Phylogenetic Analysis Using Lévy Processes: Finding Jumps in the Evolution of Continuous Traits

MICHAEL J. LANDIS, JOSHUA G. SCHRAIBER*, AND MASON LIANG

Department of Integrative Biology, University of California, Berkeley, CA 94720-3140, USA;
*Correspondence to be sent to: Department of Integrative Biology, University of California, 3060 VLSB #3140, Berkeley, CA 94720-3140, USA;
E-mail: jgschraiber@berkeley.edu.
Michael J. Landis and Joshua G. Schraiber contributed equally to this article.

Abstract.—Gaussian processes, a class of stochastic processes including Brownian motion and the Ornstein–Uhlenbeck process, are widely used to model continuous trait evolution in statistical phylogenetics. Under such processes, observations at the tips of a phylogenetic tree have a multivariate Gaussian distribution, which may lead to suboptimal model specification under certain evolutionary conditions, as supposed in models of punctuated equilibrium or adaptive radiation. To consider non-normally distributed continuous trait evolution, we introduce a method to compute posterior probabilities when modeling continuous trait evolution as a Lévy process. Through data simulation and model testing, we establish that single-rate Brownian motion (BM) and Lévy processes with jumps generate distinct patterns in comparative data. We then analyzed body mass and endocranial volume measurements for 126 primates. We rejected single-rate BM in favor of a Lévy process with jumps for each trait, with the lineage leading to most recent common ancestor of great apes showing particularly strong evidence against single-rate BM. [Continuous traits; saltational evolution; Lévy processes; Bayesian inference.]

Morphological variation in continuous characters, such as the body mass of theropod or the height of kelp, is one of the most visible examples of the diversity of life on Earth. A number of theoretical frameworks have been put forth to explain this variety of sizes and shapes seen in the natural world (Darwin 1859, Simpson 1953, Eldredge and Gould 1972, Stanley 1975). Gaussian processes—a class of stochastic processes that includes Brownian motion (BM) and the Ornstein-Uhlenbeck process—have been used extensively to model continuous trait evolution, for example, body mass evolution (Freckleton et al. 2003) or gene expression level evolution (Brawand et al. 2011). These processes are a natural model for continuous character evolution because they are the continuum limit of a broad range of discrete-time character evolution models (Cavalli-Sforza and Edwards 1967; Lande 1976; Felsenstein 1985).

However, not all discrete-time models have a Gaussian process as their limit; many evolutionary processes may result in changes in a continuous character too abrupt to be accounted for by any Gaussian process. For example, rapid changes in population size can dramatically affect rates of allele fixation, and thus introduce abrupt changes in quantitative traits (Lande 1976). The ecological release of selective constraints may induce an adaptive radiation that increases disparity unevenly across a clade (Simpson 1953; Stanley 1975). Through cladogenesis under a punctuated equilibrium model of trait evolution, divergence events are paired with sudden trait change (Eldredge and Gould 1972). If cladogenetic evolutionary processes are present, continuous trait patterns seen in extant taxa may mislead inference due to speciation events "hidden" by extinction events (Bokma 2002).

Two main routes have been taken to account for the extra variation that these micro- and macro-evolutionary processes produce. One approach pioneered by O'Meara et al. (2006) is to allow for shifts in the rate of BM in different places on the phylogeny. This method is similar in spirit to models of rate shifts in molecular evolution (Thorne et al. 1998; Huelsenbeck et al. 2000; Drummond and Suchard 2010). A number of refinements have since been proposed, such as the use of reversible jump Markov chain Monte Carlo (MCMC) to infer the timing and intensity of rate shifts (Eastman et al. 2011), which identified rate shifts in the evolution of primate body mass. Harmon et al. (2010) introduced an "early-burst" process to model rapid trait evolution following cladogenesis in which the rate of BM decreases exponentially along a branch, such that the rate of change is fastest immediately when a new lineage diverges and then decreases as the lineage grows older. For size and shape data across 49 clades of animals, they reported that their early-burst model was favored in two data sets over BM and Ornstein–Uhlenbeck processes. Although these models relax the time-homogeneity assumption of Gaussian process models, they remain fundamentally gradual, in the sense that the changes in traits cannot be too large in a short period of time. This results in the existence of intermediate forms, the hallmark of gradualism.

The other route explicitly models non-gradual evolution by augmenting BM with a process of "jumps." In a seminal work on models of continuous trait evolution, Hansen and Martins (1996) compared

the covariance structure of models of punctuated equilibrium with other models of phenotypic trait evolution and found that one could not distinguish between punctuational models and BM models from covariance alone. Bokma (2008) described a method to identify punctuated evolution by modeling continuous trait evolution as the sum of BM and normally distributed jumps resulting from speciation events. The Bokma model accounts for hidden speciation events by first estimating the speciation and extinction rates, then conditioning on the rates as part of a Bayesian MCMC analysis. In a study on mammalian body mass evolution, this model inferred that cladogenetic, rather than anagenetic, processes produced the majority of trait diversity we see today (Mattila and Bokma 2008).

However, jumps in trait evolution may not be linked directly to cladogenesis. Using a pure-jump model, Uyeda et al. (2011) identified a once-per-million-year jump periodicity in vertebrate evolution by modeling trait evolution as the sum of white noise and normally distributed jumps drawn at the times of a Poisson process. Such pure-jump models may be appropriate for traits that are thought to have weak or no gradual evolution component, such as gene expression, which may depend only on the discrete events of transcription factor binding site recruitment and degradation. Khaitovich et al. (2005) introduced a pure-jump model of evolution in which gene expression levels evolve through jumps drawn from a skewed normal distribution at the times of a Poisson process. They reported evidence of skewness in primate gene expression evolution, a biologically interesting signal that could not have been explained by simple BM models (also see Chaix et al. 2008).

This evidence of jumps motivates us to introduce a class of models to account for the wide range of modes of non-gradual evolution. Both BM and the compound Poisson processes of Khaitovich et al. (2005) and Uyeda et al. (2011) (but not the Ornstein-Uhlenbeck process) are members of a broader class of stochastic processes whose motion may be thought of as "drift and diffusion with jumps," namely the class of Lévy processes. A Lévy process is the sum of 3 components: a directional drift (also called trend in the biology literature, not to be confused with genetic drift), a BM, and a pure-jump process. The last component allows Lévy processes to have jumps in their sample paths and, in the context of continuous trait evolution, account for abrupt shifts in continuous characters that pure diffusion models cannot easily explain. Qualitatively, these jumps give the distribution of trait change "fat tails," reflecting that there is a higher probability of larger amounts of trait change than under a BM. In the mathematical finance literature, Lévy processes have been successfully used to capture the "fat-tailed" behavior of stock prices (Li et al. 2008). We developed a Bayesian method that determines whether a Lévy process with jumps explains the data better than a single-rate BM and effectively infers the parameters of that Lévy process.

Model

Lévy Processes

Stochastic processes with stationary and independent increments whose sample paths are right continuous with left limits are called Lévy processes. We will highlight the key properties of this class of processes and state some important results. Kallenberg (2010, Chapter 15) provides a more detailed and technical exposition.

Let $\{X_t, t > 0\}$ be a Lévy process. There are 2 equivalent ways of characterizing X_t, by its transition density $\mathbb{P}(X_t = y | X_0 = x)$, or by its characteristic function, given by

$$\phi(k; t) = \mathbb{E}\left(e^{ikX_t} | X_0 = 0\right), \qquad (1)$$

where $i = \sqrt{-1}$ is the imaginary unit and $\mathbb{E}(\cdot)$ is the expected value. Note that k is the variable on which the characteristic function acts. As an example, the transition density of a BM is

$$\mathbb{P}(X_t = y | X_0 = x) = \frac{1}{\sqrt{2\pi\sigma^2 t}} e^{-\frac{(y-x)^2}{2\sigma^2 t}}, \qquad (2)$$

so the corresponding characteristic function is given by

$$\begin{aligned} \phi(k; t) &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2 t}} e^{-\frac{y^2}{2\sigma^2 t}} e^{iky} dy \\ &= e^{-t\frac{1}{2}\sigma^2 k}. \end{aligned} \qquad (3)$$

A result known as the Lévy–Khinchine representation asserts that all Lévy processes have characteristic functions of the form

$$\phi(k; t) = \exp\left\{ t\left(aik - \frac{1}{2}\sigma^2 k^2 + \int \left(e^{ikj} - 1 - ikj\mathbb{I}_{|j|<1} \right) \nu(dj) \right) \right\}, \qquad (4)$$

where a and σ^2 are constants and $\nu(\cdot)$ is the so-called Lévy measure. Intuitively, the Lévy–Khinchine representation provides a mathematical decomposition of a Lévy process into its 3 constituent parts:

1. A constant directional drift (or trend) with rate a.

2. A BM with rate σ^2.

3. A pure-jump process that draws jumps from the Lévy measure $\nu(\cdot)$.

The processes we consider have no long-term directional trend, so $a = 0$. To get a better understanding of the Lévy measure, one can imagine that the process has probability $\nu(dj)dt$ of making a jump of size j during the time dt. If there are no jumps, then ν is identically 0 and Equation (4) becomes Equation (3). This shows that the only Lévy process with continuous sample paths is a single-rate BM.

