On the reliability of Bayesian posterior clade probabilities in phylogenetic analysis

Graham Jones

January 10, 2008

email: art@gjones.name
web site: www.indriid.com

Abstract

This article discusses possible reasons why posterior clade probabilities obtained from Bayesian phylogenetic analyses might be inaccurate. It attempts to list all possible sources of uncertainty and error in Bayesian phylogenetic analysis. The choice of priors on trees has been suggested by several authors as a cause of inaccurate posterior clade probabilities. I argue strongly for using priors based on biological knowledge. Two possible sources of bias which do not seem to have been published before are also discussed. One is a computational problem and the other related to majority rule concensus trees.

1 Introduction

There has been much discussion as to whether posterior clade probabilities obtained from Bayesian phylogenetic analyses are "inflated", and if so, why. Some authors [17], [18] believe "that a conservative measure of support is, in general, preferable to an inflated one". I disagree: the correct place for conservatism (or its opposite) is in deciding what actions should be taken based on the results. If the results of an analysis are to become 'facts' to be printed in a text book, you'd certainly want to be conservative; if they are taken as a hint to guide further research, you would not. Whatever decision is to be made, it will surely be better to base it on accurate information, instead of using arguments like "the support is only .95 but this measure is known to be conservative so...". I therefore take the view that the goal should be to make posterior clade probabilities as accurate as possible, without building in any conservatism.

It *is* important to carefully assess the accuracy of the posterior probabilities of trees and clades, that is, to assess the variance (and more generally the distribution) of the posterior probabilities as well as any bias. It is easy to forget variance when the main problem seems to be bias. For example, in a region where it rains on 40% of days, one forecast might say there is a 40% chance of rain each day, while another says at random it will definitely rain on 40% of days, and won't rain the other 60%. They are both bad forecasts although neither is biased.

Much of the discussion about Bayesian posterior clade probabilities (BPCPs) has been along the lines of "why are they higher than other measures of support?". I don't attempt to address that issue here. I concentrate on why BPCPs might be wrong.

In this article I attempt to include all sources of uncertainty and error in Bayesian phylogenetic analysis, not just those that relate specifically to the Bayesian approach, starting with the data that is fed into the phylogenetic analysis. Then I turn to problems with the Bayesian analysis itself: models, priors, and computational problems. I end with a possible problem with consensus trees. Theory predicts, and simulations like [8] confirm, that if all elements of a Bayesian phylogenetic analysis are correct, then the posterior probabilities for trees will be unbiased estimates of the true tree probabilities. I describe some preliminary experiments which suggest that the same cannot be said for posterior clade probabilities that appear on the majority rule consensus tree.

2 Data problems

One reason for mentioning these problems in relation to “inflated” BPCPs is that it could be that in some phylogenetic analyses, problems with the data have been masked by the “cautious” or “pessimistic” nature of other support values such as maximum likelihood bootstrap support (MLBS). In such a case, MLBS may produce an acceptable “don’t know” result while BPCPs show strong support for clades which conflicts with other data. Thus a problem with the data might be mis-interpreted as a problem with the Bayesian approach.

2.1 Species issues

I will leave any discussion about the precise definition of a species to those better qualified than me, and assume that there is a consensus for the particular organisms involved in a given phylogenetic analysis. There remains the issue of which species are chosen. I have not seen any published discussion about this aspect (though I know some biologists argue about it informally). If adding some more species changes the posterior probability of a particular clade of interest, how should that be interpreted? The choice of species is also relevant to the choice of prior, a point I touch on later.

2.2 Laboratory problems

These include:

1. Within-species variation, since the molecular sequences depend on which individuals are selected.
2. Possible formation of chimeric sequences during amplification, by the artefactual joining of DNA’s from contaminant organisms.
3. Sequencing errors (ambiguities).

2.3 Alignment problems

There seem to be two possible sources of error here.

