

## Potential Applications and Pitfalls of Bayesian Inference of Phylogeny

JOHN P. HUELSENBECK,<sup>1</sup> BRET LARGET,<sup>2</sup> RICHARD E. MILLER,<sup>3</sup> AND FREDRIK RONQUIST<sup>4</sup>

<sup>1</sup>Section of Ecology, Behavior and Evolution, Division of Biology, University of California–San Diego, La Jolla, CA 92093-0116, USA

<sup>2</sup>Department of Mathematics and Computer Science, Duquesne University, Pittsburgh, PA 15282, USA

<sup>3</sup>Department of Biological Sciences, Southeastern Louisiana University, Hammond, LA 70402, USA

<sup>4</sup>Department of Systematic Zoology, Evolutionary Biology Centre, Uppsala University, Norbyg. 18D, SE-752 36 Uppsala, Sweden

**Abstract.**—Only recently has Bayesian inference of phylogeny been proposed. The method is now a practical alternative to the other methods; indeed, the method appears to possess advantages over the other methods in terms of ability to use complex models of evolution, ease of interpretation of the results, and computational efficiency. However, the method should be used cautiously. The results of a Bayesian analysis should be examined with respect to the sensitivity of the results to the priors used and the reliability of the Markov chain Monte Carlo approximation of the probabilities of trees. [Bayesian inference; Markov chain Monte Carlo; phylogeny; posterior probability.]

In 1996, three groups independently proposed using Bayesian inference of phylogeny. Rannala and Yang (1996), then at Berkeley, showed how posterior probabilities of trees, the central quantity in Bayesian inference of phylogeny, could be calculated under a birth–death prior for a small number of species. At about the same time, two groups of statisticians were doing much the same thing; Mau (1996) and Li (1996), at Wisconsin and Ohio State, respectively, published their dissertations that year, both showing how posterior probabilities of trees could be calculated under simple priors. The next few years saw the publication of a number of important papers, most of which concerned the nuts and bolts of calculating posterior probabilities for trees (Mau and Newton, 1997; Yang and Rannala, 1997; Larget and Simon, 1999; Mau et al., 1999; Newton et al., 1999). The method of choice for calculating posterior probabilities, in fact the only numerical method currently available that can approximate posterior probabilities for trees, is a technique firmly grounded in probability theory (Tierney, 1994) called Markov chain Monte Carlo (MCMC; Metropolis et al., 1953; Hastings, 1970; Green, 1995). The next few years also saw a number of papers that used Bayesian inference in a phylogenetic context to address questions such as “Is an amino acid under positive or diversifying selection?” (Nielsen and Yang, 1998), “When did a group of species diverge?” (Thorne et al., 1998; Huelsenbeck et al., 2000a), “Did hosts and parasites cospeciate?”

(Huelsenbeck et al., 2000b), and “What is the ancestral sequence for some clade?” (Yang et al., 1995b; Huelsenbeck and Bollback, 2001). Once the theory had been described and a number of computer programs became available (BAMBE: Simon and Larget, 1998; MrBayes: Huelsenbeck and Ronquist, 2001), it was only a matter of time before the method became more widely used.

The phylogenetics literature is full of new methods for estimating phylogeny, so the introduction of yet another method of phylogenetic inference does not seem like a portentous event. Moreover, Bayesian inference is not even new. Bayesian estimation is one of the oldest methods of statistical inference, dating back to the 18 century. A number of earlier papers had hinted at Bayesian inference of phylogeny. Felsenstein (1968), in his Ph.D. thesis, introduced a number of Bayesian ideas, such as posterior probabilities of trees (Felsenstein, 1968: 21–26) and a credible set of trees (Felsenstein, 1968:25–26; also see Wheeler, 1991). However, Felsenstein was not able to calculate posterior probabilities of trees at the time; Hastings’ seminal paper appeared two years later (1970), and it was another two decades before statisticians began to fully realize the power of the MCMC approach.

Bayesian inference of phylogeny represents a significant advance for a number of reasons. First, like the maximum likelihood method, Bayesian estimation of phylogeny is based on the likelihood function. The likelihood, a quantity that is proportional to the probability of observing the data conditional

on a tree, is the vehicle that carries the phylogenetic information contained in the data in both maximum likelihood and Bayesian estimation, and the same models of DNA, amino acid, or morphological evolution can be used in both methods. Maximum likelihood is known to outperform other methods of phylogenetic estimation under a range of conditions (Huelsenbeck and Hillis, 1993; Kuhner and Felsenstein, 1994; Huelsenbeck, 1995a, 1995b), and statistical theory suggests that this should be so. Because Bayesian inference is based on the likelihood function, it should inherit many of the nice statistical properties of the maximum likelihood method. Second, unlike maximum likelihood, Bayesian inference of phylogeny can incorporate a systematist's prior information about phylogeny through the specification of a prior probability distribution of trees. This is also where some of the different formulations of Bayesian estimation of phylogeny diverge. For example, Rannala and Yang (1996) placed a birth-death prior on trees, giving equal weight to labeled histories. Other formulations of the problem typically give equal weight to rooted or unrooted phylogenies. In any case, the use of a prior probability distribution on trees can be viewed as either a strength or a weakness of the method. It seems a strength when the systematist has prior information about the phylogeny of a group. Why not incorporate such information when it is available? However, when the systematist does not have strong prior beliefs, specifying a prior seems more difficult, with the usual solution being that all trees are given equal weight, *a priori*. Third, MCMC provides an elegant and computationally efficient way of approximating posterior probabilities of trees or other model parameters, the quantities of interest in Bayesian inference. MCMC is not limited to Bayesian analysis and has been used to account for uncertainty in genealogy for several coalescence problems in a maximum likelihood framework (Kuhner et al., 1994; Beerli and Felsenstein, 1999). However, the MCMC method more naturally fits a Bayesian framework, and the growing use of Bayesian methods in other fields dates from the widespread application of MCMC starting in the middle 1980s. Bayesian inference of phylogeny is interesting because it brings a totally new perspective to a number of problems of long standing in both evo-

lutionary biology and phylogenetics. To the evolutionary biologist interested in comparative analysis, Bayesian inference suggests a natural way to accommodate uncertainty in phylogeny (Huelsenbeck et al., 2000c). To the practicing systematist, Bayesian analysis provides an intuitive measure of support for trees and a practical way to estimate large phylogenies using a statistical approach.

The recent developments in Bayesian inference of phylogeny are bound to generate debate in the systematics community. Phylogenetics has a long history of heated arguments about the relative merits of different methods—researchers in the field seem preadapted for ideological warfare—and the introduction of notions such as prior and posterior probabilities should spur new battles. In fact, the debate has been intense for some time already in the statistics literature. Many statisticians hold strong beliefs about the relative merit of so called frequentist (e.g., maximum likelihood) and Bayesian methods of inference, and the tone of the debate sometimes sounds peculiarly familiar to a systematist. Ultimately, of course, the biologist will have to decide for him- or herself whether Bayesian inference of phylogeny is sensible. This introduction is designed to help this decision process by describing Bayesian inference of phylogeny and the methods used to approximate posterior probabilities of trees. We also outline some of the pitfalls and remaining unsolved mysteries of Bayesian analysis of phylogeny and highlight some potential applications.

#### WHAT IS BAYESIAN ESTIMATION OF PHYLOGENY?

As mentioned above, Bayesian estimation of phylogeny is based on a quantity called the posterior probability distribution of trees. To calculate the posterior probability of a tree, we start by labeling all trees from 1 to  $B(s)$ , where  $B(s)$  is the number of possible trees for  $s$  species; the notation we use is  $\tau_1, \tau_2, \dots, \tau_{B(s)}$  for the first, second, third, etc., tree. We also have some observations, such as an alignment of DNA sequences or a matrix of morphological characters, which we denote  $\mathbf{X}$ . The posterior probability of trees is the probability of the  $i$ th tree conditional on the observations and is calculated using

Bayes's theorem as

$$\Pr[\tau_i | \mathbf{X}] = \frac{\Pr[\mathbf{X} | \tau_i] \times \Pr[\tau_i]}{\sum_{j=1}^{B(s)} \Pr[\mathbf{X} | \tau_j] \times \Pr[\tau_j]},$$

where  $\Pr[\tau_i | \mathbf{X}]$  is the posterior probability of tree  $i$ ,  $\Pr[\mathbf{X} | \tau_i]$  is the likelihood of tree  $i$ , and  $\Pr[\tau_i]$  is the prior probability of tree  $i$ . The denominator is a normalizing constant that involves a summation over all  $B(s)$  possible trees:  $B(s) = (2s - 3)!/[2^{s-2}(s - 2)!]$  for rooted trees and  $B(s) = (2s - 5)!/[2^{s-3}(s - 3)!]$  for unrooted trees. Bayes's formula shows how a person who started out with one set of beliefs, formulated in the prior probability distribution of trees, should modify his or her beliefs in the light of new observations.

