Reverend Bayes meets Darwin: Bayesian inference of phylogeny and its impact on evolutionary biology

(Supplemental Material)

John P. Huelsenbeck[1], Fredrik Ronquist[2], Rasmus Nielsen[3] and Jonathan P. Bollback[1]

[1]*Department of Biology, University of Rochester, Rochester, NY 14627, U.S.A.*

[2]*Department of Systematic Zoology, Evolutionary Biology Centre, Uppsala University, Norbyv. 18D, SE-752 36 Uppsala, Sweden*

[3]*Department of Biometrics, Cornell University, Ithaca, NY 14853-1643, U.S.A.*

Bayesian inference[1]

A simple example of Bayesian inference

We will illustrate Bayesian inference using a simple example involving dice. Consider a box with 100 dice, 90 of which are fair and 10 of which are biased. The probability of observing some number of pips after rolling a fair or biased die is given in the following table:

Observation	Fair	Biased
⚀	$\frac{1}{6}$	$\frac{1}{21}$
⚁	$\frac{1}{6}$	$\frac{2}{21}$
⚂	$\frac{1}{6}$	$\frac{3}{21}$
⚃	$\frac{1}{6}$	$\frac{4}{21}$
⚄	$\frac{1}{6}$	$\frac{5}{21}$
⚅	$\frac{1}{6}$	$\frac{6}{21}$

The probability of a high roll is larger for the biased dice than for the fair dice. Suppose that you draw a die at random from the box and roll it twice, observing a four on the first roll and a six on the second roll. What is the probability that the die is biased?

A Bayesian analysis combines ones prior beliefs about the probability of a hypothesis with the likelihood. The likelihood is the vehicle that carries the information about the hypothesis contained in the observations. In this case, the likelihood is simply the probability of observing a four and a six given that the die is biased or fair. Assuming independence of the tosses, the probability of observing a four and a six is

$$\Pr[⚃, ⚅ \mid \text{Fair}] = \frac{1}{6} \times \frac{1}{6} = \frac{1}{36}$$

for a fair die and

$$\Pr[⚃, ⚅ \mid \text{Biased}] = \frac{4}{21} \times \frac{6}{21} = \frac{24}{441}$$

[1]This section of the supplemental material is taken from the MrBayes manual.

for a biased die. The probability of observing the data is 1.96 times greater under the hypothesis that the die is biased. In other words, the ratio of the likelihoods under the two hypotheses suggests that the die is biased.

Bayesian inferences are based upon the posterior probability of a hypothesis. The posterior probability that the die is biased can be obtained using Bayes's *(1)* formula:

$$\Pr[\text{Biased} \mid \boxdot, \boxplus] = \frac{\Pr[\boxdot, \boxplus \mid \text{Biased}] \times \Pr[\text{Biased}]}{\Pr[\boxdot, \boxplus \mid \text{Biased}] \times \Pr[\text{Biased}] + \Pr[\boxdot, \boxplus \mid \text{Fair}] \times \Pr[\text{Fair}]}$$

where Pr[Biased] and Pr[Fair] are the prior probabilities that the die is biased or fair, respectively. As we set up the problem, a reasonable prior probability that the die is biased would be the proportion of the dice in the box that were biased. The posterior probability is then

$$\Pr[\text{Biased} \mid \boxdot, \boxplus] = \frac{\frac{24}{441} \times \frac{1}{10}}{\frac{24}{441} \times \frac{1}{10} + \frac{1}{36} \times \frac{9}{10}} = 0.179$$

This means that our opinion that the die is biased changed from 0.1 to 0.179 after observing the four and six.

Depending upon one's viewpoint, the incorporation of prior beliefs about a parameter is either a strength or a weakness of Bayesian inference. It is a strength in as much as the method explicitly incorporates prior information in inferences about a hypothesis. However, it can often be difficult to specify a prior. For the dice example, it is easy to specify the prior as we provided information on the number of fair and biased dice in the box and also specify that a die was randomly selected. However, if we were to simply state that the die is either fair or biased, but did not specify a physical description of how the die was chosen, it would have been much more difficult to specify a prior specifying the probability that the die is biased. For example, one could have taken the two hypotheses to have been *a priori* equally probable or given much more weight to the hypothesis that had the die fair as severely biased dice are rarely encountered (or manufactured) in the real world.

Bayesian inference of phylogeny

Bayesian inference of phylogeny is based upon the posterior probability of a phylogenetic tree, τ. The posterior probability of the ith phylogenetic tree, τ_i, conditioned on the observed matrix of aligned DNA sequences ($\boldsymbol{X}$) is obtained using Bayes's formula:

$$f(\tau_i|\boldsymbol{X}) = \frac{f(\boldsymbol{X}|\tau_i)f(\tau_i)}{\sum_{j=1}^{B(s)} f(\boldsymbol{X}|\tau_j)f(\tau_j)}$$

[throughout, we denote conditional probabilities as $f(\cdot|\cdot)$]. Here, $f(\tau_i|\boldsymbol{X})$ is the posterior probability of the ith phylogeny and can be interpreted as the probability that τ_i is the correct tree given the DNA sequence data. The likelihood of the ith tree is $f(\boldsymbol{X}|\tau_i)$ and the prior probability of the ith tree is $f(\tau_i)$. The summation in the denominator is over all $B(s)$ trees that are possible for s species. This number is $B(s) = \frac{(2s-3)!}{2^{s-2}(s-2)!}$ for rooted trees, $B(s) = \frac{(2s-5)!}{2^{s-3}(s-3)!}$ for unrooted trees, and $B(s) = \frac{s!(s-1)!}{2^{s-1}}$ for labelled histories. Typically, an uninformative prior is used for trees, such that $f(\tau_i) = \frac{1}{B(s)}$

DNA sequence data.—We consider an aligned matrix of s DNA sequences:

$$\boldsymbol{X} = \{x_{ij}\} = \begin{array}{l} \text{Species 1} \\ \text{Species 2} \\ \text{Species 3} \\ \vdots \\ \text{Species } s \end{array} \left\{ \begin{array}{ccccc} A & A & C & C & T \\ A & A & C & G & G \\ A & C & C & C & T \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ A & C & C & C & T \end{array} \right\}$$

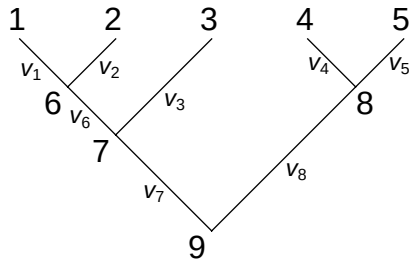


Figure 1.—An example of a phylogenetic tree for $s = 5$ species. The branch lengths are denoted v_i.

