

DETECTING CORRELATION BETWEEN CHARACTERS IN A COMPARATIVE ANALYSIS WITH UNCERTAIN PHYLOGENY

JOHN P. HUELSENBECK¹ AND BRUCE RANNALA²

¹*Section of Ecology, Behavior and Evolution, Division of Biological Sciences, University of California, San Diego, La Jolla, California 92093-0116*

E-mail: johnh@biomail.ucsd.edu

²*Department of Medical Genetics, University of Alberta, Edmonton, Alberta T6G 2H7, Canada*

Abstract.—The importance of accommodating the phylogenetic history of a group when performing a comparative analysis is now widely recognized. The typical approaches either assume the tree is known without error, or they base inferences on a collection of well-supported trees or on a collection of trees generated under a stochastic model of cladogenesis. However, these approaches do not adequately account for the uncertainty of phylogenetic trees in a comparative analysis, especially when data relevant to the phylogeny of a group are available. Here, we develop a method for performing comparative analyses that is based on an extension of Felsenstein's independent contrasts method. Uncertainties in the phylogeny, branch lengths, and other parameters are accommodated by averaging over all possible trees, weighting each by the probability that the tree is correct. We do this in a Bayesian framework and use Markov chain Monte Carlo to perform the high-dimensional summations and integrations required by the analysis. We illustrate the method using comparative characters sampled from *Anolis* lizards.

Key words.—Bayesian inference, comparative method, independent contrasts, Markov chain Monte Carlo, multivariate normal.

Received January 16, 2001. Accepted December 16, 2002.

Evidence of natural selection can often be detected by comparing characters across a number of species. This approach—the comparative method—remains one of the most powerful tools available to the evolutionary biologist. In fact, there is no one comparative method, and a large number of competing techniques have been devised to analyze comparative data with the goal of inferring correlation among characters (Cheverud et al. 1985; Felsenstein 1985; Grafen 1989; Maddison 1990; Harvey and Pagel 1991; Martins and Garland 1991; Janson 1992; Pagel 1992; Sanderson 1993; Baum and Donoghue 2000; also see Pagel 1999; Rohlf 2001). These methods generally assume that the phylogenetic history of the species is known without error (Martins 2000). Unfortunately, this assumption is rarely realized for actual data; examination of bootstrap proportions or posterior probabilities for individual clades on phylogenetic trees from the literature reveals that estimated trees are subject to potentially large errors. How can comparative studies be performed that accommodate uncertainty in phylogenetic trees?

A number of approaches exist that attempt to accommodate phylogenetic uncertainty in comparative analyses. One approach is to estimate the species phylogeny using several different phylogenetic methods. Inferences are based on those parts of the tree that do not vary among the trees obtained using the different methods. However, this approach is not satisfactory because phylogenetic methods are highly correlated, often making the same or similar assumptions; hence, one cannot conclude that the phylogenetic trees are accurate if identical or very similar trees are obtained using different phylogenetic methods. Moreover, the differences among the trees produced by different phylogenetic methods cannot be expected to reflect the relative uncertainties of the clades on the tree. Donoghue and Ackerly (1996) suggested that the sensitivity of comparative analyses be examined by performing the analysis on a set of plausible trees. Conclusions that are insensitive to the trees are then considered robust. How-

ever, they did not present any method for generating a set of plausible trees, concentrating entirely on a set of equally parsimonious trees; these do not necessarily reflect all the underlying uncertainties in the phylogeny. Felsenstein (1988; also see Richman and Price 1992) suggested that comparative analyses could be performed on the set of trees produced by a bootstrap analysis. The set of bootstrap trees (and branch lengths) should reflect the uncertainty in the phylogeny of the group (Efron et al. 1996). For large numbers of species, however, bootstrapping can be a time-consuming procedure. Moreover, as commonly implemented bootstrapping phylogenetic data only accommodates the uncertainty in the tree; uncertainties in other parameters affecting the comparative model, such as the relative branch lengths, are not accommodated.

Another approach for accommodating uncertainty in phylogeny is to average the inferences over a large number of phylogenetic trees that are produced under some stochastic process of cladogenesis (such as the birth-death process; Losos 1994, 1995; Martins 1996; Houseworth and Martins 2001). Inferences that are robust to a large range of trees are considered to be well supported. Although this approach can be expected to work well in cases where the correlation between two characters is strong, it may perform poorly in detecting correlations that are subtle or highly tree dependent (Donoghue and Ackerly 1996). In fact, such methods do not allow phylogenetic data to contribute to a comparative analysis at all. The effect of examining phylogenetic data such as aligned DNA sequences is to make some trees more likely than others. When phylogenetic data are in hand, a comparative analysis should accommodate the fact that some trees are more likely than others, even if uncertainty in the species phylogeny remains. Currently, only the bootstrap method bridges the gap between perfect knowledge and complete ignorance of the species phylogeny.

Here we present a Bayesian method for the analysis of

comparative characters that is an extension of the independent contrasts method. Independent contrasts is a particularly useful comparative method first proposed by Felsenstein (1985). The independent contrasts method infers the correlation between two continuous characters and effectively accommodates covariation caused by a common phylogenetic history. Not only is the method widely used, it was one of the first attempts to explicitly account for the phylogenetic history of the species included in a comparative analysis. Moreover, the method has been shown to perform well relative to other methods in computer simulation (when the assumptions of the method are satisfied; Martins and Garland 1991). Yet, independent contrasts, like other comparative methods, assumes that the phylogenetic tree and branch lengths are known without error. Hence, the performance of the method typically depends upon an accurate knowledge of the phylogeny. Our method directly estimates the correlation among multiple characters while integrating over uncertainty in the phylogenetic tree and branch lengths as well as over uncertainty in the model of evolution for the comparative characters.

METHODS

We are interested in inferring the correlation between two or more continuous characters while accommodating uncertainty in the species phylogeny. In this section, we describe a Bayesian approach to the problem. Specifically, we describe the data, model, inference criterion, and numerical methods that make the approach feasible.

It is worth outlining here how a Bayesian approach can be used to combine tree building and comparative analysis. Simply put, a Bayesian approach is useful because all quantities in the analysis (both the directly observable quantities, such as the observations made by the biologist, and the hidden quantities, such as the phylogenetic tree relating the species) are treated as random variables. This is not to say that the actual phylogeny relating the species is random, but rather that our knowledge of the phylogeny is uncertain, and it makes sense to then treat the phylogeny (and other hidden variables of the analysis) as uncertain quantities. Bayes's theorem is used to express uncertainty in the unobservable model parameters conditioned on the observations made:

$$\begin{aligned} & \Pr[\text{model parameters} \mid \text{observations}] \\ &= (\Pr[\text{observations} \mid \text{model parameters}] \\ & \times \Pr[\text{model parameters}]) / \Pr[\text{observations}]. \end{aligned} \tag{1}$$

In the current study, the observations include the comparative characters of interest and DNA sequence information from the same organisms that can be used to determine the uncertainty in the phylogenetic tree. We concentrate on determining the probability distribution of the correlation parameters between the different comparative characters of interest.

Data

We assume that the biologist has sampled DNA sequence and comparative data from s species, one of which is the outgroup species. The DNA sequences are contained in a matrix $\mathbf{X} = \{x_{ij}\}$, where $i = 1, 2, \dots, s, j = 1, 2, \dots, c$,

and c is the length of the DNA sequences. For example, a matrix of $s = 7$ species might be

species α	ACG ... T
species β	ACG ... T
species γ	ATG ... G
species δ	ATG ... G
species ϵ	ACG ... G
species ζ	ACG ... G
species η	ACG ... G

(2)

The nucleotide information for the s species at the i th site in the sequence is denoted \mathbf{x}_i . For example, the data at the second position in the above matrix is $\mathbf{x}_2 = (C, C, T, T, C, C, C)'$.