The above description of the Bayesian approach to phylogenetics is actually incomplete because the tree,  $\tau$ , provides insufficient information with which to calculate likelihoods. One also needs a substitution model describing how characters evolve on the tree and information on the value of the parameters in this model,  $\theta$ , and on the lengths of the branches on the tree,  $v$ . Branch lengths are typically in terms of expected number of substitutions per site. The parameters of the substitution model vary with the details of the analysis but typically include variables related to the frequency of the character states, biases in rates of change among character states, and rate variation across sites. The likelihood of the  $i$ th tree is obtained by integrating over all possible combinations of branch lengths and substitution model parameters:

$$\Pr[\mathbf{X} | \tau_i] = \int_{v_i, \theta} f(\mathbf{X} | \tau_i, v_i, \theta) f(v_i, \theta) dv_i d\theta,$$

where  $f(v_i, \theta)$  is the prior probability density of the branch lengths and substitution model parameters. (Tree topology, here denoted  $\tau$ , is a discrete parameter, and it makes sense to think about the probability of the tree taking some specific topology  $\tau_i$ . However, many of the parameters of the phylogenetic model are continuous, and the probability of any specific parameter value is 0. For continuous parameters, the idea of infinitesimal probabilities is used instead; the probability of a parameter,  $\Theta$ , taking a spe-

cific value,  $\theta$ , is  $f(\theta)d\theta$ , where  $f(\theta)$  is a probability density and  $d\theta$  is an infinitesimal interval). The likelihood  $[f(\mathbf{X} | \tau_i, v_i, \theta)]$  can be calculated under the same models of evolution used by maximum likelihood. The prior probability density distribution  $[f(v_i, \theta)]$  is more difficult to specify. Usually, so-called flat priors are placed on the parameters over a range of values likely to contain the true value of the parameter. Flat priors place equal probability on all possible values within the specified range. In the Bayesian literature, sometimes improper uniform priors are advocated. Here, uniform probability is given to all possible values of the parameter. The prior is "improper" because it is not a probability distribution, failing to integrate to 1. Often, but not always, the use of improper priors leads to an improper posterior probability distribution. The use of improper priors has not been explored for the phylogeny problem, but for some parameters, such as branch lengths, the use of improper priors is obviously inappropriate.

How do we derive phylogenetic conclusions from the posterior probability distribution of trees? One possibility is to use the most probable tree as a point estimate of phylogeny (Rannala and Yang, 1996). This is called the maximum a posteriori probability (MAP) estimate of phylogeny. A standard Bayesian approach to summarize results is to form a 95% credibility interval for the parameter of interest based on the posterior distribution. Note the Bayesian use of the term credibility interval for an entity that is similar, at least in some ways, to the confidence interval of classical statistics. In phylogeny construction, a 95% credible set of trees can be constructed by starting with the MAP tree and adding trees in order of decreasing probability until the cumulative probability is 0.95 (Felsenstein, 1968). The most powerful approach is perhaps to summarize the results of a Bayesian analysis on a majority rule consensus tree or on the MAP tree (Larget and Simon, 1999), as is typically done when summarizing the results of a bootstrap analysis (Felsenstein, 1985). However, the numbers on the branches of the tree now represent the posterior probability that the clade is true. Although the MAP, credible set, and majority rule consensus tree are the usual ways to summarize the results of a Bayesian analysis of phylogeny, other methods, equally valid, also could be devised.

# HOW ARE POSTERIOR PROBABILITIES OF TREES APPROXIMATED?

Calculating the posterior probability of a tree involves a summation over all possible trees and, for each tree, integration over all combinations of branch lengths and substitution-model parameter values. The summation over trees and integration over all combinations of branch lengths and parameter values is impossible to perform analytically, except perhaps in a few of the very simplest cases. By necessity, then, posterior probabilities of trees must be approximated. There are a number of methods available for approximating integrals, some of which involve sampling procedures. Ideally, one would like to approximate the posterior probability distribution of trees by randomly drawing from the posterior probability distribution of trees. The fraction of the time any tree was drawn would be a valid approximation of its posterior probability. This method, called Monte Carlo integration, has the advantages that each draw is independent and the investigator has control over the error in the approximation (if one wants a better approximation, more trees are drawn from the posterior probability distribution of trees). Unfortunately, this ideal cannot yet be realized. However, one can obtain information equivalent to that from independent samples by instead taking much larger dependent samples using the MCMC method. This method appears to be very efficient for approximating posterior probabilities of trees.

MCMC works as follows. First, start the Markov chain with a tree. This tree might be one that is randomly chosen or one that is likely to be a good description of the data. We designate this tree (with branch lengths)  $\Psi = (\tau, \nu)$ . Second, a new tree, designated  $\Psi'$ , is proposed. The proposal mechanism that changes  $\Psi$  into  $\Psi'$  must satisfy a few conditions: (1) the proposal mechanism must be stochastic, i.e., the probability of proposing the new tree given the old tree is  $f(\Psi' | \Psi)$  and the probability of the reverse move, which is not actually made, is  $f(\Psi | \Psi')$ ; (2) every possible tree must be accessible by repeated application of the proposal mechanism (i.e., the Markov chain must be irreducible); and (3) the chain must be aperiodic. Other than these requirements, the details of the proposal mechanism are up to the investigator. Unfortunately, the proposal details may

greatly influence the efficiency of the chain in approximating the posterior probability distribution. There is an art to devising proposal mechanisms that efficiently explore the space of trees. Third, the new tree is accepted with a probability described by Metropolis et al. (1953) and Hastings (1970):

$$R = \min \left[ 1, \frac{f(\Psi' | X)}{f(\Psi | X)} \times \frac{f(\Psi | \Psi')}{f(\Psi' | \Psi)} \right].$$

The acceptance probability,  $R$ , is the probability that the proposed state,  $\Psi'$ , becomes the next state of the chain. At first glance, it would not appear that the above formula allows approximation of the posterior probability. After all,  $f(\Psi | X)$  is the posterior probability of a tree and involves, minimally, a summation over all possible trees. However, the beauty of the Metropolis–Hastings algorithm is that the denominator of Bayes's theorem, the complex summation and integration, cancels out. Thus, the above formula reduces to the product of three ratios, each of which can be readily calculated:

$$\begin{aligned} R &= \min \left[ 1, \frac{f(\Psi' | X)}{f(\Psi | X)} \times \frac{f(\Psi | \Psi')}{f(\Psi' | \Psi)} \right] \\ &= \min \left[ 1, \frac{f(X | \Psi')f(\Psi')/f(X)}{f(X | \Psi)f(\Psi)/f(X)} \times \frac{f(\Psi | \Psi')}{f(\Psi' | \Psi)} \right] \\ &= \min \left[ 1, \underbrace{\frac{f(X | \Psi')}{f(X | \Psi)}}_{\text{likelihood ratio}} \times \underbrace{\frac{f(\Psi')}{f(\Psi)}}_{\text{prior ratio}} \right. \\ &\quad \left. \times \underbrace{\frac{f(\Psi | \Psi')}{f(\Psi' | \Psi)}}_{\text{proposal ratio}} \right], \end{aligned}$$

where  $f(X | \Psi')/f(X | \Psi)$  is the likelihood ratio,  $f(\Psi')/f(\Psi)$  is the prior ratio, and  $f(\Psi | \Psi')/f(\Psi' | \Psi)$  is the proposal ratio. The proposal ratio is also referred to as the Hastings ratio. Fourth, a uniform random number on the interval (0,1) is generated. If this number is less than  $R$ , then the new state is accepted, and  $\Psi = \Psi'$ . Otherwise, the chain remains in state  $\Psi$ . Fifth, go back to step 2. Steps 2–5 are repeated many thousands or millions of times. The fraction

of the time that the chain visits any particular tree is a valid approximation of the posterior probability of that tree. If the chain is designed so that the proposal ratio is 1, proposals of trees with higher posterior probabilities are always accepted, whereas proposals to trees with lower posterior probabilities are accepted with a probability that depends on the probability of the proposed tree relative to the probability of the current tree.

