

A NEW PHYLOGENETIC METHOD FOR IDENTIFYING EXCEPTIONAL PHENOTYPIC DIVERSIFICATION

Liam J. Revell,^{1,4,5} D. Luke Mahler,^{1,2} Pedro R. Peres-Neto,³ and Benjamin D. Redelings¹

¹National Evolutionary Synthesis Center, Duke University, Durham, North Carolina 27705

²Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, Massachusetts 02138

³Département des sciences biologiques, Université du Québec à Montréal, Montréal, Québec H3C 3P8, Canada

⁴E-mail: liam.revell@umb.edu

Received December 24, 2010

Accepted July 19, 2011

Data Archived: Dryad doi:10.5061/dryad.vj310

Currently available phylogenetic methods for studying the rate of evolution in a continuously valued character assume that the rate is constant throughout the tree or that it changes along specific branches according to an a priori hypothesis of rate variation provided by the user. Herein, we describe a new method for studying evolutionary rate variation in continuously valued characters given an estimate of the phylogenetic history of the species in our study. According to this method, we propose no specific prior hypothesis for how the variation in evolutionary rate is structured throughout the history of the species in our study. Instead, we use a Bayesian Markov Chain Monte Carlo approach to estimate evolutionary rates and the shift point between rates on the tree. We do this by simultaneously sampling rates and shift points in proportion to their posterior probability, and then collapsing the posterior sample into an estimate of the parameters of interest. We use simulation to show that the method is quite successful at identifying the phylogenetic position of a shift in the rate of evolution, and that estimated rates are asymptotically unbiased. We also provide an empirical example of the method using data for *Anolis* lizards.

[This article was published online on September 20, 2011. An error in a co-author's name was subsequently identified. This notice is included in the online and print versions to indicate that both have been corrected September 21, 2011.]

KEY WORDS: Adaptive radiation, Bayesian, comparative method, Markov chain, phenotypic evolution, phylogenetics, rates of evolution.

Research on adaptive radiation is often focused on determining whether a particular clade of interest has exhibited exceptional evolutionary diversification during its history (Schluter 2000; Freckleton and Harvey 2006; Mahler et al. 2010). For instance, many concepts of adaptive radiation involve a shift in adaptive zone due to the evolutionary origin of a key innovation (Simpson 1953). This shift may be accompanied by exceptional species proliferation, exceptional phenotypic evolution, or both (Givnish 1997; Schluter 2000; Glor 2010; Losos and Mahler 2010).

Several methods have already been devised to test if a pre-specified radiation has been the subject of exceptional net species diversification (e.g., Slowinski and Guyer 1993; Rabosky 2006; Stadler 2011), and the BiSSE method by Maddison et al. (2007) can be used to fit a model in which the rates of speciation and extinction vary simultaneously with the state of binary character. In addition, recent methods developed by Alfaro et al. (2009) and Stadler (2011) can be used to identify clades or time periods with exceptional net species diversification, even when no such groups or intervals have been hypothesized a priori.

However, a key feature of adaptive radiations is that they involve not only species proliferation, but also phenotypic evolution (Schluter 2000; Losos and Miles 2002; Glor 2010;

⁵Current address: Department of Biology, University of Massachusetts Boston, Boston, Massachusetts 02125.

Losos and Mahler 2010; Mahler et al. 2010; Slater et al. 2010). In particular, by most accounts adaptive radiations are accompanied by an exceptional rate or amount of adaptive phenotypic divergence (Olson and Arroyo-Santos 2009). In this regard, phylogenetic methods for the study of adaptive radiation have lagged somewhat behind. So far, we have methods that can estimate a single rate of evolution throughout the tree (e.g., Garland 1992; O'Meara et al. 2006), and methods that can fit two or more rates of evolution for continuously valued characters, so long as the branches or subclades in each rate regime are specified a priori by the user (e.g., McPeck 1995; O'Meara et al. 2006; Thomas et al. 2006; Revell 2008). There are also methods in which we can fit different adaptive regimes to different a priori assigned branches in the tree (Butler and King 2004), and we can fit a model in which evolution is initially rapid and slows through time (Blomberg et al. 2003; Freckleton and Harvey 2006; Mahler et al. 2010). However, no method has yet been developed in which the branches of the tree associated with each rate regime, and rates of phenotypic change themselves, are estimated simultaneously from a phylogenetic tree and phenotypic trait data.

In the present article, we describe exactly such a method for continuously valued characters. We use a Markov Chain Monte Carlo (MCMC) approach to sample rate shift positions and evolutionary rates from their Bayesian posterior distribution. We explore the properties of this method via computer simulation, examining in particular the capacity of our method to locate the phylogenetic position of a shift in the evolutionary rate and to estimate the rates themselves. In addition, we test the method on phylogenies with various numbers of terminal taxa. Finally, we apply the method to an example dataset for Caribbean *Anolis* lizards.

The primary innovation of our method is that it can be used to identify rate variation naively—that is, without a specific and fixed prior hypothesis for how heterogeneity in the rate of evolution is structured among taxa or has evolved on the tree. This is a substantial advance because in theory it should allow us to identify clades with exceptional adaptive phenotypic diversification (i.e., adaptive radiations, by some definition) that are unknown a priori. We also believe that this is among the first articles to use a Bayesian MCMC approach to fit a model for the evolution of a continuously valued character to the data and a phylogenetic tree (but see Bokma 2008 for an earlier application to the problem of punctuational and gradual evolution of a continuous trait).

Here, we use a relatively simple model of rate variation: one in which the evolutionary rate has changed (from low to high or vice versa) once and only once in the history of the study group. We hope that this method will motivate further research in this area. For instance, we envision using the general approach applied in this paper to test more complex hypotheses involving multiple

rate shifts, more than two rates, different models of evolution, or multiple continuous traits.

Methods and Results

MATHEMATICAL AND COMPUTATIONAL DETAILS

We programmed all analyses presented herein in the flexible scientific computing language, R (R Development Core Team 2010). The code we used for the analyses of this paper is available as an Appendix S1 and updated versions will be distributed as part of the R phylogenetics package “phytools” (Revell 2011). The simulation, MCMC, and MCMC diagnostics code, all provided in Appendix S1, call functions from the phylogenetic packages “ape” and “geiger” (Paradis et al. 2004; Harmon et al. 2008), and from the MCMC diagnostics package “coda” (Plummer et al. 2010).

The model presented herein is for the evolution of a single, continuously valued character on a rooted phylogenetic tree with branch lengths in units proportional to time. Under this model, evolution proceeds by a Brownian motion process on the tree (Cavalli-Sforza and Edwards 1967; Felsenstein 1985). The instantaneous variance of the evolutionary process in this model (the evolutionary rate) changes from low to high, or high to low, once and only once in the tree. Accordingly, the model has four parameters: the two evolutionary rates (σ_1^2 and σ_2^2) that prevail on either side of the rate shift; θ , a 2×1 vector containing the branch identity and the position along the branch at which the evolutionary rate transitions from σ_1^2 to σ_2^2 (or vice versa); and finally, the ancestral trait value at the root node of the tree (α). The data consist of values for a continuously distributed character for all tip species, and a bifurcating or multifurcating rooted phylogeny with branch lengths. We focus on estimating σ_1^2 , σ_2^2 , and θ from the tree and character data.

