

The Posterior and the Prior in Bayesian Phylogenetics

Michael E. Alfaro¹ and Mark T. Holder²

¹School of Biological Sciences, Washington State University, Pullman, Washington 99164-4236; email: alfaro@wsu.edu

²School of Computational Science, Florida State University, Tallahassee, Florida 32306-4120; email: mholder@scs.fsu.edu

Annu. Rev. Ecol. Evol. Syst. 2006. 37:19–42

First published online as a Review in Advance on July 21, 2006

The *Annual Review of Ecology, Evolution, and Systematics* is online at <http://ecolsys.annualreviews.org>

This article's doi:
10.1146/annurev.ecolsys.37.091305.110021

Copyright © 2006 by Annual Reviews.
All rights reserved

1543-592X/06/1201-0019\$20.00

Key Words

bootstrap, Markov chain Monte Carlo, PP, prior probability, phylogenetic confidence

Abstract

Bayesian analysis has enjoyed explosive growth in phylogenetics over the past five years. Accompanying this popularity has been increased focus on the meaning of the posterior probability (PP) and the role of the prior in phylogenetic inference. Here we discuss the behavior of the PP in Bayesian and frequentist terms and its relationship to parametric and nonparametric bootstrapping. We also review the use of priors in phylogenetics and the issues surrounding the specification of informative and minimally informative prior distributions.

INTRODUCTION

Over the past five years, Bayesian analysis has established itself as a major methodological innovation in the field of phylogenetics (Huelsenbeck et al. 2001, Lewis 2001). However, serious criticisms have been leveled against the method since near its inception (e.g., Suzuki et al. 2002). In this review we focus on two controversial aspects of Bayesian inference: (a) the use of posterior probabilities (PPs) as measures of clade support and (b) the specification of priors distributions. In particular we discuss the PP in relation to the bootstrap. Although the two support measures are superficially similar, empirical observations, simulation studies, and theoretical work clearly show they are not (and should not be) identical.

Bayesian phylogenetic analyses are conducted using a simulation technique known as Markov chain Monte Carlo (MCMC). Verifying that the MCMC analysis has been sufficiently thorough is clearly a major concern for practitioners, but coverage of the MCMC convergence issues is beyond the scope of this review.

Applying Bayes' Theorem to Phylogenetics

Bayesian inference focuses on a quantity known as the posterior probability (PP, also denoted as $\Pr[Tree|Data]$), which is the probability that a tree is the true tree based on the prior beliefs and the likelihood. If we use θ to represent the parameters of a model, then

$$\Pr(Tree|Data) = \frac{\int \Pr[Data|Tree, \theta] \times \Pr[Tree, \theta] d\theta}{\Pr[Data]} \quad 1.$$

The PP is calculated by integrating the PP density over all possible parameters of a model. This density is the product of the likelihood of the tree ($\Pr[Data|Tree, \theta]$, the probability of the data given the tree and model) and the prior probability density of the tree and model ($\Pr[Tree, \theta]$, simply referred to as the prior) divided by the probability of the data ($\Pr[Data]$).

The prior encapsulates knowledge from sources other than the data at hand, such as information from earlier studies. The likelihood used in Bayesian calculations is a measure of how well a hypothesis fits the observed data. In maximum likelihood inference, trees (or other hypotheses) are chosen solely on the basis of the likelihood function. In contrast, Bayesians choose the most probable tree, which is a function of both the likelihood and the prior. The same models of character evolution used in maximum likelihood can be used in the Bayesian framework simply by specifying prior distributions for the model parameters.

The PP cannot be calculated analytically because it requires a complex integration over the model parameters (branch lengths, rates, base frequencies) and a summation over all possible trees to calculate $\Pr[Data]$ (the denominator of Equation 1). Thus phylogeneticists must use MCMC to approximate the PP (for a review of MCMC in phylogenetics, see Yang 2005). In MCMC, a so-called walk through tree/model space is simulated, and trees are periodically sampled. These samples approximate draws from the PP distribution. We can manipulate the MCMC samples to infer trees (and support for the trees).

Rise of Bayesian Phylogenetic Methods

The development of modern Bayesian applications in phylogenetics had multiple origins in the mid-1990s, with work by Rannala and Yang (Rannala & Yang 1996, Yang & Rannala 1997), Li (1996; Li et al. 2000), and Mau, Newton, Larget, and Simon (Larget & Simon 1999, Mau 1996, Mau & Newton 1997, Mau et al. 1999, Newton et al. 1999, Simon & Larget 2001). Yang (2005) compares these approaches in his recent review. The rapid acceptance of Bayesian phylogenetics can be explained by several factors: the availability of user-friendly software implementations, MRBAYES (Huelsenbeck & Ronquist 2001, Ronquist & Huelsenbeck 2003) and BAMBE (Simon & Larget 2001); the incorporation of complex models of character evolution; the perceived computational advantages of Bayesian MCMC over maximum likelihood; the capacity to accommodate uncertainty in model parameter values, topologies, and ancestral state reconstructions (e.g., Lutzoni et al. 2001, Pagel & Lutzoni 2002); and the ability to simultaneously infer trees and support for clades.

BAYESIAN POSTERIOR PROBABILITIES AS MEASURES OF CLADE SUPPORT

Although PP statements can be made about any parameter of the model, systematists are usually interested in support for trees or clades. In current practice, workers decide on models of evolution (see review in Sullivan & Joyce 2005), specify prior distributions, perform an MCMC run, and produce a majority-rule consensus tree of all the sampled topologies. The numbers on the branches of the consensus tree (the proportion of the MCMC samples that contain a particular split) approximate the PP of the split. Because all the MCMC samples are used in calculating the PP, support for some clades may come from trees that are poorly supported. The MCMC samples can also be summarized using tree-based approaches, which find the 95% credible set of trees and summarize the topological similarities of the trees within this set (Holmes 2003, 2005).

Early developers of Bayesian phylogenetic methods (Larget & Simon 1999, Rannala & Yang 1996, Yang & Rannala 1997) noted that the PP of a clade was often higher than the associated bootstrap proportion (BP). Subsequent empirical studies confirmed this as a general phenomenon, and the first major controversy of Bayesian phylogenetics emerged (Buckley et al. 2002, Karol et al. 2001, Leache & Reeder 2002, Miller & Andrew 2005, Miller et al. 2002, Murphy et al. 2001, Streelman et al. 2002, Whittingham et al. 2002, Wilcox et al. 2002). On the heels of these studies came a small flurry of simulations and theoretical papers that sought to explain the statistical behavior of the PP, especially in relation to the BP (Alfaro et al. 2003, Cummings et al. 2003, Douady et al. 2003, Erixon et al. 2003, Miller & Andrew 2005, Simmons et al. 2004, Suzuki et al. 2002). Some authors suggested the BP was too conservative and argued for the PP as the preferred support measure (Murphy et al. 2001, Wilcox et al. 2002). They noted that PPs of 95% were generally associated with bootstrap values of 70% or greater and suggested this result was in agreement with earlier work arguing that BPs of 70% actually indicated confidence levels of 95%

(Hillis & Bull 1993; but see Efron et al. 1996, Felsenstein & Kishino 1993, Newton 1996). Others (e.g., Buckley et al. 2002) suggested that a less-conservative PP might be more susceptible to model misspecification than the bootstrap. Consensus has yet to emerge on the merit of using the PP as a support measure in phylogenetics. In part this is because the results of these studies appear to be contradictory. Some of the controversy arises from the unfortunate notion that Bayesian methods should be judged by their frequentist behavior (Huelsenbeck & Rannala 2004).

Frequentist and Bayesian Statistics

Most biologists have been trained in frequentist statistical approaches and are relatively unfamiliar with Bayesian methods. Fundamentally, the approaches differ in their treatment of parameters: The frequentist approach has fixed but unknown values, whereas the Bayesian approach has random variables. Hypothesis testing and model-selection strategies also differ between these frameworks (e.g., Gelman et al. 2000; for phylogenetic applications, see Bollback & Huelsenbeck 2001, Huelsenbeck et al. 2001). In frequentist statistics, the decision to accept or reject a null hypothesis involves ascertaining whether an observed summary of the data (a test statistic) falls within the tails of the null distribution. In contrast, in Bayesian statistics, probabilities are assigned directly to hypotheses themselves, and no hypothesis is designated the null. Both Bayesian and frequentist approaches are defensible, but because their underlying approaches differ, tests in different frameworks may sometimes produce different results (Holmes 2005).

