Version dated: January 10, 2017

RH: ESTIMATING MACROEVOLUTIONARY LANDSCAPES

A GENERAL MODEL FOR ESTIMATING MACROEVOLUTIONARY LANDSCAPES

Florian C. Boucher[1,2], Vincent Démery[3], Elena Conti[1], Luke J. Harmon[4,5], and Josef Uyeda[4]

[1]*Department of Systematic and Evolutionary Botany (ISEB), University of Zurich, Zurich, Switzerland*

[2]*Department of Botany and Zoology, University of Stellenbosch, Stellenbosch, South Africa*

[3]*Gulliver, CNRS, ESPCI Paris, PSL Research University, 10 rue Vauquelin, Paris, France*

[4]*Department of Biological Sciences and Institute for Bioinformatics and Evolutionary Studies (IBEST), University of Idaho, Moscow, Idaho, USA*

[5]*Department of Fish Ecology and Evolution, Swiss Federal Institute of Aquatic Science and Technology (Eawag), Center for Ecology, Evolution, and Biogeochemistry, 6047 Kastanienbaum, Switzerland*

Corresponding author: Florian C. Boucher, Department of Botany and Zoology, University of Stellenbosch, Private Bag X1, Matieland 7602, South Africa; E-mail: flofloboucher@gmail.com.

Abstract.—

The evolution of quantitative characters over long timescales is often studied using stochastic diffusion models. The current toolbox available to students of macroevolution is however limited to two main models: Brownian motion and the Ornstein-Uhlenbeck process, plus some of their extensions. Here we present a very general model for inferring the dynamics of quantitative characters evolving under both random diffusion and deterministic forces of any possible shape and strength, which can accommodate interesting evolutionary scenarios like directional trends, disruptive selection, or macroevolutionary landscapes with multiple peaks. This model, which we call BBM+V, is based on a general stochastic differential equation widely used in statistical mechanics: the Fokker-Planck equation. We first explain how this model can be used to describe macroevolutionary landscapes over which quantitative traits evolve and, more importantly, we detail how it can be fitted to empirical data. Using simulations, we show that the model has good behavior both in terms of discrimination from alternative models and in terms of parameter inference. We provide R code to fit the model to empirical data using either maximum-likelihood or Bayesian estimation, and illustrate the use of this code with two empirical examples of body mass evolution in mammals. BBM+V should greatly expand the set macroevolutionary scenarios that can be studied since it opens the way to estimating macroevolutionary landscapes of any conceivable shape.
(Keywords: adaptation, bounds, macroevolution, maximum-likelihood estimation, MCMC methods, selection, phylogenetic comparative data)

Introduction

Understanding the evolution of phenotypes over geological timescales is one of the fundamental goals of macroevolution (Simpson (1944)). Phenotypic evolution is typically inferred either from time series of measurements obtained in the fossil record (Hunt (2007)) or from the distribution of phenotypic characters at the tips of a phylogenetic tree (O'Meara (2012)). In both cases, one then fits stochastic models for the evolution of traits on single lineages or on phylogenies, all of which treat trait evolution as a diffusion process that may or may not be influenced by deterministic forces (O'Meara (2012)).

Many approaches attempt to bridge the gap between microevolutionary process and macroevolutionary pattern by interpreting model parameters in the context of the dynamics of evolution on adaptive landscapes (Wright (1932); Simpson (1944); Arnold et al. (2001)). Recent years have revitalized this connection with the development of numerous methodological tools specifically aimed at inferring 'macroevolutionary landscapes' (Hansen et al. (2008); Ingram and Mahler (2013); Eastman et al. (2013); Uyeda and Harmon (2014)). Such landscapes almost certainly do not reflect static landscapes upon which populations evolve over long time scales; instead, these landscapes are most productively described as representing the movements of adaptive peaks over million-year time scales (Hansen (1997); Uyeda et al. (2011); Uyeda and Harmon (2014)). In particular, a peak on such a landscape might not be a phenotypic optimum in any particular generation of evolution of a lineage, but instead might represent a long-term average peak location on a dynamic landscape. Throughout this paper, we will refer to these as 'macroevolutionary landscapes', which summarize patterns of trait evolution averaged over many generations.

Comparative methods to infer macroevolutionary landscapes are all based on the Ornstein-Ulhenbeck (OU) process (Hansen (1997)), which was itself strongly inspired by the original concept of adaptive landscape in population genetics (Lande (1976)). Under

the OU process, a continuous trait evolves under both random diffusion (*i.e.* Brownian motion, Edwards and Cavalli-Sforza (1964)) and a force that brings back the trait close to an optimal value. Following models from quantitative genetics (e.g. Lande (1976)), these two components of the macroevolutionary OU process are sometimes interpreted as genetic drift and stabilizing selection. However, such an interpretation is almost always overly simplistic. First, many other processes can generate evolution following an OU model. For example, the shape of the peak and, in turn, the dynamics of selection and drift within populations may be less important for long-term patterns than the movement of the peak itself. Under such a scenario, both the diffusion and deterministic components of OU reflect peak movement, and both are strongly dependent on the dynamics of both selection and drift. Second, the actual parameters of OU models are almost always incompatible with Lande's model of evolution on a static adaptive landscape (Estes and Arnold (2007); Uyeda and Harmon (2014)). However, even if we do not interpret *diffusion* as *drift* and *determinism* as *selection*, it is still useful to divide macroevolutionary dynamics into these two components. Any factor that leads to trait change that is random in direction from one generation to the next (e.g. drift, randomly varying selection, plasticity due to random environmental noise) will affect the diffusion component, and any factor that leads to predictable change towards some particular value (e.g. selection, predictable patterns of peak movement over time, developmental constraints towards certain values) will be seen in the deterministic component. Various extensions of the OU model have been proposed in recent years, including different optima in different clades, either determined a priori (Butler and King (2004)) or not (Uyeda and Harmon (2014)), varying rates of diffusion and intensities of attraction towards optima in different clades (Beaulieu et al. (2012)), or evolution of the optimum itself (Hansen et al. (2008)). The complexity of all of these extensions, however, leads to difficulties in model identifiability and parameter estimation (e.g. Khabbazian et al. (2016)). While these models cover a wide range of possible scenarios, they are restricted to two main kinds of adaptive landscapes: (i)

adaptive landscapes with a single peak moving in time (Hansen et al. (2008)) and (ii) adaptive landscapes with one or several static peaks, eventually of varying heights (varying attraction strengths) and widths (varying ratios of diffusion rate to attraction strength). Importantly, in the second case the different peaks are experienced by different clades, so that no single lineage evolves in an adaptive landscape with multiple peaks (see below). In addition, another type of adaptive landscape might be described by the bounded Brownian motion model (BBM, Boucher and Démery (2016)). BBM was developed as a model of neutral evolution between bounds, but it could also describe adaptive landscapes in which one part of phenotypic space (*i.e.*, traits values between the bounds) has high but constant fitness while other phenotypes have null fitness, a scenario related to holey adaptive landscapes (Gavrilets (1997)).

As can be seen from this short overview of existing methods, the types of macroevolutionary landscapes that can be estimated from comparative data are still rather limited. The case of disruptive selection is particularly alarming. Disruptive selection is indeed central to the theories of ecological speciation (Doebeli (1996); Nosil (2012)) and adaptive radiation (Schluter (2000)), in which diverging lineages adapt to different ecological niches. These theories are being increasingly appreciated in the macroevolution community and ecological speciation and/or adaptive radiation are frequently invoked as explanations for the diversity of extant clades (Soulebeau et al. (2015)). However, despite recent theoretical advances in modeling interspecific interactions over macroevolutionary timescales (e.g. Nuismer and Harmon (2015); Drury et al. (2016)), we still lack proper tools to infer disruptive selection on phylogenies. As a result of this, macroevolutionary landscapes that contain multiple local optima cannot be inferred from phylogenetic comparative data. The current alternative is to model phenotypic evolution in multimodal macroevolutionary landscapes using OU models with multiple optima (Butler and King (2004); Uyeda and Harmon (2014)). In this framework, cases in which transitions between peaks are frequent can be interpreted as evidence of

a changing of the adaptive environment, rather than the existence of multiple, simultaneously existing peaks among which lineages alternate. Such a model of adaptive evolution in which several adaptive peaks are simultaneously present in the adaptive landscape experienced by all species in the clade would be a realistic possibility.

In this paper we introduce a general model for the evolution of continuous characters on phylogenies that can accommodate macroevolutionary landscapes of any shape and thus attempts to provide solutions to the limitations mentioned above. In this model the continuous trait of interest evolves under random diffusion but is also subject to deterministic change following a macroevolutionary landscape of any possible shape and strength. We present algorithms for both maximum-likelihood and Bayesian estimation of model parameters, *i.e.* the value of the trait at the root of the tree, the position of the two bounds, the diffusion rate and the shape of the macroevolutionary landscape. Using simulations, we show that this model is easily distinguishable from other models of trait evolution like BM, OU, and BBM. Parameter estimation is also generally reliable, and in particular the shape of the macroevolutionary landscape can be accurately estimated as long as it has been fully explored by the clade evolving on it. We also show how alternative hypotheses can be statistically tested in empirical datasets. Our approach opens new avenues for macroevolutionary research: it renders possible the detection of evolutionary trends from neontological data only, but also inference of disruptive selection or of even more complex scenarios in which the macroevolutionary landscape contains multiple peaks.

Development and implementation of the model

General presentation of the BBM+V model

We introduce a general model for the evolution of continuous traits on phylogenies, in which the trait undergoes a random walk constrained between two reflective bounds and is biased by a deterministic force that can be of any shape and strength. This model is an extension of the bounded Brownian motion model for the evolution of continuous traits on phylogenies (hereafter BBM, Boucher and Démery (2016)). Under BBM, the character of interest evolves under Brownian motion (BM, Edwards and Cavalli-Sforza (1964)) with a constant evolutionary rate σ^2, but is confined between two reflective bounds, $B_{\min}$ and $B_{\max}$. The probability density of a trait evolving under BBM can be calculated using the method of images (Jackson (1998)), *i.e.* by cutting and reflecting the Gaussian probability density of a standard BM model an infinite number of times at each one of the two bounds. The characteristic time of the process, τ, is given by

$$\tau = \frac{(B_{\max} - B_{\min})^2}{\sigma^2}. \tag{1}$$

Over short time periods, *i.e.* for $t \ll \tau$, the trait does not have time to hit the bounds and effectively evolves under BM. Over long time periods, *i.e.* for $t \gg \tau$, the trait crosses the whole interval many times and the probability density reaches a uniform distribution (the stationary distribution of BBM).

