Estimating a Binary Character's Effect on Speciation and Extinction

WAYNE P. MADDISON,[1,2,3,4] PETER E. MIDFORD,[1] AND SARAH P. OTTO[1,2]

[1]Department of Zoology, University of British Columbia, Vancouver, BC V6T 1Z4, Canada; E-mail: wmaddisn@interchange.ubc.ca (W.P.M.)
[2]Biodiversity Research Centre, University of British Columbia, Vancouver, BC V6T 1Z4, Canada
[3]Department of Botany, University of British Columbia, Vancouver, BC V6T 1Z4, Canada
[4]Wissenschaftskolleg zu Berlin, Wallotstraße 19, Berlin 14193, Germany

Abstract.—Determining whether speciation and extinction rates depend on the state of a particular character has been of long-standing interest to evolutionary biologists. To assess the effect of a character on diversification rates using likelihood methods requires that we be able to calculate the probability that a group of extant species would have evolved as observed, given a particular model of the character's effect. Here we describe how to calculate this probability for a phylogenetic tree and a two-state (binary) character under a simple model of evolution (the "BiSSE" model, binary-state speciation and extinction). The model involves six parameters, specifying two speciation rates (rate when the lineage is in state 0; rate when in state 1), two extinction rates (when in state 0; when in state 1), and two rates of character state change (from 0 to 1, and from 1 to 0). Using these probability calculations, we can do maximum likelihood inference to estimate the model's parameters and perform hypothesis tests (e.g., is the rate of speciation elevated for one character state over the other?). We demonstrate the application of the method using simulated data with known parameter values. [Birth-death process; branching process; cladogenesis; extinction; key innovation; macroevolution; phylogeny; speciation; speciose; statistical inference.]

The pattern of branching of a phylogenetic tree contains information about the processes of speciation and extinction (Nee et al., 1994b; Barraclough and Nee, 2001). For instance, extinction may be revealed by an upturn near the present in a plot of species lineages through time (Nee et al., 1994a). Of special interest is whether phylogenetic trees can be used to demonstrate that certain characteristics of a lineage, such as ecological niche or mating system, affect the rate of speciation or extinction (Mitter et al., 1988; Barraclough et al., 1998; Gittleman and Purvis, 1998). Often used to answer these questions are sister-clade analyses (Mitter et al. 1988; Farrell et al. 1991; Barraclough et al., 1998; Vamosi and Vamosi 2005). For example, Mitter et al. (1988) showed that herbivorous clades of beetles were more speciose than their carnivorous sister clades; this pattern indicates that herbivory confers either a higher speciation and/or a lower extinction rate. Comparison of sister clades is a simple and relatively nonparametric approach (Slowinski and Guyer, 1993; Barraclough et al., 1996) and has had a broad impact on macroevolutionary studies. However, it has some drawbacks that prompt us to explore alternatives. Sister-clade comparisons cannot distinguish differential speciation from differential extinction (Barraclough and Nee, 2001). Also, when the character of interest is a simple categorical variable, clades with mixed states cannot easily participate in the test. Then, the choice of clades can be arbitrary, and information is discarded when collapsing the phylogenetic tree into a set of clade pairs. In principle it should be possible to find a method considering the whole tree that is more powerful than those selecting a subset of clade pairs (Chan and Moore, 2002; Ree, 2005; Paradis, 2006).

Ideally, we would like a likelihood-based approach to estimate the effect of a character on diversification rates. The underlying likelihood model would allow speciation and extinction rates to depend on the character state of a lineage at each point in time and allow the character state to change. Inferences about speciation and extinc-tion rates as a function of character state could then be made based on their likelihood: the probability of observing the data (the phylogeny and the current character states) given proposed values for the rate parameters. In this paper, we describe a method for calculating this likelihood when the character controlling diversification has two states (i.e., is binary). We show how this calculation leads to new methods for parameter estimation and hypothesis testing about a binary character's effect on diversification. The calculations can also be readily applied to a broader set of questions about character evolution or indeed about tree inference itself, although we will not address these questions here.

Likelihood calculations for models involving speciation and extinction rates (Nee et al., 1994b, Moore et al., 2004) and rates of character state change (Pagel, 1994) have been described separately, but have not yet been fully integrated. Notable contributions to this effort have been made by Pagel (1997), Paradis (2005), and Ree (2005). Pagel's (1997) model permits different character states to confer different rates of speciation, but it assumes no extinction and that character change occurs only at speciation events. Paradis (1995) and Ree (2005) present likelihood-based methods that use reconstructed ancestral states to compare speciation rates between states (again, ignoring extinction). Most importantly, these previous methods have all assumed that ancestral states can be assessed without accounting for the effects of the character on speciation/extinction processes. This is problematic, because the reconstruction of a particular ancestral state depends critically on how the character affects speciation and extinction rates. For example, imagine a phylogeny in which one character state promotes speciation compared to a second state, resulting in a greater proportion of extant species with the first state. If we attempted to understand character change without taking into account the character's effect on speciation, we would interpret the state's abundance to reflect a higher rate of change toward it, even if in

fact the rates of change were equal to and from the state (Maddison, 2006). This in turn would bias the ancestral state reconstruction and mislead us about the process of diversification. Our approach here is the first to incorporate explicitly the process of character state change directly and simultaneously into the likelihood assessment of speciation and extinction rates.

Here we describe how to calculate the probability of the tree and observed states given a basic model with six parameters: the instantaneous rates of speciation and extinction when the lineage is in state 0 (e.g., herbivory), the rates when the lineage is in state 1 (e.g., carnivory), and the instantaneous rates of character state change (0 to 1 and 1 to 0). We will occasionally abbreviate this model as the BiSSE (binary-state speciation and extinction) model. Following a presentation of the model and likelihood calculations, we apply the method to some simulated data sets.

LIKELIHOOD OF THE BiSSE MODEL

We assume that an accurate rooted phylogenetic tree with branch lengths is known (the "inferred tree") and that the character state is known for each of the terminal taxa. (Alternatively, our methods could be applied to each of the fully resolved trees coming from a Bayesian MCMC analysis—Yang and Rannala, 1997; Larget and Simon, 1999.) The tree is assumed complete: all extant species in the group have been found and included. We consider only binary characters, but extending the method to multistate characters would be straightforward. All terminal taxa are contemporaneous, and the tree is ultrametric (i.e., the total root-to-tip distance is the same for all tips). Except where noted, we measure time backwards, with 0 being the present.