The data matrix consists of the sequences for s species for $c = 5$ sites from a gene (c is the length of the aligned DNA sequences). The observations at the first site are $\boldsymbol{x}_1 = \{A, A, A, \ldots, A\}'$. In general, the information at the ith site in the matrix is denoted $\boldsymbol{x}_i$.

Phylogenetic models.—What is the probability of observing the data at the ith site? To calculate this probability, we assume a phylogenetic model. A phylogenetic model consists of a tree (τ_i) with branch lengths specified on the tree ($\boldsymbol{v}_i$) and a stochastic model of DNA substitution. Figure 1 shows an example of a phylogenetic tree of $s = 5$ species. The tips of the tree are labeled $1, 2, \ldots, s$ and the internal nodes of the tree are labeled $s + 1, s + 2, \ldots, 2s - 1$; the root of the tree is always labeled $2s - 1$. The lengths of the branches are denoted v_i and are in terms of the number of substitutions expected to occur along the ith branch. In general, the ancestor of node k will be denoted $\sigma(k)$; the ancestor of node 4 is $\sigma(4) = 8$. The ancestor of the root is $\sigma(2s - 1) = \emptyset$.

The second part of the phylogenetic model consists of a stochastic model of DNA substitution. Here, the typical assumption is that DNA substitution is a continuous-time Markov process. The heart of the model is a matrix specifying the instantaneous rate of substitution from one nucleotide state to another:

$$\boldsymbol{Q} = \{q_{ij}\} = \begin{pmatrix} \cdot & \pi_C r_{AC} & \pi_G r_{AG} & \pi_T r_{AT} \\ \pi_A r_{AC} & \cdot & \pi_G r_{CG} & \pi_T r_{CT} \\ \pi_A r_{AG} & \pi_C r_{CG} & \cdot & \pi_T r_{GT} \\ \pi_A r_{AT} & \pi_C r_{CT} & \pi_G r_{CT} & \cdot \end{pmatrix}$$

where the matrix specifies the rate of change from nucleotide i (row) to nucleotide j (column). The nucleotides are in the order A, C, G, T. The diagonals of the matrix are specified such that the rows each sum to 0. The equilibrium (or stationary) frequencies of the four nucleotides are denoted π_i ($\boldsymbol{\pi} = \{\pi_A, \pi_C, \pi_G, \pi_T\}$). This matrix specifies the most general time-reversible model of DNA substitution and is referred to as the GTR model *(2)*. Because the rate of substitution and time are confounded, the $\boldsymbol{Q}$ matrix is rescaled such that $-\sum \pi_i q_{ii} = 1$ for all i (making the average rate of substitution 1). Over a branch of length v the transition probabilities are calculated as $\boldsymbol{P}(v, \boldsymbol{\theta}) = \{p_{ij}(v, \boldsymbol{\theta})\} = e^{\boldsymbol{Q}v}$. The parameters of the substitution model are contained in a vector $\boldsymbol{\theta}$.

The likelihood of a phylogeny.—The phylogenetic model consists of a tree (τ_i) with branch lengths ($\boldsymbol{v}_i$) and a stochastic model of DNA substitution that is specified by a matrix of instantaneous rates. The probability of observing the data at the ith site in the aligned matrix is a sum over all possible assignments of nucleotides to the internal nodes of the tree:

$$f(\boldsymbol{x}_i|\tau_j, \boldsymbol{v}_j, \boldsymbol{\theta}) = \sum_{\boldsymbol{y}} \left[\pi_{y_{2s-1}} \left(\prod_{k=1}^{s} p_{y_{i\sigma(k)}, x_{ik}}(v_k, \boldsymbol{\theta}) \right) \left(\prod_{k=s+1}^{2s-2} p_{y_{i\sigma(k)}, y_{ik}}(v_k, \boldsymbol{\theta}) \right) \right]$$

Here, y_{ij} is the (unobserved) nucleotide at the jth node for the ith site. The summation is over all 4^{s-1}

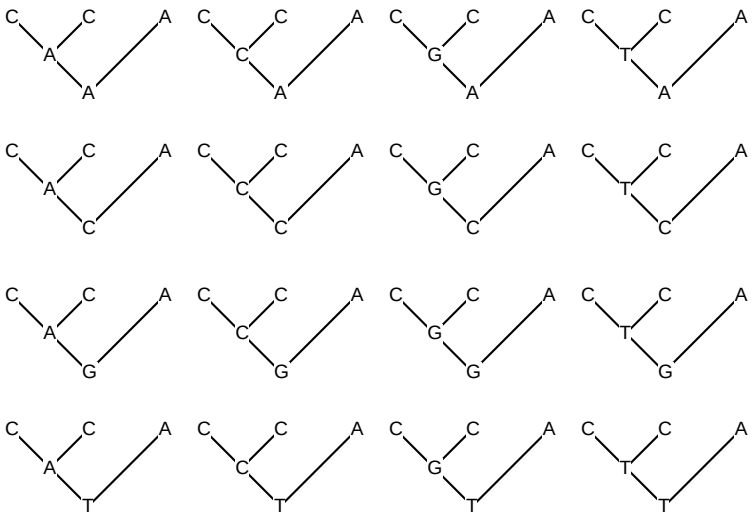


Figure 2.—The 16 possible assignments of nucleotides to the internal nodes of a tree of $s = 3$ species. The observations at the site are $\boldsymbol{x}_i = \{C, C, A\}$ and the unobserved nucleotides at the internal nodes of the tree are denoted $\boldsymbol{y}$.

