University of Connecticut
DigitalCommons@UConn

EEB Articles | Department of Ecology and Evolutionary Biology

4-1-2005

Polytomies and bayesian phylogenetic inference

Paul O. Lewis
University of Connecticut, paul.lewis@uconn.edu

Mark T. Holder
Florida State University

Kent E. Holsinger
University of Connecticut, kent.holsinger@uconn.edu

Follow this and additional works at: http://digitalcommons.uconn.edu/eeb_articles

Recommended Citation
Lewis, Paul O.; Holder, Mark T.; and Holsinger, Kent E., "Polytomies and bayesian phylogenetic inference" (2005). *EEB Articles.* Paper 9.
http://digitalcommons.uconn.edu/eeb_articles/9

This Article is brought to you for free and open access by the Department of Ecology and Evolutionary Biology at DigitalCommons@UConn. It has been accepted for inclusion in EEB Articles by an authorized administrator of DigitalCommons@UConn. For more information, please contact digitalcommons@uconn.edu.

Submitted Sept. 18, 2003, revised Sept. 20, 2004

Polytomies and Bayesian Phylogenetic Inference

Paul O. Lewis[1], Mark T. Holder[2], and Kent E. Holsinger[1]

[1] *Department of Ecology and Evolutionary Biology, University of Connecticut, 75 N. Eagleville Road, Unit 3043, Storrs, Connecticut 06269-3043, U.S.A.*

[2] *School of Computational Science and Information Technology (CSIT), 150-F Dirac Science Library, Florida State University, Tallahassee, Florida 32306-4120, U.S.A.*

Running head: POLYTOMIES AND BAYESIAN PHYLOGENETIC INFERENCE

Keywords: Bayesian, phylogenetics, unresolved, star tree, polytomy, resolution

Corresponding author: Paul O. Lewis, Department of Ecology and Evolutionary Biology, University of Connecticut, 75 N. Eagleville Road, Unit 3043, Storrs, Connecticut 06269-3043, U.S.A.; Tel: +01 860 486-2069; FAX: +01 860 486-6364; E-mail: `paul.lewis@uconn.edu`

Abstract — Bayesian phylogenetic analyses are now very popular in systematics and molecular evolution because they allow the use of much more realistic models than currently possible with maximum likelihood methods. There is, however, a growing number of examples in which large Bayesian posterior clade probabilities are associated with very short edge lengths and low values for non-Bayesian measures of support such as non-parametric bootstrapping. For the four-taxon case when the true tree is the star phylogeny, Bayesian analyses become increasingly unpredictable in their preference for one of the three possible resolved tree topologies as data set size increases. This leads to the prediction that hard (or near-hard) polytomies in nature will cause unpredictable behavior in Bayesian analyses, with arbitrary resolutions of the polytomy receiving very high posterior probabilities in some cases. We present a simple solution to this problem involving a reversible-jump MCMC algorithm that allows exploration of all of tree space, including unresolved tree topologies with one or more polytomies. The reversible-jump MCMC approach allows prior distributions to place some weight on less-resolved tree topologies, which eliminates misleadingly-high posteriors associated with more-or-less arbitrary resolutions of hard polytomies. Fortunately, assigning some prior probability to polytomous tree topologies does not appear to come with a significant cost in terms of the ability to assess the level of support for edges that do exist in the true tree. Methods are discussed for applying arbitrary prior distributions to tree topologies of varying resolution, and an empirical example showing evidence of polytomies is analyzed and discussed.

Bayesian phylogenetic analyses were recently criticized (Suzuki *et al.* , 2002) for producing high posterior clade probabilities under simulation conditions in which no such clade exists. High posterior probabilities for clades with relatively low bootstrap support have also been noted in real data (Alfaro *et al.* , 2003; Douady *et al.* , 2003). Data simulated from a star tree (Figure) and analyzed with current Markov chain Monte Carlo (MCMC) approaches to Bayesian phylogenetic analysis produces a surprising result (Table 1). When only data sets containing a single simulated site are analyzed, the analysis produces, as expected, nearly equal posterior probabilities for each of the three possible fully-resolved tree topologies. When much larger data sets are simulated (e.g. 100,000 sites), disturbingly large posterior probabilities can be seen in the first few simulation replicates (note 0.9892 for replicate 11 in Table 1). This result is paradoxical because one might reasonably expect the posterior distribution to approach $(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$ with increasing sample size, reflecting the fact that there is in reality no true support for any of the three resolved tree topologies. Instead, the expected distribution occurs only with the smallest possible sample sizes in which there truely is no phylogenetic information. For large data sets, the phylogenetic uncertainty generated by the true polytomy manifests itself as unpredictability in the level of estimated posterior support for arbitrary resolutions of the polytomy, not as increased homogeneity of support for all possible resolutions.

This Bayesian star tree paradox was first noted by Suzuki *et al.* (2002), who were primarily investigating unduly high Bayesian posteriors in cases of model violation. It is unclear from their study whether the paradox results from simulating from a star tree or from using a different model to generate the data than was used in the analysis. Our simulations (Table 1) and the much more extensive simulations of Cummings *et al.* (2003) show that the paradox can result without model violation, although certainly some types of model violation would exacerbate the problem. Others noted unusually high posterior probabilities for edges with relatively low bootstrap proportions and very short estimated edge lengths in real data sets (Alfaro *et al.* , 2003; Douady *et al.* , 2003) and many more undocumented examples can be found by searching through papers in which both boostrapping and Bayesian analyses were performed, and trees were shown with branch lengths proportional to evolutionary distance (Collin, 2003; DeBry, 2003; Jordan *et al.* , 2003; Koepfli & Wayne, 2003; Steppan *et al.* , 2004). These examples suggest that a similar result can occur with realistic amounts of data when the expected number of changes on an edge is very small. It is worth noting that data sets used to address phylogeographic questions contain many closely related sequences and, because of the abundance of very short edges and perhaps true hard polytomies, may be highly susceptible to this sort of artifact.

One solution to the problem of occasional extreme posterior clade probabilities was suggested by Douady *et al.* (2003), who proposed using a bootstrap analysis in which individual bootstrap replicate data sets were analyzed phylogenetically using a Bayesian MCMC analysis. While such a procedure would have the desired effect of making analyses more conservative (Waddell *et al.* , 2002), it requires many times the computational effort of a

single MCMC analysis and, given time constraints, sacrifices precision and reliability in the MCMC analysis for increased conservatism.

We further explore the Bayesian star tree paradox, and present a direct, fully-Bayesian solution that sacrifices little of the computational tractability of conventional Bayesian analyses while accurately reporting weak support for very short edges when they exist.

The Nature of the Problem

The basic problem underlying the Bayesian star tree paradox is that the true (star) tree from which data were generated is not among the hypotheses being considered in the Bayesian analysis. While it is true that each of the three possible resolutions of the 4-taxon star tree can be made arbitrarily close to the star tree by shrinking the central edge, this fact only adds to the problem because it is then completely arbitrary which of the three tree topologies is chosen to represent the polytomy. The central problem is thus the identifiability of the tree topology.

A simple coin flipping example serves to illustrate this point. Suppose that a perfectly fair coin is flipped 1000 times and a Bayesian analysis is used to determine whether the coin is head-biased or tail-biased. Just as in the star tree paradox, the truth (perfectly fair coin) is not allowed as a possibility in this analysis; the analysis must choose between one of two incorrect hypotheses. The model used has a single parameter p, which represents the probability of landing heads on any given flip of the coin, and flips are considered to be conditionally independent given the value of p. The outcome of a series of flips follows a binomial distribution in which p is the probability of "success" on any given trial. The tail-biased hypothesis (H_T) is represented by $p < 0.5$, whereas the head-biased hypothesis (H_H) is indicated when $p > 0.5$, with the fair coin hypothesis H_F, if it were allowed, coinciding with $p = 0.5$. When $p = 0.5 - \epsilon$, where ϵ is some arbitrarily small but positive real number, the hypothesis H_T still holds, although the value of p is very close to the fair-coin value. Likewise, $p = 0.5 + \epsilon$ is very close to the fair-coin value, but nevertheless lies inside the realm of hypothesis H_H. Letting the prior distribution of p be the uniform(0,1) distribution assigns equal prior probability to the two hypotheses, H_H and H_T (hypothesis H_F is not considered, in keeping with the star tree paradox analogy).

Figure shows that the posterior distribution for p becomes concentrated and the mode approaches 0.5 as the number of flips increases. This behavior would be ideal if we were interested in estimating p, because with increasing data any reasonable estimate of p derived from the posterior distribution approaches the true value of 0.5 with increasing sample size. Unfortunately, we are instead interested in the posterior probabilities of the two mutually-exclusive hypotheses, H_H and H_T. The posterior probability of hypothesis H_T can be obtained by integrating the posterior distribution of p from 0.0 to 0.5, and the posterior probability of H_H is just $1 - \Pr(H_T)$. As the sample size increases, it becomes very easy for the spike-like posterior distribution to shift slightly from one side of the 0.5 value to the other (see for

example Figure c). Even though it is clear that the data overwhelmingly favor values of p very close to the fair-coin value 0.5, the fair-coin hypothesis is not an option and thus the distribution comes down (sometimes heavily) on one side or the other. While this is an informal description, it can be formally shown that the marginal posterior probability of either hypothesis has a uniform(0,1) distribution in the limit as $n \to \infty$. This means, with enough data, the posterior probabilities of either H_H or H_T would be completely unpredictable.

