A variable-rate quantitative trait evolution model using penalized-likelihood

Liam J. Revell[1,2]

[1]**Department of Biology, University of Massachusetts Boston, Boston, Massachusetts 02125**
[2]**Facultad de Ciencias, Universidad Católica de la Santísima Concepción, Concepción, Chile**

Corresponding author:
Liam J. Revell[1,2]

Email address: liam.revell@umb.edu

ABSTRACT

In recent years it's become increasingly popular to use phylogenetic comparative methods to investigate heterogeneity in the rate or process of quantitative trait evolution across the branches or clades of a phylogenetic tree. Here, I present a new method for modeling variability in the rate of evolution of a continuously-valued character trait on a reconstructed phylogeny. The underlying model of evolution is stochastic diffusion (Brownian motion), but in which the instantaneous diffusion rate (σ^2) *also* evolves by Brownian motion on a log-scale. Unfortunately, it's not possible to simultaneously estimate the rates of evolution along each edge of the tree *and* the rate of evolution of σ^2 itself using Maximum Likelihood. As such, I propose a penalized-likelihood method in which the penalty term is equal to the log-transformed probability density of the rates under a Brownian model, multiplied by a 'smoothing' coefficient, λ, selected by the user. λ determines the magnitude of penalty that's applied to rate variation between edges. Lower values of λ penalize rate variation relatively little; whereas larger λ values result in minimal rate variation among edges of the tree in our fitted model, eventually converging on a single value of σ^2 for all of the branches of the tree. In addition to presenting this model here, I've also implemented it as part of my *phytools* R package in the function `multirateBM`. Using different values of the penalty coefficient, λ, I fit the model to simulated data with: Brownian rate variation among edges (the model assumption); uncorrelated rate variation; rate changes that occur in discrete places on the tree; and no rate variation at all among the branches of the phylogeny. I then compare the estimated values of σ^2 to their known true values. In addition, I use the method to analyze a simple empirical dataset of body mass evolution in mammals. Finally, I discuss some applications and limitations of the approach.

INTRODUCTION

As the quantity and quality of data for phylogenetic inference multiplies rapidly in the current genomic age (Philippe et al., 2005), so too has grown the popularity of using estimated phylogenetic trees to make inferences about the history of life on Earth (Harvey, 1996; Nunn, 2011; O'Meara, 2012). This endeavor, usually referred to as phylogenetic comparative biology, involves taking an evolutionary tree or set of trees obtained from phylogenetic inference (often combined with phenotypic trait data for the species in the tree), and then making some kind of quantitative inference about evolution (Garamszegi, 2014; Harmon, 2019). A wide variety of different methodologies now comprise the field of phylogenetic comparative biology. These are often collectively referred to as 'phylogenetic comparative methods' (Felsenstein, 1985; Nunn, 2011; O'Meara, 2012; Garamszegi, 2014; Harmon, 2019).

A significant advance that's occurred over about the past 15 years has been the development and implementation in software of statistical methodologies that allow the user to model, or explicitly take into consideration, heterogeneity in the tempo and mode of evolution across the branches or clades of our evolutionary tree (e.g., Butler and King, 2004; summarized in Harmon, 2019). For instance, in a now classic paper, O'Meara et al. (2006) published an innovative model in which the rate of evolutionary change for a trait (normally denominated σ^2 in this type of study) was permitted to differ between different parts of the phylogeny, such as specific branches or clades, as determined *a priori* by the investigator.

In the years since the publication of that seminal paper, a wide variety of different related approaches have been developed and implemented in software, such as methods that permit the process or mode of evolution (and not just its tempo) to differ in different parts of a reconstructed phylogenies (e.g., Revell and Harmon, 2008; Revell and Collar, 2009; Revell et al., 2011; Beaulieu et al., 2012; Uyeda and Harmon, 2014; Caetano and Harmon, 2017).

The predominant model that's been used to study quantitative trait evolution on phylogenetic trees is called *Brownian motion* (Felsenstein, 1985; O'Meara et al., 2006; Revell et al., 2008; Harmon, 2019). Brownian motion is a model of stochastic diffusion in which the trait changes randomly and continuously through time. The expected value of the trait is constant and equal to the starting value, x_0; and the *variance* (either between two or more lineages diverging by the process, or between the current state of a lineage and its ancestor) increases linearly with a rate of σ^2 (O'Meara et al., 2006; Harmon, 2019). Because σ^2 is the rate of increase in variance with time under our evolutionary process, our *estimate* of σ^2 is typically referred to as the *rate of evolution* of the trait (O'Meara et al., 2006). Brownian motion is described in much greater detail in other articles and books, such as Felsenstein (2004), O'Meara et al. (2006), Revell et al. (2008), and Harmon (2019).

Herein, I propose a new method for modeling variation in the *rate* of character evolution across the branches and clades of a phylogeny. Under this model, we assume that Brownian evolution has a different value (σ_i^2) on each of the branches of the tree, and that the log-values of these rates, in turn, have evolved via a separate Brownian process (theoretically with a separate rate, σ_{BM}^2 – although I'll describe later why it may not be possible to estimate this rate). I'll show how to fit this model to data using a penalized-likelihood approach (Sanderson, 2002) in which the penalty term is proportional to the logarithm of the probability of the log rates ($\log(\sigma_i^2)$) across edges under our model. I'll also apply the model to a classic empirical dataset for mammalian body size evolution from Garland et al. (1992), and discuss both applications and limitations of the approach. I envision the primary use of this method to be in *exploratory* data analysis – although it may also be possible to employ it in examining specific alternative hypotheses for evolutionary change through time.

METHODS AND RESULTS

In this section I'll begin by introducing our model of evolution, then I'll apply it to several different simulated datasets, and, finally, I'll use the method to analyze a simple empirical dataset and tree.

The model

The model of this article is as follows. We assume that the phenotypic trait of interest, x, evolves by a process of Brownian motion (Felsenstein, 1985; O'Meara et al., 2006; Harmon, 2019). Under standard Brownian motion evolution, changes are random and normally distributed, with a mean value of 0 and a variance of $\sigma^2 t$, in which σ^2 is the instantaneous variance of the Brownian process and t is the elapsed time over which change is presumed to have occurred (Felsenstein, 2004; Harmon, 2019).

In this scenario, we expect that the values for our trait at the tips of the phylogeny will have a multivariate normal distribution with an expectation equal to the value of the trait at the start of the process, x_0, and a variance-covariance matrix given by $\sigma^2\mathbf{C}$ (e.g., O'Meara et al., 2006; Revell and Collar, 2009). Here, $\mathbf{C}$ is an $n \times n$ matrix for n species containing the height above the root of the common ancestor of each pair of terminal taxa, i and j, in the matrix position $C_{i,j}$ (e.g., Rohlf, 2001; Revell et al., 2008). σ^2 is the rate of evolution, as previously defined.

With just data and our phylogeny, we can find Maximum Likelihood estimates of the two parameters of the Brownian model, σ^2 and x_0, simply by maximizing the following expression giving the likelihood (O'Meara et al., 2006). This equation may be familiar to many readers because it's based on the multivariate normal probability density (O'Meara et al., 2006; Revell and Harmon, 2008).

$$l(\sigma^2, x_0|\mathbf{x}, \mathbf{C}) = \frac{\exp[-\frac{1}{2}(\mathbf{x} - \mathbf{1}x_0)'(\sigma^2\mathbf{C})(\mathbf{x} - \mathbf{1}x_0)]}{\sqrt{2\pi^n|\sigma^2\mathbf{C}|)}} \tag{1}$$

In our equation (1), $\mathbf{x}$ is an $n \times 1$ vector containing species values of x for each of the n taxa in our phylogeny. $\mathbf{1}$ is an $n \times 1$ vector of 1.0s. Finally, $|\sigma^2\mathbf{C}|$ is the determinant of $\sigma^2\mathbf{C}$.

