

A NOVEL COMPARATIVE METHOD FOR IDENTIFYING SHIFTS IN THE RATE OF CHARACTER EVOLUTION ON TREES

Jonathan M. Eastman,^{1,2} Michael E. Alfaro,^{3,4} Paul Joyce,^{5,6} Andrew L. Hipp,^{7,8,9} and Luke J. Harmon^{10,11}

¹Department of Biological Sciences, Institute for Bioinformatics and Evolutionary Studies (IBEST), University of Idaho, Moscow, Idaho 83843-3051

²E-mail: jonathan.eastman@gmail.com

³Department of Ecology and Evolutionary Biology, University of California, Los Angeles, Los Angeles, California 90095

⁴E-mail: michaelalfaro@ucla.edu

⁵Departments of Statistics and Mathematics, University of Idaho, Moscow, Idaho 83843-3051

⁶E-mail: joyce@uidaho.edu

⁷The Morton Arboretum, 4100 Illinois Route 53, Lisle, Illinois 60532

⁸The Field Museum of Natural History, 1400 South Lake Shore Drive, Chicago, Illinois 60605

⁹E-mail: AHipp@mortonarb.org

¹⁰Department of Biological Sciences, University of Idaho, Moscow, Idaho 83843-3051

¹¹E-mail: lukeh@uidaho.edu

Received February 11, 2011

Accepted June 19, 2011

Evolutionary biologists since Darwin have been fascinated by differences in the rate of trait-evolutionary change across lineages. Despite this continued interest, we still lack methods for identifying shifts in evolutionary rates on the growing tree of life while accommodating uncertainty in the evolutionary process. Here we introduce a Bayesian approach for identifying complex patterns in the evolution of continuous traits. The method (*auteur*) uses reversible-jump Markov chain Monte Carlo sampling to more fully characterize the complexity of trait evolution, considering models that range in complexity from those with a single global rate to potentially ones in which each branch in the tree has its own independent rate. This newly introduced approach performs well in recovering simulated rate shifts and simulated rates for datasets nearing the size typical for comparative phylogenetic study (i.e., ≥ 64 tips). Analysis of two large empirical datasets of vertebrate body size reveal overwhelming support for multiple-rate models of evolution, and we observe exceptionally high rates of body-size evolution in a group of emydid turtles relative to their evolutionary background. *auteur* will facilitate identification of exceptional evolutionary dynamics, essential to the study of both adaptive radiation and stasis.

KEY WORDS: Bayesian inference, Chelonia, comparative methods, primates, rate heterogeneity, reversible-jump MCMC, trait evolution.

Explaining and identifying variation in rates of evolution among lineages has long been a focus for evolutionary biologists (e.g., Darwin 1859; Simpson 1953; Gingerich 1983; Eldredge and Stanley 1984; Estes and Arnold 2007). Simpson (1944, 1953) coined a set of terms for the very purpose of distinguishing exceptional

rates of evolution in particular lineages (i.e., tachytely, bradately, horotely). Current approaches require that shifts in the evolutionary process are identified a priori (e.g., Butler and King 2004; O'Meara et al. 2006; Revell and Collar 2009). What these and related methods (Harmon et al. 2003; Freckleton and Jetz 2009)

lack is the acknowledgment and accommodation of uncertainty in the evolutionary processes that give rise to trait values observed in extant taxa (but see Revell et al., in revision). The quality of inference is bounded by the quality of the models that we apply in the comparative framework: even the best among a set of inadequate models is still inadequate. Relaxing the assumption of static points in the tree where rates have shifted should diminish bias in branchwise estimates of evolutionary rates. Most ideal would be methods that limit influence of a priori expectations, while fairly comparing fit among the largest possible set of models and allowing robust inference of rate variation.

Drawing on recent progress in modeling molecular evolution (e.g., Huelsenbeck et al. 2004; Pagel and Meade 2008; Drummond and Suchard 2010), we model trait evolution by allowing phylogenetically localized shifts in the rate of a random-walk process of trait change. In a novel Bayesian approach, here referred to as AUTEUR, we allow the data to directly inform uncertainty in the estimated evolutionary process. AUTEUR (Accommodating Uncertainty in Trait Evolution Using R) samples across a broad set of possible trait evolution scenarios, considering models differing in number and topological position of local rate shifts.

We model trait evolution as a Brownian-motion process, which describes a broad set of neutral and nonneutral models of phenotypic evolution (Felsenstein 1973), and may be a reasonably adequate approximation of the evolutionary process in some lineages (Harmon et al. 2010). Under the modeled random-walk process, the trajectory of trait evolution (magnitude and directionality) is independent of the current state of the character. Whereas the expected value at the end of a random walk is simply the starting value, variance in traits accumulates in proportion to both the extent of independent evolution in lineages and the evolutionary rate of the character (see Felsenstein 1973; O'Meara et al. 2006; Revell et al. 2008). We thus expect little trait variance between sister species who have just diverged, especially for a slowly evolving trait. Its mathematical tractability makes Brownian motion an ideal framework in which to develop the model-fitting approach described here.

AUTEUR applies a Bayesian approach to modeling rate heterogeneity on a phylogenetic tree using reversible-jump Markov Chain Monte Carlo (Metropolis et al. 1953; Hastings 1970). This reversible-jump approach is implemented to assess fit of models of differing complexity, which in this context is the number of rate shifts in the tree (Green 1995; Huelsenbeck et al. 2004; Drummond and Suchard 2010). We construct a Markov chain such that, upon convergence, distinct Brownian-motion models are sampled according to their posterior probability (Bartolucci et al. 2006). Of main interest, and inherently tied to the set of sampled models, are the marginalized distributions of relative rates for each branch in the tree.

General Approach: Bayesian Sampling of a Multirate Brownian-Motion Process

AUTEUR takes as input a phylogenetic tree and character states for sampled species. Trait data are assumed to have evolved by Brownian motion, and evolutionary rates are presumed to be phylogenetically heritable. That is, the ancestral value for the evolutionary rate of the trait is inherited in each descendant unless the data provide evidence for a rate shift (see below). AUTEUR allows exploration from the simplest Brownian-motion process (a global evolutionary rate) to a highly complex model where each branch evolves at a rate independent from all others (i.e., the “free model” of Mooers et al. 1999).

In the following paragraphs, we detail the mechanics of our Bayesian approach. Consider an ultrametric phylogenetic tree with N species. Excluding the stem, this tree will have $2(N - 1)$ branch lengths, which we will refer to as $\vec{b} = [b_1, b_2, \dots, b_{2(N-1)}]$, listed in some arbitrary order. We will call the vector of tip phenotypes $\vec{y} = [y_1, y_2, \dots, y_N]$. Under a Brownian-motion model, these tip phenotypes come from a multivariate normal distribution (Felsenstein 1973). In the case where rates across every branch in the phylogeny are equal, this distribution will have mean α and variance-covariance matrix $\mathbf{S} = [s_{ij}] = \sigma^2[c_{ij}]$, where α is the state at the root of the tree and σ^2 is the rate parameter for the Brownian-motion process. Each c_{ij} is the shared phylogenetic path length for species i and j . We note that each s_{ij} (i.e., an element within the variance-covariance matrix) can be written as the rate-scaled shared path length for species i and j . Thus, $s_{ij} = \sum_{k=1}^{2(N-1)} \sigma^2 a_k b_k$, where $a_k = 1$ if branch k is shared by both i and j in the path from the root to each species and 0 otherwise; b_k is simply the length of the k th branch as indicated above.

