A Bayesian Approach for Inferring the Impact of a Discrete Character on Rates of Continuous-Character Evolution in the Presence of Background-Rate Variation

MICHAEL R. MAY* AND BRIAN R. MOORE

Department of Evolution and Ecology, University of California, Davis, Storer Hall, One Shields Avenue, Davis, CA 95616, USA
*Correspondence to be sent to: Department of Evolution and Ecology, University of California, Davis, Storer Hall, One Shields Avenue, Davis, CA 95616, USA; E-mail: mrmay@ucdavis.edu.

Abstract.—Understanding how and why rates of character evolution vary across the Tree of Life is central to many evolutionary questions; for example, does the trophic apparatus (a set of continuous characters) evolve at a higher rate in fish lineages that dwell in reef versus nonreef habitats (a discrete character)? Existing approaches for inferring the relationship between a discrete character and rates of continuous-character evolution rely on comparing a null model (in which rates of continuous-character evolution are constant across lineages) to an alternative model (in which rates of continuous-character evolution depend on the state of the discrete character under consideration). However, these approaches are susceptible to a "straw-man" effect: the influence of the discrete character is inflated because the null model is extremely unrealistic. Here, we describe `MuSSCRat`, a Bayesian approach for inferring the impact of a discrete trait on rates of continuous-character evolution in the presence of alternative sources of rate variation ("background-rate variation"). We demonstrate by simulation that our method is able to reliably infer the degree of state-dependent rate variation, and show that ignoring background-rate variation leads to biased inferences regarding the degree of state-dependent rate variation in grunts (the fish group Haemulidae). [Bayesian phylogenetic comparative methods; continuous-character evolution; data augmentation; discrete-character evolution.]

Variable rates of continuous-character evolution are central to many evolutionary questions. These questions may involve changes in the rate of character evolution over time (time-dependent scenarios) or among lineages (lineage-specific scenarios). Such questions may be pursued by means of agnostic surveys to detect rate variation (data-exploration approaches; Harmon et al. 2010; Eastman et al. 2011; Venditti et al. 2011) or by testing predictions regarding factors hypothesized to influence rates of character evolution (hypothesis-testing approaches; O'Meara et al. 2006; Collar et al. 2009). A particular type of hypothesis posits that the rate of continuous-character evolution depends on the state of a discrete trait, for example, the evolutionary rate of the feeding apparatus (a set of continuous traits) in a lineage depends on the habitat type (the discrete trait) of its members.

Testing hypotheses regarding state-dependent rates of continuous-character evolution is currently pursued using a computational procedure (e.g., Collar et al. 2009, 2010; Price et al. 2011, 2013) comprised of four steps: 1) fit a Brownian motion model to the observations (the tree and continuous-trait values at its tips), where the rate of continuous-character evolution is assumed to be constant across all branches of the tree (the "null" or constant-rate model); 2) generate a sample of discrete-character histories ("stochastic maps", Nielsen 2002; Huelsenbeck et al. 2003); 3) for each stochastic map, fit a Brownian motion model to the observations, where the instantaneous rate of continuous-character evolution at a given point on a given branch depends on the corresponding state of the discrete-character mapping (the "state-dependent model" O'Meara et al. 2006), and; 4) compare the fit of the state-dependent model (averaged

over the sample of stochastic maps) to the constant-rate model using the Akaike Information Criterion (AIC). If the state-dependent model is preferred, we infer that rates of continuous-character evolution are correlated with the state of the discrete character.

The current approach has two potential problems. First, stochastic maps of the discrete character are generated without reference to the continuous characters. By construction, however, the state-dependent model specifies that the discrete and continuous characters are evolving *jointly*. The continuous characters therefore possess information about the history of the discrete character; disregarding this mutual information will lead to biased parameter estimates (Revell 2012). Second, the null model—where the continuous characters are assumed to evolve at a constant rate across lineages—is extremely unrealistic. *Any* variation in the rate of continuous-character evolution—whether or not it is associated with the discrete character under consideration—is apt to be interpreted as evidence against the overly simplistic null model. This "straw-man effect" has the potential to mislead our inferences regarding the factors that impact rates of continuous-character evolution.

We describe a Bayesian approach for inferring the impact of a discrete trait on rates of continuous-character evolution that addresses the problems described above. We begin by developing a stochastic process that explicitly models the joint evolution of the discrete and continuous characters; this stochastic process can accommodate one or more continuous characters evolving under a state-dependent multivariate Brownian motion process. We refer to this new model as `MuSSCRat` (for

Multiple State-Specific Rates of continuous-character evolution). We then develop an inference model that accommodates variation in the background rate of continuous-character evolution (i.e., rate variation across lineages that is independent of the discrete character under consideration). We implement this model in a Bayesian framework, which accommodates uncertainty in the phylogeny, discrete-character history, and parameters of the state-dependent model. We show by simulation that the method is able to reliably infer the state-dependent rates of continuous-character evolution, and that ignoring background-rate variation leads to an inflated false-positive rate. Finally, we demonstrate the new method with an empirical analysis of grunts (a group of haemulid fish) to illustrate the impacts of background-rate variation and prior specification on inferences about state-dependent rate variation.

METHODS

Our goal is to develop a state-dependent multivariate Brownian motion model, `MuSSCRat`, in a Bayesian statistical framework. We begin with a simple simulation to describe the parameters and basic properties of the `MuSSCRat` model. We then show how to calculate the probability of observing the discrete and continuous characters across the tips of a phylogeny under the model (i.e., how to compute the likelihood). Finally, we describe the relevant details—the priors and Markov chain Monte Carlo (MCMC) machinery—required to perform Bayesian inference under the `MuSSCRat` model.

The State-Dependent Multivariate Brownian Motion Process

A simulation example.—We introduce the salient properties of the `MuSSCRat` model by describing a simple simulation with a binary discrete character, X, and a single continuous character, Y, over a single branch. The discrete character has two states, 0 and 1; the continuous character can be any real number. The state of the simulation at time t is the pair of discrete- and continuous-character values, (x_t, y_t). (We use capital letters—X and Y—to represent random variables, and lowercase letters—x and y—to represent specific values of those random variables.)

The discrete trait evolves under a continuous-time Markov process, changing from state 0 to state 1 with rate q_{01}, and from state 1 to state 0 with rate q_{10}. The continuous character evolves under a state-dependent Brownian motion process, where the diffusion rate, s_i^2, measures the rate of continuous-character evolution when the discrete character is in state i (i.e., $s_1^2 > s_0^2$ indicates that the continuous character evolves faster in discrete state 1 than in discrete state 0). In a small time interval of duration Δt, where the discrete character begins in state i, the continuous character changes by a

normally distributed random variable with mean 0 and variance $s_i^2 \Delta t$, and the discrete character changes state with probability $q_{ij}\Delta t$.

We begin the simulation at time $t = 0$, with the discrete character in state x_0 and the continuous character with value y_0. We then increment the simulation forward in time by a small time interval, Δt, applying the above rules describing how the state of the process changes during each time interval. We continue to increment the simulation forward in time until we reach the end of the branch (at time T). The outcome of the simulation is a *sample path* that records the state of the process from the beginning to the end of the branch (Fig. 1a).

The *transition-probability density* specifies the probability that the process ends in some state, (x_t, y_t), given an initial state (x_0, y_0), after a certain amount of time, T, has elapsed. The resulting frequency histogram of end states provides a Monte Carlo approximation of the *transition-probability* density for a branch of duration T. Note that the transition-probability densities of standard (i.e., state-independent) Brownian motion processes are normal densities. By contrast, it is clear from our simulations that the transition-probability densities under the state-dependent process are not normal densities (Fig. 1b).

Parameters of the `MuSSCRat` model.—We model the joint evolution of a discrete binary character, X, and a set of c continuous characters, Y, as a stochastic process. The discrete trait has two possible values, which we arbitrarily label 0 and 1; $x \in (0, 1)$. The continuous characters are a vector of real-valued random variables, where $y[i]$ is the value of the ith continuous character ($y \in \mathbb{R}^c$). Variables of the `MuSSCRat` model are summarized in Table 1.

We assume that the discrete character evolves under a continuous-time Markov process, and that the continuous characters evolve under a multivariate Brownian motion process with rates that depend on the state of the binary character. These model components collectively describe how the set of characters, (x, y), evolve together over a single branch; we detail the evolutionary dynamics of this process over an entire tree when we describe the likelihood function.

The instantaneous-rate matrix, Q, describes the rates at which the binary character evolves: q_{ij} describes the instantaneous rate of change from state i to state j:

$$Q = \begin{pmatrix} -q_{01} & q_{01} \\ q_{10} & -q_{10} \end{pmatrix}.$$

The complete history of the discrete trait on branch l—which we represent as κ_l—specifies the state of the character at the beginning and end of the branch, and also the times of any state changes along the branch.