On a log-scale, this is equivalent to the following – in which L (now and henceforward) is used to indicate the *log*-likelihood, as opposed to the likelihood on its original scale.

$$L = -(\mathbf{x} - \mathbf{1}x_0)'(\sigma^2\mathbf{C})(\mathbf{x} - \mathbf{1}x_0)/2 - \log(|\sigma^2\mathbf{C}|)/2 - \log(2\pi^n)/2 \tag{2}$$

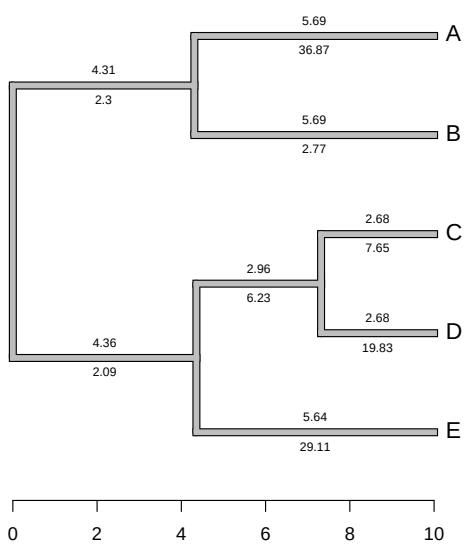
According to my penalized-likelihood approach, instead of computing the log probability density of just the species data at the tips of the tree under our model, we allow each edge of the tree to have a different value of the Brownian rate parameter, σ^2. Then we compute the penalized log-likelihood of the set of rates (σ_0^2, σ_1^2, σ_2^2, and so on), and the root state of our trait (x_0), given our tip data and reconstructed phylogeny.

We do this by maximizing the following expression – in which the penalty coefficient, λ, is set by the user, rather than estimated from the data. This is exactly equal to the log-likelihood of our model of evolution given the data; *minus* the negative log probability density of the log rates ($\log(\sigma_0^2)$, $\log(\sigma_1^2)$. $\log(\sigma_2^2)$, etc.) under a model in which the logarithm of the rate of evolution evolves by Brownian motion.

$$\begin{aligned} L(\sigma_0^2, \sigma_1^2, ..., x_0|\mathbf{x}, \mathbf{C}_{ext}, \lambda) = -(\mathbf{x} - \mathbf{1}x_0)'\mathbf{T}(\mathbf{x} - \mathbf{1}x_0)/2 - \log(|\mathbf{T}|)/2 - \log(2\pi^n)/2 - \\ \lambda(\mathbf{s} - \mathbf{1}s_0)'\mathbf{C}_{ext}(\mathbf{s} - \mathbf{1}s_0)/2 - \lambda\log(|\mathbf{C}_{ext}|)/2 - \lambda\log(2\pi^{n+m-1})/2 \end{aligned} \tag{3}$$

Reviewing equation (3), we can see that the first part of the expression is almost exactly the same as equation (2), *except* that we've substituted $\mathbf{T}$ for $\sigma^2\mathbf{C}$. The second part is *also* similar (because it's based on a multivariate normal probability density), and gives λ times the negative log probability density of the set of values for σ_i^2 at all of the nodes and tips of our phylogenetic tree. This latter part of the equation is our 'roughness penalty' (e.g., Sanderson, 2002) as it *increases* for a given value of λ (that is, penalizing the likelihood more) when σ^2 varies more from edge to edge, or decreases when it varies less.

a)



b)

	A	B	C	D	E
A	219.76	9.93	0	0	0
B	9.93	25.7	0	0	0
C	0	0	48.07	27.53	9.1
D	0	0	27.53	80.74	9.1
E	0	0	9.1	9.1	173.33

Figure 1. a) Simple phylogeny of five taxa with the total edge length (above each branch) and σ^2 rates (below it). b) Calculation of the expected variance covariance matrix of tip values for the trait x, $\mathbf{T}$, given the branch length and rates of panel a). Each i,jth element corresponds to the sum of the products above and below each edge along all the branches leading from the root to the common ancestor of taxa i and j.

More specifically, in equation (3), $\mathbf{T}$ is a matrix giving the sum of the edge lengths leading from the root to each common ancestor of each pair of species, i and j (in position $T_{i,j}$), multiplied by the branch-specific rates for each corresponding edge on that path (Figure 1). $\mathbf{C}_{ext}$ is an *extended* version of the matrix $\mathbf{C}$, above, but of dimension $n+m-1$ for n species and m internal nodes. $\mathbf{s}$ is an $n+m-1 \times 1$ vector of the logarithms of each rate at each node or terminal taxon, except the root node (that is to say,

$\mathbf{s} = [\log(\sigma_1^2), \log(\sigma_2^2), ..., \log(\sigma_{n+m-1}^2)]$). Lastly, s_0 is the logarithm of the rate of evolution at the root of the tree, $\log(\sigma_0^2)$. Other terms are defined as in equations (1) and (2).

Although we compute $\mathbf{T}$ by multiplying the *average* rate of evolution for each edge in the tree by the length of the edge, as noted above, and then by summing these quantities up the phylogeny, our penalized-likelihood equation (3) is given in terms of *node* and *tip* (as opposed to edge) values of $\log(\sigma_i^2)$. As such, to calculate $\mathbf{T}$ we need to first compute the average rate for each edge as a function of the rates at the nodes that subtend it. To do this, we integrate the log-linear function between each pair of ancestor-descendant nodes (or between an internal node and a tip, for a terminal edge of the phylogeny). This is done by calculating $a \cdot \exp(b)/b - a/b$, in which $a = \sigma_2^2$ and $b = \log(\sigma_1^2) - \log(\sigma_2^2)$ for each pair of two rates (σ_1^2 and σ_2^2, in this example) that begin and end a specific edge of our tree.

We define $\mathbf{C}_{ext}$ as a matrix of dimension $n+m-1 \times n+m-1$ containing the height above the root of the most recent common ancestor of every pair of terminal taxa (of which the tree contains n) and internal nodes (of which the tree contains m), except for the global root of the tree (hence the -1). This is proportional to the expected variance-covariance matrix of the tips and internal nodes under a Brownian evolutionary process (O'Meara et al., 2006; Harmon, 2019).

Simulated example 1: Brownian rates

For the first example of applying the method of this article, I'll imagine that evolution of our phenotypic trait, x, occurs by Brownian motion, and that the logarithm of the rate of evolution *also* evolves in a Brownian fashion. This type of evolution can be seen in Figure 2.

For panel a) of the figure, I simulated different rates of Brownian evolution for each edge of the tree under our model. In panel b), I've obtained a trait vector ($\mathbf{x}$) based on the rates simulated for panel a). Then, I projected my phylogeny into this trait space using a visualization technique called a 'traitgram' plot (Schluter et al., 1997; Evans et al., 2009; Revell, 2013), but in which each edge has been colored by the generating rate.

