Beyond Brownian Motion and the Ornstein-Uhlenbeck Process: Stochastic Diffusion Models for the Evolution of Quantitative Characters

Simone P. Blomberg,[1,*] Suren I. Rathnayake,[1,2] and Cheyenne M. Moreau[1]

1. School of Biological Sciences, University of Queensland, Saint Lucia 4072, Australia;   2. School of Mathematics and Physics, University of Queensland, Saint Lucia 4072, Australia

ABSTRACT: Gaussian processes, such as Brownian motion and the Ornstein-Uhlenbeck process, have been popular models for the evolution of quantitative traits and are widely used in phylogenetic comparative methods. However, they have drawbacks that limit their utility. Here we describe new, non-Gaussian stochastic differential equation (diffusion) models of quantitative trait evolution. We present general methods for deriving new diffusion models and develop new software for fitting non-Gaussian evolutionary models to trait data. The theory of stochastic processes provides a mathematical framework for understanding the properties of current and future phylogenetic comparative methods. Attention to the mathematical details of models of trait evolution and diversification may help avoid some pitfalls when using stochastic processes to model macroevolution.

Keywords: stochastic differential equations, diffusions, continuous traits, comparative methods.

> Brownian motion is a poor model, and so is Ornstein-Uhlenbeck, but just as democracy is the worst method of organizing a society "except for all the others", so these two models are all we've really got that is tractable. Critics will be admitted to the event, but only if they carry with them another tractable model. (J. Felsenstein, r-sig-phylo email list, April 8, 2008)

Introduction

The parametric estimation of phylogenies depends on an appropriate model of character evolution (Posada and Cran-

* Corresponding author; email: s.blomberg1@uq.edu.au.
  ORCIDs: Blomberg, https://orcid.org/0000-0003-1062-0839.

dal 2001). Molecular systematists are spoiled for choice in this regard. For example, the program jModelTest2 can fit 1,624 models of DNA sequence evolution (Darriba et al. 2012). The situation for the comparative analysis of continuous traits is quite different. Here we have mainly two analytical models in popular use: Brownian motion (BM) and the Ornstein-Uhlenbeck (OU) process. Other related models, such as "early burst," are also used (e.g., Blomberg et al. 2003; Ingram et al. 2012). See Uyeda et al. (2015) for the relationship between early-burst models and the OU model. Boucher and Démery (2016) developed a BM model with hard bounds. There are several extensions to the OU model (see below). Other approaches to phylogenetic comparative analysis do not use explicit models of evolution (in terms of being able to write down the appropriate equations). Some nonanalytical models can be used to estimate sampling distributions for regression parameters using computer simulation (Garland et al. 1993). The evolutionary model for continuous traits can also be altered by applying branch-length transformations (e.g., Grafen 1989; Pagel 1999; Freckleton et al. 2002; Blomberg et al. 2003; but see Slater 2014). Lynch (1991) introduced an approach to phylogenetic comparative analysis based on quantitative genetics (see also Hadfield 2010). We do not consider these approaches to phylogenetic comparative analysis here. Instead, we focus on providing an approach to comparative analysis based on the theory of stochastic processes, which unites BM, OU, and other processes in a common statistical and probabilistic framework.

Beginning with the work of Bachelier (1900), finance has been a prominent application of stochastic processes, where models have been developed for stock prices, derivatives, options, and other products. In that domain, the model of Black and Scholes (1973) has been particularly successful (in terms of citations, if not profits); however, research into the theory of stochastic processes is

still thriving across a wide range of disciplines, especially the physical sciences (e.g., Uhlenbeck and Ornstein 1930; Einstein 1956; Freund and Pöschel 2000; Gardiner 2009; Du Toit Strauss and Effenberger 2017). Although diffusion models are common in epidemiology and other life sciences (Fuchs 2013), applications in evolutionary biology are rare. The Wright-Fisher model and the Moran model (with various extensions) in population genetics are well-known exceptions (Fisher 1922; Wright 1931; Feller 1951; Moran 1958; Ewens 2004; Hofrichter et al. 2017) See also De Maio et al. (2015) for a recent example that uses allele frequency estimation to estimate species trees. Population geneticists have used these stochastic processes to model the microevolution of gene frequencies. Here we examine the possible uses of stochastic processes in studies of trait macroevolution, that is, evolution above the species level (Simpson 1953; Rensch 1959; Stanley 1975; Benton 2015; Serrelli and Gontier 2015), with the aim to provide new models and methods for the phylogenetic comparative analysis of non-Gaussian traits. Such models are necessary not only because current evolutionary models for quantitative traits are somewhat limited (Pennell et al. 2015) but because a greater abundance of potential models may make the analysis of cross-species data more flexible and make our understanding of macroevolution more nuanced.

As a way forward, we do not incorporate genetics into any of the following macroevolutionary models; instead, we assume the "phenotypic gambit" (Grafen 1984). That is, we assume that the genetic control of the phenotype depends on many alleles of small effect. We assume that there is ample genetic variation and that genetic correlations with other traits are unimportant. This is a necessary assumption, as we rarely have information on the genetic architecture from the fossil record (but see Schraiber et al. 2016). Some evidence suggests that genetic constraints do not play a strong role in slowing the rate of adaptation (Agrawal and Stinchcombe 2009). However, Schluter (1996) found evidence for the role of genetic constraints over timespans of the order of ~4 million years. Analysis of a large data set of body sizes, combining paleontological, phylogenetic comparative data sets of extant taxa, and historical studies across many taxa have concluded that on relatively short timescales (<1 million years), evolution appears to be bounded; over longer time spans, evolution is strongly divergent (Estes and Arnold 2007; Uyeda et al. 2011). Arnold (1992) synthesized our knowledge of evolutionary constraints. At that time, there was little evidence for the persistence of constraints over deep time, and subsequent reviews, concentrating on the quantitative genetic approach to constraints via the **G** matrix, failed to shed light on this issue (Blows and Hoffmann 2005; Pigliucci 2007; Futuyma 2010). While a role for genetic constraints on macroevolutionary timescales cannot be ruled out, there is little supporting evidence, and what evidence we have is contradictory.

If modern macroevolutionary models are found to be at odds with the established facts of genetics and microevolution, the macroevolutionary models should be treated with caution. However, even models that are "wrong" in that they entail unrealistic biological assumptions can still be useful. For example, diffusion models (such as OU and the models outlined below) are well approximated by BM over short timescales, even though BM is an unlikely model of evolution for most organisms over their entire evolutionary history.

After a brief introduction to the mathematics of BM and OU, we introduce two new models that have some similarities to OU yet have different distributional properties ("Diffusions as Models of Trait Evolution"). We show how the distributional properties of the models arise from solutions to the Fokker-Planck (Kolmogorov forward) equation ("New Evolutionary Models"). We then discuss issues related to the implementation of the models ("Fitting Models to Data") and describe new software to fit them to trait data on phylogenies ("Implementation"). We provide evidence from simulations that our implementation works in that it can accurately recover parameter estimates from data simulated using the models ("Simulation Study"). We provide a case study ("Case Study" sections) that shows how the new methods work in practice. We conclude with a discussion of stationary and transition distributions associated with the new models, along with possibilities for further extensions to the models and directions for future research.

Mathematical Background

To fully understand the mathematics of stochastic processes, some background is required. At least some knowledge of Riemann-Stieltjes integrals, as well as an appreciation of measure-theoretic probability theory, is necessary. Introductory books, such as Øksendal (2007) or Klebaner (2012), may be helpful. Gardiner (2009) provides an excellent practical approach that largely ignores the measure-theoretic foundations but concentrates mainly on applications in the physical sciences. For many purposes, one can ignore the measure-theoretic probability foundations of stochastic processes. However, we urge a basic grasp of the concepts of probability in order to understand where the correspondences lie between probability theory and evolutionary theory, where evolutionary interpretations of concepts in probability theory are justifiable, and, importantly, where the structure of probability theory as a basis for macroevolutionary theory may be insufficient.

Brownian Motion

BM is named for the movement of pollen grains suspended in water, as first observed by the botanist Robert Brown in 1837 and is observed in many other multiparticle settings. The mathematics of BM were first analyzed by Bachelier (1900), who anticipated almost all of the mathematical results of Einstein's work in 1905 in the context of molecular movement (see Einstein 1956). Wiener (1923) was the first to rigorously characterize BM as a stochastic process, and hence BM is sometimes also known as the Wiener process. BM was introduced as a model of gene frequency evolution for phylogeny estimation by Edwards and Cavalli-Sforza (1964) and as a model of quantitative character evolution for phylogeny estimation by Felsenstein (1973), who also introduced this model into phylogenetic comparative regression analyses (Felsenstein 1985).

Let $B(t)$ be the trait value of a BM process at time t. BM has the following defining properties (e.g., Klebaner 2012). BM has independent increments, and $B(t) - B(s)$ for $t > s$ is independent of the past $B(u)$, where $0 \leq u \leq s$. The increments are also Gaussian, and $B(t) - B(s)$ has a standard normal (Gaussian) distribution with a mean $\mu = 0$ and variance equal to $t - s$. This means we can use all of the powerful mathematical machinery appropriate to Gaussian distributions. Furthermore, the variance of $B(t)$ increases linearly with t. There does not seem to be any biological reason why the variance of a trait should increase linearly with time. Furthermore, this property implies that there are no bounds to evolution and that traits have no physical limits. This is unlikely to be true for any trait (e.g., McGhee 2015; but see Conway Morris et al. 2015; Vermeij 2015, ibid.). More complicated models, such as that of Boucher and Démery (2016), impose limits on the variance of sample paths.

BM is useful as a simple model of trait evolution. Its simplicity comes from the above-described properties, as well as from the fact that it has the Markov property. Also, BM is a martingale, which means that the expectation of the process at time $t + s$ is the value of the process at time t. That is, $E(B(t + s)|\mathcal{F}_t) = B(t)$. The term $\mathcal{F}_t$ represents the entire evolutionary history of the process (i.e., the trait values) up to and including time t. The martingale property can be used to make predictions for BM. If it is known that BM has a value $B(t)$ at time t, then the expected value of the process conditional on knowing $B(t)$ is also $B(t)$. The Markov and martingale properties simplify the mathematics of working with BM processes. BM lends itself to at least two evolutionary interpretations. BM can be viewed as implying no selection, and evolution occurs just by random drift. It can also be viewed as a model of very strong selection in a randomly varying environment (see Hansen and Martins 1996). Both of these

interpretations cannot be simultaneously correct, and both are likely to be wrong for any real quantitative trait. In addition, other evolutionary processes may also be well approximated by BM, as BM is used to model the pattern of evolution, not the underlying process.

The use of simple BM models for the analysis of comparative data from extant species begins with Felsenstein's (1985) phylogenetically independent contrasts method (Huey et al. 2019). Many stochastic processes use BM as a building block, and BM is included in most quantitative evolutionary models in some way. Felsenstein (2012) has developed a comparative method that allows for the BM evolution of latent traits.

The Ornstein-Uhlenbeck Process

The OU process was introduced as an improved model for physical BM, which incorporates the effect of friction (Uhlenbeck and Ornstein 1930). It also has a long history in evolutionary biology. OU can be derived from the consideration of stabilizing selection and genetic drift (Lande 1976). Its use in phylogenetic comparative methods has been promoted by many authors (Felsenstein 1988; Hansen and Martins 1996; Hansen 1997; Martins and Hansen 1997; Butler and King 2004; Hansen et al. 2008; Beaulieu et al. 2012). OU has the following form:

$$X(t) = \mu + e^{-\alpha t}\left(X(0) - \mu + \sigma \int_0^t e^{\alpha s}\, dB(s))\right), \quad (1)$$

where μ is the attracting constant and the mean of the process at stationarity. Note that $X(t)$ in equation (1) depends on $B(s)$. That is, BM is one building block of the OU process. The biological interpretation of the OU process is controversial. Most authors have interpreted α as the strength of a restraining force corresponding to stabilizing selection and the sample paths as trajectories of evolution of organisms' traits (e.g., Butler and King 2004; Beaulieu et al. 2012). However, an alternative interpretation is that the sample paths are of an evolutionary optimum subject to an overall central tendency with strength α and stochastic perturbations (Hansen 1997; Hansen et al. 2008).

