Syst. Biol. 52(2):131–158, 2003
DOI: 10.1080/10635150390192780

Stochastic Mapping of Morphological Characters

JOHN P. HUELSENBECK,[1] RASMUS NIELSEN,[2] AND JONATHAN P. BOLLBACK[1]

[1]*Section of Ecology, Behavior and Evolution, Division of Biology, University of California–San Diego, La Jolla, California 92093-0116, USA*
[2]*Department of Biometrics, Cornell University, 439 Warren Hall, Ithaca, New York 14853-7801, USA*

Abstract.— Many questions in evolutionary biology are best addressed by comparing traits in different species. Often such studies involve mapping characters on phylogenetic trees. Mapping characters on trees allows the nature, number, and timing of the transformations to be identified. The parsimony method is the only method available for mapping morphological characters on phylogenies. Although the parsimony method often makes reasonable reconstructions of the history of a character, it has a number of limitations. These limitations include the inability to consider more than a single change along a branch on a tree and the uncoupling of evolutionary time from amount of character change. We extended a method described by Nielsen (2002, Syst. Biol. 51:729–739) to the mapping of morphological characters under continuous-time Markov models and demonstrate here the utility of the method for mapping characters on trees and for identifying character correlation. [Bayesian estimation; character correlation; character mapping; Markov chain Monte Carlo.]

The footprint of natural selection on organisms can often be detected using phylogenetic methods. Correlation in either molecular or morphological characters is taken as evidence of natural selection acting on those characters (Harvey and Pagel, 1991). The correlation might be between a character and the environment, with the repeated evolution of the character in a particular environment indicating that the trait confers an advantage, or the correlation may be between one character and another. In ribosomal RNA sequences, for example, correlated changes occur in nucleotides paired in the stem structures; natural selection is acting to maintain Watson–Crick pairing of nucleotides in the functionally important stem structures. In either case—correlation between different characters or the repeated evolution of a character in a particular environment—phylogenetic methods provide the best framework for the analysis of correlation because they allow the effects of a common phylogenetic history that simultaneously acts on all of the characters to be partitioned from the evolutionary processes generating the character patterns (Felsenstein, 1985).

Despite the importance of phylogenetic analysis of character change in evolutionary biology, detection of correlation in characters is fraught with difficulties. One dilemma involves how characters should be mapped onto a phylogenetic tree. Many methods for detecting correlations rely on mapping character changes on a phylogenetic tree using the parsimony method (Ridley, 1983; Maddison, 1990). The parsimony method provides the minimum number of transformations required to explain the evolution of the character on the tree and therefore necessarily underestimates the total number of changes. Furthermore, some methods treat the parsimony mapping of a character as an observation in further statistical analyses (Ridley, 1983; Maddison, 1990). Although the parsimony method is expected to provide a reasonable mapping of a character when the rates of evolution are low, the fundamental problem with the method is that it does not account for the uncertainty in the process of character change. In effect, the parsimony method wagers all on the mapping requiring the fewest changes, when in reality many other perhaps slightly less parsimonious mappings may be nearly as good or better.

This problem with the parsimony method has long been realized (see Harvey and Pagel, 1991), and a number of methods have been proposed that avoid mapping characters on trees using parsimony. One approach that avoids some of the pitfalls of the parsimony method is to model character change as a continuous-time Markov chain; when modeling character change as a stochastic process, all possible character histories are considered, with each weighted by its probability of occurrence under the model. Typically, specific character mappings are no longer considered; parameters of the Markov process are examined instead. For example, different character transformations can have different rates, and hypotheses of biased character change can then be tested (see Pagel, 1999a). Similarly, the evolution of two characters can be modeled simultaneously, and a parameter that is related to the correlation of the characters can be estimated (Pagel, 1994).

A related method for dealing with the uncertainty of a character's history assigns probabilities to specific ancestral states on a tree (Schluter, 1995; Schluter et al., 1997; Mooers and Schluter, 1999; Pagel, 1999b). Often, methods for estimating the ancestral states on a tree are used with the parsimony criterion to map characters on a tree (Pagel, 1999a). Specifically, these methods implicitly use the parsimony criterion to assign a history along the individual branches of a tree. If, for example, the character states at either end of a branch are estimated with a high degree of certainty, then the parsimony reconstruction of the character on that branch might be accepted. Conversely, if the ancestral states at either end of the branch are uncertain, then any assignment of the history along that branch is also considered uncertain. The problem with this method is that it does not account for the residual uncertainty in the character's history along that branch; e.g., nonparsimonious histories might also be reasonable, especially if the branch is very long.

Uncertainty in the phylogenetic history of the species being examined is another factor that must be accounted for in comparative studies. Accommodating

phylogenetic uncertainty is important for at least two reasons: (1) the conclusions of a study might change depending upon the tree used to map the characters, and (2) it is not appropriate to consider the tree fixed in a statistical analysis of character evolution because the tree is not known with certainty; the danger is that the researcher will have too much confidence in the results, even if correct, because only a single tree was examined. A common approach to account for phylogenetic uncertainty is to examine the evolution of a character on a number of reasonable trees, under the assumption that the set of trees examined reflects uncertainty in the tree. Another approach involves examining the evolution of the character on a set of trees generated under a stochastic process, such as the birth–death process (Losos, 1994; Martins, 1996) or examining the evolution of a character on all trees with the results from each tree weighted by the probability of the tree being correct (Losos and Miles, 1994; Pagel, 1994; Huelsenbeck et al., 2000). The advantage of the last approach is that different trees contribute differentially to the final results of the comparative analysis; trees that better explain the data contribute more to the final results of the comparative analysis than do trees that poorly explain the data. Algorithms exist for approximating the probabilities of trees (Larget and Simon, 1999; see Huelsenbeck et al., 2000, 2001).

To date, it has been difficult to combine methods for accounting for uncertainty in the phylogenetic tree (and other parameters of the statistical model) with mappings of a character's history. The problem of explicitly mapping characters on a tree can be avoided by constructing a stochastic model of character evolution and then examining the parameters of this model. However, mapping character histories on phylogenies has a number of advantages if the traditional problems associated with determining a character's history can be overcome. For example, the number of changes, the specific places on the tree where the changes occurred, and the timing of the character transformations can all be discerned by examining the history of a character.

Here, we apply a method described by Nielsen (2002) for stochastically mapping characters on a tree to the problem of determining a morphological character's history. We account for uncertainty in model parameters by adopting a Bayesian perspective; results are averaged over all parameter values and weighted by the probability of the parameters. This approach accommodates uncertainty in the phylogenetic history of the group when mapping characters. The method for mapping character histories does not rely on the parsimony method. Rather, character change is assumed to follow a continuous-time Markov process, just as with other approaches for examining morphological character change. However, this method allows specific character histories that are consistent with the observations to be sampled according to their probability under the model. We also suggest an intuitive measure of the degree to which two characters covary and suggest a test of independence based on Bayesian posterior predictive P values.

METHODS

Sampling Character Histories

A common assumption of many phylogenetic methods and the approach described here is that character change follows a continuous-time Markov chain. At the heart of a continuous-time Markov chain is a matrix, $\mathbf{Q}$, specifying the rates of change from one character state to another. The rate matrix provides all of the information necessary to describe the process. For example, consider the following matrix describing the substitution process among three character states, 0, 1, and 2:

$$\mathbf{Q} = \{q_{ij}\} = \begin{pmatrix} -(a+b) & a & b \\ d & -(c+d) & c \\ e & f & -(e+f) \end{pmatrix}.$$

The rows and columns of this matrix are in the order 0, 1, 2. The off-diagonal elements of the matrix give the rate of change from state i (the row) to state j (the column). For example, d is the rate of change from state 1 to state 0, and b is the rate of change from state 0 to state 2. The diagonal elements give the rate of change away from state i; these numbers are negative because they are rates away from a state.

The rate matrix carries the information for a complete description of the substitution process. When the process is in state i, an exponentially distributed amount of time with parameter $-q_{ii}$ passes until the next character state transformation occurs. When the next character state change occurs, we change to state j with probability $-(q_{ij}/q_{ii})$. For example, if the process is currently in state 1, then the second row of the matrix $\mathbf{Q}$ specifies the rates of change for the next event of substitution. The elements of only the second row of the rate matrix are

$$\mathbf{Q} = \{q_{ij}\} = \begin{pmatrix} . & . & . \\ d & -(c+d) & c \\ . & . & . \end{pmatrix},$$

and an exponentially distributed amount of time with parameter $c + d$ passes until the next character state change occurs. The rate matrix then specifies the relative probabilities of the different changes. With probability $d/(c + d)$ the change is from 1 to 0, and with probability $c/(c + d)$ the change is from 1 to 2. If the next substitution is from 1 to 2, then the process starts over again. The third row of the rate matrix is now the relevant one:

$$\mathbf{Q} = \{q_{ij}\} = \begin{pmatrix} . & . & . \\ . & . & . \\ e & f & -(e+f) \end{pmatrix},$$

and an exponentially distributed amount of time with parameter $e + f$ passes until the next character change occurs. The change is to state 0 with probability $e/(e + f)$ and to state 1 with probability $f/(e + f)$, and so on.

The rate matrix also contains the information necessary to calculate the transition probabilities and stationary distribution of the process. The transition probabilities are the probability of the process ending in state j after time v conditional on having started in state i. The transition probabilities are critical to calculating the likelihood function for maximum likelihood and Bayesian analyses of phylogeny and can be calculated by exponentiating the product of the rate matrix and time: $\mathbf{P}(v) = \{p_{ij}(v)\} = e^{s\mathbf{Q}v}$. The stationary probability distribution of the process is obtained by considering the probabilities that the process ends in state j after a long (infinite) period of time. When the Markov chain is at stationarity, the process has "forgotten" its starting state; the probability of finding the process in state j is independent of its starting state. The stationary probability of state i is denoted π_i.

Modeling character change as a continuous-time Markov process has a number of advantages. First, the number of character state changes increases with time or substitution rate, in contrast to the parsimony method, which allows at most one character state transformation along a single branch of a phylogeny. To illustrate this behavior, consider the simplest possible phylogeny: a tree of two species where two different character states have been observed in the two species. Moreover, consider a simple continuous-time Markov chain that describes the evolution of the character. This Markov chain has only two states, with rate matrix

$$\mathbf{Q} = \{q_{ij}\} = \begin{pmatrix} -\pi_1 & \pi_1 \\ \pi_0 & -\pi_0 \end{pmatrix} \mu,$$

where π_0 and π_1 are the stationary frequencies of states 0 and 1, respectively, and $\mu = 1/(2\pi_0\pi_1)$ is a scaling parameter that ensures divergence is measured in terms of expected number of transformations per character (instead of in units of time). This scaling is done because divergence depends only upon the product of rate ($\mathbf{Q}$) and time (t). Unless the measures of one are independent, the other cannot be determined, so time on the tree is conveniently measured as divergence in characters. For this matrix, the stationary frequencies have been built into the rate matrix, as is a common practice for many phylogenetic models.

In our example, we make the further restriction that $\pi_0 = \pi_1 = 1/2$. The rate matrix is then

$$\mathbf{Q} = \{q_{ij}\} = \begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix},$$

and the number of character transformations over a branch of length v is a Poisson-distributed random variable with parameter v. (The number of character transformations follows a Poisson distribution only when the diagonals of the rate matrix are equal; for models of DNA sequence substitution, this is true for the Jukes–Cantor [1969] and Kimura [1980] models.) Figure 1 shows some possible character histories that could describe the pattern of observations at the tip of the two-species tree. Here, there can only be 1, 3, 5, etc., changes; the model allows only two states, 0 or 1, which forces an odd number of changes to explain the different states observed. If the character states observed in the two species had been the same, then 0, 2, 4, 6, etc., character changes would have been necessary.

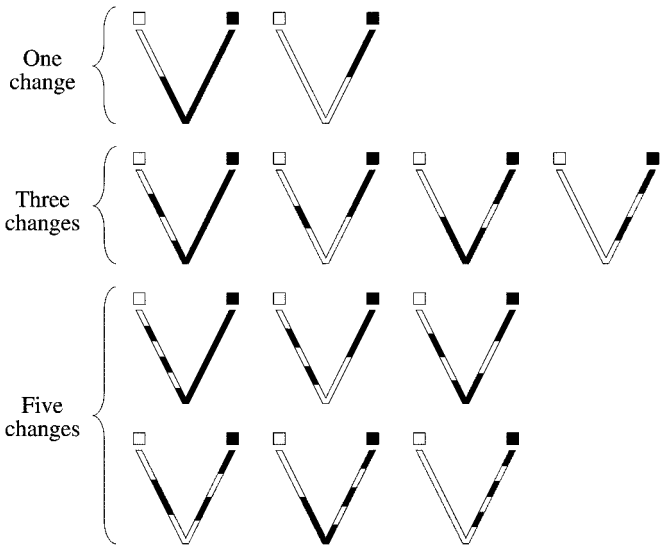


FIGURE 1. Patterns of character change on a tree of two species when the observed character states differ and there are two possible character states. In this situation, the observations at the tips of the tree can only be explained with an odd number of changes.

The probability of observing the data at the tips of the tree, $\boldsymbol{x}$ (state 0 in the left species and state 1 in the right species) conditional on the length of the two branches (v; the total tree length is $2v$) can be calculated in two equivalent ways. The first way takes advantage of the transition probabilities $p_{ij}(v)$,

$$f(\boldsymbol{x} \mid v) = \pi_0 p_{00}(v) p_{01}(v) + \pi_1 p_{10}(v) p_{11}(v),$$

and the probability of observing the data is averaged over the two possible states at the root of the tree. The other way takes advantage of the fact that, for this specific case, the number of character changes follows a Poisson distribution:

$$f(\boldsymbol{x} \mid v) = \sum_{i=0}^{\infty} \frac{(2v)^i e^{-2v}}{i!} \times z$$

(for this example, $z = 1/2$ when i is odd, and 0 otherwise). The probability of observing the data is $f(\boldsymbol{x} \mid v = 0.1) = 0.082419988$, $f(\boldsymbol{x} \mid v = 0.5) = 0.216166179$, and $f(\boldsymbol{x} \mid v = 1.0) = 0.24542109$ when the branch lengths are 0.1, 0.5, and 1.0 expected character changes in length, respectively. Table 1 shows the probability of observing the data conditional on there having been i total changes. When the branch length is small, the most-parsimonious

TABLE 1. The probabilities of i changes for a tree of two species. The observed character states, $\boldsymbol{x}$, are different for the two species, and there are only two possible states for the character. In this case, there must be an odd number of changes. The length of the branch leading to each species is v expected changes. The total tree length then is $2 \times v$. Rates of change are the same for each character transformation. The probability of observing the data at the tips of the tree conditioned on the branch length is $f(\boldsymbol{x} \mid v = 0.1) = 0.082419988$, $f(\boldsymbol{x} \mid v = 0.5) = 0.216166179$, and $f(\boldsymbol{x} \mid v = 1.0) = 0.24542109$; these numbers are obtained by integrating over all possible character histories.

i	$f(\boldsymbol{x} \mid v = 0.1, i)$	$f(\boldsymbol{x} \mid v = 0.5, i)$	$f(\boldsymbol{x} \mid v = 1.0, i)$
0	0.0 (0.00%)	0.0 (0.00%)	0.0 (0.00%)
1	0.081873075 (99.34%)	0.18393972 (85.09%)	0.135335283 (55.14%)
2	0.0 (0.00%)	0.0 (0.00%)	0.0 (0.00%)
3	0.000545820 (0.66%)	0.03065662 (14.18%)	0.090223522 (36.76%)
4	0.0 (0.00%)	0.0 (0.00%)	0.0 (0.00%)
5	0.000001092 (0.00%)	0.00153283 (0.71%)	0.018044704 (7.35%)
6	0.0 (0.00%)	0.0 (0.00%)	0.0 (0.00%)
7	$<10^{-8}$ (0.00%)	0.00003649 (0.02%)	0.001718543 (0.70%)

reconstruction accounts for most of the total probability; when $v = 0.1$, 99.34% of the probability is on the most-parsimonious reconstruction. However, when the branch length is long, then the most-parsimonious reconstruction is less probable. The most-parsimonious reconstruction accounts for only 85.09% and 55.14% of the probability when the branch lengths are 0.5 and 1.0, respectively.

