Estimating Trait-Dependent Speciation and Extinction Rates from Incompletely Resolved Phylogenies

RICHARD G. FITZJOHN[1,2,*], WAYNE P. MADDISON[1,2,3], AND SARAH P. OTTO[1,2]

[1]*Department of Zoology,* [2]*Biodiversity Research Centre and* [3]*Department of Botany, 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.

Abstract.—Species traits may influence rates of speciation and extinction, affecting both the patterns of diversification among lineages and the distribution of traits among species. Existing likelihood approaches for detecting differential diversification require complete phylogenies; that is, every extant species must be present in a well-resolved phylogeny. We developed 2 likelihood methods that can be used to infer the effect of a trait on speciation and extinction without complete phylogenetic information, generalizing the recent binary-state speciation and extinction method. Our approaches can be used where a phylogeny can be reasonably assumed to be a random sample of extant species or where all extant species are included but some are assigned only to terminal unresolved clades. We explored the effects of decreasing phylogenetic resolution on the ability of our approach to detect differential diversification within a Bayesian framework using simulated phylogenies. Differential diversification caused by an asymmetry in speciation rates was nearly as well detected with only 50% of extant species phylogenetically resolved as with complete phylogenetic knowledge. We demonstrate our unresolved clade method with an analysis of sexual dimorphism and diversification in shorebirds (Charadriiformes). Our methods allow for the direct estimation of the effect of a trait on speciation and extinction rates using incompletely resolved phylogenies. [Bayesian inference; birth–death process; BISSE; extinction; phylogenetics; sampling; speciation.]

Just as differences in traits may affect the relative survival and reproductive success of individuals, traits may affect the relative rate at which lineages go extinct or speciate (Stanley 1975; Coyne and Orr 2004; Ricklefs 2007). Sister-clade comparisons (Barraclough et al. 1998) have been widely used to detect traits that are correlated with differential diversification. Using this method, traits that have been found to have a significant impact on diversification rates include diet in insects (Mitter et al. 1988; Farrell 1998), latitude in birds and butterflies (Cardillo 1999), mating system in birds (Mitra et al. 1996), and sex allocation in flowering plants (Heilbuth 2000). These analyses have often been framed as tests of whether a character is a "key innovation;" that is, has a particular character state lead to elevated rates of diversification? More recently, a variety of statistical approaches that directly estimate speciation rates have been developed that incorporate phylogenetic tree topology and the pattern of branching times (e.g., Pagel 1997; Paradis 2005; Ree 2005; Maddison et al. 2007). These approaches allow for greater statistical power than sister-clade comparisons because they incorporate more information about the patterns of diversification. Among these is the binary-state speciation and extinction (BISSE) method (Maddison et al. 2007), a whole-tree likelihood method that can be used to detect the effect of a trait on diversification, where the trait can be classified into 2 states.

The BISSE method as formulated by Maddison et al. (2007) assumes that the phylogenetic tree is complete and fully resolved; that is, the tree must include every extant species. It also assumes that all character state information is known. These assumptions currently restrict its applicability, as few published phylogenies are both complete to the species level and large enough to detect differential diversification. Without appropriate correction, BISSE will not produce valid likelihoods for incompletely resolved trees. Incomplete phylogenetic coverage decreases the apparent number of events over a phylogeny; there are fewer inferred speciation and character change events. Because of this, the BISSE likelihood surface shifts to favor lower rates of diversification and character change. Furthermore, inferred phylogenies that include only a fraction of extant species tend to have longer terminal branches (Fig. 1b), and as a result, the estimated extinction rates approach 0 because there is a smaller increase in the number of lineages in time near the present (Nee et al. 1994).

Similar limitations have been overcome in likelihood approaches that estimate speciation and extinction rates when these rates do not depend on a character (character-independent diversification). The character-independent likelihood method of Nee et al. (1994) includes corrections that assume that the species present in a phylogeny represent a random sample of extant species from a clade by incorporating the sampling process into the likelihood calculations. Recently, Bokma (2008) developed a Bayesian approach for estimating character-independent diversification rates that treats the branching times for missing taxa as additional parameters to be estimated. In these studies, because speciation and extinction rates do not depend on a species' character state, only the branching times are required. However, if speciation and extinction rates depend on a character's state, then branching times are insufficient because the topology of the tree will depend on how the character evolves.