The obvious solution is to allow the fair-coin hypothesis onto the playing field so that it can compete with the other two hypotheses. The fair coin hypothsesis is represented by the single point, $p = 0.5$, whereas the other two hypotheses each correspond to exactly half the range of p. Assuming a flat prior for p means that the fair coin hypothesis receives zero prior probability mass, and no amount of data would be able to turn this around to produce a substantial posterior probability for this one point. The key is to use a "point mass" prior probability on the point $p = 0.5$. For example, if the H_F hypothesis should have equal standing with the other two hypotheses, it could be given a point mass prior of 1/3, with H_H and H_T also receiving prior probabilities of 1/3 each.

In the analogous phylogenetic case, the solution involves treating unresolved tree topologies as hypotheses in their own right. In the simple case of four taxa, this amounts to viewing tree space as comprising four tree topologies, not three, with the star tree being the heretofore overlooked fourth candidate. This makes it possible for the posterior distribution to become concentrated on the true tree, whether that true tree is the star tree or one of the fully-resolved tree topologies.

Allowing tree topologies with polytomies to contend with fully-resolved topologies in a Bayesian MCMC analysis is complicated by the fact that unresolved tree topologies have fewer edge length parameters than fully-resolved topologies. Thus, the MCMC analysis must jump between models of differing dimension. Such situations are ideal candidates for reversible jump MCMC (rjMCMC) (Green, 1995), and the remainder of this paper will be devoted to describing just such a solution.

Generalizing Bayesian phylogenetics to allow polytomous tree topologies

Suchard *et al.* (2001) observed that distinct tree topologies in likelihood-based phylogenetic analyses are most properly treated as separate models rather than distinct values of a single discrete parameter. Taking this view, a four-taxon star tree is a model with one fewer parameter than any unrooted bifurcating four-taxon tree, and choosing among tree topologies is more appropriately viewed as an exercise in model selection rather than parameter estimation.

One possible goal is to place unresolved tree topologies on equal footing with fully-resolved tree topologies, and a sensible starting point is to give all tree topologies equal prior probability, whether they represent a fully-resolved tree, the star tree, or something lying between these extremes. Doing this considerably expands treespace, as shown by Felsenstein

(2003), who described a simple algorithm for obtaining the number of tree topologies $T_{n,m}$ having n taxa and m internal nodes. As an example, there are 105 unrooted, bifurcating tree topologies for $n = 6$ taxa, but there are more than twice this number of tree topologies (236) if less-resolved tree topologies are counted. There is one tree having $m = 1$ internal node (the star tree), and there are 25 tree topologies having $m = 2$ internal nodes, 105 tree topologies having $m = 3$ internal nodes, and 105 fully-resolved tree topologies having $m = 4$ internal nodes.

Rather than using a flat topology prior, one might choose to give tree topologies with fewer internal edges greater prior weight than more resolved tree topologies. If a particularly controversial edge is short but has high posterior probability under a conventional analysis, it may be desirable to choose a prior that favors polytomies in order to be conservative. Critics may be convinced if the high posterior support for the edge persists despite strong pressure from the prior to collapse the edge.

Bayesian MCMC analyses

Bayesian approaches to phylogenetics, as performed by programs such as MrBayes (Huelsenbeck & Ronquist, 2003) use MCMC to approximate the posterior probability of different hypotheses. Specifically, they employ the Metropolis-Hastings (Metropolis *et al.* , 1953; Hastings, 1970) algorithm which consists of stochastically proposed changes to the parameters of the model (edge lengths and parameters in the substitution model) and a simple set of rules for deciding whether or not to accept these moves. The acceptance/rejection decision is based upon the product of the Metropolis ratio (ratio of posterior densities at the proposed point versus the current point) and the Hastings ratio (which accounts for any asymmetry in the proposal distribution).

If the data are ignored, the prior distribution equals the posterior distribution and an MCMC analysis results in an approximation of the prior distribution. The Markov chain should visit the star tree with roughly the same frequency as any of the three resolved tree topologies. But consider the effect of removing the central edge from a 4-taxon tree (creating the star tree) on the ratio of prior densities before and after the move. For simplicity, assume that edge lengths are the only parameters in the model. Before the move, the prior distribution comprises only five terms, corresponding to the prior probability density evaluated at each of the five edge lengths. On the other hand, the prior distribution of the star tree comprises only four terms because there are only four edges. The ratio of prior densities for the move must be $1/f(\nu)$, where $f(\nu)$ is the prior density of the edge length ν that corresponds to the edge missing from the star tree topology. This prior ratio is not a "legitimate" ratio for use in the Metropolis-Hastings algorithm because the probability densities have different dimensions; using such a ratio is analagous to trying to compare a meter to a square-meter. Fortunately, Green (1995) described a general way to correct for moves between models of different dimension, and this further modification of the Metropolis-Hastings algorithm is called reversible-jump MCMC, or the Metropolis-Hastings-Green algorithm.

Reversible-jump MCMC for moving between tree topologies of differing dimension

Reversible-jump MCMC can be employed to explore the space of all unrooted tree topologies, including less-resolved tree topologies such as the star tree. The basic algorithm described here involves four types of moves: 1) the "Larget-Simon" move; 2) the "Change-edge" move; 3) the "Add-edge" move ; and 4) the "Delete-edge" move. With probability ψ, one of the two dimension-changing moves (either Add-edge or Delete-edge) is attempted, and with probability $1 - \psi$ either the Larget-Simon or Change-edge move is attempted (depending on whether the current tree is the star tree). If a decision is made to attempt a dimension-changing move, the Add-edge move and Delete-edge moves are both chosen with probability 0.5 unless the current tree is either the star tree (in which case only Add-edge moves can be attempted) or a fully-resolved tree (in which case only Delete-edge moves can be attempted).

The Larget-Simon move — The Larget-Simon move is identical to the move labeled "LOCAL move without a molecular clock" by Larget & Simon (1999). This move always proposes a length change to three contiguous edges in the tree, and sometimes changes the tree topology by creating a nearest-neighbor interchange (NNI) rearrangement. This move was described in detail by Larget & Simon (1999).

The Change-edge move — The Change-edge move represents a simple modification of the Larget-Simon move for the special case of a star phylogeny. Star trees do not have three contiguous edges and thus Larget-Simon proposals cannot strictly be applied to them. The Change-edge move instead simply modifies the length of a single edge using the same proposal mechanism employed in the Larget-Simon move to change the length of the entire 3-edge tree segment. The acceptance probability for this move is:

$$\min\left\{1, \begin{pmatrix} \text{likelihood} \\ \text{ratio} \end{pmatrix} \cdot \begin{pmatrix} \text{prior} \\ \text{ratio} \end{pmatrix} \cdot \begin{pmatrix} \text{Hastings} \\ \text{ratio} \end{pmatrix}\right\} \tag{1}$$

The likelihood ratio depends on the particular substitution model chosen, and is the ratio of the likelihood of the tree after the proposed change has been made to the likelihood of the tree before the proposed change has been made. The prior ratio is the ratio of the prior probability density after the move to the prior density before the move. Only a single parameter is involved in this move (the edge length chosen at random to be modified), so for example if the prior on edge lengths were exponential with mean θ^{-1}, the prior ratio would be:

$$\frac{\theta e^{-\theta\nu^*}}{\theta e^{-\theta\nu}} = e^{-\theta(\nu^* - \nu)} \tag{2}$$

The Hastings ratio for the Change-edge move is the probability of proposing a move that would exactly reverse the Change-edge move divided by the probability of proposing the

Change-edge move itself. The new edge length ν^* is proposed as follows:

$$\nu^* = \nu e^{\lambda(u-0.5)} \tag{3}$$

where u is a uniform(0,1) random deviate. This expands or contracts the original edge length by an amount determined by u and the tuning parameter λ. The probability of proposing ν^* given λ and ν is $(\lambda\nu^*)^{-1}$, and the probability of exactly reversing this move is $(\lambda\nu)^{-1}$, so the Hastings ratio for this move is simply ν^*/ν. The Hastings ratio for the Larget-Simon move, for comparison, is $(m^*/m)^3$ (Holder *et al.* , in prep.), where m^* is the length of the three-edge segment after the proposed move and m is its length before the proposed move.

The Add-edge move — The Add-edge move (Figure a) results in the addition of a parameter to the model. One existing polytomy is chosen at random and the edges attached to the central node of that polytomy are split up into two groups (also randomly), and separated with a new edge. Because the Add-edge and Delete-edge moves are dimension-changing moves, the calculation of the acceptance probability requires not only the likelihood ratio, prior ratio and Hastings ratio but also a Jacobian term that corrects for the fact that the model dimension changes:

$$\min\left\{1, \begin{pmatrix}\text{likelihood}\\ \text{ratio}\end{pmatrix} \cdot \begin{pmatrix}\text{prior}\\ \text{ratio}\end{pmatrix} \cdot \begin{pmatrix}\text{Hastings}\\ \text{ratio}\end{pmatrix} \cdot (\text{Jacobian})\right\} \tag{4}$$

The prior ratio for the Add-edge move is

$$\frac{\pi_{n,m+1} f(\nu^*)}{\pi_{n,m}}, \tag{5}$$

where $\pi_{n,m}$ is the prior probability for a tree topology with n taxa and m internal nodes (m is the number of internal nodes *before* the Add-edge move) and $f(\nu^*)$ is the edge length prior density evaluated at the new edge length ν^*. All other edge lengths and substitution model parameters maintain the same values before and after the proposed Add-edge move, so all other terms in the prior ratio cancel.

