

METHODOLOGY ARTICLE

Open Access

Inferring ancestral states without assuming neutrality or gradualism using a stable model of continuous character evolution

Michael G Elliot^{1*} and Arne Ø Mooers²

Abstract

Background: The value of a continuous character evolving on a phylogenetic tree is commonly modelled as the location of a particle moving under one-dimensional Brownian motion with constant rate. The Brownian motion model is best suited to characters evolving under neutral drift or tracking an optimum that drifts neutrally. We present a generalization of the Brownian motion model which relaxes assumptions of neutrality and gradualism by considering increments to evolving characters to be drawn from a heavy-tailed stable distribution (of which the normal distribution is a specialized form).

Results: We describe Markov chain Monte Carlo methods for fitting the model to biological data paying special attention to ancestral state reconstruction, and study the performance of the model in comparison with a selection of existing comparative methods, using both simulated data and a database of body mass in 1,679 mammalian species. We discuss hypothesis testing and model selection. The stable model outperforms Brownian and Ornstein-Uhlenbeck approaches under simulations in which traits evolve with occasional large “jumps” in their value, but does not perform markedly worse for traits evolving under a truly Brownian process.

Conclusions: The stable model is well suited to a stochastic process with a volatile rate of change in which biological characters undergo a mixture of neutral drift and occasional evolutionary events of large magnitude.

Keywords: Comparative methods, Ancestral state reconstruction, Evolutionary models

Background

Statistical methods that take into account the dependencies introduced into comparative data by phylogenetic relatedness are fundamental to hypothesis testing and exploration in comparative biology [1,2]. Each comparative method implicitly imputes to the evolutionary process some specific stochastic model [3]. The validity of phylogenetic comparative methods depends on the degree to which historical events can be accommodated by the underlying stochastic model of trait evolution, and a mismatch between model and reality can yield erroneous statistical results, especially in the reconstruction of inaccurate ancestral character states [4-7]. For this reason it is important to develop realistic stochastic models of

character evolution (and to constrain those models using empirical data where possible).

The Brownian motion model of evolution – in which the value of a continuous trait evolves by accruing incremental changes drawn from a random distribution with zero mean and finite constant variance, such that the sum of many increments is distributed according to a normal density [1] – was introduced to model changes in gene frequencies by Cavalli-Sforza and Edwards [8] but now underlies (directly or indirectly) a range of popular methods for the analysis of continuous traits distributed over phylogenetic trees. In the realm of ancestral state reconstruction the value of a trait evolving across a phylogenetic tree can at all times be shown to be probabilistically distributed according to a normal distribution with variance depending only on tree topology and branch lengths [9-11], while in the realm of regression analysis the model yields normally distributed residuals and a covariance matrix depending only on tree topology and

*Correspondence: me378@cam.ac.uk

¹St. John's College, University of Cambridge, Cambridge CB2 1TP, UK
Full list of author information is available at the end of the article

branch lengths [12], both cases giving rise to simple and analytically-tractable solutions. The Brownian walk of a trait value can be regarded as a model of gradualistic neutral evolution since variation in the trait arises from a process of random drift over the branches of a phylogeny at a constant rate and without directionality. A further application thus involves identifying significant departures from the expectation of the Brownian motion model as a means of detecting adaptive variation in the tempo and mode of trait evolution [13-15].

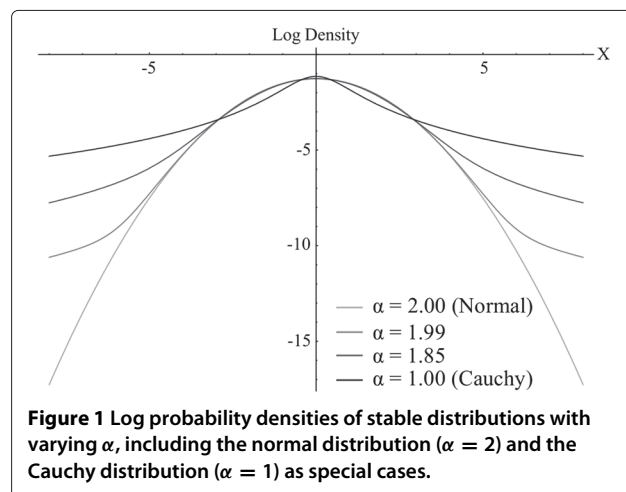
We here describe a simple generalization of the Brownian motion model of continuous character evolution which extends the model to include cases where increments to an evolving character may arise from a symmetric stochastic process but without assuming constant finite variance. Not only is the Brownian assumption of constant finite variance, to the best of our knowledge, unverified in existing biological systems, but its relaxation may better suit the domain of application of the standard Brownian motion model, namely biological characters likely subject to some degree of selection, and in many cases it may offer a more robust form of statistical inference with respect to outliers in the data. Broadly speaking, we conceptualize diversifying selection on continuous characters as causing an increase and purifying selection on such characters a decrease in the rate of evolutionary change. The relative frequencies of these forms of natural selection along with neutral drift are expected to generate — for sums of evolutionary increments over long periods of time — limit distributions with heavier tails than expected under the Brownian motion model.

Just as the limit distribution of sums of variates drawn from a distribution with constant finite variance is the normal distribution, so the limit distribution of sums of variates drawn from a distribution *without* fixed finite variance is the stable distribution [16-18], which has the normal distribution as a special case and which otherwise is characterized by heavy tails, closure under convolution and, potentially, by skew [19]. Stochastic processes and random walks with volatile variance and heavy tails have previously been modelled robustly using stable distributions in areas as diverse as the study of fractional diffusion [20], earthquake forecasting [21], signal processing [22], animal foraging [23], rainfall modelling [24], commodity pricing [25], real estate markets [26], foreign exchange rates [27], financial statistics [28], image processing [29] and telecommunications management [30]. Landis *et al.* [31] recently used a Brownian motion model with Lévy stable jumps to model jumps in the evolution of continuous traits. According to our view of selection resulting in the generation of evolutionary rate volatility we limit our attention in this paper to the symmetric zero-centred stable distribution parameterized by α , the index of stability and c , the scale. We model evolution using the stable

generalization of Brownian motion, the stable random walk [17]. The stable random walk shares some attractive properties of Brownian motion, the most important being closure under convolution, such that the sum of several stable distributions is itself a stable distribution with the same α parameter. This closure under linear transformation contrasts with all other non-stable heavy tailed distributions, which may yield complex and analytically intractable mixtures in convolution. Thus, after accumulating increments from a symmetrical zero-centred stable distribution, the value of an evolving trait is always probabilistically distributed according to a stable distribution with mean equal to the trait value prior to accumulation and with scale proportional to the number of increments, or in phylogenetic parlance the branch length.

Figure 1 illustrates the log probability densities of some unit-scale zero-centred symmetrical stable distributions that differ in the value of α , including the special cases of the normal distribution ($\alpha = 2$) and the Cauchy distribution ($\alpha = 1$). Figure 2 illustrates some stable random walks driven by accumulation of increments from stable noise with various α values. It is clear from inspection of these figures that declining α is associated with increasingly heavy tails, translating into increasingly volatile random walks which exhibit Brownian-like drift (associated with increments drawn from the high-probability modal region of the underlying distribution) interspersed with occasional rapid jumps in trait value (associated with increments drawn from the low-probability heavy tails).

As a motivating example of how the accommodation of evolutionary rate volatility may affect the inference of evolutionary processes, we present a simple toy model of continuous character evolution on a phylogenetic tree with six tips. The values of an evolving character were simulated under the Brownian motion model and values at internal nodes were reconstructed based on values at the tips only, first by fitting a Brownian motion model and



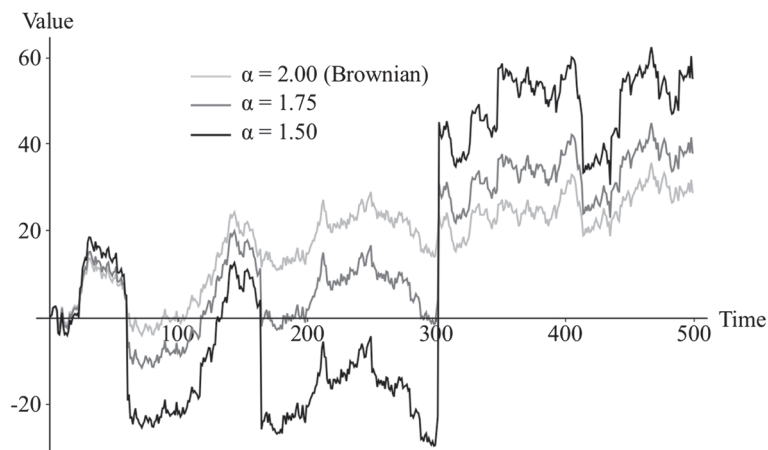


Figure 2 Stable random walks with varying α , driven by accumulation of a single sample of 500 increments drawn from a uniform distribution between zero and one, transformed onto stable distributions using the inversion method [32].

second by fitting a stable model using the method presented later in this paper. Next, the value of the character on a single tip was artificially inflated by a factor of ten, to represent a bout of adaptive evolution (or other event such as measurement error) on the branch leading to that tip. Again, values at internal nodes were reconstructed under both models. Results are illustrated in Figure 3.

