant correlation between clade priors and Bayesian support for data sets in which a single synapomorphy unites each clade.

Brandley et al. further demonstrate that when base composition priors do not reflect the base composition observable in the data, posteriors approach 1.0 at all nodes, and the correlation between clade prior and posterior disappears (Brandley et al., table 2). However, priors for base composition can be estimated empirically, from the data, and this, of course, is what should be done. In the case of the contrived data of Pickett and Randle, $\pi_{A,T} = 0.01$, $\pi_{C,G} = 0.49$ is the most appropriate base composition prior offered by Brandley et al. because our data were comprised entirely of C and G. The priors that result in decreased correlation ($\pi_{A,C,G,T} = 0.25$ and $\pi_{A,T} = 0.49$, $\pi_{C,G} = 0.01$) are poor reflections of the data. The demonstration that erroneous base frequency data can cause MrBayes to report perfect support for all nodes is a tangential matter.

Increasing the number of synapomorphies per node predictably results in an increase in Bayesian clade support at all nodes, resulting in the decrease in the correlation between clade posteriors and priors for all models tested. The increase would of course also be true of resampling support. However, clades are rarely supported by as many as three uncontradicted synapomorphies each in real data sets. In fact, we doubt there are any datasets one might point to that show the kind of decisive signal afforded by two or more uncontradicted synapomorphies per clade for every clade on the tree. It is the effect of nonuniform priors on clade posteriors in *real* data that concerns us.

REAL VERSUS REALISTIC

Brandley et al. criticize our use of contrived matrices as unrealistic—that they serve as a "poor proxy for the sorts of data typically encountered by systematists." This is not much of a criticism, however, as we intended these matrices to be unrealistic; contrived matrices only served the purpose of hypothesis testing, and were not intended to show anything about natural data, but rather, homoplasy-free data. In fact, these data sets were sufficiently unnatural that they had to be *contrived* rather than simulated.

Nevertheless, if knowledge about the performance of natural data had been the goal, Brandley et al. could have addressed more seriously our analysis of empirical studies. Instead, they propose that the cause of the correlation between priors and posteriors is suspect because of the results of allegedly realistic simulations. If the premise of realism of the simulations is in doubt, however, the conclusions that they inform natural data are similarly dubious. These data were simulated under the JC69 model, in which all base frequencies are stipulated to be equal, and all transition probabilities are stipulated to be equal, across and among all sites, everywhere in the matrix; Brandley et al. further simplified their simulations by stipulating equal branch lengths. Real data can rarely (if ever) be expected to conform to these simplifications. We consider this uncontroversial. Furthermore, the discovery that Bayesian methods find the correct solution with few sites is unimpressive when the precise model used to generate the data is used to infer clade support. Brandley et al. claim that merely 42 simulated sites are needed to overwhelm clade priors. Again, this stands in stark contrast to at least 17 empirical studies, which all used far more than 42 characters, but for which, the correlation between priors and posteriors persists.

We suspect that there is no universal solution to the question "How much data is required to overwhelm prior assumptions?" The quality of the data is almost certainly as important as the quantity. However, the quantity of information employed in many real data sets does not appear sufficient to achieve this end. While simulation studies may be useful in determining the amount of data required to overwhelm priors *under very specific model assumptions*, such simulations should be more

sophisticated than that employed by Brandley et al. if the goal is to approximate natural data.

INAPPROPRIATELY INFORMATIVE PRIORS AND THE INTERPRETATION OF POSTERiors

Brandley et al. ask "Are unequal clade priors problematic for Bayesian phylogenetics?," a question that neither our study nor their critique of it has addressed. This is a fundamentally different question than "Do unequal clade priors affect the estimation of Bayesian support in real data?" Our analysis of empirical data addressed this hypothesis and failed to refute it. Brandley et al.'s simulations only show that for data that are not real, nor to our mind, realistic, this correlation disappears with increasing evidential support. This does not address the title question nor does it provide the answer, "unequivocally, no."