Here, we extend BISSE to allow estimation of character-dependent speciation and extinction rates from incompletely resolved phylogenies. We develop likelihood calculations that compensate for incomplete phylogenetic knowledge in 2 cases: 1) where the species

a) Complete phylogeny

b) Skeleton phylogeny (random sampling)

c) Terminally unresolved phylogeny
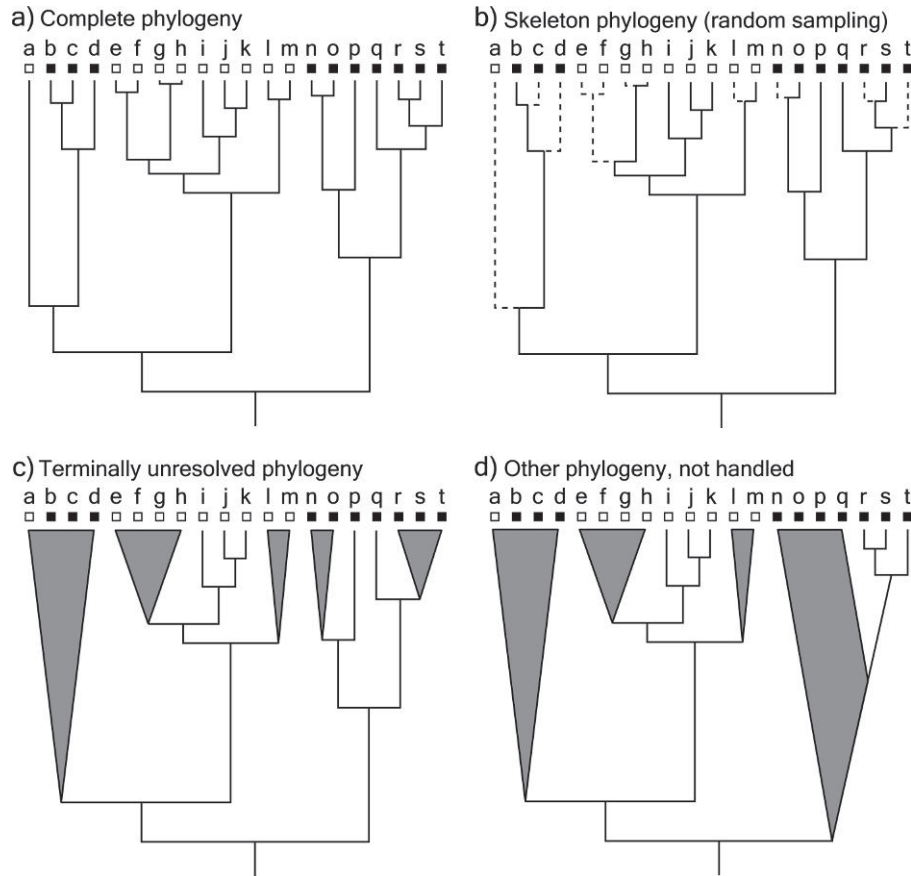
d) Other phylogeny, not handled

FIGURE 1. Different ways that phylogenetic information may be incomplete. Tree (a) is complete; every extant species is included and the tree is fully resolved. Black and white boxes above the tips refer to different character states. Tree (b) is a "skeletal tree"; species are included randomly from the full tree in (a). Sampled taxa are indicated by solid lines, and missing taxa are indicated by dashed lines. In general, nothing is known about the placement of these taxa. Tree (c) is a "terminally unresolved tree"; in this case, the species not explicitly included as tips in the phylogeny are all known to belong to terminal unresolved clades. This tree is therefore "complete" in that it includes all extant taxa but is incompletely resolved. This tree has the same branching structure as (b). Tree (d) contains a paraphyletic unresolved group and cannot be directly handled by either of the methods presented here. The relationships among species $n - q$ are not resolved, and this group is known to be paraphyletic (see panel a). To convert this tree into a terminally unresolved tree, the known relationships within the $r - t$ clade would have to be discarded to create an unresolved clade spanning species $n - t$.

in a phylogeny represent a random sample of all extant species within a group (Fig. 1b) and 2) where species not directly represented as tips in the phylogeny can be assigned to terminal unresolved clades (Fig. 1c). We also develop methods to allow for incomplete character state knowledge for both complete and incompletely resolved trees. We describe how these likelihoods can be used in Bayesian inference and apply our methods to simulated data sets. Finally, we demonstrate our method by applying it to the correlation between diversification and sexual dimorphism in shorebirds (Charadriiformes).

BISSE FOR COMPLETE PHYLOGENIES

Because our aim is to generalize the BISSE model of Maddison et al. (2007), we start with a brief description of this method. BISSE computes the probability of a phylogenetic tree and the observed distribution of character states among extant species, given a model of character evolution, speciation, and extinction. The character

states must be binary; we denote the possible character states as 0 or 1 (e.g., herbivorous or nonherbivorous insects). The likelihood calculation tracks 2 variables for each character state i along branches in a phylogeny: $D_{Ni}(t)$—the probability that a lineage in state i at time t would evolve into the extant clade N as observed and $E_i(t)$—the probability that a lineage in state i at time t would go completely extinct by the present, leaving no extant members. (For compactness, we will often refer to the clade whose most recent common ancestor is node N as "clade N.") Time is measured backward with the present at $t = 0$ and $t > 0$ representing some time in the past. The changes in these quantities over time are described by a set of ordinary differential equations