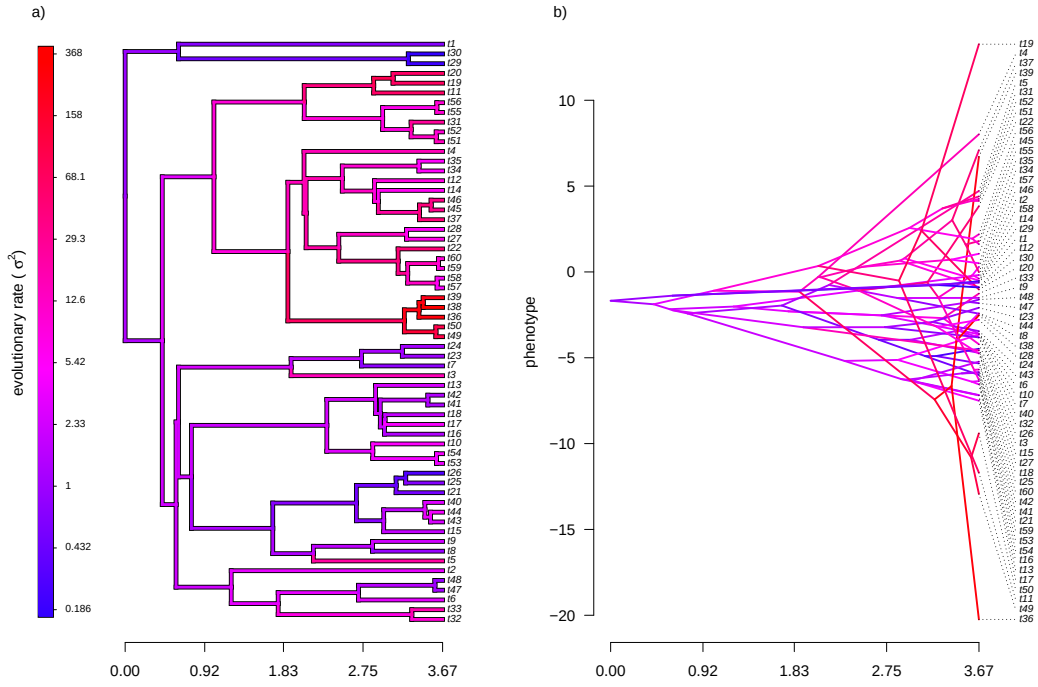


Figure 2. a) Simulated evolutionary rates, σ^2, in which the logarithm of the rate of evolution evolves by Brownian motion on the tree. b) The phylogeny of panel a) projected into a space defined by time (on the horizontal axis) and a simulated trait vector, $\mathbf{x}$, obtained using the rates of panel a).

From this figure it can be pretty clearly seen that *redder* branches, with higher evolutionary rates in panel a), tend to be associated with *larger* changes on the traitgram of panel b) – precisely as one would

expect under our model (Figure 2).

We can then proceed to fit a rate-variable model to the data of Figure 2 using our approach of penalized-likelihood. Our penalized-likelihood method, remember, requires that we select a value of λ, the penalty coefficient or 'smoothing' parameter. Higher values of λ correspond to more smoothing, such that adjacent branches of the tree will tend to have more similar values of σ^2. If we do this for four different values of λ ($\lambda = 0.01$, 0.1, 1, and 10), and then each time compare our estimated rates to their known true values from Figure 2, we obtain the result shown in Figure 3.

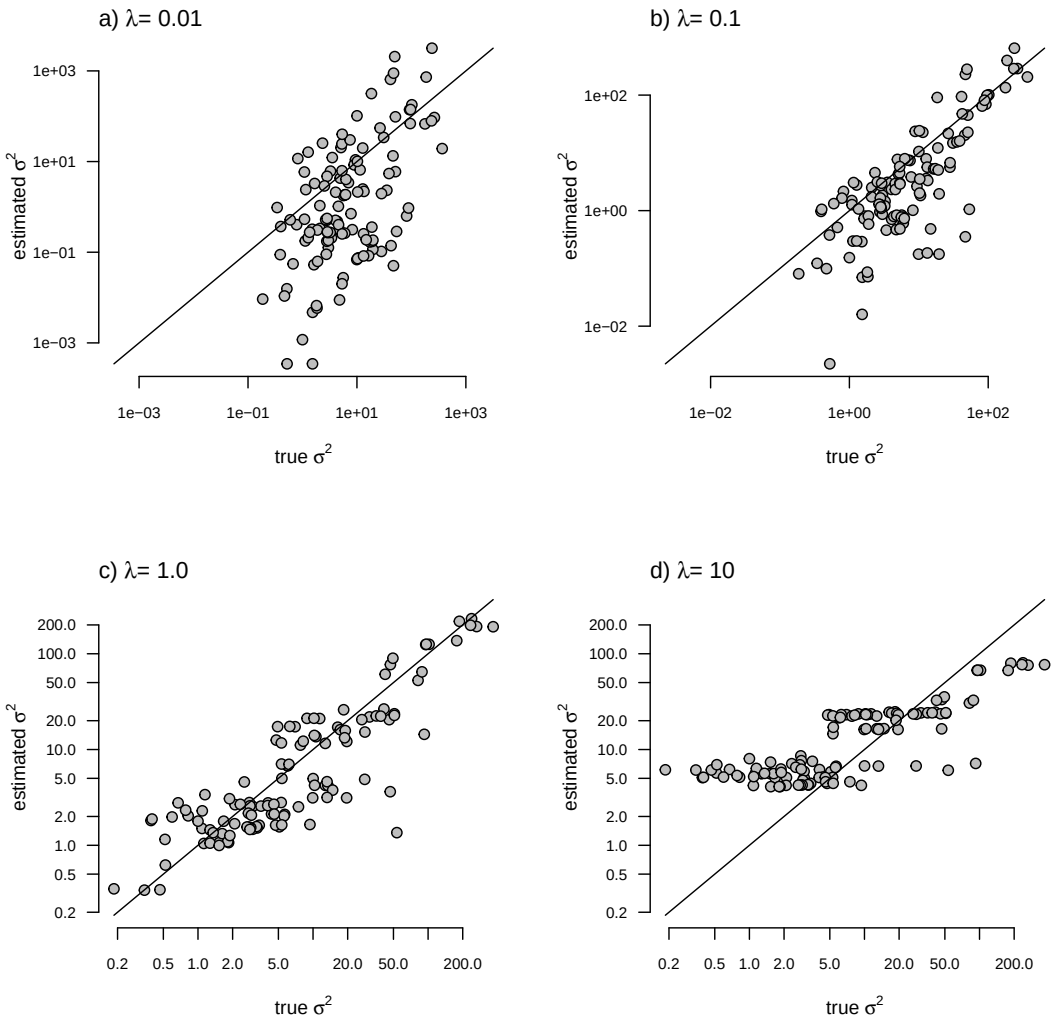


Figure 3. Estimated values of σ^2 (y) compared to their known true values (x) for the data of Figure 2 obtained using the penalized-likelihood approach with four different values of the penalty term coefficient, λ: a) $\lambda = 0.01$; b) $\lambda = 0.1$; c) $\lambda = 1$; and d) $\lambda = 10$. The 1:1 line is indicated.

We can see that the estimated values of σ^2 are in all cases quite strongly correlated with their generating values; however, (at least here) the strongest relationship is for an 'intermediate' value of λ ($\lambda = 1$), on the range that we chose to use. Readers should keep in mind, of course, that this should not generally be expected to be the case even if our rates have indeed evolved under the assumed process. I'll explain more about selecting appropriate values of λ in the Discussion.

Simulated example 2: Uncorrelated rates

In addition to generating data under our assumed model of Brownian evolution of the rates on a log scale, I also simulated *uncorrelated* rate variation among edges. In this case, I simply drew values for the log-rate of the ith node or tip, $\log(\sigma_i^2)$, randomly from a normal distribution with a standard deviation of $\sigma = 2.5$. I chose this particular value of σ *entirely* because it resulted in variability of σ^2 among edges that was magnitudinally similar to the variation obtained for Figure 2, in which I had simulated the evolutionary rate by Brownian motion. To simulate data on the tree using my new rates, I simply computed the mean rate for the subtending edge of each pair of nodes or tips using the same procedure as outlined above.

My simulated data for this example are given in Figure 4. Just as in Figure 2, above, panel a) shows the simulated rates on each edge of the phylogeny, while panel b) gives a projection of the tree into phenotype space in which the data vector used for the projection, **x**, was obtained using the rates of panel a). It should be evident from both figure panels that in this case there is little autocorrelation of nearby rates of evolution on the tree, just as we'd expect based on our simulation procedure (Figure 4).