Given what Pickett and Randle (2005) have shown about the prior assertions used in the vast majority of phylogenetic analyses thus far, and the finding of bias introduced by some uniform priors on model parameters (Zwickl and Holder, 2005; Yang and Rannala, 2005), the systematics community cannot ignore the importance of prior stipulation. Prior assertion is the critical difference between Bayesian and frequentist schools of statistics. A Bayesian believes that the knowledge that the sun rose yesterday should, rightly, influence our prediction of its rising tomorrow. Frequentists disagree (or, at least, they disagree that such prior events can be properly represented). This is no small disagreement, as the most cursory perusal of the statistical literature of the last century will bear out. Therefore, the widely held view (among phylogeneticists) that Bayesian methods amount to a fast way to compute ML trees is quite inappropriate, as it ignores this fundamental, philosophical, and mathematical divide. To address this, phylogeneticists must ask themselves, "Why should I use Bayesian methods?" If the answer is efficiency, other methods are far more efficient, such as parsimony and neighbor joining. However, efficiency *alone* can never be a justification for any method. If the answer is the simultaneous estimation of support and topology, then phylogeneticists must accept a method that is fundamentally different than traditional resampling measures, because the coincidence of these with Bayesian support is at least partly predicated on the use of ignorance priors (Efron et al., 1996, as discussed in Pickett and Randle, 2005). If, however, the investigator actually is a Bayesian, and not a frequentist, then Brandley et al.'s assertion should be unsettling.

Brandley et al. offer, "... advocates [of Bayesian phylogenetics] proceed with the hope that the likelihood will overwhelm inappropriately informative priors when they cannot be avoided." Here, they concede that inappropriately informative priors are a problem to be avoided, and the solution is to diminish, entirely, their influence. No prior on trees has been proposed that is not inappropriately informative for clades. Therefore, it seems quite clear that until that is accomplished, pri-

ors will remain an inappropriate influence that must be overwhelmed. Brandley et al. continue, "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." Here they suggest that, no matter what the meaning, biological or otherwise, of the prior statement, phylogeneticists need not worry, because the prior assertions eventually vanish. Others have made similar statements about priors (as discussed in Pickett and Randle, 2005), but always as a justification of empirical or ignorance priors. This is sharply different from the generally held statistical view that posteriors can always be altered by application of a sufficiently extreme prior. Accordingly, this justification requires some scrutiny. Pickett and Randle (2005) discussed the diminution of prior influence, though we pointed out that its phylogenetic limits are uninvestigated. The point of Bayesian statistics, however, is precisely to involve priors, whether they are empirically derived (actual prior knowledge) or conventional (which are generally distributional).

Actual prior knowledge, such as from morphological studies, can serve as empirically derived prior stipulations. When this is done, it can be analogous to the simultaneous analysis (Kluge, 1989; Nixon and Carpenter 1996) of the "prior" data and the new data (Wheeler, 1991). When no informed estimate of a prior can be made, uniform priors are employed with the goal of modeling ignorance. Such priors have been called "objective" because practitioners using them should agree that all competing hypotheses are given the same prior probability; in other words, they are obtained by convention. The reason that uniform, ignorance priors are selected is precisely to eliminate the undesirable influence that would result were the prior unjustifiably informative. An entire school of Bayesian thought is predicated on this assertion. If prior probability statements were as weak in their influence as Brandley et al. suggest—if they can safely be informative or not, meaningful or not, justified or not, because given even a paltry amount of data they will quickly be rendered moot—why would this large school even exist? In this school, uniform prior stipulations are seen only as formulaic "place holders" (i.e., not as actual prior knowledge) that allow for the approximation of Bayes theorem. This raises an often overlooked consequence of prior selection that impacts the interpretation of the posteriors: If prior views are empirically derived (actual knowledge), posteriors can be viewed as having some connection to the probability of the truth of the real world; if prior stipulations are the "place holder" variety, no such claims can logically be asserted (because the prior is neither a statement of factual "truth" about the world, nor a statement that entails a rational degree of belief).

This raises the question, "Are clade posteriors equivalent to the probability of a clade being true, or are they interpretable only as relative measures of evidentiary support, as are more traditional bootstrap and jackknife

supports?" Because uniform topological priors imply inappropriately informative clade priors, the latter interpretation is required. The Bayes factor (Jeffreys, 1960) allows an interpretation of such posteriors; it estimates how the evidence at hand has borne out prior assertions, regardless of how meaningful these priors are. Pickett and Randle (2005) advocated the use of Bayes factors when non-uniform clade priors are used, for this exact reason. If we do not interpret clade posteriors as a measure of evidentiary support, but they are also not reflections of actual prior knowledge or ignorance, one must ask what is the benefit of invoking a prior distribution at all.

Brandley et al. claim that the undesirable influence of priors, regardless of their design, is unimportant, because the prior is eventually overwhelmed. Brandley et al. suggest that advocates "hope" that "inappropriately informative" priors will either be overwhelmed or, if possible, avoided altogether. And so, it seems, they want their phylogenetic hypothesis to be free of prior influence, one way or another. Why make prior assertions if the goal is to eliminate their influence? If this goal is desired, then "we may solve the real problem directly" (Fisher, 1912:156), by employing a method that avoids the confounding effects of priors from the start; that method is maximum likelihood.