The Hastings ratio for an Add-edge move is the probability of proposing a Delete-edge move that exactly reverses the proposed Add-edge move, divided by the probability of proposing the Add-edge move itself. Proposing an Add-edge move involves the following steps:

1. Choose to perform the Add-edge move rather than the Delete-edge move
2. Choose a polytomy at random from among all polytomies present in the tree
3. Choose one of the possible ways of dividing the k edges of the polytomy between the two nodes

4. Choose a length for the newly-created edge

The probability of the first step is 0.5 if there is a choice to be made. If the current state of the chain is the star tree, then this probability will be 1.0 because the Add-edge move is the only one possible. The probability of the second step is simply $\frac{1}{n_p}$, where n_p is the number of polytomies present in the tree before the move. The number of distinct ways of dividing k edges into two groups, making sure that at least three edges are attached to each node afterwards, is $2^{k-1} - k - 1$, so the probability of step three is $\frac{1}{2^{k-1}-k-1}$. To choose a new length, a uniform random deviate u is drawn and transformed into a random deviate from the edge length prior distribution:

$$\nu^* = -\frac{1}{\theta}\ln(u) \tag{6}$$

Thus, the fourth step has no effect on the probability of the Add-edge move because the value u has uniform probability density 1.0.

Proposing the corresponding Delete-edge move involves only two steps:

1. Choose to perform the Delete-edge move rather than the Add-edge move
2. Choose the particular internal edge in the tree to delete in order to restore the polytomy broken up by the proposed Add-edge move

The probability of the first step is again 0.5 unless the proposed Add-edge move yields a fully-resolved tree, in which case the probability of the first step is 1.0. The probability of the second step is $\frac{1}{1+n_b}$, where n_b is the number of internal edges present *before* the Add-edge move. The Hastings ratio is therefore

$$\begin{aligned}
\text{Hastings ratio for Add-edge move} &= \frac{\Pr(\text{reverse Add-edge move})}{\Pr(\text{Add-edge move})} && (7)\\
&= \gamma_b \cdot \frac{\frac{1}{1+n_b}}{\frac{1}{n_p}\cdot\frac{1}{2^{k-1}-(k+1)}\cdot 1} && (8)\\
&= \frac{\gamma_b n_p(2^{k-1}-k-1)}{1+n_b} && (9)
\end{aligned}$$

where

$$\gamma_b = \begin{cases} 0.5 & \text{if current tree is the star tree and proposed tree is not fully resolved} \\ 2 & \text{if proposed tree is fully resolved and the current tree is not the star tree} \\ 1 & \text{otherwise} \end{cases} \tag{10}$$

The Jacobian term for this move is

$$\frac{d\nu^*}{du} = \frac{1}{\theta(1-u)} \tag{11}$$

where ν^* is the proposed new edge length, u is the uniform random deviate chosen to generate the value ν^*, and θ^{-1} is the mean of the exponential prior distribution assumed here for edge lengths, which also serves as the distribution from which new edge lengths are generated during Add-edge moves.

The Delete-edge move — This move (Figure b) is called the Delete-edge move because it results in the deletion of an edge length parameter. One internal edge chosen at random from all existing internal edges is removed entirely from the tree, either creating a new polytomy or enlarging an already existing polytomy. The prior ratio for the Delete-edge move is

$$\frac{\pi_{n,m-1}}{f(\nu)\ \pi_{n,m}} \tag{12}$$

Note that the edge length prior, $f(\nu)$, is evaluated at the current length of the edge being deleted, and that m refers to the number of internal nodes in the tree before the Delete-edge move. The Hastings ratio for the Delete-edge move is calculated as follows:

$$\begin{aligned}
\text{Hastings ratio for Delete-edge move} &= \frac{\Pr(\text{reverse Delete-edge move})}{\Pr(\text{Delete-edge move})} && (13)\\
&= \gamma_d \cdot \frac{\frac{1}{n_p^*} \cdot \frac{1}{2^{k^*-1}-k^*-1} \cdot 1}{\frac{1}{n_b}} && (14)\\
&= \frac{\gamma_d n_b}{n_p^*\ (2^{k^*-1}-k^*-1)} && (15)
\end{aligned}$$

where

$$\gamma_d = \begin{cases} 0.5 & \text{if current tree is fully resolved and the proposed tree is not the star tree} \\ 2 & \text{if proposed tree is the star tree and the current tree is not fully resolved} \\ 1 & \text{otherwise} \end{cases} \tag{16}$$

k^* refers to the the number of edges emanating from the polytomy created (or enlarged) by the move, n_p^* is the number of polytomies in the tree after the move, and n_b is the number of internal edges in existence before the move.

The Delete-edge move, like the Add-edge move, changes the dimension of the tree model, and the Jacobian term needed to correct for this is:

$$\frac{du}{d\nu} = \theta e^{-\theta\nu} \tag{17}$$

This is exactly the inverse of the prior ratio. The Jacobian term cancels with the prior ratio in both the Add-edge and Delete-edge moves as long as the same distribution is used for the edge length prior and the generation of new edge lengths in Add-edge moves.

Prior distributions for tree topologies

The tree topology prior ratio $\pi_{n,m-1}/\pi_{n,m}$ (or its inverse) which appears in equations 5 and 12 can take on many forms subject to the goals of the investigation. We believe the three forms below represent reasonable choices, although there are certainly many other possible choices that are also defensible:

"Flat" prior on tree topologies — Setting $\pi_{n,m-1}/\pi_{n,m} = 1$ results in a flat prior on tree topologies, where the star tree has just as much prior weight as any particular fully-resolved tree. This induces a non-flat prior on resolution classes, where a resolution class is defined as the set of all tree topologies having a particular number of internal nodes m. For example, the prior on the resolution class having $m = n-2$ internal nodes (containing all fully-resolved tree topologies) will be $T_{n,n-2}$ times larger than the prior on the resolution class having $m = 1$ internal node (containing only the star tree).

"Polytomy" prior on tree topologies — This prior sets $\pi_{n,m-1}/\pi_{n,m} = C$, where C is a constant that determines the informativeness of the prior. If $C > 1$, this prior penalizes tree topologies for each additional edge length over more polytomy-rich topologies with which it is compared.

"Geometric" prior on resolution classes — Setting $\pi_{n,m-1}/\pi_{n,m} = \frac{T_{n,m}}{pT_{n,m-1}}$ yields a prior that is effectively geometric with parameter p over resolution classes. Choosing p close to 1 produces a steep prior with the star tree having the bulk of the prior probability, while choosing p close to 0 produces a relatively flat prior over resolution classes. One pitfall of this prior is that it can only be calculated for trees containing relatively few taxa because of the necessity of calculating the numbers of trees in the two resolution classes involved.

An empirical example

Shoup & Lewis (2003) performed Bayesian analyses as well as maximum likelihood bootstrapping, uncovering several instances of conflict between these two approaches to measuring edge support. We chose to reanalyze one clade in particular from this study (the Chlamydomonadales) because it exhibited four examples of edges with high posterior probabilities but low bootstrap support. The original study included 46 taxa, but our reanalysis included only the 16 taxa in the Chlamydomonadales plus *Oedogonium cardiacum* as the outgroup. The original study used the GTR+Γ model and results from two independent MrBayes 4-chain runs, each of length 10^6 generations (after 10^4 burnin), sampling every 100 generations to yield 20,000 sampled tree topologies total. The original sequences comprised 3341 nucleotide sites (474 parsimony-informative, 380 variable but parsiomony-uninformative) from combined 18S and 28S ribosomal RNA genes.

Our reanalysis of the Chlamydomonadales clade involved three primary components: 1) maximum likelihood bootstrapping; 2) a conventional MCMC Bayesian analysis moving among fully-resolved tree topologies only; and 3) a reversible-jump MCMC analyses moving among fully-resolved as well as polytomous tree topologies as described above. The third component actually involved 5 separate analyses, one each using the Flat prior, the Polytomy prior, and the Geometric prior with $p = 0.1$, $p = 0.5$ and $p = 0.9$. Except where otherwise indicated, all analyses used software written by POL and MTH.

ML Bootstrapping — A tree obtained by neighbor-joining was used to obtain maximum likelihood estimates of the parameters of the HKY+Γ model. These estimates were: transition/transversion rate ratio $\kappa = 3.317$; shape parameter of the 4-category discrete gamma distribution assumed for relative rates among sites $\alpha = 0.182$; relative base frequencies $\pi_A = 0.247$, $\pi_C = 0.221$, $\pi_G = 0.268$ and $\pi_T = 0.264$. These parameter values were fixed for the duration of the 500-replicate ML bootstrap analysis. This entire analysis was performed using PAUP* 4.0b10 (Swofford, 2001).