To be able to sample from the posterior probability distribution, we need to start with an expression for the likelihood of our model and parameters given the data and tree. The expression we used is based on the multivariate normal equation and has been applied previously to related problems (Felsenstein 1973; O'Meara et al. 2006):

$$L(\sigma_1^2, \sigma_2^2, \alpha, \theta) = \frac{\exp\left[-\frac{1}{2}(\mathbf{x} - \alpha\mathbf{1})'(\sigma_1^2\mathbf{C}_1 + \sigma_2^2\mathbf{C}_2)(\mathbf{x} - \alpha\mathbf{1})\right]}{(2\pi)^{n/2} |\sigma_1^2\mathbf{C}_1 + \sigma_2^2\mathbf{C}_2|^{1/2}},$$

where \mathbf{x} is an $n \times 1$ vector containing our trait data for each of n species; $\mathbf{1}$ is a conformable vector of 1.0s; and \mathbf{C}_i is an $n \times n$ covariance matrix containing all the branch lengths in the i th group on the tree (e.g., Revell 2008). The break point between \mathbf{C}_1 and \mathbf{C}_2 is determined by θ , as illustrated in Figure 1. The manner in which we have computed the \mathbf{C}_i s in Figure 1 assumes that evolution proceeds by random diffusion (Brownian motion)

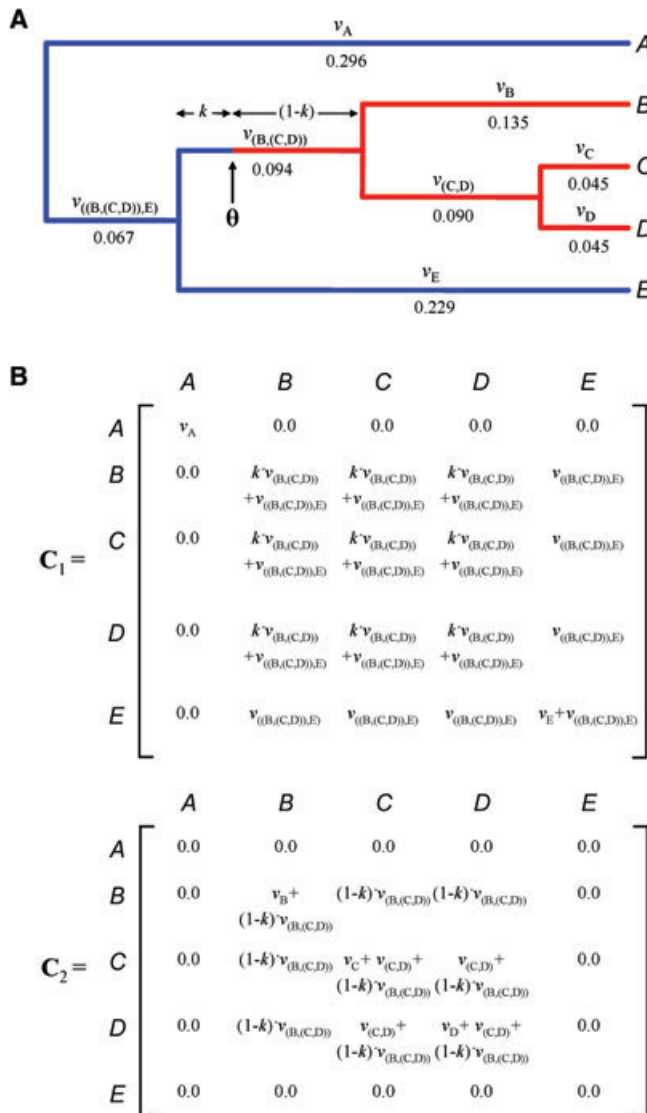


Figure 1. (A) Stochastic five taxon tree. Branch lengths (v) are shown above each edge, denoted by the node that they precede. The number below each edge is the fraction of total branch length in the tree represented by the overlying edge. (B) Calculation of C_1 and C_2 for branches painted with blue or red, respectively. Note the hypothesized shift from blue to red occurs fraction k along the branch leading to descendant species B, C, and D, and is indicated by θ .

with instantaneous rates σ_1^2 or σ_2^2 ; however, we will not focus on this assumption because other evolutionary models are certainly possible within this general framework.

We designed our MCMC run as follows. We first initialized the chain with starting values for the parameters in the model. These parameter values can optionally be supplied by the user; but to improve computational performance, by default our implementation is programmed to choose reasonable starting values for the parameters (described below). For each generation of the

chain, we proceeded to cyclically update each parameter in the model (i.e., to say, updating all four parameters took four generations) with a random step from a proposal distribution for that parameter. We used Gaussian proposal distributions, centered on zero, for changes to the rates (σ_1^2 and σ_2^2) and the ancestral value (α), and we used a symmetric exponential distribution (i.e., an exponential distribution whose density has been halved and reflected across the ordinate) for changes to the shift point (θ). Negative values, in this case, meant a change toward the root of the tree. We also tried a Gaussian proposal distribution for θ , although this made little difference on the test datasets that we analyzed. We have left the symmetric exponential as the default because we believe that it will allow for a more thorough exploration of the tree by the MCMC chain. In our implementation of this MCMC algorithm, the variances of each proposal distribution can be specified by the user.

Changes to θ that are larger than the remaining length of the current edge (i.e., branch) also require one or multiple decisions, according to the following algorithm: (1) if going rootward down the tree, and not at the root, we proceeded to the parent edge or the other daughters (allowing for multifurcation) all with equal probability; (2) if going tipward up the tree and not at a tip, we proceeded to either daughter edge with equal probability; (3) if at the root, we proceeded to any other daughter edge with equal probability; and, finally, (4) if at a tip, we reflected the change back along the tip edge exactly the distance it would have otherwise exceeded the terminal node.

If allowed to proceed as an unhindered random walk on the tree, this algorithm will eventually sample all edges on the tree with a probability directly proportional to their lengths (not shown). To ensure proper mixing, we also allowed a small fraction of steps (say, 5%, but this can also be modified by the user) to result in a move to a randomly selected branch with probability proportional to its length. Half of the time that such a move was performed we also switched the values of σ_1^2 and σ_2^2 .

Our proposal distributions for σ_1^2 , σ_2^2 , and α are symmetric. Proof of symmetry of the proposal distribution for θ is given in Appendix S2. Symmetry of the proposal distributions is an important property because it allows us to set the Hastings ratio to 1.0 (Hastings 1970; Yang 2006; see below).

Given a new proposed value for a parameter, $\sigma_1^2 t$, we accepted this change under two conditions: (1) if it increased the posterior probability; or (2) if it decreased the posterior probability but satisfied the following inequality:

$$\frac{L(\sigma_1^2 t, \sigma_2^2, \alpha, \theta) \Pr(\sigma_1^2 t, \sigma_2^2, \alpha, \theta)}{L(\sigma_1^2, \sigma_2^2, \alpha, \theta) \Pr(\sigma_1^2, \sigma_2^2, \alpha, \theta)} > r \sim U(0, 1),$$

where $r \sim U(0, 1)$ is a random variate drawn from a uniform distribution on the interval (0,1), and $\Pr(\sigma_1^2, \sigma_2^2, \alpha, \theta)$ is the prior

probability of our parameter values. The chance of accepting a change that decreases the posterior probability of the model is thus exactly equal to the ratio of posterior probabilities of the current and updated models. This is the standard form of the Metropolis–Hastings MCMC algorithm (Metropolis et al. 1953; Yang 2006) with the Hastings ratio set to 1.0.