We will consider a set of models where characters evolve under a Brownian-motion model, but the rate of evolution might vary on each branch of the tree. As this model is highly dimensional, with $2(N - 1) - N = N - 2$ more parameters than trait values, we use reversible-jump Markov chain Monte Carlo sampling for efficient searching of this space, and our prior is constructed so to favor models with few distinct rate classes. We will specify branch-specific relative rates as $\vec{\sigma}^2 = [\sigma_1^2, \sigma_2^2, \dots, \sigma_{2(N-1)}^2]$, ordered as the vector of branch lengths, \vec{b} . In this case, characters will still follow a multivariate normal distribution, but the elements of the variance-covariance matrix, \mathbf{S} , will be $s_{ij} = \sum_{k=1}^{2(N-1)} \sigma_k^2 a_k b_k$, using each branch-specific σ_k^2 to scale the path lengths. The set of evolutionary rates scalars, $\vec{\sigma}^2$, thus bears on the stochastic variance of trait values expected to accumulate within the tree.

Standard proposal mechanisms, described elsewhere (see, e.g., Green 1995), are used to update parameters in the model. We use multiplier and sliding-window proposals to update the vector

of relative rates; sliding-window proposals are used to update the root state. A critical proposal mechanism in our reversible-jump Markov chain is the splitting or merging of relative rates across the tree (see also Huelsenbeck et al. 2004). Steps in the chain that involve this “split-or-merge” operation randomly choose a single branch. If the relative rate of that branch is currently equivalent to the relative rate of its immediate ancestor, a new local rate is proposed for the descendant branches of the selected branch. This new rate cascades through the descendants of that branch until descendant tips are reached or other local rates, different from the new rate, are encountered. Merge operations perform the reverse: if in the current state of the Markov chain the immediate ancestral relative rate differs from that of the randomly chosen branch, these rates may be combined into one by the merge move. Both merge and split operations are conducted such that the mean of relative rates is preserved (Green 1995; Huelsenbeck et al. 2004).

The Markov chain is a succession of outcomes of model comparisons, where the adoption of a particular model in each state of the chain is driven by the posterior odds of that model in light of the observed data. A newly proposed model, θ' , is formulated for each state by updating some facet of the current state (e.g., number of shifts, or values for relative rates), which is compared to the current state of the chain, $\theta^{(t)}$. In broadest terms, the acceptance probability of the newly proposed state given the current state is,

$$\theta^{(t+1)} = \begin{cases} \theta' & \text{with probability: } \min \left\{ 1, \frac{P(\theta' | \vec{y})}{P(\theta^{(t)} | \vec{y})} \frac{Q(\theta^{(t)} | \theta')}{Q(\theta' | \theta^{(t)})} \right\}, \\ \theta^{(t)} & \text{otherwise.} \end{cases} \tag{1}$$

Proposed models are thus adopted in inverse proportion to the poorness of the model fit. This not only ensures that proposed models with higher posterior odds will always be accepted, but also enables the chain to adequately explore model space without being confined to local optima of posterior support. The first term of the acceptance probability involves the ratio of posterior densities of the proposed and current states; the second term describes the transition kernel between adjacent states of the chain. Beginning with the first term of the acceptance probability from above,

$$\frac{P(\theta' | \vec{y})}{P(\theta^{(t)} | \vec{y})} = \frac{P(\theta')}{P(\theta^{(t)})} \cdot \frac{P(\vec{y} | \theta')}{P(\vec{y} | \theta^{(t)})}. \tag{2}$$

This term comprises the ratio of prior odds, $\frac{P(\theta')}{P(\theta^{(t)})}$, multiplied by the likelihood ratio. An important factor in Bayesian analysis is the set of priors used in model comparisons. We describe details of formulating the ratio of prior probabilities in the section to follow.

The second term of the acceptance probability involves the Hastings ratio, which is the ratio of probabilities of traversing

either direction between proposed and current states. If state-traversal asymmetries exist in the proposal kernel, we thus need to correct for the potential of biased sampling of the chain (Hastings 1970). For instance, if in the current state of the chain a single rate governs evolution across the tree, only proposals that split the global rate into two independent rates are permissible. This constraint would serve to bias the transition probability toward higher dimension models. Analogously, if the current state involves a unique rate for each branch in the tree, no further rate shifts would be possible. We implement Green’s (1995) method for calculating the correction factor (i.e., Hastings ratio), which is

$$\frac{Q(\theta^{(t)} | \theta')}{Q(\theta' | \theta^{(t)})} = \frac{Q(K | K')}{Q(K' | K)} = \begin{cases} \frac{K + 1}{2n - 2 - K} & \text{if } K' = K + 1, \\ \frac{2n - 2 - K + 1}{K} & \text{if } K' = K - 1, \\ 1 & \text{if } K' = K. \end{cases} \tag{3}$$

As shown, the proposal kernel is rewritten in terms of the number of parameters in the models, K and K' , which describe dimensionality, or number of relative-rate shifts, in each model. Where compared models do not differ in number of parameters and where transitions between current and proposed states are symmetric, the Hastings ratio reduces to one and does not contribute to the posterior odds of the proposed model.

PROPOSALS AND PRIORS

Using the sliding-window proposal mechanism, proposal densities for σ_k^2 and on the state at the root, y_0 , are symmetric between current and proposed values. Proposed values that exceed bounds (e.g., a negative relative-rate parameter) are simply reflected back into the allowable range. The proposal kernels are: $\sigma_k^2 \sim \sigma_k^2 + U(-v, v)$ and $y'_0 \sim y_0 + U(-v, v)$, where v is a proposal width and U is a uniform distribution. Priors on these parameters are uninformative and span the following ranges: $\sigma_k^2 \sim U[0, \infty)$ and $y_0 \sim U(-\infty, \infty)$ (see Schluter et al. 1997). A multiplier proposal mechanism is also used to update σ_k^2 and may be asymmetric; this asymmetry is accommodated by the Hastings ratio. The multiplier proposal draws new rates from $\sigma_k^2 \sim \sigma_k^2 \cdot \exp^{\lambda(u-0.5)}$, where $u \sim U(0, 1)$ and where $\lambda = 2\log(\delta)$ and δ is either the proposal width from above (if $v \geq 1$) or its reciprocal (if $v < 1$). The Hastings ratio for this multiplier proposal is simply u .

We place high prior weight on few shifts in the evolutionary process. Using uninformative priors for the other parameters, we only need to consider the ratio of prior probabilities of having K and K' rate shifts in the phylogeny, computed using a Poisson distribution that is truncated beyond the maximum number of rate

shifts for a given tree ($K_{\max} = 2n - 2$; Drummond and Suchard 2010). For chain states that have at least one rate shift in the tree, we introduce a proposal mechanism that locally moves the shift one branch, with equiprobable chances of the shift moving tipward or rootward. Because in a binary tree, there are always two descendants at each node, this proposal is as well asymmetric: our proposal mechanism disfavors moving toward the more tipward branch. When used, we account for the asymmetric probabilities of moving to and from a particular branch by the Hastings ratio.

PROPOSAL-WIDTH CALIBRATION

To achieve reasonable rates of mixing when Markov chain sampling occurs, an initial sampling period is used to calibrate the proposal width (v from above). An initial geometric progression of widths (from $\frac{1}{8}, \frac{1}{4}, \dots, 8$) is used in these calibrations. Using the maximum-likelihood estimate for a global-rate model as a starting point, we compute a weighted average of proposal widths to initiate sampling where weights are based on acceptance rates for 1000 proposals under each proposal width. Thereafter, one-tenth of the total run length is used to adjust the initially chosen proposal width in the geometric progression. Updates to the proposal width occur at progressively longer intervals in the recalibration period (e.g., at generation $10^1, 10^2, \dots, 10^5$ for a total run length of 10^6 generations). For each recalibration, 1000 proposals are again considered for a pair of adjusted proposal widths, which are one-half and twice the current proposal width. As previously, proposal width is readjusted using acceptance rates as weights for sampling conducted with the pair of newly proposed widths. When Markov chain sampling begins, the proposal width is constrained to the calibrated value, and samples from the calibration period are obligatorily excluded from our estimate of the posterior distribution of models.

