Stochastic Character Mapping of Continuous Traits on Phylogenies

B. S. Martin[1,2,*], and M. G. Weber[2]

[1] *Department of Plant Biology, Ecology, Evolution, and Behavior Program, Michigan State University, East Lansing, MI 48824, USA*

[2] *Department of Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, MI 48109, USA*

[*] *Correspondence to be sent to: bruce.stagg.martin@gmail.com*

Abstract

Living and fossilized organisms represent only a tiny fraction of Earth's evolutionary history, motivating "ancestral state reconstruction" techniques that aim to infer the unobserved phenotypes of evolving lineages based on measurements of their relatives. Stochastic character mapping ("simmapping") methods perform ancestral state reconstruction by randomly sampling maps ("simmaps") of probable phenotypic evolutionary histories along phylogenies, allowing researchers to conveniently and flexibly analyze macroevolutionary patterns and processes while accounting for the inherent uncertainty of ancestral state estimates. Here, we introduce a flexible and efficient algorithm for simmapping continuous phenotypes evolving under Brownian Motion models, which we term continuous simmaps or "contsimmaps", thereby generalizing existing simmapping methods which only work with discrete phenotypes. To demonstrate potential applications of contsimmaps, we develop a pipeline that uses contsimmaps to test for associations between rates of continuous trait evolution and continuously-varying factors (e.g., generation time, climatic niche)–a difficult statistical problem for which few methods are currently available. Through an extensive simulation study, we show that this novel pipeline can accurately and robustly infer factor-rate relationships from phylogenetic comparative data, albeit with low power under certain conditions. Lastly, we apply this pipeline to an empirical dataset, showing that rates of leaf and flower trait evolution are

highly variable yet unrelated to height in a clade of eucalyptus trees spanning roughly 1 to 100 meters in maximum height. Ultimately, contsimmaps provide a valuable new tool for macroevolutionary biology by allowing researchers to more flexibly analyze the evolutionary dynamics of continuous phenotypes and test complex evolutionary hypotheses involving continuous variables.

Key words: Eucalyptus; macroevolution; comparative methods; stochastic character mapping; ancestral state reconstruction; continuous trait evolution; Brownian Motion; rate heterogeneity

Understanding the patterns by which phenotypes evolve over deep time is central for many fundamental questions in evolutionary biology, yet researchers are limited in this endeavor by the necessarily "sparse" distribution of phenotypic measurements across the tree of life. For example, to determine whether temperature is meaningfully correlated with body size variation across a clade, one needs to infer how size evolved in response to temperature fluctuations based on the evolutionary histories of both climatic niche and size in the group. However, such histories are generally impossible to directly observe and must be inferred from living and/or (in lucky cases) fossilized organisms, which inevitably only represent a narrow slice of a clade's entire evolutionary history. Thus, to make robust and accurate inferences regarding how traits are correlated with one another and/or environmental variables across the tree of life, or even how different factors constrain or promote trait diversity, evolutionary biologists require advanced techniques to somehow account for the extensive gaps in our knowledge of Earth's evolutionary past.

Perhaps the oldest and most straightforward approach to "filling in" these gaps is "ancestral state reconstruction" (ASR)–estimating the unobserved phenotypes of (typically extinct) lineages based on observed phenotypes in their relatives (Dobzhansky and Sturtevant, 1938; Sanger et al., 1955; Pauling et al., 1963; Witmer, 1995; Schultz et al.,

1996; Sumrall and Brochu, 2008). With the development of phylogenetic comparative methods that provide a rigorous statistical framework for performing ASR under various trait evolution models, ASR has become ubiquitous in macroevolutionary research (Schluter et al., 1997; Groussin et al., 2016; Joy et al., 2016). One technique for performing ASR, called stochastic character mapping (or "simmapping"), is particularly popular, allowing researchers to randomly sample evolutionary histories (often termed "simmaps") of a character on a phylogeny according to the probability of a given reconstruction under some trait evolution model (Nielsen, 2002; Huelsenbeck et al., 2003; Bollback, 2006). By sampling hundreds or thousands of simmaps, researchers can generate distributions of possible evolutionary histories which may be used in various macroevolutionary analyses–for example, determining the timing/frequency of evolutionary events (e.g., Baliga and Law, 2016; Tornabene et al., 2016; Freyman and Höhna, 2019; Hughes et al., 2021; Landis et al., 2021; Siqueira et al., 2023) or how past life history/environmental factors affected evolution (e.g., de Alencar et al., 2017; Borstein et al., 2019; Burns and Bloom, 2020; Fabre et al., 2020; Rincon-Sandoval et al., 2020; Drury et al., 2021; Nations et al., 2021; Friedman and Muñoz, 2023). In this way, simmaps allow researchers to flexibly conduct analyses over a range of possible histories and effectively account for our inherently incomplete and uncertain knowledge of evolutionary history.

Unfortunately, while simmaps have revolutionized statistical pipelines for studying macroevolution, current simmapping implementations are limited to sampling histories of discrete variables. The lack of simmapping methods for continuous variables constrains approaches for analyzing macroevolutionary data and leaves a conspicuous methodological gap in the field. For example, testing whether the tempo and/or mode of trait evolution varies according to some "explanatory factor" like habitat or diet is generally straight-forward provided the factor's entire evolutionary history is explicitly known (e.g., Revell, 2013a; Clavel et al., 2015; Beaulieu and O'Meara, 2023). When a factor's history is unknown (as is generally the case), researchers are currently unable to generate simmaps of

continuous factor histories to use in their analyses like they would for discrete factors. Consequently, testing how continuously-varying factors like body size (e.g., Friedman et al., 2019), generation time (e.g., Gingerich, 2001), or climatic niche (e.g., Tribble et al., 2023) affect trait evolution processes is much more difficult, often requiring the development of novel methods and/or analysis pipelines tailored for testing specific hypotheses (Cooper and Purvis, 2009; Hansen et al., 2008; Felsenstein, 2012; Baker et al., 2015; Weir and Lawson, 2015; Clavel and Morlon, 2017; Hansen et al., 2022; Boyko et al., 2023a; Tribble et al., 2023; Uyeda et al., 2021).

Here, we introduce an efficient method for simmapping continuous variables under Brownian Motion models of evolution. These continuous simmaps, or "contsimmaps" for short, may be used to directly visualize and analyze the dynamics of phenotypic evolution inferred under trait evolution models, determine the timing and phylogenetic locations of major evolutionary transitions in continuous phenotypes while incorporating uncertainty, and explore how continuous factors may have affected evolutionary processes. First, we present an algorithm for generating contsimmaps by sampling values of continuous variables at arbitrary points on a phylogeny conditional on observed values under a given Brownian Motion model. Second, to showcase potential uses of contsimmaps, we outline a general approach for inferring relationships between aspects of continuous trait evolution processes and continuous factors based on contsimmapped factor histories. Using an extensive simulation study, we verify the proposed approach's ability to accurately infer relationships between continuous factors and rates of trait evolution. We go on to apply this pipeline to an empirical case study, asking whether divergence in plant height is associated with variation in rates of leaf and flower trait evolution in a clade of *Eucalyptus* trees that range from ∼1 to 100 meters tall (Brooker et al., 2015; Thornhill et al., 2019; Falster et al., 2021).

Materials and Methods

We designed contsimmapping to be a flexible framework for sampling evolutionary histories of continuous variables on a phylogeny under Brownian Motion models, analogous to conventional discrete simmapping methods that sample evolutionary histories of discrete variables under continuous time Markov chain models. An important difference between contsimmaps and conventional simmaps is how evolutionary histories are represented. Conventional simmaps assign parts of a phylogeny to regimes representing different values of the discrete variable, providing reconstructed evolutionary histories in continuous time. This is impractical for continuous variables evolving under Brownian Motion models, which have infinitely many possible values and constantly change over time. Accordingly, contsimmaps instead sample values of continuous variables at a finite number of time points evenly distributed across a phylogeny, with the number of time points controlled by a user-specified "resolution". We implemented this framework in an R package called *contsimmap*, which supports contsimmaping under a flexible Brownian Motion modeling framework capable of accommodating multiple correlated variables, multiple measurements per tip and/or internal node with associated intraspecific variation/measurement error, missing measurements, and multiple evolutionary trends/rates that differ according to regimes mapped onto a phylogeny. The R package additionally provides tools for transforming, summarizing, and visualizing contsimmaps (Fig. 1), as well as fitting Brownian Motion models with parameters dependent on contsimmaped variables.

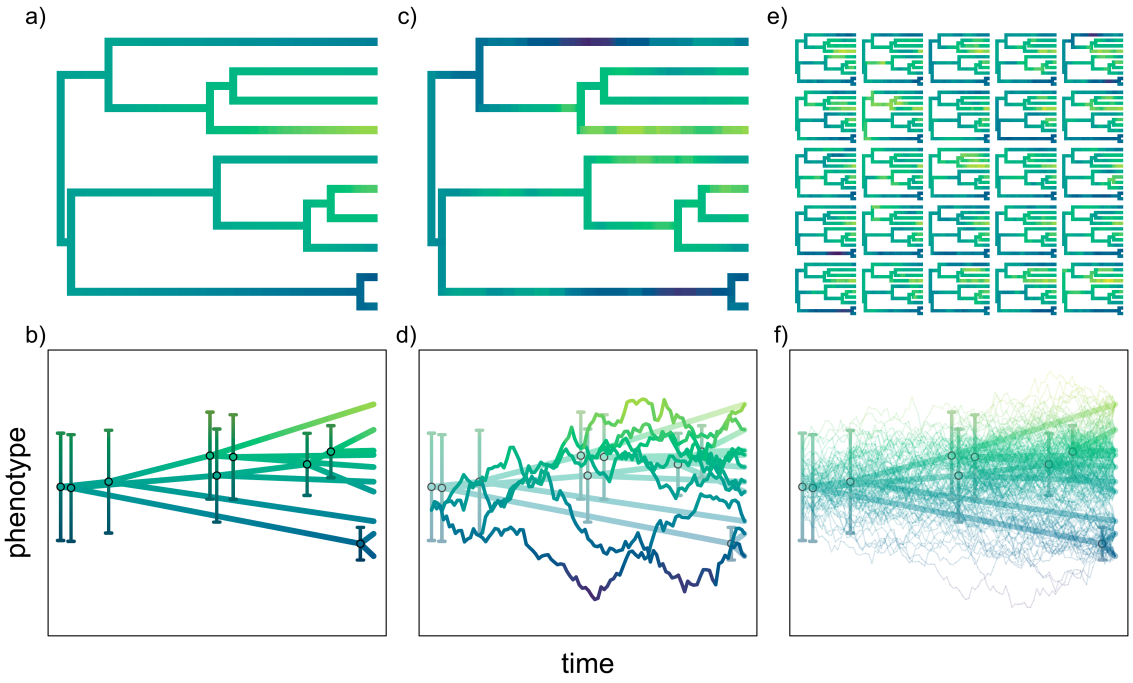

Figure 1. Phylogram and phenogram-based (Revell, 2013b) visualizations of contsimmaps, colored according to reconstructed phenotypic values. Panels a-b: conventional ancestral state reconstructions of a continuously-varying phenotype under a Brownian Motion model. The phylogram in panel a is colored according to mean phenotypic estimates, which are also depicted by the phenogram in panel b along with error bars representing 95% confidence intervals on the estimates at each node. Panels c-d: a single contsimmap generated using the same data and Brownian Motion model, again represented as both a colored phylogram and phenogram in panels c and d, respectively. Panels e-f: a sample of 25 contsimmaps representing the overall distribution of generated contsimmaps also depicted as phylograms and phenograms in panels e and f, respectively. Panels c-f also include conventional ancestral state reconstructions for reference.

Generating Contsimmaps

Here, we present a general algorithm for sampling (multivariate or univariate) phenotypes across a phylogeny under a Brownian Motion model while conditioning on phenotypic measurements at the nodes of the phylogeny. Our algorithm, similarly to other rapid algorithms for ASR under Brownian Motion models (e.g., Goolsby, 2017; Hassler et al., 2022), consists of three main steps: 1) traversing the phylogeny from tips to root (i.e., in postorder), 2) handling the root, and 3) traversing the phylogeny from root to tips (i.e., in preorder). The first step calculates distributions of probable phenotypes at all "critical" time points on the phylogeny (i.e., time points corresponding to either nodes or

shifts between mapped regimes) based *only* on any measurements associated with the descendants of each time point. The second step samples phenotypes from the computed distribution at the root. The third and final step recursively updates phenotypic distributions at each critical time point according to the sampled phenotypes for their ancestors. Additionally, during the third step, phenotypes of lineages between critical time points are interpolated by sampling from Brownian Motion bridges (i.e., Brownian Motion processes conditioned to start and end at fixed values).

We will now establish some notation to more effectively clarify the details of the algorithm. Assume we have e edges in our phylogeny (including a 0-length root edge indexed 0) with s mapped regimes and data on m continuous phenotypic variables. Let τ be an e-length vector of edge lengths assumed to be in units of time. To outline our algorithm, it is useful to focus on the quantities associated with an individual edge denoted i. First, we define two sets of time points along edge i: n_i "interpolant" points at equally-spaced times t_i, excluding times corresponding to i's ancestral/descendant nodes, and n_i^* "critical" points at (strictly increasing) times t_i^*, including times corresponding to i's descendant node (the ancestral node is excluded because it is equivalent to the descendant node of the edge ancestral to i) as well as any regime shifts along i. The number of interpolant points, n_i, is chosen by rounding $\frac{\tau_i \xi}{T} - 1$ to the next largest integer, where T represents the total height of the phylogeny and ξ a user-specified resolution. Specifically, ξ corresponds to approximate number of time points across the height of a phylogeny by defining a "target" time interval for all edges given by $\frac{T}{\xi}$. Now let X_i and X_i^* be $n_i \times m$ and $n_i^* \times m$ matrices, respectively, of phenotypic values at each time point along i. Further, let r_i^* be an n_i^*-length vector of regimes associated with the preceding critical time interval for each entry of t_i^*. Edge i may also have measurements associated with its descendant node: let Y_i be an $o_i \times m$ matrix of these measurements (in the case that there are no measurements, $o_i = 0$).