Ideally, one would like to randomly sample character histories that are consistent with the observations at the tips of a phylogenetic tree. Moreover, one would like to do this for any continuous-time Markov process describing the evolution of the character on the tree (i.e., one would not want to be limited to considering only Poisson process models). Nielsen (2002) described a method for sampling character histories consistent with the observations at the tips of the tree. The method takes advantage of the description of the process as exponential waiting times between events of character change. The method works as follows. First, calculate the probabilities of the ancestral states at each interior node on the phylogenetic tree (Felsenstein, 1981). Using these probabilities, randomly sample a combination of states for the ancestral nodes on the tree. The states at either end of each branch of the phylogeny are now fixed. The state at one end of a branch is denoted i and the state at the other end is denoted j. Next, visit a branch on the phylogeny and simulate a realization of the substitution process that is consistent with the states i and j. Specifically, starting with state i at one end of the branch, simulate an exponentially distributed random variable with parameter $-q_{ii}$. When an event of character change occurs, the change is to state k with probability $-(q_{ik}/q_{ii})$. Generate exponential random variables until the next event of character change exceeds the length of the branch, v (specifically, the scaled branch length). If the state at the end of the process is j, then a single realization of the Markov process has been successfully generated for the branch. Otherwise, the process is repeated until a realization consistent with the beginning and ending states is made. Generating exponentially distributed random variables is easy. Specifically, a uniformly distributed random variable, U, on the interval $(0, 1)$ can be converted to an exponentially distributed random variable, T, using the transformation $T = -1/\lambda \ln(1 - U)$, where λ is the rate parameter of the exponential distribution. Nielsen (2002) pointed out that computational efficiency can be improved when the states at either end of the branch are not the same by conditioning on there having been at least a single character change. Finally, repeat the process for all branches on the tree. This procedure appears to be efficient when the Markov chain does not have many character states.

Figure 2 shows 50 realizations of character histories for the simple two-taxon tree. The rates of change differ for the character mappings. For the 25 mappings to the left of the vertical line in Figure 2, the tree length is 1.0. For the 25 mappings to the right of the vertical line, the tree length is 2.0. In many instances, the character mapping is not the most-parsimonious one; in 4 of the 25 low-rate cases and in 11 of the 25 high-rate cases, the character history involves more than a single character change.

Priors on Tree Length and Transition Bias

Here, we restrict attention to two models of character change, one appropriate for two-state and the other for three-state characters. The matrix of instantaneous rates for the two-state model is

$$\mathbf{Q} = \{q_{ij}\} = \begin{pmatrix} -\pi_1 & \pi_1 \\ \pi_0 & -\pi_0 \end{pmatrix} \mu,$$

where $\mu = 1/(2\pi_0\pi_1)$, and the instantaneous rate matrix for the three- state model is

$$\mathbf{Q} = \{q_{ij}\} = \begin{pmatrix} -1 & 1/2 & 1/2 \\ 1/2 & -1 & 1/2 \\ 1/2 & 1/2 & -1 \end{pmatrix}.$$

Lewis (2001) suggested the use of such models for analysis of morphological characters; our three-state model is identical to the M_3 model of Lewis (2001). The transition probability matrices for these two models, obtained by exponentiating the product of the rate matrix and the branch length, are as follows:

$$\mathbf{P}(v) = \{p_{ij}(v)\} = \begin{pmatrix} \pi_0 + \pi_1 e^{-\mu v} & \pi_1 - \pi_1 e^{-\mu v} \\ \pi_0 - \pi_0 e^{-\mu v} & \pi_1 + \pi_0 e^{-\mu v} \end{pmatrix}$$

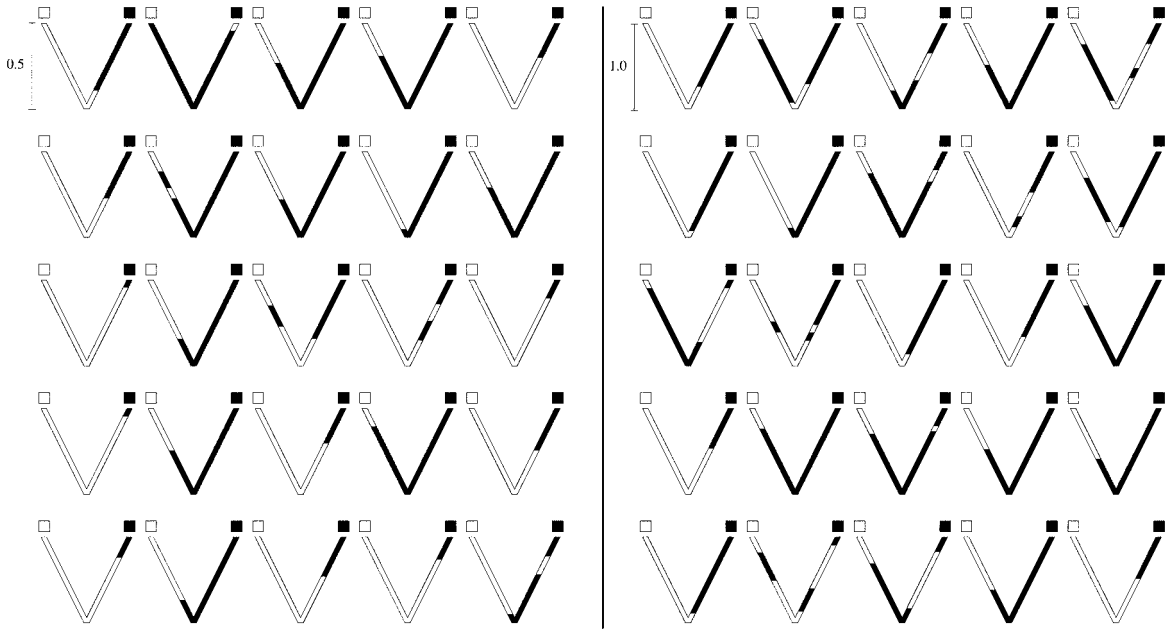


FIGURE 2. Fifty random realizations of character histories consistent with the observations at the tips of the tree when there are two character states with equal rates of change from ■ → and □ → ■. The tree length for the 25 realizations to the left of the line is one substitution, and the tree length for the 25 realizations to the right of the line is two substitutions.

for the two-state model, where $\mu = 1/(2\pi_0\pi_1)$, and

$$\mathbf{P}(v) = \{p_{ij}(v)\} = \begin{pmatrix} \frac{1}{3} + \frac{1}{3}e^{-\frac{3}{2}v} & \frac{1}{3} - \frac{1}{3}e^{-\frac{3}{2}v} & \frac{1}{3} - \frac{1}{3}e^{-\frac{3}{2}v} \\ \frac{1}{3} - \frac{1}{3}e^{-\frac{3}{2}v} & \frac{1}{3} + \frac{1}{3}e^{-\frac{3}{2}v} & \frac{1}{3} - \frac{1}{3}e^{-\frac{3}{2}v} \\ \frac{1}{3} - \frac{1}{3}e^{-\frac{3}{2}v} & \frac{1}{3} - \frac{1}{3}e^{-\frac{3}{2}v} & \frac{1}{3} + \frac{1}{3}e^{-\frac{3}{2}v} \end{pmatrix}$$

for the three-state model. The two-state model allows the rate of change to differ between the two states. Ultimately, we are interested only in mappings of characters consistent with the character states observed at the tips of the tree, i.e., the probability distribution of character histories, h, conditional on the observations, $\boldsymbol{x}$. This probability distribution is called the posterior probability distribution of character histories and can be approximated using the simulation scheme presented by Nielsen (2002). However, to simulate character histories, many other parameters must be fixed; these parameters include the topology of the phylogenetic tree (τ), the branch lengths of the tree ($\boldsymbol{v}$), and parameters of the continuous-time Markov chain (here, there are the bias parameters π_0 and π_1 for two-state models). Ideally, we would accommodate uncertainty in all of the parameters.

Our general approach was to approximate the posterior probability of the parameters of the model using the Markov chain Monte Carlo method (MCMC; Metropolis et al., 1953; Hastings, 1970; Mau 1996; Rannala and Yang, 1996; Mau and Newton, 1997; Yang and Rannala, 1997; Larget and Simon, 1999; Mau et al., 1999). We then simulated character histories on the trees sampled using the MCMC algorithm. We used MrBayes (version 3.0; Huelsenbeck and Ronquist, 2001) to sample trees using the MCMC procedure from the available DNA sequence information. For all of the DNA sequence data sets analyzed in this study, we assumed the so-called HKY85 + Γ_4 model of DNA substitution (Hasegawa et al., 1984, 1985). This model allows transitions to occur at a different rate than transversions, different stationary frequencies for the nucleotides, and rate variation across sites. We modeled rate variation across sites by assuming that the rate at a site is a random variable drawn from a gamma distribution with shape and scale parameters equal to α (Yang, 1993). Specifically, we used the discrete approximation of the gamma distribution described by Yang (1994), breaking the gamma distribution into four categories. The output of the program MrBayes is a file of trees that were sampled by the chain with branch lengths in terms of expected number of nucleotide substitutions per site. We discarded a fraction of the early trees sampled by the MCMC algorithm as the burn-in of the chain. The fraction of the time a tree is represented in the file is a valid approximation of the posterior probability of the tree.

The branch lengths of the trees sampled by the MCMC algorithm are relevant to the alignment of DNA sequences but not to the morphological characters we are interested in. We rescaled the branch lengths of the trees sampled by the MCMC procedure such that the total tree length was 1.0. The branch length proportions from the DNA sequence were maintained. The question of the appropriate rate for the morphological characters was then addressed. We adopted the approach advocated by Schultz and Churchill (1999) in which prior probability distributions are placed on the overall rate of change for the morphological characters and on the bias parameter (π_0 or π_1 for the two-state model). Specifically, we placed a gamma prior on the rate of substitution and a beta prior on the bias parameter. We explored the consequences of

assuming different gamma priors on the mapping of the characters.

Each stochastic mapping of a character, then, is preceeded by sampling a specific tree length and a specific bias parameter. The tree length had a gamma prior in all of our analyses. The gamma distribution has two parameters, the shape (α) and the scale (β). The mean of the gamma distribution is α/β and the variance is α/β^2. In the following analyses, we tried different values for the mean prior tree length and the SD of tree length. For example, in at least one analysis we set the mean prior tree length (T) to 10 and the SD to 1. We can obtain such a prior on tree length by setting $\alpha = 100$ and $\beta = 10$ ($E[T] = 100/10 = 10$ and $SD[T] = \sqrt{100/10^2} = 1$). We then calculated the posterior probability distribution of tree length using Bayes's theorem to combine the gamma prior on tree length with the character observations. Hence, even though the prior mean of the tree length is $E(T) = \alpha/\beta$, the posterior mean may be smaller or larger. For example, if the prior mean is set to a large number but the character of interest is invariable, then the posterior mean is likely to be much smaller than the prior mean. The only computational consideration is that it is very difficult to calculate the posterior probability distribution of tree length using a continuous gamma prior. We simplified the problem by breaking the gamma distribution into 50 rate categories, thereby creating a discrete distribution (Yang, 1994). Instead of picking from a continuous distribution, which is very difficult to do, we end up picking from 1 of 50 rate categories. The mean rate for a category was used to represent the tree length.

Once the tree length was sampled, we chose a bias parameter from the (continuous) beta prior. Like the gamma distribution, the beta distribution also has two parameters. Unfortunately, these parameters also are usually denoted α and β. The mean of the beta distribution is $\alpha/(\alpha + \beta)$, and the variance is $\alpha\beta/(\alpha + \beta)^2(\alpha + \beta + 1)$. In all analyses, we used a "flat" beta distribution, which has $\alpha = \beta = 1$. We calculated the posterior probability distribution of the bias parameter on transition rates by breaking the continuous beta distribution into 11 categories.

We accommodated uncertainty in parameters of the evolutionary model using the method advocated by Huelsenbeck et al. (2000). Specifically, for each tree and set of branch lengths sampled by the MCMC algorithm, we selected a rate category and, if appropriate, a bias category from the posterior probability distribution for the character. We then sampled ancestral character states and mapped the character using the simulation method described in the previous section. This process was repeated for all of the trees sampled. We can summarize the results by examining the posterior probability distribution of the number of character transformations or the expected value for the number of transformations. Our estimates are not conditioned on any single tree being correct because the MCMC procedure assures that trees are sampled according to their posterior probability. Moreover, many different character histories are generated, one for each tree sampled using MCMC, so our inferences do not rely on a single mapping of characters, such as the most parsimonious character history.

Testing for Correlation

One question that naturally arises when considering the history of more than one character is whether the characters are correlated. This has been an area of active research in statistical phylogenetics, and many approaches have been suggested for discrete (Pagel, 1994) and continuous (Felsenstein, 1985) characters. Here, we concentrate on correlation of discrete characters. Specifically, we examine how often combinations of character states in two characters are associated.

The following description illustrates the basis for our test. Imagine that we have generated a realization of the history for two characters. The states for character 1 are denoted A, B, C, etc., and the states for character 2 are denoted a, b, c, etc. After following the procedure described by Nielsen (2002) and outlined here, a realization of the history for both characters is obtained. For any single realization, the probability of finding the character in state i is simply the fraction of the time that the character was in state i over the phylogeny. Similarly, with a single realization of the history for both characters, the probability of finding state i of character 1 associated with state j of character 2 can be calculated. The results for the realizations of the history for two characters, each of which has three states, might be summarized as in Table 2. This table has a simple interpretation. For this specific stochastic realization of the history for both characters, state A was associated with state a 10% of the time, state A was associated with state b 17% of the time, state A was associated with state c 9% of the time, state B was associated with state a 8% of the time, and so on. Moreover, character 1 was in state A 36% of the time, in state B 39% of the time, and in state C 25% of the time. The numbers along the margins of this table, i.e., the marginal probability of finding characters 1 or 2 in the various states, can be used to make predictions about the frequency of association for the two characters. Specifically, given the marginal probabilities of finding character 1 in state i ($p_1[i]$) and character 2 in state j ($p_2[j]$) and assuming independence of the characters, the probability of finding i associated with j is simply $p_1(i) \times p_2(j)$. Following our example of a single realization of two characters, each with three states, the expected frequencies with which the states are associated are given in Table 3.

TABLE 2. The realization of the history for two characters, each with three states. Totals represent the marginal probabilities of finding characters 1 and 2 in the various sates.

	Character 1			
Character 2	A	B	C	Total
a	0.10	0.08	0.07	0.25
b	0.17	0.09	0.04	0.30
c	0.09	0.22	0.14	0.45
Total	0.36	0.39	0.25	

TABLE 3. Expected frequency of association for three states of two characters. Totals represent the marginal probabilities of finding characters 1 and 2 in the various states.

	Character 1			
Character 2	A	B	C	Total
a	0.09	0.10	0.06	0.25
b	0.11	0.12	0.07	0.30
c	0.16	0.18	0.11	0.45
Total	0.36	0.39	0.25	

Figure 3 shows five realizations of stochastic mappings for two characters, both of which have two states: black and white. The states appear to be correlated; the black state for character 1 always is found in association with the black state in character 2, and the white states are also perfectly associated. This appearance is confirmed by examining summaries of the character histories for each of the five realizations.

The match between the observed association of characters states and the association expected if the characters were independent can be summarized in a number of ways. One would think that the covariance between the characters would be adequate; unfortunately, the labeling for the characters states is completely arbitrary, and the covariance has little meaning for this situation. Here, we use a number of summary statistics. First, we examine the difference between the observed and expected values for each combination of states in characters 1 and 2:

$$d_{ij}(\tau, v, h_1, h_2) = a_{ij}^{(o)} - a_{ij}^{(e)},$$

where $a_{ij}^{(o)}$ is the observed coincidence of states i and j and $a_{ij}^{(e)}$ is the expected coincidence of states i and j. The value for d_{ij} is negative if states i and j are found together less frequently than would be expected under independence and positive if they are found together more frequently than expected. The statistic d_{ij} depends upon the tree, branch lengths, and character mappings for characters 1 and 2 (h_1 and h_2). Second, we use an overall measure of the disagreement between the observed and expected associations of the states for the two characters:

$$D(\tau, v, h_1, h_2) = \sum_{i=1}^{n} \sum_{j=1}^{m} \left| d_{ij}(\tau, v, h_1, h_2) \right|,$$

where n is the number of states for character 1 and m is the number of states for character 2.