The comparative character information is contained in a matrix $\mathbf{Y} = \{y_{ij}\}$ where $i = 1, 2, \dots, s$ and $j = 1, 2, \dots, d$ (d is the number of [real valued] comparative characters sampled). An example matrix of $d = 3$ comparative characters for $s = 7$ species might be

species α	0.12	1.96	100.69
species β	0.11	2.01	96.34
species γ	0.09	2.71	111.04
species δ	0.19	16.94	82.22
species ϵ	0.26	1.26	156.90
species ζ	0.06	3.14	105.31
species η	0.13	6.28	127.11

(3)

The d characters sampled for the i th species are contained in the vector $\mathbf{y}_i = (y_{i1}, y_{i2}, \dots, y_{id})$. For example, the observations on the third species are $\mathbf{y}_3 = (0.09, 2.71, 111.04)$. Note that \mathbf{x}_i represents the character states observed for all of the species at the i th site in the DNA sequence, whereas \mathbf{y}_i represents the states observed for all of the comparative characters for a single species. In other words, \mathbf{x}_i represents a column of the DNA sequence matrix, whereas \mathbf{y}_i represents a row of the comparative data matrix. The comparative data are measured on a continuous scale and may include both positive and negative numbers. These data should not be normalized as such a transformation removes potential information about the relative rates of evolution of the different characters.

Phylogenetic Tree

We assume that the s species are related to one another through a bifurcating tree, τ . Figure 1 provides an example of an arbitrary tree of $s = 7$ species. The tips of the tree represent the sampled species and are labeled 1, 2, ..., s , whereas the internal nodes of the tree are labeled $s + 1, s + 2, \dots, 2s - 2$. In general, the ancestor of node k is denoted $\Psi(k)$. The ancestor of node $\Psi(2s - 2)$ is the outgroup species s and the ancestor of node s is $\Psi(s) = \emptyset$. The length of the k th branch (the branch below the k th node on the tree) is denoted v_k . The length of the branch below the outgroup node, s , is $v_s = 0$. For DNA sequences, the branch lengths are in terms of the number of substitutions per site that are expected to occur along the branch. The tree length, V , is the sum of the branch lengths ($V = v_1 + v_2 + \dots + v_{2s-2}$). The set of

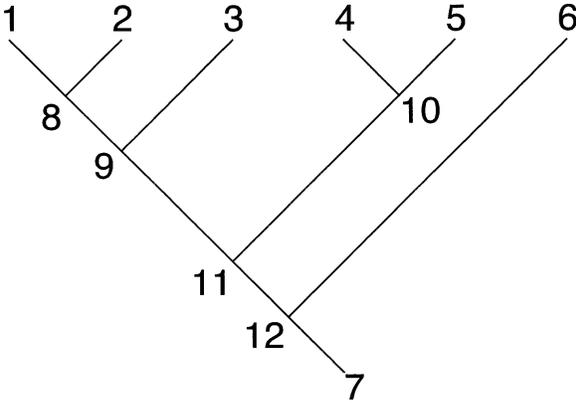


FIG. 1. An unrooted tree of $s = 7$ species.

branch lengths for a tree is denoted $\mathbf{v} = \{v_1, v_2, \dots, v_{2s-2}\}$. There are a total of

$$B(s) = \frac{(2s - 5)!}{2^{s-3}(s - 3)!}, \quad (4)$$

unrooted trees that could explain the phylogeny of the s species. Each tree is labeled with an integer $(\tau_1, \tau_2, \dots, \tau_{B(s)})$. The set of branch lengths associated with the i th tree is denoted \mathbf{v}_i .

Brownian Motion Model of Character Evolution

We assume that the comparative characters evolve according to a Brownian motion process. A character evolving according to Brownian motion makes independent moves of infinitesimal size in each infinitesimal time interval. At each instant, the process moves to the left or the right with equal probability; therefore it is not biased in terms of the direction of character evolution. Let z be the displacement of the character from its starting position after time v . For example, if $z = 0.1$, then if the character is x at time t it is (with equal probability) either $z + x$ or $z - x$ at time $t + v$. Under a Brownian motion process, the probability density of a character's displacement after time v is normally distributed with mean zero and variance $\sigma^2 v$. Here, the variance parameter, σ^2 , scales the size of the jumps made in an instant of time.

A Brownian motion model of character evolution underlies Felsenstein's (1985) independent contrasts method. Under Felsenstein's model, two characters are each considered to evolve under a Brownian motion process. For two characters undergoing (possibly correlated) Brownian motion, the probability density of the displacements of the characters from their original positions (z_1 and z_2) follows a bivariate normal distribution. If the covariance (correlation) of two or more characters is zero, they are usually considered to have evolved independently. Strictly speaking, however, although zero covariance is a necessary condition for two variables to be independent it is not a sufficient condition. Here we consider an application of Felsenstein's model that allows an arbitrary number (d) of comparative characters. Each character evolves according to a Brownian motion process and is potentially correlated with other characters. The probability density of the displacement of all d comparative characters, $\mathbf{z} = (z_1, z_2,$

$\dots, z_d)$, after time v has elapsed follows a multivariate normal distribution:

$$f(\mathbf{z} | \Sigma, v) = (2\pi)^{-d/2} |\Sigma v|^{-1/2} \exp\left[-\frac{1}{2} \mathbf{z}' (\Sigma v)^{-1} \mathbf{z}\right]. \quad (5)$$

Here, Σ is a positive definite variance-covariance matrix ($|\Sigma|$ is the determinant and Σ^{-1} is the inverse of the matrix). The diagonal entries of the variance-covariance matrix contain the d variances that scale the Brownian motion process for each comparative character and the off-diagonal entries contain the covariances:

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \rho_{12}\sigma_1\sigma_2 & \rho_{13}\sigma_1\sigma_3 & \cdots & \rho_{1d}\sigma_1\sigma_d \\ \rho_{12}\sigma_1\sigma_2 & \sigma_2^2 & \rho_{23}\sigma_2\sigma_3 & \cdots & \rho_{2d}\sigma_2\sigma_d \\ \rho_{13}\sigma_1\sigma_3 & \rho_{23}\sigma_2\sigma_3 & \sigma_3^2 & \cdots & \rho_{3d}\sigma_3\sigma_d \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_{1d}\sigma_1\sigma_d & \rho_{2d}\sigma_2\sigma_d & \rho_{3d}\sigma_3\sigma_d & \cdots & \sigma_d^2 \end{pmatrix}, \quad (6)$$

where ρ_{ij} is the correlation coefficient for the i th and j th characters ($-1 \leq \rho_{ij} \leq +1$; $\rho_{ij} = \rho_{ji}$). The main goal of this paper is to estimate the correlation coefficients for comparative data. We will also consider the relative variance for each character as this is informative about the evolvability of the character (its typical rate of change over time). As well, we will develop a Bayesian test of the hypotheses of complete independence ($\rho_{ij} = 0$ for all $i \neq j$) and complete dependence ($\rho_{ij} = 1$ for all $i \neq j$) of characters. The first test is analogous to a test of the hypothesis that the slope of the regression analysis of independent contrasts is zero in Felsenstein's method. Because biologists can often obtain very large numbers of characters by taking many morphological measurements on species and the number of correlation coefficients is $d(d - 1)$, which may become very large, it is useful to have a powerful method for identifying blocks of characters that are effectively completely correlated and can be reduced to a single representative in this matrix. The second test we develop is aimed at achieving this.

