

Quantitative Traits and Diversification

RICHARD G. FITZJOHN*

Department of Zoology and Biodiversity Research Centre, University of British Columbia, Vancouver, BC V6T 1Z4, Canada;

*Correspondence to be sent to: *Department of Zoology, University of British Columbia, Vancouver, BC V6T 1Z4, Canada; E-mail: fitzjohn@zoology.ubc.ca.*

Received 24 December 2009; reviews returned 28 February 2010; accepted 9 August 2010

Guest Associate Editor: Daniel Rabosky

Abstract.—Quantitative traits have long been hypothesized to affect speciation and extinction rates. For example, smaller body size or increased specialization may be associated with increased rates of diversification. Here, I present a phylogenetic likelihood-based method (quantitative state speciation and extinction [QuaSSE]) that can be used to test such hypotheses using extant character distributions. This approach assumes that diversification follows a birth–death process where speciation and extinction rates may vary with one or more traits that evolve under a diffusion model. Speciation and extinction rates may be arbitrary functions of the character state, allowing much flexibility in testing models of trait-dependent diversification. I test the approach using simulated phylogenies and show that a known relationship between speciation and a quantitative character could be recovered in up to 80% of the cases on large trees (500 species). Consistent with other approaches, detecting shifts in diversification due to differences in extinction rates was harder than when due to differences in speciation rates. Finally, I demonstrate the application of QuaSSE to investigate the correlation between body size and diversification in primates, concluding that clade-specific differences in diversification may be more important than size-dependent diversification in shaping the patterns of diversity within this group. [Birth–death process; Cope’s rule; extinction; macroevolution; speciation.]

Species selection may be responsible for much of the variation in diversity among clades. Species fulfill the (Lewontin 1970) definition of units of selection; different species have different traits, differential “fitness” (rates of speciation and extinction) that may be attributable to these traits, and the traits and their fitness differences are heritable. Although the concept of species selection has been controversial since its inception (e.g., Stanley 1975; Vrba and Gould 1986), there now seems to be reasonable agreement that the process may operate (recently discussed in Okasha 2006; Jablonski 2008; Rabosky and McCune 2009). Many traits have been proposed to affect rates of speciation and extinction, such as body size (Gittleman and Purvis 1998), sexual system (Heilbut 2000), and dispersal ability (Phillimore et al. 2006). In addition, hypotheses that invoke “key innovations” (e.g., floral nectar spurs; Hodges and Arnold 1995) or evolutionary “dead ends” (e.g., asexuality; Schwander and Crespi 2009) generally invoke trait-dependent variation in rates of diversification.

Phylogenies contain information about the timings of speciation events and patterns of diversification (Nee et al. 1994) and have been used extensively in comparative analyses to attempt to identify correlates of elevated speciation or extinction rates. Sister clade analyses have been widely used for detecting correlates of diversification for binary traits (e.g., Mitter et al. 1988; Heilbut 2000; Vamosi and Vamosi 2005). These require that clades are characterized by a single character state and assume that all lineages within the clade have taken this value for the majority of its evolutionary history. More recently, likelihood approaches such as binary state speciation and extinction (BiSSE; Maddison et al. 2007) have allowed this assumption to be relaxed, allowing any distribution of characters among extant

species, by using the entire pattern of branching in a phylogeny. For example, BiSSE has recently been used to demonstrate a correlation between live-bearing (vs. egg-laying) snake species and elevated speciation rates (Lynch 2009).

Sister clade analyses will not generally be appropriate for detecting correlation between diversification rates and continuous traits because a clade does not have a single body size, geographic range, or latitude (but see Gittleman and Purvis 1998, for an approach that uses mean clade traits). Several recent methods have been developed explicitly for continuous traits. Clauset and Erwin (2008) used a diffusion model to calculate equilibrium trait frequency distributions under a model where species selection is opposed by individual-level selection. However, this approach cannot incorporate phylogenetic information. The methods developed by Paradis (2005) and Freckleton et al. (2008) are explicitly phylogenetic, but they assume that ancestral character states can be estimated without accounting for the effect of the character on speciation and extinction (see below). None of these methods can distinguish between differential speciation and differential extinction. However, speciation and extinction rates may be correlated (high extinction rates may accompany high speciation rates; e.g., Gilinsky 1994; Coyne and Orr 2004; Liow et al. 2008), and many traits are thought to change diversification rates through their effect on extinction rather than speciation (e.g., Harcourt et al. 2002; Cardillo et al. 2005). Approaches that allow differential speciation to be distinguished from differential extinction will therefore allow testing of a broader array of evolutionary hypotheses.

Inferring the states of ancestral nodes is problematic when the character affects speciation or extinction

(Maddison 2006; Paradis 2008). For example, trait values might be inferred at ancestral nodes that are unlikely because they would lead to high rates of extinction if the processes of character evolution and speciation/extinction are treated separately. More subtly, the trait values estimated in this way will be incorrect if the evolution of the trait has a directional tendency, as it is not possible to detect directional shifts in a character under a Brownian motion model of character evolution on an ultrametric tree (Schluter et al. 1997). However, where speciation, extinction, and character evolution are treated simultaneously, and where speciation rates are state dependent, these directional changes can potentially be inferred. This would allow detection of cases where species selection and individual-level selection oppose (i.e., directional character change toward trait values that are disfavored by species selection). For example, it has been suggested that large-bodied individuals tend to have higher fitness (leading to an increase in mean size over time), whereas populations of large-bodied individuals are more prone to extinction (Clauset and Erwin 2008).

Here, I describe a new comparative phylogenetic method, quantitative state speciation and extinction (QuaSSE), for inferring the effect of quantitative traits on speciation and extinction rates. I first derive likelihood equations that can be used to calculate the probability of a phylogenetic tree and distribution of character states among species under a general model of cladogenesis and character evolution. I then investigate the power of this method to detect differential diversification by applying it to simulated trees. Finally, I demonstrate the method and illustrate some potential pitfalls by investigating the correlation between body size and diversification in primates.

CHARACTER EVOLUTION AND DIVERSIFICATION

I model speciation and extinction as a birth–death process (similar to Nee et al. 1994), allowing the rates of speciation and extinction to vary with a simultaneously evolving character. Assume that a species can be characterized by its mean value of some character trait, x , which varies on the interval $(-\infty, \infty)$, and that this character affects diversification through its effect on the rate of speciation or extinction (or both). Let the rate of speciation for a lineage in state x be $\lambda(x)$ and the rate of extinction be $\mu(x)$. These may be arbitrary nonnegative functions of x , and I do not assume anything about their form. In the most general case, these can also be functions of time, so that the speciation rate for a lineage in state x at time t is $\lambda(x, t)$, but for notational brevity, I will omit this time dependence. Incorporating time dependence allows modeling of clade-wide changes in diversification rates (e.g., Rabosky 2006).

It is convenient to model character evolution along lineages using a diffusion process (Allen 2003). Diffusion processes are attractive for modeling character evolution because they allow for stochasticity while being

mathematically tractable. I will measure time backward, with the present at time 0, and $t > 0$ representing some time in the past. Let $g(z, t|x, t + \Delta t)$ be the transition probability density function for the diffusion process; the probability density that a character state changes from x at time $t + \Delta t$ to state z at time t , where t is closer to the present than $t + \Delta t$ ($0 < t < t + \Delta t$). The diffusion assumptions state that

$$\phi(x, t) = \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} \int_{-\infty}^{\infty} (z - x)g(z, t|x, t + \Delta t)dz, \quad (1a)$$

$$\sigma^2(x, t) = \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} \int_{-\infty}^{\infty} (z - x)^2 g(z, t|x, t + \Delta t)dz, \quad (1b)$$

$$0 = \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} \int_{-\infty}^{\infty} (z - x)^k g(z, t|x, t + \Delta t)dz, \quad k > 2, \quad (1c)$$

where the integral is taken over all possible character transitions (Allen 2003). I will refer to $\phi(x, t)$ as the “directional” term, which captures the deterministic or directional component of character evolution; this is the expected rate of change of the character over time and may be due to selection or other any other within-lineage process that has a directional tendency. This term is typically referred to as the “drift” term (e.g., Allen 2003), but I avoid this terminology to prevent confusion with genetic drift. The term $\sigma^2(x, t)$ is the “diffusion” term and is the expected squared rate of change; this captures the stochastic elements of character evolution. The condition (1c) formally captures the assumption that large changes are unlikely by asserting that character evolution is described entirely the first two moments of the transition probability density function. Note that both $\phi(x, t)$ and $\sigma^2(x, t)$ may be functions of both the character state and time. I assume that the character state is perfectly inherited by both daughter species during speciation (e.g., speciation does not lead to character displacement).

The above diffusion process generalizes other models of character evolution. Brownian motion can be modeled by setting the functions $\phi(x, t)$ and $\sigma^2(x, t)$ to the constants ϕ and σ^2 . Where $\phi = 0$, this is standard Brownian motion, and where ϕ is nonzero, this is Brownian motion with a directional tendency (Felsenstein 1988). The Ornstein–Uhlenbeck process captures stabilizing selection that pulls the a character toward a long-term mean \hat{x} (Hansen and Martins 1996). This can be modeled by setting the directional term, $\phi(x, t)$, to the linear function $\alpha(\hat{x} - x)$, where α is the strength of this stabilizing force and setting the diffusion term, $\sigma^2(x, t)$, to the constant σ^2 . Note that when the model of character evolution is Brownian motion (with no directional tendency), this birth–death–diffusion process is essentially that described by Paradis (2005) and Freckleton et al. (2008) and by Slatkin (1981) and Clauset and Erwin (2008) but disallowing character changes at nodes.

Before describing the approach, it is worth emphasizing some limitations that follow from the above

assumptions. The birth–death process leads to an exponential growth in the number of species (at rate $\lambda - \mu$ when these rates are character independent; Nee et al. 1994), which is clearly not sustainable indefinitely. In addition, no interaction is possible between any of the lineages in the phylogeny; the rates of speciation, extinction, and character evolution cannot depend on the number of extant species or the character states of those species. This prevents modeling of density-dependent diversification (Phillimore and Price 2008) or frequency-dependent character evolution.