We treat the statistics d_{ij} and D as random variables; they can differ from one realization of character mappings to another on a given tree and also can differ among trees. We use the expected values of these test statistics, calculated by summing over all trees and over multiple realizations of character histories:

$$E(d_{ij} \mid \boldsymbol{X}) = \sum_{\tau} \int_{v,h_1,h_2} d_{ij}(\tau, v, h_1, h_2)\, f(\tau, v, h_1, h_2 \mid \boldsymbol{X})\, dv\, dh_1\, dh_2.$$

and

$$E(D \mid \boldsymbol{X}) = \sum_{\tau} \int_{v,h_1,h_2} D(\tau, v, h_1, h_2)\, f(\tau, v, h_1, h_2 \mid \boldsymbol{X})\, dv\, dh_1\, dh_2.$$

We approximate $E(d_{ij} \mid \boldsymbol{X})$ and $E(D \mid \boldsymbol{X})$ by using the trees and branch lengths sampled using the MCMC procedure. Specifically, for each sampled tree, we generate one mapping for characters 1 and 2, evaluating d_{ij} and D. The average d_{ij} and D for all trees is an approximation of $E(d_{ij} \mid \boldsymbol{X})$ and $E(D \mid \boldsymbol{X})$. We use these approximations as test statistics and calculate the posterior predictive P value for each test statistic.

The posterior predictive P value is calculated as follows. For each tree sampled in the original MCMC procedure, we simulate a large number of character histories under the assumption that the two characters are independent. The histories simulated in this manner may not match the realized pattern of character states observed at the tips of the tree. For the kth tree, we calculate $E_k(d_{ij} \mid \boldsymbol{X}^*)$ and $E_k(D \mid \boldsymbol{X}^*)$ by averaging over many random realizations of character histories for the two characters. The posterior predictive P value is calculated over all trees:

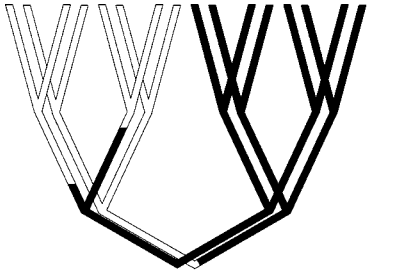
$$P = Pr[E(D \mid \boldsymbol{X}^*) \geq E(D \mid \boldsymbol{X})] = \frac{1}{N} \sum_{k=1}^{N} I[E_k(D \mid \boldsymbol{X}^*) \geq E(D \mid \boldsymbol{X})],$$

where (N is the number of sampled trees and $I(\cdot)$ is 1 if $E_k(D \mid \boldsymbol{X}^*) \geq E(D \mid \boldsymbol{X})$ and 0 otherwise. Thus, the posterior predictive P value is the probability of a value for the test statistic as extreme as was observed under a model in which the characters evolve independent.

Figure 4 illustrates the simulation method we used to approximate posterior predictive P values. The observed value for the test statistic is averaged over all of the trees sampled using the MCMC procedure. The predictive distribution is obtained by simulating two characters with independent histories for each tree (and associated parameter values) sampled by the chain. We then calculate the predicted value for each simulated data set exactly as we calculated the test statistic for the observed data. The only deviation in the procedure outlined in Figure 4 (which is the ideal way to calculate the predictive distribution) is that we did not go through a second round of MCMC because it was unnecessary in this study.

Data

We examined the utility of the methods described here with three data sets. The first data set includes 34 species of aphids of the tribe Cerataphidini (Stern, 1998). Ten of the aphid species form horned soldiers,

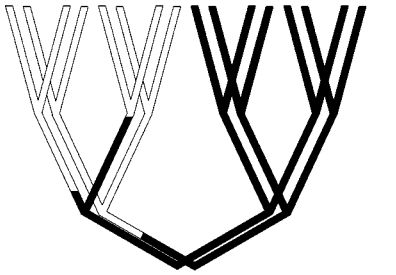


Observed

L \ U	0	1
0	0.36	0.14
1	0.00	0.50

Expected

L \ U	0	1
0	0.18	0.32
1	0.18	0.32

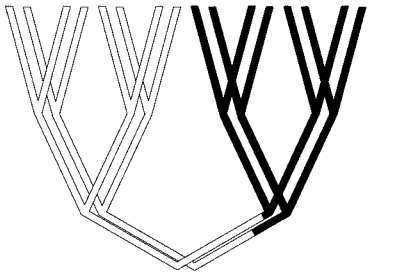


Observed

L \ U	0	1
0	0.35	0.11
1	0.00	0.54

Expected

L \ U	0	1
0	0.16	0.30
1	0.19	0.35

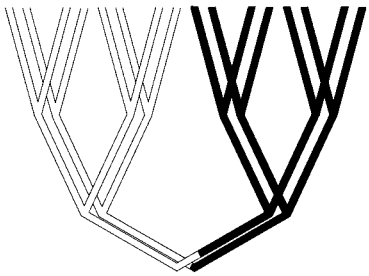


Observed

L \ U	0	1
0	0.55	0.00
1	0.02	0.43

Expected

L \ U	0	1
0	0.31	0.24
1	0.26	0.19

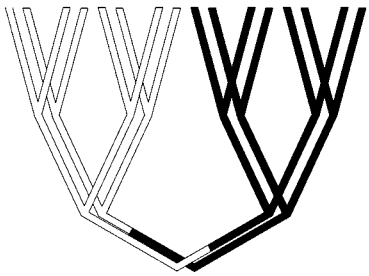


Observed

L \ U	0	1
0	0.50	0.02
1	0.00	0.48

Expected

L \ U	0	1
0	0.26	0.26
1	0.24	0.24



Observed

L \ U	0	1
0	0.45	0.07
1	0.00	0.48

Expected

L \ U	0	1
0	0.23	0.29
1	0.22	0.26

FIGURE 3. Five realizations of the histories for two characters. The characters (white and black) are mapped onto a tree with all of the branches being equal in length. A gamma prior with expectation $E(T) = 1$ and an SD = 1 was assumed. The two characters appear to covary: The white state from one character is frequently associated with the white character from the other character, and the black state from one character is frequently associated with the black state from the other character. U = upper tree character; L = lower tree character. The observed matrix gives the fraction of the time state i from one character was associated with state j from the other. The expected matrix gives the fraction of the time the states should be associated if they were independent.

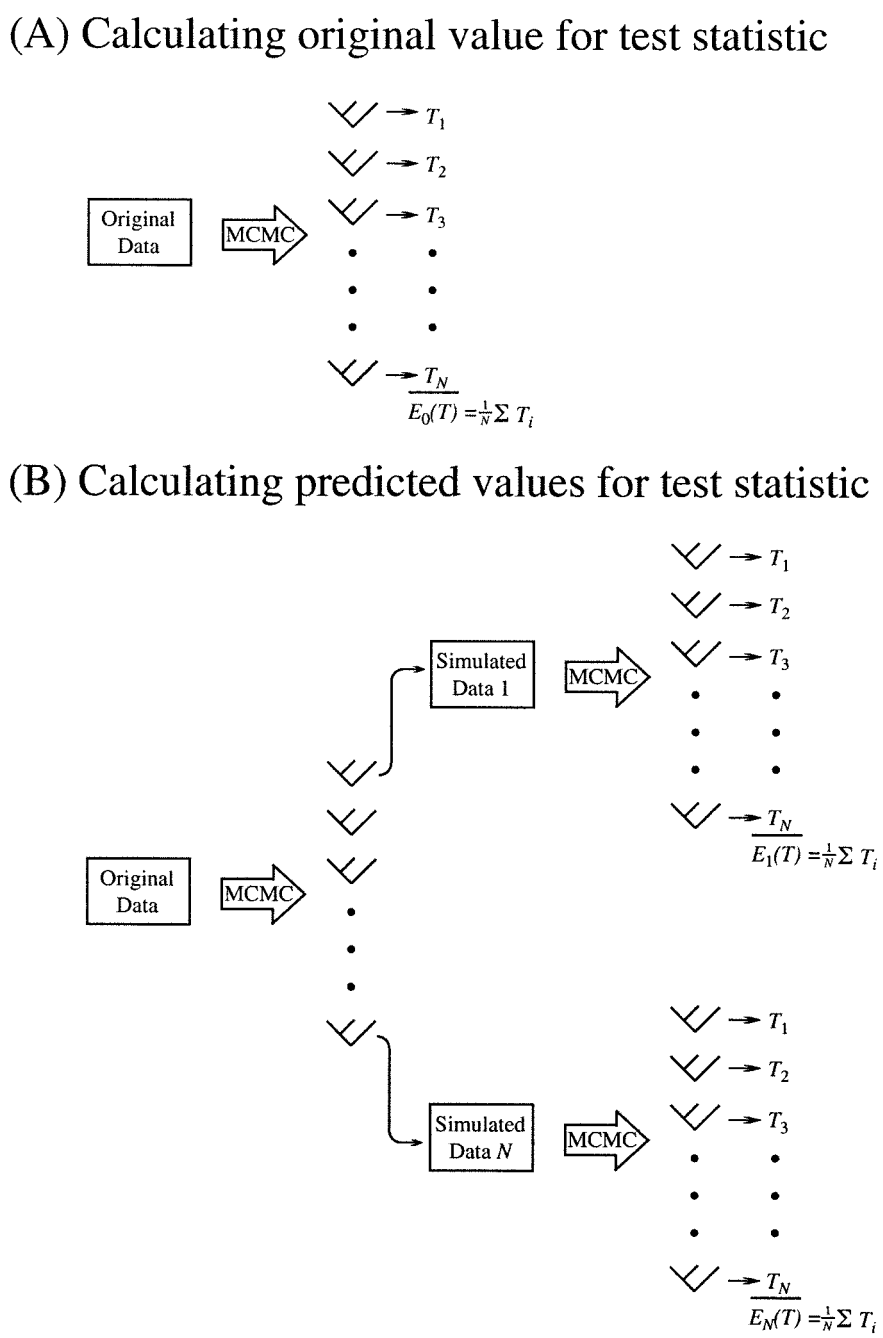


FIGURE 4. An example of how posterior predictive P values are calculated for a test statistic, T. The observed value for the test statistic is calculated by averaging over the posterior probability distribution of parameters. We use MCMC to draw parameter values from the posterior probability distribution. The predictive distribution is calculated by simulating new data using parameter values from the posterior probability distribution of parameters. Each simulated data set is treated exactly as was the original data. The predictive P value is the proportion of the test statistics from the simulated data that exceed the observed value.

who are used for nest defense. Stern (1998) estimated the phylogeny of the aphid tribe using mitochondrial cytochrome oxidases I and II (COI and COII) sequences. The phylogeny suggested a single origin and repeated loss of horned soldiers in the tribe Cerataphidini. The second data set includes 12 sea stars of the family Asterinidae (Hart et al., 1997). Hart et al. (1997) examined the phylogeny of the seastars using the mitochondrial cytochrome oxidase I gene and five mitochondrial tRNA genes (alanine, leucine, asparagine, glutamine, and proline). Hart et al. (1997) examined a number of life-history traits, but we concentrated on whether the larvae are feeding or nonfeeding (see Cunningham, 1999). We also examined the evolution of flower morphology and self-incompatibility in the family Pontederiaceae (Kohn et al., 1996; Graham et al., 1998). The phylogeny estimates for

the Pontederiaceae were based on DNA sequences from the chloroplast *rbcL* and *ndhF* genes.

RESULTS

Bayesian Analysis

We used MrBayes (version 3.0; Huelsenbeck and Ronquist, 2001) to approximate the posterior probabilities of trees for the aphid, seastar, and Pontederiaceae data using the MCMC procedure. Posterior probabilities were calculated under the HKY85 + Γ_4 model of DNA substitution. A single chain was run for 1,000,000 cycles for each data set, and the states of the chain were sampled every 100th cycle, for a total of 10,000 sampled trees and substitution model parameter values. We also calculated the maximum likelihood value for each data set under the HKY85 + Γ_4 model using PAUP* (Swofford, 2002), which allowed us to evaluate how close the chain came to the maximum likelihood values. Figure 5 shows plots of the log likelihood over the course of the MCMC analysis for the aphid data. The log likelihood was initially very small because the chain was initiated with a random tree. However, it rapidly increased as more reasonable parameter values were sampled, eventually reaching a plateau near the maximum likelihood value (indicated by the dashed line). Figure 6 shows the Bayesian majority rule consensus tree for the aphid data. The majority rule consensus tree is based on the last 8,000 trees sampled by the Markov chain; the first 2,000 trees sampled by the chain were discarded as the burn-in (that portion of the chain sampled before apparent stationarity was reached). The numbers at the interior branches of the tree of Figure 6 indicate the posterior probability of the clade. Figure 7 shows the majority rule consensus tree of Figure 6 drawn as a phylogram. The lengths for each branch were averaged over all trees that had the branch. Figures 8, 9, and 10 show plots of the log likelihood through time, a majority rule consensus tree, and a majority rule phylogram, respectively, for the seastars. Similarly, Figures 11, 12, and 13 show plots of the log likelihood through time, a majority rule consensus tree, and a majority rule phylogram, respectively, for the Pontederiaceae. The trees for the seastars and Pontederiaceae were based on the last 8,000 trees sampled by the chain.

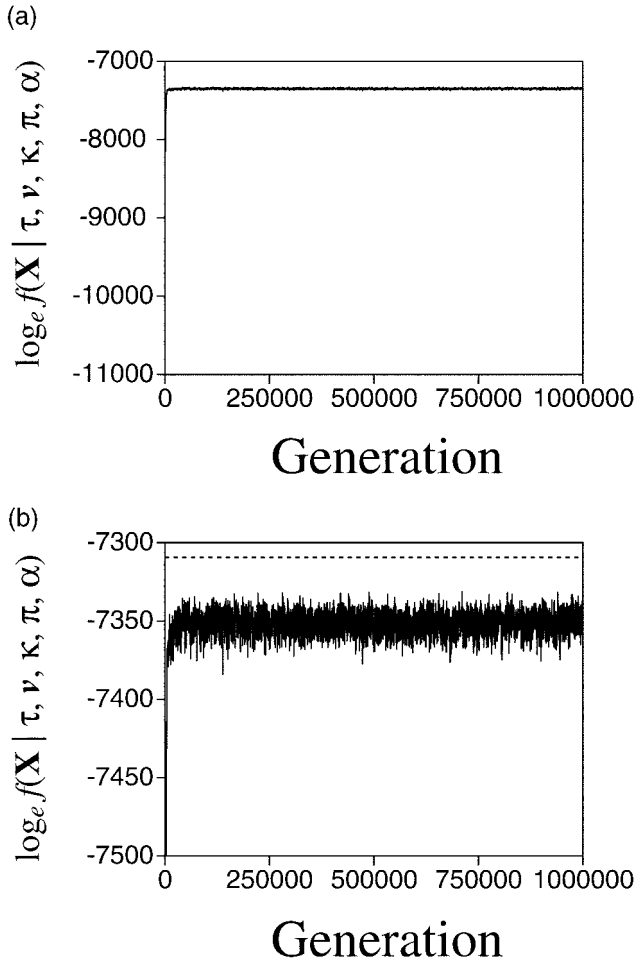


FIGURE 5. The log likelihood of the states visited by the Markov chain for the aphid COI and COII sequences. The likelihood is proportional to the probability of observing the alignment of sequences ($\boldsymbol{X}$) conditional on the tree (τ), branch lengths ($\boldsymbol{v}$), transition/transversion rate ratio (κ), base frequencies ($\boldsymbol{\pi}$), and gamma shape parameter (α). (a) Log likelihoods for all states visited by the chain, including the early states visited by the chain before apparent stationarity was reached. (b) Log likelihoods of states visited when the chain was near stationarity. The dashed line indicates the maximum likelihood under the HKY85 + Γ_4 model.

The Bayesian analysis integrates over uncertainty in the substitution model parameters. Table 4 shows the mean and 95% credible interval for the tree length, transition/transversion rate ratio, base frequencies, and gamma shape parameter. Bayesian analysis requires the specification of a prior on the parameters of the model. Here, we considered all trees equally probable a priori and placed a uniform(0,10) prior on branch lengths of the tree. Moreover, the transition/transversion rate ratio had a uniform(0,50) prior, the gamma shape parameter had a uniform(0,50) prior, and the base frequencies had a flat dirichlet(1,1,1,1) prior. As indicated in the figures and in Table 4, these relatively flat priors were strongly modified by the data; most of the posterior probability density was focused on a relatively small number of trees and on a narrow range of substitution model parameters.