Studied Properties of the Posterior Probability of a Clade

Simulation studies and real-data analyses have been used to study several properties of the PP of a clade. Two possible interpretations of a support probability for an observed clade are apparent:

1. The Bayesian perspective is the probability that the clade is present on the true tree.
2. The frequentist perspective is 1.0 minus the p-value. If we take the null hypothesis to be that any particular clade is not in the tree, then the p-value is the probability a clade would be supported at least as strongly by the data even though the clade is not present in the true tree.

Despite the relevance of the former (Bayesian) interpretation to PPs, studies in theoretical systematics have evaluated how well the posterior from a Bayesian MCMC serves to estimate both of these properties (see **Figure 1**). We discuss several properties of the PP that have been studied in theoretical papers below.

Probability that the clade is correct. In simulation studies many data sets can be produced, and the true tree is known. Thus one can test how well the PP from a Bayesian analysis estimates the actual probability that a clade is correct (Hillis & Bull 1993). Ninety-five percent of all clades assigned a PP of 0.95 should be found on

Empirical result	Property	Generating model	Procedure
	<p>a) Type I error</p> <p>How often do support values 78 or higher occur by chance?</p>	<p>NULL</p>	<p>Simulate data sets on tree NULL. Compare frequency of support for chimp + human in NULL distribution to observed support.</p> <p>Frequency = Type I error rate</p>
	<p>b) Type II error</p> <p>At a cut-off value of 78 how often will true clades go unrecognized?</p>	<p>TRUE TREE</p>	<p>Simulate data sets on tree TRUE TREE. Record frequency of support values for chimp + human \leq observed value.</p> <p>Frequency = Type II error rate</p>
	<p>c) Phylogenetic accuracy</p> <p>What percent of clades with support 78 or higher are correct?</p>	<p>TRUE TREE</p>	<p>Simulate data sets on tree TRUE TREE. Examine all clades that receive support value equal to observed (78). Then calculate the proportion of those clades present on TRUE TREE.</p> <p>Proportion = Phylogenetic accuracy</p>
	<p>d) Support for wrong clades</p> <p>What percent of clades with support 78 or higher are incorrect?</p>	<p>TRUE TREE</p>	<p>Simulate data sets on tree TRUE TREE. Examine proportion of wrong clades only that receive support \geq observed support value.</p> <p>Proportion = Support for wrong clades</p>

Figure 1

Four properties of phylogenetic support methods. The meaning of a 78% bootstrap proportion (BP) or posterior probability (PP) attached to the clade human + chimp depends on statistical context. (a) Type I error rate determination requires a null model in which the clade does not appear. (b) Type II error measures the rate of falsely rejecting human + chimp when it is true. (c) Phylogenetic accuracy (Hillis & Bull 1993) measures the probability that clades with a given support value are correct under generating conditions when the clade appears. (d) Support for wrong clades measures the strength of support assigned to clades that are not present under the generating conditions. Panels a and b are cornerstones of frequentist statistical testing. Panel c has become a common benchmark for phylogenetic confidence methods, although statements about accuracy are dependent on the generating conditions. Panel d is sometimes measured in studies comparing the BP with the PP.

the true tree. Theoretically, the PP is an ideal estimator, provided that the model is correct and the parameters for the simulation (including the tree shape) are drawn from the same probability distribution that is used as a prior. However, because PPs in phylogenetics are estimated with MCMC, inaccurate PPs may be obtained if the chain fails to reach the stationary distribution. This has not been shown to be a problem with real phylogenetic data. Simulation studies under relatively simple conditions confirm that MCMC performs well (Alfaro & Huelsenbeck 2006, Huelsenbeck & Rannala 2004). When the model is known and the prior is taken seriously, the PP perfectly matches phylogenetic accuracy (Huelsenbeck & Rannala 2004). Several simulation studies have compared phylogenetic accuracy to the PP using a fixed tree. Given the true model and a fairly easy tree to estimate, the PP is a conservative measure of phylogenetic accuracy, although it is less conservative than the nonparametric BP (Alfaro et al. 2003, Wilcox et al. 2002).

A number of studies have demonstrated that PP is sensitive to violations to the model (Buckley 2002, Douady et al. 2003, Erixon et al. 2003, Huelsenbeck & Rannala 2004, Waddell et al. 2002). Interestingly there seems to be little downside to slight overparameterization of the model, although the expectation is that all the parameters should have increased variance (Cunningham et al. 1998). In contrast, underparameterization can lead to greatly inflated estimates of the PP (Erixon et al. 2003, Huelsenbeck & Rannala 2004, Lemmon & Moriarty 2004, Suzuki et al. 2002). In light of these results, it appears prudent to use sufficiently complex models in performing Bayesian MCMC and err on the side of complexity rather than simplicity. The results about the relatively minor ill effects of overparameterization come from analyses in which the analysis model is slightly too parameter-rich; these conclusions do not imply that one should avoid model-selection techniques and always prefer the most-complex model available.

Posterior probability interpreted as a function of the p-value. Assessing the probability of a Type I error (i.e., rejecting the null hypothesis when it is true) forms the cornerstone of most frequentist statistical tests. Phylogenetic analyses fit naturally into the estimation or model-selection frameworks, whereas p-values are integral to hypothesis testing. To convert clade support statements into a hypothesis test, one must choose a threshold (critical value); support higher than the threshold is taken to mean the rejection of the null hypothesis that the tree lacks the clade. In phylogenetics, assigning a p-value of 0.05 to a branch should mean that the corresponding clade is only seen with greater support from the data in 5% of data sets drawn from trees that do not contain the clade (the null model).

It may seem intuitive to treat the one minus the PP of a clade as a p-value. However, practitioners working with real data generally do not interpret PPs (or bootstrap values for that matter) attached to nodes on a phylogenetic tree as indications of Type I error. Nevertheless, theoretical studies that generate data from star trees and focus only on the incidence of strongly supported clades are implicitly examining how well Bayesian procedures produce p-values. Frequentist p-values are usually calculated under the least-favorable conditions—i.e., under a null model that is most

likely to lead to a false rejection. In testing the significance of a clade, for example, the null hypothesis is the maximum likelihood tree constrained to not have that clade. Thus frequentist p-values are defined with respect to Type I error and lead to conservative inference because they are based on a worst-case scenario. In contrast, Bayesian methodology does not place a premium on controlling Type I error. Thus, unsurprisingly, converting Bayesian PPs into a frequentist test can result in inflated Type I error rates (Suzuki et al. 2002). When the model of evolution is severely misspecified, the PP leads to dramatically high rates of Type I error (Buckley 2002, Suzuki et al. 2002). Type II error (failing to reject an incorrect hypothesis) for a given cut-off value appears to be lower for the PP than for the BP under simple conditions (Alfaro et al. 2003).

Predictor of the best empirical estimate. If we examine the behavior of methods on real data, we do not have the luxury of knowing the true tree. One can use the best tree from an empirical study as a proxy for the true tree. Tree inference procedures can be studied by subsampling taxa and/or characters from the original matrix, analyzing these subsets, and comparing the results with the analysis of the full data set (Simmons et al. 2004, Taylor & Piel 2004). Because the true generating model is unknown, one interpretation of these studies is that they show the effects of biologically relevant model misspecification.

Testing clade support based on a data subset approximates the probability that the clade will be found in the full data analysis and gives systematists a sense of whether they can interpret a support measure as “the probability of the clade being found if we collected a much larger data set.” Bootstrapping or jackknifing significantly underestimates this probability, whereas the Bayesian PP significantly overestimates it (Simmons et al. 2004, Taylor & Piel 2004). If the conditions in these studies are phylogenetically representative, these results suggest a clade that receives an 85% bootstrap or jackknife value is more likely to appear after more-extensive data collection than one that receives an 85% PP. Strongly supported nodes (PP > 95%), however, appear to be only slightly inflated (Taylor & Piel 2004)—they do not appear to be as extremely overestimated as some simulation studies (Suzuki et al. 2002) indicate.

Correlation with the bootstrap proportion. Some studies have simply quantified the correlation coefficient of the PP and the BP, typically calculated across all clades in an analysis. The PP shows moderate correlation with the nonparametric BP, although the strength of the relationship can vary considerably depending on the simulation conditions (Alfaro et al. 2003, Cummings et al. 2003, Douady et al. 2003). Importantly, even in cases where the two measures differ substantially, a staggeringly large number of splits will have received support equal to zero. Thus the overall correlation between the PP and BP is generally high.