Methods

EVALUATING STATISTICAL PROPERTIES BY SIMULATION

We use two general classes of simulations to investigate the properties of this method: under a global rate and under a heterogeneous process of evolution in which a single shift in relative rate occurs. Within each class of simulation, we assess efficiency of the method by comparing true to estimated parameters under a range of tree sizes (16, 32, 64, and 128 taxa); second, we measure sensitivity by using a range of evolutionary rates across trees (8-, 16-, 32-, and 64-fold reductions of a unit rate). Trees were generated by the TREESIM package (version 1.3, Stadler 2010) under a pure-birth model of diversification and with a stochastic speciation rate of one new lineage per time unit. For multiple-rate simulations, we confined tree space to those trees that bore at least one clade with half the number of total tips in the tree

(i.e., clades with 8, 16, 32, or 64 tips). We positioned the single rate shift to occur at the base of a selected subclade of appropriate size: the ancestral rate (σ_A^2) was 1.0 and the rate descending from the shifted node drew from one of four factor reductions of this ancestral rate: $\sigma_S^2 \in (\frac{1}{8}, \frac{1}{16}, \frac{1}{32}, \frac{1}{64})$. The sequence of tree sizes and this latter set of reduced rates were used for simulations conducted under a constant and global rate. Code from GEIGER (version 1.0, Harmon et al. 2008) was used to simulate trait data for all simulations. For simulations investigating the effect of clade size, a shifted rate (σ_S^2) of $\frac{1}{16}$ was used. We used a tree size of 64 species for simulations assessing the influence of magnitude of rate differences. Chain sampling in our simulations occurred over 10^6 generations, of which the first half was discarded as a burn-in.

We use several measures to evaluate performance and statistical properties of the method, for both global-rate and multiple rate simulations. We summarize simulations by evaluating posterior probabilities of the true complexity of the generating model as well as the posterior probability of a model at least as complex as the simulated process. We consider error in estimated relative rates in a branchwise fashion, where we weight the proportion of the estimated rate to the true rate by the length of each branch. The average of these weighted branchwise proportions is used as a summary of error across the entire tree. For instance, if each relative rate, σ_k^2 , is twice the true rate at each corresponding branch, the tree-wide summary of the proportion of the true rate would also be two. For simulations involving a shift in evolutionary rate, we further assess rate-shift error by determining the inferred posterior probability of a shift occurring at the truly rate-shifted branch. Retaining every hundredth sample from the Markov chain, we use a thinned posterior sample in evaluating each measure of performance.

EMPIRICAL EXAMPLES

We investigate support for a heterogeneous process in evolution for a set of reasonably large datasets: body size for turtles (see Jaffe et al. 2011; 226 species) and female body mass for primates (see Redding et al. 2010; complete sampling of 233 species). We use log-transformed measures of body size for each dataset. Straight-line carapace lengths in *Chelonia* span well over an order of magnitude. If we linearize by the cube root, raw body masses for extant primates similarly range just over a single order of magnitude: from the mouse lemur (*Microcebus rufus*) to the gorilla (*Gorilla gorilla*). We ask whether this trait variance is ascribable to a rate-homogenous process or if observed variation is attributable to lineage-localized shifts in the evolutionary process. For each analysis, we combine results from three independent Markov chains, discarding the first quarter of 10^6 generations of sampling as burn-in.

Results

Over the range of examined simulations, our introduced Bayesian sampler reliably and efficiently estimates the generating process of trait evolution. Simulation results suggest that despite some imprecision in identifying branch-specific locations of rate shifts in our smallest trees, marginalized rate estimates at branches are very reasonable across the range of datasets used herein. Our empirical datasets appear to confirm this result: while it may be seldom that a particular node will receive substantial posterior support for a rate shift, phylogenetically local patterns in rate heterogeneity may be readily distinguishable if the marginal posterior densities of rate estimates are used for purposes of rates comparison.

The range of overall acceptance rates within simulations exhibited little variance between the two classes of simulations (global-rate simulations: [0.38, 0.49]; multiple-shift simulations: [0.32, 0.53]). Split-or-merge operations exhibited the weakest rates of acceptance (from 0.01 to 0.04), which is often characteristic of reversible-jump Markov chains (see Green and Hastie 2009). Yet for a dataset that exhibits strong signal for models of a particular complexity, we should expect very few split-or-merge operations to be accepted upon convergence of the chain. Including calibration periods for the proposal width, three independent chains for each empirical dataset were completed in ca. 11.5 h on a machine with dual quad-core 2.4 GHz Xeon processors with 8 GB RAM running OS 10.6.7.

GLOBAL RATE SIMULATIONS

Simulations under a global rate for trees differing in the number of species suggest that AUTEUR performs well in recovering the generating model complexity (on average, $PP \geq 0.6$), with a modest increase in posterior probability with increasing tree size (Fig. 1A). Posterior weight for the generating, global-rate model slightly exceeded prior weight of ca. 0.5, yet inferred model complexity was generally overparameterized (Fig. 1A,C). Regardless of tree size, rate estimates across the tree did not appear to deviate substantially from the generating rates of evolution (Fig. 1B). For all global-rate simulations, tree-wide summaries of rate estimates fell well within the band between one-half and twice the true rate (Fig. 1B,D), and variance of these estimates appeared to decrease with increasing tree size (Fig. 1B). Under these simulations, any deviation of rate estimates from the true rates would be indicated where the tree-wide proportion of the true rate lies on either side of one. Recall that this statistic accommodates branch length, such that longer branches whose relative rates are estimated with high error contribute greater weight in the calculation of our tree-wide measure of error (see Methods).

As we should expect, the magnitude of the constant rate of evolution appeared to exert little influence on the ability to recover either the true model complexity or true relative rates in

these global rate simulations (Fig. 1C,D). Even where posterior models were overly complex, relative rate estimates were very near the true values and appear unbiased (Fig. 1B,D).

MULTIPLE RATE SIMULATIONS

Posterior densities of relative rates at branches appear to be reasonably well estimated despite relatively low performance in detecting specific branches that experienced a rate shift in simulation. Yet, the largest trees in our simulations (≥ 64 tips) achieved very high posterior probabilities in recovering adequate model complexity (Fig. 2A,D), precision in topological identification of the simulated rate shift (Fig. 2B,E), and overall estimate accuracy of the underlying evolutionary rates (Fig. 2C,F). Marked improvement in regard to identification of the rate-shifted branch is observed between simulations conducted with 32 versus 64 species, where the latter simulations exhibited little variance around complete posterior support for a shift occurring at the truly shifted branch (Fig. 2B). A similar improvement in recovering position of the truly shifted branch is evident between simulations involving eightfold and 16-fold reductions of the ancestral rates (Fig. 2E). Even where the posterior sample of model complexity was underparameterized, rate estimates at branches were generally quite reasonable (e.g., compare panels from Fig. 2A with 2C). Substantial improvement in the accuracy of rate estimates is apparent with increasing tree size (Fig. 2C).

For the smallest of trees in our simulations, we were unlikely to recover proper model complexity (Fig. 2A), and elevated variance in rate estimates is likely a result of this underparameterization (Fig. 2C). Accuracy of rate estimates appears strongly related to ability in recovering adequate model complexity where at least some branches are allowed independent relative rates (Fig. 2A,C).

Disparities between the ancestral and shifted rates (hereafter, “effect sizes”) appear strongly related to method performance. For eightfold or larger reductions in the ancestral rate at the simulated rate shift, the primary mass of posterior density for model complexity was strongly centered on multiple-rate models (Fig. 2D). Precision in inferred placement of the rate shift improved with increasingly large effect size, and an eightfold rate difference may lie at the threshold of sensitivity for this method with 64 taxa (Fig. 2E). Noting that the potential for rate-estimate error increases along the abscissa in Fig. 2F, we find branchwise rate estimates to exhibit little error and bias: all effect-size simulation-replicates fall well within a band of error ranging between one-half and twice the true relative rates (Fig. 2F).