1. Alignment ambiguities. One cannot be certain that sites which are claimed to be “unambiguously aligned” are genuinely so.
2. Does the alignment program (ClustalW for example) - or the researcher if the alignment is done manually - introduce bias? Phylogenetic trees will be involved, either explicitly (eg the “guide tree” in ClustalW) or implicitly. It could be that the alignment method settles on a particular result which effectively rules out certain trees.

Note that these alignment problems are double-edged and present opportunities as well as difficulties. In principle it should be possible to use ambiguously aligned sites in the analysis if the ambiguity is taken into account. And a tree constructed during alignment, being based on insertions and deletions as well as substitutions, uses information not available to the later phylogenetic analysis. In principle it should be possible to combine the information to get a better analysis, but as is currently done in most cases, the alignment analysis and phylogenetic analysis overlap awkwardly and it is not clear to me whether this could be a source of serious problems. Some recent work [11], [16] has begun to address this issue.

3 Problems with Bayesian MCMC analysis

There are four sources of problems here [23]. (I split their first one into two.)

3.1 Model misspecification

The main issues seem to be:

1. Lateral gene transfer may undermine the suitability of basing the probability model on a tree.
2. There may be dependence between sites.
3. A molecular clock is usually too strong an assumption, while an independent clock for each branch is too weak.
4. The model for the way the mutation rate varies over sites and branches may be too restricted.
5. The substitution model may be wrong or too inflexible.
6. Different substitution models may be required at different sites (within user-supplied partitions).
7. Changes in DNA replication and repair machinery mean that different substitution models are appropriate in different parts of the tree. For example, there may be a change from CG-favouring to AT-favouring substitutions.

To simplify discussion of these issues, I will consider what is happening within a single user-defined partition. A 'substitution model' means a rate matrix R which has been normalised so that one mutation is expected per unit time. This is typically multiplied by a branch length (in nonclocklike models) or by a difference in node times and by a mutation rate μ (in clocklike models). Usually μ is allowed to vary between sites. A relaxed clock model also allows μ to vary between branches. The variation of mutation rates over branches and sites is normally modeled using a branch rate model and a site rate model, producing two numbers which are then multiplied together for a particular site on a particular branch. Although this allows rates to vary with sites and with branches, a site cannot change its rate over time, but can only change from branch to branch in the same way as all other sites: this restriction is what is referred to in point (4).

I have listed lateral gene transfer for the sake of completeness, but for the rest of this article, I will assume that lateral gene transfer is not an issue for the analysis in question, and that evolution can be modeled as a bifurcating tree with nodes representing speciation events, and with species evolving independently along branches.

On point (2), it seems to me that the usual assumption of independence between sites is almost always wrong, at least slightly, and will in most cases produce a sharper likelihood function than the true one, resulting in a sharper posterior and overestimated posterior probabilities for trees and clades. If the true nature of the dependence between sites is difficult to model, one might replace the likelihood L by L^s where $s < 1$. I suggest this as a "quick fix" which is very easy to implement. The choice of s would need to be justified. On point (5), the most flexible substitution model that is generally provided is the general time reversible (GTR) model although molecular evolution is not time reversible. On point (7), R is almost always taken as constant over branches.

Some of the other concerns are being addressed by phylogenetic analysis programs. For example BEAST [3], [4] has recently introduced a relaxed molecular clock, dealing with point (3), and others [13], [10] are tackling point (6). Covarion models [19] address point (4). But there is still plenty of scope for inappropriate models to be a cause of inflated BPCPs.

3.2 Prior misspecification

The specification of suitable priors on trees in Bayesian phylogenetic analysis is in its infancy. In most Bayesian phylogenetic analyses the thought process does not seem to go beyond accepting the defaults in a program such as MrBayes [7]. In the next two sub-sections I deal with two approaches which I believe are misguided.

