

Testing for heterogeneous rates of discrete character evolution on phylogenies

Liam J. Revell^{1,2}, Klaus P. Schliep³, D. Luke Mahler⁴, Travis Ingram⁵

¹Department of Biology, University of Massachusetts Boston, Boston, MA, United States

²Facultad de Ciencias, Universidad Católica de la Santísima Concepción, Concepción, Chile

³Institute of Molecular Biotechnology, Graz University of Technology, Graz, Austria

⁴Department of Ecology and Evolutionary Biology, University of Toronto, Toronto, Ontario, Canada

⁵Department of Zoology, University of Otago, Dunedin, New Zealand

Handling Editor: Luke Holman, Associate Editor: Masahito Tsuboi

Corresponding author: Liam J. Revell, Department of Biology, University of Massachusetts Boston, Boston, MA, United States. Email: liam.revell@umb.edu

Abstract

Many hypotheses in the field of phylogenetic comparative biology involve specific changes in the rate or process of trait evolution. This is particularly true of approaches designed to connect macroevolutionary pattern to microevolutionary process. We present a method to test whether the rate of evolution of a discrete character has changed in one or more clades, lineages, or time periods. This method differs from other related approaches (such as the “covarion” model) in that the “regimes” in which the rate or process is postulated to have changed are specified a priori by the user, rather than inferred from the data. Similarly, it differs from methods designed to model a correlation between two binary traits in that the regimes mapped onto the tree are fixed. We apply our method to investigate the rate of dewlap colour and/or caudal vertebra number evolution in Caribbean and mainland clades of the diverse lizard genus *Anolis*. We find little evidence to support any difference in the evolutionary process between mainland and island evolution for either character. We also examine the statistical properties of the method more generally and show that it has acceptable type I error, parameter estimation, and power. Finally, we discuss some general issues of frequentist hypothesis testing and model adequacy, as well as the relationship of our method to existing models of heterogeneity in the rate of discrete character evolution on phylogenies.

Keywords: *Anolis*, comparative methods, maximum likelihood, phenotypic evolution

Introduction

Over recent decades, phylogenetic comparative methods have grown steadily in their importance and now assume a relatively central role in evolutionary research (Harmon, 2018; Harvey & Pagel 1991; Revell & Harmon, 2022). The majority of phylogenetic comparative methods take a phylogenetic tree and phenotypic trait data for the constituent species of that tree with the aim of combining the two to better understand trait evolution over the period of history represented by the phylogeny (Nunn, 2011; Revell & Harmon, 2022). These inferences might include, for instance, that two or more phenotypic attributes of a clade tend to evolve in a correlated fashion, or that the phenotypic characteristics of an ancestral species were most likely to have been more similar to one of its extant descendants than to others (Revell & Harmon, 2022). In the former case, we might use the phylogenetically independent contrasts algorithm of Felsenstein (1985) or Grafen’s (1987) phylogenetic regression; whereas in the latter we could employ maximum likelihood or Bayesian ancestral state estimation (e.g., Pagel et al., 2004; Revell, 2024; Revell & Harmon, 2022; Schluter et al., 1997).

Lately, there has been a substantial increase in the diversity of phylogenetic comparative methods that are designed to capture complexity and heterogeneity in the process of evolutionary change among lineages or through time (e.g., Beaulieu

et al., 2013; Boucher & Démary, 2016; Boucher et al., 2018; Butler & King, 2004; Felsenstein, 2005, 2012; O’Meara et al., 2006; Revell, 2014, 2024; Revell & Collar, 2009; Revell & Harmon, 2022). A subset of these new approaches were developed with an explicit underlying relationship to the microevolutionary process within species (e.g., Felsenstein, 2005; Revell & Harmon, 2008; Uyeda & Harmon, 2014). For example, Hohenlohe and Arnold (2008) described a new approach for testing hypotheses about lineage or species phenotypic diversification using the quantitative genetic additive variance-covariance matrix (the G-matrix; also see Revell & Harmon, 2008). Around the same time, Felsenstein (2005, 2012; also see Revell, 2014) presented a model for studying discrete character evolution, as well as the evolutionary correlation between discrete and continuous traits, that is based on Wright’s (1934) microevolutionary quantitative genetic threshold model. Most recently, Machado et al. (2023) developed a conceptual and methodological framework that combines comparative methods and microevolutionary process models to explicitly contrast hypotheses based on microevolution (in which the phenotypic divergence among species is predicted in terms of evolutionary genetics and developmental ontogeny), with competing non-microevolutionary scenarios.

Apart from this vital work to directly link micro- and macroevolution via comparative methods, other more

phenomenological approaches and models have also enabled a better understanding of underlying evolutionary processes from macroevolutionary data. For example, [Revell and Harmon \(2022\)](#) presented a polymorphic trait evolution model for discrete characters in which the polymorphic condition is intermediate between the two (or more) corresponding monomorphic states. Even though this model was not explicitly designed to capture the population genetic phenomenon of allelic polymorphism and fixation, it (in many ways) conceptually matches this microevolutionary process. Likewise, [Boucher and Démary \(2016\)](#) described a new model for quantitative trait evolution under stochastic diffusion (i.e., Brownian motion) with hard, reflective bounds. Although this model is not linked to an explicit underlying microevolutionary process, it may nonetheless better approximate the microevolutionary phenomenon of hard developmental constraints than do other existing “constraint” models, such as the Ornstein–Uhlenbeck model of [Hansen \(1997; Butler & King, 2004\)](#).

An important attribute of many of these innovations in phylogenetic comparative biology is that they permit heterogeneity in the evolutionary process from edge to edge or clade to clade of the tree ([Beaulieu et al., 2013; Butler & King, 2004; O’Meara et al., 2006; Revell & Collar, 2009](#)). Some of the simplest (where simplicity is defined in terms of the number of parameters to be estimated) of these heterogeneous process models in phylogenetic comparative biology, such as the ‘early-burst’ model (also known as the ACDC model; [Blomberg et al., 2003](#)), allow the evolutionary rate to change as a continuous function of time since the root of the phylogeny. Other more sophisticated approaches (e.g., [Rabosky, 2014](#)) permit the rate of evolution to vary continuously among lineages, but not necessarily as a function of elapsed time since the global root (also see [Martin et al., 2023; Revell, 2021; Revell & Harmon, in review](#)). Others still model heterogeneity in the evolutionary rate or process via discrete regime shifts, but in which the probability that any lineage is in any regime is either inferred explicitly from the data or integrated over during the analysis (e.g., [Beaulieu et al., 2013; Mahler et al., 2013; Marazzi et al., 2012; Revell et al., 2012; Uyeda & Harmon, 2014](#)).

Finally, an important class of approach involves the explicit a priori specification by the user of the different regimes that are hypothesized to evolve heterogeneously (e.g., [Butler & King, 2004; O’Meara et al., 2006; Revell & Collar, 2009; Revell et al., 2022](#)). In this case, the idea is fairly straightforward. The investigator first begins by fixing or “painting” a set of regimes on the tree based on a biological hypothesis for how evolution may have proceeded in their group of interest. These regimes could correspond to clades, time periods, individual branches of the phylogeny, or the postulated history of a discrete character mapped onto the nodes and branches of the phylogeny. One then proceeds to fit a model in which the process of evolution, or the parameter values that describe evolution via this process, is permitted to differ between regimes ([Revell, 2013; Revell & Harmon, 2022](#)). For a number of years, this type of approach has been quite popular among comparative biologists with numerous models and methods having been developed to investigate heterogeneity in the evolutionary process of continuously valued character traits across the branches or clades of the phylogeny (e.g., [Beaulieu et al., 2012; Butler & King, 2004; Caetano & Harmon, 2019; O’Meara et al., 2006; Revell & Collar, 2009; Revell et al., 2022; Thomas et al., 2006](#)).

Although discretely valued character traits are often studied in phylogenetic comparative research, to our surprise no precisely comparable set of methodologies had yet been developed for discrete characters. Perhaps the most similar approach is one denominated the codon “branch model” by [Yang \(1998; Yang & Nielsen, 1998, 2002\)](#). According to this model, different pre-specified edges of an independently estimated phylogeny are permitted to evolve with different nonsynonymous/synonymous (d_N/d_S) nucleotide substitution rate ratios. This scenario could then be compared to a simpler model in which d_N/d_S ratio does not vary among the edges and nodes of the phylogeny. To our knowledge, however, this model has never been extended beyond its original intent as a test for positive selection on gene sequences ([Yang, 1998](#)).

That being said, several related approaches have been proposed for discrete phenotypic traits and are in wide use. For instance, [Pagel \(1994\)](#) presented a method in which the rate of one binary trait is permitted to vary as a function of the state of a second (and, possibly, vice versa). Since this model will fit well when two different discrete characters have a disproportionate tendency to evolve towards certain trait combinations, the model is often interpreted as one that can be used to measure the correlated evolution of discrete characters (but see [Maddison & Fitzjohn, 2015; Revell & Harmon, 2022](#)). Later, [Tuffley and Steel \(1997\)](#) described a model for molecular evolution which they dubbed the “covarian model.” According to the covarian model, a discrete character evolves via a process with two hidden (that is, unobserved and unknown a priori) rate categories: either “on” (in which it can change state) or “off” (in which it cannot; [Penny et al., 2001; Tuffley & Steel, 1997](#)). This was extended by [Galtier \(2001\)](#) to permit any number of hidden-rate categories and subsequently generalized and applied to phylogenetic comparative analysis by [Beaulieu et al. \(2013; also see Marazzi et al., 2012\)](#). All of these latter methods suppose that the rate or process of one discrete trait varies among the edges and nodes of the phylogeny either as a function of the observed (e.g., [Pagel, 1994](#)) or unobserved state of a second discretely valued character (e.g., [Beaulieu et al., 2013; Marazzi et al., 2012; Penny et al., 2001](#)).