Analysis of Character History

Figure 14 shows mappings of the horned-soldier caste character on six arbitrarily chosen trees. The horned-soldier caste is indicated by bold lines on the trees. Only half of the six character mappings matched the numbers of gains and losses that would be produced using the parsimony method. Table 5 shows the posterior probabilities for different numbers of gains and losses for two different priors on the tree length. We considered two priors: a low-rate prior with a mean tree length of 1 and a high-rate prior with an average tree length of 10 character transformations. The maximum posterior probability estimate of the number of gains and losses is one gain and two losses. This holds true for the low- and high-rate priors. However, the 95% credible set of character

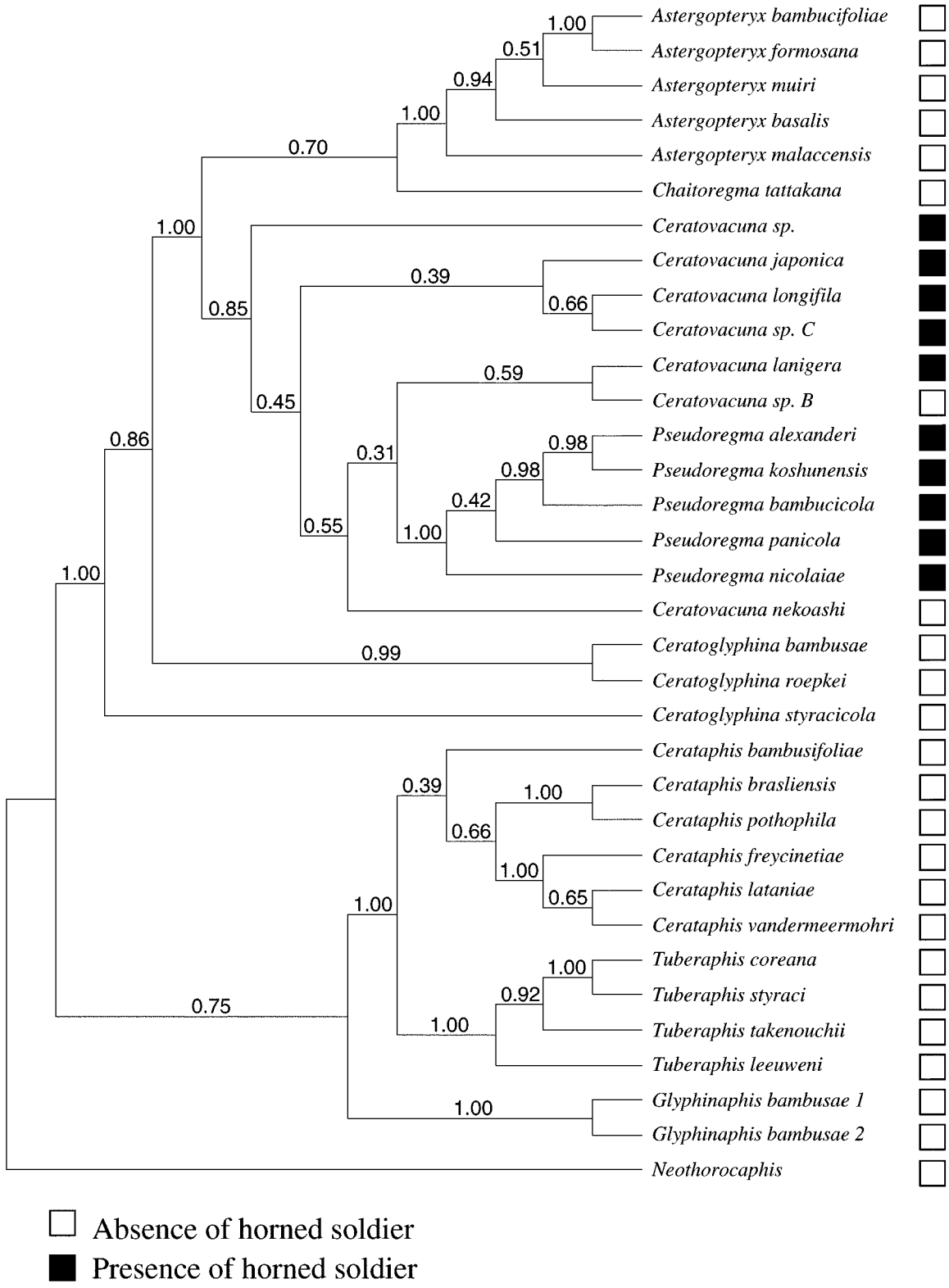


FIGURE 6. The majority rule consensus tree of the 8,000 post-burn-in trees sampled by the Markov chain for the aphid data. The numbers at the interior branches indicate the posterior probability of the clade. The distribution of aphid species with horned soldiers (solid squares) and without horned soldiers (open squares) are shown along the tips of the tree.

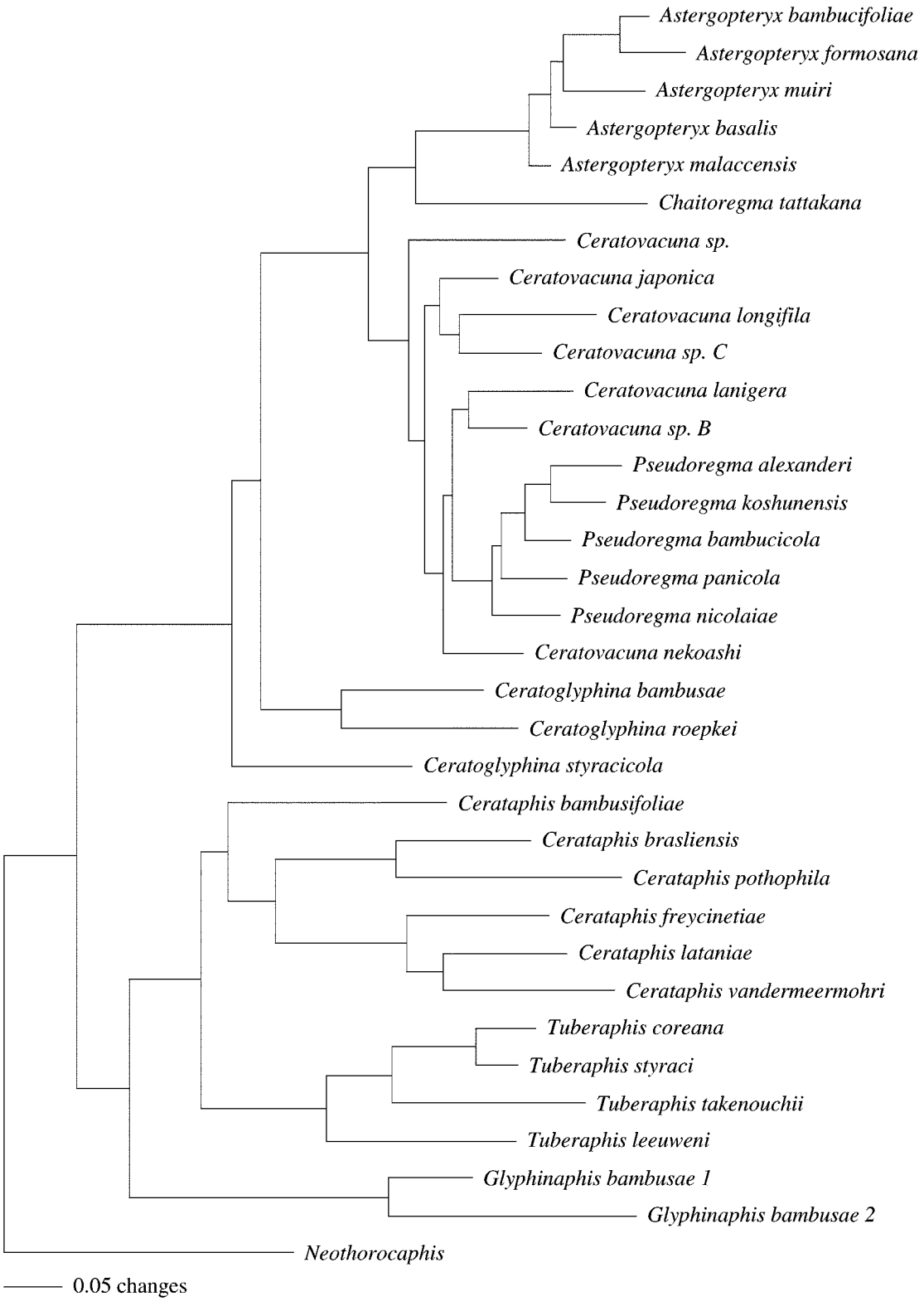


FIGURE 7. A phylogram of the majority rule consensus tree of the 8,000 post-burn-in trees sampled by the Markov chain for the aphid data. The lengths of the branches are proportional to the mean of the posterior probability density for each branch. The bar indicates 0.05 expected substitutions per site.

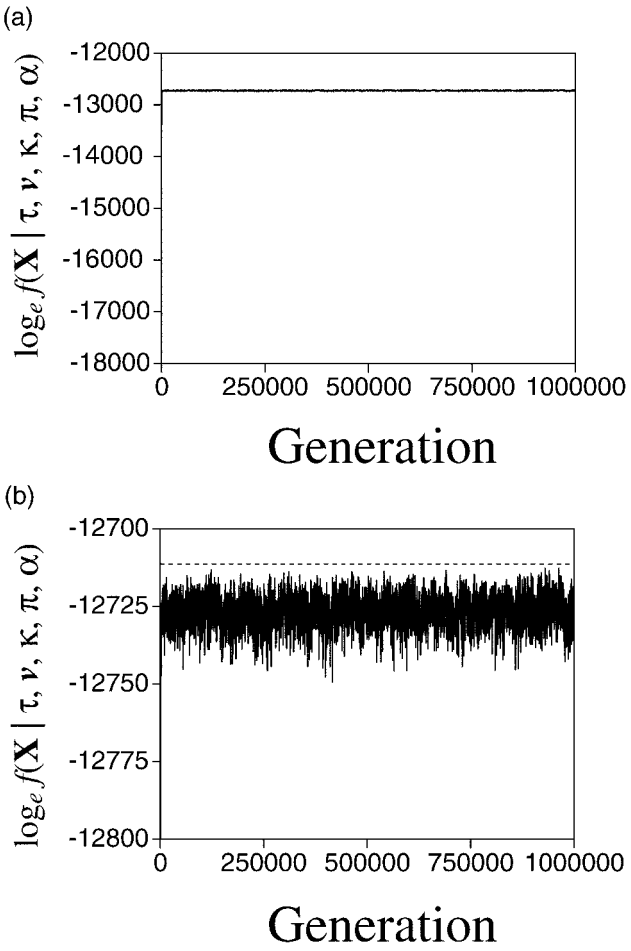


FIGURE 8. The log likelihood of the states visited by the Markov chain for the seastar mitochondrial sequences. The likelihood is proportional to the probability of observing the alignment of sequences ($\boldsymbol{X}$) conditional on the tree (τ), branch lengths (ν), transition/transversion rate ratio (κ), base frequencies (π), and gamma shape parameter (α). (a) Log likelihoods for all states visited by the chain, including the early states visited by the chain before apparent stationarity was reached. (b) Log likelihoods of states visited when the chain was near stationarity. The dashed line indicates the maximum likelihood under the HKY85 + Γ_4 model.

mappings is larger for the high-rate prior (see Table 5). The 95% credible set of reconstructions was obtained by ordering the number of gains/losses from the highest posterior probability to the lowest posterior probability. Character mappings were included in the credible set, starting from the highest posterior probability mapping, until the cumulative probability was 0.95.

Prior probability densities were placed on the two parameters of the stochastic model of character change for the aphid data. The first prior was a gamma prior on the tree length. The other prior was a flat beta (1,1) distribution on the rate of loss of the horned-soldier caste. In the course of mapping a character, both a gamma rate category and a beta category were chosen from the posterior probability distribution of rates and transformation bias, respectively. Figure 15 shows the prior and posterior probability distribution of choosing a rate and bias category. For the low-rate prior on tree length, the posterior probability was centered on a few of the highest rate categories, whereas for the high-rate prior, the posterior probability was more evenly spread among the 50 tree-length categories. Thus, there is some information in the morphological data about the rate of character transformation. We only considered an uninformative prior for the bias in rate of gain/loss. The posterior probability distribution for the bias parameter did not change if a low- or high-rate prior was placed on tree length. As expected, the posterior probability distribution for the bias parameter is consistent with a higher rate of loss than of gain of the horned-soldier caste.

For the seastars, the reconstructions with the highest posterior probability had four gains and no losses of the larval feeding trait (Table 6). These reconstructions were the same regardless of whether a low-rate or a high-rate prior were placed on tree length, but the 95% credible set was larger for the high-rate tree-length prior. Although the reconstructions with the highest posterior probability were highly skewed toward gains, the posterior probability was more evenly spread over reconstructions that had a large number of gains and those that had a large number of losses. Part of the reason for this can be seen in the six character mappings shown in Figure 16. A large fraction of the time, the ancestor of the group was reconstructed as having larval feeding, in which case there were more losses of the character than gains. Most of the time, however, non-feeding larval was reconstructed as being the ancestor, which forced more gains of the trait than losses.

The more even reconstruction of gains and losses resulted in a symmetrical posterior probability distribution for the bias parameter (Fig. 17). The probability

TABLE 4. Parameter estimates[a] for the three data sets under the HKY85 + Γ_4 model: V = the tree length; κ = the transition/transversion rate ratio; π_A = the frequency of nucleotide A; π_C = the frequency of nucleotide C; π_G = the frequency of nucleotide G; π_T = the frequency of nucleotide T; α = the gamma shape parameter.

Parameter	Pontederiaceae	Aphids	Seastars
V	0.31 (0.27, 0.35)	5.06 (4.11, 6.18)	5.02 (3.96, 6.10)
κ	5.59 (4.50, 6.94)	14.69 (11.19, 19.33)	10.28 (8.43, 12.34)
π_A	0.291 (0.272, 0.311)	0.411 (0.384, 0.437)	0.356 (0.339, 0.372)
π_C	0.175 (0.159, 0.191)	0.107 (0.098, 0.118)	0.262 (0.249, 0.274)
π_G	0.211 (0.195, 0.228)	0.017 (0.014, 0.021)	0.099 (0.091, 0.106)
π_T	0.323 (0.304, 0.344)	0.464 (0.441, 0.487)	0.283 (0.270, 0.296)
α	0.047 (0.002, 0.111)	0.206 (0.183, 0.231)	0.174 (0.159, 0.189)

[a]Mean (95% credible interval).

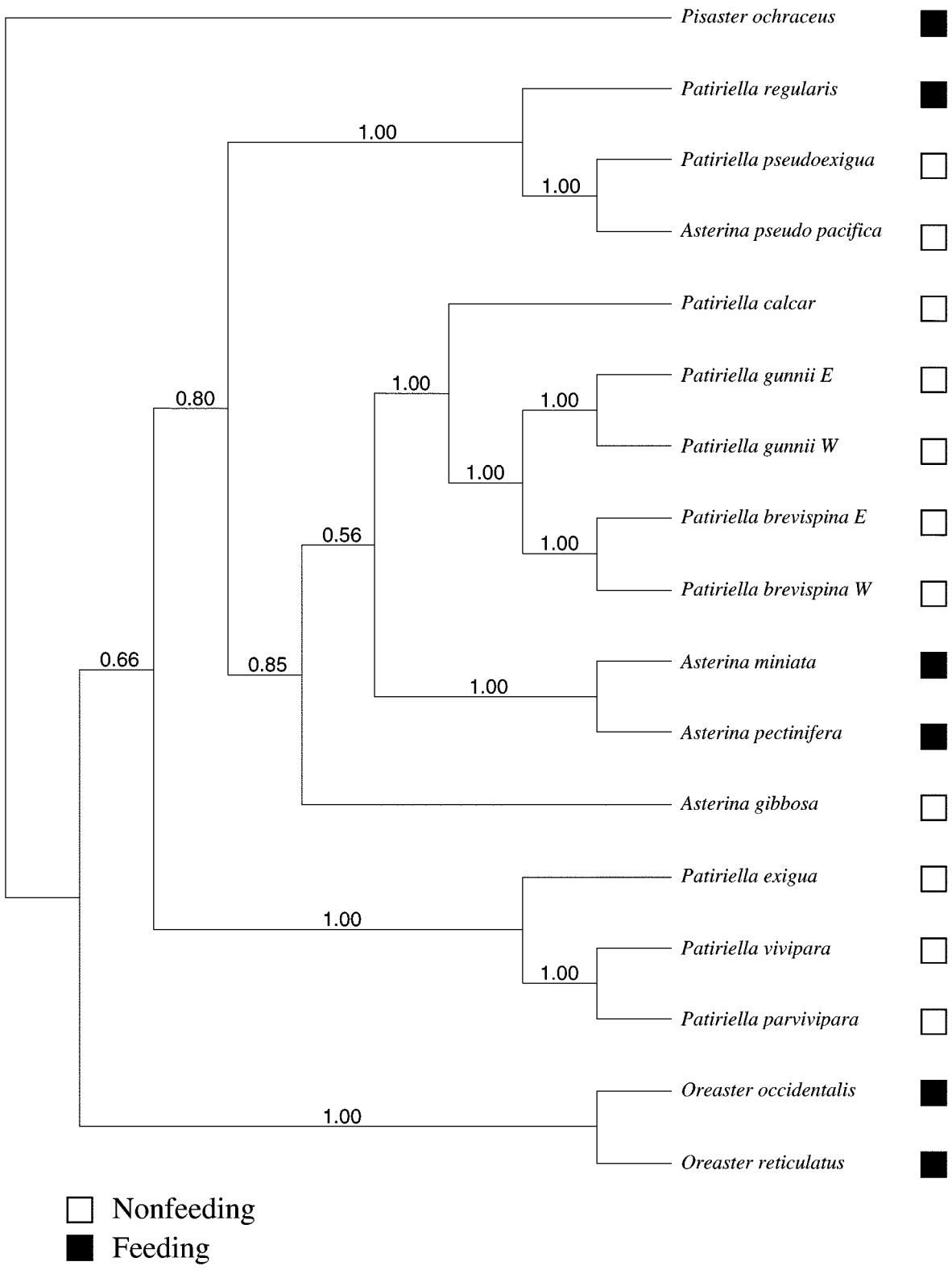


FIGURE 9. The majority rule consensus tree of the 8,000 post-burn-in trees sampled by the Markov chain for the seastar data. The numbers at the interior branches indicate the posterior probability of the clade. The distribution of seastar species with larval feeding (solid squares) and without larval feeding (open squares) are shown along the tips of the tree.

