doi:10.1111/j.1558-5646.2012.01645.x

A NEW BAYESIAN METHOD FOR FITTING EVOLUTIONARY MODELS TO COMPARATIVE DATA WITH INTRASPECIFIC VARIATION

Liam J. Revell[1,2] and R. Graham Reynolds[1]

[1] *Department of Biology, University of Massachusetts Boston, Boston, Massachusetts 02125*

[2] *E-mail: liam.revell@umb.edu*

Received December 31, 2011
Accepted March 13, 2012
Data Archived: Dryad: doi:10.5061/dryad.7fv08k72

Phylogenetic comparative methods thatincorporate intraspecific variability are relatively new and, so far, not especially widely used in empirical studies. In the present short article we will describe a new Bayesian method for fitting evolutionary models to comparative data that incorporates intraspecific variability. This method differs from an existing likelihood-based approach in that it requires no a priori inference about species means and variances; rather it takes phenotypic values from individuals and a phylogenetic tree as input, and then samples species means and variances, along with the parameters of the evolutionary model, from their joint posterior probability distribution. One of the most novel and intriguing attributes of this approach is that jointly sampling the species means with the evolutionary model parameters means that the model and tree can influence our estimates of species mean trait values, not just the reverse. In the present implementation, we first apply this method to the most widely used evolutionary model for continuously valued phenotypic trait data (Brownian motion). However, the general approach has broad applicability, which we illustrate by also fitting the λ model, another simple model for quantitative trait evolution on a phylogeny. We test our approach via simulation and by analyzing two empirical datasets obtained from the literature. Finally, we have implemented the methods described herein in a new function for the R statistical computing environment, and this function will be distributed as part of the 'phytools' R library.

KEY WORDS: **Comparative method, interspecific data, phylogenetic tree.**

Comparative methods for the analysis of continuously valued character data have progressed forward in a dramatic fashion in recent years. For instance, there are now methods to fit multiple evolutionary rates, selection regimes, and correlations between characters to different parts of a phylogenetic tree (Butler and King 2004; O'Meara et al. 2006; Thomas et al. 2006; Revell and Collar 2009). In addition, there are methods to fit a punctuational model of evolutionary change while taking extinct lineages into account (Bokma 2002, 2008); as well as methods to fit multiple rates of evolution without specifying the positions of the rate shifts and in situations where taxon sampling is incomplete (Eastman et al. 2011; Revell et al. 2012; Slater et al. 2012; Thomas and Freckleton 2012). Most of the aforementioned approaches are based on maximum likelihood, but Bayesian Markov chain Monte Carlo (MCMC) methods, now a mainstay of phylogenetic inference (Yang and Rannala 1997), have also begun to infiltrate comparative biology (Bokma 2008; Eastman et al. 2011; Revell et al. 2012).

Although two different but closely related methods have been proposed to account for intraspecific variation in comparative analyses (Ives et al. 2007; Felsenstein 2008), most approaches for evolutionary model fitting still essentially ignore sampling error—treating the estimated means for individual species as the population means known without error. This can have significant consequences including causing bias in the estimation of the parameters of our evolutionary model. For instance, estimates of phylogenetic signal are downwardly biased on average (Ives et al. 2007; Revell et al. 2008); whereas estimates of the

EVOLUTION
INTERNATIONAL JOURNAL OF ORGANIC EVOLUTION

evolutionary rate will be biased upwards if sampling error has been ignored.

In a 2007 article, Ives et al. (2007) developed a maximum likelihood technique for incorporating intraspecific variability into the phylogenetic analysis of interspecific data. This method is built around the relatively straightforward premise that if the intraspecific variances for species are treated as known, then one can simply add the extra sampling variance for the means of species to the total variance expected between species in the tree (Ives et al. 2007). The remaining variance not due to sampling error is, of course, due to the evolutionary process and elapsed time. Then one can just optimize whatever evolutionary model is being fit (for instance, Brownian motion), identifying the values of the evolutionary parameters in the model for the process of divergence between species that maximize the probability of the data conditioned on the estimated sampling variances and the phylogeny (Ives et al. 2007). In addition to the MATLAB (MathWorks 2011) implementation of the original authors (Ives et al. 2007), this method has also been incorporated into the 'geiger' (Harmon et al. 2008) and 'phytools' (Revell 2012) R phylogenetics libraries.

In the Bayesian method presented herein, we take a very different approach. In our method, the phylogeny, the evolutionary model, and the phenotypic values for individuals serve as input. We then sample species means, intraspecific phenotypic variances, and the parameters of the evolutionary process from their joint posterior probability distribution. The consequence of taking this approach will be that, in theory, information flow is bidirectional in the sense that the evolutionary model and its parameters can influence the estimates of our species means and variances (as well as the reverse). In the end, we obtain a posterior sample for both our evolutionary model parameters and our species means and variances.

In addition to presenting our new method, we will ask the following questions: (1) Does our Bayesian inference procedure produce satisfactory estimates of underlying evolutionary parameters; and, in general, do 95% credible intervals from the posterior sample include the generating parameter values? (2) Are the mean phenotypic trait values for terminal species estimated unbiasedly by our method; and, furthermore, do these estimates have lower error than parametric estimates obtained from the data but ignoring the tree and evolutionary model? In the present article, we will address these questions with simulation and also present two different example analyses of published datasets.

Methods

MATHEMATICAL AND COMPUTATIONAL DETAILS

We programmed all the analyses of this study in the scientific computing language R (R Development Core Team 2011). Sample simulation code is provided in an Appendix S1 and the novel functions of this study will be distributed as part of the 'phytools' R library (Revell 2012). The 'phytools' package is a multifunctional R library that depends heavily on the core phylogenetics package, 'ape' (Paradis et al. 2004).

In this article, we present a new Bayesian approach to the phylogenetic analysis of species data with intraspecific variability. We illustrate this approach by fitting a single rate Brownian model for evolutionary change. Under a Brownian process, sometimes called a "random walk" model, evolutionary changes along the branches of the tree are drawn randomly from a normal distribution with variance equal to the product of the evolutionary rate (σ^2_{BM}) and the length of the branch (Cavalli-Sforza and Edwards 1967; Felsenstein 1985). Under this model of evolution the true species means have a multivariate normal density with a variance–covariance matrix given by $\sigma^2_{BM}\mathbf{C}$, in which $\mathbf{C}$ is an $n \times n$ matrix containing in each position i, j the height above the root node of the most recent common ancestor of the species pair i and j (Rohlf 2001; Revell 2008). Thus, the diagonal of $\mathbf{C}$ is populated with the total tree length from the root node to any tip; and the off-diagonals are populated by the heights of internal nodes on the tree.