Both reconstruction methods yield similar ancestral states for the original Brownian motion data, but differ for the manipulated data. The Brownian motion model exhibits an “averaging effect” in which the apparently high rate of evolution resulting from the manipulation is distributed somewhat evenly over all the branches, causing a large increase in estimated ancestral states for several internal

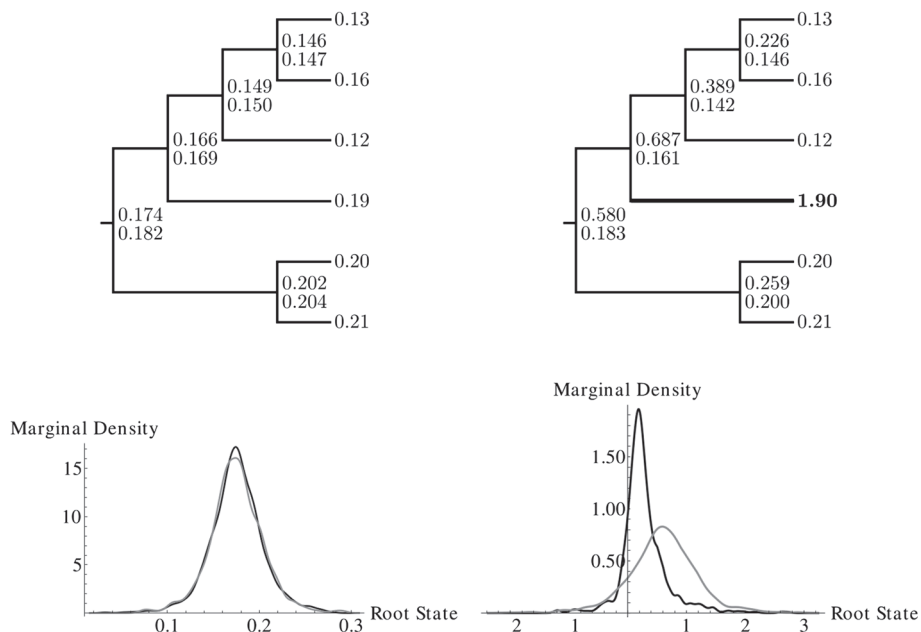


Figure 3 Ancestral state reconstruction of data simulated under Brownian motion. Top row: phylogenies exhibiting ancestral state reconstructions; each tip is labeled with the known character state, while at each internal node, upper values represent the Brownian motion maximum likelihood reconstruction and lower values represent the reconstruction under the stable Markov chain Monte Carlo model to be described in this paper (left: original simulated data; right: the character value of a single tip has been multiplied by ten to simulate rapid evolution (or measurement error) on a single branch). Bottom row: the marginal probability density derived from MCMC for the ancestral state assignment of the root node using unmodified (left) and modified (right) data; the Brownian motion marginal probability density is grey and the stable margin probability density is black.

nodes. The stable model, however, can entertain rare increments of large magnitude, and so is not strongly affected by the manipulation; the apparently high rate of evolution on a single branch can be accommodated as a rare event in a heavy-tailed stochastic process.

In this paper we describe the stable model and the procedures used to fit it to empirical data, with an emphasis on ancestral state reconstruction. We outline strategies for hypothesis testing and model selection. We apply the stable model to simulated data in order to estimate error rates of the model selection procedure, and also use it, alongside a number of alternative models, in an illustrative case study of ancestral state reconstruction based on a large dataset of mammalian body masses. Finally we discuss the relationship between the stable model of trait evolution and a number of alternative extensions of the Brownian motion model that have been proposed in the literature, and suggest avenues for further development.

Methods

Model

Consider a rooted phylogenetic tree \mathcal{T} which may or may not contain polytomies. A continuous character \mathbf{X} evolves along the branches of \mathcal{T} , taking values b_1 and b_2 at the beginning and end, respectively, of each branch b . Under the standard Brownian motion model of evolution, the continuous character evolves by accumulating random independent increments drawn from a probability distribution with constant mean zero and constant finite variance σ^2 . According to the central limit theorem, the sum of such increments along a branch b of length t_b is probabilistically distributed according to a normal density with mean zero and variance $t_b\sigma^2$, a density which we denote $\phi(b_2 - b_1; t_b\sigma^2)$. Given the independence of increments, and therefore of branches, the likelihood of an ancestral state reconstruction of a continuous character evolving under Brownian motion is given by the product:

$$L(\mathbf{X}, \sigma; \mathcal{T}) = \prod_b \phi(b_2 - b_1; t_b\sigma^2) \quad (1)$$

If the variance of the increment generating distribution is not constant and finite (as we suppose to be the case under departures from neutrality and gradualism) then according to the generalized central limit theorem the limit distribution for the sum of random independent variates is not normal but falls into the more general class of stable distributions, parameterized by an index of stability α and a scale c . The symmetric stable distribution has probability density denoted $S(x; \alpha, c)$. Following

Matsui and Takemura [33], the unit stable density with $c = 1$ may be defined as:

$$S(x; \alpha, 1) = \frac{\alpha}{\pi|\alpha - 1|x} \int_0^{\frac{\pi}{2}} G(\kappa; \alpha, x) \exp(-G(\kappa; \alpha, x)) d\kappa \quad (2)$$

where

$$G(\kappa; \alpha, x) = \left(\frac{x \cos \kappa}{\sin \alpha \kappa} \right)^{\frac{\alpha}{\alpha-1}} \frac{\cos(\alpha - 1)\kappa}{\cos \kappa}, \quad (3)$$

and the general symmetrical stable density $S(x; \alpha, c)$ is a transformation of the unit stable density $S(\frac{x}{c}; \alpha, 1) / c$.

One important property of the stable distribution is that the normal distribution is a special case with $\alpha = 2$. For the zero-centred symmetrical cases treated here, we note that:

$$\phi(x; \sigma^2) = S\left(x; 2, \frac{\sigma}{\sqrt{2}}\right) \quad (4)$$

Furthermore, the sum of t variates drawn from a stable distribution $S(\alpha, c)$ is distributed as $S\left(\alpha, (tc^\alpha)^{\frac{1}{\alpha}}\right)$. Thus, under a stable model of evolution, the likelihood of an ancestral state reconstruction of \mathbf{X} is given by:

$$L(\mathbf{X}, \alpha, c; \mathcal{T}) = \prod_b S\left(b_2 - b_1; \alpha, (t_b c^\alpha)^{\frac{1}{\alpha}}\right) \quad (5)$$

which is functionally identical to Eq. 1 when $\alpha = 2$.

Unfortunately, there is no analytical solution to the stable probability density function in Eq. 2, so it is necessary to employ numerical methods [19,33-36] to calculate likelihoods of stable models. The model does not lend itself to direct maximum likelihood estimation of parameters due to the existence of a highly multi-modal likelihood surface and because arbitrarily high likelihoods can be obtained by setting $b_1 = b_2$ for any single internal branch b and having $c \rightarrow 0$, a problem exhibited by other statistical models with non-constant variance [37], and circumvented through the placing of an appropriate prior on the scale parameter that penalizes the approach to zero, and fitting the model using a Bayesian approach. Since numerical estimation methods are unreliable under extremely heavy tails (i.e. $\alpha < 0.2$) [33] we apply flat or triangular priors to the index of stability on the domain $0.2 < \alpha \leq 2$, and a loose inverse gamma prior on the scale parameter which has $Pr(c \rightarrow 0) \rightarrow 0$ (see section on Choice of Priors below).