For the moment assume that the states of the comparative characters at the internal nodes of the tree are known. The states for the internal nodes of the tree will be contained in a matrix \mathbf{W} . This applies to all of the internal nodes except for the root node of the ingroup subtree (node $2s - 2$). The states for this node are contained in a vector μ . The d characters for the i th internal node ($i = s + 1, \dots, 2s - 3$) are contained in the vector $\mathbf{w}_i = (w_{i1}, w_{i2}, \dots, w_{id})$. The probability of observing the comparative characters at the tips of the tree for the ingroup species is then

$$f(Y | \tau_i, \mathbf{v}_i, \Sigma, \mathbf{W}, \mu) = f(\mu) \left[\prod_{j=1}^{s-1} f(y'_j - w'_{\psi(j)} | \Sigma, v_j/V_I) \right] \times \left[\prod_{j=s+1}^{2s-3} f(w'_j - w'_{\psi(j)} | \Sigma, v_j/V_I) \right]. \quad (7)$$

Note that the lengths of the branches are scaled by the tree length for the ingroup species, V_I . The prior probability of the states of the comparative characters at the root of the ingroup portion of the tree is $f(\mu)$. Of course, the states of the comparative characters at the internal nodes of the tree are generally unknown. Hence, the probability of observing

the comparative characters at the tips of the tree is a weighted average over all possible assignments of states to the interior nodes of the tree:

$$f(Y|\tau_i, v_i, \Sigma, \mu) = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} f(Y|\tau_i, v_i, \Sigma, W) d\mathbf{w}_{s+1} \dots d\mathbf{w}_{2s-3}. \quad (8)$$

The prior probability of the states at the root of the ingroup subtree weights each possible configuration of unobserved states.

Model of DNA Substitution

As is typical in phylogenetic analysis of DNA sequences, we assume that substitutions occur according to a continuous-time Markov chain. The instantaneous rates of change from one nucleotide to another are contained in the matrix \mathbf{Q} :

$$\mathbf{Q} = \{q_{ij}\} = \begin{pmatrix} - & \pi_C & \kappa\pi_G & \pi_T \\ \pi_A & - & \pi_G & \kappa\pi_T \\ \kappa\pi_A & \pi_C & - & \pi_T \\ \pi_A & \kappa\pi_C & \pi_G & - \end{pmatrix} \quad (9)$$

and corresponds to the rate matrix for the model of Hasegawa et al. (1984, 1985). The nucleotides are in the order A, C, G, and T. The matrix gives the instantaneous rate of change from nucleotide i (row) to nucleotide j (column). The stationary base frequencies are $\pi = (\pi_A, \pi_C, \pi_G, \pi_T)$. The transition/transversion rate bias, κ , allows transitions to occur at a different rate than transversions (when $\kappa > 1$, transitions occur at a higher rate than transversions). The parameters of the substitution process, π and κ , are contained in the vector $\Theta = (\pi, \kappa)$. The matrix \mathbf{Q} is rescaled such that $-\sum_{i=1}^4 \pi_i q_{ii} = 1$; this rescaling allows the lengths of the branches on the tree to be interpreted as the number of substitutions per site that are expected to occur along the branch. The probability of a change from nucleotide i to nucleotide j over a branch of length v is calculated by exponentiating the product of the instantaneous rate matrix (\mathbf{Q}) and the branch length (v), $\mathbf{P}(v, \Theta) = \{p_{ij}(v, \Theta)\} = e^{\mathbf{Q}v}$.

The probability of observing the data at the i th site (\mathbf{x}_i) is a summation over all possible assignments of nucleotides to the ancestral nodes of the tree:

$$f(x_i|\tau_j, v_j, \theta, r) = \sum_u \pi_{y_{si}} P_{u_{\Psi(2s-2)_i} x_{si}}(v_{2s-2}r, \theta) \left[\prod_{k=1}^{s-1} P_{u_{\Psi(k)_i} x_{ki}}(v_k r, \theta) \right] \times \left[\prod_{k=s+1}^{2s-3} P_{u_{\Psi(k)_i} u_{ki}}(v_k r, \theta) \right], \quad (10)$$

where μ is a generic vector of ancestral states. Felsenstein (1981) described an efficient pruning algorithm for calculating the summation over ancestral states. As described, this model has a parameter, r , that describes the rate of substitution for the site. If $r = 1$ for all sites, then rates are not allowed to vary across the sequence. This assumption can be relaxed in a number of ways. A commonly used method for relaxing the equal rates assumption allows the rate of sub-

stitution, r , to be a random variable drawn from a gamma distribution (Yang 1993). The probability of observing the data at the i th site is then:

$$f(x_i|\tau_j, v_j, \theta, \alpha) = \int_0^{\infty} f(x_i|\tau_j, v_j, \theta, r) f(r|\alpha, \alpha) dr, \quad (11)$$

where $f(r|\alpha, \alpha)$ is the gamma density with the shape and the scale parameters of the distribution both set to α . The gamma distribution, parameterized in this manner, then has a mean of one and a variance of $1/\alpha$ ($\alpha > 0$). Typically, the above integral cannot be evaluated and an approximation, suggested by Yang (1994), is used instead; the continuous gamma distribution is broken into K categories, each with equal probability. The mean rate from the k th category, r_k , is used to represent the rate for the entire category. The probability of observing the data at the i th site is then:

$$f(x_i|\tau_j, v_j, \theta, \alpha) = \sum_{k=1}^K f(x_i|\tau_j, v_j, \theta, r_k) \frac{1}{K}. \quad (12)$$

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

$$f(X|\tau_j, v_j, \theta, \alpha) = \prod_{i=1}^c f(x_i|\tau_j, v_j, \theta, \alpha). \quad (13)$$

Bayesian Inference of Character Correlation

Our goal is to infer the variance-covariance matrix for the comparative characters while integrating over uncertainty in the phylogenetic tree and branch lengths. In a Bayesian analysis, the approach is to calculate the joint posterior probability of parameters. Inferences of specific parameters are then based upon the marginal probability of the parameter. The joint posterior probability of the parameters is obtained using Bayes's rule:

$$f(\tau_i, v_i, \theta, \alpha, \Sigma, \mu | X, Y) = [f(X|\tau_i, v_i, \theta, \alpha) f(Y|\tau_i, v_i, \Sigma, \mu | X, Y) \times f(\tau_i, v_i, \theta, \alpha) f(\Sigma, \mu)] \div \left[\sum_{j=1}^{B(s)} \int f(X|\tau_j, v_j, \theta, \alpha) f(Y|\tau_j, v_j, \Sigma, \mu | X, Y) \times f(\tau_j, v_j, \theta, \alpha) f(\Sigma, \mu) dv_j d\theta d\alpha d\Sigma d\mu \right], \quad (14)$$

where the summation is over all possible trees and the single integral represents integration over the space of $v_i, \theta, \alpha, \Sigma$, and μ . The prior probability of the tree, branch lengths, gamma shape parameter, and substitution parameters is $f(\tau_i, v_i, \theta, \alpha)$. Here, we take the priors for each parameter to be independent. Each phylogeny is equally probable ($1/B[s]$), the lengths of each branch are taken to be uniformly distributed on the interval (0, 10), the transition/transversion rate ratio, κ , is uniformly distributed on the interval (0, 50), and the base frequencies (π) follow a flat Dirichlet distribution.

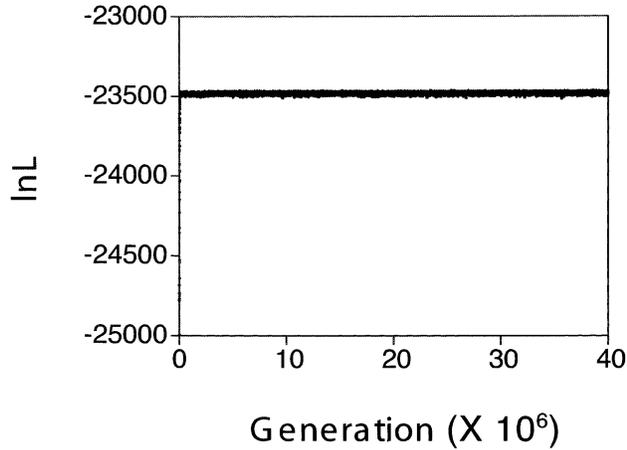


FIG. 2. The Markov chains were run for 4×10^7 generations. This plot shows the log probability of observing the DNA sequences versus time for both chains. The log likelihood plateaus after a about a million cycles. The chain was sampled every 2000 cycles, resulting in a total of 20,000 samples from each chain. The first 10,000 samples were discarded as the burn-in for each chain.