In the present study, we used a log-normal prior probability distribution centered on 0.0 for the ratio σ_1^2/σ_2^2 , and a uniform prior on the log-scale for the geometric mean of σ_1^2 and σ_2^2 . We also used an unbounded uniform prior for α , and a uniform prior for θ . For θ , this means that the prior probability of the shift point being on any edge of the tree is exactly proportional to the length of that edge. We explored various distributions for the prior on σ_1^2 and σ_2^2 and our analyses did not seem to be especially sensitive to the prior, except for under one specific set of conditions that we will discuss at greater length below (see Discussion).

In our implementation of this method, the user can supply randomly or nonrandomly chosen starting parameter values for σ_1^2 , σ_2^2 , and α to initialize the MCMC run. However, by default, we used a previously derived analytic solution for the maximum likelihood estimate (MLE) of the evolutionary rate under a single rate regime (i.e., $\sigma_1^2 = \sigma_2^2 = \text{MLE}(\sigma^2)$; for example, O’Meara et al. 2006) and we used the MLE of the root node ancestral value under a single rate for α (Rohlf 2001; O’Meara et al. 2006). This was done mainly to improve computational performance and decrease burn-in by starting with values for σ_1^2 , σ_2^2 , and α that we might expect to be roughly of the correct magnitude. To initialize θ , we randomly selected a location for the shift point between rate regimes on the tree. The probability of choosing a shift point along any branch on the tree was set to be proportional to the branch length; meaning, for instance, that one would be exactly twice as likely to start with a random shift point on a branch with length $2v$ than on one with length v . This is essentially equivalent to choosing a random value of θ from our prior probability distribution for θ .

Each posterior sample for this analysis consists of values for σ_1^2 , σ_2^2 , α , and θ ; a value for the log-likelihood; and a list of the tip labels for the set of tips in state σ_2^2 . Because σ_1^2 and σ_2^2 depend both on θ and the set of tips in state σ_2^2 (i.e., whether σ_1^2 or σ_2^2 is the derived rate), we propose the following algorithm for preprocessing the posterior sample from our MCMC run. First, we found the median shift point in the posterior sample. This was done by identifying the sampled point with the minimum summed distance to all the other points in the sample (although other options for this are certainly possible, see Discussion). Next, we went through each sample in the posterior, splitting the tree at the shift point for that particular sample, and then assigning the derived and ancestral rates to edges or fractions of edges in the ancestral and derived subtrees, respectively (this might be σ_1^2 and σ_2^2 , or σ_2^2 and σ_1^2 , depending on the membership to our list

of labels for that sample). Finally, we reattached the two subtrees and computed the average rates rootward and tipward from the median shift point. Note that the collection of edges and fractions of edges rootward or tipward of the median shift point can include one or both rate categories depending on how the estimated shift point for that sample differs from the median shift point. For consistency across samples, we now assigned σ_2^2 always as the derived rate and σ_1^2 as the ancestral rate.

SIMULATED EXAMPLE

We generated and analyzed a simulated phylogenetic tree and phenotypic dataset to illustrate the application and results of our method. This simulation also forms the basis for our performance analysis of the method, below (see section Performance analysis).

We first simulated a stochastic pure-birth phylogeny with 100 terminal species. We then randomly selected a position on the tree as the location of the rate shift for our quantitative character. This evolutionary scenario is illustrated by the colored branches of the phylogenetic tree of Figure 2. The rate shift is located in a random position on the labeled branch “147,” where branches are identified by the number of the descendant node, and nodes are numbered according to the conventions of “phylo” objects in the “ape” phylogenetics package for R (Paradis et al. 2004; Paradis 2006). We next evolved a continuous character on the phylogeny under Brownian motion with the starting value $\alpha = 0.0$ at the root of the tree. The simulated evolutionary process had instantaneous rates $\sigma_1^2 = 1.0$ on edges rootward of the shift point (blue branches in Fig. 2), and $\sigma_2^2 = 10.0$ on edges tipward of this point (red branches).

For the MCMC run, we set the following control parameters. We set the standard deviations of the Gaussian proposal distributions for σ_1^2 and α to 0.5, and the standard deviation of the proposal distribution for σ_2^2 to 1.0. We set the rate parameter (λ) for the exponential proposal distribution for shift point moves to $\lambda = 5.0$. Because random deviates from the exponential proposal distribution for tree moves were also assigned random sign with equal probability, the realized proposal distribution for tree moves has the following density: $f(x) = \frac{\lambda}{2}e^{-\lambda \cdot x}$ for $x \geq 0$ and $f(x) = \frac{\lambda}{2}e^{\lambda \cdot x}$ otherwise. We also set the probability of proposing a move to a random point in the tree to 0.05. We set the variance of the log-normal prior for the rate ratio σ_1^2/σ_2^2 to 2.0; finally, we used a uniform prior for α and θ .

We ran the Metropolis–Hastings MCMC algorithm for 100,000 generations, sampling every 10 generations. Figure 3A shows the trace of the log-likelihood sampled every 100 generations (i.e., every 10 samples) from the entire MCMC run. In this example, we can see that the chain converges rapidly. We then preprocessed the posterior sample, as discussed above. Figures 3B and 3C show the frequency histograms of the posterior

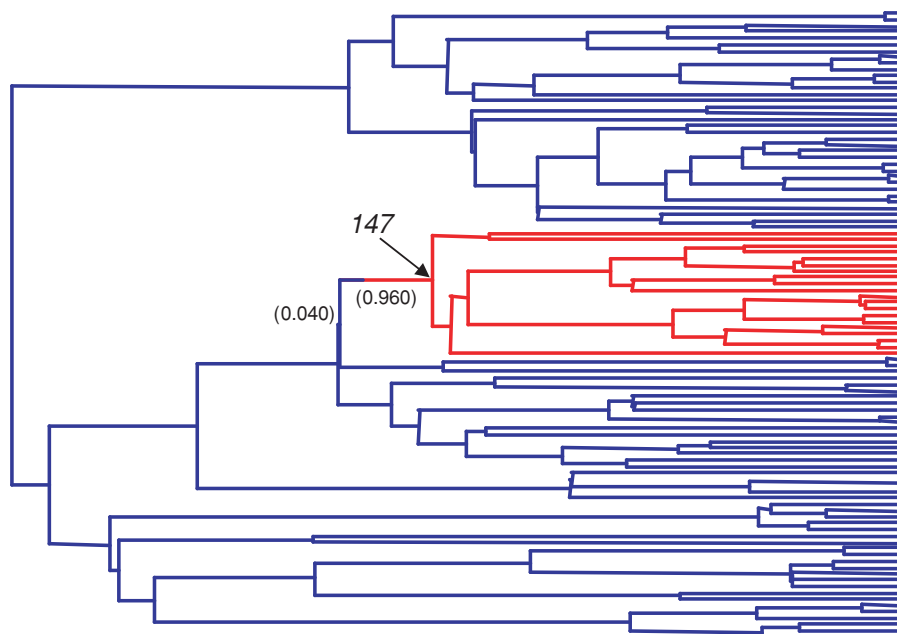


Figure 2. Stochastic, 100 taxon tree used for the simulated example. Phenotypic data were generated on this tree with a 10-fold higher evolutionary rate along the branches painted in red. The node tipward of the rate shift (numbered “147,” by the “phylo” convention in “ape”) is also indicated. Numbers presented in parentheses below or adjacent to branches are the posterior probabilities that the rate shift occurred on each labeled edge from the illustrative example. Only posterior probabilities ≥ 0.001 are reported.