Using the Lévy–Khinchine formula, it is possible to compute the moments of a Lévy process, assuming that they exist. Because we will only consider symmetric Lévy processes, we are only interested in the process' variance and excess kurtosis, the latter of which is a measure of the relative frequency of large evolutionary changes compared with a BM. These 2 moments are given by

$$V(t) = \mathbb{E}(X_t^2) = -\phi^{(2)}(0; t) \qquad (5)$$
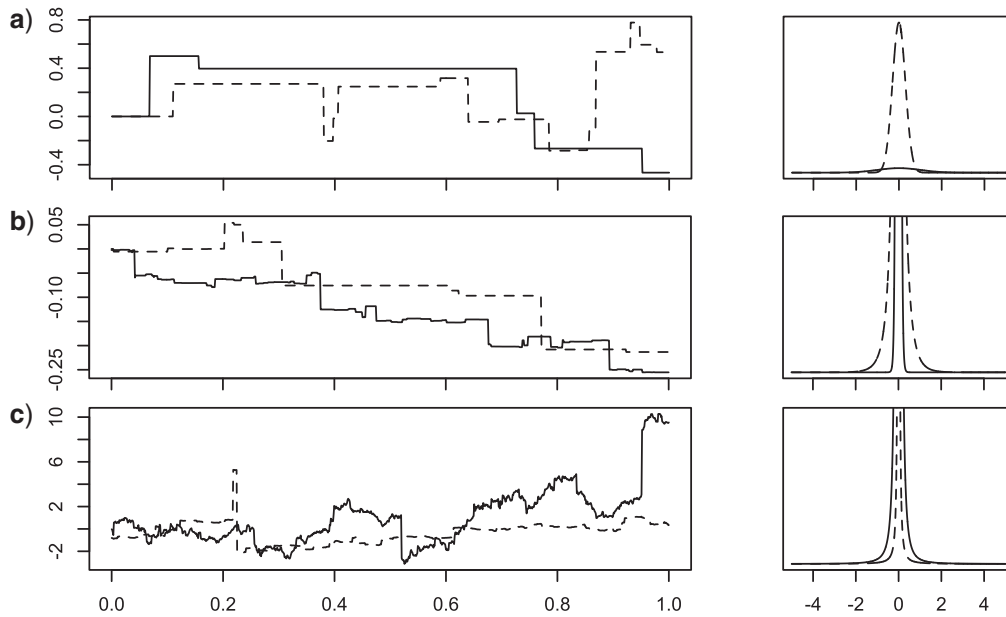
FIGURE 1. Sample paths of Lévy processes without BM (left panel) and their corresponding Lévy measures (right panel). Compound Poisson process with normally distributed jump JN (a) paths were sampled with parameters $\lambda = 2$, and $\delta = 1$ (solid line) and $\lambda = 20$, and $\delta = 0.3$ (dashed line). VG (b) paths were sampled with parameters $\kappa = 0.1$, and $\tau = 0.2$ (solid line) and $\kappa = 1$, and $\tau = 0.6$ (dashed line). AS (c) paths were sampled with parameters $\alpha = 1.5$, and $\beta = 0.1$ (solid line) and $\alpha = 0.9$, and $\beta = 0.005$ (dashed line).

and

$$K(t) = \frac{\mathbb{E}(X_t^4)}{V(t)^2} - 3 = \frac{\phi^{(4)}(0; t)}{V(t)^2} - 3, \qquad (6)$$

where $\phi^{(n)}$ is the nth derivative of ϕ.

Three Examples of Lévy Processes

In addition to a single-rate BM, we implemented 3 different models that are representative of the range of behavior possible with Lévy processes. These models are a compound Poisson process with normally distributed jumps (jump normal, abbreviated JN), the variance gamma (VG) process, and the α-stable (AS) process. To gain an intuition for the behavior of each process, Figure 1 shows representative pure-jump sample paths and the corresponding jump measures for JN, VG, and AS, whose properties we examine in turn. Note, to accentuate the flavor of each jump measure under each parameterization in Figure 1, the BM rate was assigned to $\sigma = 0$. Parameters of each model are summarized in Table 1.

The compound Poisson process—The JN model has Lévy measure

$$\nu(dj) = \lambda \frac{1}{\sqrt{2\pi\delta^2}} e^{-\frac{j^2}{2\delta^2}} dj.$$

With rate λ, the process makes jumps with values drawn from a centered normal distribution with standard deviation δ. As Figure 1a shows, the paths of the JN process are characterized by periods of stasis interrupted by bursts of rapid change. Looking at the Lévy measure, a process with more jumps will have a taller Lévy measure while a process with larger jumps will have a fatter Lévy measure. The transition density of the JN process with no BM is

TABLE 1. Model parameters and interpretations for all implemented models

Model	Parameter	Interpretation
Brownian motion (BM)	σ	Rate of BM
Jump normal (JN)	σ	Rate of BM
	λ	Rate of jumps
	δ	Standard deviation of jump size
Variance gamma (VG)	σ	Rate of BM
	κ	Relative rate of large jumps
	τ	Size of jumps
α-stable (AS)	σ	Rate of BM
	α	Relative rate of small jumps
	β	Size of jumps

known and is given by

$$\mathbb{P}(J_t = j | J_0 = 0, \delta, \lambda) = \sum_{n=0}^{\infty} \frac{(\lambda t)^n}{n!} e^{-\lambda t} \frac{1}{\sqrt{2\pi n \delta^2}} e^{-\frac{j^2}{2n\delta^2}}. \qquad (7)$$

The variance and excess kurtosis of a process with both BM, with rate σ^2, and JN motion are

$$V(t) = (\sigma^2 + \lambda \delta^2)t \qquad (8)$$

and

$$K(t) = \frac{3\lambda \delta^4}{(\sigma^2 + \lambda \delta^2)^2 t}, \qquad (9)$$

respectively.

The VG process—The VG model has Lévy measure

$$\nu(dj) = \frac{1}{\kappa |j|} e^{-\sqrt{\frac{2}{\kappa \tau^2}} |j|} dj.$$

Here, τ controls the size of jumps while κ controls the relative probability of large versus small jumps. The Lévy measure has infinite mass, and thus the VG process is infinitely active, meaning that in any finite period of time, the process makes infinitely many jumps. However, as can be seen in Figure 1b most of those jumps are arbitrarily small. When κ is large, the VG process only makes very large or very small jumps.

Like the JN process, the transition density of the VG process with no BM is known analytically,

$$\mathbb{P}(J_t = j | J_0 = 0, \tau, \kappa) = \frac{2^{\frac{2t-3\kappa}{4\kappa}} \kappa^{-\frac{2t+\kappa}{4\kappa}}}{\Gamma(t/\kappa)\sqrt{\pi\tau^2}} \left(\frac{\tau^2}{j^2}\right)^{\frac{-2t+\kappa}{4\kappa}}$$
$$\times K_{|t/\kappa - 1/2|}\left(\sqrt{\frac{2j^2}{\kappa\tau^2}}\right), \quad (10)$$

where $\Gamma(\cdot)$ is the gamma function and $K_\epsilon(\cdot)$ is the modified Bessel function of the second kind with index ϵ (Abramowitz and Stegun 1964; Chapters 9 and 10).