The usual procedure for the phylogeny problem is to update parameters of the model in blocks. For example, a number of different proposal mechanisms might be used, each of which changes a different parameter; some might change the tree and branch lengths and others might change parameters of the substitution model.

#### AN APPLICATION OF BAYESIAN INFERENCE

In this section, we present the results of a small survey study in which people (both systematists and amateurs) were asked to evaluate the monophyly of a plant genus, *Ipomoea*; it shows that people can easily formulate prior opinions and modify their beliefs in a reasonable manner in light of new observations, although we do not prove that the individuals in the study updated their opinions using Bayes's rule. The use by Bayesian methods of subjective probabilities simply formalizes something that systematists do on a regular basis. Systematics has a tradition of using very strong priors in phylogenetic analysis in the form of constraints on trees; these constraints place prior probabilities of zero on large classes of trees. However, the survey also demonstrates some of the potential pitfalls of explicitly incorporating prior probabilities of trees.

*Ipomoea* is a large genus within the Convolvulaceae containing over 600 species of vines and shrubs distributed throughout the tropics and subtropics (McDonald, 1991; Austin and Huaman, 1996; Wilkin, 1999). *Ipomoea* has been placed within the tribe of morning glories with spiny pollen, Ipomoeae (sensu lato), which also includes the genera *Argyreia*, *Lepistemon*, *Stictocardia*, and *Turbina*. The monophyly of *Ipomoea* is consistent with the taxonomy of this genus and its close relatives (Van Ooststroom, 1953; Verdcourt, 1963; McDonald, 1991; Austin and Huaman, 1996) (Table 1). The monophyly of

TABLE 1. A classification of the species included in the phylogenetic analysis of this study.

---

Family Convolvulaceae
Tribe Merremieae
<i>Merremia tuberosa</i> (L.) Rendle
<i>Operculina brownii</i> Ooststr.
Tribe Ipomoeae
Genus <i>Argyreia</i>
<i>Argyreia nervosa</i> (Burm. F.) Bojer
Genus <i>Lepistemon</i>
<i>Lepistemon owariense</i> (P. Beauv.) Hallier f.
Genus <i>Stictocardia</i>
<i>Stictocardia tiliifolia</i> (Desr.) Hallier f.
Genus <i>Turbina</i>
<i>Turbina holubii</i> (Baker) A. Meeuse
Genus <i>Ipomoea</i> L. (all species listed below this point are in the genus <i>Ipomoea</i> )
Subgenus <i>Ipomoea</i>
Section <i>Ipomoea</i>
<i>Ipomoea arachnosperma</i> Welw.
<i>Ipomoea pes-tigridis</i> L.
Section <i>Pharbitis</i> (Choisy) Griseb.
<i>Ipomoea hederacea</i> Jacq.
<i>Ipomoea purpurea</i> (L.) Roth
Subgenus <i>Quamoclit</i> (Moench) Clarke
Section <i>Mina</i> (Cerv.) Griseb.
<i>Ipomoea lobata</i> (Cerv.) Thell.
<i>Ipomoea quamoclit</i> L.
Section <i>Calonyction</i> (Choisy) Griseb.
<i>Ipomoea alba</i> L.
Section <i>Tricolores</i> J. A. McDonald
<i>Ipomoea tricolor</i> Cav.
Section <i>Orthipomoea</i> Choisy
<i>Ipomoea eriocarpa</i> R. Br.
<i>Ipomoea plebeia</i> R. Br.
Subgenus <i>Eriospermum</i> (Hallier f.) Verdcourt ex Austin
Section <i>Eriospermum</i> Hallier f.
<i>Ipomoea amnicola</i> Morong.
<i>Ipomoea arborescens</i> (Humb. & Bonpl. ex Willd.) G. Don
<i>Ipomoea batatas</i> (L.) Lam.
<i>Ipomoea carnea</i> Jacq.
<i>Ipomoea leptophylla</i> Torr.
<i>Ipomoea pedicellaris</i> Benth.
<i>Ipomoea setosa</i> Ker Gawl.
<i>Ipomoea sumatrana</i> (Miq.) Ooststr.
<i>Ipomoea umbraticola</i> House
Section <i>Erpipomoea</i> Choisy
<i>Ipomoea aquatica</i> Forssk.
<i>Ipomoea cairica</i> (L.) Sweet
<i>Ipomoea ochracea</i> (Lindl.) G. Don
<i>Ipomoea pes-caprae</i> R. Br.

---

the group was addressed here through a sample of 23 *Ipomoea* species and one species of each of the genera *Argyreia*, *Lepistemon*, *Stictocardia*, and *Turbina*. *Merremia* and *Operculina* from the tribe Merremieae provided out-group taxa. Participants in the survey were asked to state their prior beliefs about the monophyly of the *Ipomoea* species, using as information the classification. The histogram of the responses is shown in Figure 1A.

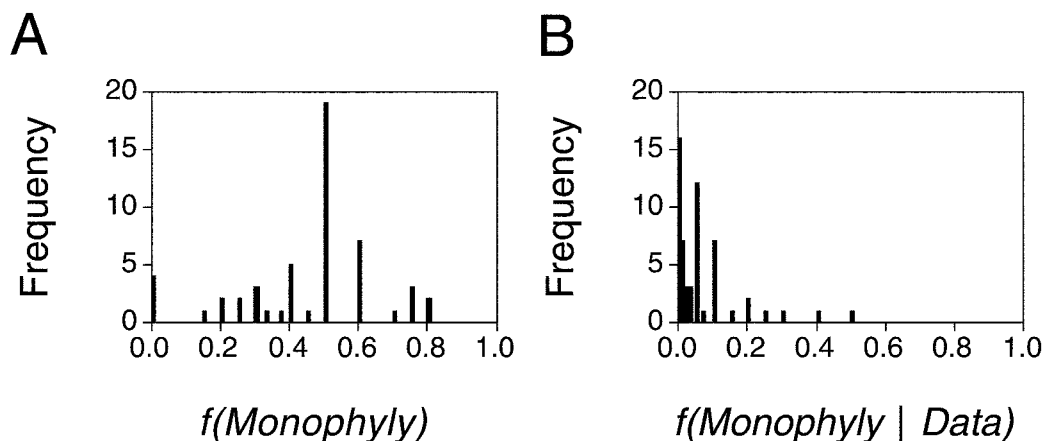


FIGURE 1. Frequency histograms of the responses concerning *Ipomoea* monophyly. (A) Prior beliefs. (B) Updated beliefs.

Miller et al. (1999) and Manos et al. (2001) collected ITS and *waxy* DNA sequences for 27 species of morning glories and two out-group taxa. The sequences were 649 and 651 sites in length for *waxy* and ITS, respectively. Figure 2 shows a phylogenetic tree estimated using maximum likelihood under the GTR +  $\Gamma$  model of DNA substitution. This model allows different nucleotide changes to have different rates, different nucleotide frequencies, and among-site rate variation. The numbers on the internal branches represent the nonparametric bootstrap support for individual clades. This tree is inconsistent with *Ipomoea* monophyly (as is the maximum parsimony tree). Moreover, in the 100 nonparametric bootstrap replicates, not one tree consistent with *Ipomoea* monophyly was observed. The log likelihood for the combined data for the maximum likelihood tree was  $-7859.76$ . The log likelihood of the best tree consistent with *Ipomoea* monophyly was much lower,  $-7910.46$ .

Figure 1B shows the modified probabilities of the participants. Virtually all participants revised their assessment of *Ipomoea* monophyly after observing the results of the ITS and *waxy* analysis by lowering the probability of monophyly (two participants did not change their beliefs at all). Many of the participants gave a lot of credence to the taxonomy, on which they based their prior of *Ipomoea* monophyly; most placed the prior probability of *Ipomoea* monophyly at about 0.5, with the argument that this probability was fair, giving no preference to *Ipomoea* monophyly

or nonmonophyly. This prior could be interpreted as a belief that about half of all phylogenies determined by morphological characteristics are, in fact, incorrect. A less intuitive consequence is that, because there are so few trees consistent with *Ipomoea* monophyly a prior probability of 0.5 places a tremendous amount of prior weight on individual trees consistent with *Ipomoea* monophyly. There are a total of  $5.33 \times 10^{29}$  trees consistent with *Ipomoea* monophyly and  $1.58 \times 10^{35}$  trees inconsistent with *Ipomoea* monophyly. By specifying a prior that has the probability of *Ipomoea* monophyletic as 0.5, individual trees consistent with *Ipomoea* monophyly have 296,366 times more prior probability than do individual trees inconsistent with *Ipomoea* monophyly; thus, the prior odds are weighted tremendously in favor of *Ipomoea* monophyly. A prior that places uniform probabilities on all possible tree topologies completely discounts the information that previous systematists had classified *Ipomoea* as a monophyletic group.