The parameters of the model are as follows. While a lineage has character state 0, the instantaneous speciation rate is λ_0, the extinction rate is μ_0, and the rate of transition to state 1 is q_{01}. Similarly, while a lineage has character state 1, the speciation, extinction, and transition rates are λ_1, μ_1, and q_{10}. These parameters are assumed constant throughout the tree, although it would be straightforward to extend the model to explore hypotheses about changes to these parameters. We assume that the transitions happen instantaneously over the time scales considered (i.e., we ignore periods of time during which a species is polymorphic). We also assume that these events are independent of one another; in particular, we assume that the character state change does not, in and of itself, cause speciation (or vice versa).

Probability of the Tree and Character States (D)

Although the rates reflect probabilities of events moving forward in time, our calculations will move backward in time, from the tips of the tree to the root. This is the well-established "pruning" (Felsenstein, 1981) or "downpass" (Maddison and Maddison, 1992) approach and is used to handle more compactly all the possibilities of ancestral states at various parts of the tree (Felsenstein, 1981). This approach uses a simple principle: if we are
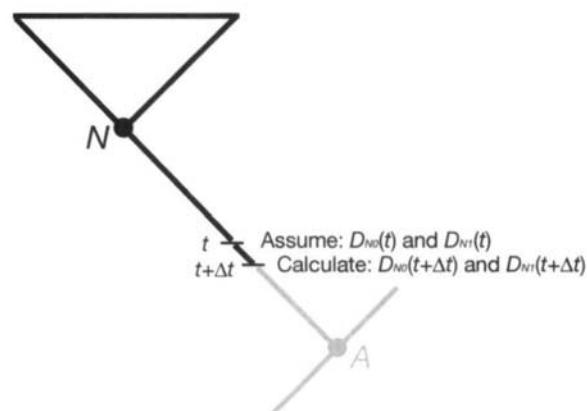

FIGURE 1. Calculation of the probabilities (D) of the observed tree and character states, along a branch of the tree. We assume that we know the D's for time t on the branch and attempt to calculate them for time $t+\Delta t$.

able to use key probabilities at any point on a tree to derive corresponding probabilities immediately ancestral (i.e., closer to the root), then it must be possible to move step-by-step down the tree toward the root. When the root is reached, the calculated probabilities will apply to the whole tree.

Our calculations make use of the full tree topology, unlike those of Nee et al. (1994b), which use only timing of branching events. We need to use the full tree topology because we are considering simultaneously the evolution of character states. In Appendix 1 we describe how removing the dependence on character states reduces our equations to those of Nee et al.

In the BiSSE model, the key probabilities, $D_{N0}(t)$ or $D_{N1}(t)$, describe the chance that a lineage beginning at time t with state 0 or 1 would evolve into a clade like that observed to have descended from node N, closer to the present (Fig. 1). We track these probabilities back by a very short amount of time, Δt, toward the root, accounting for all possible events that could have happened along the way. If we use a small enough time interval Δt, we can ignore the possibility that more than one event happens during the time interval. Once we have derived equations for the change in the probability over Δt, we then shrink the time interval and use the definition of a derivative to obtain differential equations describing the change in these probabilities as we descend toward the root. By integrating these differential equations along the branches, we are able to solve for the overall probability of the data given the BiSSE model. In the following, we carry out these calculations, first along a branch and then across nodes in the tree.

Calculations within a branch.—Let us assume that we have already calculated the probability $D_{N0}(t)$ that the clade of node N and its stem lineage would have evolved exactly as represented by the inferred tree (including branch lengths and terminal character states) given that the lineage started in state 0 at time t on the branch below node N (Fig. 1). We will describe the calculations focused on state 0; the calculations for state 1 are parallel and indeed need to be performed simultaneously.
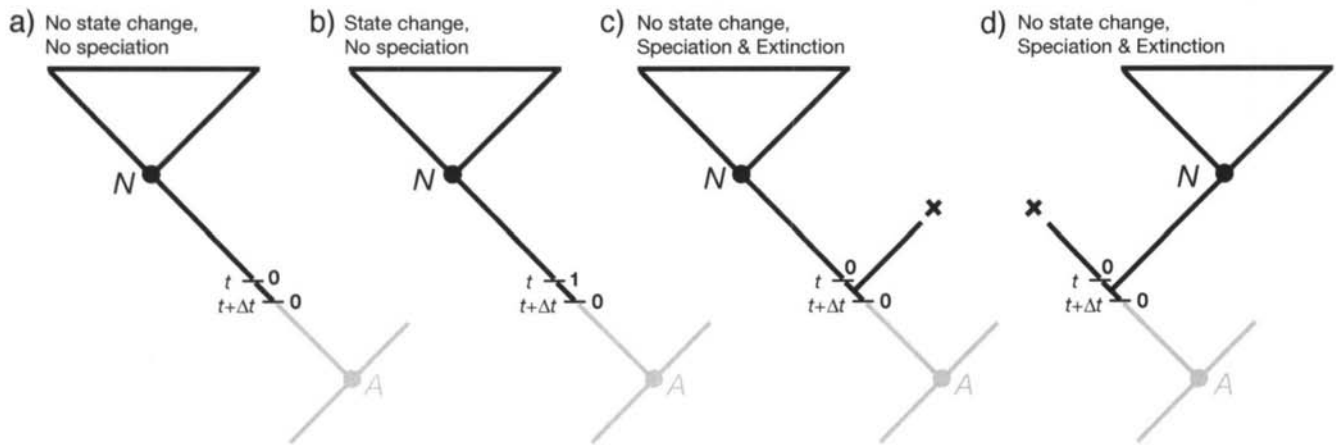
FIGURE 2. Alternative scenarios by which a lineage with state 0 at time $t+\Delta t$ on the branch might yield clade decended from node N but no other living descendants.

To move the next step down the tree (Fig. 1), we need to calculate corresponding probabilities for one small time step (Δt) further down the branch; i.e., $D_{N0}(t + \Delta t)$ and $D_{N1}(t + \Delta t)$. To calculate $D_{N0}(t + \Delta t)$, we enumerate all of the ways that the observed clade could have arisen by considering all possible events that could occur in the Δt time interval. Specifically, we use the law of total probability to write $D_{N0}(t + \Delta t)$ as the sum of all possible transitions forward in time from $t + \Delta t$ to t times the probability that the clade would have evolved in the manner observed from time t to the present. There are four cases, each of which requires that the lineage of interest did not go extinct during the Δt interval (Fig. 2). In the first case (Fig. 2a), nothing happens in the Δt interval. In the second case (Fig. 2b), the character changes. In the third case (Fig. 2c), a speciation event occurs and the left lineage generates node N; the right lineage must thus go extinct sometime before the present, which occurs with probability $E_0(t)$, described further in the next section. The fourth case (Fig. 2d) is similar, except that the left lineage goes extinct and the right lineage generates node N. No other cases contribute to the probability $D_{N0}(t + \Delta t)$ because of the assumptions that the events are independent and the clade had to survive to the present.