Conventional MCMC — A single Markov chain was run 2×10^6 generations after a 10^5 generation burnin period. The HKY+Γ model was used, with all model parameters except edge lengths updated using the slice sampler (Neal, 2003) every 50 generations. Branch lengths and the tree topology were updated using the Metropolis-Hastings proposal known as the Larget-Simon LOCAL move (Larget & Simon, 1999; Holder *et al.* , in prep.). Trees and model parameters were sampled every 100 generations, yielding 2×10^4 samples. The topology prior for this analysis was flat across all fully-resolved tree topologies and zero for tree topologies with any number of polytomies. An exponential prior (mean 1.0) was applied to each edge length, the transition/transversion rate ratio parameter κ, and the inverse of the shape parameter of the gamma distribution assumed for relative rates across sites α^{-1}. The inverse of the shape parameter equals the variance of relative rates among sites because the scale parameter of the gamma distribution is chosen to make the mean relative rate 1.0. Using the variance (α^{-1}) is more natural than using the shape (α) to

parameterize the gamma distribution in a Bayesian setting because standard choices of prior distributions allow for the possibility of rate homogeneity (rate homogeneity corresponds to $\alpha = \infty$, which has infinitesimal density under typical choices for prior distributions). Variance is also a more intuitive measure of rate heterogeneity because it increases (rather than decreases) with increasing rate heterogeneity. Because they are constrained to sum to 1.0, a different strategy was taken for updating base frequency parameters. Four continuous, non-zero parameters, f_A, f_C, f_G, and f_T, were each given exponential priors (mean 1.0) and updated individually by slice sampling. Base frequencies needed for calculation of the likelihood were generated by normalizing the f_i parameters. That is,

$$\pi_i = \frac{f_i}{f_A + f_C + f_G + f_T} \tag{18}$$

The exponential prior (mean 1.0) on each f_i is equivalent to placing a (flat) Dirichlet(1.0, 1.0, 1.0, 1.0) prior directly on the relative base frequencies. The advantage to treating relative base frequencies in this way is that no special treatment is required (i.e. one does not need to devise a multivariate proposal); the f_i parameters can be updated independently in exactly the same way as any other substitution model parameter.

Reversible-jump MCMC — The run conditions and choices of substitution model priors for this analysis were identical to those of the conventional MCMC analysis; however, this analysis attempted one of the two described dimension-changing moves with probability 0.5 each generation. Five separate MCMC analyses were performed, each using a distinct prior distribution on tree topologies. One run employed the "Flat" prior (all tree topologies have equal prior probability regardless of resolution). A second run used the "Polytomy" prior with constant C equal to $e \approx 2.71828$. This value for C makes the cost of adding an edge equal to one log-likelihood unit. The three remaining runs used the "Geometric" prior with $p = 0.1$, $p = 0.5$ and $p = 0.9$, respectively. The five prior probability distributions over the 15 resolution classes for this 17-taxon analysis are given in Table 2.

Despite the differences in taxon sampling, the use of a simpler substitution model, and differences in the MCMC analysis, the results for the Chlamydomonadales clade are very similar between the original analysis of Shoup & Lewis (2003) and our reanalysis. For the bootstrap and conventional MCMC analyses, the only notable difference is that the posterior probability for the *Carteria/Heterochlamydomonas* subclade (node 3 in Figure) was substantially lower (0.87) in the original study than it is in our reanalysis (0.99). The remaining discussion will focus exclusively on the reanalysis results because these are all based on the same data and substitution model and are thus comparable.

Five edges (1, 2, 3, 7 and 11 in Figure) had very high (>93%) posterior probabilities with conventional MCMC and low (<72%) ML bootstrapping support. The posterior probabilities for two of these edges (7 and 11) dropped substantially when the reversible-jump

MCMC approach was used, regardless of which topological prior was used. Two of the remaining three edges (2 and 3) maintained high posterior probabilities, even when using prior distributions that strongly favor polytomous tree topologies (Table 3). Edge 1 was the only edge strongly influenced by the choice of topological prior: its posterior probability varied from 0.41 to 0.91 depending on which topology prior was used. The three "Geometric" prior distributions explored range from weakly favoring polytomies ($p = 0.1$) to very strongly favoring polytomies ($p = 0.9$). The "Geometric" prior with $p = 0.5$ and the "Polytomy" prior fall between these two extremes. The "Flat" prior does not favor any particular polytomous tree more than any particular full-resolved tree, but we note that this prior is not flat across resolution classes because of the different numbers of tree topologies composing each resolution class. For this 17-taxon example, this prior on resolution classes induced by assuming the Flat prior across tree topologies actually favors the class of tree topologies having $m = 12$ internal nodes most (fully resolved tree topologies have $m = 15$ internal nodes in this case).

While it is of course impossible to know for sure, these results suggest that two of the five edges of interest may in fact be hard polytomies (or edges too short to detect with the particular genes chosen). Two of the three other edges of interest maintain high posterior probabilities in the face of strong prior pressure to collapse into polytomies. It is reassuring to see that biasing the prior toward polytomous tree topologies does not reduce support in all cases. Such behavior would indicate that biasing the prior simply reduces the power of the analysis to detect nonzero-length edges when they do in fact exist. Instead, this example suggests that it is possible to identify hard polytomies as such. We note that this study does not identify or correct other potential problems. Branches that remain strong after such an analysis may represent overcredibility due to model violations (see Buckley, 2002), and using reversible-jump MCMC cannot ameliorate any effects caused by model inadequacy.

Convergence and mixing

Our analysis of the Chlamydomonadales example might be challenged in at least two ways. First, if the run length used (2.1×10^6) in the reversible-jump MCMC analysis was not sufficient to accurately approximate the posterior distribution, then the two polytomies identified may be artifacts of the failure to sample from a stationary chain. Second, the choice of prior distributions, especially with respect to edge lengths, may have been the primary reason for the high posterior probabilities obtained for nodes 7 and 11. In this case, it is important to check whether other reasonable prior distributions for edge lengths produce the same results.

Assessment of mixing and convergence in the reversible-jump MCMC analysis — We investigated the samples produced from the MCMC analyses of the Chlamydomonadales example several ways to make sure that the two polytomies identified with our new approach were not simply an artefact of inadequate mixing or failure of the chain to converge to the stationary distribution. To assess convergence with respect to substitution model parameters (excluding edge lengths), we computed Potential Scale Reduction Factors (Brooks & Gelman, 1998).

This measure (symbolized R here) assesses convergence by constructing a ratio of estimators of the variance of sampled values from several independent MCMC runs. The numerator is a pooled variance estimator that combines within-chain and among-chain variation, whereas the denominator is based only on within-chain variation. If all of the parallel chains have converged, samples taken from these chains will be statistically indistinguishable, and R will be close to 1.0. If the chains have not been run long enough, estimates of the variance of sampled values made from within-chain variation will be smaller than estimates of the same quantity made using variation among-chains. In this case, R will be greater than 1.0. Values of R less than 1.2 are generally considered acceptable (Gelman *et al.* , 1995, p. 332). The R values computed for 10 replicate reversible-jump MCMC analyses (using the "Flat" topology prior) are all very close to 1.0, consistent with convergence. In each analysis, the first half of the 2.1×10^6 samples were discarded, yielding these potential scale reduction factors: $R_\kappa = 0.9997592$, $R_{\alpha^{-1}} = 0.9998576$, $R_{f_A} = 1.0108441$, $R_{f_C} = 1.0111435$, $R_{f_G} = 1.0109623$ and $R_{f_T} = 1.0110756$.

Potential scale reduction factors only address convergence in substitution model parameters, but perhaps of more concern is mixing and convergence with respect to topology. It is unlikely, but nevertheless possible, for an MCMC analysis to become quickly stuck on a relatively improbable tree topology. All indications (time-series plots of the log-likelihood or log-posterior, potential scale reduction factors, etc.) may point to convergence in such a case; however, the chain is far from converged. Li *et al.* (2000) introduced SRQ plots to assess mixing and convergence for topologies. From an SRQ plot, it is possible to see graphically whether subsequences of arbitrary length and starting position would provide reaonable approximations to the posterior probabilty obtained from the complete sequence of samples. In an SRQ plot, the abscissa records the ratio T_i/T_N, where T_i is the number of times the focal tree topology has been sampled up to iteration i. T_N is the total number of times the focal tree topology was sampled; the MCMC estimate of the posterior probability of the focal tree topology is thus T_N/N. The ordinate records the ratio i/N. The slope of the line in an SRQ plot should ideally be close to 1.0 throughout. Departures from this indicate that at some points the chain was on a trajectory that should have led to a different final posterior probability. Especially worrying would be a plot in which the slope was some value s for most of the run, where $s \neq 1.0$, then changed abruptly at the very end. Figure shows SRQ plots for the maximum posterior probability (MPP) tree topology in the reversible-jump MCMC analysis utilizing the "Flat" topology prior. This tree topology, which is identical to the majority rule consensus tree shown in Figure , captured nearly 80% of the posterior probability. The SRQ plot produced for the entire run (Figure c) is close to ideal, and the plot created for the first 200 and 2000 post-burn-in samples (Figure a and b, respectively) are very reasonable.

Relevant to the question of mixing is how many times the chain returned to the focal tree after having left it to visit a different topology. Table 4 provides summary information about sojourns made by the MPP tree topology as well as the other four topologies composing

the 95% credible set. This shows that the chain returned to the MPP topology at least 588 separate times, and stayed on the MPP topology on average only 27.1 samples before switching to a different topology. This ability to leave and then return repeatedly to the same topology suggests good mixing with respect to topologies.

Assessment of alternative edge length priors in the conventional MCMC analysis — In many previous studies in which suspiciously large posterior probablities have been estimated for relatively short edges(Alfaro *et al.* , 2003; Douady *et al.* , 2003; Cummings *et al.* , 2003), MrBayes (Huelsenbeck & Ronquist, 2003) was used with the default edge length priors, which were uniform distributions with large means (e.g. 5.0 for version 1.1). The concern is that such a large prior mean might amplify weak support for one possible resolution of a polytomy, producing what seems to be surprisingly high estimates of the posterior probability of an edge. To address the concern that using a different edge length prior might have produced similar results to those obtained using the reversible-jump approach, we performed additional conventional MCMC analyses using different means for the exponential prior on edge lengths. The results detailed in Table 2 used a prior mean edge length of 1.0. Because 1.0 is longer than most empirical edge length estimates, we tried setting the prior mean to shorter values, namely 0.1 and 0.01. These had no effect on the majority rule consensus topology, and did not substantially reduce the posterior probabilities of the two edges (7 and 11) of interest. For edge 7, the posterior probabilities for means 1.0, 0.1 and 0.01 were 0.95, 0.95 and 0.96, respectively. For edge 11, the posterior probabilities for means 1.0, 0.1 and 0.01 were 0.94, 0.94 and 0.95, respectively. All aspects of these analyses were identical to the original with the exception of the prior distribution assumed for edge lengths.