ways that nucleotides can be assigned to the internal nodes of the tree. Figure 2 illustrates the possible nucleotide assignments for a simple tree of $s = 3$ species. Felsenstein *(3)* introduced a pruning algorithm that efficiently calculates the summation. Often, the rate at the site is assumed to be drawn from a gamma distribution. This allows one to relax the assumption that the rate of substitution is equal across all sites. If gamma-distributed rate variation is assumed, then the probability of observing the data at the ith site becomes:

$$f(\boldsymbol{x}_i|\tau_j, \boldsymbol{v}_j, \boldsymbol{\theta}, \alpha) = \int_0^\infty \left\{ \sum_{\boldsymbol{y}} \left[\pi_{y_{2s-1}} \left(\prod_{k=1}^{s} p_{y_{i\sigma(k)}, x_{ik}}(v_k r, \boldsymbol{\theta}) \right) \left(\prod_{k=s+1}^{2s-2} p_{y_{i\sigma(k)}, y_{ik}}(v_k r, \boldsymbol{\theta}) \right) \right] \right\} f(r|\alpha)\, dr$$

where $f(r|\alpha)$ is the density of the rate r under the gamma model *(4)*. The parameter α is the shape parameter of the gamma distribution (here, the shape and the scale parameters of the gamma distribution are both set to α). Typically, this integral is impossible to evaluate. Hence, an approximation first suggested by Yang *(5)* is used in which the continuous gamma distribution is broken into K categories, each with equal weight. The mean rate from each category represents the rate for the entire category. The probability of observing the data at the ith site then becomes:

$$f(\boldsymbol{x}_i|\tau_j, \boldsymbol{v}_j, \boldsymbol{\theta}, \alpha) = \sum_{n=1}^{K} \left\{ \sum_{\boldsymbol{y}} \left[\pi_{y_{2s-1}} \left(\prod_{k=1}^{s} p_{y_{i\sigma(k)}, x_{ik}}(v_k r_n, \boldsymbol{\theta}) \right) \left(\prod_{k=s+1}^{2s-2} p_{y_{i\sigma(k)}, y_{ik}}(v_k r_n, \boldsymbol{\theta}) \right) \right] \right\} \frac{1}{K}$$

Assuming independence of the substitutions across sites, the probability of observing the aligned matrix of DNA sequences is

$$f(\boldsymbol{X}|\tau_j, \boldsymbol{v}_j, \boldsymbol{\theta}, \alpha) = \prod_{i=1}^{c} f(\boldsymbol{x}_i|\tau_j, \boldsymbol{v}_j, \boldsymbol{\theta}, \alpha)$$

Importantly, the likelihood can be calculated under a number of different models of character change. For example, the codon model describes the substitution process over triplets of sites (a codon) and allows the estimation of the nonsynonymous/synonymous rate ratio *(6, 7)*. Similarly, models of DNA substitution have been described that allow nonindenpendent substitutions to occur in stem regions of rRNA genes *(8)*. Finally,

one can calculate likelihoods for amino acid *(9)*, restriction site *(10)*, and, more recently, morphological data *(11)*.

Bayesian inference of phylogeny.—As described so far, the likelihood depends upon several unknown parameters; generally, the phylogeny, branch lengths, and substitution parameters are unknown. The method of maximum likelihood estimates these parameters by finding the values of the parameters which maximize the likelihood function. Currently, programs such as PAUP* *(12)*, PAML *(13)*, and PHYLIP *(14)* estimate phylogeny using the method of maximum likelihood.

Bayesian inference is based instead upon the posterior probability of the parameter. As described above, the posterior probability of the ith tree is

$$f(\tau_i|\boldsymbol{X}) = \frac{f(\boldsymbol{X}|\tau_i)f(\tau_i)}{\sum_{j=1}^{B(s)} f(\boldsymbol{X}|\tau_j)f(\tau_j)}$$

where the likelihood function is integrated over all possible values for the branch lengths and substitution parameters:

$$f(\boldsymbol{X}|\tau_i) = \int_{v_i}\int_{\theta}\int_{\alpha} f(\boldsymbol{X}|\tau_i, \boldsymbol{v}_i, \boldsymbol{\theta}, \alpha) f(\boldsymbol{v}_i) f(\boldsymbol{\theta}) f(\alpha)\, d\boldsymbol{v}_i\, d\boldsymbol{\theta}\, d\alpha$$

Markov chain Monte Carlo.—Typically, the posterior probability cannot be calculated analytically. However, the posterior probability of phylogenies can be approximated by sampling trees from the posterior probability distribution. Markov chain Monte Carlo (MCMC) can be used to sample phylogenies according to their posterior probabilities. The Metropolis-Hastings (MH) algorithm *(15, 16, 17)* is an MCMC algorithm that has been used successfully to approximate the posterior probabilities of trees *(18, 19)*.

The MH algorithm works as follows. Let $\boldsymbol{\Psi} = \{\tau, \boldsymbol{v}, \boldsymbol{\theta}, \alpha\}$ be a specific tree, combination of branch lengths, substitution parameters, and gamma shape parameter. The MH algorithm constructs a Markov chain that has as its stationary frequency the posterior probability of interest (in this case, the joint posterior probability of τ, $\boldsymbol{v}$, $\boldsymbol{\theta}$, and α). The current state of the chain is denoted $\boldsymbol{\Psi}$. If this is the first generation of the chain, then the chain is initialized (perhaps by randomly picking a state from the prior). A new state is then proposed, $\boldsymbol{\Psi}'$. The probability of proposing the new state given the old state is $f(\boldsymbol{\Psi}'|\boldsymbol{\Psi})$ and the probability of making the reverse move (which is never actually made) is $f(\boldsymbol{\Psi}|\boldsymbol{\Psi}')$. The new state is accepted with probability