The Posterior Probability and the Bootstrap Proportion

The nonparametric bootstrap has a long and controversial history in phylogenetics (Berry & Gascuel 1996; Efron et al. 1996; Felsenstein 1985; Felsenstein & Kishino

1993; Hillis & Bull 1993; Li & Zharkikh 1994; Newton 1996; Sanderson 1995; Sanderson & Wojciechowski 2000; Zharkikh & Li 1992, 1995). Despite the controversy surrounding the meaning and interpretation of the BP, the procedure is easy to describe: Pseudoreplicate data matrices are created from the original data by sampling the columns of site patterns, and the optimal tree for the pseudoreplicate is found. The process is repeated hundreds of times, and support for a clade is expressed as the proportion of times the clade appeared in the pseudoreplicates (**Figure 2a**).

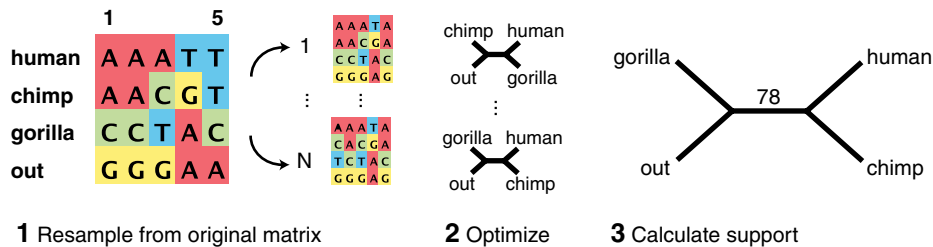
Several papers have suggested there are theoretical reasons to expect the PP and the BP to be the same (Cummings et al. 2003, Erixon et al. 2003, Randle et al. 2005). A key difference between the methods, however, is that bootstrapping generates new data sets (by subsampling the original matrix), whereas Bayesian MCMC treats the data matrix and relies on a model to account for sampling variability. Given the similarities between the methods, we predict maximum likelihood and Bayesian analysis of any given bootstrap pseudoreplicate to be strongly congruent. In fact, when PPs are calculated for the bootstrap pseudoreplicates rather than the original matrix, the BP and the averaged PP are similar (Erixon et al. 2003, Waddell et al. 2002).

In a highly cited paper, Efron et al. (1996) state that the BP can be interpreted as an approximate PP if a noninformative prior is used on site patterns. Although we do not dispute this interpretation, this unconstrained model of site patterns differs substantially from the typical Bayesian phylogenetic models. Therefore, Efron et al.'s (1996) statement does not imply that the BP and the PP obtained from a MRBAYES run should be the same (see also Efron 2003). In the following section we elaborate on the differences between the Bayesian perspective on the BP and a typical Bayesian phylogenetic analysis.

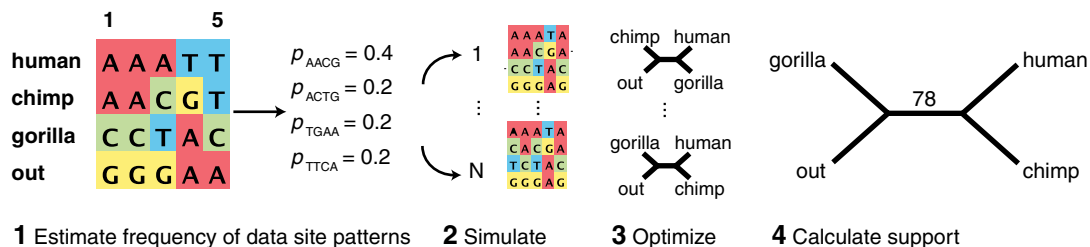
Figure 2

Three perspectives of the bootstrap. (a) Typically, bootstrapping is performed by sampling columns of the data matrix (site patterns), finding the optimal tree for the pseudoreplicates, and recording the proportion of times a clade of interest appears. (b) The multinomial bootstrap mimics the resampling procedure in panel *a* by parameterizing the frequency of occurrence of the site patterns. Data sets can be simulated by drawing site patterns from this model, producing results that are identical to sampling with replacement from the original data matrix. (c) A nonhypothesis testing form of the parametric bootstrap would involve fitting the best tree, instead of a null or constraint tree as in the SOWH test (Goldman et al. 2000), and model of sequence evolution to the original data. New data sets can then be simulated under this model and analyzed, and the presence or absence of clades can be assessed. Note that procedures in panels *b* and *c* may yield different bootstrap proportions (BPs). Although they are both parametric procedures, the underlying models are quite different. Bayesian Markov chain Monte Carlo (MCMC) makes use of models that are similar to those used in parametric bootstrapping with the addition of prior probabilities on model parameters. Thus the Bayesian bootstrap posterior probability (PP) and the Bayesian MCMC PP may also differ owing to the difference in the underlying models (Svennblad et al. 2006).

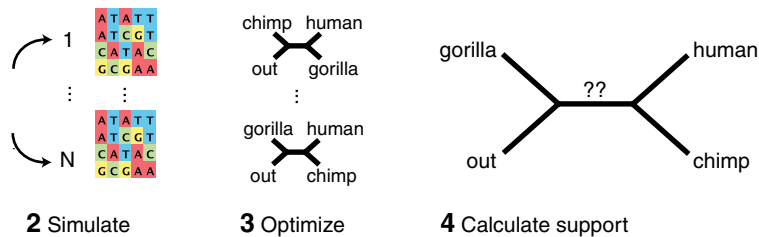
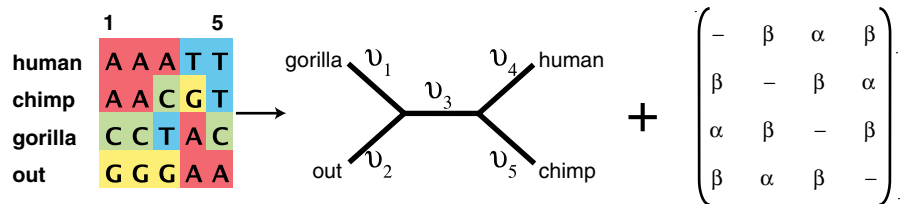
a Bootstrap



b Multinomial bootstrap



c Phylogenetic parametric bootstrap



The Parameters of the Nonparametric Bootstrap

Parametric bootstrapping involves estimating parameters of a model from the data, then generating data sets according to this fitted model. The typical description of nonparametric bootstrapping (i.e., resampling the original data with replacement data) sounds like a different procedure. It is possible to describe the nonparametric bootstrap as a procedure of estimating parameters from the original data under a peculiar model, then generating new data from this model (and parameters)—this is exactly the same strategy taken in parametric bootstrapping. The peculiar model is the unconstrained model of Goldman (1993). No tree is assumed in this model; instead one simply estimates the frequencies of every site pattern from the original matrix.

The unconstrained model gets its name from the fact that it enforces no relationships between the frequencies of different character types. In essence, the only assumptions made are that the sites are independent draws from the same underlying distribution—the multinomial distribution with expected character frequencies exactly matching the proportions seen in the observed data (**Figure 2b**) (Holmes 2003, 2005). Simulating from this multinomial distribution is identical to resampling the original data with replacement data because the probability of a site pattern generated at any position in the matrix is simply the proportion of sites in the original matrix that displayed that pattern. Describing the nonparametric bootstrapping as fitting the unconstrained model and then sampling from it may seem awkward and unnecessarily formal, but this description is the key to understanding Efron et al.'s (1996) Bayesian view of the BP.

A Bayesian View of the Nonparametric Bootstrap

Efron et al.'s (1996) Bayesian interpretation of the nonparametric bootstrap is framed in the context of estimating the means of normal distributions. However, one can apply the argument to phylogenetics using the unconstrained model with uninformative prior probability on the parameters. (Efron et al., however, do not explicitly describe this prior; see Sennblad et al. 2006 for possible implementations).