For our method, we first need to obtain a function to calculate the probability of our data, which consists of between one and a large number of observations per species for a single continuously valued phenotypic trait, conditioned on the means and variances of individual species, on the parameters of our evolutionary model, and on our phylogenetic tree. To do this, we will compute two different probabilities—first, the probability that a particular proposed mean vector arose under the modeled evolutionary process and parameter values; and second, the probability of our observed data conditioned on the proposed mean vector and intraspecific variances. We will consider two different models for the variance within species in this study (not to be confused with our two different evolutionary models, discussed later). In the first model, henceforward the "reduced" model, we assume that intraspecific variability is the same for all species. In this case, the likelihood of our evolutionary model parameters, mean vector, and within-species phenotypic variance, conditioned on the data and tree, is given by the following equation:

$$\begin{aligned} &L(\sigma^2_{BM}, \alpha, \bar{\mathbf{x}}, \sigma^2 | \mathbf{x}, \mathbf{C}) \\ &= \frac{\exp[-\frac{1}{2}(\bar{\mathbf{x}} - \alpha\mathbf{1})'(\sigma^2_{BM}\mathbf{C})^{-1}(\bar{\mathbf{x}} - \alpha\mathbf{1})]}{(2\pi)^{n/2}\left|\sigma^2_{BM}\mathbf{C}\right|^{1/2}} \\ &\quad \times \prod_{i=1}^{n}\prod_{j=1}^{m_i}\frac{\exp[-(x_{ij} - \bar{x}_i)^2/2\sigma^2]}{\sqrt{2\pi\sigma^2}}. \end{aligned}$$

The first part of this equation provides the probability that the data in the mean vector ($\bar{\mathbf{x}}$) arose on the phylogenetic tree

($\mathbf{C}$), given the proposed Brownian motion rate (σ^2_{BM}) and state for the root node of the tree (α). This expression is based on the multivariate normal equation and assumes Brownian evolution as the generating process (O'Meara et al. 2006). The second part is the product over all observations in $\mathbf{x}$ (which is indexed by species, i, and sample per species, j) of the probabilities that the observed data arose given the mean vector $\bar{\mathbf{x}}$ and a proposed common intraspecific variance, σ^2. This expression merely assumes that the data are normally distributed within species and that the observations within each species are independently sampled. Data that are not normally distributed within species can often be rendered so by transformation (e.g., log-transformation) (Log-transformation may also normalize the within-species variability, thus helping to satisfy the uniform variance assumption of this model).

For reasons of computation, most of the analyses of this article focus on the reduced model. However, if we have reason to suspect that the interspecific variance varies strongly among species, we might prefer to avoid the constant variance assumption of the model given above. In this case, we substitute vector $\mathbf{v}$ for intraspecific variance σ^2, and use the following equation for the likelihood:

$$L(\sigma^2_{BM}, \alpha, \bar{\mathbf{x}}, \mathbf{v}|\mathbf{x}, \mathbf{C}) = \frac{\exp[-\frac{1}{2}(\bar{\mathbf{x}} - \alpha\mathbf{1})'(\sigma^2_{BM}\mathbf{C})^{-1}(\bar{\mathbf{x}} - \alpha\mathbf{1})]}{(2\pi)^{n/2}\left|\sigma^2_{BM}\mathbf{C}\right|^{1/2}} \times \prod_{i=1}^{n}\prod_{j=1}^{m_i}\frac{\exp[-(x_{ij} - \bar{x}_i)^2/2v_i]}{\sqrt{2\pi v_i}}.$$

The only difference between this version and the prior equation is that here we compute the probability of each observation within species, conditioned on both a species-specific mean (contained in $\bar{\mathbf{x}}$) and a species-specific variance (contained in $\mathbf{v}$).

Finally, to illustrate the generality of our approach we also considered one variant of the standard Brownian model: the λ model of Pagel (1999). According to this model the off-diagonals of the phylogenetic covariance matrix $\mathbf{C}$ are multiplied by the coefficient λ, which is typically bounded on the interval (0, 1). A useful way to think of λ might be as a measure of the fraction of variability among species that is attributable to the tree and a Brownian model of evolutionary change (whereas $1 - \lambda$ is the fraction that is independent among species). The likelihood of the parameters of the evolutionary model (in this case σ^2_{BM}, α, and λ) and the mean vector and intraspecific variance ($\bar{\mathbf{x}}$ and σ^2), conditioned on the data and tree, is given by the following equation in which $\mathbf{I} \cdot \mathbf{C}$ indicates the Hadamard (i.e., element wise) product of a diagonal matrix of 1.0s ($\mathbf{I}$) and phylogenetic covariance matrix, $\mathbf{C}$, as previously defined

$$L(\sigma^2_{BM}, \lambda, \alpha, \bar{\mathbf{x}}, \sigma^2|\mathbf{x}, \mathbf{C}) = \frac{\exp\{-\frac{1}{2}(\bar{\mathbf{x}} - \alpha\mathbf{1})'[\sigma^2_{BM}(\lambda\mathbf{C} + (1-\lambda)\mathbf{I}\cdot\mathbf{C})]^{-1}(\bar{\mathbf{x}} - \alpha\mathbf{1})\}}{(2\pi)^{n/2}\left|\sigma^2_{BM}[\lambda\mathbf{C} + (1-\lambda)\mathbf{I}\cdot\mathbf{C}]\right|^{1/2}} \times \prod_{i=1}^{n}\prod_{j=1}^{m_i}\frac{\exp[-(x_{ij} - \bar{x}_i)^2/2\sigma^2]}{\sqrt{2\pi\sigma^2}}.$$

To sample from the posterior probability distribution, we used MCMC according to the following procedure. First, we initiated the chain with some set of values for the parameters of the evolutionary model, $\bar{\mathbf{x}}$, and σ^2 (or $\mathbf{v}$). We then used Gaussian proposal distributions to update the state of the chain for each parameter. For proposed changes to σ^2_{BM} or σ^2 (or $\mathbf{v}$) in which the proposed value of the parameter, x, was less $x' < 0$, we set $x' = -x'$ (that is, we reflected across $x = 0$ by changing the sign of x'). For proposed changes to λ such that $\lambda' > \max(\lambda)$ or $\lambda' < 0.0$, in which $\max(\lambda)$indicates the maximum value of λ such that the likelihood equations given above are defined, we set $\lambda' = 2\max(\lambda) - \lambda'$ or $\lambda' = -\lambda'$ (for the former and latter, respectively). For an ultrametric tree, $\max(\lambda)$ can be computed as $\max(\lambda) = t/\max(t_I)$, in which t is the total tree height and $\max(t_I)$ is the maximum height of any internal node on the tree. Reflecting across the boundary condition for the parameters maintains symmetry of the proposal distribution as $P(x'|x) = P(x|x')$ for all x and x' under this protocol. Every generation, we proposed an updated value of σ^2_{BM}, α, $\bar{\mathbf{x}}$, σ^2 (or $\mathbf{v}$), and λ. Because $\bar{\mathbf{x}}$ and $\mathbf{v}$ are both vectors, updating all the elements of each was done sequentially in separate generations and according to the same procedure.

