

Effects of Branch Length Uncertainty on Bayesian Posterior Probabilities for Phylogenetic Hypotheses

Bryan Kolaczkowski and Joseph W. Thornton

Center for Ecology and Evolutionary Biology, University of Oregon

In Bayesian phylogenetics, confidence in evolutionary relationships is expressed as posterior probability—the probability that a tree or clade is true given the data, evolutionary model, and prior assumptions about model parameters. Model parameters, such as branch lengths, are never known in advance; Bayesian methods incorporate this uncertainty by integrating over a range of plausible values given an assumed prior probability distribution for each parameter. Little is known about the effects of integrating over branch length uncertainty on posterior probabilities when different priors are assumed. Here, we show that integrating over uncertainty using a wide range of typical prior assumptions strongly affects posterior probabilities, causing them to deviate from those that would be inferred if branch lengths were known in advance; only when there is no uncertainty to integrate over does the average posterior probability of a group of trees accurately predict the proportion of correct trees in the group. The pattern of branch lengths on the true tree determines whether integrating over uncertainty pushes posterior probabilities upward or downward. The magnitude of the effect depends on the specific prior distributions used and the length of the sequences analyzed. Under realistic conditions, however, even extraordinarily long sequences are not enough to prevent frequent inference of incorrect clades with strong support. We found that across a range of conditions, diffuse priors—either flat or exponential distributions with moderate to large means—provide more reliable inferences than small-mean exponential priors. An empirical Bayes approach that fixes branch lengths at their maximum likelihood estimates yields posterior probabilities that more closely match those that would be inferred if the true branch lengths were known in advance and reduces the rate of strongly supported false inferences compared with fully Bayesian integration.

Introduction

A central goal of statistics is to help us express how certain we should be when we draw an inference from evidence. Phylogenies provide the framework for all valid comparative biology, so reliable measures of statistical confidence in evolutionary trees have long been sought. Bayesian phylogenetics (Rannala and Yang 1996; Huelsenbeck et al. 2001) expresses statistical support in terms of posterior probability, defined as the probability that a tree is correct given the data, a model of the evolutionary process, and prior probability distributions over trees and model parameters (Huelsenbeck et al. 2002; Alfaro and Holder 2006). Bayesian Markov Chain Monte Carlo (BMCMC) algorithms allow large phylogenies to be efficiently inferred using complex evolutionary models, and posterior probabilities seem to provide an intuitively meaningful measure of statistical confidence. As a result, Bayesian methods are now widely applied in phylogenetics and have resolved difficult and long-standing problems (Karol et al. 2001; Murphy et al. 2001).

Concerns have been raised that posterior probabilities may not provide reliable indicators of statistical confidence. Although a number of studies support the general reliability of the Bayesian approach—while cautioning that results can be sensitive to model misspecification (Buckley 2002; Wilcox et al. 2002; Alfaro et al. 2003; Erixon et al. 2003; Huelsenbeck and Rannala 2004; Lemmon and Moriarty 2004; Mar et al. 2005)—others have directly challenged these findings, concluding that posterior probabilities are regularly “overcredible” and produce high rates of false inferences even when the correct model is used

(Suzuki et al. 2002; Cummings et al. 2003; Douady et al. 2003; Misawa and Nei 2003; Simmons et al. 2004; Taylor and Piel 2004; Lewis et al. 2005). As a result of these conflicting findings, the confidence we should have in phylogenies inferred using Bayesian techniques is unclear. Further, if posterior probabilities are unreliable indicators of statistical confidence even when the correct model is used, the reasons for this behavior are not known.

Much of the controversy surrounding the reliability of posterior probabilities in phylogenetics may stem from a misunderstanding of what posterior probabilities mean and how they should be interpreted. Bayesian posterior probability represents the degree of subjective belief a rational agent should have in a hypothesis, given the data and his or her prior beliefs not only about the hypothesis in question but also about the values of nuisance parameters. Numerous studies have compared posterior probabilities on trees with the frequentist chance that those trees are correct—the fraction of inferred trees, over a large number of analyses, that are true (Wilcox et al. 2002; Erixon et al. 2003; Misawa and Nei 2003; Huelsenbeck and Rannala 2004; Simmons et al. 2004; Taylor and Piel 2004; Yang and Rannala 2005). Posterior probabilities are not necessarily expected to match the long-run proportion of correct inferences, because 1) posterior probabilities are conditioned on prior assumptions that may be incorrect or incomplete, whereas the proportion of correct trees is not, and 2) posterior probabilities are calculated directly from the data at hand and do not require replication, whereas calculating the percent of correctly inferred trees requires a large number of inferences to be made from different data sets.

Only under restricted conditions is the average posterior probability of a group of inferences expected to equal the proportion of correct inferences in the group. The first condition is replication. Although evolutionary history happened once, computer simulations allow multiple replicate data sets to be generated from any conceivable set of evolutionary conditions, so the proportion of correct inferences can be calculated. The second condition is that the chance

Key words: Bayesian phylogenetics, posterior probability, branch length prior, empirical Bayes.

E-mail: joet@uoregon.edu.

Mol. Biol. Evol. 24(9):2108–2118. 2007

doi:10.1093/molbev/msm141

Advance Access publication July 17, 2007

of choosing each set of evolutionary parameter values to generate data must be known in advance and used as prior information (i.e., the “true priors” must be used). When both conditions hold, the average posterior probability of a group of inferences is equal to the proportion of those inferences that are correct. This correspondence follows directly from Bayes’ theorem (see Supplementary Material online) and has been empirically demonstrated using phylogenetic simulations (Huelsenbeck and Rannala 2004; Yang and Rannala 2005). When the true priors are known, posterior probability can therefore be safely interpreted as the frequentist probability that a hypothesis is correct.

In real analyses, the true values of model parameters are never known in advance and cannot be used as priors. Bayesian methods incorporate this uncertainty by integrating over many possible parameter values, weighted by prior beliefs about the probability of each. As a result, the posterior probability calculated by integration over parameter values may differ from the frequentist probability that a tree is correct. Little is known about the effects of different prior assumptions on posterior probabilities. In the only study to date addressing this question, Yang and Rannala (2005) simulated binary sequence data on rooted 3-taxon trees assuming a molecular clock with branch lengths drawn from exponential distributions. They analyzed these data assuming separate exponential priors for terminal and internal branch lengths and compared the average posterior probability of a group of trees with the proportion correct. When the same distributions used to generate data were used as priors, posterior probability predicted the proportion correct. However, when the true distribution was used as a prior on terminal branches but the mean of the prior distribution on the internal branch length was greater than the true mean, posterior probabilities were higher than the proportion of correct trees; when the mean of the internal branch length prior was less than the true mean, posterior probabilities were lower than the fraction correct.