$$\frac{dD_{Ni}}{dt} = -(\lambda_i + \mu_i + q_{ij})D_{Ni}(t) + q_{ij}D_{Nj}(t)$$
$$+ 2\lambda_i E_i(t)D_{Ni}(t), \tag{1a}$$

$$\frac{dE_i}{dt} = \mu_i - (\lambda_i + \mu_i + q_{ij})E_i(t) + q_{ij}E_j(t) + \lambda_i E_i(t)^2, \tag{1b}$$
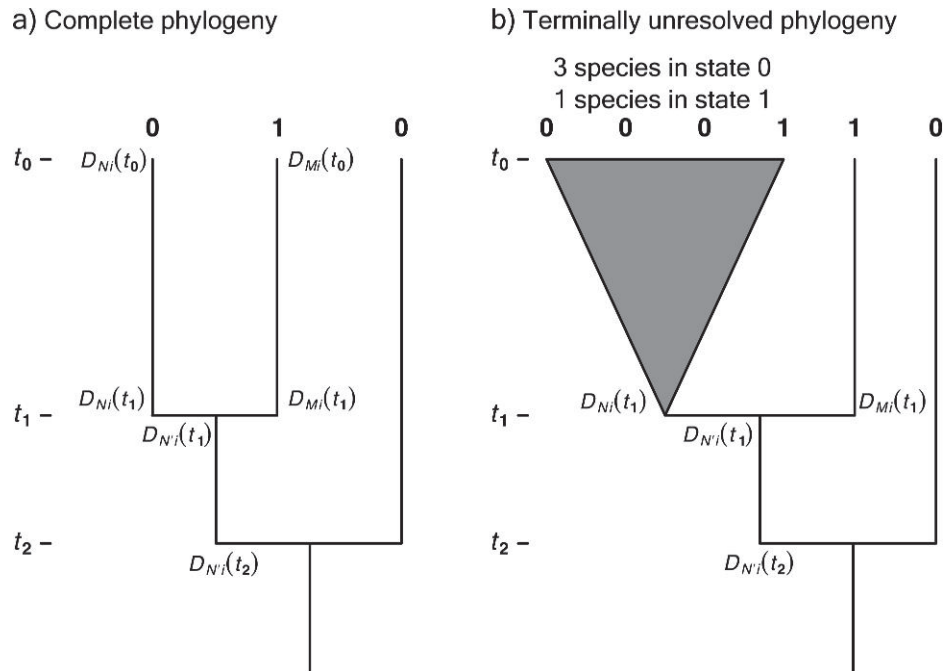
FIGURE 2. BISSE with (a) and without (b) full phylogenetic knowledge. In panel (a), the values at the base of the nodes leading to the first tips ($D_{Ni}(t_1)$ and $D_{Mi}(t_1)$) are calculated backward in time using Equations (1) and then combined with Equation (2) to become the initial condition $D_{N'i}(t_1)$ for calculating $D_{N'i}(t_2)$. In panel (b), the four species on the left are unresolved but can be assigned to a tip that branches at time t_1. $D_{Ni}(t)$ would be calculated forward in time using our new method, and $D_{Mi}(t)$ with BISSE, with these values combined as above.

where λ_i is the speciation rate in state i, μ_i is the extinction rate in state i, and q_{ij} is the rate of transition from state i to j forward in time (Maddison et al. 2007). These equations are solved numerically along each branch backward in time to compute $D_{Ni}(t)$ (Fig. 2).

On each branch, the character state at the tip provides the initial conditions for Equations (1). $D_{Ni}(0) = 1$ if the tip N is in state i and 0 otherwise because the lineage must be in its observed state. Similarly, $E_0(0) = E_1(0) = 0$ as a lineage cannot go extinct in zero time. At a node joining the lineages leading to clades N and M, the probability of generating both daughter clades given that the node is in state i is

$$D_{N'i}(t) = D_{Ni}(t)D_{Mi}(t)\lambda_i, \qquad (2)$$

where N' represents the union of clades N and M (see Fig. 2). The likelihood calculation proceeds backward in time down the tree from the tips until it reaches the root. At the root, R, we have the two probabilities, D_{R0} and D_{R1}, corresponding to the possible character states at the root. The overall likelihood, D_R, must sum over the probabilities that the root was in each state (see Appendix 1).