The probabilities for these four cases can then be summed to yield $D_{N0}(t + \Delta t)$:

$D_{N0}(t+\Delta t) =$

$(1 - \mu_0 \Delta t) \times$ (in all cases no extinction in Δt)

$\quad [(1 - q_{01}\Delta t)(1 - \lambda_0 \Delta t)D_{N0}(t)$ (see Fig. 2a: No state change, no speciation)

$\quad + (q_{01}\Delta t)(1 - \lambda_0 \Delta t)D_{N1}(t)$ (see Fig. 2b: State change, no speciation)

$\quad + (1 - q_{01}\Delta t)(\lambda_0 \Delta t)E_0(t)D_{N0}(t)$ (see Fig. 2c: No state change, speciation, extinction)

$\quad + (1 - q_{01}\Delta t)(\lambda_0 \Delta t)E_0(t)D_{N0}(t)]$ (see Fig. 2d: No state change, speciation, extinction)

$\quad + (\mu_0 \Delta t) \times 0$ (if the lineage went extinct, the clade has zero probability of being observed)

$$(1)$$

For clarity in Equation (1), we have ignored several possible transitions that involve multiple events within the time interval (e.g., speciation and extinction), because these occur with a probability (of order Δt^2) that is neg-

ligibly small if Δt is small. Dropping all terms of order Δt^2, we get:

$$D_{N0}(t + \Delta t) = [1 - (\lambda_0 + \mu_0 + q_{01})\Delta t]D_{N0}(t) + (q_{01}\Delta t)$$
$$D_{N1}(t) + 2(\lambda_0 \Delta t)E_0(t)D_{N0}(t) \qquad (2a)$$

Similarly,

$$D_{N1}(t + \Delta t) = [1 - (\lambda_1 + \mu_1 + q_{10})\Delta t]D_{N1}(t) + (q_{10}\Delta t)$$
$$D_{N0}(t) + 2(\lambda_1 \Delta t)E_1(t)D_{N1}(t) \qquad (2b)$$

Dividing $[D_{N0}(t + \Delta t) - D_{N0}(t)]$ and $[D_{N1}(t + \Delta t) - D_{N1}(t)]$ by the time interval, Δt, and taking the limit as Δt goes to zero, we can derive two coupled differential equations:

$$\frac{dD_{N0}}{dt} = -(\lambda_0 + \mu_0 + q_{01})D_{N0}(t) + q_{01}D_{N1}(t)$$
$$+ 2\lambda_0 E_0(t)D_{N0}(t) \qquad (3a)$$

$$\frac{dD_{N1}}{dt} = -(\lambda_1 + \mu_1 + q_{10})D_{N1}(t) + q_{10}D_{N0}(t)$$
$$+ 2\lambda_1 E_1(t)D_{N1}(t) \qquad (3b)$$

We have not found analytical solutions for these equations. Nevertheless, given solutions for $E_0(t)$ and $E_1(t)$ described below, they can be numerically integrated

along a branch. This permits us to derive $D_{N0}(t)$ and $D_{N1}(t)$ at the bottom of the branch (rootward) from the probabilities at the top of the branch.

Initial conditions.—If N is a terminal node with state 0, then the initial conditions are $D_{N0}(0) = 1$ and $D_{N1}(0) = 0$. That is, the state must be that observed at time 0. Likewise, if the taxon has state 1, $D_{N0}(0) = 0$ and $D_{N1}(0) = 1$.

Calculations at nodes.—When we arrive at the bottom of the branch, i.e., at the immediate ancestral node A, we need to combine the probability from the branch with the probability from its sister branch before continuing the journey down the ancestral branch. For A's clade to have evolved as represented in the inferred tree, the speciation event must have occurred, and each of the daughter lineages must have survived to the present day. The probability of the A clade evolving as in the inferred tree equals the probability that the left daughter lineage evolved as in the tree, times the probability that the right daughter lineage evolved as in the tree, times the probability of speciation across a small interval of time, Δt. This probability of speciation, $\lambda \Delta t$, thus introduces a Δt term at each node. These Δt terms therefore contribute a factor $(\Delta t)^n$ to the probability over the whole tree, where n is the number of internal nodes. Because this factor does not depend on the parameter values, we can safely ignore it without changing the proportionality of the likelihoods under different parameter values. Technically, ignoring it converts our probabilities to probability densities, but we will continue to speak of probabilities for brevity.

Thus, right before the speciation event has occurred, the probability that the lineage is in state 0 or 1 and evolves into a clade like that observed to have descended from node A (including the speciation event at node A) is given by:

$$D_{A0}(t_A) = D_{N0}(t_A)D_{M0}(t_A)\lambda_0 \qquad (4a)$$

$$D_{A1}(t_A) = D_{N1}(t_A)D_{M1}(t_A)\lambda_1 \qquad (4b)$$

where M is the sister node to N. Note that these equations include only the case where the character state is the same for the ancestor A and its two descendant lineages. That is, we treat speciation and character state changes as independent events, so that the probability is zero for both changes to occur simultaneously.

Now, having passed node A, we can continue toward the root, using $D_{A0}(t_A)$ and $D_{A1}(t_A)$ as starting points for the numerical integration down A's branch.

At the root.—When we get to the root R we will have calculated $D_{R0}(t_R)$ and $D_{R1}(t_R)$, which describe the probability of observing the phylogeny and the extant character states, given that the root was in state 0 or 1, respectively. To obtain a single likelihood to serve as the basis for inference we need to account for the probability that the root was in state 0 or 1. We could, following Schluter et al. (1997) and Pagel (1999), add D_{R0} and D_{R1} together. This effectively assigns a probability of 0.5 to the root being in state 0 and to the root being in state 1, even if the transition probabilities of the BiSSE model would make it much more likely that the character is in one state or the other. Alternatively, we could weight these root states by the equilibrium frequencies of 0 and 1 implicit in the model (see Appendix 2), as done by default in Mesquite version 1.1 and later for similar likelihood calculations for ancestral states (Maddison and Maddison, 2006). We take this latter approach here.