The monophyly of *Ipomoea* can also be evaluated using a computer program; the computer uses Bayes's theorem to update the posterior probabilities of phylogenetic trees based on DNA sequence data. The probability that *Ipomoea* is monophyletic is simply the sum of the posterior probabilities of trees having *Ipomoea* monophyletic. If an equal-weight prior is placed on all phylogenetic trees (i.e., the prior probability of *Ipomoea* monophyly is  $3.37 \times 10^{-6}$ ), the posterior probability of monophyly is essentially 0.

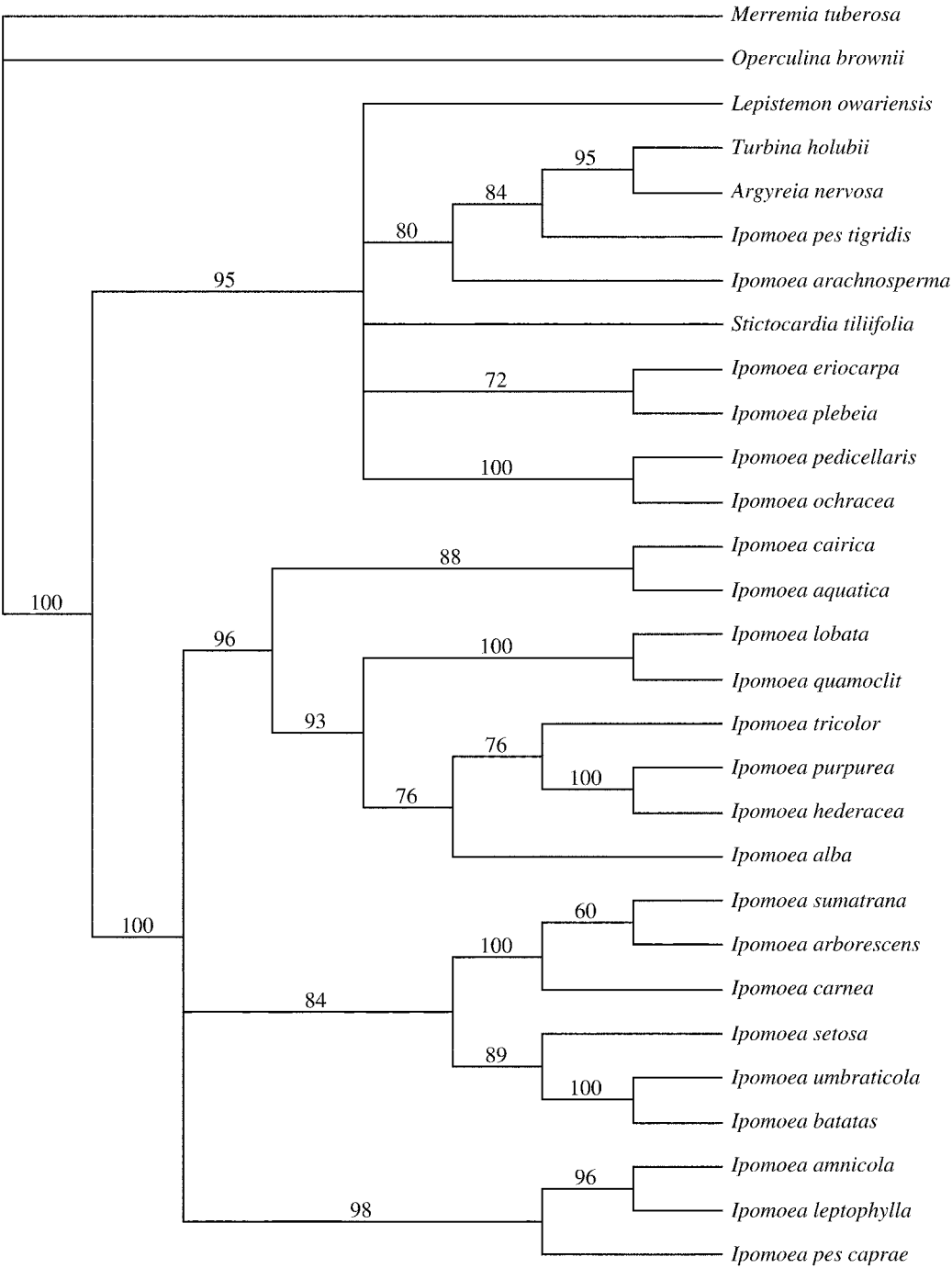


FIGURE 2. The majority rule consensus tree of maximum likelihood analyses of 100 nonparametric bootstrap replicates under the GTR +  $\Gamma$  model of DNA substitution.

Even if the prior probability of *Ipomoea* monophyly is as high as 0.9, the posterior probability of monophyly is so small as to be virtually immeasurable using Monte Carlo techniques (i.e.,  $<0.01$ ). For this example, the likelihood function overwhelms the prior placed on *Ipomoea* monophyly.

A strict application of Bayes's theorem more radically changes belief about *Ipomoea* monophyly than do the people in the survey. Does this mean that people do not behave like Bayesians? Not necessarily. The participants are likely accounting for more sources of uncertainty than does the model used for the MCMC calculation. These other influences include the possibilities that the alignment of the DNA sequences is incorrect and that the assumptions of the phylogenetic analysis are wrong. One person even considered the (small) chance that the data were fraudulent. The result was that the participants were more conservative in how they modified their beliefs. These additional factors are difficult to accommodate in a computer program; computers have a strong tendency to treat data and models quite literally.

#### PITFALLS

For many systematists, the most exciting aspect of Bayesian inference using MCMC is its computational efficiency. Problems that have previously been impossible to analyze with statistical methods because of their size or the complexity of the relevant models, even given the resources of a supercomputer center, can now be tackled within days on an ordinary desktop computer. It is easy to be carried away by the possibilities and forget that there are also pitfalls of Bayesian analysis. For beginners, the most important concern is often the sensitivity of the results, the posterior distribution, to the chosen prior distributions. Indeed, some critics emphasize the subjective nature of Bayesian probabilities and contrast it with the general striving for objectivity in science. A more practical but potentially serious problem is related to the MCMC technique: How can we determine when chains have been run long enough to produce a reasonable sample from the posterior distribution?

#### *Sensitivity to Priors*

Bayesian estimation is unique among phylogenetic methods because, along with the

usual assumptions about the evolutionary process that generated the observations, it forces the investigator to specify any prior beliefs. All parameters in the evolutionary model, including phylogeny, branch lengths, and substitution model parameters, must be associated with prior probability distributions. But why is it necessary to specify prior beliefs? Doesn't the explicit incorporation of prior beliefs make all Bayesian probabilities subjective?

One can turn the first question around and ask, "Why should background knowledge be ignored?" It is not only sensible to base conclusions on all the available information (total evidence); in some cases, it is even critical. Assume, for instance, that you are asked to bet money on the relative probability,  $P$ , of turning up heads with a particular coin. You are only allowed to toss the coin twice, and it turns up heads both times. Maximum likelihood estimation based on these observations would give the estimated value  $P = 1.0$ . Of course, you cannot exclude the possibility that the coin is so heavily biased that it always turns up heads. However, the coin was taken from your wallet and it looks and feels just like a normal coin. If the coin were fair, there would still be a probability of 0.5 of turning up either heads or tails twice in two tosses. Given this information, it seems unlikely that the maximum likelihood estimate is close to the true value. In situations like this, the only sensible approach is to combine the information from your observations with the available background information, in this case suggesting that the true value of  $P$  is likely to be closer to 0.5 than to 1.0.