LIKELIHOOD CALCULATIONS FOR INCOMPLETELY RESOLVED PHYLOGENIES

Incompleteness in phylogenetic information can come in many forms. A species may be entirely unplaced phylogenetically or placed into a clade but not into a precise relationship within the clade. Its character state may be known or unknown. We will derive methods for two situations: "skeletal trees," where we have a fully resolved tree for a random sample of species whose states are fully known, and "terminally unresolved trees," where trees include all extant species and are fully resolved except for terminal clades that are completely unresolved phylogenetically and whose character states are known to varying degrees. Skeletal trees (Fig. 1b) could arise when a biologist samples species simultaneously for their presence in a phylogenetic analysis and having data for the character of interest. For these trees, we assume that nothing is known about the phylogenetic placement of the missing taxa. Terminally unresolved trees arise frequently when the species included in a molecular phylogeny are exemplars and where information on the nonincluded (unplaced) species is available (e.g., from previous systematic studies). If the unplaced species can be assigned to terminal clades containing the exemplar species, then our method can be used (Fig. 1c). Here, we assume that every species can be assigned to an unresolved clade. Note that terminally unresolved trees are phylogenetically complete, in that they include all extant taxa, but are incompletely resolved, in that not all phylogenetic relationships are known. A broader class of incomplete phylogenies do not match either of these cases, and the methods we describe below cannot be used directly. This includes paraphyletic unresolved groups (Fig. 1d).

Skeletal Trees: Unplaced Missing Taxa

First, we consider skeletal trees, where a given phylogenetic tree represents a random sample of all extant species in a taxonomic group. To account for incomplete phylogenies, we model a "sampling" event at the present that corresponds to a biologist obtaining data for the species. This event occurs during an infinitesimally small time period during which a species in state i has a probability f_i of being sampled for inclusion in a phylogeny. The f_i values should be determined from estimates of the numbers of species having each character state that are unsampled versus sampled. If the character states of unsampled species are unknown, then the f_i values could be set equal for all states and reflect the proportion of all extant species that have been sampled. With this sampling event, $E_i(t)$ can be interpreted as the probability of a lineage not being present in the phylogeny, either by going extinct or not being sampled. The initial condition $E_i(0)$ is therefore $(1 - f_i)$. Similarly, rather than representing the probability that a lineage in state i at time t would evolve into the full extant clade N at the present, $D_{Ni}(t)$ includes the probability that the tip taxa present in the phylogeny are sampled. The initial conditions become $D_{Ni}(0) = f_i$ if the sampled tip is in state i and 0 otherwise. After these modifications to the initial conditions, the calculations continue as described in Maddison et al. (2007). This is similar to the method used by Nee et al. (1994) to correct likelihood calculations for inferring character-independent speciation and extinction rates from incomplete phylogenies. Indeed, in Appendix 2, we show that when the speciation and extinction rates are independent of character state, the calculations are equivalent.

This approach assumes that the taxon sampling process is independent of the position in the phylogeny. However, it need not be independent of the character state as the f_i can differ between states 0 and 1. However, this approach also assumes that taxon sampling is even across the phylogeny, which in many cases it is not.

Terminally Unresolved Trees

Terminally unresolved trees contain all species, but their relationships are not fully resolved, with some species grouped into unresolved clades (Fig. 1c). We can envision this situation as comparable to the skeletal trees, but with the unplaced species not entirely unknown: we know to which terminal clades they belong and we may know their character states. This extra information about placement and character states can be used to improve inference.

We will use the word "tip" to refer to a terminal unit in the tree, which may represent either a single extant species or a terminal unresolved clade. Thus, the number of tips in the tree will be less than the number of species implied if there are unresolved terminal clades. We assume that the sampling of species is complete; that is, every extant species is either present directly

as a tip or can be assigned to a tip that represents an unresolved clade. We do not assume knowledge of the timing of the last common ancestor for a terminal unresolved clade, instead we assume that diversification happens at any point after splitting from its sister clade (Fig. 2). We also do not assume any particular topology for the unresolved clades, rather we sum over all possible phylogenetic histories, according to their probability. We initially assume complete knowledge of character states, but we relax this assumption in the next section.

If we can compute the probability of a terminal clade, $D_{Ni}(t)$, then we can combine this with the probability of their sister lineages using Equation (2) and continue with BISSE down the rest of the tree (Fig. 2b). In contrast to the backward-time approach employed by BISSE, we use a forward-time method to calculate $D_{Ni}(t)$ for an unresolved clade. Because it has no phylogenetic structure, we cannot distinguish among the different possible evolutionary histories of an unresolved clade. Consequently, we model clade evolution as a Markov process, tracking only the probability of different clade compositions over time. The possible clade compositions can be distinguished by the number of species in each state; let (n_0, n_1) represent a clade with n_0 species in state 0 and n_1 species in state 1. Even though the number of possible clade types is infinite, we truncate state space to a finite number of species. This process is similar to two birth–death processes (Nee 2006), one for each character state, but includes transitions between the processes. We term this a "birth–death-transition process."

