A discrete character evolution model for phylogenetic comparative biology with Γ-distributed rate heterogeneity among branches of the tree

Liam J. Revell[1, 2] **and Luke J. Harmon**[3]

[1]**Department of Biology, University of Massachusetts Boston, Boston, MA, USA**
[2]**Facultad de Ciencias, Universidad Católica de la Santísima Concepción, Concepción, Chile**
[3]**Department of Biological Sciences, Institute for Bioinformatics and Evolutionary Studies (IBEST), University of Idaho, Moscow, ID, USA**

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

Email address: `liam.revell@umb.edu`

ABSTRACT

Phylogenetic comparative methods are now widely used to measure trait evolution on the tree of life. Often these methods involve fitting an explicit model of character evolution to trait data and then comparing the explanatory power of this model to alternative scenarios. In this article, we present a new model for discrete trait evolution in which the rate of character change in the tree varies from edge (i.e., "branch") to edge of the phylogeny according to a discretized Γ distribution. When the edge-wise rates of evolution are, in fact, Γ-distributed, we show via simulation that this model can be used to reliably estimate the shape parameter (α) of the distribution of rate variation among edges. We also describe how this model can be employed in ancestral state reconstruction, and demonstrate via simulation how doing so will tend to increase the accuracy of our estimated states when the generating rates are, in fact, Γ-distributed. Finally, we apply the method to a real dataset of digit number in squamate reptiles, modified from Brandley et al. (2008).

1 INTRODUCTION

One of the major endeavors of phylogenetic comparative analysis is to combine an estimated phylogeny with phenotypic trait data for the species of our tree in an effort to better understand the evolution and history of our study clade (Harmon 2019; Revell and Harmon 2022). This project frequently requires that we specify an explicit mathematical model postulating how our traits, such as body size, limb lengths, the presence or absence of some feature, or a category of habitat specialization, may have evolved on the tree during macroevolutionary time. The model might take a variety of flavors, each with its own parameters and attributes. In that case, our analysis could simply involve fitting a set of such models, each capturing a slightly different scenario of trait evolution, to learn which of them best explains the data that we have in our possession.

When our trait is discrete ~~in nature~~ (that is, the character only manifests in one of a limited set of conditions), the predominant model employed by comparative biologists to approximate its evolution on the tree is one that's popularly referred to as the M*k* model (Lewis 2001; Harmon 2019; Revell and Harmon 2022). This model is so named because the underlying stochastic process it postulates is a memoryless Markov chain (the 'M'), with a *k*-dimensional state space. What we refer to here as the 'state space' is just the number of unique conditions that the Markov chain can assume: two, for example, if our character can take on two different levels; three, if it might exist in three values; etc. Though originally conceived as a model with equal back-and-forth rates of transition between the two, three, or more conditions in the space (Lewis 2001), the M*k* model has now been elaborated to permit an arbitrary

number of different rates of character transition between levels (described in Harmon 2019; Revell and Harmon 2022). Fitting such a model could involve estimating up to $k \times (k-1)$ parameters for k states (all non-diagonal elements of a $k \times k$ state transition matrix, typically called the **Q** matrix).

In addition to simply permitting the rates of evolution between character states to differ one from the other, the Mk model has also been extended in a number of additional interesting ways. For instance, Pagel (1994) described a clever manner in which an Mk-type model can be used to approximate the interdependent evolution of two different binary traits. Later, Marazzi et al. (2012) and Beaulieu et al. (2013) proposed a different model (called the precursor or hidden-rates model, depending on the way in which it's parameterized) whereby one or multiple unobserved conditions can influence our trait's transition rate to other observed values (also see Revell and Harmon 2022). Revell and Harmon (2022) described a polymorphic trait evolution model in which the polymorphic condition (e.g., $a+b$) is postulated to be evolutionarily intermediate between corresponding monomorphic states (a and b, in this case). Finally, Revell (in review) defined a model in which the rate of evolution of a discrete character is influenced by an explicit, *a priori* hypothesis of rate heterogeneity that's been painted onto the edges and nodes of the phylogeny by the investigator.

In the present article, we'll describe a new extension of the Mk model in which the edge-wise rates of evolution of our trait are sampled randomly from a (normalized) Γ distribution with shape parameter α. This model is in many ways analogous to the popular Γ rate heterogeneity model for molecular sequence evolution that's commonly used for phylogeny estimation (e.g., Yang 1994; Felsenstein 2004). In our case, however, we'll be assuming that that rate of evolution for a single discrete trait changes from edge to edge according to an uncorrelated Γ (rather than varying by nucleotide site). This article is intended to provide a description of the method, which has already been implemented in the popular R package *phytools* (Revell 2012), a preliminary exploration of its statistical properties, and an empirical example based on the evolution of digit number in squamate reptiles using data adapted from Brandley et al. (2008).

# 2 MODEL, METHODS, AND RESULTS

2.1 The model

This article describes a new discrete phenotypic trait evolution model in which the rate of character evolution varies from edge to edge according to an uncorrelated Γ distribution with shape parameter α and mean of 1.0. (A mean of 1.0 is obtained by setting the second parameter of the Γ distribution, β, to $\beta = \alpha$.) The Γ distribution is a flexible probability distribution that takes a shape resembling that of an exponential function for low values of α (e.g., $\alpha \leq 1.0$), and an increasingly bell-like conformation for larger and larger values of α (e.g., $\alpha \geq 4$). Figure 1A shows a set of Γ distributions for various values of α.

Under a standard Mk model, the rates of transition between character levels are dictated by a $k \times k$ matrix **Q** in which each element, $q_{i,j}$ (for all $i \neq j$), gives the instantaneous transition rate from condition i to condition j, and the diagonal is equal to the negative row sum (Revell and Harmon 2022). For a binary trait (i.e., $k = 2$), this transition matrix (**Q**) would appear as follows.

$$\mathbf{Q} = \begin{bmatrix} -q_{0,1} & q_{0,1} \\ q_{1,0} & -q_{1,0} \end{bmatrix} \tag{1}$$

To compute the probability of character change under our ~~stochastic process~~ given **Q** and any arbitrary elapsed time, t, we simply calculate $\mathbf{P} = \exp(\mathbf{Q} \times t)$, in which $\exp(\mathbf{X})$ indicates the matrix exponential of **X**. This operation gives us a new $k \times k$ matrix, **P**, in which $p_{i,j}$ is equal to the probability of beginning the time interval, t, in condition i and ending in condition j under our modeled process.