EMPIRICAL EXAMPLES

Posterior odds overwhelmingly support multiple-rate models of body-size evolution in both turtles and primates. As shown by Bartolucci et al. (2006), posterior odds of one model against

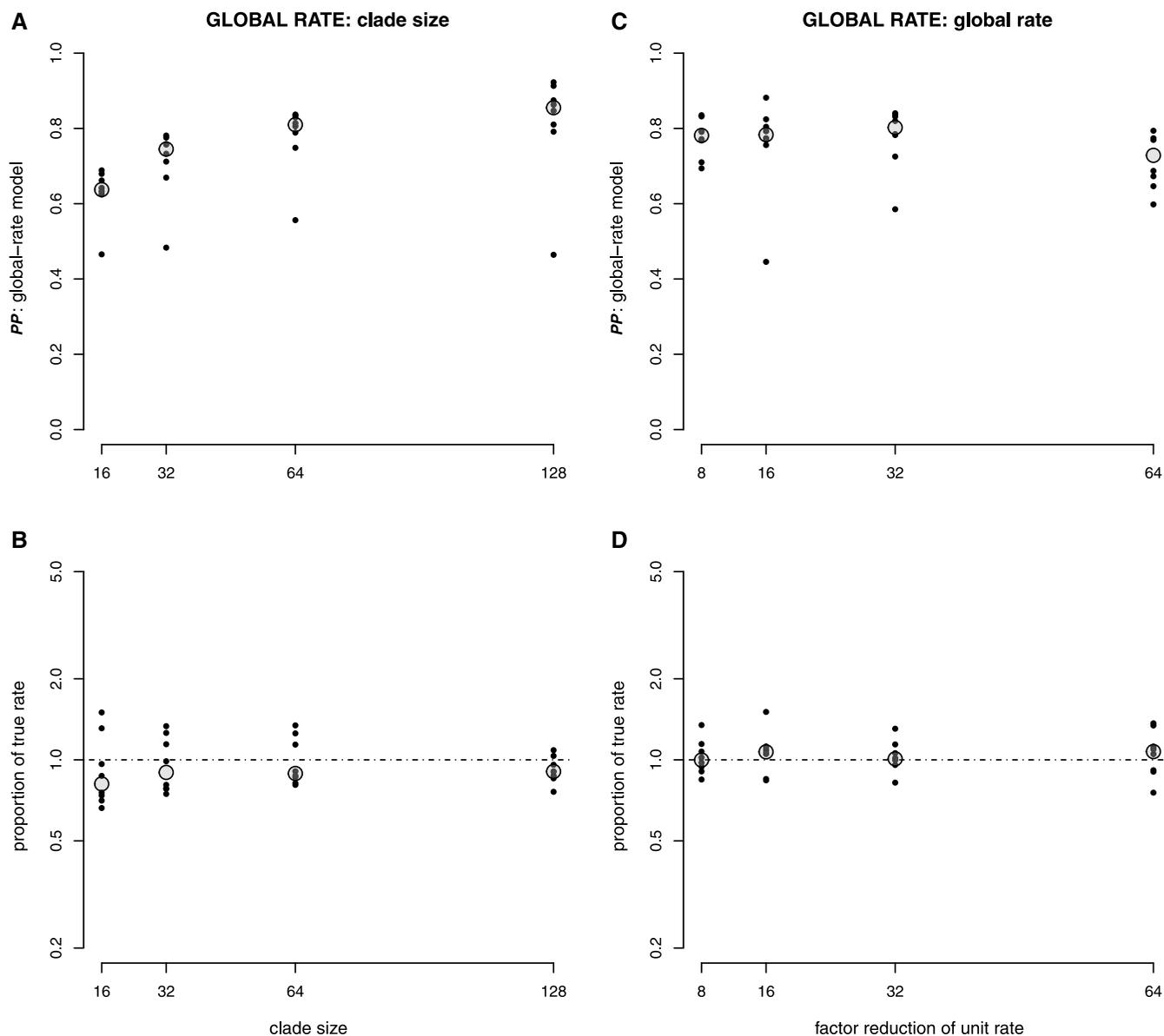


Figure 1. Panels summarize method performance for global-rate simulations in relationship to clade size (“clade size”; leftmost panels) and evolutionary rate (“global rate”; rightmost panels). All simulations were conducted with a global rate: for “clade size” simulations, evolutionary rate (σ^2) was $\frac{1}{16}$, and tree size is indicated along the abscissa; for “global rate” simulations, σ^2 was drawn from $(\frac{1}{8}, \frac{1}{16}, \frac{1}{32}, \frac{1}{64})$ and trees each had 64 tips. Throughout, the large shaded circles correspond to medians of eight simulation replicates. Ability of auteur to recover the generating model used in simulation is plotted as a function of tree size (A) and evolutionary rate (C). Values in (A) and (C) are posterior probabilities, *PP*, of a global-rate model. Tree-wide error in rate estimates are shown in relationship to tree size (B) and evolutionary rate (D). In (B) and (D), the axis of ordinates spans a range of error, from average branchwise proportions from $\frac{1}{5}$ the true rate to a fivefold average difference between estimates and true rates. Points represent averaged proportions of the true relative rate across all branches in the tree (see Methods). We consider error on a semi-log plot inasmuch as estimates of rates that are half or double the true value are equally (un)reasonable. The expectation for the method, if rates are estimated without error, is a proportion of the true rate of one (indicated by dot-dashed line).

another can be efficiently estimated from reversible-jump Markov chains by computing the ratio of frequencies in the chain sampling M_1 versus M_2 , where each M represents a model of particular complexity (e.g., a global-rate model of trait evolution). Not a single sample from the posterior density of model com-

plexities had fewer than two rate parameters, rendering Bayes factor comparisons impossible from our Markov samples for these taxa. To evaluate convergence of the Markov chains, we used several diagnostics from the CODA package (version 0.14-2, Plummer et al. 2010). Effective sample sizes of the relative rate