If $\vec{x}(t)$ is a column vector representing probabilities of different clade types at time t and $\mathbf{Q}$ is a transition rate matrix describing the rates of changes between clade types, then the probability of generating any possible clade is given by

$$\vec{x}(t_0) = \exp((t_1 - t_0)\mathbf{Q}) \cdot \vec{x}(t_1), \quad t_1 > t_0, \qquad (3)$$

where t_1 represents an earlier point in time to t_0 and "exp" represents the matrix exponential (Sidje 1998). The values of $\vec{x}(t_0)$ that correspond to the observed data can then be used as the probability of the clade evolving as observed, $D_{Ni}(t_1)$, for subsequent BISSE calculations. For example, for a clade that begins in state 0 at time t_1 and ends in clade type $(3, 1)$ at the present (t_0, as in Fig. 2b), we find $\vec{x}(t_0)$ from Equation (3) and pick out the probability of generating a $(3, 1)$ clade from this vector, which is used as $D_{N0}(t_1)$. The probability of generating a clade with no species, $(0, 0)$, is the probability that the clade would have gone extinct, $E_0(t_1)$. This process is then repeated assuming that the lineage leading to the clade was initially in state 1, giving $D_{N1}(t_1)$ and $E_1(t_1)$.

Before describing the transition rate matrix $\mathbf{Q}$, we must first specify the structure of the state space $\vec{x}(t)$. Let the first element represent the probability of having zero species, the next two elements represent the two single-species clades with one species in state 0 or 1

species in state 1 (respectively) and so on. That is, probabilities are assigned to the positions of $\vec{x}(t)$ in the order:

$$\underbrace{(0,0),}_{\text{0 species}} \underbrace{(1,0),(0,1),}_{\text{1 species}} \underbrace{(2,0),(1,1),(0,2),\ldots,}_{\text{2 species}}$$

$$\underbrace{(k,0),(k-1,1),(k-2,2),\ldots,(1,k-1),(0,k),}_{k \text{ species}} \ldots, \tag{4}$$

so that a clade with k species is represented by the $k+1$ elements in positions $k(k+1)/2 + 1$ to $(k+1)(k+2)/2$. To keep the state space finite, the final element of $\vec{x}(t)$ is an absorbing state, representing the probability that a clade has at least $n_{\max}$ species. By doing this, we assume that once a clade reaches $n_{\max}$ species, it is so large that there is a negligible probability of generating the observed number of species by time t_0. In practice, $n_{\max}$ can be chosen to be large enough so that it does not significantly affect calculations (e.g., by monitoring the change in relevant values in $\vec{x}(t_0)$ as $n_{\max}$ is increased). At the base of the clade, t_1, there must have been a single ancestral lineage in either state 0 or 1. The state of the system at this time must have been a vector of zeros except for a 1 in either the second position (corresponding to $(1,0)$ to calculate $D_{N0}(t)$) or the third position (corresponding to $(0,1)$ to calculate $D_{N1}(t)$).

To calculate $\mathbf{Q}$, we assume that each time step is small enough that only a single event may happen; a lineage currently in state (n_0, n_1) may have one species:

- speciate, moving from state (n_0, n_1) to $(n_0 + 1, n_1)$ or $(n_0, n_1 + 1)$ at rate $n_0\lambda_0$ or $n_1\lambda_1$, respectively,

- go extinct, moving to $(n_0 - 1, n_1)$ or $(n_0, n_1 - 1)$ at rate $n_0\mu_0$ or $n_1\mu_1$,

- change character state, moving to state $(n_0 - 1, n_1 + 1)$ or $(n_0 + 1, n_1 - 1)$ at rate n_0q_{01} or n_1q_{10}.

Using these rules, the transition rate matrix has a block structure, involving the blocks $\mathbf{S}_k$, $\mathbf{E}_k$, and $\mathbf{C}_k$. The block $\mathbf{S}_k$ is a $(k+2) \times (k+1)$ matrix describing speciation from k to $k+1$ species:

$$\mathbf{S}_k = \begin{pmatrix} k\lambda_0 & 0 & 0 \\ 0 & (k-1)\lambda_0 & 0 \\ 0 & \lambda_1 & (k-2)\lambda_0 \\ 0 & 0 & 2\lambda_1 & \ddots \\ & & & \ddots & \lambda_0 & 0 \\ & & & & (k-1)\lambda_1 & 0 \\ & & & & 0 & k\lambda_1 \end{pmatrix}.$$

$\mathbf{E}_k$ is a $k \times (k+1)$ matrix describing extinction from k to $k-1$ species:

$$\mathbf{E}_k = \begin{pmatrix} k\mu_0 & \mu_1 & 0 \\ 0 & (k-1)\mu_0 & 2\mu_1 \\ 0 & 0 & (k-2)\mu_0 & \ddots \\ & & & \ddots & (1-k)\mu_1 & 0 \\ & & & & \mu_0 & k\mu_1 \end{pmatrix}.$$