The properties of OU are well known (e.g., Klebaner 2012). The OU process is a Gaussian process with continuous paths. It has the Markov property and is stationary, provided the initial distribution is the stationary distribution $N(\mu, \sigma^2/2\alpha)$. It is the only stochastic process that has all three properties (Gaussian, Markov, and stationarity; Breiman 1968; Klebaner 2012). OU is not a martingale, which can be seen intuitively, as the expectation of the process for $t \to \infty$ is μ and not X_t. More rigorously and for small t, $E[X_t|\mathcal{F}_t] = X_0 e^{-\alpha t} + \mu(1 - e^{-\alpha t}) \neq X_t$, where X_0 is the starting value of the process at $t = 0$ and $\mathcal{F}_t$ is

the evolutionary history of the process up to time t. The Gaussian property of both BM and OU makes them relatively simple to work with.

There are many more complex models that use OU as a starting point. For example, Bartoszek et al. (2012) have developed a method to analyze multivariate characters using OU. Ingram and Mahler (2013) use OU to detect evolutionary convergence in comparative data. Uyeda and Harmon (2014) use OU to analyze adaptive landscapes. Khabbazian et al. (2016) use OU to detect shifts in evolutionary optima on phylogenies. Hansen (1997) and Butler and King (2004) use likelihood methods to fit models with different μ values on different branches of the phylogeny. This allows for tests of a priori hypotheses of different diffusion coefficients on different branches of the phylogeny. Beaulieu et al. (2012) extend this idea by allowing σ and α to vary with time. No doubt further applications will be forthcoming.

Note that the stochastic integral in equation (1) is with respect to "white noise," implying that $B(t)$ is differentiable, whereas one of the properties of BM is that it is *not* differentiable. The meaning of such integrals is therefore not straightforward. In fact, it requires a new definition for integration. The definition adopted here is that of Itô (1944, 1946). There are other approaches to stochastic integration, most notably the Stratonovich integral (e.g., Gardiner 2009). Turelli (1977) discusses situations in which one definition may be preferred over the other. In practice, the Itô integral is the most widely used. Note that the Itô and Stratonovich stochastic differential equations for both BM and OU are identical (applying the results of Gardiner 2009, p. 98).

Diffusions as Models of Trait Evolution

Consider the stochastic differential equation (SDE)

$$dX_t = b(X_t, t)\, dt + \sigma(X_t, t)\, dB_t, \tag{2}$$

where X_t is the value of our trait at time t. Such SDEs are termed "diffusion" equations and arise as solutions to the Fokker-Planck equation (Gardiner 2009). The left-hand side of the equation represents a small change in trait variable X_t at time t. The right-hand side has two terms. The first term on the right-hand side is the deterministic part of the model, $b(X_t, t)$, termed the "drift" function. The differential of the first term is dt, which denotes a differential with respect to continuous time. Note the difference in usage of the term compared with its use in population genetics, where drift implies a stochastic process. We will retain the traditional mathematical terminology. The second term is stochastic, as the differential is $dB(t)$, commonly known as "white noise." The function

$\sigma(X_t, t)$ is termed the "diffusion" function. In financial statistics, $\sigma(X_t, t)$ is termed the "volatility" (Mikosch 1998). Note that both b and σ can depend on both X_t and t in some arbitrary way. Importantly, the only meaning of equation (2) is with respect to the Itô definition of the integral. Stochastic processes of this type are termed "Itô diffusions."

The OU diffusion process can be defined by the following SDE:

$$dX_t = \alpha(\mu - X_t)\, dt + \sigma\, dB_t, \tag{3}$$

where α and σ are real, positive constants. Here, α represents the restraining force of stabilizing selection (in units of change of fitness per unit change of the trait value), σ is the standard deviation (in trait units) of the BM process, and μ represents the mean trait value (at stationarity). The drift coefficient here is a linear function of X_t. The form of the drift is significant, as this expression controls the forcing of the trait X_t back toward μ. OU is thus said to be "mean reverting": X_t tends to return to μ over time. However, the property of mean reversion is not limited to the OU process. Eliazar and Cohen (2012) discuss the conditions where mean reversion can occur. OU is mean reverting largely because the Gaussian stationary distribution is unimodal and symmetric. Other processes with different diffusion functions but with the same OU drift may be mean reverting, but mode reversion is more common, particularly in cases where the stationary distribution is asymmetric. There are cases that can be constructed where reversion is neither to the mean nor to the mode. The drift function can be different from the OU drift and still be reverting to a constant, whether that corresponds to the mean or mode of the stationary distribution or another value. Hence, it is possible to reverse-engineer SDEs for any stationary distribution by judicious choice of either the drift or diffusion functions (Cai and Lin 1996; Eliazar and Cohen 2012).

It is also clear that equation (3) is time homogeneous, since neither b nor σ depend on t. The process is also ergodic. That is, given enough time, the time average for any particular species' trait is equal to the average trait value across species (Lebowitz and Penrose 1973). These properties suggest that a stationary distribution exists for this process. Figure 1 shows a sample evolution along a five-species tree for the OU model.

New Evolutionary Models

The main drawback of both BM and OU that we wish to highlight is the Gaussian nature of both stochastic processes. While analytically and computationally useful, this assumption limits the application of the models to traits that are normally distributed, such as body size or length
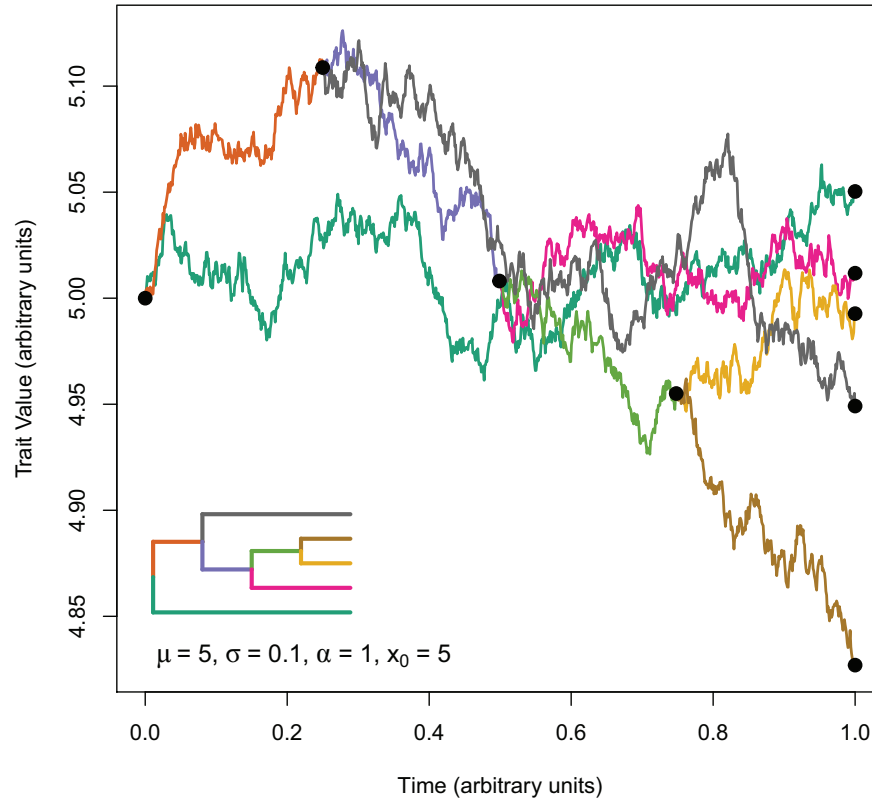
Figure 1: Ornstein-Uhlenbeck evolution along a five-species tree. μ is the mean of the process, α is the strength of the restraining force, and σ is the diffusion coefficient. x_0 is the starting value for the process. Large circles are nodes and tips.

measurements. Other traits, such as sex ratios or life spans, do not follow normal distributions, and new methods that take into account the form of the distribution of traits are desirable. Of course, one could transform the trait variable so that it is then approximately Gaussian, such as using the $\text{logit}(x)$, $\text{probit}(x)$, or $\sin^{-1} x^{1/2}$ transformations for proportions or the $\log(x)$ transformation for counts, and then use Gaussian process models (Ives 2015; Warton et al. 2016). However, shoehorning data using transformations can make interpretation of model outputs more difficult. Instead, we suggest that the direct modeling of non-Gaussian evolutionary processes provides a much more elegant view of the evolutionary process. There is a strong analogy with the development of generalized linear models, which greatly extends the analysis of non-Gaussian linear models (McCullagh and Nelder 1989). Here we outline a generalized method of constructing new stochastic process models for continuous trait evolution.

The use of non-Gaussian models in finance has been discussed by Barndorff-Nielsen and Shephard (2001). The key to the construction of new models for evolution is the solution of the Fokker-Planck (Kolmogorov forward)

equation (Risken 1996). In one dimension, it takes the form

$$
\begin{aligned}
\frac{\partial f(x,t)}{\partial t} &= -\frac{\partial}{\partial x}[b(x,t)f(x,t)] \\
&+ \frac{1}{2}\frac{\partial^2}{\partial x^2}[\sigma^2(x,t)f(x,t)].
\end{aligned}
\tag{4}
$$

Equation (4) governs the time evolution of the underlying probability law $f(x,t)$. That is, $f(x,t)$ is the density of the process at trait value x at time t. It is a partial differential equation in x and t. Note that it is *not* stochastic. If the stochastic process is time homogeneous, that is, when $df/dt = 0$, equation (4) can be written as

$$
\frac{d}{dx}[b(x)f(x)] - \frac{1}{2}\frac{d^2}{dx^2}[\sigma^2(x)f(x)] = 0.
\tag{5}
$$

Solving for $f(x)$ gives the following formula for the construction of the stationary distribution (app. A; apps. A, B are available online):

$$
f(x) = \frac{C}{\sigma^2(x)}\exp\left(\int_{x_0}^{x}\frac{2b(y)}{\sigma^2(y)}dy\right),
\tag{6}
$$

where C is a constant of integration found by solving $\int f(x)\,dx = 1$ and x_0 is the trait value at time $t = 0$. Equation (6) is sometimes known as Wright's equation (Wright 1938; Cobb 1998). Using OU as an example, equation (6) can be evaluated by plugging in the drift coefficient $b(y) = \alpha(\mu - y)$ and letting σ^2 be the square of the (constant) OU diffusion coefficient. It can easily be shown then that $f(x)$ is a Gaussian distribution with mean μ and variance $\sigma^2/2\alpha$.

Consider the following diffusion equations:

$$dX_t = \alpha(\mu - X_t)\,dt + \sqrt{\epsilon X_t}\,dB_t, \qquad (7)$$

$$dX_t = \alpha(\mu - X_t)\,dt + \sqrt{\epsilon X_t(1 - X_t)}\,dB_t, \qquad (8)$$

where $X_t > 0$ for equation (7) and $X_t \in (0, 1)$ for equation (8). The drift terms in equations (7) and (8) are of the same form as in equation (3). Hence, these processes are both reverting and will be driven by a central tendency, with a restraining force α. The difference between these two processes and OU is in the diffusion function. With a reverting process with OU drift, the form of the diffusion function determines the stationary distribution. The stationary distributions for each process described

by equations (7) and (8) are derived in appendix B. While the notation for calculating with diffusion models is powerful and elegant, stochastic processes come alive when visualized using simulation. Example plots of paths mapped onto a phylogeny with five species are shown in figures 2 and 3. Equation (7) has as its stationary distribution

$$f(x|\mu, \delta) = \left(\frac{x}{\delta}\right)^{-1+\mu/\delta} \frac{e^{-x/\delta}}{\Gamma(\mu/\delta)}, \quad \delta = \frac{\epsilon}{\alpha},$$

where Γ is the gamma function. That is, $f(x|\mu, \delta)$ is a density of a gamma distribution with mean $= \mu$, mode $= \mu - \delta$, and variance $= \delta\mu$: $x \sim \text{Gamma}(\mu/\delta, 1/\delta)$. In fact, equation (7) is the Cox-Ingersoll-Ross (CIR) model commonly used in finance (Cox et al. 1985). See figure 2.