Suchard *et al.* (2001) employed a hierarchical model in which a parameter μ governed the mean of the exponential edge length prior distribution. In this approach the μ parameter automatically tunes the edge length prior distribution so that its mean is appropriate for the data being analyzed. Although a fully Bayesian approach, this has an effect similar to an empirical Bayes analysis in which the maximum likelihood estimate of the mean edge length is used to specify the mean of the edge length prior distribution. The advantage of the fully Bayesian model is that uncertainty in μ itself is accommodated. We investigated adding the hyperparameter μ to our model and, following Suchard *et al.* (2001), using an Inverse Gamma prior on μ with mean 1 and variance 10. The posterior mean of μ was 0.059 (standard deviation 0.011), and this analysis produced posterior probabilities for edges 7 and 11 of 0.94 and 0.95, respectively. Thus, using a hyperprior for edge lengths also did not affect the results of the original analysis, which fixed the prior mean to 1.0.

A simulation example

The empirical example begs the question "Just how short can an edge be and still be detected by a reversible-jump MCMC that strongly favors polytomous tree topologies?" To address this question, a series of data sets were simulated using the maximum likelihood tree from the Chlamydomonadales example as the model tree. Maximum likelihood estimates

under the HKY model were used for all edge lengths except the deeper of the two edges that collapsed in the empirical example. This edge (i.e. edge 7 in Figure) was varied among simulations, with 100 data sets simulated (using Seq-gen version 1.2.6, Rambaut & Grassly, 1997) for each of the following five lengths for this edge: 0.0 (representing a true hard polytomy), 0.0006, 0.0012, 0.0018, 0.0024 and 0.003 (measured as expected number of substitutions per site). These values are multiples of 0.0003 ($\approx 1/3341$), which is the edge length providing for, on average, one change per *data set* along the edge of interest. Note that because edge lengths represent expected amounts of change, an edge length of 0.0003 does not guarantee that a change will occur on the edge of interest in every simulated data set, nor does it guarantee that sites where changes do occur will be free from homoplasy. The simulated data sets were of the same size as the original data (17 taxa, 3341 sites), and the "Polytomy" prior (with $C = e$) was used with reversible-jump MCMC for analysis. The results (Figure) show that for an edge to be reliably detected, it must be long enough that several sites are expected to change across the edge. However, Figure demonstrates that our method is virtually certain of being able to detect edge lengths as short as 0.003 expected substitutions per site, while still correctly identifying hard polytomies as such.

Conclusion

The use of Bayesian MCMC methods that do not allow tree topologies with polytomies to have non-zero posterior probability can lead to unjustifiably high posterior probabilities for arbitrary resolutions of hard polytomies, or for internal edges in which the expected number of changes is very small. It is important to point out that high posteriors in such cases are not the rule, but such edges are enriched in majority rule consensus trees (a very common way to summarize the posterior distributions in Bayesian phylogenetics), which attracts the attention of investigators and makes this phenomenon of some concern.

Fortunately, there is a simple solution to the problem. The Green modification to the Metropolis-Hastings algorithm used widely in Bayesian MCMC analyses allows less-resolved tree topologies containing one or more polytomies to compete with fully-resolved tree topologies for posterior probability. If tree topologies containing polytomies can attain high posterior probabilities, support will be removed from fully-resolved tree topologies representing arbitrary resolutions of the polytomy. Furthermore, the solution described here requires only slightly more computational effort than the current approaches that consider fully-resolved tree topologies only. Using prior distributions that strongly favor polytomous tree topologies over fully-resolved tree topologies provides a conservative approach that reduces spurious high posterior probabilities in cases of true polytomies (or edges too short to be detected) without compromising the ability to correctly identify and strongly support short but real edges in the true tree.

The Bayesian approach to phylogenetic inference has been characterized by an exceptional growth in popularity. It is to be expected that once a new methodology becomes widely used, problems will surface that previously went unnoticed simply because of their low frequency

of occurrence. The first reflex is often to point out such problems and argue that the entire methodology be abandoned (e.g. Suzuki *et al.* , 2002), but the fact that Bayesian analyses have made possible the use of the most realistic evolutionary models ever developed provide ample reason to work instead toward finding solutions to problems when they arise.

Acknowledgements

We gratefully acknowledge the advice and encouragement provided by Alan Gelfand. We also thank Alexei Drummond, an anonymous reviewer, and Associate Editor Thomas Buckley for their very thorough and constructive reviews. MTH and POL would like to acknowledge the support provided by the Alfred P. Sloan Foundation/National Science Foundation award 98-4-5 ME to POL.

References

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

Brooks, S. P., & Gelman, A. 1998. General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics* **7**(4): 434–455.

Buckley, T. R. 2002. Model misspecification and probabilistic tests of topology: evidence from empirical data sets. *Syst. Biol.* **51**: 509–523.

Collin, R. 2003. Phylogenetic relationships among calyptraeid gastropods and their implications for the biogeography of marine speciation. *Systematic Biology* **52**(5): 618–640.

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

DeBry, R. W. 2003. Identifying conflicting signal in a multigene analysis reveals a highly resolved tree: the phylogeny of Rodentia. *Systematic Biology* **52**(5): 604–617.

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

Felsenstein, J. 2003. *Inferring Phylogenies.* Sunderland, Massachusetts: Sinauer Associates.

Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. 1995. *Bayesian Data Analysis.* Chapman & Hall.

Green, P. J. 1995. Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika* **82**: 711–732.

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

Holder, M. T., Lewis, P. O., Swofford, D. L., & Larget, B. in prep.. Hastings ratios of multi-parameter MCMC proposals used in phylogenetics. *Systematic Biology.*

Huelsenbeck, J. P., & Ronquist, F. R. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**: 1572–1574.

Jordan, S., Simon, C., & Polhemus, D. 2003. Molecular Systematics and Adaptive Radiation of Hawaii's Endemic Damselfly Genus Megalagrion. *Systematic Biology* **52**(1): 89–109.

Jukes, T. H., & Cantor, C. R. 1969. Evolution of protein molecules. Pages 21–132 *in*: *Mammalian Protein Metabolism* (Munro, H. N., ed.). Academic Press, New York.

Koepfli, K.-P., & Wayne, R. K. 2003. Type I STS markers of more informative than cytochrome b in phylogenetic reconstruction of the mustelidae (Mammalia: Carnivora). *Systematic Biology* **52**(5): 571–593.

Larget, B., & Simon, D. L. 1999. Markov chain monte carlo algorithms for the Bayesian analysis of phylogenetic trees. *Mol. Biol. Evol.* **16**: 750–759.

Li, S., Pearl, D. K., & Doss, H. 2000. Phylogenetic tree construction using Markov Chain Monte Carlo. *Journal of the American Statistical Association* **95**(450): 493–508.

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

Neal, R. M. 2003. Slice sampling. *Ann. Stat.* **31**: 705–767.

Rambaut, A, & Grassly, N. C. 1997. Seq-Gen: an application for the Monte Carlo simulation of DNA sequence evolution along phylogenetic trees. *CABIOS* **13**: 235–238.

Shoup, S., & Lewis, L. 2003. Polyphyletic origin of parallel basal bodies in swimming cells of chlorophycean green algae (Chlorophyta). *J. Phycol.* **39**: 789–796.

Steppan, S. J., Adkins, R. M., & Anderson, J. 2004. Phylogeny and divergence-date estimates of rapid ratiations in muroid rodents based on multiple nuclear genes. *Systematic Biology* **53**(4): 533–553.

Suchard, M. A., Weiss, R. E., & Sinsheimer, J. S. 2001. Bayesian selection of continuous-time markov chain evolutionary models. *Mol. Biol. Evol.* **18**: 1001–1013.

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

Swofford, D. L. 2001. *PAUP*: Phylogenetic Analysis Using Parsimony (and other methods), version 4.0b10*. Sinauer Associates, Sunderland, Massachusetts.

Waddell, P. J., Kishino, H., & Ota, R. 2002. Very fast algorithms for evaluating the stability of ML and Bayesian phylogenetic trees from sequence data. *Genome Informatics* **13**: 82–92.