ACKNOWLEDGEMENTS

We thank Matt Brandley, Frank Anderson, Mikael Tholleson, Rod Page, and an anonymous reviewer for comments that improved the manuscript.

REFERENCES

- Efron, B., E. Halloran, and S. Holmes. 1996. Bootstrap confidence levels for phylogenetic trees. *Proc. Natl. Acad. Sci. USA* 93:13429–13434.
- Farris, J. S., V. A. Albert, M. Källersjö, D. Lipscomb, and A. Kluge. 1996. Parsimony jackknifing outperforms neighbor-joining. *Cladistics* 12:99–124.
- Fisher, R. A. 1912. On an absolute criterion for fitting frequency curves. *Messenger of Mathematics* 41:155–160.
- Jeffreys, H. 1960. *Theory of probability*. Clarendon Press, Oxford.
- Jordan, S., C. Simon, and D. Polhemus. 2003. Molecular systematics and adaptive radiation of Hawaii's endemic damselfly genus *Megalagrion* (Odonata: Coenagrionidae). *Syst. Biol.* 52:89–109.
- Kiefer, A., F. Mayer, J. Kosuch, O. von Helversen, and A. Veith. 2002. Conflicting molecular phylogenies of European long-eared bats (*Plecotus*) can be explained by cryptic diversity. *Mol. Phyl. Evol.* 25:557–566.
- Kluge, A. G. 1989. A concern for evidence and a phylogenetic hypothesis of relationships among *Epicrates* (Boidae, Serpentes). *Sys. Zool.* 38:7–25.
- Nixon, K. C., and J. M. Carpenter. 1996. On simultaneous analysis. *Cladistics* 12:221–241.
- Pickett, K. M., and C. P. Randle. 2005. Strange Bayes indeed: Uniform topological priors imply non-uniform clade priors. *Mol. Phyl. Evol.* 34:203–211.
- Rokas, A., G. Melika, Y. Abe, J. L. Nieves-Aldrey, J. M. Cook, and G. N. Stone. 2003. Lifecycle closure, lineage sorting, and hybridization revealed in a phylogenetic analysis of European oak gallwasps (Hymenoptera: Cynipidae: Cynipini) using mitochondrial sequence data. *Mol. Phyl. Evol.* 26:36–45.
- Rydin, C., and M. Källersjö. 2002. Taxon sampling and seed plant phylogeny. *Cladistics* 18:485–513.
- Simmons, M. P., K. M. Pickett, and M. Miya. 2004. How meaningful are Bayesian support values? *Mol. Biol. Evol.* 21:188–199.
- Wheeler, W. C. 1991. Congruence among data sets: a Bayesian approach. Pages 334–346 in *Phylogenetic analysis of DNA sequences* (M. M. Miyamoto, and J. Cracraft, eds.). Oxford University Press, Oxford.
- Yang, Z. H., and B. Rannala. 2005. Branch-length priors influences Bayesian posterior probability of phylogeny. *Syst. Biol.* 54:455–470.
- 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 submitted 7 August 2005; reviews returned 29 September 2005;
final acceptance 28 October 2005
Associate Editor: Frank Anderson

Syst. Biol. 55(1):151–159, 2006
Copyright © Society of Systematic Biologists
ISSN: 1063-5157 print / 1076-836X online
DOI: 10.1080/10635150500431205

Time, Species, and the Generation of Trait Variance in Clades

ROBERT E. RICKLEFS

Department of Biology, University of Missouri-St. Louis, 8001 Natural Bridge Road, St. Louis, Missouri 63121-4499, USA; E-mail: ricklefs@umsl.edu

No consensus has developed concerning the mechanism of phenotypic diversification among species within clades. Eldredge and Gould (1972) and Gould and Eldredge (1977) proposed an alternative to the then prevailing view among evolutionary biologists that evolution occurred by frequent small steps along branches of a phylogeny (anagenesis). Instead of this phyletic, or gradual evolution, they pointed to evidence from the paleontological record suggesting occasional

dramatic episodes of morphological change separated by periods of stasis. They associated shifts in their "punctuated equilibrium" model with speciation events (cladogenesis), and invoked a mechanism of "genetic revolution" linked to species formation in small, isolated peripheral populations (Mayr, 1963) to produce sudden change unrelated to selective pressures in the environment. Similar mechanisms referred to as "founder flush" and "transilience" also require population