The experiments of Yang and Rannala (Y&R) established that the choice of branch length priors can affect posterior probabilities and that the special correspondence between posterior probability and proportion correct is not always robust to the use of certain priors. Several key questions remain unresolved, however. First, Y&R considered only a single pattern of branch lengths; different branch length patterns might interact with prior assumptions to produce different effects. Second, although Y&R examined various priors for the internal branch, a separate prior—always the true distribution used to simulate data—was independently assigned to terminal branches. Most widely available Bayesian phylogenetics software use a single prior distribution for all branches on the tree; how the kinds of branch length priors used in most real analyses affect posterior probabilities is unknown. Third, it has been common to use a uniform prior distribution with a large upper bound on branch lengths to represent ignorance about this parameter; because such a prior will usually overestimate mean branch lengths, Y&R predicted that flat priors would produce excessively high posterior probabilities. Whether flat branch length priors actually yield high posterior probabilities was not tested, however. Finally, the simulations of Y&R represent a peculiar situation in which the phyloge-

netic trees and branch lengths on which sequences were simulated were drawn from sampling distributions. In reality, there is a single-correct evolutionary history; how different prior assumptions affect posterior probabilities under such circumstances is not known.

Here, we use a simulation-based approach to determine how integrating over uncertainty affects posterior probabilities using a range of prior assumptions available in commonly used software packages and under a variety of evolutionary conditions.

Materials and Methods

Bayesian Analyses

Posterior probabilities were estimated by Markov Chain Monte Carlo using MrBayes v3.1 (Ronquist and Huelsenbeck 2003). Four incrementally heated chains (temp = 0.2) were run until the average standard deviation in posterior probability estimates across 2 independent runs was <0.01. Prior distributions on branch lengths were uniform on (0, M] with upper bound $M = 1, 5, 10$, or 100 or exponential with mean $\mu = 10^{-5}, 10^{-3}, 0.01, 0.1, 1.0$, or 10.0. (MrBayes uses $1/\mu$ to parameterize the exponential distribution, so $\mu = 10^{-5}$ corresponds to an input value of 100,000.) The true generating model was used for all BMCMC analyses.

We also conducted Bayesian analyses using 2 approaches that do not require integrating over branch length uncertainty. First, we used an empirical Bayes approach that places prior probability 1.0 on the maximum likelihood branch lengths for each possible topology, which were estimated using PAUP* v4.0b10 (Swofford 2002). Posterior probabilities were then calculated directly from Bayes’ theorem using these branch lengths. Second, for Bayesian simulations (see below), we also used the true point priors on branch lengths, which place prior probability 1.0 on the branch lengths actually used to simulate data.

Simulations

We first performed “Bayesian simulations” (Huelsenbeck and Rannala 2004; Yang and Rannala 2005) using 4-taxon phylogenies with fixed branch lengths (terminals 0.5 substitutions/site, internal 0.01). At each sequence length (100, 1000, and 10,000 nt), 500 replicates were simulated using the JC69 model and a randomly selected topology for each replicate.

To assess the effects of different branch length patterns on inferred posterior probabilities, we simulated 500 replicate data sets of 100 or 10,000 nt on 4-taxon topologies with internal branch length 0.01 and 6 different terminal branch length combinations: 1) all short branches (0.01), 2) 1 long branch (0.75) and 3 short, 3) 3 long and 1 short, 4) all long branches, 5) inverse-Felsenstein-zone lengths, with 2 long sister branches and 2 short sister branches, and 6) Felsenstein-zone branch lengths, with 2 long nonsister branches and 2 short nonsister branches.

To assess the impact of typical branch length prior assumptions on clade probabilities under more realistic conditions, we simulated 100 data sets of various sequence

lengths (1000, 10,000, 25,000, and 50,000) using parameter values drawn from an analysis of real sequence data (Murphy et al. 2001). We estimated the tree topology, branch lengths, and parameters of the General Time Reversible model with invariant sites and gamma-distributed rate variation (GTR+I+G) by BMCMC using the original nucleotide data and then used these conditions to simulate replicate data sets.

Comparing Posterior Probabilities

For each branch length prior, we collected posterior probabilities on trees into 10 equally sized bins and compared the average posterior probability of each bin with the proportion of correct trees in the bin; these values should be equal if nuisance parameters are known with certainty (Huelsenbeck and Rannala 2004; Yang and Rannala 2005). Bins with fewer than 20 trees were excluded to avoid stochastic error in estimating the proportion correct. We also considered inferred trees with posterior probability ≥ 0.95 as strongly supported and calculated the proportion of replicate data sets producing incorrect inferences with strong support using each prior distribution.

Results

Effects of Integrating over Uncertainty Depend on Prior Assumptions and Sequence Length

We first performed “Bayesian simulations” using challenging 4-taxon phylogenies with equal terminal branch lengths and a short internal branch. Data were analyzed using flat priors with various upper bounds and exponential priors with various means. To isolate the effects of integrating over prior uncertainty, we compared these results with 2 kinds of control experiments in which prior distributions that do not require integration were applied. First, we used the true point prior distribution, which places prior probability 1.0 on the actual branch lengths used to simulate the data, producing posterior probabilities given perfect prior knowledge. Second, we used an empirical Bayes approach that places prior probability 1.0 on the maximum likelihood branch length estimates obtained from the data. Empirical Bayes analysis allows us to separate the effects of integrating over uncertainty from those caused by a lack of perfect prior knowledge.

We found that integrating over uncertainty has a strong effect on posterior probabilities compared to control methods in which integration is not employed. When the true values of model parameters were known in advance, obviating the need to integrate over uncertainty, the average posterior probability of a group of hypotheses equalled the fraction of correct hypotheses in the group (fig. 1A), as predicted by Bayes’ theorem, and the rate of false inferences with high posterior probability was close to zero, irrespective of sequence length (fig. 1D).

When branch lengths were not known in advance but were estimated from the data using maximum likelihood, posterior probabilities deviated slightly from the fraction of correct inferences. When sequences were very short, the empirical Bayes technique produced average posterior probabilities for the best-supported trees (posterior proba-

bility $>1/3$) that were higher than the proportion of those trees that were correct; still, posteriors were almost never >0.70 . With longer sequences, posterior probabilities were very close to those inferred when the true branch lengths were known in advance. False inferences with strong support were rare at all sequence lengths (fig. 1A and D).

In contrast, when uncertainty was integrated over using typical prior distributions, posterior probabilities deviated strongly from the fraction of correct inferences. The magnitude of this effect was determined by sequence length and the specific prior distribution used (fig. 1B and C). When sequences were of short or moderate length, all prior distributions produced average posterior probabilities for the best-supported tree that were considerably greater than the proportion of correct trees. When long sequences were analyzed, diffuse branch length priors—uniform distributions or exponential distributions with moderate to large means—produced posterior probabilities that closely matched those inferred when branch lengths were known in advance. Uniform branch length priors with upper bounds from 1 to 100 all produced similar posterior probabilities (fig. 1C). In contrast, even with long sequences, small-mean exponential distributions (10^{-5} – 10^{-2}) produced average posterior probabilities considerably greater than the proportion of correct trees (fig. 1B).