The stationary distribution of the process described by equation (8) is

$$f(x) = \frac{1}{\text{B}(\mu/\delta, (1 - \mu)/\delta)}$$
$$\times x^{\mu/\delta - 1}(1 - x)^{(1-\mu)/\delta - 1}, \quad \delta = \frac{\epsilon}{\alpha},$$

where B is the beta function. That is, $f(x|\mu, \delta)$ is the density of a beta distribution with $x \sim \text{Beta}(\mu/\delta, (1 - \mu)/\delta)$ (fig. 2).
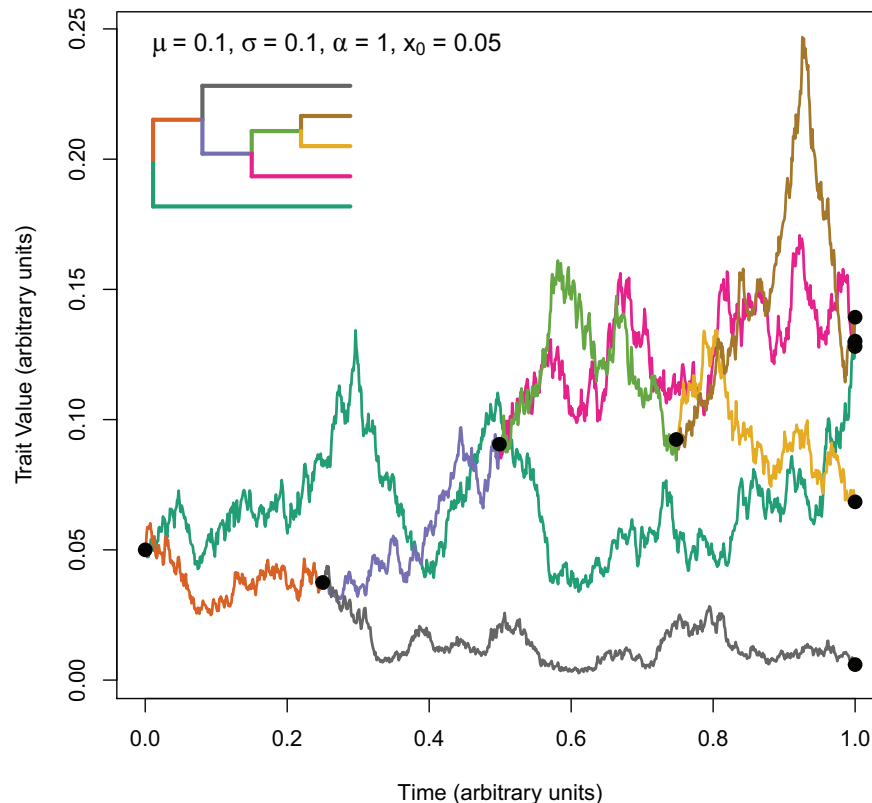

Figure 2: Cox-Ingersoll-Ross evolution along a five-species tree. μ is the reverting level of the process, α is the strength of the restraining force, and σ is the scaling constant for the diffusion coefficient. x_0 is the starting value for the process. Large circles are nodes and tips.

Figure 3: Beta evolution along a five-species tree. μ is the reverting level of the process, α is the strength of the restraining force, and σ is the scaling constant for the diffusion coefficient. x_0 is the starting value for the process. Large circles are nodes and tips.
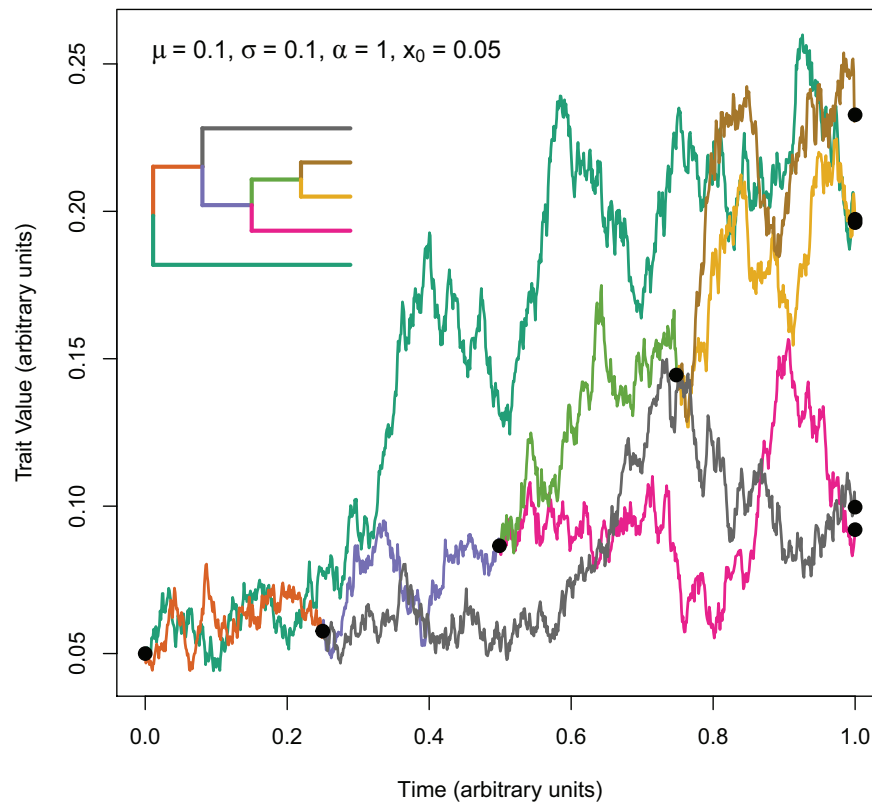
The analysis of equations (7) and (8) and several other examples have been provided by Cobb (1998). It is interesting that in both cases the substitution $\delta = \epsilon/\alpha$ was necessary in order to correctly recognize the distributions as gamma or beta. This suggests that the separate estimation of ϵ and α is difficult if estimation is based solely on the stationary distribution. The same stationary distributions occur for arbitrary α and ϵ, so long as their ratio (δ) remains constant. A sample evolutionary path from this beta process is presented in figure 3.

Fitting Models to Data

To be useful, theory must be confronted with data (Hilborn and Mangel 2013). The evolutionary models discussed here (of which BM and OU are special cases) therefore require methods to fit them to comparative data in order to estimate parameters and test hypotheses about those parameters. For BM, OU, and other Gaussian processes, we can use the machinery developed for normal distributions. In particular, there are simple relationships between branch lengths on a phylogeny and covariances for linear Gaussian processes (Hansen and Martins 1996). Methods have recently been proposed for the calculation

of likelihoods for continuous characters on a tree where the transition density of the evolutionary model is known (Hiscott et al. 2016). Boucher et al. (2018) have recently introduced a new evolutionary model based on the Fokker-Plank equation that discretizes the state space assuming a Gaussian, OU-like model, closely related to the models presented here. Their method involves the inversion of transition matrices along a discretized trait path. However, in cases when the OU approximation is undesirable, as in the present study, simulation methods that do not rely on the inversion of the transition matrix may be preferred. Indeed, often the transition density is not known in closed form or not known at all. If there was no phylogenetic dependence (i.e., a star phylogeny), we could estimate model parameters, as the existence of a stationary distribution implies that at any time point independent samples will follow the stationary distribution. Unfortunately in the case of comparative data with "phylogenetic signal," the data are not independent. We cannot make use of this result.

Simulation is currently the most popular option when the transition density of the process is intractable, although numerical solutions to SDEs are possible, particularly for large data sets (Durham and Gallant 2002; Sørensen 2004; Kloeden and Platen 2011; Malaspinas et al.

2012). Simulation-based methods for estimating parameters for stochastic processes are widely available, largely based on theory developed for use in statistical finance (Iacus 2008). For example, stock prices may be observed every fraction of a second, resulting in a large amount of high-frequency data with which to make inferences. See Schraiber et al. (2016) for an example of analyzing high-frequency data using path augmentation in population genetics. Methods to deal with missing data in the high-frequency setting have been developed (e.g., Roberts and Stramer 2001). In the present context, the formidable problem is that data in many comparative studies are observed only at the tips of the phylogeny. Rarely, internal branches may be calibrated with fossils. The entire evolutionary history of the trait for each species is thus missing and unknown. This is worse than low-frequency data (addressed by Fuchs 2013): it is almost "no-frequency" data. The simulation of the entire evolutionary history, except for the tip and fossil data, is necessary.

The notion of using fossil phenotypes and dates to fix points in the trait-time space is attractive but may contain grave difficulties. Recent studies have emphasized that the inclusion of fossil data can enhance our understanding of trait evolution (Slater et al. 2012), although cladistic criticisms of the use of fossils to establish ancestor-decendent relationships have never been refuted (Engelmann and Wiley 1977; Patterson 1981). The recent development of tip-dating methods may avoid such criticism (Ronquist et al. 2012; O'Reilly et al. 2015; Zhang et al. 2016), and recent approaches to studying models of evolution using fossils are also promising (Hunt 2007; Hunt et al. 2015). However, we may have to be content to build quantitative trait models that incorporate ancestor-descendent relationships as ancillary hypotheses, recognizing that tests of such hypotheses may be impossible for any real data set (but see Stadler 2010). Simulation studies may be valuable in assessing the sensitivity of trait model parameter estimation to fossil placement (as an ancestor or as a sister taxon). Perhaps inferring a fossil as a direct ancestor rather than as a close sister taxon will make little difference to parameter estimates for models of quantitative trait evolution. However, this has yet to be established.

Methods

Implementation

We have developed an R package, menura, for the fitting of non-Gaussian SDE models to trait data on phylogenies using a hybrid data augmentation (DA) and Metropolis-Hastings Markov chain Monte Carlo (MCMC) procedure in a Bayesian context (Tanner and Wong 1987; Papaspiliopoulos et al. 2013). Our scheme alternates between the augmentation (i.e., simulation) of the evolutionary trajectories and

the sampling of parameters via the Metropolis-Hastings algorithm (Fuchs 2013). Such an approach requires the update of small sections (i.e., subtrees) of the simulated trait history at each iteration of the MCMC procedure. Acceptance rates during MCMC are higher when only small parts of the tree are updated at a time (Elerian 1999; Elerian et al. 2001; Roberts and Stramer 2001; Kalogeropoulos 2007). DA is the default setting for path updates (method = "DA" to the main model-fitting function fit_model). We also provide the option of a full Metropolis-Hastings scheme for updating and estimating the evolutionary trajectories (method = "Fuchs"). In practice, we find the simpler DA method performs well. Path simulations are performed using functions from package sde for R (Iacus 2016).

For the "canned" models (OU, CIR, and beta), the Lamperti transformation is used to improve the simulation of trait trajectories by transforming the SDE to one with a unit diffusion coefficient (Lamperti 1962; Burnecki et al. 1997; Møller and Madsen 2010; Fuchs 2013). Consider equation (2). The Lamperti transformation is

$$Y_t = g(X_t) = \int^{X_t} \frac{\mathrm{d}u}{\sigma(u)}.$$

Provided the transformation $g(\cdot)$ exists and is invertible, Y fulfils the diffusion equation:

$$\mathrm{d}Y_t = \left(\frac{b(g^{-1}(Y_t), t)}{\sigma(g^{-1}(Y_t))} - \frac{1}{2}\frac{\partial \sigma}{\partial x}(g^{-1}(Y_t)) \right) \mathrm{d}t + \mathrm{d}B_t,$$

with $Y_{t_0} = g(x_0)$.

Transforming the model to remove any dependence of the diffusion coefficient on $X(t)$ and on t makes the transformed process "more Gaussian," but at the cost of increasing the complexity of the drift coefficient (Iacus 2008). However, there are grave difficulties even with fitting Gaussian models, where the transition density is known in closed form. OU has significant problems (Cooper et al. 2015), including problems with the identifiability of parameters (Hansen 1997; Ho and Ané 2014). Many simulated likelihood methods have been proposed for fitting models where the transition density is unknown (Brandt and Santa-Clara 2002; Durham and Gallant 2002; Sørensen 2004; Cano et al. 2006; Hurn et al. 2007; Kalogeropoulos 2007), including phylogenetic comparative methods (Kutsukake and Innan 2012). These methods often include a discretized, "locally Gaussian" approximation method, such as the Euler scheme or the Milstein scheme (Elerian 1998; Iacus 2008).

In the current version (0.4.4.4), menura can estimate μ, α, or σ using the three canned models. In addition, the user can define other arbitrary diffusion models. The program allows different fixed values of model parameters to occur in different parts of the tree. However, estimated parameters

must be applicable to the whole tree. We do not currently allow for separate rate shifts on subclades of the tree. For example, if one is interested in estimating σ, different values of α and μ can be set arbitrarily in different parts of the tree, but only one value of σ can be estimated for the whole tree. In addition, menura also samples from the joint likelihood of the model, which allows, for example, the post hoc construction of likelihood ratio tests and Bayes factors. A vignette is available in menura that demonstrates the use of the package, and comprehensive help documentation is also provided.

Construction of the model likelihood is necessary to determine the acceptance probability of the Metropolis-Hastings step. In menura, we use the Euler method, a first-order locally Gaussian approximation to the model likelihood (Iacus 2008). In addition, a second-order approximation to the likelihood using the Milstein scheme is available as an option (Milstein 1978). However, the Milstein scheme is more computationally intensive. We note that for the OU model, the Euler and Milstein schemes coincide (Iacus 2008).