Phylogenies cannot be inferred using the unconstrained model because trees are not parameters of the model—the only parameters are the expected frequencies of each site pattern. The link between phylogenetic inference and this unconstrained model is made by associating every point in the space of pattern frequencies with a tree. Every possible data matrix can be represented as a vector of site pattern frequencies (where elements indicate the proportion of columns from the original matrix that shows a particular site pattern). If we view the elements of the vector as coordinates in multidimensional space, then each data set maps to a point in the space of possible pattern frequencies (Kim 2000). The choice of an optimality criterion (e.g., specification of a maximum likelihood model) induces a partition of this space. Each tree corresponds to a region of the parameter space. Any data set that displays pattern frequencies within this region will favor the tree. Thus, the data matrix that a researcher collects can be summarized as a location in the space of possible pattern

frequencies, and tree searching is an attempt to determine which tree's region contains the observed data location.

Of course, the presence of sampling error implies that the observed pattern frequencies will deviate from the so-called true pattern frequencies (i.e., those one would obtain from infinitely long sequences). Assuming just the unconstrained model and taking the partitioning of pattern frequency space as a given, a Bayesian assessment of confidence attempts to estimate the probability that the true pattern frequencies are within a particular tree's region of parameter space. Efron et al.'s (1996) Bayesian interpretation of the nonparametric BP is then the probability that the true pattern frequency falls within the region of parameter space that would favor a given tree. Thus, the nonparametric BP can be considered an approximation to a PP, although only when using a multinomial distribution over site pattern frequencies.

The Parametric Bootstrap

In the phylogenetic parametric bootstrap, the parameters are not site pattern frequencies but instead are components of a probabilistic model of sequence evolution—substitution rate parameters, branch lengths, and base frequencies. In the nonparametric bootstrap, new data sets are generated using the multinomial distribution centered at the observed data. In the parametric bootstrap, new data sets are generated from a phylogenetic model that has been displaced from the observed pattern frequencies to the closest point on a “model manifold,” to use Kim's (2000) terminology. Note that closeness here is measured by the Kullback-Leibler distance, which is a form of likelihood ratio between a phylogenetic model and the unconstrained model (**Figure 2c**). In the hypothesis testing form of parametric bootstrapping [e.g., the SOWH test (Goldman et al. 2000)], the pattern frequencies are moved to the closest model manifold for a tree that does not have the clade of interest. Data are simulated using a multinomial distribution (as in the nonparametric bootstrap), but the parameters of the multinomial (the probabilities assigned to each site pattern) are not free parameters—they are determined by the phylogenetic model. Given these differences, we do not expect parametric bootstrap and multinomial bootstrap support calculated from the same data to always correspond. Although they are both bootstrap procedures, the underlying sampling distributions are not the same. A similar case can be made for an expectation of differences between the nonparametric bootstrap and Bayesian MCMC, even though both may be viewed as PPs (see below).

Contrasting the Bayesian Interpretation of the Bootstrap with Markov Chain Monte Carlo

Although one could interpret the nonparametric BP as a PP derived from a unconstrained model (as discussed above), this PP should not always be equal to the PP from Bayesian MCMC. This is because Bayesian MCMC more closely resembles the parametric bootstrap in that both procedures fully exploit a model of character evolution as the estimate support. Thus, even with a Bayesian interpretation of the nonparametric BP, we should not expect the BP and the Bayesian MCMC PP to

always be the same because the underlying models are quite different. Sennblad et al. (2006) provide a more in-depth, analytical discussion.

Summary of the Discussion of Posterior Probabilities

The proliferation of Bayesian studies demands that phylogeneticists carefully consider the statistical meaning of the posterior in Bayesian and frequentist contexts. A Bayesian perspective provides the natural framework for interpreting the PP as the probability that the clade is correct given the prior, the data, and the model of evolution. We believe this view has utility in phylogenetic systematics, and PPs are a natural and useful way for comparing alternative hypotheses. However, even within the Bayesian framework, phylogeneticists must be aware of several factors that can influence the PP. Most important among these is the sensitivity of the PP to model misspecification. Assessments of model adequacy (Bollback 2002) and model averaging procedures (Huelsenbeck et al. 2004, Lewis et al. 2005) should accompany Bayesian analyses so that model-sensitive inferences can be identified. Fortunately, the flexibility of the Bayesian framework and MCMC computations allows complex models to be implemented in a fairly straightforward fashion (Castoe et al. 2004, Nylander et al. 2004).

It is unwise to automatically use the PP as a guide to rates of Type I error because the Bayesian procedures have not been constructed with frequentist performance in mind (though it is theoretically possible to define Bayesian models and priors that give good frequentist performance). Similarly, one should not expect PPs to correspond well with nonparametric BPs any more than one would expect the *p*-values from a parametric and a nonparametric test to be the same (Sennblad et al. 2006).

ISSUES SURROUNDING THE PRIOR IN BAYESIAN PHYLOGENETICS

Prior probability distributions (referred to as priors hereafter) are often a main focus of debates between Bayesians and advocates of other forms of statistical inference. The use of explicit prior distributions is unique to the Bayesian perspective, and the elicitation of priors is clearly a difficult and subjective step in an analysis. Even within the Bayesian school, practitioners disagree about the specification of priors when an obvious, uncontroversial prior is not available.

The Subjective Bayesian Approach

In subjective Bayesian analysis, prior probabilities are assigned according to the researcher's degree of belief in different hypotheses (or parameter values). Advocates of subjective Bayesianism (de Finetti 1974, 1975; Goldstein 2006; Lindley 2000) usually cite the coherence of the methodology as one of its strengths. If the priors accurately reflect one's initial beliefs, and the model is good, then the PP distribution is the

ideal guide to the researcher: Conclusions based on the posterior will be neither too conservative nor too liberal. The primary objection to the subjective Bayesian approach is that, although analyses to update personal beliefs may be useful in many situations, the scientific process demands objective statements about the data and hypotheses. Explicitly incorporating one's personal beliefs makes the analysis unnecessarily limited to the scientist publishing the data and makes the results possibly unconvincing to other researchers.

Objective Bayesian Analyses

Objective Bayesian analysis has been a major focus of statistics in the past few decades. Simply put, the goals of an objective Bayesian analysis are to allow the data (via the likelihood) to dominate the conclusions that are drawn (for a readable review, see Berger 2004). Objective Bayesian priors are often called noninformative priors; however, minimally informative may be a more-accurate description (i.e., all probability distributions are informative about some hypotheses). Priors are chosen to have a minimal effect on the parameter of interest, and interpreting a prior as a statement of one's beliefs is shunned (Bernardo 1997).

Objective Bayesian priors (e.g., Bernardo 1979) are not always uniform (flat) priors; instead they take the form of the model's likelihood function into account to allow the data to dominate the analysis. Objective Bayesian priors have a number of desirable properties: Many avoid problems that arise from priors that are sensitive to arbitrary choices (such as how to parameterize a model), and they often result in analyses that behave well from a frequentist viewpoint (Bayarri & Berger 2004). Unfortunately, they are not used in phylogenetics because the likelihood function for character data on trees cannot be expressed in a tractable form. Instead, priors in Bayesian phylogenetics are usually chosen to be vague over a wide range of parameter values, in the hope that such priors will have minimal effect on the results of the analysis. Some refer to this kind of approach as pseudo-Bayes because the priors are neither truly subjective nor objective (Berger 2004). Despite the lack of a formal justification, many Bayesians regard it as an effective way of analyzing real data (Berger 2004).

Overview of the Use of Priors in Bayesian Phylogenetics

All calculations in a Bayesian analysis are performed using the joint probability density of all parameters. Systematists are usually interested in tree topologies, but prior probabilities must be specified for every aspect of the model. In practice, systematists specify the joint prior probability density by multiplying the prior densities for several groups of parameters. This is equivalent to treating these parameter groups as if they were independent of each other. When parameters are tightly correlated (such as the proportion of invariants sites and the parameter governing the amount of rate heterogeneity among varying sites), the assumption of independence is probably unwarranted, but is still generally made.

Priors on the Model of Sequence Evolution

In many cases, priors on parameters of the sequence evolution model may be of less concern than tree and edge-length priors. Sequence evolution parameters are estimated from all sites in the data matrix and thus tend to have sharp likelihood profiles and are relatively insensitive to the choice of prior. Yang et al. (1995) noted that, in the context of maximum likelihood, the parameters of the sequence evolution model were often similar from tree to tree. If this general pattern holds, then even poorly chosen priors on the parameters of the sequence model may have little effect on tree inference because all trees may be affected in similar ways.