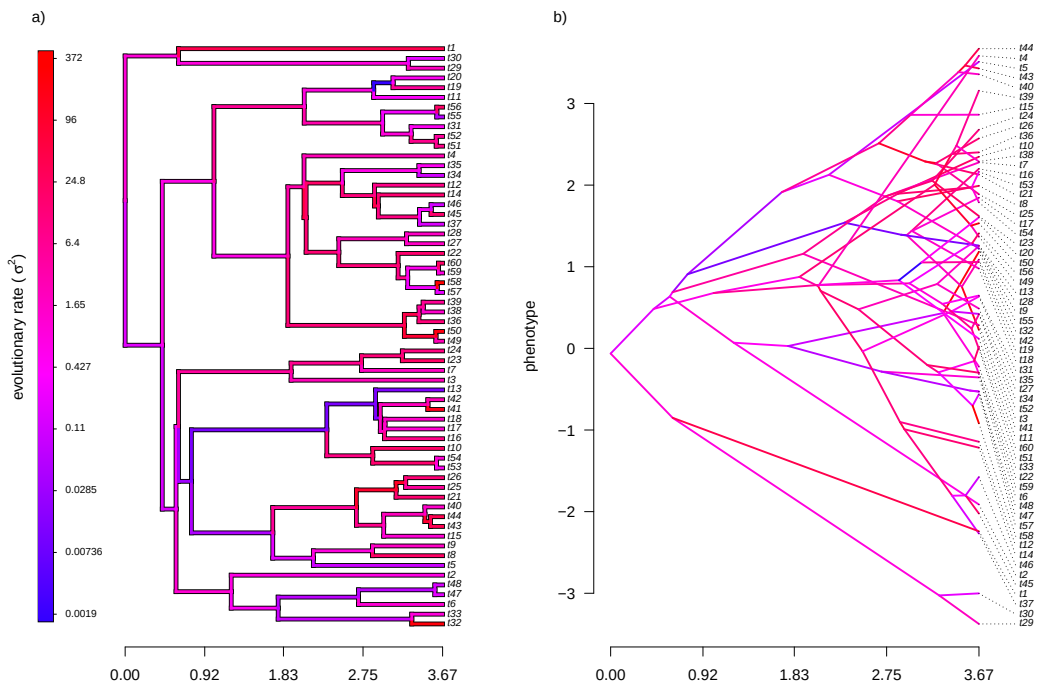


Figure 4. a) Simulated evolutionary rates, σ^2, in which the logarithm of the rate of evolution is uncorrelated between nodes and tips on the phylogeny. b) The tree projected into a space defined by time (on the horizontal axis) and a simulated trait vector, **x**, obtained using the rates of panel a).

With these data in hand, I next repeated exactly the same analysis that I undertook for the data of Figure 2 (once again, with four different values of λ: 0.01, 0.1, 1, and 10). I then, likewise, compared the *estimated* values of σ^2 to each of the values used in simulation. The results of this comparison are shown in Figure 5. Strikingly, the correlations between our generating and estimated values of σ^2 are much lower than they were in Figure 3 when the generating process for the rates was Brownian.

Simulated example 3: Discrete rate shifts

In addition to Brownian evolution of σ^2 (from the first simulation example), and uncorrelated σ^2 (from the second), I also simulated discrete *shifts* in the rate of Brownian evolution on the phylogeny. This is analogous to the fitted model in O'Meara et al. (2006; also see Revell and Collar, 2009; Beaulieu et al., 2012). Here, however, when I fit my model using penalized-likelihood, I will not inform my model-fitting method of the specific locations of the rate shifts on our phylogeny.

To generate data of this type, I first simulated the evolution of discrete regimes on the same tree as I used to create the data of Figures 2 and 4 under a continuous-time Markov chain (the Mk model of

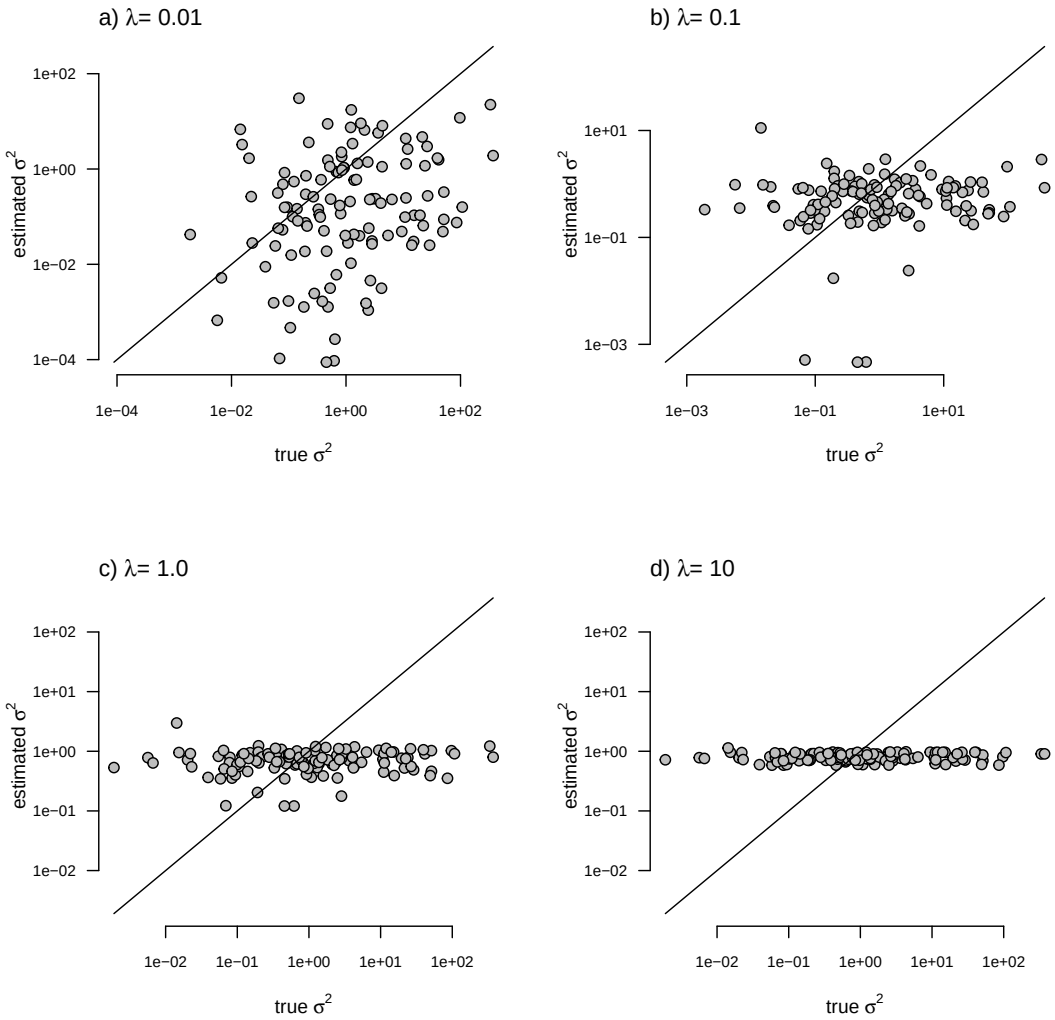


Figure 5. Estimated values of σ^2 (y) compared to their known true values (x) for the data of Figure 4 obtained using the penalized-likelihood approach using four different values of the penalty term coefficient, λ: a) $\lambda = 0.01$; b) $\lambda = 0.1$; c) $\lambda = 1$; and d) $\lambda = 10$. The 1:1 line is indicated.

Lewis, 2001; Harmon, 2019). Simply for the sake of convenience in subsequent calculations, I filtered any simulation in which more than one transition between regimes occurred on a single branch. In addition, I moved the transition between regimes to the precise midpoint of each edge. (Since our model for rate evolution is one in which nodes and tips have values of σ^2, this makes for an easier comparison between our simulation and our estimated rates.) I also constrained the substitution model of my rate regimes to be *ordered*. That is, transitions were only allowed to occur from low rate to medium rate, and *vice versa*; and from medium to high rate, and *vice versa*.

Having generated different stochastic regimes on the phylogeny, I then proceeded to simulate phenotypic trait evolution given these regimes. For my low, medium, and high rate regimes, I set σ^2 to be equal to 0.1, 1, and 10, respectively. The resultant phylogeny and simulated data are given in Figure 6.

Just as for the previous two examples, I fit our multi-rate Brownian motion model to these data using penalized-likelihood for four different values of the smoothing penalty coefficient, λ ($\lambda = 0.01$, 0.1, 1, and 10). I then compared the estimated values from our fitted models to the generating values from simulation. The results can be seen in Figure 7.