The prior for the states of the comparative characters at the root of the ingroup subtree (μ) are assumed to follow a uniform distribution with an upper and lower range that contains all reasonable values for the comparative characters (which in the following analysis are size measurements). The variance parameter, σ , is also assumed to follow a uniform distribution, with a range that contains reasonable values for the parameters. Finally, the correlation parameters, ρ_{ij} , are assumed to follow a uniform(-1,+1) distribution.

The priors, described above, are meant to quantify the biologist's beliefs about a parameter before observation of the data. The inclusion of priors that assign probabilities of assuming particular values to parameters before examining the data has the potential to be a strength or a weakness in a Bayesian analysis. Priors are advantageous when they allow the biologist to explicitly incorporate prior knowledge about a parameter. They also greatly simplify estimation when parameters are constrained to a particular range of values and maximum likelihood estimation would require a constrained optimization, whereas the Bayesian analysis can simply introduce a prior imposing the appropriate constraints. Priors can be a disadvantage if they have a subjective influence on estimates when no prior knowledge is available. One way to avoid subjectivity when no prior information is available is to employ an uninformative prior. For example, in this study we assume that all trees are a priori equally probable; this prior does not bias the analysis toward any single tree. Similarly, we place uniform priors on other parameters, such as the transition/transversion rate ratio and branch lengths, with a range that accommodates reasonable values for these parameters. When uniform priors are used (as is the case for almost all of the parameters in this study), the likelihood function mainly determines the posterior density of the parameters. In fact, as more observations are collected, the priors become less important and the posterior distribution is dominated by the likelihood function.

Markov Chain Monte Carlo

The joint posterior probability density of the parameters, formulated above, cannot be analytically solved. Hence, we resort to a simulation method to approximate the posterior probability. Specifically, we use Markov chain Monte Carlo (MCMC) to approximate the joint posterior density. We use a variant of MCMC called the Metropolis-Hastings algorithm (MH; Metropolis et al. 1953; Hastings 1970; also see Green 1995). The MH algorithm works by constructing a Markov chain that has as its stationary distribution the posterior probability distribution of interest. As an example, consider Bayesian inference of a generic parameter θ from a set of observations, x . The likelihood function is $l(\theta) = cf(x|\theta)$ and is proportional to the probability of observing the data given a specific value for the parameter, the constant of proportionality, c , being arbitrary. The posterior probability of the parameter is computed using Bayes's formula:

$$f(\theta|x) = \frac{f(x|\theta)f(\theta)}{\int_{\theta} f(x|\theta)f(\theta) d\theta}, \tag{15}$$

where integration is over the space for the parameter and the prior probability distribution of the parameter is $f(\theta)$. MCMC approximation of the posterior distribution would work as follows. First, the current state of the chain is denoted θ . If this is the first cycle of the chain, then θ is initialized using a (perhaps) arbitrarily chosen value. Second, propose a new state (θ') for the parameter using a stochastic process. The probability of proposing the new state is $f(\theta'|\theta)$. The probability of the reverse move (which is not actually made) is $f(\theta|\theta')$. Finally, calculate the probability that the new state is accepted. The probability is calculated using the formula described by Hastings (1970):

$$R = \min\left[1, \frac{f(\theta'|x)}{f(\theta|x)} \times \frac{f(\theta|\theta')}{f(\theta'|\theta)}\right]. \tag{16}$$

The first term in this equation is a ratio of the posterior probability of interest (the target distribution) and the second term is a ratio of the proposal probabilities (typically called the Hastings ratio, as Hastings was the first to allow asymmetric proposal mechanisms). As described, MCMC would not appear to allow one to approximate the posterior distribution because the ratio of the posterior probabilities contains integrals that may be impossible to evaluate. However, the denominator of Bayes's rule cancels, and the acceptance probability becomes:

$$\begin{aligned} R &= \min\left[1, \frac{f(\theta'|x)}{f(\theta|x)} \times \frac{f(\theta|\theta')}{f(\theta'|\theta)}\right] \\ &= \min\left[1, \frac{f(x|\theta')f(\theta')/f(x)}{f(x|\theta)f(\theta)/f(x)} \times \frac{f(\theta|\theta')}{f(\theta'|\theta)}\right] \\ &= \min\left[1, \frac{f(x|\theta')}{f(x|\theta)} \times \frac{f(\theta')}{f(\theta)} \times \frac{f(\theta|\theta')}{f(\theta'|\theta)}\right]. \end{aligned} \tag{17}$$

In words, the acceptance probability equals the likelihood ratio times the prior ratio times the proposal ratio. All of these quantities can be readily calculated.

In this study, we construct a Markov chain that has as its state space the parameters of the evolutionary model. We update the state of the chain by considering one parameter at a time. The details of the mechanisms for proposing new states and the acceptance probabilities are described in the Appendix.

Tests of Specific Models

Model choice in a Bayesian analysis is often guided by the Bayes factor. Consider two models, M_1 and M_2 , having posterior probabilities $f(M_1 | \mathbf{X})$ and $f(M_2 | \mathbf{X})$, respectively. The posterior probability of model 1 is obtained using Bayes's theorem as

$$f(M_1 | X) = \frac{f(X | M_1)f(M_1)}{f(X | M_1)f(M_1) + f(X | M_2)f(M_2)}, \quad (18)$$

and the posterior probability of M_2 is $1 - f(M_1 | \mathbf{X})$. The term $f(\mathbf{X} | M_i)$ is the marginal likelihood of model i and is integrated over the parameter space of the model [$f(M_i | \mathbf{X}) = \int_0 f(X | \theta)f(\theta | M_i) d\theta$]. The Bayes factor is defined as a ratio of the posterior odds of the two models to the prior odds,

$$\begin{aligned} \frac{f(M_1 | X) / f(M_1)}{f(M_2 | X) / f(M_2)} &= \frac{f(X | M_1)f(M_1) / f(M_1)}{f(X | M_2)f(M_2) / f(M_2)} = \frac{f(X | M_1)}{f(X | M_2)}. \end{aligned} \quad (19)$$

This equals the posterior odds ratio if a uniform prior is used such that $f(M_1) = f(M_2)$. The Bayes factor plays a critical role in model choice in a Bayesian analysis and is interpreted as measuring "the change in the odds in favor of the hypothesis when going from the prior to the posterior" (Lavine and Schervish 1999, p. 120). Twice the log of the Bayes factor is on roughly the same scale as the more familiar likelihood ratio.

In addition to generating the posterior density of the elements of the variance-covariance matrix, we were interested in explicitly testing the hypothesis that none of the characters are associated with one another (M_R : $\rho_{ij} = 0$ for all $i \neq j$). A test of this model is of general interest as many comparative studies are interested in testing for significant correlation between characters. Rejection of this model suggests an association for at least some pairs of characters and can motivate further study of the association. Here, we will contrast the model M_R with an alternative model (M_G) that allows the correlation coefficients to range freely on the interval $(-1, +1)$. Hence, our first model is a restriction of the general model. A second possible test imposes the constraints $p_{ij} = 1$ under M_R . This tests for complete dependence of characters.

Anolis Data

We illustrate the Bayesian method for inferring correlated evolution using data from 30 lizards (Losos et al. 1998; Jackman et al. 1999). The molecular data are an aligned matrix of $s = 30$ mitochondrial DNA sequences of the complete NADH dehydrogenase subunit II (ND2) gene and five tRNA genes (tryptophan, alanine, asparagine, cysteine, and tyro-

TABLE 1. Estimates of the parameters of the substitution model. The mean of the posterior distribution and the interval containing 95% of the posterior density (credible interval) are shown κ , transition/transversion rate ratio; π_i , frequency of the i th nucleotide.