Conversely, increases in model complexity should lead to a flattening of likelihood profiles because the data are used to estimate more parameters. If this happens, prior influence may become a concern, especially in regions of parameter space where models are relatively insensitive to parameter change (Felsenstein 2004, Rannala 2002, Zwickl & Holder 2004).

Priors on the branching process. Trees represent the results of real biological processes (e.g., replication, anagenesis, cladogenesis); thus it seems natural to model the process of forming trees itself—to place a prior on different tree shapes according to models of how trees are generated. As Yang et al. (1995) note, probability distributions over tree shapes were employed in some of the earliest attempts to use likelihood techniques to reconstruct genealogies (Edwards & Cavalli-Sforza 1963, Thompson 1975). The Yule process models tree formation by repeated division (speciation) without lineage extinction, whereas the birth-death process includes lineage extinction. Priors on both processes do not place uniform probability on all trees, but instead place them on all labeled histories. Historically, birth-death priors have been used in Bayesian phylogenetics (Rannala & Yang 1996), although nearly all recent analyses use a uniform prior on topologies. The effect of choosing different tree priors on inference remains uninvestigated.

Informative topological priors. At the outset of any new investigation, most systematists have some prior expectation for tree shapes because few groups remain completely unstudied in terms of their evolutionary history. Therefore, a default uniform prior on topologies is usually not appropriate. Effectively, the failure to accommodate relevant prior information may produce a conservative analysis, especially if the tree agrees with the previous estimates (the reported clade support will probably be lower than an estimate using an informative prior).

Instead of ignoring prior information, one could attempt to include it in the analysis. One approach is to use a super matrix, wherein all previous data are included and reanalyzed. Alternatively, one could use measures of clade support from previous studies as the topological prior in the analysis of new data. Ideally, one would use the joint PP distribution of trees and all parameters from the first analysis directly as the prior for the subsequent analysis. Unfortunately, this is not practical for two reasons: First, the computer memory required to represent this distribution is generally prohibitive, and second, sufficiently accurate estimates of the probability density are

not available. MCMC sampling may provide reasonable estimates of quantities such as clade PPs, but a detailed view of the entire parameter space of trees and models of character evolution is beyond our grasp. Ronquist et al. (2004) have pioneered an approach to accommodate measures of clade support from one analysis into a second analysis. From a tree with estimates of branch support (ideally PPs of the clades), they propose two sets of factors that can be used to approximate the entire joint distribution on trees. These approximations appear to capture the relative probabilities of trees that are good (close to optimal). This implies that if the final analysis recovers a tree similar to previous results, then the approximate topology prior has probably mimicked the ideal prior—the posterior distribution of trees from the previous analysis. One outstanding problem is how to deal with partial overlap of characters among the input trees (if some of them are based, in part, on the same sequence data). Downweighting the trees to account for redundant use of data is in order, but it is difficult to give precise guidelines on how such weighting should be done (Ronquist et al. 2004)

Clade Priors and Tree Priors

The prior probabilities of clades are usually not specified as independent inputs to the Bayesian inference. In practice, a prior on different topologies is specified, and this prior induces a prior distribution on splits. The most common prior in real analysis is a uniform prior over all topologies or over all labeled histories. A broader class of priors on tree topology includes label-invariant distributions (Steel & Penny 1993, Steel & Pickett 2006). In this class of distributions, the tree shape alone determines the prior probability on a topology. Any prior that seeks to be uninformative about the relationships among a group of taxa belongs to the class of label-invariant priors. Steel & Pickett (2006) prove it is impossible for any leaf-invariant prior to imply a uniform prior on all possible clades (assuming that analysis includes over four taxa for rooted inference or over five taxa for unrooted trees). This property arises from the compatibility of clades of differing sizes with differing numbers of topologies.

Pickett and Randle clearly view this property as a flaw in Bayesian phylogenetic analysis and as a reason to distrust clade PP: “[W]hen all trees are treated as equally probable, a priori, Bayesian clade support distributions are affected by priors that fail to model ignorance accurately” (Pickett & Randle 2005, Randle & Pickett 2006). However, unequal priors on clades of different sizes do not imply a problem with clade PPs—if the model and prior are correct, the PP will accurately reflect the probability that both exist in the true tree. The unequal prior arises from an intrinsic relationship between clades and trees: Smaller clades are compatible with more trees than larger clades. Thus the phenomenon reflects a truth about probability statements on bifurcating trees (and not about Bayesian analyses, per se). As an analogy, let us consider the problem of estimating the number of beans in a jar. We could assign a prior probability to every possible value, but it would not be reasonable to expect this prior to assign equal probabilities to hypotheses such as “the number of beans is divisible by two” and “the number is divisible by three.” Although it may seem desirable to express a priori ignorance about every conceivable hypothesis, it is not

possible to do this and make coherent probability statements. For large trees, all clades have a low prior, and empirically, only modest amounts of data appear necessary to result in high PPs, even for clades that are assigned relatively low prior probabilities (Brandley et al. 2006).

Priors on Branches

Most Bayesian phylogenetic analyses follow the tradition of maximum likelihood methods and treat each branch length as an independent parameter. In practice, a single prior distribution is applied to each branch in the tree (implicitly assuming independence of branch lengths). A disadvantage of treating each branch length as an independent parameter is that it loses some information relevant to the phylogenetic inference. For example, assuming that the sampled tips are contemporaneous and the rate of evolution does not vary wildly over short timescales, branch lengths should be close to ultrametric. Models of tree generation can be used to specify that prior distributions are branch lengths that conform to the molecular clock hypothesis.

Suchard et al. (2001) employed a hierarchical approach to the assignment of priors on branch lengths that partially accommodates the information implied by contemporaneous taxonomic sampling. In their implementation, the prior probability for a particular edge had an exponential distribution. However, instead of specifying a fixed value for the mean of this exponential distribution, they treat the mean of the branch length prior as a type of parameter (a hyperparameter). A hyperprior distribution is assigned to the mean of the branch length prior, producing a model that is hierarchical—some parameters only affect predictions about the data that will be seen (and hence affect the likelihood), whereas hyperparameters only affect the probability density of other parameters. In the context of a joint prior over branch lengths, this hierarchical approach effectively expresses the intuition that the branch lengths will be roughly the same scale as each other.

Most practitioners are interested in the topology of the tree returned by a phylogenetic analysis, not in the other values inferred for other aspects of the models of evolution used during the inference. Despite the fact that the topology is the focal parameter, the prior distributions placed on other nuisance parameters can affect the topology inferred or the strength of support for aspects of the tree (Felsenstein 2004, Yang 2005, Zwickl & Holder 2004). The prior distribution placed on branch lengths warrants particular scrutiny for several reasons: There is no unambiguous upper limit on branch lengths; the length of internal branches is correlated with the difference in likelihoods between different topologies; and the prior distribution applied to branches influences $2N-3$ factors (the prior densities of each branch) that are multiplied to contribute to the prior probability density.

Branch lengths are usually expressed in terms of expected number of changes per site. In this parameterization, branch lengths are unbounded parameters. Clearly enormous branch lengths are not realistic, but it is difficult to justify any particular upper limit on the branch length. As Felsenstein (2004) points out, the likelihood does not go to zero as branch lengths become infinite. As all the branches in the tree

become large, the likelihood approaches a number determined by the equilibrium state frequencies of the model. This likelihood is usually extremely small, but not zero. Thus, the posterior distribution may be particularly sensitive to the upper bound placed on prior used for the branch length. Of course, the prior on branch lengths does not need to be a distribution that requires the specification of an upper limit (e.g., the default branch length in version 3.1 of MRBAYES is an exponential distribution). The use of a prior that puts decreasing density on higher branch length is justifiable because the likelihood is more informative about short branch lengths—it requires less data to discriminate between branch lengths of 0.01 and 0.02 than to discriminate between lengths of 1.01 and 1.02.

The potential for the branch length prior to affect clade probabilities is apparent when one considers extreme priors. If a strong prior were used to force all internal branch lengths to be infinitesimal, then all trees would have similar likelihoods (similar to a star tree's likelihood). A prior emphasizing longer internal branch lengths may make Bayesian phylogenetic analyses too liberal (Yang & Rannala 2005). This causes concern in light of the tendency to use priors with unrealistically high means for branch lengths. Yang and Rannala give several suggestions for making Bayesian analyses more conservative, in particular using priors with small internal means.