Herein, we imagine a scenario in which the process of evolution for a discrete trait varies among a series of a priori postulated regimes. These could be clades of the phylogeny, geological or historical time periods, specific edges in which an increase or decrease in rate is hypothesized, or the known (rather than reconstructed) history of a second discrete character trait. Although this is not a model with a direct mathematical connection to an underlying theoretical model of microevolution, in many cases, our hypothesized regimes might derive from microevolutionary theory. For instance, we might hypothesize a shift in the evolutionary process due to a change in genetic variance, the population size, or the origin of a new ecological or developmental constraint, among other factors.

Note that this approach should not be taken as an alternative to the aforementioned methods which either seek to identify heterogeneity in the evolutionary process for one character due to an unobserved factor or trait; or integrate over uncertainty in the history of an observed discrete character that affects the rate or evolutionary process of a second. Instead, the method of this study should be employed to investigate scenarios in which a particular set of regimes are hypothesized a priori (on biologically justifiable grounds),

at which point our model can be fit and compared to, for instance, a simpler scenario of homogeneous evolutionary rates through time, or perhaps to an alternative hypothesis for rate heterogeneity on the phylogeny (such as a hidden-rates model [HRM], as we will discuss below; Beaulieu et al., 2013).

After we describe our new approach to analyzing the evolution of a discretely valued character on the tree, we will proceed to present some relatively simple simulations examining its general statistical performance in terms of type I error, parameter estimation, and power. We will then use the method to test specific, a priori hypotheses about character evolution for two different discrete character traits in the neotropical lizard genus, *Anolis*. First, we will test the hypothesis that the rate or process of dominant dewlap colour evolution differs between island and continental *Anolis* lizards, premised on a theory that greater syntopy (mainland) or higher population densities (Caribbean) could affect the evolutionary tempo of this important mate recognition trait. Second, we will test the hypothesis that the rate or process of caudal vertebra number evolution is higher in the Caribbean compared to continental *Anolis*, theorizing that this might arise from stronger diversifying ecological selection (due to the more diverse specialized ecologies of the insular *Anolis*; e.g., Losos, 2009) in island compared to mainland lizards.

Model, methods, and results

The model

Like nearly all modern methods for studying the evolution of discretely valued character states on the phylogeny (but see Felsenstein, 2005, 2012; Revell, 2014), the model of this study is a flavour of the *Mk* model of Lewis (2001). The *Mk* model is so called because it describes a continuous-time discrete-state Markov chain with a total of k possible states. (Thus an *Mk* model with two states is sometimes called an M2 model; a model with three states an M3 model; and so on.) Under this model a set of non-negative real numbers ($q_{i,j}$) gives the instantaneous transition rates between states i and j for all $i \neq j$.

In the simplest case, we could imagine an M2 process in which $q_{0,1} = q_{1,0}$: that is to say, the rates of transition from state 0 to 1 and from state 1 to 0 are equal. In this scenario, the probability of beginning in state 0 and ending in state 1 after time t can be written as:

$$P_{0,1}(t) = \frac{1}{2} - \frac{1}{2}e^{-2q_{0,1}t}$$

(Lewis, 2001). This probability integrates over all the ways in which a time period of length t can begin in state 0 and change to state 1: that is, by changing state once, $0 \rightarrow 1$; by changing state, reversing back, and then changing again ($0 \rightarrow 1 \rightarrow 0 \rightarrow 1$); and so on (Lewis, 2001). For conditions in which the probability of multiple changes during time t is very small (e.g., low $q_{0,1}$, small t , or $q_{1,0} = 0$), this expression will converge on the simple integral of an exponential distribution with shape parameter $q_{0,1}$ from $0 \rightarrow t$, as in this case, we are merely computing the probability that a rare event has occurred after time t which should have an exponential probability density if the rare event occurs randomly at a constant rate.

Obviously, the probability of starting and ending in the same state 0 is merely one minus our previous expression, or:

$$P_{0,0}(t) = \frac{1}{2} + \frac{1}{2}e^{-2q_{0,1}t}$$

(Lewis, 2001). More generally, in a k -state process in which transition rates need not be equal nor symmetrical among states, the matrix of transition probabilities between states can be obtained by exponentiating a transition matrix, \mathbf{Q} , multiplied by the elapsed time, t , in which each i, j th element of \mathbf{Q} (for $i \neq j$) contains the instantaneous transition rate between i and j , and in which the diagonal elements are set to $-\sum_{j=1, i \neq j}^k q_{i,j}$, such that each row-sum of \mathbf{Q} is equal to zero. In other words:

$$\mathbf{P} = \exp(\mathbf{Q}t)$$

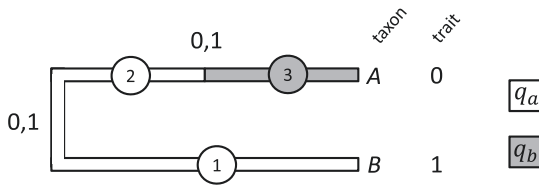
Here, $\exp(x)$ denotes the matrix exponential of x and \mathbf{P} is a matrix containing the probabilities of changing (or not, on the diagonal of the matrix) between any pair of states. It is straightforward to then proceed and compute the likelihood of any particular value of \mathbf{Q} , given the data at the tips of the phylogeny and our set of edge lengths, which can be done efficiently using the pruning algorithm of Felsenstein (1973, 1981). For any tree and vector of observations for a discrete character at the tips of the tree, the value of \mathbf{Q} that maximizes this likelihood would be our maximum likelihood estimate (MLE) of the transition matrix \mathbf{Q} . (Note that the model called the *Mk* model by Lewis (2001) explicitly assumes that all transitions between states occur at the same rate. Although the more general model in which transitions between different pairs of states are permitted to occur at different rates is now often also called the “*Mk* model”—in some places, it has been more precisely described as the “extended” *Mk* model; e.g., Harmon, 2018; Revell & Harmon, 2022. Herein, we will follow the more typical contemporary convention and refer to the general, k -state Markov model as the *Mk* model.)

Elaborating on this model but slightly, we imagine two different transition processes, a and b , operating simultaneously in different parts of the tree in which (for simplicity) $q_{0,1}^a = q_{1,0}^a$ and $q_{0,1}^b = q_{1,0}^b$, but $q_{0,1}^a \neq q_{0,1}^b$. In this case, for a time interval t consisting first of time in condition a , t_a , followed by time in condition b , t_b , the probability of starting in state 0 and ending in state 1 after time $t = t_a + t_b$ will be equal to the probability of starting in state 0 and ending time t_a in state 1, then beginning and ending time t_b in state 1, plus the probability of starting and ending time t_a in state 0, times the probability of starting time interval t_b in state 0 and ending it in state 1. Simply following our equations from earlier, this can be written as:

$$P_{0,1}(t) = \left(\frac{1}{2} - \frac{1}{2}e^{-2q_a t_a}\right) \left(\frac{1}{2} + \frac{1}{2}e^{-2q_b t_b}\right) + \left(\frac{1}{2} + \frac{1}{2}e^{-2q_a t_a}\right) \left(\frac{1}{2} - \frac{1}{2}e^{-2q_b t_b}\right)$$

in which we have substituted q_a for $q_{0,1}^a = q_{1,0}^a$ and q_b for $q_{0,1}^b = q_{1,0}^b$. This calculation is illustrated for a simplified “phylogeny” with two taxa in Figure 1, in which the likelihood that is computed (at the root of the tree) is equal to probability of the observed data at the tips (taxon A in state “0” and taxon B in state “1”), given the phylogeny and the model of evolution in which a regime shift (from transition rate q_a to q_b) has been postulated along the edge leading to taxon A.