Given any new proposed value for a parameter, for instance $\sigma^{2\prime}_{BM}$ in the reduced Brownian model, we accepted this change with probability

$$\min\left(1, \frac{L(\sigma^{2\prime}_{BM}, \alpha, \bar{\mathbf{x}}, \sigma^2)\Pr(\sigma^{2\prime}_{BM}, \alpha, \bar{\mathbf{x}}, \sigma^2)}{L(\sigma^2_{BM}, \alpha, \bar{\mathbf{x}}, \sigma^2)\Pr(\sigma^2_{BM}, \alpha, \bar{\mathbf{x}}, \sigma^2)}\right).$$

In other words, any proposal that increased the posterior probability was invariably accepted; however, proposals decreasing the posterior probability were also accepted with a probability equivalent to the posterior odds ratio. This is the standard form of MCMC for symmetric proposal distributions, as in this study (Metropolis et al. 1953; Yang 2006).

We used Gaussian prior probability densities for all parameters except for σ^2_{BM} and σ^2 (or $\mathbf{v}$), and for λ. In the case of σ^2_{BM} and σ^2 (or $\mathbf{v}$), we used an exponential prior probability density function, whereas for λ we used a uniform density.

For all the analyses of this study, we used prior probability densities for the species means centered on 0.0 and with a variance of 1000 (in other words, essentially a flat prior on the range of our

data). For σ^2_{BM} and σ^2, we again used a very uninformative prior density: exponential with a mean of 1000; and for the uniform prior density of λ we used the interval [0, max(λ)]. Other prior probability densities are also possible; however, for relatively uninformative priors, such as those chosen for this study, we found that the method was fairly insensitive to the choice of prior.

SIMULATION TESTS

We assessed the performance of this method using four sets of simulation analyses. First, we simulated 100 pure-birth phylogenetic trees, each containing 50 taxa. We rescaled all trees to have a total length of 1.0 from the root node to any tip. Next, we simulated the true species means on the tree under Brownian motion, with several different values of the Brownian rate parameter ($\sigma^2_{BM} = 1.0$, 2.0, 4.0, and 8.0). Then, we simulated datasets containing between one and 10 samples per species, with constant intraspecific variability ($\sigma^2 = 0.2$). In addition, we simulated species means on the tree with constant Brownian rate parameter ($\sigma^2_{BM} = 1.0$); but then sampled between one and 10 observations per species with different intraspecific phenotypic variability ($\sigma^2 = 0.2$, 0.4, 0.6, 0.8, 1.0) for different simulations. Finally, we analyzed the simulated dataset and tree using the Bayesian MCMC method of this study with constant intraspecific variance across all species (the Brownian reduced model).

Second, we simulated data on trees with varying numbers of terminal species ($N = 20$, 50, 100, 200, 500) and then simulated data using the procedure for the simulations with constant intraspecific variance described above. We analyzed these data using our MCMC method and the reduced model of this study.

Third, using the same set of 100 trees as in the first set of simulations, we simulated true species means and datasets containing between 10 and 30 individuals per species using a Brownian rate of $\sigma^2_{BM} = 1.0$. In this case, however, we did not assume that the variances were constant, but assigned the intraspecific variance for each species by drawing random values from a χ^2 distribution with df = 4, but rescaled to have a mean of 0.2. This distribution for the phenotypic variances of species was chosen arbitrarily. Then, we analyzed the simulated dataset and tree using the Bayesian MCMC method of this study, but in which we sampled both means and variances for the tip species of the tree (i.e., the full model).

Finally, fourth, we simulated datasets using the λ model of Pagel (1999). λ is normally considered on the range of (0, 1), although we sampled λ up to its theoretical maximum (which varies by tree). We simulated data using the following values for λ: $\lambda = 0.0, 0.2, 0.4, 0.6, 0.8$, and 1.0. Other aspects of this simulation were the same as in the reduced model for constant evolutionary variance ($\sigma^2_{BM} = 1.0$) and phenotypic variance within species ($\sigma^2 = 0.2$). We then analyzed this datasets using the reduced λ model.

We ran each MCMC chain for 200,000 generations, sampling every 100 generations. We invariably excluded the first 100,000 generations (1001 samples, including the initial state of the chain) as burn-in. Before proceeding, we computed the effective sample sizes for each parameter, and repeatedly reran the chain for 200,000 additional generations until the effective sample sizes for all parameters were above 30 (or until the number of loops reached 10, in other words, 2,000,000 generations, although this was very rare in our analyses). For each repetition, the final state of the MCMC chain from the previous run was used to initiate the subsequent run. We set this low effective sample size threshold ($N_e > 30$) because it was difficult to automate the MCMC across all simulation conditions and stochastic datasets in a manner that guaranteed high effective sample sizes for all the parameters of the model. Normally, we would recommend that users tune proposal distributions to achieve effective sample sizes of at least $N_e = 100$ (as we will do with our empirical examples, described later). In most of our MCMC runs, the majority of parameters had effective sample sizes of $N_e = 1000$ or more. We computed our parameter estimates as the arithmetic means of the post-burn-in posterior sample. We also computed 95% credible intervals for σ^2_{BM}, the Brownian motion rate parameter, for each simulation. We evaluated the performance of the method for evolutionary inference by computing the mean value of the evolutionary rate parameter (σ^2_{BM}) across simulations, and by calculating the fraction of times that the 95% credible interval for σ^2_{BM} included the underlying generating value of this parameter. In addition, we evaluated the hypothesis that the estimated species means from this analysis would be closer to the true species means than were the arithmetic within-species means. To measure this, we computed the mean squared difference between the estimated means and their true values; and we compared this mean squared error to the mean squared error obtained from the arithmetic within-species means. We measured bias in the estimation of the means by taking the mean difference between each set of estimated means and their true values, and then averaging across all simulations. For analyses in which we also sampled the intraspecific phenotypic variances for species, we calculated the mean squared error between the estimated and generating values and mean bias in the same way. Finally, for analyses of the λ model, we computed the mean accuracy of the parameter estimate across simulations, as well as the fraction of times where the confidence interval for λ included the generating value.

EMPIRICAL EXAMPLES

In addition to the simulation tests, we also investigated the performance of this method using empirical data. We identified two datasets from the literature that offer contrasting scenarios for the application of empirical data to this method. The first dataset (Hulsey and Hollingsworth 2011) consisted of head length

measurements from three individuals of each of 36 species (i.e., an even number of samples for each species) of cyprinid minnows (Family: Cyprinidae) and a phylogenetic tree inferred from one mtDNA and one nuclear gene using a concatenated and partitioned Bayesian MCMC analysis. Head lengths were measured from the anterior of the upper jaw to the posterior of the preopercle (Hulsey and Hollingsworth 2011). The second dataset (Meachen-Samuels and Van Valkenburgh 2009) consisted of between one and 55 individuals measured for skull length from each of 35 species (i.e., an uneven number of samples per species) of felids (Family: Felidae), and a phylogenetic tree inferred from a concatenated 4620 base pair alignment (eight mtDNA genes and one nuclear SRY3′ gene), which is a subset of the alignments published in Johnson et al. (2006). Skull lengths were measured from the anteriormost incisor to the posterior of the sagittal crest (Van Valkenburgh and Koepfli 1993; Meachen-Samuels and Van Valkenburgh 2009). We estimated the phylogenetic tree in this group from the available alignment using Bayesian MCMC in MrBayes 3.2 (Ronquist and Hulsenbeck 2003). We first selected the best-fit model of sequence evolution (GTR+I+Γ) using the ModelTest Server 3.8 (Posada 2006), and we set the other conditions for the MCMC following Johnson et al. (2006). We ran the MCMC for 10^6 generations after a burn-in of 2×10^5 generations using the default temperature (0.2) with four chains, sampling trees every 100 generations. We repeated this analysis twice to verify that the MCMC chain had converged and recovered a consistent topology across runs (Huelsenbeck and Bollback 2001).