Our algorithm samples values of X and X^* given the phenotypic

measurements/regime mappings for each edge described above (Y and r, respectively), e $m \times m$ variance-covariance matrices describing error in measurements at descendant nodes for each edge (Γ, commonly termed "node error"), s m-length vectors describing deterministic changes in phenotypes per unit time for each regime (μ, commonly termed "evolutionary trends"), and s $m \times m$ matrices describing stochastic changes in phenotypes per unit time for each regime (Σ, commonly termed "evolutionary rate matrices"). To do this, we additionally keep track of the n_i^* variance-covariance matrices for each edge i, denoted V_i^*, which describe the uncertainty of phenotypic estimates at critical time points. Below, we adopt a general notation of using subscripts to denote edge/regime indices first, followed by time points next (if applicable), and lastly specific phenotypic variables. For example, X_i^* refers to the matrix of phenotypic values at critical time points along the ith edge, $X_{i,j}^*$ to its jth row (corresponding to the jth entry of t_i^* or time $t_{i,j}^*$), and $X_{i,j,k}^*$ to kth value in this row (corresponding to the kth phenotypic variable). Similarly, μ_i would refer to trends in evolution for the ith regime and $\mu_{i,k}$ to the trend for the kth phenotypic variable specifically. We use this notation to describe the three main steps of our algorithm in more detail below:

1) Complete a postorder traversal over all edges. For each edge denoted i with p_i immediate descendant edges:

   1a) Initialize calculations for edge i by defining Z as an $(o_i + p_i) \times m$ matrix and setting the first o_i rows of Z to Y_i. Similarly, define W as $o_i + p_i$ $m \times m$ matrices and set the first o_i matrices to Γ_i. Any missing entries in Z are set to 0, and corresponding diagonal and off-diagonal entries of the associated W matrices are set to ∞ and 0, respectively (Hassler et al., 2022). Accordingly, if edge i is a tip with no associated measurements (i.e., $o_i + p_i = 0$), create a "dummy observation" by setting o_i to 1 and initializing Z as a $1 \times m$ matrix of 0s and W as a single $m \times m$ diagonal matrix with ∞ along the diagonal. Otherwise, if edge i has immediate descendant edges (i.e., not a tip with $p_i > 0$), for each

descendant edge indexed l from 1 to p_i:

$$Z_{o_i+l} = X^*_{d_l,1} - \mu_{r^*_{d_l,1}}(t^*_{d_l,1} - t^*_{i,n^*_i}) \tag{1}$$

$$W_{o_i+l} = V^*_{d_l,1} + \Sigma_{r^*_{d_l,1}}(t^*_{d_l,1} - t^*_{i,n^*_i}) \tag{2}$$

Where Z_l denotes the lth row of Z, W_l denotes the lth matrix of W, and d is a p_i-length vector of indices corresponding to the edges immediately descending from i.

1b) Calculate the uncertainty and mean of the phenotypic values at i's descendant node:

$$V^*_{i,n^*_i} = \left(\sum_{l=1}^{o_i+p_i} W_l^{-1}\right)^{-1} \tag{3}$$

$$X^*_{i,n^*_i} = V^*_{i,n^*_i}\left(\sum_{l=1}^{o_i+p_i} Z_l W_l^{-1}\right) \tag{4}$$

Where W_l^{-1} specifically denotes the "pseudo-inverse" of W_l (Hassler et al., 2022). Notably, Eq. (4) may be undefined when one or more measurements are missing or assumed to be exact, which can cause entries of V^*_{i,n^*_i} and W^{-1}, respectively, to be ∞. The former case can be solved by simply defining $\infty * 0 = 0$ (Hassler et al., 2022). For the latter case, we can generalize Eq. (4) by thinking of the expression as a weighted average of phenotypic value vectors Z_l with weights given by $V^*_{i,n^*_i}W_l^{-1}$. From this perspective, infinite entries along the diagonals of W^{-1} correspond to "infinitely large" weights in this average. Accordingly, we define a "normalization" procedure for matrices in W^{-1}: let $W^{-1}_{.,k,k}$ refer to the diagonal entries in the kth row and column across all W^{-1} matrices. Then, for each phenotypic variable k, check whether $W^{-1}_{.,k,k}$ contains ∞. If it does, set finite and infinite elements of $W^{-1}_{.,k,k}$ to 0 and 1, respectively,

and change corresponding off-diagonal entries to 0. If we denote these normalized matrices $\tilde{W}^{-1}$, we can write a more "robust" version of Eq. (4) which can handle exact trait measurements:

$$X^*_{i,n^*_i} = \left(\sum_{l=1}^{o_i+c_i} \tilde{W}_l^{-1}\right)^{-1} \left(\sum_{l=1}^{o_i+c_i} Z_l \tilde{W}_l^{-1}\right) \tag{5}$$

Intuitively, this means that inexact measurements are "superseded" by exact measurements, which themselves are simply averaged together. We note that this solution is more pragmatic than rigorous when there are two or more exact measurements–paradoxically, averages of multiple exact measurements are still considered exact under this framework. While conceptually problematic, situations where multiple measurements are assumed to be exact may arise in practice due to estimation of 0-length edges during phylogenetic inference and/or simplifying comparative analyses by assuming no measurement error. In any case, the above expression allows our algorithm to yield predictable, defined results in such cases.

1c) If edge i includes any regime shifts, calculate expected phenotypic values and uncertainty at each preceding critical time point. For each critical point j from $n^*_i - 1$ to 1:

$$V^*_{i,j} = V^*_{i,j+1} + \Sigma_{r^*_{i,j+1}}(t^*_{i,j+1} - t^*_{i,j}) \tag{6}$$

$$X^*_{i,j} = X^*_{i,j+1} - \mu_{r^*_{i,j+1}}(t^*_{i,j+1} - t^*_{i,j}) \tag{7}$$

2) Simulate phenotypic values at the root of the phylogeny by sampling new values for $X^*_{0,1}$ from a multivariate normal distribution with mean $X^*_{0,1}$ and variance-covariance matrix $V^*_{0,1}$. This works because all nodes in the phylogeny descend from the root by definition, so $X^*_{0,1}$ and $V^*_{0,1}$ already account for all observed phenotypic data after completing the postorder traversal in step 1.

3) Complete a preorder traversal over all edges except the root ($i = 0$). For each critical time point j along edge i:

3a) Initialize calculations by defining Z as an m-length row vector and W as a single $m \times m$ matrix. Z and W represent counterparts to $X^*_{i,j}$ and $V^*_{i,j}$. While $X^*_{i,j}$ and $V^*_{i,j}$ define the distribution of phenotypic values at the jth critical time point along edge i based *only* on all descendants of edge i, Z and W are based *only* on all *non*-descendants of i. Because this is a preorder traversal, phenotypic values at the immediately previous critical time point have already been simulated and Z and W are given by:

$$Z = X^*_{i,j-1} + \mu_{r^*_{i,j}}(t^*_{i,j} - t^*_{i,j-1}) \tag{8}$$

$$W = \Sigma_{r^*_{i,j}}(t^*_{i,j} - t^*_{i,j-1}) \tag{9}$$

In the case $j - 1 = 0$, we must instead use the trait values and time associated with the last critical time point along the edge directly ancestral to i, denoted a. Specifically, $X^*_{i,j-1}$ and $t^*_{i,j-1}$ are substituted with X^*_{a,n^*_a} and t^*_{a,n^*_a}, respectively, in the expression above.

3b) Update the uncertainty and mean of the phenotypic values at the jth critical time point along edge i:

$$V^*_{i,j} = \left(W^{-1} + {V^*_{i,j}}^{-1}\right)^{-1} \tag{10}$$

$$X^*_{i,j} = \left(\tilde{W}^{-1} + {\tilde{V}^*_{i,j}}^{-1}\right)^{-1} \left(Z\tilde{W}^{-1} + X^*_{i,j}{\tilde{V}^*_{i,j}}^{-1}\right) \tag{11}$$

Where $\tilde{W}^{-1}$ and ${\tilde{V}^*_{i,j}}^{-1}$ again denote normalized versions of these matrices. This normalization procedure is identical to the one used for Eq. (5), with W^{-1} and ${V^*_{i,j}}^{-1}$ treated as a set of two matrices, with one key exception. Notably, W^{-1} will only contain ∞ when one or more phenotypic variables evolve with a rate of

0 (i.e., at least one diagonal entry of $\Sigma_{r^*_{i,j}}$ is 0). To keep resulting contsimmaps consistent with this condition, in the case that $W^{-1}_{k,k}$ and $({V^*_{i,j}}^{-1})_{k,k}$ are *both* ∞, *only* $W^{-1}_{k,k}$ is set to 1 and $({V^*_{i,j}}^{-1})_{k,k}$ is instead set to 0. This critically ensures that phenotypic variables with an evolutionary rate of 0 only exhibit changes over time due to trends defined by μ.

3c) Simulate phenotypic values at the jth critical time point along edge i by replacing $X^*_{i,j}$ with new values sampled from a multivariate normal distribution with mean $X^*_{i,j}$ and variance-covariance matrix $V^*_{i,j}$.

3d) Simulate phenotypic values at all interpolant time points along edge i lying between critical time points $j-1$ and j. For each interpolant time point l for which $t^*_{i,j-1} < t_{i,l} < t^*_{i,j}$ (if $j-1=0$, swap $t^*_{i,j-1}$ for t^*_{a,n^*_a} as above), sample $X_{i,l}$ from a multivariate normal distribution with mean Z and variance-covariance matrix W given by:

$$Z = \frac{(X^*_{i,j} - X_{i,l-1})(t_{i,l} - t_{i,l-1})}{t^*_{i,j} - t_{i,l-1}} + X_{i,l-1} \tag{12}$$

$$W = \frac{(t^*_{i,j} - t_{i,l})(t_{i,l} - t_{i,l-1})}{t^*_{i,j} - t_{i,l-1}} \Sigma_{r^*_{i,j}} \tag{13}$$

In the case $l-1=0$, we must instead use the phenotypic values and time associated with the immediately previous critical time point $j-1$. Specifically, $X_{i,l-1}$ and $t_{i,l-1}$ are swapped with $X^*_{i,j-1}$ and $t^*_{i,j-1}$, respectively, in the expressions above (if $j-1=0$, these are in turn swapped with X^*_{a,n^*_a} and t^*_{a,n^*_a}, respectively, as above).

Using Contsimmaps to Analyze Factor-Dependent Trait Evolution

Simmaps are particularly useful for "sequential inference" pipelines, whereby the impact of some "explanatory factor" on an evolutionary process is inferred by first generating simmaps of factor histories, then fitting factor-dependent evolutionary models

conditional on the simmaps. For conventional discrete simmaps, arbitrary evolutionary models can be rendered factor-dependent by simply allowing the parameters of the evolutionary model–such as evolutionary rates or trends in the case of Brownian Motion models–to vary among lineages in different discrete states (e.g., lineages in mountain versus lowland habitats, annual versus perennial plant lineages; see Revell, 2013a for further description of this approach). For contsimmaps, with an infinite continuum of states, models must instead use "parameter functions" to map factor values to associated parameter values (see FitzJohn, 2010 for an example of this approach). For example, a researcher could test whether increased temperatures are associated with accelerated body size evolution by fitting a Brownian Motion model assuming an exponential relationship, $\sigma^2 = e^{\beta_0+\beta_1 T}$, between contsimmapped thermal niche values (T) and rates of change in body size over time (σ^2). Here, β_0 and β_1 are free parameters inferred by fitting the model to data, with β_1 directly quantifying the relationship between temperature and rates.

We implemented tools in our R package for fitting factor-dependent Brownian Motion models to trait data conditional on samples of contsimmapped factor histories. Specifically, our implementation constructs a likelihood function for a given trait dataset based on a sample of n contsimmaps and user-defined parameter functions that map contsimmapped values to evolutionary rates, trends, and/or node error variances. Given estimates for the free parameters in all parameter functions, the outputted likelihood function automatically transforms contsimmapped values into parameter values and uses the pruning algorithm outlined by Hassler et al., 2022 to calculate likelihoods conditional on each contsimmap. To derive a single overall likelihood, $\tilde{L}$, from the n conditional likelihoods, L, the likelihood function marginalizes over the contsimmaps by assuming either a "flat" or "nuisance" prior. Under the flat prior, each contsimmap is assumed equally likely, with the overall likelihood given by the average of all conditional likelihoods:

$$\tilde{L} = \frac{1}{n}\sum_{i=1}^{n} L_i \tag{14}$$

Under the nuisance prior, the conditional likelihoods for each contsimmap are weighted by the probability they gave rise to the trait data among all other contsimmaps and summed (often termed the "Fitzjohn root prior" in the context of marginalizing over root states; see FitzJohn et al., 2009):

$$\tilde{L} = \frac{1}{\sum\limits_{i=1}^{n} L_i} \sum_{i=1}^{n} L_i^2 \tag{15}$$

Intuitively, the nuisance prior allows the trait data and model to influence which contsimmapped factor histories are considered most likely by taking a weighted average of the conditional likelihoods, with higher weights assigned to contsimmaps that explain the observed data relatively well under a given model. The rough contribution of each contsimmap's conditional likelihood to the overall likelihood, which we term "normalized conditional likelihoods", can be quantified as $L/\sum_{i=1}^{n} L$ and $L^2/\sum_{i=1}^{n} L^2$ in the cases of flat and nuisance priors, respectively.