3.2.1 Clade priors

Pickett and Randle [14] suggest that the commonly assumed uniform prior on topologies is problematic because it makes the prior on clades non-uniform, with a higher prior probability and therefore a higher posterior probability for small and large clades compared to medium-sized ones. Here, small, medium and large are relative to the number of species in the analysis, so a medium-sized clade means one that includes about half the species in the analysis. Brandley et al [2] argue against Pickett and Randle, who reply in [15]. None of these articles points out a fundamental fact, which is that a prior on clades which is anywhere near being uniform is impossible for large numbers of taxa. Nor do they consider the possibility that a non-uniform prior on clades is a good thing, which it is in my opinion.

I will deal with unrooted trees, and therefore with partitions rather than clades. By a partition I mean a subset of species which belong one of the two components obtained by removing a branch from an unrooted tree. Note that the existence of a partition is equivalent to the existence of the complementary partition which contains all the remaining species, so probabilities for large partitions behave in the same way as those for small partitions. Since [14] deals with rooted trees, the numbers in the following calculations are slightly different to theirs.

Suppose there is an even number $n = 2m \geq 6$ of taxa. A prior on topologies takes the form of a non-negative weight w_i for each topology τ_i, with the w_i summing to 1. Any topology will have at least 2 branches corresponding to a partition of size 2, and at most one branch which divides the taxa exactly in half, corresponding to 2 partitions of size m. Since there are $n(n-1)/2 = m(2m-1)$ possible partitions of size 2, and $(2m)!/m!m!$ of size m, the prior probability of a partition of size 2 will be much greater than that for a partition of size m for large m, regardless of how the w_i are chosen. The ratio must be at least $(2m)!/(m(2m-1)m!m!)$. To give some figures, suppose $n = 50$. The prior on a partition of size 25 given a uniform prior on topologies is about 5×10^{-15}. The most one can favour partitions of size 25 is to set $w_i = 0$ for all τ_i which do not have a branch which divides the taxa exactly in half. The prior on a partition of size 25 is then about 1.6×10^{-14}, only about 3 times bigger than for the uniform topology. A huge "bias" against medium-sized partitions is inevitable, and no choice of w_i has much effect on this. The same general conclusion is true of clades as well as partitions.

Once you have accepted the basic probability model you must accept the logical consequences of this. You can favour a particular medium-sized clade if you wish, but you cannot treat medium-sized and small-sized clades equally. Your ignorance about clades is limited. One might make an analogy with choosing lottery numbers. Getting exactly three numbers right is more likely than getting exactly five numbers right, and this remains true even if the lottery is 'fixed' so that some numbers are more likely than others.

Pickett and Randle [14] say "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)." I would say instead: researchers should be aware that assertions about the existence of medium-sized clades are more detailed statements about the real world than assertions about the existence of small clades and therefore require more evidence to justify them. The Bayesian approach to phylogenetics takes this into account, and I regard that as a good thing.

See [20] for a much more detailed and philosophical discussion of this issue.

3.2.2 Branch length priors

Yang and Rannala investigate the influence of branch length priors [23] for unrooted trees. This seems a useful analysis but I disagree with their conclusions and recommendations. I am happy to regard their results as a sensitivity analysis, and while they say that "posterior probabilities change drastically with the prior for internal branch lengths", I'd say: only if the priors are varied drastically. They are considering priors with extremely small means (down to 0.0001, ie one mutation expected per 10,000 sites in the interior branches).

Their justification for using branch length priors for interior branches with very small means is that for any particular phylogenetic analysis the vast majority of tree topologies are wrong, and so the prior that a biologist would wish to place on most trees would favour many zero length branches. They appear to be assuming, without explicitly saying so, that the prior takes the form P(topology) × P(branch lengths), ie that the topology and branch lengths are independent. This does not sit well with their justification for a small-mean branch length prior which is based on some topologies requiring very different branch lengths to others.