Next, we imported the data and trees into R and then natural logarithm transformed the morphological data. We rooted each tree using the closest outgroup taxa, after which we proceeded to prune the outgroup tips and associated internal nodes. We then used a semiparametric method based on penalized likelihood (Sanderson 2002) to render each phylogeny ultrametric using a conservative smoothing parameter of 0.1.

For each tree and dataset, we ran the Bayesian MCMC analysis for the Brownian reduced model of this study for 10^6 generations, sampling every 100 generations and excluding the first 250,000 generations as burn-in to achieve effective sample sizes for all parameters of at least $N_e > 100$. We confirmed stationarity of the MCMC chain using Heidelberger and Welch's (1983) convergence diagnostic implemented in the R package "coda" (Plummer et al. 2006) with $\alpha = 0.05$. We also computed 95% credible intervals from the posterior samples for each parameter using "coda."

Results

SIMULATION TESTS

We conducted four sets of simulation analyses in this study. First, we simulated Brownian evolution on 100 species trees with various values for the Brownian rate (σ^2_{BM}) or intraspecific variance (σ^2), for small sample size per species. We then analyzed these data using the reduced Brownian model. Second, we simulated data with constant σ^2_{BM} and σ^2, but for various numbers of terminal species. We analyzed these data and trees using the reduced Brownian model. Third, we simulated Brownian evolution with variable intraspecific variances (σ^2), for larger sample sizes per species; and we analyzed these data using the full Brownian model. Finally, we simulated under the λ model of Pagel (1999), but for constant σ^2_{BM} and σ^2. We analyzed these data under a reduced λ model. Although we highlight only the most salient aspects of our simulation analyses below, a detailed summary of all simulation results is given in Appendix S2.

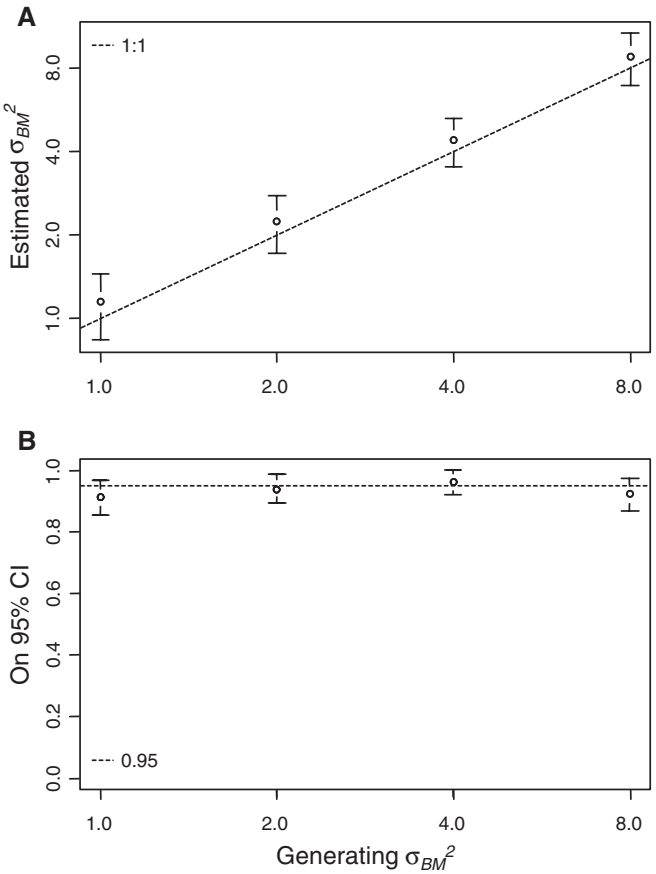

Figure 1. Results from the simulation analysis of the reduced Brownian model with constant intraspecific variance. (A) Generating and estimated values of the Brownian rate parameter, σ^2_{BM}. Error bars show the standard deviation (SD) across simulations. Diagonal dashed line indicates the 1:1 line. (B) Fraction of times that the credible interval for σ^2_{BM} included the generating value of σ^2_{BM}. Ninety-five percent, the expected rate, is indicated by the horizontal dashed line. Error bars show the standard error of the fraction.

From the first simulation analysis, with constant intraspecific variance and variable σ^2_{BM}, Figure 1A shows the mean value

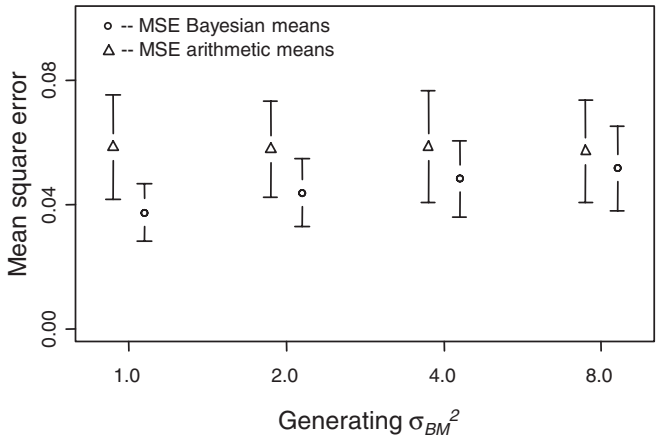

Figure 2. Mean squared deviation between the estimated and generating species means. Circles denote estimates obtained via Bayesian MCMC. Triangles indicate arithmetic means. Error bars represent the standard deviation (SD) across simulations.

of the evolutionary rate parameter (estimated from the posterior sample in each analysis) plotted against the generating value of σ^2_{BM}. The error bars represent ± 1 standard deviation (SD) from the variance in parameter estimates among simulations. The dashed line shows the line along which estimated and generating σ^2_{BM} are equal; clearly the mean evolutionary rate estimate is slightly upwardly biased. Figure 1B shows the frequency with which the 95% credible interval for σ^2_{BM} included the generating value for σ^2_{BM} across simulation conditions. Error bars are the 95% confidence limits on these frequencies. The observed fraction is not significantly different from the expected fraction (0.95) under any of the simulation conditions explored here (Fig. 1B). In addition to evaluating the accuracy of evolutionary parameter estimation, we also computed the mean square deviation (MSE) between the model estimated and generating species means at the tips of the tree. We compared this MSE to the MSE between the generating and arithmetic means calculated in the standard way. The MSE from the Bayesian MCMC is considerably smaller than the MSE from the arithmetic means (Fig. 2), indicating that the tree and evolutionary model provide useful information in estimating the species means that are consequently more accurate than when the tree is ignored. This effect is largest for the largest value of σ^2/σ^2_{BM} (i.e., when the intraspecific variance is relatively large compared to the variance of the Brownian evolutionary process). Note that the error bars shown in this figure are SDs among simulation results; not confidence intervals on the average MSEs, which would have been much smaller. Bias in both the Bayesian and arithmetic species means was nonexistent for this simulation, and indeed for all the simulations of this study, and will not be discussed further.