Notably, the conditional likelihoods under each contsimmap also depend on the trait values at the root of the phylogeny, which are often inferred as additional free parameter in other phylogenetic comparative methods for modeling continuous trait evolution (e.g., Revell, 2012; Pennell et al., 2014; Boucher et al., 2018). However, because conditional likelihoods under any single root trait value will vary drastically across different contsimmaps (e.g., see Figure 4 in Boyko et al., 2023a), we calculate conditional likelihoods for each contsimmap while marginalizing over root trait values by assuming either a flat or nuisance prior over the root trait values as well. Note that the likelihood of observed data under any (potentially multivariate) Brownian Motion model conditional on a vector of root trait values x is given by $r\Phi(x; \hat{x}, V)$, where r is a proportionality constant (called the "remainder" in Hassler et al., 2022), $\hat{x}$ and V denote the expected root trait values and associated uncertainty (in the form of a variance-covariance matrix), and $\Phi(x; \mu, \Sigma)$ represents the probability density function of a multivariate normal distribution with mean μ and variance-covariance matrix Σ evaluated at x. Because the integral of any

probability density function is 1 by definition, the integral of this expression is r–thus, the overall likelihood for a given contsimmap is r under a flat prior. Under a nuisance prior, the overall likelihood is equal to the proportionality constant r multiplied by the integral of the squared multivariate normal probability density function ($r/\sqrt{|V|(4\pi)^k}$, where k denotes the number of traits/dimensions of the multivariate normal distribution). Assuming a nuisance prior will deflate overall likelihoods conditional on any given contsimmap when root trait values are highly uncertain (i.e., $|V|$ is large) and vice versa.

To fit these factor-dependent Brownian Motion models to data, optimization or Bayesian inference algorithms may be used to infer values of the free parameters defining parameter functions. Our R package currently provides tools for using the C++ library NLOPT (Johnson, 2021), interfaced through the R package *nloptr* (Ypma et al., 2022), to find free parameter values that maximize the likelihood of an outputted likelihood function. In general, we found that likelihood surfaces under continuous factor-dependent Brownian Motion models may be quite complex, exhibiting multiple optima and/or relatively flat "ridges" that present challenges for numerical optimization. While our implementation allows users to apply any algorithm available in the NLOPT library and customize optimizer settings as they see fit, we developed a convenient default optimization procedure, consisting of three phases that leverage the complementary strengths/weaknesses of three distinct algorithms. The initial "warmup" phase uses NLOPT's gradient-based truncated Newton algorithm (Dembo and Steihaug, 1983) with preconditioning and random restarts (gradients are stochastically approximated using finite differences; see online appendix for details), which rapidly improves model fit but often gets stuck on suboptimal peaks of the likelihood surface. The subsequent "exploratory" phase uses NLOPT's sbplx (based on subplex; see Rowan, 1990) to search for higher-likelihood regions in the vicinity of this initial estimate. However, because the sbplx algorithm tends to terminate in relatively flat areas and/or saddle points of the likelihood surface rather than true maxima, a final "polish" phase uses NLOPT's principal

axis algorithm (Brent, 2013) to find a local maximum either within or close to the high-likelihood region. Ultimately, this procedure appears to offer a good compromise between robustness and speed, finding and converging on peaks associated with relatively high likelihoods in a practical amount of time. Notably, we found that the algorithms used in the warmup and polish phases do not work well in (and are largely unnecessary for) searching low-dimensional parameter spaces, so our implementation skips these phases by default when fitting models with two parameters or less.

Simulation Study

To assess the performance of our approach for inferring relationships between trait evolution processes and continuous factors, we tested whether our method could reliably detect and quantify factor-dependent rates of continuous trait evolution from simulated data. Broadly speaking, we simulated the evolution of a single continuous trait, Y, with rates depending on a simulated continuous factor, either an "observed" factor X_o or "unobserved" factor X_h, and applied our approach to the simulated data to infer relationships between rates of Y evolution and X_o. In an empirical context, X_h represents an unobserved or "hidden" factor that nonetheless affected rates of trait evolution and may thus mislead hypothesis testing for factor-dependent rates (May and Moore, 2020; see also Beaulieu and O'Meara, 2016; Boyko and Beaulieu, 2023; Boyko et al., 2023b; Tribble et al., 2023). After outlining our simulation procedure below, we outline our analysis procedure, which includes a pragmatic technique for constructing additional null models that help account for hidden factors and thereby mitigate their potential effects on hypothesis testing.

We allowed rates of Y evolution, σ^2, to vary with factor values (denoted X below) according to one of four parameter functions: 1) a "simple" function whereby rates exponentially increase/decrease, 2) a "threshold" function whereby rates shift between some minimum and maximum value, 3) a "sweetspot" function whereby rates peak or dip

around a particular factor value, and 4) a "null" function where rates stay constant. We define the simple function as:

$$\sigma^2 = \mathrm{e}^{\beta_0 + \beta_1 X} \tag{16}$$

Where β_0 and β_1 represent the intercept and slope of the factor-rate relationship on a natural logarithmic scale. This simple function can be applied in scenarios where rates are hypothesized to exponentially increase or decrease in associated with a factor like temperature (Clavel and Morlon, 2017; Slater et al., 2017) or generation time (Gingerich, 2001)). We us an exponential–as opposed to linear–relationship to ensure estimated rates are always positive while also better reflecting empirical findings that evolutionary rates tend to vary multiplicatively rather than additively (Limpert et al., 2001; Gingerich, 2009).

The threshold function is applicable to cases where rates are hypothesized to follow a logistic relationship with another factor. A benefit of the threshold function is that, unlike the simple function, it is bounded between a minimum and maximum rate. Under the simple function, rates of trait evolution are allowed to become arbitrarily close to 0 and/or grow without bound. In reality, various factors like mutation rates, integration among traits, and/or physical limits to trait variation will constrain how slow or fast rates can be. We define the threshold function as:

$$\sigma^2 = \mathrm{e}^{\alpha} \left(\tanh \delta \left(2 \left(1 + \mathrm{e}^{-2\pi\sqrt{3}\frac{X-\theta}{\mathrm{e}^{\omega}}} \right)^{-1} - 1 \right) + 1 \right) \tag{17}$$

Here, α alters the overall scale of rates, specifically corresponding to the natural log of the rate halfway between the minimum and maximum possible rates or "mid rate", while θ determines the factor value or "location" at which rates shift (i.e., the value of X at which $\sigma^2 = \mathrm{e}^{\alpha}$). We denote δ, which controls the direction and magnitude of the rate shift, the "rate deviation". The fold-difference between minimum and maximum rates is explicitly given by $\mathrm{e}^{2|\delta|}$, and positive and negative values of δ yield upward and downward shifts with increasing factor values, respectively. Lastly, ω adjusts the "width" of the shift,

with rates roughly reaching their minimum/maximum values at factor values of $\theta \pm \mathrm{e}^{\omega}/2$.

The sweetspot function can be used when rates are hypothesized to peak or dip at intermediate factor values (Cooper and Purvis, 2009; FitzJohn et al., 2009; Feldman et al., 2016; Amado et al., 2021). For example, Cooper and Purvis (2009) showed that rates of cranial shape evolution are slowest among Paltyrrhine monkeys and Palangeriforme possums of intermediate size, increasing among both smaller and larger-bodied lineages. Under the sweetspot function, factor-rate relationships follow a Gaussian curve of arbitrary height and orientation that only takes strictly positive values:

$$\sigma^2 = \mathrm{e}^{\alpha}\left(\tanh\delta\left(2\mathrm{e}^{-18\left(\frac{X-\theta}{\mathrm{e}^{\omega}}\right)^2} - 1\right) + 1\right) \tag{18}$$

Where α, θ, δ, and ω largely have the same effects and interpretations as they do for the threshold function. Under the sweetspot function, however, rates will peak to their maximum at θ and roughly reach their minimum at factor values of $\theta \pm \mathrm{e}^{\omega}/2$ if δ is positive. Conversely, if δ is negative, rates will instead dip to their minimum at θ and roughly reach their maximum at $\theta \pm \mathrm{e}^{\omega}/2$. Lastly, we define the null parameter function, whereby rates stay constant regardless of factor variation, as $\sigma^2 = \mathrm{e}^{\alpha}$ for consistency with the α parameter under threshold and sweetspot functions.

For each simulation, we used *phytools* (Revell, 2012) to simulate an ultrametric, pure-birth phylogeny of height 1 with either 50, 100, or 200 tips. To simulate the continuous factors X_o and X_h, we generated two densely-sampled continuous factor histories (with a resolution or ξ value of 500) evolving under Brownian Motion processes with root trait values of 0, no trends, and constant rates of 4. For the trait Y, we simulated an additional Brownian Motion process with starting value 0 and no trends, but with rates varying according one of 12 possible factor-rate relationships which differed both in the overall magnitude (i.e., "strength") of rate variation as well as how gradually rates changed with respect to factor values (i.e., "width"). Specifically, we used one null parameter function; six "strong" versions of the simple, threshold, or sweetspot functions

depending on either X_o or X_h whereby rates varied $\sim$ 20-fold between factor values of -2 and 2; three "weak" versions of the simple, threshold, or sweetspot functions depending on X_o only whereby rates only varied $\sim$ 5-fold on the same interval; and two "wide" versions of the threshold or sweetspot functions depending on X_o only whereby rates again varied $\sim$ 20-fold but between factor values of -4 and 4. See Table S1 for the specific parameter values used.

Notably, our approach assumes a given factor perfectly predicts rates of trait evolution, which is unlikely for empirical data. To test the robustness of our approach to imperfect correspondences between factors and rates, we additionally modified some simulations by adding "noise" to factor-rate relationships. We added noise by multiplying rate values at each time point with a random variable sampled from a gamma distribution with shape and rate $\frac{t}{\nu}$, where t is the length of the time interval preceding a time point. These multipliers represent the average value of a white noise gamma process with mean 1 and unit variance ν over a time interval of length t. By multiplying the rates over each time increment with these random variables, the simulated trait evolution process is transformed from Brownian Motion to variance-gamma (a type of Lévy or "pulsed" trait evolution process explored in some prior work; see Landis et al., 2013; Landis and Schraiber, 2017). This procedure ensures that random noise around rates tends to "average out" over sufficiently long periods of time, such that rates converge to what would be expected under the simulated factor-rate relationship given enough data. Thus, this kind of rate noise only weakens the factor-rate relationship rather than completely altering it. For our simulations, we set ν to 0.05, corresponding to rates ranging between roughly 10%-300% their expected value over a time interval spanning a tenth of the phylogeny's height, 40%-200% for a third of the height, and 60%-150% for the entire height.

Ultimately, the three phylogeny sizes (50, 100, or 200 tips), 12 possible factor-rate relationships, and presence/absence of noise around factor-rate relationships yield $3 \times 12 \times 2 = 72$ simulation conditions. We repeated the simulation process 100 times for

each condition, yielding a grand total of 7,200 simulations. For the analysis procedure, we retained the phylogeny and tip factor/trait values of X_o and Y for each simulation (assuming no node error in X_o/Y values to render the simulation study more manageable). We analyzed the factor/trait datasets by first fitting Brownian Motion models to the X_o data (assuming no node error or evolutionary trends) and using the fitted model to generate 100 contsimmaps of X_o with a resolution or ξ value of 100. We then fit four Brownian Motion models to the Y data conditioned on these contsimmaps (again assuming no node error or evolutionary trends)–three non-null models assuming rates depend on mapped X_o values through either a simple, threshold, or sweetspot function with unknown free parameters, plus a null model assuming an unknown, constant rate (given by the mid rate/α parameter). Trait data simulated with rates varying due to noise and/or the hidden factor, X_h, are likely to exhibit a poor fit to the null model and provide spurious support for non-null models that at least allow rates to vary across the phylogeny (May and Moore, 2020; see also Beaulieu and O'Meara, 2016; Boyko and Beaulieu, 2023; Boyko et al., 2023b; Tribble et al., 2023). Thus, we fit three additional null Brownian Motion models to the Y data assuming rates depend on a mapped "dummy factor" D, which is simulated under the Brownian Motion model fitted to the X_o data, through either a simple, threshold, or sweetspot relationship. Because D exhibits the same evolutionary dynamics as X_o but with random tip values, support for D-dependent over X_o-dependent models provide evidence that rates of Y evolution vary, but in a way not necessarily related to the observed factor X_o (see also Tribble et al., 2023, which employs a similar approach in testing for relationships between discrete trait evolution and continuous factors).

We fit all Brownian Motion models to simulated data by running the default optimization procedure (see previous section) 10 times from initial parameter values sampled from a uniform distribution spanning from -5 to 5 and retaining the inferred parameter values from whichever optimization run found the highest likelihood. To save on the time needed to run the simulation study, we only allowed the warmup and

exploratory/polish phases to run for a maximum of 1,000 and 10,000 iterations, respectively. We always assumed flat priors over root trait values. On the other hand, we assumed nuisance priors over contsimmaps for D-dependent models and flat priors for all other models. Intuitively, this forces non-null models to "fairly" integrate over probable histories of observed factors while allowing null models to assign higher weights to simulated dummy factor contsimmaps which are particularly likely to explain the trait data. Preliminary tests of our approach showed that this "mixed prior" technique renders null models more competitive with non-null models and greatly reduces false positive errors at a modest cost to statistical power. Preliminary tests also demonstrated that the default optimization procedure sometimes gets stuck in likelihood ridges when fitting threshold and sweetspot function-based models, which may collapse to effectively constant rate models as the location (θ) and/or width (ω) parameters become arbitrarily small/large. To prevent optimization algorithms from spending too much time exploring these regions of parameter space, we imposed boundaries on these parameters by defining a "factor interval" spanning from $\min X - (\max X - \min X)/2$ to $\max X + (\max X - \min X)/2$, where X represents whatever factor a given parameter function depends on (either the observed factor X_o or dummy factor D depending on the model). We constrained the location of inferred shifts and peaks/dips to lie within the factor interval and the corresponding widths (given by e^{ω}) to be between $1/100^{\mathrm{th}}$ and 5 times the range of the factor interval. We also limited the rate deviation (δ) parameter to be between -10 and 10, as rate deviations with an absolute value of 10 or greater cause minimum and maximum rates to effectively equal 0 and $2\mathrm{e}^{\alpha}$ (i.e., the maximum allowable rate for a given mid rate or α value), respectively.