$$\begin{aligned}
R &= \min\left(1, \frac{f(\boldsymbol{\Psi}'|\boldsymbol{X})}{f(\boldsymbol{\Psi}|\boldsymbol{X})} \times \frac{f(\boldsymbol{\Psi}|\boldsymbol{\Psi}')}{f(\boldsymbol{\Psi}'|\boldsymbol{\Psi})}\right) \\
&= \min\left(1, \frac{f(\boldsymbol{X}|\boldsymbol{\Psi}')f(\boldsymbol{\Psi}')/f(\boldsymbol{X})}{f(\boldsymbol{X}|\boldsymbol{\Psi})f(\boldsymbol{\Psi})/f(\boldsymbol{X})} \times \frac{f(\boldsymbol{\Psi}|\boldsymbol{\Psi}')}{f(\boldsymbol{\Psi}'|\boldsymbol{\Psi})}\right) \\
&= \min\left(1, \underbrace{\frac{f(\boldsymbol{X}|\boldsymbol{\Psi}')}{f(\boldsymbol{X}|\boldsymbol{\Psi})}}_{\text{Likelihood Ratio}} \times \underbrace{\frac{f(\boldsymbol{\Psi}')}{f(\boldsymbol{\Psi})}}_{\text{Prior Ratio}} \times \underbrace{\frac{f(\boldsymbol{\Psi}|\boldsymbol{\Psi}')}{f(\boldsymbol{\Psi}'|\boldsymbol{\Psi})}}_{\text{Proposal Ratio}}\right)
\end{aligned}$$

A uniform random variable between 0 and 1 is drawn. If this number is less than R, then the proposed state is accepted and $\boldsymbol{\Psi} = \boldsymbol{\Psi}'$. Otherwise, the chain remains in the original state. This process of proposing a new state, calculating the acceptance probability, and either accepting or rejecting the proposed move is repeated many thousands of times. The sequence of states visited forms a Markov chain. This chain is sampled (either every step, or the chain is "thinned" and samples are taken every so often). The samples from the Markov chain form a valid, albeit dependent, sample from the posterior probability distribution *(20)*. As described here, the Markov chain samples from the joint probability density of trees, branch lengths, and substitution parameters. The marginal probability of trees can be calculated by simply printing to a file the trees that

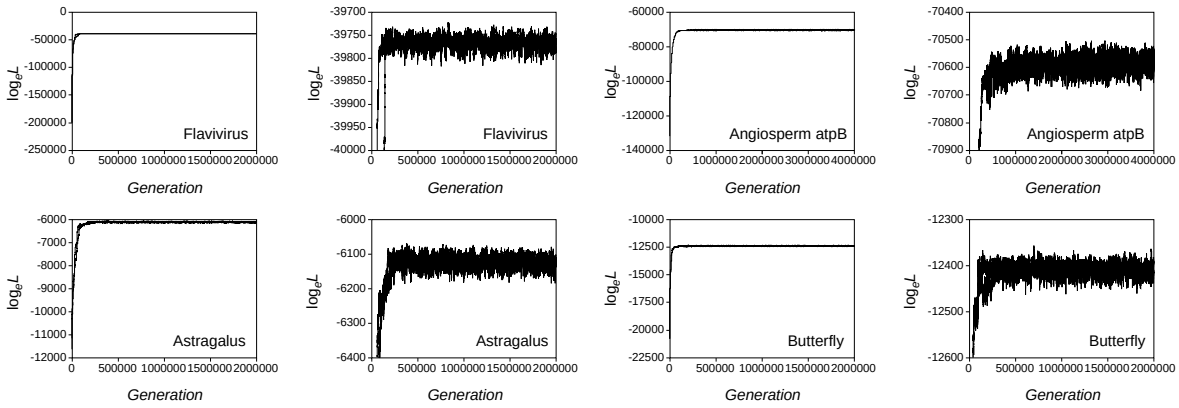


Figure 3.—Plots of the log probability of observing the data ($\log_e L$) through time for each of the four data sets.

are visited during the course of the MCMC analysis. The proportion of the time any single tree is found in this sample is an approximation of the posterior probability of the tree.

Programs for Bayesian inference of phylogeny.—There are a few programs for the Bayesian analysis of phylogenetic trees. BAMBE *(22)* and MrBayes (*23*) approximate the posterior probability of phylogenies using MCMC (specifically, the MH algorithm). BAMBE assumes uniform priors on phylogenies and branch lengths. The program uses an improved method for calculating likelihoods that is very fast. MrBayes implements a larger number of substitution models and priors for parameters. Another program, MCMCTREE in the PAML package of programs *(13)* calculates posterior probabilities of trees using a combination of Monte Carlo and MCMC integration. The program works for up to $s = 11$ species. Besides the algorithm for approximating posterior probabilities, the program differs from BAMBE in assuming a birth-death process prior on phylogenies. This prior places equal weight on labelled histories (where a labelled history differs from a rooted tree in considering the relative speciation times).

Inferring large trees

Evidence for convergence in topology

In all of our analyses we ran at least two Markov chains, each of which started from a random tree. Each chain consists of a total of $n = 4$ chains, all except one of which are heated. Each of these chains was also started from a randomly chosen tree. We examined a number of diagnostics to determine if the chains had reached stationarity. For one, we examined the correlation between the posterior probabilities of individual clades found in both chains. These results are depicted in the full paper and show that the inferences that would be drawn from different chains are largely the same. Another way of examinining convergence is to look at the behavior of one or more parameters through time. For example, we examined the log probability of observing the data through time for the chains. These results are shown in Figure 3. The chains start with trees that poorly explain the data but rapidly find better trees, eventually sampling trees which fluctuate around a specific log probability.

We also examined the topological variance between and within chains to check for convergence. Before convergence on the target distribution is reached, the trees from separate chains will tend to be more different than trees from a single chain. When the chains have reached the target distribution, the mean topological distance will be the same within and among chains. Thus, convergence can be monitored by comparing the

mean topological distance within chains (W) and between chains (B). Initially, the ratio B/W is likely to be considerably larger than 1. When the ratio is close to 1.0, it is likely that the chains have converged to the target distribution. Robinson and Foulds measure of topological distance was used (*24*); it is equivalent to twice the number of unique taxon bipartitions found in one of the two trees but not in both. The mean topological distance within each chain was estimated by drawing 1000 or 5000 random pairs of trees from the chain. The mean topological distance within chains is simply the mean of the means for each chain. The mean topological distance between chains was estimated by drawing 1000 or 5000 random pairs of trees from different chains. All chains were started from random trees. This means that the topological distance between chains was initially close to the maximum value, equivalent to twice the number of taxa in the analysis minus 6. Thus, the initial topological distance between chains was close to 206 for the butterfly analysis, 240 for the flavivirus analysis, 274 for the Astragalus analysis, and 708 for the angiosperm analysis.