Parameter	Mean	Credible interval
κ	2.863	(2.696, 3.044)
π_A	0.283	(0.270, 0.296)
π_C	0.271	(0.259, 0.283)
π_G	0.115	(0.108, 0.123)
π_T	0.331	(0.319, 0.344)

sine; Jackman et al. 1999). The aligned matrix is $c = 1456$ sites in length. The outgroup species is *Diplolaemus darwini* and the ingroup consists of 29 lizards of the genus *Anolis*. The comparative data were collected by Losos et al. (1998). The snout-vent length, mass, forelimb length, hindlimb length, tail length, and number of toe lamellae were collected for a sample of the lizards. Our analysis is performed on the mean (untransformed) value for each character.

RESULTS

We ran two Markov chains, constructed as described in the Appendix, for 4×10^7 generations each. The chains were sampled every 2000 cycles. Figure 2 shows the log probability of observing the DNA sequences through time. Note that the likelihood plateaus after about a million generations; we discarded the 10,000 samples collected in the first 2×10^7 generations as the burn-in. This is typical of MCMC analyses as inferences should be based on samples taken from the chain when at stationarity. All of our inferences are based on the 20,000 samples combined from both chains that were taken when the chains were at apparent stationarity.

Figure 3 shows the consensus tree of the 20,000 trees sampled from the chains. The numbers at the interior nodes of the tree represent the posterior probability of the clade. The posterior probability of a clade is a measure of uncertainty of the group. Many of the clades on the tree were well supported, with posterior probabilities over 0.90. However, some parts of the tree were more poorly resolved. Moreover, there was uncertainty in the branch lengths on the trees as well as in the parameters of the substitution model. Table 1 shows the mean of the posterior distribution and the interval containing 95% of the posterior density for parameters of the model of DNA substitution. Bayesians call such an interval a "credible interval," interpreting it as the probability that the true value of the parameter is inside of the interval.

Our main interest was in the correlation coefficients between all ($\#$) = 15 pairs of characters. We based inferences of the correlation coefficient between character i and j (ρ_{ij}) on the marginal posterior probability for that coefficient. Figures 4 and 5 show the marginal posterior probability distributions for the correlation parameters and the variance parameters, respectively. Table 2 shows the mean and 95% credible intervals for the 15 correlation coefficients. The credible intervals for all 15 correlation coefficients do not overlap zero, suggesting that there is an association among all of the characters. However, the association was strongest between the following pairs of characters: snout-

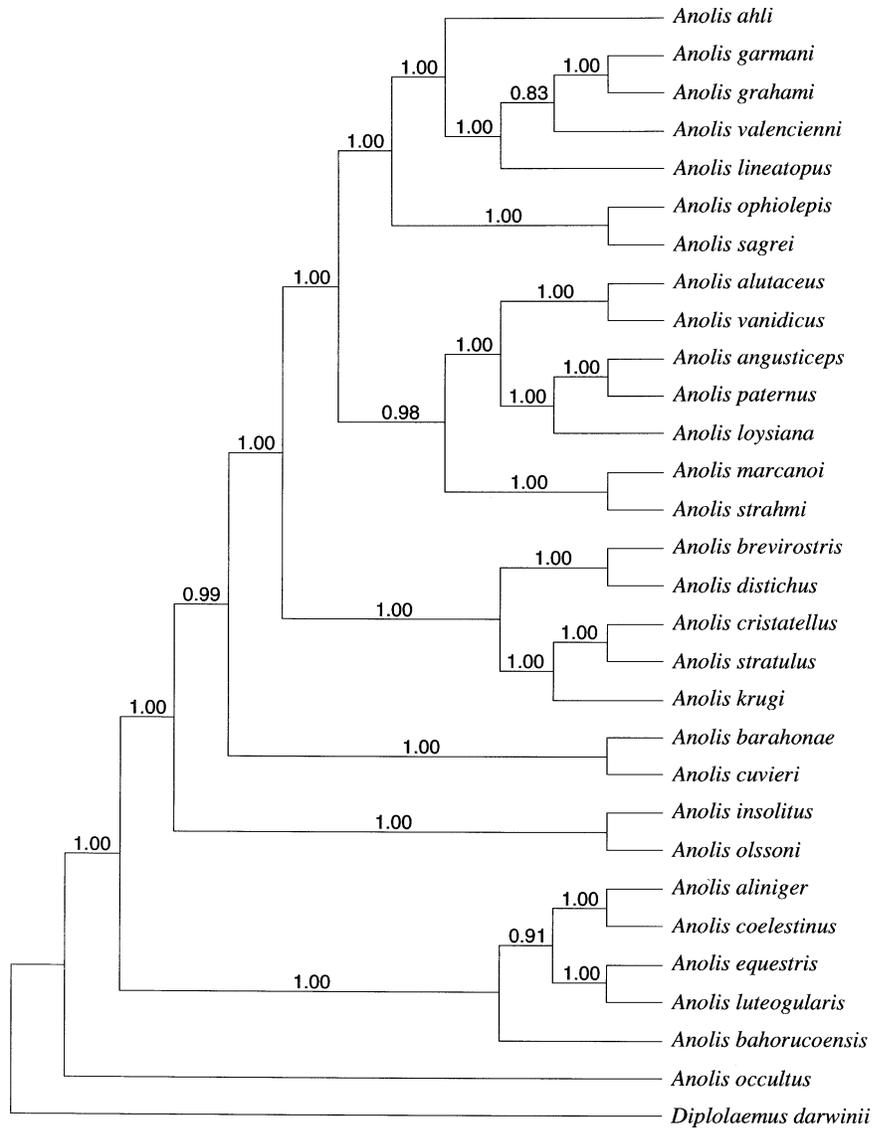


FIG. 3. A consensus tree of the 20,000 trees sampled from the chain. The numbers at the interior nodes represent the posterior probability that the clade is true.

vent length versus mass, snout-vent length versus forelimb length, hindlimb length versus tail length, mass versus forelimb length, and forelimb length versus hindlimb length. The analysis also provided information on the parameter σ_i , which scales the rate of evolution for the i th character. Table 3 shows the mean and 95% credible interval for the six scaling parameters of the Brownian motion model.

We examined two specific models for the pattern of correlation among the characters. Specifically, we examined the hypotheses M'_R : $\rho_{ij} = 0$ for all $i \neq j$ and M''_R : $\rho_{ij} = 1$ for all $i \neq j$. We contrast these specific models to a general model (M_G) in which the correlation coefficients range freely on the interval $(-1, +1)$. The Bayes factor for a comparison of a restricted point hypothesis (M_R) to a more general alternative hypothesis (M_G) can be measured by comparing the posterior density at the restriction under the general model to the prior density at the restriction:

$$B_{RG} = \frac{f(\theta = \theta_0 | X, M_G)}{f(\theta = \theta_0 | M_G)}, \quad (20)$$

where θ_0 represents the restriction of the parameters. This ratio is called the Savage-Dickey ratio (Dickey 1971). We approximated the posterior density at the restriction ($\rho_{12} = \rho_{13} = \dots = \rho_{56} = 0$) by fitting the posterior probability of the 15 correlation coefficients to a multivariate normal distribution. We then evaluated the density at $\rho_{12} = \rho_{13} = \dots = \rho_{56} = 0$. The prior density at $\rho_{12} = \rho_{13} = \dots = \rho_{56} = 0$ is simply $(\frac{1}{2})^{15} = 3.05 \times 10^{-5}$ (each of the 15 correlation coefficients have independent uniform priors on the interval $[-1, +1]$). The Bayes factor for the *Anolis* data is approximately $\log_e B_{RG} = -1.2 \times 10^8$, which is very strong evidence against the restricted model. There is a strong association among the *Anolis* morphological characters.