samples obtained after preprocessing the posterior samples for σ_1^2 and σ_2^2 , with the first 10,000 generations excluded as burn-in. We computed effective sample sizes (ESSs) and 95% credible intervals (CIs) for the mean of the posterior distribution for σ_1^2 , σ_2^2 , and α (Table 1). This can be done quite easily using the R package “coda” (Plummer et al. 2010) or, alternatively, in the Java program “Tracer” (Rambaut and Drummond 2009) that has the benefit of a very user friendly graphical interface. We recommend ESSs for the evolutionary rates of at least 100. If an ESS less than 100 is obtained, then the MCMC can be rerun and the post burn-in samples combined (e.g., Ho et al. 2007). We computed the estimated values of σ_1^2 , σ_2^2 , and α as the mean of the preprocessed posterior sample (excluding the burn-in). All were very close to the generating conditions here (Table 1). The choice of the posterior arithmetic mean as an estimator is arbitrary. We might instead compute the posterior median or geometric mean (although in this example, the arithmetic means of the posterior sample are quite close to the generating parameter values). The ESSs for σ_1^2 and α were quite high, indicating relatively low autocorrelation in the posterior samples for these parameters; however, the ESS was considerably lower for σ_2^2 , suggesting that we might do better by adjusting the variance of the proposal distribution for this parameter.

We also computed an approximate median rate shift point by computing all pairwise distances between shift points in the post burn-in posterior sample, and then selecting the shift point with the minimum summed distance to all the other points in the sample,

as described above. This value, which corresponds very closely to the generating shift point in this example, is also reported in Table 1. This procedure will not be computationally feasible for very long MCMC runs; however, in that case one could instead use a sparser sample of shift points from the posterior (taken, say, every 100 or 1000 generations, instead of every 10 generations as in this example). Finally, we computed the posterior probability of the shift point being on each edge of the tree. For all edges with posterior probability ≥ 0.001 , we have plotted these probabilities below or adjacent to the corresponding branches in Figure 2. Nearly, all (96%) of the posterior density for the location of the rate shift is on the generating edge in this case (Fig. 2).

PERFORMANCE ANALYSIS

To assess the performance of the method more generally, we conducted two sets of simulation tests of the method. First, we conducted the following simulation 80 times in total (20 times for each of the four sets of generating rates, described below): (1) We simulated a stochastic, pure-birth, $N = 100$ species phylogenetic tree with branch lengths. (2) We picked a shift point at random on the tree. Although in theory our method should be appropriate to detect evolutionary rate shifts in subclades of any size, we anticipate that the method will suffer from low power when the number of species rootward versus tipward of the rate shift is extremely unbalanced. Thus, to avoid this issue in our early performance analysis of the method, we decided to exclude randomly chosen shift points with fewer than 20 or more than

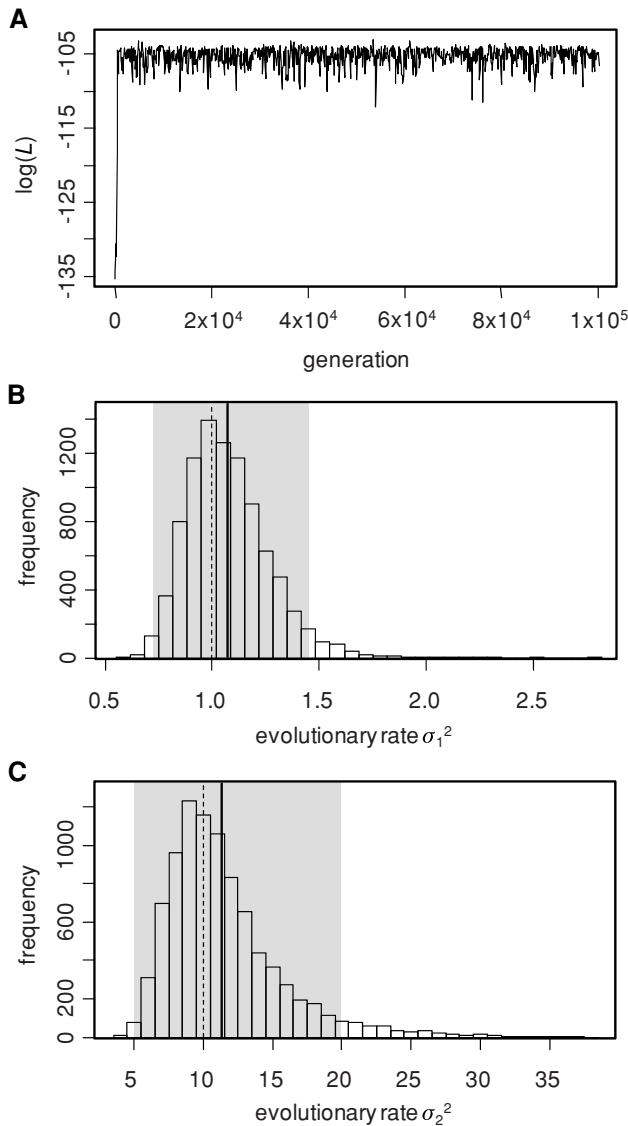


Figure 3. Results from the simulated example. (A) Trace of $\log(L)$ by generation number for the 100,000 generation MCMC analysis. The $\log(L)$ was sampled every 100 generations. (B) Frequency histogram of the post burn-in posterior sample for σ_1^2 . The generating value of $\sigma_1^2 = 1.0$ is indicated by the vertical dashed line. The mean from the posterior sample is given by the vertical solid line, whereas the 95% credible interval (CI) is given by the shaded area. (C) Same as (B), but for σ_2^2 . The generating condition, in this case, was $\sigma_2^2 = 10.0$.

80 descendant species (in other words, we excluded splits for simulation in which more than 80% of the taxa in the phylogeny were on one side of the split). (3) We then simulated data on the tree with generating conditions as follows: the branches ancestral to the shift point evolved with rate $\sigma_1^2 = 1.0$, whereas the derived branches evolved with rates $\sigma_2^2 = 0.1, 1.0, 5.0,$ or 10.0 (20 simulations each). (4) We initiated the MCMC chain as in the illustrative example, above, and ran the chain for 100,000 generations. Nor-

mally users would probably run multiple MCMC chains, adjusting the control parameters to ensure proper mixing and convergence. Here, we merely adjusted the control parameters (mainly the variances of the proposal distributions or the number of generations in the chain) and reran any MCMC for which the ESSs for σ_1^2 or σ_2^2 was less than 100. We did this until we obtained ESSs greater than 100 for all runs. (5) We computed summary statistics on the posterior sample, excluding the first 20,000 generations of the sample. In addition to the summary measures reported in Table 1, we also computed the patristic distance between the inferred shift point ($\hat{\theta}$) and the generating value of θ for each replicate (i.e., the minimum edge distance connecting the two points in the tree).