For each simulation, we selected the model with lowest Bayesian Information Criterion (BIC; see Schwarz, 1978; Dziak et al., 2020) as the best-fitting model. We chose to use BIC over the more widely used small sample size corrected Akaike Information Criterion (AICc) after preliminary investigations demonstrated that AICc-based model selection exhibits somewhat elevated error rates among simulations with noisy rates on

large phylogenies (Table S2). BIC penalizes model complexity more harshly than AICc–particularly for large sample sizes–and substantially reduced the error rates of our method at little cost to overall power and accuracy. We calculated error ("false positive") and power ("true positive") rates as the percent of simulations for which X_o-dependent models were selected as the best-fitting model for simulations with constant/X_h-dependent rates and X_o-dependent rates, respectively. Among the simulations yielding true positive results, we additionally calculated "differentiation rates" as the percent for which the best-fitting model assumed the same X_o-rate relationship used to simulate the data. To further explore how accurately our approach can estimate the relationship between X_o and rates of Y evolution, we also calculated BIC weights and generated model-averaged rate estimates across values of X_o for all simulations with either constant or X_o-dependent rates, ignoring simulations with X_h-dependent rates because they lack a straight-forward expected X_o-rate relationship to compare with rate estimates. We measured the overall accuracy, bias, and precision of model-averaged rates estimates across values of X_o by calculating the inverse of the median absolute fold difference between estimated and simulated rates (i.e., the median absolute difference between estimated and simulated rates on the log scale, negated such that higher values correspond to increased accuracy), the percent of estimated rates greater than simulated rates, and the fold difference between the 2.5% and 97.5% quantiles of estimated rates, respectively.

Empirical Example

We applied our approach for inferring continuous factor-dependent rates of trait evolution to test whether rates of phenotypic evolution are associated with size in the *Eucalyptus* subgenus *Eucalyptus*, sometimes called "Monocalypts", a clade of ∼124 Australian woody plant species ranging from shrubby "mallees" only reaching a few meters in height (e.g., *E. acies* Brooker, *E. cunninghamii* Sweet, *E. erectifolia* Brooker & Hopper) to gargantuan trees growing to nearly 100 meters in height (e.g., *E. obliqua* L'Hér, *E.*

regnans F.Muell.) (Ladiges et al., 2010; Brooker et al., 2015; Thornhill et al., 2019; Falster et al., 2021; Nicolle, 2022). Body size tends to be associated with many aspects of life history, as larger organisms generally live in smaller populations with slower generational turnover (Calder, 2001; Niklas and Enquist, 2001; White et al., 2007; Sibly and Brown, 2007; Adler et al., 2014; Salguero-Gómez et al., 2016; Bakewell et al., 2020), leading to the hypothesis that evolutionary rates are slower among lineages of larger organisms compared to their smaller relatives. Nonetheless, it remains unclear how rates of phenotypic evolution generally scale with size in many groups, particularly plants (Cooper and Purvis, 2009; Baker et al., 2015; Chira et al., 2018; Friedman et al., 2019; Zimova et al., 2023; see also Lanfear et al., 2013). We obtained a phylogeny of 108 Monocalypt species by pruning the clade out from a recently inferred, time-calibrated phylogeny of over 700 species in *Eucalyptus* and related genera (the "Maximum Likelihood 1" tree in Thornhill et al., 2019). We used the online *Eucalyptus* encyclopedia EUCLID (Brooker et al., 2015) and the AusTraits database (Falster et al., 2021) to gather data on 10 continuous traits for these species, which we categorized into four modules: 1) juvenile leaves consisting of juvenile leaf widths/lengths, 2) adult leaves consisting of petiole lengths and adult leaf widths/lengths, 3) inflorescences consisting peduncle and inflorescence pedicel lengths, and 4) flowers consisting of bud, fruit, and seed lengths. We used the same sources to also aggregate data on height for these species, which acts as a proxy for size and plays the role of a factor in our analyses. All of these traits exhibit substantial within-species variation. Accordingly, trait values for a given species were generally reported as ranges rather than average values, though only maximum values were reported for some 0-29 species depending on the trait (e.g., "Grows up to 8 meters in height"). The only average trait values reported were bud lengths for 13 species. Ultimately, for each trait, we obtained data for between 102 and 105 out of the 108 species in the phylogeny, with full value ranges (i.e., both minimum and maximum values) known for 73-105 species.

We assumed trait values were log-normally distributed within species with provided

minimum and maximum trait values corresponding to the 2.5% and 97.5% quantiles of this distribution, such that averages and associated intraspecific variances (on the natural log scale) are given by the midpoints and $1/4^2$ the squared differences, respectively, of log-transformed minimum and maximum trait values. Notably, some value ranges for petiole, peduncle, and pedicel lengths had maximum and/or minimum values of 0 in some species, which is undefined on a log scale. Because these traits were measured to a precision of 1 mm (Brooker et al., 2015), we rounded up maximum trait values of 0 to 0.05 mm (i.e., the average assuming a flat prior over measurements from 0 to 1 mm). For minimum trait values of 0, we either rounded them up to 0.05 mm if paired with a non-zero maximum value or treated them as missing if paired with a maximum value also equal to 0. For all cases where only the maximum trait value for a given species was known, we developed a crude but simple procedure for imputing minimum trait values to mitigate bias resulting from comparing average trait values in some species to maximum trait values in others. Specifically, for each trait, we regressed log-transformed minimum trait values on maximum values for all species with fully-known value ranges, and used the fitted relationship to predict minimum trait values for species with known maximum values only. For species with only average bud length measurements available, we simply log-transformed the given averages and treated associated intraspecifc variance as an unknown parameter to be estimated during model fitting.

Ultimately, we generated 349 contsimmaps (resolution/$\xi = 100$) of Monocalypt height under an evolving rates or "evorates" model whereby rates of Monocalypt height evolution were allowed to gradually change over time and across lineages (Martin et al., 2023; see online appendix for further details). We elected to use this relatively complicated model of height evolution as a constant-rate Brownian Motion model exhibited a rather poor fit to our Monocalypt height data by comparison, reflecting substantial variation in rates of height evolution among Monocalypts. We then generated 349 corresponding dummy factor contsimmaps by simulating trait evolution with starting values and rates

identical to those of the height contsimmaps. Using these contsimmaps, we fit seven multivariate Brownian Motion models to each trait module–six models assuming rates depend either on height or the dummy factor through a simple, threshold, or sweetspot relationship, plus a null model assuming constant rates. All models allowed each trait to have independent mid rate (α) or intercept (β_0) parameters to account for overall differences in rates among traits within a given module, though we assumed all other free parameters were constant across traits for statistical tractability (i.e., a single parameter function "shape" for each module which is rescaled among different traits). We fixed node error variances based on calculated intraspecific variances (excepting bud lengths in some species for which only average trait values were known). Because our data were largely derived from minimum-maximum ranges within species, we lacked any meaningful signal of trait correlations within species and thus set all intraspecific correlations among traits to 0 in our models, though we did estimate evolutionary correlations among traits. We used the correlation matrix transform outlined in Lewandowski et al., 2009 and Stan Development Team, 2019 to avoid issues with exploring correlation parameters that form invalid correlation matrices. To fit these models, we used the same optimization procedure and parameter boundaries which were used in the simulation study, but ran the optimization procedure 20 rather than 10 times.

Results

Simulation Study

To illustrate potential uses of contsimmaps, we performed an extensive simulation study to investigate whether contsimmaps of continuous variables can be used to infer relationships between rates of continuous trait evolution and continuously-varying factors, in analogy with popular approaches for inferring relationships between rates and discretely-varying factors (e.g., Revell, 2013a). Overall, our contsimmap-based pipeline

exhibited appropriate error and modest power rates when applied to simulated data (Table 1; Fig. 2). Across all simulation conditions, error rates remained consistent and conservative at around 2-4%, only exceeding 5% in 2 out of the 24 null simulation conditions. On the other hand, power rates varied dramatically across simulation conditions, ranging from only ∼5-25% for simulations exhibiting subtler patterns of rate variation (i.e., weak and wide factor-rate relationships) on phylogenies with 50 tips, to ∼75-90% and ∼90-100% for those exhibiting strong rate variation on phylogenies with 100 and 200 tips, respectively. Unsurprisingly, higher power rates were generally associated with more accurate, unbiased, and precise model-averaged estimates of factor-rate relationships, while lower power rates were associated with increased bias towards inference of overly conservative and "flat" relationships (Fig. 3; Figs. S1-3). Differentiation rates–that is, the ability to distinguish among different kinds of factor-rate relationships–largely depended on the kind of factor-rate relationship used to simulate data. Specifically, our method could easily dinstiguish sweetspot (i.e., Gaussian) and especially simple (i.e., exponential) factor-rate relationships from other kinds of relationships, yet often mistook threshold (i.e., logistic) relationships for either simple or sweetspot ones.

Random variation around simulated factor-rate relationships, or "noise", had minor yet noticeable effects on our method's performance. Specifically, noise was associated with slightly increased error rates, albeit inconsistently, as well as ∼5-15% decreases to power rates across all simulation conditions. In terms of differentiation rates, noise appeared to increase support for more complex models assuming threshold or sweetspot factor-rate relationships at the expense of those assuming simple relationships. Overall, however, phylogeny size and/or the factor-relationship used to simulate data had a much stronger effect on both power and differentiation rates compared to noise. Additionally, low error rates across the board indicate that our method is fairly robust to random rate variation due to noise and/or unobserved factors. Interestingly, noise generally had the least severe effects on our method's performance for simulations on both small and large phylogenies.

This pattern is particularly apparent in the estimated factor-rate relationships for simulations with strong relationships in Fig. 3. While noise had virtually no effect on estimated relationships for simulations on phylogenies with 100 and 200 tips, it tended to conservatively bias estimated relationships for simulations with 100 tips. Roughly speaking, we hypothesize that simulations on smaller phylogenies often failed to yield enough signal of factor-dependent rates to support factor-dependent models regardless of noise, while simulations on larger phylogenies yielded enough signal to "overcome" the effect of noise on inference. Indeed, for simulation conditions with weak or wide factor-rate relationships–which generated weaker signals of factor-dependent rates–noise's effect on estimated relationships was most pronounced for phylogenies with 200 rather than 100 tips, presumably reflecting greater amount of data needed to reliably infer subtler factor-rate relationships.

Table 1. *Proportions of times the contsimmap-based pipeline for detecting continuous factor-dependent rates of trait evolution selected a model assuming a given kind of factor-rate relationship (i.e., simple, threshold, or sweetspot) as the best-fitting one (based on having the lowest Bayesian Information Criterion) across all simulation conditions. Proportions for simulations with and without random variation around factor-rate relationships ("noise") are given within and without parentheses, respectively. All models assuming either constant or dummy factor-dependent rates were considered "null" models.*

	constant	hidden factor-dependent			simple		threshold			sweetspot		
		simple	threshold	sweetspot	strong	weak	strong	weak	wide	strong	weak	wide
						50 tips						
null	0.98 (0.99)	1.00 (0.99)	0.99 (0.97)	0.98 (0.99)	0.51 (0.71)	0.80 (0.87)	0.60 (0.62)	0.83 (0.92)	0.68 (0.83)	0.74 (0.76)	0.94 (0.95)	0.90 (0.82)
simple	0.00 (0.00)	0.00 (0.00)	0.01 (0.01)	0.00 (0.00)	0.48 (0.26)	0.18 (0.11)	0.34 (0.24)	0.15 (0.07)	0.29 (0.12)	0.06 (0.05)	0.02 (0.01)	0.06 (0.08)
threshold	0.02 (0.01)	0.00 (0.01)	0.00 (0.02)	0.00 (0.01)	0.01 (0.02)	0.01 (0.01)	0.02 (0.06)	0.01 (0.00)	0.02 (0.01)	0.02 (0.02)	0.00 (0.01)	0.00 (0.01)
sweetspot	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.02 (0.00)	0.00 (0.01)	0.01 (0.01)	0.04 (0.08)	0.01 (0.01)	0.01 (0.04)	0.18 (0.17)	0.04 (0.03)	0.04 (0.09)
						100 tips						
null	0.98 (0.96)	0.98 (0.96)	1.00 (0.98)	0.96 (0.94)	0.11 (0.24)	0.56 (0.64)	0.19 (0.33)	0.60 (0.69)	0.43 (0.48)	0.10 (0.25)	0.68 (0.73)	0.60 (0.67)
simple	0.02 (0.00)	0.01 (0.00)	0.00 (0.00)	0.00 (0.01)	0.87 (0.71)	0.44 (0.27)	0.49 (0.24)	0.34 (0.15)	0.51 (0.38)	0.04 (0.00)	0.10 (0.05)	0.13 (0.07)
threshold	0.00 (0.02)	0.00 (0.02)	0.00 (0.01)	0.01 (0.02)	0.02 (0.00)	0.00 (0.03)	0.13 (0.21)	0.02 (0.09)	0.01 (0.06)	0.09 (0.07)	0.00 (0.03)	0.04 (0.06)
sweetspot	0.00 (0.02)	0.01 (0.02)	0.00 (0.01)	0.03 (0.03)	0.00 (0.05)	0.00 (0.06)	0.19 (0.22)	0.04 (0.07)	0.05 (0.08)	0.77 (0.68)	0.22 (0.19)	0.23 (0.20)
						200 tips						
null	0.98 (0.93)	0.96 (0.97)	0.98 (0.96)	0.96 (0.98)	0.00 (0.07)	0.13 (0.37)	0.07 (0.06)	0.16 (0.34)	0.10 (0.23)	0.00 (0.02)	0.29 (0.44)	0.21 (0.22)
simple	0.02 (0.00)	0.01 (0.00)	0.01 (0.02)	0.01 (0.00)	0.99 (0.86)	0.87 (0.55)	0.17 (0.17)	0.57 (0.36)	0.65 (0.51)	0.00 (0.00)	0.07 (0.04)	0.08 (0.05)
threshold	0.00 (0.04)	0.00 (0.02)	0.00 (0.00)	0.01 (0.00)	0.00 (0.05)	0.00 (0.04)	0.39 (0.44)	0.09 (0.11)	0.13 (0.14)	0.04 (0.07)	0.01 (0.02)	0.08 (0.13)
sweetspot	0.00 (0.03)	0.03 (0.01)	0.01 (0.02)	0.02 (0.02)	0.01 (0.02)	0.00 (0.04)	0.37 (0.33)	0.18 (0.19)	0.12 (0.12)	0.96 (0.91)	0.63 (0.50)	0.63 (0.60)