LIKELIHOOD CALCULATIONS

In this section, I will derive equations to compute the probability of a phylogenetic tree and character state distribution among extant species under the above model of character evolution and character-dependent speciation and extinction. I will assume that the calculations are carried out on a single ultrametric phylogenetic tree that has branch lengths proportional to time. It is straightforward to extend this analysis to integrate over a family of trees (e.g., bootstrapped trees or samples from a Bayesian analysis). I also assume that the tree is complete and fully resolved; that is, that it includes every extant species above a common ancestor and contains no polytomies. Later, I will relax this assumption slightly to allow for partial taxon sampling.

The calculations follow the same general structure as those of BiSSE (Maddison et al. 2007). Following the notation of BiSSE, let $E(x, t)$ be the probability that a lineage in state x at time t goes completely extinct, leaving no descendants by the present (time 0). This is a continuous function in both trait space and time in contrast to the analogous quantities in BiSSE that were continuous only in time. Similarly, let $D_N(x, t)$ be the probability that a lineage in state x at time t would evolve into the extant clade N as observed, including branch lengths and present-day character states. The subscript N denotes that this function applies to a particular lineage N .

Probability of Extinction

Assume that we know the function $E(x, t)$ at some time t in the past. If we can express E at a time immediately prior to this, $t + \Delta t$, in terms of its values at time t , then we can continue to do this until reaching the origin of a branch (Felsenstein 1981). To do this, consider all the events that could occur over a very short period of time, Δt , and write $E(x, t + \Delta t)$ by multiplying the probability of each event happening by the probability of extinction given that a particular event happened (Fig. 1). I assume that in this small period of time, at most one speciation or extinction event may occur (specifically, I assume that the probability of two or more events occurring is of order $(\Delta t)^2$ and therefore negligible with sufficiently small Δt). Over this period of time, there are three possibilities: 1) the lineage goes extinct with probability

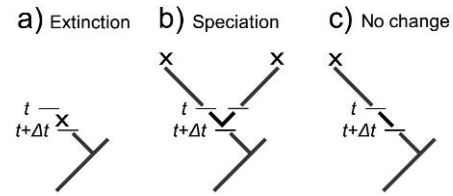


FIGURE 1. Possible ways a lineage extant at time $t + \Delta t$ might go extinct. If at most a single lineage-changing event occurs, then a) extinction happens with probability $\mu(x)\Delta t$, leading to total extinction with probability 1, b) a speciation event happens with probability $\lambda(x)(1 - \mu(x))\Delta t$, leading to total extinction with probability $E(x, t)^2$, or c) no speciation or extinction happens with probability $(1 - \mu(x)\Delta t - \lambda(x)\Delta t)$, leading to total extinction with probability $E(x, t)$. We must integrate over the character change that might occur during this period of time; lineages in which the character may change are indicated in black.

$\mu(x)\Delta t$, 2) speciates with probability $\lambda(x)\Delta t(1 - \mu(x)\Delta t)$, or 3) neither speciates nor goes extinct with probability $(1 - \lambda(x)\Delta t)(1 - \mu(x)\Delta t)$ (see Maddison et al. 2007). If extinction does not occur in this time interval, character change may have occurred along the branch, and we must account for all possible character transitions that might have occurred. Where speciation occurred, this character change occurs independently on both lineages and both lineages must be extinct by the present (Fig. 1). Summing over these possibilities gives

$$\begin{aligned}
 E(x, t + \Delta t) = & \mu(x)\Delta t \times 1 + (1 - \mu(x)\Delta t)\lambda(x)\Delta t \\
 & \times \left[\int_{-\infty}^{\infty} g(z, t|x, t + \Delta t)E(z, t)dz \right]^2 \\
 & + (1 - \lambda(x)\Delta t)(1 - \mu(x)\Delta t) \\
 & \times \int_{-\infty}^{\infty} g(z, t|x, t + \Delta t)E(z, t)dz \\
 & + O(\Delta t^2), \tag{2}
 \end{aligned}$$

where $O(\Delta t^2)$ includes terms of order $(\Delta t)^2$ or higher (see Fig. 1). Subtracting $E(x, t)$ from both sides, dividing by Δt , taking the limit $\Delta t \rightarrow 0$, and using the diffusion conditions above, the following partial differential equation can be derived (see Appendix 1 for details):

$$\begin{aligned}
 \frac{\partial E(x, t)}{\partial t} = & \mu(x) + \lambda(x)E(x, t)^2 - (\lambda(x) + \mu(x))E(x, t) \\
 & + \phi(x, t)\frac{\partial E(x, t)}{\partial x} + \frac{\sigma^2(x, t)}{2}\frac{\partial^2 E(x, t)}{\partial x^2}. \tag{3}
 \end{aligned}$$

Probability of the Data

Next, consider the probability of a lineage including topology, branch lengths, and character states among its extant descendants in clade N , $D_N(x, t)$. Because the calculations here assume that we are not at a node, only a single lineage can be present in the reconstructed phylogeny. The three possible events that could occur over

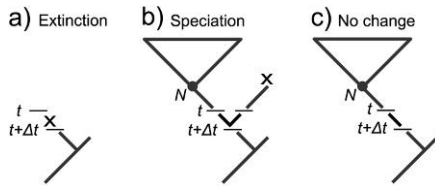


FIGURE 2. Possible ways a lineage extant at time $t + \Delta t$ might lead to exactly the clade N as observed. If at most a single lineage-changing event occurs, then a) extinction happens with no chance of explaining the data, b) a speciation event requiring the extinction of either lineage (probability of $2D_N(x, t)E(x, t)$ of explaining the data), or c) no speciation or extinction, with probability $D_N(x, t)$ of explaining the data. See Figure 1 for other details.

the period of time Δt that are consistent with this are 1) extinction, with no chance to explain the data (the extant clade N), 2) speciation, requiring the eventual extinction of either of the resulting lineages (with the other becoming clade N), and 3) no speciation or extinction, leaving a single lineage to become clade N (Fig. 2). Incorporating the possible character transitions in all nonextinct lineages as for $E(x, t)$ gives

$$\begin{aligned}
 D_N(x, t + \Delta t) &= \mu(x)\Delta t \times 0 + 2\lambda(x)(1 - \mu(x)\Delta t)\Delta t \\
 &\times \left[\int_{-\infty}^{\infty} g(z, t|x, t + \Delta t)D_N(z, t)dz \right] \\
 &\times \left[\int_{-\infty}^{\infty} g(z, t|x, t + \Delta t)E(z, t)dz \right] \\
 &+ (1 - \lambda(x)\Delta t)(1 - \mu(x)\Delta t) \\
 &\times \int_{-\infty}^{\infty} g(z, t|x, t + \Delta t)D_N(z, t)dz \\
 &+ O(\Delta t^2).
 \end{aligned} \quad (4)$$

The 2 in equation (4) appears because either of the two lineages that are extant after a speciation event could be consistent with the observed data, provided the other goes extinct (Maddison et al. 2007). Using the same logic as was used to derive $E(x, t)$ gives the partial differential equation

$$\begin{aligned}
 \frac{\partial D_N(x, t)}{\partial t} &= 2\lambda(x)D_N(x, t)E(x, t) - (\lambda(x) + \mu(x))D_N(x, t) \\
 &+ \phi(x, t)\frac{\partial D_N(x, t)}{\partial x} + \frac{\sigma^2(x, t)}{2}\frac{\partial^2 D_N(x, t)}{\partial x^2}.
 \end{aligned} \quad (5)$$

Equations (3) and (5) form the core of QuaSSE.

Initial and Boundary Conditions

Equations (3) and (5) do not have known analytic solutions. However, given appropriate initial and boundary conditions, they may be integrated numerically along a branch toward the root of the tree. For the initial condition for E , note that a lineage cannot go extinct in zero time, so $E(x, 0) = 0$ for all x . The initial condition

for $D_N(x, t)$ must be a probability distribution function; that is, it must integrate to 1 over all x because at time 0 a lineage does exist. If we knew with absolute certainty that an extant species had state x_{obs} , we could use a Dirac delta function,

$$D_N(x, 0) = \delta(x - x_{\text{obs}}),$$

which concentrates the probability distribution on the observed character state x_{obs} and integrates to 1. However, in contrast with discrete data, a species' state is never known without error due to both within-species variation and measurement error. In the examples below, I will use a normal distribution centered on x_{obs} , with standard deviation σ_{obs} , but any probability distribution could be used.

To integrate these equations numerically, a finite domain and boundary conditions need to be specified. Suppose that the range (x_l, x_r) is modeled; I assume that at these boundaries, the derivative of $E(x, t)$ and $D_N(x, t)$ with respect to x is zero (i.e., Neumann boundary conditions). This requires that the derivative of $\lambda(x)$ and $\mu(x)$ with respect to x is approximately zero at the boundaries and that the region is sufficiently wide that $D_N(x, t)$ is very close to zero at the boundaries so that the probability of explaining the data from states beyond these boundaries is negligible.

Calculations at the Nodes

Given the initial and boundary conditions above, equations (3) and (5) can be integrated along a branch to give distributions at the base of nodes. At the node N' that joins the branches leading from nodes N and M , the initial condition is

$$D_{N'}(x, t) = D_N(x, t)D_M(x, t)\lambda(x). \quad (6)$$

This is the probability of the lineage at the node being in state x at time t speciating, then giving rise to both the N and M clades. This value is then used as the initial condition for the integration along the branch leading down from this node.

Calculations at the Root

At the base of the tree, we have a function $D_R(x, t_R)$, where t_R is the time at the root. To get a single likelihood value, D_R , we must integrate over all possible character states x . This has been discussed elsewhere for the binary case (Goldberg and Igić 2008; FitzJohn et al. 2009). The simplest approach is to integrate over the possible states:

$$D_R = \int_{x_l}^{x_r} D_R(x, t_R)dx. \quad (7)$$

This is equivalent to assigning a flat prior to the character state at the root (e.g., Pagel 1994). However, the tree and model contain some information about the likely state at the root, and we can use this by weighting the

state x by its relative probability of yielding the observed data

$$D_R = \int_{x_l}^{x_r} D_R(x, t_R) \frac{D_R(x, t_R)}{\int_{x_l}^{x_r} D_R(y, t_R) dy} dx. \quad (8)$$

The latter approach is used in the calculations in this paper.