A summary of the results from these analyses is given in Table 2. Results for all of the 80 simulations are in Appendix S3. For each set of simulation conditions, Table 2 gives the arithmetic means of σ_1^2 , σ_2^2 , and α (the geometric means and medians for σ_1^2 and σ_2^2 are reported in Appendix S3), the proportion of simulations in which the correct node was inferred, and the fraction of simulations in which the 95% CIs for σ_1^2 did not overlap $\hat{\sigma}_2^2$ (our estimate of σ_2^2) and vice versa. This latter frequency is analogous to the statistical “power” of the method (or its type I error, for the generating conditions $\sigma_1^2 = \sigma_2^2 = 1.0$; alternatively, 1.0 minus this fraction is the type II error rate of the method if $\sigma_1^2 \neq \sigma_2^2$). This procedure is somewhat ad hoc as the model itself explicitly assumes that $\sigma_1^2 \neq \sigma_2^2$; however, the results of Table 2 and Appendix S1 suggest that we will only be infrequently misled to believe that $\sigma_1^2 \neq \sigma_2^2$ if they are in fact equal. We also report the mean distance between $\hat{\theta}$ and θ for each simulation condition. In general, parameter estimates are pretty good, and 95% CIs nearly always included the generating parameter value. The method also has excellent success in identifying the position of the rate shift to a specific edge in the tree, particularly when the proportional difference between σ_1^2 and σ_2^2 was high (Table 2).

Second, we also explored the performance of the method on smaller and larger phylogenies than the stochastic $N = 100$ trees described above. To do this, we used the following procedure: (1) We simulated 20 pure-birth phylogenies with each of the following sizes $N = 30, 50, 70,$ and 200 . (2) On each tree, we chose a random rate shift location such that no less than 20% and no more than 80% of the species in the tree were found tipward from that point. (3) We then simulated the evolution of a continuously valued character with $\sigma_1^2 = 1.0$ and $\sigma_2^2 = 10.0$ as the ancestral and derived rates, respectively. (4) We ran our MCMC chain on each simulated dataset and tree, using the conditions described previously, and then computed summary measures from the posterior sample. Again, we reran MCMCs for which ESSs of either σ_1^2 or σ_2^2 was less than 100.

The results from these analyses are summarized in Table 3 and specific results from all analyses are given in Appendix S3.

Table 1. Results from analysis of the simulated example. The generating and inferred shift point (θ) is indicated by the tipward node ("147" in this case) and (absolute, not relative) position of the rate shift along the edge.

Parameter	Generating value	Estimate (mean from posterior sample)	Effective sample size	95% credible interval
σ_1^2	1.0	1.074	2172	(0.7373,1.4440)
σ_2^2	10.0	11.43	247.0	(5.0812,19.8603)
α	0.0	0.042	3520	(-0.7087,0.7484)
θ	[147,0.0084]	[147,0.0536]	–	–

In general, we found that the method had remarkable success in identifying the location of the rate shift in the tree. Evolutionary rates were biased on small trees (in particular, such that $\hat{\sigma}_1^2$ and $\hat{\sigma}_2^2$ were more similar) but this bias nearly vanishes for the larger simulated phylogenies in the study.

EMPIRICAL TEST

Finally, we also examined the performance of the method using an empirical dataset and tree. We analyzed the evolution of body size (measured as log-SVL: "snout-to-vent length") in a 32 species subtree extracted from the 100 taxon *Anolis* phylogeny of Mahler et al. (2010). We chose to analyze this subtree rather than the whole Caribbean *Anolis* phylogeny because the results of prior studies (e.g., Butler and King 2004) suggest that more than two different evolutionary processes might govern the evolution of this lizard group in the Caribbean. We focused on the subtree given

in Figure 4, which contains the *Anolis sagrei* group on Cuba; the *A. distichus* group on Hispaniola; *A. cristatellus* and related Puerto Rican lizards; and, finally, the endemic radiation of six *Anolis* species on Jamaica.

We optimized the MCMC as follows. We used Gaussian proposal distributions for σ_1^2 , σ_2^2 , and α , with variances 0.015, 0.050, and 0.90. We used a prior on the ratio of the log-transformed rates with variance of 4.0. A summary of the results from this analysis is given in Figure 4. We found that the vast majority of the posterior density for a rate shift in our model was found either at the base of or within the Jamaican radiation. The estimated rate tipward of the shift was about 8.5 times higher than the estimated rate rootward of the shift. Note that Jamaica is the only island in this phylogeny that is hypothesized to have been colonized de novo in this subtree (see Mahler et al. 2010). Consequently, this result is consistent with our impression based on prior studies that

Table 2. Summary of results from the performance analysis. $\bar{\sigma}_1^2$ and $\bar{\sigma}_2^2$ are the mean parameter estimates across simulations. "ESS" denotes the mean effective sample size across simulations. "On CI" indicates the fraction of times that the generating value was on the 95% credible interval for σ_1^2 or σ_2^2 . "No CI overlap" means the proportion of simulations where the 95% CI for σ_1^2 does not include $\hat{\sigma}_2^2$ and vice versa. "Correct edge" indicates the fraction of times in which the correct edge in the tree was identified by this method. Finally, "distance" indicates the mean distance from the true shift point to its inferred value. For this calculation, the total tree was rescaled to unit length. Values in parentheses are standard deviations across simulations. Obviously, "correct edge" and "distance" are meaningless when no rate heterogeneity is simulated; however, we nonetheless include these measures (marked with an asterisk) because they serve to demonstrate that the MCMC algorithm does not converge to a randomly selected branch in the absence of a simulated rate shift on that branch.

	Simulation 1: $\sigma_1^2=1.0$, $\sigma_2^2=0.1$	Simulation 2: $\sigma_1^2=1.0$, $\sigma_2^2=1.0$	Simulation 3: $\sigma_1^2=1.0$, $\sigma_2^2=5.0$	Simulation 4: $\sigma_1^2=1.0$, $\sigma_2^2=10.0$
$\bar{\sigma}_1$ (mean) ² (SD)	1.05 (0.175)	1.06 (0.172)	1.241 (0.389)	1.16 (0.158)
$\bar{\sigma}_2$ (mean) ² (SD)	0.129 (0.0618)	1.08 (0.173)	4.326 (1.463)	10.66 (3.522)
$\bar{\alpha}$ (SD)	-0.137 (0.132)	0.013 (0.409)	0.0400 (0.0331)	-0.083 (0.057)
ESS(σ_1^2)	679.6	691.0	324.0	746.4
ESS(σ_2^2)	618.1	633.8	368.8	419.4
ESS(α)	773.8	780.2	806.5	780.8
On CI(σ_1^2)	1.00	0.95	0.80	1.0
On CI(σ_2^2)	0.85	0.95	0.80	0.9
No CI overlap	1.00	0.00	0.75	1.0
Correct edge	0.95	0.15*	0.70	0.85
Distance (SD)	0.0531 (0.0473)	0.195* (0.146)	0.0821 (0.0533)	0.0533 (0.0587)

Table 3. Summary of the results from the test of power. $\bar{\sigma}_1^2$, $\bar{\sigma}_2^2$, “ESS,” “on CI,” “no CI overlap,” “correct edge,” and “distance” are defined as in Table 2. Values in parentheses are standard deviations across simulations.