The package is named for the Australian superb lyrebird *Menura novaehollandiae*. The lyrebird is the world's largest passerine and is known for its elaborate tail and ability for mimicry (Lill 2004). The lyrebird's talent for mimicry alludes to our use of simulation to "mimic" stochastic diffusions. The R package menura is available at https://github.com/simone66b/Menura.

Simulation Study

To test the efficacy of our new methods implemented in menura, we performed simulations across various tree sizes (number of species), prior specifications, and evolutionary models (i.e., CIR, OU , or beta). We simulated tip data along a tree with known α, μ, and σ (really ϵ as in eqq. [7] and [8], but we shall refer to ϵ as σ). We used menura to estimate various combinations of the model parameters. We ran each simulation for 200,000 iterations. We used balanced trees for all of our simulations. Tree sizes (number of tips) were 32, 64, and 128. Priors were either weakly informative or uninformative. We used Gamma(1, 2) and half-normal (SD = 10) priors for α and σ. Parameters for each prior and tree size are presented in table 1. We plotted the posterior means and standard deviations for each simulation to look for evidence of bias and to examine the variance in our estimates of each parameter. For each combination of tree size, priors, and model we attempted to run 30 simulations, but we discarded simulations that had clearly not converged, as judged by visual inspection of the MCMC trace and standard convergence diagnostics as implemented in the R package coda (Plummer et al. 2006). All analyses were performed using R version 3.5.2 (R Core Team 2018).

Case Study

To demonstrate our methods in practice, we performed an analysis of the longevity of carnivores and ungulates using the data and phylogeny of Tidière et al. (2016) with Grafen's arbitrary branch lengths (Grafen 1989). Longevity is typically skewed and is strictly positive, like a gamma distribution. Hence, this makes the CIR model a possible candidate model. Histograms of the raw data indicated the reasonableness of the CIR model for longevities. We estimated σ, often interpreted as the "rate of evolution" setting $\alpha = 1$. Specifically, we estimated σ separately for both the ungulate and the carnivore clade. To test the hypothesis that these σ's are equal, we randomly sampled from the posterior distributions of each and constructed the ratio of the random samples. Under the null hypothesis, the ratio should be close to unity. Hence, we used the posterior distribution of the ratios to construct a hypothesis test. We used both Gamma(1, 2) and half-normal (SD = 10) priors, although results were qualitatively similar so we only report the analysis with the half-normal prior here. We ran 10 chains each of length 2 million for each clade. We excluded the first million iterations as burn-in and applied a thinning interval of 100. We checked for stationarity of each chain using the Heidelberg and Welch (1983) stationarity test as implemented in the coda package for R (Plummer et al. 2006). In addition, we checked for cross-chain convergence using the Gelman-Rubin test (Gelman and Rubin 1992) in the coda package.

Although menura can estimate α and σ jointly, we found that estimation was much easier if we fixed one of the parameters. This is because when based only on tip data, the OU model (and, by extension, the CIR and beta models) is not identifiable, such that it is not possible to estimate both parameters jointly. We recognize that users would prefer to be able to estimate all model parameters from a given data set; however, we have to work within the mathematical limitations of the methods and data. See the articles by Ho and Ané (2014) and Hansen (1997) for in-depth discussions of this issue. We note that separate estimates of σ and α can be made if a good fossil time series is available.

Results

Simulations

Results from our simulations show that in most cases, estimates are unbiased and close to the "true" values, except when the true values of σ or α are not supported by the prior distributions (figs. 4, 5). For practical purposes, we recommend rerunning analyses with different priors in order to quantify the sensitivity of the results to prior misspecification. If wildly different results are produced using

Table 1: Conditions for the model simulations and estimation tests

Model	α	σ	μ	N_{tips}	Priors	Estimated parameter	Runs
OU	.5	1	0	32	Gamma(2, 1)	α	27
	.5	1	0	64	Gamma(2, 1)	α	10
	.5	1	0	128	Gamma(2, 1)	α	8
	1	1	0	32	Gamma(2, 1)	α	30
	1	1	0	64	Gamma(2, 1)	α	24
	1	1	0	128	Gamma(2, 1)	α	16
	10	1	0	32	Half-normal(0, 10)	α	27
	10	1	0	64	Half-normal(0, 10)	α	4
	10	1	0	128	Half-normal(0, 10)	α	6
	1	.5	0	32	Gamma(2, 1)	σ	8
	1	.5	0	64	Gamma(2, 1)	σ	10
	1	.5	0	128	Gamma(2, 1)	σ	21
	1	1	0	32	Gamma(2, 1)	σ	16
	1	1	0	64	Gamma(2, 1)	σ	21
	1	1	0	128	Gamma(2, 1)	σ	19
	1	10	0	32	Half-normal(0, 10)	σ	16
	1	10	0	64	Half-normal(0, 10)	σ	19
	1	10	0	128	Half-normal(0, 10)	σ	17
CIR	.5	1	1	32	Gamma(2, 1)	α	24
	.5	1	1	64	Gamma(2, 1)	α	24
	.5	1	1	128	Gamma(2, 1)	α	13
	1	1	1	32	Gamma(2, 1)	α	28
	1	1	1	64	Gamma(2, 1)	α	23
	1	1	1	128	Gamma(2, 1)	α	8
	10	1	1	32	Half-normal(0, 10)	α	14
	10	1	1	64	Half-normal(0, 10)	α	4
	10	1	1	128	Half-normal(0, 10)	α	11
	1	.5	1	32	Gamma(2, 1)	σ	20
	1	.5	1	64	Gamma(2, 1)	σ	16
	1	.5	1	128	Gamma(2, 1)	σ	19
	1	1	1	32	Gamma(2, 1)	σ	12
	1	1	1	64	Gamma(2, 1)	σ	18
	1	1	1	128	Gamma(2, 1)	σ	20
	1	5	6	32	Half-normal(0, 10)	σ	14
	1	5	6	64	Half-normal(0, 10)	σ	12
	1	5	6	128	Half-normal(0, 10)	σ	12
Beta	.5	1	.5	32	Gamma(2, 1)	α	16
	.5	1	.5	64	Gamma(2, 1)	α	8
	.5	1	.5	128	Gamma(2, 1)	α	9
	1	1	.6	32	Gamma(2, 1)	α	13
	1	1	.6	64	Gamma(2, 1)	α	7
	1	1	.6	128	Gamma(2, 1)	α	8
	10	1	.6	32	Half-normal(0, 10)	α	21
	10	1	.6	64	Half-normal(0, 10)	α	11
	10	1	.6	128	Half-normal(0, 10)	α	6

Note: For each model, simulations were conducted for the values of α, σ, and μ, and then either α or σ were estimated. N_{tips} is the tree size. Priors were Gamma(a, b), where a is the shape parameter and b is the rate parameter, or Half-normal(m, s), where m and s are the mean and standard deviation of the underlying normal distribution. The maximum number of runs for any simulation condition was 30; however, we retained only the runs that had clearly converged (column "Runs"). CIR = Cox-Ingersoll-Ross; OU = Ornstein-Uhlenbeck.

different (but equally plausible) priors, this is an indication that there may not be sufficient data for reliable estimation. In our experience, priors that "agree" with the data also have better convergence properties in MCMC runs. Based on the standard deviations for each simulation, σ was estimated with more precision than α. However, we had difficulty in achieving convergence for many simulations, so we recommend running multiple (more
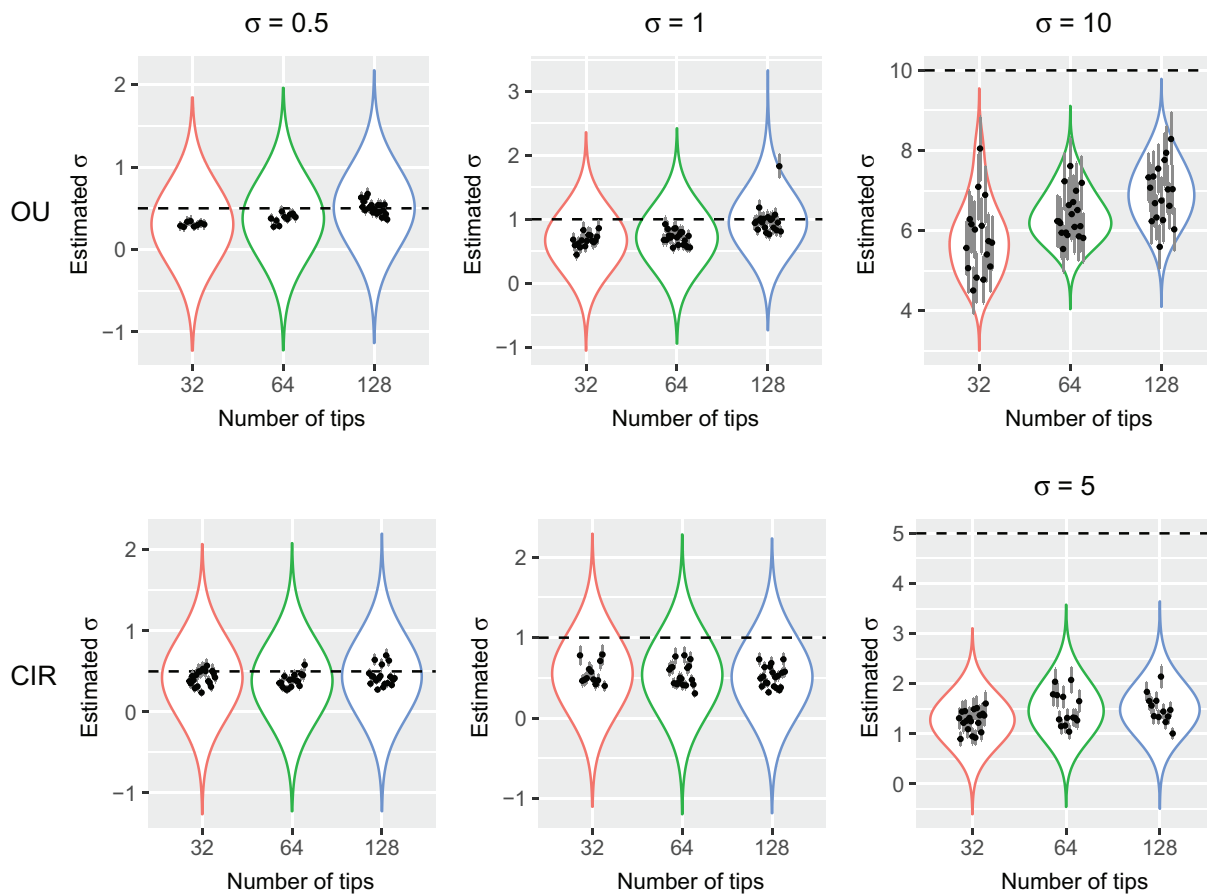
Figure 4: Results of simulations for the Ornstein-Uhlenbeck (OU) and Cox-Ingersoll-Ross (CIR) models for different-sized trees and σ values. Dashed lines represent the "true" target value. Black circles represent the posterior mean for each simulation (chain). The gray vertical bars represent ± 2 posterior standard deviations.

than three chains) long MCMC chains (>1 million iterations) and examining the plots of the traces by eye as well as using the standard convergence diagnostics (e.g., Gelman and Rubin 1992).

Case Study

While the Gelman-Rubin test with the carnivore data was reasonable ($R = 1.01$ with an upper 95% confidence limit of 1.02), this was not true for the ungulate data ($R = 1.93$ with an upper 95% confidence limit of 2.58). Inspection of the traces revealed, as with the simulations, that posterior distributions for individual chains seemed reasonable but that the posterior means were slightly different for each chain. We subsequently pooled the chains for the ungulate data. However, we wish to emphasize that doing so results in an overly conservative test, as the posterior variance of the pooled chain will be larger than if the chains had truly converged to the same posterior mean. As both pooled chains for ungulates and carnivores were unimodal, this

problem appeared to make little difference to the pooled mean estimates.

Results of our estimation of σ are presented in figure 6. Posterior mean estimates of σ were 2.36 ± 0.23 (SD) and 1.87 ± 0.21 for ungulates and carnivores, respectively. Our test of the null hypothesis that the two σ's were equal resulted in a P value of .062. That is, although the mean ratio of ungulate σ to carnivore σ was 1.28 ± 0.19, the probability of the ratio being less than unity was 0.062, slightly greater than the conventional level for significance (see fig. 7).