While the process is in a particular discrete state, we assume the continuous characters evolve under a multivariate Brownian motion model. We allow the background rate of evolution to vary among lineages
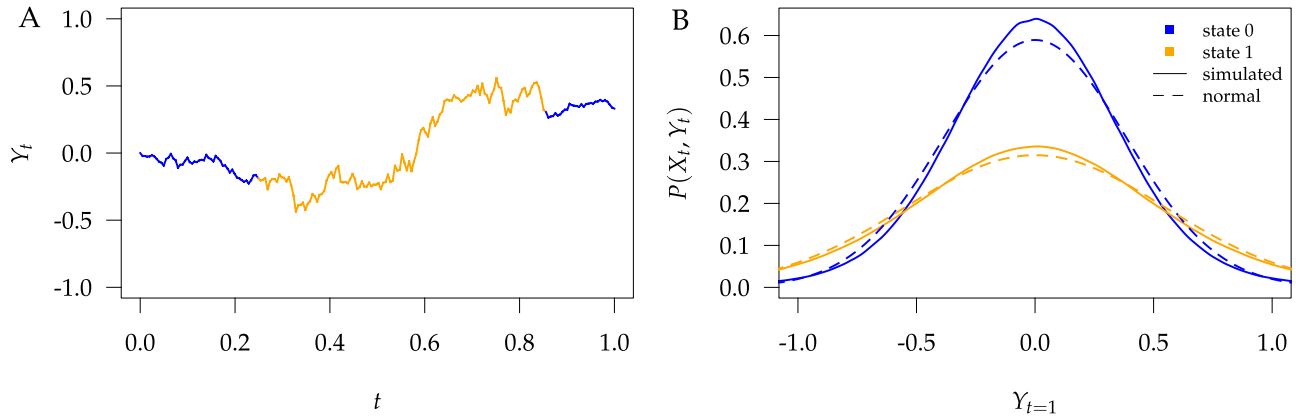
FIGURE 1.   Simulations under the state-dependent multivariate Brownian motion process. The process is either in discrete state 0 (blue) or discrete state 1 (orange), where the rate of change between discrete states is equal ($q_{01} = q_{10} = 1$), and the rate of continuous-character evolution is higher when the process is in the orange state ($s_0^2 = 0.1, s_1^2 = 0.5$). A) A single sample path from $t = 0$ to $t = 1$. The process begins and ends in the blue state, but spends some time in the orange state. Note that there is more evolution in the orange state than in the blue state. B) The distribution of end states for 10 million simulated realizations. Solid lines represent the simulated joint probability densities of the discrete and continuous states. Dashed lines represent the normal densities with parameters estimated from the simulated end states. Note that the simulated densities depart from the normal densities (both Kolmogorov–Smirnov P-value $< 2E{-}16$).

TABLE 1.   The variables of the MuSSCRat model and their interpretation.

Variable	Interpretation
Y	The continuous characters for one lineage
y_t	The state of the continuous characters at time t
c	The number of continuous characters
$\mathcal{Y}$	An $n \times c$ matrix containing the c continuous characters for the n species in the tree
X	The discrete character for one lineage
x_t	The state of the discrete character at time t
$\mathcal{X}$	An $n \times 1$ matrix containing the discrete character for the n species in the tree
κ_l	The complete history of the discrete character (the state at the beginning and end of the process, and all of the changes in between) along the lineage l
Q	The instantaneous-rate matrix of the discrete-character CTMC model
q_{ij}	The instantaneous rate of change from state i to state j
$\boldsymbol{\beta^2}$	The background rate of continuous-character evolution among all lineages
β_l^2	The background rate of continuous-character evolution for lineage l
$\boldsymbol{\sigma^2}$	The relative rates of continuous-character evolution for each continuous character
$\boldsymbol{\zeta^2}$	The relative rates of continuous-character evolution for each discrete state
R	The evolutionary correlation matrix
ρ_{ij}	The evolutionary correlation between continuous characters i and j
Σ	The discrete-state-independent evolutionary variance–covariance matrix

in the phylogeny by letting each branch have its own rate parameter, β_l^2 (we call this "background-rate variation"). While in discrete state i, the background rate is multiplied by the state-specific relative rate, ζ_i^2. We also allow the relative rate of evolution to vary among the continuous characters. The vector $\boldsymbol{\sigma^2}$ contains these relative rates; σ_i^2 is the relative rate at which continuous character i evolves. The evolutionary correlations between characters are contained in the $c \times c$ symmetric correlation matrix, R, where $[R]_{ij} = \rho_{ij}$ specifies the correlation between characters i and j.

We assume that the relative rates of change between characters, $\boldsymbol{\sigma^2}$, and the evolutionary correla-

tions between characters, R, are independent of the discrete state; in other words, we assume that the state of the discrete trait affects only the *overall rate* of continuous-character evolution, but not the *nature* of the evolutionary process (as represented by $\boldsymbol{\sigma^2}$ and R). We combine the relative rates among characters and the correlation matrix to form the overall evolutionary variance-covariance matrix, Σ:

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho_{12} & \cdots & \sigma_1\sigma_c\rho_{1c} \\ \sigma_2\sigma_1\rho_{12} & \sigma_2^2 & \cdots & \sigma_2\sigma_c\rho_{2c} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_c\sigma_1\rho_{1c} & \sigma_c\sigma_2\rho_{2c} & \cdots & \sigma_c^2 \end{pmatrix}.$$

Bayesian Inference

We implement the MuSSCRat model with background-rate variation as a Bayesian model to infer the joint posterior density of the parameters given the observed data. We must specify both the likelihood function and also the joint prior density to compute the joint posterior distribution. We describe each of these components below.

Data.—We imagine that we have sampled one discrete character and c continuous characters for each of n species; relationships among these species are defined by the phylogeny, Ψ. We store the discrete characters in an $n \times 1$ column vector, $\mathcal{X}$, and the continuous characters in an $n \times c$ matrix, $\mathcal{Y}$. We assume that the discrete and continuous characters evolve independently along each of the $2n-2$ branches of the tree. We index the internal nodes according to their sequence in a post-order traversal of the tree, starting from the root (which has index 1).

Likelihood function.—We simplify likelihood calculations by including the vector of character histories, κ, where κ_l is the discrete-character history along branch l (including the state at the beginning and end of the branch), as variables in the model. In effect, we are "augmenting" the discrete-character data observed at the tips of the tree with unobserved discrete-character histories over the entire tree; this technique is referred to as *data augmentation* (Tanner and Wong 1987; Robinson et al. 2003; Mateiu and Rannala 2006; Lartillot 2006; Landis et al. 2013).

The *augmented likelihood* is a product of the joint probability of $(\mathcal{X}, \kappa)$ and the conditional probability density of $\mathcal{Y}$ given κ:

$$P(\mathcal{X}, \mathcal{Y}, \kappa \,|\, \theta, \Psi) = P(\mathcal{X}, \kappa \,|\, \theta, \Psi) P(\mathcal{Y} \,|\, \kappa, \theta, \Psi),$$

where θ are the parameters of the MuSSCRat model. We compute the joint probability of $(\mathcal{X}, \kappa)$ as a product of independent probabilities across each of the $2n-2$ branches:

$$P(\mathcal{X}, \kappa \,|\, \theta, \Psi) = P(x_1 \,|\, \theta) \prod_{l=1}^{2n-2} P(\kappa_l \,|\, \theta, t_l) \mathbb{1}(\kappa_l),$$

where $\mathbb{1}(\kappa_l)$ is an indicator function that ensures that character histories are consistent across ancestor-descendant branches (i.e., that the state at the end of one branch matches the state at the beginning of its descendant branches), and ensures that character histories for terminal branches end in the observed state. We compute $P(\kappa_l \,|\, \theta, t_l)$ as illustrated in Figure 2. By convention, we use the stationary distribution of the Markov chain defined by Q as the probability of the root state, $P(x_1 \,|\, \theta)$.

To calculate the conditional probability density of the continuous characters, we first consider how the continuous characters, $\mathbf{y}$, evolve along a single branch of length t when the history of the discrete character along
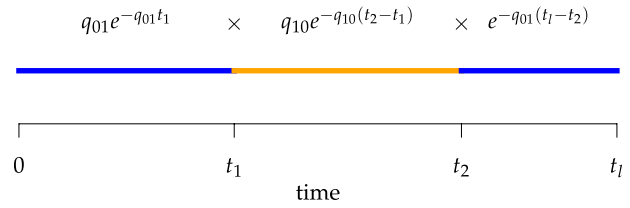
$$q_{01}e^{-q_{01}t_1} \quad \times \quad q_{10}e^{-q_{10}(t_2-t_1)} \quad \times \quad e^{-q_{01}(t_l-t_2)}$$

FIGURE 2. Computing the probability of a character history, κ_l. Blue and orange segments correspond to discrete states 0 and 1, respectively. The probability of the history is the product of the probabilities of waiting times between events (or the probability of no event in the final segment) given the current rate of change.

that branch is known. Given that the discrete character is in state i for duration t_i, changes in the continuous character follow a multivariate normal distribution with mean $\mathbf{0}$ and variance-covariance matrix $t_i \times \beta_l^2 \varsigma_i^2 \Sigma$. This implies that the changes in $\mathbf{y}$ while in discrete state i, Δ_i, are multivariate-normally distributed:

$$\Delta_0 \sim \mathrm{MVN}(\mathbf{0}, \beta_l^2 \tau(\kappa_l, 0) \varsigma_0^2 \Sigma)$$

$$\Delta_1 \sim \mathrm{MVN}(\mathbf{0}, \beta_l^2 \tau(\kappa_l, 1) \varsigma_1^2 \Sigma),$$

where $\tau(\kappa_l, i)$ is the amount of time the history spends in discrete state i and $\mathbf{0}$ is a $1 \times c$ vector of zeros (indicating that the expected amount of change for each continuous character is 0). Because it is the sum of multivariate-normally distributed random variables, $\Delta_0 + \Delta_1$ is also multivariate-normally distributed:

$$\Delta_0 + \Delta_1 \sim \mathrm{MVN}(\mathbf{0}, \Sigma_l),$$

where $\Sigma_l = \beta_l^2 [\tau(\kappa_l, 0) \varsigma_0^2 + \tau(\kappa_l, 1) \varsigma_1^2] \Sigma$ is the branch-specific variance-covariance matrix given the discrete-character history and the background rate of evolution for the branch, β_l^2.