The extension that we propose in this article is bounded Brownian motion in a non-trivial macroevolutionary landscape, a model that we label BBM+V. In this model, the character of interest evolves under BBM but there is also a force that is exerted in each point of the trait interval. This force derives from a potential $V(x)$, defined over the interval

$[B_{\min}, B_{\max}]$. Differences in the values of the potential generate a force that attracts the trait towards regions of trait space in which the potential is the lowest: at each point x, the BBM process is biased by a force proportional to $-V'(x)$. Like BBM, the BBM+V model can be modeled using the Fokker-Planck equation, a partial differential equation widely used in statistical mechanics to describe the time evolution of the probability density of an observable under the influence of both random and deterministic forces (Risken (1984)). Under BBM+V, the probability density $p(x, x_0, t)$ of the position of a trait x with initial value x_0 after time t has elapsed follows:

$$\frac{\partial p}{\partial t}(x, x_0, t) = \frac{\sigma^2}{2}\frac{\partial}{\partial x}\left[\frac{\partial p}{\partial x}(x, x_0, t) + p(x, x_0, t)\frac{\partial V}{\partial x}(x)\right]. \tag{2}$$

In this equation, the evolution of the probability density of the trait (left hand side) is determined by both random diffusion (*i.e.*, BM; first term on the right hand side) and a deterministic force set by the derivative of the potential (second term on the right hand side). The factor $\sigma^2/2$ on the front of the second term in the right hand side is chosen so that the stationary distribution for the probability density is

$$p^*(x) = \lim_{t\to\infty} p(x, x_0, t) = \mathcal{N}\exp(-V(x)), \tag{3}$$

where $\mathcal{N}$ is a normalization factor. Finally, the initial position of the trait, x_0, gives the following initial condition:

$$p(x, x_0, 0) = \delta(x - x_0), \tag{4}$$

where δ is the Dirac distribution, which is zero everywhere except when $x = x_0$ and has an integral of one.

In summary, the potential $V(x)$ determines the force $-\sigma^2 V'(x)/2$ that is exerted on the trait over the interval, and the process has a stationary distribution, which is proportional

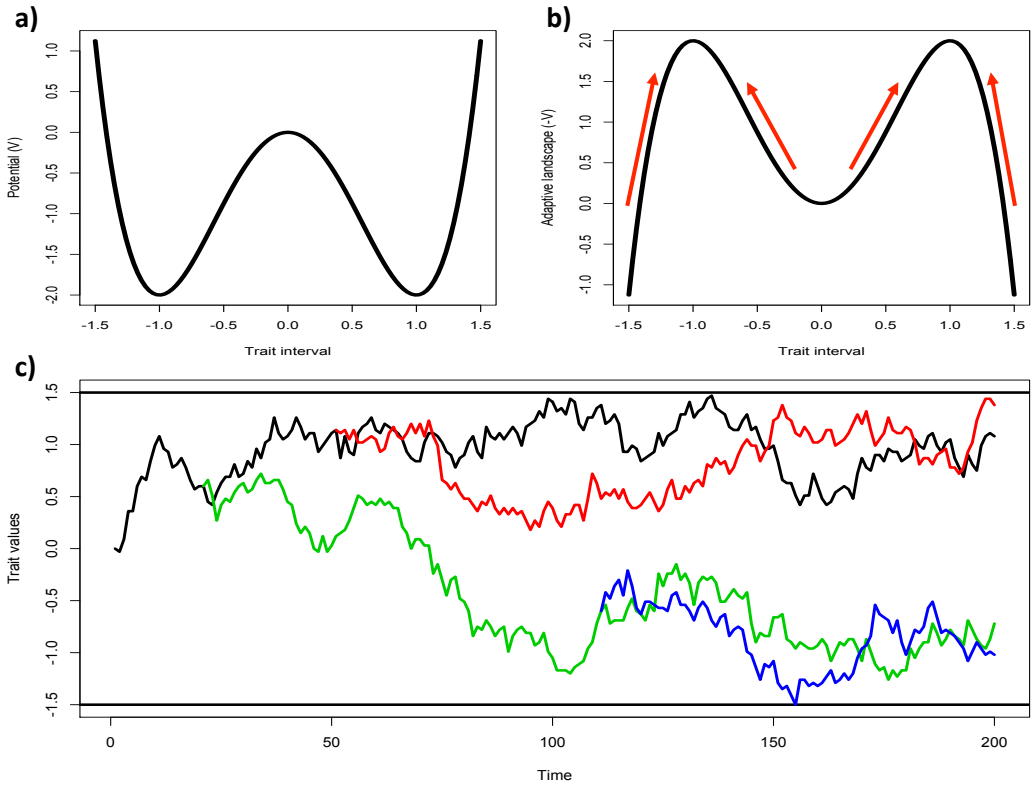

Figure 1: Behaviour of the BBM+V model. a) Example of a potential with $V(x) = 2x^4 - 4x^2$. b) The opposite of the potential, $-V(x)$, can be interpreted as a macroevolutionary landscape. Wells of the potential become peaks in the macroevolutionary landscape. Differences in the potential generate a force $-V'(x)$, which attracts trait values towards the two peaks of the macroevolutionary landscape, as indicated by the arrows. c) One simulation of the evolution of a trait in a clade of four species in this macroevolutionary landscape: the x-axis shows time and the y-axis the trait value of each species, in different colors. The ancestral trait value lies between the two peaks, and species' traits get attracted towards one of the peaks, which have equal heights. The traits of three of the species hit one of the bounds of the trait interval during their evolution.

to $\exp(-V(x))$ but does not depend on σ^2. The force represents the deterministic component of trait evolution, since it pulls traits towards specific values. In turn, the opposite of the potential, $-V(x)$, can be interpreted as a macroevolutionary landscape because trait values are attracted towards regions of trait space with the lowest potential, which themselves corresponds to peaks in $-V(x)$. Figure 1 shows how the potential, the deterministic forces, and the macroevolutionary landscape are related. In the remainder of this article, we will use the term *macroevolutionary landscape* to refer to $-V(x)$ and will avoid mentioning the potential ($V(x)$) as much as possible. It is important to note here that the evolutionary rate, σ^2, is not a measure of the strength of the random component of the process. Indeed, σ^2 determines both the intensity of the random diffusion component (first term in the right side of Eq. 2) and of the deterministic force exerted on the trait (second term in the right side of Eq. 2). The relative strengths of the random and deterministic components of the process are better captured by the variations in the stationary distribution of the process, $p^*_{\max}/p^*_{\min} = \exp(V_{\max} - V_{\min})$: a ratio close to one means that diffusion dominates, while deterministic forces are important if this ratio is large.

As in the BBM, we can define a characteristic time T_c representing the time it takes for the trait to explore the macroevolutionary landscape. Its precise definition is given below, but its most important feature is that very low values of the macroevolutionary landscape in the middle of the trait interval can considerably slow down the exploration. For instance, if two maxima are separated by a local minimum with potential difference ΔV, the characteristic time follows the Arrhenius law for large ΔV: $T_c \sim \exp(\Delta V)$ (Gardiner (1985)). Over short time periods, *i.e.* for $t \ll T_c$, the trait neither has time to fully explore the macroevolutionary landscape nor does it hit the bounds of the trait interval. Over long time periods, *i.e.* for $t \gg T_c$, the trait crosses the whole interval many times and explores the whole macroevolutionary landscape; the probability density thus reaches its stationary distribution (Eq. 3).

BBM is a special case of BBM+V in which the potential is constant over the whole trait interval: no force is exerted on the trait and it simply diffuses randomly between the two bounds of the interval. Thus, under BBM the macroevolutionary landscape is flat.

Calculation of the likelihood

The likelihood of the BBM+V model given a phylogenetic tree and observed values of the trait at the tips of the tree is obtained by multiplying the probability densities along each branch of the tree and integrating over all possible trait values at the internal nodes:

$$\mathcal{L} = \int_{B_{\min}}^{B_{\max}} \left(\prod_{i \in I \cup T} p(x_i, x_{parent(i)}, t_i - t_{parent(i)}) \right) \prod_{i \in I} dx_i, \tag{5}$$

where I is the set of internal nodes (excluding the root), T is the set of tips, x_i is the value of the trait at the node i, t_i is the time at node i and $parent(i)$ is the parent of the node i. Computing the likelihood numerically thus requires integrating over all possible values of the trait at interior nodes of the tree, which makes it computationally challenging. Since the distribution of the trait at the tips of the tree is not multivariate normal, fast methods like Generalized Least-Squares (Grafen (1989)), phylogenetic independent contrasts (Felsenstein (1973); Freckleton (2012)), or the 3-point algorithm (Ho and Ané (2014b)) cannot be used either.

To compute the likelihood of BBM+V, we instead discretize the trait interval by considering only a set of n points equally spaced between $B_{\min}$ and $B_{\max}$, a procedure already used for BBM (Boucher and Démery (2016)). In the following of this article, we call this regular set of points the *grid*. The continuous evolution equation for the probability density (2) can then be cast in a matrix form

$$\frac{dP}{dt}(t) = \frac{(n-1)^2}{2\tau} M P(t), \tag{6}$$

where $\tau = (B_{\max} - B_{\min})^2/\sigma^2$, $P(t)$ is a matrix whose components $P_{ij}(t)$ describe the probability of the trait being in position i on the grid at time t starting from position j at time 0, and M is a transition matrix specifying the probability for the trait to evolve from one of these n values on the grid to another during an infinitesimal time step according to the BBM+V model. We will now show how to relate the coefficients of M to the continuous equation (2).