Discussion

Simulations

Given the lack of information in the data regarding the model dynamics and in the absence of fossil information, we found that the judicious choice of prior for each parameter was important. Especially if the priors placed little
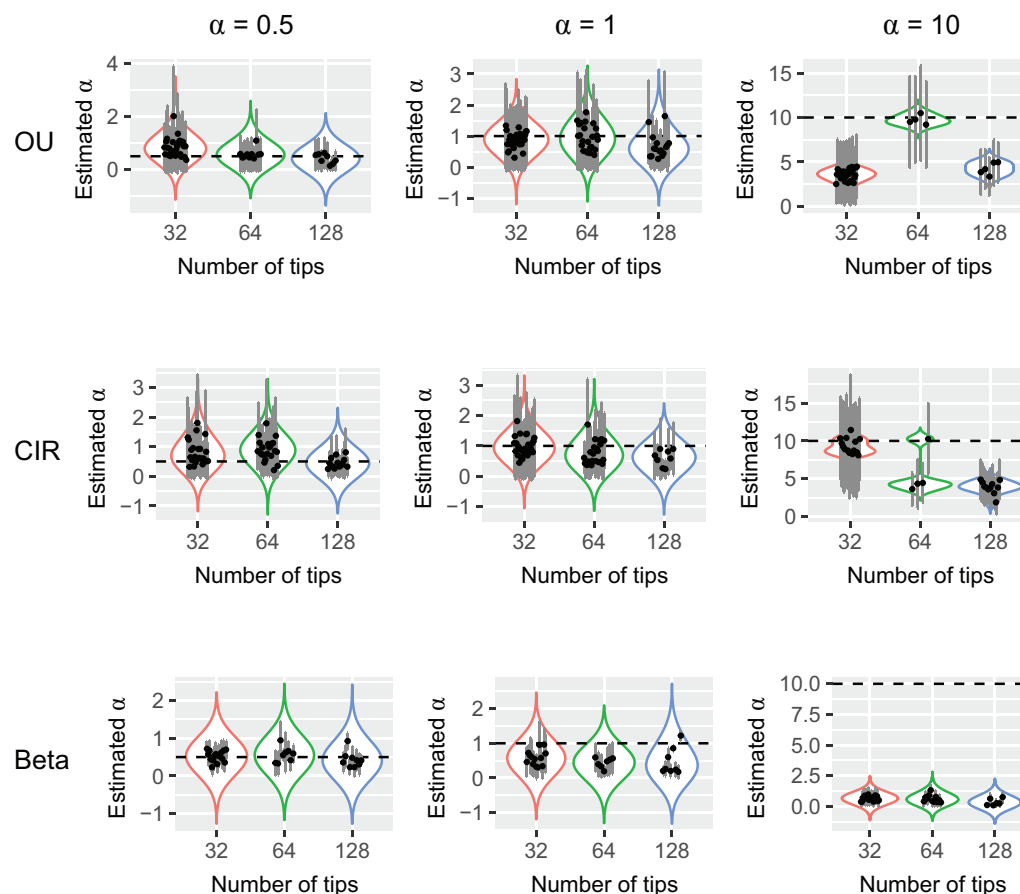
Figure 5: Results of simulations for the Ornstein-Uhlenbeck (OU), Cox-Ingersoll-Ross (CIR), and beta models, for different-sized trees and α values. Dashed lines represent the "true" target value. Black circles represent the posterior mean for each simulation (chain). The gray vertical bars represent ± 2 posterior standard deviations.

probability mass near the true value, then estimates were particularly poor (e.g., right column of fig. 5). Hence, we recommend experimenting with priors (Gelman et al. 2017) and in particular using weakly informative priors—that is, priors that regularize the problem but still provide enough flexibility to correctly estimate the model parameters (Stan Development Team 2017).

Case Study

There appears to be a difference in σ for longevity between ungulates and carnivores (figs. 6, 7), although more detailed work is necessary to arrive at solid conclusions. We experienced difficulty in getting chains to converge for the ungulate data (see fig. 6), even after 2 million iterations. This is a serious problem, as inferences are more difficult or impossible and conclusions are weaker when based on nonconvergent chains. We speculate that some of the problem may be due to rate heterogeneity (i.e., within clade variability in σ). The results for carnivores were much tighter

even though they were based on only 17 species (compared with 30 for the ungulates), and therefore there was less opportunity for rate heterogeneity to be a problem (fig. 6). The analysis presented here is only for illustration of the types of analyses that menura can perform and should not be considered in any way definitive. Nevertheless, fitting the CIR model to mammal longevity is novel and suggests that further exploration of the model in this context may be fruitful.

Model Extensions

An obvious extension of univariate stochastic processes is to recast them in a multivariate or multidimensional framework. There has been some research into multivariate phylogenetic comparative methods, including several software packages, largely based on BM, OU, and early-burst models (Zheng et al. 2009; Bartoszek 2011; Klingenberg 2011; Bartoszek et al. 2012; Klingenberg and Marugán-Lobón 2013; Adams 2014*a*, 2014*b*, 2014*c*; Clavel et al. 2015,
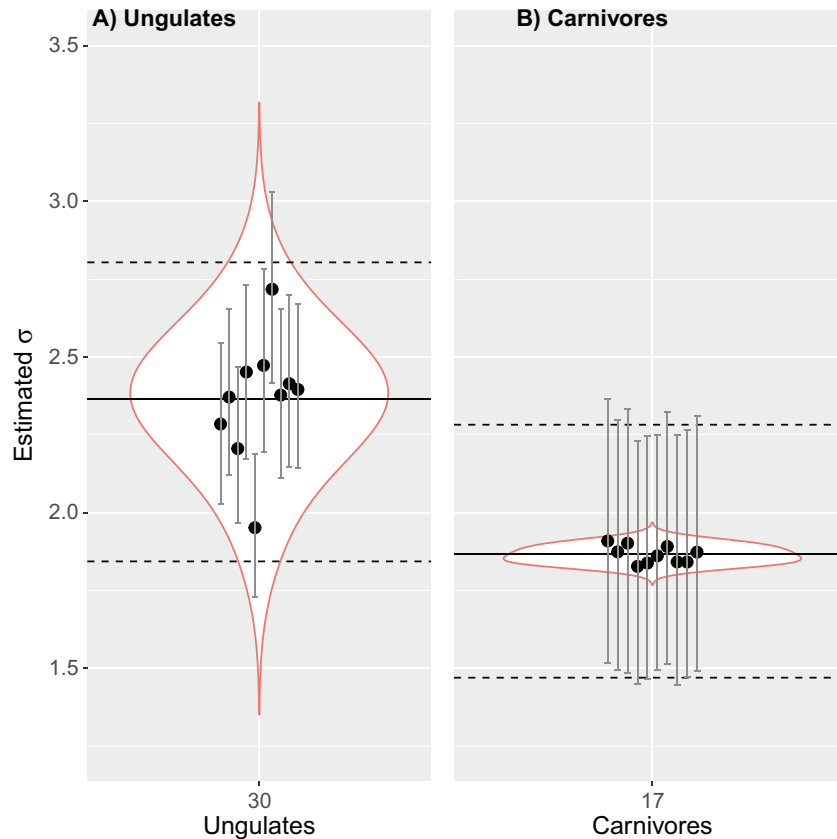
Figure 6: Results for the case study comparing σ for ungulates (*A*) and carnivores (*B*). For ungulates and carnivores, violin plots are of the posterior means for each chain. Error bars are the 95% highest posterior density interval for each chain. The solid line is the pooled posterior mean, and the dashed lines are the 95% highest posterior density intervals for the pooled chains. The bandwidth was 0.2 for violin plot *A* and 0.02 for plot *B*.

2018; Adams and Collyer 2018; Caetano and Harmon 2018). Certainly, multivariate diffusions are necessary to understand the correlation among characters (Bartoszek et al. 2012). However, the properties of univariate diffusion models do not always carry over to the multivariate setting. In particular, there are well-known differences between the recurrence and transience properties of BM in multiple dimensions (Mörters and Peres 2010).

A further extension of diffusion models is to the case where evolution is not strictly continuous but is punctuated by "jumps" using Lévy processes (Landis et al. 2012). Lévy processes are stochastic processes with independent, stationary increments. They can be thought of as consisting of three superimposed processes:

$$X_t = \sigma B_t + J_t + M_t,$$

where B_t is a BM (possibly with drift), J_t is a compound Poisson point process, and M_t is a (square-integrable) martingale with jumps. Hence, simple BM is a special case of a Lévy process with no discrete jumps. Note that OU is not a

Lévy process. Landis et al. (2012) estimate parameters for a Lévy process fitted to data for body mass and brain volume in primates and found evidence for some jumps in each trait, rejecting a simple BM model. Duchen et al. (2017) used Lévy process models to look for jumps in *Anolis* lizard body size and Australasian lories (Aves: Loriini) gut morphology. Both taxa showed evidence for evolutionary jumps. It may be difficult to choose between jump models and models that estimate rapid changes in evolutionary rate (large differences in σ) for particular clades (e.g., Alfaro et al. 2009; Rabosky et al. 2013, 2014; Shi and Rabosky 2015), although jump models tend to have fewer parameters, which may make them more parsimonious.

Models with jumps may be a good approximation in situations where evolution, though continuous, has been so rapid that on the macroevolutionary timescale change appears to be instantaneous (Duchen et al. 2017). Landis and Schraiber (2017) find evidence for jumps in the evolution of vertebrate body size and interpret this as rapid evolution, followed by stasis.
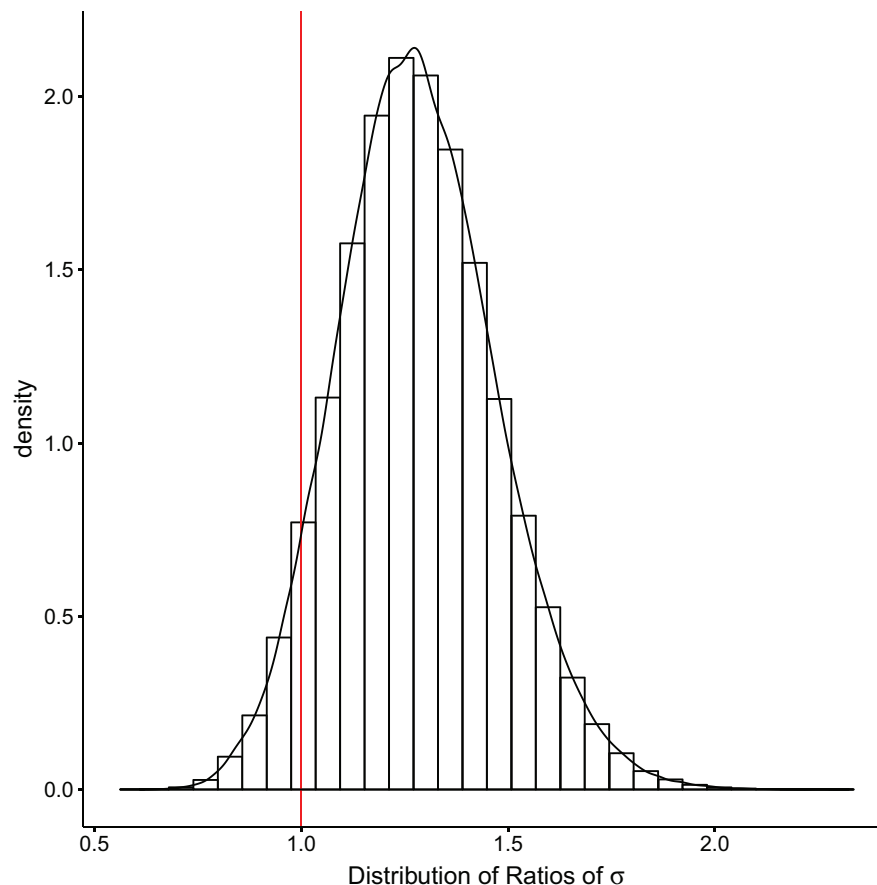
Figure 7: Bayesian test of the null hypothesis of equal ratios for ungulate σ over carnivore σ, showing some evidence for a larger σ for ungulates ($P = .062$). The red line is at unity (the null hypothesis).

Stochastic Differential Equations from Stationary Distributions

Deriving an SDE given a stationary distribution is more difficult, since the correspondence between SDEs and their stationary distribution (if it exists) is not unique. The problem has been addressed by Cai and Lin (1996). Extra information is needed, specifically the form of the spectral density of the process that affects the structure of the drift coefficient in the SDE. Equivalently, the autocorrelation function (ACF) of the process can be estimated and if the integral of the absolute value of the ACF is finite, the spectral density is the Fourier transform of the ACF.