One substantive difference between Figure 7 and those prior to it is that in this case our nodes and tips

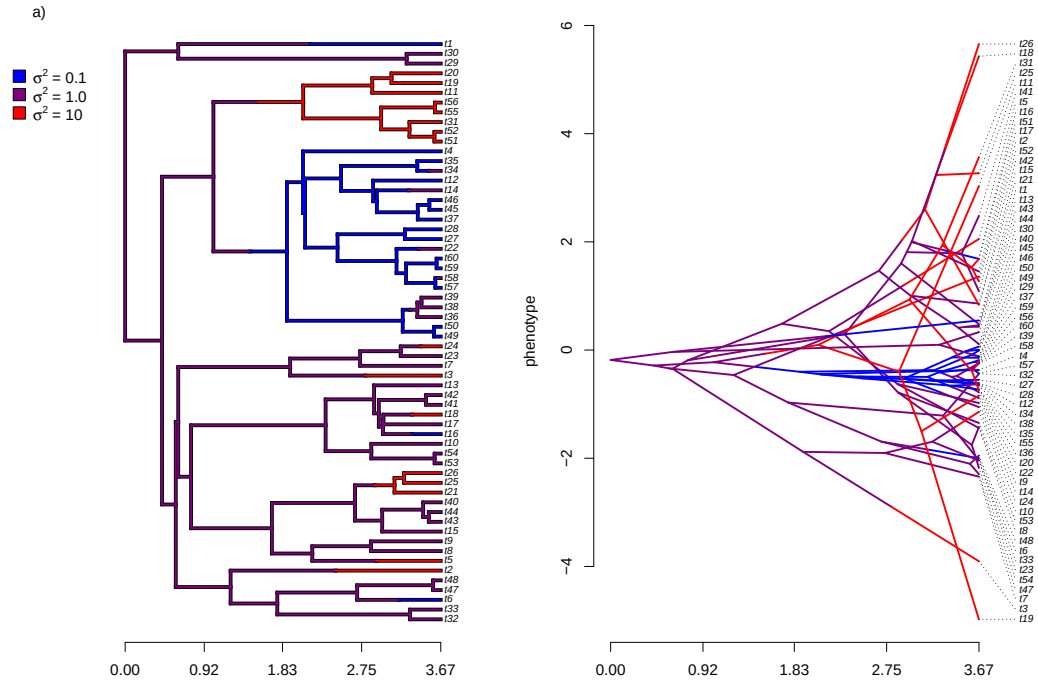


Figure 6. a) Simulated evolutionary rates, σ^2, in which rate of evolution shifts discretely under a continuous-time Markov process. b) The tree of panel a) projected into a space defined by time (on the horizontal axis) and a simulated trait vector, $\mathbf{x}$, obtained using the rates of panel a).

all belong to one of only three, distinct rate categories: $\sigma^2 = 0.1$, 1, or 10. As such, I though it would also be reasonable to compute the mean estimated value of σ^2 for each of the regimes, and then graph these mean values onto our plot. This analysis is also included in the figure. What's immediately striking about the result is the close alignment between the mean estimated and generating values of σ^2, particularly for lower values of the smoothing parameter λ. On the other hand, many of the individual edge-specific σ^2 estimates are very *far* from the true, generating rate values (Figure 7).

Simulated example 4: Invariant rates

For my last simulated example, I generated data for a single trait in which the rate of evolution, σ^2, was *constant* (and set to $\sigma^2 = 1$) over all the nodes and branches of the phylogeny. I then proceeded to fit a multi-rate model using penalized-likelihood, just as I did for all three of the previous simulations. Finally, I compared the estimated values, and their means, to the known, *true* value of σ^2 for the same four levels of our smoothing coefficient, λ ($\lambda = 0.01$, 0.1, 1, and 10). The result of this analysis can be seen in Figure 8.

In this figure, and in contrast to the equivalent plots from prior simulations, I've intentionally kept the x and y axes constant between each figure panel. What the plots show is, firstly, that the *mean* σ^2 across all edges of the phylogeny under our multi-rate penalized-likelihood analysis (shown using the heavy black dot on each panel) tends to be quite similar to the *true* single rate across each of the four values of λ explored here, as well as to the Maximum Likelihood estimate of σ^2 for a single rate model (shown using the horizontal dotted line; Figure 8), as estimated using the *geiger* R package (Pennell et al., 2014). Secondly, the plot shows that for increasing values of λ, variation among edges of the tree decreases and eventually *converges* on the Maximum Likelihood estimate for a single rate (Figure 8). Both of these results are properties that we expected for our estimation procedure.

Analysis of body mass evolution in mammals

Finally, I also analyzed a simple dataset for overall body mass evolution in a small phylogeny of mammals. These data, which consist of values for mass in kg transformed to the log scale and a phylogeny of 49

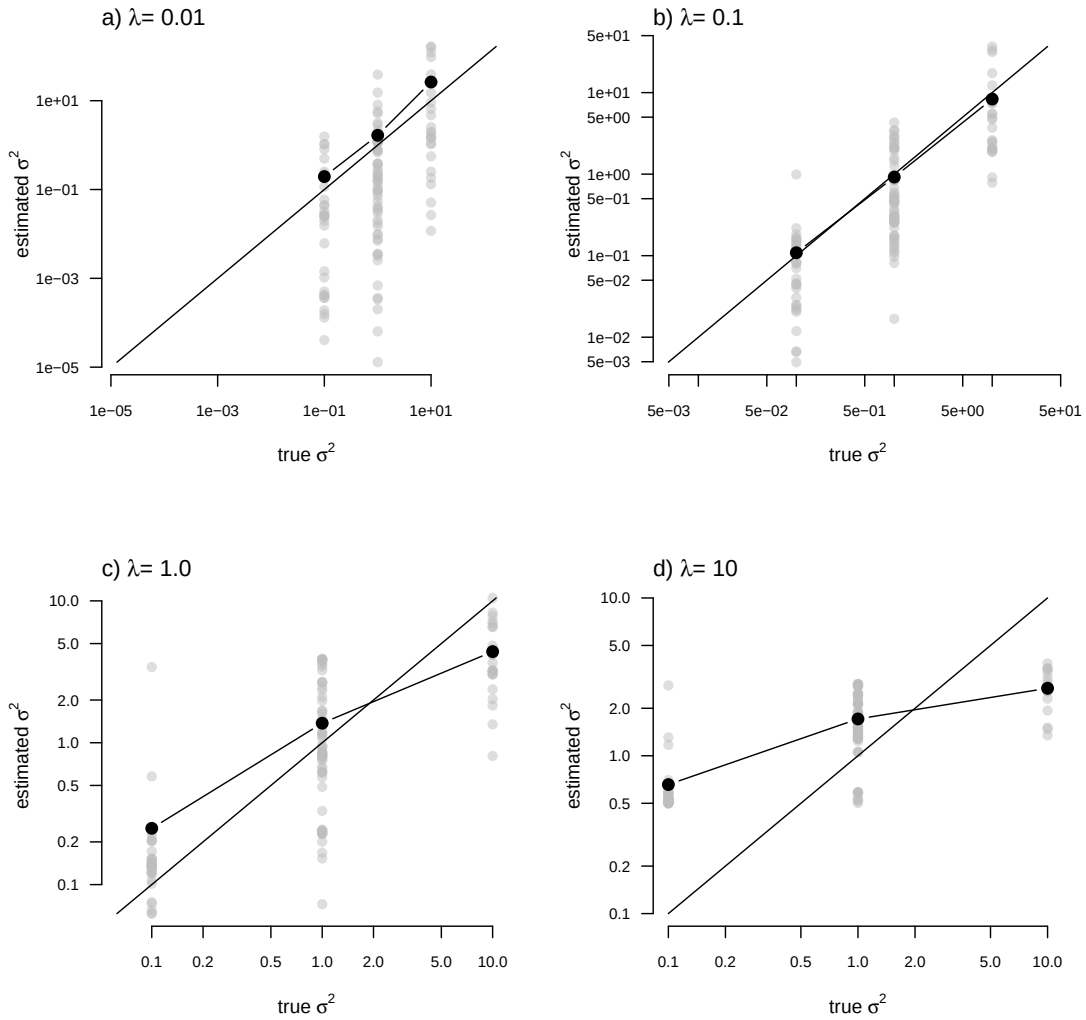


Figure 7. Estimated values of σ^2 (y) compared to their known true values (x) for the data of Figure 6 obtained using the penalized-likelihood approach using four different values of the penalty term coefficient, λ: a) $\lambda = 0.01$; b) $\lambda = 0.1$; c) $\lambda = 1$; and d) $\lambda = 10$. The 1:1 line is indicated.

mammalian species, are from Garland et al. (1992) and come packaged with the R phylogenetics library *phytools* (Revell, 2012). I fit our variable-rate Brownian model with penalized-likelihood using the same four values of λ as I employed with simulated data ($\lambda = 0.01$, 0.1, 1, and 10). Then, I graphed the fitted models using the corresponding *phytools* plotting method for our fitted model object. This maps the estimated rates onto the edges of the tree and can use a color gradient that is specified by the user. Here, I used an inverted `heat.colors` gradient from the R package *grDevices* (R Core Team, 2020).