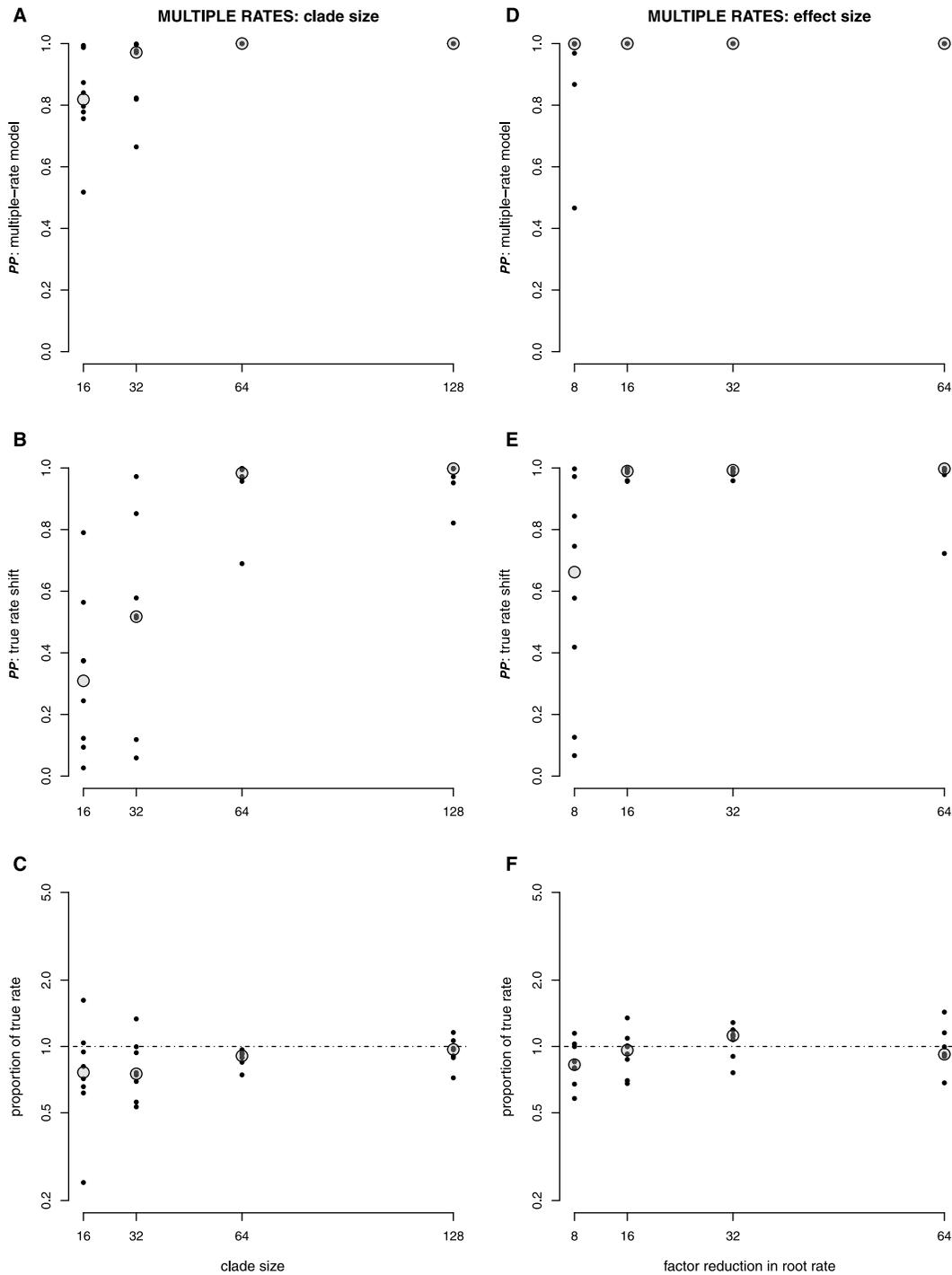


Figure 2. Panels summarize method performance for multiple-rate simulations in relationship to clade size (“clade size”; leftmost panels) and disparity between ancestral and shifted evolutionary rates (“effect size”; rightmost panels). All simulations were conducted in which a shifted rate (σ_5^2) differed from that at the root of the tree ($\sigma_A^2 = 1$) and where the shifted rate affected one-half of all branches in the phylogeny. For “clade size” simulations, evolutionary rate at the shifted branch (σ_5^2) and descendants thereof was $\frac{1}{16}$, and tree size is indicated along the abscissa; for “effect size” simulations, σ_5^2 was drawn from $(\frac{1}{8}, \frac{1}{16}, \frac{1}{32}, \frac{1}{64})$, and trees each had 64 tips. Shown is the ability of our method to recover adequate model complexity under a rate-shift scenario, as related to tree size (A) and factor difference between σ_A^2 and σ_5^2 (D). Median *PP* for two-rate models were 0.49, 0.58, 0.73, and 0.77 with increasing clade size; similarly, for increasing effect size, median posterior probabilities (*PP*) of two-rate models were 0.69, 0.77, 0.71, and 0.72. Power to precisely recover the branch at which the simulated shift occurred is shown as a function of tree size (B) and effect size (E). Plotted posterior probabilities in (B) and (E) are conditioned on there being at least one rate shift in the evolutionary model. Shown is the relationship between branchwise accuracy of relative-rate estimates in relation to tree size (C) and effect size (F). See Figure 1 for further detail.

parameters were generally well in excess of 10^3 , and samples from the thinned chains exhibited little autocorrelation. Model likelihoods appeared to reach stationarity within several thousand generations of sampling postcalibration.

Moderate support for several rate shifts is apparent within *Graptemys* (Emydidae) and *Geochelone* (Testudinidae; $0.42 < PP < 0.60$; Fig. 3). A pair of shifts receives similar support for two geoemydid turtles of likely intergeneric hybrid origin, *Mauremys iversoni* and *Ocadia philippeni* ($0.40 < PP < 0.56$; Fig. 3). We focus model-averaged rate comparisons within the three polytypic families of the Testudinoidea: Emydidae, Geoemydidae, and Testudinidae (Rhodin et al. 2010). Although rates of body-size evolution within the tortoises (Testudinidae) appear to be in excess of those of the Geoemydidae ($P = 0.013$), emydid body sizes appear to evolve at truly exceptional rates (Fig. 3). Rates for the Emydidae are distinguishable from both the phylogenetic background ($P = 0.014$; Fig. 3) as well as from their sister group, comprising the Testudinidae and the Geoemydidae ($P = 0.042$).

Although inferred evolutionary rates in the New World monkeys (Platyrrhini) were below the estimated median rate for primates, we find no evidence to conclude that rates of body-size evolution within this group are distinct from other primates ($P = 0.42$). In fact, we find insufficient evidence for exceptional evolution in any of the primary groups within the Primates using model-averaged rate comparisons ($0.21 < P < 0.78$). These results are reinforced by the lack of posterior support for branch-associated shifts in rate. Within primates, a pair of downturns in evolutionary rate are weakly supported for the Simiiformes and Platyrrhini ($0.20 < PP < 0.45$), and an upturn within macaques (*Macaca*) and in the lineage leading to the pygmy marmoset (*Callicebus pygmaea*) receive similar weak support ($0.32 < PP < 0.33$; Fig. 3).

Discussion

In general, the Bayesian sampler of evolutionary rates presented here provides accurate posterior samples of the modes and tempo of trait evolution across phylogenetic trees. Although the method appears susceptible to over-fitting of model complexity (Figs. 1 and 2), relative rates that are marginalized across model complexity are estimated with very little error (Fig. 2C). Inasmuch as true evolutionary processes may be quite discrepant from the saltational shifts in rates modeled here, we advocate an emphasis on model-averaged branchwise estimates of relative rates rather than positions of inferred shifts. Such a focus should be more resistant to ad hoc interpretations of the posterior results. This model-averaging framework still allows comparisons of evolutionary hypotheses at specific points or intervals on the tree, yet provides a more accurate representation of the evidence for rate

shifts along nonfocal branches. We focus our discussion on performance and potential extensions of this method.

CONSIDERATIONS

Although we model the evolutionary process as inherently saltational—rates change discretely between branching events—we expect that this method may as well be capable of capturing more gradualistic change in rate. Support for shifted rates that is diffusely spread over the backbone of a subtree (e.g., Fig. 3) may be suggestive of a more gradual process of change that must necessarily be discretized by the method introduced here. The lack of overwhelming posterior support for a rate shift concentrated at a particular branch may not imply that evolutionary rates within the lineage are not atypical. The exceptional rates of body-size evolution in the Emydidae appear to provide an example of just this sort, where the posterior mass of support for rate shifts is quite diffuse (Fig. 3). The converse also appears to hold: strong posterior support for a shift at a given branch may not provide sufficient evidence for distinguishable rates, but rather should be formally tested (e.g., see Fig. 3 caption).

Model-averaged rates from the posterior summary of the reversible-jump sampler provide a natural foundation for hypothesis testing. The evolutionary rate within a lineage in comparison to the phylogenetic background is one such testable hypothesis (e.g., Fig. 3), although similar tests could be quite varied and need not involve comparisons between monophyletic or paraphyletic groups. Even where there is weak support for a rate shift at any particular branch, model-averaged rates may provide sufficient signal in detecting localized and distinguishable patterns of incongruous evolution in distinct areas of tree. As shown here (Fig. 3), heterogeneity in the process of evolution may well be correlated with phylogeny (see also O'Meara et al. 2006) but also with traits extrinsic to those whose rates are being modeled (e.g., selective regimes, as in Butler and King 2004). Correlating posterior rate estimates with the states of extrinsic characters would require trustworthy estimates of states at each branch in the phylogeny (e.g., ecological niches, biogeographical regions, etc.), an issue that requires careful consideration (e.g., Goldberg and Igic 2007; Li et al. 2008; Ekman et al. 2008).