Extensions

Incomplete phylogeny.—Where a phylogenetic tree does not include all extant relatives, the above calculations may not be used, as they will produce incorrect likelihoods (see FitzJohn et al. 2009). However, if the species included in a phylogeny represent a random sample from the extant taxa, a simple modification to the calculations above allows the tree to be used, following Nee et al. (1994) and FitzJohn et al. (2009). Suppose that a species in state x has probability $f(x)$ of being sampled for inclusion in the tree. We can model this sampling event similar to a mass extinction at the present (Nee et al. 1994). If the probability of being included in a phylogenetic tree is thought to be independent of the character of interest, then we may set $f(x) = f$. Above, $E(x, t)$ was defined as the probability that a lineage in state x at time t would have no extant descendants. For incomplete trees, we can interpret this as the probability of failing to appear in the phylogenetic tree either by extinction or by not being sampled. The initial conditions then become $E(x, 0) = 1 - f(x)$. Likewise, we can interpret $D_N(x, t)$ as the probability that the lineage evolves into a clade N and is sampled. The initial condition $D(x, 0)$ is then the product of a distribution describing uncertainty in the extant species state and $f(x)$.

Multiple characters.—Multiple traits may affect diversification rate, and these may not evolve independently. For example, body size and latitude are likely to be correlated and are thought to both have effects on speciation and/or extinction rates (Jablonski 2008). Suppose that we are tracking k traits. Let \mathbf{x} be a vector of character states of length k , and let x_i be the i th trait ($i = 1, 2, \dots, k$). The speciation and extinction functions become $\lambda(\mathbf{x})$ and $\mu(\mathbf{x})$. As above, we retain just the first two moments of character evolution (which now includes covariances) so that $\phi_i(x, t)$ is the rate of directional evolution of the i th trait, $\sigma_{i,i}(x, t)$ is the rate of diffusion of the i th trait, and $\sigma_{i,j}(x, t)$ is the instantaneous covariance between the i th and j th traits ($i \neq j$). In Appendix 2, I derive the multivariate analogues to equations (3) and (5):

$$\begin{aligned} \frac{\partial E(\mathbf{x}, t)}{\partial t} &= \mu(\mathbf{x}) + \lambda(\mathbf{x})E(\mathbf{x}, t)^2 - (\lambda(\mathbf{x}) + \mu(\mathbf{x}))E(\mathbf{x}, t) \\ &+ \sum_{i=1}^k \phi_i(\mathbf{x}, t) \frac{\partial E(\mathbf{x}, t)}{\partial x_i} \\ &+ \sum_{i=1}^k \sum_{j=1}^k \frac{\sigma_{i,j}(\mathbf{x}, t)}{2} \frac{\partial^2 E(\mathbf{x}, t)}{\partial x_i \partial x_j} \end{aligned} \quad (9)$$

and

$$\begin{aligned} \frac{\partial D_N(\mathbf{x}, t)}{\partial t} &= 2\lambda(\mathbf{x})D_N(\mathbf{x}, t)E(\mathbf{x}, t) - (\lambda(\mathbf{x}) + \mu(\mathbf{x}))D_N(\mathbf{x}, t) \\ &+ \sum_{i=1}^k \phi_i(\mathbf{x}, t) \frac{\partial D_N(\mathbf{x}, t)}{\partial x_i} \\ &+ \sum_{i=1}^k \sum_{j=1}^k \frac{\sigma_{i,j}(\mathbf{x}, t)}{2} \frac{\partial^2 D_N(\mathbf{x}, t)}{\partial x_i \partial x_j}, \end{aligned} \quad (10)$$

where the single sums are taken over the directional parameters, and the double sums are taken over the diffusion parameters (when $i = j$) and the covariances between characters ($i \neq j$). A similar approach can be used where the second character is a binary state (see Appendix 2).

Implementation & Technical Details

To integrate equations (3) and (5) numerically, I used an implicit integration scheme, where the propagation of the values in character space are performed using future values of E and D_N . To do this, I discretized both the character space and time. In each time step of size Δt , the changes in E and D_N through time are initially set to the character-independent solutions to equations (3) and (5),

$$\begin{aligned} E(x, t + \Delta t) &= \frac{\mu(x) - \lambda(x)E(x, t) + e^{(\lambda(x) - \mu(x))\Delta t}(1 + E(x, t) - \mu(x))}{\mu(x) - \lambda(x)E(x, t) + e^{(\lambda(x) - \mu(x))\Delta t}(1 + E(x, t) - \lambda(x))}, \end{aligned} \quad (11a)$$

$$\begin{aligned} D_N(x, t + \Delta t) &= D_N(x, t) \\ &\times \left(\frac{e^{(\lambda(x) - \mu(x))\Delta t}(\lambda(x) - \mu(x))}{e^{(\lambda(x) - \mu(x))\Delta t}\lambda(x)(1 - E(x, t)) - \mu(x) + \lambda(x)E(x, t)} \right)^2. \end{aligned} \quad (11b)$$

(Although these are character independent, these equations need to be evaluated for each of the discretized x positions.) Following this, for each step, I integrate over the character evolution that may have occurred during this period of time. For constant directional and diffusion terms (i.e., when the character evolves under Brownian motion), this can be done efficiently by convolving the functions $E(x, t)$ and $D_N(x, t)$ with a normal distribution with mean $\phi\Delta t$ and variance $\sigma^2\Delta t$, that is, the solution to the diffusion process described by the partial derivatives on the right-hand side of equations (3) and (5). I implemented these calculations in R (R Development Core Team 2008), using the fast Fourier transform routines in the package FFTW (Frigo and Johnson 2005) to perform the convolutions. I focus on maximum likelihood (ML) estimation in the analyses in this paper, using the subplex algorithm in R to maximize the likelihood function with respect to the parameters of $\lambda(x)$, $\mu(x)$, and the directional and diffusion coefficients. However, the likelihoods computed could be used in Bayesian calculations

(e.g., FitzJohn et al. 2009), although the choice of appropriate priors may not be trivial and the number of calculations required to draw samples from the posterior using Markov chain Monte Carlo will make this fairly slow in practice. This implementation (currently allowing only one character) is available in the R package “diversitree” (available from <http://www.zoology.ubc.ca/prog/diversitree>).

Tree Simulation

I tested the performance of QuaSSE on simulated trees. To simulate a tree, I started with a single lineage in state x_0 . Each time step, I allowed at most one lineage to speciate or go extinct and then updated the character state of every lineage stochastically following a Brownian motion model of character evolution. I scaled time so that on average 500 time steps would occur between lineage-changing events by setting the time step equal to $1/(500(\sum_i \lambda(x_i) + \mu(x_i)))$, where x_i is the character state of the i th lineage and the sum is taken over all extant lineages in the tree. This ensured that the character had adequate time to evolve between speciation events to approximate continuous character evolution.

I simulated trees where there was no effect of a character on speciation or extinction (i.e., $\lambda(x) = \lambda$, $\mu(x) = \mu$), and where speciation or extinction were sigmoidal functions of the character state,

$$y_0 + \frac{(y_1 - y_0)}{1 + \exp(r(x_{\text{mid}} - x))},$$

where y_0 and y_1 are the asymptotic values at low and high x , r describes the steepness of the sigmoid, and x_{mid} is the inflection point. I chose a sigmoidal function as this captures a directional effect of a character on speciation or extinction while preventing negative or extremely large speciation or extinction rates. When constant, the speciation rate was 0.1 and the extinction rate was 0.03. For the differential speciation simulations, the speciation rates varied with x from 0.1 to 0.15 (low difference), 0.1 to 0.2 (medium), or 0.1 to 0.3 (high). For differential extinction, the rates varied from 0.03 to 0.045 (low difference), 0.03 to 0.06 (medium), or 0.03 to 0.09 (high). For all simulations, I set $x_{\text{mid}} = 0$ and $r = 2.5$. I also used two rates of character diffusion; low ($\sigma^2 = 0.01$) or high ($\sigma^2 = 0.025$). For these simulations, there was no directional tendency ($\phi = 0$). Note that the scale used for x is arbitrary (making the choice of x_{mid} arbitrary), and changing r is equivalent to changing σ^2 when only one of $\lambda(x)$ and $\mu(x)$ varies with x and $\phi = 0$. For the parameter values used, most of the variation in speciation rate with respect to the character occurred over the region $[-2, 2]$; I started simulated trees in a character state chosen randomly from a uniform distribution on this range.

Sigmoidal functions require four parameters (plus parameters for extinction and character evolution), and

there may not be sufficient signal in the data to be able to fit such complicated models (see Simulation Results). Therefore, I fit models where speciation and extinction were constant, linear, or sigmoidal functions of the trait. To make the linear models satisfy the boundary condition that $\partial E(x, t)/\partial x$ is effectively zero at extreme values of x , I set the slope to zero for $\lambda(x)$ and $\mu(x)$ once the character was 20 times the character-independent ML diffusion coefficient away from the extant character distribution. I also set the functions to zero if they became negative.

To test if speciation or extinction functions that vary with character state fit better than constant functions, I used likelihood ratio tests (LRTs) where model comparisons were nested. Even with large trees, the appropriate cutoff value for the χ^2 statistic may not be the expected 3.84 (e.g., Maddison et al. 2007). I used simulated trees where speciation and extinction rates were independent of any character states to estimate the false-positive rate. For all tree sizes, the false-positive rate for tests of differential speciation was close to 5% at the 5% level (Kolmogorov–Smirnov test of observed distribution vs. χ^2 distribution with 1 df, $P > 0.45$). However, for differential extinction, the false-positive rate was higher (12% significant results at the 5% level), and the distribution of likelihood ratios significantly deviated from a χ^2_1 distribution ($P < 0.008$). I therefore used empirically determined cutoff values for the three tree sizes of 6.50 (125 species), 5.98 (250 species), and 5.72 (500 species) based on these simulations for the power calculations reported below.