We obtained similar results when we varied the intraspecific variability among simulations. Figure 3A shows the generating and mean estimated values for the intraspecific variance parameter, σ^2, across simulation conditions, along with the SD among simulations obtained for a given value of σ^2. Here, the within-species variance (σ^2) is estimated with high accuracy for all simulation conditions. In addition, the accuracy (MSE) of the model-fitted species means is considerably better than for their arithmetic counterparts, particularly as the intraspecific variability (σ^2) is increased (Fig. 3B).

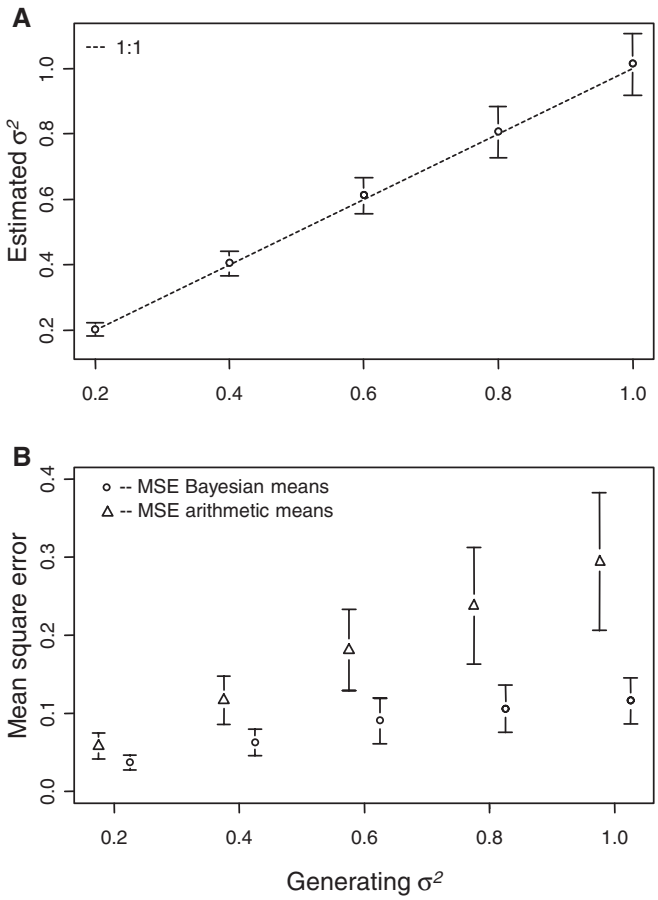

Figure 3. Results from simulation analysis of the reduced Brownian model with constant evolutionary rate, σ^2_{BM}. (A) Generating and estimated values for the common within-species variance, σ^2 (error bars as in Fig. 1A and B) Mean squared deviation between the estimated and generating species means. Circles, triangles, and error bars as in Figure 2.

In our second set of simulations we also varied the size (i.e., the number of tip taxa) of the stochastic phylogenetic trees used in our simulation and analyses, now holding σ^2_{BM} and σ^2 constant. Figure 4 shows the mean parameter estimate for σ^2_{BM} across the range of tree sizes explored in these simulations. We can see from this analyses that the positive bias found for σ^2_{BM} for trees of size $N = 50$ decreases substantially for larger N. Thus, the Bayesian estimator seems to be asymptotically unbiased. Of additional interest is the fact that the bias in σ^2_{BM} is in the opposite direction

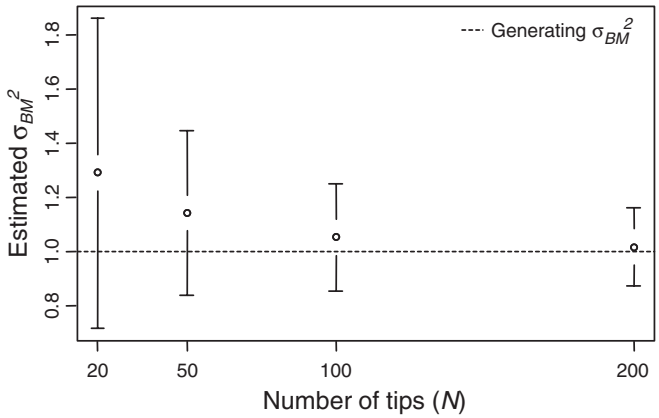

Figure 4. Mean estimated σ^2_{BM} for phylogenies with various numbers of terminal species.

(upward instead of downward) from maximum-likelihood (ML) estimation of the same parameter when the true species means are known (O'Meara et al. 2006).

In addition to these analyses of the reduced model, for our third set of simulations we simulated and analyzed under the full model—that is, a model in which each tip taxon is assigned a separate intraspecific phenotypic variance for the character under study. Representative results are shown in Figure 5. Figure 5A shows the generating and model-fitted species means for one simulation, and Figure 5B shows the generating and model-fitted intraspecific variances. Although the species means are estimated with little error, the estimated within species variances are slightly biased in an upward direction both in this individual simulation and across simulations (Table S5). In contrast to the Bayesian estimates of the species means, our equivalent estimates of the variances are less accurate than the standard population variances (Table S5). This is not particularly surprising because the tree can provide only limited help in improving the accuracy of our estimated variances (by improving the accuracy of the species means) as the variances did not evolve on the tree, but were sampled randomly from a χ^2 distribution.