$\mathbf{C}_k$ is a $(k+1) \times (k+1)$ square matrix describing character state changes, leaving the number of species constant at k:

$$\mathbf{C}_k = \begin{pmatrix} \cdot & q_{10} & 0 \\ kq_{01} & \cdot & 2q_{10} \\ 0 & (k-1)q_{01} & \cdot & \ddots \\ 0 & 0 & (k-2)q_{01} & \ddots & (k-1)q_{10} & 0 \\ & & & \ddots & \cdot & kq_{10} \\ & & & & q_{01} & \cdot \end{pmatrix},$$

where the dotted elements along the diagonal of $\mathbf{C}_k$ are chosen so that the columns of $\mathbf{Q}$ sum to zero. Denoting matrices of zeros with 0, the transition rate matrix $\mathbf{Q}$ is

$$\mathbf{Q} = \begin{pmatrix} \mathbf{C}_0 & \mathbf{E}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{C}_1 & \mathbf{E}_2 \\ \mathbf{0} & \mathbf{S}_1 & \mathbf{C}_2 & \ddots \\ \mathbf{0} & \mathbf{0} & \mathbf{S}_2 & \ddots & \mathbf{E}_{n_{\max}-1} & \mathbf{0} \\ & & & \ddots & \mathbf{C}_{n_{\max}-1} & \mathbf{0} \\ & & & & \mathbf{S}_{n_{\max}-1} & \mathbf{0} \end{pmatrix}. \tag{5}$$

The final speciation block $\mathbf{S}_{n_{\max}-1}$, describing speciation into the absorbing state is an n_t element row vector:

$$((n_{\max} - 1)\lambda_0, \ (n_{\max} - 2)\lambda_0 + \lambda_1, \ldots, (n_{\max} - 1)\lambda_1).$$

As a special case, this approach can be used to calculate likelihoods for terminally unresolved trees where speciation and extinction do not depend on a character's state, as described in Appendix 2.

Incomplete Character State Knowledge

Regardless of the level of phylogenetic completeness and resolution, character state information may be unknown for some species. Here, we describe corrections

to the BISSE likelihood for missing character state information for fully resolved phylogenies, skeletal trees, and terminally unresolved trees.

For fully resolved phylogenies, if no information on a character for a tip is available, then the "data" become the presence of the tip only. On the single branch leading to this tip, we can then interpret $D_{Ni}(t)$ as the probability of giving rise to a single species, regardless of its character state. The initial conditions must therefore be $D_{Ni}(0) = 1$ for both states because with no time for extinction, there is a 100% probability that the branch will lead to the observed data. Using this logic, for skeletal trees, the initial conditions are $D_{Ni}(t) = f_i$.

For terminally unresolved trees, character state information may not be known for all members of an unresolved clade. In this case, we can calculate the joint probability that a clade evolved to a particular composition and that it was sampled as observed. Say that the unresolved clade of interest truly has x_0 species in state 0 and x_1 species in state 1 but that we know the state information only for a sample of these species so that s_i species are known to be in state i. If the probability of a species' state being known is independent of its state, then we can assume that the $s_N = s_0 + s_1$ known species represent samples without replacement from a pool of $x_N = x_0 + x_1$ species and compute the sampling probability using the hypergeometric distribution. Of the $\binom{x_N}{s_N}$ ways of sampling s_N species from this pool, there are $\binom{x_i}{s_i}$ ways of sampling s_i species in state i. The sampling probability is therefore given by:

$$\Pr(s_0, s_1 | x_0, x_1) = \frac{\binom{x_0}{s_0} \binom{x_1}{s_1}}{\binom{x_N}{s_N}}. \qquad (6)$$

Although we do not know the true number of species in each state, we can use Equation (3) to compute the probability that the clade composition is (x_0, x_1) and then use Equation (6) to give the probability of knowing that s_i species are in state i. To do this, we multiply the probability of generating the clade by the probability of sampling the clade as observed and sum over all possible clade compositions:

$$D_{Ni}(t) = \sum_{j=s_0}^{x_N - s_1} \Pr(j, x_N - j) \frac{\binom{j}{s_0} \binom{x_N - j}{s_1}}{\binom{x_N}{s_N}}. \qquad (7)$$

Here, $\Pr(x_0, x_1)$ is the probability of a clade with x_0 species in state 0 and x_1 species in state 1, calculated from Equation (3). This calculation assumes that we know that there are x_N species in the clade, but this calculation can be generalized if x_N itself is not known exactly but can be described by a probability distribution. Where we have full state information (i.e., $s_0 + s_1 = x_N$), Equation (7) reduces to $\Pr(s_0, s_1)$.