MCMC estimation

Markov chain Monte Carlo (MCMC) methods are widely used to estimate complex multivariate probability

densities in numerous biological fields. The goal of such methods is to generate a sample from a probability distribution by constructing a Markov chain that has the desired distribution as its equilibrium density. A common strategy is to utilize a Metropolis-Hastings sampler [38] in which the statistical model is initialized with some set of parameter values θ , a new candidate parameter θ' is generated by a symmetrical proposal distribution, and accepted as the next step of the Markov chain with probability equal to $P(\theta')/P(\theta)$. We found that the Metropolis-Hastings sampler performed poorly in estimating ancestral states and parameters of the stable model due to the multi-modality of the local likelihood surface (Figure 4) and the non-independence of ancestral state values, which together generate numerous very small “islands” of high likelihood which are unlikely to be explored by the Markov chain in a reasonable amount of time. This results from the fact that proposals generated by the Metropolis-Hastings algorithm are constrained by the proposal distribution to be close in value to the current parameter value. Modified versions of the procedure, such as Metropolis-coupled Markov chain Monte Carlo [39] did not yield any benefit.

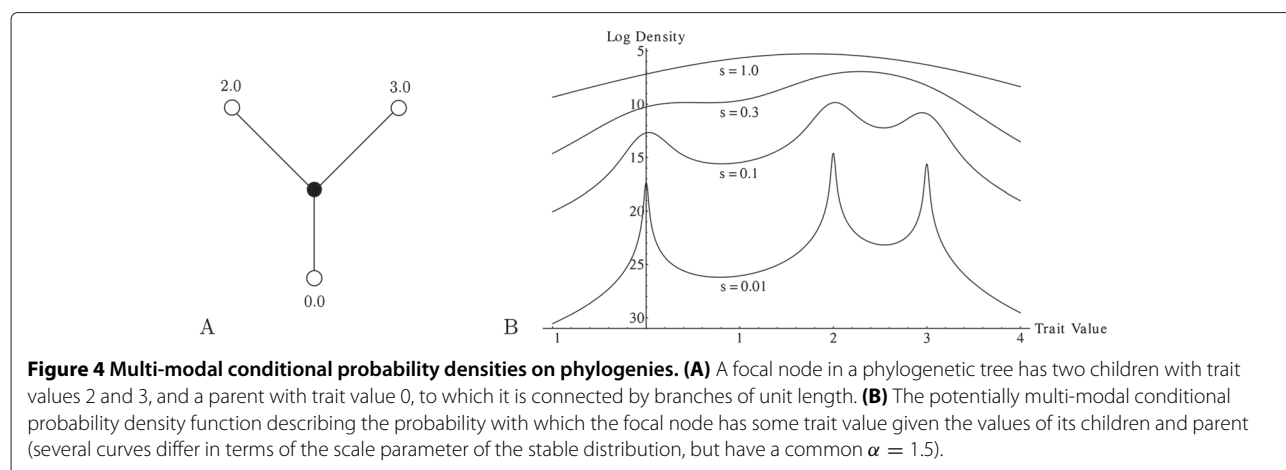
We found that implementation of a slice sampler [40] manifestly improved the mixing of Markov chains. New proposals under a slice sampler are drawn from the entire range of possible values of the parameter and are not restricted by a proposal distribution to be close in value to the current estimate. Instead of accepting new proposals according to the likelihood ratio criterion, slice samplers accept all new proposals but tend to make proposals that have likelihood similar to the current likelihood of the chain. This allows large jumps across widely dispersed peaks in the multi-modal likelihood surface at each step of the Markov chain, offering an “escape” from local sub-optimal peaks in the likelihood surface. To be specific, each step of the chain involves the value of each individual parameter θ_i being replaced by a new value θ'_i drawn

randomly from the conditional probability distribution $P(\theta'_i|\theta_j, \theta_k, \dots)$.

The procedure for generating new proposals under slice sampling is a three-step process. First, the conditional probability of the current parameter, given fixed values for all other parameters in the model, $P(\theta_i|\theta_j, \theta_k, \dots)$, is estimated. This is a numerical value between zero and one that is proportional to the likelihood of the current model. Second, a random number y is drawn from the uniform distribution between zero and the calculated conditional probability. Third, we identify the set of all possible parameter values which would have conditional probability $P(\theta'_i|\theta_j, \theta_k, \dots)$ greater than y and accept a new proposal at random from this set. If y is very close to the current conditional probability, the set of proposals with conditional probability greater than y will tend to include large numbers that fit the data better than the current model. Selecting a proposal at random from this set would tend to increase the likelihood of the Markov chain over time. However, when y is much less than the current conditional probability, the set of proposals with conditional probability higher than y will also include many candidates with lower likelihood, and selecting a proposal from the set would permit the chain to decline in likelihood over time. Neal (2003) has shown that the stationary distribution of such a Markov chain constitutes a sample from the complete posterior distribution that we are attempting to characterize. Critically, for a highly multi-modal posterior distribution, slice sampling permits large jumps away from suboptimal likelihood peaks even when the sampled distribution exhibits widely separated modes. The procedure is illustrated graphically in Figure 5.

Choice of priors

We specify prior distributions on the stable parameters α and c , denoted $\text{Pr}(\alpha)$ and $\text{Pr}(c)$ respectively. We have found that choice of priors on the index of stability have little effect on the ancestral state reconstruction, while



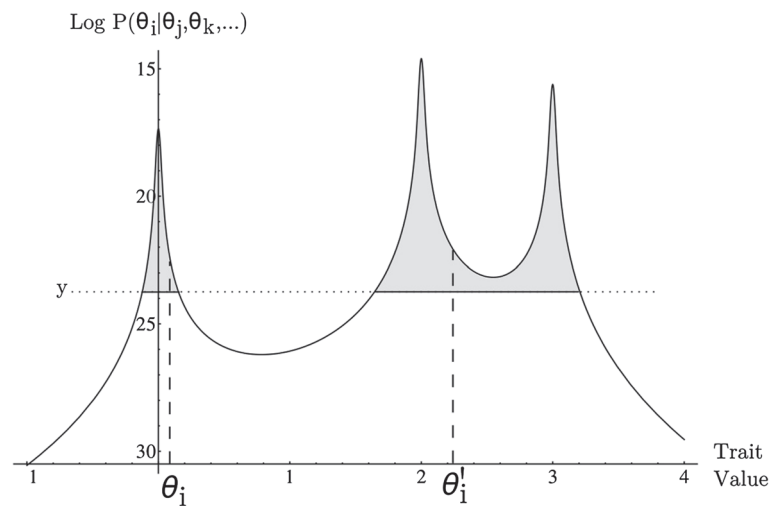


Figure 5 Slice sampling. An example of slice sampling updating the value of the focal node from Figure 4(A) from θ_i to θ'_i . First, the conditional probability that the value of the evolving character at the focal node is equal to θ_i – given the other ancestral states and stable model parameters – is calculated. A random number y is drawn from the uniform distribution between zero and this conditional probability (marked with a dotted line). The set of possible values of θ'_i is estimated by bracketing regions around the modes of the distribution for which the conditional probability is greater than y (solid lines and shaded region of the distribution). The new value θ'_i is drawn as a uniform random variable from this set.

the prior on the scale parameter effects the smoothness of the posterior density (Figure 4, right panel). How should we choose appropriate priors for the Monte Carlo simulation? We believe that the Brownian motion model is a useful null model for the evolutionary process. Since the Brownian motion model is a special case of the stable model it is conservative to choose priors that maximize the *a priori* likelihood of the Brownian model fit. Consider Equations 1 and 5. If we set $\alpha = 2$ and $c = \sigma/\sqrt{2}$ then the stable model will be identical to the inferred Brownian motion model. So, priors with maximum likelihood at $\alpha = 2$ and $c = \sigma/\sqrt{2}$ are conservative with respect to the null hypothesis.

We use a uniform or triangular distribution between 0.2 and 2 for $\text{Pr}(\alpha)$. We do not permit $\alpha < 0.2$ since numerical estimation methods are unreliable at this extreme of heavy tails [33]; if a triangular rather than uniform prior is selected then its maximum is set equal to 2.