Although this article focuses primarily on a Brownian model of evolutionary change, to illustrate the generality of our approach we also conducted a final set of analyses in which we simulated and estimated under Pagel's (1999) λ model. This model is a simple extension of Brownian motion in which the expected covariances between related species are downweighted by factor λ. Figure 6 shows a plot of the generating values for λ against the mean estimated value, with the error bars showing the SD among simulations. λ is estimated with an upward bias when the generating value of λ is small, and a downward bias when the generating value of λ is large. This is particular evident for simulations in which the generating value of λ was $\lambda = 0.0$ (Fig. 6). This bias does not characterize ML estimation of λ and

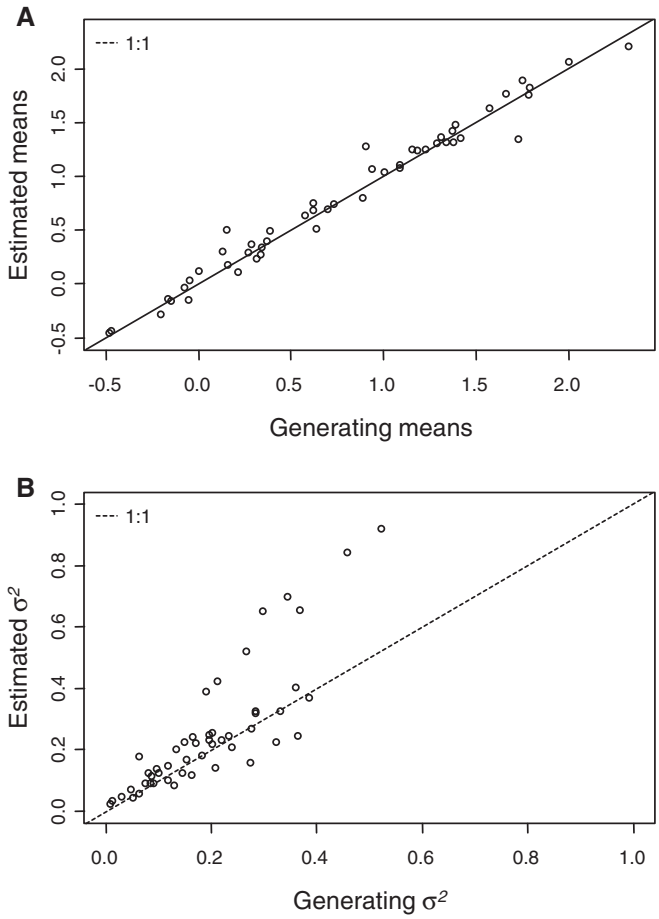

Figure 5. Representative result from simulation analysis of the full Brownian model. (A) Generating and estimated species means. (B) Generating and estimated species-specific variances.

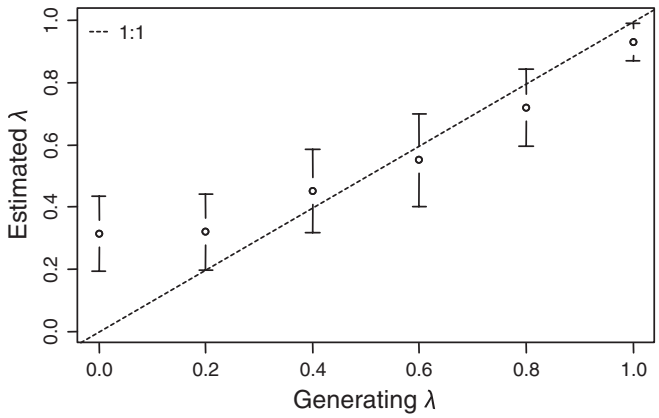

Figure 6. Results from simulation analysis of the reduced λ model. Generating values of λ are on the abscissa, and estimated values on the ordinate. Error bars represent the standard deviation (SD) across simulations.

it may be that as there are bounds on the parameter space for λ beyond which the likelihood is not well defined, the posterior sample for λ cannot be symmetric when the generating value is near these bounds. This would explain the observed pattern and might suggest that another metric for summarizing the posterior sample for λ (the median or mode, for instance) might be more appropriate.

As with the Brownian model, we found that Bayesian estimates of the species means were generally better than their arithmetic counterparts (Table S5). This was even true for simulated $\lambda = 0.0$, although the improvement in this case is nonsignificant. We strongly suspect that the improved accuracy of the Bayesian species means when $\lambda = 0.0$ (and thus the tree contains no phylogenetic information about the species mean trait values) is actually an artifact of our study design. This is because we simulated data assuming a constant intraspecific variance for all species. Sampling from the posterior distribution assuming this model (as we have done with the reduced λ model) better approximates the true sampling distribution of $\bar{\mathbf{x}}$—even if the tree contains no additional information about the true species means. This improvement is very minor and much smaller than the improvement in accuracy of species means estimation that is achieved by incorporating phylogenetic information when phylogenetic signal is high (Table S5).

EMPIRICAL EXAMPLES

We analyzed two datasets using the methods of this article. First, we analyzed a dataset for cyprinid head length and a tree published by Hulsey and Hollingworth (2011). Second, we inferred the Bayesian maximum clade credibility tree from a published felid alignment (Johnson et al. 2006), and analyzed this with existing data for felid head length (Van Valkenburgh and Koepfli 1993; Meachen-Samuels and Van Valkenburgh 2009). Morphological datasets, the minnow phylogenetic tree (Hulsey and Hollingsworth 2011), and the felid alignment (Johnson et al. 2006) are publicly available.

Our estimated felid tree was topologically congruent with the published tree of Johnson et al. (2006). We have appended our inferred felid tree in the Appendix S3. We analyzed each dataset and tree using the reduced Brownian model and MCMC conditions described in the Methods section, above. We tuned the conditions of the MCMC to ensure that all parameter effective sample sizes were larger than $N_e > 100$ (in most cases $N_e > 1000$, see Table 1). We have reported individual parameter effective sample sizes, mean values of σ^2_{BM}, α, and the uniform intraspecific variance, σ^2, for each analysis, along with the 95% highest posterior density (HPD) intervals in Table 1. Bayesian species means from the fitted model and arithmetic means are also reported in Appendix S4. Because we do not know the true generating model in either of these cases, we will not comment on the accuracy of the fitted models; however, we will note that the estimated species means from the Bayesian posterior sample are very close to the arithmetic species means. This is expected when intraspecific variance is low, and inspection of our empirical data suggests that indeed intrapsecific variance is low compared to the differences among species for these datasets.

Discussion

Comparative methods for the study of phenotypic change in the context of phylogenetic trees have evolved considerably in recent years. Where the study of continuously valued traits is concerned, investigators are typically interested in the evolution of species mean trait values over time. Yet, in nearly all empirical datasets the true species mean is unknown and one is forced to substitute a population mean (sometimes known by as few as one or two specimens, see Harmon and Losos 2005).

Two recent articles (Ives et al. 2007; Felsenstein 2008) presented alternative (although closely related, see Felsenstein 2008, p. 714) approaches to the problem of incorporating uncertainty in the estimation of species means. In the method of Ives et al. (2007), for instance, the means and sampling variances of species are first calculated; and then we compute the likelihood of the evolutionary parameters conditioned on the sampled means and variances, evolutionary model, and phylogenetic tree. In the present article, we offer a different paradigm. Here, the tree and the data for individuals serve as input to the analysis. Then, we use Bayesian MCMC to simultaneously sample from the joint posterior distribution of the parameters of our evolutionary model, species means, and intraspecific phenotypic variances.