The variance and excess kurtosis of a process with both BM and VG motion are

$$V(t) = (\sigma^2 + \tau^2)t \quad (11)$$

and

$$K(t) = \frac{3\kappa\tau^4}{(\sigma^2 + \tau^2)^2 t}, \quad (12)$$

respectively.

The AS process—The AS model has Lévy measure

$$\nu(dj) = \frac{\beta^\alpha}{|j|^{1+\alpha}} dj,$$

where β is a scale parameter, controlling the magnitude of jumps taken and $0 \le \alpha \le 2$ is the so-called stability parameter. For every $\alpha < 2$, the Lévy measure has infinite mass, so the AS process is infinitely active. However, Figure 1c shows that the behavior of the AS process is quite different from the VG process. In particular, the AS process does not experience as strong a trade-off between small and large jumps as the VG process does. As $\alpha \to 0$, the tails of the Lévy measure become heavier and heavier, but the relative proportion of probability for medium-sized jumps remains nearly constant, as opposed to the VG process. This is manifested in the fact that the AS process has infinite pth moment for $p > \alpha$ when $\alpha < 2$; thus, the variance and the excess kurtosis of the process do not exist for $\alpha < 2$. In addition, unlike the JN and VG processes, the transition density is not known in closed form. However, the characteristic function of the AS process without BM is known to be

$$\phi(k; t) = e^{t|\beta k|^\alpha}, \quad (13)$$

and so we can make use of the Fourier inversion theorem to numerically compute the transition density of the AS process without BM,

$$\mathbb{P}(J_t = j | J_0 = 0, \beta, \alpha) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-ikj}\phi(k; t)dk$$
$$= \frac{1}{2\pi} \int_{-\infty}^{\infty} \cos(kj)\phi(k; t)dk, \quad (14)$$

where the second equality follows because $\phi(k; t)$ is real and even.

METHODS

Inference of Lévy Processes

We use a Bayesian framework to analyze Lévy processes evolving on a phylogeny. Let $p(\theta)$ be the prior density for the parameters of the Lévy process model and $L(D|\theta)$ be the likelihood of the observed data given the parameters. We want to compute the posterior density,

$$p(\theta|D) \propto L(D|\theta)p(\theta). \quad (15)$$

To compute the likelihood of a Lévy process on a phylogeny, we use Felsenstein's pruning algorithm (Felsenstein 1981). To calculate $L_i(y_i)$, the likelihood of the data observed in all species that are descended from node i, given that the trait value at node i equals y_i, we use the likelihood at the descendent nodes j and k. Letting $\{X_t, t > 0\}$ be the Lévy process under consideration,

$$L_i(y_i|\theta) = \left(\int \mathbb{P}(X_{t_j} = y_j | X_0 = y_i)L_j(y_j|\theta)dy_j\right)$$
$$\times \left(\int \mathbb{P}(X_{t_k} = y_k | X_0 = y_i)L_k(y_k|\theta)dy_k\right), \quad (16)$$

where t_j and t_k are the branch lengths leading to nodes j and k, respectively. At the root (node 0), we assume an improper uniform prior for the trait value y_0, and we integrate over all possible values of the root node to obtain

$$L(D|\theta) = \int_{-\infty}^{\infty} L_0(y_0|\theta)dy_0. \quad (17)$$

However, the integrals in Equations (16) and (17) are intractable for most Lévy processes. To get around this, we exploit the fact that if X is a Lévy process consisting of a BM with no directional drift and diffusion rate σ^2, and a pure-jump process, the Lèvy–Khinchine representation guarantees that $X = B + J$, where B is a BM and J is the pure-jump process, and B and J are independent. Then conditional on $J = j$, the transition density of X is given by

$$\mathbb{P}(X_t = y | X_0 = x, J = j) = \frac{1}{\sqrt{2\pi\sigma^2 t}} e^{-\frac{((y-j)-x)^2}{2\sigma^2 t}}. \quad (18)$$

This follows because the BM has to get to $y - j$ and then the jump process will do the rest. Thus, conditioned on all the jumps on the branch leading up to a node, $\mathbf{J} = \{J^{(n)}, \ldots, J^{(1)}\}$ for a tree with n nonroot nodes, $L(D|\theta, \mathbf{J})$ is the likelihood of the data under BM where branch i has branch-specific offset $J^{(i)}$. Then,

$$p(\theta, \mathbf{J}|D) \propto L(D|\theta, \mathbf{J})p(\mathbf{J}|\theta)p(\theta), \quad (19)$$

where

$$p(\mathbf{J}|\theta) = \prod_i \mathbb{P}(J_{t_i}^{(i)} = j^{(i)} | J_0 = 0, \theta)$$

is the joint probability of the jumps along each branch (determined by the specific jump model adopted).

We want to integrate over the jumps to get

$$p(\theta|D) = \int p(\theta, \mathbf{J}|D)d\mathbf{J}, \qquad (20)$$

but this integral remains intractable. Instead, we approximate the integral by using MCMC to obtain samples from the joint posterior distribution of the parameters and the jumps. Marginalizing over the sampled jumps approximates the integral in the right-hand side of Equation (20).

To obtain posterior samples of the jumps, we serially update each branch in a post-order traversal of tree by proposing a new value $J^{(i)\prime}$ from a normal distribution centered at the current sampled $J^{(i)}$ and with variance 0.5. This variance lead to good mixing for the data we considered, but should be specified by the user as appropriate. We then accept or reject the proposed jump update using the Metropolis–Hastings ratio,

$$\mathbb{P}(\text{Accept } J^{(i)\prime}) = \frac{L(D|\theta, \mathbf{J}')}{L(D|\theta, \mathbf{J})}\frac{p(\mathbf{J}'|\theta)}{p(\mathbf{J}|\theta)},$$

where $\mathbf{J}' = \{J^{(n)}, \ldots, J^{(i+1)}, J^{(i)\prime}, J^{(i-1)}, \ldots, J^{(1)}\}$ is the vector of jumps with only one branch updated. Note that the proposal ratio is equal to 1 because of the symmetry of the normal distribution and the prior ratio is equal to 1 because no parameters are updated. This method is similar to the path sampling method of Robinson et al. (2003), in that we use MCMC to sample and integrate over hidden states (the unobserved jumps).

During the MCMC run, we randomly choose to update either the jumps or the model parameters. When we choose to update a model parameter, we randomly choose a model parameter to update. All parameters except for α from the AS process are positive and real and so were assigned scaling proposal distributions. Because $0 < \alpha < 2$, we use a truncated normal proposal distribution to update α. Parameter updates are accepted or rejected according to the Metroplis–Hastings ratio,

$$\mathbb{P}(\text{Accept } \theta') = \frac{L(D|\theta', \mathbf{J})}{L(D|\theta, \mathbf{J})}\frac{p(\mathbf{J}|\theta')}{p(\mathbf{J}|\theta)}\frac{p(\theta')}{p(\theta)}\frac{q(\theta|\theta')}{q(\theta'|\theta)},$$

where θ is the randomly selected parameter, θ′ is the proposed update, and $q(\cdot|\cdot)$ is the proposal distribution.

Data

We log-transformed the male–female means of body mass, endocranial volume (ECV), and mass-to-ECV ratio data reported in Isler et al. (2008). The branches of the phylogeny provided by Isler et al. (2008) were measured in increments of half-million years. In favor of higher resolution of branch lengths, we substituted the Isler et al. phylogeny with the Redding et al. (2010) primate phylogeny included in the R package auteur (Eastman et al. 2011). We intersected the Isler et al. data set with the Redding et al. phylogeny, which resulted in 126 taxa with data present in the phylogeny. This phylogeny has 1267 myr of total branch and a root height of 65 myr. The resulting phylogeny was used for all analyses and simulations reported in this article. The primate data and phylogeny are hosted on the Dryad data repository (doi:10.5061/dryad.0n761).