Unfortunately for models of trait evolution, we rarely have detailed information on the evolutionary trajectory of a trait (i.e., the true historical realization of the process), and hence we cannot analyze the spectral density of the trajectory in order to infer an appropriate drift function. If we are to proceed, we need to make extra assumptions.

Fortunately, if we assume that the spectral density is of the low-pass filter type—that is,

$$\Phi_{XX}(\omega) = \frac{\alpha \delta^2}{\pi(\omega^2 + \alpha^2)}, \tag{9}$$

where Φ_{XX} is the spectral density at frequency ω and δ^2 is the mean-square value of the process $X(t)$—then the drift coefficient will be of the reverting OU type in equation (3), with α in equation (9) being identical to α in equation (3). The low-pass filter assumption implies that the drift function is determined mainly by the low-frequency (long-wavelength) characteristics of the evolutionary trajectory. That is, the form of the drift is mainly determined by long-lasting slow deviations from μ, and short-term (high-frequency) excursions are less important. To our knowledge, this assumption has never been made explicit in the literature on the application of the OU model in phylogenetic comparative methods.

Calculation of the diffusion coefficient comes directly from the application of the time-homogeneous Fokker-Planck equation (5), except instead of solving for $f(x)$, we now solve for $\sigma(x)$ (Cai and Lin 1996). The expression for $\sigma^2(x)$ becomes

$$\sigma^2(x) = -\frac{2\alpha}{f(x)} \int^x yf(y)\,\mathrm{d}y.$$

Stationarity

The notions of stationarity and stationary distributions have been central to this study. In the absence of an excellent fossil record of trait evolution for most traits and most taxa, it seems to be a necessary (although strong) assumption for evolutionary stochastic process models that are more complicated than BM. Indeed, the success of the OU process in evolutionary studies is almost as much based on its stationarity as its Gaussian properties. Several authors have constructed nonstationary evolutionary models based on OU (e.g., Bartoszek 2012; Beaulieu et al. 2012; Jhwueng and Maroulas 2014). Nonstationarity can arise because of time dependence in the drift coefficient, time dependence in the diffusion coefficient, or both. For mean-reverting processes, the mean of the process μ and/or the strength of the restraining force α may be time dependent (Beaulieu et al. 2012); σ may vary with time smoothly over the tree (Bartoszek 2012).

Aside from the problem of overparameterization (Bartoszek 2012), different parameters on different clades of the tree imply at least a short period of nonstationarity as species evolve from an ancestral evolutionary regime to the new conditions. Some OU-based models assume immediate stationarity after the change in evolutionary regime (e.g., Butler and King 2004). If the old regime is almost the same as the new conditions, then stationarity in the new conditions may be achieved relatively quickly. However, if the old regime is very different from the new one, the length of the nonstationary period may be considerable and the underlying "instantaneous" stationary model will be wrong. Only fossil evidence can help in this regard because fossils can provide fixed points in the morphospace-time that can anchor the model and provide evidence of non-stationary trait evolution or stasis. Of course, if the ancestral and derived stationary distributions are very similar so that stationarity is achieved quickly, it will be difficult to tell these two scenarios apart.

Transition Distributions

The stationary distribution is not the only distribution associated with a Markov diffusion process. The transition, or conditional, distribution is important for simulation and likelihood calculations (Iacus 2008). It can be found as a solution to the Fokker-Planck equation (Klebaner 2012) and is defined as

$$P(y, t, x, s) = P(X(t) \le y | X(s) = x). \qquad (10)$$

Equation (10) defines the probability distribution of y and the value of X occurring at time t, given that the process X has reached x at time s, where $s < t$. Unfortunately, for most processes the transition distribution is unknown or intractable. For Gaussian processes, the conditional density is usually straightforward and can be used to construct phylogenetic variance-covariance matrices for use in phylogenetic generalized least squares (PGLS) analyses (Hansen and Martins 1996). BM has as its transition distribution the normal distribution with mean $\mu = \mathrm{E}(X_t | X_s = x) = x$, by the martingale property. The conditional variance of BM is $\mathrm{Var}(X_t | X_s = x) = \sigma^2 t$. That is, the variance is independent of the trait value and depends only on t.

For the OU process (3) and $t > s \ge 0$, the transition density is Gaussian with mean $\mathrm{E}(X_t | X_s = x) = xe^{-\alpha(t-s)} + \mu(1 - e^{-\alpha(t-s)})$ and variance $\mathrm{Var}(X_t | X_s = x) = (\sigma^2/2\alpha)(1 - e^{-2\alpha(t-s)})$. The complexity of transition distributions increases quickly with the complexity of the corresponding SDE. Equation (7), the CIR model, has the following transition density (Cox et al. 1985):

$$f(s, x, t, y) = c\left(\frac{u}{v}\right)^{\nu/2} \exp(-(u+v))I_\nu(2\sqrt{uv})$$

for $t > s \ge 0$, where

$$c = \frac{2\alpha}{\epsilon^2(1 - e^{-\alpha(t-s)})}, \quad u = cx^{-\alpha(t-s)},$$

$$v = cy, \quad \nu = \frac{2\alpha\mu}{\epsilon^2} - 1.$$

The term I_ν is the modified Bessel function of the first kind of order ν:

$$I_\nu(z) = \sum_{k=0}^{\infty} \left(\frac{z}{2}\right)^{2k+\nu} \frac{1}{k!\,\Gamma(k + \nu + 1)},$$

where $z \in \mathbb{R}^+$ and $\Gamma(\cdot)$ is the gamma function. The expectation and variance of this distribution are

$$\mathrm{E}(X_t | X_s = x) = 2\frac{\alpha\mu}{c\epsilon^2} + xe^{-\alpha(t-s)},$$

$$\mathrm{Var}(X_t | X_s = x) = \frac{2}{c}\left(\frac{\alpha\mu}{c\epsilon^2} + xe^{-\alpha(t-s)}\right),$$

respectively. The transition density of equation (8) is even more complicated and involves infinite sums of hyper-geometric functions (Abundo 1997). Transition densities are of extreme importance for phylogenetic comparative methods, as they determine the structure of the evolutionary

covariance matrix and the relationship between branch lengths and covariances (Hansen and Martins 1996). Hence, estimation approaches such as PGLS (Martins and Hansen 1997; Grafen 1989; Blomberg et al. 2012) are intimately dependent on knowing transition densities. However, even quite simple models like the non-Gaussian models presented here are likely to present formidable problems with calculating evolutionary covariances, as the formulae for the transition densities for these models are extremely difficult to work with, if they are known at all.

Evolutionary Models for Phylogenetic Comparative Analysis

Scientific models may be developed with several different motivations (Gavrilets 1999). The scientist may build models to make a decision (e.g., to reject a null hypothesis), summarize evidence (e.g., calculate the likelihood of observing the data, given a model), or quantify their beliefs (using Bayes theorem). Another important property of a model is its predictive ability, and predictive models have long been the favorite approach in the physical sciences: models predict future observations, which then test the validity of the model. Biologists, and especially evolutionary biologists, have never put much faith in predictive models (Hillis 1993; Gavrilets 1999). So many factors affect the evolution of organisms—and over such a long time span—that one is tempted to give up hope of developing mathematical models that have any predictive value in the real world. It is true that it would be foolish to make predictions of where in the phylomorphospace (sensu Sidlauskas 2008) species will evolve to in some future deep time. We have no hope of making the necessary observations.

Although we may not be able to predict the precise long-term evolutionary trajectory of any particular species, we can perhaps predict (or postdict) the probability distribution of traits across species. We may use cross validation (Efron and Gong 1983) to assess the predictive ability of our models, or in a Bayesian context we may use posterior predictive simulation (Gelman et al. 2013). Given the traits from a newly discovered species (fossil or extant), we can predict that the new trait values fit well within the distribution of the known species' trait values, where the parameters of the distribution are estimated from extant species using a particular model of evolution. If the values for the new species' traits are more extreme so that they fall into the tails of the stationary distribution, we may reject our model of evolution for that set of traits and species. Predictions of the models may not be precise, but they are predictions nonetheless (see Pigliucci 2007). This "gray box" approach to model identification (Kristensen et al. 2004) gives up the possibility of knowledge of the microevolutionary processes leading to species diversification

and trait evolution. This is replaced with a tractable stochastic process that summarizes the evolution of the statistical distribution of trait values over deep time. Given the poor quality of many comparative data sets (little or no fossil data, incomplete sampling of extant taxa, low intraspecific replication), this may be the best that can be achieved.

The modeling approach and the new models described here involve a considerable amount of mathematical sophistication in their derivation and in the analysis of their properties. Computational skill is necessary in developing algorithms to fit the models to data. Critics may object that the approach outlined here is too complex or unnecessary. However, diffusions are already the most popular model for phylogenetic comparative studies, in the form of BM and OU. We hope simply to widen horizons and provide a unifying framework. It is true that all models are wrong but some are useful (Box 1976). Nevertheless, mathematics (and its sister taxon, computation) are the best tools we have to precisely describe both the nature of macroevolutionary phenomena and our assumptions about them. A small amount of precise mathematics can sometimes cut through imprecise verbal arguments. For example, the microevolutionary genetic theory developed by Fisher, Haldane, and Wright effectively silenced the arguments between naturalists and Mendelians on the importance of natural selection and the nature of genetic variation, leading to the evolutionary synthesis (Mayr and Provine 1998). A mathematical theory of macroevolution that unites stochastic models of trait evolution with models of phylogenesis, speciation, and extinction may allow us to better statistically model the course of phenotypic evolution (e.g., Maddison et al. 2007; FitzJohn 2010; Goldberg et al. 2011), although estimating parameters for these models may be difficult without fossil trait data. Analytical models of trait-dependent diversification might be useful. Recent applications of trait-mediated diversification models based only on extant trait data may be misleading (Rabosky and Goldberg 2015; Beaulieu and O'Meara 2016). A more sophisticated understanding of the mathematics of diffusions and other stochastic processes may allow the critical appraisal of macroevolutionary models for biological phenomena in deep time.

Conclusion

Currently, popular models of trait evolution rely heavily on Gaussian processes and their useful mathematical properties. However, non-Gaussian models are possible and may have some advantages over Gaussian models in certain situations where the data are likely to be nonnormal. The present study describes new, non-Gaussian models of trait evolution, together with methods for building new models and a discussion of the mathematical and computational difficulties in working with diffusion models in a more

generalized setting. Several new avenues for investigation are suggested. In particular, the role of fossils in improving the identifiability of models and the extension of models to multivariate trait space seem especially timely. These areas are not without challenges. Including fossils as ancestors rather than as sister taxa has been a difficult problem for many years, as the early cladists were well aware. The extension of univariate models to multivariate trait space is likely to be more difficult than expected, as even the simplest evolutionary model, BM, has different properties in multiple dimensions. Another important research direction is to establish the expected covariances for traits in terms of the transition distributions for non-Gaussian models. This is likely to be difficult but would pay off immensely, allowing estimation of regression models by PGLS and permitting the construction of a new generalized phylogenetic model by analogy with generalized linear models (McCullagh and Nelder 1989).

Is Felsenstein still correct in his quotation at the beginning of this article? Certainly, non-Gaussian diffusions are harder to work with and are less tractable. However, some advances have been made. There are some interesting analytical results, as outlined in the present study. Simulation is now much easier and can be used to explore the properties of diffusion processes on trees. Research into the application of diffusion models to the evolution of quantitative traits appears to hold great promise.

Acknowledgments

Literature Cited

Abundo, M. 1997. On some properties of one-dimensional diffusion processes on an interval. Probability and Mathematical Statistics 17:277–310.

Adams, D. C. 2014a. A generalized *K* statistic for estimating phylogenetic signal from shape and other high-dimensional multivariate data. Systematic Biology 63:685–697.

———. 2014b. A method for assessing phylogenetic least squares models for shape and other high-dimensional multivariate data. Evolution 68:2675–2688.

———. 2014c. Quantifying and comparing phylogenetic evolutionary rates for shape and other high-dimensional phenotypic data. Systematic Biology 63:166–177.

Adams, D. C., and M. L. Collyer. 2018. Multivariate phylogenetic comparative methods: evaluations, comparisons, and recommendations. Systematic Biology 67:14–31.

Agrawal, A. F., and J. R. Stinchcombe. 2009. How much do genetic covariances alter the rate of adaptation? Proceedings of the Royal Society B 276:1183–1191.

Alfaro, M. E., F. Santini, C. Brock, H. Alamillo, A. Dornburg, D. L. Rabosky, G. Carnevale, and L. J. Harmon. 2009. Nine exceptional radiations plus high turnover explain species diversity in jawed vertebrates. Proceedings of the National Academy of Sciences of the USA 106:13410–13414.