What's most striking about the result (shown in Figure 9) is that estimated rates are highly correlated between analyses using different values of λ, even if their absolute magnitudes differ. Qualitatively, this analysis consistently reveals that the *highest* rates of body mass evolution in our tree are found among the Artiodactyla (even toed ungulates, such as deer, moose, bison, etc.), while the *lowest* rates of body mass evolution are observed for the Perissodactyla (odd-toed ungulates, such as horses and tapirs; Figure 9).

Notes on implementation

The model and methods of this study have been implemented for the R statistical computing environment (R Core Team, 2020), and all simulations and analysis were conducted in R. The penalized-likelihood

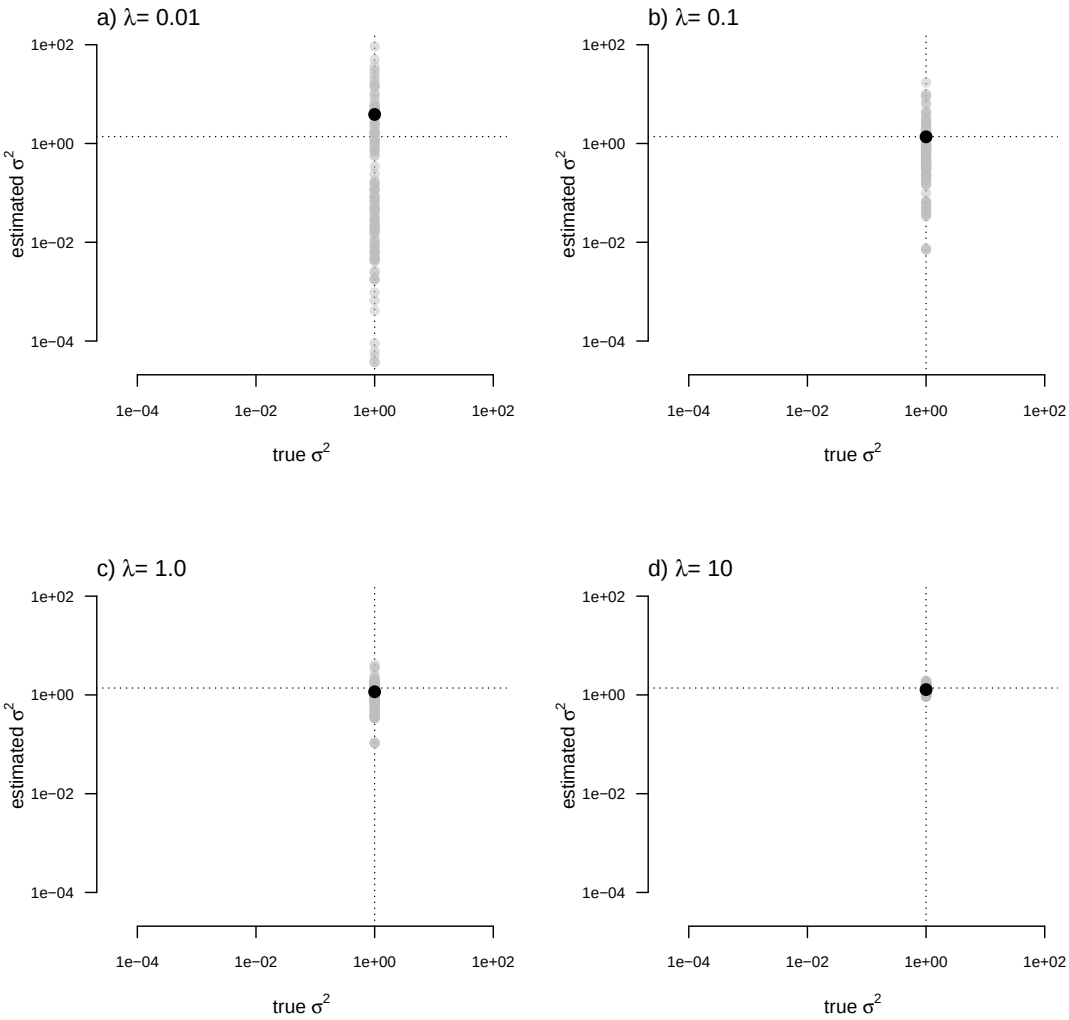


Figure 8. Estimated values of σ^2 (y) compared to their known true values (x) for data simulated with a constant rate of evolution, $\sigma^2 = 1$. Vertical and horizontal lines show the true rate of evolution from the simulation, and the Maximum Likelihood estimate of the rate obtained in a single-rate analysis.

method that I describe in the article is implemented as the function `multirateBM` of my *phytools* R package (Revell, 2012).

phytools in turn depends on the important R phylogenetics packages *ape* (Paradis and Schliep, 2019) and *phangorn* (Schliep, 2011), as well as on a number of other R libraries (Venables and Ripley, 2002; Ligges and Mächler, 2003; Lemon, 2006; Plummer et al., 2006; Chasalow, 2012; Becker et al., 2018; Gilbert and Varadhan, 2019; Azzalini and Genz, 2020; Qiu and Joe., 2020; Warnes et al., 2020; Goulet et al., 2021; Pinheiro et al., 2021). Finally, some analyses in the latter part of the article were done with the help of the *geiger* package (Pennell et al., 2014).

DISCUSSION

Phylogenetic comparative analysis is one of the leading weapons in the arsenal of evolutionary biologists interested in understanding and characterizing the history of life on this planet (Nunn, 2011; Garamszegi, 2014; Harmon, 2019). Many phylogenetic comparative methods have been developed in recent years, and a number of these are specifically designed to model heterogeneity in the tempo or mode of evolution among branches and clades of the tree (e.g., O'Meara et al., 2006; Revell and Harmon, 2008; Revell et al.,