During an infinitesimal time step the trait can only move between neighboring sites, leading to our first condition: $M_{ij} = 0$ if $|i - j| > 1$. The next condition comes from the fact that the probability flux between two sites vanishes for the stationary distribution, which reads in its discrete form $P_i^* = \mathcal{N} \exp(-V_i)$. The probability flux between the neighboring sites i and $i + 1$ is $M_{i+1,i}P_i^* - M_{i,i+1}P_{i+1}^*$, leading to a second condition for the coefficients of M:

$$\frac{M_{i+1,i}}{M_{i,i+1}} = \exp(V_i - V_{i+1}) \quad \forall i. \tag{7}$$

We must also ensure that M conserves probabilities, *i.e.* that the probabilities for a trait to evolve from site i to any of the sites on the grid sums to one. This can be written as $1 = \sum_i P_{ij}(t) = U^{\mathrm{T}} P_j(t)$, where U is a vector whose components are all equal to 1 ($\forall i$, $U_i = 1$) and $P_j(t)$ is a vector whose components are $[P_j(t)]_i = P_{ij}(t)$. Taking the time derivative of $U^{\mathrm{T}} P_j(t) = 1$ and using the evolution equation 6 leads to $U^{\mathrm{T}} M P_j(t) = 0$. This relation being satisfied for all j and t, we end up with a second general condition on M, $U^{\mathrm{T}} M = 0$ or $\sum_i M_{ij} = 0$, which can be used to set the last condition:

$$M_{ii} = -(M_{i-1,i} + M_{i+1,i}) \quad \forall i. \tag{8}$$

It is straightforward to show that

$$M_{ij} = \exp\left(\frac{V_j - V_i}{2}\right) \quad \text{for} \quad |i - j| = 1, \tag{9}$$

satisfies the conditions obtained above. Eq. 8 is used to determine M_{ii} and all others terms are null. Online Appendix III in Boucher and Démery (2016) shows how these discretized equations converge to the continuous one (2) as we increase the number of points used to discretize the trait interval (*i.e.*, $n \to \infty$).

The expression of the evolution of probability density as a first order matricial differential equation (Eq. 6) allows us to give a precise definition of the characteristic time. We can diagonalize the matrix M, noting its eigenvalues $\lambda_1 \geq \cdots \geq \lambda_n$. The largest one corresponds to the stationary state, $\lambda_1 = 0$, and all others are strictly negative. Online Appendix I shows that the characteristic time of the BBM+V process is defined by the second largest eigenvalue, λ_2, via $T_c = \frac{2\tau}{(n-1)^2|\lambda_2|}$.

Note that the use of the discretization procedure presented above is the only reason why we impose that trait evolution is bounded. Indeed, Eqs. 3,5 actually describe analytical solutions for the model without bounds, but it would not be possible to calculate its likelihood numerically. The precision of the discretization procedure (*i.e.*, the number of points used to discretize the trait interval) naturally influences the precision of the numerical calculation of the likelihood (Boucher and Démery (2016)) and the accuracy in the estimation of the shape of the macroevolutionary landscape, but on the other hand calculation grows quadratically with the number of points (the transition matrix M has n^2 terms). In the rest of this article all simulations have been run with $n = 50$ points between the minimum and maximum values of the trait at the tips of the tree (but the actual number of points used to discretize the trait interval might be higher if bounds are inferred to lie far away from the observed trait interval, see below). This value was used because of the large number of simulations we ran, but we generally recommend people working on a single empirical case to increase this number if they have the computational facilities to do so.

Once the transition matrix between different points on the grid is calculated, we calculate the likelihood of the model as is done for the evolution of discrete characters (*i.e.*,

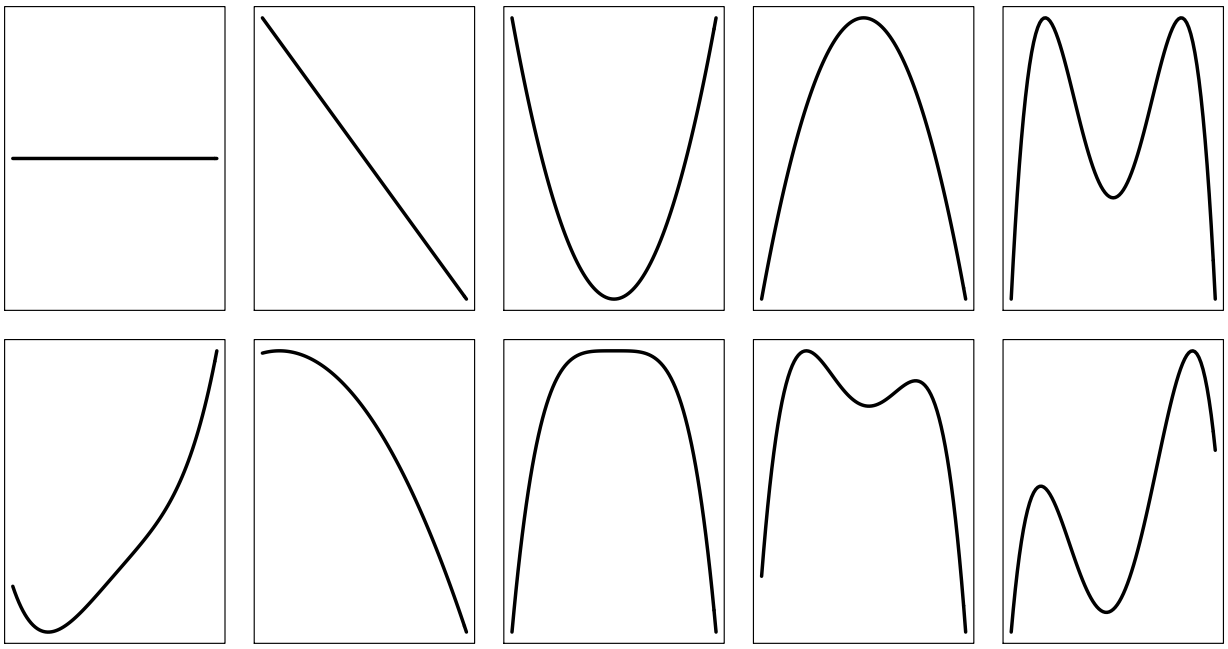

Figure 2: Ten contrasted shapes of macroevolutionary landscapes that can be obtained using a parametric form of the potential with only three parameters: $V(x) = ax^4 + bx^2 + cx$. The first row shows five typical evolutionary scenarios, from left to right: random drift between bounds (*i.e.*, BBM), an evolutionary trend towards small trait values, disruptive selection, a single peak, and two peaks. The second row shows five other possible macroevolutionary landscapes. The x-axis shows the trait value and the y-axis the opposite of the potential, $-V(x)$. Units have been omitted for clarity.

the Mk model, Lewis (2001)). We use the pruning algorithm (Felsenstein (1973)) to calculate the likelihood of the model, starting from the probability density of the trait at the tips, and propagating it down to the root of the tree. Finally, we treat the value of the trait at the root of the phylogenetic tree, x_0, as a parameter of the BBM+V model. This makes comparison with other models of evolution possible, since most implementations include the root value as a parameter.

Shape of the macroevolutionary landscape

The model we have presented above can accommodate any shape of the macroevolutionary landscape. However, in order to infer macroevolutionary landscapes from empirical

data we need to specify a parametric shape for it and optimize its parameters. One possibility to do so would have been to use step functions with different values at each point in the trait grid, but this would have lead to a very large number of parameters to estimate. Using combinations of sine functions of various periods and amplitudes would also have been possible, but their periodicity renders optimization difficult. Instead we chose a polynomial function with only three terms, such that $V(x) = ax^4 + bx^2 + cx$, with x taken on the interval $[-1.5, +1.5]$. This interval was chosen arbitrarily but any symmetric interval would give a similar macroevolutionary landscape (see Online Appendix II). This function can indeed approximate a variety of scenarios, including flat landscapes (*i.e.*, BBM, $V(x) = 0$), linear trends (e.g., $V(x) = x$), domed (e.g., $V(x) = -x^2$) or U-shaped (e.g., $V(x) = x^2$) macroevolutionary landscapes, but also macroevolutionary landscapes with two central peaks of equal (e.g., $V(x) = x^4 - x^2$) or different heights (e.g., $V(x) = x^4 - x^2 + x$). Figure 2 shows a variety of shapes of the macroevolutionary landscape ($-V(x)$) that can be obtained with this parametric function. No constant term needs to be added to this polynomial function, since it is the derivative of $V(x)$ that controls the dynamics of the model (see above).

Maximum-likelihood inference of model parameters

We have implemented maximum-likelihood estimation of BBM+V model in the R statistical environment (R Core Team (2016)). All functions needed to fit the model to empirical data are freely available from the following Github repository: `https://github.com/fcboucher/BBMV`, and only depend on functions from the ape package (Paradis et al. (2004)). We have verified that the likelihoods for BBM+V obtained from our code are compatible with likelihoods for other models of trait evolution implemented in the *fitContinuous* function of package geiger (Pennell et al. (2014)). This makes comparison between BBM+V and other evolutionary models like BM or OU possible, using the Akaike information criterion (AIC) for example. Maximum-likelihood inference of model parameters is conducted

using the *optim* function in R with the Nelder-Mead optimization routine (*i.e.*, the simplex method, although our code also allows for optimization using the BFGS method with box constraints), which was found to perform better than other optimization routines following preliminary tests. Optimization of the values of the two bounds is complicated: allowing them to vary continuously in trait space leads to the grid of points used in the discretization procedure changing, which can lead to numerical errors. Rather, we decided to fix a step size and allow the bounds to move on this grid further and further away from the minimum and maximum value of the observed trait, up to ten times the width of the observed trait interval. For better numerical precision, we do not directly optimize the evolutionary rate, σ^2, but rather $\log(\sigma^2/2)$.

Maximum-likelihood estimation is complex since the model has seven parameters: the value of the trait at the root of the tree x_0, the evolutionary rate σ^2, the values of the two bounds of the traits interval, and the three coefficients determining the shape of the macroevolutionary landscape. In order to prevent failure of the optimization procedure, we proceed in three steps: (i) we first run six optimizations of the likelihood from different starting points (varying σ^2 and the position of bounds, always starting from a flat potential), plus one in which bounds are fixed to the minimum and maximum values of the observed trait (*i.e.*, they are not optimized); (ii) from these seven optimization runs, we then select the one that reached the highest likelihood, and run a final optimization step starting from its maximum-likelihood parameters; (iii) in cases where this final optimization run did not reach convergence we run another optimization step starting from the final values of the parameters obtained in previous one - this process is repeated until convergence is reached, up to a maximum of ten times. This implementation is at least eight times slower than running just one optimization, but helps to avoid erroneous inferences as much as possible.