We use an inverted gamma distribution for our prior on c , though of course in principle the stable model accommodates any reasonable priors that maintain $0 < \alpha \leq 2$ and $c > 0$. Our choice of $\text{Pr}(c)$ is motivated by the fact that the inverted gamma distribution has an appropriate shape for the scale of a stable distribution and, as the conjugate prior for a normal variance, it is commonly used as the prior on the unknown variance of a normal distribution with fixed mean [41]. It is defined as:

$$g(x; a, b) = \frac{b^a x^{-1-a} \exp\left(-\frac{b}{x}\right)}{\Gamma(a)} \quad (6)$$

with $a > 0$ and $b > 0$. As suggested above, it seems conservative to suppose that the rate of evolutionary change with highest *a priori* likelihood should be the rate of change imputed by the neutral Brownian motion reconstruction, denoted σ in Equation 1. This rate of evolution is readily calculated from ancestral states generated by a squared change parsimony ancestral state reconstruction or other methods; hence, the inverted gamma prior on the scale of the evolutionary process should have its mode at $\sigma/\text{sqrt}2$. The mode of an inverted gamma distribution is defined as $x = b/(a - 1)$, so the values of the prior hyperparameters a and b should obey the following constraints: $a = -1 + b\sqrt{2}/\sigma$ and $b > \sigma/\sqrt{2}$. So, the value of a is entirely determined and the appropriate value of b is easily obtained by linear search or Newton-Raphson optimization, maximizing the likelihood of Equation 1.

In this way, our use of the Brownian motion model as a null hypothesis allows us to choose directly some reasonable and conservative priors on the more general stable model, in which the Brownian motion ancestral state reconstruction has highest prior likelihood. By default, the software accompanying this paper uses this approach, but can use custom values of the prior hyperparameters a and b upon request.

Hypothesis testing and model selection

It is desirable to formulate a statistical model selection criterion to determine whether the stable model of continuous character evolution (with $\alpha < 2$) fits the data better than than the Brownian motion model (with $\alpha = 2$), as a means of estimating the best possible ancestral state

estimates and of testing the hypothesis that a set of character data at the tips of a phylogeny exhibits patterns consistent with departure of the evolutionary process from neutrality. For maximum likelihood approaches, Akaike's Information Criterion (AIC) [42] is a popular choice. However, since AIC depends upon maximized likelihood, while our Markov chain Monte Carlo procedure only generates a sample from the posterior (conditional) distribution of stable parameters and ancestral states, it is difficult to calculate Akaike's Information Criterion in this case.

There exist a number of Bayesian generalizations of AIC which may be calculated from posterior MCMC samples, most notably the Deviance Information Criterion (DIC) [43]. Statistical deviance is a quantity that, for our purposes, can be defined as a measure of goodness of fit that is calculated at each step in the Markov chain, and is equal to minus twice the log likelihood of the stable model specified at the focal step of the Markov chain, i.e., $D = -2\log L(\mathbf{X}, \alpha, c; \mathcal{T})$. Hence, a lower deviance indicates a higher likelihood and better fit to the data. Akaike's Information Criterion is equal to the deviance of the *maximum likelihood* model, offset by a penalty for complexity equal to twice the number of parameters in the model. In the absence of a maximum likelihood estimate of the deviance, DIC makes use of \bar{D} , the average deviance across all posterior samples from the Markov chain, as its primary measure of goodness of fit. It is also possible to calculate a "base-line" measure of goodness of fit by calculating, for each parameter, the average value over all steps in the Markov chain, and then calculating the deviance of this average model, denoted \hat{D} . If \bar{D} is much higher than \hat{D} , the implication is that the model being fit is highly complex, since parameter estimates during the chain must be oscillating around values far removed from their mean across the chain as a whole. The Deviance Information Criterion encapsulates this notion of complexity in an "effective number of parameters" or p_D equal to $\bar{D} - \hat{D}$. So, while AIC is defined as the deviance of the maximized likelihood plus the number of parameters, DIC is defined as the average deviance across all posterior samples plus the effective number of parameters, or, $\bar{D} + p_D$.

Our simulation studies indicated that in phylogenetic datasets p_D did not increase sufficiently in line with tree size, resulting in over-fitting of stable models on large trees. This is likely because tree size is not incorporated as a component of model complexity in the information criterion. A Bayesian Predictive Information Criterion (BPIC) developed by Ando [personal communication [44]], which amounts to DIC with a increased multiplier on the p_D penalty, resolved this problem of over-fitting (see Results below). Specifically, we use the criterion $\bar{D} + 2p_D$.

Software implementation

Efficient software for fitting the stable model to phylogenetic trees and their associated data was written in C++ and is available at <http://www.sfu.ca/~micke/stabletraits.html> as source code and also compiled for various operating systems. The software reports a posterior sample of ancestral state reconstructions and stable parameter values in a format compatible with the Tracer software application [45], along with the proportional scale reduction factor convergence diagnostic [46] and Bayesian Predictive Information Criterion for assessment of model fit [44]. Multiple chains are run on independent threads, or on independent processors in a cluster computing environment (for which a torque job submission script is also available).

Application to simulated and natural data

In order to assess the improvement (if any) in quality of ancestral state reconstruction and the ability of statistical tests to identify biological characters that have evolved under a stable rather than Brownian stochastic process, data were simulated under a variety of conditions. Evolutionary increments were generated randomly from a stable model with unit scale and index of stability ranging from 1.0 to 2.0 in steps of 0.2, on random phylogenetic trees with 25, 40, 60, 90, 130, 175, 235, 325, 440 and 600 tips generated under the Yule model in Mesquite [47]. This procedure gave rise to 60 experimental conditions, each consisting of 250 trees and simulated datasets, and varying in tree size and index of stability of simulated trait values. Based on data at the tips, ancestral states were reconstructed using the modal posterior density estimate from MCMC, first with α fixed at 2.0 (representing a Brownian motion model) and second with a free α (representing a general symmetric stable model). Ancestral states were also estimated using the homogenous Ornstein-Uhlenbeck model for comparison. Estimates of Type I and Type II error rates under the BPIC criterion were made for each tree size/stability condition. Accuracy was assessed for Brownian, Stable and OU models by calculating variance of the inferred ancestral states from the true simulated states for each simulated dataset under each condition. We report here the median variance ratio of stable/Brownian and stable/Ornstein-Uhlenbeck reconstructions within each experimental condition.

To provide a demonstration of the model's application to real biological datasets, we made use of a supertree of mammalian species [48] and fitted the stable model of character evolution to data on the log adult body mass of 1,679 eutherian mammals [combined data from [49,50]]. For purposes of comparison we also estimated ancestral states under two homogenous evolutionary models, the Brownian motion model and the Ornstein-Uhlenbeck model, the latter estimates obtained by maximum

Table 2 Sum of squared errors in ancestral state reconstruction, median ratio of stable/Brownian reconstruction error, with data simulated on trees of various sizes under stable processes with varying indices of stability

α	Tree size (tips)									
	25	40	60	90	130	175	235	325	440	600
2.0	1.00	1.02	1.01	1.00	1.01	1.00	1.00	1.00	1.00	1.00
1.8	0.97	0.96	0.94	0.92	0.84	0.82	0.81	0.77	0.76	0.73
1.6	0.85	0.79	0.70	0.64	0.61	0.56	0.47	0.47	0.42	0.37
1.4	0.66	0.50	0.49	0.35	0.28	0.29	0.20	0.24	0.16	0.16
1.2	0.48	0.36	0.26	0.15	0.14	0.12	0.11	0.09	0.08	0.07
1.0	0.21	0.16	0.08	0.07	0.06	0.04	0.03	0.02	0.02	0.01

convergent evolution. The resulting patterns may be at odds with the neutral drift modelled by Brownian motion [4]. Indeed, studies of the performance of ancestral state reconstruction using known ancestral states derived from fossil estimates [6,7,70,71] or from taxa evolving sufficiently rapidly to be observed in real time [5] indicate that a mismatch between the stochastic model and historical reality can result in incorrect estimates [but see also [72]].