BAYESIAN INFERENCE

The above equations can be used to calculate the likelihood from an incomplete phylogeny, that is, the probability of the data given a model of speciation, extinction, and character evolution. This method can then be used to estimate rates using maximum likelihood and to compare models using likelihood ratio tests. Here, we will discuss their application to Bayesian inference so that measures of parameter uncertainty can be simultaneously obtained. For a general introduction to Bayesian inference in phylogenetics, see Huelsenbeck et al. (2002). We will focus on the posterior probability distribution of the model parameters; that is, the probability of the parameters given the data.

To compute the posterior probability, we need to specify the prior probability distribution for the parameters. We use an exponential prior for the six parameters (see Churchill 2000). This choice reflects the philosophical preference for explanations requiring fewer events, all else being equal (Occam's razor). For example, if few species are present in state 0, then there is no information about the extinction rate for species in state 0 (μ_0). An exponential prior would then generate a posterior distribution with the same mean as the prior. Other common priors include a uniform prior and a uniform prior on the log of each parameter (e.g., on $\ln(\mu_0)$). These are both "improper priors" because they do not integrate to a finite value over the possible range of the parameters $(0, \infty)$. Because of this, in the case where little signal is present in the data, the posterior will not integrate to a finite value and cannot easily be interpreted (Gelman et al. 1995). Because it is itself proper, the exponential prior always produces a proper posterior probability distribution, and it has the additional benefit that its influence on the posterior distribution can be easily detected by comparing the mean of prior and posterior distributions. The prior probability density associated with the parameter θ_j is set to:

$$\Pr(\theta_j) = c_j \, e^{-c_j \theta_j}, \qquad (8)$$

where θ_j is the value of the jth parameter and c_j is a rate parameter. The posterior probability of the model given the data is proportional to

$$D_R(t_R) \prod_j c_j \, e^{-c_j \theta_j}, \qquad (9)$$

where the product is taken over the six model parameters. An exploration of the alternative priors indicated that the priors generally had negligible influence except where there were very few extant species in a given character state.

To choose values for c_j, we use a preliminary measure of the rate of diversification from the tree. Ignoring state changes and asymmetries in speciation or extinction rates, the expected number of species in a tree of length t_R is $n = e^{(\lambda - \mu)t_R}$, where λ and μ are the character-independent speciation and extinction rates (Nee et al.

1994). Rearranging, the diversification rate $(\lambda - \mu)$ that would produce n species at time t_R is $\ln(n)/t_R$. We chose the prior rates so that the mean of the exponential distribution was twice this value (i.e., $c_j = t_R/2\ln(n)$). The same prior was used for all model parameters.

NUMERICAL METHODS AND APPLICATION

To test our method, we followed the same approach as Maddison et al. (2007) by simulating trees and character states using known rates and then attempting to infer those rates from the tree. We simulated trees containing 500 species with rates $\lambda_0 = \lambda_1 = 0.1$, $\mu_0 = \mu_1 = 0.03$, and $q_{01} = q_{10} = 0.01$ (equal rate trees) or with $\lambda_1 = 0.2$ (unequal rate trees). These are the same rates as Maddison et al. (2007) for comparison, and the trees were simulated using their method.

We generated random incomplete phylogenies from these complete simulated phylogenies. To perform random taxonomic sampling to create skeletal trees (Fig. 1b), we sampled a proportion of all tips independently of tip state. The per-state fraction of species in each state that were present in the final sample was calculated and used to specify f_0 and f_1 when calculating likelihoods. To simulate terminally unresolved trees, a similar sampling routine can be used. Insofar as terminally unresolved trees can arise when character data are available for all species but detailed phylogenetic placement is available for only a sample of species, we can simulate this by choosing which species were sampled for detailed phylogenetic placement. The remaining unsampled species would be assigned to terminally unresolved clades represented by a single species that was sampled, the exemplar of the clade. However, this sampling requires some additional care because every extant species must be either present in the phylogeny or assigned to an unresolved clade (cf. Fig. 1c,d). Simply sampling species can leave orphaned species that fall below resolved clades and so cannot be placed into fully unresolved clades. For example, suppose that species j and k were chosen to have resolved placement from the phylogeny in Fig. 1a, but species i left unresolved. The species i cannot be placed into an unresolved clade represented by a single sampled exemplar species and is thus "orphaned." As a way of guaranteeing that there were no orphans in the final tree, we included a fraction of the orphan species in the sample and reassessed which species remained orphans, repeating until no orphan species were present. Note that this sampling approach does not generate a random sample of species, as assumed in our skeletal tree approach. For the results reported in this paper, we assumed that the character states of all species were known.