SIMULATION RESULTS

On simulated trees, there was often power to detect differential speciation, whereas differential extinction was always difficult, but not impossible, to detect (Fig. 3). Because a linear function was significant for almost all cases where a sigmoidal function was significant, I interpret a significant fit of the linear function as detection of differential speciation or extinction. As tree size increased, the power to detect differential speciation grew from 10% to 40% on 125 species trees to up to 70% on 500 species trees. As the difference between minimum and maximum speciation or extinction rates increased, differential speciation and extinction became easier to detect. However, there was essentially no power to detect differential extinction for the simulated trees unless the character had a large effect on rates of extinction (Fig. 3f–h). Differential speciation was easier to detect in simulations with higher diffusion parameters. Increasing the diffusion parameter increases the sampling of the character space where $\lambda(x)$ changes most with respect to x , effectively increasing the sampling of the relevant character states. Power to detect trait-dependent speciation also depended strongly on the starting position of the simulation (data not shown). The power to detect differential speciation was highest on trees where the simulation started slightly below the

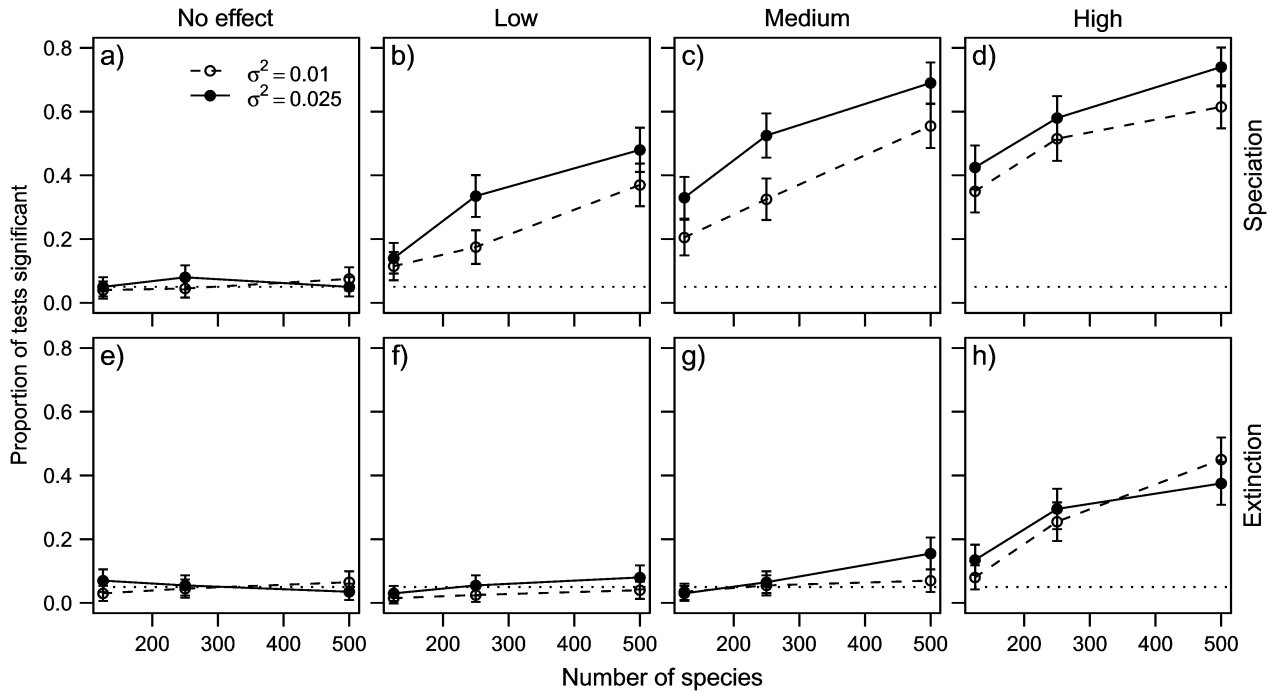


FIGURE 3. Power to detect differential speciation (a–d) and differential extinction (e–h) with QuaSSE on simulated phylogenies. Trees were evolved with 125, 250, or 500 species, and no, low, medium, or high effect of a character state on speciation or extinction. The character state evolved under Brownian motion at a low or high rate, as indicated by dashed or solid lines. Error bars are 95% binomial confidence intervals over the 200 replicate simulated trees. Horizontal dotted lines indicate the 5% significance level.

mean diversification rate, again reflecting the amount of sampling of the informative region of differential speciation.

Even though the trees were simulated using a sigmoidal effect of character on speciation or extinction, linear models were often preferred to the full sigmoidal model when fits were compared using the Akaike information criterion (AIC) (Fig. 4). Sigmoidal models were rarely preferred for the extinction function. As tree size increased, the sigmoidal speciation models were more often preferred over linear models (data not shown). However, even for 500 species trees, sigmoidal functions were preferred in less than 40% of significant cases. This is probably due to the fact that for most smaller trees, most species occupy the roughly linear part of the sigmoidal function (Fig. 4). In addition, the sigmoidal model that was fit was often a step function rather than a smooth sigmoid (Fig. 4), showing that although differences in extreme speciation rates are detectable, the exact pattern may not be.

To investigate if there is power to detect directional changes in character states, I ran some simulations that included a nonzero directional parameter, ϕ . I used only the set of parameters with the highest power in the absence of directional evolution (500 species, high rate of character evolution, large effect of the trait on speciation). I ran simulations where this directional tendency was negative and opposed species selection (i.e., the trait tended to decrease along a lineage, whereas species with larger trait values had higher rates of speciation) and where the tendency was positive and

reinforced species selection. When rates of this directional tendency were very high in either direction, the power to detect differential diversification was reduced as character states tended to evolve into flatter regions of the speciation function (Fig. 5). When the directional tendency opposed species selection, there was some power to detect the trend, but this power was never high for the parameter values explored (Fig. 5). There was essentially no power to detect the presence of the directional tendency where it reinforced species selection, as both processes moved most character states into regions where speciation was constant with respect to the character state.

APPLICATION TO PRIMATE BODY SIZE DATA

Several studies have suggested that speciation and/or extinction are correlated with animal body size. Typically, smaller bodied species have been hypothesized to have higher speciation rates or lower extinction rates than larger bodied species (e.g., [Cardillo et al. 2005](#); [Clauset and Erwin 2008](#)). In some groups, there is also paleontological evidence for increases in body size over evolutionary time, with species tending to be larger than their ancestors (Cope's rule; [Alroy 1988](#); [Jablonski 1997](#)). I investigated body size evolution in primates, testing whether body size is a correlate of speciation or extinction. I used a recent primate supertree ([Vos and Mooers 2006](#)). This tree contains several polytomies that need to be resolved before running the analysis (213 of 232

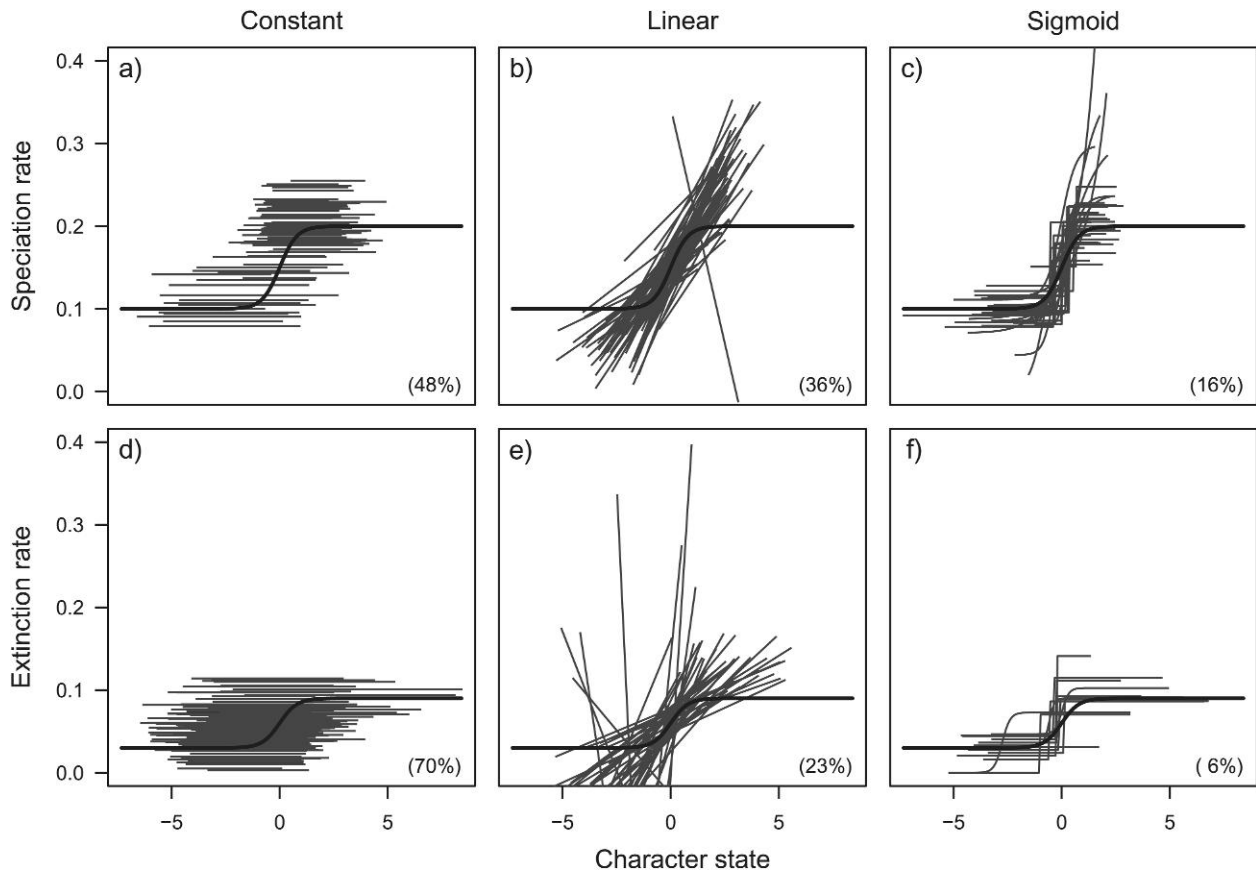


FIGURE 4. Representative speciation (a–c) and extinction (d–f) function fits. Functions were fit to 200 trees containing 250 taxa with a high rate of character evolution ($\sigma^2 = 0.025$) and a medium effect of speciation ($\lambda(x)$ ranged from 0.1 to 0.2) or a high effect of extinction ($\mu(x)$ ranged from 0.03 to 0.09). Thick black lines show the true values used in the simulations. Gray lines show fits for individual trees and span the range of observed character data for each tree. Only the best fit chosen by LRT or AIC is displayed (see text for details).

internal nodes are resolved; Fig. 6). The polytomies reflect phylogenetic uncertainty, but the likelihood calculations above would interpret them as bursts of speciation followed by relatively low rates of speciation. It is not sufficient to randomly resolve the nodes and leave branch lengths as effectively zero, as branch lengths need to be specified in some way to prevent this misinterpretation. To do this, I used a bifurcating tree generated by T. Kuhn in which he randomly resolved the topology of the polytomies and then used BEAST (Drummond and Rambaut 2007) to simulate unknown branch lengths under a constant rates birth-death model (Kuhn T., Thomas G.H., and Mooers A.Ø., personal communication). I used log-transformed female body mass from a recent collection of primate trait data as a measure of size (Redding et al. 2010).