Because changes to the continuous characters follow a multivariate normal distribution, we can compute the conditional probability density of the continuous characters, $P(\mathcal{Y} \,|\, \kappa, \theta, \Psi)$, using standard algorithms to integrate over the distributions of the states at internal nodes. Specifically, we use Felsenstein's REML algorithm (Felsenstein 1973, 2004), extended to multivariate Brownian motion (Huelsenbeck and Rannala 2003; Freckleton 2012), to compute the conditional probability density of $\mathcal{Y}$. This algorithm assumes a uniform prior over all possible continuous states at the root.

Incorporating background-rate variation in the MuSSCRat model does not complicate the computation of the augmented likelihood, since it simply "rescales" the variance-covariance matrices on each branch. However, including background-rate variation *does* complicate inference; specifically, it causes the MuSSCRat model to become nonidentifiable (see Supplementary Material available on Dryad at http://dx.doi.org/10.5061/dryad.499c4j2). A model is nonidentifiable when multiple combinations of parameters have identical likelihoods (Rannala 2002). Consequently, parameters of a nonidentifiable model cannot be estimated by standard maximum-likelihood methods

because there may be no unique "maximum" likelihood. In this case, it is necessary to apply constraints on nonidentifiable parameters that "penalize" different combinations of parameters that have identical likelihoods. Bayesian models provide a natural solution to nonidentifiability, as the prior distributions on the parameters naturally penalize combinations of parameters that might have identical likelihoods (i.e., the joint posterior probability of parameter combinations with identical likelihoods will differ if their joint prior probabilities are different). We describe our assumed prior distribution on background rates in the next section.

Priors.—We assume that the background-rate parameters, $\boldsymbol{\beta^2}$, are drawn from a hierarchical model with parameters μ and ν, and the remainder of the parameters are drawn from independent prior distributions, so that the joint prior density becomes:

$$P(\boldsymbol{\zeta^2}, \boldsymbol{\beta^2}, \mu, \nu, \boldsymbol{\sigma^2}, R, Q) =$$
$$P(\boldsymbol{\beta^2} \mid \mu, \nu) P(\mu) P(\nu) P(\boldsymbol{\zeta^2}) P(\boldsymbol{\sigma^2}) P(R) P(Q).$$

We describe our prior distributions in the following paragraphs; these parameterizations reflect our "baseline" model, but we explore alternative priors and prior sensitivity in our empirical analyses.

We draw the lineage-specific background rates of continuous-character evolution, $\boldsymbol{\beta^2}$, iid from a shared lognormal distribution with mean μ and standard deviation ν. We use a uniform prior on $\log_{10}(\mu)$, such that μ is drawn from a $\log_{10}$-uniform distribution between 10^{-3} and 10. We draw the standard deviation, ν, from an exponential distribution with mean H. The constant H is the standard deviation for a lognormal distribution that indicates that our 95% prior belief ranges over one order of magnitude (see Supplementary Material available on Dryad). This model is the continuous-character analog of the uncorrelated lognormal (UCLN) relaxed-clock model used to describe variation in rates of molecular evolution across lineages (e.g., Drummond et al. 2006; Lemey et al. 2010). Accordingly, we refer to this extension of the MuSSCRat model with background-rate variation as the MuSSCRat + UCLN model. A convenient property of the UCLN model is that—as ν shrinks to 0—it collapses to a "strict" morphological-clock model, where $\beta_l^2 = \mu$ for all lineages. Our prior on ν specifies that we expect the values of $\boldsymbol{\beta^2}$ to range over about one order of magnitude, but the exponential prior allows the standard deviation to shrink to 0 if the data prefer a strict morphological clock. In summary, we specify the background-rate-variation component of the prior model as:

$$\beta_l^2 \sim \text{Lognormal}(\mu, \nu)$$
$$\mu \sim \log_{10}\text{-uniform}(L = 10^{-3}, U = 10)$$
$$\nu \sim \text{Exponential}(1/H_\nu).$$

The parameter vector $\boldsymbol{\zeta^2}$ describes the relative rate of continuous-character evolution for each of the discrete states. We specify a Dirichlet distribution on *half* the values of $\boldsymbol{\zeta^2}$. Specifying the prior on half the values of $\boldsymbol{\zeta^2}$ ensures that the mean value of $\boldsymbol{\zeta^2}$ is 1, which allows us to interpret these parameters as the *relative* rate of continuous-character evolution in the alternative discrete states. We assume the concentration parameters of the Dirichlet distribution are the same, so this is a symmetric Dirichlet distribution with parameter $\omega = 1$:

$$\boldsymbol{\zeta^2}/2 \sim \text{symmetric Dirichlet}(\omega = 1).$$

The average rate of change for each of the c continuous characters may vary; we allow the relative rate of continuous characters to vary by including a parameter vector, $\boldsymbol{\sigma^2}$, where σ_i^2 is the rate of evolution of the ith continuous character. We specify a Dirichlet distribution on $1/c$th of the values of $\boldsymbol{\sigma^2}$. We adopt the same logic as above for the prior on $\boldsymbol{\zeta^2}$, specifying a symmetric Dirichlet distribution with parameter $\gamma = 1$ such that the mean value of $\boldsymbol{\sigma^2}$ is 1:

$$\boldsymbol{\sigma^2}/c \sim \text{symmetric Dirichlet}(\gamma = 1).$$

The symmetric matrix R determines the evolutionary correlation between each pair of continuous characters; $\rho_{ij} = \rho_{ji}$ is the correlation between characters i and j. The matrix R has a special constraint (it must be positive semidefinite) that makes it difficult to specify, for example, iid priors on each matrix element, ρ_{ij}. We use the LKJ distribution as a prior on R, which defines a prior over positive-semidefinite correlation matrices (Lewandowski et al. 2009). Correlation matrices drawn from this distribution have prior density:

$$P(R \mid \eta) \propto \det(R)^{\eta - 1},$$

where $\eta > 0$ is inversely related to the variance of the correlation parameters: larger values of η result in marginal distributions on ρ_{ij} that are concentrated closer to 0, while smaller values of η result in distributions that are more diffuse. We choose $\eta = 1$, which indicates a uniform distribution over all possible positive-semidefinite correlation matrices:

$$R \sim \text{LKJ}(\eta = 1).$$

Finally, the matrix Q describes the rates of change between the discrete-character states. We parameterize the stationary frequencies, $\boldsymbol{\pi}$, and the average rate of change, λ. We build Q from these parameters as follows:

$$Q = \lambda r \begin{pmatrix} - & \pi_1 \\ \pi_0 & - \end{pmatrix},$$

where the diagonal elements are specified so that the sum of each row is 0, and the scalar r is an arbitrary value that guarantees that the expected number of transitions over a tree of length T is λT.

Assuming that the rates of change are symmetric ($\pi_0 = \pi_1$) or asymmetric ($\pi_0 \neq \pi_1$) may have some impact

on our analysis through the distribution on κ. Moreover, inferring whether rates of change are (a)symmetric is often of direct interest to researchers studying discrete-character evolution. We therefore specify a mixture distribution on π, so that π may be symmetric or asymmetric. Specifically, we draw π, from a degenerate distribution concentrated on equal rates, $\delta(\frac{1}{2},\frac{1}{2})$, with probability $p = 0.5$, and from a Dirichlet distribution with parameter $\phi = 1$ with probability $1 - p = 0.5$. We draw λ from a lognormal prior distribution with standard deviation H, and specify the mean such that the expected number of transitions over the entire phylogeny is k. The prior expected number of transitions reflects an empirical prior, and should be specified differently for different data sets; for the simulations and analyses we describe later, we use $\mathbb{E}(k) = 5$. Our overall prior on Q is:

$$\pi \sim \begin{cases} \delta\left(\frac{1}{2},\frac{1}{2}\right) & \text{w.p. } p \\ \text{symmetric Dirichlet}(\phi = 1) & \text{w.p. } 1 - p \end{cases}$$

$$\lambda \sim \text{Lognormal}(\mathbb{E}(k) = 5, H).$$

Posterior.—Having specified the augmented likelihood function and the joint prior density, we can write down the joint posterior density of the model parameters and the discrete-character histories:

$$P(\zeta^2, \beta^2, \mu, \nu, \sigma^2, R, Q, \kappa \mid \mathcal{X}, \mathcal{Y}, \Psi) \propto \qquad (1)$$
$$P(\mathcal{X}, \kappa \mid Q, \Psi) P(\mathcal{Y} \mid \zeta^2, \beta^2, \sigma^2, R, \kappa, \Psi)$$
$$P(\zeta^2, \beta^2, \mu, \nu, \sigma^2, R, Q),$$

where the first two terms on the right-hand side are the augmented likelihood, and the third term is the joint prior density on the model parameters.