The preliminary results revealed by applying this approach to empirical and simulated data in this study are quite encouraging. First, although perhaps least notably, we found that the method provides asymptotically unbiased estimates of the generating evolutionary model parameters (Fig. 4). The bias for small phylogenetic trees was in the opposite direction of the well-known bias in maximum likelihood estimation when the species means are known without error (O'Meara et al. 2006; Revell 2008). In particular, the Brownian evolutionary rates estimated using our Bayesian method were too high (at least given the prior probability densities that we chose for the analyses of this article) for trees with few taxa (Figs. 1A, 4). Second, we found that the accuracy of the Bayesian estimated species means is considerably better than the accuracy of the arithmetic species means under both the reduced and full models (e.g., Figs. 2, 3B). This indicates that accounting for the tree and evolutionary model can improve our estimation of the true species means. This is particularly true when the intraspecific variance is high relative to the divergence between species (Fig. 3B). Although this result is very intriguing, if we look closely at the results from the analysis of our empirical

Table 1. Results from analyses of empirical datasets. Sample sizes are given as the effective sample size (N_e) from the posterior sample after 10^6 generations of MCMC minus a 250,000 generation burn-in, sampled every 100 generations. Means for σ^2_{BM}, α, and intraspecific variance (σ^2) are given as averages with the 95% bounds of the highest posterior density (HPD) interval (lower, upper) from the MCMC analysis.

Dataset	Parameter	Mean from posterior sample	95% HPD	Effective sample size
Minnows	σ^2_{BM}	0.0992	(0.0544, 0.1526)	1731.9
	α	2.5476	(2.2008, 2.8872)	130.97
	σ^2	0.0022	(0.0015, 0.0031)	1061.6
Felids	σ^2_{BM}	0.1207	(0.0655, 0.1803)	1871.1
	α	4.9400	(4.6464, 5.219)	215.00
	σ^2	0.0081	(0.0068, 0.0092)	1283.7

datasets we find that the Bayesian fitted species means can be very close to the arithmetic means (Table S6). This will generally be the case when the within-species variance is small because under those conditions the log-likelihood is quickly dominated by the second product for sampled species means that differ substantially from the arithmetic within-species means. Indeed, in our empirical datasets the fitted within species variances were very small (Table 1).

Herein we focus on the ever popular, sometimes maligned (e.g., Price 1997; Butler and King 2004), Brownian model for evolutionary change on the tree (Cavalli-Sforza and Edwards 1967; Felsenstein 1985; O'Meara et al. 2006; Revell 2008); however, we emphasize that the general paradigm that we present in this article is applicable to any model for the evolution of continuously valued character traits in the context of a phylogeny. To illustrate this point, we apply the method to the λ model of Pagel (1999). We found that estimated values for λ taken as the mean of the posterior distribution were slightly biased. This bias tended to be in an upward direction for small generating λ and downward for high λ (Fig. 6). This is most likely due to asymmetry in the posterior probability density near the bounds of λ and suggests that the posterior sample mean might be a poor metric with which to summarize the posterior distribution of λ.

As the approach outlined herein is applicable to multiple models of evolutionary change in quantitative traits on the tree (e.g., Blomberg et al. 2003; Butler and King 2004; O'Meara et al. 2006), it is appropriate to consider comparison among models. According to our method the posterior distribution is sampled by MCMC, rather than optimized directly. Consequently, we cannot compare among models using AIC or BIC, in which an optimized value of the likelihood function is required. One option is to compare among models by way of the DIC or "Deviance Information Criterion" (Spiegelhalter et al. 2002). DIC is computed as follows:

$$\mathrm{DIC} = p_D - \bar{D}$$

in which $\bar{D}$ is merely the posterior mean log-likelihood. The measure of model complexity, p_D, is computed as $p_D = \bar{D} - D(\bar{\theta})$ in which $D(\bar{\theta})$ is the value of the likelihood function at the posterior mean for the parameters of our model. We would choose, in this case, the model with the lowest DIC. Comparison of DIC values for two different models might tell us whether we should prefer the reduced or full Brownian motion models; or whether we should prefer the Brownian or λ model. Ando (2007) has also proposed the Bayesian Predictive Information Criterion (BPIC), which he shows has superior statistical properties to the DIC.

In the models presented herein, we assume either: that the intraspecific variance is the same across all species; or that intraspecific variance differs among species, but is random with respect to the tree. An alternative intriguing possibility is that intraspecific variance differs among species in a phylogenetically autocorrelated manner. We expect that this will be an important area of future research based on our approach; however, any attempt to phylogenetically model the evolution of intraspecific variance on the tree should be performed with caution. This is because the expected evolutionary dynamics of the additive genetic variance are equilibrial and non-Gaussian (and thus very non-Brownian; Bürger et al. 1989; Jones et al. 2003; Revell 2007); and we expect a part–whole correlation between the additive genetic variance and the phenotypic variance within species (Falconer and MacKay 1996; Lynch and Walsh 1998).

Phylogenetic methods and "tree-thinking" have become ubiquitous in modern evolutionary biology over the past 25 years. Phylogenetic comparative methods have considerable potential in helping us to understand more about the history of life on this planet. Although method development in this field has advanced rapidly, there remains considerable new progress to be made. In this short article, we present both a novel method and paradigm for the analysis of interspecies data with intraspecific variability. We hope this new approach to the analysis of interspecific data with intraspecific sampling variation in comparative studies stimulates additional research in this area.

ACKNOWLEDGMENTS

Thanks are due to D. Hulsey and P. Hollingsworth for providing raw data and a tree file from their datasets and to L. Mahler, G. Slater, and K. Winchell for reading earlier versions of this article. Helpful comments were also given by E. Kramer, L. Schmitz, and an anonymous reviewer.

LITERATURE CITED

Ando, T. 2007. Bayesian predictive information criterion for the evaluation of hierarchical Bayesian and empirical Bayes models. Biometrika 94:443–458.

Blomberg, S. P., T. Garland, Jr., and A. R. Ives. 2003. Testing for phylogenetic signal in comparative data: Behavioral traits are more labile. Evolution 57:717–745.

Bokma, F. 2002. Detection of punctuated equilibrium from molecular phylogenies. J. Evol. Biol. 15:1048–1056.

———. 2008. Detection of "punctuated equilibrium" by Bayesian estimation of speciation and extinction rates, ancestral character states, and rates of anagenetic and cladogenetic evolution on a molecular phylogeny. Evolution 62:2718–2726.

Bürger, R., G. P. Wagner, and F. Stettinger 1989. How much heritable variation can be maintained in a finite population by mutation-selection balance. Evolution 43:1748–1766.

Butler, M. A., and A. A. King. 2004. Phylogenetic comparative analysis: a modeling approach for adaptive evolution. Am. Nat. 164:683–695.

Cavalli-Sforza, L. L., and A. W. F. Edwards. 1967. Phylogenetic analysis: models and estimation procedures. Am. J. Hum. Genet. 19:233–257.

Eastman, J. M., M. E. Alfaro, P. Joyce, A. L. Hipp, and L. J. Harmon. 2011. A novel comparative method for identifying shifts in the rate of character evolution on trees. Evolution 65:3578–3589.