Integrating over uncertainty also increased the rate of strongly supported false inferences (fig. 1D). At all sequence lengths, exponential priors and flat priors produced false inferences more frequently than either of the control priors that do not require integration. Exponential priors performed more poorly in this regard than uniform priors. However, the frequency of such inferences was high only when exponential priors with small means were used. For any given prior, the proportion of false inferences with strong support declined as sequence length increased. Even with 10,000 sites, however, the small-mean exponential priors continued to produce high posterior probabilities for false trees at appreciable rates.

Branch Length Pattern Determines the Effects of Uncertainty

To examine how different evolutionary conditions might effect posterior probabilities, we investigated the role of the true tree’s branch length pattern in modulating the effect of integrating over uncertainty. We simulated data on 4-taxon trees with all possible combinations of long/short terminal branches. We analyzed these data using a flat branch length prior ($U(0, 10)$), the default exponential prior used in MrBayes v3.1 ($\mu = 0.1$), a small-mean exponential prior ($\mu = 10^{-5}$), or the empirical Bayes approach that uses fixed maximum likelihood estimates.

We found that the pattern of branch lengths on the true phylogeny determines both the direction and the magnitude of the effect that prior uncertainty has on posterior probabilities (figs. 2–4). This effect appears to arise from the amount and structure of convergent evolution produced under various branch length patterns. When 1 or none of the 4 terminal branches was long—making convergent state patterns unlikely—inferred posterior probabilities were close to the proportion of correct inferences, and strongly

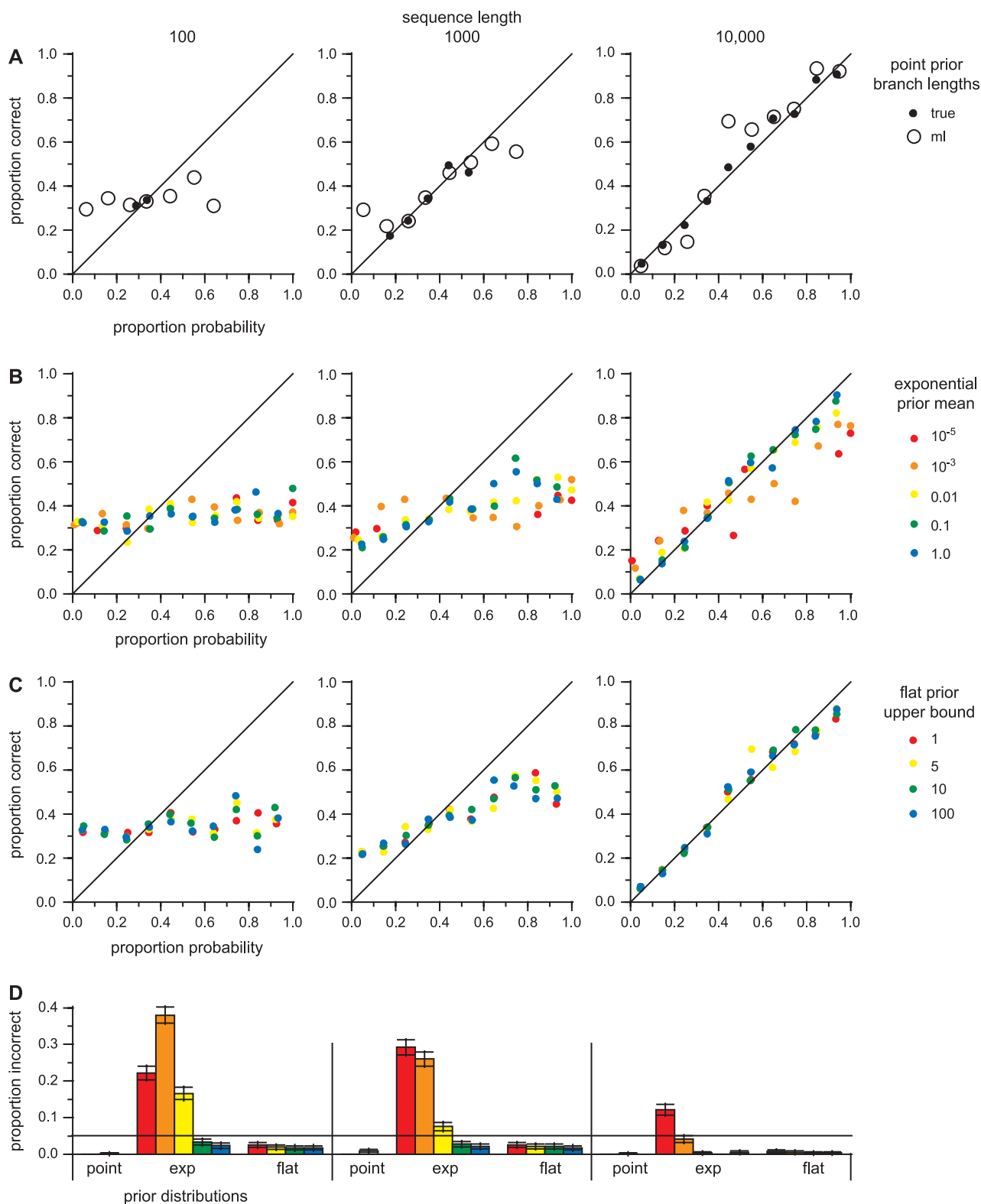


FIG. 1.—Effect of branch length uncertainty on posterior probabilities. Data were generated by “Bayesian simulation” on 4-taxon trees with equal terminal branch lengths. (A–C) Trees inferred from replicate data sets were binned by posterior probability; the average posterior probability of a group of trees is plotted against the proportion of correct trees in the group for various sequence lengths and prior branch length distributions. (A) Point priors placed probability 1.0 on the true branch lengths used to simulate data (black dots) or on the maximum likelihood branch length estimates (open circles). (B) Exponential priors with various means were used. (C) Flat priors with various upper bounds were used. (D) The proportion of replicate data sets giving strong support (≥ 0.95 posterior probability) for incorrect trees is shown for each prior distribution and sequence length. Colors are the same as in panels (A–C); bars indicate standard error (SE), and horizontal line indicates an error rate of 0.05.