Flavivirus		Mean topological distance				
Trees sampled	Draws	Chain 1	Chain 2	Within (W)	Between (B)	Ratio B/W
625-1250	1000	71.24	43.10	57.17	98.33	1.720
1250-2500	1000	48.15	31.62	39.89	44.65	1.119
2500-5000	1000	34.34	33.95	34.14	35.00	1.025
5000-10000	1000	36.41	42.01	39.21	40.74	1.039
10000-20000	1000	37.61	44.15	40.88	41.89	1.025
5000-20000	1000	38.15	44.45	41.30	41.79	1.011

Butterfly		Mean topological distance				
Trees sampled	Draws	Chain 1	Chain 2	Within (W)	Between (B)	Ratio B/W
625-1250	1000	87.10	68.20	77.65	100.99	1.300
1250-2500	1000	85.36	63.96	74.66	85.05	1.139
2500-5000	1000	74.27	68.71	71.49	75.19	1.052
5000-10000	1000	71.05	64.44	67.75	68.63	1.013
10000-20000	1000	74.19	69.38	71.78	72.95	1.016
5000-20000	1000	73.90	67.84	70.87	71.67	1.011

Astragalus		Mean topological distance				
Trees sampled	Draws	Chain 1	Chain 2	Within (W)	Between (B)	Ratio B/W
625-1250	1000	124.51	152.62	138.56	150.55	1.087
1250-2500	1000	121.26	113.45	117.35	122.38	1.043
2500-5000	1000	115.64	114.69	115.16	117.46	1.020
5000-10000	1000	116.02	118.80	117.41	118.53	1.009
10000-20000	1000	117.42	117.55	117.49	117.62	1.001
5000-20000	5000	117.41	118.69	118.05	118.27	1.002

Angiosperm atpB		Mean topological distance				
Trees sampled	Draws	Chain 1	Chain 2	Within (W)	Between (B)	Ratio B/W
1250-2500	1000	178.16	273.99	226.08	348.42	1.541
2500-5000	1000	119.18	212.42	165.80	223.42	1.348
5000-10000	1000	123.27	161.71	142.49	156.23	1.096
10000-20000	1000	126.21	134.63	130.42	135.22	1.037
20000-40000	1000	123.20	123.67	123.43	126.13	1.022
10000-40000	1000	125.98	128.75	127.36	129.52	1.017

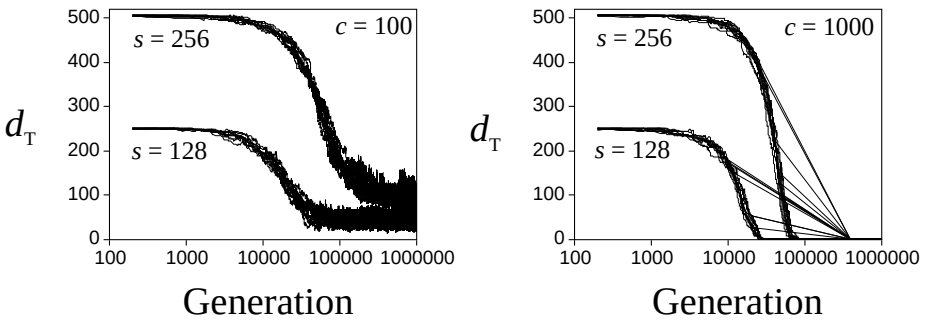


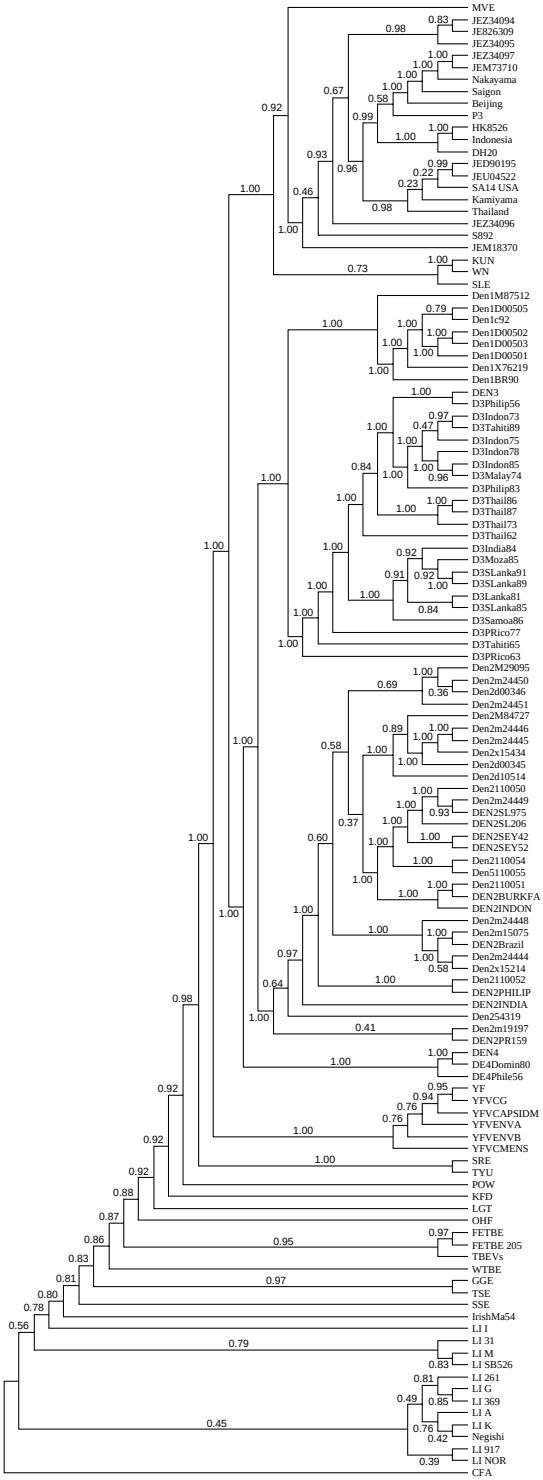
Figure 4.—MCMC converges to the correct tree in simulation. DNA sequences were simulated under the Jukes-Cantor model on symmetrically branching trees of $s = 128$ *or* $s = 256$ *species. All of the branches of the trees were 0.1 expected substitution per site in length. The graphs show the topological distance,* d_T *(1), from the true tree through time for chains that started from randomly chosen trees. The topological distance either plateaus near the true tree for the short sequences (*$c = 100$*) or to the true tree for the long sequences (*$c = 1000$*). Bayesian analysis assumed the correct model of DNA substitution.*