More generally, for arbitrarily different transition matrices \mathbf{Q}_a and \mathbf{Q}_b , the matrix of transition probabilities along



$$\begin{aligned}
 P[0,1] &= \pi_0 P_{0,1}(t_1) P_{0,0}(t_{2,3}) + \pi_1 P_{1,1}(t_1) P_{1,0}(t_{2,3}) \\
 &= \pi_0 P_{0,1}(t_1) P_{0,0}(t_2) P_{0,0}(t_3) + \pi_0 P_{0,1}(t_1) P_{0,1}(t_2) P_{1,0}(t_3) + \\
 &\quad \pi_1 P_{1,1}(t_1) P_{1,1}(t_2) P_{1,0}(t_3) + \pi_1 P_{1,1}(t_1) P_{1,0}(t_2) P_{0,0}(t_3) \\
 &= \pi_0 \left(\frac{1}{2} - \frac{1}{2}e^{-2q_a t_1}\right) \left(\frac{1}{2} + \frac{1}{2}e^{-2q_a t_2}\right) \left(\frac{1}{2} + \frac{1}{2}e^{-2q_b t_3}\right) + \\
 &\quad \pi_0 \left(\frac{1}{2} - \frac{1}{2}e^{-2q_a t_1}\right) \left(\frac{1}{2} - \frac{1}{2}e^{-2q_a t_2}\right) \left(\frac{1}{2} - \frac{1}{2}e^{-2q_b t_3}\right) + \\
 &\quad \pi_1 \left(\frac{1}{2} + \frac{1}{2}e^{-2q_a t_1}\right) \left(\frac{1}{2} + \frac{1}{2}e^{-2q_a t_2}\right) \left(\frac{1}{2} - \frac{1}{2}e^{-2q_b t_3}\right) + \\
 &\quad \pi_1 \left(\frac{1}{2} + \frac{1}{2}e^{-2q_a t_1}\right) \left(\frac{1}{2} - \frac{1}{2}e^{-2q_a t_2}\right) \left(\frac{1}{2} + \frac{1}{2}e^{-2q_b t_3}\right)
 \end{aligned}$$

Figure 1. Illustration of the calculation of the probability of observing the data pattern [0, 1] for taxa A and B, respectively, on a two-taxon tree with a rate shift hypothesized a priori along one of the two edges of the tree. For a given set of values for the transition parameters of our model, this is equivalent to the likelihood at the root of the phylogeny. For purposes of simplifying the calculation, we assume that for a given regime (a or b) the rate of transition (q) between binary states 0 and 1 is the same as between states 1 and 0 (that is to say, that $q_a = q_{0,1}^a = q_{1,0}^a$ and $q_b = q_{0,1}^b = q_{1,0}^b$); and that π_0 and π_1 gives the prior probability of 0 and 1 at the root, respectively.

a given edge is divided into time intervals t_a and t_b can be written as:

$$P = \exp(Q_a t_a) \exp(Q_b t_b).$$

To see how this might be so, let's once again revert back to our binary character in which we said:

$$P = \exp(Q t) = \begin{bmatrix} P_{0,0}(t) & P_{0,1}(t) \\ P_{1,0}(t) & P_{1,1}(t) \end{bmatrix}.$$

Thus for two successive time periods (t_a and t_b) on an edge, we have:

$$P = \begin{bmatrix} P_{0,0}(t_a) & P_{0,1}(t_a) \\ P_{1,0}(t_a) & P_{1,1}(t_a) \end{bmatrix} \begin{bmatrix} P_{0,0}(t_b) & P_{0,1}(t_b) \\ P_{1,0}(t_b) & P_{1,1}(t_b) \end{bmatrix}$$

and consequently:

$$P = \begin{bmatrix} P_{0,0}(t_a)P_{0,0}(t_b) + P_{0,1}(t_a)P_{1,0}(t_b) & P_{0,0}(t_a)P_{0,1}(t_b) + P_{0,1}(t_a)P_{1,1}(t_b) \\ P_{1,0}(t_a)P_{0,0}(t_b) + P_{1,1}(t_a)P_{1,0}(t_b) & P_{1,0}(t_a)P_{0,1}(t_b) + P_{1,1}(t_a)P_{1,1}(t_b) \end{bmatrix}$$

(via normal matrix multiplication), which clearly enumerates all the ways of getting between the two character states over the time intervals t_a and t_b . For instance, $P_{1,2}$ (the element of P in the first row and second column) contains the probability of starting t in state 0 and ending in state 1, which is equal to the sum of the probability of starting and ending segment a of length t_a in state 0, but then changing from 0 to 1 along segment b of length t_b , and the probability of starting segment a in state 0 and ending state 1, and then not changing state along segment b . Note that the calculation of the probability $P(i|i)$ integrates over the possibility of not changing at all and the possibility of one or more changes to other states, followed by a reversal to i . Naturally, this can be generalized to any number of character states or rate regimes.

As before, it is straightforward to compute the likelihood for any set of values in Q_a and Q_b , using the pruning algorithm of Felsenstein (1973, 1981), and if we identify the values of Q_a and Q_b (and so on, for more than two regimes) that maximize the likelihood, we have found our MLEs of the two or various transition matrices of our model. Lastly, although we have included the detailed explanation above on a two-taxon tree with two mapped regimes useful to conceptualize the calculation of the likelihood, in practice, it is much more straightforward to split any edge with a regime shift into two or more edges with one or more unbranching nodes. Then one can simply apply the pruning algorithm of Felsenstein (1981) to compute the likelihood, but in which each edge is allowed to have a different value for the matrix of transition rates between states, Q , depending on its (single) mapped regime. Indeed, this is precisely how these calculations have been implemented in software.

Notes on implementation

All the models and methods of this study have been implemented for the R statistical computing environment (R Core Team, 2024), and all simulations and analyses were conducted in R. The statistical method described herein has been implemented in the function *fitmultiMk* of the *phytools* R package (Revell, 2012, 2024). *phytools* itself in turn depends on the important R phylogenetics package *ape* (Paradis & Schliep, 2019; Paradis et al., 2004; Popescu et al., 2012), as well as on a number of other R packages (Azzalini & Genz, 2022; Becker et al., 2022; Chasalow, 2012; Gerber & Furrer, 2019; Gilbert & Varadhan 2016; Lemon, 2006; Ligges & Mächler, 2003; Maechler et al., 2023; Microsoft Corp. and Weston, 2022a, 2022b; Mullen et al., 2011; Pinheiro et al., 2023; Plummer et al., 2006; Qiu & Joe, 2023; Schliep, 2011; Venables & Ripley, 2002). Finally, we used the packages *future.apply* (Bengtsson, 2021), *geiger* (Pennell et al., 2014), and *lmtest* (Zeileis & Hothorn, 2002), in various analyses and simulations of this study.

Simulation tests of the method and results

To examine the type I error rate of the method, we first simulated sets of 200 pure-birth phylogenies, each containing either $N = 25, 50, 100, 200, 400,$ or 800 taxa, all of which were rescaled to have a total depth of 1.0 unit. This total depth is arbitrary, but we chose to keep it constant across simulations so that (for instance) our 800 taxon tree would not have a much greater average root-to-tip distance than, say, our 25 or 50 taxon stochastic phylogenies. We then randomly selected two non-nested clades of each tree with the intent of assigning the tips and edges of these two clades into the same second derived regime, to be then compared against the basal regime on the rest of the three. (Though we appreciate that the term “basal” is somewhat controversial when applied to extant taxa, here we use it specifically to refer to a regime mapped to the genuine root of the tree, so we think it is appropriate.) We thus selected the two clades such that: (1) each selected clade contained at least two taxa; and (2) collectively the two clades comprised neither less than 25% nor more than 75% of the tips of the tree. We simulated a constant rate of character evolution on the tree in which $q_{0,1} = q_{1,0} = 0.5$, corresponding to an expected number of changes from the root to any tip of 0.5, regardless of the number of taxa in the tree. (Obviously, larger trees nonetheless possessed more changes on average, because they have more total edge length for a given depth.) We then proceeded to fit the heterogeneous rate model presented in this

article along with a standard Mk (i.e., $M2$) model with symmetrical transition rates between states. We compared the likelihoods of the two models using a likelihood-ratio test with one degree of freedom, for the one additional parameter estimated in the multi-regime model. The distribution of p -values for each simulation condition is given in [Supplementary Appendix Figure S1](#); whereas [Supplementary Appendix Table S1](#) shows the measured type I error rates for each size of tree. On average, we found that type I error was just very slightly above its nominal level (average type I error = 0.054). Furthermore, in no simulation conditions was type I error significantly greater than 0.05 according to a one-tailed binomial test, with only a relatively weak tendency for type I error to be higher in smaller compared to larger phylogenetic trees ([Supplementary Figure S1](#); [Supplementary Table S1](#)).

We also undertook a set of simulations to examine the power and parameter estimation of the method. For this analysis, we used only the set of 200 simulated 100-taxon trees from the type I error analysis. On these trees, we simulated two different rate regimes for the binary character, in which the regime was determined by our a priori mapping and where $q_a = q_{0,1}^a = q_{1,0}^a = 0.5$, whereas $q_b = q_{0,1}^b = q_{1,0}^b = 0.5, 1, 2, 4, \text{ and } 8$. As one might expect, the power to reject the null increased with the difference in rate between regimes ([Supplementary Table S2](#)). The median estimated value of the transition rates was largely unbiased across all differences in rate ([Supplementary Table S2](#); [Figure 2](#)); however, the mean estimated value of q_b was upwardly biased for higher generating values of q_b , evidently due to a small fraction of simulations in which q_b was badly overestimated ([Figure 2](#)).