Probability of Extinction (E)

In order to use the differential Equations (3) to calculate the probability, $D_{N0}(t)$ or $D_{N1}(t)$, that a lineage evolves as represented in the inferred tree, we must determine the probability that a lineage alive at time t goes extinct before the present time. Here, we derive $E_0(t)$ and $E_1(t)$ using a similar procedure, except that the extinction probabilities do not depend on tree structure of the surviving lineages, only on time.
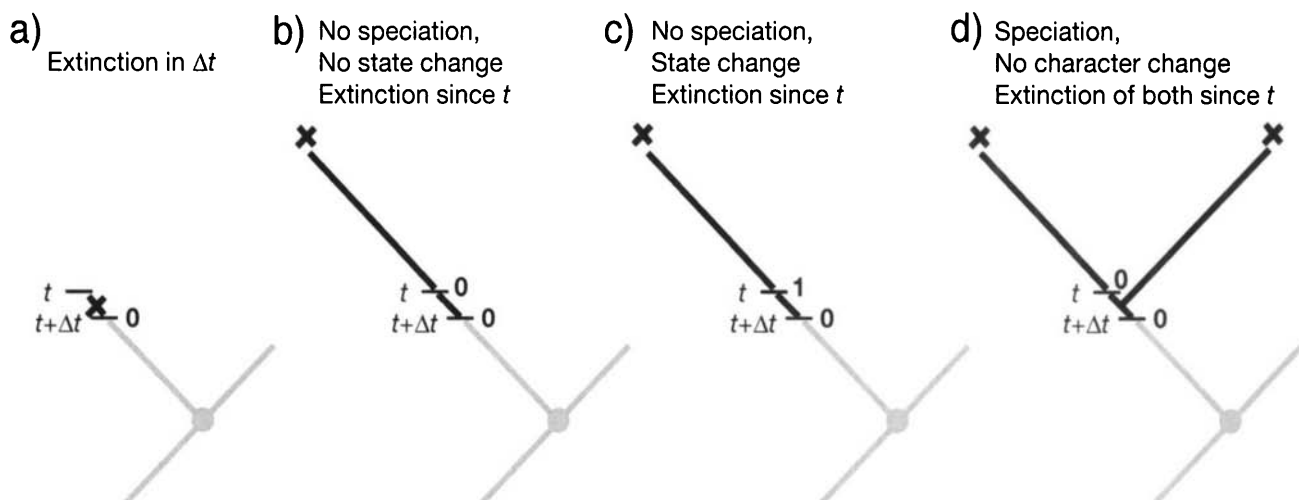

FIGURE 3.   Alternative scenarios by which a lineage at time t with state 0 might go extinct.

Assume that we have calculated $E_0(t)$, the probability that a lineage starting at time t in state 0 leaves no descendants at the present day. Then $E_0(t + \Delta t)$ can be obtained by considering the four different possible events in the Δt interval consistent with the lineage (and all of its descendants) going extinct ultimately (Fig. 3). In the first case (Fig. 3a), the lineage goes extinct during the Δt time interval. In the second case (Fig. 3b), the lineage neither goes extinct nor changes state nor speciates between t and $t + \Delta t$, but nevertheless it goes extinct eventually. In the third case (Fig. 3c), the lineage changes state during the Δt time interval and then goes extinct. Finally, in the fourth case (Fig. 3d), the lineage speciates during the Δt time interval, but now both descendant lineages must go extinct before the present day. We assume in the fourth case that the extinction events are independent of one another, thus contributing a term $E_0(t)^2$ to the probability. No other cases need be considered because of the assumption that the events are independent and Δt is extremely small.

The probabilities for these four cases can then be summed to yield the probability of extinction, $E_0(t + \Delta t)$:

$$E_0(t + \Delta t) =$$
$$\mu_0 \Delta t$$ (see Fig. 3a: Extinction in Δt)
$$+ (1 - \mu_0 \Delta t)(1 - q_{01} \Delta t)(1 - \lambda_0 \Delta t) E_0(t)$$ (see Fig. 3b: No state change, no speciation)
$$+ (1 - \mu_0 \Delta t)(q_{01} \Delta t)(1 - \lambda_0 \Delta t) E_1(t)$$ (see Fig. 3c: State change, no speciation)
$$+ (1 - \mu_0 \Delta t)(1 - q_{01} \Delta t)(\lambda_0 \Delta t) E_0(t)^2$$ (see Fig. 3d: No state change, speciation)

$$(5)$$

Again, we have included some terms in Equation (5) that are negligibly small (of order Δt^2). Dropping these terms, we get:

$$E_0(t + \Delta t) = \mu_0 \Delta t + [1 - (\mu_0 + q_{01} + \lambda_0) \Delta t] E_0(t)$$
$$+ (q_{01} \Delta t) E_1(t) + (\lambda_0 \Delta t) E_0(t)^2 \qquad (6a)$$

Similarly,

$$E_1(t + \Delta t) = \mu_1 \Delta t + [1 - (\mu_1 + q_{10} + \lambda_1) \Delta t] E_1(t)$$
$$+ (q_{10} \Delta t) E_0(t) + (\lambda_1 \Delta t) E_1(t)^2 \qquad (6b)$$

Dividing the change in E_0 and E_1 by the time interval, Δt and taking the limit as Δt goes to zero yields the coupled differential equations:

$$\frac{dE_0}{dt} = \mu_0 - (\mu_0 + q_{01} + \lambda_0) E_0(t) + q_{01} E_1(t) + \lambda_0 E_0(t)^2$$
$$(7a)$$

$$\frac{dE_1}{dt} = \mu_1 - (\mu_1 + q_{10} + \lambda_1) E_1(t) + q_{10} E_0(t) + \lambda_1 E_1(t)^2$$
$$(7b)$$

As with the equations for $D(t)$, these equations for the extinction probabilities can be solved using numerical integration.

Initial conditions.—At time $t = 0$, there is no time for extinction, and thus $E_0(0) = E_1(0) = 0$.

APPLICATION TO ESTIMATION AND TESTING

Maximizing the BiSSE likelihood, calculated as described above, can yield estimates for all six parameters. This means that the method could be used not just to understand speciation and extinction, but the entire process of diversification including character state change. With such a flexible model, a variety of hypothesis tests are possible. For example, is speciation under state 0 more rapid than under state 1? Is net diversification rate (speciation − extinction) higher under one state than the other? What causes an excess of species with one state (Maddison, 2006): asymmetrical character change (rate 0 to 1 different from rate 1 to 0), asymmetrical extinction, or asymmetrical speciation? To answer these questions, one can perform likelihood ratio tests by comparing likelihoods given unconstrained versus appropriately constrained models.

In the following, we provide an initial exploration of BiSSE using simulated phylogenetic trees and characters. First, we estimate rate parameters when true speciation, extinction, and character transition rates either do or do not depend on the character state. Second, we examine whether the null hypothesis of equal rates can be rejected using likelihood-ratio tests.

These examples do not attempt to provide a full exploration of the method. To understand the method's capabilities and limitations, we would want to study the power of the method to reject false null hypotheses under a variety of parameter values and assumptions, including cases with multiple simultaneous asymmetries. We would also like to know the size of the tree needed for adequate power, as well as the accuracy and precision of estimated parameter values. These explorations we leave for a subsequent paper.