In a similar manner, we tested the restricted model that all of the characters are perfectly correlated (M'_R : $\rho_{ij} = 1$ for all

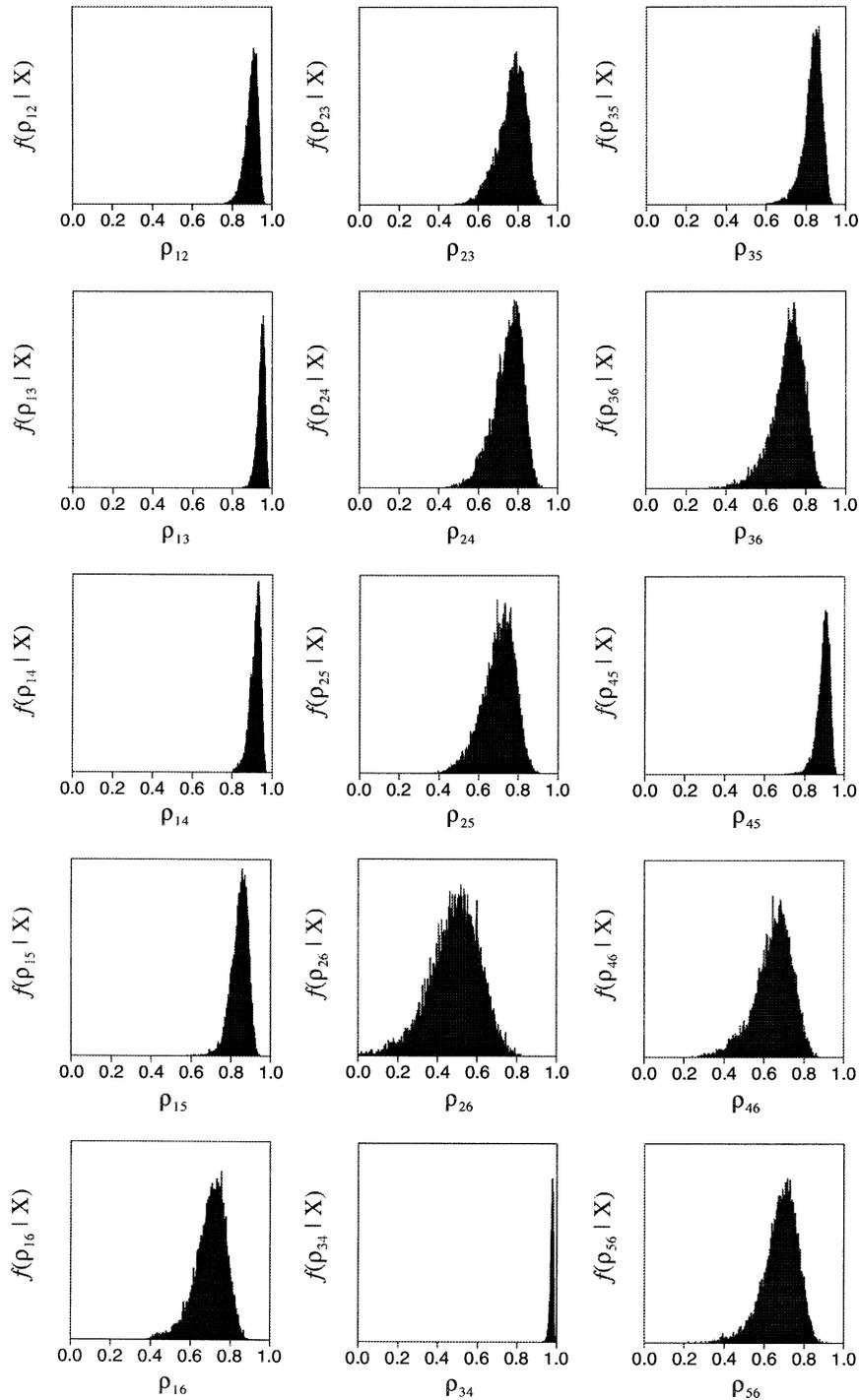


FIG. 4. The 15 character correlation parameters examined in this study. Each frequency histogram, an approximation of the continuous posterior probability density distribution, is based on a total of 20,000 points sampled from both chains.

$i \neq j$). The Bayes factor for a comparison of this model to the general (unrestricted) model is $\log_e B_{R'G} = -1.3 \times 10^5$. Although there is strong evidence that there is an association among the characters, there is also decisive evidence against the idea that the characters are perfectly correlated with one another.

DISCUSSION

Our method for inferring correlation among characters does not formally depend upon an exact knowledge of the phylogeny of a group. While it is assumed that the species are related by a phylogenetic tree, the method does not depend

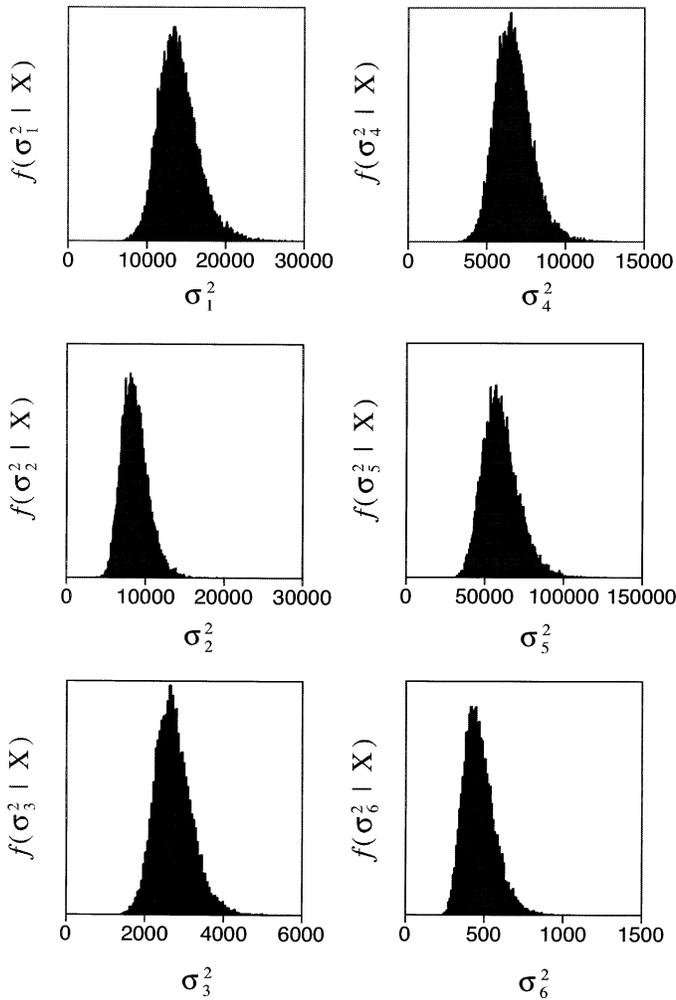


FIG. 5. The variance parameters of the Brownian motion process. Each frequency histogram is based on a total of 20,000 points sampled from both chains.

upon any single tree being correct. Instead, inferences are made on all possible trees, with the result from each tree weighted by the probability that the tree is correct. Not only is uncertainty in the phylogeny accommodated, but uncertainty in the branch lengths, substitution model parameters, and parameters of the Brownian motion process are accounted for by integrating over possible values. The analysis of the lizard data illustrated how the method can be applied to continuously varying characters and revealed strong correlation among all of the morphological features. The original analysis by Losos et al. (1998), however, examined the relationship between the morphology of the lizards and their ecological niche. The method we describe, while appropriate for examining the relationship among the morphological characters, is not appropriate without modification to examining the relationship between continuous and discrete characters.

The typical approach taken in comparative analysis has been to treat the phylogenetic tree as known. The tree is then effectively treated as an observation (pseudodata) in analyses of the comparative characters. The obvious problem with this approach is that the phylogenetic history of a group is rarely

TABLE 2. Estimates of the correlation coefficients for the six characters: (1) snout-vent length; (2) mass; (3) forelimb length; (4) hindlimb length; (5) tail length; (6) number of toe lamellae. The mean of the posterior distribution and the interval containing 95% of the posterior density (credible interval) are shown.