Table 1: *Results for 25 replicate simulations using the star tree of Figure . Data sets were composed of a single nucleotide site or 100,000 nucleotide sites. The Jukes & Cantor (1969) model (no rate heterogeneity) was used for the simulations as well as the analysis, and each of the four edges had length 0.05 expected substitutions per site. For each replicate, the largest posterior probability of the three is in bold.*

Sequence	1 site			100,000 sites		
Replicate	(1,2,(3,4))	(1,3,(2,4))	(1,4,(2,3))	(1,2,(3,4))	(1,3,(2,4))	(1,4,(2,3))
1	**0.3405**	0.3242	0.3353	0.2990	0.3288	**0.3722**
2	**0.3410**	0.3279	0.3311	0.3172	0.0464	**0.6364**
3	0.3308	0.3327	**0.3365**	0.1584	**0.7969**	0.0447
4	0.3318	0.3329	**0.3353**	**0.4625**	0.3600	0.1775
5	0.3302	**0.3367**	0.3331	**0.7077**	0.0881	0.2042
6	**0.3380**	0.3343	0.3277	0.0884	0.0262	**0.8854**
7	0.3324	**0.3357**	0.3319	**0.9551**	0.0422	0.0027
8	0.3348	0.3293	**0.3359**	0.1826	**0.5511**	0.2663
9	0.3344	0.3291	**0.3365**	0.3043	**0.4224**	0.2733
10	**0.3367**	0.3320	0.3313	**0.6559**	0.0707	0.2734
11	0.3355	0.3262	**0.3383**	0.0073	**0.9892**	0.0035
12	**0.3367**	0.3318	0.3315	0.3029	0.2922	**0.4049**
13	**0.3358**	0.3310	0.3332	**0.4607**	0.1362	0.4031
14	**0.3466**	0.3211	0.3323	**0.6704**	0.0975	0.2321
15	**0.3429**	0.3295	0.3276	**0.6120**	0.1852	0.2028
16	0.3216	0.3390	**0.3394**	**0.3605**	0.3570	0.2825
17	0.3318	0.3321	**0.3361**	**0.5455**	0.2505	0.2040
18	0.3369	**0.3374**	0.3257	0.4253	**0.4254**	0.1493
19	**0.3348**	0.3332	0.3320	0.1595	**0.7465**	0.0940
20	0.3284	**0.3407**	0.3309	**0.4436**	0.1697	0.3867
21	0.3288	**0.3370**	0.3342	**0.3994**	0.3904	0.2102
22	0.2660	**0.4605**	0.2735	0.1151	**0.5912**	0.2937
23	0.3345	**0.3358**	0.3297	**0.8333**	0.0951	0.0716
24	0.3355	0.3267	**0.3378**	**0.8317**	0.0736	0.0947
25	**0.3430**	0.3339	0.3231	0.2703	**0.4112**	0.3185

Table 2: *The numbers of tree topologies and prior distribution across resolution classes for the three prior strategies "Flat", "Polytomy" and "Geometric". m is the number of internal nodes characterizing tree topologies in each resolution class, and ranges from 1 (star tree) to $n-2=17-2=15$ (fully-resolved tree topologies). The first column contains the logarithm (base 10) of the number of distinct, unrooted tree topologies $T_{n,m}$ in each resolution class for $n=17$ taxa. The total number of tree topologies for $n=17$ is $10^{17.377514}$. For the "Geometric" prior, three separate values of the parameter p were considered: 0.1 (weakly favors polytomies), 0.5 and 0.9 (strongly favors polytomies).*

Resolution class (m)	$\log_{10}(T_{n,m})$	Flat	Polytomy	Geometric $p=0.1$	Geometric $p=0.5$	Geometric $p=0.9$
1	0.000000	0.000000	0.000000	0.125927	0.500015	0.900000
2	4.816361	0.000000	0.000000	0.113335	0.250008	0.090000
3	7.801126	0.000000	0.000001	0.102001	0.125004	0.009000
4	10.002016	0.000000	0.000047	0.091801	0.062502	0.000900
5	11.729312	0.000002	0.000931	0.082621	0.031251	0.000090
6	13.118317	0.000055	0.008388	0.074359	0.015625	0.000009
7	14.240733	0.000730	0.040907	0.066923	0.007813	0.000001
8	15.138396	0.005766	0.118896	0.060231	0.003906	0.000000
9	15.836311	0.028761	0.218166	0.054208	0.001953	0.000000
10	16.348281	0.093490	0.260894	0.048787	0.000977	0.000000
11	16.679054	0.200235	0.205562	0.043908	0.000488	0.000000
12	16.824003	0.279569	0.105584	0.039517	0.000244	0.000000
13	16.765475	0.244321	0.033945	0.035566	0.000122	0.000000
14	16.460717	0.121117	0.006190	0.032009	0.000061	0.000000
15	15.791711	0.025954	0.000488	0.028808	0.000031	0.000000

Table 3: *Posterior split probabilities resulting from applying five different prior distributions for topologies (Table 2) to the Chlamydomonadales data. The first column identifies edges by the numbers used in Figure . The second and third columns are support values (expressed as percentages) from a maximum likelihood bootstrap analysis (500 replicates) and conventional Bayesian MCMC analysis, respectively. In the two edges 7 and 11, the posterior probability is substantially less under the reversible-jump MCMC method introduced here compared to conventional MCMC that does not allow polytomies. All analyses used the HKY+Γ model.*

Edge	ML Bootstrap	Conventional MCMC	Flat prior	Polytomy prior	Geometric resolution class priors		
					$p = 0.1$	$p = 0.5$	$p = 0.9$
1	62	99	88	83	91	87	41
2	50	100	100	100	100	100	100
3	56	99	96	91	96	92	68
4	100	100	100	100	100	100	100
5	100	100	100	100	100	100	100
6	100	100	100	100	100	100	100
7	45	95	5	3	7	4	2
8	99	100	100	100	100	100	100
9	100	100	100	100	100	100	100
10	100	100	100	100	100	100	100
11	71	94	2	1	4	2	0
12	91	100	100	100	100	100	100
13	100	100	100	100	100	100	100
14	100	100	100	100	100	100	100

Table 4: *Summary of the tree topologies composing the 95% credible set from the analysis involving a flat prior over all topologies (including unresolved topologies). The posterior probabilities of the 5 tree topologies listed sum to 0.9587. The number of samples (out of 20,000 total) in which a topology was found is given in the column labeled Frequency. A sojourn is a consecutive series of samples in which only the topology of interest was found. The number of sojourns is thus the number of times the topology appeared in the sequence of samples immediately following a different topology. All of the topologies in the 95% credible set were identical to the topology shown in Figure except for one or more missing edges, which are listed in the last column.*

Topology	Frequency	Posterior Probability	Number of sojourns	Maximum sojourn	Average sojourn	Edges missing
1	15958	0.798	588	148	27.1	7, 11
2	1996	0.100	166	68	12.0	1, 7, 11
3	489	0.024	286	8	1.7	11
4	374	0.019	162	12	2.3	7
5	357	0.018	37	40	9.6	3, 7, 11

Figure 1 — Star tree used for simulations to exemplify the Bayesian Star Tree Paradox.

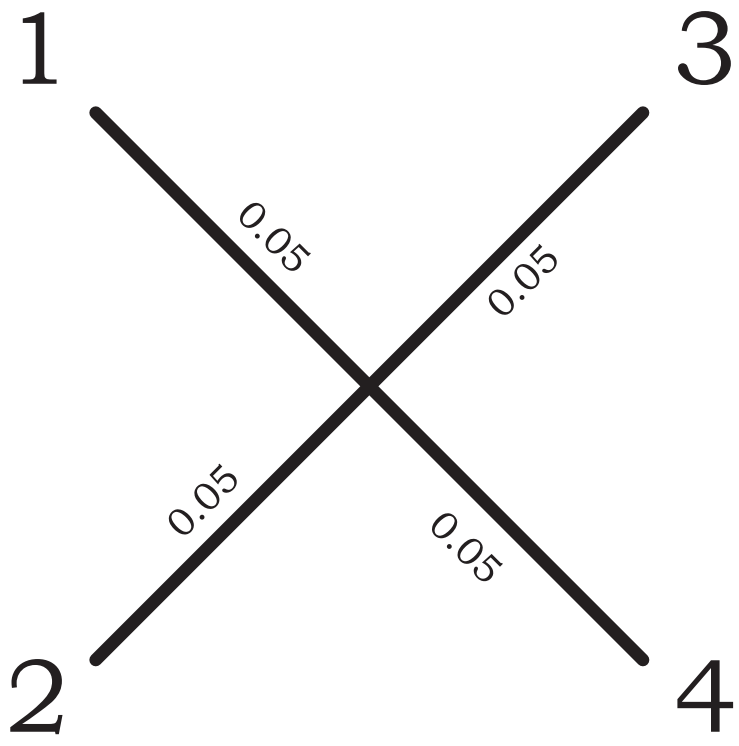
Figure 2 — Series of simulations showing concentration of the posterior distribution for the parameter p near the true value 0.5 as sample size increases. (a) 6 heads observed in 10 flips. (b) 48 heads observed in 100 flips. (c) 475 heads observed in 1000 flips. In each case a flat (uniform) prior distribution for p was assumed.