An interesting aspect of the explicit specification of priors on all model parameters is that when drawing inferences about one parameter (topology for instance), it allows us to integrate out all other parameters. The posterior probability of a tree, for instance, represents the probability summed (integrated) over all possible branch lengths and substitution model parameter values. Integrating out irrelevant parameters is likely to be associated with good statistical properties, such as robustness (Berger et al., 1999). In maximum likelihood estimation, however, the usual approach is to use the profile likelihood instead of the integrated likelihood; for profile likelihood, likelihoods are calculated by maximizing the likelihood with respect to the parameters that are not of direct interest to the



biologist. Such parameters are referred to as nuisance parameters, and they must be estimated along with the parameter(s) of interest (for discussions of nuisance parameters and integrated likelihood, see Goldman, 1990; Berger et al., 1999). Nuisance parameters typically include the branch lengths as well as a number of substitution model parameters.

As more and more data are added to a Bayesian analysis, the influence of the prior beliefs on the posterior distribution decreases. In a typical Bayesian analysis of phylogeny, the results are likely to be rather insensitive to the prior. In the *Ipomoea* study, in the most extreme cases persons placed 2.5 million times (15 log likelihood units) more weight on trees consistent with *Ipomoea* monophyly than on trees inconsistent with *Ipomoea* monophyly. Nonetheless, this extreme prior was overwhelmed by the data: The difference in likelihood between the best trees with and without *Ipomoea* monophyly was about 50 log likelihood units. The prior would have to be much more in favor of *Ipomoea* monophyly to substantially change the posterior probability distribution of trees.

Even if it is unlikely to be a problem in most Bayesian analyses of phylogeny, the sensitivity to the chosen priors remains a legitimate concern. It is good practice to describe the priors of a Bayesian analysis carefully, particularly if they are nonstandard. If there is reason to believe that the data are weak, the influence of the priors on the posterior distribution can be examined by rerunning the analysis with different priors. A good way of illustrating the relative contributions of priors and data to the posterior probability distribution is to formulate the support for a particular hypothesis before and after analysis in terms of odds. In the *Ipomoea* example, for instance, the most extreme persons placed the prior odds 2.5 million to 1 on *Ipomoea* being monophyletic. Updating these odds with the data, the posterior odds become  $1.6 \times 10^{15}$  to 1 against a monophyletic *Ipomoea* (for another example, see Huelsenbeck and Imennov, 2002).

### *Convergence and Mixing*

An appropriately constructed Markov chain will eventually produce dependent samples from the posterior probability distribution of trees. However, the chain is typically started from a point in parameter space

far removed from the regions with high posterior probability, and it will take some time before the chain finds the probable parameter values and produces a reasonable approximation of the posterior distribution. In MCMC parlance, there is a burn-in period before the chain converges onto the stationary or target distribution, the posterior probability distribution in this case. Even after convergence, the chain may move extremely slowly between different regions of the parameter space with high posterior probability, in which case the chain is said to mix poorly. How can we know that the chain we are sampling from has converged and mixes well? The disappointing answer is that it is impossible to know for certain. However, there are several good methods for examining both convergence and mixing.

The obvious way of monitoring convergence is to examine plots of log likelihood values. Initially, likelihood values tend to increase from generation to generation, but eventually the likelihood values cease to climb and instead fluctuate randomly up and down. This is taken as a sign of convergence, and the initial climbing phase is recognized as the burn-in period. The samples drawn from the chain before convergence are discarded, and our inferences are based exclusively on the remaining samples.

Unfortunately, log likelihood plots are notoriously unreliable for convergence monitoring (Gilks et al., 1996). For example, it is quite possible that a chain oscillates around what appears to be a stable likelihood value for many generations but then suddenly starts to climb again. Furthermore, the likelihood plot provides little information about the mixing behavior of the chain; two chains that are indistinguishable in their likelihood plots may sample from widely separate regions of the posterior distribution. Perhaps the most powerful approach to addressing these concerns is to compare independent MCMC runs. If several chains started from widely different places in parameter space end up producing indistinguishable samples, then that is a strong indication of convergence and appropriate mixing.

In MCMC estimation of posterior probability distributions resulting from phylogenetic analysis, the topology parameter is likely to be particularly problematic. Therefore, it makes sense to start independent runs with random trees and focus on the integration

over tree space when evaluating the results. For instance, the topological variance within and between chains can be compared using a suitable measure of tree-to-tree distance (Huelsenbeck et al., 2001). Before chains have converged, they will sample from different parts of tree space and the average tree-to-tree distance within chains will be considerably smaller than that among chains. After convergence, and given appropriate mixing, all chains will be sampling from the same distribution and the average tree-to-tree distance will be the same within and among chains. An alternative way of monitoring topological convergence is to compare the posterior clade probabilities among independent runs started from random trees. Again, similar clade probabilities in independent runs constitute strong evidence for convergence.

Even though topology is a particularly critical parameter, it is of course possible to examine convergence in the estimate of any parameter in the model. If the systematist were particularly interested in drawing inferences about the shape of the gamma distribution of rate variation, for example, it would be appropriate to monitor convergence for that parameter. Regardless of the parameter in focus, independent runs allow estimation of the so-called Monte Carlo error. Five or 10 independent MCMC runs may be sufficient to obtain a reasonable estimate of the error. In this way, we might for instance give the precision of the clade probabilities in a Bayesian MCMC analysis in terms of the observed Monte Carlo standard deviation around the mean.

The size and complexity of the problems that are possible to analyze with MCMC are determined by the convergence and mixing of the chains. Therefore, techniques that improve convergence and mixing are likely to be particularly important in extending the limits of MCMC analysis. A good illustration is provided by the Metropolis-coupling technique originally described by Geyer (1991). The idea is to run  $n$  Markov chains in parallel (the  $n$  chains are labeled  $i = 0, 1, \dots, n - 1$ ). One chain samples from the posterior distribution of interest,  $\Pr[\tau|X]$ , also referred to as the cold distribution. The other chains are sampling from heated distributions obtained by raising the cold distribution by  $\beta_i$ , where  $0 \leq \beta_i \leq 1$ . Thus, the heated distributions have the form  $\Pr[\tau|X]^{\beta_i}$ . Usually, if  $n$

chains are used then incremental heating of the form  $\beta_i = 1/(1 + iT)$  is applied, where  $T$  is a "temperature" parameter that is set to an appropriate value. At regular intervals, two chains are picked randomly and an attempt is made to change their states (parameter values) using a normal Metropolis step.

The effect of the heating is to decrease the difference between hilltops and valleys in the posterior distribution, such that the heated chains move more readily between isolated hilltops. The only function of the heated chains, however, is to provide the cold chain with intelligent proposals of new states. Because the heated chains move around more readily among the available hilltops, they can significantly improve mixing of the cold chain by letting it jump from one hill to another in a single step, a move that might otherwise take millions of generations because of the improbability of the intermediate states. Metropolis-coupled MCMC, or MCMCMC, or (MC)<sup>3</sup> for short, has extended the limits of MCMC analysis of phylogeny considerably; the largest successful (MC)<sup>3</sup> analyses, comprising more than 350 sequences (Huelsenbeck et al., 2001), integrate over a tree space that is several hundred orders of magnitude larger than the tree spaces that have been successfully analyzed without Metropolis coupling.

Despite the success of Metropolis coupling, convergence and mixing remain important problems for MCMC analysis of phylogeny. Undoubtedly, these problems will be a major focus of future research, and considering how young the field is, we may expect to see some significant advances that will extend the capabilities of MCMC in addressing phylogeny problems. For the time being, we suggest that the user, now aware of a few of the problems of convergence in MCMC algorithms, exercise caution. Keep in mind that "MCMC is a complex mixture of computer programming, statistical theory, and practical experience. When it works, it does things that cannot be done any other way, but it is good to remember that it is not fool-proof" (Geyer, 1999:80). We can also make a few general recommendations. First, we advocate that several long chains be run, especially when several processors are available. Although some statisticians have argued that convergence can be monitored from a single long chain (e.g., Geyer, 1992), the results from several chains can often lead to the discovery

of problems (e.g., when different chains give obviously different answers). Second, whenever possible, we advocate running multiple chains, each of which starts from a random tree. The disadvantage of using random trees (instead of, say, a neighbor-joining tree) to start the chain is that it may take a long time for the chain to find trees with high probability. However, convergence seems more assured if several chains, each of which started from different random trees, all give the same result. Third, many or all of the parameters of the model should be monitored for convergence. Different parameters of the phylogenetic model probably will converge at different rates.