Even if one accepts that the priors for the topology and branch lengths should be independent, and that branch length priors should take account of the fact that the vast majority of tree topologies are wrong, an appropriate branch length prior would still give a moderate probability for some branch lengths to be substantially bigger than zero - even among wrong topologies, some internal branches will be correct. This would result in a branch length prior along the lines of a point mass at zero of weight $(1-b)$ with a continuous function for the rest of the density such as $ba\exp(-at)$. Here a is realistic for a real branch, perhaps 10, and b is small, but not absurdly so: b represents the proportion of correct internal branches in an randomly chosen topology. Yang and Rannala mention such a prior and say “This should have an effect similar to that achieved by assuming a small μ_0 [mean] in the exponential prior and will reduce high posterior probabilities for trees”. I doubt that it would have much effect with a sensible choice for b. The mean of such a distribution is ba, so one would have to assume that $b = 0.001$, that is, only one in 1000 interior branches of a random topology was expected to be nonzero, in order to achieve a branch length prior mean of 0.0001. Also note that evaluated at $t = 0.1$, the value of $ba\exp(-at)$ is about $3.7b$ if $a = 10$ while $a\exp(-at)$ is only about 10^{-430} if $a = 10000$. So the prior used in their simulations is biasing against realistic branch lengths to an enormous extent, and this would not be the case with the mixed distribution.

Having said all that, my real objection to their prior is that it is not based on biology. For typical choices of taxa, a tree with mostly zero length branches is biological nonsense and should have zero (or mind-bogglingly small) probability in the prior. Note that assuming that $b = .001$ in the above formulation means (among other things) that with 100 taxa, a star topology has probability just over 0.9 in the prior. Bayesian priors are subjective, but only a creationist would want to use that one!

3.2.3 Biological priors

A sensible prior on trees should be based on assumptions about evolution: speciation and extinction events, mutation and selection mechanisms, and the choice of taxa in the analysis. The researcher has some say in the latter of course, so one might adjust the taxa to suit the prior or vice-versa. I have seen very little discussion about biologically justified priors except where there is other information that can be used to favour particular trees or clades. About ten years ago, in two articles [21], [22] Yang and Rannala developed a model based on a birth-death process for rooted trees. This is the sort of approach I would advocate, but I haven’t seen any development of this, and their more recent article discussed above seems a step backwards to me.

[12] shows that there is a remarkable similarity in the duration of species measured in years over very diverse taxonomic groups (Bacteria, Marine diatoms, Dinoflagellates, Planktic foraminifers, Benthic foraminifers, Insects (beetles and Drosophila), Higher plants, Bryophytes, Lizards, Turtles and crocodiles, Rodents, Carnivores). According to [12], all these groups have an average species duration (time from origin to extinction) in the range 1 to 30 million years. Since the overall number of species changes slowly (a 10% increase every 25 million years is an approximate overall average since life began) the time between successive speciations (that is, the branch lengths measured in years) must be similar to species duration. Putting it a bit crudely, one could say that speciation is fairly clocklike, although mutation can be very non-clocklike. This suggests that a birth-death-sampling process as used in [22] is quite appropriate for assigning prior probabilities to trees.

However, real phylogenetic trees are well known to be less balanced than those generated by a birth-death process ([1] and many earlier references). It is not hard to put a prior on tree topologies which matches the observed data well in relation to tree balance, but as far as I know, there is no satisfactory model for a process in time which (a) generates trees with the observed degree of imbalance (b) is mathematically tractable in the sense that a probability for an arbitrary tree can be readily calculated during the MCMC algorithm and (c) gives plausible node times. There does not yet seem to be enough data to derive a prior on node times empirically.

Clearly, there is much work to be done in converting knowledge about macro-evolution into good tree priors for phylogenetic analysis. This, I suggest, is what biologists should be arguing about, rather than the ‘problems’ dealt with in 3.2.1 and 3.2.2. I don’t however think that this is the mostly likely place to look for an explanation for inflated BPCPs: the rest of this article shows that there are plenty of other possible culprits.

3.3 Program errors

There's not much to say about program errors. They certainly exist in a program as complex as MrBayes, but whether they are bugs that matter to most analyses is less clear. Simulation studies such as [8] provide some reassurance. The fact that many researchers use at least two programs to analyse their data is also a good thing. The other sort of programming errors are the ones made by users of programs. [2] points out that [14] makes such an error. I wonder if anyone would have noticed if the authors had not been having an argument.