To see how we then go about calculating the probability of a set of data at the tips of the tree for any particular value of **Q**, we can begin by considering the simplified tree and data pattern of Figure 1B. In this tree, the two nominal taxa of our phylogeny, A and B, each exhibit different conditions for the character: taxon A has the state 0 (shown in black); while taxon B is found in condition 1 (shown in white). Logically, the probability of observing this set of states must be equal to the probability of starting (at the root) in state 0, here written as π_0, multiplied by the probability of finding the single internal node of our tree in state 0 after elapsed time t_{AB}, $P(0|0, t_{AB})$, multiplied by the probability of the tip A being observed in state 0 after elapsed time t_A, $P(0|0, t_A)$, multiplied in turn by the probability of tip B being
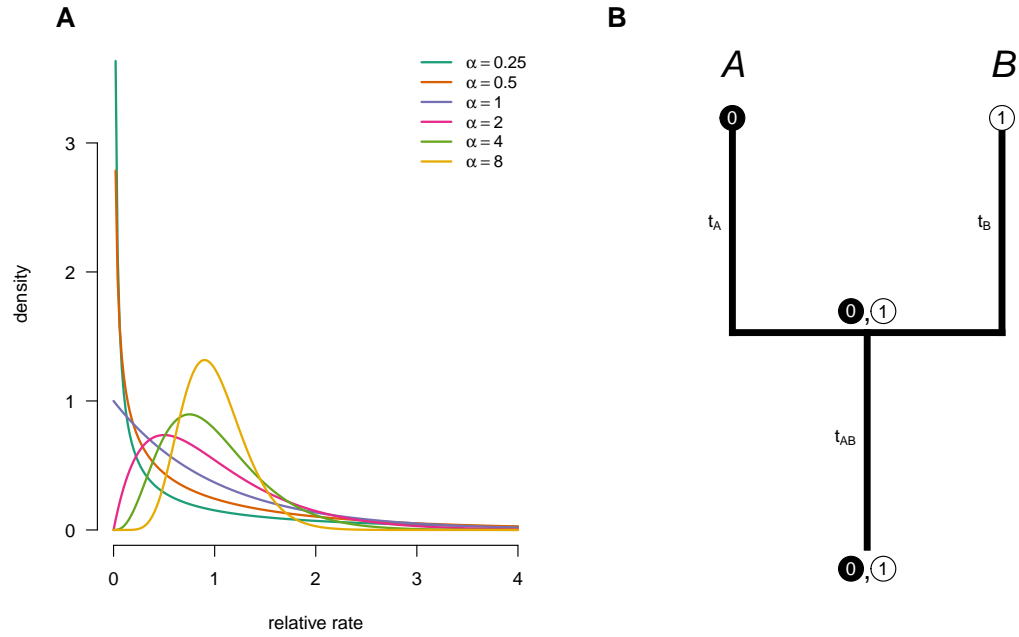
Figure 1. (**A**) A set of Γ probability distribution with different values of the shape parameter, α. By setting $\beta = \alpha$ the mean of each distribution is equal to 1.0, while α determines the shape of the distribution in which $\alpha < 1$ correspond to distributions with high among edge rate heterogeneity. (**B**) An example phylogeny with two taxa (*A* and *B*) and two trait values of a discrete character (0 and 1). See main text for more details.

observed in state 1 after elapsed time t_B, $P(1|0, t_B)$ – and then summed across both possible root state and all four possible internal node state combinations. When written down, this probability ~~looks~~ as follows.

$$
\begin{aligned}
P(0, 1) = {} & \pi_0 \times P(0|0, t_{AB}) \times P(0|0, t_A) \times P(1|0, t_B) \\
& + \pi_0 \times P(1|0, t_{AB}) \times P(0|1, t_A) \times P(1|1, t_B) \\
& + \pi_1 \times P(1|1, t_{AB}) \times P(0|1, t_A) \times P(1|1, t_B) \\
& + \pi_1 \times P(0|1, t_{AB}) \times P(0|0, t_A) \times P(1|0, t_B)
\end{aligned}
\tag{2}
$$

Though for a two-taxon tree, such as the one shown in Figure 1B, the value of equation (2) is straightforward to compute, doing so in the same way for a tree with three, four, or more species would be increasingly onerous if it involved the explicit enumeration of all possible states at all internal nodes of the phylogeny as we've done here ~~for illustrative purposes only.~~ Fortunately for us, Felsenstein (1981) described an efficient "pruning" algorithm to compute this probability via a single post-order (tip to root) traversal of the tree.

Now to account for variation in the rate of evolution from edge to edge of our phylogeny, we'll use a discretization of the Γ distribution following Yang (1994). Directly following Yang (1994), we first set m rate categories (k in Yang 1994, but m here to avoid confusion with our k different character states). We then subdivide the range of r, our relative rates (which could theoretically vary from 0 to ∞, the range of a Γ-distributed random variable), into m evenly spaced bins. To compute the median of each bin, we need to find the values of each r_i (that is, $r_1, r_2, \ldots, r_m$) at the percentiles of our Γ distribution that correspond to the midpoints of each of m bins: $1/(2m), 3/(2m), \ldots, (2m-1)/(2m)$ (Yang 1994): quantities that are straightforward to calculate using a computer. Referring back, once again, to our two-taxon case of Figure 1B, the probability of our data in which the relative edge rates vary as a random variable drawn from an uncorrelated Γ distribution can be approximated as follows.
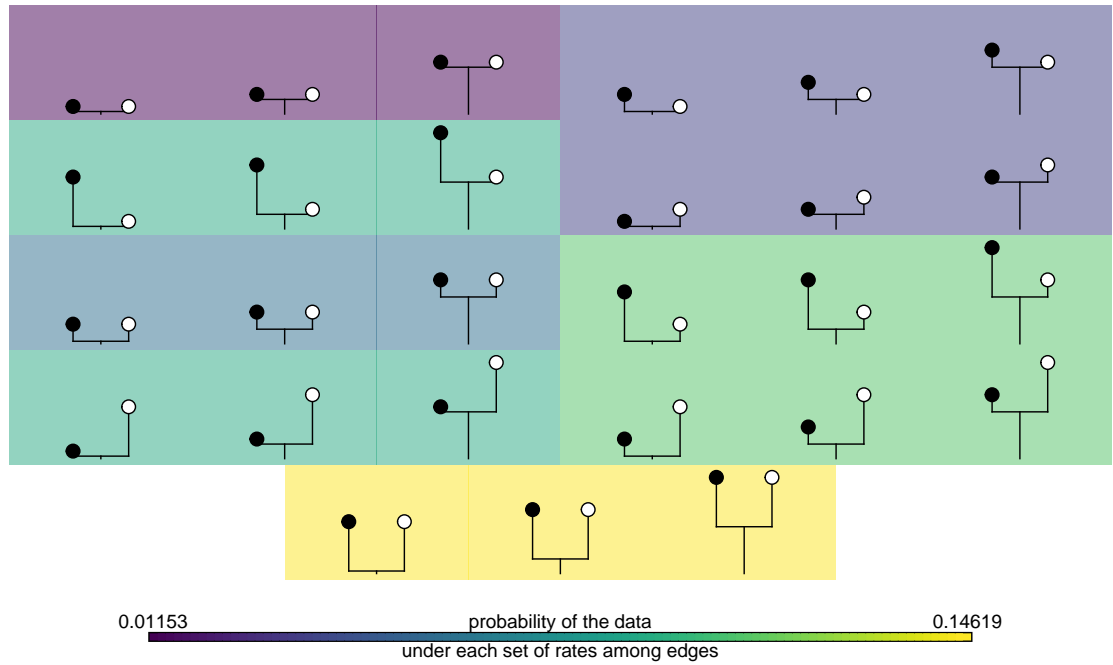
Figure 2. The probability of the tip data from Figure 1B given all the 27 possible ways in which three rate categories (i.e., $m = 3$) could be distributed across the three edges of our two taxon tree, in which $q_{0,1} = q_{1,0} = 0.1$ and $\alpha = 0.7$. To compute the total probability of the data at the tips of the phylogeny in Figure 1B whilst integrating over discretized Γ-distributed rates across edges, we can simply sum these values and divide by 27.