Falconer, D. S., and T. F. C. Mackay. 1996. Introduction to quantitative genetics. Prentice Hall, Essex, U.K.

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

———. 2008. Comparative methods with sampling error and within-species variation: contrasts revisited and revised. Am. Nat. 171:713–725.

Harmon, L. J., and J. B. Losos. 2005. The effect of intraspecific sample size on type I and type II error rates in comparative studies. Evolution 59:2705–2710.

Harmon, L. J., J. T. Weir, C. D. Brock, R. E. Glor, and W. Challenger. 2008. GEIGER: investigating evolutionary radiations. Bioinformatics 24:129–131.

Heidelberger, P., and P. D. Welch. 1983. Simulation run length control in the presence of an initial transient. Oper. Res. 31:1109–1144.

Huelsenbeck, J. P., and J. P. Bollback. 2001. Empirical and hierarchical Bayesian estimation of ancestral states. Syst. Biol. 50:351–366.

Hulsey, C. D., and P. J. Hollingsworth Jr. 2011. Do constructional constraints influence cyprinid (Cyprinidae: Leuciscinae) craniofacial coevolution? Biol. J. Linn. Soc. 103:136–146.

Ives, A. R., P. E. Midford, and T. Garland Jr. 2007. Within-species variation and measurement error in phylogenetic comparative methods. Syst. Biol. 56:252–270.

Johnson, W. E., E. Eizirik, J. Pecon-Slattery, W. J. Murphy, A. Antunes, E. Teeling, and S. J. O'Brien. 2006. The late Miocene radiation of modern Felidae: a genetic assessment. Science 311:73–77.

Jones, A. G., S. J. Arnold, and R. Bürger. 2003. Stability of the G-matrix in a population experiencing pleiotropic mutation, stabilizing selection, and genetic drift. Evolution 57:1747–1760.

Lynch, M., and B. Walsh. 1998. Genetics and analysis of quantitative traits. Sinauer, Sunderland, MA.

MathWorks. 2011. MATLAB 7.13 R2011b. The MathWorks, Inc., Natick, MA.

Meachen-Samuels, J., and B. Van Valkenburgh. 2009. Craniodental indicators of prey size preference in the Felidae. Biol. J. Linn. Soc. 96:784–799.

Metropolis, N., A. W. Rosenbluth, M. N. Rosenbluth, and A. H. Teller. 1953. Equations of state calculations by fast computing machines. J. Chem. Phys. 21:1087–1092.

O'Meara, B. C., C. Ané, M. J. Sanderson, and P. C. Wainwright. 2006. Testing for different rates of continuous trait evolution using likelihood. Evolution 60:922–933.

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

Paradis, E., J. Claude, and K. Strimmer 2004. APE: analyses of phylogenetics and evolution in R language. Bioinformatics 20:289–290.

Plummer, M., N. Best, K. Cowles, and K. Vines. 2006. CODA: convergence diagnostics and output analysis for MCMC. R. News 6:7–11.

Posada, D. 2006. ModelTest Server: A web-based tool for the statistical selection of models of nucleotide substitution online. Nucleic Acids Research 34:W700–W703.

Price, T. 1997. Correlated evolution and independent contrasts. Philos. Trans. R. Soc. Lond. Ser. B 352:519–529.

R Development Core Team. 2011 R: a language and environment for statistical computing. R foundation for statistical computing, Vienna, Austria.

Revell, L. J. 2007. Testing the genetic constraint hypothesis in a phylogenetic context: a simulation study. Evolution 61:2720–2727.

———. 2008. On the analysis of evolutionary change along single branches in a phylogeny. Am. Nat. 172:140–147.

———. 2012. Phytools: An R package for phylogenetic comparative biology (and other things). Methods in Ecology and Evolution 3:217–223.

Revell, L. J., and D. C. Collar. 2009. Phylogenetic analysis of the evolutionary correlation using likelihood. Evolution 63:1090–1100.

Revell, L. J., L. J. Harmon, and D. C. Collar. 2008. Phylogenetic signal, evolutionary process, and rate. Syst. Biol. 57:591–601.

Revell, L. J., D. L. Mahler, P. R. Peres-Neto, and B. D. Redelings. 2012. A new phylogenetic method for identifying exceptional phenotypic diversification. Evolution 66:135–146.

Rohlf, F. J. 2001. Comparative methods for the analysis of continuous variables: geometric interpretations. Evolution 55:2143–2160.

Ronquist, F., and J. P. Huelsenbeck. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19:1572–1574.

Sanderson, M. J. 2002. Estimating absolute rates of molecular evolution and divergence times: a penalized likelihood approach. Mol. Biol. Evol. 19:101–109.

Slater, G. J., L. J. Harmon, D. Wegmann, P. Joyce, L. J. Revell, and M. E. Alfaro. 2012. Fitting models of continuous trait evolution to incompletely sampled comparative data using Approximate Bayesian Computation. Evolution 66:752–762.

Spielgelhalter, D. J., N. G. Best, B. P. Carlin, and A. van der Linde. 2002. Bayesian measures of model complexity and fit. J. R. Stat. Soc. B 64:583–639.

Thomas, G. H., and R. P. Freckleton. 2012. MOTMOT: models of trait macroevolution on trees. Methods Ecol. Evol. 3:145–151.

Thomas, G. H., R. P. Freckleton, and T. Székely. 2006. Comparative analyses of the influence of developmental mode on phenotypic diversification rates in shorebirds. Proc. R. Soc. B 273:1619–1624.

Van Valkenburgh, B., and K. P. Koepfli. 1993. Cranial and dental adaptations to predation in Canids. Symp. Zool. Soc. Lond. 65:15–37.

Yang, Z. 2006. Computational molecular evolution. Oxford Univ. Press, Oxford, U.K.

Yang, Z., and B. Rannala. 1997. Bayesian phylogenetic inference using DNA sequences: a Markov chain Monte Carlo method. Mol. Biol. Evol. 14:717–724.

Associate Editor: E. Kramer

Supporting Information

The following supporting information is available for this article:

Appendix S1. Sample simulation code.
Appendix S2. Summary of results for all simulations.
Appendix S3. Inferred tree for felid dataset in nexus format.
Appendix S4. Fitted and arithmetic means from empirical analyses.
Table S1. Summary of results from simulation analyses of the reduced Brownian model in which the rate of evolution, σ^2_{BM}, was varied across simulations, but the intraspecific variance, σ^2, was held constant.
Table S2. Summary of results from simulation analyses of the reduced Brownian model in which the rate of evolution, σ^2_{BM}, was held constant while the intraspecific variance, σ^2, was varied among simulations.
Table S3. Summary of results from simulation analyses of the reduced Brownian model for varying tree size (i.e., number of tip taxa, N).
Table S4. Summary of results from the full Brownian model.
Table S5. Summary of results from the reduced λ model.
Table S6. Fitted and arithmetic means from the empirical analyses.

Supporting Information may be found in the online version of this article.