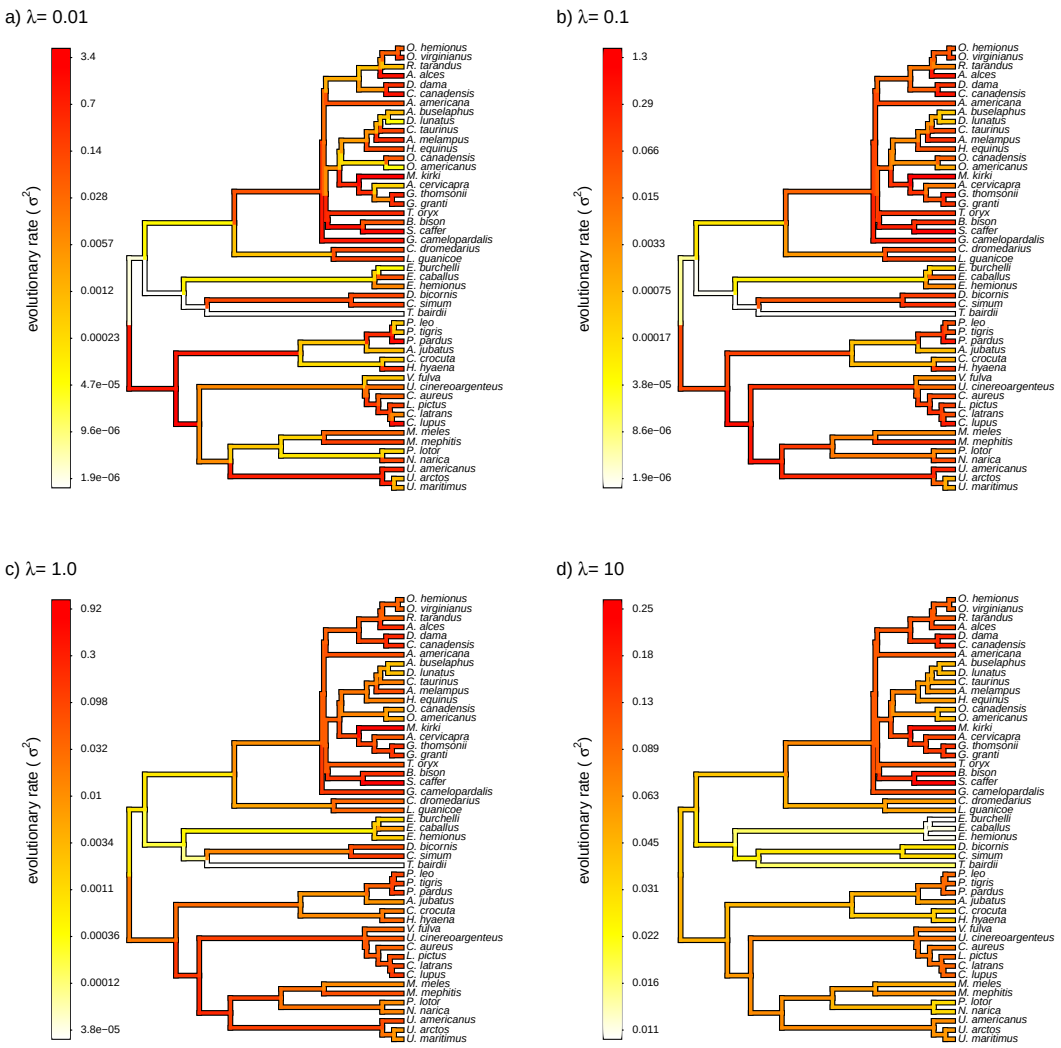


Figure 9. Fitted variable-rate Brownian motion model for log(body mass) evolution in 49 species of mammals. Different values of λ correspond to different penalty coefficients for rate variation among edges of the tree. Note that each panel has a different scale.

2011; Beaulieu et al., 2012; Uyeda and Harmon, 2014; Caetano and Harmon, 2017).

In the present article, I describe a new phylogenetic comparative method – implemented as a function in my *phytools* R package (Revell, 2012). In this method, which I've developed to analyze the evolution of a single continuously-valued trait on the phylogeny, the rate of evolution (σ^2) is free to vary from branch to branch in the tree. Model-fitting is done using the technique of penalized-likelihood, in which the penalty term is a function of the negative log probability density of the rates, assuming that the logged rates themselves evolve by a Brownian process on the phylogeny.

I've applied this method to datasets simulated under a variety scenarios of potential interest to biologists that might consider using this approach with their own data. The first of these, in which the logarithm of the rate of evolution ($\log(\sigma^2)$) evolves by Brownian motion, is the underlying model that is *assumed* by the analysis method of this article. I show that for a range of different values of the smoothing parameter, λ, estimated evolutionary rates are correlated with the generating rates of the simulation (Figure 3).

Under other simulation conditions we see different results. For instance, when the evolutionary rates are random and *uncorrelated* among edges, our estimated rates from the penalized-likelihood method also

tend to be correlated with the generating rates used for simulation (Figure 5) – though this correlation is substantially weaker than when the generating rates evolved via a Brownian process. When data were generated with discrete rate shifts on the tree, the mean estimated rate across edges in each regime closely matched the generating value of the rate (Figure 7). Finally, when there was *no* variation at all in rate among the edges of the phylogeny, the arithmetic mean estimated rate tended to quite closely match the Maximum Likelihood value of σ^2 from a simple, one-rate model (Figure 8).

In addition to this simulation, I applied the model to a classic empirical dataset of log body mass in 49 species of mammals (Garland et al., 1992). Although the variation in estimated rates among edges decreased with increasing λ, the qualitative pattern of variation was quite similar among different λ values – generally showing the highest rates of evolution of body mass in artiodactyls (even-toed ungulates, such as deer, cows, and antelopes), and the *lowest* estimated values of σ^2 for the perissodactyls (odd-toed ungulates, such as horses and tapirs).

Why not just estimate σ^2_{BM} (or λ)?

Earlier in the article, I mentioned that it was probably not possible to simultaneously estimate the evolutionary rates for each edge or node in the tree (σ^2_0, σ^2_1, and so on) *and* the rate of evolution of the rates themselves (let's call it σ^2_{BM}); however, at the time I didn't say why this seemed likely to be the case. An imaginable (but ultimately untractable, as we'll see) solution to this problem would be to simply take the following expression for the likelihood (in which the first part is the probability density of our trait observations given our rates; and the second part is the probability of our rates, given an underlying model for their evolution) and maximize it.

$$l(\sigma^2_0,\sigma^2_1,...,\sigma^2_{BM},x_0|\mathbf{x},\mathbf{C}_{ext}) = \frac{\exp[-\frac{1}{2}(\mathbf{x}-\mathbf{1}x_0)'\mathbf{T}(\mathbf{x}-\mathbf{1}x_0)]}{\sqrt{2\pi^n|\mathbf{T}|)}}\times \frac{\exp[-\frac{1}{2}(\mathbf{s}-\mathbf{1}s_0)'(\sigma^2_{BM}\mathbf{C}_{ext})(\mathbf{s}-\mathbf{1}s_0)]}{\sqrt{2\pi^n|\sigma^2_{BM}\mathbf{C}_{ext}|)}} \tag{4}$$

In equation (4), all the terms are as previously defined for equations (1), (2), and (3); plus σ^2_{BM} is the rate of evolution of the log-rates (s_0, s_1, etc.), and $\mathbf{1}$ is a conformable vector of 1.0s ($n\times 1$ in the first part of the equation, and $n+m-1\times 1$ in the second).

The problem with this logic, unfortunately, is that equation (4) will *always* go to ∞ as $\sigma^2_{BM}\rightarrow 0$. This is simply because a proper probability density function integrates to 1.0, and since the s_is are not observed, the second part of equation (4) will have zero width and infinite height when all values of s_i are equal and $\sigma^2_{BM}=0$. A similar difficulty would apply to any scaling coefficient (such as λ) that applied to the second term in our likelihood expression but not the first.

How to choose a value of λ

I've so far presented a new, penalized-likelihood approach for modeling heterogeneity in the rate of evolution for a continuously-valued trait on the phylogeny. Unfortunately, since neither λ nor the rate parameter of the Brownian evolutionary process of the rates are estimable, our method requires that we simply assign a value for the penalty or smoothing coefficient of our fitted model. Here, I employed the relatively simplistic tactic of trying multiple values of λ for each simulated or empirical dataset, and then comparing the results. Analyses like those shown in Figure 9 suggest that it might be possible to draw some credible inferences about evolution using this strategy.

Nonetheless, most readers are probably interested in a more rigorous approach to identifying reasonable values of the penalty or smoothing coefficient, λ. Although I've not yet explored it for the model of this paper, a tactic that seems to work well for other penalized-likelihood methods is *cross-validation* (e.g., Sanderson, 2002; Smith and O'Meara, 2012). In cross-validation we would first select a value of λ and fit the model using this value. Next, we would successively drop one or more observations (tips) from our tree (either at random or one-by-one), re-estimate rates for all nodes, and then compare our estimated rates to their values from the model when it had been fit using the full dataset. Our preferred value for λ would be the one that minimized the sum of squared differences between the measured rates in the global estimate and those computed from our subsampling procedure (Sanderson, 2002; Smith and O'Meara, 2012). I've not yet implemented this method in *phytools*, but it wouldn't be difficult to

script in R for *phytools* users already adept at such things. Unfortunately, model-fitting using the *phytools* function in which this method is implemented (`multirateBM`) is *extremely* time-consuming, and so this cross-validation procedure would take a very long time to run on even a single dataset as it requires that the model be fit to our data hundreds of times or more in one analysis (Smith and O'Meara, 2012).