In order to test complex macroevolutionary landscapes against simpler ones, we have also written R functions for fitting the BBM+V model with simpler macroevolutionary

landscapes, *i.e.* $V(x) = bx^2+cx$, $V(x) = cx$, and $V(x) = 0$ (BBM). Alternatives shapes of the macroevolutionary landscape can then be statistically compared based on their likelihoods, using likelihood ratio tests or any information criterion.

Finally, for all versions of the macroevolutionary landscape, confidence intervals containing the 95% highest probability density around parameter estimates while fixing other parameters to their maximum likelihood estimate can be calculated. This is technically done by removing the lowest 2.5% density regions on each side of the MLE for σ^2, the parameters describing the shape of the macroevolutionary landscape, and the root value when its MLE does not lie in one of the bounds of the trait interval or removing the lowest 5% density region for the bounds and the root value when its MLE lies in one of the bounds of the trait interval. These confidence intervals can be returned along with likelihood profile plots around parameter estimates.

MCMC algorithm

In addition to maximum-likelihood optimization, we present a MCMC algorithm to estimate parameters in BBM+V, which is also written in R. Since the aim of this MCMC algorithm will often be to get an idea of the distribution of parameter estimates, we have focused on the full model with three polynomial terms. However, nested models with simpler macroevolutionary landscapes can also be fit by setting the probability of update for unnecessary parameters to zero. Numerical calculation of the likelihood of BBM+V in our MCMC implementation is done as for the maximum-likelihood case and we use the Metropolis-Hastings algorithm to create a Markov chain of parameter estimates.

Parameters of the BBM+V model have different natures: the three coefficients determining the shape of the macroevolutionary landscape (a, b, and c) as well as the diffusion coefficient $\log(\sigma^2/2)$ are continuous variables, while the root value of the trait, x_0, and the positions of the bounds of the trait interval, $B_{\min}$ and $B_{\max}$, are only allowed to vary on a

regular grid of points (see above). These different parameters thus have different kinds of prior and proposal functions.

We have implemented two prior distributions for continuously varying parameters: either normal or uniform ones. However, one should keep in mind that very large values of a and b in particular (*i.e.*, the coefficients of the x^4 and x^2 terms) can lead to extremely steep macroevolutionary landscapes, which will be unrealistic in most cases. For these two parameters at least, a normal prior centered on zero thus seems to be the most sensible choice. For the three remaining parameters (x_0, $B_{\min}$, and $B_{\max}$) we have only implemented discrete uniform priors, which by default cover a regular grid extending from 10 points to the left of the minimum value of the trait observed at the tips of the tree to 10 points to the right of the maximum value of the trait observed at the tips of the tree.

As for proposal functions, both normal deviates and sliding windows are available for continuously distributed parameters (but other proposals could easily be implemented by modifying our R code). Note that since we actually update $\log(\sigma^2/2)$, this corresponds in fact to multiplier proposals for σ^2. For other parameters that are forced to be on the grid of points (*i.e.*, x_0, $B_{\min}$ and $B_{\max}$), only discrete sliding windows are possible and a move is forced to occur. The sensitivity of these proposal functions can be set by the user, but we recommend that discrete sliding windows only allow moves of one step on the grid. Parameters of the model are updated independently, but in order to speed up convergence of the Markov chain to its stationary distribution the relative frequencies of update of the different parameters can be modified.

In practice, we have observed two contrasting behaviors for convergence: (i) when T_c is larger than the depth of the phylogenetic tree, σ^2 will converge rapidly in the MCMC chain while parameters setting the shape of the macroevolutionary landscape will not ; (ii) on the opposite, when T_c is smaller than the depth of the phylogenetic tree, parameters setting the shape of the macroevolutionary landscape will converge rapidly while σ^2 will

be slow to converge. These two behaviors simply reflect the fact that when T_c is large relative to the total time span of trait evolution, the distribution of traits at the tips of the phylogeny will not have converged to the stationary distribution of the BBM+V model: the macroevolutionary landscape is poorly explored and tip values retain very little information regarding its shape. On the contrary, when T_c is small, the macroevolutionary landscape will have been thoroughly explored by the clade, but it is difficult to determine the value of the evolutionary rate with precision (see below). Manipulating the relative frequencies at which these different parameters are updated can have dramatic effects on the speed of convergence of the MCMC chain. A good way to set this tuning parameter would be either to first run a maximum-likelihood estimation of the model or to do an initial quick MCMC run (for example with a very low number of points to discretize the trait interval) in order to get an idea of the values of σ^2 and V. From our experience, convergence of our MCMC algorithm takes a long time, even on rather small datasets (c. 200,000 to 1 million steps). We recommend running at least two independent chains in order to make sure that they have converged to the same stationary distribution.

Performance of the BBM+V model

In order to assess the performance of BBM+V in terms of parameter inference and model discrimination, we have conducted a large number of simulations. Our focus was on the likelihood of BBM+V, thus we restricted our simulations to the maximum-likelihood optimization procedure since it is much faster than MCMC estimation.

As we have seen above, the macroevolutionary landscape that we infer can take a variety of shapes (Figure 2). We restricted our simulations to four contrasted scenarios that cover a broad range of interesting cases: (a) a flat macroevolutionary landscape (BBM, $V(x) = 0$), (b) a directional trend ($V(x) = 1.5x$), (c) disruptive selection (*i.e.*, a U-shaped macroevolutionary landscapes with extreme trait values being favored, $V(x) = -1.2x^2$), and (d) two peaks of the same height ($V(x) = 3x^4 - 6x^2$). For each one of these four scenarios, we fit four different versions of BBM+V: the full model ($V(x) = ax^4 + bx^2 + cx$), a model with only quadratic and linear terms ($V(x) = bx^2 + cx$), a model with only a linear term ($V(x) = cx$), and a model with a flat macroevolutionary landscape (*i.e.*, BBM, $V(x) = 0$). Note that these four versions of BBM+V being fitted exactly correspond to the four scenarios simulated. In addition, we also fit BM and an OU model with a single optimum (hereafter OU1) to each simulated dataset using the *fitContinuous* function in the geiger package (Pennell et al. (2014)).

Phylogenetic trees were simulated under a pure birth model, with unit birth rate. Trees were grown until the desired number of tips was obtained and were then rescaled to a total depth of 100 arbitrary time units in order to enable comparison between simulations. All simulations were done with $B_{\min} = -5$, $B_{\max} = 5$, and $x_0 = 0$. However, for each of the four scenarios described above we used two different values of σ^2 which were calculated so that: (i) $T_c = 5$ (*i.e.*, 1/20 of tree depth), which should ensure that the distribution of the trait at the tips of the tree has converged to the stationary distribution of the model,

and (ii) $T_c = 2,000$ (*i.e.*, 20 times tree depth), in which stationarity should not have been reached at the end of the simulation. In addition, we also explored the effect of tree size on parameter estimation and model discrimination using trees of 50, 75, 100, and 200 tips. For each combination of the shape of the potential, the value of σ^2, and tree size, we conducted 20 different simulations (640 simulations in total).

Model discrimination

We first focused on whether BBM+V can be distinguished from other models of evolution using the Akaike Information Criterion corrected for small sample size (AICc). Our simulations showed that when stationarity was reached ($T_c = 5$), all four scenarios of the BBM+V model that we simulated could easily be discriminated from BM and OU1, even in trees with only 50 tips: the median ΔAICc with the simulated model in this least favorable case was 76.6 for BM and 37.5 for OU1 (Online Appendix Figure 1). Under these conditions, a flat macroevolutionary landscape (BBM) was the scenario that was the least discriminated from both BM and OU1 (median ΔAICc=56.8 and 17.2, respectively). Discrimination between BBM+V and BM or OU1 increased with the number of tips in the tree (Online Appendix Figure 1). In simulations where stationarity was not reached ($T_c = 2,000$), only a scenario with two peaks could still be discriminated from BM and OU1 (median ΔAICc=8.4 and 10.6, respectively, for trees of 50 tips). However, simulations indicated that other shapes of the macroevolutionary landscape in the BBM+V model cannot be confidently discriminated from BM and OU1 (median ΔAICc=0.2 and 1.7, respectively, for trees of 50 tips) when these were simulated but stationarity was not reached. The number of tips in the tree did not strongly influence discriminatory power in these three last scenarios (Online Appendix Figure 1).

Discrimination between BBM+V models with different shapes of the macroevolutionary landscape was also generally satisfactory (Figure 3, see also Online Appendix Figure 2).