#### UNSOLVED PROBLEMS

Because Bayesian analysis using MCMC is a new method in systematics, it is not surprising that there are a number of unsolved mysteries and potential problems that have not yet been addressed. Here, we limit ourselves to discussing three of these issues. Perhaps the most vexing mystery is the observed discrepancy between Bayesian posterior probabilities and nonparametric bootstrap support values. Another important problem is model validation. The use of existing phylogenetic results as priors in subsequent analyses or as the basis for the construction of supertrees also is an important field where additional research is necessary.

##### *Bootstrap Values and Posterior Probabilities*

A common empirical observation is that bootstrap support values, calculated using parsimony or maximum likelihood, tend to be lower than the corresponding Bayesian posterior probabilities. This raises a number of questions: Does the bootstrap measure something qualitatively different than is indicated by the posterior clade probabilities? Is the bootstrap more robust than the Bayesian posterior probabilities? Do the Bayesian posterior probabilities lead to overconfidence in the results?

This is a field where more research is clearly needed, and it seems impossible to provide any fast and hard answers currently. Although bootstrap values and Bayesian posterior probabilities both measure phylogenetic uncertainty, they are calculated very differently. The bootstrap, or the nonparametric bootstrap to be exact, is a measure

of uncertainty based on resampling from the original data matrix. Thus, it evaluates precision based on a multinomial sampling distribution, which does not in itself have any relation to an evolutionary model. The evolutionary model only enters into the calculations when a phylogenetic estimate is derived for each resampled matrix. Bayesian analysis, however, explicitly measures uncertainty based on the specified evolutionary model and the available data; it seems much more similar to the parametric bootstrap in this respect. These simple technical differences undoubtedly explain the discrepancies between nonparametric bootstrap support values and posterior clade probabilities. However, the prime concern is to find the underlying cause for the effect, because it may have profound implications.

Maximum likelihood analysis of phylogeny has been around for a while. Yet, some of the fundamental properties are still poorly known because standard statistical theory does not apply. For instance, analysis of phylogeny is not based on true likelihoods but on profile likelihoods, in which some nuisance parameters are estimated, as we have discussed. Problems arise because the profile likelihood does not incorporate the variability of the estimated nuisance parameters. A possible result is inadvertent imprecision in the phylogenetic estimate. Bayesian phylogenetic inference is not based on profile likelihoods but on marginal distributions, in which the uncertainty concerning all parameters except topology is integrated out. These considerations, then, suggest that one cause of the discrepancy may be that maximum likelihood tends to underestimate the confidence in phylogenetic results because of methodological differences.

The interpretation of the bootstrap itself, as applied to phylogenetics, is also under discussion. The standard bootstrap values do not accurately measure confidence in the traditional, hypothesis-testing context, but it is possible to calculate corrected bootstrap values (Efron et al., 1996; Sanderson and Wojciechowski, 2000). Unfortunately, the correction is computationally expensive, and it can result in either an increase or a decrease in the original bootstrap proportions. However, some arguments raised by Sanderson and Wojciechowski (2000) suggest that the correction is generally upwards for moderate to large data sets. In their 140-taxon

analysis, the bootstrap support for *Neo-Astragalus* rose from 67% to 93% after correction. Bayesian analysis of the same data produced a posterior probability for the *Neo-Astragalus* clade that agrees well with the corrected bootstrap proportion (Huelsenbeck et al., 2001). It is premature to base any far-reaching conclusions on these observations, but a major cause of discrepancy between bootstrap proportions and clade probabilities may be an inherent statistical bias in uncorrected bootstrap proportions.

A third potential explanation for the discrepancy is that Bayesian analysis is more sensitive than other methods to model misspecification. This possibility is difficult to address, and more research is clearly warranted. However, there do not appear to be any obvious reasons to expect Bayesian analysis to be more sensitive than other methods, such as maximum likelihood, to model misspecification. On the contrary, the use of marginal posterior probabilities instead of profile likelihoods could actually infer greater robustness of Bayesian methods.

A distinct advantage of the Bayesian posterior probabilities is that their interpretation is so clear-cut and intuitively appealing; they represent the probability that the corresponding clade is true given the model, the priors, and the data. Obviously, the Bayesian approach is not a panacea for problems in statistical inference of phylogeny, and there are already several anecdotal cases in which obviously erroneous or conflicting clades receive high posterior probabilities in Bayesian analyses. The most likely explanation is that the model is inappropriate for these data. These conflicting results can be considered as either a strength or a weakness of the method. We tend to see it as a strength; if the model is inappropriate then more work on the model is needed before we can hope to make substantial progress. Perhaps there is real incongruence between the gene tree and the species tree, perhaps there is time heterogeneity in the evolution of the characters, or perhaps there are other factors that distort the results. In any case, it seems important to discover the reason for the problem and take that into account in the analysis. Others will prefer to hide the uncertainty in an inconclusive support value. However, inappropriate modeling assumptions are likely to be problematic for all parametric statistical methods, and hiding potential problems in

inconclusive support values can be counterproductive in the long term if it slows down the development of more realistic evolutionary models.

### *Complex Priors on Trees*

Many scientists are attracted to the Bayesian approach because it seems to mimic how scientists behave in the real world; a scientist starts off with some belief about a hypothesis, collects data relevant to the hypothesis, and then modifies his or her beliefs in the light of the observations. Bayesian analysis simply formalizes this process. However, in practice, the Bayesian ideal is often not realized. For example, for the phylogeny problem, practicing systematists are forced to use the priors of the computer programmers who wrote the software package that performs the Bayesian phylogenetic analysis. Many people currently use a "Simon and Larget" prior when using the program BAMBE (Simon and Larget, 1998) or a "Huelsenbeck and Ronquist" prior when using the program MrBayes (Huelsenbeck and Ronquist, 2001). A practical problem in the phylogenetic context, then, is how to formulate the topological results of one analysis in a way that is suitable as a prior in a subsequent analysis or, for that matter, how to even specify complex priors based on previous taxonomic work. We need a prior probability distribution on trees, but even for a small phylogenetic problem the number of possible trees  $B(s)$  is so large that we cannot expect to be able to estimate the posterior probability of all of them, particularly not the ones with low posterior probabilities. Even if we could, we would have to deal with an enormous amount of information to specify the prior for the next analysis. For large analyses, it may even be difficult to estimate the posterior probability for the most likely trees, because even small clades that are poorly resolved will result in a large number of fully resolved trees of approximately equal probability.

### *Model Testing*

The results of a Bayesian analysis of phylogeny are contingent upon the chosen model being correct, so it is important to test whether the model provides an adequate description of the data. In a Bayesian analysis, the problem of model testing comes down to two related issues: checking a model and

examining the sensitivity of posterior probability distributions to the choice of model (Gelman et al., 1995).

There is no single agreed-upon method for checking models in a Bayesian analysis. Typically, the ability of a model to predict future observations (or more practically the ability of part of the collected data to predict the remaining data that were collected), Bayes factors (the ratio of the marginal likelihoods under two models, a quantity similar to the more familiar likelihood ratio test statistic) or posterior predictive  $P$  values, are used to check the model used in a Bayesian analysis. To date, in the phylogenetics literature, a more practical approach has been taken in which model choice is performed using likelihood ratio tests and then inferences are based on posterior probabilities of trees; little attention has been given to the methods of model choice developed in the Bayesian literature. However, the Bayesian methods might be useful in several contexts. Some of the methods do not require strict nesting of hypotheses, as is a requirement when using the  $\chi^2$  approximation of the null distribution for likelihood ratio tests. Moreover, the Bayesian methods integrate over uncertainty in many of the model parameters that are not of direct interest. However, the interpretation of Bayes factors is not easy, with one solution being to give a verbal description (e.g., instead of a  $P$  value) of the strength of the support of one hypothesis over another (Jeffreys, 1961). Posterior predictive  $P$  values, however, may be useful in providing a way to measure the reliability of a model in a phylogenetic analysis (Bollback, 2002).

Model choice is another area that deserves more study in phylogenetics. Several interesting questions remain. For example, most methods of Bayesian model choice have not yet been implemented for comparing phylogenetic models. Implementation of many of these methods may be challenging, because they could use complicated forms of MCMC. Once the Bayesian methods have been developed, the behavior of these methods, compared with more commonly used methods such as likelihood ratios tests, should be explored.