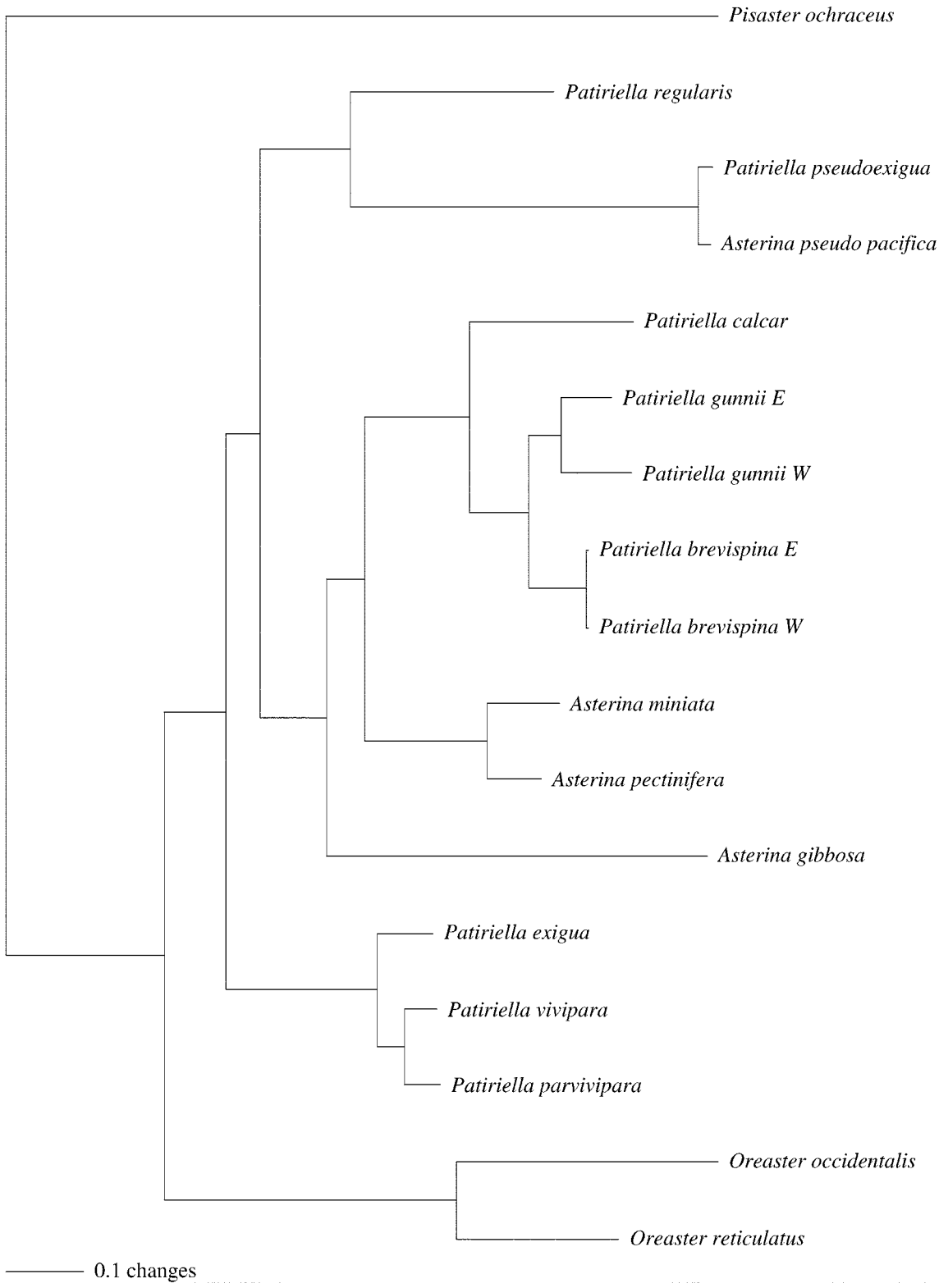


FIGURE 10. A phylogram of the majority rule consensus tree of the 8,000 post-burn-in trees sampled by the Markov chain for the seastar data. The lengths of the branches are proportional to the mean of the posterior probability density for each branch. The bar indicates 0.1 expected substitutions per site.

TABLE 5. The probability distribution for the number of gains and losses for the aphid data. The maximum posterior probability estimates of the number of gains and losses are underlined. The 95% credible set of reconstructions is indicated by the bold numbers. Results for the low-rate prior on the tree length: $E(T) = 1$, $\mathrm{SD}(T) = 5$. Results for the high-rate prior on the tree length: $E(T) = 10$, $\mathrm{SD}(T) = 10$.

No. gains	No. losses										
	0	1	2	3	4	5	6	7	8	9	10
Low-rate prior											
0	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.000	0.000
1	0.000	**0.050**	**0.292**	**0.071**	**0.035**	**0.004**	**0.003**	0.001	0.001	0.001	0.000
2	**0.003**	**0.013**	**0.019**	**0.112**	**0.042**	**0.030**	**0.005**	**0.004**	0.001	0.001	0.001
3	**0.003**	**0.002**	**0.003**	**0.008**	**0.054**	**0.029**	**0.022**	**0.006**	**0.002**	0.001	0.001
4	**0.003**	0.000	0.001	0.001	**0.002**	**0.028**	**0.017**	**0.012**	**0.003**	**0.002**	0.001
5	0.001	0.001	0.000	0.000	0.001	**0.002**	**0.018**	**0.010**	**0.009**	0.001	**0.002**
6	0.000	0.000	0.000	0.000	0.000	0.000	0.001	**0.009**	**0.007**	**0.004**	0.001
7	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	**0.003**	**0.003**	**0.002**
8	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	**0.002**	**0.002**
9	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001
10	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
High-rate prior											
0	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	**0.001**	0.000	0.000
1	0.000	**0.036**	**0.213**	**0.057**	**0.036**	**0.003**	**0.004**	**0.002**	**0.001**	0.000	0.000
2	**0.002**	**0.010**	**0.019**	**0.110**	**0.040**	**0.035**	**0.006**	**0.005**	**0.003**	**0.001**	0.000
3	**0.002**	**0.002**	**0.005**	**0.010**	**0.058**	**0.026**	**0.023**	**0.005**	**0.003**	**0.002**	**0.001**
4	**0.001**	**0.001**	**0.001**	**0.002**	**0.006**	**0.034**	**0.020**	**0.016**	**0.003**	**0.003**	**0.001**
5	0.000	**0.001**	**0.001**	**0.001**	**0.001**	**0.003**	**0.020**	**0.014**	**0.012**	**0.003**	**0.003**
6	0.000	0.000	0.000	0.000	0.000	0.000	**0.001**	**0.013**	**0.008**	**0.008**	**0.002**
7	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	**0.008**	**0.006**	**0.006**
8	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	**0.001**	**0.004**	**0.005**
9	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	**0.001**	**0.004**
10	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

TABLE 6. The probability distribution for the number of gains and losses for the seastar data. The maximum posterior probability estimates of the number of gains and losses are underlined. The 95% credible set of reconstructions is indicated by the bold numbers. Results for the low-rate prior on the tree length: $E(T) = 1$, $\mathrm{SD}(T) = 5$; 5% of the posterior probability lies outside of the table. Results for the high-rate prior on the tree length: $E(T) = 10$, $\mathrm{SD}(T) = 10$; 4% of the posterior probability lies outside of the table.

No. gains	No. losses												
	0	1	2	3	4	5	6	7	8	9	10	11	12
Low-rate prior													
0	0.000	0.000	0.000	**0.010**	**0.037**	**0.006**	0.000	0.000	0.000	0.000	0.000	0.000	0.000
1	0.000	0.000	0.000	**0.005**	**0.008**	**0.037**	**0.005**	0.000	0.000	0.000	0.000	0.000	0.000
2	0.000	**0.045**	**0.007**	**0.003**	**0.004**	**0.006**	**0.022**	**0.005**	0.000	0.000	0.000	0.000	0.000
3	0.000	**0.002**	**0.025**	**0.005**	**0.003**	**0.004**	**0.007**	**0.014**	**0.002**	0.000	0.000	0.000	0.000
4	**0.202**	**0.027**	**0.005**	**0.012**	**0.006**	**0.003**	**0.003**	**0.005**	**0.008**	0.001	0.000	0.000	0.000
5	**0.038**	**0.099**	**0.024**	**0.004**	**0.008**	**0.004**	**0.003**	**0.002**	**0.003**	**0.004**	0.001	0.000	0.000
6	**0.002**	**0.022**	**0.050**	**0.013**	**0.004**	**0.005**	**0.003**	**0.003**	**0.001**	0.001	0.001	0.000	0.000
7	0.000	**0.001**	**0.011**	**0.028**	**0.008**	**0.003**	**0.003**	**0.001**	0.000	0.000	0.001	0.001	0.000
8	0.000	0.000	**0.001**	**0.008**	**0.012**	**0.004**	**0.001**	**0.002**	0.000	0.000	0.000	0.001	0.000
9	0.000	0.000	0.000	0.000	**0.003**	**0.006**	**0.002**	**0.002**	0.001	0.000	0.000	0.000	0.000
10	0.000	0.000	0.000	0.000	0.000	**0.001**	**0.002**	**0.001**	0.000	0.001	0.001	0.000	0.001
11	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.001	0.000	0.000	0.000	0.000	0.000
12	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	**0.001**	0.001	0.001	0.001	0.000
High-rate prior													
0	0.000	0.000	0.000	**0.006**	**0.028**	**0.004**	0.000	0.000	0.000	0.000	0.000	0.000	0.000
1	0.000	0.000	0.000	**0.003**	**0.009**	**0.033**	**0.006**	0.001	0.000	0.000	0.000	0.000	0.000
2	0.000	**0.026**	**0.006**	**0.002**	**0.005**	**0.009**	**0.025**	**0.004**	0.001	0.000	0.000	0.000	0.000
3	0.000	0.000	**0.019**	**0.006**	**0.004**	**0.004**	**0.008**	**0.015**	**0.003**	0.001	0.000	0.000	0.000
4	**0.150**	**0.024**	**0.005**	**0.013**	**0.007**	**0.004**	**0.005**	**0.007**	**0.008**	**0.002**	0.000	0.000	0.000
5	**0.029**	**0.095**	**0.023**	**0.006**	**0.008**	**0.004**	**0.004**	**0.004**	**0.004**	**0.007**	**0.002**	0.000	0.000
6	0.001	**0.023**	**0.055**	**0.015**	**0.005**	**0.005**	**0.005**	**0.003**	**0.003**	**0.003**	**0.004**	**0.002**	0.000
7	0.000	0.001	**0.016**	**0.033**	**0.011**	**0.005**	**0.003**	**0.003**	**0.003**	**0.002**	**0.003**	**0.002**	0.001
8	0.000	0.000	0.000	**0.008**	**0.017**	**0.007**	**0.004**	**0.002**	**0.002**	**0.002**	**0.002**	**0.002**	**0.002**
9	0.000	0.000	0.000	0.001	**0.006**	**0.011**	**0.005**	**0.003**	**0.002**	**0.002**	0.001	**0.002**	0.001
10	0.000	0.000	0.000	0.000	0.000	**0.004**	**0.006**	**0.002**	**0.002**	**0.001**	0.001	0.001	0.001
11	0.000	0.000	0.000	0.000	0.000	0.000	**0.002**	**0.003**	**0.003**	**0.001**	0.001	0.001	0.001
12	0.000	0.000	0.000	0.000	0.000	0.000	0.000	**0.001**	**0.002**	**0.002**	0.001	0.000	0.001

TABLE 7. The values for the test statistics for the coincidence of floral morphologies with self-incompatibility. The results of the analysis were largely robust to a range of priors for the tree lengths for character 1 (T_1) and character 2 (T_2).

Rate Prior				Statistic[a]						
$E(T_1)$	$SD(T_1)$	$E(T_2)$	$SD(T_2)$	D	d■●	d■○	d▩●	d■○	d□●	d□○
1	1	1	1	0.219	0.054	−0.054	−0.031	0.031	−0.022	0.022
10	1	10	1	0.242	0.058	−0.058	−0.040	0.040	−0.018	0.018
10	50	10	50	0.205	0.048	−0.048	−0.033	0.033	−0.014	0.014
15	15	5	5	0.218	0.052	−0.052	−0.034	0.034	−0.017	0.017
15	5	5	5	0.225	0.054	−0.054	−0.036	0.036	−0.018	0.018
15	5	5	2	0.225	0.054	−0.054	−0.035	0.035	−0.019	0.019
20	10	20	10	0.225	0.053	−0.053	−0.038	0.038	−0.014	0.014
20	10	10	5	0.227	0.054	−0.054	−0.036	0.036	−0.017	0.017
30	20	30	20	0.212	0.049	−0.049	−0.033	0.033	−0.015	0.015
30	30	30	30	0.207	0.049	−0.049	−0.034	0.034	−0.014	0.014
50	50	50	50	0.195	0.046	−0.046	−0.031	0.031	−0.014	0.014

[a]■ = tristylous; ▩ = enantiostylous; □ = monomorphic; ● = self-incompatible; ○ = self-compatible.

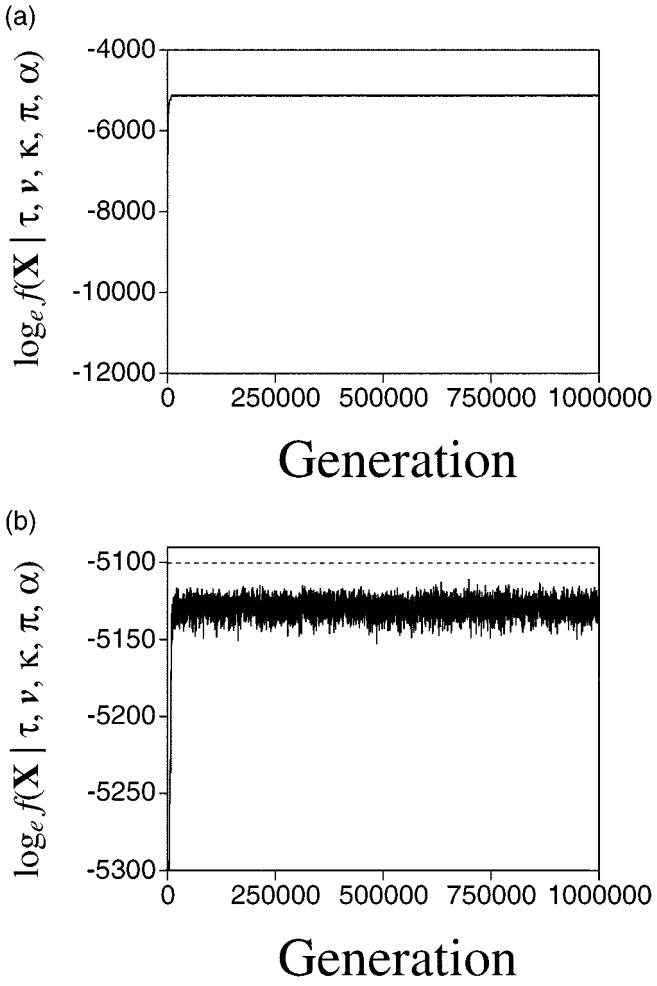


FIGURE 11. The log likelihood of the states visited by the Markov chain for the Pontederiaceae *rbcL* and *ndhF* sequences. The likelihood is proportional to the probability of observing the alignment of sequences ($\mathbf{X}$) conditional on the tree (τ), branch lengths (v), transition/transversion rate ratio (κ), base frequencies (π), and gamma shape parameter (α). (a) Log likelihoods for all states visited by the chain, including the early states visited by the chain before apparent stationarity was reached. (b) Log likelihoods of states visited when the chain was near stationarity. The dashed line indicates the maximum likelihood under the HKY85 + Γ_4 model.

distribution of the bias parameter was rather insensitive to the rate prior placed on tree length. The low-rate prior did not seem appropriate for the data because most of the posterior probability mass was placed on high-rate categories (Fig. 17).