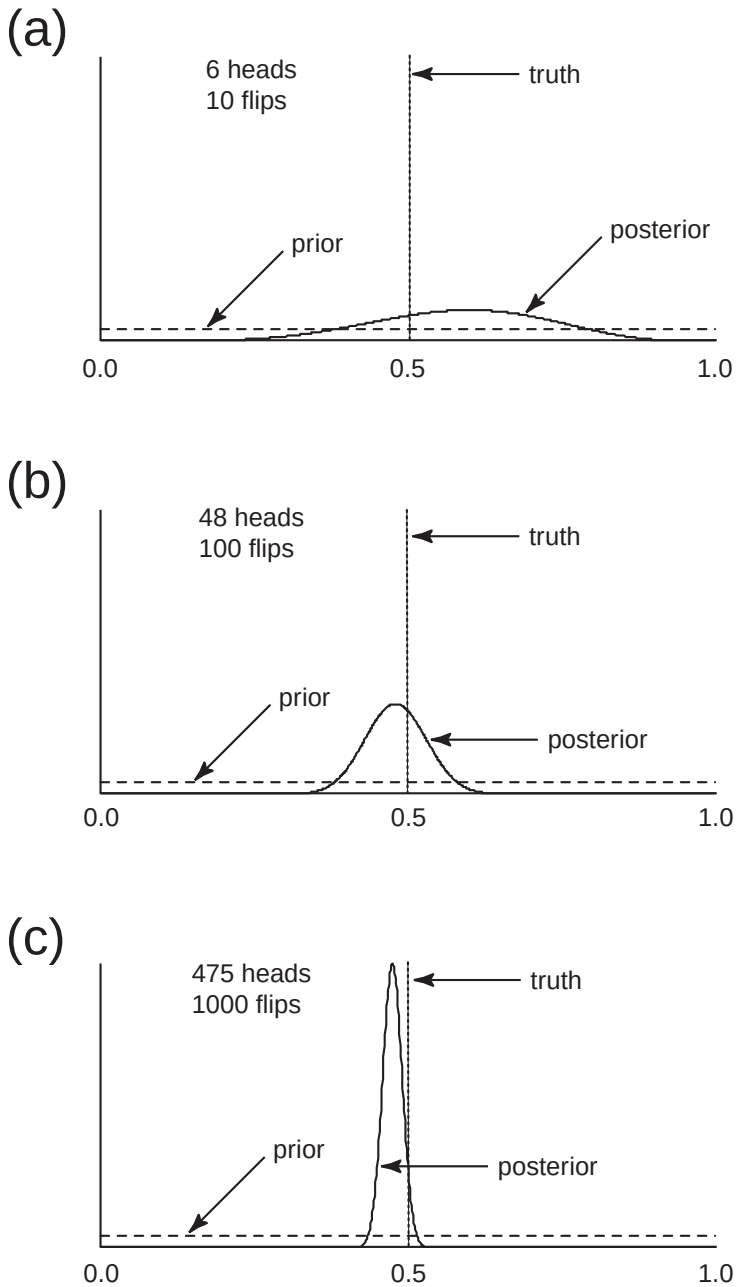
Figure 3 — The two dimension-changing moves. (a) An Add-edge move creates a new internal edge and breaks up an existing polytomy. (b) A Delete-edge move deletes an existing internal edge and either creates a new polytomy or enlarges an existing one.

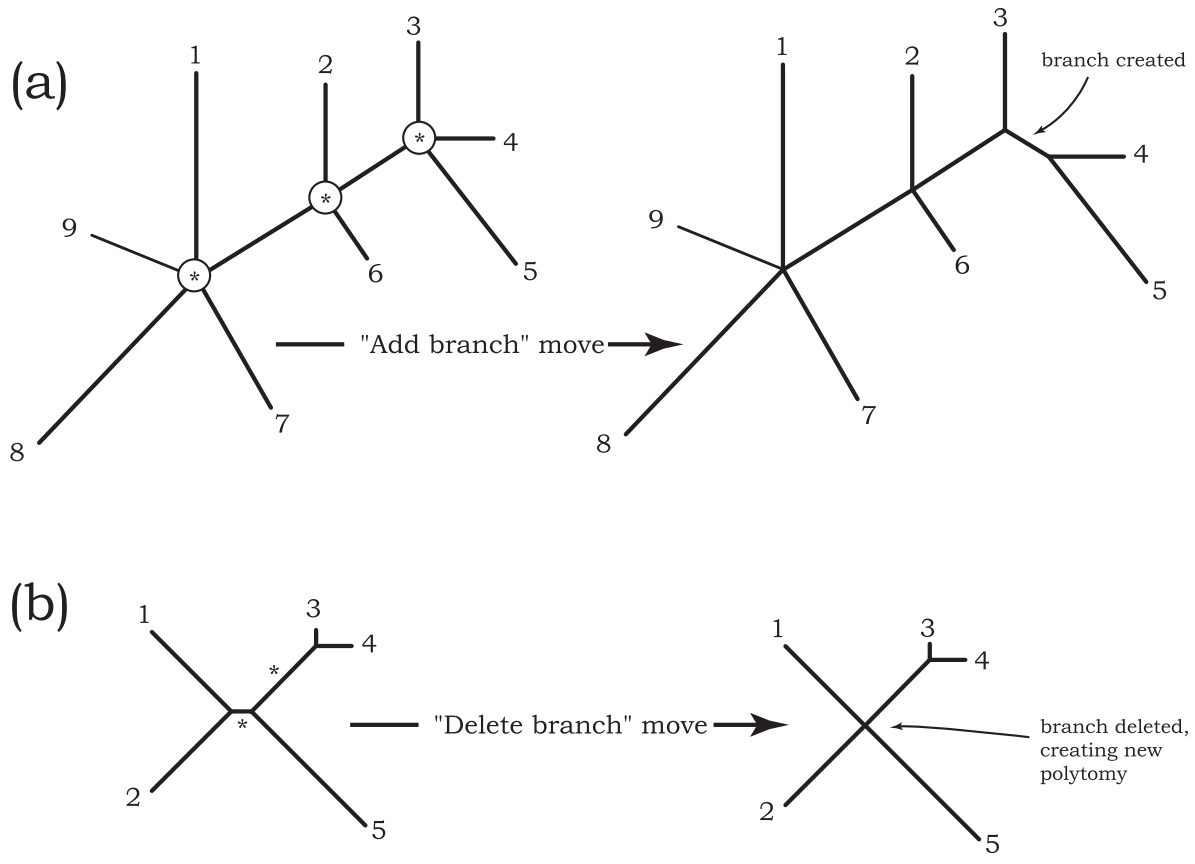
Figure 4 — Tree resulting from the conventional Bayesian analysis of the Chlamydomadales clade from the (Shoup & Lewis, 2003) study. The edge lengths are posterior means, and the topology is the 50% majority rule consensus tree. Node numbers indicated for the interior nodes are the same as the node numbers in Table 3. The two nodes circled are edges that are strongly favored only by the conventional Bayesian MCMC analysis; nonparametric maximum likelihood bootstrapping or Bayesian reversible-jump MCMC analyses (under a variety of topology prior distributions) provide only weak support for these two edges.

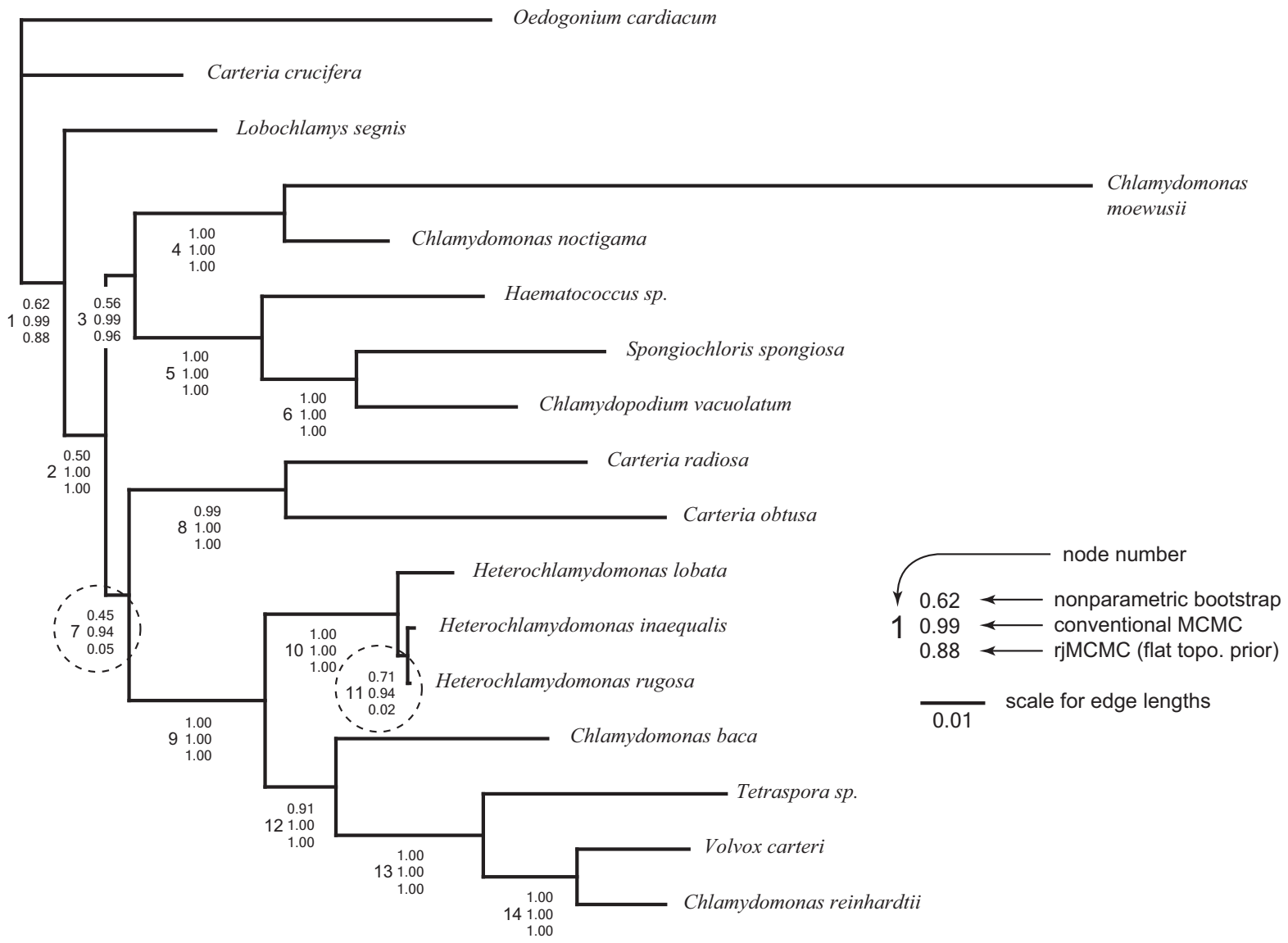
Figure 5 — Scaled Regeneration Quantile (SRQ) plots showing adequacy of topological mixing in the analysis involving a flat prior over all topologies (including unresolved topologies). The topology tracked is the maximum posterior probability topology, which is topology 1 in Table 4. This topology had posterior probability 0.789, and was missing edges 7 and 11 from the tree topology illustrated in Figure . The horizontal axis in each case is T_i/T_N, where T_i is the number of times the tree topology was sampled in the first i samples. The vertical axis in each case is i/N, where i is the sample and N the total number of samples considered. The first 1000 samples were discarded as burn-in, leaving 20000 samples total. (a) First 200 samples after burn-in. (b) First 2000 samples after burn-in. (c) All 20000 samples after burn-in.