To facilitate comparisons among models differing in complexity, we allow for Bayes factor comparisons in the provided software (Jeffreys 1935; Kass and Raftery 1995). Within a single chain (or from pooled chains), the Bayes factor provides a posterior measure of model support while integrating over uncertainty in the precise placement of rate shifts in the tree, if the model so allows. We further allow sampling to be parametrically constrained (e.g., considering only k -rate models, rather than the full range of model complexities) for alternative statistical comparisons between models with differing complexities. For instance, sampling can be constrained to maximal model complexity to

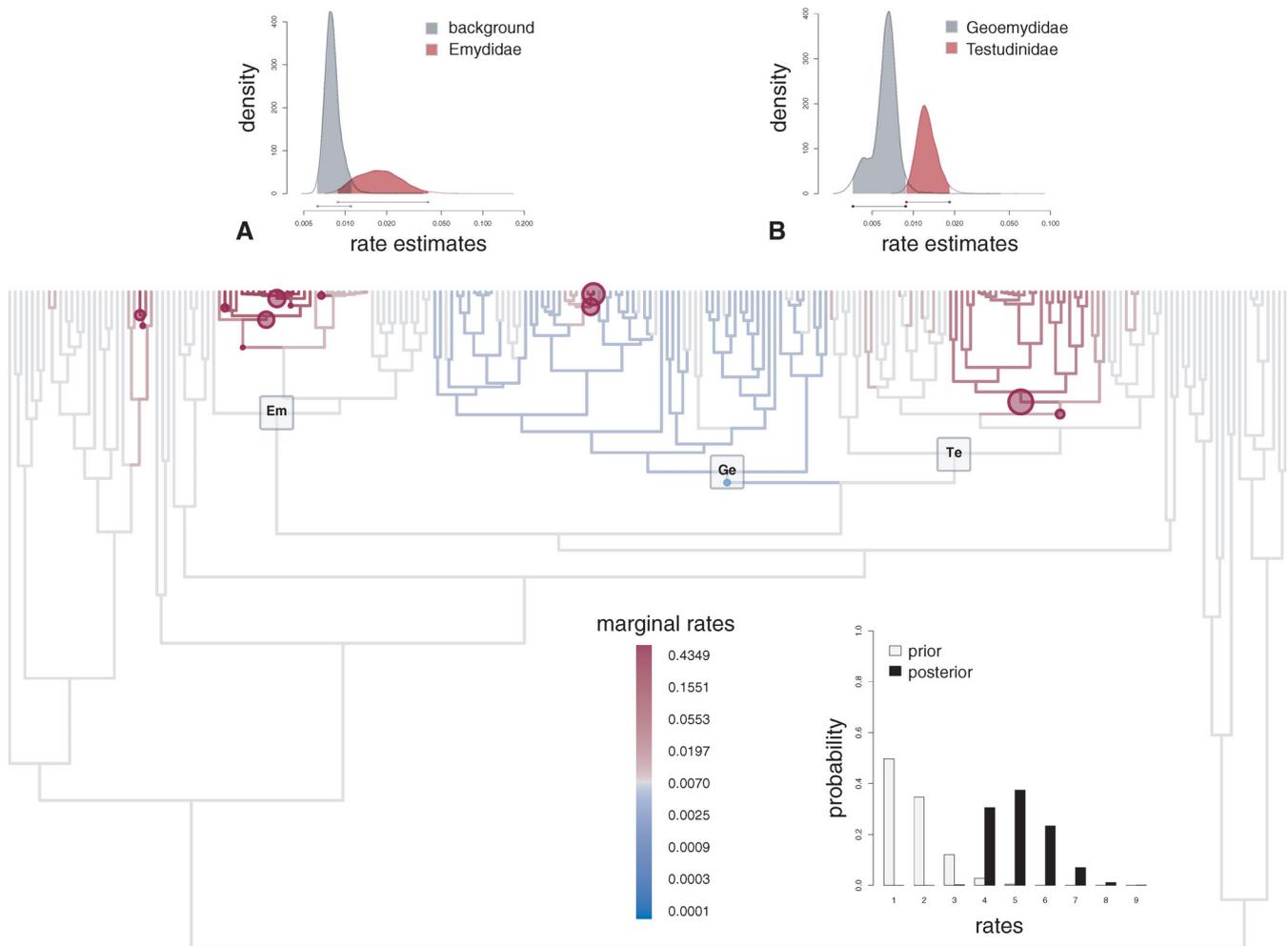


Figure 3. Posterior comparisons of rate of body-size evolution in turtles using model-averaging of rate estimates. Shown are posterior densities of rates comparing (A) those between the Emydidae (red; Em) and the phylogenetic background (gray); and (B) rates between the Geoemydidae (gray; Ge) and the Testudinidae (red; Te). Bars represent the highest posterior density (0.95) of rate estimates for each group. Rates are computed as a weighted average of posterior rate estimates, where weighting is determined by branch length (see Methods); note that rate densities are plotted on a log-scale. To test whether rates were distinguishable between lineages, we conducted 10^4 Monte Carlo sampling iterations, each of which was a comparison between randomly sampled draws from the posterior distributions of lineage-specific relative rates. Expecting the sign of these comparisons to be random if the two posterior distributions were truly identical, we interpret the proportion of comparisons in a particular direction to be an approximate probability value in a one-tailed test. We report two-tailed probability values under this resampling procedure, given that we proposed no prior expectation for the sign of these differences. Comparisons in panels (A) and (B) both strongly indicate rate differences ($P = 0.014$ and $P = 0.013$, respectively). The depicted turtle phylogeny (C) is from Jaffe et al. (2011). Hue and size of circles at branches denote posterior support for a rate shift at the indicated branch. Larger and redder circles suggest higher posterior support for an upturn in evolutionary rate (see text for details). Branches in the phylogeny are colored such that rates not deviant from the median are shaded gray; rates below (or above) the median are shaded blue (or red). Rates corresponding to each hue are indicated in the legend.

examine posterior support for the “free model” (sensu Mooers et al. 1999), where each branch evolves at a unique rate. Although AUTEUR may be capable of sampling from the ‘free model’ in unconstrained analyses, it is likely that few datasets will approach the number of rate categories as there are branches in the tree.

We note a potential difficulty in distinguishing a heterogeneous process of evolution where a shift occurs on a rootmost

branch. If we were to simulate a downturn in evolutionary rate at the base of the tree (on a particular branch), we might expect posterior support for rate shifts to be spread between this and its sister branch (i.e., with some posterior weight supporting an upturn on the sister branch and some posterior support for a downturn on the opposite branch). In simulations conducted on trees perfectly balanced at the rootmost split, we find this to be the case: posterior support for rate shifts tends to be spread