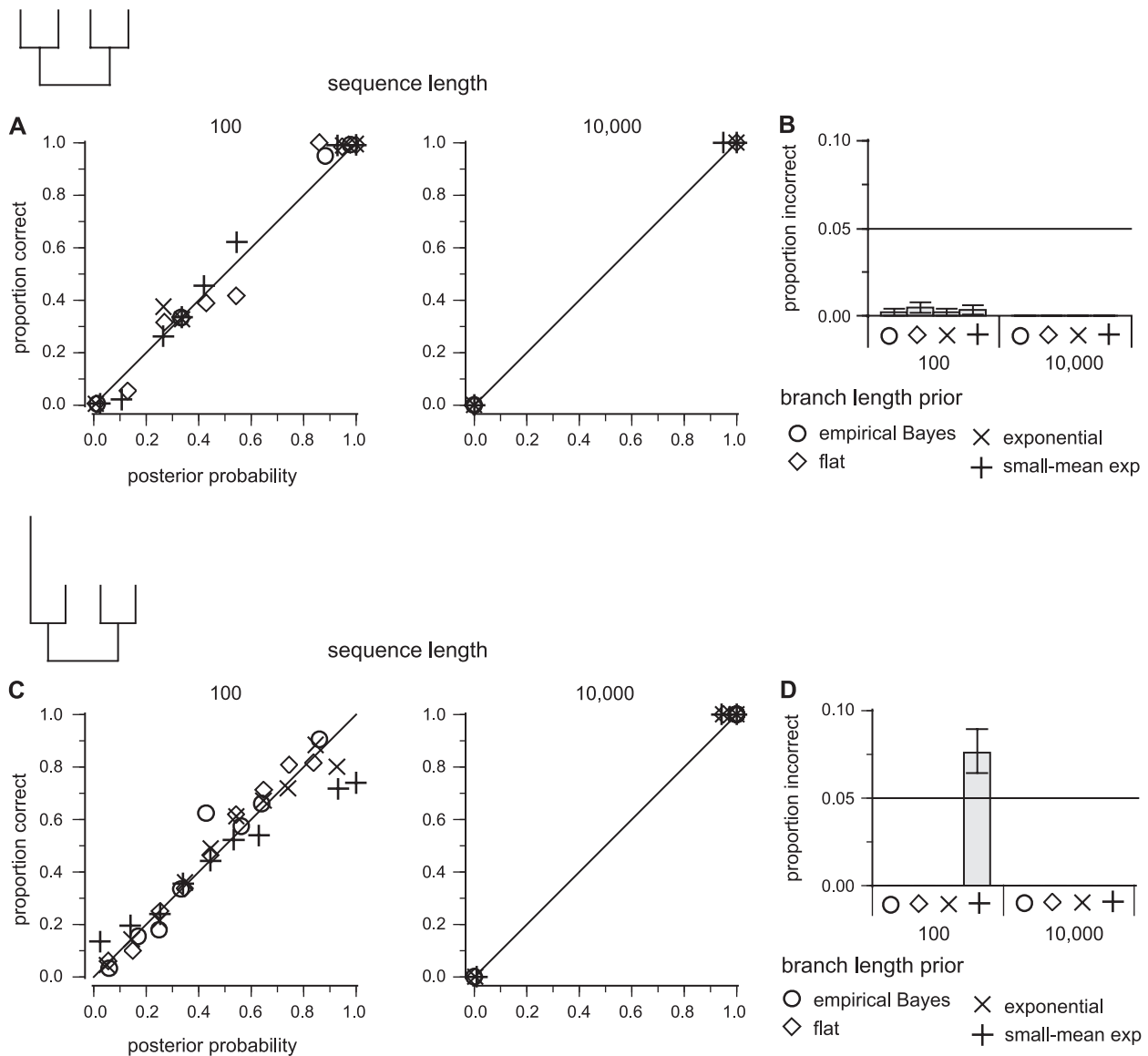


FIG. 2.—Effect of integrating over uncertainty on posterior probability when few terminal branches are long. (A, C) The average posterior probability of each bin of inferred trees is plotted against the proportion of correct trees in the group. (B, D) The proportion of trees falsely resolved with posterior probability ≥ 0.95 is shown; bars indicate SE, and an error rate of 0.05 is indicated by a horizontal line. In (A) and (B), sequences of length 100 or 10,000 were generated on a tree with a short internal branch (0.01 substitutions/site) and all short terminal branches (0.01); in (C) and (D), 1 of the 4 terminal branches is long (0.75). Branch length priors used were an empirical Bayes prior that places prior probability 1.0 on the maximum likelihood branch length estimates (open circles), a flat prior with uniform probability over branch lengths from 0 to 10 substitutions/site (open diamonds), an exponential prior with mean 0.1 (Xs), and a small-mean exponential prior ($\mu = 10^{-5}$, crosses).

supported false inferences were rare (fig. 2). When there was ample opportunity for convergent evolution but no expected structure to the convergence—that is, when 3 or 4 of the terminal branches were long—average posterior probabilities were higher than the proportion of correct trees, and the frequency of strongly supported false inferences increased (fig. 3). In contrast, when convergence was structured to favor a particular tree—as on trees with 2 long and 2 short branches—posterior probabilities were lower than the proportion of correct trees and the rate of false inferences was dependent on the relationship between taxa with long terminal branches (fig. 4). When the true tree was in the inverse-Felsenstein zone (2 sister long branches), the

true tree was always recovered, but posterior probability was reduced (fig. 4A). When the true tree was in the Felsenstein zone (2 nonsister long branches), integrating over incorrect branch lengths reduced support for the correct tree and inflated support for an incorrect tree (fig. 4C–D).

Several of the inferences drawn from the experiments in figure 1 applied across branch length patterns. First, the effect of integrating over uncertainty was again strongest when short sequences were analyzed. Presumably, longer sequences cause the likelihood function over branch lengths to be more narrowly peaked around the true values, so integrating over that function more closely approximates knowing the true values in advance.