For each tree, I fit several functions: constant rates for both speciation and extinction, and models where the speciation or extinction function was linear, sigmoidal, or modal. For the modal function, I used a vertically offset Gaussian

$$y_0 + (y_1 - y_0) \exp\left(-\frac{(x - x_{\text{mid}})^2}{2\omega_x^2}\right),$$

where y_0 is the rate at low and high x , y_1 is the rate at the mid point x_{mid} , and ω_x^2 is the width (variance) of the Gaussian kernel. To test for the presence of directional body size evolution (e.g., Cope's rule), I also ran models where there was a nonzero, but constant, directional term (ϕ).

Among models with a directional relationship between log body size and speciation, there was strong support for a positive linear relationship between log body size and speciation rates (LRT against the constant rate model: $\chi_1^2 = 10.8$, $P = 0.001$), with a step-shaped sigmoidal curve preferred ($\Delta \text{AIC} = 2.8$, Fig. 7). Contrary to the predictions above, speciation rates were inferred to increase with increasing body size. However, the best-fit model was a modal-speciation model (Table 1), where species with body masses around 2.5–13.4 kg had elevated speciation rates (Fig. 7). Including a positive directional term, consistent with increasing average body size along lineages, improved model fit significantly (Table 1).

The body size with elevated speciation rates inferred using the modal-speciation function is concentrated in the Cercopithecoidea and Hominoidea (Old World

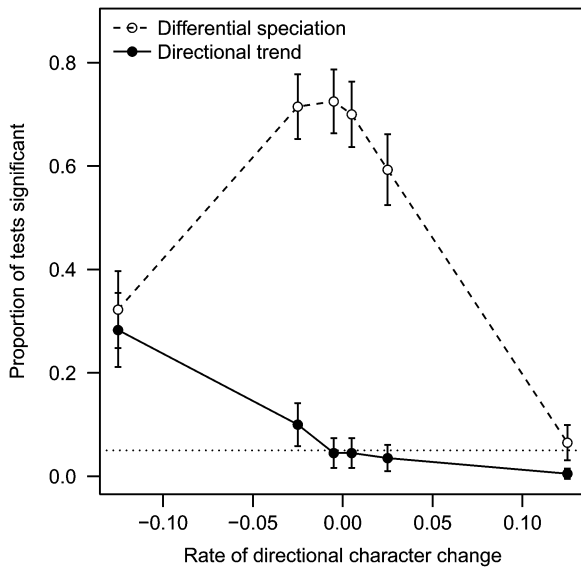


FIGURE 5. Power to detect trait-dependent speciation (dashed lines/open circles) and directional character change (solid lines/filled circles) at different rates of the directional tendency on simulated 500 species phylogenies.

monkeys and apes), and this section of the tree does appear to have undergone a recent burst of relatively rapid diversification compared with the rest of the tree (Fig. 6). It is possible that any character that is concentrated in this clade could lead to a significant correlation with diversification, so this body size result could be spurious. To explore this further, I used modelling evolutionary diversification under stepwise AIC (MEDUSA; Alfaro et al. 2009) to test for clade-specific differences in diversification across the tree. Using the suggested AIC difference of 4, there was support for a single partition that separated the tree into the Old World monkey clade (superfamily Cercopithecoidea; Fig. 6), and the rest of the tree (LRT $\chi^2_3 = 24.9$, $P < 0.0001$).

I modified QuaSSE to allow this partition. I allowed a “background” group (all clades except for the Old World monkeys) to have one set of speciation and extinction functions and a “foreground” group (the Old World monkeys) to have another. The two groups shared a common diffusion coefficient and I set the directional term to zero. A model with constant speciation and extinction functions that could differ between the partitions had a lower (better) AIC and fewer parameters than the unpartitioned modal-speciation model (Table 1). I found no support for any relationship between body size and either speciation or extinction for the “background” group (Table 1). However, there was support for a model where speciation was a decreasing linear function of log-body size among the Old World monkeys (LRT vs. the constant rate partitioned model: $\chi^2_1 = 5.1$, $P = 0.024$) or where extinction was an increasing function of log-body size within this group ($\chi^2_1 = 9.8$, $P = 0.002$), suggesting decreasing diversification with

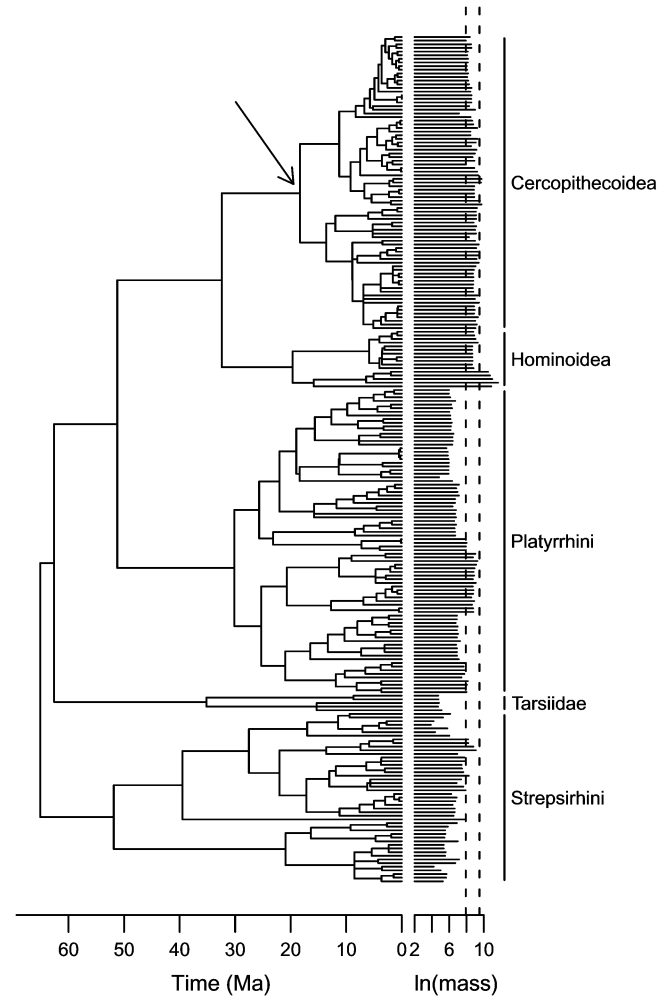


FIGURE 6. Phylogenetic tree of the primates from Vos and Mooers (2006). Log body size (in grams) is shown by the horizontal bar for each species. The vertical dashed lines indicate the approximate range of body masses for which QuaSSE inferred elevated speciation rates under the “modal” speciation model (the lines indicate masses that are 10% above the base speciation rate). The arrow indicates where MEDUSA inferred a shift in speciation and extinction rates compared with the rest of the tree.

increased body size in this group. However, the latter fit suggested an extremely high rate of extinction among large Old World monkeys (Fig. 7).

The results presented here are in broad accord with the analysis of body size evolution in primates by Paradis (2005), Gittleman and Purvis (1998), and Freckleton et al. (2008), despite using different methods, phylogenetic trees, and data sets. These studies all initially inferred a relationship of increasing diversification rates with increasing body size, though this was not significant in Gittleman and Purvis (1998). Paradis (2005) also looked at a partitioned data set and found support for decreasing diversification with increasing body size after allowing clade-specific diversification rates.

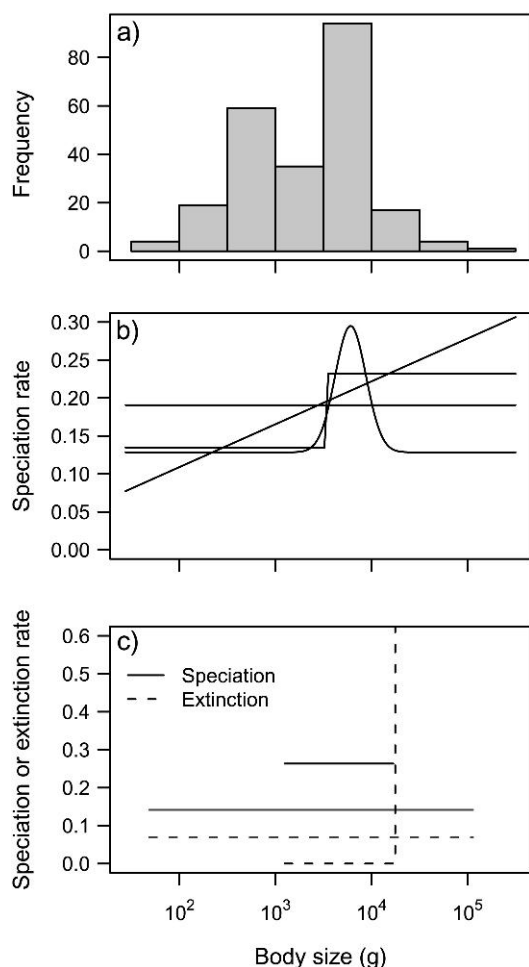


FIGURE 7. Primate speciation and extinction rate model fits. a) Primate body mass distribution. b) ML speciation rate model fits for the complete tree, showing constant, linear, sigmoidal, and modal functions. The modal model provided the best fit to the data (Table 1). c) ML speciation and extinction functions on the partitioned tree: Gray lines are fits for the “background group,” and the black lines are fits for the Cercopithecoidea. The reverse-L shape of the Cercopithecoidea extinction rate function (black dashed line in panel c) indicates a zero extinction rate for all observed body sizes and an extremely high extinction rate for species with body masses slightly greater than the largest observed mass in the Cercopithecoidea.