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

Bachelier, L. 1900. Theorie de la speculation. DPhil thesis. Ecole Normale Superieure, Paris.

Barndorff-Nielsen, O. E., and N. Shephard. 2001. Non-Gaussian Ornstein-Uhlenbeck-based models and some of their uses in financial economics. Journal of the Royal Statistical Society B 63:167–241.

Bartoszek, K. 2011. Multivariate aspects of phylogenetic comparative methods. Licentiate thesis. University of Gothenburg.

———. 2012. The Laplace motion in phylogenetic comparative methods. Pages 25–30 in Proceedings of the XVIII National Conference on Applications of Mathematics in Biology and Medicine. 403.1562 (q-bio.PE).

Bartoszek, K., J. Pienaar, P. Mostad, S. Andersson, and T. F. Hansen. 2012. A phylogenetic comparative method for studying multivariate adaptation. Journal of Theoretical Biology 314:204–215.

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.

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

Benton, M. J. 2015. Exploring macroevolution using modern and fossil data. Proceedings of the Royal Society B 282:20150569.

Black, F., and M. Scholes. 1973. The pricing of options and corporate liabilities. Journal of Political Economy 81:637–654.

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

Blomberg, S. P., J. G. Lefevre, J. A. Wells, and M. Waterhouse. 2012. Independent contrasts and PGLS regression estimators are equivalent. Systematic Biology 61:382–391.

Blows, M. W., and A. A. Hoffmann. 2005. A reassessment of genetic limits to evolutionary change. Ecology 86:1371–1384.

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

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

Box, G. E. P. 1976. Science and statistics. Journal of the American Statistical Association 71:791–799.

Brandt, M. W., and P. Santa-Clara. 2002. Simulated likelihood estimation of diffusions with an application to exchange rate dynamics in incomplete markets. Journal of Financial Economics 63:161–210.

Breiman, L. 1968. Probability. Addison-Wesley, Reading, MA.

Burnecki, K., M. Maejima, and A. Weron. 1997. The Lamperti transformation for self-similar processes. HSC Research Report no. HSC/97/02. Hugo Steinhaus Center, Wrocław University of Technology.

Butler, M. A., and A. A. King. 2004. Phylogenetic comparative analysis: a modeling approach for adaptive evolution. American Naturalist 164:683–695.

Caetano, D. S., and L. J. Harmon. 2018. Estimating correlated rates of trait evolution with uncertainty. Systematic Biology 68:412–429.

Cai, G. Q., and Y. K. Lin. 1996. Generation of non-Gaussian stationary stochastic processes. Physical Review E 54:299–303.

Cano, J., M. Kessler, and D. Salmerón. 2006. Approximation of the posterior density for diffusion processes. Statistics and Probability Letters 76:39–44.

Clavel, J., L. Aristide, and H. Morlon. 2018. A penalized likelihood framework for high-dimensional phylogenetic comparative methods and an application to new-world monkeys brain evolution. Systematic Biology 68:93–116.

Clavel, J., G. Escarguel, and G. Merceron. 2015. mvmorph: an R package for fitting multivariate evolutionary models to morphometric data. Methods in Ecology and Evolution 6:1311–1319.

Cobb, L. 1998. Stochastic differential equations for the social sciences. Mathematical Frontiers of the Social and Policy Sciences.

Conway Morris, S., J. F. Hoyal Cuthill, and S. Gerber. 2015. Hunting Darwin's snark: which maps shall we use? Interface Focus 5(6).

Cooper, N., G. H. Thomas, C. Venditti, A. Meade, and R. P. Freckleton. 2015. A cautionary note on the use of Ornstein Uhlenbeck models in macroevolutionary studies. Biological Journal of the Linnean Society 118:64–77.

Cox, J. C., J. E. Ingersoll, and S. A. Ross. 1985. A theory of the term structure of interest rates. Econometrica 53:385–407.

Darriba, D., G. Taboada, R. Doallo, and D. Posada. 2012. jmodeltest2: more models, new heuristics and parallel computing. Nature Methods 9:772.

De Maio, N., D. Schrempf, and C. Kosiol. 2015. PoMo: an allele frequency–based approach for species tree estimation. Systematic Biology 64:1018–1031.

Du Toit Strauss, R., and F. Effenberger. 2017. A hitch-hiker's guide to stochastic differential equations: solution methods for energetic particle transport in space physics and astrophysics. Space Science Reviews 212:151–192.

Duchen, P., C. Leuenberger, S. M. Szilágyi, L. Harmon, J. Eastman, M. Schweizer, and D. Wegmann. 2017. Inference of evolutionary jumps in large phylogenies using Lévy processes. Systematic Biology 66:950–963.

Durham, G. B., and A. R. Gallant. 2002. Numerical techniques for maximum likelihood estimation of continuous-time diffusion processes. Journal of Business and Economic Statistics 20:297–316.

Edwards, A. W. F., and L. L. Cavalli-Sforza. 1964. Reconstruction of evolutionary trees. Pages 67–76 in V. H. Heywood and J. McNeill, eds. Phenetic and phylogenetic classification. Systematics Association Publication no. 6. Systematics Association, London.

Efron, B., and G. Gong. 1983. A leisurely look at the bootstrap, the jackknife, and cross-validation. American Statistician 37:36–48.

Einstein, A. 1956. Investigations on the theory of the Brownian movement. Dover, Mineola, NY.

Elerian, O. 1998. A note on the existence of a closed form conditional transition density for the Milstein scheme. Economics Series Working Paper no. 1998-W18, Department of Economics, University of Oxford.

———. 1999. Simulation estimation of continuous-time models with applications to finance. PhD thesis, University of Oxford.

Elerian, O., S. Chib, and N. Shephard. 2001. Likelihood inference for discretely observed nonlinear diffusions. Econometrica 69:959–993.

Eliazar, I. I., and M. H. Cohen. 2012. The misconception of mean-reversion. Journal of Physics A: Mathematical and Theoretical 45:332001.

Engelmann, G. F., and E. O. Wiley. 1977. The place of ancestor-descendant relationships in phylogeny reconstruction. Systematic Biology 26:1–11.

Estes, S., and S. J. Arnold. 2007. Resolving the paradox of stasis: models with stabilizing selection explain evolutionary divergence on all timescales. American Naturalist 169:227–244.

Ewens, W. J. 2004. Mathematical population genetics. 2nd ed. Springer, Berlin.

Feller, W. 1951. Diffusion processes in genetics. Pages 227–246 in Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability. University of California Press, Berkeley.

Felsenstein, J. 1973. Maximum-likelihood estimation of evolutionary trees from continuous characters. American Journal of Human Genetics 25:471–492.

———. 1985. Phylogenies and the comparative method. American Naturalist 125:1–15.

———. 1988. Phylogenies and quantitative characters. Annual Review of Ecology and Systematics 19:445–471.

———. 2012. A comparative method for both discrete and continuous characters using the threshold model. American Naturalist 179:145–156.

Fisher, R. A. 1922. On the dominance ratio. Proceedings of the Royal Society of Edinburgh 42:321–341.

FitzJohn, R. G. 2010. Quantitative traits and diversification. Systematic Biology 59:619–633.

Freckleton, R. P., P. H. Harvey, and M. Pagel. 2002. Phylogenetic analysis and comparative data: a test and review of the evidence. American Naturalist 160:712–726.

Freund, J. A., and T. Pöschel, eds. 2000. Stochastic processes in physics, chemistry, and biology. No. 0075-8450 in Lecture Notes in Physics. Springer, Berlin.

Fuchs, C. 2013. Inference for diffusion processes: with applications in life sciences. Springer, Berlin.

Futuyma, D. J. 2010. Evolutionary constraint and ecological consequences. Evolution 64:1865–1884.

Gardiner, C. 2009. Stochastic methods: a handbook for the natural and social sciences. 4th ed. Springer Series in Synergetics. Springer, Berlin.

Garland, T., A. W. Dickerman, C. Janis, and J. A. Jones. 1993. Phylogenetic analysis of covariance by computer simulation. Systematic Biology 42:265–292.

Gavrilets, S. 1999. A dynamical theory of speciation on holey adaptive landscapes. American Naturalist 154:1–22.

Gelman, A., J. Carlin, H. Stern, D. Dunson, A. Vehtari, and D. Rubin. 2013. Bayesian data analysis. 3rd ed. Texts in Statistical Science. CRC, Boca Raton, FL.

Gelman, A., and D. Rubin. 1992. Inference from iterative simulation using multiple sequences. Statistical Science 7:457–511.

Gelman, A., D. Simpson, and M. Betancourt. 2017. The prior can often only be understood in the context of the likelihood. Entropy 19:555.

Goldberg, E. E., L. T. Lancaster, and R. H. Ree. 2011. Phylogenetic inference of reciprocal effects between geographic range evolution and diversification. Systematic Biology 60:451–465.

Grafen, A. 1984. Natural selection, kin selection and group selection. Behavioural Ecology 2:62–84.

———. 1989. The phylogenetic regression. Philosophical Transactions of the Royal Society B 326:119–157.

Hadfield, J. D. 2010. MCMC methods for multi-response generalized linear mixed models: the MCMCglmm R package. Journal of Statistical Software 33:1–22.

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

Hansen, T. F., and E. P. Martins. 1996. Translating between microevolutionary process and macroevolutionary patterns: the correlation structure of interspecific data. Evolution 50:1404–1417.

Hansen, T. F., J. Pienaar, and S. H. Orzack. 2008. A comparative method for studying adaptation to a randomly evolving environment. Evolution 62:1965–77.

Heidelberger, P., and P. D. Welch. 1983. Simulation run length control in the presence of an initial transient. Operations Research 31:1109–1144.

Hilborn, R., and M. Mangel. 2013. The ecological detective. Monographs in Population Biology. Princeton University Press, Princeton, NJ.

Hillis, W. D. 1993. Why physicists like models and why biologists should. Current Biology 3:79–81.

Hiscott, G., C. Fox, M. Parry, and D. Bryant. 2016. Efficient recycled algorithms for quantitative trait models on phylogenies. Genome Biology and Evolution 8:1338–1350.

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

Hofrichter, J., J. Jost, and T. D. Tran. 2017. Information geometry and population genetics: the mathematical structure of the Wright-Fisher model. Understanding Complex Systems. Springer International Publishing, Cham.

Huey, R. B., T. Garland, and M. Turelli. 2019. Revisiting a key innovation in evolutionary biology: Felsenstein's "Phylogenies and the Comparative Method." American Naturalist 193:755–772.

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 of the USA 104:18404–18408.

Hunt, G., M. J. Hopkins, and S. Lidgard. 2015. Simple versus complex models of trait evolution and stasis as a response to environmental change. Proceedings of the National Academy of Sciences of the USA 112:4885–4890.

Hurn, A. S., J. I. Jeisman, and K. A. Lindsay. 2007. Seeing the wood for the trees: a critical evaluation of methods to estimate the parameters of stochastic differential equations. Journal of Financial Econometrics 5:390–455.

Iacus, S. M. 2008. Simulation and inference for stochastic differential equations: with R examples. Springer Series in Statistics. Springer, Berlin.

———. 2016. sde: simulation and inference for stochastic differential equations. R package version 2.0.15.

Ingram, T., L. J. Harmon, and J. B. Shurin. 2012. When should we expect early bursts of trait evolution in comparative data? predictions from an evolutionary food web model. Journal of Evolutionary Biology 25:1902–1910.

Ingram, T., and D. 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.

Itô, K. 1944. Stochastic integral. Proceedings of the Imperial Academy 20:519–524.

———. 1946. On a stochastic integral equation. Proceedings of the Japan Academy 22:32–35.

Ives, A. R. 2015. For testing the significance of regression coefficients, go ahead and log-transform count data. Methods in Ecology and Evolution 6:828–835.

Jhwueng, D.-C., and V. Maroulas. 2014. Phylogenetic Ornstein-Uhlenbeck regression curves. Statistics and Probability Letters 89:110–117.

Kalogeropoulos, K. 2007. Likelihood-based inference for a class of multivariate diffusions with unobserved paths. Journal of Statistical Planning and Inference 137:3092–3102.

Khabbazian, M., R. Kriebel, K. Rohe, and C. Ané. 2016. Fast and accurate detection of evolutionary shifts in Ornstein-Uhlenbeck models. Methods in Ecology and Evolution 7:811–824.

Klebaner, F. C. 2012. Introduction to stochastic calculus with applications. 3rd ed. Imperial College Press, London.