Software Configuration

The software used in this study was programmed in C++, borrowing code from GNU Scientific Library (G.P. Contributors 2010) and MrBayes (Ronquist et al. 2012). The source code may be found at http://github.com/mlandis/creepy-jerk (Last accessed: 10/29/12). With one exception, all parameters were assigned half-Cauchy distributions with scale parameters of 1 as prior densities. Under the AS processes, $0 \le \alpha \le 2$, so we used a uniform distribution on $[0, 2]$ as its prior. Each posterior distribution was computed by running MCMC for 2×10^6 cycles, sampling every 10^3 cycles, where the first 10^5 cycles were discarded as part of the burn-in. The R package coda (Plummer et al. 2006) was used to verify MCMC convergence. For the BM, JN, VG, and AS models, one MCMC run took 0.5, 8, 6, and 48 h, respectively. This discrepancy results from the fact that, while the JN and VG models have analytical solutions for their jump densities, we had to approximate the AS jump density using time-consuming numerical integration.

Analysis

We characterized how Lévy processes perform in the context of phylogenetic inference for both simulated data and real data. We used simulated data to test the accuracy of parameter inference and quantify the power to reject BM when the true model is a Lévy process with jumps. We then analyzed the primate data set to both estimate parameters and determine whether a BM model is rejected in favor of a Lévy process with jumps in biological data. Our analysis examines the aforementioned 4 Lévy processes: BM, JN, VG, and AS.

To test the BM model, we performed a 3-step procedure similar to a parametric bootstrap. First, data were analyzed under a pure BM model, resulting in an estimate of the BM rate, σ^2. Then, 20 "jump-absent" data sets were simulated under BM with the inferred rate. Finally, each simulated data set was analyzed using a "jump-present" model, and the average posterior distributions of either the variance and excess kurtosis (for JN and VG) or the parameter α (for AS) were compared with the posterior distribution of those parameters inferred from the original data. Note that the variance, excess kurtosis, and α calculated here do *not* describe the data observed at the tips, but rather their expected values as a function of time, see Equations (8), (9), (11), and (12).

By inspecting the posterior distribution of the variance and excess kurtosis for JN and VG between data and the BM simulations, we determined whether there was evidence for non-Gaussian evolution. Under the BM model, the expected excess kurtosis is 0, so if the posterior of the excess kurtosis placed significant mass away from 0, we interpreted that as strong evidence for non-Gaussian evolution. For the AS model, these moments are not defined; however, when $\alpha = 2$ the AS process is equivalent to BM, so if the posterior distribution of α placed significant mass away from 2, we took that as evidence for non-Gaussian evolution.

When we were able to reject BM in favor of a Lévy process with jumps, we characterized the amount of trait change attributable to those jumps by

computing the signal-to-noise ratio, defined as the ratio of the mean to standard deviation, of the posterior distributions of the sampled jumps for each branch in the phylogeny. To normalize for branch-length effects, we further divided all signal-to-noise ratios by their respective branch lengths. When the signal-to-noise ratio equals zero, the BM component of the model alone is capable of producing the observed trait changes along that branch. A non-zero signal-to-noise ratio was interpreted as evidence that traits along the branch evolved faster than could be explained by the model's BM component.

RESULTS

Simulated Data

We simulated 20 data sets for each model (JN, VG, and AS) on the primate phylogeny (see Methods) and computed posterior distributions under the true model for each simulated data set. Figure 2 presents boxplots of the maximum a posteriori estimates for each simulation, with the horizontal line indicating the true parameter value. Inference under the JN and AS models recovered the true parameters with minor error. Inference under the VG model recovered σ and τ reasonably well while
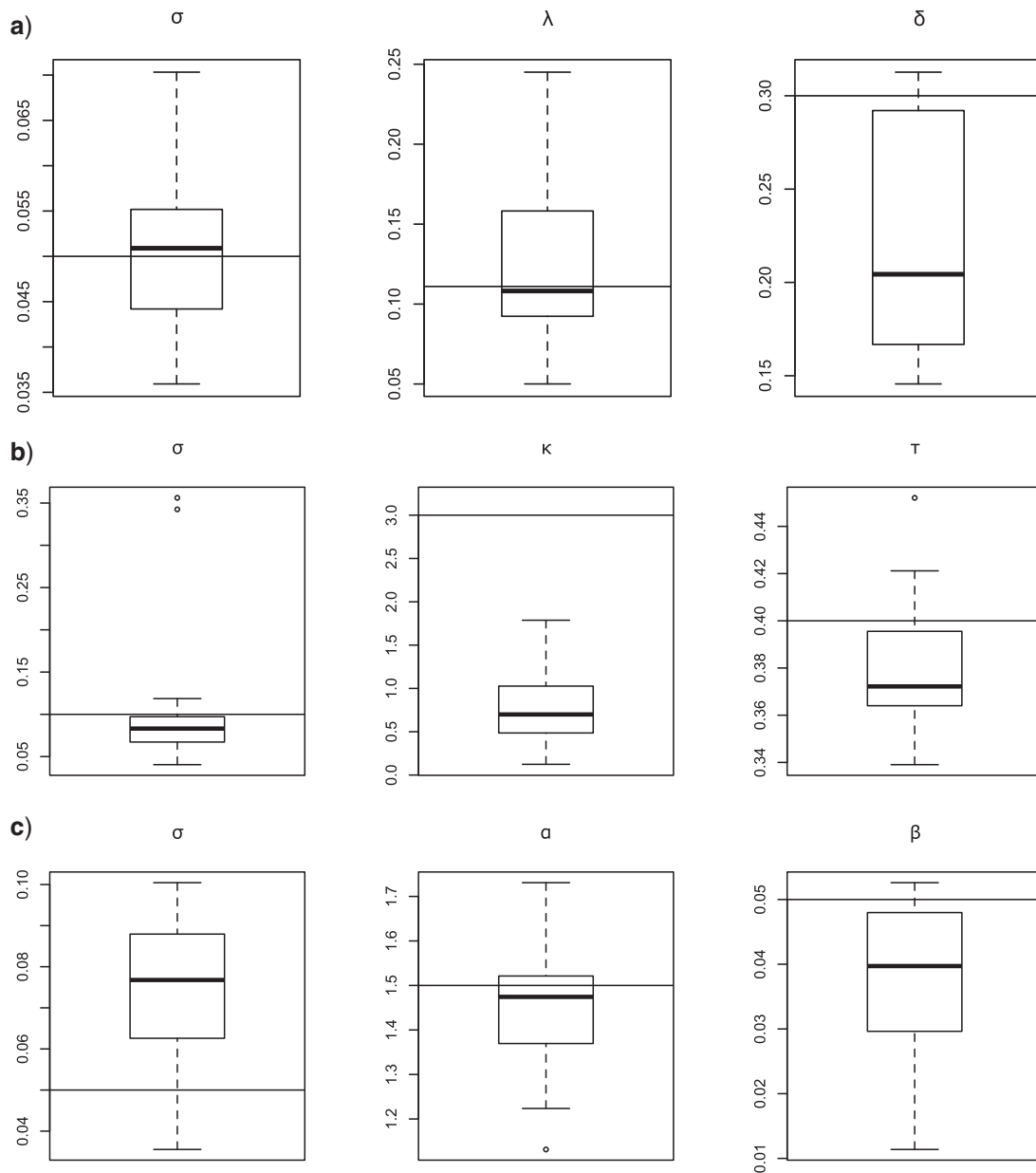