3.4 Computational problems

The main issues are a failure of the MCMC algorithm to reach convergence, or a failure to obtain a representative sample from the posterior in a reasonable time. "Poor mixing" is usually described as a failure for the chain to visit all local maxima when they are separated by regions of very low posterior probability. There is also the possibility that there are enormous regions of low posterior probability which are not being explored by the chain. I will describe a 1-dimensional example to illustrate the problem.

Suppose the posterior is a mixture of two normals such as $0.9N(0,1) + 0.1N(0,s^2)$ where $s \gg 1$. If the proposal distribution is narrow, this can result in an MCMC algorithm sampling the narrow component of the distribution for an extremely long time. Unfortunately, if the starting point is quite distant from the origin, the typical behaviour of the algorithm often looks like a satisfactory convergence. The log-posterior values stay low until the chain finds the peak, then rapidly climbs and can stay high for millions of generations. Eventually, the chain will wander again down and explore the broad component of the distribution, but it would be easy to be fooled if the true distribution was unknown. Figure 1 shows a run with $s = 100$ and starting point 400, with proposal distribution a uniform distribution on $[-0.5, 0.5]$.

The enormous space of trees explored by the MCMC algorithm in phylogenetic analyses may be prone to similar problems. Suppose there are 10^{50} tree topologies whose individual contribution to the posterior is never more than 10^{-25} but whose combined contribution to the posterior is 0.1. These cannot be sampled in a reasonable amount of time, but their inclusion in estimates of posterior clade probabilities might well result in lower values (you might get 0.9 instead of 1.0). In [9] it is stated that "Runs were discarded if they failed to reach the same likelihood plateau observed in other independent runs". Such comments suggest that this issue is not currently taken seriously. Perhaps it should be. I would like at least to see some experiments using wide proposal distributions (especially for topologies) which stand a chance of exploring regions of low probability in a reasonable time.

It should be added that *if* this is a problem, it is likely to be an even greater problem for maximum likelihood methods which do not attempt to find anything but the (supposed) peak.

4 A problem with consensus trees

Results are usually reported as a single consensus tree with support for individual branches. It may be that these support values are biased upwards even when there are no problems with the data, model, prior or the computation. I have not seen this possibility suggested by anyone else. Suppose the posterior probabilities for trees are unbiased but suffer from high variance (due to limited data), so that some posterior tree probabilities are too big and others too small. The concern is that the consensus method will pick a tree with clades that have accidentally high probability due to the variance and then report those high probabilities as strong support for the clades, thus converting variance into bias. This particularly applies to methods of constructing a consensus tree which use clade probabilities as opposed to tree probabilities, including of course, the commonly used majority rule consensus tree. The result would be that while the posterior probabilities for trees have the desired frequentist behaviour, the posterior probabilities for clades do not.

I have investigated a mockup of the process whereby a set of trees is converted into a majority rule consensus tree. The mockup assumed five 6-taxon unrooted trees are involved, these being derived from Felsenstein's example on p341 of [5]. These five trees were supposed to have a true distribution of probabilities given by the