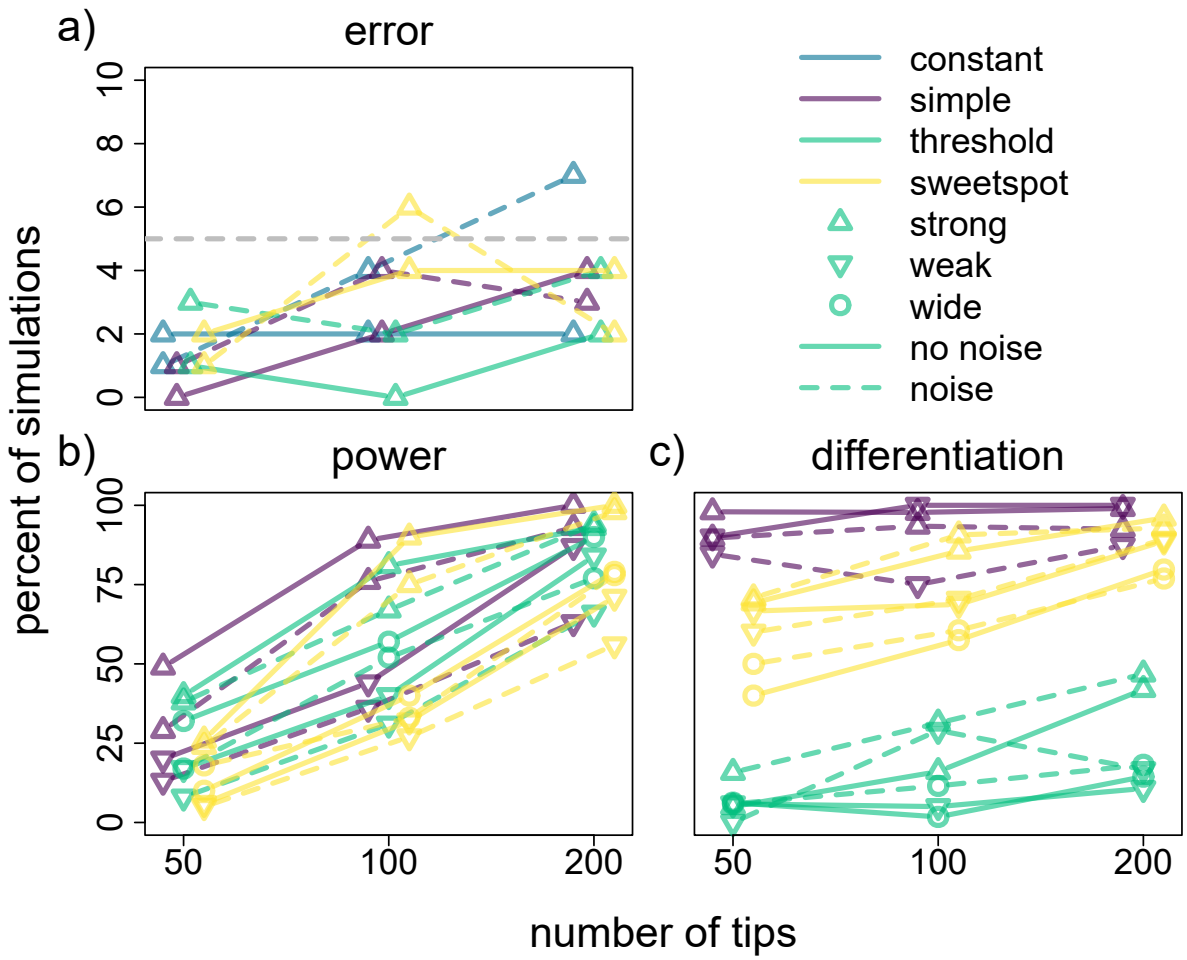

Figure 2. Error, power, and differentiation rates of the contsimmap-based pipeline for detecting continuous factor-dependent rates of trait evolution. Different colors correspond to simulations with different factor-rate relationships; different symbols to simulations with differing relationship strength and, in the case of threshold and sweetspot models, width; and dashed versus solid lines to simulations with versus without random variation in rates ("noise") around simulated factor-rate relationships. Panel a: percent of simulations with either constant or hidden factor-dependent rates for which the best-fitting model (i.e., lowest Bayesian Information Criterion) was an observed factor-dependent model (i.e., error rates). Panel b: percent of simulations with observed factor-dependent rates for which the best-fitting model was also an observed factor-dependent model (i.e., power rates). Panel c: percent of observed-factor dependent simulations supporting models with observed factor-dependent rates for which the best fitting model assumed the very same kind factor-rate relationship (i.e., simple, threshold, or sweetspot) used to simulate the data (i.e., differentiation rates).

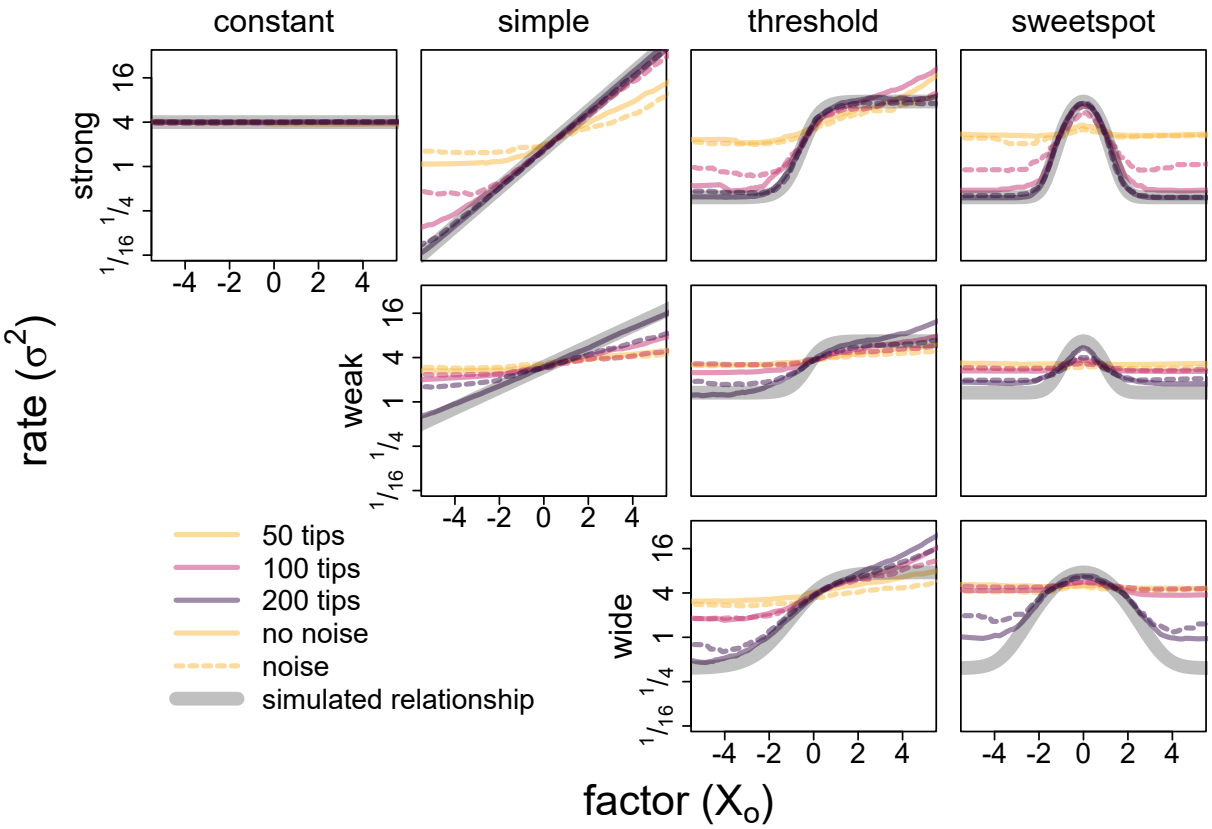

Figure 3. Mean model-averaged factor-rate relationships inferred using the contsimmap-based pipeline for detecting continuous factor-dependent rates of trait evolution in comparison to simulated factor-rate relationships under all simulation conditions with either constant or observed factor-dependent rates. Different colors correspond to different phylogeny sizes (i.e., number of tips) and dashed versus solid lines to simulation conditions with versus without random variation in rates ("noise") around simulated factor-rate relationships. Simulated factor-rate relationships are represented by thick gray lines for reference.

Empirical Example

In addition to the simulation study, we also applied our contsimmap-based pipeline for inferring factor-dependent rates of continuous trait evolution to test whether rates of leaf and flower trait evolution are associated with height variation in the *Eucalyptus* subgenus *Eucalyptus*, or Monocalypts. Ultimately, we found no evidence of height-rate associations for any of the trait modules investigated, instead finding overwhelming support for associations between rates and simulated dummy factors. These results strongly suggest rates of leaf and flower trait evolution in Monocalypts are variable but not

closely related to height (Fig. 4). To further investigate these results, we calculated marginal rate estimates for each trait module across the Monocalypt phylogeny by first taking the weighted average of estimated rates across all contsimmaps under each fitted model, with weights given by the normalized conditional likelihoods for each contsimmap (i.e., prior-weighted conditional likelihoods for each contsimmap rescaled to have a sum to 1), then model-averaging the resulting rate estimates across all models using BIC weights. Notably, marginal rate variation patterns for each trait module appear similar to those for height evolution as inferred under an evorates model (see online appendix for further details on this model)–particularly for juvenile leaves and flowers–suggesting that overall rates of phenotypic evolution among Monocalypts vary according to some common yet currently unknown factor.

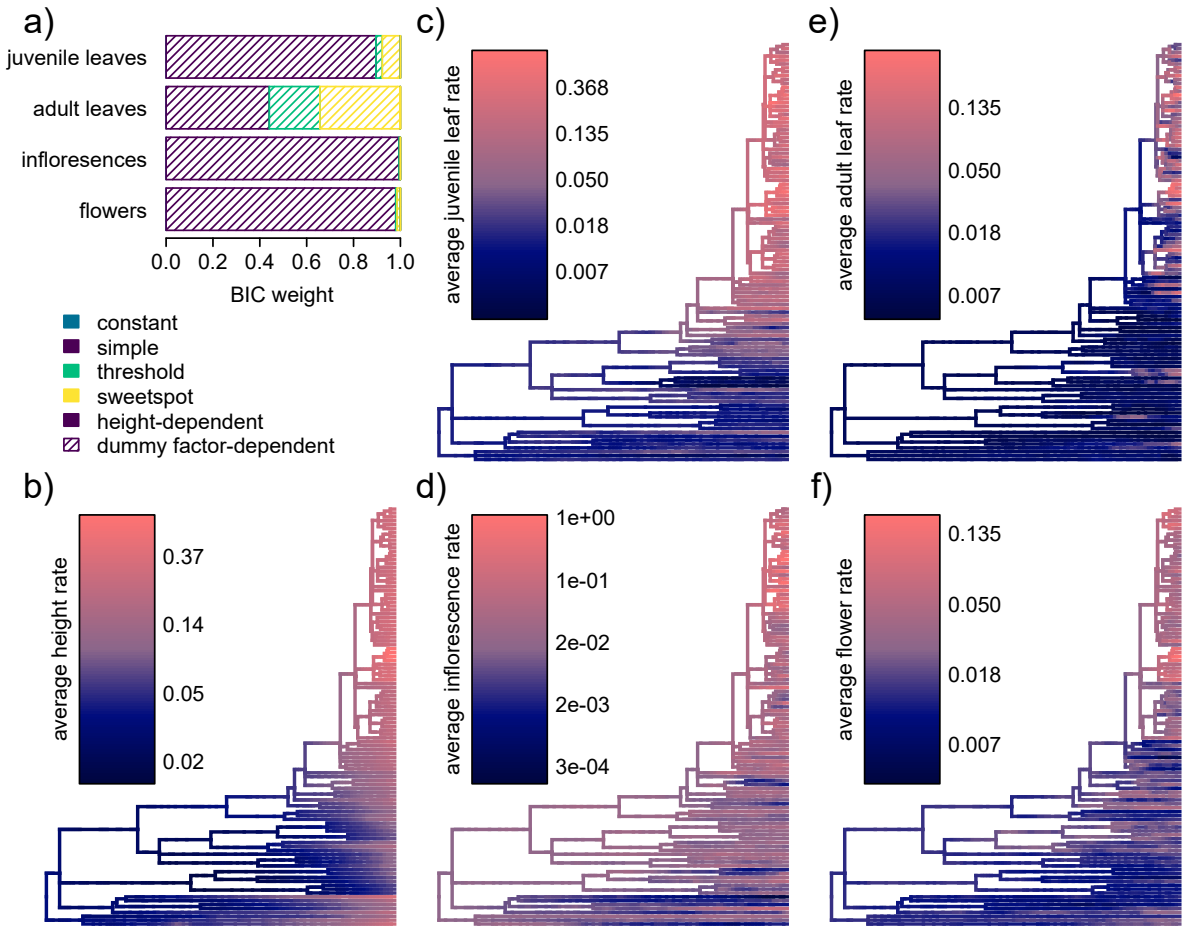

Figure 4. Variation in rates of leaf and flower trait evolution in the *Eucalyptus* subgenus *Eucalyptus* ("Monocalypts"). Panel a: bar graph displaying Bayesian Information Criterion weights for the contsimmap-based models fit to each trait module. Different colors correspond to different factor-rate relationships, while solid and hatched bars correspond to models assuming rates depend on height and simulated dummy factors, respectively. Panel b: Phylogram colored according to the average of 349 rate contsimmaps sampled from the posterior of an evolving rates model fit to the Monocalypt height data, with dark blue and light red corresponding to slow and fast rates, respectively. Panels c-f: phylograms depicting marginal rates estimated under contsimmap-based models for each trait module, with dark blue and light red once again corresponding to slow and fast rates, respectively.