FIGURE 2.   Boxplots of maximum a posteriori model parameter estimates under JN (a), VG (b), and AS (c) for 20 replicates of jump-present data simulated under each model. The horizontal line shows the true parameter value underlying the simulated data. The true parameters for JN are $\sigma = 0.05$, $\lambda = 0.111$, and $\delta = 0.3$. The true parameters for VG are $\sigma = 0.1$, $\kappa = 3$, and $\tau = 0.4$. The true parameters for AS are $\sigma = 0.05$, $\alpha = 1.5$, and $\beta = 0.05$.
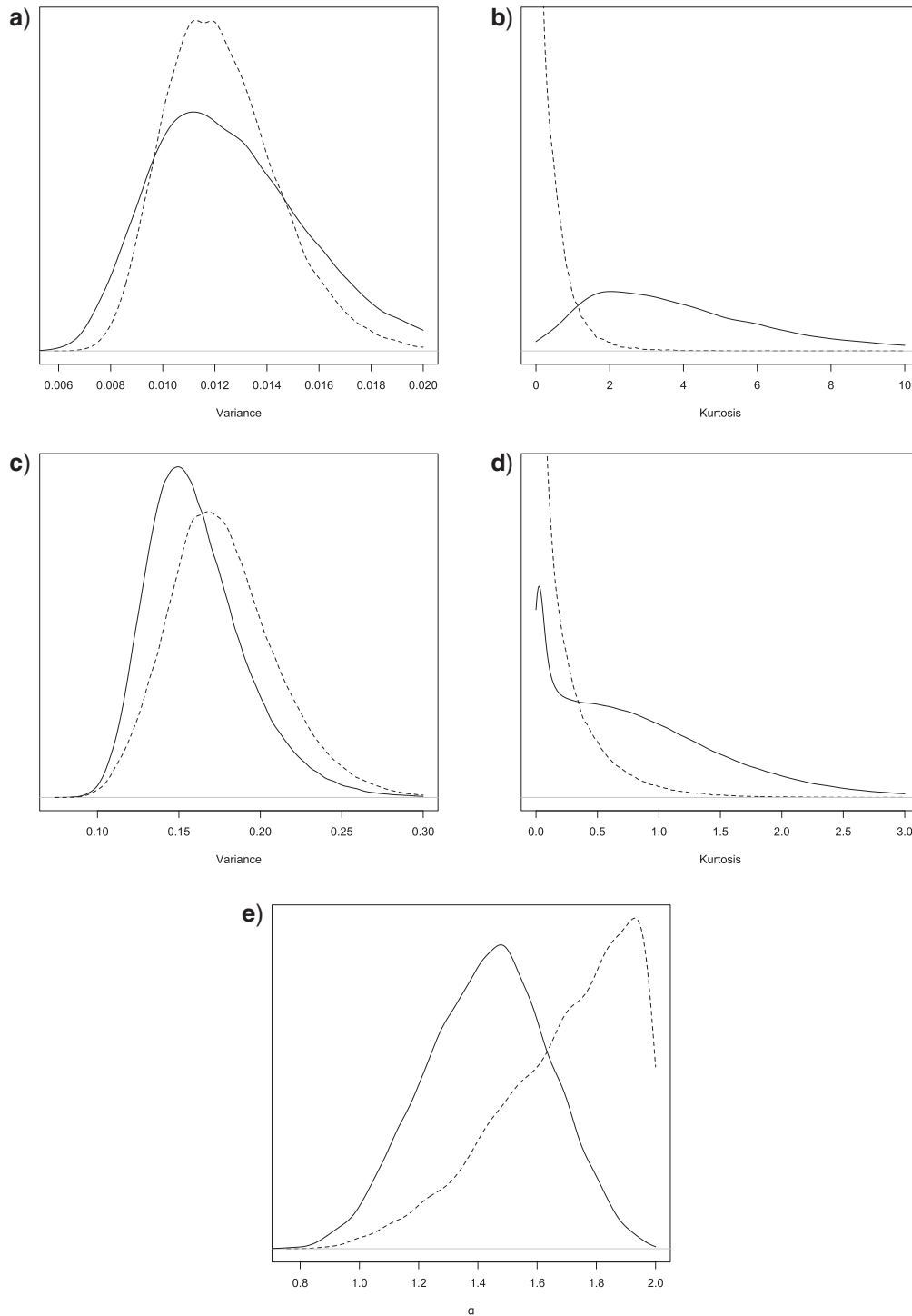
FIGURE 3. Average posteriors of model summary statistics under JN (a), VG (b), and AS (c) upon simulated data. Solid lines indicate average posteriors from 20 replicates of jump-present data simulated under the same model. Dashed lines indicate average posteriors from 20 replicates of jump-absent data simulated under pure BM parameterized with equivalent variance per unit time ($\sigma = 0.1118$, $\sigma = 0.4050$, and $\sigma = 0.2389$ for analysis by JN, VG, and AS, respectively).

underestimating κ by an order of magnitude. The mean and root mean square errors of the posteriors are recorded in Supplementary Table S1.

We then applied our method to test for the presence of jumps to the simulated data sets. The results are shown in Figure 3. When the true model is either JN or VG, the inferred variance was approximately equal between the jump present and BM simulations, but the excess kurtosis was different. For the AS model, the inferred α deviated significantly from 2 only in the
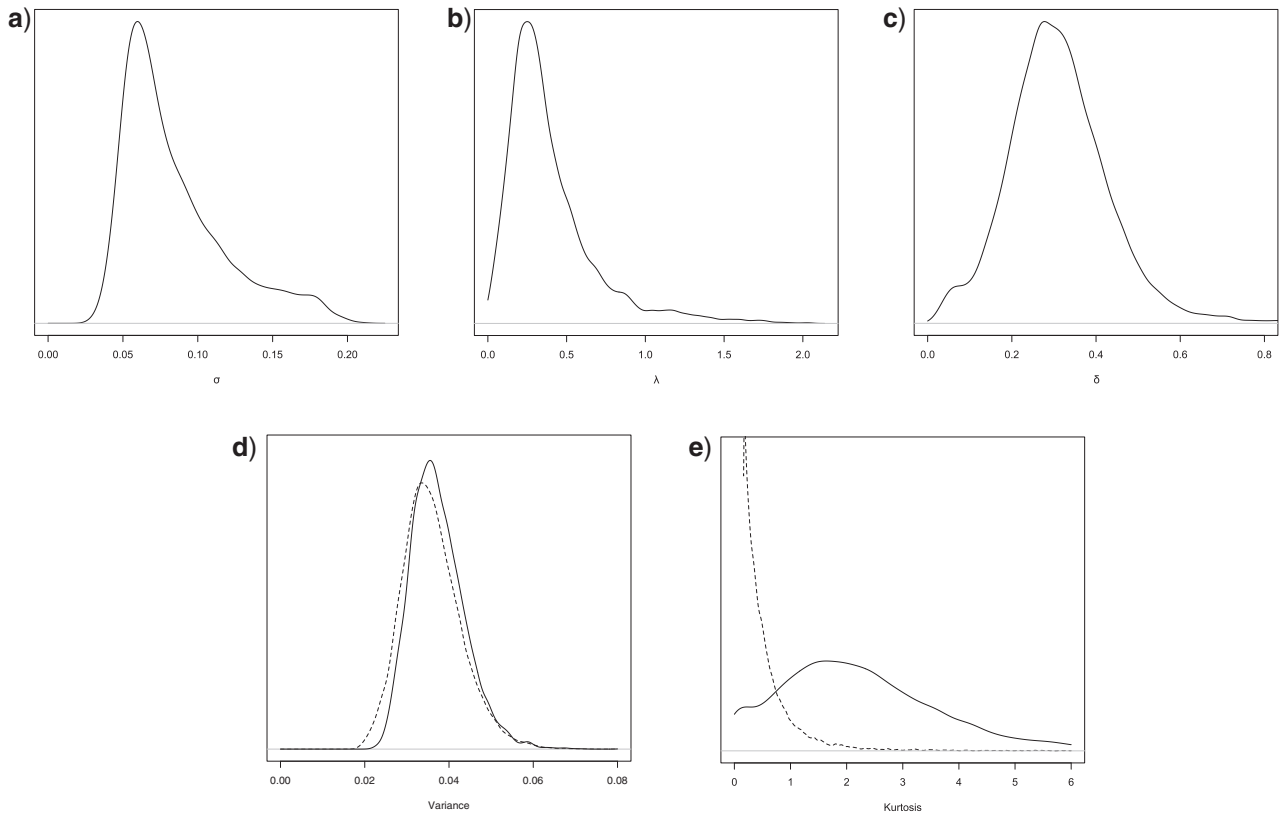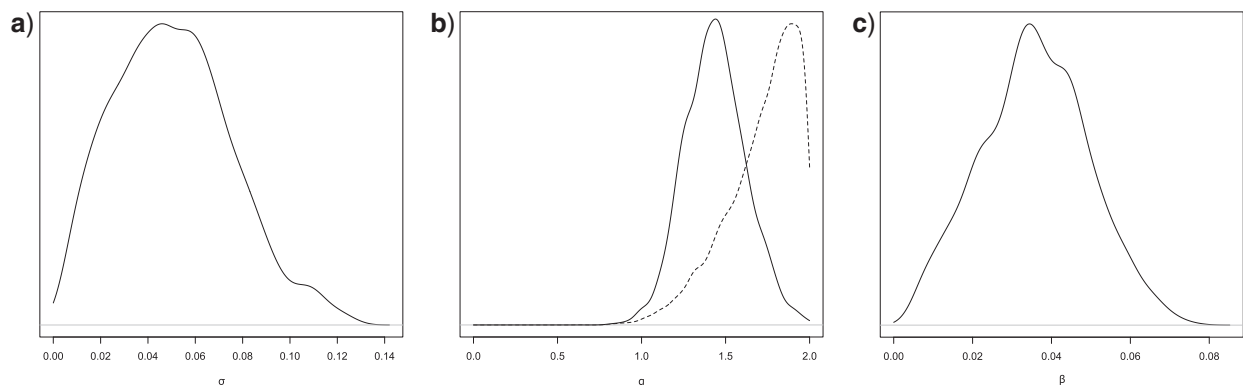
FIGURE 4.   Posteriors of model summary statistics under JN upon primate body mass data. (a–c) the model parameters with maximum a posteriori estimates $\hat{\sigma}=0.0596$, $\hat{\lambda}=0.2497$, and $\hat{\delta}=0.2929$, respectively. (d and e) The model variance and kurtosis per unit time. Solid lines indicate posteriors from the empirical data. Dashed lines indicate average posteriors from 20 replicates of jump-absent data simulated under pure BM parameterized with equivalent variance per unit time ($\sigma=0.18$).