Interestingly, Yang & Rannala (2005) also note that if one uses informative priors on topologies, one can reasonably make the branch length prior specifically tailored to the branch in question. If one wishes to express the prior belief that a clade is particularly unlikely, then one can also reasonably assume that if this is wrong (and the clade does exist), it is not separated by the other taxa by a long branch. (If it were, then the clade probably would have been recovered in previous analyses.)

Inferring Trees from Star Trees

Interestingly, data sets simulated from an unresolved model tree occasionally show strong support of one resolution of the tree (Cummings et al. 2003, Lewis et al. 2005, Suzuki et al. 2002). The effect can be particularly pronounced if the model of character evolution used to generate the data is more complex than the model used for inference (Suzuki et al. 2002). Simulations from star trees can be used to detect a bias in a method, such as long-branch attraction (Bruno & Halpern 1999). However, in some cases (Cummings et al. 2003, Lewis et al. 2005, Suzuki et al. 2002), the effect occurs even if external branches in the simulations are equal length and there is no bias for one tree over another. Given that Bayesian methods are consistent (if the assumed model is accurate and the prior does not rule out the true tree), one might expect the effect to decrease as more data are collected. Interestingly, it does not (Lewis et al. 2005). At first this result seems paradoxical. However, Bayesian theory guarantees the posterior will accurately reflect the uncertainty about a hypothesis that is true or false (eventually leading to complete confidence in or rejection of the hypothesis as more data is collected) (Cummings et al. 2003, Lewis et al. 2005). Theory does not guarantee that conditions on the border of multiple hypotheses (multiple trees in this

case) will return equal support for all bordering hypotheses, even as the amount of data increases.

On one level, these results are not too troubling. First, we do not expect true hard polytomies to occur frequently. Even small internal branches can be recovered if there is sufficient data, so the spurious support given to essentially infinite data is confined to hard polytomies. Second, the percentage of data sets showing strong support for a resolution is not high (unless the model of character evolution is also misspecified). Third, most researchers are accustomed to using moderate α -levels in frequentist statistics. If one views a polytomy as a null hypothesis, then it is not surprising that one occasionally rejects a null inappropriately. We do not advocate interpreting Bayesian PPs in these frequentist terms. Instead, we merely want to point out that most practitioners accept some false positives from a statistical method.

Conversely, returning strong support for nonexistent clades is clearly an undesirable property. Even if false resolutions only occasionally receive strong support, a large tree could contain multiple spurious groupings.

Ideally, if data came from a polytomy, an inference method would return equal support for all possible resolutions. Because of sampling error, this ideal cannot be achieved in all cases. The frequencies of data patterns generated by a star tree are similar to those generated from a tree with a short internal branch. The frequencies are so similar that it is not feasible to discriminate between them, given a small sample of data. This fundamental reality means that if one were to engineer an analysis that avoids incorrectly resolved polytomies, there would have to be some loss of power (i.e., an inability to detect short branches when they exist).

Priors for More Conservative Bayesian Markov Chain Monte Carlo

Given that Bayesian methods (and other highly parametric approaches) tend to have high power, it seems appropriate to consider modifications of standard Bayesian approaches to make the results more conservative. Lewis et al. (2005) propose considering unresolved trees during the MCMC analysis and even using priors that prefer unresolved trees over more-resolved trees. Favoring polytomies may be done for the sake of conservatism, not realism. Thus, we may favor polytomies because errors that return unresolved trees are less troublesome than incorrectly supporting a false group. The consideration of polytomies during the MCMC requires a new set of proposals utilizing reversible-jump MCMC (Green 1995) to switch between models (trees in this case) that differ in the number of parameters they contain (for details, see Lewis et al. 2005). Simulations and real-data analyses indicate that considering polytomies during the MCMC does not dramatically affect the power of the analysis to detect short branches. Yang & Rannala (2005) discuss several approaches to using priors that favor short internal branches as a means of making Bayesian analysis. An ad hoc approach could also be used to mimic the behavior of a conservative Bayesian analysis (e.g., one could collapse branches that are shorter than some cut-off value before summarizing the trees sampled during the MCMC).

Summary of the Discussion of Priors

The requirement that practitioners of Bayesian methods must specify prior probability statements for every aspect of the model may seem daunting. Prior specification merits careful consideration, but for many aspects of the model, the data speak strongly. In these cases, a broad range of priors will lead to similar conclusions. Recent work suggests that the use of priors favoring short branch lengths, or polytomies, may prove to be an effective technique for allowing users to make Bayesian support statements conservative (without sacrificing too much power). Improving the feasibility of inference under complex models of character evolution will undoubtedly be one of the most important contributions of Bayesian phylogenetics to the field of systematics. Using sound priors with these parameter-rich models is particularly important because many possible parameter values will be compatible with the data.

Specifying priors that have no effect on the analysis is not an attainable goal. Thus if one wishes to utilize minimally informative priors, it is prudent to examine analyses using multiple combinations of priors to demonstrate that the priors are not having an unexpectedly strong effect on the conclusions.

CONCLUSIONS

Perhaps the greatest justification for the continued use and development of Bayesian methodology is the potential it offers as a model-based framework for analyzing multigene or genomic data sets. Bayesian methods are not, however, a panacea, and workers must familiarize themselves with the strengths and limitations of the approach when interpreting their phylogenetic analysis (reviewed in Huelsenbeck et al. 2002). Bayesian phylogenetics does not always produce statistics that have a good frequentist interpretation. Research into the development of probability distributions on tree space (Holmes 2005) has the potential to bridge this difference to some degree by improving the basis of all phylogenetic statistical tests. Beyond the issues we discuss here, additional work is also needed to understand the adequacy of current Bayesian MCMC samplers for yielding reliable PP under difficult tree conditions (Mossel & Vigoda 2005). Continued efforts in the areas of phylogenetic model development, model selection, and model adequacy are also necessary to ensure the adequate performance of Bayesian methods on the increasingly complex data sets of the future.

ACKNOWLEDGMENTS

Paul Lewis provided helpful discussions of several issues, in particular the use of hierarchical priors on edge lengths. Susan Holmes clarified several points regarding the Bayesian interpretation of the nonparametric bootstrap and statistical incoherency in phylogenetics. Devin Drown and Jessica Lynch Alfaro provided helpful comments on an earlier version of the manuscript. NSF DEB-0445453 (M.E.A.) provided funding for this project. M.T.H. was funded by a CIPRES grant (NSF EF 0331495 to David Swofford).

LITERATURE CITED

- Alfaro ME, Huelsenbeck JP. 2006. Comparative performance of Bayesian and AIC-based measures of phylogenetic model uncertainty. *Syst. Biol.* 55:89–96
- 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.* 20:255–66
- Bayarri MJ, Berger JO. 2004. The interplay of Bayesian and frequentist analysis. *Stat. Sci.* 19:58–80
- Berger JO. 2004. The case for objective Bayesian analysis. *Bayesian Anal.* 1:1–17
- Bernardo JM. 1979. Reference posterior distributions for Bayesian inference. *J. R. Stat. Soc. Ser. B* 41:113–47
- Bernardo JM. 1997. Non-informative priors do not exist: a dialogue with José M. Bernardo. *J. Stat. Plan. Inference* 65:159–89
- Berry V, Gascuel O. 1996. On the interpretation of bootstrap trees: appropriate threshold of clade selection and induced gain. *Mol. Biol. Evol.* 13:999–1011
- Bollback JP. 2002. Bayesian model adequacy and choice in phylogenetics. *Mol. Biol. Evol.* 19:1171–80
- Bollback JP, Huelsenbeck JP. 2001. Phylogeny, genome evolution, and host specificity of single-stranded RNA bacteriophage (family Leviviridae). *J. Mol. Evol.* 52:117–28
- Brandley MC, Leache AD, Warren DL, McGuire JA. 2006. Are unequal clade priors problematic for Bayesian phylogenetics? *Syst. Biol.* 55:138–46
- Bruno WJ, Halpern AL. 1999. Topological bias and inconsistency of maximum likelihood using wrong models. *Mol. Biol. Evol.* 16:564–66
- Buckley TR. 2002. Model misspecification and probabilistic tests of topology: evidence from empirical data sets. *Syst. Biol.* 51:509–23
- Buckley TR, Arensburger P, Simon C, Chambers GK. 2002. Combined data, Bayesian phylogenetics, and the origin of the New Zealand cicada genera. *Syst. Biol.* 51:4–18
- Castoe T, Doan T, Parkinson C. 2004. Data partitions and complex models in Bayesian analysis: the phylogeny of gymnophthalmid lizards. *Syst. Biol.* 53:448–69
- 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.* 52:477–87
- Cunningham CW, Zhu H, Hillis DM. 1998. Best-fit maximum-likelihood models for phylogenetic inference: empirical tests with known phylogenies. *Evolution* 52:978–87
- de Finetti B. 1974. *Theory of Probability*, Vol. 1. Chichester: Wiley
- de Finetti B. 1975. *Theory of Probability*, Vol. 2. Chichester: Wiley
- 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.* 20:248–54
- Edwards AWF, Cavalli-Sforza LL. 1963. The reconstruction of evolution. *Heredity* 18:553