Parameter	Mean	Credible interval
ρ_{12}	0.897	(0.827, 0.943)
ρ_{13}	0.942	(0.898, 0.971)
ρ_{14}	0.916	(0.852, 0.955)
ρ_{15}	0.844	(0.752, 0.909)
ρ_{16}	0.701	(0.523, 0.826)
ρ_{23}	0.769	(0.618, 0.871)
ρ_{24}	0.748	(0.586, 0.859)
ρ_{25}	0.702	(0.526, 0.828)
ρ_{26}	0.489	(0.229, 0.692)
ρ_{34}	0.974	(0.955, 0.986)
ρ_{35}	0.834	(0.728, 0.903)
ρ_{36}	0.712	(0.530, 0.833)
ρ_{45}	0.896	(0.820, 0.942)
ρ_{46}	0.649	(0.439, 0.794)
ρ_{56}	0.683	(0.498, 0.817)

known with certainty. In fact, phylogenetics remains a controversial field with debate not only on the phylogenetic history of specific groups but debate about appropriate methods of analysis when different methods result in different trees. A number of methods have already been proposed to accommodate phylogenetic uncertainty in comparative analyses. One interesting approach is to perform comparative analyses on trees generated under a stochastic model of cladogenesis, such as the birth-death process, and base inferences on a summary of the results on many such trees (Losos 1994, 1995; Martins 1996). The weakness of this approach, of course, is that it fails to account for observations that are relevant to the phylogeny of the group. More often than not, some knowledge of the phylogenetic history of a group is available and a model that makes all such histories equally probable is inappropriate. However, these methods do bear an interesting relationship to the one proposed in this paper; in the absence of any information on the phylogeny of a group, the inference that would be drawn about the relationships among comparative characters by applying our method (for only two correlated characters) is essentially similar to the inferences that would result from an application of the method of Losos (1994, 1995) or that of Martins (1996). That is, inferences would be based on an equal consideration of all possible trees. The difference between the result of a com-

TABLE 3. Estimates of the parameter σ_i that scales the rate of evolution for the i th character: (1) snout-vent length; (2) mass; (3) forelimb length; (4) hindlimb length; (5) tail length; (6) number of toe lamellae. The mean of the posterior distribution and the interval containing 95% of the posterior density (credible interval) are shown.

Parameter	Mean	Credible interval
σ_1	118.3	(98, 140)
σ_2	93.1	(77, 112)
σ_3	52.2	(44, 62)
σ_4	81.8	(69, 96)
σ_5	244.9	(203, 292)
σ_6	21.2	(17, 27)

parative analysis using the method of Losos (1994, 1995) or Martins (1996) and the result obtained by our method would represent the information arising from the collection of new phylogenetic data.

Our method assumes that amino acid or DNA sequence data are available for the species sampled in a comparative analysis. Often, however, such data will not be available. If morphological data (besides the comparative characters) are available, one could use the method by applying a stochastic model to the characters, using the method of Lewis (2001), for example. Otherwise, one could average inferences over a sample of trees generated under a stochastic process.

The model of morphological evolution that we consider is a generalization of the earlier model of Felsenstein (1985). An interesting property of this model is that it assumes the variance-covariance of characters in a specific time interval of fixed duration is constant over all lineages in the tree. This is, in essence, a morphological clock hypothesis. Thus, if we were to measure time in units of generations and let Σ be the variance-covariance matrix of characters within a single generation, then the variation observed within a species within a single generation could be used to calibrate the clock and estimate the ages of nodes in the tree. Of course, it is unrealistic to assume that Σ will be constant. In reality, Σ will depend on the particular combinations of genes present in a population affecting the traits under consideration and their combined additive and epistatic effects. The level of additive genetic variance in the population will determine the rate of evolution (see Lynch and Walsh 1998), and this will change over time under forces of drift and natural selection. A better model would therefore allow $\Sigma(t)$ to be a function of time, evolving over the branches of the tree. In this study, this was accounted for by allowing different branches of the tree to differ in length.

The method we describe can be extended to accommodate uncertainty in other aspects of the phylogenetic model. For example, the comparative characters themselves are rarely known with certainty, and typically only the mean and the variance of a character for a species is available. Uncertainty of this sort can be accommodated using the approach we describe by integrating over all possible values for a character for each species. This would require that the uncertainty in each comparative character for a species be described by a parametric distribution, such as the normal distribution. For example, instead of assuming that the snout-vent length of a lizard species is represented by a single number, as was done in this study, the length could be assumed to be normally distributed with a mean and variance determined by a sample of individual lizards sampled from the species. MCMC could then be used to integrate over the uncertainty in the measured characters.

Finally, any analysis such as the one described in this paper, makes some arbitrary choices about how to parameterize the model. In this case, for example, we assume that the branch lengths for the continuous characters of interest are proportional to the lengths of the branches for the sequence data. However, other parameterizations could have been chosen. For example, the lengths of all of the branches of the continuous characters could have been considered to be equal in length, and then the DNA sequence data would

have been used only to describe the uncertainty in the phylogeny. It is probably worth exploring different parameterizations of the general approach described here to find which are most robust and generally applicable.

ACKNOWLEDGMENTS

We thank J. Hartigan for helpful comments and J. Losos, who gave us permission to use the data on his *Anolis* lizards. BR was supported by National Institutes of Health grant HG01988 and JPH was supported by National Science Foundation grants DEB-0075406 and MCB-0075404.

LITERATURE CITED

- Baum, D. A., and M. J. Donoghue. 2000. A likelihood framework for the study of adaptation. Pp. 24–44 in S. Orzack and E. Sober, eds. *Adaptationism and optimality* Cambridge Univ. Press, New York.
- Cheverud, J. M., M. M. Dow, and W. Leutenegger. 1985. The quantitative assessment of phylogenetic constraints in comparative analyses: sexual dimorphism in body weight among primates. *Evolution* 39:1335–1351.
- Dickey, J. 1971. The weighted likelihood ratio, linear hypotheses on normal location parameters. *Ann. Stat.* 42:204–223.
- Donoghue, M. J., and D. D. Ackerly. 1996. Phylogenetic uncertainties and sensitivity analyses in comparative biology. *Philos. Trans. R. Soc. Lond. B* 351:1241–1249.
- Efron, B., E. Halloran, and S. Holmes. 1996. Bootstrap confidence levels for phylogenetic trees. *Proc. Natl. Acad. Sci.* 93:7085–7090.
- Felsenstein, J. 1981. Evolutionary trees from DNA sequences: a maximum likelihood approach. *J. Mol. Evol.* 17:368–376.
- . 1985. Phylogenies and the comparative method. *Am. Nat.* 125:1–15.
- . 1988. Phylogenies and quantitative characters. *Annu. Rev. Ecol. Syst.* 19:445–471.
- Grafen, A. 1989. The phylogenetic regression. *Philos. Trans. R. Soc. Lond. B* 326:119–157.
- Green, P. J. 1995. Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika* 82:711–732.
- Harvey, P. H., and M. D. Pagel. 1991. *The comparative method in evolutionary biology*. Oxford Univ. Press, Oxford, U.K.
- 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.
- 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.
- Hastings, W. K. 1970. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* 57:97–109.
- Houseworth, E. A., and E. P. Martins. 2001. Random sampling of constrained phylogenies: conducting phylogenetic analyses when the phylogeny is partially known. *Syst. Biol.* 50:628–639.
- Jackman, T. R., A. Larson, K. de Queiroz, and J. B. Losos. 1999. Phylogenetic relationships and tempo of early diversification in *Anolis* lizards. *Syst. Biol.* 48:254–285.
- Janson, C. H. 1992. Measuring evolutionary constraints: a Markov model for phylogenetic transitions among seed dispersal syndromes. *Evolution* 46:136–158.
- Larget, B., and D. Simon. 1999. Markov chain Monte Carlo algorithms for the Bayesian analysis of phylogenetic trees. *Mol. Biol. Evol.* 16:750–759.
- Lavine, M., and M. J. Schervish. 1999. Bayes factors: what they are and what they are not. *Am. Stat.* 53:119–122.
- Lewis, P. L. 2001. Maximum likelihood phylogenetic inference: modeling discrete morphological characters. *Syst. Biol.* 50:913–925.
- Li, S. 1996. *Phylogenetic tree construction using Markov chain Monte Carlo*. Ph.D. diss., Ohio State University, Columbus, OH.