FIGURE 5.   Posteriors of model summary statistics under AS upon primate ECV data. (a–c) The model parameters with maximum a posteriori estimates $\hat{\sigma}=0.1541$, $\hat{\alpha}=1.670$, and $\hat{\beta}=0.0698$, respectively. Solid lines indicate posteriors from the empirical data. Dashed lines indicate average posteriors from 20 replicates of jump-absent data simulated under pure BM parameterized with equivalent variance per unit time ($\sigma=0.12$).

jump-present data. The maximum a posteriori estimates and 95% highest posterior density intervals may be found in Supplementary Table S2.

Empirical Data (Primates)

Next, we computed the posterior distributions for body mass, ECV, and mass-to-ECV ratios for the BM, JN, VG, and AS models. The maximum a posteriori estimates and 95% highest posterior density intervals for each data set are provided in Supplementary Table S3. We applied our test to detect evolution that cannot be explained by BM to each data set. For the sake of brevity, we only present results for body mass under the JN model (Figure 4) and ECV under the AS model (Figure 5), although several models showed evidence of non-Gaussianity in the evolution of these traits. For the mass-to-ECV data,
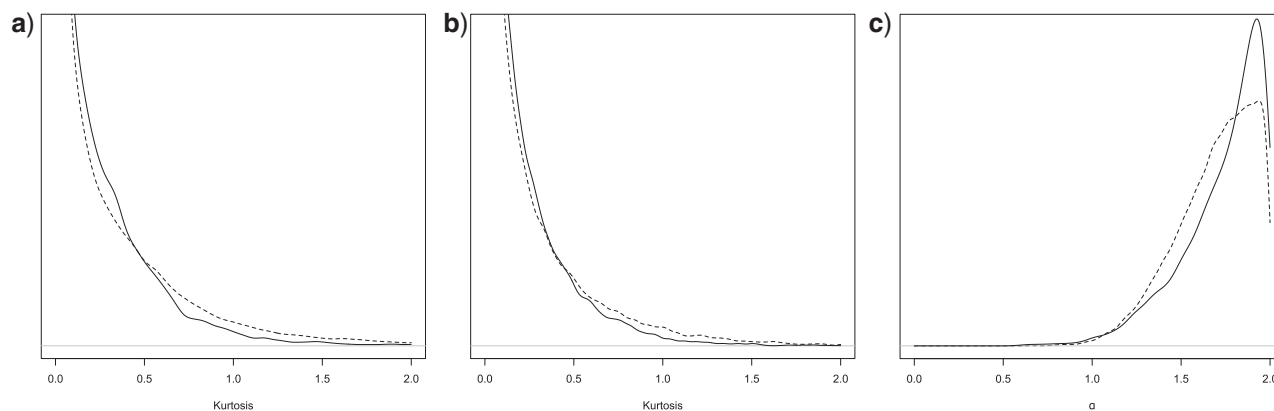
FIGURE 6.   Posteriors of model summary statistics under JN (a), VG (b), and AS (c) upon primate body mass-to-ECV ratio data. Solid lines indicate posteriors from the empirical data. Dashed lines indicate average posteriors from 20 replicates of jump-absent data simulated under pure BM parameterized with equivalent variance per unit time ($\sigma = 0.096$).

no Lévy process with jumps was preferred over BM (Figure 6). Supplementary Table S4 has more detailed numbers, including parameter estimates for each model.

For body mass under the JN model, the estimates of both the jump rate λ and the jump size δ are nonzero. This is seen when comparing the posterior estimates of the excess kurtosis, which are qualitatively different between the BM simulations and the real data (Figure 4). In addition, the posterior estimates of the variance of the process were nearly identical between the BM simulations and the real data. Together, these provided evidence that the evolution of primate body mass is not well explained by BM alone.

For ECV under the AS model, the posterior density of α inferred from the ECV data placed extremely little mass on 2, while the BM simulations consistently resulted in maximum a posteriori estimates of $\alpha = 2.0$ (Figure 5), evidence of non-Gaussian evolution of primate ECV.

For the mass-to-ECV ratio, we found no remarkable deviation from BM (Figure 6). This is reflected in the fact that the posteriors of the kurtosis for the JN and VG models, as well as the posterior of α for the AS model, were extremely close to the posteriors inferred from the BM simulations.

Figure 7 shows the primate phylogeny with branches colored according to their branch-normalized signal-to-noise ratios. Because we rejected the BM in favor of a Lévy process with jumps in the body mass and ECV data, nonzero signal-to-noise ratios are possibly explained by jumps in trait evolution.

DISCUSSION

Darwin (1859) first proposed that what is now called continuous character evolution occurs gradually, with species changing very little over short time periods. Since then, some (Simpson 1953; Eldredge and Gould 1972; Stanley 1975) have suggested that evolution occasionally happens more quickly, with rapid changes in characters occurring over short periods of evolutionary time. However, most studies of continuous trait evolution that use comparative data rely on a Brownian motion (BM) model. Because the path of a BM is continuous; that is, the value of the trait at the next moment in time

is necessarily very close to the value of the trait at the current moment, the most natural interpretation of these models excludes the possibility of saltational change. Moreover, even though some saltational processes can produce the same distribution of tip data as a BM, these are highly restricted—for example, if jumps occur only at nodes in the tree that lead to extant taxa.

A natural generalization of BM that allows for paths that are not strictly continuous is the class of Lévy processes. The discontinuities in the path can be thought of as "jumps," in which the character changes instantly without any intermediate forms. These jumps approximate rapid changes in character value over a short time scale and result in distributions of character change that have "fat tails"; in statistical literature, distributions with fat tails are said to be leptokurtic.

We examined 3 specific Lévy processes: a compound Poisson with normally distributed jumps (JN), variance gamma (VG), and α-stable (AS). All processes also include a BM component, and hence can be interpreted as modeling gradual evolution punctuated by large, sudden changes in trait value. The JN process waits an exponentially distributed amount of time with rate λ before making a jump whose size is drawn from a normal distribution with standard deviation δ. The VG and AS processes are so-called infinitely active processes that jump infinitely often. However, most of the jumps are arbitrarily small, and so the processes are well behaved. An important difference between the VG and AS processes is that the AS process is much more likely to take extremely large jumps, compared with the VG process (as reflected in the fact that the variance of the AS process is infinite). For the VG process, the parameter κ corresponds to the rate of very large jumps and the parameter τ controls the variance of the jumps that are taken. In the AS process, the parameter α is confined between 0 and 2. As α approaches 2, the process converges in distribution to a BM, while as it approaches 0 the process makes larger jumps more frequently. The parameter β controls the scale of jumps that are taken.