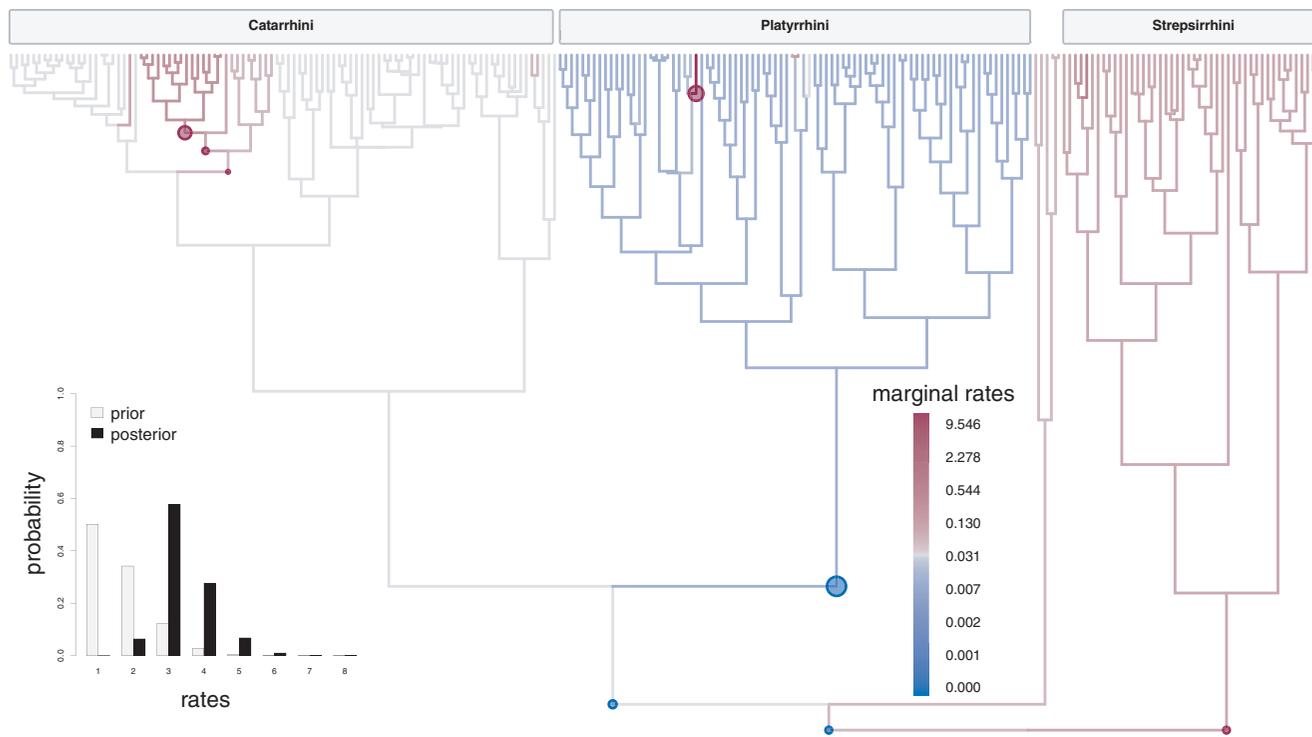


Figure 4. Marginalized rate estimates of body-mass evolution in primates and posterior support for a multiple-rate model by comparison of prior and posterior odds of varying model complexities. Three primary clades within the Primates are indicated at top (Catarrhini: Old World monkeys and apes; Platyrrhini: New World monkeys; and Strepsirrhini: lemurs and allies). Although multiple-rate models are clearly favored (inset), permutation tests did not indicate strong support for disparate rates of evolution between these major lineages. The primate tree used here is from FitzJohn (2010), which is based primarily on work by Vos and Mooers (2006). See Figure 3 for further detail on symbols and shading.

between each of the first descendant branches of the root (data not shown). In excluding the root branch from the constructed evolutionary models, the ability to resolve the nature of a shift at the rootmost branches will be limited. We note, however, that relative rates in these balanced-tree simulations do tend to accurately reflect the underlying known rates, which reemphasizes the importance of using marginal distributions of branch-specific rate estimates.

Some of the poorest estimates from this method occur where data were few (Figs. 1B and 2C). Under multiple-rate trait evolution, relative rates across the tree may be estimated with high variance as a result of the inability to recover proper model complexity; in simulation, this was most apparent for our smallest datasets (i.e., 16 and 32 taxa; Fig. 2C). If error is due to underparameterization of model complexity, we expect that as the difference between σ_B^2 and σ_A^2 becomes small (thereby converging toward a single rate), error in rate estimates will approach zero (as is the case where the difference between σ_B^2 and σ_A^2 is large and thus detectable; Fig. 2C,F). There may, however, be a range of disparities between σ_B^2 and σ_A^2 within which a true rate difference may be difficult to confirm and rate-estimate error may accumulate in the tree.

EXTENSIONS

To improve mixing in the Bayesian sampler, we note the potential benefit of Metropolis coupling of chains (e.g., Huelsenbeck and Ronquist 2001). We have observed that chains may occasionally become resistant to slight model perturbation in the topological position of a rate shift, especially where a set of branches provides strong signal for an exceptional rate. This can be problematic if the chain becomes trapped by sampling models with suboptimal placement of a rate shift. We suspect Metropolis coupling of hot and cold chains will enable more precise estimation of the topological locations of rate shifts. At the very least, we stress the importance of combining posterior estimates from multiple independent chains, which we facilitate with the provided software. Phylogenetic error, especially where error is nonrandom with respect to the true phylogeny, is certain to contribute to erroneous results in AUTEUR (e.g., see Martins and Hansen 1997 and Revell et al. 2005). In fact, results that weakly support shifted rates for two putative intergeneric hybrids (see Fig. 3 and Results) may be attributable to violation of the assumption of a nonreticulating phylogeny. A difficult scenario, distinct from hybridization, could be where the estimated phylogeny does not contain the bipartition whose stem has truly experienced a shift toward an exceptional

rate of trait diversification. A reasonable approach in dealing with phylogenetic error might be to integrate inferences from AUTEUR over a sample of credible trees. Sensitivity of results to phylogenetic error could otherwise be investigated by simulation. We further note that measurement error can be readily incorporated into these analyses where the extent of error associated with each trait value is known (Martins and Hansen 1997; Ives et al. 2007; Harmon et al. 2010).

An extension of the Revell and Collar (2009) likelihood method of testing evolutionary shifts in correlated evolution of a pair of characters could be readily implemented in the reversible-jump Bayesian framework. These authors ask whether the evolution of two functional–morphological characters in centrarchid fish is correlated, and if so, whether the relationship is uniform across evolutionary history. Rather than (or in addition to) fitting models where the rates of evolution of quantitative characters changes at particular nodes in the tree, the evolutionary matrix among characters would be the parameter of interest in a reversible-jump approach. In the most highly complex model, each branch in the phylogeny would have an independent evolutionary matrix among characters. Reversible-jump Markov sampling would provide a natural framework in which to simultaneously fit a broad range of character–correlation models and without the limitation of needing to select the most appropriate model(s) for comparison a priori.

The reversible-jump approach of AUTEUR is sufficiently flexible as to allow for broader extensions into other commonly implemented models of trait evolution and for which model likelihoods are calculable (e.g., Ornstein–Uhlenbeck process, Hansen 1997; Early-burst evolution, Blomberg et al. 2003). We expect development toward an implementation of AUTEUR under a generalized Ornstein–Uhlenbeck model of constrained trait evolution. In such a model, trait values are largely bounded to optima such that values strongly deviant from the optimum are expected to be drawn to fitted medial values (Hansen 1997; Butler and King 2004). Rather than comparing fit of potentially numerous a priori models with distinct histories of transitions between selective regimes, one could foresee the utility in a reversible-jump sampler of which one purpose is to “identify” these shifts between selective regimes.

Conclusions

The Bayesian sampler of phenotypic rates, presented here and provided as an R software package (AUTEUR), greatly broadens the field of possibilities in modeling continuous-trait evolution and identifying exceptional rates of change in lineages. AUTEUR can be used as a framework in which to test existing hypotheses for rate heterogeneity or as a method for generating hypotheses of

rate variation. We emphasize the use of model-averaged rates in posterior analyses, as estimate error across a wide range of simulation conditions appears reasonably small. We find that the ability to recover model complexity is highly related to the disparity in rates across the tree and dataset size. Thus, where the potential for large absolute error in rate estimates is most menacing, this method appears consistently reliable.

ACKNOWLEDGMENTS

We are indebted to M. Pennell, J. Rosindell, D. Jochimsen, G. Slater, two anonymous reviewers, and P. Lindenfors for editorial and general comments on a previous draft of the present manuscript. We thank A. Jaffe and G. Slater for their generosity in enabling our use of the chelonian dataset. We thank J. Brown for his comments on an earlier version of this paper and for his advice leading us toward the analytic implementation adopted for AUTEUR. Financial support for MEA and LJH was provided by NSF DEB 0918748 and 0919499, respectively.