Empirical example

In addition to these simulations, we also fit our model to two different empirical cases. In both of these, we examined the rate of evolution of a discrete character in mainland vs. island lineages of lizards in the genus *Anolis*, known as anoles ([Losos, 2009](#)). Since the number of transitions from mainland to island in anoles (and vice versa) is relatively few, we treated these for our purposes as having occurred at known locations on our phylogeny, which we obtained from [Gamble et al. \(2014\)](#). In particular, we assumed that the global ancestral node of all anoles was present on the continental mainland, that occupancy of the Caribbean islands from mainland lineages (or vice versa) occurred via colonization, and then we proceeded to place colonization events precisely halfway along the edge leading to each clade in which descendants were present in the islands. We also reconstructed one island-to-mainland colonization event and within this clade a further secondary colonization of islands. (Though this lattermost colonization is not recovered in several other phylogenetic analyses of anoles, e.g., [Alföldi et al., 2011](#); [Poe et al., 2017](#), we do not expect that this detail will affect our analyses in a significant way.) Our phylogeny thus includes a total of 120 islands and 64 mainland taxa, with four changes between mainland and island or the reverse. The mainland/island history that we assumed for the purposes of this analysis is mapped onto the tree in [Figure 3](#).

Using this mainland/island history as a basis for all subsequent inferences, we next analyzed dewlap colour evolution. Data for this analysis were originally obtained for a study on dewlap size evolution in anoles ([Ingram et al., 2016](#)). The dewlap is an extensible gular fan used by anoles (and some other lizards) for both intra- and interspecific displays.

Importantly, the colour of the dewlap has been shown to be a critical mate recognition cue in many species ([Losos, 2009](#)). Our logic in comparing the rate of anole dewlap colour evolution between island and continental faunas was the following. Although little studied, at a coarse scale mainland and island anole faunas differ in that mainland communities support higher syntopic species richness but lower average abundances compared to island communities ([Anderson & Poe, 2019](#); [Andrews, 1979](#)). The ultimate causes of dewlap colour diversity remain poorly known ([Losos, 2009](#); [Nicholson et al., 2007](#)), but if either sexual selection at high densities or selection for species recognition at high diversities is responsible for phylogenetic transitions in dewlap colour, we might expect mainland and island anole lineages to exhibit different rates of dewlap colour evolution. Since we do not know whether syntopy or density is liable to create more pressure on anole dewlap colours to diversify one from the other, we merely hypothesize that the rate of evolution of dewlap colour should differ between mainland and island faunas, rather than pre-supposing that island lineages will evolve at a higher rate than their continental counterparts or the converse.

Dewlaps come in many colours and colour combinations, but to keep the discrete character models tractable, we coded dewlap colour by placing the dominant (by dewlap surface area) colour of each species' dewlap as being subjectively closest to one of the following six states: *black* (anoles with black, blue, or purple dewlaps); *orange* (orange or brown dewlaps); *pink* (pink and peach dewlapped anoles); *red*; *white* (white and grey dewlaps); or, finally, *yellow* (yellow and green dewlaps; [Figure 3](#)). Obviously, considerable nuance will be lost first in reducing the often complex dewlap colouration to a single dominant colour for each species, and further still by reducing the genuine interspecific variation in dominant colour to our minimal, six-state set. We would not, however, expect this simplification of our dataset to lead to an elevated type I error in the method. (To the contrary, we suspect that the opposite is more likely to be true.)

We then proceeded to fit a series of six models to the data and tree. These six models consisted of an evolutionary process in which: transitions occurred at the same rate between all pairs of states (ER); transitions occurred at the same backward and forward rate between each pair of states, but could occur at different rates between different state pairs (SYM); and transitions occurred at different rates between each pair of states (ARD). We fit each of these models either allowing for different rates between mainland and island lineages (-M) or forcing them to have the same rates of change between character states (-S), thus resulting in six models of varying complexity in all (ER-S, ER-M, SYM-S, SYM-M, ARD-S, and ARD-M). We assumed a flat prior probability distribution at the root for all models. Results from this analysis are given in [Table 1](#). In general, although in the best-fitting multi-rate models (SYM-M) the average transition rate between states was higher on islands than in mainland lineages—penalizing for the number of parameters to be estimated, the best-supported model was clearly a model in which both mainland and island fauna dewlap dominant colour evolved under the same set of rates of transition between states (SYM-S; [Table 1](#)).

In addition to this character, we also analyzed island and mainland caudal (i.e., tail) vertebra number evolution. These data were obtained by simply counting the number of vertebrae from the pelvic girdle to the tip of the tail in a specimen in which the tail was previously deemed to be

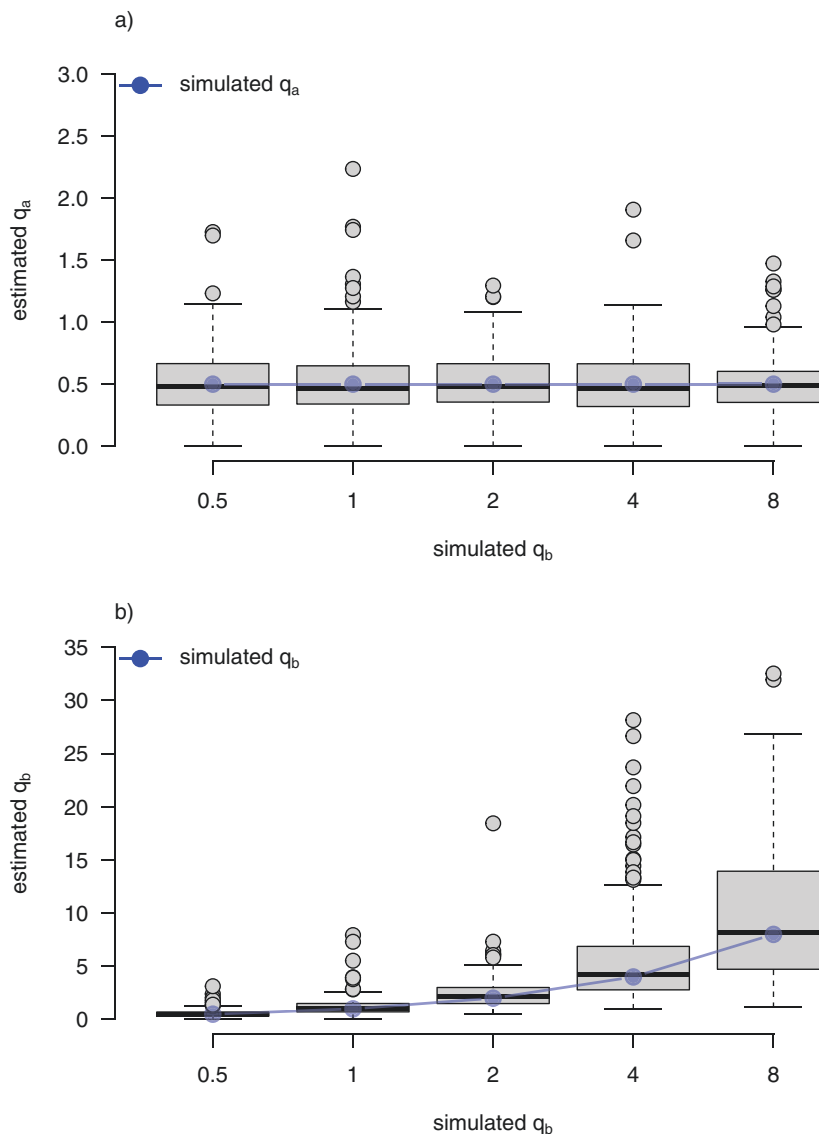


Figure 2. Results from an analysis of parameter estimation. Each panel shows a boxplot for estimated values of the transition rate between states under regime *a* (panel a) and under regime *b* (panel b). As in a standard boxplot, the horizontal black lines show the median value of \hat{q}_a or \hat{q}_b for each simulation, the grey box shows the 25–75% interquartile range, and the whiskers extend 1.5 times the interquartile range or to the maximum or minimum value of \hat{q}_a or \hat{q}_b (whichever is smaller or larger, respectively). Values outside 1.5 times the interquartile range are plotted as points. In panel b, 2.9% of points (29 in 1,000 simulations) had estimates, \hat{q}_b , greater than the vertical height of y and were thus left out of the figure; however, these values were included in the calculation of the median and interquartile ranges of the boxplot, and in the mean and medians presented in [Supplementary Appendix Table S2](#). The solid points and connecting lines show the simulated values of q_a and q_b for panels a and b, respectively.

completely intact (e.g., [Figure 4](#)). Our logic in comparing the rate of anole caudal vertebra evolution between island and mainland lineages is simply that conventional wisdom suggests that Caribbean anoles exhibit greater arboreal microhabitat diversification and specialization than do their mainland congeners ([Losos, 2009](#); whilst acknowledging that mainland anoles are also ecologically and morphologically diverse, e.g., [Pinto et al., 2008](#)). The tail is an appendage that can play an important role in locomotion, particularly in an arboreal setting. Consequently, it seemed reasonable to imagine that it might be under stronger divergent selection in the Caribbean, where anoles may fill a wider array of arboreal and semi-arboreal ecological roles than in continental clades. As such, and in contrast to our non-directional hypothesis for dewlap colour evolution, we hypothesized that the island *Anolis* lizard caudal vertebra

number should evolve at a higher rate than for mainland lineages.