DISCUSSION

How much signal is there within a phylogeny about the evolutionary processes that generated it? On the simulated trees used here, it was generally possible to infer the correct trend in the character dependence of speciation but difficult to infer the exact functional form of the trend. For instance, both the linear and sigmoidal functions capture the tendency of speciation to increase or decrease with increasing character state and the inferred linear speciation function was often a rough characterization of the true function (Fig. 4). It is often difficult to infer ancestral states with confidence (which are needed to identify a speciation-trait correlation, even though this is only done implicitly here), as the information provided by the tips attenuates deeper

TABLE 1. Summary of model fits for the correlation between body size and diversification for primates

Model type	$\ln L$	n	AIC	ΔAIC
Constant	-834.8	3	1675.7	29.6
Linear λ	-829.4	4	1666.9	20.8
Sigmoidal λ	-826.0	6	1664.1	18.0
Modal λ	-822.4	6	1656.8	10.7
With directional tendency				
Linear λ	-826.0	5	1661.9	15.8
Sigmoidal λ	-823.8	7	1661.7	15.6
Modal λ	-818.8	7	1651.7	5.6
Partitioned tree (no directional tendency)				
Constant	-822.0	5	1654.0	7.8
Linear λ (fg)	-819.4	6	1650.9	4.7
Linear λ (bg)	-821.6	6	1655.2	9.0
Linear λ (both)	-819.1	7	1652.2	6.1
Linear μ (fg)	-817.1	6	1646.1	0.0
Linear μ (bg)	-821.7	6	1655.4	9.3
Linear μ (both)	-816.8	7	1647.7	1.6

Notes: “Fg” and “bg” refer to the Cercopithecoidea clade (Old World monkeys) and the rest of the tree, respectively. “Both” is where the functions were fit to both groups separately. $\ln L$ is the log likelihood of the ML fit, n is the number of parameters, and ΔAIC is the AIC difference relative to the best model (linear μ (fg)).

into the past. Here, adding more species improved the ability to recover the more specific model, but this may be through the larger number of shallow nodes rather than through more accurate information about deep ancestral states (Mossel and Steel 2005).

It is possible that extinction is not possible to reliably detect on real (nonsimulated) molecular phylogenies. Accurate detection of extinction requires that we determine the rate at which species fail to appear in our phylogeny, which is a difficult task. ML estimates of the extinction rates are frequently zero despite fossil evidence of nonzero extinction (e.g., Nee 2006; Purvis 2008). However, even when ML estimates are zero, the confidence intervals around extinction rate estimates may be large, allowing potentially high levels of extinction to be consistent with the observed data. Where we have strong independent evidence of high extinction rates, perhaps our analyses would be improved by including these rates directly either through a prior distribution on extinction rates in a Bayesian analysis or by using this estimated rate and not attempting to directly estimate it from the phylogeny. The likelihood calculations proposed here would hold in either case.

Many phylogenies appear to show some sort of slowdown in lineage accumulation toward the present, which will generate low extinction rate estimates. The response to this has generally been to alter the model of diversification. Most commonly, slowdowns have been interpreted as evidence that speciation rates may be density dependent (e.g., McPeck 2008; Phillimore and Price 2008), and various alternative models of cladogenesis have been proposed and tested based on this pattern (e.g., McPeck 2008; Rabosky 2009). Because of its use of the birth-death model, which does not allow interaction among lineages, it would not be straightforward to incorporate these types of dynamics directly into QuaSSE, though it is possible that they may be

approximated (Rabosky and Lovette 2008, but see Bokma 2009). Care should be taken to interpret results from QuaSSE and other birth–death-based models (e.g., Nee et al. 1994; Paradis 2005; Rabosky 2006; Maddison et al. 2007; Freckleton et al. 2008; Alfaro et al. 2009) in light of these limitations.

An alternative explanation for the observed “slow-down,” and consequent problems for estimating speciation and extinction rates, is that our methods of tree construction and ultrametricization creates trees that are incongruous with the model. Extinction rate estimates will always be sensitive to the precise lengths of terminal branches, and any consistent bias toward lengthening the terminal branches will cause problems (Purvis 2008). Furthermore, our delineation of species is generally retrospective, with lineages counted as species once both morphological changes and reproductive isolation have occurred. However, many isolated lineages may be considered “species” in that they will never again exchange genes. Some of these would eventually become recognized species, but most will go extinct. However, simple birth–death models do not include this sort of process; incorporating such lags in species recognition into tree construction or diversification models, along with information from the fossil record where available, may help with efforts to infer meaningful speciation and extinction rates.

The likelihood equations derived here provide exact solutions to the forward-time dynamics described by Paradis (2005) and Freckleton et al. (2008), and also to the early model of Slatkin (1981), but ignoring character evolution at nodes. The key advance of this work is that it treats character evolution and cladogenesis simultaneously. Though the equations cannot be solved directly, likelihoods computed using this approach will correspond exactly to those under this model of character evolution and cladogenesis. Because the likelihood method here uses all the available phylogenetic and character data, it should have higher statistical power than methods based on approximations, such as first inferring ancestral states and ignoring the character-dependent diversification process when doing so. Run on the same trees, the model of Freckleton et al. (2008) had approximately 26% of the power of QuaSSE at detecting differential speciation (data not shown). However, the factors that affect power were the same as identified by Freckleton et al. (2008); increased rates of character evolution, stronger effects of a character on speciation, and larger trees all increased power (Fig. 3). QuaSSE does retain some ability to detect differential extinction in contrast to the method of Freckleton et al. (2008), but this power appears to be limited and parameter dependent (Fig. 3). QuaSSE was also robust to the levels of background extinction used here (cf. Paradis 2005).

Despite their assumptions, diffusion models of character evolution and birth–death models of cladogenesis have given us insights over the last few decades into correlated character evolution (Felsenstein 1985), evolutionary constraints (Hansen and Martins 1996), and

patterns of diversification (Alfaro et al. 2009). Although the combination of the birth–death and diffusion methods used in QuaSSE may inherit the limitations of both methods, it presents a tractable and powerful method that will help to answer long standing questions about the correlates of diversification from phylogenetic data and current character distributions. As Freckleton et al. (2008) noted, we have no general expectation of what the relationship between speciation or extinction and character states might look like. Because QuaSSE can use arbitrary speciation and extinction functions, it allows investigation of alternative functions. However, we should not generally expect to extract more than general trends from the data, especially where variation in extinction is important in affecting patterns of diversification.

FUNDING

This work was supported by a University Graduate Fellowship from the University of British Columbia, a Capability Fund Grant from Manaaki Whenua/Landcare Research, and a Vanier Commonwealth Graduate Scholarship from the Natural Sciences and Engineering Research Council of Canada (NSERC to R.G.F.) and an NSERC discovery grant to Sarah P. Otto.

ACKNOWLEDGMENTS

I would like to thank Will Cornwell, Leithen M’Gonigle, Wayne Maddison, Arne Mooers, Dolph Schluter, and the UBC Delta-Tea group for feedback on the manuscript and helpful discussions. Folmer Bokma, Dan Rabosky, and an anonymous reviewer also provided feedback on an earlier version of the manuscript. I especially thank Sally Otto for encouragement and invaluable advice during the development of this work. Arne Mooers, David Redding, and Rutger Vos generously shared data for the primate analysis, and Tyler Kuhn generated the primate trees with simulated branch lengths. The UBC Zoology Computing Unit and Wayne Maddison provided computer time.

REFERENCES

- Alfaro M.E., Santini F., Brock C., Alamillo H., Dornburg A., Rabosky D.L., Carnevale G., Harmon L.J. 2009. Nine exceptional radiations plus high turnover explain species diversity in jawed vertebrates. *Proc. Natl. Acad. Sci. U.S.A.* 106:13410–13414.
- Allen L.J.S. 2003. An introduction to stochastic processes with applications to biology. Upper Saddle River (NJ): Pearson Prentice Hall.
- Alroy J. 1988. Cope’s rule and the dynamics of body mass evolution in North American fossil mammals. *Science*. 280:731–734.
- Bokma F. 2009. Problems detecting density-dependent diversification on phylogenies. *Proc. R. Soc. Lond. B Biol. Sci.* 276:993–994.
- Cardillo M., Mace G.M., Jones K.E., Bielby J., Bininda-Emonds O.R.P., Sechrest W., Orme C.D.L., Purvis A. 2005. Multiple causes of high extinction risk in large mammals. *Science*. 309:1239–1241.
- Clauset A., Erwin D.H. 2008. The evolution and distribution of species body size. *Science*. 321:399–401.
- Coyne J.A., Orr H.A. 2004. *Speciation*. Sunderland (MA): Sinauer Associates.
- Drummond A.J., Rambaut A. 2007. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol. Biol.* 7:214.