The above Bayesian model conditions on a known tree, Ψ. It is straightforward to relax the assumption of a fixed tree by including it as a parameter in the model. In this case, we may include a sequence alignment and specify a subsitution model, and jointly infer the phylogeny and the parameters of the `MuSSCRat` and substitution models.

Constant background rates.—The `MuSSCRat` model with constant-background rates is nested within the model with background-rate variation described above: as $\nu \to 0$, the lognormal prior on β^2 collapses to a point centered on μ (so that all values of β^2 for all branches become increasingly similar).

To specify the constant-background-rate model explicitly, we draw a single value for β^2 from a $\log_{10}$-uniform distribution between 10^{-3} and 10. Otherwise, we use the same prior distributions for the constant-background-rate model as we described for the variable-background-rate model, above.

Markov chain Monte Carlo.—The joint posterior probability density cannot be calculated analytically because we

cannot evaluate the normalizing constant of equation 1 (the marginal likelihood). We therefore approximate the joint posterior probability density numerically using MCMC; specifically, we draw samples from the joint posterior distribution using the Metropolis–Hastings and Green algorithm (Metropolis et al. 1953; Hastings 1970; Green 1995). We use standard proposal distributions for the majority of the model parameters; for brevity, we only provide details for two of our more uncommon proposal distributions—for moves between symmetric and asymmetric Q matrices, and for the discrete-character histories—in the Supplementary Material available on Dryad.

Our data-augmentation strategy involves including the complete history of the discrete character, κ, as a variable in the Markov chain. As such, the MCMC procedure includes proposals that change the discrete-character history. When a new character history, κ', is proposed, it is accepted with probability A, computed as:

$$A = \min\left[1, \frac{P(\mathcal{X}, \kappa' \mid \theta)}{P(\mathcal{X}, \kappa \mid \theta)} \times \frac{P(\mathcal{Y} \mid \kappa', \theta)}{P(\mathcal{Y} \mid \kappa, \theta)} \times \frac{f(\kappa')}{f(\kappa)}\right],$$

where $f(\kappa)$ is the distribution from which the new character history is drawn. Note that the probabilities of the discrete characters, $P(\mathcal{X}, \kappa \mid \theta)$, and continuous characters, $P(\mathcal{Y} \mid \kappa, \theta)$, both contribute to the probability that the proposed discrete-character history is accepted. Importantly, this means that the continuous characters are able to correctly influence the discrete-character histories; that is, we are correctly modeling the joint distribution of the discrete and continuous characters.

Implementation.—We implemented our `MuSSCRat` model in the open-source Bayesian phylogenetic software, `RevBayes` (Höhna et al. 2016). Our implementation relies upon the data-augmentation functionality developed in `RevBayes` by Michael J. Landis and Sebastian Höhna for discrete characters (unpublished), extended to accommodate phylogenetic uncertainty. Owing to the flexibility of `RevBayes`, our implementation allows users to explore the impact of binary or multistate discrete traits on rates of continuous-character evolution, provides tremendous flexibility for specifying priors, enables simultaneous inference of ancestral states for both discrete and continuous characters, and allows joint inference of the phylogeny, divergence times, and parameters of the `MuSSCRat` model. We provide `Rev` scripts for performing analyses under the `MuSSCRat` model in `RevBayes` (see Data Dryad repository http://dx.doi.org/10.5061/dryad.499c4j2 and GitHub repository https://github.com/mikeryanmay/musscrat_supp_archive/releases/tag/1.1.

STATISTICAL BEHAVIOR

The `MuSSCRat` model has many parameters relative to the number of observations, $\mathcal{X}$ and $\mathcal{Y}$. It is therefore

unclear how well this complex model can detect rate variation, or distinguish between state-dependent and background sources of rate variation. Accordingly, we performed a simulation study to characterize the statistical behavior of the state-dependent model. Specifically, we performed experiments to understand: 1) its ability to detect state-dependent rate variation in the absence of background-rate variation; 2) its ability to detect state-dependent rate variation in the presence of background-rate variation; 3) the cost of including background-rate variation in the model when background rates are actually constant, and; 4) the consequences of assuming background rates are constant when they are actually variable.

For the following analyses, we approximated the joint posterior probability density by running two replicate MCMC simulations for each simulated data set. We performed MCMC diagnosis to ensure that the joint posterior density was adequately approximated. We provide details of the MCMC simulations and MCMC diagnoses in the Supplementary Material available on Dryad.

Measures of Performance

Frequentist properties.—The frequentist interpretation of a Bayesian credible interval (CI) is that the true value of a parameter has a $100(1-\alpha)\%$ chance of being within the $100(1-\alpha)\%$ CI of its corresponding marginal posterior distribution (assuming the model is true, see Huelsenbeck and Rannala 2004). With this interpretation in mind, we assessed the frequentist properties of the 95% CI inferred for our simulated data sets (assuming the conventional significance level, $\alpha=0.05$). We define the *coverage probability* as the frequency with which the true value of a parameter is contained in the 95% CI, the *false-positive rate* as the frequency with which the true value is excluded from the 95% CI (one minus the coverage probability), and the *power* as the frequency with which a state-independent model is excluded from the 95% CI when the state-dependent model is true.

Accuracy and bias.—We assess the accuracy and bias of the posterior-mean estimate of ζ_1^2 using the percent-error statistic, defined as:

$$\text{PE} = 100\% \times \frac{\widehat{\zeta_1^2} - \zeta_1^2}{\zeta_1^2},$$

where ζ_1^2 is the true value of state-dependent rate for discrete state 1, and $\widehat{\zeta_1^2}$ is estimated value of the state-dependent rate for discrete state 1 (we use the mean of the corresponding marginal posterior distribution). Values of PE < 0 indicate an underestimate; conversely, values of PE > 0 indicate an overestimate.

Simulation Experiments

Experiment 1: Constant background rates.—We simulated data sets of different sizes (with $c=\{1,2,4,8\}$ continuous characters), over a variety of tree sizes (with $N=\{25,50,100\}$ species), and state-dependent rate-ratios, $\zeta_1^2/\zeta_0^2=\{1,2,4,8\}$ (where $\zeta_1^2/\zeta_0^2=1$ corresponds to the case when rates do not depend on the state of the discrete trait). For each combination of c, N, and ζ_1^2/ζ_0^2, we simulated 100 trees under a constant-rate birth–death process with a speciation rate of 1 and an extinction rate of 0.5 using the R (R Core Team 2017) package TESS (Höhna et al. 2015); we then rescaled each tree to have a root height of 1. We simulated discrete-character histories under a symmetric continuous-time Markov chain with a rate specified such that the expected number of transitions was 5. We then drew correlation matrices from an LKJ distribution with $\eta=1$, and relative rates for each continuous character from a symmetric Dirichlet distribution with $\gamma=1$. Finally, we simulated c continuous characters under a multivariate Brownian motion model assuming a background rate of 1. We provide more details for the simulation procedure in the Supplementary Material available on Dryad.

We analyzed each simulated data set in RevBayes under the MuSSCRat model, assuming the true phylogeny was known. We constrained the model so that the background rate was equal for each lineage (i.e., the constant-rate model), and excluded the standard deviation parameter, v, from the model. We estimated the remaining model parameters from the simulated data using the priors described in the Bayesian inference section, above. Since the generating model and the inference model both exclude background-rate variation, this simulation scenario reflects the performance of the method when the model is correctly specified.

The false-positive rate for this simulation experiment was 5.3% (i.e., when $\zeta_1^2/\zeta_0^2=1$; Fig. 3, the top row of the top three panels), which is indistinguishable from the expected 5% (two-tailed binomial test $P\neq0.05$, P-value≈0.6). Next, we computed the power when rates of continuous-character evolution varied among discrete states ($\zeta_1^2>\zeta_0^2$). The power ranged from $\approx21\%$ to 100% from the worst-case to the best-case scenarios. Predictably, power improved as the number of continuous characters and species increased; overall, the average power was 80.4% (Fig. 3, top three panels, excluding the top row). The posterior-mean estimate of ζ_1^2 was slightly biased for small numbers of continuous characters ($c\leq2$) but quickly converged to the true value as the number of characters and species increased (Fig. 4, top row of panels).