In this paper we have described a stochastic model of continuous character evolution based on a generalization of the Brownian model of evolution that does not assume that the rate of evolutionary change is constant and finite. Under these relaxed assumptions, the sum of increments accruing to an evolving character along each branch of a phylogeny is known to tend toward a stable limit distribution, which is identical to a normal distribution in the special case of Brownian motion but otherwise has heavier tails (Figure 1). These heavy tails allow rare evolutionary increments of large magnitude to occur, resulting in a volatile evolutionary process characterized by occasional “adaptive” evolutionary shifts interspersed with neutral-like patterns of variation (Figure 2). Stable parameters and ancestral states can be fit to biological data distributed over a phylogenetic tree using Markov chain Monte Carlo methods. We have implemented software that makes use of a slice sampler [40] to sample the posterior probability distribution of ancestral state assignments at each node and the values of stable parameters (the index of

stability α , which is equal to 2 under Brownian motion, and scale c). The slice sampler is able to take advantage of our knowledge of the approximate location of modal regions to move across a multi-modal likelihood surface without becoming trapped in locally but not globally optimal regions (Figure 5). An additional benefit of the slice sampler is its adaptive step size, requiring no tuning of proposal distributions, which makes practical application of the method straightforward. Our analysis of simulated data indicates that the Bayesian Predictive Information Criterion (BPIC) provides a conservative test of the hypothesis of departures from neutrality (i.e., the existence of heavy tails) in an evolutionary process (Table 1), and that the stable model estimates ancestral states with reduced error in comparison with the Brownian motion model, when traits evolve by accumulating increments from a probability distribution without constant finite variance (Table 2).

We found the stable model to fit the eutherian body-size data significantly better than the Brownian motion model (Table 4), suggesting the existence of departures from neutrality. Under the stable model, the ancestral eutherian is relatively small; in line with fossil evidence (also presented in Table 4), low body mass persists through early diversification of the Superorders Afrotheria, Euarchotheria and Laurasiatheria and the origin of various orders of small size such as Primates, Rodentia, Lagomorpha, Scandentia, Afrosoricida and Macroscelidea; large

Table 3 Sum of squared errors in ancestral state reconstruction, median ratio of stable/Ornstein-Uhlenbeck reconstruction error, with data simulated on trees of various sizes under stable processes with varying indices of stability

α	Tree size (tips)									
	25	40	60	90	130	175	235	325	440	600
2.0	0.36	0.35	0.35	0.33	0.43	0.52	0.54	0.71	0.75	0.83
1.8	0.34	0.33	0.36	0.36	0.38	0.44	0.50	0.48	0.56	0.47
1.6	0.31	0.28	0.29	0.31	0.31	0.46	0.27	0.24	0.27	0.24
1.4	0.28	0.24	0.25	0.18	0.16	0.17	0.13	0.14	0.11	0.11
1.2	0.23	0.17	0.16	0.11	0.09	0.08	0.07	0.06	0.06	0.04
1.0	0.17	0.11	0.08	0.06	0.04	0.03	0.03	0.02	0.02	0.01

Table 4 Ancestral state reconstruction of adult female body mass (kg) based on a dataset of 1,679 eutherian mammal species means, under Brownian Motion (BM; maximum likelihood), Ornstein-Uhlenbeck (OU; maximum likelihood), Early Burst (EB; maximum likelihood), Eastman *et al*'s (2011) heterogeneous multi-rate (maximum posterior probability) and stable (maximum posterior probability) models

Clade	Reconstructed body mass (kg)					Fossil estimates (kg)
	BM	OU	EB	Eastman	Stable	
Eutheria	1.24	1.23	1.14	0.97	0.33	0.015-0.017 (<i>Juramaia</i>) 0.02-0.025 (<i>Eomaia</i>) Shrew-sized (<i>Prokennalestes</i>)
Euarchoptoglires	0.72	0.43	0.73	0.60	0.34	0.0003 (<i>Tribosphenomys</i>)
Rodentia	0.37	0.26	0.39	0.34	0.24	0.0014 (<i>Microparamys</i>) 0.1-0.2 (Lagomorpha indet.)
Lagomorpha	0.53	0.36	0.56	0.50	0.37	0.1 (<i>Purgatorius</i>)
Primates	0.76	0.49	0.77	0.64	0.45	
Scandentia	0.25	0.15	0.27	0.25	0.25	
Laurasiatheria	1.13	0.61	1.10	0.60	0.25	
Cetartiodactyla	23.6	80.29	17.60	18.34	79.22	2.2-3.9 (<i>Diacodexidae</i>) 2.6-3.9 (<i>Homacodontidae</i>) 6.6-9.6 (<i>Laredochoerus</i>) 158.2 (<i>Hyrachyus</i>)
Perissodactyla	86.5	152.51	57.17	72.75	273.76	740.1 (<i>Amyrnodon</i>) 0.15-10 (<i>Viveravidae</i>)
Carnivora	6.34	8.77	5.45	5.23	15.20	1-10 (<i>Miacidae</i>)
Pholidota	4.61	5.54	4.48	4.47	4.28	
Chiroptera	0.08	0.12	0.09	0.02	0.02	0.012-0.015 (<i>Icaronycteris</i>)
Eulipotyphla	0.58	0.41	0.62	0.36	0.22	0.011 (<i>Paranyctoides</i>)
Xenarthra	3.04	3.13	2.60	2.8	2.15	
Pilosa	4.17	3.94	3.87	2.80	4.49	
Cingulata	3.30	3.36	3.12	3.23	2.40	
Afrotheria	1.73	8.95	1.50	1.51	0.36	
Afrosoricida	0.72	0.17	0.77	0.65	0.15	
Macroscelidea	0.30	0.44	0.34	0.65	0.17	
Hyracoidea	3.40	9.14	3.38	3.38	3.56	> 4 (<i>Heterohyrax</i>)
Proboscidea	1,478.15	949.76	1039.24	1,460.62	2,928.50	> 2,000 (<i>Palaeomastodon</i>)
Sirenia	90.33	204.84	53.41	87.24	341.12	488.1 (<i>Halitherium</i>)

Reconstructed values are provided for the most recent common ancestors of extant taxa in the specified clades. The right column details a selection of oldest fossil taxa within each clade for which body mass estimates are available [data from [56-65]].

reductions in body size are rare, occurring in Chiroptera, while large increases in body size occur with the origin of several modern Orders of relatively large species including the ungulates (Perissodactyla + Artiodactyla), Carnivora, Cetartiodactyla, Proboscidea and Sirenia. The Brownian motion model differs in several respects. First the Brownian motion reconstruction of the ancestral eutherian's body mass is an order of magnitude greater than the stable reconstruction; the Brownian motion reconstruction thus posits significant reductions of body size prior to the evolution of orders of small body size including Rodentia,

Scandentia, Chiroptera, Eulipotyphla and Macroscelidea. As expected, the Brownian motion model exhibits an "averaging effect" more generally, in which transformations in body mass are distributed somewhat evenly over the phylogeny, while the stable model permits large transformations in body mass to occur on a smaller subset of branches. For this reason, the ancestral state reconstruction of taxa ancestral to typically large species (i.e., Cetartiodactyla, Perissodactyla, Proboscidea) are smaller under the Brownian motion model, and the ancestral state reconstruction of taxa ancestral to typically small

species (i.e., Rodentia, Lagomorpha, Chiroptera, Eulipotyphla, Afrosoricida, and Macroscelidea) are larger under the Brownian motion model. The tendency for ancestral state reconstructions to be weighted toward intermediate values is stronger under the Brownian motion model than under the stable model since the former model vitiates against the inference of directional tendency.

A desire to model the adaptive evolution of continuous traits has given rise to a number of approaches that refine or extend the Brownian model. We categorize these as homogenous approaches, in which the stochastic process underlying the generation of evolutionary increments to an evolving character does not vary across branches of a phylogenetic tree, versus heterogenous processes, in which the stochastic process varies across branches. One popular homogenous model is the global Ornstein-Uhlenbeck model of trait evolution [73], in which the direction and rate of evolution at any time depends upon a selection coefficient and the degree of deviation of the trait's current value from some global optimum or "phylogenetic mean". This so-called "mean-reverting" process has been used as a model of stabilizing selection since deviation away from the phylogenetic mean is penalized under maximum likelihood reconstruction. Our simulation studies indicate that the stable model estimates ancestral states with reduced error in comparison to the homogenous Ornstein-Uhlenbeck model, when traits evolve by accumulating increments from a probability distribution without fixed variance (Table 3). The homogenous Ornstein-Uhlenbeck reconstruction of mammalian body mass (Table 4) is in some respects intermediate between the Brownian motion and stable reconstructions, with relatively small ancestors of Orders with small body size and relatively large ancestors of Orders with large body size. This can be interpreted in terms of the stabilizing selection model with an intermediate phylogenetic mean: the proposal of a small ancestral rodent (for example) permits a tendency to evolve back toward the mean within the rodent clade, while the proposal of a large ancestral carnivore (for example) permits the same tendency in the opposite direction within the carnivore clade. This pattern is most striking in the ancestral state inferred for Afrotheria (8.95 kg versus 1.73 kg under Brownian motion and 0.36 kg under the stable model), where a large ancestor reduces the rate of evolution on branches leading ultimately to the large elephants and manatees while permitting many high-likelihood reductions of size toward the phylogenetic mean in small taxa such as Afrosoricida and Macroscelidea. This reconstruction for Afrotheria seems unlikely and may lead us to suppose that a single phylogenetic mean for the entire Eutheria does not form a realistic model of stabilizing selection.