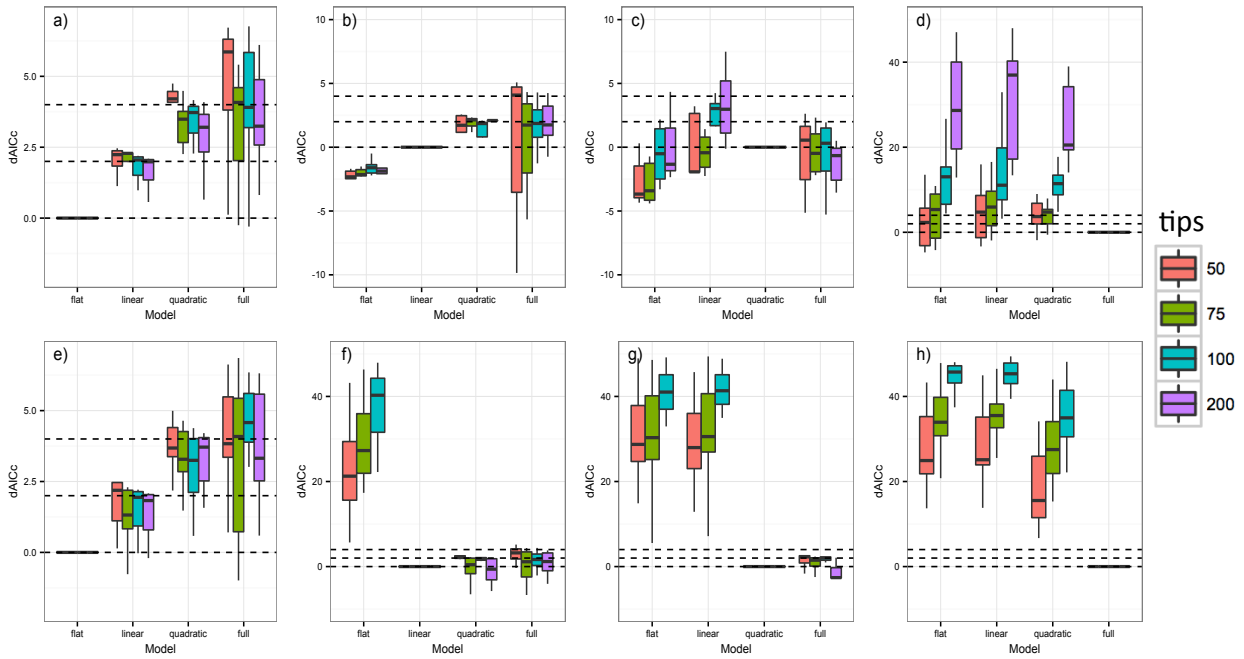

Figure 3: Discrimination between BBM+V models with different shapes of the macroevolutionary landscape. Each plot shows the distribution of ΔAICc values between the generating model and the three alternative ones over 20 simulations. a-d) Simulations with $T_c = 2,000$, in which stationarity was not reached. e-h) Simulations with $T_c = 5$, in which stationarity was reached. Generating models are presented as follow: a) and e) show a flat macroevolutionary landscape (BBM); b) and f) show a linear landscape (directional trend); c) and g) show a quadratic landscape (disruptive selection; d) and h) show a landscape with two peaks. Dashed horizontal lines show values of ΔAICc=0, 2, and 4. ΔAICc values have been truncated to improve visualization of small values and plots for a BBM model are on a scale that differs from the one used for all other scenarios. In some cases, ΔAICc values for one model and for trees with 200 tips do not appear on the plot because all values exceeded the maximum y-value shown. The complete figure, with no values truncated, is available in Online Appendix Figure 2.

In cases where stationarity was reached ($T_c = 5$), power to distinguish between alternative shapes of the macroevolutionary landscape was high and generally increased with the number of tips in the phylogeny. However, simulations showed a tendency for the most complex forms of BBM+V (*i.e.*, $V(x) = bx^2 + cx$ and $V(x) = ax^4 + bx^2 + cx$) to fit equally or better than the simulated model in cases in which traits were simulated under the linear model, and for the most complex form of BBM+V (*i.e.*, $V(x) = ax^4 + bx^2 + cx$) to fit equally or better than the simulated model in cases in which traits were simulated under the quadratic model (Figure 3). This is probably because complex forms of the macroevolutionary landscape can easily accommodate simpler scenarios, and perhaps more importantly because they can accommodate small deviations from the stationary distribution that inevitably occur in these stochastic simulations. In contrast, when stationarity was not reached ($T_c = 2,000$) different shapes of the macroevolutionary landscape could not be discriminated and simpler forms were always preferred over more complex ones, especially so in trees with few tips (Figure 3). The only exception came from simulations under a macroevolutionary landscape with two peaks, which could be discriminated from simpler versions of the BBM+V model even when stationarity was not reached for trees with 100 or 200 tips (Figure 3).

Parameter inference

Accuracy of BBM+V models in parameter estimation was assessed by comparing the maximum-likelihood estimates of parameters with values used in simulations. For simulations in which stationarity was reached ($T_c = 5$), we found that estimates of the position of the two bounds of the trait interval are generally accurate and rarely exceed the actual values of the bounds used in simulations, regardless of the number of tips in the tree (Online Appendix Figures 3 & 4). Estimation accuracy was of course lower when stationarity had not been reached (Online Appendix Figures 3 & 4). The most accurate estimates of bounds are obtained in cases where peaks lie at one or both of the bounds, *i.e.*, for the bound towards

which selection is exerted in the directional trend model (case b, $B_{\min}$ in our simulations) and for both bounds in the disruptive selection model (case c). In these cases all estimates always equal the simulated value, which is expected since some traits in the clade will inevitably lie very close to the bound and none of the trait values exceed the bound by definition, even in cases where stationarity has not been reached. Accuracy in the estimation of x_0 is much less satisfactory and usually has a huge variance, especially so when T_c is small (Online Appendix Figure 5). The estimation of σ^2 is very accurate when T_c is large but has much larger variance when T_c is large (Online Appendix Figure 6).

Finally, we compared the precision in the estimation of the macroevolutionary landscape as a whole, and not in the estimation of the three parameters that determine the potential separately. This is because these three coefficients can sometimes be redundant and lead to very similar shapes of the macroevolutionary landscape: for example, a (the coefficient of the x^4 term) and b (the coefficient of the x^2 term) are highly correlated. Simulations showed that macroevolutionary landscapes are generally accurately estimated by BBM+V in cases where stationarity has been reached since the actual shape that was simulated is most often recovered (Figure 4). Accuracy was much worse in simulations that had not reached stationarity (Figure 4), and increased with the number of tips in the tree (results for 50, 100, and 200 tips can be found in Online Appendix Figures 7-14). Importantly, we found that models which are more complex than the generating model (for example when the full model with three coefficients was fitted to data simulated under a linear potential) still recover the shape of the macroevolutionary landscape with good accuracy (Online Appendix Figures 7-14).

Empirical examples

We demonstrate the utility of BBM+V using two examples of body mass evolution in mammals. We decided to study Carnivora and Primates since both of these orders are large

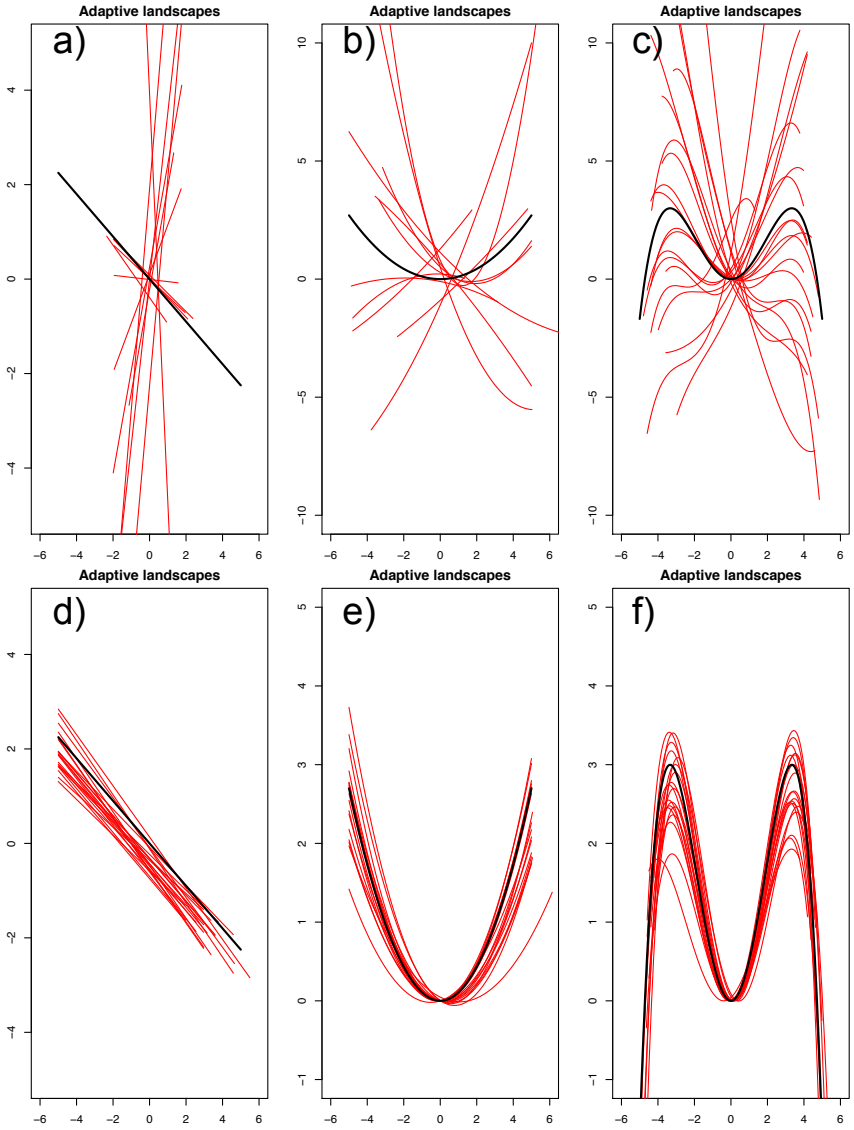

Figure 4: Estimation of the macroevolutionary landscape in different versions of the BBM+V model. Thin red lines in each plot show the macroevolutionary landscapes estimated in 20 simulations, while the simulated macroevolutionary landscape is shown by the thick black line. Only results for trees with 75 tips are shown. a-c) Simulations with $T_c = 2,000$, in which stationarity was not reached. d-f) Simulations with $T_c = 5$, in which stationarity was reached. Generating models are presented as follow: a) and d) show a linear landscape (directional trend); b) and e) show a quadratic landscape (disruptive selection; c) and f) show the full landscape (two peaks). Results for the flat landscape (BBM) are not shown since the macroevolutionary landscape is fixed in this case.

(> 200 spp.), thus ensuring very good statistical power, and because species sampling in these two groups for both phylogenetic and body mass information is almost complete in publicly available databases. In both cases, a time-calibrated phylogeny was obtained by pruning the mammalian phylogeny of Fritz et al. (2009) to the clade of interest. Adult body mass came from the Pantheria database (Jones et al. (2009)) and was log-transformed prior to analysis. In each clade, we fitted six alternative models for body mass evolution using maximum-likelihood: BM, OU1, BBM, BBM+V with $V(x) = cx$, BBM+V with $V(x) = bx^2 + cx$, and BBM+V with $V(x) = ax^4 + bx^2 + cx$. Model fits to the data were compared using AICc. In addition, we used our MCMC algorithm to obtain Bayesian estimates of the shape of the macroevolutionary landscape in both clades (see details in Online Appendix V). Convergence of MCMC chains was assessed both visually by looking at the trace plots of the parameters, likelihood, prior, and posterior, and by measuring the effective sample sizes of these different quantities using the R package coda (Plummer et al. (2006)).