Klingenberg, C. P. 2011. MorphoJ: an integrated software package for geometric morphometrics. Molecular Ecology Resources 11:353–357.

Klingenberg, C. P., and J. Marugán-Lobón. 2013. Evolutionary covariation in geometric morphometric data: analyzing integration, modularity, and allometry in a phylogenetic context. Systematic Biology 62:591–610.

Kloeden, P., and E. Platen. 2011. Numerical solution of stochastic differential equations. Stochastic Modelling and Applied Probability. Springer, Berlin.

Kristensen, N. R., H. Madsen, and S. B. Jørgensen. 2004. Parameter estimation in stochastic grey-box models. Automatica 40:225–237.

Kutsukake, N., and H. Innan. 2012. Simulation-based likelihood approach for evolutionary models of phenotypic traits on phylogeny. Evolution 67:355–367.

Lamperti, J. 1962. Semi-stable stochastic processes. Transactions of the American Mathematical Society 104:62–78.

Lande, R. 1976. Natural selection and random genetic drift in phenotypic evolution. Evolution 30:314–334.

Landis, M. J., and J. G. Schraiber. 2017. Pulsed evolution shaped modern vertebrate body sizes. Proceedings of the National Academy of Sciences of the USA 114:13224–13229.

Landis, M. J., J. G. Schraiber, and M. Liang. 2012. Phylogenetic analysis using Lévy processes: finding jumps in the evolution of continuous traits. Systematic Biology 62:193–204.

Lebowitz, J. L., and O. Penrose. 1973. Modern ergodic theory. Physics Today 26:23–29.

Lill, A. 2004. Superb lyrebird (*Menura novaehollandiae*). Pages 484–495 *in* J. del Hoyo, A. Elliott, J. Sargatal, D. A. Christie, and E. de Juana, eds. Handbook of the birds of the world. Vol. 9. Lynx Edicions, Barcelona.

Lynch, M. 1991. Methods for the analysis of comparative data in evolutionary biology. Evolution 45:1065–1080.

Maddison, W. P., P. E. Midford, and S. P. Otto. 2007. Estimating a binary character's effect on speciation and extinction. Systematic Biology 56:701–710.

Malaspinas, A.-S., O. Malaspinas, S. N. Evans, and M. Slatkin. 2012. Estimating allele age and selection coefficient from time-serial data. Genetics 192:599–607.

Martins, E. P., and T. F. Hansen. 1997. Phylogenies and the comparative method: a general approach to incorporating phylogenetic information into the analysis of interspecific data. American Naturalist 149:646–667.

Mayr, E., and W. B. Provine. 1998. The evolutionary synthesis: perspectives on the unification of biology. Harvard University Press, Cambridge, MA.

McCullagh, P., and J. Nelder. 1989. Generalized linear models. 2nd ed. Chapman & Hall/CRC Monographs on Statistics and Applied Probability. CRC, Boca Raton, FL.

McGhee, G. R. 2015. Limits in the evolution of biological form: a theoretical morphologic perspective. Interface Focus 5(6).

Mikosch, T. 1998. Elementary stochastic calculus with finance in view. Advanced Series on Statistical Science and Applied Probability. Vol. 6. World Scientific, Hackensack, NJ.

Milstein, G. N. 1978. A method of second-order accuracy integration of stochastic differential equations. Theory of Probability and Its Applications 23:396–401.

Møller, J. K., and H. Madsen. 2010. From state dependent diffusion to constant diffusion in stochastic differential equations by the Lamperti transform. IMM Technical Report no. 2010-16. DTU Informatics.

Moran, P. A. P. 1958. Random processes in genetics. Mathematical Proceedings of the Cambridge Philosophical Society 54:60–71.

Mörters, P., and Y. Peres. 2010. Brownian motion. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge, MA.

Øksendal, B. 2007. Stochastic differential equations: an introduction with applications. 6th ed. Springer, Berlin.

O'Reilly, J. E., M. dos Reis, and P. C. Donoghue. 2015. Dating tips for divergence-time estimation. Trends in Genetics 31:637–650.

Pagel, M. 1999. Inferring the historical patterns of biological evolution. Nature 401:877–884.

Papaspiliopoulos, O., G. O. Roberts, and O. Stramer. 2013. Data augmentation for diffusions. Journal of Computational and Graphical Statistics 22:665–688.

Patterson, C. 1981. Significance of fossils in determining evolutionary relationships. Annual Review of Ecology and Systematics 12:195–223.

Pennell, M. W., R. G. FitzJohn, W. K. Cornwell, and L. J. Harmon. 2015. Model adequacy and the macroevolution of angiosperm functional traits. American Naturalist 186:E33–E50.

Pigliucci, M. 2007. Finding the way in phenotypic space: the origin and maintenance of constraints on organismal form. Annals of Botany 100:433–438.

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

Posada, D., and K. A. Crandal. 2001. Selecting the best-fit model of nucleotide substitution. Systematic Biology 50:580–601.

R Core Team. 2018. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna.

Rabosky, D. L., S. C. Donnellan, M. Grundler, and I. J. Lovette. 2014. Analysis and visualization of complex macroevolutionary dynamics: an example from Australian scincid lizards. Systematic Biology 63:610–627.

Rabosky, D. L., and E. E. Goldberg. 2015. Model inadequacy and mistaken inferences of trait-dependent speciation. Systematic Biology 64:340–355.

Rabosky, D. L., F. Santini, J. Eastman, S. A. Smith, B. Sidlauskas, J. Chang, and M. E. Alfaro. 2013. Rates of speciation and morpho-logical evolution are correlated across the largest vertebrate radiation. Nature Communications 4:1958.

Rensch, B. 1959. Evolution above the species level. Methuen, London.

Risken, H. 1996. The Fokker-Planck equation. Springer Series in Synergetics. Vol. 18. 2nd ed. Springer, Berlin.

Roberts, G. O., and O. Stramer. 2001. On inference for partially observed nonlinear diffusion models using the Metropolis-Hastings algorithm. Biometrika 88:603–621.

Ronquist, F., S. Klopfstein, L. Vilhelmsen, S. Schulmeister, D. L. Murray, and A. P. Rasnitsyn. 2012. A total-evidence approach to dating with fossils, applied to the early radiation of the Hymenoptera. Systematic Biology 61:973–999.

Schluter, D. 1996. Adaptive radiation along genetic lines of least resistance. Evolution 50:1766–1774.

Schraiber, J. G., S. N. Evans, and M. Slatkin. 2016. Bayesian inference of natural selection from allele frequency time series. Genetics 203:493–511.

Serrelli, E., and N. Gontier. 2015. Macroevolution: explanation, interpretation and evidence. Springer, Berlin.

Shi, J. J., and D. L. Rabosky. 2015. Speciation dynamics during the global radiation of extant bats. Evolution 69:1528–1545.

Sidlauskas, B. 2008. Continuous and arrested morphological diversification in sister clades of characiform fishes: a phylomorphospace approach. Evolution 62:3135–3156.

Simpson, G. G. 1953. The major features of evolution. Columbia University Press, New York.

Slater, G. J. 2014. Correction to "phylogenetic evidence for a shift in the mode of mammalian body size evolution at the Cretaceous-Palaeogene boundary", and a note on fitting macroevolutionary models to comparative paleontological data sets. Methods in Ecology and Evolution 5:714–718.

Slater, G. J., L. J. Harmon, and M. E. Alfaro. 2012. Integrating fossils with molecular phylogenies improves inference of trait evolution. Evolution 66:3931–3944.

Sørensen, H. 2004. Parametric inference for diffusion processes observed at discrete points in time: a survey. International Statistical Review 72:337–354.

Stadler, T. 2010. Sampling-through-time in birth-death trees. Journal of Theoretical Biology 267:396–404.

Stan Development Team. 2017. Prior choice recommendations. https://github.com/stan-dev/stan/wiki/prior-choice-recommendations.

Stanley, S. M. 1975. A theory of evolution above the species level. Proceedings of the National Academy of Sciences of the USA 72:646–650.

Tanner, M., and W. Wong. 1987. The calculation of posterior distributions by data augmentation (with discussion). Journal of the American Statistical Association 82:528–550.

Tidière, M., J.-M. Gaillard, V. Berger, D. W. Müller, L. B. Lackey, O. Gimenez, M. Clauss, and J.-F. Lemaître. 2016. Comparative analyses of longevity and senescence reveal variable survival benefits of living in zoos across mammals. Scientific Reports 6:36361.

Turelli, M. 1977. Random environments and stochastic calculus. Theoretical Population Biology 12:140–178.

Uhlenbeck, G. E., and L. S. Ornstein. 1930. On the theory of Brownian motion. Physical Review 36:823–841.

Uyeda, J. C., D. S. Caetano, and M. W. Pennell. 2015. Comparative analysis of principal components can be misleading. Systematic Biology 64:677–689.

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 of the USA 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.

Vermeij, G. J. 2015. Forbidden phenotypes and the limits of evolution. Interface Focus 5(6).

Warton, D. I., M. Lyons, J. Stoklosa, and A. R. Ives. 2016. Three points to consider when choosing a LM or GLM test for count data. Methods in Ecology and Evolution 7:882–890.

Wiener, N. 1923. Differential space. Journal of Mathematical Physics 2:131–174.

Wright, S. 1931. Evolution in Mendelian populations. Genetics 16:97–159.

———. 1938. The distribution of gene frequencies under irreversible mutation. Proceedings of the National Academy of Sciences of the USA 24:253–259.

Zhang, C., T. Stadler, S. Klopfstein, T. A. Heath, and F. Ronquist. 2016. Total-evidence dating under the fossilized birth-death process. Systematic Biology 65:228–249.

Zheng, L., A. R. Ives, T. Garland, B. R. Larget, Y. Yu, and K. Cao. 2009. New multivariate tests for phylogenetic signal and trait correlations applied to ecophysiological phenotypes of nine *Manglietia* species. Functional Ecology 23:1059–1069.

Associate Editor: Travis Ingram
Editor: Daniel I. Bolnick
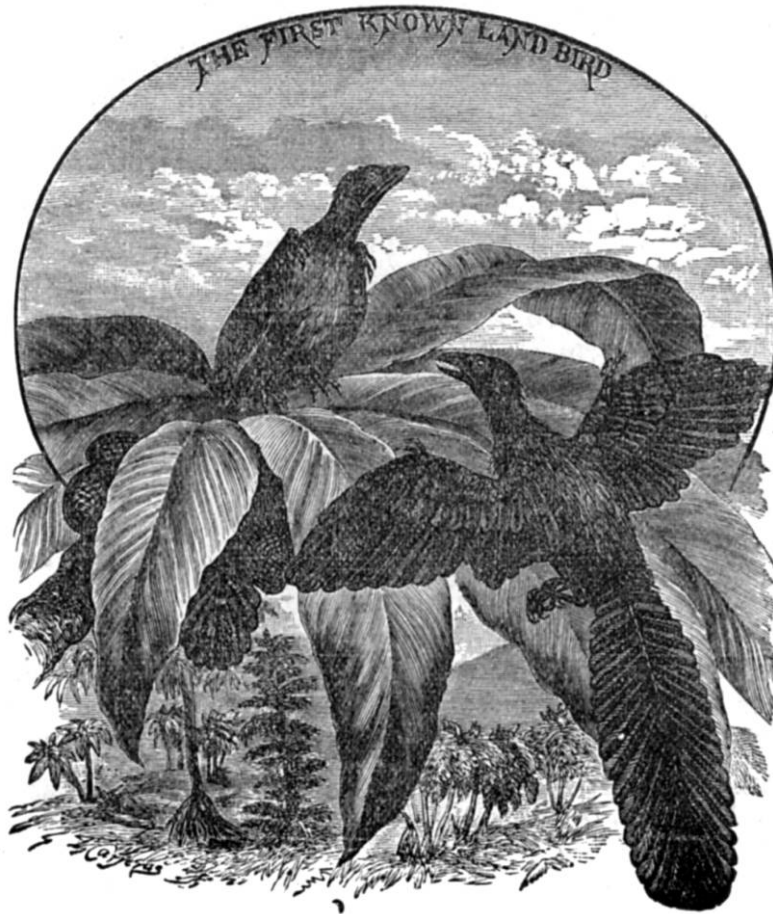
"The story is most skillfully told; it will charm the grown-up naturalist, and, as we have reason to know, interest an intelligent lad. The wood-cuts, some of which we reproduce, are by Mr. Carreras, and are geological restorations of unusual excellence." From the review of Buckley's *Winners in Life's Race* (*The American Naturalist*, 1884, 18:47–50).