#### A BAYESIAN FUTURE

Whether Bayesian phylogenetic methods become accepted and widely used will prob-

ably come down to a single consideration: Does the Bayesian method allow the biologist to address new and interesting evolutionary questions? Here, the news is encouraging. Bayesian methods have already been applied to a number of interesting questions, such as estimating ancestral states on a phylogeny (Yang et al., 1995b), searching for positively selected sites (Nielsen and Yang, 1998), inferring a history of cospeciation (Huelsenbeck et al., 2000b), estimating divergence times under a relaxed molecular clock (Thorne et al., 1998), inferring phylogeny using gene order data (Larget et al., 2002), and stochastically reconstructing a character's history (Nielsen and Huelsenbeck, 2002). Bayesian inference also suggests a natural way to accommodate the uncertainty in phylogenies when performing comparative analyses (Losos and Miles, 1994; Huelsenbeck et al., 2000c).

Even the most basic question of phylogenetics—how to estimate the phylogeny of a group of taxa—has benefited from a Bayesian perspective. Topology is a strange parameter, and how to deal with it in a statistical framework has generated discussion (Yang et al., 1995a). Bayesian estimation provides the first different perspective of topology as a parameter; a Bayesian analysis treats the phylogeny as a random variable. Bayesian methods also provide an easy method for assessing the reliability of a phylogenetic tree. All inferences about phylogeny are simply based upon the posterior probability distribution of trees, something that can be approximated in principle using a single run of a Markov chain. Summarizing the posterior probability distribution of trees, however, is an open question. So far, majority rule consensus trees (Larget and Simon, 1999), credible sets (Felsenstein, 1968; Wheeler, 1991), and the maximum posterior probability tree (Rannala and Yang, 1996) have been suggested as ways to summarize the results of a Bayesian analysis of phylogeny. However, there may be other aspects of the posterior distribution that are of interest. For example, perhaps the systematist is only interested in the subtree(s) defined by a subset of the taxa included in the analysis. In this case, it is easy to summarize the results of the subset of species, while allowing all of the species to contribute to the phylogenetic analysis. (Simply prune away the species that are not

of interest after running a Markov chain; the fraction of the time that any subtree is in the pruned list of trees is an approximation of the posterior probability of that subtree.) Another way to assess the support for a clade is to examine the degree to which your beliefs changed after observing a matrix of characters. The Bayes factor measures "the change in the odds in favor of the hypothesis when going from the prior to the posterior" (Lavine and Schervish, 1999). This approach has been used to examine the support for the monophyly of a group of insects (Huelsenbeck, 2001).

This issue of *Systematic Biology* includes several papers that were presented at a symposium on Bayesian inference of phylogeny held at the annual meeting of the Society of Systematic Biologists in Knoxville, Tennessee, during the summer of 2001. Together, they represent some of the potential applications of Bayesian inference. Thorne and Kishino (2002) have extended their method for estimating divergence times of clades to multigene data and devised a method for examining correlation in rates across lineages. Aris-Brosou and Yang (2002) also have examined divergence time estimation, considering several models for relaxing the molecular clock and exploring the sensitivity of divergence time estimates to the model of rate variation used in the analysis. Suchard et al. (2002) have devised a method for detecting recombination using Bayes factors. This approach is especially timely given the increased use of phylogenetic methods on (potentially) recombining viral sequences. Nielsen (2002) has shown how character transformations can be mapped onto phylogenies using an approach that is essentially Bayesian (also see Nielsen and Huelsenbeck, 2002). Nielsen's method is especially important, because it allows characters to be mapped onto a tree under stochastic models of character change. This approach brings a new perspective to the problem of determining a character's history and conjures images of a MacClade-like program (Maddison and Maddison, 2000) for visualizing the history of a character under stochastic models. Previously, only the parsimony method could be used to map a character onto a phylogeny. Now, characters can be mapped onto a tree under the same stochastic models of character evolution used in maximum likelihood, Bayesian, and many distance meth-

ods. Nielsen's approach will force evolutionary biologists to consider the uncertainty in character histories. Miller et al. (2002) have shown a nice application of Bayesian phylogenetic analysis to the question of morning glory monophyly. They paid especially close attention to monitoring convergence of chains and showed some examples where Markov chains have failed to converge. Finally, Rannala (2002) has added a cautionary note to Bayesian analysis using MCMC. The MCMC procedure is so general and easy to apply that it can even be applied to problems where the parameters are not identifiable.

#### ACKNOWLEDGMENTS

We thank R. Nielsen and M. Holder for suggestions on improvements to this manuscript. J.P.H. was supported by NSF grants DEB-0075406 and MCB-0075404, F.R. was supported by Swedish Research Council grant 621-2001-2963, B.L. was supported by NSF grant DEB-0075406, P.S.M. was supported by NSF grant DEB-9707945, and R.E.M. was supported by NSF grant DEB-9707223 awarded to M. Rausher.

#### REFERENCES

- ARIS-BROUSOU, S., AND Z. YANG. 2002. Effects of models of rate evolution on estimation of divergence dates with special reference to the metazoan 18S ribosomal RNA phylogeny. *Syst. Biol.* 51:703-714.
- AUSTIN, D. F., AND Z. HUAMAN. 1996. A synopsis of *Ipomoea* (Convolvulaceae) in the Americas. *Taxon* 45: 3-38.
- BEERLI, P., AND J. FELSENSTEIN. 1999. Maximum likelihood estimation of migration rates and effective population numbers in two populations using a coalescent approach. *Genetics* 152:763-773.
- BERGER, J. O., B. LISEO, AND R. L. WOLPERT. 1999. Integrated likelihood methods for eliminating nuisance parameters. *Stat. Sci.* 14:1-28.
- BOLLBACK, J. P. 2002. Bayesian model adequacy and choice in phylogenetics. *Mol. Biol. Evol.* 19: 1171-1180.
- EFRON, B., E. HALLORAN, AND S. HOLMES. 1996. Bootstrap confidence levels for phylogenetic trees. *Proc. Nat. Acad. Sci. USA* 93:13429-13434.
- FELSENSTEIN, J. 1968. Statistical inference and the estimation of phylogenies. Ph.D. Dissertation, Univ. Chicago, Chicago.
- FELSENSTEIN, J. 1985. Confidence limits on phylogenies: An approach using the bootstrap. *Evolution* 39:783-791.
- GELMAN, A., J. B. CARLIN, H. S. STERN, AND D. B. RUBIN. 1995. Bayesian data analysis. Chapman & Hall, New York.
- GEYER, C. J. 1991. Markov chain Monte Carlo maximum likelihood. Pages 156-163 in *Computing science and statistics: Proceedings of the 23rd symposium on the interface* (E. M. Keramidas, ed.). Interface Foundation, Fairfax Station, Virginia.
- GEYER, C. J. 1992. Practical Markov chain Monte Carlo. *Stat. Sci.* 7:473-511.