$$P(0,1) = \frac{1}{m^3} \sum_{i=1}^{m} \sum_{j=1}^{m} \sum_{k=1}^{m} P(0,1|r_A = r_i, r_B = r_j, r_{AB} = r_k) \tag{3}$$

Figure 2 illustrates a complete enumeration of the $3^3 = 27$ possible ways that $m = 3$ rates could be distributed across the three edges of our tree of Figure 1B, along with the probability of our observed data at the tips of the tree for $q_{0,1} = q_{1,0} = 0.1$ and $\alpha = 0.7$. To calculate the total probability under this model, we would just sum all of these quantities and divide by 27, as given in equation (3), above.

As before, this sum would be very onerous to compute (and, indeed, impossible for even relatively modest-sized trees) were it not for the pruning algorithm of Felsenstein (1981). To implement pruning with (discrete) Γ-distributed rate heterogeneity, we also depend on the further equality that for evolution along a single edge of the tree of length t, the probability of ending in state 1 given that the edge began in state 0 can be written as follows.

$$P(1|0,t) = \frac{1}{m} \sum_{i=1}^{m} P(1|0,t,r = r_i) \tag{4}$$

That is to say, the probability of starting in condition 0 and then after some time t being in state 1, given that the rate of evolution over time t is a discretized Γ-distributed random variable with conditions $r = r_1, r_2, \ldots, r_m$ each with probability $\frac{1}{m}$, is just the *average* probability across each of our *m* rate categories. (We can't prove this here, but if the reader doesn't believe us, we ask them consider a room illuminated by a single, incandescent bulb. We know of a special class of genie, a 'light-flicker,' that, upon entering a room, will immediately beginning to flick the light switch of this bulb randomly off, and on, and off again. Light-flicker genies exist in two types, the 'hare' & the 'tortoise,' both about equally common in nature. One type, the hare, will flick the light switch off and on with near-infinite quickness;

while the other, the tortoise, turns the room light randomly off and on again with a rhythm of painful slowness, perhaps needing thousands or millions of years, on average, to flick a light from off to on or *vice versa*. Imagining, once again, that a room starts in the lightened condition, having seen a genie enter then leave ahead of us, but not knowing which type we saw, the probability that the room is darkened, should we dare to follow the genie in after some modest, but non-trivial, elapsed time, will be closer to 0.25 than 0 or 1... in other words, the *average* of 50:50 and 0:100 – the odds that the room would've been darkened by our hare and tortoise light-flicker genies, respectively. As a corollary to this parable, it's worth noting that if we entered the room and found it darkened, we could be near certain that a hare light-flicker had preceded us: showing how the static property of the room contains information about which genie it once held!)

More generally, the *matrix* of transition probabilities between states after time *t* can be expressed as ~~the average of~~ the exponentiated product $\mathbf{Q}r_i t$, in which $\exp(\mathbf{X})$ is the matrix exponential of $\mathbf{X}$.

$$\mathbf{P}_t = \frac{1}{m} \sum_{i=1}^{m} \exp(\mathbf{Q} \times r_i \times t) \tag{5}$$

As we can compute the probability of any data pattern on the tree for a given shape parameter (α), number of rate categories (*m*), and transition matrix $\mathbf{Q}$, we can likewise find the set of such values that maximize this probability – in other words, the Maximum Likelihood (ML) estimate. That's what we will do here. As a side note to this observation, we'll also take this opportunity to observe that it would be ~~a relatively a straightforward thing~~ to fit the model of this article using Bayesian MCMC instead of ML. Indeed, the *phytools* function in which this model has been implemented (fitgammaMk), also exports the likelihood function of the model that *phytools* users could easily pass along their favorite MCMC algorithm. (This would, of course, require the investigator to supply appropriate prior probability distributions for α and $\mathbf{Q}$, a task ~~which~~ has been contemplated in other literature.)

In addition to simply identifying the value of the Γ distribution α shape parameter that maximizes the probability of our observed data, our new model can be used in at least two additional ways. Firstly, it can be employed to calculate (conditioning on a transition model and ML value for α) the marginal likelihoods that each edge is in each of the *m* rates of our fitted model. This is done by simply traversing the tree, edge by edge, holding the rate on a given edge constant to each of the *m* rate categories, and then computing the marginal probability of the data (integrating over all *m* rates across all of the *other* edges of the tree) for each rate. Since these marginal likelihoods can also be interpreted as the posterior probabilities that each edge is in each of our *m* rate conditions, we can in turn proceed to use them as weights and calculate marginal edge rates (as the weighted mean of our rate categories) for each branch of the phylogeny. Secondly, we can use the model to perform marginal or joint ancestral state reconstruction at the nodes of our phylogeny. This is undertaken in exactly the usual way (see Yang 2014; Revell and Harmon 2022), but in which we integrate over each of the *m* rate categories as we traverse the tree. We'll illustrate both of these applications below.

2.2 Estimation of α

Now that we've expressed, in principle, how one might go about computing the probability of a set of data for any particular value of the (discretized) Γ distribution shape parameter, α, and state transition matrix, $\mathbf{Q}$, we used numerical simulations to establish whether or not we can, in fact, estimate α in practice. Note that these initial simulations were not designed to explore the range of conditions (in taxon number, transition models, character state levels, or values of α) that might be encountered in an empirical study. To the contrary, we decided to use a small number of 'maximally-powered' simulations that we anticipated might give us the *best* chance of reliably estimating rate heterogeneity. We broaden our simulation conditions in the sections below (focused on statistical power and marginal ancestral state estimation), and also hope to see a wider variety of numerical simulations required to understand the study of rate heterogeneity from a single character in future articles by us or others.