	$N=30$	$N=50$	$N=70$	$N=200$
$\bar{\sigma}_1(\text{mean})^2$ (SD)	1.93 (0.909)	1.33 (0.414)	1.17 (0.353)	1.066 (0.146)
$\bar{\sigma}_2(\text{mean})^2$ (SD)	8.80 (5.20)	10.17 (5.09)	10.02 (3.20)	9.54 (1.34)
$\bar{\alpha}$ (SD)	0.0955 (0.401)	-0.156 (0.404)	-0.029 (0.466)	0.020 (0.493)
ESS(σ_1^2)	483.6	526.7	683.8	763.0
ESS(σ_2^2)	393.1	277.4	314.6	558.2
ESS(α)	815.0	858.1	859.1	828.3
On CI(σ_1^2)	0.90	1.00	0.95	0.95
On CI(σ_2^2)	0.85	0.85	0.90	1.00
No CI overlap	0.55	0.85	0.95	1.00
Correct edge	0.80	0.85	0.95	0.90
Distance (SD)	0.173 (0.232)	0.133 (0.144)	0.062 (0.060)	0.057 (0.048)

the evolutionary rate is higher on newly colonized islands when ecological opportunity is high (Mahler et al. 2010).

Discussion

Herein, we develop a Bayesian MCMC approach for the analysis of rate variation in continuously valued characters in the context of a phylogenetic tree. This method is an innovation over previous related maximum likelihood techniques (e.g., O’Meara et al. 2006; Thomas et al. 2006) because it allows us to remain naive about how evolutionary rate variation is distributed among the branches of the phylogeny, and specifically, when in time a change in the evolutionary rate has occurred. However, the model we present herein is quite simple: one and only one rate shift is allowed on the phylogeny (in contrast to O’Meara et al. 2006 where multiple shifts are permitted, although they need to be specified a priori).

The method works quite well in general as we were able to estimate the branch of the shift in evolutionary rate for a continuous character with quite high accuracy. Unsurprisingly, we were able to estimate the correct edge with higher success under conditions where the evolutionary rate shift was proportionally largest (Table 2). Notably, though, for a rate shift of a given size the probability of inferring the correct branch did not increase for larger trees (Table 3). Although the mean patristic distance from the inferred shift point to its generating location did in fact decline for larger tree sizes, one should keep in mind that as we scaled all trees to have a common length of 1.0, merely identifying the correct edge in the tree virtually guarantees that one will have identified a phylogenetic position closer to the generating shift point.

Unfortunately, although we were usually able to narrow down the position of the rate shift to the correct edge, the method had much less success in locating the shift to a specific point along that

edge. For instance, Figure 5 shows our estimate of the posterior density for the shift position conditioned on the edge being the correct edge from the illustrative example. The posterior density is essentially flat on the interval (although it inclines somewhat towards the tipward side of the edge—opposite to the generating shift point in this case). We are not sure why our data for tip species can contain so much information regarding the edge on which the evolutionary rate changes, but so little about where on that edge it changes; however, we plan to explore this issue in greater depth with future studies.

Our parameter estimates for the evolutionary rates appear to be asymptotically unbiased for large sample sizes and large evolutionary rate shifts. That is to say, for large trees and a single large rate shift, the arithmetic mean of the posterior sample of evolutionary rates rootward and tipward of this shift will be unbiased estimates of the generating rates. However, for other circumstances, the evolutionary rates estimated in this way range from being slightly to substantially biased. We suspect that the bias in these situations comes from two sources.

First, the bias tends to cause the estimated rates, $\hat{\sigma}_1^2$ and $\hat{\sigma}_2^2$, to be more similar than their underlying generating values. We see this in all of the simulations with fewer than 100 species, as well as for our simulations with $\sigma_2^2 = 5.0$ (Table 2, 3). We believe that this is just an inherent quality of integrating over uncertainty in the location of the evolutionary rate. Consider, for instance, the evolutionary scenario illustrated in Figure 2. In this case, if the rate shift for a given sample is located one edge tipward of the true shift point (and thus the ancestral rate contains some red branches), then the posterior probability will be highest, and thus the sample more likely to be retained, if the ancestral rate σ_1^2 is relatively high; however, σ_2^2 , the derived rate, contains only red branches and will be unaffected. Conversely, if the shift point for a different sample is located one edge rootward of the true shift point (and thus the derived rate contains some

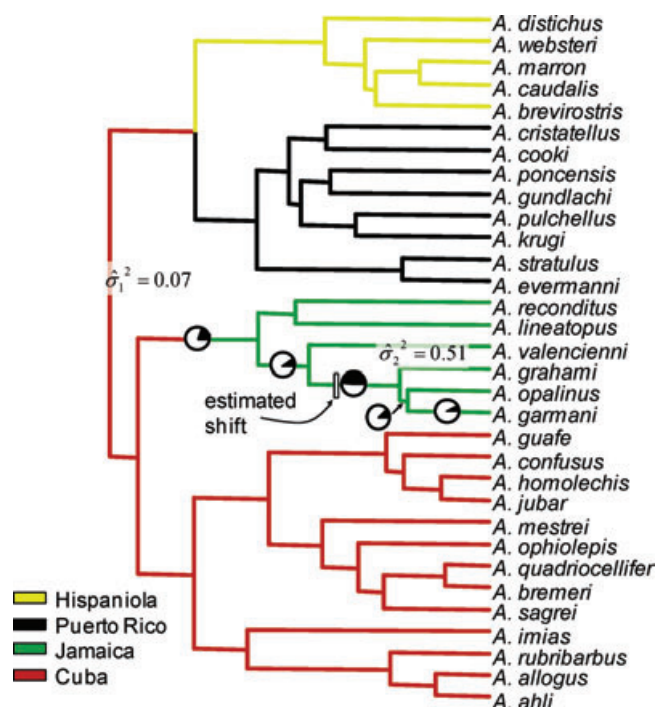


Figure 4. Phylogeny of a subtree from the radiation of Caribbean *Anolis*. Rates and posterior densities are from an analysis of body size evolution on this tree. Posterior probabilities of the rate shift being on each edge of the tree (if > 0.01) are represented by the filled fraction of each pie graph. Most of the posterior density for a rate shift suggests an increase in the evolutionary rate at the base of or within the Jamaican diversification of *Anolis*. This finding is consistent with prior studies showing that the rate of evolution is increased on newly colonized islands, as Jamaica is the only island that was colonized de novo in this subtree (all other inferred colonizations are secondary or back-colonizations; see Mahler et al. 2010). The total tree length is scaled to 1.0 in this example.

blue branches), then the posterior probability will be highest if the derived rate σ_2^2 is relatively low; however, the ancestral rate contains only blue branches and will be unaffected. Given that both of these conditions will cause the sampled rates σ_1^2 and σ_2^2 to be more similar, this effect could account for the pattern we see in Tables 2 and 3. To avoid this, we might be tempted to average the posterior sample of rates but condition on the inferred edge; however, we advise against this because for all empirical studies the generating rate shift location is unknown and conditioning on the edge effectively ignores uncertainty in the evolutionary rates that is due to uncertainty about where the rates have changed over time. Very large rate shifts and large tree sizes effectively concentrate the posterior density for the rate shift on one or a very small number of edges (e.g., Fig. 2), thus mitigating this issue.