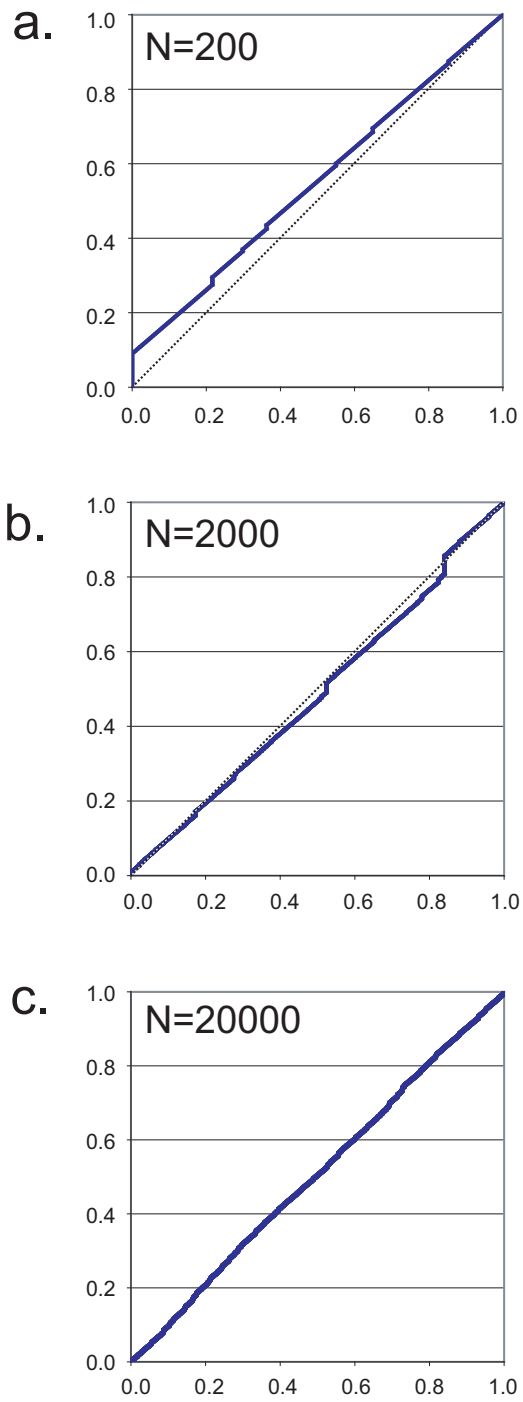
Figure 6 — Box plot showing performance of the reversible jump MCMC approach with the "Polytomy" prior at detecting short edge lengths. The topology prior was parameterized such that ln(prior for topology with m internal nodes) - ln(prior for topology with $m + k$ internal nodes) = k. The length of edge 7 in the tree of Figure was varied from 0.0 to 0.003 in increments of 0.0006. For each of these edge lengths, 100 simulated data sets were generated using the HKY85 model with all other parameters (i.e. base frequencies, other edge lengths, transition/transversion rate ratio) set to their maximum likelihood estimates based on the actual sequence data. Sequence length for the simulated dataset was the same (3341 sites) as the actual data. The open box for each edge length depicts the 25th (bottom edge), 50th (middle line) and 75th (top edge) percentile, while the whiskers extend down to the 10th and up to the 90th percentiles. The filled box represents the mean posterior clade probability over the 100 replicates.

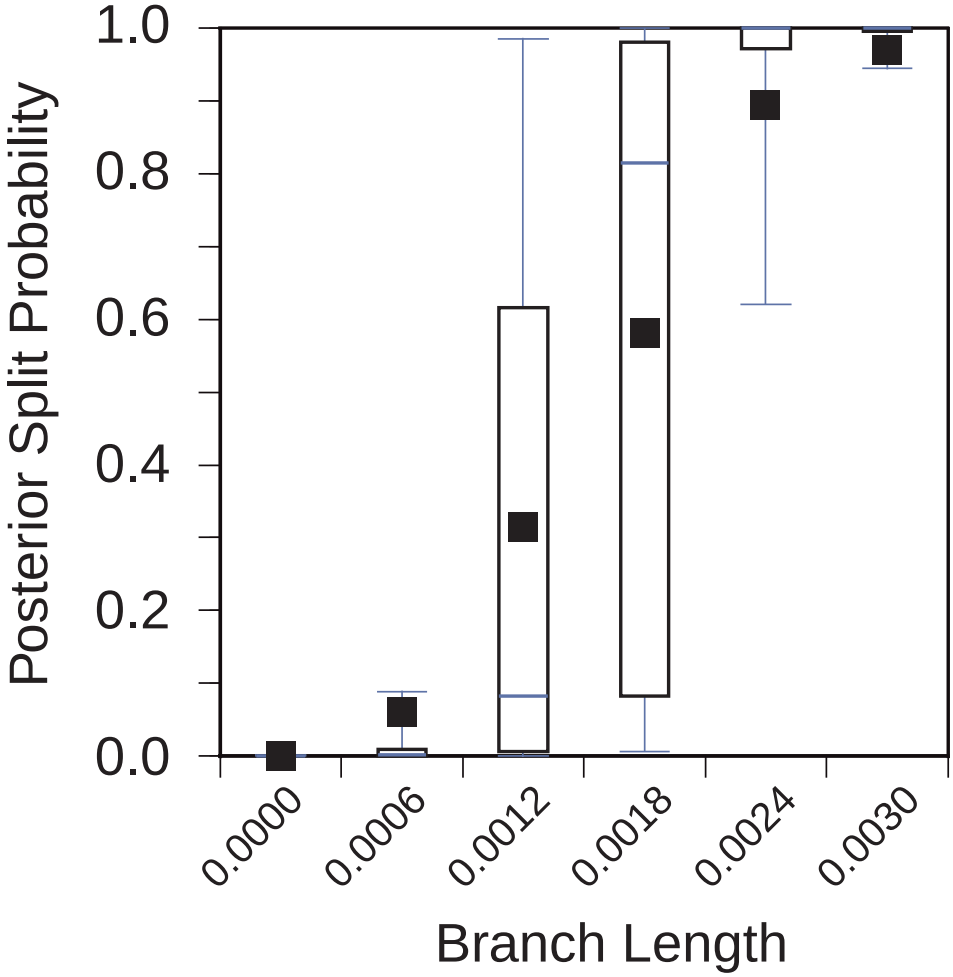

1
3
0.05
0.05
0.05
0.05
2
4


(a)
6 heads
10 flips
truth
prior
posterior
0.0
0.5
1.0
(b)
48 heads
100 flips
truth
prior
posterior
0.0
0.5
1.0
(c)
475 heads
1000 flips
truth
prior
posterior
0.0
0.5
1.0


(a)
1
2
3
4
5
6
7
8
9
"Add branch" move
branch created
(b)
"Delete branch" move
branch deleted,
creating new
polytomy


Oedogonium cardiacum
Carteria crucifera
Lobochlamys segnis
Chlamydomonas moewusii
Chlamydomonas noctigama
Haematococcus sp.
Spongiochloris spongiosa
Chlamydopodium vacuolatum
Carteria radiosa
Carteria obtusa
Heterochlamydomonas lobata
Heterochlamydomonas inaequalis
Heterochlamydomonas rugosa
Chlamydomonas baca
Tetraspora sp.
Volvox carteri
Chlamydomonas reinhardtii
1 0.62 0.99 0.88
2 0.50 1.00 1.00
3 0.56 0.99 0.96
4 1.00 1.00 1.00
5 1.00 1.00 1.00
6 1.00 1.00 1.00
7 0.45 0.94 0.05
8 0.99 1.00 1.00
9 1.00 1.00 1.00
10 1.00 1.00 1.00
11 0.71 0.94 0.02
12 0.91 1.00 1.00
13 1.00 1.00 1.00
14 1.00 1.00 1.00
node number
0.62 nonparametric bootstrap
1 0.99 conventional MCMC
0.88 rjMCMC (flat topo. prior)
scale for edge lengths
0.01


a.
N=200
b.
N=2000
c.
N=20000
0.0
0.2
0.4
0.6
0.8
1.0


1.0
0.8
0.6
0.4
0.2
0.0
Posterior Split Probability
0.0000
0.0006
0.0012
0.0018
0.0024
0.0030
Branch Length