Thus, to (as much as reasonably possible) *maximize* our power to estimate α, we simulated a total of twenty 2,000 species, pure-birth (Yule) phylogenies, rescaling each tree to have a total depth of 1.0 time unit. We next sampled the edge rates of each tree from a Γ distribution with a total of *six* different values of the Γ shape parameter, α: 0.25, 0.5, 1.0, 2.0, 4.0, and 8.0. (Note that although our model for estimation is a discretized Γ distribution, we intentionally elected to use a continuous Γ distribution for

A

B
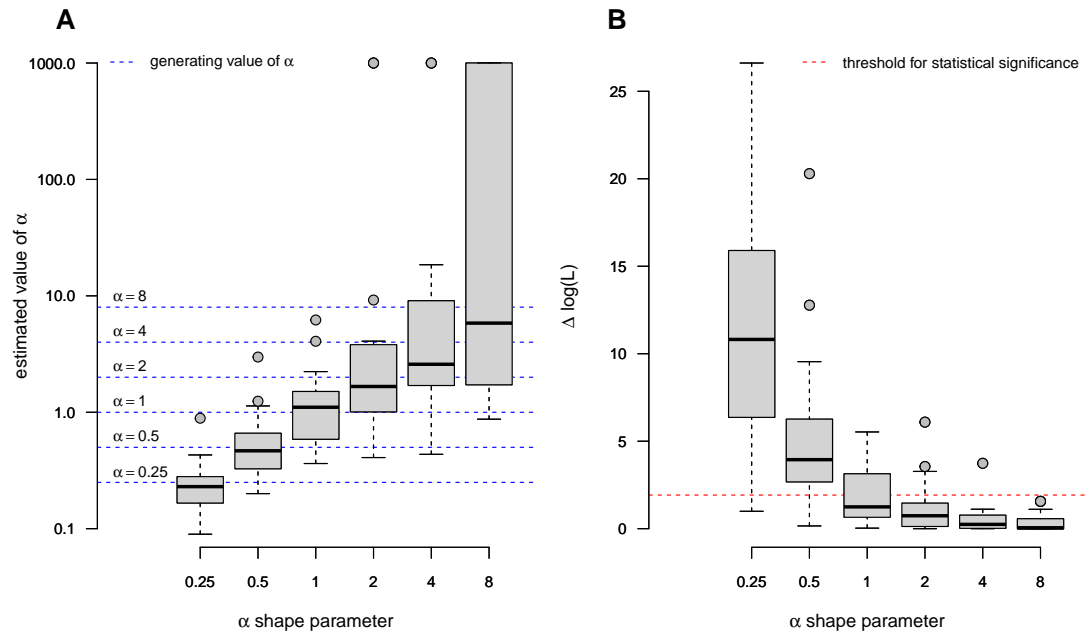
Figure 3. Results from an analysis of simulated datasets generated with Γ-distributed rate heterogeneity under various values of α. (**A**) The distribution of estimated values of α for each generating value. (**B**) The difference in log-likelihood between the best-fitting Γ model and the standard Mk rate-homogeneous model. In both panels, the grey box gives the 25% to 75% inter-quartile range, the horizontal black line the median, and the whiskers the range, excluding outliers (shown with grey dots). The maximum upper-bound for α during optimization was set to 1,000.

179 simulation both here and in subsequent sections of this article. We're thus also testing how reasonably-
180 well our discretization of Γ approximates underlying continuous rate variation among edges.) These
181 simulated values of α varied from very high rate heterogeneity among edges ($\alpha \leq 1.0$) to relatively low
182 rate heterogeneity ($\alpha \geq 4.0$). Having sampled edge rates, we next simulated a single, ordered, three-state
183 character (with trait levels *a*, *b*, and *c*, and a single rate of transition between adjacent character levels) on
184 each tree. We then proceeded to fit our variable-rate model using ML, specifying the known *structure*
185 of the transition process, but simultaneously estimating the instantaneous transition rate (q) and Γ shape
186 parameter (α), arbitrarily fixing the number of rate categories (m) to $m = 8$.

187   In addition to fitting our Γ rate heterogeneous model, for each tree and dataset we *also* fit a rate
188 homogenous (i.e., 'standard') Mk model, likewise fixing the model structure to its generating value. We
189 then measured the difference in log-likelihood between the generating rate heterogeneous (Γ) model and
190 our fitted rate homogeneous (Mk) model. For each phylogeny and set of simulated data, we undertook
191 a total of ten optimization iterations of both the Γ and standard Mk models, each with random starting
192 values for α (if applicable) and the transition rate between states (q), to help ensure convergence to the
193 ML solution in each case. We invariably assumed a prior distribution at the root (π) following FitzJohn et
194 al. (2009).

195   Figure 3A shows the distribution of ML estimated values of α for simulated level of rate heterogeneity.
196 In general, for relatively low values of α (that is, high rate heterogeneity among edges: $\alpha \leq 1.0$), the
197 mean and median estimated values of α were quite close to the generating quantities: highlighted by the
198 horizontal dashed blue lines (Figure 3A). For increasing generating values of α (that is, *decreasing* rate
199 heterogeneity among edges), estimated α tended to be upwardly biased. This was especially marked
200 for $\alpha = 8.0$, in which many optimization iterations hit the pre-specified upper-bound of optimization
201 ($\alpha = 1,000$; Figure 3A).

202   Measurement of the difference in model fit between the Γ rate heterogeneous model and the Mk model
203 showed a largely consistent trend (Figure 3B). In particular, low α (that is, *high* rate heterogeneity among
204 edges) tended to result in much higher model fit of the variable-rate Γ model compared to the standard Mk
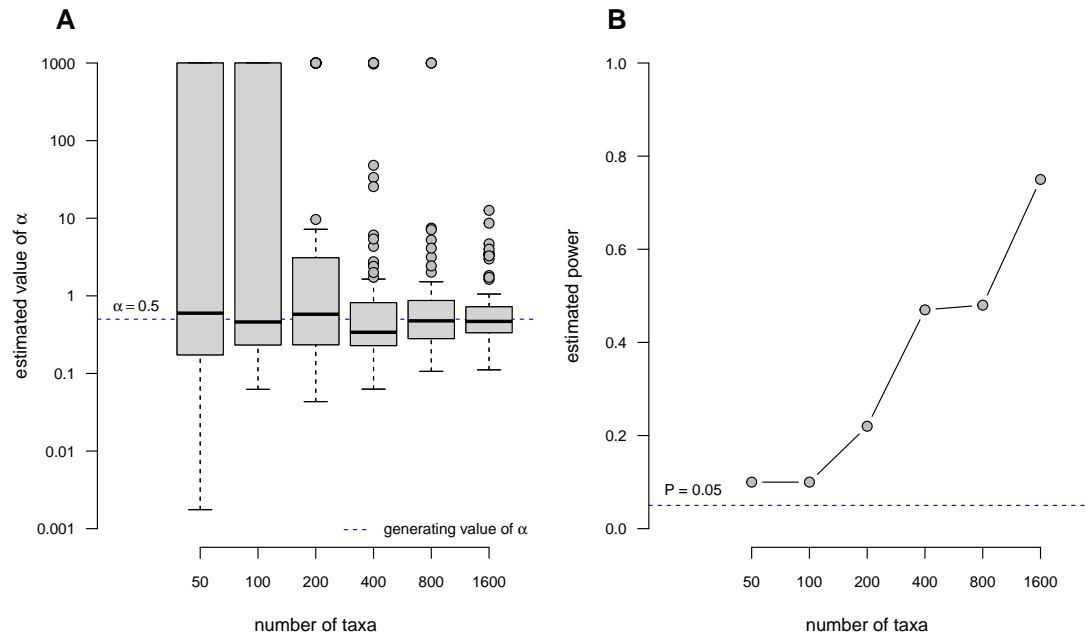
Figure 4. Results from an analysis of simulated datasets generated with Γ-distributed rate heterogeneity with $\alpha = 0.5$ for phylogenies containing various numbers of taxa. (**A**) The distribution of estimated values of α for each tree size. As in Figure 3, the grey box gives the 25% to 75% inter-quartile range, the horizontal black line the median, and the whiskers the range, excluding outliers.(**B**) Estimated power: the proportion of statistical test in which Γ rate heterogeneity was significant in a likelihood-ratio test with one degree of freedom, the difference in parameter complexity between the Mk and Γ models.

model, whereas this advantage declined for increasing α (decreasing rate heterogeneity among edges). Notably, even on a tree with 2,000 tips, none of 20 simulation replicates met Δlog-likelihood criterion for statistical significance at the 0.05 level when generating α was equal to $\alpha = 8.0$: the lowest level of rate heterogeneity simulated (Figure 3B).