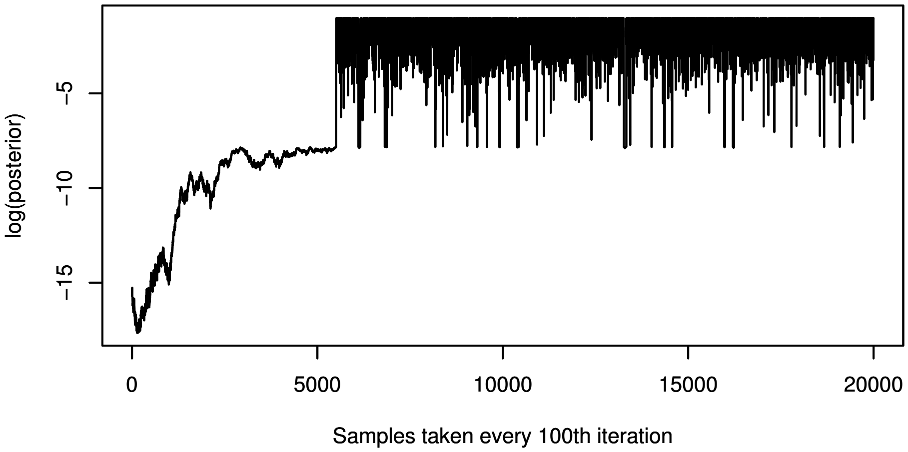

Figure 1. MCMC algorithm over two million iterations.

vector $(q_1, ..., q_5)^T = \mathbf{q}$. The "noisy" versions of the tree probabilities were modeled using a Dirichlet distribution with parameters $\alpha\mathbf{q}$ so that the mean is $\mathbf{q}$ and the variance of the i'th component is $q_i(1 - q_i)/(\alpha + 1)$ where α is a parameter which increases with the concentration of the distribution. For each value of $\mathbf{q}$, the clade probabilities for the majority rule consensus tree were calculated for $\mathbf{q}$ itself, and compared to the clade probabilities for the majority rule consensus trees from a sample (size 100) of noisy versions of $\mathbf{q}$. Since the majority rule removes clade probabilities below 0.5, a statistic which counts $(x - 0.5)$ for a clade probability x if $x > 0.5$, and zero otherwise, was used. I denote by W this statistic calculated for $\mathbf{q}$, and by $\hat{W}$ the mean value calculated for noisy versions of the tree probabilities.

The values for $\mathbf{q}$ were sampled from a Dirichlet distribution with parameters (1,1,1,1,1) (that is, uniform on the unit simplex). The results (for 1000 values of $\mathbf{q}$) show clearly that the noisy samples tend to have clade probabilities greater than those for $\mathbf{q}$. The effect increases with the dispersion of the Dirichlet distribution, that is, there is more bias for smaller values of α. With $\alpha = 50$, around 85% of values of $\mathbf{q}$ led to $\hat{W}$ being bigger than W with a mean difference of 0.02. For $\alpha = 10$, $\hat{W}$ was bigger than W 95% of the time with a mean difference of 0.09. For $\alpha = 2$ these figures were 100% and 0.35. To give a rough idea of what this bias means, an increase in clade probabilities from (0.60, 0.80, 0.98) to (0.64, 0.83, 1.00) would result in W increasing by about 0.09.

This preliminary experiment raises more questions than answers. Will the effect be bigger or smaller for actual sets of trees produced by an MCMC algorithm? Is the distribution of tree probabilities over different data sets for the same problem reasonably well modeled using a Dirichlet distribution? If so, what is a typical value for α in a real phylogenetic analysis? Can the problem be fixed using a different type of consensus tree?

5 Conclusion

I do not believe there is anything fundamentally wrong with the Bayesian approach to phylogenetic analysis. There are however many areas where there is room for improvement. More flexible models are needed, more care is required in choosing priors based on biological knowledge, and there is work to be done in detecting and eliminating computational problems. The variance of posterior probability estimates needs to be assessed, and the consequences of that variance determined. Much of the controversy surrounding the Bayesian approach has centered around the use of priors. In my opinion they are not the main cause of inflated posterior probabilities. The main reason for finding better priors and basing them on biological knowledge is that it should improve accuracy, not that it might reduce a confidence score.

Acknowledgment I would like to thank Alexei Drummond and Joel Velasco for useful comments and suggestions.

References

[1] Michael G.B. Blum and Olivier François. *Which Random Processes Describe the Tree of Life? A Large-Scale Study of Phylogenetic Tree Imbalance.* Systematic Biology, 2006, 55(4): 685-691.