Finally, we examined the ability of the MCMC method to converge to the correct tree(s) by performing simulations of 128 and 256 sequences. Sequences were simulated under the Jukes-Cantor model of DNA substitution. In all simulation replicates, the chains converged to trees that were either identical to or very close to the true tree (Figure 4).

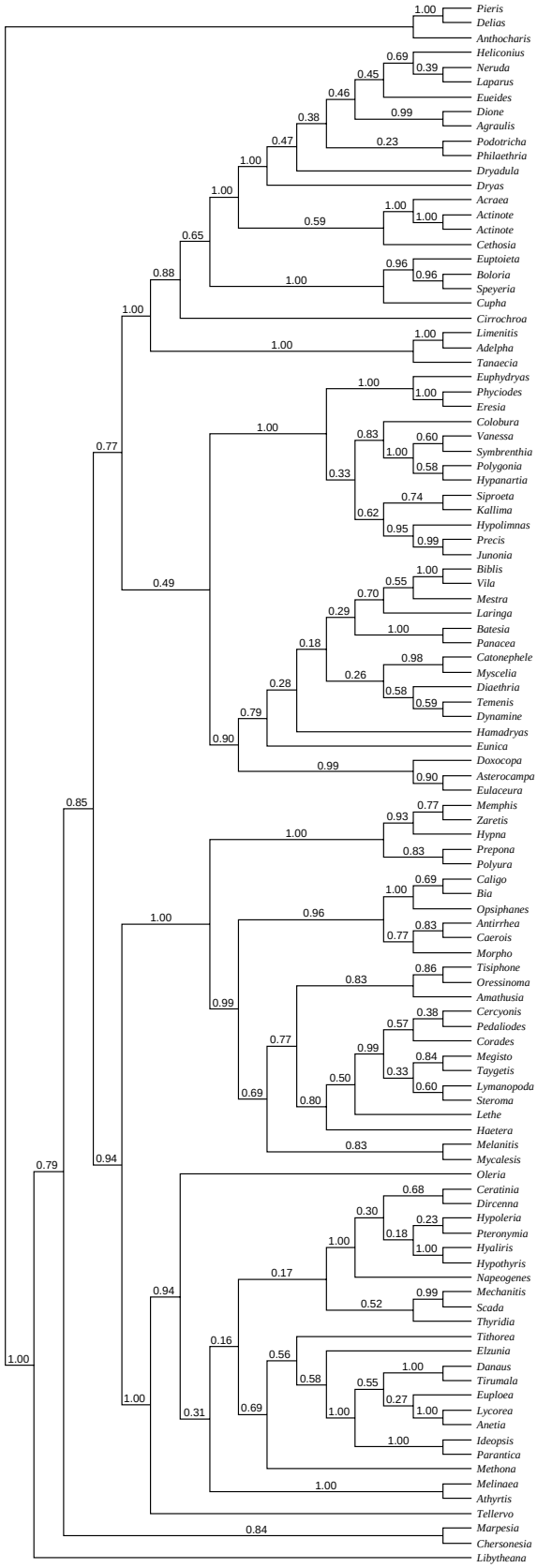
Phylogenetic trees

For each data set, we constructed a 50% majority rule consensus tree of the trees sampled by the MCMC procedure. The numbers at the interior nodes indicate the posterior probability that the clade is correct under the model. Polytomious nodes indicate that various resolutions of the polytomy have posterior probabilities less than 0.5.

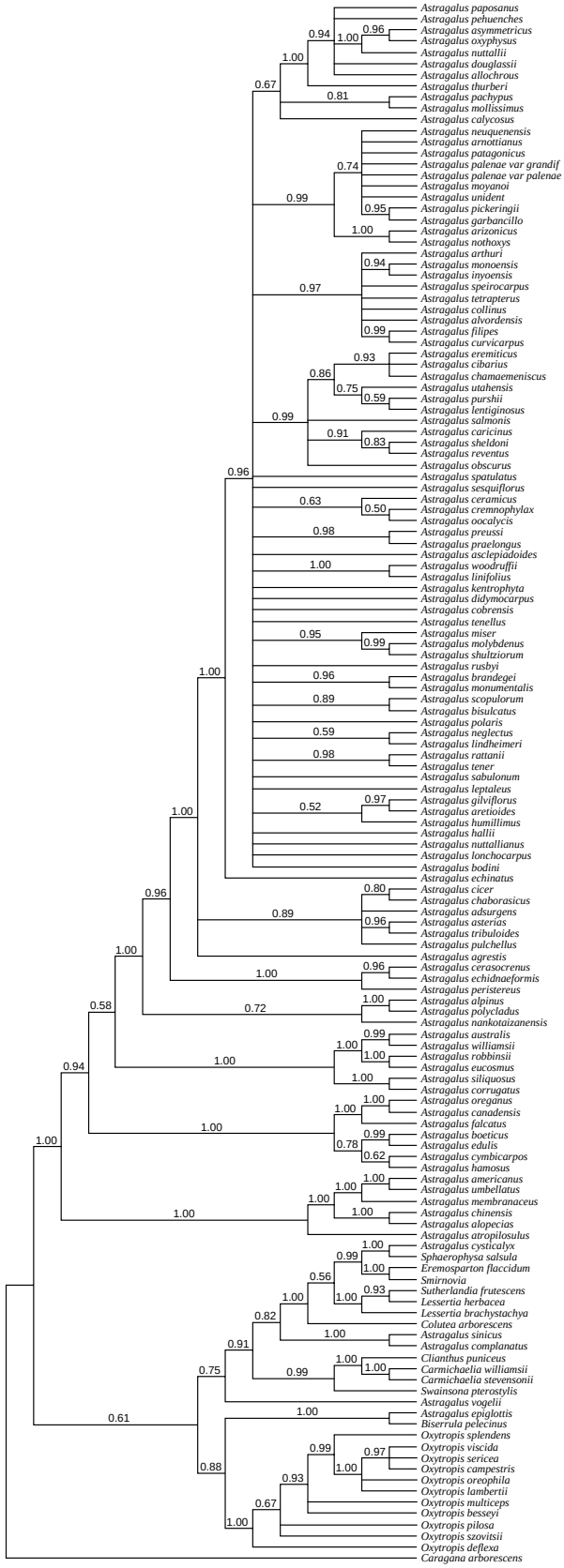
Flavivirus (*25*)