Experiment 2: Variable background rates.—In this simulation scenario, we reused all of the simulated trees, discrete-character histories, correlation matrices, and relative-rate parameters describing the degree of variation among continuous characters from the first simulation experiment (with constant background rates). In
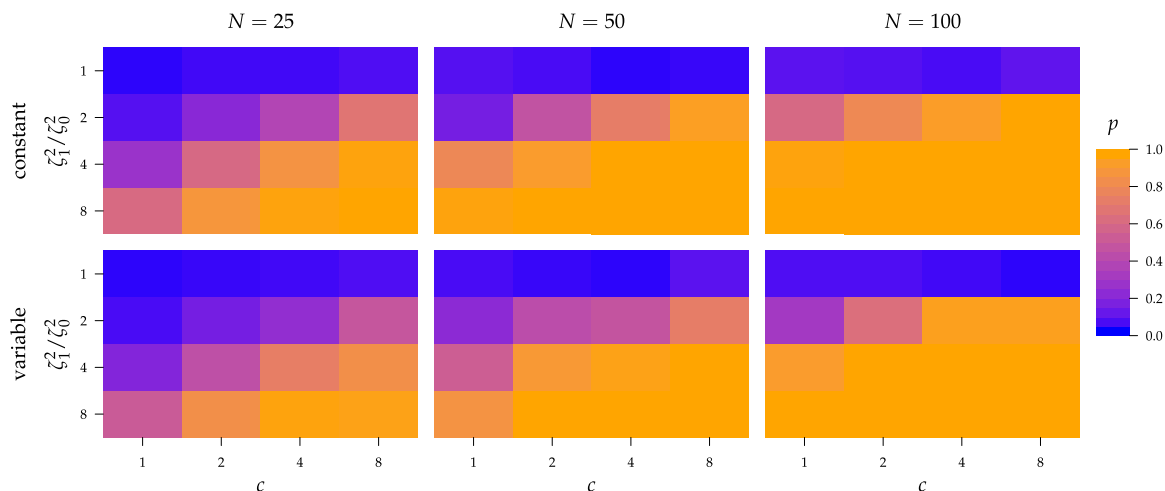
FIGURE 3. The frequency with which $\zeta_1^2/\zeta_0^2 = 1$ was excluded from the 95% CI when background rates were constant (top row of panels) or variable (bottom row of panels). Each panel corresponds to simulations for a given number of species, N. Within each panel, rows correspond to different degrees of state-dependent rate variation, ζ_1^2/ζ_0^2, and columns correspond to different numbers of continuous characters, c. Each cell represents the fraction of the 95% CI that exclude $\zeta_1^2/\zeta_0^2 = 1$, colored according to the scale (at right).

this simulation, however, we simulated lineage-specific rates of continuous-character evolution by drawing the background rate for each lineage, β_l^2, from a lognormal distribution with mean $\mu = 1$ and standard deviation $\nu = H$.

For this scenario, we analyzed each simulated data set using the MuSSCRat model with background-rate variation, by allowing β_l^2 to vary among branches, as described in the Bayesian inference section (MuSSCRat + UCLN). Again, this simulation scenario reflects the performance of the method when the model is correctly specified, since the data-generating model and the inference model both allow background rates to vary among lineages.

The false-positive rate for this experiment was 4.9% (Fig. 3, top row of bottom three panels), again indistinguishable from the expected 5% (two-tailed binomial test $P \neq 0.05$, P-value ≈ 1). The power was only slightly lower than that of experiment 1: the power was 14.7% in the worst case, and 100% in the best case. On average, the power was 73.7% (Fig. 3, bottom three panels, excluding top row). Again, the posterior-mean estimate of ζ_1^2 was only modestly biased for analyses based on a small number of continuous characters (Fig. 4, bottom row of panels).

Experiment 3: Cost of background-rate variation.—When background rates of continuous-character evolution are constant, we expect that including unnecessary parameters (i.e., to accommodate background-rate variation) in the inference model should decrease our ability to detect state-dependent rate variation. The goal of this simulation experiment is to understand the cost of accommodating background-rate variation when it is

absent. To achieve this, we reused the data sets from Experiment 1 (simulated under constant background rates) with $N = 50$, $c = 8$, and $\zeta_1^2/\zeta_0^2 = \{1, 2, 4, 8\}$, but analyzed these data sets under the MuSSCRat + UCLN model.

For Experiment 1—where we correctly assumed that background rates were constant—the coverage probability was 95.3% (two-tailed binomial test $P \neq 0.95$, P-value ≈ 0.4). By contrast, in this experiment—when we incorrectly assumed that background rates are variable—the coverage probability was 97.4% (two-tailed binomial test $P \neq 0.95$, P-value $= 0.02$). Overall, the cost of accommodating background-rate variation when absent was therefore quite modest (97.4% − 95.3% = 2.1%). Additionally, the posterior distributions of the background-rate-variation parameter, ν, shrunk strongly toward the true value, $\nu = 0$ (the constant-background-rate model): the average posterior-mean estimate of ν across these simulations was ≈ 0.21 (compared to a prior mean of ≈ 0.59; see Supplementary Material available on Dryad).

Experiment 4: Consequences of ignoring background-rate variation.—When background rates of continuous-character evolution vary among lineages, we expect that *excluding* background-rate variation from the inference model may be positively misleading. The goal of this simulation experiment is to understand the consequences of failing to accommodate background-rate variation on inferences about state-dependent rates of continuous-character evolution. To achieve this, we reused the data sets from Experiment 2 (simulated under variable background rates), with $N = 50$, $c = 8$, and $\zeta_1^2/\zeta_0^2 = \{1, 2, 4, 8\}$, but analyzed these data sets using the "constrained" MuSSCRat model (i.e., that assumes a
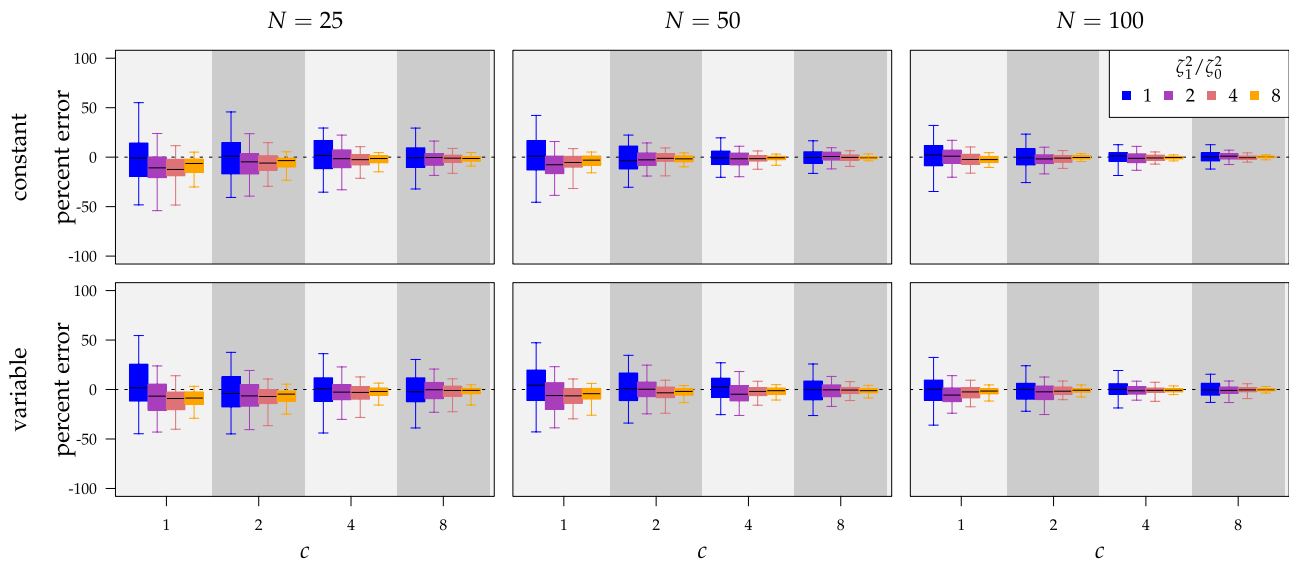
FIGURE 4.   The percent error of the posterior-mean estimates of ζ_1^2 when background rates were constant (top row of panels) or variable (bottom row of panels). Each column of panels corresponds to simulations for a given number of species, N. Within each panel, boxplots depict the distribution of percent error across 100 simulated data sets for each of the c continuous characters (along the x-axis), colored by the true state-dependent rates, ζ_1^2/ζ_0^2 (see inset legend). Boxplots represent the middle 50% (boxes) and the middle 95% (whiskers) of simulations.

constant background rate of evolution by forcing β_l^2 to be the same for all lineages).

For Experiment 2—where we correctly assumed that background rates are variable—the coverage probability was 94.5% (two-tailed binomial test $P \neq 0.95$, P-value ≈ 0.19). By contrast, in this experiment—when we incorrectly assumed that background rates are constant—the coverage probability decreased to 85.4% (two-tailed binomial test $P \neq 0.95$, P-value $< 1e-10$). This decreased coverage probability implies that we are very confident in the wrong answer about 10% more often than we should be. For example, when state-dependent rates are truly equal ($\zeta_1^2/\zeta_0^2 = 1$), we will incorrectly—but confidently—infer that state-dependent rates differ 17.7% of the time.

EMPIRICAL ANALYSES

Haemulids (grunts) are a group of percomorph fishes that have previously been used to explore state-dependent rates of continuous-character evolution (Price et al. 2013). Specifically, the hypothesis posits that—owing to the increased habitat complexity of reefs—the feeding apparatus (comprising several continuous traits) of reef-dwelling grunt species should evolve at a higher rate than that of their non-reef-dwelling relatives. We revisit this hypothesis by analyzing the haemulid data from Price et al. (2013) under the MuSSCRat model, using a phylogeny estimated from the more extensive molecular data set from Tavera et al. (2018).

Phylogenetic Analyses

We assembled a molecular data set by subsampling the alignments from Tavera et al. (2018) to include only the 49 species represented in our morphological data set. We estimated a chronogram under a partitioned substitution model assuming an uncorrelated lognormal branch-rate prior model and a sampled birth–death node-age prior model. We performed posterior-predictive tests to ensure that the substitution model provided an adequate description of the substitution process. We computed the maximum *a posteriori* (MAP) chronogram from the posterior distribution of sampled trees and conditioned on this tree in our comparative analyses. We provide details of these analyses in the Supplementary Material available on Dryad.