These processes can be interpreted in a biological context. The JN process reflects the classic idea of stasis punctuated by rapid character change and has some history in the literature (Hansen and Martins 1996; Bokma 2008; Uyeda et al. 2011). VG and AS are more exotic
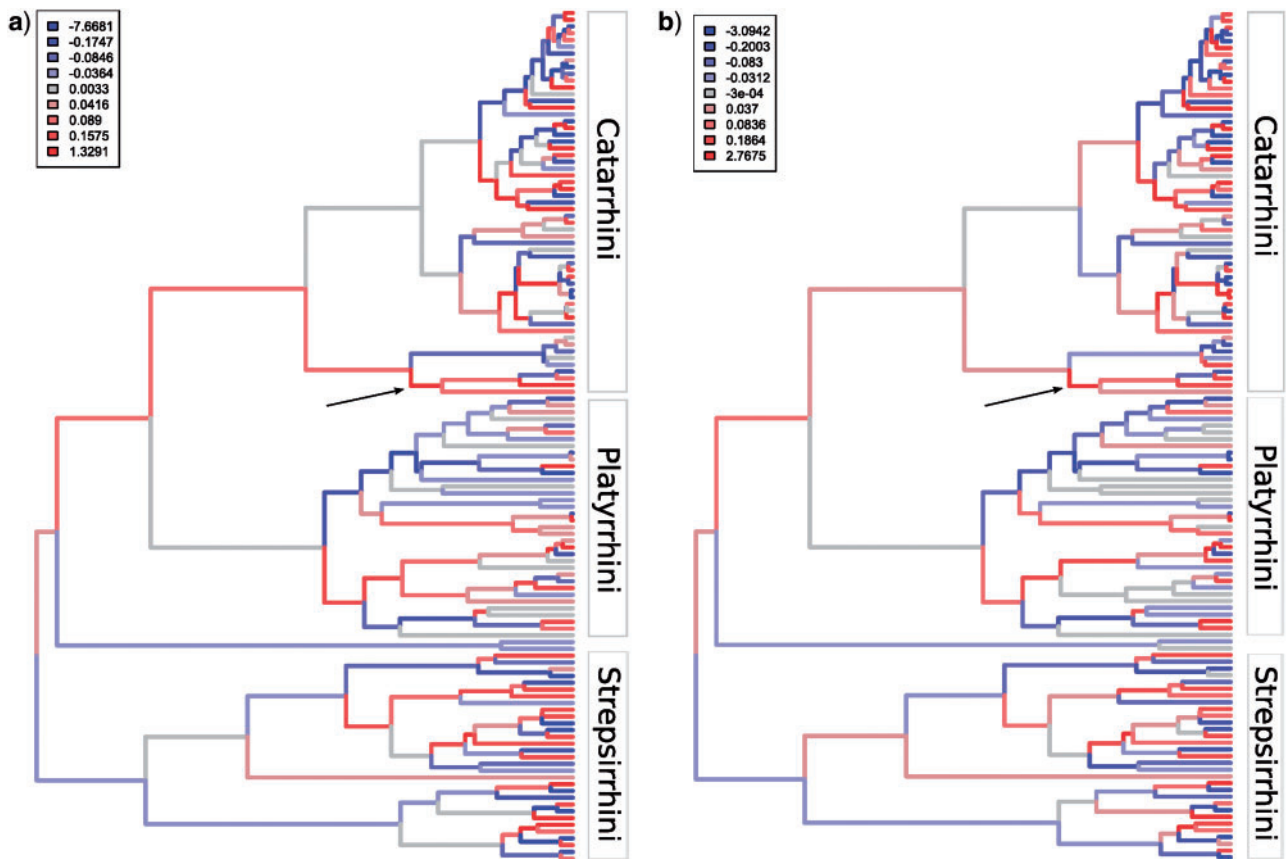
FIGURE 7.   Branch-normalized signal-to-noise ratios of posterior jump distributions. The primate phylogeny with inferred evolutionary histories for body mass under JN (a) and for ECV under AS (b) are shown. Branches are shaded according to the quantile containing their branch length-normalized signal-to-noise ratios. A value of approximately zero indicates trait evolution explained predominantly by the BM component of the fitted model. Uncolored branches in light gray indicate the tendency for the model to explain trait evolution with jumps valued according to the figure legend. The arrow points to the most recent common ancestor of great apes.

models; however, they may capture certain aspects of evolution that would otherwise be impossible to model. For example, in the Lande (1976) description of the impact of genetic drift on quantitative traits, trait evolution is a BM on a time scale determined by the effective population size: evolution works more slowly in large population and more quickly in small populations. Because the VG process arises as a time change of a BM (Madan et al. 1998), it can capture the impact of fluctuating population size on continuous character evolution. The AS process, on the other hand, is a natural generalization of BM that has many of the same features, but allows for fatter tails and erratic sample paths.

Because analytic computation of the likelihood using Felsenstein's pruning algorithm is not possible for the Lévy processes that we considered, we developed a MCMC method to estimate the parameters of a Lévy process. The MCMC algorithm samples possible jump histories along each branch of the phylogeny. Using data augmentation for ancestral states, similar in spirit to that of Robinson et al. (2003), we numerically integrate over the history of jumps. Because any Lévy process can be split into a BM and pure-jump components, our method is applicable to any Lévy process outside of the examples we considered here.

To determine whether a phylogeny contains sufficient information to reject single-rate BM in favor of a more

general Lévy process, we conducted simulation studies using each of the models that we implemented. Figure 2 shows that we were able to recover the parameters of the JN and AS processes with high accuracy. However, for the VG process, τ, which controls the variance of the jumps, was well estimated, but the rate of large jumps, κ, is underestimated. We are uncertain why κ is consistently underestimated but suspect that tree shape plays an important role.

We then made use of the fact that non-Brownian Lévy processes have more frequent large deviations in short time periods than BM. This large deviation is manifested as excess kurtosis. Using the characteristic function of a Lévy process (i.e., the Lévy–Khinchine formula), we calculated the posterior distribution of the variance and excess kurtosis per unit time. Because the Gaussian distribution has zero excess kurtosis, this posterior estimate should be close to zero when BM is a good model for trait evolution and as significant mass away from zero when the trait evolution is non-Gaussian. In the case of the AS process, the excess kurtosis is not defined and so we focused our attention on the parameter α. As $\alpha \rightarrow 2$, the AS process becomes a BM; thus, if the posterior distribution of α was not very close to 2, the evolution of the continuous character was inferred to be non-Brownian.

We applied our MCMC method to data from 126 primate species (Isler et al. 2008; Eastman et al. 2011) to uncover evidence of non-Gaussian evolution in a large group of mammals. For each species, we obtained measurements of body mass, ECV and also examined the ratio of mass-to-ECV. For the body mass and ECV data, we found evidence supporting Lévy process with jumps over BM and highlighted results under the JN and AS models, respectively, while the mass-to-ECV ratio appeared to evolve as a BM. The parameters inferred for body mass suggest that there is a burst of body size evolution equivalent to 5–6 myr of gradual evolution approximately once every 4 myr, which is within the same order of magnitude of jump periodicity as reported by Uyeda et al. (2011). ECV evolution was fit by an AS process, with an intermediate value of $\alpha = 1.7$, consistent with a mode of evolution in which character changes are mostly gradual but punctuated by infrequent, extremely large jumps.

We also obtained a posterior distribution on the amount of trait change in excess to the BM component of the Lévy process on every branch of the phylogeny. Using these data, we identified branches of the primate phylogeny that showed evidence for evolution that was faster than the BM component of the Lévy process could explain. In Figure 7, we colored the branches of the primate phylogeny according to the signal-to-noise ratio of the jump size on that branch, normalized by the branch length. Because the jumps account for the "extra" distance that the BM component of the model cannot explain, large magnitudes of this ratio correspond to branches where there is relatively strong evidence for trait evolution faster than the average BM rate on the tree. This signal weakens deep in the tree, as well as for long branches, although it is interesting to note that some deep branches show excess evolution relative to their branch length (e.g., body mass in the common ancestor of old world monkeys and apes). We identified several clades that showed strong evidence of unexpectedly rapid evolution prior to diversification. For example, the ancestor of the great apes (indicated by an arrow) shows evidence of unexpectedly rapid evolution in both body mass and ECV, while evolution in the ancestors of the Old World and New World monkeys is well explained by the average rate of BM.