Cooper and Purvis [55] report success in fitting more complex heterogenous models to a larger set of

mammalian body mass data. The Ornstein-Uhlenbeck model is easily extended to the heterogenous case by permitting more than one clade-specific phylogenetic means [51,73-75], the number and phylogenetic position of such means being specified *a priori* or estimated from the data. Various transformations of the phylogeny, such as raising all branch lengths to a constant power in order to approximate speciation change [67] or somewhat *ad hoc* transformation of branch lengths to maximize the likelihood of a Brownian model [12,76,77] have also been used to generate implicitly heterogenous models. The Early Burst model [14,52], also applied by Cooper and Purvis [55], is interesting in generating rate heterogeneity by allowing the rate of a Brownian motion process to vary over time, rather than across branches or clades. The rate of evolution is taken to be an exponentially increasing or decreasing function of node height, allowing a greater proportion of evolutionary change to occur early in the phylogeny or late in the phylogeny depending on the choice of an exponential scaling factor r . We applied the Early Burst model to our mammalian body mass dataset but did not identify a significant deviation from $r = 0$; ancestral state estimates in Table 4 are derived from a reconstruction based on $r = -0.009$ estimated by Cooper and Purvis [55]. The concentration of more evolutionary change in basal branches of the phylogeny appears to allow rapid early deviation from the phylogenetic mean value, with the majority of ancestral taxa exhibiting marginally smaller body sizes, often more consistent with the fossil evidence, than under the Brownian motion reconstruction, yet surprisingly with considerably less ordinal-level diversification than imputed by the stable model which does not build early diversification in to the stochastic process itself.

Homogenous approaches offer a number of advantages over heterogenous approaches in that the latter category must not only infer the parameters of the stochastic process but also must infer the structure of rate heterogeneity over the phylogeny. Especially when heterogeneity is associated with clades, over-fitting heterogenous models by positing too many rate shifts or clade-specific evolutionary regimes may become a danger. Eastman *et al.* [53] have proposed a heterogenous model of trait evolution that explicitly avoids over-fitting by sampling over model parameter values and the number of model parameters simultaneously. The model is an extension of the standard Brownian motion model in which the rate of evolution is "inherited" over time but may undergo occasional shifts in value. Each shift introduces a new parameter that is penalized in a reversible jump Markov chain Monte Carlo algorithm. We found that the complexity of the reversible jump algorithm considerably increases the computational burden of fitting the model: for the body mass data considered here, the stable slice sampler accomplished around

70,000 steps per minute on a modest dual-core home laptop, versus around 2,500 per minute for the multi-rate model, without the need for a lengthy calibration of proposal densities beforehand. Ancestral states reconstructed under the Eastman model (Table 4) were in broad agreement with other non-Brownian methods presented here in imputing smaller early mammals, and were in close agreement with results of the Early Burst analysis.

In general, the stable model suggests a greater degree of ordinal-level diversification of mammalian body masses, and appears to accommodate a more volatile evolutionary process, than any of the other models considered here. For 13 of the 22 nodes listed in Table 4 the stable model reconstructs the smallest body masses of any model, and for 4 nodes it reconstructs the largest body mass of any model, making the stable reconstruction consistent with the hypothesis of small early mammals and occasional marked ordinal-level enlargement. The ability of the stable model to accommodate striking variation in evolutionary rate, even more so than approaches such as that of Eastman *et al.* [53] explicitly designed to model such variation, is most apparent in the highly diverse and species-poor Afrotheria, where the stable reconstruction involves the largest ancestral elephants and manatees yet the smallest Afroinsectivores of any of the methods considered here. While the heterogenous Ornstein-Uhlenbeck model binds rate volatility to the structure of the phylogeny through the assumption of clade-specific phylogenetic means, and the Eastman *et al.* RJ-MCMC model binds rate volatility to the structure of the phylogeny through the “inheritance” of rate shifts from ancestral to descendant branches, the stable model through its homogeneously heavy tails provides unstructured volatility that is able to concentrate the production of evolutionary variation onto relatively few branches scattered across the phylogeny.

The stable model is the simplest non-Brownian model considered here, requiring only a single parameter in addition to the standard Brownian motion model. The relative efficiency of the estimation procedure used to fit the stable model may make it attractive for analysis of very large trees or large sets of trees derived from Bayesian phylogenetics. Furthermore, deviation from the Brownian model according to the BPIC criterion may be used to provide independent statistical support for the adoption of one of the more complex heterogenous models currently available. Rates imputed by the stable model may guide appropriate selection of branches for independent rates in such cases. In the mammal body mass data examined here, for example, the frequency distribution of standardized trait changes along branches of the phylogeny (reported by the accompanying software) indicates accelerated evolution at the origin of a number of clades including *Hyomys* (white-eared giant rats), Tragulidae (mouse deer), Manidae (pangolins), Megachiroptera

(megabats), Megadermatidae (false vampire bats), Solenodontidae (solenodons), Orycteropodidae (armadillo) and Hyracoidea (hyraxes), suggesting that these clades may merit their own phylogenetic mean values under a heterogenous Ornstein-Uhlenbeck approach. In order to determine whether such a model is useful in any particular case it is necessary, as with any stochastic model of evolution, to rigorously constrain the model empirically, and while the results presented here are primarily illustrative and to provide comparison across unconstrained models, we note that the low ancestral state inferences for extinct taxa at the root of Rodentia, Lagomorpha, Primates, Chiroptera and Lipotyphla, and the high ancestral state inferences for taxa at the root of Sirenia and Elephantidae, appear broadly in line with fossil evidence.

While our homogenous approach may be associated with some advantages with respect to efficiency, simplicity and unstructured volatility, the heterogenous models such as Early Burst have the benefit of imposing an explicit evolutionary narrative on the process of trait diversification which may be useful for exploring and testing general hypotheses about historical processes [14]. Heterogenous models typically involve the elaboration of a simple Gaussian kernel to accommodate phylogeny- or time-structured variation in the evolutionary process. We suggest that in future work heterogenous stable models analogous to those considered above may be readily generated by directly replacing the Gaussian kernel with the more general stable kernel, at the expense of a single parameter. Stable Ornstein-Uhlenbeck processes, for example, are already well-characterized [17]. The stable model we introduce to phylogenetic evolutionary biology here may find other uses, for example in assigning substitution rates to edges on phylogenetic trees under relaxed clock models [78]. One primary obstacle to the replacement of Gaussian kernels by stable kernels in models of continuous character evolution is that stable distributions have undefined variance [17]. Methods making direct use of variance are typically used to detect correlated evolution between multiple continuous characters evolving on the same phylogenetic tree [2]. Independent contrasts [79] for example, generates standardized data points for each univariate character by scaling the increments accruing along paired branches of a phylogeny by the square root of the sum of branch lengths, which is proportional to the expected standard deviation of a Brownian process. Methods of phylogenetic regression [12,80,81] extend least squares methods to multivariate phylogenetic data by incorporating branch length and topological information into the model's covariance matrix. The fact that stable variance is undefined means that there is no stable equivalent to standard deviation or the covariance matrix. We note that regression and correlation models based

on stochastic processes driven by non-Gaussian stable perturbations have been implemented successfully in non-phylogenetic fields [i.e., [82-84]]. These approaches raise the prospect that likelihood-based analysis of heavy tailed multivariate distributions may offer useful insights into future studies of correlated evolution of multiple continuous characters in evolutionary biology, since correlated evolution is precisely the kind of problem domain in which the putative Brownian assumptions of neutrality and gradualism are likely to be invalid.