- Efron B. 2003. Second thoughts on the bootstrap. *Stat. Sci.* 18:135–40
- Efron B, Halloran E, Holmes S. 1996. Bootstrap confidence levels for phylogenetic trees. *Proc. Natl. Acad. Sci. USA* 93:13429–34
- Erixon P, Sennblad B, Britton T. 2003. Reliability of Bayesian posterior probabilities and bootstrap frequencies in phylogenetics. *Syst. Biol.* 52:665–73
- Felsenstein J. 1985. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* 39:783–91
- Felsenstein J. 2004. *Inferring Phylogenies*. Sunderland, MA: Sinauer
- Felsenstein J, Kishino H. 1993. Is there something wrong with the bootstrap on phylogenies? A reply to Hillis and Bull. *Syst. Biol.* 42:193–200
- Gascuel O, ed. 2005. *Mathematics of Evolution and Phylogeny*. Oxford, UK: Oxford Univ. Press
- Gelman A, Carline JB, Stern HS, Rubin DB. 2000. *Bayesian Data Analysis*. Boca Raton, FL: Chapman & Hall. 526 pp.
- Goldman N. 1993. Statistical tests of models of DNA substitution. *J. Mol. Evol.* 36:182–98
- Goldman N, Anderson JP, Rodrigo AG. 2000. Likelihood-based tests of topologies in phylogenetics. *Syst. Biol.* 49:652–70
- Goldstein M. 2006. Subjective Bayesian analysis: principles and practice. *Bayesian Anal.* 1(3):403–20
- Green PJ. 1995. Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika* 82:711–32
- Hillis DM, Bull JJ. 1993. An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analysis. *Syst. Biol.* 42:182–92
- Holmes S. 2003. Bootstrapping phylogenetic trees: theory and methods. *Stat. Sci.* 18:241–55
- Holmes S. 2005. Statistical approach to tests involving phylogenies. See Gascuel 2005, pp. 91–120
- Huelsenbeck JP, Larget B, Alfaro ME. 2004. Bayesian phylogenetic model selection using reversible jump Markov chain Monte Carlo. *Mol. Biol. Evol.* 2004:1123–33
- Huelsenbeck JP, Larget B, Miller RE, Ronquist F. 2002. Potential applications and pitfalls of Bayesian inference of phylogeny. *Syst. Biol.* 51:673–88
- Huelsenbeck JP, Rannala B. 2004. Frequentist properties of Bayesian posterior probabilities of phylogenetic trees under simple and complex substitution models. *Syst. Biol.* 53:904–13
- Huelsenbeck JP, Ronquist F. 2001. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* 17:754–55
- Huelsenbeck JP, Ronquist F, Nielsen R, Bollback JP. 2001. Bayesian inference of phylogeny and its impact on evolutionary biology. *Science* 294:2310–14
- Karol KG, McCourt RM, Cimino MT, Delwiche CF. 2001. The closest living relatives of land plants. *Science* 294:2351–53
- Kim J. 2000. Slicing hyperdimensional oranges: the geometry of phylogenetic estimation. *Mol. Phylogenet. Evol.* 17:58–75
- Larget B, Simon DL. 1999. Markov chain Monte Carlo algorithms for the Bayesian analysis of phylogenetic trees. *Mol. Biol. Evol.* 16:750–59

- Leache AD, Reeder TW. 2002. Molecular systematics of the Eastern Fence Lizard (*Sceloporus undulatus*): a comparison of parsimony, likelihood, and Bayesian approaches. *Syst. Biol.* 51:44–68
- Lemmon AR, Moriarty EC. 2004. The importance of proper model assumption in Bayesian phylogenetics. *Syst. Biol.* 53:265–77
- Lewis PO. 2001. Phylogenetic systematics turns over a new leaf. *Trends Ecol. Evol.* 16:30–37
- Lewis PO, Holder MT, Holsinger KE. 2005. Polytomies and Bayesian phylogenetic inference. *Syst. Biol.* 54:241–53
- Li S, Pearl D, Doss H. 2000. Phylogenetic tree construction using Markov chain Monte Carlo. *J. Am. Stat. Assoc.* 95:493–508
- Li W. 1996. *Phylogenetic tree construction using Markov chain Monte Carlo*. PhD thesis. Ohio State Univ., Columbus
- Li W, Zharkikh A. 1994. What is the bootstrap technique? *Syst. Biol.* 43:424–30
- Lindley DV. 2000. The philosophy of statistics. *Statistician* 49:293–337
- Lutzoni F, Pagel M, Reeb V. 2001. Major fungal lineages are derived from lichen symbiotic ancestors. *Nature* 411:937–40
- Mau B. 1996. *Bayesian phylogenetic inference via Markov chain Monte Carlo methods*. PhD thesis. Univ. Wis., Madison. 106 pp.
- Mau B, Newton MA. 1997. Phylogenetic inference for binary data on dendrograms using Markov chain Monte Carlo. *J. Comput. Graph. Stat.* 6:122–31
- Mau B, Newton MA, Larget B. 1999. Bayesian phylogenetic inference via Markov chain Monte Carlo methods. *Biometrika* 55:1–12
- Miller AN, Andrew NMA. 2005. Multi-gene phylogenies indicate ascomal wall morphology is a better predictor of phylogenetic relationships than ascospore morphology in the sordariales (ascomycota, fungi). *Mol. Phylogenet. Evol.* 35:60–75
- Miller RE, Buckley TR, Manos P. 2002. An examination of the monophyly of morning glory taxa using Bayesian phylogenetic inference. *Syst. Biol.* 51:740–53
- Mossel E, Vigoda E. 2005. Phylogenetic MCMC algorithms are misleading on mixtures of trees. *Science* 309:2207–9
- Murphy WJ, Eizirik E, O'Brien SJ, Madsen O, Scally M, et al. 2001. Resolution of the early placental mammal radiation using Bayesian phylogenetics. *Science* 294:2348–51
- Newton MA. 1996. Bootstrapping phylogenies: large deviations and dispersion effects. *Biometrika* 83:315–28
- Newton MA, Mau B, Larget B. 1999. Markov chain Monte Carlo for the Bayesian analysis of evolutionary trees from aligned molecular sequences. In *Statistics in Molecular Biology and Genetics*, ed. F Seiller-Moseiwitch, TP Speed, M Waterman, 33:143–62. Hayward, CA: Monogr. Ser. Inst. Math. Stat.
- Nylander J, Ronquist F, Huelsenbeck JP. 2004. Bayesian phylogenetic analysis of combined data. *Syst. Biol.* 53:47–67
- Pagel M, Lutzoni F. 2002. Accounting for phylogenetic uncertainty in comparative studies of evolution and adaptation. In *Biological Evolution and Statistical Physics*, ed. M Laessig, A Valleriani, pp. 151–64. Berlin: Springer-Verlag
- Pickett KM, Randle CP. 2005. Strange Bayes indeed: uniform topological priors. *Mol. Phylogenet. Evol.* 34:203–11