Comparative Analyses

We analyzed the continuous morphological data under the MuSSCRat model, with habitat type (reef/non-reef) as the discrete character. In these analyses, we conditioned on the MAP chronogram estimated above. We performed a series of analyses to understand: 1) the impact of including or excluding background-rate variation, and; 2) the sensitivity of posterior estimates to the specified priors. For the following analyses, we approximated the joint posterior density by running four replicate MCMC simulations for each analysis using RevBayes. Again, we provide details of the MCMC simulations and MCMC diagnoses in the Supplementary Material available on Dryad.
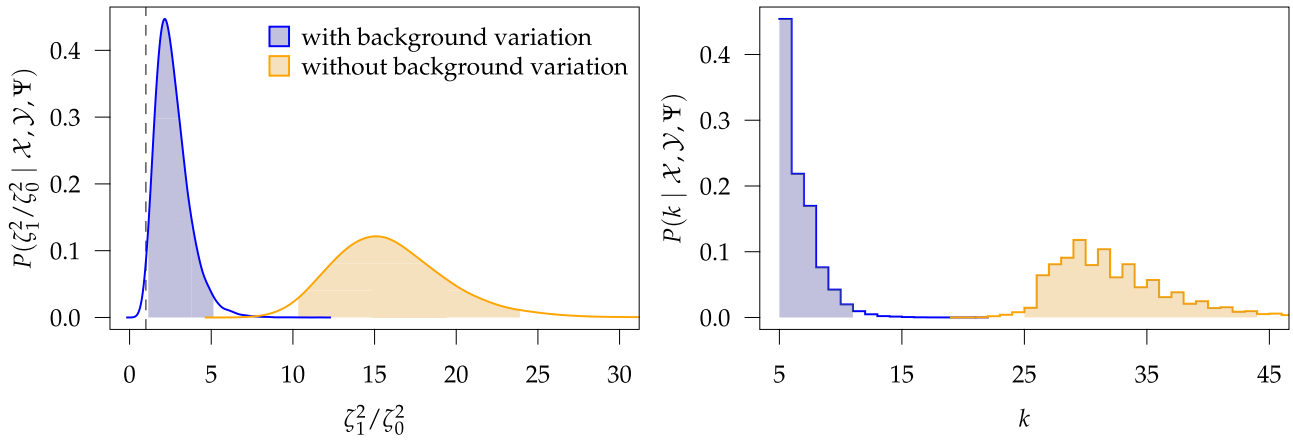
FIGURE 5.   At left, the posterior densities (curves) and the 95% CI (shaded regions) for the state-dependent rate-ratio, ζ_1^2/ζ_0^2, when the background-rates are constant (orange), or when they vary among lineages (blue), inferred for the haemulid data set. The dashed vertical line corresponds to $\zeta_1^2/\zeta_0^2 = 1$. At right, the posterior distribution (lines) and the 95% CI (shaded regions) for the number of habitat transitions, k, assuming the background-rates are constant (orange), or vary among lineages (blue).

Character data.—We used eight continuous morphological characters related to the feeding apparatus from Price et al. (2013); we included species that also had molecular sequence data from Tavera et al. (2018), resulting in a total of 49 species. The continuous characters include: 1) the mass of the adductor mandibulae muscle; 2) the length of the ascending process of the premaxilla; 3) the length of the longest gill raker; 4) the diameter of the eye; 5) the length of the buccal cavity; 6) the width of the buccal cavity; 7) the height of the head, and; 8) the length of the head. Rather than size correcting these characters, we included body size as an additional character (for a total of nine continuous characters). Following Price et al. (2013), we log-transformed each character before the analyses (and cube-rooted the adductor mass prior to log transformation). We used the habitat data from Price et al. (2013) to score each species for the binary discrete character; we coded non-reef-dwelling species and reef-dwelling species as states 0 and 1, respectively.

Inferring state-dependent rates.—To understand the impact of background-rate variation, we estimated the posterior distribution of the MuSSCRat model parameters with and without background-rate variation using the prior settings described in the Bayesian inference section.

The treatment of background-rate variation had a profound impact on both the habitat-specific rate of continuous-character evolution, and also on the inferred history of habitat evolution (Fig. 5). Under the MuSSCRat model without background-rate variation, we inferred that the feeding apparatus of reef-dwelling haemulids evolved ≈ 16 times faster than that of their non-reef-dwelling relatives; under the MuSSCRat + UCLN model, we inferred a ≈ 2.6-fold increase in the evolutionary rate of reef-dwelling species (95% CIs [10.3−23.9] and [1.1−5.1], respectively).

Examining the posterior distribution of habitat transitions reveals that excluding background-rate variation implies biologically implausible scenarios of habitat evolution. When we disallowed background-rate variation, we inferred ≈ 32 transitions between reef- and non-reef habitats across the phylogeny; when we allowed background rates to vary, we inferred a more reasonable ≈ 6.2 transitions (95% CIs [25−43] and [5−10], respectively). The inferred history of the habitat across the branches of the phylogeny was similarly distorted when we assumed that background rates did not vary (see Supplementary Material available on Dryad).

Prior sensitivity.—We assessed the prior sensitivity of inferences by performing a series of analyses using different prior values for various parameters of the model. Specifically, we explored the following prior values:

$$\omega \in \{1/4, 1/2, 1, 2, 4\}$$
$$\gamma \in \{1/4, 1/2, 1, 2, 4\}$$
$$\eta \in \{1/2, 1, 2, 5, 10\}$$
$$\mathbb{E}(k) \in \{1, 3, 5, 10, 20\}$$
$$\mathbb{E}(\nu) \in \{H \div 2, H, H \times 2\},$$

where $\mathbb{E}(\nu)$ is the prior expected standard deviation of the background-rate variation model. We varied a single prior setting at a time, rather than testing all possible combinations of these priors; we left the remaining priors as described in the Bayesian inference section, for a total of 23 prior combinations.

Most prior settings appear to have little impact on the posterior distribution of the focal parameter, ζ_1^2/ζ_0^2 (Fig. 6). Unsurprisingly, the prior on the focal parameter, ω, had the greatest influence on the state-dependent rate estimates: the posterior-mean estimate ranged from
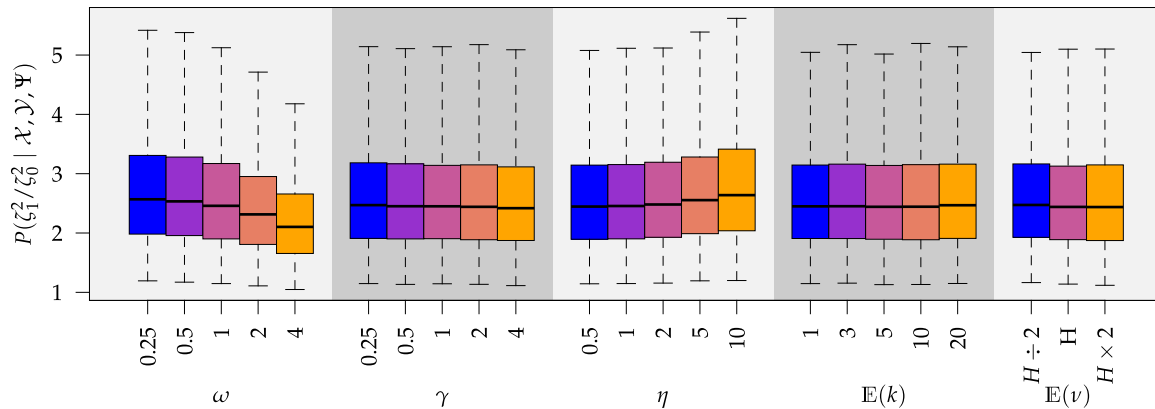
FIGURE 6. The posterior densities of the state-dependent rate-ratio, ζ_1^2/ζ_0^2, for the haemulid data set under various priors. Each band of boxplots corresponds to a different prior-sensitivity experiment. Within each band, boxplots represent the 50% CI (box) and 95% CI (whiskers) for the posterior density under a particular value of that prior.

2.23 to 2.75 over the priors that we tested (Fig. 6, left band); in all cases $\zeta_1^2/\zeta_0^2 = 1$ was excluded from the 95% CI. We discuss the (negligible) prior sensitivity of the remaining model parameters in the Supplementary Material available on Dryad.

DISCUSSION

Understanding the factors that drive variation in rates of character evolution is a fundamental goal for evolutionary comparative biologists. Current approaches for assessing the influence of a discrete character on rates of continuous-character evolution suffer from two problems: 1) they do not correctly characterize the mutually informative relationship between the discrete and continuous characters, and; 2) they compare against a simple—and likely unrealistic—null model, potentially misleading inferences about state-dependent rates due to rate variation that is unrelated to the discrete character of interest, which we term "background-rate variation". This second problem is especially concerning, given that rates of evolution are likely to vary greatly across the Tree of Life, and for many reasons not related to the discrete character a particular researcher is investigating.