Given that the number of caudal vertebrae varies over quite a broad range (from 34 to 55 in these data), one might intuitively assume that the number of parameters to estimate in this model would be impossibly large. In fact, if we make some relatively reasonable simplifying assumptions (keeping in mind that all models are, by definition, intended to be simplifications of reality) the dimensionality of the problem can be quite manageable, even though the state-space is big. Specifically, we decided to treat the acquisition and loss of caudal vertebrae as an ordered process—in which gain and loss were free to proceed with different tempos, but where changes in the same direction between any pair of adjacent states should occur with the same rate ([Supplementary Appendix Figure S2](#)). Once again, though we found that the

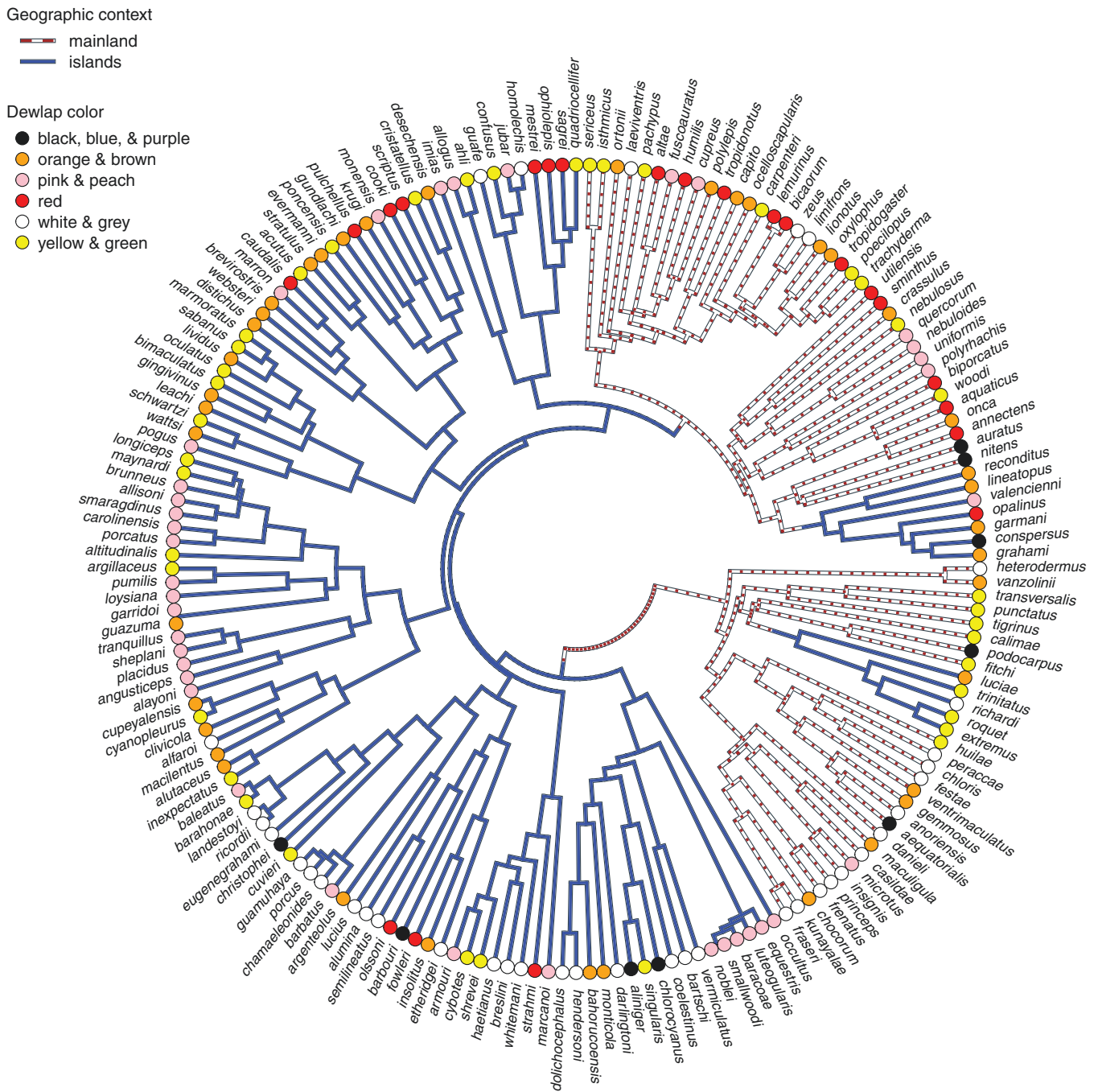


Figure 3. Dominant dewlap colour mapped onto a phylogeny of Caribbean (solid branches) and mainland (dashed branches, see caption) *Anolis* lizards.

estimated rate of character evolution in the best-fitting model was higher in islands than in mainland anole lineages, the best-supported model (accounting for parameterization) was, as before, the ordered, single-rate model (Table 2).

In addition to comparing our multi-rate models to standard, homogeneous-rate M_k models, as detailed above, for illustrative purposes only we also fit a total of three additional two-regime HRMs (Beaulieu et al., 2013; ER, SYM, ARD, but with hidden rates) to our dewlap dominant colour data, and a single additional HRM (ordered, but with hidden rates) to our caudal vertebra number dataset. We argue in the Discussion section, below, why it might be important to include HRMs in an analysis of regime-based rate heterogeneity in cases in

which the null hypothesis of rate homogeneity is rejected (although perhaps less so when it is not, as in this study). In addition to standard HRMs, it is also theoretically (and practically, in some software) possible to fit a model with both mapped regimes and a hidden character. We do not explore this category of model here, but it might be the subject of future study.

Model likelihoods, parameter complexity, Akaike information criterion (AIC), and model weights for all models are given in Supplementary Appendix Tables S3 and S4. The inclusion of these additional HRMs did not substantively alter our main result in that we still did not find evidence of rate heterogeneity in dominant dewlap colour nor in caudal

Table 1. Mean transition rates, log-likelihoods, number of fitted parameters, and AIC for the six fitted models of dewlap dominant colour evolution described in the main text. The best-supported model (SYM-S) is highlighted in bold text. Also, see [Supplementary Appendix Table S3](#).

Model	\bar{q}_{islands}	$\bar{q}_{\text{mainland}}$	log(L)	k	AIC
ER-S	0.009	0.009	-308.6	1	619.3
ER-M	0.009	0.009	-308.6	2	621.3
SYM-S	0.073	0.073	-289.8	15	609.6
SYM-M	0.071	0.012	-285.6	30	631.3
ARD-S	0.009	0.009	-282.7	30	625.4
ARD-M	0.013	0.011	-272.6	60	665.3



Figure 4. Digital x-ray image of an anole (*Anolis sagrei*) showing the caudal vertebrae of an intact tail. Image courtesy K. Winchell.

vertebra number linked to island or mainland geographic regimes. For dominant dewlap colour, the best-supported model (by AIC and model weight) continued to be the symmetric (SYM) single-regime model ([Supplementary Appendix Table S3](#)). By contrast, with the addition of HRMs, the best-supported model for caudal vertebra number (by the same criteria) became the HRM ([Supplementary Appendix Table S4](#)). An illustration of the structure and parameter estimates of this HRM is given by [Supplementary Appendix Figure S3](#).

Discussion

In recent decades, phylogenetic comparative methods have exploded in popularity within evolutionary research. This growth has generally been accompanied by an appreciation that models assuming a homogeneous process of evolution

Table 2. Rate of caudal vertebra loss and gain on the mainland [M] and islands [I], log-likelihoods, number of fitted parameters, and AIC for the two fitted models caudal vertebra number evolution described in the main text. As in [Table 1](#), the best-supported model (ordered-single) is highlighted in bold text. Also, see [Supplementary Appendix Table S4](#).

Model	q_{loss}	q_{gain}	log(L)	k	AIC
Ordered single	0.494	0.318	-305.1	2	614.2
Ordered multiple	0.523 [I] 0.373 [M]	0.333 [I] 0.237 [M]	-304.4	4	616.7

may be overly simplistic for many phylogenies, particularly those that contain many taxa or that span vast periods of evolutionary time. A number of methods have been developed to explicitly model heterogeneity in the rate or process of phenotypic evolution on the tree. An important class of such method involves the investigator “painting” different a priori regimes onto the branches or clades of a phylogeny, and then testing the hypothesis that evolutionary change for one or more characters differs between different painted regimes ([Butler & King, 2004](#)). This approach was perhaps best formalized by [O’Meara et al. \(2006\)](#) in the context of testing hypotheses about changes in the rate of evolution for a continuous character evolving by Brownian motion on the tree.

Surprisingly, and in spite of the considerable popularity of this type of analysis for continuous character data, no precisely analogous approach has ever been proposed for discretely valued traits. Herein, we present just such an approach. According to the method of this article, and just as in the method of [O’Meara et al. \(2006\)](#), the user must propose a priori a specific hypothesis for where the rate or process of character evolution is thought to vary on their phylogeny. This hypothesis can then be fit using likelihood and compared to other hypotheses such as a null hypothesis of constant rates of evolution, or other alternative hypotheses about how the rate varies among clades or branches of the phylogeny. We show that the method, which has been implemented in the R package *phytools* ([Revell, 2012](#)), has an acceptable type I error when the null is true, as well as reasonable power and parameter estimation when it is not ([Supplementary Figures S1 and S2](#); [Supplementary Tables S1 and S2](#)).