Implementation

The BiSSE likelihood calculations described above have been programmed in the Diverse package of modules (Midford and Maddison, 2007) for Mesquite (Maddison and Maddison, 2007). These calculations use a fourth-order Runge-Kutta method of numerical integration (Ralson and Rabinowitz, 1978) for proceeding along the branches and Mesquite's implementation of Brent's (1973) optimizer for seeking maximum likelihood estimates. If the choice is made to condition the

probabilities on the survival of the entire clade, the module uses Nee at al.'s (1994b) convention (i.e., the condition used is that at least two descendants of the clade survive).

In our examination of the method, maximum likelihood optimization was first attempted using 10 random starting parameter values, with the numerical integrator using a coarse division of each branch into segments (an average-length branch was assigned 100 segments; other branches were assigned up to a maximum of 400 or a minimum of 50 segments in proportion to their length). Error in the numerical integration is expected to decrease as segment length decreases, and therefore the most likely parameter set resulting from the 10 attempts was then used as a starting point for a final optimization using a finer division of branches (average 1000 segments, maximum 4000, minimum 500). Details of the numerical integration method will be considered in a subsequent paper, but we note that the reasonable estimates obtained in our results suggest that the method is having some success. Likelihoods were conditional on clade survival. The prior used for the character state at the root (i.e., the probability of state 0 versus 1 at the root) was the set of equilibrium frequencies implicit in the proposed model parameter values.

Simulated trees and terminal character states were generated in Mesquite via a prerelease version of the "BiSSE Trees & Characters" module of the Diverse package (Midford and Maddison, 2007). This module divided time into small steps, such that if an event's occurrence has an instantaneous rate r the probability of the event in the time slice is $r/1000$. In each time slice on each lineage, a uniform random number was drawn to determine whether the character changed, followed by a random number draw to determine if extinction occurred. If extinction did not occur on a lineage, then a random number was drawn to determine if speciation had occurred. If the tree went extinct, the simulation was restarted. A simulation continued until the tree first had the requested number of species (e.g., 500). This is not ideal, because extinction and speciation could have caused the tree to achieve the requested number of species several times; stopping the simulation at the first time biases the result toward shorter terminal branches. The bias should, however, be very small for the parameters investigated below, where speciation rates are high relative to rates of extinction and character state change and where the size of the tree is large. (For example, under the scenario with the highest extinction rate explored below, the average height of the tree when 500 species was first reached was 0.99916 that of the height of the tree when 500 species was last reached, as determined by 100 simulations in which trees were allowed to grow to 600 species. See also Appendix 2 for a deterministic approximation to average times to reach a certain number of species.)

Parameter Estimation

To explore parameter estimation, we simulated trees under each of four parameter combinations. The first combination was entirely symmetrical: $\lambda_0 = \lambda_1 = 0.1$, $\mu_0 = \mu_1 = 0.03$, $q_{01} = q_{10} = 0.01$. The other three combinations introduced an asymmetry in each process (speciation, extinction, character change) by altering one parameter at a time in the perfectly symmetrical model. Specifically, the three alternative models involved raising the speciation rate with state 1 ($\lambda_1 = 0.2$), raising the extinction rate with state 0 ($\mu_0 = 0.06$), or lowering the rate of character change to state 0 ($q_{10} = 0.005$). Five hundred trees of 500 extant species were simulated under each model.

Maximum likelihood estimates of parameters from the resulting trees were obtained from the Diverse package. The results are plotted in Figure 4, which focuses on the estimate of the asymmetric parameter in each simulation. In general, the estimates of speciation rates λ_0 and λ_1 are fairly good, falling near the correct parameter values. As well, the estimates for the symmetrical case ($\lambda_0 = \lambda_1 = 0.1$) are easily distinguished from the estimates in the asymmetrical case ($\lambda_0 = 0.1$, $\lambda_1 = 0.2$). For extinction rates and rates of character state change, the estimates do not so closely match the simulated rates. Whether this reflects general difficulties in estimating extinction (Nee et al., 1994; Kubo and Iwasa, 1995; Paradis, 2005) or state change or instead depends on the particular parameter values chosen remains to be explored in a future paper.

Hypothesis Testing

We examined the same simulated trees to explore what power the method would have to reject the null hypothesis that a rate (e.g., of speciation) was equal for the two character states, when in fact it was different. In addition to the unconstrained six-parameter likelihood estimates described above, we also calculated the maximum likelihood given constrained five-parameter models, holding either the speciation rates equal ($\lambda_0 = \lambda_1$), the extinction rates equal ($\mu_0 = \mu_1$), or the rates of state change equal ($q_{01} = q_{10}$). The log-likelihood difference between the constrained five-parameter likelihood model and the unconstrained six-parameter likelihood model was used as a test statistic. If this difference was larger than a critical value (based on the simulated trees with truly symmetrical rate parameters), then we rejected the null hypothesis of symmetrical rates for the parameter of interest.