2.3 Power analysis

To measure power in a statistical test for significant evidence of rate heterogeneity, we conducted the following simulation. First, we simulated 100 stochastic phylogenies under a Yule (pure-birth) model containing each of 50, 100, 200, 400, 800, and 1,600 species (600 simulations in total), and then rescaled each tree to have a total root to tip depth of 1.0. Next, on each tree we simulated a single discrete character with Γ-distributed rate heterogeneity among edges using an α shape parameter of $\alpha = 0.5$, along with the same three-state **Q** matrix as in our prior analysis.

For each simulated tree and data vector, we then proceeded to fit our variable-rate model using ML, specifying the known structure of the model, as before, but simultaneously estimating the instantaneous transition rate (q) and shape parameter of Γ, while setting the number of rate categories (m) to $m = 8$. In addition to fitting our Γ rate heterogeneous model, for each tree and dataset we also fit a rate homogenous Mk model, likewise fixing the model structure to its generating value. We then computed a P-value based on a likelihood-ratio test with one degree of freedom: the difference in the number of estimated parameters between our two models. As before, to help ensure convergence to the ML solution we invariably undertook a total of ten optimization iterations of both the Γ and standard Mk models, each with random starting values for α (if applicable) and the transition rate between states (q). Finally, we assumed a prior distribution at the root (π) following FitzJohn et al. (2009).

The results from our power analysis are shown in Figure 4. We found that the median estimated value of α tended to be centered relatively close to the generating value of $\alpha = 0.5$ across all simulation conditions (Figure 4A). The variance among simulations, however, was *very* high for relatively small trees: such as $N = 50$ and $N = 100$ (Figure 4A). Commensurate with this observation, we found that power

to reject the null hypothesis of rate homogeneity tended to be barely above the nominal α (acceptable type I error) level of 0.05 for small trees, and only increased to around 80% for simulated phylogenies containing as many as 1,600 tip taxa (Figure 4B).

2.4 Ancestral state reconstruction

Lastly, we measured the accuracy of ancestral state estimation under the Γ model of rate heterogeneity among edges. To do this, we first simulated two hundred 500 taxon pure-birth (Yule) phylogenetic trees. In all previous simulations (as noted in prior sections), we scaled our simulated trees to have a total root to tip depth of 1.0. Here, however, we decided to scale to a depth of 10.0 because preliminary simulations revealed very high average accuracy of ancestral state estimation when trees were scaled to a depth of 1.0 for both our fitted Mk and Γ models. (Greater tree depth tends to reduce the accuracy of ancestral state estimation of discrete traits for a given transition matrix, $\mathbf{Q}$, because more character changes will accumulate over longer evolutionary time.) As for prior sections, on each tree we simulated the evolution of a single, three-state discrete character evolving under the same stochastic process (with Γ rate heterogeneity $\alpha = 0.5$) as in our power analysis of the prior section, but in which we also recorded the simulated ancestral states for all internal nodes of the phylogeny.

Next, for each simulated tree and data vector we proceeded to fit our variable-rate model using ML, specifying the known structure of the model (as in our other simulation analyses), but simultaneously estimating the instantaneous transition rate (q) and α shape parameter of Γ. As before, we fixed the number of rate categories of our discrete Γ distribution (m) to be $m = 8$. In addition to fitting our Γ rate heterogeneous model, for each tree and dataset we also fit a rate homogenous Mk model, likewise fixing the model structure to its generating value. As before, we undertook a total of ten optimization iterations of both the Γ and standard Mk models, each with random starting values for α (if applicable) and the transition rate between states (q), to help ensure convergence. As before, we assumed a FitzJohn et al. (2009) root node prior distribution.

For each fitted Mk and Γ model of each simulated tree and dataset, we undertook marginal ancestral state estimation following Pagel (1999; Yang 2006). Marginal ancestral state estimation involves traversing the tree by node, and at each node computing the probability of the data given each state of the character, conditioning on the tree and fitted character model, but integrating across all possible states at all other nodes of the tree. These node likelihoods are then normalized (scaled) to sum to 1.0, at which point they can be properly interpreted as the (empirical Bayes) posterior probabilities that each node is in each of the observed conditions of our character (Yang 2006, 2014). To measure the accuracy of our ancestral state reconstructions under both the standard Mk and Γ models, we simply computed the average, node-wise product of the known, simulated states (represented as vectors of 0s and 1s) and the marginal scaled likelihoods from our procedure of ancestral state estimation. This average product can range from 0 (if all nodes are estimated confidently but completely inaccurately) to 1 (reflecting high confidence in the correct ancestral conditions across all nodes). Intermediate measured accuracy comprises both the scenario of an intermediate portion of nodes inferred correctly with high confidence (and the remainder confidently wrongly), or equally low confidence across all nodes. Since here we're focused on identifying any difference in accuracy between Mk and Γ models that might arise for data generated under a Γ model of rate heterogeneity, we're not too preoccupied in differentiating these latter possibilities. Lastly, we computed the difference in raw accuracy and divided it by the accuracy of the Mk estimates to measure the proportional improvement in accuracy of the Γ compared to Mk model when the former was true. (A positive value of 0.05, for example, would indicate that measured accuracy was about 5% better for the Γ model than the Mk model, whereas a value of -0.05 would suggest the converse.)

Figure 5 shows the distribution of scaled differences in accuracy between the Γ and Mk models in marginal ancestral state estimation from our 200 simulations. Invariably, estimates obtained using the Γ model, which was the generating evolutionary scenario of the simulations, were more accurate than those obtained from the Mk model (Figure 5), with a mean improvement in accuracy of about 5.3%.