Discussion

Here, we introduced contsimmapping, a flexible and efficient method for sampling histories of continuous variables on phylogenies under Brownian Motion models of trait evolution. Contsimmapping provides an intuitive framework for reconstructing and

analyzing the evolution of continuous phenotypes while accounting for our inherently incomplete knowledge of the evolutionary past. Further, like conventional simmapping for discrete variables, contsimmapping provides a way to generate distributions of probable evolutionary histories for some variable, enabling straight-forward development of rigorous statistical pipelines for determining how evolutionary processes are affected by continuous variables like body size (e.g., Friedman et al., 2019), generation time (e.g., Gingerich, 2001), or climatic niche (e.g., Tribble et al., 2023). To this end, we implemented a contsimmap-based pipeline capable of robustly detecting and accurately quantifying relationships between continuously-varying factors and rates of trait evolution. Using this novel approach, we show that rates of leaf and flower trait evolution are generally unrelated to height in a clade of *Eucalyptus* trees spanning nearly two orders of magnitude in height variation (Brooker et al., 2015; Thornhill et al., 2019; Falster et al., 2021). Nonetheless, our analysis uncovered striking similarities in patterns of evolutionary rate variation across different *Eucalyptus* traits (Figure 4), demonstrating how contsimmaps provide a powerful toolkit for flexibly interrogating the evolutionary dynamics of continuous traits.

Our implementation of contsimmapping is already quite comprehensive, allowing for contsimmapping of multiple correlated continuous variables while incorporating within-species variation/measurement error and even regime-dependent rates and trends. Additionally, the *contsimmap* R package provides a variety of tools for transforming, summarizing, and visualizing contsimmaps, which may be used to conveniently explore patterns in phylogenetic comparative data on continuous traits while accounting for uncertainty as well as produce publication-quality figures (e.g., Figs. 1,4; see also Figs. S6,8 in online appendix). While we largely focused on using contsimmaps to infer associations between rates of continuous trait evolution and continuously-varying factors in the current study, our pipeline could be easily adapted to infer factor-dependent evolutionary trends and node error variances using tools already available in our R package. Looking forward, contsimmapping could be extended to sample evolutionary histories under other popular

models of continuous trait evolution such as Ornstein-Uhlenbeck and Lévy processes, which are generally interpreted as models of adaptive and pulsed evolution, respectively. Another promising future direction would be the development of methods for fitting Ornstein-Uhlenbeck models, discrete trait evolution models, or lineage diversification models with factor-dependent parameters conditional on contsimmaps. Such methods would further enhance macroevolutionary biologists' ability to explore and answer questions regarding the interplay of evolutionary processes with continuous variables.

Using contsimmaps to infer associations between continuous variables and rates of trait evolution

Overall, our contsimmap-based pipeline provides an effective and practical way to detect and quantify relationships among rates of trait evolution and continuous variables–at least given a sufficiently large effect size and phylogeny. Our pipeline is not the first statistical method developed for inferring relationships between rates of trait evolution and continuous factors, though it is currently the most flexible and thoroughly-tested one to our knowledge. Indeed, many previous methods were developed for particular empirical studies to circumvent the lack of a standard, general-purpose approach. Accordingly, the general statistical performance of these methods often remains largely unknown (e.g., Cooper and Purvis, 2009; Uyeda et al., 2021). The *evorag* method developed by Weir and Lawson, 2015 constitutes a notable exception as a well-studied method suitable for a variety of purposes, though it only uses sister pair contrasts to infer factor-rate relationships. By using entire phylogenies, our method theoretically more fully utilizes the information available in phylogenetic comparative data to infer factor-rate relationships. More recently, Hansen et al., 2022 developed an elegant regression-based approach for inferring correlations between rates and continuous variables, though this method is only capable of inferring linear or exponential relationships between a single continuous factor and rates for a single continuous trait (or node error variances, which are

interpreted as "microevolutionary rates" by the authors). In comparison to this method, our contsimmap-based pipeline sacrifices computational simplicity for flexibility, allowing researchers to incorporate explicit reconstructions of discrete and multiple continuous factors into analyses, fit multivariate models of continuous trait evolution, and/or simultaneously link multiple trait evolution process parameters (i.e., rates, node error variances, and/or evolutionary trends) with reconstructed factors via arbitrary parameter functions (computational/statistical tractability notwithstanding). Overall, our new pipeline is a substantial generalization of previous methods, greatly expanding the hypothesis-testing capabilities of macroevolutionary biologists and allowing researchers to shed light on many famous long-standing questions regarding how phenotypic evolutionary processes are affected by continuous variables.

While our pipeline for inferring associations between rates of trait evolution and continous factors is generally quite accurate and robust, it notably struggles to detect relatively weak factor-rate relationships from smaller phylogenetic comparative datasets. Based on the simulation conditions examined here, our method generally seems to require phylogenies with around 100 tips or more to reliably detect ∼20-fold differences in rates associated with a continuous factor, and at least 200 tips for 5-fold differences. Nonetheless, our approach is still able to reliably detect and reject factor-rate associations even in the face of rate variation due to unobserved factors and/or general "noise" around relationships. Thus, empirical applications of this approach may fail to detect weak factor-rate associations in some cases yet are importantly unlikely to infer spurious associations (though, as with any correlative statistical method, associations may in fact reflect unobserved factors correlated/confounded with observed factors). Interestingly, while the error rates of our method remained low across all simulation conditions, we do note subtly elevated error among simulations on larger phylogenies involving noise and/or sweetspot (i.e., Gaussian) relationships between rates and unobserved factors. We hypothesize these patterns ultimately reflect "poorly replicated bursts" of trait evolution

(*sensu* Maddison and FitzJohn, 2015; Uyeda et al., 2018), whereby apparent yet transient rate fluctuations occur in a few subclades that happen to exhibit similar observed factor values by chance. In practice, one could investigate marginalized rate estimates (see Figure 4) to detect such scenarios and/or incorporate node error into analyses to reduce the impact of recent, transient rate fluctuations on the inference of trait evolutionary processes (Landis and Schraiber, 2017). Notably, support for constant rate models was rather low across all simulation conditions, including those with truly constant rates unaffected by noise and/or unobserved factors (Table S3). We suspect the highly stochastic nature of trait divergence under Brownian Motion processes (i.e., subclades exhibiting somewhat anomalous levels of trait disparity by chance), along with the phylogenetic autocorrelation of factor data, may lead to phylogenetic pseudoreplication of spurious factor-rate associations with surprising frequency. Such phenomena would be naturally exacerbated by actual rate variation driven by unobserved factors and/or within-species variation/measurement error creating the illusion of unexpectedly large disparity among recently-diverged lineages (Felsenstein, 2008; Landis and Schraiber, 2017). In any case, dummy factor-dependent models were absolutely critical for controlling our method's error rates by providing more "competitive" null models that effectively account for such spurious correlations generated either by chance or unobserved factors. Our work adds to a growing body of literature demonstrating that accounting for "background" or "residual" heterogeneity in evolutionary processes is absolutely critical for robust phylogenetic comparative inference and hypothesis testing (Maddison and FitzJohn, 2015; Beaulieu and O'Meara, 2016; Uyeda et al., 2018; May and Moore, 2020; Boyko and Beaulieu, 2023; Boyko et al., 2023b; Tribble et al., 2023).

Beyond the magnitude of rate variation and phylogeny size, differences in the shapes of factor-rate relationships had a profound influence on our method's power and accuracy. Perhaps unsurprisingly, "wider" factor-rate relationships that featured more gradual changes in rates over factor values were generally harder to both detect and

accurately infer compared to narrower relationships with the same effect size. Thus, our method's ability to infer factor-rate relationships importantly depends not only on the overall magnitude of rate differences, but also how frequently apparent rate shifts tend to occur across a phylogeny, which is primarily determined by how abruptly rates change with respect to factor values. Additionally, sweetspot factor-rate relationships, whereby rates peak or dip at intermediate factor values, appear to require particularly strong effect sizes and/or large phylogenies to reliably detect, likely due to difficulties in correlating the transient rate fluctuations that occur under such relationships with factor values only observed at the tips of a phylogeny. By comparison, simple (i.e., exponential) and threshold (i.e., logistic) factor-rate relationships, whereby rates strictly decrease or increase with respect factor values, appear to generate more persistent rate shifts and thus yield stronger signals of factor-rate associations. However, similarly to methods for detecting relationships between continuous factors and lineage diversification rates (FitzJohn, 2010), despite consistently detecting factor-dependent rates from data simulated under more complex threshold relationships, our approach largely failed to distinguish such relationships from sweetspot and especially simple relationships–even with large phylogenies including 200 tips.

Notably, some recent work suggests detecting correlations between rates and factors from phylogenetic comparative data is generally quite challenging (Hansen et al., 2022). Thus, the low power documented in our simulation study may well reflect fundamental limits to phylogenetic comparative inference of factor-rate relationships rather than shortcomings particular to our approach. Shrinkage of estimated rates, whereby high and low rates tend to be under- and overestimated, respectively, has also been found in previous studies investigating methods for inferring variation in rates of trait evolution (Revell, 2013a; Martin et al., 2023). Notably, the shrinkage of factor-rate relationships inferred from larger phylogenies was generally mild compared to the shrinkage of relationships inferred from smaller phylogenies–particularly when the simulated factor-rate relationship

was weak–suggesting the shrinkage observed in our simulation study primarily arises from our method's limited power with small effect and sample sizes. Nonetheless, as noted by Revell (2013a), uncertainty in factor histories may cause factor values associated with low rates to be inferred in lineages evolving at relatively high rates and vice versa, weakening apparent factor-rate associations and further exacerbating shrinkage of estimated rates. A joint inference approach, whereby the factor history and its effect on trait evolution processes are inferred simultaneously rather than sequentially, could help mitigate these effects and enable more accurate inference of factor-rate associations. However, such methods entail significant mathematical, statistical, and computational challenges due to their complexity (e.g., FitzJohn, 2010, May and Moore, 2020, Boyko et al., 2023b).

Size is unrelated to rates of trait evolution in Monocalypts

Using our contsimmap-based pipeline for inferring associations between rates of trait evolution and continuous factors, we show that rates and size appear unrelated in the *Eucalyptus* subgenus *Eucalyptus* (i.e., Monocalypts), despite numerous theoretical predictions and empirical findings suggesting body size is intertwined with many aspects life history and thus evolutionary rates (Calder, 2001; Niklas and Enquist, 2001; White et al., 2007; Sibly and Brown, 2007; Adler et al., 2014; Salguero-Gómez et al., 2016; Bakewell et al., 2020). Broadly, lineages of larger organisms are thought to accumulate genetic and phenotypic variation more slowly due to their relatively small population sizes and long generation times. Indeed, larger animals and plants do generally exhibit slower rates of molecular evolution (Gillooly et al., 2005; Fontanillas et al., 2007; Bromham, 2011; Lanfear et al., 2013; Weber et al., 2014; but see Thomas et al., 2006; Wright et al., 2011; Lourenço et al., 2013). However, smaller populations can also evolve rapidly due to genetic drift, and empirical relationships between body size and rates of phenotypic evolution/lineage diversification have notably proven inconsistent compared to relationships between size and molecular rates. In line with our results, a recent analysis

suggests size and lineage diversification rates are unrelated within the genus *Eucalyptus* (Vasconcelos et al., 2022), agreeing with patterns found in several animal groups (Cardillo et al., 2003; Feldman et al., 2016; Rainford et al., 2016; but see Amado et al., 2021) yet contrasting with a broader pattern of slower lineage diversification among larger plants (Boucher et al., 2017; Igea et al., 2017; see also Wollenberg et al., 2011; Tedesco et al., 2017). Previous work on phenotypic rates predominantly focuses on vertebrates and has also yielded mixed results, suggesting that body shape evolution is slower among larger fish and birds (Friedman et al., 2019; Zimova et al., 2023) while body size and cranial shape evolution is generally faster among larger mammals (Cooper and Purvis, 2009; Baker et al., 2015).

Ultimately, the lack of general scaling relationships between body size and rates of phenotypic evolution (or lineage diversification rates) likely reflects idiosyncratic adaptations in some clades weakening or completely altering expected relationships among body size, life history traits, and evolutionary rates. For example, *Eucalyptus* species may exhibit an inverse relationship between size and lifespan, as smaller, shrubby species, termed "mallees", tend to be highly fire-tolerant and live for many centuries through countless cycles of burning followed by resprouting. On the other hand, the tallest *Eucalyptus* species generally live in more mesic environments and rely on banks of serotinous seed capsules to sprout and replace the previous generation following intense, mature tree-killing fires (Nicolle, 2006). These peculiarities of *Eucalyptus* biology may account for differences in size-lineage diversification rate correlations between *Eucalyptus* and other plant clades (Boucher et al., 2017; Igea et al., 2017; Vasconcelos et al., 2022), though further investigations of size-phenotypic rate correlations are needed to gauge whether the results of our study are truly anomalous or reflect broader patterns across plants.