When investigating whether the character affected speciation rates, 5% of the symmetric simulations yielded a 2 × log-likelihood difference of 3.60 or more between the constrained ($\lambda_0 = \lambda_1$) and the unconstrained ($\lambda_0 \neq \lambda_1$) models even though there was no difference in speciation rate. Thus, we used 3.60 as our 5% cutoff for significance. Of the asymmetric simulations, 58% resulted in a greater log-likelihood difference than this cutoff, leading to the rejection of the null hypothesis of equal speciation rates ($\lambda_0 = \lambda_1$). This shows at least some power to reject the null hypothesis with trees of 500 species when there is a twofold difference in speciation rates. Rejection rates with asymmetries in extinction or character state change were, however, much lower. Exploring extinction rate differences, 21.2% of the asymmetric simulations exceeded the cutoff 2 × log-likelihood difference of 4.08 based on the symmetric simulations. Exploring rates
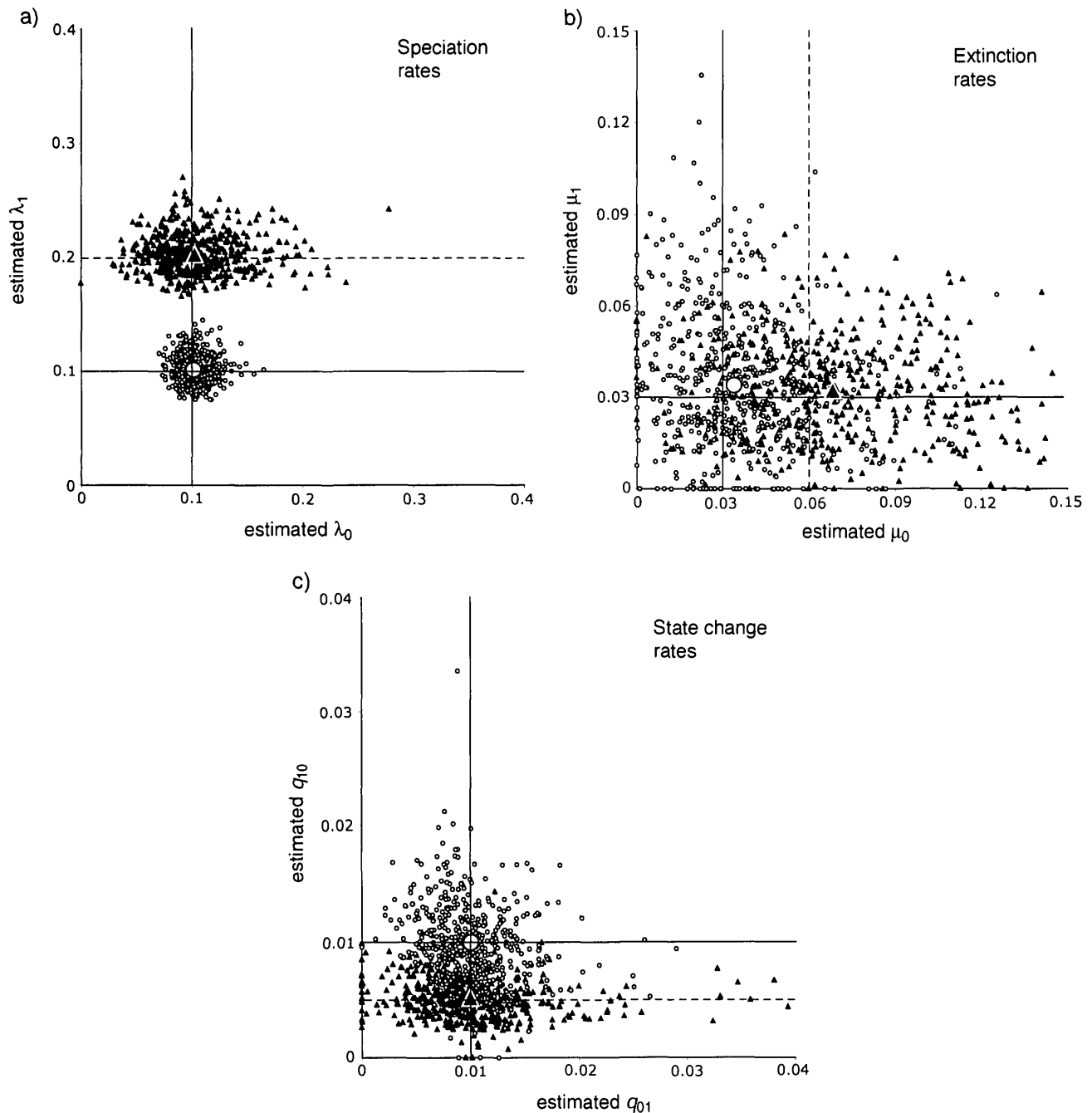
FIGURE 4. Estimated parameter values from simulated trees and characters. Lines indicate true parameter values of the simulations (solid, equal rates for the two states; dashed, unequal). Small symbols represent estimates; large symbols represent mean values of estimates. Open circles represent parameter estimates from symmetrical simulations; i.e., equal parameter values for states 0 and 1 (speciation rates 0.1, extinction rates 0.03, character change 0.01). Closed triangles represent estimates from simulations with asymmetries in the respectively estimated parameters. Asymmetries shown are (a) speciation rate asymmetry ($\lambda_1 = 0.1$ versus 0.2), (b) extinction rate asymmetry ($\mu_0 = 0.03$ versus 0.06), and (c) character state change rate asymmetry ($q_{10} = 0.01$ versus 0.005). In the maximum likelihood analysis, all six parameters were free to vary, but each scatterplot focuses only on the parameters of interest in each case.

of character state change, only 16.4% of the asymmetric simulations exceeded the cutoff $2 \times$ log-likelihood difference of 4.52 based on the symmetric simulations. As we noted above with respect to parameter estimation, these results do not necessarily indicate that extinction or character changes are generally harder to study, only that they are under the parameter values studied.

Given sufficient data, we would expect twice the log likelihood difference from constrained versus unconstrained models to follow a χ^2 distribution with one degree of freedom, under the null hypothesis of equal rates for the two states. The 5% cutoff for significance under this asymptotic distribution would be 3.841. The 5% cutoff values established by our equal-rate simulations are

close to this value, suggesting that the χ^2 approximation may be a reasonable one for likelihood ratio tests involving trees of this size.

It remains to be studied how the probability of rejecting a false null hypothesis (power) varies with number of species and with the degree of difference in rates. From our initial exploration, however, we suspect that large phylogenetic trees will generally be needed to have sufficient power. Rather than being a sign of a weakness in the method, we believe that this low power stems from the fact that there are many similar ways to generate an observed phylogeny and observed character states (Maddison, 2006)—and only with much data can these be distinguished.

Extensions

In this paper, we have used these likelihood calculations to explore our ability to detect asymmetrical rates of speciation, extinction, or character state change. The calculations could equally be applied to exploring how uncertainty in some of these parameters (e.g., speciation and extinction rates) influences the likelihood estimates for other parameters (e.g., character state changes). They could also be used to infer ancestral states using likelihood (Schluter et al., 1997). Maddison (2006) has explained why methods are needed that consider a character's effect on diversification when interpreting the character's evolutionary history; our likelihood calculations can provide such methods. They could also be used in tree inference itself, permitting models in which character states affect diversification processes.

The general approach of describing differential equations for likelihood and integrating down branches should be applicable to a broad array of questions. Using numerical integration methods frees us from the constraint of considering only those models that can be solved analytically. We could, for example, extend the method to include multiple correlated characters or other models of character evolution or other models of speciation (e.g., speciation that occurs simultaneously with character change). Another useful extension would be to deal with continuous valued data, perhaps incorporating a model such as Slatkin's (1981). Paradis's (2005) method is already generalized for multiple characters of different types, although as noted it does not integrate state change with speciation/extinction into a common model.

A final caution about the BiSSE method: even if it proves to have adequate ability to estimate parameters and test hypotheses, it will suffer the same limitation (Read and Nee, 1995; Maddison, 2000) shared by Pagel's (1994) and Maddison's (1990) tests of character correlation and Ree's (2005) test of diversification. If we determine, for example, that $\lambda_1 > \lambda_0$, we cannot on this basis alone say that the character of interest is controlling speciation rate. Another character, with parallel distribution of character states, could in fact be responsible. This would be of little concern if our character's states were scattered with much homoplasy on the tree, because then we

could argue that any other character is unlikely to have a parallel distribution, unless it were causally linked to our examined character, and so indirect causality could still be argued. But, the issue of codistributed characters would be a concern if all of the species with state 1 (for example) occurred in a single clade. Our method, as well as the others cited, could give a significant result in such a case, even though any other synapomorphy of the clade could actually be responsible for the effect, and such a synapomorphy might be unrelated to our character of interest except by coincidence of origin on the same single lineage (Read and Nee, 1995; Maddison, 2000). Thus, the correct conclusion given a significant result using our method is that the character examined *or a codistributed character* appear to be controlling diversification rates.