Character Correlation

We analyzed the coincidence of the states for two characters, flower morphology and self-incompatibility, for the Pontederiaceae data (Kohn et al., 1996). The tree length for each character was assumed to follow a gamma prior distribution. We examined the robustness of the results to a number of different gamma prior distributions for both characters (Tables 7, 8). The results on covariation in the characters are largely robust to the range of priors examined in this study.

We first examined the test statistic D, our overall measure of the disagreement between the observed and expected associations of the states for the two characters. Table 7 shows the values of D for the 11 priors examined in this study; D ranged from 0.195 to 0.242. The posterior predictive P values for D also varied, from 0.832 to 0.952 (Table 8; Fig. 18). However, the D test statistic accounts for association in all of the pairwise comparisons of states and may mask strong associations between individual characters. Examination of the test statistics d_{ij} provides information on the nature of the covariation. Specifically, tristylous flowers and self-incompatibility coincide more frequently than expected by chance. The posterior predictive distributions for all of the test statistics examined here are shown in Figure 18.

DISCUSSION

The parsimony method has been the only method available for mapping characters onto a phylogenetic tree. Here, we present an application of Nielsen's (2002) method for mapping characters under continuous-time Markov models that is fundamentally different from the parsimony method in that it does not share the fundamental flaws of the parsimony method for mapping characters. For example, under continuous-time Markov

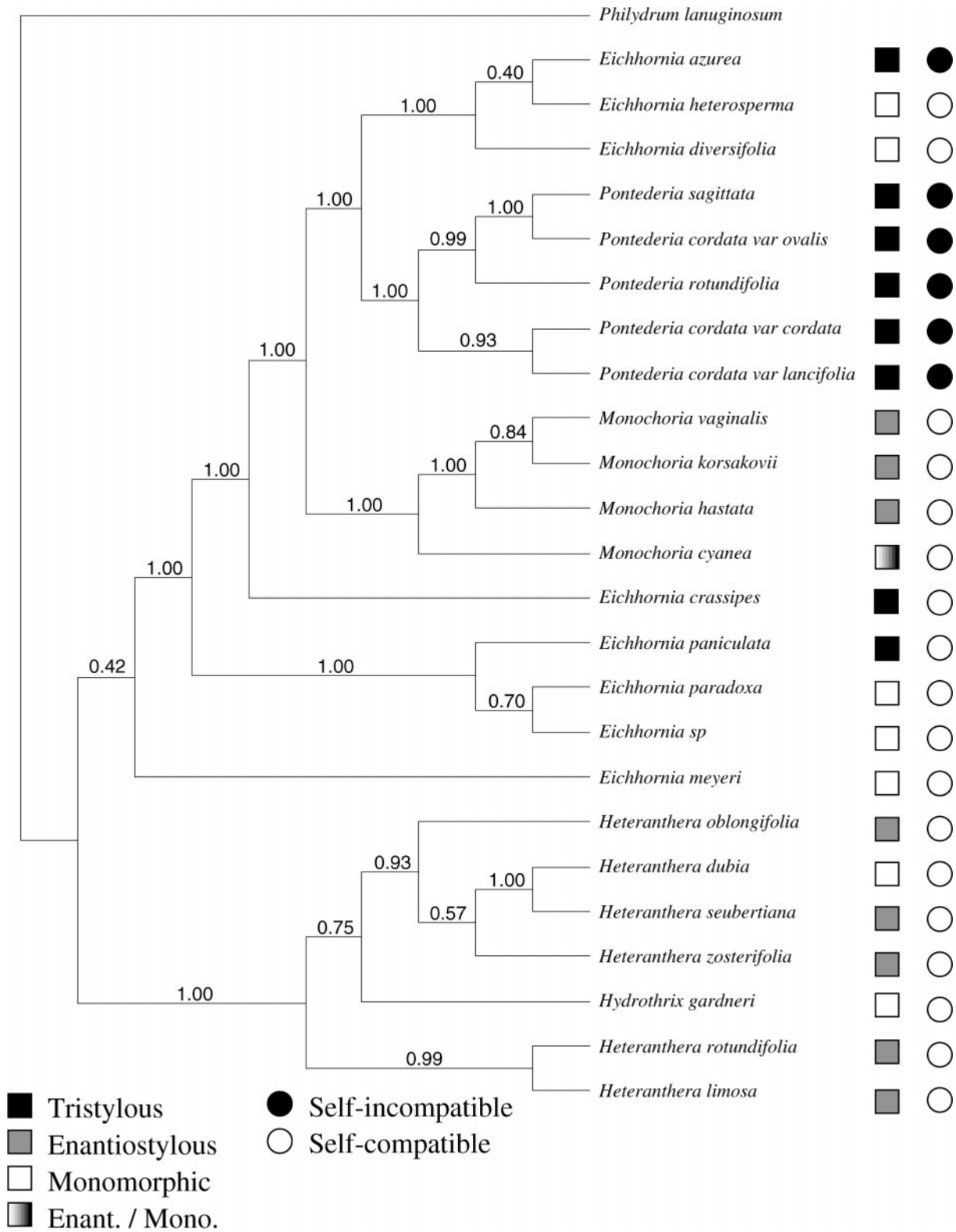


FIGURE 12. The majority rule consensus tree of the 8,000 post-burn-in trees sampled by the Markov chain for the Pontederiaceae data. The numbers at the interior branches indicate the posterior probability of the clade. The distribution of character states for floral morphology (squares) and self-incompatibility (circles) are shown along the tips of the tree.

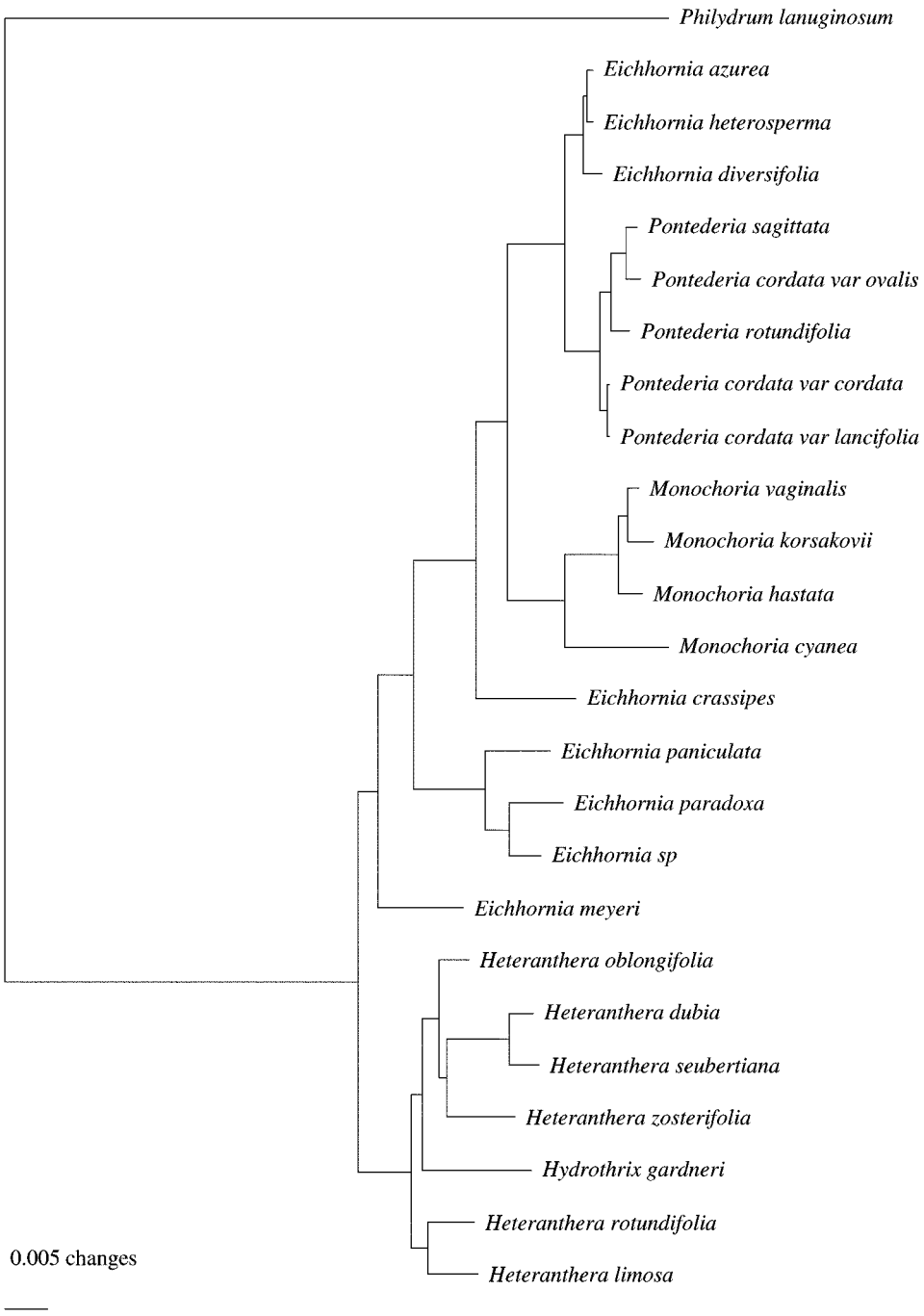


FIGURE 13. A phylogram of the majority rule consensus tree of the 8,000 post-burn-in trees sampled by the Markov chain for the Pontederiaceae data. The lengths of the branches are proportional to the mean of the posterior probability density for each branch. The bar indicates 0.005 expected substitutions per site.

models, more than a single change is allowed on a branch and the probability of a change on a branch increases with its length. For the data sets analyzed in this study, the most-parsimonious reconstructions (i.e., those involving the fewest number of changes) accounted for a relatively small fraction of the total probability distribution for number of changes (Fig. 19). This makes intuitive sense: there is always the chance that histories other than the most-parsimonious one produced the distribution of character states in the taxa. Changes in excess of the parsimonious reconstructions should not be considered ad hoc as has been argued by Farris (1983); as shown here, such changes are quite probable under simple yet reasonable models of evolution.

This application of continuous-time Markov models to morphological characters is not novel; the procedure has

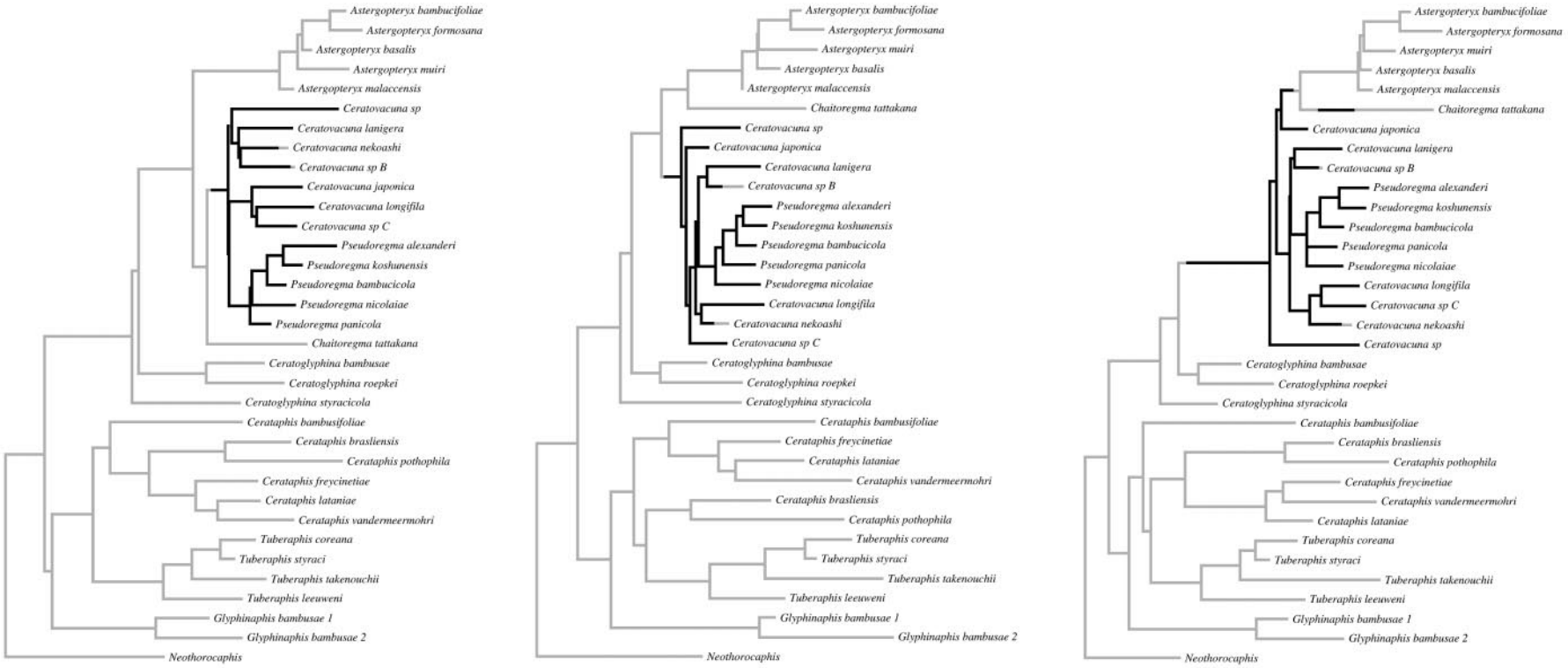


FIGURE 14. *(Continued on next page)*

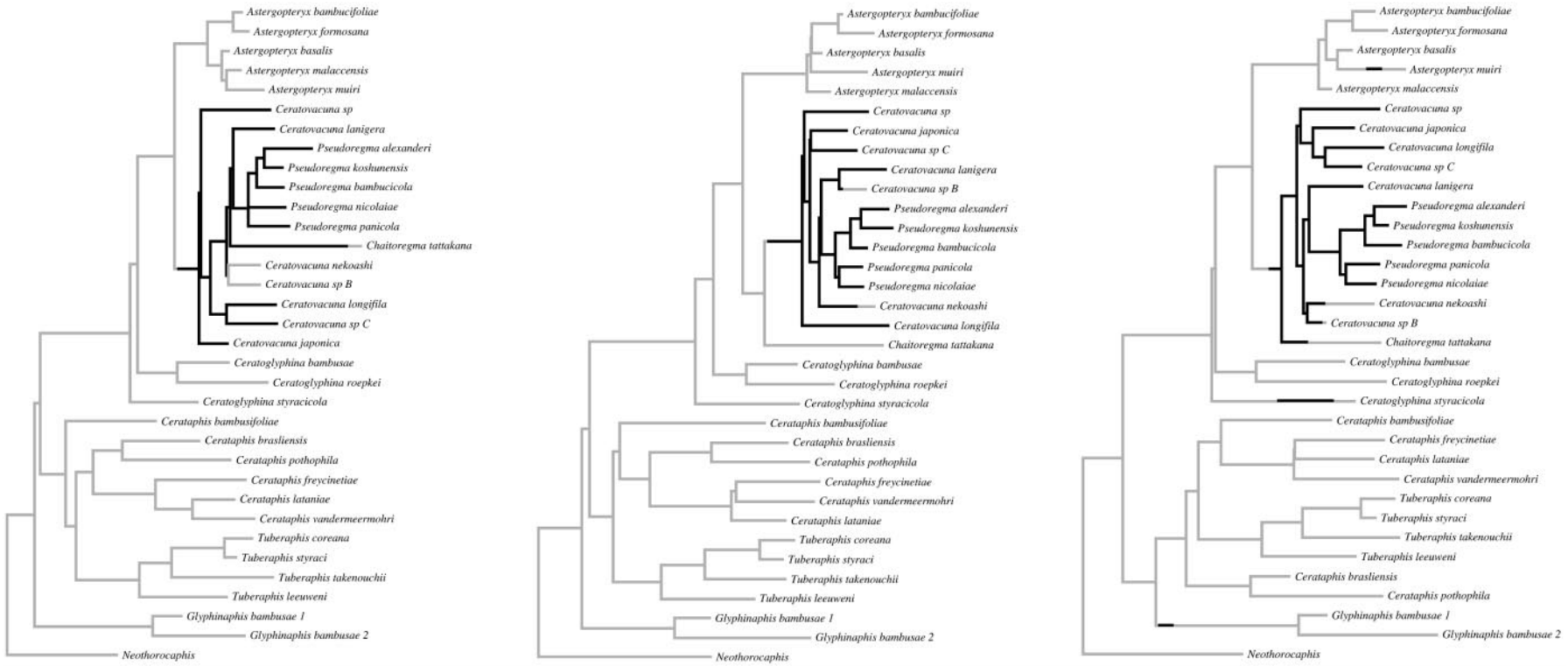


FIGURE 14. Six random character histories for the aphid data. Bold branches indicate the horned-soldiers caste. Histories were mapped under the two-state continuous-time Markov process. The trees and branch lengths were sampled by the Markov chain. Several of the character histories do not correspond to the most-parsimonious reconstruction. A gamma prior, $E(T) = 1; \mathrm{SD}(T) = 5$, was placed on the tree length (T) and a flat beta prior was placed on the bias parameter, π_0.