- Randle CP, Mort ME, Crawford DJ. 2005. Bayesian inference of phylogenetics revisited: developments and concerns. *Taxon* 54:9–15
- Randle CP, Pickett KM. 2006. Are nonuniform clade priors important in Bayesian phylogenetic analysis? A response to Brandley et al. *Syst. Biol.* 55:147–50
- Rannala B. 2002. Identifiability of parameters in MCMC Bayesian inference of phylogeny. *Syst. Biol.* 51:754–60
- Rannala B, Yang Z. 1996. Probability distribution of molecular evolutionary trees: a new method of phylogenetic inference. *J. Mol. Evol.* 43:304–11
- Ronquist F, Huelsenbeck JP. 2003. MRBAYES 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19:1572–74
- Ronquist F, Huelsenbeck JP, Britton T. 2004. Bayesian supertrees. In *Phylogenetic Supertrees: Combining Information to Reveal the Tree of Life*, ed. ORP Bininda-Emonds, pp. 193–224. Amsterdam: Kluwer
- Sanderson MJ. 1995. Objections to bootstrapping phylogenies: a critique. *Syst. Biol.* 44:299–320
- Sanderson MJ, Wojciechowski MF. 2000. Improved bootstrap confidence limits in large-scale phylogenies, with an example from the Neo-Astragalus (Leguminosae). *Syst. Biol.* 49:671–85
- Simmons M, Pickett K, Miya M. 2004. How meaningful are Bayesian support values? *Mol. Biol. Evol.* 21:188–99
- Simon D, Larget B. 2000. Bayesian analysis in molecular biology and evolution (BAMBE), version 2.03 beta. Dept. Math. Comput. Sci., Duquesne Univ., Pittsburgh
- Steel M, Pickett KM. 2006. On the impossibility of uniform priors on clades. *Mol. Phylogenet. Evol.* 39:585–86
- Steel MA, Penny D. 1993. Distributions of tree comparison metrics: some new results. *Syst. Biol.* 42:126–41
- Streelman JT, Alfaro ME, Westneat MW, Bellwood DR, Karl SA. 2002. Evolutionary history of the parrotfishes: biogeography, ecomorphology, and comparative diversity. *Evolution* 56:961–71
- Suchard MA, Weiss RE, Sinsheimer JS. 2001. Bayesian selection of continuous-time Markov chain evolutionary models. *Mol. Biol. Evol.* 18:1001–13
- Sullivan J, Joyce P. 2005. Model selection in phylogenetics. *Annu. Rev. Ecol. Syst.* 36:445–66
- Suzuki Y, Glazko GV, Nei M. 2002. Overcredibility of molecular phylogenies obtained by Bayesian phylogenetics. *Proc. Natl. Acad. Sci. USA* 99:16138–43
- Svennblad B, Erixon P, Oxelman B, Britton T. 2006. Fundamental differences between the methods of maximum likelihood and maximum posterior probability in phylogenetics. *Syst. Biol.* 55:116–21
- Taylor D, Piel W. 2004. An assessment of accuracy, error, and conflict with support values from genome-scale phylogenetic data. *Mol. Biol. Evol.* 21:1534–37
- Thompson EA. 1975. *Human Evolutionary Trees*. Cambridge, UK: Cambridge Univ. Press
- Waddell PJ, Kishino H, Ota R. 2002. Very fast algorithms for evaluating the stability of ML and Bayesian phylogenetic trees from sequence data. *Genome Inform.* 13:82–92

- Whittingham LA, Slikas B, Winkler DW, Sheldon FH. 2002. Phylogeny of the tree swallow genus, *Tachycineta* (Aves: Hirundinidae), by Bayesian analysis of mitochondrial DNA sequences. *Mol. Phylogenet. Evol.* 22:430–41
- Wilcox T, Zwick D, Heath T, Hillis D. 2002. Phylogenetic relationships of the dwarf boas and a comparison of Bayesian and bootstrap measures of phylogenetic support. *Mol. Phylogenet. Evol.* 25:361–71
- Yang Z. 2005. Bayesian inference in molecular phylogenetics. See Gascuel 2005, pp. 63–90
- Yang Z, Goldman N, Friday A. 1995. Maximum likelihood tree from DNA sequences: a peculiar statistical estimation problem. *Syst. Biol.* 44: 384–99
- Yang Z, Rannala B. 1997. Bayesian phylogenetic inference using DNA sequences: a Markov chain Monte Carlo method. *Mol. Biol. Evol.* 14:717–24
- Yang Z, Rannala B. 2005. Branch-length prior influences Bayesian posterior probability of phylogeny. *Syst. Biol.* 54:455–70
- Zharkikh A, Li W. 1992. Statistical properties of bootstrap estimation of phylogenetic variability from nucleotide sequences. I. Four taxa with a molecular clock. *Mol. Biol. Evol.* 9:1119–47
- Zharkikh A, Li W. 1995. Estimation of confidence in phylogeny: the complete-and-partial bootstrap technique. *Mol. Phylogenet. Evol.* 4:44–63
- Zwickl D, Holder MT. 2004. Model parameterization, prior distributions, and the general time-reversible model in Bayesian phylogenetics. *Syst. Biol.* 53:877–88



Contents

Birth-Death Models in Macroevolution <i>Sean Nee</i>	1
The Posterior and the Prior in Bayesian Phylogenetics <i>Michael E. Alfaro and Mark T. Holder</i>	19
Unifying and Testing Models of Sexual Selection <i>Hanna Kokko, Michael D. Jennions, and Robert Brooks</i>	43
Genetic Polymorphism in Heterogeneous Environments: The Age of Genomics <i>Philip W. Hedrick</i>	67
Ecological Effects of Invasive Arthropod Generalist Predators <i>William E. Snyder and Edward W. Evans</i>	95
The Evolution of Genetic Architecture <i>Thomas F. Hansen</i>	123
The Major Histocompatibility Complex, Sexual Selection, and Mate Choice <i>Manfred Milinski</i>	159
Some Evolutionary Consequences of Being a Tree <i>Rémy J. Petit and Arndt Hampe</i>	187
Late Quaternary Extinctions: State of the Debate <i>Paul L. Koch and Anthony D. Barnosky</i>	215
Innate Immunity, Environmental Drivers, and Disease Ecology of Marine and Freshwater Invertebrates <i>Laura D. Mydlarz, Laura E. Jones, and C. Drew Harvell</i>	251
Experimental Methods for Measuring Gene Interactions <i>Jeffery P. Demuth and Michael J. Wade</i>	289
Corridors for Conservation: Integrating Pattern and Process <i>Cheryl-Lesley B. Chetkiewicz, Colleen Cassady St. Clair, and Mark S. Boyce</i>	317

The Population Biology of Large Brown Seaweeds: Ecological Consequences of Multiphase Life Histories in Dynamic Coastal Environments <i>David R. Schiel and Michael S. Foster</i>	343
Living on the Edge of Two Changing Worlds: Forecasting the Responses of Rocky Intertidal Ecosystems to Climate Change <i>Brian Helmuth, Nova Mieszkowska, Pippa Moore, and Stephen J. Hawkins</i>	373
Has Vicariance or Dispersal Been the Predominant Biogeographic Force in Madagascar? Only Time Will Tell <i>Anne D. Yoder and Michael D. Nowak</i>	405
Limits to the Adaptive Potential of Small Populations <i>Yvonne Willi, Josh Van Buskirk, and Ary A. Hoffmann</i>	433
Resource Exchange in the Rhizosphere: Molecular Tools and the Microbial Perspective <i>Zoe G. Cardon and Daniel J. Gage</i>	459
The Role of Hybridization in the Evolution of Reef Corals <i>Bette L. Willis, Madeleine J.H. van Oppen, David J. Miller, Steve V. Vollmer, and David J. Ayre</i>	489
The New Bioinformatics: Integrating Ecological Data from the Gene to the Biosphere <i>Matthew B. Jones, Mark P. Schildbauer, O.J. Reichman, and Shawn Bowers</i>	519
Incorporating Molecular Evolution into Phylogenetic Analysis, and a New Compilation of Conserved Polymerase Chain Reaction Primers for Animal Mitochondrial DNA <i>Chris Simon, Thomas R. Buckley, Francesco Frati, James B. Stewart, and Andrew T. Beckenbach</i>	545
The Developmental, Physiological, Neural, and Genetical Causes and Consequences of Frequency-Dependent Selection in the Wild <i>Barry Sinervo and Ryan Calsbeek</i>	581
Carbon-Nitrogen Interactions in Terrestrial Ecosystems in Response to Rising Atmospheric Carbon Dioxide <i>Peter B. Reich, Bruce A. Hungate, and Yiqi Luo</i>	611
Ecological and Evolutionary Responses to Recent Climate Change <i>Camille Parmesan</i>	637

Indexes

Cumulative Index of Contributing Authors, Volumes 33–37	671
Cumulative Index of Chapter Titles, Volumes 33–37	674