REFERENCES

- Bartolucci, F., L. Scaccia, and A. Mira. 2006. Efficient Bayes factor estimates from the reversible jump output. *Biometrika* 93: 41–52.
- Blomberg, S. P., T. Garland, and A. R. Ives. 2003. Testing for phylogenetic signal in comparative data: behavioral traits are more labile. *Evolution* 57: 717–745.
- Bookstein, F.L.. 1987. Random walk and the existence of evolutionary rates. *Paleobiology* 13: 446–464.
- Bonett, R. M., P. T. Chippindale, P. E. Moler, R. W. Van Devender, and D. B. Wake. 2009. Evolution of gigantism in amphiumid salamanders. *PLoS one* 4(5):e5615.
- Butler, M. A., and A. A. King. 2004. Phylogenetic comparative analysis: a modeling approach for adaptive evolution. *Am. Nat.* 164: 683–695.
- Drummond, A. J., and M. A. Suchard. 2010. Bayesian random local clocks, or one rate to rule them all. *BMC Biol.* 8:114.
- Ekman, S., H. L. Andersen, and M. Wedin. 2008. The limitations of ancestral state reconstruction and the evolution of the ascus in the Lecanorales (lichenized Ascomycota). *Syst. Biol.* 57: 141–156.
- Eldredge, N., and S. M. Stanley. (eds.). 1984. *Living fossils*. Springer Verlag: New York.
- Estes S., and S. J. Arnold. 2007. Resolving the paradox of stasis: models with stabilizing selection explain evolutionary divergence on all timescales. *Am. Nat.* 169: 227–244.
- Felsenstein, J. 1973. Maximum likelihood estimation of evolutionary trees from continuous characters. *Am. J. Hum. Genet.* 25: 471–492.
- FitzJohn, R. G. 2010. Quantitative traits and diversification. *Syst. Biol.* 59: 619–633.
- Freckleton, R. P., and W. Jetz. 2009. Space versus phylogeny: disentangling phylogenetic and spatial signals in comparative data. *Proc. R. Soc. Lond. B* 276: 21–30.
- Gingerich, P. D. 1983. Rates of evolution: effects of time and temporal scaling. *Science* 222: 159–161.
- Goldberg, E. E., and B. Iqic. 2008. On phylogenetic tests of irreversible evolution. *Evolution* 62: 2727–2741.
- Green, P. J. 1995. Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika* 82: 711–732.

- Green, P. J., and D. Hastie. 2009. Reversible jump MCMC. Available at: <http://www.maths.bris.ac.uk/mapjg/Papers.html>.
- Hansen, T. F. 1997. Stabilizing selection and the comparative analysis of adaptation. *Evolution* 51: 1341–1351.
- Harmon, L. J., J. A. Schulte II, A. Larson, and J. B. Losos. 2003. Tempo and mode of evolutionary radiation in iguanian lizards. *Science* 301: 961–964.
- Harmon, L. J., J. B., Losos, T. J., Davies, R. G., Gillespie, J. L., Gittleman, W. B., Jennings, K. H., Kozak, M. A., McPeck, F. Moreno-Roark, T. J. Near, et al. 2010. Early bursts of body size and shape evolution are rare in comparative data. *Evolution* 64: 2385–2396.
- Harmon, L. J., J. Weir, C. 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 applications. *Biometrika* 57: 97–109.
- Huelsenbeck, J. P., and F. Ronquist. 2001. MrBayes: Bayesian inference of phylogeny. *Bioinformatics* 17: 754–755.
- Huelsenbeck, J. P., B. Larget, and M. E. Alfaro. 2004. Bayesian phylogenetic model selection using reversible jump Markov chain Monte Carlo. *Mol. Biol. Evol.* 21: 1123–1133.
- 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.
- Jaffe, A. L., G. J. Slater, and M. E. Alfaro. 2011. The evolution of island gigantism and body size variation in tortoises and turtles. *Biol. Lett.* 7: 558–561.
- Jeffreys, H. 1935. Some tests of significance, treated by the theory of probability. *Proc. Cambridge Philos. Soc.* 31: 203–222.
- Kass, R. E., and A. E. Raftery. 1995. Bayes factors. *J. Am. Stat. Assoc.* 90: 773–795.
- Li, G., M. Steel, and L. Zhang. 2008. More taxa are not necessarily better for the reconstruction of ancestral character states. *Syst. Biol.* 57: 647–653.
- Martins, E. P., and T. F. Hansen. 1997. Phylogenies and the comparative method: a general approach to incorporating phylogenetic information into the analysis of interspecific data. *Am. Nat.* 149: 646–667.
- Metropolis, N., A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller. 1953. Equation of state calculations by fast computing machines. *J. Chemical Phys.* 21: 1087–1092.
- Mooers, A. Ø., S. M. Vamasi, and D. Schluter. 1999. Using phylogenies to test macroevolutionary hypotheses of trait evolution in cranes (Gruinae). *Am. Nat.* 154:249–259.
- O'Meara, B. C., C. Ane, M. J. Sanderson, and P. C. Wainwright. 2006. Testing for different rates of continuous trait evolution using likelihood. *Evolution* 60: 922–933.
- Pagel, M., and A. Meade. 2008. Modelling heterotachy in phylogenetic inference by reversible jump Markov chain Monte Carlo. *Philos. Trans. R. Soc. Lond. B* 363: 3955–3964.
- Plummer, M., N. Best, K. Cowles, and K. Vines. 2010. coda: output analysis and diagnostics for MCMC. R package version 0.14-2. <http://CRAN.R-project.org/package=coda>
- Redding, D. W., C. DeWolff, and A. Ø. Mooers. 2010. Evolutionary distinctiveness, threat status and ecological oddity in primates. *Conserv. Biol.* 24: 1052–1058.
- 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 R. E. Glor. 2005. Under parameterized model of sequence evolution leads to bias in the estimation of diversification rates from molecular phylogenies. *Syst. Biol.* 54: 973–983.
- Revell, L. J., and L. J. Harmon. 2008. Testing quantitative genetic hypotheses about the evolutionary rate matrix for continuous characters. *Evol. Ecol. Res.* 10: 311–321.
- 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., L. J. Harmon, R. B. Langerhans, and J. J. Kolbe. 2007. A phylogenetic approach to determining the importance of constraint on phenotypic evolution in the neotropical lizard, *Anolis cristatellus*. *Evol. Ecol. Res.* 9: 261–282.
- Revell, L. J., D. Mahler, P. Peres-Neto, and R. Redelings. A new method for identifying exceptional evolutionary diversification. *Evolution: in revision*.
- Rhodin, A. G. J., P. P. van Dijk, J. B. Iverson, and H. B. Shaffer. 2010. Turtles of the World, 2010 Update: annotated checklist of taxonomy, synonymy, distribution, and conservation status. *Chelonian Research Monographs* 5: 85–164.
- Schluter, D., T., Price, A. Ø. Mooers, and D. Ludwig. 1997. Likelihood of ancestor states in adaptive radiation. *Evolution* 51: 1699–1711.
- Simpson, G. G.. 1944. *Tempo and Mode in Evolution*. Columbia University Press, New York.
- . 1953. *The major features of evolution*. Columbia University Press, New York.
- Stadler, T. 2010. TreeSim: simulating trees under the birth-death model. R package version 1.0. <http://CRAN.R-project.org/package=TreeSim>.
- Vos, R. A., and A. Ø. Mooers. 2006. A new dated supertree of the Primates. Chapter 5. In: Vos R.A., editor. *Inferring large phylogenies: the big tree problem*. [Ph.D. thesis]. Burnaby BC, Canada: Simon Fraser University.

Associate Editor: P. Lindenfors