While rates of phenotypic evolution among Monocalypts showed no relationship to size, they are highly variable based on our analysis' strong support for models assuming

rates are associated with simulated dummy factors. By further interrogating the patterns of rate variation inferred under the dummy factor-dependent models, we show that variation in rates of phenotypic evolution across the Monocalypt phylogeny are relatively consistent across traits. We hypothesize these patterns reflect some common as-of-yet unknown factor modulating the apparent pace of phenotypic evolution among Monocalypts. Given that the highest rates consistently occur in the subclade roughly corresponding to the section Eucalyptus–a relatively young radiation concentrated in southeast Australia (Ladiges et al., 2010; Nicolle, 2022)–such rate variation may reflect a genuine shift in evolutionary dynamics among this subclade or a more general acceleration in rates over time. Notably, ultrametric phylogenetic comparative data cannot distinguish accelerating rate models from Ornstein-Uhlenbeck models (Blomberg et al., 2003), so these results may simply stem from Monocalypt trait evolution generally following Orstein-Uhlenbeck-like processes, whereby traits fluctuate around "optimum values" owing to adaptive processes and/or constraints on trait variation (see also Harmon et al., 2021). Alternatively, such results may reflect inaccurate estimation of phylogenetic topology and/or branch lengths, as accurate phylogenetic inference is challenging for *Eucalyptus* due to frequent hybridization and incomplete lineage sorting (Rutherford et al., 2016; Thornhill et al., 2019; McLay et al., 2023). In any case, contsimmapping provides an effective tool for more thoroughly exploring the interplay of body size and rates of phenotypic evolution, potentially helping researchers develop more cohesive, generalized theories explaining how the rates of different evolutionary processes interrelate to one another and various life history attributes across the tree of life.

Conclusion

Here, we developed a novel method for sampling the histories of continuous variables on phylogenies under Brownian Motion models of trait evolution, generalizing conventional discrete stochastic character mapping or "simmapping" methods to work

with continuous variables. We further show how these continuous stochastic character maps or "contsimmaps" may be used to accurately and robustly infer relationships between evolutionary processes and continuously-varying factors such as body size, generation time, or climatic niche. In the process, we notably developed pragmatic techniques to account for the influence of unobserved factors on evolutionary processes in macroevolutionary hypothesis testing. Lastly, we used contsimmaps to test whether height is associated with rates of leaf and flower trait evolution in a clade of eucalyptus trees. Despite finding no evidence for height-rate associations, our empirical case study nonetheless demonstrates the empirical utility of contsimmaps in characterizing general patterns of variation in evolutionary processes across clades. Ultimately, contsimmapping will empower researchers with new and innovative strategies for analyzing the evolutionary dynamics of continuous phenotypes and testing macroevolutionary hypotheses that require knowing how continuous variables have changed over evolutionary time.

Funding

B.S.M. was partially supported by a Michigan State University fellowship (Summer Fellowship in Ecology, Evolution, and Behavior) in this work, while M.G.W. was partially supported by two National Science Foundation grants [grant numbers DEB-2236747 and DUE-2012014]. Additional support to both B.S.M. and M.G.W. was provided by another National Science Foundation grant [grant number DEB-1831164] awarded to M.G.W.

Acknowledgements

We would like to thank the members of the corresponding author's former dissertation committee–Luke Harmon, Gideon Bradburd, William Wetzel, and Jeffrey Conner–for their invaluable advice throughout the development of this work. We are especially grateful for their writing suggestions, which greatly improved the clarity of this manuscript. Past and present members of the Weber and Bradburd labs–formerly at

Michigan State University and now at the University of Michigan–provided additional feedback which was particularly helpful in refining how we communicated the work to broader audiences. Lastly, we would like to thank participants and organizers of the 2019 Nantucket phylogeny developeR workshop–in particular Nicholas Bone, Fabio Machado, and Josef Uyeda–for their helpful feedback on the very first implementation of continuous stochastic character mapping.

Supplementary Material

Data will be made available from a Dryad Digital Repository upon publication. In the meantime, please reach out B.S.M. (bruce.stagg.martin@gmail.com) to request any of the data and/or scripts used in this work. The current version of the *contsimmap* R package is available at the GitHub repository: `https://github.com/bstaggmartin/contsimmap`.

References

Adler P.B., Salguero-Gómez R., Compagnoni A., Hsu J.S., Ray-Mukherjee J., Mbeau-Ache C., and Franco M. 2014. Functional traits explain variation in plant life history strategies. Proc Natl Acad Sci USA 111:740–745.

Amado T.F., Martinez P.A., Pincheira-Donoso D., and Olalla-Tárraga M.Á. 2021. Body size distributions of anurans are explained by diversification rates and the environment. Glob Ecol Biogeogr 30:154–164.

Baker J., Meade A., Pagel M., and Venditti C. 2015. Adaptive evolution toward larger size in mammals. Proc Natl Acad Sci USA 112:5093–5098.

Bakewell A.T., Davis K.E., Freckleton R.P., Isaac N.J.B., and Mayhew P.J. 2020. Comparing life histories across taxonomic groups in multiple dimensions: How mammal-like are insects? Am Nat 195:70–81.

Baliga V.B. and Law C.J. 2016. Cleaners among wrasses: Phylogenetics and evolutionary patterns of cleaning behavior within Labridae. Mol Phylogenet Evol 94:424–435.

Beaulieu J. and O'Meara B. 2023. OUwie: Analysis of Evolutionary Rates in an OU Framework. R package version 2.10.

Beaulieu J.M. and O'Meara B.C. 2016. Detecting hidden diversification shifts in models of trait-dependent speciation and extinction. Syst Biol 65:583–601.

Blomberg S.P., Garland T. Jr, and Ives A.R. 2003. Testing for phylogenetic signal in comparative data: Behavioral traits are more labile. Evolution 57:717–745.

Bollback J.P. 2006. SIMMAP: Stochastic character mapping of discrete traits on phylogenies. BMC Bioinformatics 7:88.

Borstein S.R., Fordyce J.A., O'Meara B.C., Wainwright P.C., and McGee M.D. 2019. Reef fish functional traits evolve fastest at trophic extremes. Nat Ecol Evol 3:191–199.

Boucher F.C., Démery V., Conti E., Harmon L.J., and Uyeda J. 2018. A general model for estimating macroevolutionary landscapes. Syst Biol 67:304–319.

Boucher F.C., Verboom G.A., Musker S., and Ellis A.G. 2017. Plant size: A key determinant of diversification? New Phytol 216:24–31.

Boyko J.D. and Beaulieu J.M. 2023. Reducing the biases in false correlations between discrete characters. Syst Biol 72:476–488.

Boyko J.D., Hagen E.R., Beaulieu J.M., and Vasconcelos T. 2023a. The evolutionary responses of life-history strategies to climatic variability in flowering plants. New Phytol 240:1587–1600.

Boyko J.D., O'Meara B.C., and Beaulieu J.M. 2023b. A novel method for jointly modeling the evolution of discrete and continuous traits. Evolution 77:836–851.

Brent R.P. 2013. Algorithms for Minimization Without Derivatives. Dover Publications, Mineola, NY.

Bromham L. 2011. The genome as a life-history character: Why rate of molecular evolution varies between mammal species. Philos Trans R Soc B 366:2503–2513.

Brooker I., Slee A., Connor J., Duffy S., and West J. 2015. EUCLID Eucalypts of Australia. 4 ed. Identic Pty Ltd, Brisbane, QLD.

Burns M.D. and Bloom D.D. 2020. Migratory lineages rapidly evolve larger body sizes than non-migratory relatives in ray-finned fishes. Proc R Soc B 287:20192615.

Calder W.A. 2001. Ecological consequences of body size. Pages 1–7 *in* Encyclopedia of Life Sciences. John Wiley & Sons, Ltd, Chichester, UK.

Cardillo M., Huxtable J.S., and Bromham L. 2003. Geographic range size, life history and rates of diversification in Australian mammals. J Evol Biol 16:282–288.

Chira A.M., Cooney C.R., Bright J.A., Capp E.J.R., Hughes E.C., Moody C.J.A., Nouri L.O., Varley Z.K., and Thomas G.H. 2018. Correlates of rate heterogeneity in avian ecomorphological traits. Ecol Lett 21:1505–1514.

Clavel J., Escarguel G., and Merceron G. 2015. mvMORPH: An R package for fitting multivariate evolutionary models to morphometric data. Methods Ecol Evol 6:1311–1319.

Clavel J. and Morlon H. 2017. Accelerated body size evolution during cold climatic periods in the Cenozoic. Proc Natl Acad Sci USA 114:4183–4188.

Cooper N. and Purvis A. 2009. What factors shape rates of phenotypic evolution? A comparative study of cranial morphology of four mammalian clades. J Evol Biol 22:1024–1035.

de Alencar L.R.V., Martins M., Burin G., and Quental T.B. 2017. Arboreality constrains morphological evolution but not species diversification in vipers. Proc R Soc B 284:20171775.

Dembo R.S. and Steihaug T. 1983. Truncated-Newton algorithms for large-scale unconstrained optimization. Math Program 26:190–212.

Dobzhansky T. and Sturtevant A.H. 1938. Inversions in the chromosomes of *Drosophila pseudoobscura*. Genetics 23:28–64.

Drury J.P., Clavel J., Tobias J.A., Rolland J., Sheard C., and Morlon H. 2021. Tempo and mode of morphological evolution are decoupled from latitude in birds. PLoS Biol 19:e3001270.

Dziak J.J., Coffman D.L., Lanza S.T., Li R., and Jermiin L.S. 2020. Sensitivity and specificity of information criteria. Brief Bioinform 21:553–565.

Fabre A.C., Bardua C., Bon M., Clavel J., Felice R.N., Streicher J.W., Bonnel J., Stanley

E.L., Blackburn D.C., and Goswami A. 2020. Metamorphosis shapes cranial diversity and rate of evolution in salamanders. Nat Ecol Evol 4:1129–1140.

Falster D., Gallagher R., Wenk E.H., Wright I.J., Indiarto D., Andrew S.C., Baxter C., Lawson J., Allen S., Fuchs A., Monro A., Kar F., Adams M.A., Ahrens C.W., Alfonzetti M., Angevin T., Apgaua D.M.G., Arndt S., Atkin O.K., Atkinson J., Auld T., Baker A., von Balthazar M., Bean A., Blackman C.J., Bloomfield K., Bowman D.M.J.S., Bragg J., Brodribb T.J., Buckton G., Burrows G., Caldwell E., Camac J., Carpenter R., Catford J.A., Cawthray G.R., Cernusak L.A., Chandler G., Chapman A.R., Cheal D., Cheesman A.W., Chen S.C., Choat B., Clinton B., Clode P.L., Coleman H., Cornwell W.K., Cosgrove M., Crisp M., Cross E., Crous K.Y., Cunningham S., Curran T., Curtis E., Daws M.I., DeGabriel J.L., Denton M.D., Dong N., Du P., Duan H., Duncan D.H., Duncan R.P., Duretto M., Dwyer J.M., Edwards C., Esperon-Rodriguez M., Evans J.R., Everingham S.E., Farrell C., Firn J., Fonseca C.R., French B.J., Frood D., Funk J.L., Geange S.R., Ghannoum O., Gleason S.M., Gosper C.R., Gray E., Groom P.K., Grootemaat S., Gross C., Guerin G., Guja L., Hahs A.K., Harrison M.T., Hayes P.E., Henery M., Hochuli D., Howell J., Huang G., Hughes L., Huisman J., Ilic J., Jagdish A., Jin D., Jordan G., Jurado E., Kanowski J., Kasel S., Kellermann J., Kenny B., Kohout M., Kooyman R.M., Kotowska M.M., Lai H.R., Laliberté E., Lambers H., Lamont B.B., Lanfear R., van Langevelde F., Laughlin D.C., Laugier-Kitchener B.A., Laurance S., Lehmann C.E.R., Leigh A., Leishman M.R., Lenz T., Lepschi B., Lewis J.D., Lim F., Liu U., Lord J., Lusk C.H., Macinnis-Ng C., McPherson H., Magallón S., Manea A., López-Martinez A., Mayfield M., McCarthy J.K., Meers T., van der Merwe M., Metcalfe D.J., Milberg P., Mokany K., Moles A.T., Moore B.D., Moore N., Morgan J.W., Morris W., Muir A., Munroe S., Nicholson Á., Nicolle D., Nicotra A.B., Niinemets Ü., North T., O'Reilly-Nugent A., O'Sullivan O.S., Oberle B., Onoda Y., Ooi M.K.J., Osborne C.P., Paczkowska G., Pekin B., Guilherme Pereira C., Pickering C., Pickup M., Pollock L.J., Poot P., Powell J.R., Power S.A., Prentice I.C., Prior L., Prober S.M., Read J.,

Reynolds V., Richards A.E., Richardson B., Roderick M.L., Rosell J.A., Rossetto M., Rye B., Rymer P.D., Sams M.A., Sanson G., Sauquet H., Schmidt S., Schönenberger J., Schulze E.D., Sendall K., Sinclair S., Smith B., Smith R., Soper F., Sparrow B., Standish R.J., Staples T.L., Stephens R., Szota C., Taseski G., Tasker E., Thomas F., Tissue D.T., Tjoelker M.G., Tng D.Y.P., de Tombeur F., Tomlinson K., Turner N.C., Veneklaas E.J., Venn S., Vesk P., Vlasveld C., Vorontsova M.S., Warren C.A., Warwick N., Weerasinghe L.K., Wells J., Westoby M., White M., Williams N.S.G., Wills J., Wilson P.G., Yates C., Zanne A.E., Zemunik G., and Ziemińska K. 2021. AusTraits, a curated plant trait database for the Australian flora. Sci Data 8:254.

Feldman A., Sabath N., Pyron R.A., Mayrose I., and Meiri S. 2016. Body sizes and diversification rates of lizards, snakes, amphisbaenians and the tuatara. Glob Ecol Biogeogr 25:187–197.