In Carnivora (249 spp. included in our dataset), BBM received the lowest AICc but was closely followed (ΔAICc=0.10) by BBM+V with a linear trend towards larger body mass, a pattern consistent with Cope's rule (Stanley (1973)). More complex versions of BBM+V received higher AICc values (ΔAICc with BBM higher than 0.81), while BM and OU1 had a much lower fit to the data (ΔAICc with BBM of 5.32 and 6.57, respectively). In agreement with these maximum-likelihood results, the posterior distribution of the macroevolutionary landscape for body mass evolution in Carnivora that we estimated using our MCMC algorithm was centered around a flat macroevolutionary landscape (Online Appendix Figure 15).

In Primates (256 spp. included in our dataset), the model that received the lowest AICc was BBM+V with a quadratic term ($V(x) = bx^2 + cx$). Maximum-likelihood estimates of parameters describing the macroevolutionary landscape under this model were $b = -194.9$ and $c = 189.8$, which together resulted in a macroevolutionary landscape typical of disrup-

tive selection, *i.e.* selection towards both small and large body mass values (Figure 5). The second best model was BBM+V with $V(x) = ax^4 + bx^2 + cx$ (ΔAICc=3.38), which gave a macroevolutionary landscape with a very similar shape. Simpler versions of the BBM+V model which cannot accommodate disruptive selection, *i.e.* BBM and BBM+V with a linear trend, received much lower support (ΔAICc with the best model of 26.95 and 28.93, respectively). Finally, BM and OU1 received much lower support than all BBM+V models, with ΔAICc with the best model of 81.02 and 83.07, respectively. The two MCMC chains that we ran also strongly supported a scenario in which body mass in Primates would have evolved towards either small or large values, as indicated by a macroevolutionary landscape with a deep valley for intermediate body masses (Figure 5). Interestingly, the mode of the posterior distribution of the body mass of the most recent common ancestor of Primates (*i.e.*, the root value x_0) was around 1,000 grams, a value that lies in the valley of the macroevolutionary landscape that we estimated (Figure 5).

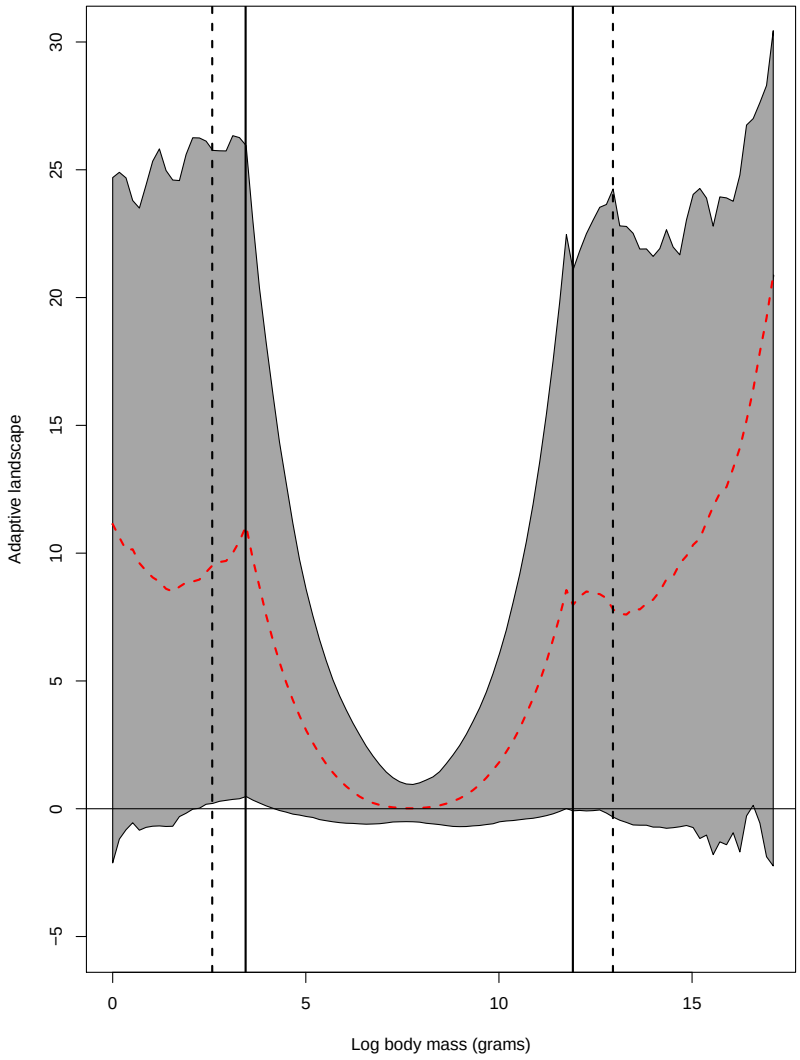

Figure 5: Posterior distribution of the macroevolutionary landscape estimated for body mass evolution in Primates. This posterior distribution was obtained by concatenating the two MCMC chains after the first 20% of samples were discarded as burnin (1.12 million MCMC steps in total). The figure shows the value of the macroevolutionary landscape ($-V(x)$) on the y-axis as a function of logarithm of body mass measured in grams. The dashed red line shows the median value of the macroevolutionary landscape over the posterior, while the grey area ranges from the 5% to the 95% quantiles. The horizontal line shows a flat macroevolutionary landscape, *i.e.*, $V(x) = 0$. The two vertical solid lines show the minimum and maximum values of body mass observed in extent Primates, while the dashed vertical lines show the median estimates of the bounds over the posterior. Quantiles of the posterior distribution of the macroevolutionary landscape become less smooth outside of these values because less and less of the MCMC samples contained a macroevolutionary landscape stretching this far on the trait interval.

DISCUSSION

In this article we have presented equations for a very general model of evolution for continuous traits, as well as its implementation. This opens new possibilities for estimating macroevolutionary landscapes from phylogenetic comparative data. Below we discuss the strengths and weaknesses of this model. We note that Blomberg (2016) has recently introduced a family of essentially similar models for continuous trait evolution, which are in addition freed from the assumption that trait evolution is bounded. However, no framework exists yet to infer parameters of these models from phylogenetic comparative data and the most promising avenue to do this seems to be simulation based methods (Blomberg (2016)). As already explained in the methods, our inclusion of bounds on trait evolution in the BBM+V model is primarily motivated by the fact that this renders the model numerically tractable since trait space can be divided into a finite number of states.

New avenues for studying phenotypic evolution from comparative data

The flexibility of the new model that we propose is its greatest strength. BBM+V can indeed be used to infer macroevolutionary landscapes of any conceivable shape (Figure 2), and thus recover processes such as evolutionary trends, disruptive selection, or diversifying selection leading to macroevolutionary landscapes with several peaks. These last three scenarios lie at the core of modern (macro)evolutionary theory, but could not yet be inferred from phylogenetic comparative data (O'Meara (2012)).

In the BBM+V model, most of the information needed to infer the shape of the macroevolutionary landscape ultimately comes from the distribution of the phenotypic character at the tips of the phylogeny. Indeed, the trait distribution is expected to converge to the stationary distribution of BBM+V, set aside the fact that recently diverged species will have similar trait values. For example, evolutionary trends can be recovered in the absence

of fossil data from a highly skewed trait distribution for contemporaneous species. In the same vein, the simultaneous presence of two peaks in the macroevolutionary landscape can be inferred from a bimodal trait distribution. The characteristic time, T_c, that we have introduced gives the typical time needed to reach stationarity and should always be compared to the total depth of the phylogenetic tree for the clade under study, T_{tot}. Unsurprisingly, model performance will increase with T_{tot}/T_c and in the extreme case where $T_{tot} \ll T_c$, traits will not have explored much of the macroevolutionary landscape. In practice, we have observed than even in cases where $T_{tot}/T_c = 0.1$, all versions of the BBM+V are already discernible from other models (not shown). The characteristic time of the BBM+V process bears much similarity with the phylogenetic half-life of the OU process, which describes the time necessary for the trait value to move halfway from its initial position to the optimum, and thus also gives a measure of the typical time required to reach stationarity in this model. In agreement with our findings, previous studies have shown that accuracy in parameter estimation of the OU model increases with decreasing phylogenetic half-lives (Uyeda and Harmon (2014); Ho and Ané (2014a)).

Using simulations, we have demonstrated that BBM+V can be distinguished from other classic models of trait evolution, and also that distinct shapes of the macroevolutionary landscape can be distinguished from each other based on their likelihoods. This is of course only true when the process has had time to reach stationarity, *i.e.* when the macroevolutionary landscape has been explored by the trait. Indeed, our simulations also show that the danger of over-fitting is quite low with BBM+V since simpler models will often be preferred when stationarity has not been reached. Our focus on AIC to discriminate between alternative models was motivated by the fact that it is the most commonly used in the macroevolutionary community. However, AIC might be prone to overfitting in parameter rich macroevolutionary models and other measures that penalize more for extra parameters, like the Bayesian Information Criterion or its modified version (Zhang and Siegmund (2007)),

might be preferable (Ho and Ané (2014a)). Another solution to diagnose overfitting would be to use parametric bootstrapping techniques (Boettiger et al. (2012)), which is readily implementable for BBM+V since we provide an R function to simulate the model.

Our results also show that estimation accuracy under BBM+V drastically differs between parameters (Online Appendix IV). Estimation is generally accurate when estimating the bounds of the trait interval and the shape of the macroevolutionary landscape. Accuracy in the estimation of these parameters increases with tree size and with T_{tot}/T_c and our results suggest that trees with 75 tips should most often be large enough to obtain reliable estimates (Figure 4). Importantly the three coefficients that determine the shape of the macroevolutionary landscape should not be analyzed separately since numerous different combinations of a, b, and c can give similar shapes. Rather, we recommend to interpret the general shape of the macroevolutionary landscape by focusing on a few important features: (i) whether the macroevolutionary landscape is flat or not, (ii) whether it features a single trend towards one of the bounds of the trait interval, (iii) whether it contains one or several peaks, and if relevant (iv) where are these peaks located in the trait interval.