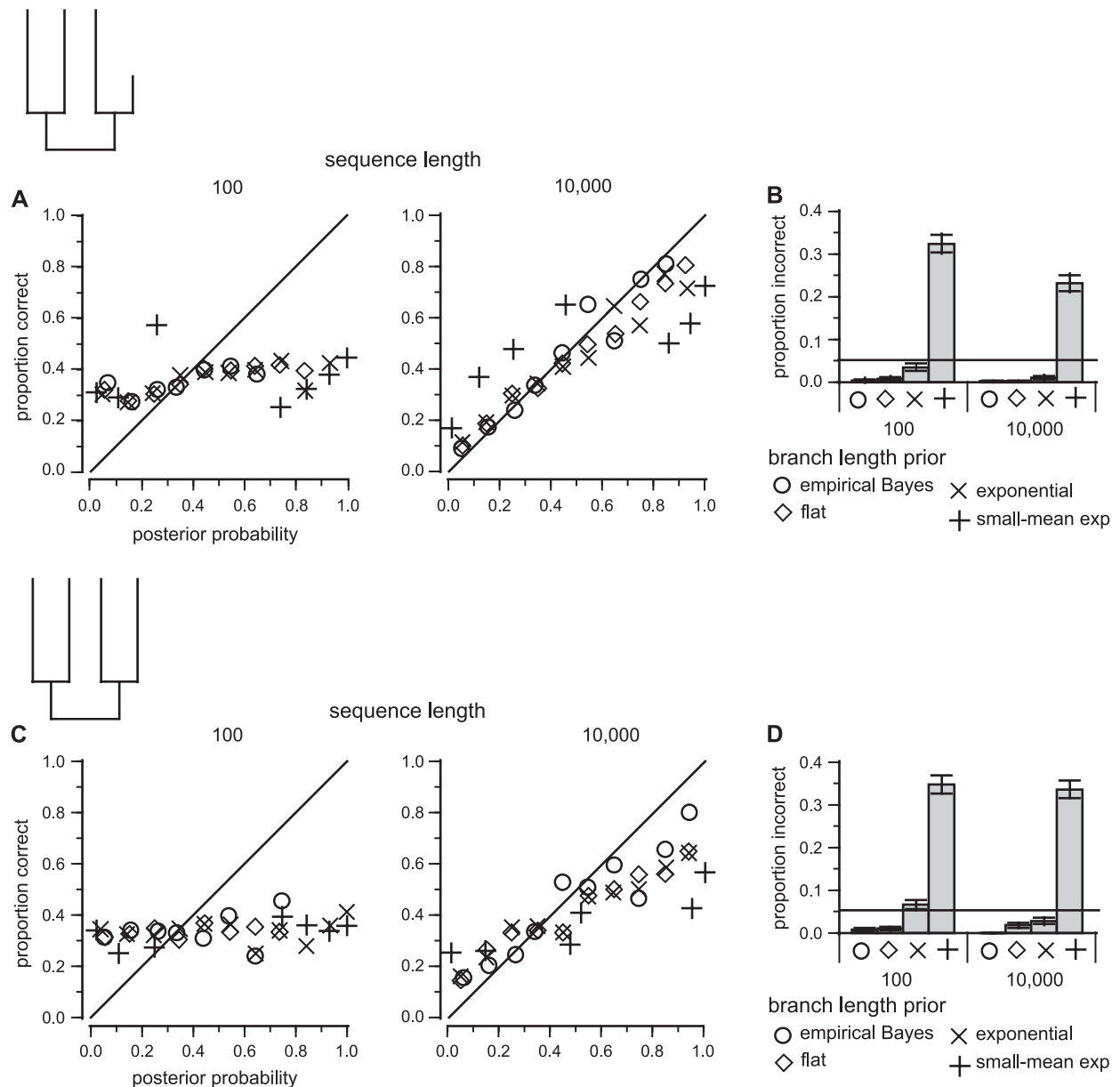


FIG. 3.—Effect of integrating over uncertainty when many terminal branches are long. (A, C) The average posterior probability of each bin of inferred trees is plotted against the proportion of correct trees in the bin. (B, D) The proportion of trees falsely resolved with posterior probability ≥ 0.95 is shown. Sequences of length 100 or 10,000 were generated on 4-taxon phylogenies with short internal branches (0.01 substitutions/site) and either 3 (top) or all 4 (bottom) long terminal branches (0.75). Branch length priors examined were the empirical Bayes prior (open circles), a flat prior uniform over (0,10] (open diamonds), an exponential prior with mean 0.1 (Xs), and a small-mean exponential prior with mean 10^{-5} (crosses).

Second, the small-mean exponential prior was particularly problematic. This prior consistently produced more radical deviations from the proportion of correct inferences than the flat or moderate-mean exponential priors did. Under virtually all conditions, the frequency of strongly supported false inferences was notably higher when the small-mean exponential prior was used than when the flat and moderate-mean exponential priors were applied. Whereas analyzing longer sequences mitigated this effect for the flat and moderate priors, adding more data had a much weaker beneficial effect on posteriors using the small-mean exponential prior. For example, when long sequences were simulated along a tree with 4 long terminal branches and an-

alyzed using the small-mean exponential prior, only 50% of trees with posterior probability > 0.94 were correct, whereas 65% of such trees were correct when more diffuse priors were used (fig. 3C–D). On Felsenstein-zone trees, small-mean exponential priors led to a strong long-branch attraction artifact, with the incorrect tree being inferred with high posterior probability in virtually all replicates, and this effect was most severe with long sequences (fig. 4C–D). Flat and moderate exponential priors suffered from a much weaker bias; the wrong tree was occasionally recovered with strong support, but only from short sequences.

Finally, empirical Bayes analysis produced posterior probabilities that more closely matched the chance that