Felsenstein J. 2008. Comparative methods with sampling error and within-species variation: Contrasts revisited and revised. Am Nat 171:713–725.

Felsenstein J. 2012. A comparative method for both discrete and continuous characters using the threshold model. Am Nat 179:145–156.

FitzJohn R.G. 2010. Quantitative traits and diversification. Syst Biol 59:619–633.

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

Fontanillas E., Welch J.J., Thomas J.A., and Bromham L. 2007. The influence of body size and net diversification rate on molecular evolution during the radiation of animal phyla. BMC Evol Biol 7:95.

Freyman W.A. and Höhna S. 2019. Stochastic character mapping of state-dependent diversification reveals the tempo of evolutionary decline in self-compatible Onagraceae lineages. Syst Biol 68:505–519.

Friedman S.T., Martinez C.M., Price S.A., and Wainwright P.C. 2019. The influence of size on body shape diversification across Indo-Pacific shore fishes. Evolution 73:1873–1884.

Friedman S.T. and Muñoz M.M. 2023. A latitudinal gradient of deep-sea invasions for marine fishes. Nat Commun 14:773.

Gillooly J.F., Allen A.P., West G.B., and Brown J.H. 2005. The rate of DNA evolution: Effects of body size and temperature on the molecular clock. Proc Natl Acad Sci USA 102:140–145.

Gingerich P.D. 2001. Rates of evolution on the time scale of the evolutionary process. Genetica 112-113:127–144.

Gingerich P.D. 2009. Rates of evolution. Annu Rev Ecol Evol Syst 40:657–675.

Goolsby E.W. 2017. Rapid maximum likelihood ancestral state reconstruction of continuous characters: A rerooting-free algorithm. Ecol Evol 7:2791–2797.

Groussin M., Daubin V., Gouy M., and Tannier E. 2016. Ancestral reconstruction: Theory and practice. Pages 70–77 *in* Encyclopedia of Evolutionary Biology (R. M. Kliman, ed.). Academic Press, Waltham, MA.

Hansen T.F., Bolstad G.H., and Tsuboi M. 2022. Analyzing disparity and rates of morphological evolution with model-based phylogenetic comparative methods. Syst Biol 71:1054–1072.

Hansen T.F., Pienaar J., and Orzack S.H. 2008. A comparative method for studying adaptation to a randomly evolving environment. Evolution 62:1965–1977.

Harmon L.J., Pennell M.W., Francisco Henao-Diaz L., Rolland J., Sipley B.N., and Uyeda J.C. 2021. Causes and consequences of apparent timescaling across all estimated evolutionary rates. Annu Rev Ecol Evol Syst 52:587–609.

Hassler G., Tolkoff M.R., Allen W.L., Ho L.S.T., Lemey P., and Suchard M.A. 2022. Inferring phenotypic trait evolution on large trees with many incomplete measurements. J Am Stat Assoc 117:678–692.

Huelsenbeck J.P., Nielsen R., and Bollback J.P. 2003. Stochastic mapping of morphological characters. Syst Biol 52:131–158.

Hughes J.J., Berv J.S., Chester S.G.B., Sargis E.J., and Field D.J. 2021. Ecological selectivity and the evolution of mammalian substrate preference across the K-Pg boundary. Ecol Evol 11:14540–14554.

Igea J., Miller E.F., Papadopulos A.S.T., and Tanentzap A.J. 2017. Seed size and its rate of evolution correlate with species diversification across angiosperms. PLoS Biol 15:e2002792.

Johnson S.G. 2021. The NLopt nonlinear-optimization package. Version 2.7.1.

Joy J.B., Liang R.H., McCloskey R.M., Nguyen T., and Poon A.F.Y. 2016. Ancestral reconstruction. PLoS Comput Biol 12:e1004763.

Ladiges P.Y., Bayly M.J., and Nelson G.J. 2010. East-west continental vicariance in: *Eucalyptus*: Subgenus: *Eucalyptus*. Pages 267–302 *in* Beyond Cladistics: The Branching of a Paradigm (D. M. Williams and S. Knapp, eds.). University of California Press, Oakland, CA.

Landis M.J., Eaton D.A.R., Clement W.L., Park B., Spriggs E.L., Sweeney P.W., Edwards E.J., and Donoghue M.J. 2021. Joint phylogenetic estimation of geographic movements and biome shifts during the global diversification of *Viburnum*. Syst Biol 70:67–85.

Landis M.J. and Schraiber J.G. 2017. Pulsed evolution shaped modern vertebrate body sizes. Proc Natl Acad Sci USA 114:13224–13229.

Landis M.J., Schraiber J.G., and Liang M. 2013. Phylogenetic analysis using Lévy processes: finding jumps in the evolution of continuous traits. Syst Biol 62:193–204.

Lanfear R., Ho S.Y.W., Jonathan Davies T., Moles A.T., Aarssen L., Swenson N.G., Warman L., Zanne A.E., and Allen A.P. 2013. Taller plants have lower rates of molecular evolution. Nat Commun 4:1879.

Lewandowski D., Kurowicka D., and Joe H. 2009. Generating random correlation matrices based on vines and extended onion method. J Multivar Anal 100:1989–2001.

Limpert E., Stahel W.A., and Abbt M. 2001. Log-normal distributions across the sciences: Keys and clues. Bioscience 51:341–352.

Lourenço J.M., Glémin S., Chiari Y., and Galtier N. 2013. The determinants of the molecular substitution process in turtles. J Evol Biol 26:38–50.

Maddison W.P. and FitzJohn R.G. 2015. The unsolved challenge to phylogenetic correlation tests for categorical characters. Syst Biol 64:127–136.

Martin B.S., Bradburd G.S., Harmon L.J., and Weber M.G. 2023. Modeling the evolution of rates of continuous trait evolution. Syst Biol 72:590–605.

May M.R. and Moore B.R. 2020. A Bayesian approach for inferring the impact of a discrete character on rates of continuous-character evolution in the presence of background-rate variation. Syst Biol 69:530–544.

McLay T.G.B., Fowler R.M., Fahey P.S., Murphy D.J., Udovicic F., Cantrill D.J., and Bayly M.J. 2023. Phylogenomics reveals extreme gene tree discordance in a lineage of dominant trees: Hybridization, introgression, and incomplete lineage sorting blur deep evolutionary relationships despite clear species groupings in *Eucalyptus* subgenus *Eudesmia*. Mol Phylogenet Evol 187:107869.

Nations J.A., Mount G.G., Morere S.M., Achmadi A.S., Rowe K.C., and Esselstyn J.A. 2021. Locomotory mode transitions alter phenotypic evolution and lineage diversification in an ecologically rich clade of mammals. Evolution 75:376–393.

Nicolle D. 2006. A classification and census of regenerative strategies in the eucalypts (*Angophora*, *Corymbia* and *Eucalyptus*–Myrtaceae), with special reference to the obligate seeders. Aust J Bot 54:391–407.

Nicolle D. 2022. Classification of the eucalypts (*Angophora*, *Corymbia* and *Eucalyptus*) version 6. `https://www.dn.com.au/Classification-Of-The-Eucalypts.pdf` accessed: 2023-11-29.

Nielsen R. 2002. Mapping mutations on phylogenies. Syst Biol 51:729–739.

Niklas K.J. and Enquist B.J. 2001. Invariant scaling relationships for interspecific plant biomass production rates and body size. Proc Natl Acad Sci USA 98:2922–2927.

Pauling L., Zuckerkandl E., Henriksen T., and Lövstad R. 1963. Chemical paleogenetics: Molecular "restoration studies" of extinct forms of life. Acta Chem Scand 17:9–16.

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

Rainford J.L., Hofreiter M., and Mayhew P.J. 2016. Phylogenetic analyses suggest that diversification and body size evolution are independent in insects. BMC Evol Biol 16:8.

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

Revell L.J. 2013a. A comment on the use of stochastic character maps to estimate evolutionary rate variation in a continuously valued trait. Syst Biol 62:339–345.

Revell L.J. 2013b. Two new graphical methods for mapping trait evolution on phylogenies. Methods Ecol Evol 4:754–759.

Rincon-Sandoval M., Duarte-Ribeiro E., Davis A.M., Santaquiteria A., Hughes L.C., Baldwin C.C., Soto-Torres L., Acero P A., Walker H.J. Jr, Carpenter K.E., Sheaves M.,

Ortí G., Arcila D., and Betancur-R R. 2020. Evolutionary determinism and convergence associated with water-column transitions in marine fishes. Proc Natl Acad Sci USA 117:33396–33403.

Rowan T.H. 1990. Functional stability analysis of numerical algorithms. Ph.D. thesis Department of Computer Science, University of Texas at Austin, TX.

Rutherford S., Wilson P.G., Rossetto M., and Bonser S.P. 2016. Phylogenomics of the green ash eucalypts (Myrtaceae): A tale of reticulate evolution and misidentification. Aust Syst Bot 28:326–354.

Salguero-Gómez R., Jones O.R., Jongejans E., Blomberg S.P., Hodgson D.J., Mbeau-Ache C., Zuidema P.A., de Kroon H., and Buckley Y.M. 2016. Fast-slow continuum and reproductive strategies structure plant life-history variation worldwide. Proc Natl Acad Sci USA 113:230–235.

Sanger F., Thompson E.O., and Kitai R. 1955. The amide groups of insulin. Biochem J 59:509–518.

Schluter D., Price T., Mooers A.Ø., and Ludwig D. 1997. Likeliood of ancestor state in adaptive radiation. Evolution 51:1699–1711.

Schultz T.R., Cocroft R.B., and Churchill G.A. 1996. The reconstruction of ancestral character states. Evolution 50:504–511.

Schwarz G. 1978. Estimating the dimension of a model. Ann Stat 6:461–464.

Sibly R.M. and Brown J.H. 2007. Effects of body size and lifestyle on evolution of mammal life histories. Proc Natl Acad Sci USA 104:17707–17712.

Siqueira A.C., Muruga P., and Bellwood D.R. 2023. On the evolution of fish-coral interactions. Ecol Lett 26:1348–1358.

Slater G.J., Goldbogen J.A., and Pyenson N.D. 2017. Independent evolution of baleen whale gigantism linked to Plio-Pleistocene ocean dynamics. Proc R Soc B 284:20170546.

Stan Development Team . 2019. Stan Modeling Language Users Guide and Reference Manual. Version 2.21.0.

Sumrall C.D. and Brochu C.A. 2008. Viewing paleobiology through the lens of phylogeny. Paleontol Soc Papers 14:165–183.

Tedesco P.A., Paradis E., Lévêque C., and Hugueny B. 2017. Explaining global-scale diversification patterns in actinopterygian fishes. J Biogeogr 44:773–783.

Thomas J.A., Welch J.J., Woolfit M., and Bromham L. 2006. There is no universal molecular clock for invertebrates, but rate variation does not scale with body size. Proc Natl Acad Sci USA 103:7366–7371.

Thornhill A.H., Crisp M.D., Külheim C., Lam K.E., Nelson L.A., Yeates D.K., and Miller J.T. 2019. A dated molecular perspective of eucalypt taxonomy, evolution and diversification. Aust Syst Bot 32:29–48.

Tornabene L., Van Tassell J.L., Robertson D.R., and Baldwin C.C. 2016. Repeated invasions into the twilight zone: Evolutionary origins of a novel assemblage of fishes from deep Caribbean reefs. Mol Ecol 25:3662–3682.

Tribble C.M., May M.R., Jackson-Gain A., Zenil-Ferguson R., Specht C.D., and Rothfels C.J. 2023. Unearthing modes of climatic adaptation in underground storage organs across Liliales. Syst Biol 72:198–212.

Uyeda J.C., Bone N., McHugh S., Rolland J., and Pennell M.W. 2021. How should functional relationships be evaluated using phylogenetic comparative methods? A case study using metabolic rate and body temperature. Evolution 75:1097–1105.

Uyeda J.C., Zenil-Ferguson R., and Pennell M.W. 2018. Rethinking phylogenetic comparative methods. Syst Biol 67:1091–1109.

Vasconcelos T., O'Meara B.C., and Beaulieu J.M. 2022. A flexible method for estimating

tip diversification rates across a range of speciation and extinction scenarios. Evolution 76:1420–1433.

Weber C.C., Nabholz B., Romiguier J., and Ellegren H. 2014. Kr/Kc but not dN/dS correlates positively with body mass in birds, raising implications for inferring lineage-specific selection. Genome Biol. 15:542.

Weir J.T. and Lawson A. 2015. Evolutionary rates across gradients. Methods Ecol Evol 6:1278–1286.

White E.P., Ernest S.K.M., Kerkhoff A.J., and Enquist B.J. 2007. Relationships between body size and abundance in ecology. Trends Ecol Evol 22:323–330.

Witmer L. 1995. The extant phylogenetic bracket and the importance of reconstructing soft tissues in fossils. Pages 19–33 *in* Functional Morphology in Vertebrate Paleontology (J. J. Thomason, ed.). Cambridge University Press, New York, NY.

Wollenberg K.C., Vieites D.R., Glaw F., and Vences M. 2011. Speciation in little: The role of range and body size in the diversification of Malagasy mantellid frogs. BMC Evol Biol 11:217.

Wright S.D., Ross H.A., Jeanette Keeling D., McBride P., and Gillman L.N. 2011. Thermal energy and the rate of genetic evolution in marine fishes. Evol Ecol 25:525–530.

Ypma J., Johnson S.G., Stamm A., Borchers H.W., Eddelbuettel D., Ripley B., Hornik K., Chiquet J., Adler A., Dai X., and Ooms J. 2022. nlotpr: R Interface to NLOPT. R package version 2.0.3.

Zimova M., Weeks B.C., Willard D.E., Giery S.T., Jirinec V., Burner R.C., and Winger B.M. 2023. Body size predicts the rate of contemporary morphological change in birds. Proc Natl Acad Sci USA 120:e2206971120.