- 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.* 1:1–15.
- Felsenstein J. 1988. Phylogenies and quantitative characters. *Annu. Rev. Ecol. Syst.* 19:445–471.
- FitzJohn R.G., Maddison W.P., Otto S.P. 2009. Estimating trait-dependent speciation and extinction rates from incompletely resolved phylogenies. *Syst. Biol.* 58:595–611.
- Freckleton R.P., Phillimore A.B., Pagel M. 2008. Relating traits to diversification: a simple test. *Am. Nat.* 172:102–115.
- Frigo M., Johnson S.G. 2005. The design and implementation of FFTW3. *Proc. IEEE.* 93:216–231.
- Gilinsky N.L. 1994. Volatility and the phanerozoic decline of background extinction intensity. *Paleobiology.* 20:445–458.
- Gittleman J., Purvis A. 1998. Body size and species-richness in carnivores and primates. *Proc. R. Soc. Lond. B Biol. Sci.* 265:113–119.
- Goldberg E.E., Igić B. 2008. On phylogenetic tests of irreversible evolution. *Evolution.* 62:2727–2741.
- Hansen T.F., Martins E.P. 1996. Translating between microevolutionary process and macroevolutionary patterns: the correlation structure of interspecific data. *Evolution.* 50:1404–1414.
- Harcourt A.H., Copperto S.A., Parks S.A. 2002. Rarity, specialization and extinction in primates. *J. Biogeogr.* 29:445–456.
- Heilbut J.C. 2000. Lower species richness in dioecious clades. *Am. Nat.* 156:221–241.
- Hodges S.A., Arnold M.L. 1995. Spurring plant diversification: are floral nectar spurs a key innovation? *Proc. R. Soc. Lond. Ser. B Biol. Sci.* 262:343–348.
- Jablonski D. 1997. Body-size evolution in Cretaceous molluscs and the status of Cope's rule. *Nature.* 385:250–252.
- Jablonski D. 2008. Species selection: theory and data. *Annu. Rev. Ecol. Syst.* 39:501–524.
- Lewontin R. 1970. The units of selection. *Annu. Rev. Ecol. Syst.* 1:1–18.
- Liow L.H., Fortelius M., Bingham E., Lintulaakso K., Mannila H., Flynn L., Stenseth N.C. 2008. Higher origination and extinction rates in larger mammals. *Proc. Natl. Acad. Sci. U.S.A.* 105:6097–6102.
- Lynch V.J. 2009. Live-birth in vipers (Viperidae) is a key innovation and adaptation to global cooling during the Cenozoic. *Evolution.* 63:2457–2465.
- Maddison W.P. 2006. Confounding asymmetries in evolutionary diversification and character change. *Evolution.* 60:1743–1746.
- Maddison W.P., Midford P.E., Otto S.P. 2007. Estimating a binary character's effect on speciation and extinction. *Syst. Biol.* 56:701–710.
- McPeck M.A. 2008. The ecological dynamics of clade diversification and community assembly. *Am. Nat.* 172:E270–E284.
- Mitter C.B., Farrell B., Wiegmann B. 1988. The phylogenetic study of adaptive zones: has phytophagy promoted insect diversification? *Am. Nat.* 132:107–128.
- Mossel E., Steel M. 2005. How much can evolved characters tell us about the tree that generated them? In: Gascuel O., editor. *Mathematics of evolution and phylogeny.* Oxford: Oxford University Press. p. 384–412.
- Nee S. 2006. Birth-death models in macroevolution. *Annu. Rev. Ecol. Syst.* 37:1–17.
- Nee S., May R.M., Harvey P.H. 1994. The reconstructed evolutionary process. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 344:305–311.
- Okasha S. 2006. *Evolution and the levels of selection.* New York: Oxford University Press.
- Pagel M. 1994. Detecting correlated evolution on phylogenies: a general method for the comparative analysis of discrete characters. *Proc. R. Soc. Lond. B Biol. Sci.* 255:37–45.
- Paradis E. 2005. Statistical analysis of diversification with species traits. *Evolution.* 59:1–12.
- Paradis E. 2008. Asymmetries in phylogenetic diversification and character change can be untangled. *Evolution.* 61:241–247.
- Phillimore A.B., Freckleton R.P., Orme C.D.L., Owens I.P.F. 2006. Ecology predicts large-scale patterns of phylogenetic diversification in birds. *Am. Nat.* 168:220–229.
- Phillimore A.B., Price T.D. 2008. Density-dependent cladogenesis in birds. *PLoS Biol.* 6:483–489.
- Purvis A. 2008. Phylogenetic approaches to the study of extinction. *Annu. Rev. Ecol. Syst.* 39:301–319.
- R Development Core Team. 2008. *R: a language and environment for statistical computing.* Vienna (Austria): R Foundation for Statistical Computing. ISBN 3-900051-07-0.
- Rabosky D.L. 2006. Likelihood methods for detecting temporal shifts in diversification rates. *Evolution.* 60:1152–1164.
- Rabosky D.L. 2009. Heritability of extinction rates links diversification patterns in molecular phylogenies and fossils. *Syst. Biol.* 58:629–640.
- Rabosky D.L., Lovette I.J. 2008. Explosive evolutionary radiations: decreasing speciation or increasing extinction through time? *Evolution.* 62:1866–1875.
- Rabosky D.L., McCune A.R. 2009. Reinventing species selection with molecular phylogenies. *Trends Ecol. Evol.* 25:68–74.
- Redding D.W., DeWolff C., Mooers A.Ø. 2010. Evolutionary distinctiveness, threat status and ecological oddity in primates. *Conserv. Biol.* 24:1052–1058.
- Schluter D., Price T., Mooers A.Ø., Ludwig D. 1997. Likelihood of ancestor states in adaptive radiation. *Evolution.* 51:1699–1711.
- Schwander T., Crespi B. 2009. Twigs on the tree of life? Neutral and selective models for integrating macroevolutionary patterns with microevolutionary processes in the analysis of asexuality. *Mol. Ecol.* 18:28–42.
- Slatkin M. 1981. A diffusion model of species selection. *Paleobiology.* 7:421–425.
- Stanley S.M. 1975. A theory of evolution above the species level. *Proc. Natl. Acad. Sci. U.S.A.* 72:646–650.
- Vamosi S.M., Vamosi J.C. 2005. Endless tests: guidelines for analysing non-nested sister-group comparisons. *Evol. Ecol. Res.* 7:567–579.
- Vos R.A., Mooers A.Ø. 2006. A new dated supertree of the Primates. Chapter 5. In: Vos R.A., editor. *Inferring large phylogenies: the big tree problem.* [Ph.D. thesis]. Burnaby BC, Canada: Simon Fraser University.
- Vrba E.S., Gould J.S. 1986. The hierarchical expansion of sorting and selection: sorting and selection cannot be equated. *Paleobiology.* 12:217–228.

APPENDIX 1

Single Character Derivation

Here, I derive equation (3). This derivation parallels both that of BiSSE (Maddison et al. 2007) and the Kolmogorov backward differential equation (Allen 2003). Starting with equation (2), subtracting $E(x, t)$ from both sides and dropping remaining terms that are of order $(\Delta t)^2$ gives

$$\begin{aligned}
 & E(x, t + \Delta t) - E(x, t) \\
 &= \mu(x)\Delta t + \lambda(x)\Delta t \left[\int_{-\infty}^{\infty} g(z, t|x, t + \Delta t)E(z, t)dz \right]^2 \\
 &\quad - (\lambda(x) + \mu(x))\Delta t \int_{-\infty}^{\infty} g(z, t|x, t + \Delta t)E(z, t)dz \\
 &\quad + \int_{-\infty}^{\infty} g(z, t|x, t + \Delta t)E(z, t)dz - E(x, t) + O(\Delta t^2).
 \end{aligned} \tag{A.1}$$

Next, note that because g is a probability distribution function, it must integrate to 1 over all possible future character states, $\int_{-\infty}^{\infty} g(z, t|x, t + \Delta t)dz = 1$, so that

$$E(x, t) = \int_{-\infty}^{\infty} g(z, t|x, t + \Delta t)E(x, t)dz. \tag{A.2}$$

Replacing the final $E(x, t)$ term in equation (A.1) with equation (A.2) and dividing both sides by Δt gives (A.3) gives

$$\begin{aligned} & \frac{E(x, t + \Delta t) - E(x, t)}{\Delta t} \\ &= \mu(x) + \lambda(x) \left[\int_{-\infty}^{\infty} g(z, t|x, t + \Delta t) E(z, t) dz \right]^2 \\ & \quad - (\lambda(x) + \mu(x)) \int_{-\infty}^{\infty} g(z, t|x, t + \Delta t) E(z, t) dz \\ & \quad + \frac{1}{\Delta t} \int_{-\infty}^{\infty} g(z, t|x, t + \Delta t) (E(z, t) - E(x, t)) dz + O(\Delta t). \end{aligned} \tag{A.3}$$

We then take the limit $\Delta t \rightarrow 0$. In this limit, the left-hand side becomes the partial derivative $\partial E(x, t)/\partial t$. Because wild jumps are not allowed and we are considering an infinitesimally small time period, the term $E(z, t)$ can be expanded as a Taylor series in z around the point $z = x$:

$$\begin{aligned} E(z, t) &= E(x, t) + (z - x) \frac{\partial E(x, t)}{\partial x} + \frac{(z - x)^2}{2} \frac{\partial^2 E(x, t)}{\partial x^2} \\ & \quad + O((z - x)^3). \end{aligned} \tag{A.4}$$

Using this expansion, the third term on the right-hand side of equation (A.3) can then be written

$$\begin{aligned} & \lim_{\Delta t \rightarrow 0} (\lambda(x) + \mu(x)) \int_{-\infty}^{\infty} g(z, t|x, t + \Delta t) \\ & \quad \times \left(E(x, t) + (z - x) \frac{\partial E(x, t)}{\partial x} \right. \\ & \quad \left. + \frac{(z - x)^2}{2} \frac{\partial^2 E(x, t)}{\partial x^2} + O((z - x)^3) \right) dz. \end{aligned}$$

In the limit $\Delta t \rightarrow 0$, transitions any distance away from x become increasingly unlikely. That is,

$$\lim_{\Delta t \rightarrow 0} g(z, t|x, t + \Delta t) = \delta(z - x),$$

where $\delta(x)$ is the Dirac delta function, concentrating all probability density on the point $z = x$. This means that

$$\lim_{\Delta t \rightarrow 0} \int_{-\infty}^{\infty} (z - x)^k g(z, t|x, t + \Delta t) dz = 0, \quad k > 0,$$

and the third term of equation (A.3) can be rewritten as follows:

$$\begin{aligned} & \lim_{\Delta t \rightarrow 0} (\mu(x) + \lambda(x)) \int_{-\infty}^{\infty} g(z, t|x, t + \Delta t) E(z, t) dz \\ & \quad = (\mu(x) + \lambda(x)) E(x, t). \end{aligned} \tag{A.5}$$

The same logic applied to the second term of equation

$$\lim_{\Delta t \rightarrow 0} \lambda(x) \left[\int_{-\infty}^{\infty} g(z, t|x, t + \Delta t) E(z, t) dz \right]^2 = \lambda(x) E(x, t)^2. \tag{A.6}$$

After substituting in the Taylor expansion from equation (A.4), the fourth term of equation (A.3) becomes

$$\begin{aligned} & \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} \int_{-\infty}^{\infty} g(z, t|x, t + \Delta t) \\ & \quad \left((z - x) \frac{\partial E(x, t)}{\partial x} + \frac{(z - x)^2}{2} \frac{\partial^2 E(x, t)}{\partial x^2} + O((z - x)^3) \right) dz. \end{aligned}$$

This can be simplified using the diffusion conditions from equation (1) to give

$$\phi(x, t) \frac{\partial E(x, t)}{\partial x} + \frac{\sigma^2(x, t)}{2} \frac{\partial^2 E(x, t)}{\partial x^2}. \tag{A.7}$$

Substituting equations (A.5), (A.6), and (A.7) into equation (A.3) gives the partial differential equation (3).

APPENDIX 2

Multivariate Character Derivation

I start with the two-character case from which the extension to an arbitrary number of characters immediately follows. Suppose we have two characters, x and y . Let $g(a, b, t|x, y, t + \Delta t)$ be the probability density that the character changes from x, y at time $t + \Delta t$ to state a, b at time t , where t is closer to the present than $t + \Delta t$ ($0 < t < t + \Delta t$). The functions E and D_N are now functions of both character variables; $E(x, y, t)$ and $D_N(x, y, t)$.