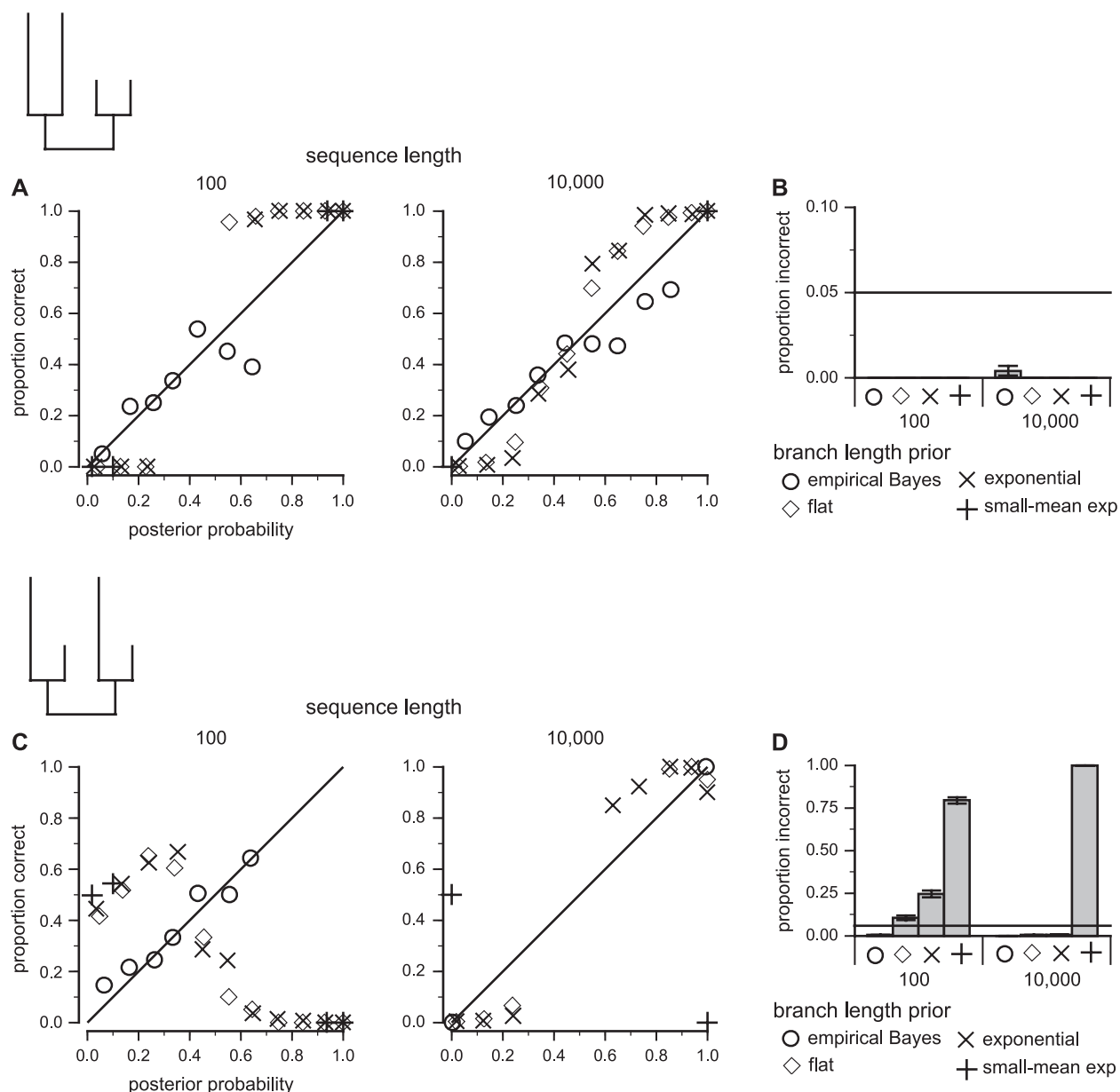


FIG. 4.—Effect of integrating over uncertainty when 2 terminal branches are long. (A, B) Sequences of lengths 100 (left) and 10,000 (right) were generated using inverse-Felsenstein-zone trees with 2 long sister branches (0.75 substitutions/site) and 2 short sister branches (0.01). (C, D) Sequences were generated on Felsenstein-zone trees with nonsister long branches. Priors examined were the empirical Bayes prior placing probability 1.0 on the maximum likelihood branch lengths (open circles), a uniform prior over (0,10] (open diamonds), an exponential prior with mean 0.1 (Xs), and a small-mean exponential prior with mean 10^{-5} (crosses).

an inference was correct than any of the nonpoint priors examined. When 3 or 4 terminal branches were long—conditions that caused posterior probabilities to be higher than the proportion of correct trees—posterior probabilities calculated using the empirical Bayes approach were closer to the proportion of correct trees than any of the nonpoint prior distributions (fig. 3). Even under challenging Felsenstein-zone conditions, the empirical Bayes approach produced posterior probabilities that closely matched the proportion of correct inferences (fig. 4C–D). The rate of strongly supported false trees using empirical Bayes was consistently lower than when branch length uncertainty was integrated over using any of the priors. These results

indicate that it is integration over parameter uncertainty—rather than lack of perfect prior knowledge—that is the chief cause of the observed effects on posterior probabilities.

Uncertainty Affects Posterior Probabilities under Realistic Conditions

To examine the potential effects of branch length uncertainty under more realistic conditions, we simulated data using parameter values inferred from the placental mammal data of Murphy et al. (2001) and analyzed them using BMCMC with various prior distributions.

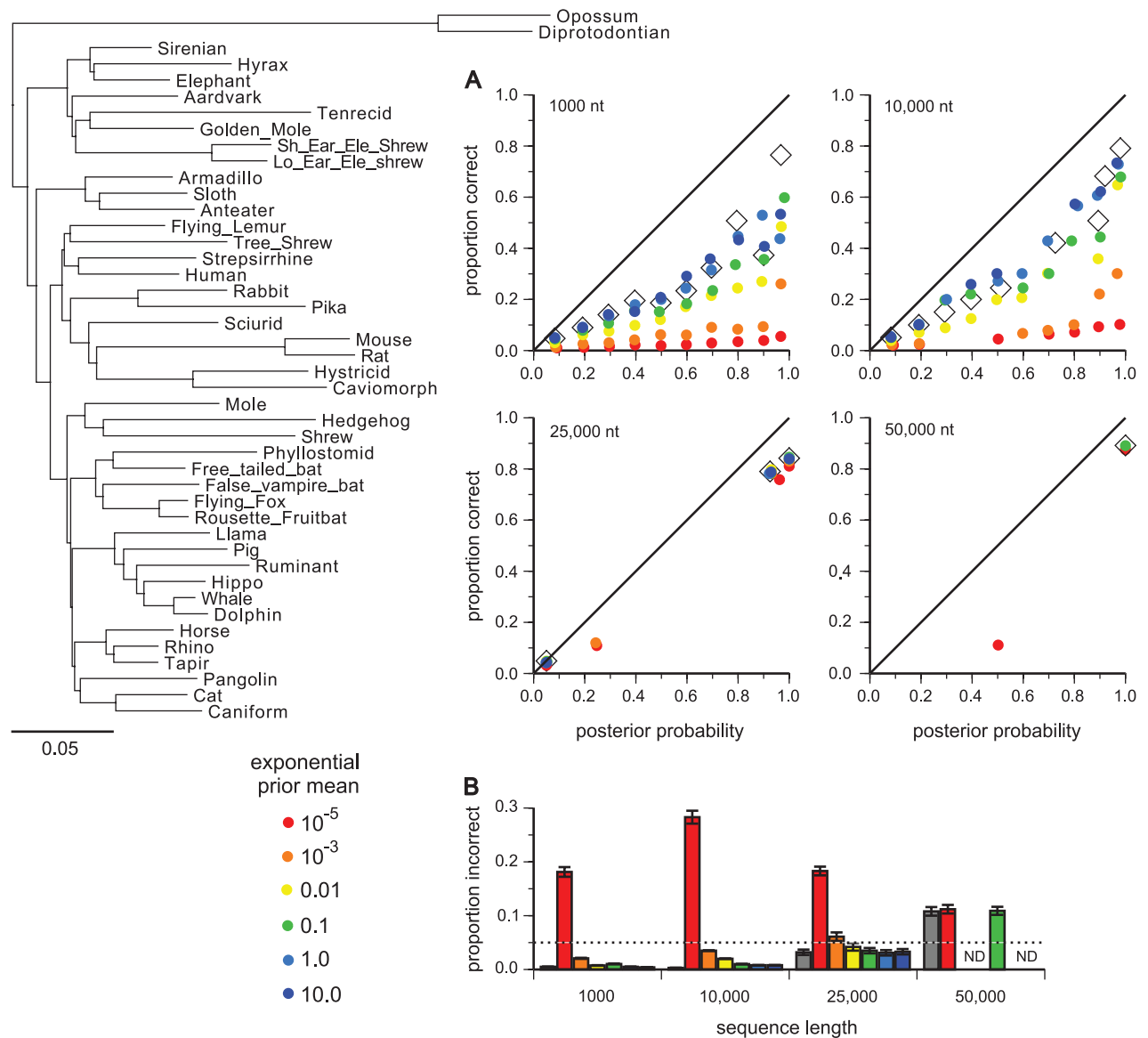


FIG. 5.—Branch length uncertainty affects posterior probabilities under empirically derived conditions. Sequences were simulated on the tree at left assuming a complex evolutionary model with branch lengths and parameters derived from the large mammalian data set of Murphy et al. (2001) and analyzed by BMCMC using the true evolutionary model with either a flat prior on branch lengths (open diamonds) or exponential priors (colored circles, see legend at bottom left) with means ranging from 10^{-5} to 10 substitutions/site. (A) All clades sampled using each prior were binned by their posterior probability, and the fraction of correct clades in each bin was calculated. (B) The proportion of false clades with posterior probability ≥ 0.95 is shown for each branch length prior. Bars indicate SE, and the horizontal line indicates an error rate of 0.05. Due to the computational demands of this experiment, 50,000-nt sequences were analyzed using only the flat prior and exponential priors with small (10^{-5}) or moderate (0.1) means. ND indicates that no data was obtained using the other prior distributions.