In contrast, the evolutionary rate σ^2 has high estimation variance, and especially so when T_{tot}/T_c is large. This same effect had already been observed for BBM (Boucher and Démery (2016)) and probably reflects the fact that when the macroevolutionary landscape has been fully explored by the clade its shape and extent are easily estimated but the speed at which the landscape is travelled is not. Given this limitation we don't see much value in interpreting the estimate of σ^2 when fitting BBM+V to an empirical dataset. Rather, the characteristic time of the process, T_c, should be the quantity that is interpreted in comparison with tree depth. Finally, we found that the estimation of the trait value at the root of the tree, x_0, is poor under all situations, a result that generalizes the one already obtained for BBM (Boucher and Démery (2016)). This stems from the fact that BBM+V is a model that retains low phylogenetic signal since it includes strong deterministic forces and wipes out

any hope of confidently inferring ancestral trait values in empirical datasets.

Interpretation of macroevolutionary landscapes inferred from BBM+V

We have introduced BBM+V as a method for the inference of macroevolutionary landscapes from phylogenetic comparative data. The adaptive landscape is a fruitful metaphor to understand phenotypic evolution on a variety of evolutionary scales (Wright (1932); Simpson (1944); Arnold et al. (2001)) but care must be taken when interpreting inferences made from phylogenetic comparative data in the light of adaptive landscape theory, which was mainly developed for population genetics (Uyeda and Harmon (2014)). Some of the models of trait evolution on phylogenies include deterministic forces that influence trait evolution in an attempt to mimic selection: the OU model includes a term that was designed to resemble stabilizing selection towards a given trait value (Hansen (1997)) and BBM+V can imitate a large variety of selection shapes. However, all macroevolutionary models for trait evolution, including OU and BBM+V, are phenomenological by nature since they rely on probabilistic diffusion equations and since they model evolution over long time-scales (typically thousands to million years) that are not amenable to observation. In other words, these models recover patterns from the data, which researchers have to interpret in terms of evolutionary processes. One example of such a confusion between long-term patterns and short-term processes has been highlighted when interpreting the good fit of an OU model to empirical datasets: several studies have indeed shown that neutral evolution between bounds produces patterns closely resembling the ones obtained under an OU process (Revell et al. (2008); Boucher et al. (2014)). Development of the BBM model has now rendered possible to distinguish between these two radically different scenarios (Boucher and Démery (2016)), but many such cases in which two different microevolutionary processes produce the same macroevolutionary pattern remain.

Macroevolutionary landscapes estimated using BBM+V should thus be interpreted with extreme caution: they reflect the general shape of deterministic forces that have been acting on the evolution of a continuous trait in a clade, but are agnostic regarding the nature of these deterministic forces, *i.e.*, they are not actual measurements of the relation between individuals' traits and fitness. The most obvious limitation of BBM+V is that it makes the hypothesis that the macroevolutionary landscape is constant through time. This is likely to be wrong in a majority of cases since environmental change or interactions with other species will lead to changes in the intensity and shape of the selection gradient acting on one trait in a given clade (Simpson (1944); Hansen (2012)). Macroevolutionary landscapes inferred using BBM+V will thus necessarily reflect some kind of average macroevolutionary landscape experienced by the clade throughout its evolutionary history.

Notwithstanding these limitations, which are inherent to all macroevolutionary models, BBM+V offers the opportunity to estimate a variety of macroevolutionary landscapes from phylogenetic comparative data. One kind of possible landscapes deserves particular mention here: macroevolutionary landscapes in which multiple peaks exist. This scenario would be especially interesting to compare to a situation in which these multiple peaks are available for different subclades, which is what is implemented in OU models with multiple optima (Butler and King (2004)). In this latter class of models, each lineage is indeed subject to attraction towards a single peak at a time, but lineages may shift between different peaks. These shifts are either determined *a priori* (Butler and King (2004); Beaulieu et al. (2012)) or inferred directly from phylogenetic patterns of trait evolution (Ingram and Mahler (2013); Khabbazian et al. (2016); Uyeda and Harmon (2014)). In BBM+V with multiple peaks on the contrary, each lineage is always influenced by the different peaks in its macroevolutionary landscape, and transitions between peaks might be frequent if the traits of most species in the clade are located in a valley of the macroevolutionary landscape. More theoretically, these two alternatives would represent two different scenarios: OU models with several op-

tima might be better at describing situations in which a lineage shifts to another adaptive zone (Simpson (1944)), while BBM+V with multiple peaks might represent more genuine diversifying selection towards alternative phenotypic optima.

Even though we have emphasized that it is difficult to connect microevolutionary processes to macroevolutionary patterns, there is one promising way in which BBM+V could be used to do so. Indeed, the Bayesian implementation of the model enables the use of informative priors based on quantitative genetic parameters. Uyeda and Harmon (2014) have demonstrated how this could be done for the OU model: using the quantitative genetic model of Lande (1976), they showed how measurements of heritability, phenotypic variance, and effective population size can inform priors on the parameters of the OU model. By connecting the parameters in Eq. 2 to quantitative genetic models, the same procedure could be carried out for BBM+V.

Finally, BBM+V need not be restricted to infer macroevolutionary landscapes. This model indeed has its roots in spatial diffusion theory and as such could be used in phylogeographic studies to model the dispersal of a set of individuals or populations for which the phylogeny is known (e.g. Grollemund et al. (2015)). This field of research has indeed seen huge methodological advances in recent years (Lemey et al. (2010); Bloomquist et al. (2012)). In this context, BBM+V could be used to infer preferred directions of dispersal (*i.e.*, BBM+V with directional trends) or even to infer particular regions that act as geographic attractors for the taxon under study (*i.e.*, BBM+V with one or several peaks), while explicitly taking into account hard boundaries on the distribution of organisms (e.g. oceans for terrestrial organisms).

Limitations of the BBM+V model

Our implementation of BBM+V does not come without limitations. The one that will be of most practical importance is computational time. BBM+V is indeed a parameter rich

model (up to seven parameters need to be estimated in the most general version with $V(x) = ax^4 + bx^2 + cx$) and in addition we have opted for a safe maximum-likelihood optimization procedure that roughly leads to an 8-fold increase in computing time (see above). In our experience, fitting the BBM+V model to a dataset of 100 species using maximum-likelihood takes about 10 to 20 minutes on a regular computer. This is likely to discourage some researchers since other models of trait evolution have much faster implementations [e.g. those from the phylolm (Ho and Ané (2014b)), phytools (Revell (2012)), or l1ou (Khabbazian et al. (2016)) R packages], but our view is that this amount of time is still orders of magnitudes shorter than that required to infer a phylogeny or collect trait data for a clade of this size.

Another technical limitation is that our implementation of BBM+V is restricted to single traits. We are fully aware that extending it to multivariate datasets would be very convenient, since multiple traits are expected to often evolve in a correlated fashion (Arnold (1992)). However, this is for the moment hampered by computational time. Indeed, the most time-consuming part in the calculation of the likelihood is to invert the instantaneous transition matrix, M. If we were to study two traits simultaneously, we would need to discretize the plane that they define into a regular grid of points, and computing time would not be multiplied by two but rather raised to the power two. The only possible solution that we can envision would be to use algorithmic tricks that avoid inverting the entire transition matrix, but rather a matrix describing transitions between a given point on the grid and its immediate neighbors, as recently proposed for inference of ancestral areas (Landis et al. (2013)). This would require much development and is out of the scope of this article.

Finally, there are limitations inherent to the formulation of the model itself. BBM+V is indeed based on a constant-rate diffusion model and as such cannot model accelerating or decelerating trait evolution (Harmon et al. (2010)) or sudden jumps in the value of the trait, as would be expected under quantum evolution (Simpson (1944); Kirkpatrick (1982)) or punctuated equilibrium (Gould and Eldredge (1977)). In addition, the parametric form of

the macroevolutionary landscape that we have chosen currently only allows for a maximum of two peaks, although this could be readily modified. Enabling more peaks to exist would require more complex forms of the macroevolutionary landscape, which would make the model even more difficult to optimize.

Conclusion

Our development and implementation of BBM+V greatly expands the set of models available for studying the evolution of continuous characters on phylogenies and enables the estimation of macroevolutionary landscapes of various shapes. We have shown that the model generally achieves good performance both in terms of parameters estimation and in terms of discrimination from alternative macroevolutionary models. R code for fitting BBM+V to empirical data is freely available from `https://github.com/fcboucher/BBMV`, and this repository also contains a detailed tutorial to the different functions for simulating and inferring the model. One promising avenue for further developments of the BBM+V model lies in the simplicity with which the model can be simulated. The discretized equation for the evolution of trait values (6) indeed enables fast simulations of the model, since they only require to diagonalize the diffusion matrix once. These simulations need not be restricted to the case we have described in this article, *i.e.* only one trait and a specific parametric form of the potential, and could be run for any macroevolutionary landscape that one can envision. Such detailed simulations could then be used in an Approximate Bayesian Computation framework (Slater et al. (2012); Lintusaari et al. (2016)) to infer the shape of the macroevolutionary landscape in which a clade has evolved. We thus hope that the development of the BBM+V model will prove useful in increasing knowledge of phenotypic evolution over macroevolutionary timescales.

SUPPLEMENTARY MATERIAL

Online appendices can be found in the Dryad repository associated with this submission.

ACKNOWLEDGEMENTS

This work was supported by a Plant Fellows and a Claraz-Schenkung grant to FB. This work was also supported by a National Science Foundation grant awarded to L.J.H. (DEB 1208912). We made heavy use of the Science Cloud computer cluster of the University of Zurich.

*

References

Arnold, S. J. 1992. Constraints on phenotypic evolution. The American Naturalist 140:S85–S107.