We also perceived that the evolutionary rates were sometimes upwardly biased even for very large rate shifts and relatively species rich trees. We are unsure of the source of this bias,

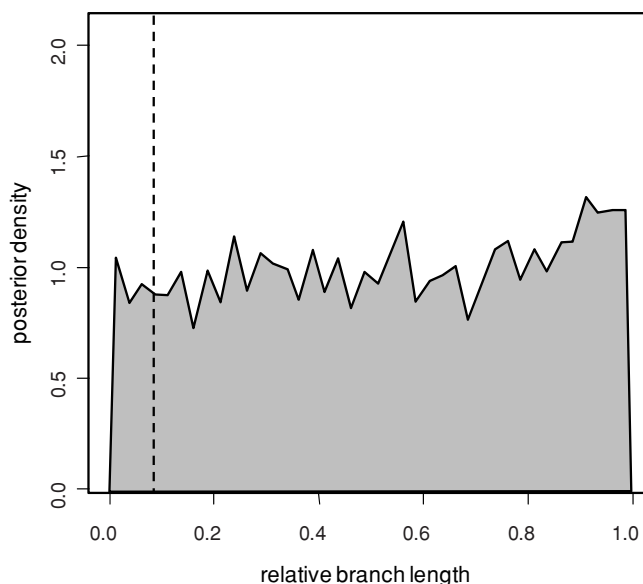


Figure 5. Posterior density for the phylogenetic position of the change in evolutionary rate, conditioned on the edge and standardized to a total branch length of 1.0. The vertical dashed bar indicates the relative position of the generating rate shift along that branch. Essentially, the posterior density is uniform on the edge length.

but note that it appears to decline (if not disappear) for larger trees (e.g., Tables 2 and 3). In this case, we suspect that the bias might be due to averaging the posterior sample of rates with the arithmetic rather than the geometric mean. The arithmetic mean “penalizes” sampled evolutionary rates that are half or double the generating value unequally—putting more weight on the doubled value in this case. Because these samples are equally bad, by some measure, computing the geometric mean (in which halved and doubled samples are penalized equally) might be a more appropriate summary of the posterior sample. Indeed, computing the geometric means of the posterior sample of rates does reduce the bias in parameter estimation (Appendix S3).

In the present study, we have focused on the simplest evolutionary model for a continuously valued trait: Brownian motion. However, the general approach developed herein should be useful in analyzing other more complex evolutionary scenarios. For instance, it might also be helpful in analyzing variability in the selection regime over time (e.g., Hansen 1997; Butler and King 2004). In that case, our model for evolutionary heterogeneity would be an Ornstein–Uhlenbeck process with an optimum that shifts at some point in the tree (Butler and King 2004). Our model parameters would then be the phenotypic locations of the adaptive optima, as well as the shift point between evolutionary regimes (as in this study).

We analyzed only a single character in the present article. It is conceivable to extend the general approach to multiple coevolving traits. For instance, Revell and Collar (2009) fit a model in

which the evolutionary correlation between two aspects of buccal morphology changes on certain branches of the phylogenetic tree of centrarchid fishes. Using the approach developed in this paper, we could instead ask if the evolutionary correlation between traits changes in the history of our group, and if so, when and in which ancestral lineage.

In addition to modeling only one trait, we also analyzed only a single shift point between rates. In the future, we might use the same general approach represented herein to model multiple shifting points between two or more than two evolutionary rates or processes. This makes biological sense to explore, because our method does not presently allow for homoplasy in the derived evolutionary rate (i.e., Fig. 2). In the real world, we certainly expect that the evolutionary rate would change multiple times or exhibit reversal in some derived lineages.

Nonetheless, we feel that this new approach takes an important step forward in many regards. In particular, it enables us to ask if a subclade of our tree has diversified exceptionally in their phenotypes, even if we had no a priori reason to have identified that group as unusual. This method also effectively circumvents the circularity associated with asking if a clade of interest (special, perhaps, in its perceived diversity) has diversified especially. Use of this method should also avoid the so-called “trickle-down effect” of Moore et al. (2004). According to this effect, a shift in the species diversification rate (or, in this case, the rate of evolution for a continuous character) that occurs in a nested position within a clade can create the illusory impression that the entire clade has diversified exceptionally, that is, if this was our a priori hypothesis for phenotypic divergence. Our method does not require a specific prior phylogenetic hypothesis for variation in the evolutionary rate through time and thus should not be subject to this problem. This method will thus be particularly useful for identifying the exceptional diversification of adaptive radiation by being able to distinguish a shift in the evolutionary rate at the root of our putative adaptive radiation from a shift located in a more nested phylogenetic position.

Exceptional diversification need not only come in the form of exceptionally large diversity. For instance, we might also ask if there is evidence of derived constraint (e.g., Revell and Collar 2009; Lavoué et al. 2011). Indeed, the method presented herein is equally well suited to identify an exceptionally low rate as it is an exceptionally high rate of evolution, and in fact, we have explored that in some of our simulations (Table 2).

Although we present only one form for the prior distribution for σ_1^2 and σ_2^2 here, we explored several different priors when developing this method. For instance, we put an exponential prior on σ_1^2 and σ_2^2 separately, with various values for λ . What we discovered is that the method is not sensitive to the prior so long as $|\log(\sigma_2^2/\sigma_1^2)|$ is large (in other words, when rates are very different). However, the method becomes extremely sensitive to the

prior as $|\log(\sigma_2^2/\sigma_1^2)|$ approaches zero (i.e., when there is no phenotypic evolutionary rate variation in the tree). In particular, for an exponential prior under these conditions, the posterior distribution becomes an approximately even mixture of reasonable values for the single global rate, and the prior distribution. We found that putting a prior on the log-ratio of σ_1^2 and σ_2^2 had reasonably good behavior under all conditions, including when no real rate variation was simulated.

In the present article, we “averaged” the shift points by simply identifying the posterior sample with the minimum summed distance to all the sampled shift points in the set of all posterior samples. This is only one possible algorithm to find the median of these points. Alternatively, we might identify the point on the tree that minimizes the summed or squared summed distances to the other shift points, or we might first identify all posterior samples on the modal branch and then average only that subset. All of these options will produce different values for the median shift point in the posterior distribution. Because the median shift point is used to preprocess the posterior sample before analysis (see Methods), different methods for calculating it will also result in a different posterior sample of rates.

Adaptive radiations, the rapid divergence and phenotypic diversification of a clade, are thought to be of central importance in the origin of new species and morphologies (Simpson 1953; Givnish and Systma 1997; Schluter 2000; Glor 2010; Losos and Mahler 2010). Phylogenetic tools have been developed that can naively identify exceptional species diversification (Alfaro et al. 2009; Stadler 2011), but no comparable method has yet been presented for phenotypic divergence. Herein, we propose such a method that is based on Bayesian MCMC. Although we focus on identifying exceptional diversification under a Brownian process, we anticipate that the general approach developed herein will be equally useful for other evolutionary models as well, such as adaptive evolution toward different fitness optima in different parts of the phylogeny (e.g., Butler and King 2004).

ACKNOWLEDGMENTS

LR, BR, and LM received considerable support for this work from the National Evolutionary Synthesis Center (NSF EF-0905606). PP-N would like to acknowledge funding support from the National Science and Engineering Research Council of Canada (NSERC).