At all sequence lengths, integrating over branch length uncertainty using any of the priors caused posterior probabilities to be skewed upward compared with those that would be inferred if branch lengths were known. With increasing sequence length, posterior probabilities for inferred clades converged toward 1.0 using any of the priors, but a large fraction of these strongly supported clades were incorrect (fig. 5). Even with very long sequences (50,000 nt), all clades inferred using flat and moderate-mean exponential priors had posterior probability 1.0, but about 10 % of these were incorrect. The small-mean exponential prior performed similarly, but an additional group of weakly supported clades were also recovered, and virtually all of these

were incorrect. These results suggest that for difficult real-world problems, even extraordinarily long sequences are not sufficient for Bayesian methods that integrate over uncertainty to reliably recover the correct tree and avoid recovering false clades with high support.

Discussion

Bayes theorem guarantees that posterior probabilities would equal the proportion of correct inferences if branch lengths and other nuisance parameters were known a priori. In real analyses, these values can never be known with

certainty. Our experiments show that integrating over these parameters using common “uninformative” priors can cause posterior probabilities to deviate radically from this equivalence. The direction of the deviation—either upward or downward—is determined by the pattern of branch lengths on the tree; the magnitude depends largely on sequence length and prior assumptions. Under realistic conditions derived from empirical sequence analysis, posterior probabilities calculated using typical prior assumptions can produce spurious support for incorrect phylogenies even when the evolutionary model is otherwise correct and sequences are very long.

Our results reinforce that, in practice, Bayesian posterior probabilities should not be misinterpreted as the unconditional probability that a hypothesis is true. In no way does this conclusion suggest that posterior probabilities are technically incorrect; Bayes’ theorem guarantees that the posterior probability of a tree is no more and no less than the probability that the tree is correct given the data, an evolutionary model, and prior distributions over trees and model parameters. Posterior probability accurately expresses the quantitative degree of belief in a hypothesis that a rational being should have, given a set of explicit starting assumptions and following analysis of a set of data. Posterior probabilities must be interpreted strictly in this conditional sense.

Numerous previous analyses have suggested that posterior probabilities are regularly “overcredible,” leading to inflated estimates of statistical confidence. These studies, however, looked at only a small subset of possible tree shapes (Suzuki et al. 2002; Cummings et al. 2003; Douady et al. 2003; Misawa and Nei 2003; Simmons et al. 2004; Taylor and Piel 2004; Lewis et al. 2005). Our experiments show that with some branch length combinations, integrating over uncertainty does cause posterior probabilities to exceed the frequentist chance that a tree or clade is correct. Other branch length combinations, however, produce posterior probabilities that are lower than the chance an inference is true. As a result, there can be no general post hoc “correction”—such as adjusting posterior probabilities downward—that can make posterior probabilities approximate the frequentist chance that a hypothesis is true.

Most phylogenetic analyses to date have used diffuse priors to indicate a lack of prior knowledge about evolutionary parameters like branch lengths. Our results indicate that various branch length distributions affect posterior probabilities in very similar ways, so long as the prior is sufficiently vague. Exponential priors with moderate to large means and flat priors with various upper bounds all produced nearly identical posterior probabilities across a range of problems. Although concern has been raised that the upper bound on the uniform distribution is arbitrary (Felsenstein 2004), we found no evidence that choosing different uniform distributions within a reasonable range appreciably affects posterior probabilities.

More informative priors have the potential to improve estimates of statistical confidence, but only if they accurately reflect prior beliefs about the true evolutionary conditions. We found that small-mean exponential priors, when applied to all branch lengths on a tree, severely underestimate the probability of convergence on long branches, resulting in radically increased posterior probabilities when

many branches are long. A partitioned prior like the one used by Y&R—with a small-mean distribution applied to internal branches and a large mean distribution for terminal branches—increases the relative prior probability of convergence, thus tending to decrease posterior probabilities (Yang and Rannala 2005). This type of prior may be appropriate when examining evolutionary radiations or other situations in which short internal and long terminal branches are expected given prior information. If used generally with the goal of causing posterior probabilities to more closely approximate the frequentist chance that inferences are correct, however, this prior is unlikely to be successful: it would reduce the high posterior probabilities induced by some branch length patterns but exacerbate the lower posterior probabilities caused by others.

When model parameters are expected to follow a particular kind of prior distribution, but the specific parameter values describing that distribution are unknown, uncertainty about the prior itself can be integrated over using a “hyperprior” (Berger et al. 2005). In our experiments using simulated mammalian sequence data, however, exponential priors with a wide range of means all produced posterior probabilities greater than the proportion of correct inferences, presumably because the true branch lengths are not exponentially distributed. These results suggest that hyperpriors may improve performance only when the family of prior distributions includes a reasonable approximation of the true distribution of branch lengths across the tree. The development of prior distributions that are flexible enough to match prior beliefs concerning real phylogenetic problems merits further research.

Compared with a fully Bayesian approach, the empirical Bayes method we investigated yielded more reliable results—a lower rate of strongly supported false inferences and posterior probabilities that more closely approximate those that would be derived with perfect prior knowledge. In nonphylogenetic applications, simple empirical Bayes approaches that do not account for uncertainty can be less reliable than fully Bayesian methods, producing overly narrow confidence intervals that do not account for stochastic error in parameter estimates (Carlin and Gelfand 1990). Phylogenetic inference differs from classical domains of statistical analysis in numerous ways, however, so classical results are not always expected to hold (Yang et al. 1995). Due to the hierarchical structure of phylogenetic trees, incorrect values of nuisance parameters—branch lengths in particular—can cause very strong biases in favor of particular topologies, because incorrect parameter values systematically misinterpret convergent state patterns as due to common descent or vice versa. Model-based phylogenetic methods are unbiased only when the probability of convergence is accurately assessed, a condition strictly fulfilled only when the model and its parameters, including branch lengths, are correct (Rogers 2001). Our results confirm that when the true branch lengths are known in advance, the probability of convergence is accurately assessed and the likelihoods calculated for each tree accurately reflect the relative probabilities that each topology would have generated the data. The empirical Bayes technique we used estimates branch lengths from the data using maximum likelihood and approximates, to the extent those estimates

are accurate, Bayesian analysis if the true lengths were known in advance. Our experiments indicate that these branch length estimates are accurate enough to yield posterior probabilities almost identical to those calculated with perfect prior knowledge, except when sequences were extremely short.

In contrast, the fully Bayesian approach integrates branch lengths over a diffuse distribution of values, virtually all of which are incorrect. Integration over incorrect branch lengths causes the probability of convergence to be systematically under- or overestimated, depending on the topology. As a result, the likelihood of one tree relative to the others is increased, skewing posterior probabilities up- or downward, depending on the phylogeny favored by convergent patterns. Our experiments suggest that the cost of ignoring uncertainty using empirical Bayes—stochastic error in estimating parameter values—is much smaller than the cost of incorporating it using Bayesian integration, which introduces strong biases that even very long sequences may not overcome. Further investigation of the performance of empirical Bayes strategies in phylogenetics is therefore warranted.