2.5 Empirical case study

To explore the behavior of our method using empirical data we elected to analyze a dataset of forefoot & hindfoot digit number in squamate reptiles originally published by Brandley et al. (2008). These data consist of a single estimated tree and observations of 0, 1, 2, ..., 5 digits in the fore and hindfeet of 257 species of squamates, plus the tuatara (*Sphenodon punctatus*). A very small number of species were reported in Brandley et al. (2008) as exhibiting polymorphism in digit number. Since the authors scored
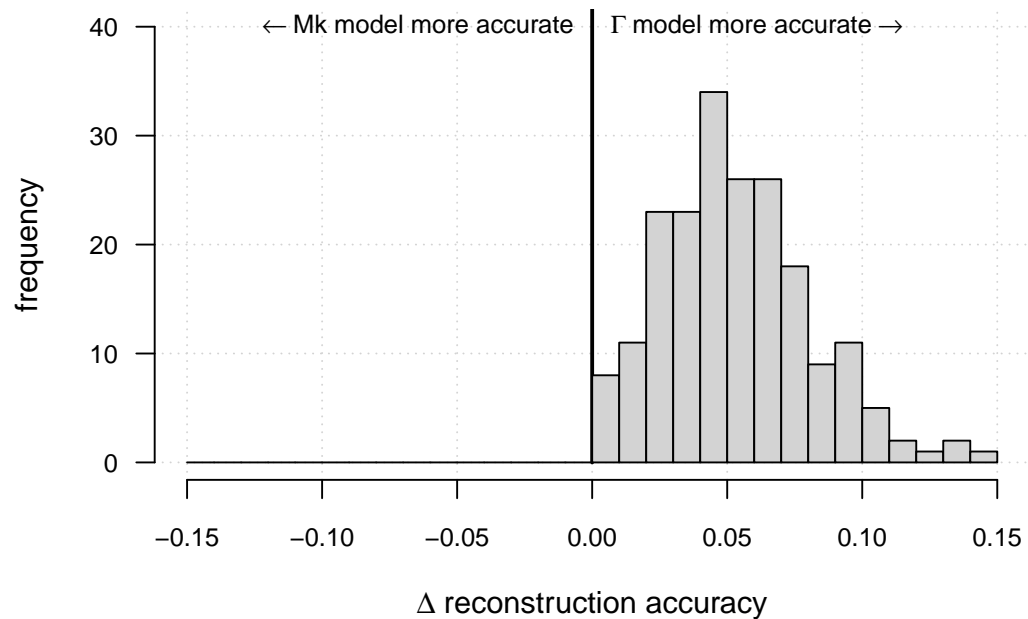
Figure 5. Normalized difference in accuracy of marginal ancestral state estimation under the M*k* and Γ model when the data were simulated under the Γ model. Raw accuracy was invariably higher for the generating model. See main text for more details.

these taxa with a digit number tally equal to the *average* number of digits (and because they consist of relatively few species of the dataset), we simply rounded polymorphic taxa to the closest whole digit.

In our recently published book (Revell and Harmon 2022) we analyzed a subset of the Brandley et al. (2008) data for hindfoot digit number only and discovered that directional (sequential loss) and ordered (sequential loss and gain) M*k* models had much better support (accounting for model parameter complexity) compared to other transition models between states. As such, here we've elected to focus on only this category of model.

For each of the fore and hindfoot digit number traits, we fit a total of eight models: four M*k* models and four models with Γ-distributed rate heterogeneity among edges. These models were as follows (in each case, the number of parameters to be estimated is indicated in parentheses): a directional (loss-only) M*k* model with one rate of digit loss (1); a directional Γ model with one rate of digit loss (2); an ordered (loss-and-gain) M*k* model with one rate of loss and a second rate of gain (2); an ordered Γ model with one rate of loss and another rate of gain (3); a directional (loss-only) M*k* model with a different rate of loss for each character level (5); a directional Γ model with a different rate of loss for each level (6); an ordered (loss-and-gain) M*k* model with different rates of loss and gain for each character level (10); finally, an ordered Γ model with different rates of loss and gain for each level (11). As in our simulations, above, we used a FitzJohn et al. (2009) root prior, ran ten optimization iterations per model with random starting values, and set the number of rate categories of the discretized Γ distribution to $m = 8$. For the best-fitting Γ model for each trait, we also computed the marginal edge rates and scaled likelihoods (marginal ancestral states) at all nodes of the tree.

Table 1 shows a summary of the model fits, estimated value of α (if applicable), model parameter complexity, and model support for each of the eight trait evolution models we fit to forefoot digit number. In comparing between each *pair* of otherwise equivalent models with and without Γ rate heterogeneity among edges, information theoretic model support (measured via AIC) was consistently superior for the rate heterogeneous model ($12.0 \leq \Delta AIC \leq 379.9$ across all model pairs; Table 1). ML estimated values of α across all models were invariably low ($0.047 \leq \alpha \leq 0.104$), indicating relatively *high* rate heterogeneity among edges in the best-fitting Γ models. The best-supported model overall was unambiguously the

Table 1. Log-likelihood, estimated value of α (if applicable), number of parameters, AIC, and model weights for five different discrete character evolution models of forefoot digit number in squamate reptiles, with and without Γ rate heterogeneity among edges. See main text for more details.

	log(L)	estimated α	d.f.	AIC	weight
directional one-rate Mk	-386.5609	–	1	775.1218	0.00000
directional one-rate Γ	-195.6262	0.06039	2	395.2524	0.00001
ordered two-rate Mk	-338.1719	–	2	680.3438	0.00000
ordered two-rate Γ	-192.0290	0.04669	3	390.0580	0.00013
directional five-rate Mk	-197.5535	–	5	405.1069	0.00000
directional five-rate Γ	-190.1696	0.08672	6	392.3392	0.00004
ordered ten-rate Mk	-182.0694	–	10	384.1389	0.00251
ordered ten-rate Γ	-175.0838	0.10363	11	372.1675	0.99731

Table 2. Log-likelihood, estimated value of α (if applicable), number of parameters, AIC, and model weights for five different discrete character evolution models of hindfoot digit number in squamate reptiles, with and without Γ rate heterogeneity among edges. See main text for more details.

	log(L)	estimated α	d.f.	AIC	weight
directional one-rate Mk	-357.9915	–	1	717.9829	0.00000
directional one-rate Γ	-210.3328	0.04761	2	424.6656	0.00323
ordered two-rate Mk	-324.5030	–	2	653.0060	0.00000
ordered two-rate Γ	-209.5280	0.03284	3	425.0559	0.00266
directional five-rate Mk	-211.7965	–	5	433.5930	0.00004
directional five-rate Γ	-202.4479	0.26559	6	416.8958	0.15709
ordered ten-rate Mk	-203.8508	–	10	427.7016	0.00071
ordered ten-rate Γ	-195.7757	0.25896	11	413.5515	0.83628

ordered, multi-rate, Γ rate heterogeneous model (Table 1). Consequently, we used this model to estimate marginal edge rates of the character and scaled likelihoods (ancestral states) at all internal nodes of the phylogeny. This reconstruction is illustrated in Figure 6A.

Table 2 shows a summary of the model fits, parameter complexity, and model support for each of the eight trait evolution models we fit to hindfoot digit number. Much as we found for forefoot digit number, comparing between each pair of models with and without Γ rate heterogeneity among edges, model support was consistently higher for the Γ model ($14.2 \leq \Delta AIC \leq 293.3$ across all model pairs; Table 2). As for forefoot digit number, rate heterogeneity in the best-fitting Γ models was substantial (all $0.033 \leq \alpha \leq 0.266$). The best-supported model overall was the ordered, multi-rate, Γ rate heterogeneous model, with an Akaike weight of around 0.84. The second best-supported model was the directional, multi-rate, Γ model (Table 2). We used the best-supported ordered model to estimate marginal edge rates of the character and scaled likelihoods (ancestral states) at all internal nodes of the tree. This reconstruction is illustrated in Figure 6B.