Relationship to other methods

Models in which regime shifts are postulated a priori by the user, painted on the tree, and then tested against alternative paintings or a null hypothesis of homogeneous evolution, have been around for nearly two decades, and are already quite popular for the analysis of one or more continuous characters on the tree (e.g., [Beaulieu et al., 2012](#); [Butler & King, 2004](#); [O’Meara et al., 2006](#); [Revell & Collar, 2009](#)). We were somewhat surprised to discover that a precise analogue did not exist for discretely valued character traits and have tried to fill that void with this contribution. To our knowledge, the sole exception is the codon branch model of [Yang \(1998](#); [Yang & Nielsen, 1998, 2002](#)), which, so far as we are aware, has never been extended beyond testing for positive selection in gene sequences. Other prior methods have been proposed to explicitly model heterogeneity in the rate or process of evolution in discrete characters on phylogenetic trees. Perhaps most significantly, [Pagel \(1994\)](#) presented a model in which the rate of evolution of one binary character depends on the state of a second character,

or vice versa; or in which the evolution of both characters is interdependent. Since this model will fit well when two characters tend to evolve towards certain character state combinations, the method of Pagel (1994) is often interpreted as a test for the correlated evolution of two binary traits. As shown by Maddison and Fitzjohn (2015; and as is obvious from the structure of the model), Pagel's (1994) method will also lead to a significant result if a change in value for one character is correlated with a change in the rate of evolution for a second (even if the former is unreplicated on the tree; Maddison & Fitzjohn, 2015). Our method too will be significant if our discrete character changes in its rate or process of evolution within a single clade that we have specified a priori. However, in our case, this is by design, not by accident. Under this circumstance, a significant result would merely indicate that the data suggest that our single focal subclade evolved by a different process or rate than did the taxa of the rest of the phylogeny, and any attribution of this finding to a specific biological cause is not implied by the result and would be left instead to the interpretation of the investigator.

Another very important category of analytical method also exists for modelling rate heterogeneity of a discrete character trait on the phylogeny. These are models in which the rate of evolution of our trait is assumed to be under the influence of a second, unobserved, discrete character trait. This class of method traces its history to the covarion model of Tuffley and Steel (1997), but was adapted to phylogenetic comparative analysis as the "hidden-rates model" (HRM) by Beaulieu et al. (2013; also see Marazzi et al., 2012). The HRM differs from what we have proposed herein in that under the HRM we allow our data to tell us where character evolution may have changed in rate, rather than hypothesizing the location of one or more rate shifts a priori. We will revisit the HRM in subsequent discussion, below.

Finally, a completely separate class of model has recently been proposed for the phylogenetic comparative analysis of discrete character evolution that is based on the threshold model from quantitative genetics originally developed by Wright (1934; Felsenstein, 2005, 2012). According to the threshold model, our observed discrete character is simply a manifestation of an underlying, unobserved continuous trait (called liability) and one or more thresholds. Under this model, whenever liability crosses a threshold, the discrete character changes state. The threshold model can also create heterogeneity in the evolutionary process through time and among lineages. For instance, lineages near the threshold may change state many times, whereas clades far from any threshold will tend not to change at all (Revell, 2014; Revell & Harmon, 2022). Indeed, there are some circumstances in which evolution under the threshold model might be well-approximated by the model of this article. For example, if the regimes hypothesized a priori by the investigator correspond to parts of the tree in which liability is near a threshold (and thus the discrete trait changes frequently) or far from any threshold (where the character would not be expected to change at all), then the multi-regime model will likely provide a good explanation for our data. We should make clear, however, that it would not be appropriate under any circumstances to paint regimes on the tree after having observed the data such that clades with many changes are in one "regime" while clades with fewer changes are in another.

Notes on the empirical case studies

In addition to presenting this method and exploring its statistical attributes via simulation, we also applied the method to a pair of case studies. In each, we hypothesized that the rate of evolution for a discrete character (dominant dewlap colour and number of caudal vertebrae) might differ between mainland and Caribbean *Anolis* lizard faunas. Transitions between the Central and South American mainland and the Caribbean have occurred so few times that they can be reasonably unambiguously reconstructed on the tree of anoles (Figure 3). We hypothesized a priori that dominant dewlap colour might evolve at different rates in the Caribbean and mainland lineages due to average (and opposite) differences in diversity and abundance in these island versus mainland environments. For example, higher densities of conspecifics might favour stronger and more frequent selection acting on dewlap colouration to compete for mates. On the other hand, higher species richness could favour the evolution of greater dewlap diversity to reduce costly heterospecific mating attempts. We also hypothesized that caudal vertebra number might evolve more rapidly in Caribbean lizards under the assumption that Caribbean anoles tend to exhibit more specialized arboreal niche use, and thus that the tail might be under stronger divergent selection on the islands due to the different locomotor demands imposed by highly specialized arboreal locomotion.

We found relatively little evidence in support of either hypothesis; however, from the perspective of methodology, perhaps our non-finding should be as much encouraging as it is disappointing. Over the past few years, it has become a relatively popular endeavour to identify circumstances in which phylogenetic comparative methods can lead us spuriously astray due to null-model inadequacy (Maddison & Fitzjohn, 2015; O'Meara & Beaulieu, 2016; Rabosky & Goldberg, 2015). Maybe we should feel encouraged by the fact that not all simple null hypotheses will be rejected in favour of more complicated alternatives, if these alternatives are not, in fact, supported by our data. On the other hand, it would be incorrect to claim that our results show that no interesting differences exist in tail or dewlap evolution between mainland and Caribbean anoles. To the contrary, our analysis has merely demonstrated that a pair of relatively simple a priori hypotheses for rate variation between continental and island anoles are not supported when confronted with data.

A brief comment on model adequacy, model selection, and type I error

In this article, we have focused on a rate heterogeneous discrete character evolution model conceptually most closely akin to the rate or process heterogeneous continuous trait models described by Butler and King (2004) or O'Meara et al. (2006). As alluded to in the prior section, however, both frequentist hypothesis testing and information-theoretic model selection (of the type that we illustrated both in our simulation and empirical analyses), using all such methods and models, are vulnerable to the inadequacy of the null and alternative models of the type made famous (in phylogenetic comparative method circles) by Maddison and Fitzjohn (2015; for the Pagel, 1994 model of correlated binary trait evolution) and Rabosky and Goldberg (2015; for state-dependent diversification models). That is to say, readers eager to apply the analysis of this article to their own data should keep in mind that, just as for any frequentist hypothesis test, rejection of the null does not mean that the alternative is correct—merely that

the null model is a poor description of our data. We would thus encourage users to include this model among a suite of possible scenarios for how their character may have evolved: for example, comparing a regime-based hypothesis of rate heterogeneity to an HRM (Beaulieu et al., 2013). Indeed, the latter has now been implemented in *phytools* (as well as in the excellent *corHMM* package by Beaulieu et al., 2023; Boyko & Beaulieu, 2021) and can be compared directly to the model fit of this method using standard statistical machinery of likelihoods. In our analysis, we failed to reject the null hypotheses in both of our cases, which forces us to conclude that the alternatives (different character transition rates between island and mainland *Anolis*) are not well-supported by our data. This conclusion would be unaffected by stronger model support for a hidden-rate scenario. For illustrative purposes, however, we have nonetheless included this analysis in a supplementary appendix to this article (see [Supplementary Appendix Tables S3 and S4](#); [Supplementary Figure S3](#)). Finally, users might also contemplate assessing the absolute model adequacy of their best-supported model, for example, using posterior predictive simulation (e.g., Brown & Thomson, 2018).

Conclusion

Here we have presented a relatively simple extension of the familiar *Mk* model of Lewis (2001), but in which the rate or process of discrete trait evolution is permitted to differ among branches, clades, or time periods that have been hypothesized a priori by the user. We show via simulation that the method has reasonable statistical attributes—such as type I error not significantly higher than its nominal value and median parameter estimates across simulations that are very close to their generating values. Finally, we apply the method to two different case studies comparing the rate of discrete character evolution between the Caribbean and mainland *Anolis* lizard clades.

Supplementary material

Supplementary material is available at *Journal of Evolutionary Biology* online.

Data availability

All of the analyses of this article were conducted using free, open-source software distributed via the Comprehensive R Archive Network, CRAN (<https://cran.r-project.org/>). The data files and code required to exactly reiterate the analyses of this article can be found at <https://doi.org/10.5281/zenodo.13208973>.