Conclusions

Stochastic process models of evolution regard an evolving trait as accumulating, over time, random increments drawn from some underlying probability distribution. We have described a generalization of the Brownian motion model in which the increment-generating function is a stable distribution characterized by heavy tails, which accommodates both the neutral drift associated with Brownian motion but also occasional burst of rapid evolutionary change. Simulation and empirical studies indicate that stable models can successfully be fit to biological data, and Bayesian model selection criteria can be used to assess goodness of fit in comparison with the Brownian motion model, which is a special case of the more general symmetrical stable distribution. The model presented in this paper is a homogenous model in which a single stochastic process, common to all branches of a phylogeny, gives rise to increments to evolving continuous traits. The approach may be contrasted with heterogenous models in which different evolutionary regimes are bound to different subtrees of a phylogeny, or arise stochastically across the branches of a phylogeny. While homogenous models offer simplicity and computational convenience, it is currently unclear whether such models – and even more highly parameterized heterogenous models of trait evolution – are capable of capturing adequately the richness and complexity of evolutionary processes in nature. We have made an empirical attempt to corroborate results from the stable model on the basis of published fossil data on extinct mammalian body masses. The various homogenous and heterogenous models are consistent in some respects but also exhibit marked differences in reconstructed ancestral states. A major line of future research should be to expand the availability of fossil and other historical data that would facilitate the empirical measurement of the distribution of evolutionary changes over time for known traits and phylogenetic trees. In general, we believe it is likely to be the case that models of continuous trait evolution should be tailored specifically for the empirical question at hand. The present research suggests that models of evolution incorporating heavy tails and volatile stochastic processes may be a useful addition to the toolset of biologists interested in traits exhibiting

heterogenous patterns of diversification driven by adaptive evolution.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

ME and AOE conceived of the model, designed simulation studies and co-authored the paper. ME wrote the software and conducted the analyses. All authors read and approved the final manuscript.

Acknowledgements

The authors thank Gavin Thomas for providing invaluable criticism of the model and performance of the software implementation for large datasets, Tomohiro Ando for providing material on BPIC, and members of the UBC and SFU evolution groups for their advice on an early version of this manuscript. This work was enabled by computational resources provided by Compute/Calcul Canada and the kind support of Brian Corrie at IRMACS. ME was financially supported by NSERC and the Human Evolutionary Studies Program at SFU.

Author details

¹St. John's College, University of Cambridge, Cambridge CB2 1TP, UK.

²Department of Biological Sciences and Human Evolutionary Studies Program, Simon Fraser University, 8888 University Drive, V5A 1S6 Burnaby, BC, Canada.

Received: 3 March 2014 Accepted: 17 October 2014

Published online: 28 November 2014

References

1. Harvey P, Pagel M: *The Comparative Method in Evolutionary Biology*. Oxford Series in Ecology and Evolution. Oxford, UK: Oxford University Press; 1991.
2. Martins E (Ed): *Phylogenies and the Comparative Method in Animal Behavior*, Oxford, UK: Oxford University Press; 1996.
3. Pagel M, Harvey P: **Comparative methods for examining adaptation depend on evolutionary models**. *Folia Primatol* 1989, **53**:203–220.
4. Cunninham C: **Some limitations of ancestral character-state reconstruction when testing evolutionary hypotheses**. *Syst Biol* 1999, **48**:665–674.
5. Oakley T, Cunninham C: **Independent contrasts succeed where ancestor reconstruction fails in a known bacteriophage phylogeny**. *Evolution* 2000, **54**:397–405.
6. Finarelli J, Flynn J: **Ancestral state reconstruction of body size in the Caniformia (Carnivora, Mammalia): the effects of incorporating data from the fossil record**. *Syst Biol* 2006, **55**:301–313.
7. Slater G, Harmon L, Alfaro M: **Integrating fossils with molecular phylogenies improves inference of trait evolution**. *Evolution* 2012, **66**:3931–3944.
8. Cavalli-Sforza L, Edwards A: **Phylogenetic analysis models and estimation procedures**. *Am J Hum Genet* 1967, **19**:233–257.
9. Maddison W: **Squared-change parsimony reconstructions of ancestral states for continuous-valued characters on a phylogenetic tree**. *Syst Zool* 1991, **40**(3):304–14.
10. McArdle B, Rodrigo A: **Estimating the ancestral state of a continuous-valued character using squared-change parsimony: an analytical solution**. *Syst Biol* 1994, **43**:573–578.
11. Schluter D, Price T, Mooers A, Ludwig D: **Likelihood of ancestor states in adaptive radiation**. *Evolution* 1997, **51**(6):1699–1711.
12. Grafen A: **The phylogenetic regression**. *Phil Trans Roy Soc Lond B* 1989, **326**:265–292.
13. Harmon L, Schulte J, Larson A, Losos J: **Tempo and mode of evolutionary radiation in iguanian lizards**. *Science* 2003, **301**:2385–2396.
14. Harmon L, Losos J, Davies T, Gillespie R, Gittleman J, Jennings W, Kozak K, McPeck M, Moreno-Roark F, Near T, Purvis A, Ricklefs R, Schluter D, Schulte J, Seehausen O, Sidlauskas B, Torres-Carvajal O, Weir J, Mooers A: **Early bursts of body size and shape evolution are rare in comparative data**. *Evolution* 2010, **64**:2385–2396.