Although our method is able to discriminate between Gaussian and non-Gaussian evolutionary models, we were not able to find a test statistic that could discriminate between the different jump processes. Bayesian methods of model testing, such as Bayes factors, require computing the marginal likelihood. However, because of the stochastic nature of our method for integrating over the large number of possible jump histories using MCMC, many methods for estimating the marginal likelihood of a model are unstable or require an unfeasibly large number of MCMC cycles. Moreover, since the method we present does not compute the marginal likelihood of the parameters alone (with the jumps integrated out), we cannot use information criteria such as the Akaike Information Criterion to conduct model fitness tests. In future work, we plan to implement a Bayesian reversible-jump MCMC method to distinguish between different jump models. This will help to identify how much signal the data contains to single out any particular Lévy process model of evolution. Although the method presented in this article conducts inference under time-homogeneous Lévy processes, nothing prevents the model from being implemented in a rate-shifting framework (see O'Meara et al. 2006; Eastman et al. 2011). This will further help to distinguish jump events from rate-shifting events.

Previous methods describing inference of Lévy processes in the mathematical finance literature have shown that it is possible to precisely infer parameters and accurately choose models with time-series data. However, the correlation structure of a phylogeny complicates inference. As noted by Ané (2008), phylogenetic inference of trait evolution is strongly affected by tree shape and proposed an effective sample size to gauge how powerful a given topology is for the inference of model parameters. Boettiger et al. (2011) explored the impact of tree shape on the ability for model tests to distinguish BM models from Ornstein–Uhlenbeck models of continuous trait evolution. Further examination of how tree shape affects inference will become particularly important as increasingly complex models of continuous character evolution are put forward (Khaitovich et al. 2005; Bokma 2008; Harmon et al. 2010; Eastman et al. 2011).

We face 2 other problems owing to the nature of the phylogeny and the data being analyzed. To illustrate these problems, consider that the clearest signal of excess kurtosis that our model captures lies in terminal sister nodes, where one lineage has evolved as expected under BM, but the other lineage has experienced an abnormally large jump in trait change. First, assigning data with measurement or sampling error to the tips could introduce (or mask) an excess of trait change and lead to the false inference of the presence of jumps in trait evolution for the phylogeny. If this is a concern, tips may be modeled with noise at the potential price of losing power to reject BM in favor of a Lévy process with jumps. Second, the phylogeny is assumed to be fully resolved with errorless branch lengths. If a trait truly evolved by single-rate BM but exhibits an excess of trait change for the specified branch length, it is possible that branch's trait evolution is simply an outlier among realizable evolution histories, but it is also possible that the true branch length is longer than indicated. A potential solution is to include posterior samples of the branch length from a Bayesian phylogenetic analysis.

When fitting models of evolution to comparative data, it is important to keep in mind the distinction between the model of evolutionary change and the joint distribution of trait values at the tips that such a model produces. This mapping is not one-to-one; many different models can result in the same joint distribution at the tips and are therefore indistinguishable from the data alone. To choose between these otherwise equivalent models, scientists must look beyond comparative data, for example, to the fossil record and mechanistic biological models. Here, we have used BM as a representative process that results in a multivariate normal distribution with a particular covariance structure. Other processes that produce this same joint distribution exist. Similarly, though we fit models with jumps, there are many gradual processes that can produce the exact same distribution at the tips as a jump model, such as models which use BM with random rate shifts (although these models may not have straight forward or desirable biological interpretations).

Many other Levy processes exist. We have only showcased a few, but our method can be applied to any Lévy process with a known characteristic function. It will be interesting to see whether different evolutionary processes, different clades, or different traits are best

modeled by certain types of Lévy processes, be it BM or the AS process.

REFERENCES

Abramowitz M. Stegun I. 1964. Handbook of mathematical functions with formulas, graphs, and mathematical tables, Vol. 55. New York: Dover Publications.

Ané C. 2008. Analysis of comparative data with hierarchical autocorrelation. Ann. Appl. Stat. 2:1078–1102.

Boettiger C., Coop G., Ralph P. 2011. Is your phylogeny informative? Measuring the power of comparative methods. Evolution 66: 2240–2251.

Bokma F. 2002. Detection of punctuated equilibrium from molecular phylogenies. J. Evol. Biol. 15:1048–1056.

Bokma, F. 2008. Detection of òpunctuated equilibriumó by Bayesian estimation of speciation and extinction rates, ancestral character states, and rates of anagenetic and cladogenetic evolution on a molecular phylogeny. Evolution 62:2718–2726.

Brawand D., Soumillon M., Necsulea A., Julien P., Csárdi G., Harrigan P., Weier M., Liechti A., Aximu-Petri A., Kircher M., Albert F.W., Zeller U., Khaitovich P., Grützner F., Bergmann S., Nielsen R., Pääbo S., Kaessmann H. 2011. The evolution of gene expression levels in mammalian organs. Nature 478:343–348.

Cavalli-Sforza L., Edwards A. 1967. Phylogenetic analysis. Models and estimation procedures. Am. J. Hum. Genet. 19:233–257.

Chaix R., Somel M., Kreil D., Khaitovich P., Lunter G. 2008. Evolution of primate gene expression: drift and corrective sweeps? Genetics 180:1379–1389.

Darwin, C. 1859. On the origin of species. London: John Murray. p. 490.

Drummond A., Suchard M. 2010. Bayesian random local clocks, or one rate to rule them all. BMC Biol. 8:114.

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

Eldredge N., Gould S. 1972. Punctuated equilibria: an alternative to phyletic gradualism. Models Paleobiol. 82:115.

Felsenstein J. 1981. Evolutionary trees from DNA sequences: a maximum likelihood approach. J. Mol. Evol. 17:368–376.

Felsenstein J. 1985. Phylogenies and the comparative method. Am. Nat. 125:1–15.

Freckleton R., Harvey P., Pagel M. 2003. Bergmann's rule and body size in mammals. Am. Nat. 161:821–825.

G.P. Contributors 2010. GSL—GNU scientific library—GNU project— free software foundation (FSF).

Hansen T. Martins E. 1996. Translating between microevolutionary process and macroevolutionary patterns: the correlation structure of interspecific data. Evolution 149:646–667.

Harmon L., Losos J., Davies T., Gillespie R., Gittleman J., Jennings W., Kozak K., McPeek M., Moreno-Roark F., Near T., Purvis A., Ricklefs R., Schluter D., Schulte II J., Seehausen O., Sidlauskas B., Torres-Carvajal O., Weir J., Mooers A. 2010. Early bursts of body size and shape evolution are rare in comparative data. Evolution 64: 2385–2396.

Huelsenbeck J., Larget B., Swofford D. 2000. A compound poisson process for relaxing the molecular clock. Genetics 154: 1879–1892.

Isler K., Kirk E., Miller J., Albrecht G., Gelvin B., and Martin R. 2008. Endocranial volumes of primate species: scaling analyses using a comprehensive and reliable data set. J. Hum. Evol. 55: 967–978.

Kallenberg O. 2010. Foundations of modern probability. New York: Springer.

Khaitovich P., Pääbo S., Weiss G. 2005. Toward a neutral evolutionary model of gene expression. Genetics 170:929–939.

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

Li H., Wells M., Yu C. 2008. A Bayesian analysis of return dynamics with Lévy jumps. Rev. Financ. Stud. 21:2345–2378.

Madan D., Carr P., Chang E. 1998. The variance gamma process and option pricing. Eur. Financ. Rev. 2:79–105.

Mattila T., Bokma F. 2008. Extant mammal body masses suggest punctuated equilibrium. Proc. R. Soc. Lond. B. 275:2195–2199.

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

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

Redding D., DeWolff C., Mooers A. 2010. Evolutionary distinctiveness, threat status, and ecological oddity in primates. Conserv. Biol. 24:1052–1058.

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

Ronquist F., Teslenko M., van der Mark P., Ayres D., Darling A., Höhna S., Larget B., Liu L., Suchard M., Huelsenbeck J. 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Syst. Biol. 61:539–542.

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

Stanley S. 1975. A theory of evolution above the species level. Proc. Natl Acad. Sci. USA 72:646.

Thorne J., Kishino H., Painter I. 1998. Estimating the rate of evolution of the rate of molecular evolution. Mol. Biol. Evol. 15: 1647–1657.

Uyeda J., Hansen T., Arnold S., Pienaar J. 2011. The million-year wait for macroevolutionary bursts. Proc. Natl Acad. Sci. USA 108: 15908–15913.