We present a Bayesian method that deals with both of these issues using a model (MuSSCRat) that correctly integrates over discrete-character histories with extensions that accommodate background-rate variation. This method involves estimating a large number of parameters, especially compared to the size of typical morphological data sets. This raises serious questions about the reliability of inferences made using the method—especially because the background-rate variation model may wash out any signal of state-dependent rate variation—and also about the sensitivity of inferences to the choice of priors. In the following sections, we describe simulation and empirical results that shed light on the statistical behavior of the method.

Statistical behavior under simulation.—We explored the ability of the MuSSCRat model to infer state-dependent rates of continuous-character evolution using simulated data. We varied these simulations over the number of species, the number of continuous characters, and the degree of state-dependent rate variation. We repeated our simulations under different background-rate models: "background-constant" simulations, where background rates were the same across lineages, and "background-variable" simulations, where background rates were allowed to vary among lineages.

When the model was correctly specified (i.e., when we inferred parameters under the true background-rate model), the method had appropriate frequentist behavior: the false-positive rate was approximately 5%, and the power increased with the number of taxa, the number of continuous characters, and the degree of state-dependent rate variation. The power was modestly reduced for background-variable simulations ($\approx74\%$) compared to the background-constant simulations ($\approx80\%$). Posterior-mean estimates of the state-dependent rate parameters were biased only for small trees ($N = 25$) or data sets with only one or two continuous characters. These results suggest that researchers should be able to reliably infer the state-dependent rate parameters for data sets with a reasonable number of species and continuous characters.

We also used our simulated data sets to assess the costs of including background-rate variation, as well as the consequences of ignoring it. Including unnecessary parameters in the model (overspecification) should lead to increased uncertainty and a concomitant decrease in power; that is, for background-constant data, allowing for background-rate variation in the inference model should dampen the signal of state dependence. Conversely, excluding parameters from the model (underspecification) should lead to artifactually increased confidence and a higher false-positive rate;

that is, for background-variable data, an inference model that assumes that background rates are constant may spuriously interpret the unmodeled rate variation as additional evidence for state dependence. In our simulations, the cost of model overspecification (an ≈2% decrease in power) was minor compared to the consequences of model underspecification (an ≈10% increase in the false-positive rate).

Empirical impact of background-rate variation.—We reanalyzed the trophic-character data for the haemulids (grunts) from Price et al. (2013) with constant and variable background rates. The inclusion (or exclusion) of background-rate variation in the inference model had a profound impact on inferences regarding the degree of state-dependent rate variation. Under the constant-background-rate model, reef-dwelling species were inferred to evolve more than 15 times faster than their non-reef-dwelling relatives. By contrast, we inferred a ≈2.6-fold increase when we allowed background-rate variation. Furthermore, the history of the discrete character inferred under the constant-background-rate model involved an implausible number of habitat transitions. Together, these results suggest that, while trophic-character evolution within the haemulids is elevated within reefs, other factors (manifest as background-rate variation) also played an important role in the evolution of continuous traits in this group.

Benefits of being Bayesian.—Our Bayesian implementation comes with all of the usual advantages of Bayesian inference: the marginal posterior distributions for parameters have a natural interpretation (the 95% CI contains the true value of the parameter 95% of the time), and these estimates are automatically averaged over uncertainty in all of the parameters. Indeed, our implementation also allows us to accommodate uncertainty in the tree topology and divergence times, although the impact of phylogenetic uncertainty appears to be relatively mild for haemulids (see Supplementary Material available on Dryad).

Of course, the need to specify prior densities for each model parameter means that posterior estimates may be sensitive to arbitrary prior choices. However, at least for haemulids, inferences about state-dependent rates appear to be quite robust over a range of reasonable priors. We note that our results regarding the impact of phylogenetic uncertainty and prior sensitivity are data set specific, and may not be generally true for all/most data sets. For this reason, we urge users to perform similar sensitivity assessments for their empirical studies.

Beyond the prosaic strengths and weaknesses of Bayesian inference, adopting a Bayesian framework allowed us to overcome two critical issues for the MuSSCRat model. First, the joint distribution of the discrete and continuous characters implied by the MuSSCRat model makes it difficult (perhaps impossible) to calculate the full likelihood analytically. Specifying our model in a Bayesian framework allowed us to use a MCMC technique—data augmentation—to simplify these likelihood calculations and correctly describe the joint evolution of discrete and continuous characters. We note that similar Monte Carlo integration techniques might be practical for maximum-likelihood applications; indeed, these Monte Carlo solutions have been used to perform maximum-likelihood inference for similar types of problems (Mayrose and Otto 2010; Levy Karin et al. 2017). We conducted experiments that suggest that such approaches would be unreliable for the haemulid data set (see Supplementary Material available on Dryad), but more work is necessary to understand the generality of these results.

Second, adopting a Bayesian framework allowed us to include background-rate variation while retaining the ability to perform reliable inference under the model. The MuSSCRat model with background-rate variation is inherently nonidentifiable: multiple combinations of background-rate and state-dependent-rate parameters can have identical likelihoods, so a unique maximum-likelihood estimate of the parameters may not exist. Within a Bayesian setting, the joint prior distribution acts to "tease apart" combinations of parameters that would otherwise have identical likelihoods, thus making it possible to infer parameters under nonidentifiable models.

Broader context.—There is mounting concern within the phylogenetic comparative community that methods for understanding relationships between evolutionary variables can be unreliable in the face of model misspecification (Beaulieu et al. 2013; Maddison and FitzJohn 2014; Rabosky and Goldberg 2015; Beaulieu and O'Meara 2016; Uyeda et al. 2018). One particular problem, identified by Rabosky and Goldberg (2015), is that our model-selection procedures often involve comparisons against an extremely unrealistic—and therefore easy-to-reject—null model.

Beaulieu and O'Meara (2016) clarify the fundamental issue raised by Rabosky and Goldberg (2015): when asked to choose between two models, it should come as no surprise when model-comparison procedures reject an overly simplistic, constant-rate null model in favor of a very specific, variable-rate alternative model. The danger is that *any* rate variation (whether or not it is associated with a focal variable of interest) will be interpreted by a model-comparison procedure as evidence against a constant-rate null model. This logic also applies to parameter-estimation procedures: when considering a variable-rate model with a single explanatory variable, it seems likely that any evidence for heterogeneity is at risk of being spuriously attributed to the factor of interest. We refer to this problem as the "straw-man effect," and we suspect that it applies whenever model comparison includes at least one constant-rate model or—in the case of parameter estimation—whenever a variable-rate model is overly simplistic.

A possible solution to the straw-man effect—one that we favor, but by no means the only conceivable solution—is to move away from null-model hypothesis testing with overly specific variable-rate models toward more general hierarchical models that include various sources of rate variation, as we have done in the present work. A justifiable concern with such complex hierarchical models is that the results may be sensitive to the assumed nature of the background variation—whether it is due to lineage-specific effects, heritable (but unobserved) biological traits, the environment, or any number of alternatives—which may be difficult to justify *a priori* or distinguish *a posteriori*.

Future prospects.—Our `MuSSCRat` model makes many simplifying assumptions about the nature of the evolutionary process under consideration. For example, we assume that the continuous characters evolve under a Brownian motion process; that the underlying evolutionary variance-covariance matrix, Σ, is the same over the entire tree; that the discrete characters evolve under a simple continuous-time Markov process; and that the background-rate variation is adequately described by an uncorrelated lognormal distribution. Each of these assumptions provides opportunities for future model development.

There are many alternatives to the Brownian motion model in the phylogenetic comparative toolkit, perhaps chief among them the Ornstein–Uhlenbeck (OU) process (Hansen 1997; Butler and King 2004). State-dependent OU process models have been widely used to detect shifts in evolutionary optima associated with discrete characters. It seems that these approaches are vulnerable to unmodeled process heterogeneity, since they typically compare state-dependent models against homogeneous models (as demonstrated by Uyeda et al. 2018). Extending the data-augmentation framework presented here to a variable-optimum OU process (e.g., as described in Uyeda et al. 2018) should be straightforward.

We have assumed that the "structure" of the evolutionary variance-covariance matrix is independent of the discrete character (in the sense that the discrete character affects only the magnitude of the variance-covariance matrix). Recently, Caetano and Harmon (2019) developed a method that allows the structure of the variance-covariance matrix to depend on the state of a discrete character. While this is an important advance over models that assume that the variance-covariance matrix is homogeneous, it too may suffer from the straw-man effect: heterogeneity in the variance-covariance matrix unrelated to the discrete character may be positively misleading. Extending the framework developed by Caetano and Harmon (2019) to incorporate background variation in the variance-covariance matrix is an important avenue for future development.

We used a relatively simple continuous-time Markov process to model the evolution of the discrete character.

In reality, it may be that the rate or process of discrete-character evolution varies over the tree (Beaulieu et al. 2013), or that the discrete character impacts rates of lineage diversification (Maddison 2006), or violates the assumptions of a Markov model in some other way. The extent to which these complex models of discrete-character evolution impact the distribution of discrete-character histories—and therefore compromise inferences about discrete-state-dependent rates of continuous-character evolution—is an open question, though we expect the impact might be mild for the haemulids (see Supplementary Material available on Dryad). Although the modeling task may be straightforward, developing the computational machinery to do inference under such models—where a discrete character affects both rates of continuous-character evolution while itself evolving under a complex process—is a nontrivial technical challenge.