LITERATURE CITED

- Alfaro, M. E., F. Santini, C. Brock, H. Alamillo, A. Dornburg, D. L. Rabosky, C. Carnevale, and L. J. Harmon. 2009. Nine exceptional radiations plus high turnover explain species diversity in jawed vertebrates. *Proc. Nat. Acad. Sci.* 106:13410–13414.
- Blomberg, S. P., T. Garland Jr., and A. R. Ives. 2003. Testing for phylogenetic signal in comparative data: behavioral traits are more labile. *Syst. Biol.* 57:717–745.
- Bokma, F. 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.
- 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.
- Felsenstein, J. 1973. Maximum likelihood estimation of evolutionary trees from continuous characters. *Am. J. Hum. Genet.* 25:471–492.
- Felsenstein, J. 1985. Phylogenies and the comparative method. *American Naturalist* 125:1–15.
- Freckleton, R. P., and P. H. Harvey. 2006. Detecting non-Brownian trait evolution in adaptive radiations. *PLoS Biol.* 4:2104–2111.
- Garland, T. Jr. 1992. Rate tests for phenotypic evolution using phylogenetically independent contrasts. *Am. Nat.* 140:509–519.
- Givnish, T. J. 1997. Adaptive radiation and molecular systematics: issues and approaches. Pp. 1–54 in T. J. Givnish and K. J. Systma, eds. *Molecular evolution and adaptive radiation*. Cambridge Univ. Press, Cambridge, UK.
- Givnish, T. J., and K. J. Systma (eds.). 1997. *Molecular evolution and adaptive radiation*. Cambridge Univ. Press, Cambridge, UK.
- Glor, R. E. 2010. Phylogenetic insights on adaptive radiation. *Annu. Rev. Ecol. Evol. Syst.* 41:251–270.
- Hansen, T. F. 1997. Stabilizing selection and the comparative analysis of adaptation. *Evolution* 51:1341–1351.
- Harmon, L. J., J. T. Weir, C. D. Brock, R. E. Glor, and W. Challenger. 2008. GEIGER: investigating evolutionary radiations. *Bioinformatics* 24:129–131.
- Hastings, W. K. 1970. Monte Carlo sampling methods using Markov chains and their application. *Biometrika* 57:97–109.
- Ho, S. Y. W., B. Shapiro, M. J. Phillips, A. Cooper, and A. J. Drummond. 2007. Evidence for time dependency of molecular rate estimates. *Syst. Biol.* 56:515–522.
- Lavoué, S., M. Miya, M. E. Arnegard, P. B. McIntyre, V. Mamonekene, and M. Nishida. 2011. Remarkable morphological stasis in an extant vertebrate despite tens of millions of years of divergence. *Proc. R. Soc. B* 278:1003–1008.
- Losos, J. B., and D. B. Miles. 2002. Testing the hypothesis that a clade has adaptively radiated: iguanid lizard clades as a case study. *Am. Nat.* 160:147–157.
- Losos, J. B., and D. L. Mahler. 2010. Adaptive radiation: the interaction of ecological opportunity, adaptation, and speciation. Pp. 381–420 in M. A. Bell, D. J. Futuyma, W. F. Eanes, and J. S. Levinton, eds. *Evolution since Darwin: the first 150 years*. Sinauer Associates, Sunderland, MA.
- Maddison, W. P., P. E. Midford, and S. P. Otto. 2007. Estimating a binary character's effect on speciation and extinction. *Syst. Biol.* 56:701–710.
- Mahler, D. L., L. J. Revell, R. E. Glor, and J. B. Losos. 2010. Ecological opportunity and the rate of morphological evolution in the diversification of Greater Antillean anoles. *Evolution* 64:2731–2745.
- 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.
- McPeck, M. A. 1995. Testing hypotheses about evolutionary change on single branches of a phylogeny using evolutionary contrasts. *Am. Nat.* 145:686–703.
- Moore, B. R., K. M. A. Chan, and M. J. Donoghue. 2004. Detecting diversification rate variation in supertrees. Pp. 487–533 in O. R. P. Bininda-Emonds, ed. *Phylogenetic supertrees: combining information to reveal the tree of life*. Kluwer Academic, Dordrecht, The Netherlands.
- Olson, M. E., and A. Arroyo-Santos. 2009. Thinking in continua: beyond the “adaptive radiation” metaphor. *Bioessays* 31:1337–1346.
- 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.
- Paradis, E. 2006. *Analysis of phylogenetics and evolution with R*. Springer, New York.
- 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 diagnosis and output analysis for MCMC. *R News* 6:7–11.
- R Development Core Team. 2010. *R: a language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Rabosky, D. L. 2006. Likelihood methods for detecting temporal shifts in diversification rates. *Evolution* 60:1152–1164.
- Rambaut, A., and A. J. Drummond. 2009. Tracer: MCMC trace analysis tool v1.5.0. Available at: <http://tree.bio.ed.ac.uk/software/tracer/> (accessed July 19, 2011).
- Revell, L. J. 2008. On the analysis of evolutionary change along single branches in a phylogeny. *Am. Nat.* 172:140–147.
- . 2011. phytools: phylogenetic tools for comparative biology (and other things). R package version 0.0–4. Available at: <http://anolis.oeb.harvard.edu/~liam/R-phylogenetics>
- Revell, L. J., and D. C. Collar. 2009. Phylogenetic analysis of the evolutionary correlation using likelihood. *Evolution* 63:1090–1100.
- Rohlf, F. J. 2001. Comparative methods for the analysis of continuous variables: geometric interpretations. *Evolution* 55:2143–2160.
- Schluter, D. 2000. *The ecology of adaptive radiation*. Oxford Univ. Press, Oxford UK.
- Slowinski, J. B., and C. Guyer. 1993. Testing whether certain traits have caused amplified diversification: an improved method based on a model of random speciation and extinction. *Am. Nat.* 142:1019–1024.
- Simpson, G. G. 1953. *The major features of evolution*. Columbia Univ. Press, New York.
- Slater, G. J., S. A. Price, F. Santini, and M. E. Alfaro. 2010. Diversity versus disparity and the radiation of modern cetaceans. *Proc. R. Soc. B* 277:3097–3104.
- Stadler, T. 2011. Mammalian phylogeny reveals recent diversification rate shifts. *Proc. Nat. Acad. Sci.* 108:6187–6192.
- 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.
- Yang, Z. 2006. *Computational molecular evolution*. Oxford Univ. Press, Oxford, UK.

Associate Editor: M. Alfaro

Supporting Information

The following supporting information is available for this article:

Appendix S1: Computer code.

Appendix S2: Proof of symmetry of the proposal distribution for randomly selecting a new point on an evolutionary tree.

Appendix S3: Complete results from simulation test of the method.

Table S1. Summary of results from simulation 1., .

Table S2. Summary of results from simulation 2.

Table S3. Summary of results from simulation 3.

Table S4. Summary of results from simulation 3.

Table S5. Summary of results from simulations of trees with number of taxa $N = 30$.

Table S6. Summary of results from simulations of trees with number of taxa $N = 50$.

Table S7. Summary of results from simulations of trees with number of taxa $N = 70$.

Table S8. Summary of results from simulations of trees with number of taxa $N = 200$.

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

Please note: Wiley-Blackwell is not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.