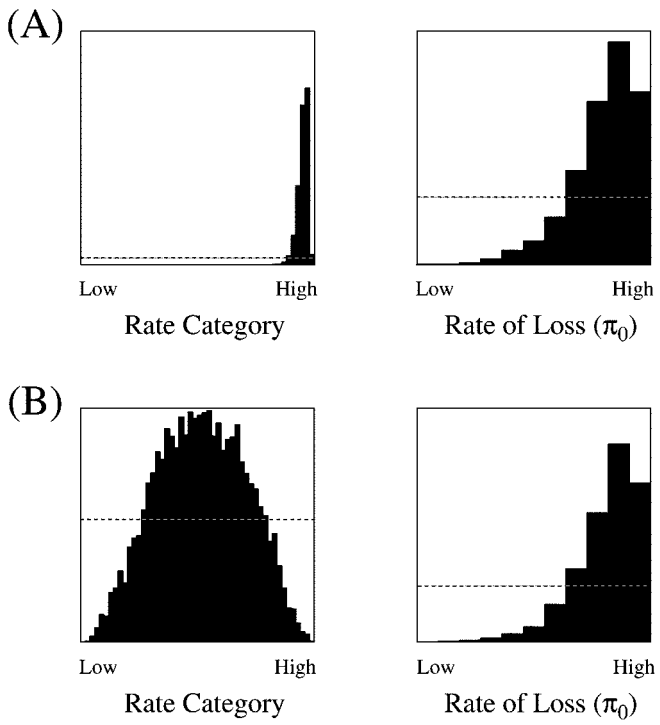


FIGURE 15. The posterior probability of being in a rate or bias category for the low- and high-rate priors on the aphid data. The gamma and beta categories were broken into 50 and 11 categories, respectively, with equal weight on each category. The prior probability distribution, indicated by the dashed line, is flat across categories. The parameters of the beta distribution resulted in a flat (uniform) prior on the bias parameter, π_0. (a) Results for the low-rate prior on the tree length: $E(T) = 1$, $\mathrm{SD}(T) = 5$. (b) Results for the high-rate prior on the tree length: $E(T) = 10$, $\mathrm{SD}(T) = 10$.

been widely applied in other contexts. Stochastic models have been used to estimate ancestral states (Schluter, 1995; Schluter et al., 1997; Cunningham, 1999; Mooers and Schluter, 1999; Pagel, 1999b; Schultz and Churchill, 1999) and to establish correlation in characters (Pagel, 1994, 1999a). More recently, Lewis (2001) argued that phylogeny should be estimated using stochastic models of morphological character change. Lewis's ideas have been implemented in computer software, and phylogeny can now be estimated using morphological characters in a Bayesian framework (Huelsenbeck and Ronquist, 2001). The method described here simply permits a more detailed study of character history for morphological characters.

We account for uncertainty in the phylogeny and other phylogenetic parameters by averaging over all possible values for the parameters, weighting each by its probability of being correct (Huelsenbeck et al., 2000). In a Bayesian analysis, the joint posterior probability distribution of all the parameters of the model is calculated. The estimate of any single parameter is based on its marginal posterior probability, which is calculated by integrating over the uncertainty in all of the other model parameters. When mapping characters on a tree, there are numerous sources of uncertainty, and a Bayesian approach seems natural: the tree, branch lengths, substitution model parameters, and character mappings all have some degree of associated uncertainty. We calculated the joint posterior probability distribution of all the parameters and could then look at the marginal posterior probability distribution of specific parameters or summary statistics.

There are numerous ways to summarize the results of a Bayesian analysis, and we have explored only a few of these. For example, a program such as MacClade (Maddison and Maddison, 1992) could allow different stochastic character mappings to be generated and explored. The degree of homoplasy also can be summarized using the method presented here. Chang and Kim (1996) suggested a new measure of homoplasy appropriate for stochastic models of character evolution. Their homoplasy measure was "the conditional probability that, given two randomly chosen taxa share the same state, they are not [identical by descent]" (Chang and Kim, 1996:194). The states for two taxa are not identical by descent if there are any transformations in the character along the path from the most recent common ancestor of the two taxa to the taxa themselves. The method proposed by Chang and Kim (1996) was limited because it was applied only

TABLE 8. The posterior predictive P values for the coincidence of floral morphologies with self-incompatibility. The results of the analysis were largely robust to a range of priors for the tree lengths for character 1 (T_1) and character 2 (T_2).

Rate prior				P[a]						
$E(T_1)$	$\mathrm{SD}(T_1)$	$E(T_2)$	$\mathrm{SD}(T_2)$	D	d■●	d■○	d ▩●	d■○	d□●	d□○
1	1	1	1	0.922	0.972	0.027	0.043	0.956	0.050	0.949
10	1	10	1	0.853	0.966	0.033	0.066	0.933	0.145	0.854
10	50	10	50	0.922	0.987	0.012	0.014	0.985	0.087	0.912
15	15	5	5	0.847	0.970	0.029	0.068	0.931	0.123	0.876
15	5	5	5	0.832	0.969	0.030	0.071	0.928	0.133	0.866
15	5	5	2	0.810	0.952	0.047	0.072	0.927	0.162	0.837
20	10	20	10	0.903	0.985	0.014	0.028	0.971	0.150	0.849
20	10	10	5	0.854	0.976	0.023	0.061	0.938	0.124	0.875
30	20	30	20	0.948	0.992	0.007	0.017	0.982	0.068	0.931
30	30	30	30	0.932	0.989	0.010	0.022	0.977	0.095	0.904
50	50	50	50	0.952	0.993	0.006	0.016	0.983	0.059	0.940

[a] ■ = tristylous; ▩ = enantiostylous; □ = monomorphic; ● = self-incompatible; ○ = self-compatible.

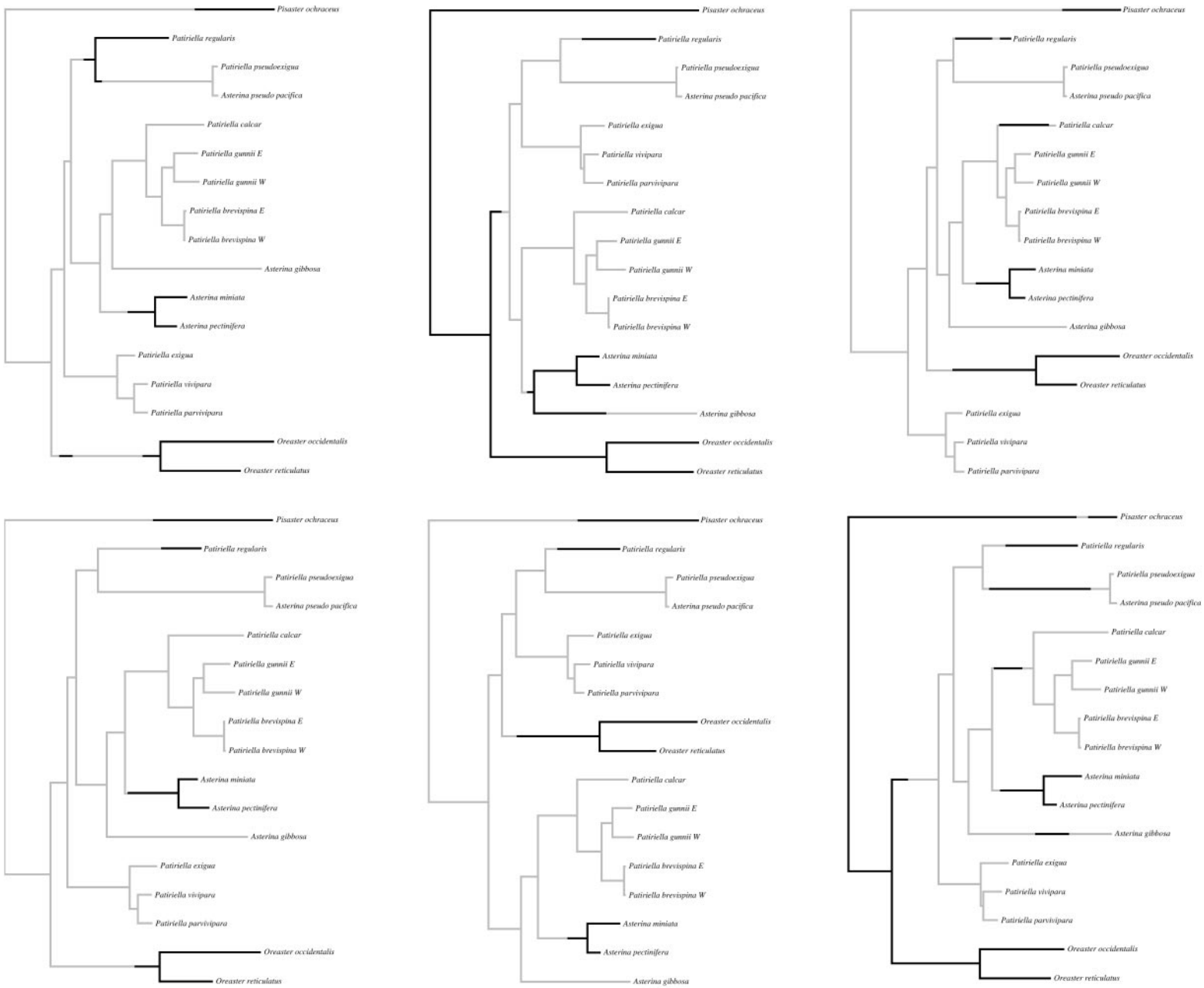


FIGURE 16. Six random character histories for the seastar data. Bold branches indicate feeding larval. Histories were mapped under the two-state continuous-time Markov process. The trees and branch lengths were sampled by the Markov chain. Several of the character histories do not correspond to the most-parsimonious reconstruction. A gamma prior, $E(T) = 1$; $\mathrm{SD}(T) = 5$, was placed on the tree length (T) and a flat beta prior was placed on the bias parameter, π_0.

TABLE 9. The probability that the states for each pair of species are not identical by descent for the Pontederiaceae flower character of Kohn et al. (1996). Only the probabilities of pairs of species that are coded the same for the character are shown. Self-comparisons, which are not relevant, are indicated by a dash. The prior for tree length had a mean of $E(T) = 1$ and an SD of 1. The homoplasy measure, H, of Chang and Kim (1996) is 0.78.

	Species																							
Species[a]	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
1	—		0.93	0.77	0.90							0.90			0.90								0.90	0.90
2		—						0.68	0.95	0.99						0.67		0.71	0.71					
3			—		0.90							0.90			0.90								0.90	0.90
4				—	0.71							0.71			0.71								0.71	0.71
5					—							0.92			0.67								0.93	0.66
6						—	0.94				0.73		0.14	0.76						0.14	0.94	0.94		
7							—				0.94		0.94	0.95						0.94	0.79	0.92		
8								—	0.90	0.99						0.68		0.35	0.36					
9									—	0.99						0.95		0.90	0.90					
10										—						1.00		0.99	0.99					
11											—		0.14	0.76						0.14	0.94	0.94		
12												—			0.67								0.92	0.66
13													—	0.22						0.72	0.94	0.94		
14														—						0.22	0.95	0.95		
15															—								0.67	0.89
16																—		0.72	0.72					
17																	—							
18																		—	0.31					
19																			—					
20																				—	0.94	0.94		
21																					—	0.79		
22																						—		
23																							—	0.66
24																								—

[a]1 = *Eichhornia azurea*; 2 = *Eichhornia heterosperma*; 3 = *Eichhornia paniculata*; 4 = *Eichhornia crassipes*; 5 = *Pontederia sagittata*; 6 = *Heteranthera oblongifolia*; 7 = *Monochoria vaginalis*; 8 = *Eichhornia meyeri*; 9 = *Hydrothrix gardneri*; 10 = *Heteranthera dubia*; 11 = *Heteranthera zosterifolia*; 12 = *Pontederia rotundifolia*; 13 = *Heteranthera rotundifolia*; 14 = *Heteranthera seubertiana*; 15 = *Pontederia cordata* var. *cordata*; 16 = *Eichhornia diversifolia*; 17 = *Monochoria cyanea*; 18 = *Eichhornia paradoxa*; 19 = *Eichhornia* sp.; 20 = *Heteranthera limosa*; 21 = *Monochoria hastata*; 22 = *Monochoria korsakovii*; 23 = *Pontederia cordata* var. *ovalis*; 24 = *Pontederia cordata* var. *lancifolia*.

to a simple two-state model. With the method proposed here, one can calculate the homoplasy measure by examining the history of character states for pairs of taxa, asking for each whether the states are identical by descent. This process could be implemented on trees sampled using the MCMC procedure, allowing the homoplasy measure to incorporate uncertainty in the tree. We used this approach for the Pontederiaceae flower morphology and self-incompatibility characters of Kohn et al. (1996). Tables 9 and 10 show the probabilities that pairs of species with the same character state are not identical by descent for the two characters. The homoplasy measure of Chang and Kim (1996) is $H = 0.78$ and $H = 0.13$ for the flower morphology and self-incompatibility characters, respectively, indicating that flower morphology is much more homoplastic than is self-incompatibility.

For many questions, explicitly mapping characters onto a phylogenetic tree under a continuous-time Markov model seems to obviate the need to examine ancestral states on phylogenies. Interest in ancestral states on phylogenies seems to be driven mostly by the desire to have a better idea of the history of a character (Cunningham, 1999). Rarely are the ancestral states of direct interest, except in a few studies (e.g., when recreating ancestral proteins for study in a laboratory; Malcolm et al., 1990; Stackhouse et al., 1990; Adey et al., 1994; Jermann et al., 1995; Chang and Donogue, 2000). Usually, the evolutionary biologist wants a picture of the history of the character on the tree. This history includes the number of changes, the type of transformation, and the timing of the transformation. Mapping characters on a

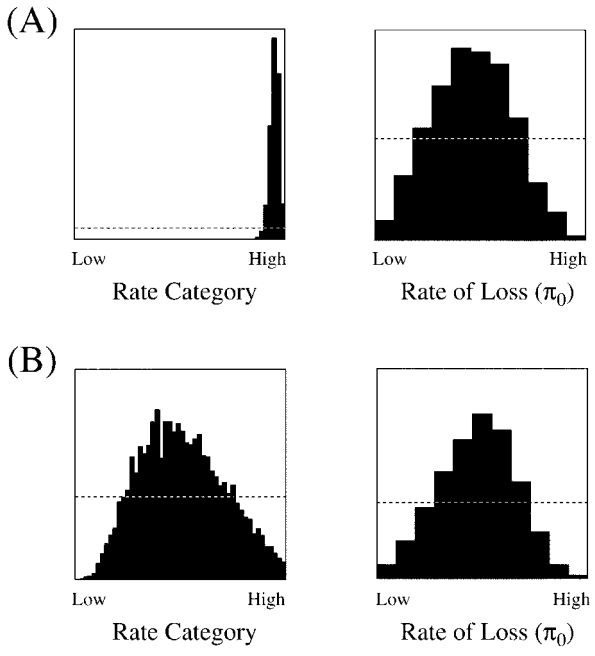


FIGURE 17. The posterior probability of being in a rate or bias category for the low- and high-rate priors on the seastar data. The gamma and beta categories were broken into 50 and 11 categories, respectively, with equal weight on each category. The prior probability distribution, indicated by the dashed line, is flat across categories. The parameters of the beta distribution resulted in a flat (uniform) prior on the bias parameter, π_0. (a) Results for the low-rate prior on the tree length: $E(T) = 1$, $\mathrm{SD}(T) = 5$. (b) Results for the high-rate prior on the tree length: $E(T) = 10$, $\mathrm{SD}(T) = 10$.