Finally, we have assumed that background-rate variation follows an uncorrelated lognormal prior distribution. A strong assumption of this model is that rates of evolution are uncorrelated between lineages; however, if the rate of character evolution is itself evolving, then we might expect rates to be correlated between ancestors and descendants (as in the model described by Huelsenbeck et al. 2000). Because the `MuSSCRat` model with background-rate variation is nonidentifiable, specifying a prior on background-rate variation is critical to teasing apart the relative contributions of background- and state-dependent effects on the overall rate variation. However, the Bayesian solution to nonidentifiability is not without caveats: since we rely on the prior to distinguish the relative effects of background-rate variation and state-dependent rates, our inferences about state-dependent rates may be sensitive to the assumed model of background-rate variation. Indeed, we performed exploratory analyses of the haemulid data set using an alternative background-rate model that allowed for autocorrelated rates (following Eastman et al. 2011) and found that posterior estimates were fairly sensitive to this aspect of the model (see Supplementary Material available on Dryad). Unfortunately, it is not clear that alternative models of background-rate variation can be reliably distinguished using standard model-selection procedures. However, Bayesian models provide a natural way for updating our prior beliefs and integrating information across disparate data sets, and we foresee a fruitful Bayesian research program that leverages information from across the Tree of Life (extinct and extant) to characterize overall heterogeneity in the evolutionary process, and to more accurately distinguish among its multifarious causes.

SUPPLEMENTARY MATERIAL

Supplementary scripts and data (including the Haemulidae data and simulated data used in this

study) can be found in the Dryad Digital repository: http://dx.doi.org/10.5061/dryad.499c4j2 and the GitHub repository https://github.com/mikeryanmay/musscrat_supp_archive/releases/tag/bioRxiv1.0.

References

Beaulieu J.M., O'Meara B.C. 2016. Detecting hidden diversification shifts in models of trait-dependent speciation and extinction. Syst. Biol. 65(4):583–601.

Beaulieu, J.M., O'Meara, B.C., Donoghue, M.J. 2013. Identifying hidden rate changes in the evolution of a binary morphological character: the evolution of plant habit in campanulid angiosperms. Syst. Biol. 62(5):725–737.

Butler, M.A., King, A.A. 2004. Phylogenetic comparative analysis: a modeling approach for adaptive evolution. Am. Nat. 164(6):683–695.

Caetano, D.S., Harmon, L.J. 2019. Estimating correlated rates of trait evolution with uncertainty. Syst. Biol. 68(3):412–429.

Collar, D., Schulte, J., O'Meara, B., and Losos, J. 2010. Habitat use affects morphological diversification in dragon lizards. J. Evol. Biol. 23(5):1033–1049.

Collar, D.C., O'Meara, B.C., Wainwright, P.C., and Near, T.J. 2009. Piscivory limits diversification of feeding morphology in centrarchid fishes. Evolution 63(6):1557–1573.

Drummond, A.J., Ho, S.Y., Phillips, M.J., and Rambaut, A. 2006. Relaxed phylogenetics and dating with confidence. PLoS Biol. 4(5):e88.

Eastman, J.M., Alfaro, M.E., Joyce, P., Hipp, A.L., and Harmon, L.J. 2011. A novel comparative method for identifying shifts in the rate of character evolution on trees. Evolution 65(12):3578–3589.

Felsenstein, J. 1973. Maximum-likelihood estimation of evolutionary trees from continuous characters. Am. J. Hum. Genetics 25(5):471.

Felsenstein, J. 2004. Inferring phylogenies Vol.. 2. Sunderland (MA): Sinauer Associates.

Freckleton, R.P. 2012. Fast likelihood calculations for comparative analyses. Methods Ecol. Evol. 3(5):940–947.

Green, P.J. 1995. Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. Biometrika 82(4):711–732.

Hansen, T.F. 1997. Stabilizing selection and the comparative analysis of adaptation. Evolution 51(5):1341–1351.

Harmon, L.J., Losos, J.B., Jonathan Davies, T., Gillespie, R.G., Gittleman, J.L., Bryan Jennings, W., Kozak, K.H., McPeek, M.A., Moreno-Roark, F., Near, T.J., Purvis, A. 2010. Early bursts of body size and shape evolution are rare in comparative data. Evolution 64(8):2385–2396.

Hastings, W.K. 1970. Monte Carlo sampling methods using Markov chains and their applications. Biometrika 57(1):97–109.

Höhna, S., Landis, M.J., Heath, T.A., Boussau, B., Lartillot, N., Moore, B.R., Huelsenbeck, J.P., and Ronquist, F. 2016. RevBayes: Bayesian phylogenetic inference using graphical models and an interactive model-specification language. Syst. Biol. 65(4):726–736.

Höhna, S., May, M.R., and Moore, B.R. 2015. TESS: an R package for efficiently simulating phylogenetic trees and performing Bayesian inference of lineage diversification rates. Bioinformatics 32(5):789–791.

Huelsenbeck, J.P., Larget, B., and Swofford, D. 2000. A compound Poisson process for relaxing the molecular clock. Genetics 154(4):1879–1892.

Huelsenbeck, J.P., Nielsen, R., and Bollback, J.P. 2003. Stochastic mapping of morphological characters. Syst. Biol. 52(2):131–158.

Huelsenbeck, J.P. and Rannala, B. 2003. Detecting correlation between characters in a comparative analysis with uncertain phylogeny. Evolution 57(6):1237–1247.

Huelsenbeck, J.P. and Rannala, B. 2004. Frequentist properties of Bayesian posterior probabilities of phylogenetic trees under simple and complex substitution models. Syst. Biol. 53(6):904–913.

Landis, M.J., Matzke, N.J., Moore, B.R., and Huelsenbeck, J.P. 2013. Bayesian analysis of biogeography when the number of areas is large. Syst. Biol. 62(6):789–804.

Lartillot, N. 2006. Conjugate Gibbs sampling for Bayesian phylogenetic models. J. Comput. Biol. 13(10):1701–1722.

Lemey, P., Rambaut, A., Welch, J.J., and Suchard, M.A. 2010. Phylogeography takes a relaxed random walk in continuous space and time. Mol. Biol. Evol. 27(8):1877–1885.

Levy Karin, E., Wicke, S., Pupko, T., and Mayrose, I. 2017. An integrated model of phenotypic trait changes and site-specific sequence evolution. Syst. Biol. 66(6):917–933.

Lewandowski, D., Kurowicka, D., and Joe, H. 2009. Generating random correlation matrices based on vines and extended onion method. J. Multivariate Anal. 100(9):1989–2001.

Maddison, W.P. 2006. Confounding asymmetries in evolutionary diversification and character change. Evolution 60(8):1743–1746.

Maddison, W.P. and FitzJohn, R.G. 2014. The unsolved challenge to phylogenetic correlation tests for categorical characters. Syst. Biol. 64(1):127–136.

Mateiu, L. and Rannala, B. 2006. Inferring complex DNA substitution processes on phylogenies using uniformization and data augmentation. Syst. Biol. 55(2):259–269.

Mayrose, I. and Otto, S.P. 2010. A likelihood method for detecting trait-dependent shifts in the rate of molecular evolution. Mol. Biol. Evol. 28(1):759–770.

Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H., and Teller, E. 1953. Equation of state calculations by fast computing machines. J. Chem. Phys. 21(6):1087–1092.

Nielsen, R. 2002. Mapping mutations on phylogenies. Syst. Biol. 51(5):729–739.

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

Price, S.A., Holzman, R., Near, T.J., and Wainwright, P.C. 2011. Coral reefs promote the evolution of morphological diversity and ecological novelty in labrid fishes. Ecol. Lett. 14(5):462–469.

Price, S.A., Tavera, J.J., Near, T.J., Wainwright, P.C. 2013. Elevated rates of morphological and functional diversification in reef-dwelling haemulid fishes. Evolution 67(2):417–428.

R Core Team 2017. R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.

Rabosky, D.L. and Goldberg, E.E. 2015. Model inadequacy and mistaken inferences of trait-dependent speciation. Syst. Biol. 64(2):340–355.

Rannala, B. 2002. Identifiability of parameters in MCMC Bayesian inference of phylogeny. Syst. Biol. 51(5):754–760.

Revell, L.J. 2012. A comment on the use of stochastic character maps to estimate evolutionary rate variation in a continuously valued trait. Syst. Biol. 62(2):339–345.

Robinson, D.M., Jones, D.T., Kishino, H., Goldman, N., and Thorne, J.L. 2003. Protein evolution with dependence among codons due to tertiary structure. Mol. Biol. Evol. 20(10):1692–1704.

Tanner, M.A. and Wong, W.H. 1987. The calculation of posterior distributions by data augmentation. J. Am. Stat. Assoc. 82(398):528–540.

Tavera, J., Acero, A., and Wainwright, P.C. 2018. Multilocus phylogeny, divergence times, and a major role for the benthic-to-pelagic axis in the diversification of grunts (Haemulidae). Mol. Phylogenet. Evol. 121:212–223.

Uyeda, J.C., Zenil-Ferguson, R., and Pennell, M.W. 2018. Rethinking phylogenetic comparative methods. Syst. Biol., 67(6):1091–1109.

Venditti, C., Meade, A., and Pagel, M. 2011. Multiple routes to mammalian diversity. Nature 479(7373):393.