- Losos, J. B. 1994. An approach to the analysis of comparative data when a phylogeny is unavailable or incomplete. *Syst. Biol.* 43: 117–123.
- . 1995. 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. Pp. 60–98 in P. C. Wainwright and S. Reilly, eds. *Ecological morphology: integrative organismal biology*. Univ. of Chicago Press, Chicago, IL.
- Losos, J. B., T. R. Jackman, A. Larson, K. de Queiroz, and L. Rodríguez-Schettino. 1998. Contingency and determinism in replicated adaptive radiations of island lizards. *Science* 279: 2115–2118.
- Lynch, M., and J. B. Walsh. 1998. *Genetics and analysis of quantitative traits*. Sinauer Associates, Sunderland, MA.
- 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.
- Martins, E. P. 1996. Conducting phylogenetic comparative studies when the phylogeny is not known. *Evolution* 50:12–22.
- . 2000. Adaptation and the comparative method. *Trends Ecol. Evol.* 15:296–299.
- Martins, E. P., and T. Garland. 1991. Phylogenetic analyses of the correlated evolution of continuous characters: a simulation study. *Evolution* 45:534–557.
- 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.
- Pagel, M. D. 1992. A method for the analysis of comparative data. *J. Theor. Biol.* 156:431–422.
- . 1999. Inferring the historical patterns of biological evolution. *Nature* 401:877–884.
- Richman, A. D., and T. Price. 1992. Evolution of ecological differences in the old world leaf warblers. *Nature* 355:817–821.
- Rohlf, F. J. 2001. Comparative methods for the analysis of continuous variables: geometric interpretations. *Evolution* 55: 2143–2160.
- Sanderson, M. J. 1993. Reversibility in evolution: a maximum likelihood approach to character gain/loss bias in phylogenies. *Evolution* 47:236–252.
- Simon, D., and B. Larget. 2000. Bayesian analysis in molecular biology and evolution (BAMBE). Ver. 2.02 beta. Dept. Mathematics and Computer Science, Duquesne University, Pittsburgh, PA.
- Yang, Z. 1993. Maximum likelihood estimation of phylogeny from DNA sequences when substitution rates differ over sites. *Mol. Biol. Evol.* 10:1396–1401.
- . 1994. Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: approximate methods. *J. Mol. Evol.* 39:306–314.

Corresponding Editor: T. Garland, Jr.

APPENDIX

We update the state of the Markov chain one parameter at a time. For each cycle of the MCMC analysis, one of six proposal mechanisms could be used; each proposal mechanism updates a specific parameter and has a fixed probability of being used in any cycle. The probability of attempting a move that updates the i th parameter is p_i . Here, we describe the proposal mechanisms and the acceptance probability for each.

Topology and branch lengths (τ and ν).—We used the LOCAL proposal mechanism, implemented in the program BAMBE (Simon and Larget 2000) to simultaneously change the tree (τ) and branch lengths (ν). The LOCAL move is described in Larget and Simon (1999). Briefly, an internal node of the tree is chosen at random. This branch is the focal branch of the LOCAL move. There are two branches that are adjacent to each end of the focal branch. Randomly choose one branch from each end of the focal branch. There are

now a total of three branches that form the linear backbone of the move (the focal branch and one branch from each end of the focal branch). There are also two branches that are adjacent to each end of the focal branch that were not chosen to be members of the backbone. Label these branches A and B. The length of the backbone is the sum of the lengths of the three branches that form the backbone. This path length is denoted m . The lengths of the branches forming the backbone are increased or decreased by choosing a new pathlength (m'), where $m' = me^{\lambda(U-1/2)}$. Here, U is a uniformly distributed random variable on the interval $[0, 1]$ and λ is a tuning parameter that determines the size of the change in the length of the backbone. The length of each branch in the backbone is changed proportionally by multiplying the old length by m'/m . Finally, one of the two branches that do not make up the backbone (branch A or B) is chosen at random and detached from the tree. That branch is randomly reattached along the backbone; every point along the backbone has equal probability of being the reattachment point. The acceptance probability of this move is

$$R = \min\left[1, \text{likelihood ratio} \times \text{prior ratio} \times \left(\frac{m'}{m}\right)^2\right]. \quad (\text{A1})$$

Transition/transversion rate ratio (κ).—We update κ using a sliding window mechanism. A window of size h is chosen before the analysis. A new state for the transition/transversion rate ratio (κ') is chosen as $\kappa' = \kappa + h(2U - 1)/2$. If the new state is outside of the prior range of κ (e.g., $\kappa' < 0$), then the parameter is reflected back into the acceptable region. The acceptance probability of this move is

$$R = \min(1, \text{likelihood ratio} \times \text{prior ratio}). \quad (\text{A2})$$

Stationary base frequencies (π).—New values for the base frequencies are proposed from a Dirichlet distribution with the expected value at the current value. The Dirichlet distribution has probability density

$$f(\pi | \alpha_A, \alpha_C, \alpha_G, \alpha_T) = \frac{\Gamma(\alpha_A + \alpha_C + \alpha_G + \alpha_T)}{\Gamma(\alpha_A)\Gamma(\alpha_C)\Gamma(\alpha_G)\Gamma(\alpha_T)} \pi_A^{\alpha_A-1} \pi_C^{\alpha_C-1} \pi_G^{\alpha_G-1} \pi_T^{\alpha_T-1}, \quad (\text{A3})$$

where α_i is the Dirichlet parameter for the i th nucleotide, and $\Gamma(\bullet)$ is the gamma function. New base frequencies are drawn from a Dirichlet distribution with $\alpha_i = \pi_i \alpha_0$. We set $\alpha_0 = 100$ in all analyses. The acceptance probability is

$$R = \min\left[1, \text{likelihood ratio} \times \text{prior ratio} \times \frac{f(\pi' | \alpha_A, \alpha_C, \alpha_G, \alpha_T)}{f(\pi | \alpha_A, \alpha_C, \alpha_G, \alpha_T)}\right]. \quad (\text{A4})$$

Character variance (σ_i^2).—We update the variance of a character by multiplying a randomly chosen character variance parameter by a random factor. One of the d comparative characters is chosen at random. The variance is changed by multiplying the old variance by $e^{\lambda(U-1/2)}$ and the variance-covariance matrix recalculated. The new variance is then $\sigma' = \sigma e^{\lambda(U-1/2)}$. The acceptance probability is:

$$R = \min\left(1, \text{likelihood ratio} \times \text{prior ratio} \times \frac{\sigma'}{\sigma}\right). \quad (\text{A5})$$

Character correlation coefficient (ρ_{ij}).—The correlation coefficient is also changed using a sliding window mechanism. The correlation coefficient for the i th and j th character is changed and the new variance-covariance matrix calculated (Σ'). The acceptance probability is:

$$R = \min(1, \text{likelihood ratio} \times \text{prior ratio}). \quad (\text{A6})$$

Ancestral states of comparative character.—The state of a randomly chosen comparative character at a randomly chosen internal node on the tree is chosen to be changed. The new state is proposed using a sliding window mechanism. The acceptance probability is:

$$R = \min(1, \text{likelihood ratio} \times \text{prior ratio}). \quad (\text{A7})$$