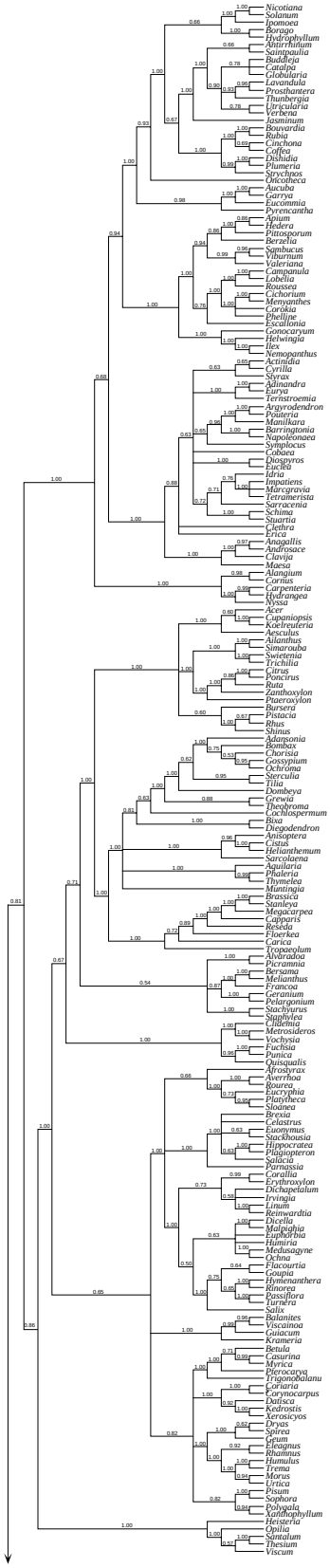
Butterfly (*26*)



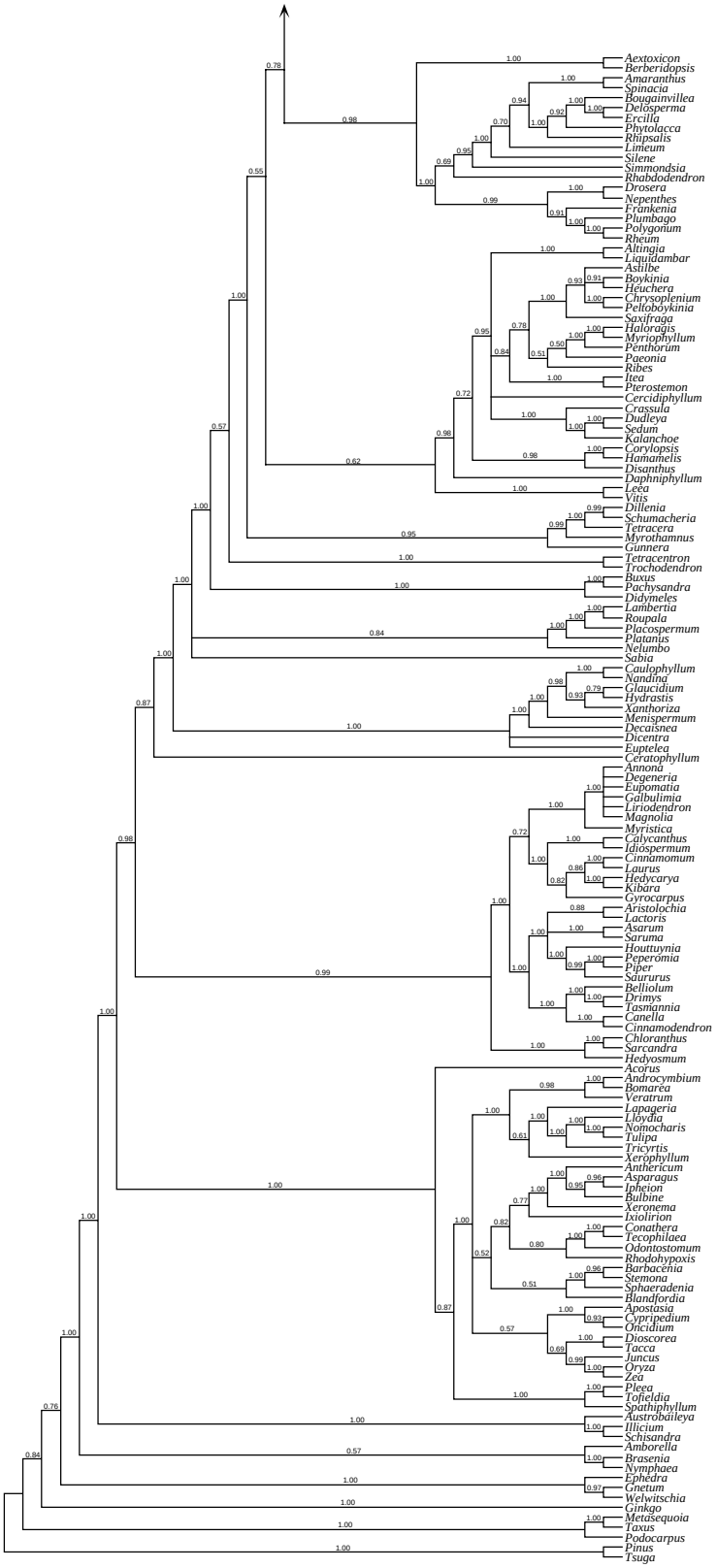
Astragalus (*27*)