In addition to empirical Bayes analysis, other strategies to characterize evidentiary support without conditioning on prior beliefs about nuisance parameters—or relying on bootstrapping, which is computationally intensive and can also be biased (Hillis and Bull 1993; Sitnikova et al. 1995)—are available. These include procedures based on evaluating the likelihood ratio of the best tree with a clade versus the best tree without it (Edwards 1992; Anisimova and Gascuel 2006). Understanding the statistical properties of these and other confidence measures under a variety of conditions deserves further study. Because no single measure is likely to provide a complete and accurate estimate of statistical confidence under all evolutionary conditions, a careful and critical application of a variety of techniques—each evaluated in light of a detailed understanding of its properties—will provide the most robust and meaningful assessments of confidence in phylogenetic hypotheses.

Supplementary Material

Supplementary appendix is available at *Molecular Biology and Evolution* online (<http://www.mbe.oxfordjournals.org/>).

Acknowledgments

We thank Iain Pardoe, Patrick Phillips, and Steve Proulx for fruitful discussions and an anonymous reviewer for many helpful comments. Supported by National Science Foundation (NSF) DEB-0516530, National Institutes of Health GM62351, NSF-Integrated Graduate Education and Research Training grant DGE-9972830, and an Alfred Sloan Research Foundation Fellowship to J.W.T.

Literature Cited

- Alfaro ME, Holder MT. 2006. The posterior and the prior in Bayesian phylogenetics. *Annu Rev Ecol Syst.* 37:19–42.
- Alfaro ME, 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(2):255–266.
- Anisimova M, Gascuel O. 2006. Approximate likelihood ratio test for branches: a fast, accurate and powerful alternative. *Syst Biol.* 55(4):539–552.
- Berger JO, Strawderman W, Tang D. 2005. Posterior propriety and admissibility of hyperpriors in normal hierarchical models. *Ann Stat.* 33(2):606–646.
- Buckley TR. 2002. Model misspecification and probabilistic tests of topology: evidence from empirical data sets. *Syst Biol.* 51(3):509–523.
- Carlin BP, Gelfand AE. 1990. Approaches for empirical Bayesian confidence intervals. *J Am Stat Assoc.* 85(409):105–114.
- Cummings MP, Handley SA, Myers DS, Reed DL, Rokas A, Winka K. 2003. Comparing bootstrap and posterior probability values in the four-taxon case. *Syst Biol.* 52(4):477–487.
- Douady CJ, Delsuc F, Boucher Y, Doolittle WF, Douzery EJP. 2003. Comparison of Bayesian and maximum likelihood bootstrap measures of phylogenetic reliability. *Mol Biol Evol.* 20(2):248–254.
- Edwards AWF. 1992. *Likelihood*. Baltimore (MD): Johns Hopkins University Press.
- Erixon P, Sennwald B, Britton T, Oxelman B. 2003. Reliability of Bayesian posterior probabilities and bootstrap frequencies in phylogenetics. *Syst Biol.* 52(5):665–673.
- Felsenstein J. 2004. *Inferring Phylogenies*. Sunderland (MA): Sinauer Associates.
- Hillis DM, Bull JJ. 1993. An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analysis. *Syst Biol.* 42(2):182–192.
- Huelsenbeck JP, Larget B, Miller RE, Ronquist F. 2002. Potential applications and pitfalls of Bayesian inference of phylogeny. *Syst Biol.* 51(5):673–688.
- Huelsenbeck JP, Rannala B. 2004. Frequentist properties of Bayesian posterior probabilities of phylogenetic trees under simple and complex substitution models. *Syst Biol.* 53(6):904–913.
- Huelsenbeck JP, Ronquist F, Nielsen R, Bollback JP. 2001. Bayesian inference of phylogeny and its impact on evolutionary biology. *Science.* 294:2310–2314.
- Karol KG, McCourt RM, Cimino MT, Delwiche CF. 2001. The closest living relatives of land plants. *Science.* 294(5550):2351–2353.
- Lemmon AR, Moriarty EC. 2004. The importance of proper model assumption in Bayesian phylogenetics. *Syst Biol.* 53(2):265–277.
- Lewis PO, Holder MT, Holsinger KE. 2005. Polytomies and Bayesian phylogenetic inference. *Syst Biol.* 54(2):241–253.
- Mar JC, Harlow TJ, Ragan MA. 2005. Bayesian and maximum likelihood phylogenetic analyses of protein sequence data under relative branch length differences and model violation. *BMC Evol Biol.* 5(1):8.
- Misawa K, Nei M. 2003. Reanalysis of Murphy et al.'s data gives various mammalian phylogenies and suggests overcredibility of Bayesian trees. *J Mol Evol.* 57:S290–S296.
- Murphy WJ, Eizirik E, O'Brien SJ, et al. (11 co-authors). 2001. Resolution of the early placental mammal radiation using Bayesian phylogenetics. *Science.* 294(5550):2348–2351.
- Rannala B, Yang Z. 1996. Probability distribution of molecular evolutionary trees: a new method of phylogenetic inference. *J Mol Evol.* 43:304–311.
- Rogers JS. 2001. Maximum likelihood estimation of phylogenetic trees is consistent when substitution rates vary according to the invariable sites plus gamma distribution. *Syst Biol.* 50(5):713–722.

- Ronquist F, Huelsenbeck JP. 2003. Mrbayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*. 19(12):1572–1574.
- Simmons MP, Pickett KM, Miya M. 2004. How meaningful are Bayesian support values? *Mol Biol Evol*. 21(1):188–199.
- Sitnikova T, Rzhetsky A, Nei M. 1995. Interior-branch and bootstrap tests of phylogenetic trees. *Mol Biol Evol*. 12(2):319–333.
- Suzuki Y, Glazko GV, Nei M. 2002. Overcredibility of molecular phylogenies obtained by Bayesian phylogenetics. *Proc Natl Acad Sci USA*. 99(25):16138–16143.
- Swofford DL. 2002. *Phylogenetic analysis using parsimony (*and other methods)*. Sunderland (MA): Sinauer Associates.
- Taylor DJ, Piel WH. 2004. An assessment of accuracy, error, and conflict with support values from genome-scale phylogenetic data. *Mol Biol Evol*. 21(8):1534–1537.
- Wilcox TP, Zwickl DJ, Heath TA, Hillis DM. 2002. Phylogenetic relationships of the dwarf boas and a comparison of Bayesian and bootstrap measures of phylogenetic support. *Mol Phylogenet Evol*. 25:361–371.
- Yang Z, Goldman N, Friday AE. 1995. Maximum likelihood trees from DNA sequences: a peculiar statistical estimation problem. *Syst Biol*. 44(3):384–399.
- Yang Z, Rannala B. 2005. Branch-length prior influences Bayesian posterior probability of phylogeny. *Syst Biol*. 54(3):455–470.

Hervé Philippe, Associate Editor

Accepted July 10, 2007