REFERENCES

Barraclough, T. G., P. H. Harvey, and S. Nee. 1996. Rate of rbcL gene sequence evolution and species diversification in flowering plants (angiosperms). Proc. R. Soc. Lond. B 263:589–591.

Barraclough, T. G., and S. Nee. 2001. Phylogenetics and speciation. Trends Ecol Evol. 16:391–399.

Barraclough, T. G., A. P. Vogler, and P. H. Harvey. 1998. Revealing the factors that promote speciation. Phil. Trans. R. Soc. Lond. B 353:241–249.

Brent, R. P. 1973. Algorithms for optimization without derivatives. Prentice Hall, Englewood Cliffs, New Jersey.

Chan, K. M. A., and B. R. Moore. 2002. Whole-tree methods for detecting differential diversification rates. Syst. Biol. 51:855–865.

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

Gittleman, J. L., and A. Purvis. 1998. Body size and species-richness in carnivores and primates. Proc. R. Soc. Lond. B 265:113–119.

Goudet, J. 1999. An improved procedure for testing the effects of key innovation on rate of speciation. Am. Nat. 153:549–555.

Kendall, D. G. 1948. On some modes of population growth leading to R. A. Fisher's logarithmic series distribution. Biometrika 35:6–15.

Larget, B., and D. L. Simon. 1999. Markov chain Monte Carlo algorithms for the Bayesian analysis of phylogenetic trees. Mol. Biol. Evol. 16:750–759.

Maddison, W. P. 1990. A method for testing the correlated evolution of two binary characters: Are gains or losses concentrated on certain branches of a phylogenetic tree? Evolution 44:539–557.

Maddison, W. P. 2000. Testing character correlation using pairwise comparisons on a phylogeny. J. Theor. Biol. 202: 195–204.

Maddison, W. P. 2006. Confounding asymmetries in evolutionary diversification and character change. Evolution. 60:1743–1746.

Maddison, W. P., and D. R. Maddison. 1992. MacClade. Version 3: Analysis of phylogeny and character evolution. Sinauer Associates, Sunderland, Massachusetts.

Maddison, W. P., and D. R. Maddison. 2007. Mesquite: A modular system for evolutionary analysis. Version 2.0. http://mesquiteproject.org.

Midford, P. E., and W. P. Maddison. 2007. Diverse package for Mesquite. Version 1.0. http://mesquiteproject.org.

Mitter, C., B. Farrell, and B. Wiegmann. 1988. The phylogenetic study of adaptive zones: Has phytophagy promoted insect diversification? Am. Nat. 132:107–128.

Moore, B. R., K. M. A. Chan, and M. J. Donoghue. 2004. Detecting diversification rate variation in supertrees. Pages 487–533 *in* Phylogenetic supertrees: Combining information to reveal the Tree of Life (O. R. P. Bininda-Emonds, ed.). Kluwer Academic, Dordrecht, The Netherlands.

Nee, S., E. C. Holmes, R. M. May, and P. H. Harvey. 1994a. Extinction rates can be estimated from molecular phylogenies. Phil. Trans. R. Soc. Lond. B 344:77–82,

Nee, S., R. M. May, and P. H. Harvey. 1994b. The reconstructed evolutionary process. Phil. Trans. R. Soc. Lond. B 344: 305–311.

Pagel, M. 1994. Detecting correlated evolution on phylogenies: A general method for the comparative analysis of discrete characters. Proc. R. Soc. Lond. Biol. Sci. B 255:37–45.

Pagel, M. 1997. Inferring evolutionary processes from phylogenies. Zool. Scripta 26:331–348.

Pagel, M. 1999. The maximum likelihood approach to reconstructing ancestral character states of discrete characters on phylogenies. Syst. Biol. 48:612–622.

Paradis, E. 2005. Statistical analysis of diversification with species traits. Evolution 59:1–12.

Ralson, A., and Rabinowitz, P. 1978. A first course in numerical analysis, 2nd edition. McGraw-Hill, New York.

Read, A. F., and S. Nee. 1995. Inference from binary comparative data. J. Theor. Biol. 173:99–108.

Ree, R. H. 2005. Detecting the historical signature of key innovations using stochastic models of character evolution and cladogenesis. Evolution 59:257–265.

Schluter, D., T. Price, A. Ø. Mooers, and D. Ludwig. 1997. Likelihood of ancestral states in an adaptive radiation. Evolution 51:1699–1711.

Slatkin, M. 1981. A diffusion model of species selection. Paleobiology 7:421–425.

Slowinski, J. B., and C. Guyer. 1993. Testing whether certain traits have caused amplified diversification: An improved method based on a model of random speciation and extinction. Am. Nat. 142:1019–1024.

Vamosi, S. M., and J. C. Vamosi. 2005. Endless tests: Guidelines for analysing non-nested sister group comparisons. Evol. Ecol. Res. 7:567–579.

Yang, Z. H., and B. Rannala. 1997. Bayesian phylogenetic inference using DNA sequences: A Markov chain Monte Carlo method. Mol. Biol. Evol. 14:717–724.

APPENDIX 1

Character-independent model.—The likelihood model described in the text can be reduced to a simple speciation-extinction (or birth-death) model if one assumes that speciation and extinction rates are not affected by a character. As the speciation-extinction model has been well studied (e.g., Nee et al., 1994b), we describe how our calculations reduce to theirs in the case that the speciation and extinction rates are constant (λ, μ).

Rederiving Equations (3) and (7) without reference to a character, we obtain differential equations for $D_N(t)$, the probability that a lineage at time t gives rise to the extant descendant lineages and no others, and $E(t)$, the probability that a lineage and all of its descendants go extinct before the present day:

$$\frac{dD_N}{dt} = -(\lambda + \mu)D_N(t) + 2\lambda E(t)D_N(t) \qquad (8a)$$

$$\frac{dE}{dt} = \mu - (\mu + \lambda)E(t) + \lambda E(t)^2 \qquad (8b)$$

Using these two differential equations, the tree can be traversed as described in the text to yield a probability of observing the extant data.

In this simple case, however, it is possible to derive an analytical solution for the calculations of probabilities along a single branch. We describe this briefly to show the relationship between our method and that of Nee et al. (1994b), and because the character-independent model describes the approximate behavior of the likelihood functions when the character state has little effect on extinction and speciation rates. The solution for $E(t)$ is:

$$E(t) = 1 - \frac{(\lambda - \mu)}{(\lambda - e^{-(\lambda-\mu)t}\mu)} \qquad \text{if } \lambda \neq \mu \qquad (9a)$$

$$E(t) = \frac{\lambda t}{1 + \lambda t} \qquad \text{if } \lambda = \mu \qquad (9b)$$

as derived by Kendall (1948). $E(t)$ rises as we go further back in time (increase t), approaching 1 if $\lambda < \mu$ and μ/λ otherwise. Basically, the further back in time we go, the more opportunity there is for the entire lineage and its descendants to go extinct.