3 SOFTWARE

The model of this article is implemented in the function `fitgammaMk` of the *phytools* (Revell 2012; Revell 2023) R package. All analyses of this study were undertaken in R (R Core Team 2023) using *phytools*. *phytools*, in turn, depends heavily on the object classes and methods of the core R phylogenetics package, *ape* (Paradis et al. 2004; Popescu et al. 2012; Paradis and Schliep 2019). At the time of writing, *phytools* also depended on or imported from a number of other R packages including *clusterGeneration* (Qiu and Joe. 2020), *coda* (Plummer et al. 2006), *combinat* (Chasalow 2012), *doParallel* (Microsoft and Weston 2022a), *expm* (Maechler et al. 2023), *foreach* (Microsoft and Weston 2022b), *maps* (Becker et al. 2022), *MASS* (Venables and Ripley 2002), *mnormt* (Azzalini and Genz 2022), *nlme* (Pinheiro and Bates 2000; Pinheiro et al. 2022), *numDeriv* (Gilbert and Varadhan 2019), *optimParallel* (Gerber and Furrer 2019), *phangorn* (Schliep 2011), *plotrix* (Lemon 2006), and *scatterplot3d* (Ligges and Mächler 2003). The *geiger* (Harmon et al. 2008; Pennell et al. 2014) and *viridisLite* (Garnier et al. 2023) packages were
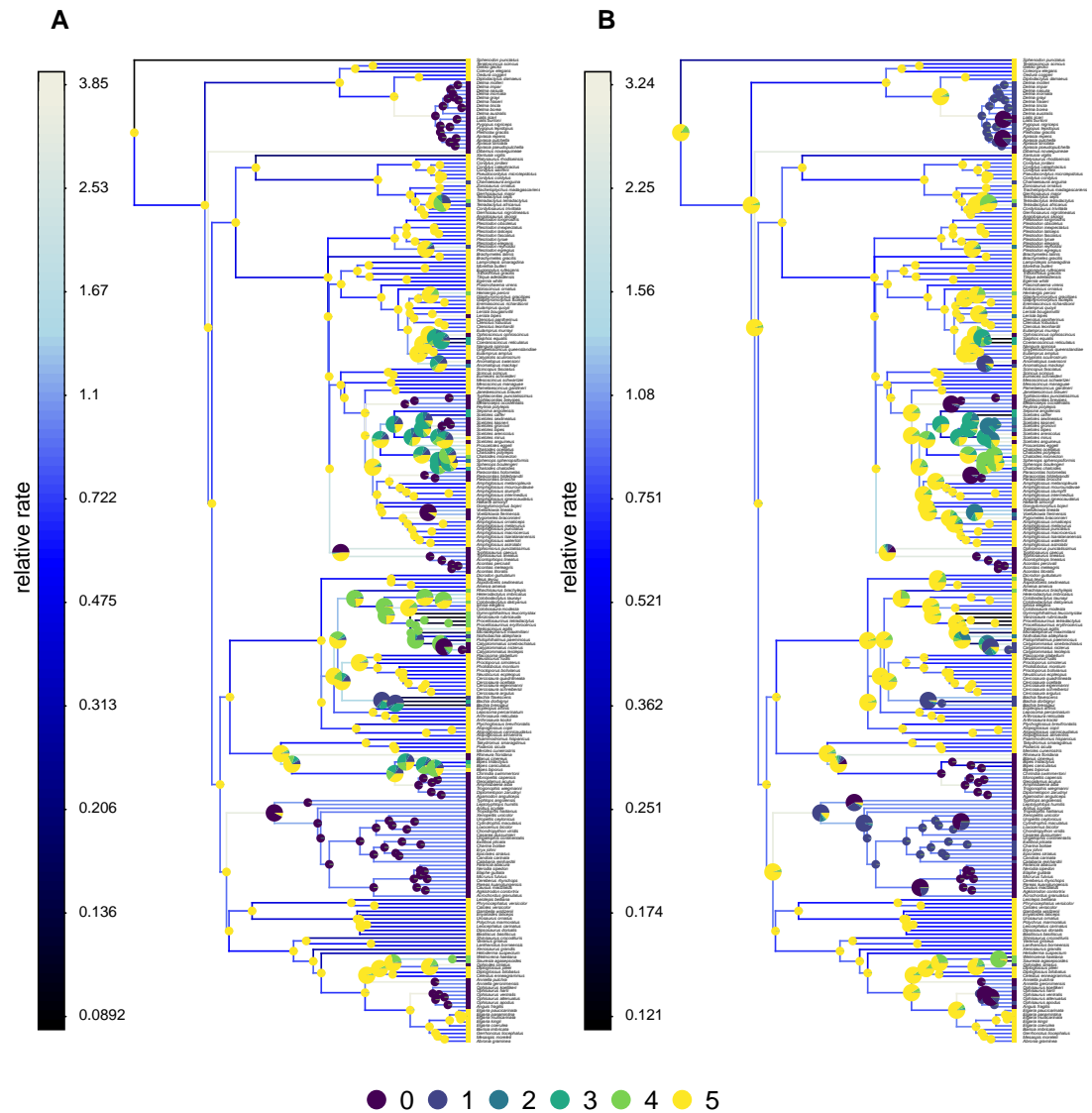
Figure 6. Phylogeny of squamate reptiles from Brandley et al. (2008) showing the number of digits in the forefoot (**A**) or hindfoot (**B**) of each species, estimated ancestral states under an ordered Γ model, and estimated edge rates. Larger pie charts were used for all nodes in which no state had a marginal scaled likelihood > 0.95. See main text for more details.

336 also used for some analyses and visualizations of this study.

4 DISCUSSION

338 Over the past several decades, phylogenies have become pervasive in many disciplines of the biological
339 sciences and beyond. For example, phylogenies are now employed extensively in biomedical research,
340 in infectious disease epidemiology, in genomics, and in anthropology, among other fields. Increasingly,
341 phylogeneticists combine the estimated tree (typically obtained from aligned nucleotide or amino acid
342 sequences) with phenotypic trait data in an effort to learn something about how the features and attributes
343 of living organisms have evolved over time. For trait data that is discretely-valued, such as the presence
344 or absence of a physical feature or a categorical behavioral trait, the typical model used to approximate
345 character evolution a phylogeny is one that's been called the Mk model (Lewis 2001). The Mk model
346 describes a continuous-time Markov chain with k states, and we normally assume that this process is
347 constant over all the branches and nodes of our reconstructed tree.