Conclusions

Numerous phylogenetic comparative methods have been developed in recent years. Many of these are designed to model heterogeneity in tempo and mode across the nodes and branches of a phylogeny. Here, I present a simple penalized-likelihood method for modeling and graphing variation in the tempo of evolutionary change across the edges of a phylogeny. The method requires that we (arbitrarily) select a penalty or smoothing coefficient, λ, that will in turn greatly affect our inference about the magnitude of rate heterogeneity among branches in our tree. Nonetheless, I suggest that a cross-validation approach (following Smith and O'Meara, 2012) might prove useful for more rigorous selection of this coefficient.

ACKNOWLEDGMENTS

Thanks to J. Berv, L. Harmon, and L. Schmitz for useful comments on this project as it was developed. This research was funded by grants from the National Science Foundation (DBI-1759940) and FONDECYT, Chile (1201869).

REFERENCES

Azzalini, A. and Genz, A. (2020). *The R package `mnormt`: The multivariate normal and t distributions*. R package verison 2.0.2.

Beaulieu, J. M., Jhwueng, D.-C., Boettiger, C., and O'Meara, B. C. (2012). Modeling stabilizing selection: Expanding the ornstein-uhlenbeck model of adaptive evolution. *Evolution*, 66(8):2369–2383.

Becker, R. A., Wilks, A. R., Brownrigg, R., Minka, T. P., and Deckmyn, A. (2018). *maps: Draw Geographical Maps*. R package version 3.3.0.

Butler, M. A. and King, A. A. (2004). Phylogenetic comparative analysis: A modeling approach for adaptive evolution. *The American Naturalist*, 164(6):683–695.

Caetano, D. S. and Harmon, L. J. (2017). ratematrix: An r package for studying evolutionary integration among several traits on phylogenetic trees. *Methods in Ecology and Evolution*, 8(12):1920–1927.

Chasalow, S. (2012). *combinat: combinatorics utilities*. R package version 0.0-8.

Evans, M. E. K., Smith, S. A., Flynn, R. S., and Donoghue, M. J. (2009). Climate, niche evolution, and diversification of the "bird-cage" evening primroses (oenothera, sections anogra and kleinia). *The American Naturalist*, 173(2):225–240.

Felsenstein, J. (1985). Phylogenies and the comparative method. *The American Naturalist*, 125(1):1–15.

Felsenstein, J. (2004). *Inferring phylogenies*. Sinauer Associates, Sunderland, Mass.

Garamszegi, L. (2014). *Modern phylogenetic comparative methods and their application in evolutionary biology : concepts and practice*. Springer, Heidelberg.

Garland, T., Harvey, P. H., and Ives, A. R. (1992). Procedures for the analysis of comparative data using phylogenetically independent contrasts. *Systematic Biology*, 41(1):18–32.

Gilbert, P. and Varadhan, R. (2019). *numDeriv: Accurate Numerical Derivatives*. R package version 2016.8-1.1.

Goulet, V., Dutang, C., Maechler, M., Firth, D., Shapira, M., and Stadelmann, M. (2021). *expm: Matrix Exponential, Log, 'etc'*. R package version 0.999-6.

Harmon, L. (2019). *Phylogenetic Comparative Methods*. Independent, Open Textbook Library, Minneapolis.

Harvey, P. (1996). *New uses for new phylogenies*. Oxford University Press, Oxford New York.

Lemon, J. (2006). Plotrix: a package in the red light district of r. *R-News*, 6(4):8–12.

Lewis, P. O. (2001). A likelihood approach to estimating phylogeny from discrete morphological character data. *Systematic Biology*, 50(6):913–925.

Ligges, U. and Mächler, M. (2003). Scatterplot3d - an r package for visualizing multivariate data. *Journal of Statistical Software*, 8(11):1–20.

Nunn, C. (2011). *The comparative approach in evolutionary anthropology and biology*. University of Chicago Press, Chicago.

O'Meara, B. C. (2012). Evolutionary inferences from phylogenies: A review of methods. *Annual Review of Ecology, Evolution, and Systematics*, 43(1):267–285.

O'Meara, B. C., Ané, C., Sanderson, M. J., and Wainwright, P. C. (2006). Testing for different rates of continuous trait evolution using likelihood. *Evolution*, 60(5):922–933.

Paradis, E. and Schliep, K. (2019). ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics*, 35:526–528.

Pennell, M. W., Eastman, J. M., Slater, G. J., Brown, J. W., Uyeda, J. C., FitzJohn, R. G., Alfaro, M. E., and Harmon, L. J. (2014). geiger v2.0: an expanded suite of methods for fitting macroevolutionary models to phylogenetic trees. *Bioinformatics*, 30(15):2216–2218.

Philippe, H., Delsuc, F., Brinkmann, H., and Lartillot, N. (2005). Phylogenomics. *Annual Review of Ecology, Evolution, and Systematics*, 36(1):541–562.

Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., and R Core Team (2021). *nlme: Linear and Nonlinear Mixed Effects Models*. R package version 3.1-152.

Plummer, M., Best, N., Cowles, K., and Vines, K. (2006). Coda: Convergence diagnosis and output analysis for mcmc. *R News*, 6(1):7–11.

Qiu, W. and Joe., H. (2020). *clusterGeneration: Random Cluster Generation (with Specified Degree of Separation)*. R package version 1.3.7.

R Core Team (2020). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Revell, L. J. (2012). phytools: An r package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution*, 3:217–223.

Revell, L. J. (2013). Two new graphical methods for mapping trait evolution on phylogenies. *Methods in Ecology and Evolution*, 4(8):754–759.

Revell, L. J. and Collar, D. C. (2009). Phylogenetic analysis of the evolutionary correlation using likelihood. *Evolution*, 63(4):1090–1100.

Revell, L. J. and Harmon, L. J. (2008). Testing quantitative genetic hypotheses about the evolutionary rate matrix for continuous characters. *Evolutionary Ecology Research*, 10:311–321.

Revell, L. J., Harmon, L. J., and Collar, D. C. (2008). Phylogenetic signal, evolutionary process, and rate. *Systematic Biology*, 57(4):591–601.

Revell, L. J., Mahler, D. L., Peres-Neto, P. R., and Redelings, B. D. (2011). A new phylogenetic method for identifying exceptional phenotypic diversification. *Evolution*, 66(1):135–146.

Rohlf, F. J. (2001). Comparative methods for the analysis of continuous variables: Geometric interpretations. *Evolution*, 55(11):2143–2160.

Sanderson, M. J. (2002). Estimating absolute rates of molecular evolution and divergence times: A penalized likelihood approach. *Molecular Biology and Evolution*, 19(1):101–109.

Schliep, K. P. (2011). phangorn: phylogenetic analysis in r. *Bioinformatics*, 27(4):592–593.

Schluter, D., Price, T., Mooers, A. Ø., and Ludwig, D. (1997). Likelihood of ancestor states in adaptive radiation. *Evolution*, 51(6):1699–1711.

Smith, S. A. and O'Meara, B. C. (2012). treePL: divergence time estimation using penalized likelihood for large phylogenies. *Bioinformatics*, 28(20):2689–2690.

Uyeda, J. C. and Harmon, L. J. (2014). A novel bayesian method for inferring and interpreting the dynamics of adaptive landscapes from phylogenetic comparative data. *Systematic Biology*, 63(6):902–918.

Venables, W. N. and Ripley, B. D. (2002). *Modern Applied Statistics with S*. Springer, New York, fourth edition. ISBN 0-387-95457-0.

Warnes, G. R., Bolker, B., and Lumley, T. (2020). *gtools: Various R Programming Tools*. R package version 3.8.2.