The above solution for $E(t)$ can be used to solve for $D_N(t)$:

$$D_N(t) = e^{-(\lambda-\mu)(t-t_N)} \frac{(\lambda - e^{-(\lambda-\mu)t_N}\mu)^2}{(\lambda - e^{-(\lambda-\mu)t}\mu)^2} D_N(t_N) \qquad \text{if } \lambda \neq \mu \qquad (10a)$$

$$D_N(t) = \frac{(1 + \lambda t_N)^2}{(1 + \lambda t)^2} D_N(t_N) \qquad \text{if } \lambda = \mu \qquad (10b)$$

where t_N is the time depth of node N (see Fig. 1). These equations permit us to jump from node N to its ancestor A without recourse to approximate numerical integration. To traverse node A, we need to account for the speciation event, as we did in Equation (4). The above process of traversing a branch and then accounting for the node can be repeated until the root is reached. Doing so, we get the probability of observing the full phylogeny:

$$D_R(t_R) = \left[\prod_{i=1}^{2n} e^{-(\lambda-\mu)\left(t_{i,b}-t_{i,t}\right)} \frac{\left(\lambda - e^{-(\lambda-\mu)t_{i,t}}\mu\right)^2}{\left(\lambda - e^{-(\lambda-\mu)t_{i,b}}\mu\right)^2} \right] (\lambda)^n \quad \text{if } \lambda \neq \mu \qquad (11a)$$

$$D_R(t_R) = \left[\prod_{i=1}^{2n} \frac{(1 + \lambda\, t_{i,t})^2}{(1 + \lambda\, t_{i,b})^2} \right] (\lambda)^n \quad \text{if } \lambda = \mu \qquad (11b)$$

where $t_{i,b}$ is the time at the base of the ith branch (rootward), $t_{i,t}$ is the time at the terminus of the ith branch length (nearest the present), n is the number of nodes (including the root node), and the product is taken over all branches in the tree.

If we condition on the existence of a root node with two surviving lineages (dividing (11a) by $\lambda\, [1 - E(t_R)]^2$), then after some algebra, Equation (11a) can be shown to equal equation (21) of Nee et al. (1994b) except for a factor, $(N - 1)!$, which does not depend on the parameters. The $(N - 1)!$ arises because Nee et al. consider the probability of all possible tree topologies with the same branching times (regardless of where the branches occur), and there are $(N - 1)!$ such tree topologies. Thus, our method generates the same likelihood surface as that of Nee et al. (1994b) when the character does not influence the rate of speciation and extinction, but as we have shown, can be more readily extended to include more complex processes of diversification.

APPENDIX 2

Equilibrium frequencies.—The expected frequency of state 0 and 1 within an assemblage of species should, over evolutionary time, reach an equilibrium that can be calculated from the model. Here, we consider the passage of time, T, in the forward direction and track the number of lineages in state 0, n_0, and the number of lineages in state 1, n_1. Given that each lineage could speciate, go extinct, or switch state, the expected number of species of each type should obey the following ordinary differential equations:

$$\frac{dn_0}{dT} = \lambda_0 n_0 - \mu_0 n_0 - q_{01} n_0 + q_{10} n_1 \qquad (12a)$$

$$\frac{dn_1}{dT} = \lambda_1 n_1 - \mu_1 n_1 - q_{10} n_1 + q_{01} n_0 \qquad (12b)$$

From these equations, we can derive a differential equation for the frequency of lineage in state 0, $x = n_0/(n_0 + n_1)$, using the quotient rule:

$$\frac{dx}{dT} = \frac{\dfrac{dn_0}{dT}n_1 - \dfrac{dn_1}{dT}n_0}{(n_0 + n_1)^2} = g\,x\,(1-x) - x\,q_{01} + (1-x)\,q_{10} \quad (13)$$

where $g = \lambda_0 - \mu_0 - \lambda_1 + \mu_1$. The equilibrium frequency of lineages in state 0, $\hat{x}$, is thus the single root of the quadratic equation, $g\,\hat{x}\,(1 - \hat{x}) - \hat{x}\,q_{01} + (1 - \hat{x})\,q_{10} = 0$, that lies between 0 and 1. When $g = 0$, this equilibrium occurs at $\hat{x} = q_{10}/(q_{01} + q_{10})$.

Average number of species over time.—We can also solve Equations (12) explicitly to determine the average number of species as we move forward in time from a single species to time T when a character affects the diversification process. Averaging over those cases where the root is in state 0 (with probability $\hat{x}$) and in state 1 (with probability $1 - \hat{x}$), the average number of lineages, $\overline{L(T)}$, at time T grows exponentially according to:

$$\overline{L(T)} = e^{[\hat{x}(\lambda_0 - \mu_0) + (1 - \hat{x})(\lambda_1 - \mu_1)]T} \quad (14)$$

If lineages in state 0 speciate more rapidly or go extinct less rapidly than lineages in state 1 (i.e., if $g > 0$), then more species are expected to exist after time T for groups that are initially in state 0. After some algebra, the average number of lineages, $\overline{L_0(T)}$, at time T assuming

that the root state is 0 can be written as:

$$\overline{L_0(T)} = \overline{L(T)}\left[1 - \frac{g\,(1-x)^2}{g(1-x)^2 - q_{01}}\right] + e^{(\lambda_0 - \mu_0 - \frac{q_{01}}{1-x})T}$$
$$\times \left[\frac{g(1-x)^2}{g(1-x)^2 - q_{01}}\right] \quad (15a)$$

Assuming instead that the root state is 1, we have:

$$\overline{L_1(T)} = \overline{L(T)}\left(1 - \frac{gx^2}{gx^2 + q_{10}}\right) + e^{(\lambda_1 - \mu_1 - \frac{q_{10}}{x})T}\left(\frac{gx^2}{gx^2 + q_{10}}\right) \quad (15b)$$

Equations (15a) and (15b) both reduce to $\overline{L(T)}$ given by Equation (14) when the net diversification rate is the same for the two character states (i.e., g is zero).

The above equations can be solved for T to estimate the amount of time required to generate a certain number of species when the character state influences diversification. It should be emphasized, however, that these equations use a deterministic model (12) to approximate the stochastic process of diversification and character state change. Thus, Equations (14) and (15) provide only a heuristic guide to the evolutionary process when character states influence speciation and extinction rates.