Continuing as for the single character case, $E(x, y, t + \Delta t)$ can be written

$$\begin{aligned} E(x, y, t + \Delta t) &= \mu(x, y) \Delta t + (1 - \mu(x, y) \Delta t) \lambda(x, y) \Delta t \\ & \quad \times \left[\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(a, b, t|x, y, t + \Delta t) E(a, b, t) da db \right]^2 \\ & \quad + (1 - \mu(x, y) \Delta t) (1 - \lambda(x, y) \Delta t) \\ & \quad \times \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(a, b, t|x, y, t + \Delta t) E(a, b, t) da db \\ & \quad + O(\Delta t^2). \end{aligned} \tag{B.1}$$

Dropping terms of $O(\Delta t^2)$, subtracting $E(x, y, t + \Delta t)$ from both sides, dividing by Δt , and rearranging using the fact that

$$E(x, y, t) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(a, b, t|x, y, t + \Delta t) E(x, y, t) da db,$$

we have

$$\begin{aligned} & \frac{E(x, y, t + \Delta t) - E(x, y, t)}{\Delta t} \\ &= \mu(x, y) + \lambda(x, y) \\ & \times \left[\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(a, b, t|x, y, t + \Delta t) E(a, b, t) da db \right]^2 \\ & - (\mu(x, y) + \lambda(x, y)) \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(a, b, t|x, y, t + \Delta t) \\ & \times E(a, b, t) da db \\ & + \frac{1}{\Delta t} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(a, b, t|x, y, t + \Delta t) \\ & \times (E(a, b, t) - E(x, y, t)) da db \\ & + O(\Delta t). \end{aligned} \quad (\text{B.2})$$

To simplify equation (B.2), we need an expression for $E(a, b, t)$ in terms of the original location (x, y) . This can be expanded as a Taylor series in two variables around the point $(a = x, b = y)$:

$$\begin{aligned} E(a, b, t) &= E(x, y, t) + (a - x) \frac{\partial E}{\partial x} + \frac{(a - x)^2}{2} \frac{\partial^2 E}{\partial x^2} \\ & + (b - y) \frac{\partial E}{\partial y} + \frac{(b - y)^2}{2} \frac{\partial^2 E}{\partial y^2} \\ & + (a - x)(b - y) \frac{\partial^2 E}{\partial x \partial y} + O((\Delta x, \Delta y)^3), \end{aligned} \quad (\text{B.3})$$

where $O((\Delta x, \Delta y)^3)$ includes the terms $O((a - x)^3)$, $O((b - y)^3)$, $O((a - x)^2(b - y))$, and $O((a - x)(b - y)^2)$.

We take the limit $\Delta t \rightarrow 0$ for equation (B.2) and consider each term in sequence. Using the same logic as the one character case,

$$\lim_{\Delta t \rightarrow 0} g(a, b, t|x, y, t + \Delta t) = \delta(a - x)\delta(b - y),$$

so

$$\begin{aligned} & \lim_{\Delta t \rightarrow 0} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (a - x)^{k_x} (b - y)^{k_y} \\ & \times g(a, b, t|x, y, t + \Delta t) da db = 0 \quad \text{for } k_x, k_y > 0. \end{aligned}$$

With this, the second and third terms of equation (B.2) can be written

$$\begin{aligned} & \lim_{\Delta t \rightarrow 0} \lambda(x, y) \left[\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(a, b, t|x, y, t + \Delta t) E(x, y, t) da db \right]^2 \\ & = \lambda(x, y) E(x, y, t)^2, \end{aligned} \quad (\text{B.4})$$

and

$$\begin{aligned} & \lim_{\Delta t \rightarrow 0} (\mu(x, y) + \lambda(x, y)) \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(a, b, t|x, y, t + \Delta t) \\ & \times E(x, y, t) da db = (\mu(x, y) + \lambda(x, y)) E(x, y, t) \end{aligned} \quad (\text{B.5})$$

After substituting in the Taylor expansion (B.3) into the fourth term of equation (B.2) and taking the limit $\Delta t \rightarrow 0$, we have

$$\begin{aligned} & \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(a, b, t|x, y, t + \Delta t) \\ & \times \left((a - x) \frac{\partial E}{\partial x} + \frac{(a - x)^2}{2} \frac{\partial^2 E}{\partial x^2} + (b - y) \frac{\partial E}{\partial y} + \frac{(b - y)^2}{2} \frac{\partial^2 E}{\partial y^2} \right. \\ & \left. + (a - x)(b - y) \frac{\partial^2 E}{\partial x \partial y} + O((\Delta x, \Delta y)^3) \right) da db. \end{aligned}$$

A similar set of assumptions to those in equation (1) can be made. Consider first the derivatives involving just one character. Rearranging the first of these gives

$$\begin{aligned} & \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(a, b, t|x, y, t + \Delta t) (a - x) \frac{\partial E}{\partial x} da db \\ & = \frac{\partial E}{\partial x} \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} \int_{-\infty}^{\infty} (a - x) \\ & \times \int_{-\infty}^{\infty} g(a, b, t|x, y, t + \Delta t) db da. \end{aligned}$$

Defining

$$\int_{-\infty}^{\infty} g(a, b, t|x, y, t + \Delta t) db = g(a, t|x, t + \Delta t),$$

so that integrating over all transitions in y gives the transition probability density function for x . The equation above then becomes

$$\frac{\partial E}{\partial x} \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} \int_{-\infty}^{\infty} (a - x) g(a, t|x, t + \Delta t) da,$$

where the term within the limit is identical to equation (1a). With similar manipulation for the other terms, the diffusion conditions become

$$\begin{aligned} \phi_x(x, y, t) &= \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (a - x) \\ & \times g(a, b, t|x, y, t + \Delta t) da db, \end{aligned} \quad (\text{B.6a})$$

$$\begin{aligned} \sigma_{x,x}(x, y, t) &= \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (a - x)^2 \\ & \times g(a, b, t|x, y, t + \Delta t) da db, \end{aligned} \quad (\text{B.6b})$$

$$\begin{aligned} \sigma_{x,y}(x, y, t) &= \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (a - x)(b - y) \\ & \times g(a, b, t|x, y, t + \Delta t) da db, \end{aligned} \quad (\text{B.6c})$$

$$\begin{aligned} 0 &= \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (a - x)^{k_x} (b - y)^{k_y} \\ & \times g(a, b, t|x, y, t + \Delta t) da db, \quad k_x + k_y > 2, \end{aligned} \quad (\text{B.6d})$$

with analogous expressions for $\phi_y(x, y, t)$ and $\sigma_{y,y}(x, y, t)$. The term $\sigma_{x,y}(x, y, t)$ represents the instantaneous covariance between x and y .

Substituting equations (B.4), (B.5), and (B.6) into equation (B.2) and applying the limit to the left-hand side gives

$$\begin{aligned} \frac{\partial E(x, y, t)}{\partial t} &= \mu(x, y) + \lambda(x, y)E(x, y, t)^2 \\ &\quad - (\lambda(x, y) + \mu(x, y))E(x, y, t) \\ &\quad + \phi_x(x, y, t) \frac{\partial E(x, y, t)}{\partial x} + \phi_y(x, y, t) \frac{\partial E(x, y, t)}{\partial y} \\ &\quad + \frac{\sigma_{x,x}(x, y, t)}{2} \frac{\partial^2 E(x, y, t)}{\partial x^2} \\ &\quad + \frac{\sigma_{y,y}(x, y, t)}{2} \frac{\partial^2 E(x, y, t)}{\partial y^2} \\ &\quad + \sigma_{x,y}(x, y, t) \frac{\partial^2 E(x, y, t)}{\partial x \partial y}, \end{aligned} \tag{B.7}$$

and a similar process leads to the partial differential equation

$$\begin{aligned} \frac{\partial D_N(x, y, t)}{\partial t} &= 2\lambda(x, y)D_N(x, y, t)E(x, y, t) \\ &\quad - (\lambda(x, y) + \mu(x, y))D_N(x, y, t) \\ &\quad + \phi_x(x, y, t) \frac{\partial D_N(x, y, t)}{\partial x} \\ &\quad + \phi_y(x, y, t) \frac{\partial D_N(x, y, t)}{\partial y} \\ &\quad + \frac{\sigma_{x,x}(x, y, t)}{2} \frac{\partial^2 D_N(x, y, t)}{\partial x^2} \\ &\quad + \frac{\sigma_{y,y}(x, y, t)}{2} \frac{\partial^2 D_N(x, y, t)}{\partial y^2} \\ &\quad + \sigma_{x,y}(x, y, t) \frac{\partial^2 D_N(x, y, t)}{\partial x \partial y} \end{aligned} \tag{B.8}$$

This can be extended to an arbitrary number of characters to give equations (9) and (10).

Similar logic can also be used to derive equations when speciation, extinction, and character transition functions also depend on the state of an additional binary character; let $\lambda_i(x)$ and $\mu_i(x)$ denote the speciation and extinction function in a continuous character state x while in the binary state i , where $i = 0$ or 1 , $\phi_i(x, t)$ and $\sigma_i^2(x, t)$ be the directional and diffusion functions, while in state i , and $q_{ij}(x)$ be the rate of transition from binary state i to j , which may depend on the continuous trait. The variables become $E_i(x, t)$ and $D_{iN}(x, t)$, which are the probability of extinction or of the lineage leading to node N (respectively) for the lineage in binary state i and continuous state x at time t . Following logic similar to above and in Maddison et al. (2007) yields the equations

$$\begin{aligned} \frac{\partial E_i(x, t)}{\partial t} &= \mu_i(x) + \lambda_i(x)E_i(x, t)^2 \\ &\quad - (\lambda_i(x) + \mu_i(x) + q_{ij}(x))E_i(x, t) \\ &\quad + q_{ij}(x)E_j(x, t) + \phi_i(x, t) \frac{\partial E_i(x, t)}{\partial x} \\ &\quad + \frac{\sigma_i^2(x, t)}{2} \frac{\partial^2 E_i(x, t)}{\partial x^2}, \end{aligned} \tag{B.9a}$$

$$\begin{aligned} \frac{\partial D_{iN}(x, t)}{\partial t} &= 2\lambda_i(x)D_{iN}(x, t)E_i(x, t) \\ &\quad - (\lambda_i(x) + \mu_i(x) + q_{ij}(x))D_{iN}(x, t) \\ &\quad + q_{ij}(x)D_{jN}(x, t) + \phi_i(x, t) \frac{\partial D_{iN}(x, t)}{\partial x} \\ &\quad + \frac{\sigma_i^2(x, t)}{2} \frac{\partial^2 D_{iN}(x, t)}{\partial x^2}. \end{aligned} \tag{B.9b}$$