Arnold, S. J., M. E. Pfrender, and A. G. Jones. 2001. The adaptive landscape as a conceptual bridge between micro- and macroevolution. Genetica 112:9–32.

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

Blomberg, S. P. 2016. Beyond brownian motion and the ornstein-uhlenbeck process: Stochastic diffusion models for the evolution of quantitative characters. bioRxiv .

Bloomquist, E. W., P. Lemey, and M. A. Suchard. 2012. Three roads diverged? routes to phylogeographic inference. Trends in Ecology & Evolution 25:626–632.

754 Boettiger, C., G. Coop, and P. Ralph. 2012. Is your phylogeny informative? measuring the
755 power of comparative methods. Evolution 66:2240–2251.

756 Boucher, F. C. and V. Démery. 2016. Inferring bounded evolution in phenotypic characters
757 from phylogenetic comparative data. Systematic Biology 65:651–661.

758 Boucher, F. C., W. Thuiller, T. J. Davies, and S. Lavergne. 2014. Neutral biogeography and
759 the evolution of climatic niches. The American naturalist 183:573–84.

760 Butler, M. A. and A. A. King. 2004. Phylogenetic Comparative Analysis : A Modeling
761 Approach for Adaptive Evolution. Evolution 164:683–695.

762 Doebeli, M. 1996. An explicit genetic model for ecological character displacement. Ecology
763 77:510–520.

764 Drury, J., J. Clavel, M. Manceau, and H. Morlon. 2016. Estimating the effect of competition
765 on trait evolution using maximum likelihood inference. Systematic Biology 65:700–710.

766 Eastman, J. M., D. Wegmann, C. Leuenberger, and L. J. Harmon. 2013. Simpsonian 'Evo-
767 lution by Jumps' in an Adaptive Radiation of Anolis Lizards. ArXiv e-prints .

768 Edwards, A. W. F. and L. L. Cavalli-Sforza. 1964. Reconstruction of evolutionary trees.
769 Pages 67–76 *in* Phenetic and phylogenetic classification (V. H. Heywood and J. McNeill,
770 eds.). Systematics Association, London.

771 Estes, S. and S. J. Arnold. 2007. Resolving the paradox of stasis: Models with stabiliz-
772 ing selection explain evolutionary divergence on all timescales. The American Naturalist
773 169:227–244 pMID: 17211806.

774 Felsenstein, J. 1973. Maximum-Likelihood Estimation of Evolutionary Trees from Continuous
775 Characters. American Journal of Human Genetics 25:471–492.

776 Freckleton, R. P. 2012. Fast likelihood calculations for comparative analyses. Methods in
777 Ecology and Evolution 3:940–947.

778 Fritz, S. a., O. R. P. Bininda-Emonds, and A. Purvis. 2009. Geographical variation in pre-
779 dictors of mammalian extinction risk: Big is bad, but only in the tropics. Ecology Letters
780 12:538–549.

781 Gardiner, C. W. 1985. Handbook of stochastic methods (1st ed.). Berlin: Springer.

782 Gavrilets, S. 1997. Evolution and speciation on holey adaptive landscapes. Trends in Ecology
783 & Evolution 12:307–312.

784 Gould, S. J. and N. Eldredge. 1977. Punctuated equilibria: the tempo and mode of evolution
785 reconsidered. Paleobiology 3:115–151.

786 Grafen, A. 1989. The Phylogenetic Regression. Philosophical transactions of the Royal So-
787 ciety of London. Series B, Biological sciences 326:119–157.

788 Grollemund, R., S. Branford, K. Bostoen, A. Meade, C. Venditti, and M. Pagel. 2015. Bantu
789 expansion shows that habitat alters the route and pace of human dispersals. Proceedings
790 of the National Academy of Sciences 112:13296–13301.

791 Hansen, T. 2012. Adaptive landscapes and macroevolutionary dynamics. Pages 205–221 *in*
792 The adaptive landscape in evolutionary biology (E. Svensson and R. Calsbeek, eds.).
793 Oxford University Press, Oxford.

794 Hansen, T. F. 1997. Stabilizing selection and the comparative analysis of adaptation. Evo-
795 lution 51:1341–1351.

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

Harmon, L. J., J. B. Losos, T. Jonathan Davies, R. G. Gillespie, J. L. Gittleman, W. Bryan Jennings, K. H. Kozak, M. a. McPeek, F. Moreno-Roark, T. J. Near, A. Purvis, R. E. Ricklefs, D. Schluter, J. a. Schulte Ii, O. Seehausen, B. L. Sidlauskas, O. Torres-Carvajal, J. T. Weir, and A. O. Mooers. 2010. Early bursts of body size and shape evolution are rare in comparative data. Evolution 64:2385–96.

Ho, L. S. T. and C. Ané. 2014a. Intrinsic inference difficulties for trait evolution with ornstein-uhlenbeck models. Methods in Ecology and Evolution 5:1133–1146.

Ho, L. s. T. and C. Ané. 2014b. A linear-time algorithm for gaussian and non-gaussian trait evolution models. Systematic Biology 63:397–408.

Hunt, G. 2007. The relative importance of directional change, random walks, and stasis in the evolution of fossil lineages. Proceedings of the National Academy of Sciences 104:18404–18408.

Ingram, T. and D. L. Mahler. 2013. SURFACE : detecting convergent evolution from comparative data by fitting Ornstein-Uhlenbeck models with stepwise Akaike Information Criterion. Methods in Ecology and Evolution 4:416–425.

Jackson, J. D. 1998. Classical Electrodynamics, 3rd Edition. John Wiley & Sons.

Jones, K. E., J. Bielby, M. Cardillo, S. A. Fritz, J. O'Dell, C. D. L. Orme, K. Safi, W. Sechrest, E. H. Boakes, C. Carbone, C. Connolly, M. J. Cutts, J. K. Foster, R. Grenyer, M. Habib, C. A. Plaster, S. A. Price, E. A. Rigby, J. Rist, A. Teacher, O. R. P. Bininda-Emonds, J. L. Gittleman, G. M. Mace, and A. Purvis. 2009. Pantheria: a species-level database of life history, ecology, and geography of extant and recently extinct mammals. Ecology 90:2648–2648.

820 Khabbazian, M., R. Kriebel, K. Rohe, and C. Ané. 2016. Fast and accurate detection of evo-
821 lutionary shifts in ornstein–uhlenbeck models. Methods in Ecology and Evolution 7:811–
822 824.

823 Kirkpatrick, M. 1982. Quantum evolution and punctuated equilibria in continuous genetic
824 characters. The American Naturalist 119:833–848.

825 Lande, R. 1976. Natural selection and random genetic drift in phenotypic evolution. Evolu-
826 tion 30:314–334.

827 Landis, M. J., N. J. Matzke, B. R. Moore, and J. P. Huelsenbeck. 2013. Bayesian analysis of
828 biogeography when the number of areas is large. Systematic Biology 62:789–804.

829 Lemey, P., A. Rambaut, J. J. Welch, and M. A. Suchard. 2010. Phylogeography takes a
830 relaxed random walk in continuous space and time. Molecular Biology and Evolution
831 27:1877–1885.

832 Lewis, P. O. 2001. A Likelihood Approach to Estimating Phylogeny from Discrete Morpho-
833 logical Character Data. Systematic biology 50:913–925.

834 Lintusaari, J., M. U. Gutmann, R. Dutta, S. Kaski, and J. Corander. 2016. Fundamentals
835 and recent developments in approximate bayesian computation. Systematic Biology .

836 Nosil, P. 2012. Ecological speciation. Oxford Series in Ecology and Evolution.

837 Nuismer, S. L. and L. J. Harmon. 2015. Predicting rates of interspecific interaction from
838 phylogenetic trees. Ecology Letters 18:17–27.

839 O'Meara, B. C. 2012. Evolutionary Inferences from Phylogenies : A Review of Methods.
840 Annual Review of Ecology, Evolution, and Systematics 43:267–285.

841 Paradis, E., J. Claude, and K. Strimmer. 2004. Ape: Analyses of phylogenetics and evolution
842 in r language. Bioinformatics 20:289–290.

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

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

848 R Core Team. 2016. R: A Language and Environment for Statistical Computing. R Founda-
849 tion for Statistical Computing Vienna, Austria.

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

852 Revell, L. J., L. J. Harmon, and D. C. Collar. 2008. Phylogenetic signal, evolutionary process,
853 and rate. Systematic biology 57:591–601.

854 Risken, H. 1984. The Fokker-Planck Equation vol. 18 of *Springer Series in Synergetics*.
855 Springer Berlin Heidelberg.

856 Schluter, D. 2000. The Ecology of Adaptive Radiation. Oxford Series in Ecology and Evolu-
857 tion.

858 Simpson, G. G. 1944. Tempo and Mode in Evolution. Columbia University Press, New York.

859 Slater, G. J., L. J. Harmon, D. Wegmann, P. Joyce, L. J. Revell, and M. E. Alfaro. 2012.
860 Fitting models of continuous trait evolution to incompletely sampled comparative data
861 using approximate bayesian computation. Evolution 66:752–762.

Soulebeau, A., X. Aubriot, M. Gaudeul, G. Rouhan, S. Hennequin, T. Haevermans, J.-Y. Dubuisson, and F. Jabbour. 2015. The hypothesis of adaptive radiation in evolutionary biology: hard facts about a hazy concept. Organisms Diversity & Evolution 15:747–761.

Stanley, S. M. 1973. An explanation for cope's rule. Evolution 27:1–26.

Uyeda, J. C., T. F. Hansen, S. J. Arnold, and J. Pienaar. 2011. The million-year wait for macroevolutionary bursts. Proceedings of the National Academy of Sciences 108:15908–15913.

Uyeda, J. C. and L. J. Harmon. 2014. A novel Bayesian method for inferring and interpreting the dynamics of adaptive landscapes from phylogenetic comparative data. Systematic biology 63:902–918.

Wright, S. 1932. The roles of mutation, inbreeding, crossbreeding and selection in evolution. Proceedings of the Sixth International Congress of Genetics 1:356–366.

Zhang, N. R. and D. O. Siegmund. 2007. A modified bayes information criterion with applications to the analysis of comparative genomic hybridization data. Biometrics 63:22–32.