Author contributions

Liam Revell (Conceptualization [lead], Data curation [supporting], Formal analysis [lead], Funding acquisition [equal], Investigation, Methodology, Project administration [lead], Resources [equal], Software [lead], Supervision [equal], Validation, Visualization, Writing—original draft, Writing—review & editing [lead]), Klaus Schliep (Conceptualization, Data curation [supporting], Formal analysis, Funding acquisition [equal], Investigation, Methodology, Project administration, Resources [supporting], Software [equal], Supervision, Validation, Visualization, Writing—original draft, Writing—review & editing [supporting]), Luke

Mahler (Conceptualization [equal], Data curation [lead], Formal analysis [supporting], Funding acquisition [equal], Investigation, Methodology, Project administration [supporting], Resources [equal], Software [supporting], Supervision [equal], Validation, Visualization, Writing—original draft, Writing—review & editing [supporting]), and Travis Ingram (Conceptualization [supporting], Data curation [lead], Formal analysis, Funding acquisition [equal], Investigation, Methodology, Project administration [supporting], Resources [equal], Software [supporting], Supervision [equal], Validation, Visualization, Writing—original draft, Writing—review & editing [supporting])

Acknowledgments

The authors would like to thank A. Harrison, R. Castañeda, R. Glor, A. Herrel, S. Poe, and Y. Stuart for their help compiling the dewlap colour data used in this study; K. Winchell for providing the image used in [Figure 4](#); and the United States National Science Foundation (DBI 1759940) and Fondo Nacional de Desarrollo Científico y Tecnológico (FONDECYT), Chile (1201869) for supporting portions of this research.

Conflicts of interest

None declared.

References

- Alföldi, J., Di Palma, F., Grabherr, M., ... Lindblad-Toh, K. (2011). The genome of the green anole lizard and a comparative analysis with birds and mammals. *Nature*, *477*, 587–591.
- Anderson, C. G., & Poe, S. (2019). Phylogeny, biogeography and island effect drive differential evolutionary signals in mainland and island lizard assemblages. *Zoological Journal of the Linnean Society*, *185*, 301–311.
- Andrews, R. M. (1979). Evolution of life histories: A comparison of *Anolis* lizards from matched island and mainland habitats. *Breviora*, *454*, 1–51.
- Azzalini, A., & Genz, A. (2022). The R package “mnormt”: The multivariate normal and “t” distributions (version 2.1.1). <http://azzalini.stat.unipd.it/SW/Pkg-mnormt>
- Beaulieu, J., O’Meara, B., Oliver, J., & Boyko, J. (2023). *corHMM*: Hidden Markov models of character evolution. R package version 2.9. <https://cran.r-project.org/package=corHMM>
- Beaulieu, J. M., Jhwueng, D. -C., Boettiger, C., & O’Meara, B. C. (2012). Modeling stabilizing selection: Expanding the Ornstein-Uhlenbeck model of adaptive evolution. *Evolution*, *66*, 2369–2383.
- Beaulieu, J. M., O’Meara, B. C., & Donoghue, M. J. (2013). Identifying hidden rate changes in the evolution of a binary morphological character: The evolution of plant habit in campanulid angiosperms. *Systematic Biology*, *62*, 725–737.
- Becker, R. A., Wilks, A. R., Brownrigg, R., Minka, T. P., Deckmyn, A., & CRAN Team (2022). *maps*: Draw geographical maps. R package version 3.4.1.1. <https://CRAN.R-project.org/package=maps>
- Bengtsson, H. (2021). A unifying framework for parallel and distributed processing in R using futures. *The R Journal*, *13*(2), 208–227. <https://doi.org/10.32614/rj-2021-048>
- Blomberg, S. P., Garland, T. Jr., & Ives, A. R. (2003). Testing for phylogenetic signal in comparative data: Behavioral traits are more labile. *Evolution*, *57*(4), 717–745. <https://doi.org/10.1111/j.0014-3820.2003.tb00285.x>
- Boucher, F. C., & Démery, V. (2016). Inferring bounded evolution in phenotypic characters from phylogenetic comparative data. *Systematic Biology*, *65*(4), 651–661. <https://doi.org/10.1093/sysbio/syw015>

- Boucher, F. C., Démary, V., Conti, E., ... Uyeda, J. (2018). A general model for estimating macroevolutionary landscapes. *Systematic Biology*, 67(2), 304–319. <https://doi.org/10.1093/sysbio/syx075>
- Boyko, J. D., & Beaulieu, J. M. (2021). Generalized hidden Markov models for phylogenetic comparative datasets. *Methods in Ecology and Evolution*, 12(3), 468–478. <https://doi.org/10.1111/2041-210x.13534>
- Brown, J. M., & Thomson, R. C. (2018). Evaluating model performance in evolutionary biology. *Annual Review of Ecology, Evolution, and Systematics*, 49(1), 95–114. <https://doi.org/10.1146/annurev-ecolsys-110617-062249>
- Butler, M. A., & King, A. A. (2004). Phylogenetic comparative analysis: A modeling approach for adaptive evolution. *The American Naturalist*, 164(6), 683–695. <https://doi.org/10.1086/426002>
- Caetano, D. S., & Harmon, L. J. (2019). Estimating correlated rates of trait evolution with uncertainty. *Systematic Biology*, 68(3), 412–429. <https://doi.org/10.1093/sysbio/syy067>
- Chasalow, S. (2012). combinat: Combinatorics utilities. R package version 0.0-8. <https://CRAN.R-project.org/package=combinat>
- Felsenstein, J. (1973). Maximum likelihood and minimum-steps methods for estimating evolutionary trees from data on discrete characters. *Systematic Biology*, 22(3), 240–249. <https://doi.org/10.1093/sysbio/22.3.240>
- Felsenstein, J. (1981). Evolutionary trees from DNA sequences: A maximum likelihood approach. *Journal of Molecular Evolution*, 17(6), 368–376. <https://doi.org/10.1007/BF01734359>
- Felsenstein, J. (1985). Phylogenies and the comparative method. *American Naturalist*, 125(1), 1–15. <https://doi.org/10.1086/284325>
- Felsenstein, J. (2005). Using the quantitative genetic threshold model for inferences between and within species. *Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences*, 360(1459), 1427–1434. <https://doi.org/10.1098/rstb.2005.1669>
- Felsenstein, J. (2012). A comparative method for both discrete and continuous characters using the threshold model. *The American Naturalist*, 179(2), 145–156. <https://doi.org/10.1086/663681>
- Galtier, N. (2001). A maximum-likelihood phylogenetic analysis under a covarion-like model. *Molecular Biology and Evolution*, 18(5), 866–873. <https://doi.org/10.1093/oxfordjournals.molbev.a003868>
- Gamble, T., Geneva, A. J., Glor, R. E., & Zarkower, D. (2014). *Anolis* sex chromosomes are derived from a single ancestral pair. *Evolution*, 68(4), 1027–1041. <https://doi.org/10.1111/evo.12328>
- Gerber, F., & Furrer, R. (2019). optimParallel: An R package providing a parallel version of the L-BFGS-B optimization method. *The R Journal*, 11, 352–358.
- Gilbert, P., & Varadhan, R. (2016). numDeriv: Accurate numerical derivatives. R package version 2016.8-1. <https://CRAN.R-project.org/package=numDeriv>
- Grafen, A. (1987). The phylogenetic regression. *Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences*, 326, 119–157.
- Hansen, T. F. (1997). Stabilizing selection and the comparative analysis of adaptation. *Evolution*, 51(5), 1341–1351. <https://doi.org/10.1111/j.1558-5646.1997.tb01457.x>
- Harmon, L. J. (2018). *Phylogenetic comparative methods: Learning from trees*. EcoEvoRxiv. 978-1719584463.
- Harvey, P. H., & Pagel, M. D. (1991). *The comparative method in evolutionary biology*. Oxford University Press.
- Hohenlohe, P. A., & Arnold, S. J. (2008). MIPoD: A hypothesis-testing framework for microevolutionary inference from patterns of divergence. *The American Naturalist*, 171(3), 366–385. <https://doi.org/10.1086/527498>
- Ingram, T., Harrison, A., Mahler, D. L., ... Losos, J. B. (2016). Comparative tests of the role of dewlap size in *Anolis* lizard speciation. *Proceedings of the Royal Society of London, Series B: Biological Sciences*, 283, 20162199.
- 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(6), 913–925. <https://doi.org/10.1080/106351501753462876>
- Ligges, U., & Mächler, M. (2003). Scatterplot3d—An R Package for visualizing multivariate data. *Journal of Statistical Software*, 8(11), 1–20.
- Losos, J. B. (2009). *Lizards in an evolutionary tree*. University of California Press.
- Machado, F. A., Mongle, C. S., Slater, G., ... Uyeda, J. C. (2023). Rules of teeth development align microevolution with macroevolution in extant and extinct primates. *Nature Ecology & Evolution*, 7(10), 1729–1739. <https://doi.org/10.1038/s41559-023-02167-w>
- Maddison, W. P., & Fitzjohn, R. G. (2015). The unsolved challenge to phylogenetic correlation tests for categorical characters. *Systematic Biology*, 64(1), 127–136. <https://doi.org/10.1093/sysbio/syu070>
- Maechler, M., Dutang, C., & Goulet, V. (2023). expm: Matrix exponential, log, “etc.” R package version 0.999-7. <https://CRAN.R-project.org/package=expm>
- Mahler, D. L., Ingram, T., Revell, L. J., & Losos, J. B. (2013). Exceptional convergence on the macroevolutionary landscape in island lizard radiations. *Science*, 341(6143), 292–295. <https://doi.org/10.1126/science.1232392>
- Marazzi, B., Ané, C., Simon, M. F., ... Sanderson, M. J. (2012). Locating evolutionary precursors on a phylogenetic tree. *Evolution*, 66(12), 3918–3930. <https://doi.org/10.1111/j.1558-5646.2012.01720.x>
- Martin, B. S., Bradburd, G. S., Harmon, L. J., & Weber, M. G. (2023). Modeling the evolution of rates of continuous trait evolution. *Systematic Biology*, 72(3), 590–605. <https://doi.org/10.1093/sysbio/syac068>
- Microsoft Corporation, & Weston, S. (2022a). doParallel: Foreach parallel adaptor for the “parallel” package. R package version 1.0.17. <https://CRAN.R-project.org/package=doParallel>
- Microsoft Corporation, & Weston, S. (2022b). foreach: Provides foreach looping construct. R package version 1.5.2. <https://CRAN.R-project.org/package=foreach>
- Mullen, K., Ardia, D., Gil, D., ... Cline, J. (2011). ‘DEoptim’: An R package for global optimization by differential evolution. *Journal of Statistical Software*, 40(6), 1–26.
- Nicholson, K. E., Harmon, L. J., & Losos, J. B. (2007). Evolution of *Anolis* lizard dewlap diversity. *PLoS One*, 2(3), e274. <https://doi.org/10.1371/journal.pone.0000274>
- Nunn, C. L. (2011). *The comparative approach in evolutionary anthropology and biology*. University of Chicago Press.
- O’Meara, B. C., Ané, C., Sanderson, M. J., & Wainwright, P. C. (2006). Testing for different rates of continuous trait evolution using likelihood. *Evolution*, 60(5), 922–933.
- O’Meara, B. C., & Beaulieu, J. M. (2016). Past, future, and present of state-dependent models of diversification. *American Journal of Botany*, 103(5), 792–795. <https://doi.org/10.3732/ajb.1600012>
- Pagel, M. (1994). Detecting correlated evolution on phylogenies: A general method for the comparative analysis of discrete characters. *Proceedings of the Royal Society of London, Series B: Biological Sciences*, 255, 37–45.
- Pagel, M., Meade, A., & Barker, D. (2004). Bayesian estimation of ancestral character states on phylogenies. *Systematic Biology*, 53(5), 673–684. <https://doi.org/10.1080/10635150490522232>
- Paradis, E., Claude, J., & Strimmer, K. (2004). APE: Analyses of phylogenetics and evolution in R language. *Bioinformatics*, 20(2), 289–290. <https://doi.org/10.1093/bioinformatics/btg412>
- Paradis, E., & Schliep, K. (2019). ape 5.0: An environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics*, 35(3), 526–528. <https://doi.org/10.1093/bioinformatics/bty633>
- Pennell, M. W., Eastman, J. M., Slater, G. J., ... Harmon, L. J. (2014). geiger v2.0: An expanded suite of methods for fitting macroevolutionary models to phylogenetic trees. *Bioinformatics*, 30(15), 2216–2218. <https://doi.org/10.1093/bioinformatics/btu181>
- Penny, D., McCormish, B. J., Charleston, M. A., & Hendy, M. D. (2001). Mathematical elegance with biochemical realism: The covarion model of molecular evolution. *Journal of Molecular Evolution*, 53, 711–723.