- GEYER, C. J. 1999. Likelihood inference for spatial point processes. Pages 79–140 in *Stochastic geometry: Likelihood and computation* (O. E. Barndorff-Nielsen, W. S. Kendall, and M. N. M. van Lieshout, eds.). Chapman and Hall/CRC, London.
- GILKS, W., S. RICHARDSON, AND D. SPIEGELHALTER. 1996. *Markov chain Monte Carlo in practice*. Chapman and Hall, New York.
- GOLDMAN, N. 1990. Maximum likelihood inference of phylogenetic trees with special reference to a Poisson process model of DNA substitution and to parsimony analyses. *Syst. Zool.* 39:345–361.
- GREEN, P. J. 1995. Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika* 82:711–732.
- HASTINGS, W. K. 1970. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* 57:97–109.
- HUELSENBECK, J. P. 1995a. Performance of phylogenetic methods in simulation. *Syst. Biol.* 44:17–48.
- HUELSENBECK, J. P. 1995b. The robustness of two phylogenetic methods: Four taxon simulations reveal a slight superiority of maximum likelihood over neighbor joining. *Mol. Biol. Evol.* 12:843–849.
- HUELSENBECK, J. P. 2001. A Bayesian perspective on the Strepsiptera problem. *Tijdschr. Entomol.* 144:165–178.
- HUELSENBECK, J. P., AND J. P. BOLLBACK. 2001. Empirical and hierarchical Bayesian estimation of ancestral states. *Syst. Biol.* 50:351–366.
- HUELSENBECK, J. P., AND D. M. HILLIS. 1993. Success of phylogenetic methods in the four-taxon case. *Syst. Biol.* 42:247–264.
- HUELSENBECK, J. P., AND N. S. IMENNOV. 2002. Geographic origin of human mitochondrial DNA: Accommodating phylogenetic uncertainty and model comparison. *Syst. Biol.* 51:155–165.
- HUELSENBECK, J. P., B. LARGET, AND D. L. SWOFFORD. 2000a. A compound Poisson process for relaxing the molecular clock. *Genetics* 154:1879–1892.
- HUELSENBECK, J. P., B. RANNALA, AND B. LARGET. 2000b. A Bayesian framework for the analysis of cospeciation. *Evolution* 54:353–364.
- HUELSENBECK, J. P., B. RANNALA, AND J. P. MASLY. 2000c. Accommodating phylogenetic uncertainty in evolutionary studies. *Science* 288:2349–2350.
- HUELSENBECK, J. P., AND F. RONQUIST. 2001. MrBayes: Bayesian inference of phylogeny. *Bioinformatics* 17:754–755.
- HUELSENBECK, J. P., F. RONQUIST, R. NIELSEN, AND J. P. BOLLBACK. 2001. Bayesian inference of phylogeny and its impact on evolutionary biology. *Science* 294:2310–2314.
- JEFFREYS, H. 1961. *Theory of probability*. Oxford Univ. Press, Oxford, U.K.
- KUHNER, M. K., AND J. FELSENSTEIN. 1994. A simulation comparison of phylogeny algorithms under equal and unequal evolutionary rates. *Mol. Biol. Evol.* 11:459–468.
- KUHNER, M., J. YAMATO, AND J. FELSENSTEIN. 1994. Estimating effective population size and mutation rate from sequence data using Metropolis–Hastings sampling. *Genetics* 149:429–434.
- LARGET, B., AND D. SIMON. 1999. Markov chain Monte Carlo algorithms for the Bayesian analysis of phylogenetic trees. *Mol. Biol. Evol.* 16:750–759.
- LARGET, B., D. SIMON, AND J. B. KADANE. 2002. Bayesian phylogenetic inference from animal mitochondrial genome arrangements. *Stat. Soc. B* 64:1–13.
- LAVINE, M., AND M. J. SCHERVISCH. 1999. Bayes factors: What they are and what they are not. *Am. Stat.* 53:119–122.
- LI, S. 1996. *Phylogenetic tree construction using Markov chain Monte Carlo*. Ph.D. Dissertation, Ohio State Univ., Columbus.
- LOSOS, J. B., AND D. B. MILES. 1994. Adaptation, constraint, and the comparative method: Phylogenetic issues and methods. Pages 60–98 in *Ecological morphology: Integrative organismal biology* (P. C. Wainwright and S. Reilly, eds.). Univ. Chicago Press, Chicago.
- MADDISON, D. R., AND W. P. MADDISON. 2000. *MacClade 4: Analysis of phylogeny and character evolution*. Sinauer, Sunderland, Massachusetts.
- MANOS, P. S., R. E. MILLER, AND P. WILKIN. 2001. Phylogenetic analysis of *Ipomoea*, *Argyrea*, *Stictocardia*, and *Turbina* suggests a generalized model of morphological evolution in morning glories. *Syst. Bot.* 26:585–602.
- MAU, B. 1996. *Bayesian phylogenetic inference via Markov chain Monte Carlo methods*. Ph.D. Dissertation, Univ. Wisconsin, Madison.
- MAU, B., AND M. NEWTON. 1997. Phylogenetic inference for binary data on dendrograms using Markov chain Monte Carlo. *J. Comput. Graph. Stat.* 6:122–131.
- MAU, B., M. NEWTON, AND B. LARGET. 1999. Bayesian phylogenetic inference via Markov chain Monte Carlo methods. *Biometrics* 55:1–12.
- MCDONALD, J. A. 1991. Origin and diversity of Mexican Convolvulaceae. *An. Inst. Biol. Univ. Nac. Auton. Mex. Ser. Bot.* 62:65–82.
- METROPOLIS, N., A. W. ROSENBLUTH, M. N. ROSENBLUTH, A. H. TELLER, AND E. TELLER. 1953. Equations of state calculations by fast computing machines. *J. Chem. Phys.* 21:1087–1091.
- MILLER, R. E., T. R. BUCKLEY, AND P. S. MANOS. 2002. An examination of the monophyly of morning glory taxa using Bayesian phylogenetic inference. *Syst. Biol.* 51:740–753.
- MILLER, R. E., M. D. RAUSHER, AND P. S. MANOS. 1999. Phylogenetic systematics of *Ipomoea* (Convolvulaceae) based on ITS and *waxy* sequences. *Syst. Bot.* 24:209–227.
- NIELSEN, R. 2002. Mapping mutations on phylogenies. *Syst. Biol.* 51:729–739.
- NIELSEN, R., AND J. P. HUELSENBECK. 2002. Detecting positively selected amino acid sites using posterior predictive *P*-values. Pages 576–588 in *Pacific Symposium on Biocomputing, proceedings* (R. B. Altman, A. K. Dunker, L. Hunter, K. Lauderdale, and T. E. Klein, eds.). World Scientific, Singapore.
- NIELSEN, R., AND Z. YANG. 1998. Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. *Genetics* 148:929–936.
- NEWTON, M., B. MAU, AND B. LARGET. 1999. Markov chain Monte Carlo for the Bayesian analysis of evolutionary trees from aligned molecular sequences. Pages 143–162 in *Statistics in molecular biology*, Vol. 33 (F. Seillier-Moseiwitch, T. P. Speed, and M. Waterman, eds.). Institute of Mathematical Statistics.
- RANNALA, B. 2002. Identifiability of parameters in MCMC Bayesian inference of phylogeny. *Syst. Biol.* 51:754–760.
- RANNALA, B., AND Z. YANG. 1996. Probability distribution of molecular evolutionary trees: A new method of phylogenetic inference. *J. Mol. Evol.* 43:304–311.

- SANDERSON, M. J., AND M. F. WOJCIECHOWSKI. 2000. Improved bootstrap confidence limits in large-scale phylogenies, with an example from *Neo-Astragalus* (Leguminosae). *Syst. Biol.* 49:671–685.
- SIMON, D., AND B. LARGET. 1998. Bayesian analysis in molecular biology and evolution (BAMBE), version 1.01 beta. Department of Mathematics and Computer Science, Duquesne Univ. Pittsburgh, Pennsylvania.
- SUCHARD, M. A., R. E. WEISS, K. S. DORMAN, AND J. S. SINSHEIMER. 2002. Oh brother, where art thou? A Bayes factor test for recombination with uncertain heritage. *Syst. Biol.* 51:715–728.
- THORNE, J. L., AND H. KISHINO. 2002. Divergence time and evolutionary rate estimation with multilocus data. *Syst. Biol.* 51:689–702.
- THORNE, J., H. KISHINO, AND I. S. PAINTER. 1998. Estimating the rate of evolution of the rate of molecular evolution. *Mol. Biol. Evol.* 15:1647–1657.
- TIERNEY, L. 1994. Markov chains for exploring posterior distributions (with discussion). *Ann. Stat.* 22:1701–1762.
- VAN OOSTSTROOM, S. J. 1953. Convolvulaceae. Pages 458–489 in *Flora Malesiana* (C. G. G. J. van Steenis, ed.). Woltors-Nordhoff, Groningen, The Netherlands.
- VERDCOURT, B. 1963. Convolvulaceae. Pages 1–161 in *Flora of tropical East Africa* (C. E. Hubbard and E. Milne-Redhead, eds.). Whitefriars Press, London.
- WHEELER, W. C. 1991. Congruence among data sets: A Bayesian approach. Pages 334–346 in *Phylogenetic analysis of DNA sequences* (M. M. Miyamoto and J. Cracraft, eds.). Oxford Univ. Press, Oxford, U.K.
- WILKIN, P. 1999. A morphological cladistic analysis of the Ipomoeae (Convolvulaceae). *Kew Bull.* 54:853–876.
- YANG, Z., N. GOLDMAN, AND A. E. FRIDAY. 1995a. Maximum likelihood trees from DNA sequences: A peculiar statistical estimation problem. *Syst. Biol.* 44:384–399.
- YANG, Z., S. KUMAR, AND M. NEI. 1995b. A new method of inference of ancestral nucleotide and amino acid sequences. *Genetics* 141:1641–1650.
- YANG, Z., AND B. RANNALA. 1997. Bayesian phylogenetic inference using DNA sequences: A Markov chain Monte Carlo method. *Mol. Biol. Evol.* 14:717–724.

First submitted 7 May 2002; reviews returned

3 June 2002; final acceptance 9 June 2002

Associate Editor: Chris Simon