15. O'Meara B, Ane C, Sanderson M, Wainwright P: **Testing for different rates of continuous trait evolution using likelihood.** *Evolution* 2006, **60**:922–933.
16. Gnedenko B, Kolmogorov A: *Limit Distributions for Sums of Independent Random Variables.* Addison-Wesley Series in Statistics. Reading MA: Addison-Wesley; 1949.
17. Samorodnitsky G, Taqqu M: *Stable Non-Gaussian Random Processes.* New York: Chapman and Hall; 1994.
18. Uchaikin V, Zolotarev V: *Chance and Stability: Stable Distributions and their Applications.* Utrecht: VSP; 1999.
19. Nolan J: **Numerical calculation of stable densities and distribution functions.** *Stoch Mod* 1997, **13**:759–774.
20. Gorenflo R, Mainardi F, Moretti D, Pagnini G, Paradisi P: **Fractional diffusion: probability distributions and random walk models.** *Physica A* 2002, **305**:106–112.
21. Lavallée D: **Stochastic modeling of slip spatial complexities for the 1979 Imperial Valley, California, earthquake.** *Geophys Res Lett* 2003, **30**:1–4.
22. Nikias C: *Signal Processing with Alpha-Stable Distributions and Applications.* New York: Wiley-Blackwell; 1997.
23. Wiswanathan G, Afanasyev V, Buldyrev S, Murphy E, Prince P, Stanley H: **Lévy flight search patterns of wandering albatrosses.** *Nature* 1996, **381**:413–415.
24. Menabde M, Sivapalan M: **Modeling of rainfall time series and extremes, using bounded random cascades and Lévy-stable distributions.** *Water Resour Res* 2000, **36**:3293–3300.
25. Weron R: *Modeling and forecasting electricity loads and prices: a statistical approach.* New York: John Wiley; 2006.
26. Young M: **Revisiting non-normal real estate return distributions by property type in the U.S.** *J Real Estate Financ Econ* 2008, **36**:233–248.
27. Fofack H, Nolan J: **Distribution of parallel exchange rates in African countries.** *J Int Money Financ* 2001, **20**:987–1001.
28. Rachev S (Ed): *Handbook of Heavy Tailed Distributions in Finance.* Handbooks in Finance, Amsterdam: North-Holland; 2003.
29. Arce G: *Nonlinear Signal Processing.* New York: Wiley; 2005.
30. Mikosch T, Resnick S, Rootzén H, Stegeman A: **Is network traffic approximated by stable Lévy motion or fractional brownian motion?** *Ann Appl Probab* 2002, **12**:23–68.
31. Landis M, Schraiber J, Liang M: **Phylogenetic analysis using Lévy processes: finding jumps in the evolution of continuous traits.** *Syst Biol* 2013, **62**:193–204.
32. Devroye L: *Non-Uniform Random Variate Generation.* New York: Springer Verlag; 1986.
33. Matsui M, Takemura A: **Improvements in numerical evaluation of symmetric stable density and its derivatives.** *CIRJE Discuss Papers* 2004, **F292**:1–20.
34. Brorsen B, Yang S: **Maximum likelihood estimates of symmetric stable distribution parameters.** *Commun Stat Simul Comp* 1990, **19**:1459–1464.
35. McCulloch J: *A Practical Guide to Heavy Tails: Statistical Techniques and Applications.* Cambridge, MA: Birkhäuser Boston; 1998. chap. Numerical approximation of the symmetric stable distribution and density.
36. Nolan J: *Lévy Processes: Theory and Applications.* Cambridge, MA: Birkhäuser Boston; 2001. chap. Maximum likelihood estimation and diagnostics for stable distributions.
37. Ciuperca G, Ridolfi A, Idris J: **Penalized maximum likelihood estimator for normal mixtures.** *Scan J Stat* 2003, **30**:45–59.
38. Metropolis M, Rosenbluth A, Rosenbluth M, Teller A, Teller E: **Equation of state calculations by fast computing machines.** *J Chem Phys* 1953, **21**:1087–1092.
39. Geyer C: *Computing Science and Statistics: Proceedings of the 23rd Symposium of the Interface.* New York: American Statistical Association; 1991. chap. Marko chain Monte Carlo maximum likelihood.
40. Neal R: **Slice sampling.** *Ann Stat* 2003, **31**:705–767.
41. McLachlan G: *Finite mixture models.* New York: Wiley; 2000.
42. Akaike H: **A new look at the statistical model identification.** *IEEE Trans Automat Contr* 1974, **19**:716–723.
43. Spiegelhalter D, Best N, Carlin B, van der Linde A: **Bayesian measures of model complexity and fit (with discussion).** *J Roy Stat Soc B* 2002, **64**:583–616.
44. Ando T, Tsay R: **Predictive likelihood for Bayesian model selection and averaging.** *Int J Forecast* 2010, **26**:744–763.
45. Rambaut A, Drummond A: **Tracer v1.4, Available from.** <http://beast.bio.ed.ac.uk/Tracer>
46. Brooks S, Gelman A: **General methods for monitoring convergence of iterative simulation.** *J Comp Graph Stat* 1998, **7**:434–455.
47. Maddison W, Maddison D: **Mesquite: a modular system for evolutionary analysis, version 2.6.** 2011. <http://www.mesquiteproject.org>.
48. Bininda-Emonds O, Cardillo M, Jones K, MacPhee R, Beck R, Grenyer R, Price S, Vos R, Gittleman J, Purvis A: **The delayed rise of present-day mammals.** *Nature* 2007, **446**:507–512.
49. Ernest S: **Life history characteristics of placental nonvolant mammals.** *Ecology* 2003, **84**:3402.
50. Pitnick S, Jones K, Wilkinson G: **Mating system and brain size in bats.** *Proc R Soc Lond B* 2006, **273**:719–724.
51. Hansen T, Pienaar J, Orzack S: **A comparative method for studying adaptation to a randomly evolving environment.** *Evolution* 2008, **62**:1965–1977.
52. Blomberg SP, Jr, Garland T, Ives AR: **Testing for phylogenetic signal in comparative data: behavioral traits are more labile.** *Evolution* 2003, **57**:717–745.
53. Eastman J, Alfaro M, Joyce P, Hipp A, Harmon L: **A novel comparative method for identifying shifts in the rate of character evolution on trees.** *Evolution* 2011, **65**:3578–3589.
54. Harmon J, Weir JT, Brock C, Glor RE, Challengier W: **GEIGER: Investigating evolutionary radiations.** *Bioinformatics* 2008, **24**:129–131.
55. Cooper N, Purvis A: **Body size evolution in mammals: complexity in tempo and mode.** *Am Nat* 2010, **175**:727–738.
56. Alroy J, Crampton J, Erwin D, Kiessling W, Miller A, Peters A, Polly D: **The Paleobiology Database.** 2011. Published online at <http://www.paleodb.org>, accessed June 2011.
57. Fleagle J: *Primate Adaptation and Evolution.* New York: Academic Press; 1999.
58. Goswami A, Friscia A (Eds): *Carnivoran Evolution: New Views on Phylogeny, Form and Function,* Cambridge, UK: Cambridge University Press; 2010.
59. Ji Q, Luo Z, Yuan C, Wible J, Zhang J, Georgi J: **The earliest known eutherian mammal.** *Nature* 2002, **416**:816–822.
60. Luo Z, Yuan C, Meng Q, Ji Q: **A Jurassic eutherian mammal and divergence of marsupials and placentals.** *Nature* 2011, **476**:443–445.
61. Rasmussen D, Pickford M, Mein P, Senut B, Conroy G: **Earliest known Procaviid Hyracoid from the late Miocene of Namibia.** *J Mammal* 1996, **77**:745–754.
62. Rose K: *The Beginning of the Age of Mammals.* Baltimore: Johns Hopkins University Press; 2006.
63. Rose K, DeLeon V, Missiaen P, Rana R, Sahni A, Singh L, Smith T: **Early Eocene lagomorph (Mammalia) from Western India and the early diversification of Lagomorpha.** *Proc Biol Sci* 2008, **275**:1203–1208.
64. Rydell J, Speakman J: **Evolution of nocturnality in bats: potential competitors and predators during their early history.** *Biol J Linn Soc Lond* 1995, **54**:183–191.
65. Wood J: **The extinction of the Multituberculates outside North America: a global approach to testing the competition model.** *Master's thesis.* Ohio State University 2010.
66. Coddington J: **Cladistic tests of adadaptation hypotheses.** *Cladistics* 1988, **4**:3–22.
67. Pagel M: **Inferring the historical patterns of biological evolution.** *Nature* 1999, **401**:877–884.
68. Martins E: **Adaptation and the comparative method.** *Trends Ecol Evol* 2000, **15**:296–299.
69. Pagel M, Meade A, Barker D: **Bayesian estimation of ancestral character states on phylogenies.** *Syst Biol* 2004, **53**:673–684.
70. Donoghue M, Doyle J, Gauthier J, Kluge A, Rowe T: **The importance of fossils in phylogeny reconstruction.** *Ann Rev Ecol Systemat* 1989, **20**:431–460.
71. Webster A, Purvis A: **Testing the accuracy of methods for reconstructing ancestral states of continuous characters.** *Proc Biol Sci* 2002, **269**:143–149.
72. Polly P: **Paleontology and the comparative method: ancestral node reconstructions versus observed node values.** *Am Nat* 2001, **157**:596–609.
73. Hansen T: **Stabilizing selection and the comparative analysis of adaptation.** *Evolution* 1997, **51**:1341–1351.

74. Butler M, King A: **Phylogenetic comparative analysis: a modeling approach for adaptive evolution.** *Am Nat* 2004, **164**:233–257.
75. Beaulieu J, Jhwueng D, Boettiger C, O'Meara B: **Modeling stabilizing selection: expanding the Ornstein-Uhlenbeck model of adaptive evolution.** *Evolution* 2012, **66**:2369–2383.
76. Gittleman J, Kot M: **Adaptation: statistics and a null model for estimating phylogenetic effects.** *Syst Biol* 1990, **39**:227–241.
77. Garland T, Harvey P, Ives A: **Procedures for the analysis of comparative data using phylogenetically independent contrasts.** *Syst Biol* 1992, **41**:18–32.
78. Thorne J, Kishino H, Painter I: **Estimating the rate of evolution of the rate of molecular evolution.** *Mol Biol Evol* 1998, **15**:1647–1657.
79. Felsenstein J: **Phylogenies and the comparative method.** *Am Nat* 1985, **125**:1–15.
80. Garland T, Dickerman A, Janis C, Jones J: **Phylogenetic analysis of covariance by computer simulation.** *Syst Biol* 1993, **42**:265–292.
81. Martins EP, Hansen TF: **Phylogenies and the comparative method: a general approach to incorporating phylogenetic information into the analysis of interspecific data.** *Am Nat* 149:646–667.
82. McCulloch J: *A Practical Guide to Heavy Tails: Statistical Techniques and Applications.* Cambridge, MA: Birkhäuser Boston; 1998. chap. Linear regression with stable disturbances.
83. Frain J: **Maximum likelihood estimates of regression coefficients with alpha-stable residuals and day of week effects in total returns on equity indices.** *Trinity Econ Papers* 2008, **tep0108**:1–27.
84. Paulauskas V, Rachev S: **Maximum likelihood estimators in regression models with infinite variance innovations.** *Stat Pap* 2003, **44**:47–65.

doi:10.1186/s12862-014-0226-8

Cite this article as: Elliot and Mooers: Inferring ancestral states without assuming neutrality or gradualism using a stable model of continuous character evolution. *BMC Evolutionary Biology* 2014 **14**:226.

Submit your next manuscript to BioMed Central
and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