- Pinheiro, J., Bates, D., & R Core Team. (2023). nlme: Linear and Non-linear Mixed Effects Models. R package version 3.1-163. <https://CRAN.R-project.org/package=nlme>
- Pinto, G., Mahler, D. L., Harmon, L. J., & Losos, J. B. (2008). Testing the island effect in adaptive radiation: Rates and patterns of morphological diversification in Caribbean and mainland *Anolis* lizards. *Proceedings of the Royal Society of London, Series B: Biological Sciences*, 275(1652), 2749–2757. <https://doi.org/10.1098/rspb.2008.0686>
- Plummer, M., Best, N., Cowles, K., & Vines, K. (2006). CODA: Convergence diagnosis and output analysis for MCMC. *R-News*, 6, 7–11.
- Poe, S., Nieto-Montes de Oca, A., Torres-Carvajal, O., ... Latella, I. (2017). A phylogenetic, biogeographic, and taxonomic study of all extant species of *Anolis* (Squamata; Iguanidae). *Systematic Biology*, 66(5), 663–697. <https://doi.org/10.1093/sysbio/syx029>
- Popescu, A. -A., Huber, K. T., & Paradis, E. (2012). ape 3.0: New tools for distance-based phylogenetics and evolutionary analysis in R. *Bioinformatics*, 28(11), 1536–1537. <https://doi.org/10.1093/bioinformatics/bts184>
- Qiu, W., & Joe, H. (2023). clusterGeneration: Random cluster generation (with specified degree of separation). R package version 1.3.8. <https://CRAN.R-project.org/package=clusterGeneration>
- R Core Team. (2024). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. <https://www.R-project.org/>
- Rabosky, D. L. (2014). Automatic detection of key innovations, rate shifts, and diversity-dependence on phylogenetic trees. *PLoS One*, 9(2), e89543. <https://doi.org/10.1371/journal.pone.0089543>
- Rabosky, D. L., & Goldberg, E. E. (2015). Model inadequacy and mistaken inferences of trait-dependent speciation. *Systematic Biology*, 64(2), 340–355. <https://doi.org/10.1093/sysbio/syu131>
- Revell, L. J. (2012). phytools: An R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution*, 3(2), 217–223. <https://doi.org/10.1111/j.2041-210x.2011.00169.x>
- Revell, L. J. (2013). A comment on the use of stochastic character maps to estimate evolutionary rate variation in a continuously valued trait. *Systematic Biology*, 62(2), 339–345. <https://doi.org/10.1093/sysbio/sys084>
- Revell, L. J. (2014). Ancestral character estimation under the threshold model from quantitative genetics. *Evolution*, 68(3), 743–759. <https://doi.org/10.1111/evo.12300>
- Revell, L. J. (2021). A variable-rate quantitative trait evolution model using penalized-likelihood. *PeerJ*, 9, e11997. <https://doi.org/10.7717/peerj.11997>
- Revell, L. J. (2024). phytools 2.0: An updated R ecosystem for phylogenetic comparative methods (and other things). *PeerJ*, 12, e16505. <https://doi.org/10.7717/peerj.16505>
- Revell, L. J., & Collar, D. C. (2009). Phylogenetic analysis of the evolutionary correlation using likelihood. *Evolution*, 63(4), 1090–1100. <https://doi.org/10.1111/j.1558-5646.2009.00616.x>
- Revell, L. J., & 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. (2022). *Phylogenetic comparative methods in R*. Princeton University Press.
- Revell, L. J., & Harmon, L. J. A discrete character evolution model for phylogenetic comparative biology with Γ -distributed rate heterogeneity among branches of the tree. bioRxiv 595896. <https://doi.org/10.1101/2024.05.25.595896>, in review.
- Revell, L. J., Mahler, D. L., Peres-Neto, P. R., & Redelings, B. D. (2012). A new method for identifying exceptional phenotypic diversification. *Evolution*, 66(1), 135–146. <https://doi.org/10.1111/j.1558-5646.2011.01435.x>
- Revell, L. J., Toyama, K. S., & Mahler, D. L. (2022). A simple hierarchical model for heterogeneity in the evolutionary correlation a phylogenetic tree. *PeerJ*, 10, e13910. <https://doi.org/10.7717/peerj.13910>
- Schliep, K. P. (2011). phangorn: Phylogenetic analysis in R. *Bioinformatics*, 27(4), 592–593. <https://doi.org/10.1093/bioinformatics/btq706>
- Schluter, D., Price, T., Mooers, A., & Ludwig, D. (1997). Likelihood of ancestor states in adaptive radiation. *Evolution*, 51(6), 1699–1711. <https://doi.org/10.1111/j.1558-5646.1997.tb05095.x>
- Thomas, G. H., Freckleton, R. P., & Székely, T. (2006). Comparative analyses of the influence of developmental mode on phenotypic diversification rates in shorebirds. *Proceedings of the Royal Society of London, Series B: Biological Sciences*, 273(1594), 1619–1624. <https://doi.org/10.1098/rspb.2006.3488>
- Tuffley, C., & Steel, M. A. (1997). Modelling the covarion hypothesis of nucleotide substitution. *Mathematical Biosciences*, 147, 63–91.
- Uyeda, J. C., & 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. <https://doi.org/10.1093/sysbio/syu057>
- Venables, W. N., & Ripley, B. D. (2002). *Modern applied statistics with S* (4th ed.). Springer.
- Wright, S. (1934). An analysis of variability in the number of digits in an inbred strain of guinea pigs. *Genetics*, 19(6), 506–536. <https://doi.org/10.1093/genetics/19.6.506>
- Yang, Z. (1998). Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. *Molecular Biology and Evolution*, 15(5), 568–573. <https://doi.org/10.1093/oxfordjournals.molbev.a025957>
- Yang, Z., & Nielsen, R. (1998). Synonymous and nonsynonymous rate variation in nuclear genes of mammals. *Journal of Molecular Evolution*, 46(4), 409–418. <https://doi.org/10.1007/pl00006320>
- Yang, Z., & Nielsen, R. (2002). Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. *Molecular Biology and Evolution*, 19(6), 908–917. <https://doi.org/10.1093/oxfordjournals.molbev.a004148>
- Zeileis, A., & Hothron, T. (2002). Diagnostic checking in regression relationships. *R-News*, 2, 7–10.