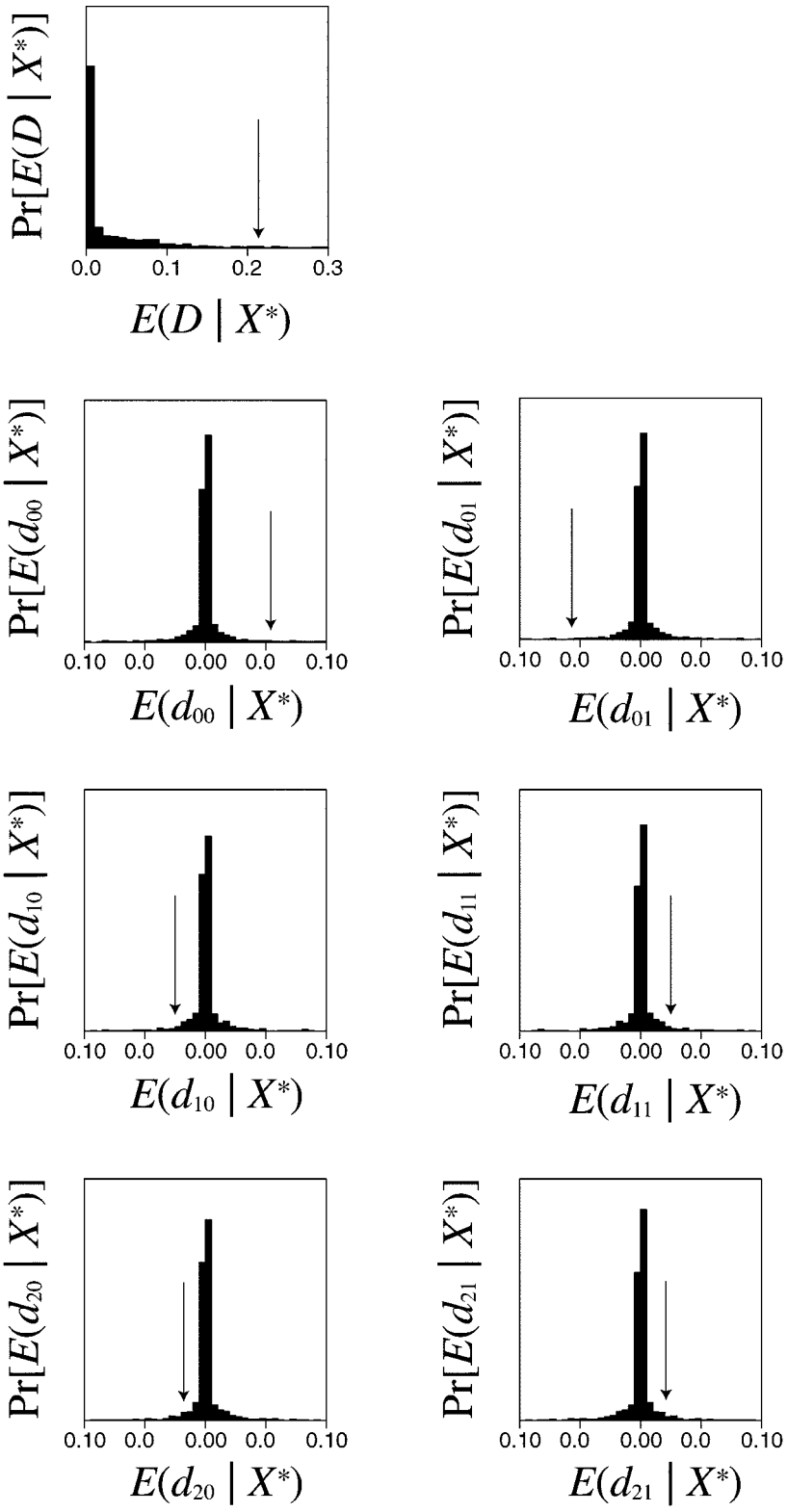


FIGURE 18. The posterior predictive probability distributions for analysis of covariation in the Kohn et al. (1996) Pontederiaceae data. In reference to Figure 12, the states are 00 = ■●; 01 = ■○; 10 = ■●; 11 = ■○; 20 = □●; and 21 = □○. The arrows indicate the observed value for the test statistics.

TABLE 10. The probability that the states for each pair of species are not identical by descent for the Pontederiaceae self-incompatibility character of Kohn et al. (1996). Only the probabilities of pairs of species that are coded the same for the character are shown. Self-comparisons, which are not relevant, are indicated by a dash. The prior for tree length had an mean of $E(T) = 1$ and an SD of 1. The homoplasy measure, H, of Chang and Kim (1996) is 0.13.

	Species																							
Species[a]	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
1	—		0.93	0.77	0.90							0.90			0.90								0.90	0.90
2		—						0.68	0.95	0.99						0.67		0.71	0.71					
3			—		0.90							0.90			0.90								0.90	0.90
4				—	0.71							0.71			0.71								0.71	0.71
5					—							0.92			0.67								0.93	0.66
6						—	0.94				0.73		0.14	0.76						0.14	0.94	0.94		
7							—				0.94		0.94	0.95						0.94	0.79	0.92		
8								—	0.90	0.99						0.68		0.35	0.36					
9									—	0.99						0.95		0.90	0.90					
10										—						1.00		0.99	0.99					
11											—		0.14	0.76						0.14	0.94	0.94		
12												—			0.67								0.92	0.66
13													—	0.22						0.72	0.94	0.94		
14														—						0.22	0.95	0.95		
15															—								0.67	0.89
16																—		0.72	0.72					
17																	—							
18																		—	0.31					
19																			—					
20																				—	0.94	0.94		
21																					—	0.79		
22																						—		
23																							—	0.66
24																								—

[a]1 = *Eichhornia azurea*; 2 = *Eichhornia heterosperma*; 3 = *Eichhornia paniculata*; 4 = *Eichhornia crassipes*; 5 = *Pontederia sagittata*; 6 = *Heteranthera oblongifolia*; 7 = *Monochoria vaginalis*; 8 = *Eichhornia meyeri*; 9 = *Hydrothrix gardneri*; 10 = *Heteranthera dubia*; 11 = *Heteranthera zosterifolia*; 12 = *Pontederia rotundifolia*; 13 = *Heteranthera rotundifolia*; 14 = *Heteranthera seubertiana*; 15 = *Pontederia cordata* var. *cordata*; 16 = *Eichhornia diversifolia*; 17 = *Monochoria cyanea*; 18 = *Eichhornia paradoxa*; 19 = *Eichhornia* sp.; 20 = *Heteranthera limosa*; 21 = *Monochoria hastata*; 22 = *Monochoria korsakovii*; 23 = *Pontederia cordata* var. *ovalis*; 24 = *Pontederia cordata* var. *lancifolia*.

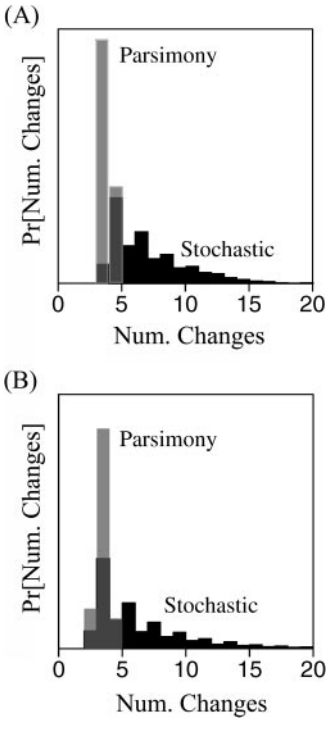


FIGURE 19. The probability distribution for the tree length under the parsimony method and the stochastic method for the seastar (a) and aphid (b) data.

tree is a direct way to visualize this history. Of course, no single realization of a character history is likely to be correct, so it makes sense to examine a large number of character histories, each drawn from the probability distribution of character mappings, and then try to summarize features that are common among the histories.

We parameterized the model of character change in a very specific way. Nucleotide sequences were used to determine the probability distribution of trees and the probability distribution of branch-length proportions. However, other parameterizations are also possible. All branch lengths could have been considered equal in length. We also assumed a model of character change in which each state can transform directly into any other. However, one could also force the characters to change in an ordered manner or allow rates to switch on or off over the history of the tree (Tuffley and Steel, 1998; Galtier, 2001).

We suggest a very simple method for determining the degree to which two characters covary, testing any covariation using posterior predictive P values. We only examined one test statistic summarizing the disagreement between the observed and expected coincidence of character states. Other possibilities for test statistics include the χ^2 statistic or measures of linkage disequilibrium. We also considered a permutation procedure for testing the observed covariation in characters. However, we did not present those results here because the results for the analysis of the Pontederiaceae were the same

when a permutation procedure was used and because the simulation procedure we used guaranteed that the characters were independent of one another, which was more in keeping with the spirit of calculating posterior predictive *P* values.

ACKNOWLEDGMENTS

We thank J. Kohn, D. Stern, and M. Hart for sending the alignments used in this study. J.P.H. was supported by NSF grants DEB-0075406 and MCB-0075404. R.N. was supported by NSF grant DEB-0089487.

REFERENCES

ADEY, N. B., T. O. TOLLEFSBOL, A. B. SPARKS, M. H. EDGELL, AND C. A. HUTCHISON. 1994. Molecular resurrection of an extinct ancestral promoter for mouse L1. Proc. Nat. Acad. Sci. USA 91:1569–1573.

CHANG, B. S., AND M. J. DONOGHUE. 2000. Recreating ancestral proteins. Trends Ecol. Evol. 15:109–114.

CHANG, J. T., AND J. KIM. 1996. The measurement of homoplasy: A stochastic view. Pages 189–203 *in* Homoplasy: The recurrence of similarity in evolution (M. J. Sanderson and L. Hufford, eds.). Adacemic Press, Chicago.

CUNNINGHAM, C. W. 1999. Some limitations of ancestral character-state reconstruction when testing evolutionary hypotheses. Syst. Biol. 48:665–674.

FARRIS, J. S. 1983. The logical basis of phylogenetic analysis. Pages 7–36 *in* Advances in Cladistics, Volume II (N. Platnick and V. Funk, eds.). Columbia Univ. Press, New York.

FELSENSTEIN, J. 1981. Evolutionary trees from DNA sequences: A maximum likelihood approach. J. Mol. Evol. 17:368–376.

FELSENSTEIN, J. 1985. Phylogenies and the comparative method. Am. Nat. 125:1–15.

GALTIER, N. 2001. Maximum-likelihood phylogenetic analysis under a covarion-like model. Mol. Biol. Evol. 18:866–873.

GRAHAM, S. W., J. R. KOHN, B. R. MORTON, J. E. ECKENWALDER, AND S. C. H. BARRETT. 1998. Phylogenetic congruence and discordance among one morphological and three molecular data sets from Pontederiaceae. Syst. Biol. 47:545–567.

HART, M. W., M. BYRNE, AND M. J. SMITH. 1997. Molecular phylogenetic analysis of life-history evolution in asterinid starfish. Evolution 51:1848– 1861.

HARVEY, P. H., AND M. PAGEL. 1991. The comparative method in evolutionary biology. Oxford Univ. Press, Oxford, U.K.

HASEGAWA, M., H. KISHINO, AND T. YANO. 1985. Dating the human–ape split by a molecular clock of mitochondrial DNA. J. Mol. Evol. 22:160–174.

HASEGAWA, M., T. YANO, AND H. KISHINO. 1984. A new molecular clock of mitochondrial DNA and the evolution of hominoids. Proc. Jpn. Acad. Ser. B 60:95–98.

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

HUELSENBECK, J. P., B. RANNALA, AND J. P. MASLY. 2000. 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.

JERMANN, T. M., J. G. OPITZ, J. STACKHOUSE, AND S. A. BENNER. 1995. Reconstructing the evolutionary history of the artiodactyl ribonuclease superfamily. Nature 374:57–59.

JUKES, T. H., AND C. R. CANTOR. 1969. Evolution of protein molecules. Pages 21–123 *in* Mammalian protein msetabolism (H. N. Munro, ed.). Academic Press, New York.

KIMURA, M. 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J. Mol. Evol. 16:111–120.

KOHN, J. R., S. W. GRAHAM, B. MORTON, J. J. DOYLE, AND S. C. H. BARRETT. 1996. Reconstruction of the evolution of reproductive characters in Pontederiaceae using phylogenetic evidence from chloroplast DNA restriction-site variation. Evolution 50:1454–1469.

LARGET, B., AND D. SIMON. 1999. Markov chain Monte Carlo algorithms for the Bayesian analysis of phylogenetic trees. Mol. Biol. Evol. 16:750–759.

LEWIS, P. O. 2001. A likelihood approach to estimating phylogeny from discrete morphological character data. Syst. Biol. 50:913–925.

LOSOS, J. B. 1994. An approach to the analysis of comparative data when a phylogeny is unavailable or incomplete. Syst. Biol. 43: 117–123.

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, W. P. 1990. A method for testing the correlated evolution of two binary characters: Are gains or losses concentrated on certain branches of a phylogenetic tree. Evolution 44:539–557.

MADDISON, W. P., AND D. R. MADDISON. 1992. MacClade: Analysis of phylogeny and character evolution, version 3.0. Sinauer, Sunderland, Massachusetts.

MALCOLM, B. A., K. P. WILSON, B. W. MATTHEWS, J. F. KIRSCH, AND A. C. WILSON. 1990. Ancestral lysozymes reconstructed, neutrality tested, and thermostability linked to hydrocarbon packing. Nature 345:86–89.

MARTINS, E. P. 1996. Conducting phylogenetic comparative studies when the phylogeny is not known. Evolution 50:12–22.

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.

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.

MOOERS, A. Ø., AND D. SCHLUTER. 1999. Support for one and two rate models of discrete trait evolution. Syst. Biol. 48:623–633.

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 (F. Seillier-Moseiwitch, T. P. Speed, and M. Waterman, eds.). Institute of Mathematical Statistics.

NIELSEN, R. 2002. Mapping mutations on phylogenies. Syst. Biol. 51:729–739.

PAGEL, M. 1994. Detecting correlated evoluton on phylogenies: A general method for the comparative analysis of discrete characters. Proc. R. Soc. Lond. B 255:37–45.

PAGEL, M. D. 1999a. Inferring the historical patterns of biological evolution. Nature 401:877–884.

PAGEL, M. 1999b. The maximum likelihood approach to reconstructing ancestral character states of discrete characters on phylogenies. Syst. Biol. 48:612–622.

RANNALA, B., AND Z. YANG. 1996. Probability distribution of molecular evolutionary trees: A new method of phylogenetic inference. J. Mol. Evol. 43:304–311.

RIDLEY, M. 1983. The explanation of organic diversity: The comparative method and adaptations for mating. Oxford Univ. Press, Oxford, UK.

SCHLUTER, D. 1995. Uncertainty in ancient phylogenies. Nature 377:108–109.

SCHLUTER, D., T. PRICE, A. Ø. MOOERS, AND D. LUDWIG. 1997. Likelihood of ancestor states in adaptive radiation. Evolution 51:1699–1711.

SCHULTZ, T. R., AND G. A. CHURCHILL. 1999. The role of subjectivity in reconstructing ancestral character states: A Bayesian approach to unknown rates, states, and transformation asymmetries. Syst. Biol. 48:651–664.

STACKHOUSE, J., S. R. PRESNELL, G. M. MCGEEHAN, K. P. NAMBIAR, AND S. A. BENNER. 1990. The ribonuclease from an ancient bovid ruminant. FEBS Lett. 262:104–106.

STERN, D. L. 1998. Phylogeny of the tribe Cerataphidini (Homoptera) and the evolution of the horned soldier aphids. Evolution 52:155–165.

SWOFFORD, D. L. 2002. PAUP*: Phylogenetic analysis using parsimony (*and other methods) version 4. Sinauer, Sunderland, Massachusetts.

TUFFLEY, C., AND M. STEEL. 1998. Modeling the covarion hypothesis of nucleotide substitution. Math. Biosci. 147:63–91.

YANG, Z. 1993. Maximum likelihood estimation of phylogeny from DNA sequences when substitution rates differ over sites. Mol. Biol. Evol. 10:1396–1401.

YANG, Z. 1994. Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: Approximate methods. J. Mol. Evol. 39:306–314.

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 10 April 2002; reviews returned 17 July 2002; final acceptance 3 December 2002

Associate Editor: Ted Schultz