348   In the present article we introduce a totally new model in which we assume that the edge-wise rate
349 of trait evolution is distributed as a random variable drawn from a Γ distribution. The Γ distribution is a
350 flexible two parameter distribution which has a general form determined by the parameter α. When the
351 second parameter of the Γ distribution, β, is set to $\beta = \alpha$, then the mean of the distribution is 1.0 and the
352 variance decreases monotonically with increasing α. Indeed, as $\alpha \to \infty$ (but in practice for $\alpha << \infty$), our
353 Γ model converges on a homogeneous rate Mk. Conversely, as $\alpha \to \frac{1}{\infty}$ (but in practice for $\alpha >> \frac{1}{\infty}$) the Γ
354 model of this paper converges on a two-rate process in which some branches evolve with a very high rate,
355 and most not at all.

356   When the generating process is Γ-distributed rates among edges, we found that it was possible to
357 estimate α reasonably well – particularly if α was low and especially for larger trees (Figures 3, 4). On
358 the other hand, we found that the variance in estimated α increased substantially for larger values of α
359 (i.e., low rate heterogeneity) and smaller phylogenetic trees (Figures 3, 4). We also found that the method
360 had modest power to reject the null hypothesis of constant rates. In particular, even for relatively high rate
361 heterogeneity ($\alpha = 0.5$) power was only around 50% for $N = 800$. We also found that taking into account
362 rate heterogeneity when present had a highly significant impact on ancestral state estimation (Figure 5).

363   Lastly, we analyzed a dataset of fore and rearfoor digit number in 257 species of squamate reptiles
364 (plus the tuatara, *Sphenodon punctatus*) using a phylogeny and dataset from Brandley et al. (2008). We
365 found significant evidence for rate heterogeneity in digit number evolution, with all Γ models fitting
366 substantively better than their rate homogeneous counterparts.

367   In conclusion, we believe that the analyses we illustrate in this article can be an important addition
368 to the growing toolbox of methodologies for studying character evolution on trees. We hope that its
369 statistical properties, use cases, and limitations can be examined further in future research.

5 ACKNOWLEDGMENTS

6 FUNDING

LITERATURE CITED

377 Azzalini, A., and A. Genz. 2022. The R package `mnormt`: The multivariate normal and *t* distributions
378   (version 2.1.1).
379 Beaulieu, J. M., B. C. O'Meara, and M. J. Donoghue. 2013. Identifying hidden rate changes in the
380   evolution of a binary morphological character: The evolution of plant habit in campanulid angiosperms.
381   Syst. Biol. 62:725–737.
382 Becker, R. A., A. R. Wilks, R. Brownrigg, T. P. Minka, and A. Deckmyn. 2022. maps: Draw geographical
383   maps.

Brandley, M. C., J. P. Huelsenbeck, and J. J. Wiens. 2008. RATES AND PATTERNS IN THE EVOLU-
        TION OF SNAKE-LIKE BODY FORM IN SQUAMATE REPTILES: EVIDENCE FOR REPEATED
        RE-EVOLUTION OF LOST DIGITS AND LONG-TERM PERSISTENCE OF INTERMEDIATE
        BODY FORMS. Evolution 62:2042–2064. Wiley.

Chasalow, S. 2012. combinat: Combinatorics utilities.

Felsenstein, J. 1981. Evolutionary trees from DNA sequences: A maximum likelihood approach. Journal
        of Molecular Evolution 17:368–376. Springer Science; Business Media LLC.

Felsenstein, J. 2004. Inferring Phylogenies. Sinauer Associates, Sunderland, MA.

FitzJohn, R. G., W. P. Maddison, and S. P. Otto. 2009. Estimating trait-dependent speciation and extinction
        rates from incompletely resolved phylogenies. Systematic Biology 58:595–611.

Garnier, Simon, Ross, Noam, Rudis, Robert, Camargo, A. Pedro, Sciaini, Marco, Scherer, and Cédric.
        2023. viridis(Lite) - colorblind-friendly color maps for r.

Gerber, F., and R. Furrer. 2019. optimParallel: An r package providing a parallel version of the L-BFGS-B
        optimization method. The R Journal 11:352–358.

Gilbert, P., and R. Varadhan. 2019. numDeriv: Accurate numerical derivatives.

Harmon, L. J. 2019. Phylogenetic comparative methods: Learning from trees. Ecoevorxiv.

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

Lemon, J. 2006. Plotrix: A package in the red light district of r. R-News 6:8–12.

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

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

Maechler, M., C. Dutang, and V. Goulet. 2023. Expm: Matrix exponential, log, 'etc'.

Marazzi, B., C. Ané, M. F. Simon, A. Delgado-Salinas, M. Luckow, and M. J. Sanderson. 2012. Locating
        evolutionary precursors on a phylogenetic tree. Evolution 66:3918–3930.

Microsoft, and S. Weston. 2022a. doParallel: Foreach parallel adaptor for the 'parallel' package.

Microsoft, and S. Weston. 2022b. Foreach: Provides foreach looping construct.

Pagel, M. 1994. Detecting correlated evolution on phylogenies: A general method for the comparative
        analysis of discrete characters. Proc. R. Soc. Lond. B Biol. Sci. 255:37–45. The Royal Society.

Pagel, M. 1999. The maximum likelihood approach to reconstructing ancestral character states of discrete
        characters on phylogenies. Systematic Biology 48:612–622. Oxford University Press (OUP).

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

Paradis, E., and K. Schliep. 2019. Ape 5.0: An environment for modern phylogenetics and evolutionary
        analyses in R. Bioinformatics 35:526–528.

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

Pinheiro, J. C., and D. M. Bates. 2000. Mixed-effects models in s and s-PLUS. Springer, New York.

Pinheiro, J., D. Bates, and R Core Team. 2022. Nlme: Linear and nonlinear mixed effects models.

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

Popescu, A.-A., K. T. Huber, and E. Paradis. 2012. Ape 3.0: New tools for distance-based phylogenetics
        and evolutionary analysis in R. Bioinformatics 28:1536–1537.

Qiu, W., and H. Joe. 2020. clusterGeneration: Random cluster generation (with specified degree of
        separation).

R Core Team. 2023. R: A language and environment for statistical computing. R Foundation for Statistical
        Computing, Vienna, Austria.

Revell, L. J. 2023. Phytools 2.0: An updated r ecosystem for phylogenetic comparative methods (and
        other things), doi: 10.1101/2023.03.08.531791. Cold Spring Harbor Laboratory.

Revell, L. J. 2012. Phytools: An R package for phylogenetic comparative biology (and other things).
        Methods Ecol. Evol. 3:217–223.

Revell, L. J., and L. J. Harmon. 2022. Phylogenetic comparative methods in R. Princeton University
        Press, Princeton, NJ.

439   Schliep, K. P. 2011. Phangorn: Phylogenetic analysis in R. Bioinformatics 27:592–593.
440   Venables, W. N., and B. D. Ripley. 2002. Modern applied statistics with s. Fourth. Springer, New York.
441   Yang, Z. 2006. Computational molecular evolution. Oxford University Press, London, England.
442   Yang, Z. 1994. Maximum likelihood phylogenetic estimation from DNA sequences with variable rates
443   over sites: Approximate methods. Journal of Molecular Evolution 39:306–314. Springer Science;
444   Business Media LLC.
445   Yang, Z. 2014. Molecular evolution. Oxford University Press, London, England.