Angiosperms (Part 1) (*28*)



Angiosperms (Part 2)



Substitution-model parameter estimates

Estimates of the parameters of the substitution model (θ). The instantaneous rate of change from nucleotide i to nucleotide j is denoted r_{ij}, and is measured relative to the rate of change between G and T ($r_{GT} = 1$). The frequency of nucleotide i is denoted π_i, the gamma shape parameter is α, and the rate multiplier for codon position i is m_i. The numbers in each column give the mean of the marginal posterior probability distribution and the 95% credible interval (in parantheses) for the parameter.

Parameter (θ)	Astragalus	Butterfly	Flavivirus	Angiosperms
r_{AC}	1.22 (0.76, 1.83)	1.14 (0.90, 1.44)	2.95 (2.50, 3.48)	1.61 (1.47, 1.74)
r_{AG}	2.77 (1.90, 3.89)	4.78 (3.91, 5.85)	6.18 (5.33, 7.22)	3.82 (3.54, 4.10)
r_{AT}	1.79 (1.20, 2.68)	1.06 (0.82, 1.36)	1.74 (1.45, 2.11)	0.27 (0.24, 0.31)
r_{CG}	0.83 (0.53, 1.24)	0.75 (0.58, 0.98)	1.28 (1.04, 1.58)	1.56 (1.41, 1.73)
r_{CT}	5.10 (3.60, 7.11)	5.38 (4.51, 6.49)	11.25 (9.76, 13.15)	4.99 (4.63, 5.38)
π_A	0.23 (0.20, 0.26)	0.25 (0.23, 0.27)	0.25 (0.24, 0.26)	0.34 (0.33, 0.35)
π_C	0.25 (0.23, 0.28)	0.27 (0.25, 0.29)	0.24 (0.23, 0.25)	0.15 (0.14, 0.15)
π_G	0.26 (0.23, 0.29)	0.23 (0.21, 0.25)	0.27 (0.26, 0.28)	0.18 (0.17, 0.19)
π_T	0.26 (0.23, 0.29)	0.26 (0.24, 0.28)	0.24 (0.23, 0.25)	0.33 (0.32, 0.34)
α	0.49 (0.27, 0.60)	—	—	—
m_1	—	0.31 (0.28, 0.35)	0.43 (0.41, 0.45)	0.35 (0.33, 0.36)
m_2	—	0.25 (0.24, 0.26)	0.23 (0.21, 0.24)	0.24 (0.23, 0.26)
m_3	—	2.45 (2.40, 2.53)	2.35 (2.32, 2.38)	2.39 (2.35, 2.42)

References and Notes

1. T. Bayes, *Phil. Trans. Roy. Soc.*, 330 (1763).

2. S. Tavaré, *Lectures in Mathematics in the Life Sciences*, **17**, 57 (1986).

3. J. Felsenstein, J, *J. Mol. Evol.*, **17**, 368 (1981).

4. Z. Yang, *Mol. Biol. Evol.* **10**, 1396 (1993).

5. Z. Yang, *J. Mol. Evol.* **39**, 306 (1994).

6. N. Goldman and Z. Yang, *Mol. Biol. Evol.* **11**, 725 (1994).

7. S. Muse and B. Gaut, *Mol. Biol. Evol.* **11**, 715 (1994).

8. M. Schöniger and A. von Haeseler, *Mol. Phyl. Evol.* **3**, 240 (1994).

9. J. Adachi and M. Hasegawa, MOLPHY: Programs for molecular Phylogenetics I—PROTML: Maximum likelihood inference of protein phylogeny. *Comp. Sci. Mono.* **27** (1992).

10. P. E. Smouse and W.-H. Li, *Evol.* **41**, 1162 (1987).

11. P. O. Lewis, Talk at the Annual meeting of the SSE, SSB, and ASN societies held in Madison, Wisconsin (1999).

12. D. L. Swofford, *PAUP*: Phylogenetic Analysis Using Parsimony and Other Methods.* Sinauer Associates, Sunderland, MA (1998).

13. Z. Yang, *CABIOS* **15**, 555 (1997).

14. J. Felsenstein, PHYLIP (Phylogeny Inference Package, version 3.5c, Dept. Genet., Univ. Wash.

15. N. Metropolis, A. W. Rosenbluth, A. H. Teller, E. Teller, *J. Chem. Phys.* **21**, 1087 (1953).

16. W. Hastings, *Biometrika* **57**, 97 (1970).

17. P. J. Green, *Biometrika* **82**, *711* (1995).

18. Z. Yang and B. Rannala, *Mol. Biol. Evol.* **14**, 717 (1997).

19. B. Larget and D. Simon, *Mol. Biol. Evol.* **16**, 750 (1999).

20. L. Tierney, *Annuals of Statistics* **22**, 1701 (1994).

21. M. H. Hasegawa, H. Kishino, T. Yano *J. Mol. Evol.* **22**, 160 (1985).

22. D. Simon and B. Larget, *Bayesian Analysis in Molecular Biology, BAMBE* (Dept. Math. Comp. Sci., Duquesne, Univ, 1999).

23. J. P. Huelsenbeck and F. Ronquist, *Bioinformatics* **17**, 754 (2001).

24. D. F. Robinson, L. R. Foulds, *Lectures notes in mathematics in the Life Sciences* **748**, 119 (1979).

25. P. M. Zanotto, *Proc. Natl. Acad. Sci. USA* **93**, 548 (1996).

26. A. Brower, *Proc. R. Soc. Lond. B* **267**, 1201 (2000).

27. M. J. Sanderson, M. F. Wojciechowski, *Syst. Biol.* **49**, 671–685 (2000).

28. V. Savolainen *et al.*, *Syst. Biol.* **49**, 306–362 (2000).