[2] M.C. Brandley, A.D. Leach, D.L. Warren, and J.A. McGuire, *Are Unequal Clade Priors Problematic for Bayesian Phylogenetics?* Syst. Biol. 55(1):138146, 2006

[3] Drummond AJ, Rambaut A (2007) *BEAST: Bayesian evolutionary analysis by sampling trees.* BMC Evolutionary Biology 7:214

[4] Drummond AJ, Ho SYW, Phillips MJ & Rambaut A (2006) *Relaxed Phylogenetics and Dating with Confidence.* PLoS Biology 4, e88.

[5] J. Felsenstein, *Inferring Phylogenies*, Sinauer Associates, Inc., 2004

[6] Huelsenbeck, J. P. and F. Ronquist. *MRBAYES: Bayesian inference of phylogeny*, Bioinformatics 17:754-755. 2001.

[7] Ronquist, F. and J. P. Huelsenbeck. *MRBAYES 3: Bayesian phylogenetic inference under mixed models*, Bioinformatics 19:1572-1574. 2003.

[8] P. Huelsenbeck and B. Rannala, *Frequentist Properties of Bayesian Posterior Probabilities of Phylogenetic Trees Under Simple and Complex Substitution Models*, Syst. Biol. 53(6):904913, 2004

[9] James et al, *Reconstructing the early evolution of Fungi using a six-gene phylogeny*, Nature Vol 443, 19 October 2006

[10] Lartillot, N. and H. Phillipe. 2004. *A Bayesian mixture model for across-site heterogeneities in the amino-acid replacement process.* Molecular Biology and Evolution 21(6): 1095-1109.

[11] Gerton Lunter, István Miklós, Alexei Drummond, Jens Ledet Jensen and Jotun Hein, *Bayesian coestimation of phylogeny and sequence alignment*, BMC Bioinformatics 2005, 6:83

[12] A. Makarieva and V. G. Gorshkov, *On the dependence of speciation rates on species abundance and characteristic population size*, J. Biosci., Vol. 29, No. 1, 119128, March 2004

[13] Pagel, M. and Meade, A. 2004. *A phylogenetic mixture model for detecting pattern-heterogeneity in gene sequence or character-state data.* Systematic Biology 53: 571-581.

[14] K.M. Pickett, C.P. Randle, *Strange bayes indeed: uniform topological priors imply non-uniform clade priors*, Molecular Phylogenetics and Evolution 34 (2005) 203211

[15] K.M. Pickett, C.P. Randle, *Are Nonuniform Clade Priors Important in Bayesian Phylogenetic Analysis? A Response to Brandley et al.*, Syst. Biol. 55(1):147151, 2006

[16] Redelings, B.D., Suchard, M.A., 2005. *Joint Bayesian estimation of alignment and phylogeny*, Syst. Biol. 54, 401418.

[17] M.P. Simmons, K.M. Pickett, and M. Miya, *How Meaningful Are Bayesian Support Values?*, Mol. Biol. Evol. 21(1):188199. 2004

[18] Y. Suzuki, G. V. Glazko, and M. Nei. *Overcredibility of molecular phylogenies obtained by Bayesian phylogenetics*, Proc. Natl. Acad. Sci. USA 99:1613816143, 2002

[19] Tuffley C, Steel MA (1997) *Modeling the covarion hypothesis of nucleotide substitution.* Math BioSci 147:63-91

[20] J. Velasco, *The Prior Probabilities of Phylogenetic Trees*, to appear in Biology & Philosophy, 35 pp.

[21] Z. Yang and B. Rannala, *Probability Distribution of Molecular Evolutionary Trees: A New Method of Phylogenetic Inference*, J Mol Evol (1996) 43:304311

[22] Z. Yang and B. Rannala, *Bayesian Phylogenetic Inference Using DNA Sequences: A Markov Chain Monte Carlo Method*, Mol. Biol. Evol. 14(7):717-724. 1997

[23] Z. Yang and B. Rannala, *Branch-Length Prior Influences Bayesian Posterior Probability of Phylogeny*, Syst. Biol. 54(3):455470, 2005