Implementation

We implemented the above methods were in the R package "diversitree" (available from http://www.

zoology.ubc.ca/prog/diversitree). The diversitree package will also be accessible through an upcoming version of Mesquite (Maddison WP and Maddison DR 2008). The matrix exponentiations were calculated numerically using the DMEXPV routine in Expokit (Sidje 1998). Because the transition rate matrix **Q** is very sparse, it is practical to use this approach for unresolved clades containing up to several hundred species. The posterior probability distribution cannot be sampled from directly, so we use Markov chain Monte Carlo (MCMC) to approximate the distribution using slice sampling for the parameter updates (MacKay 2003; Neal 2003). For each tree, we ran 3 independent MCMC chains for 10,000 steps from random starting locations, discarding the first 2500 steps of each chain. Although these chains are short compared with those used in tree inference, the sampler here is exploring a reasonably smooth continuous probability surface, rather than tree space, with disjoint regions of high probability separated by areas of low probability (data not shown). Consequently, convergence of the MCMC chains was very rapid.

RESULTS

We briefly present the results of Bayesian inference using BISSE with complete phylogenetic knowledge, then discuss how the statistical power is affected by incomplete phylogenetic knowledge.

Bayesian Inference with BISSE

Where speciation rates were equal for each character state ($\lambda_0 = \lambda_1$; equal rate trees), the mean inferred speciation rates were close to the true values used to simulate the trees, and the posterior probability density was tightly distributed around this true value (Fig. 3, solid curves). Where speciation rates were unequal ($\lambda_0 < \lambda_1$), species in state 0 were relatively rare (approximately 10% of extant species). Consequently, the rates for transitions in state 0 (λ_0, μ_0, and q_{01}) were less precisely estimated than for state 1, although still largely centered around their true values (Fig. 3, dashed curves). This pattern is consistent with that in Maddison et al. (2007), who found that the maximum likelihood estimates for the rare character state were more widely distributed than the estimates for the more common character state (their fig. 4).

Some of the model parameter estimates were correlated; in particular, the speciation and extinction rates for a particular character state were positively and linearly correlated (data not shown), indicating that a range of speciation/extinction rate combinations had similar posterior probabilities. The diversification rate is the difference between the speciation and the extinction rates ($r_i = \lambda_i - \mu_i$; Nee et al. 1994). The uncertainty around the diversification rate estimate was similar to the uncertainty around the speciation rates, even where extinction rates were poorly estimated (Fig. 4).
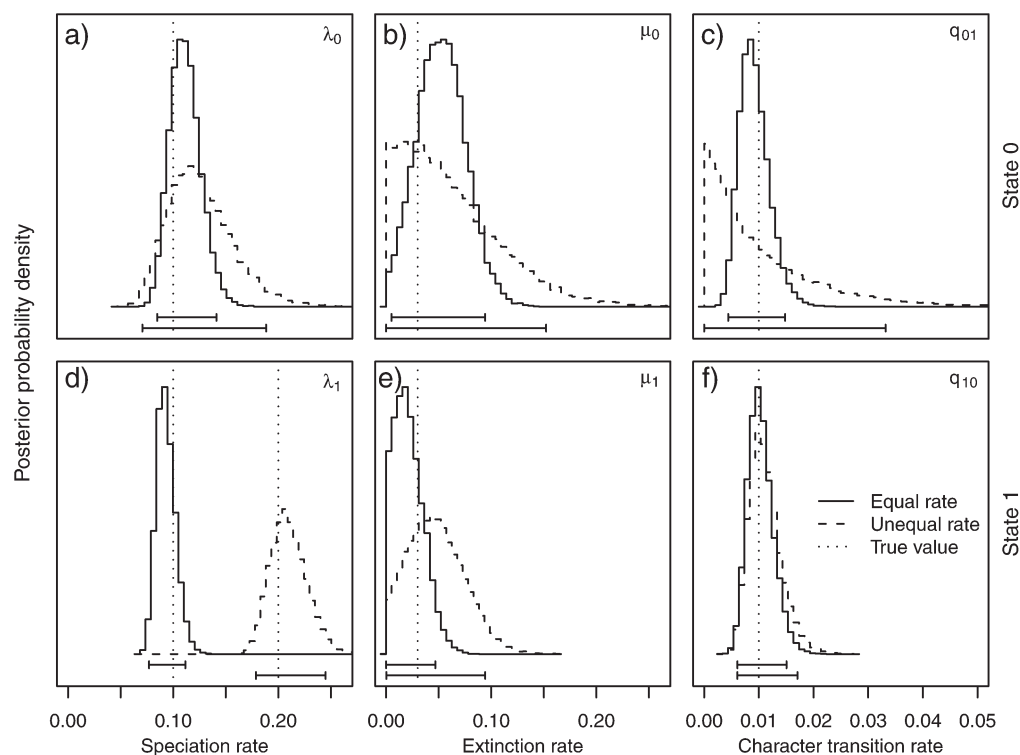
FIGURE 3.   Posterior probability densities for the 6 BISSE parameters on a fully resolved phylogeny. Two trees were generated, each containing 500 species and with either all rates equal ($\lambda_0 = \lambda_1$, $\mu_0 = \mu_1$, $q_{01} = q_{10}$; solid curves) or with unequal speciation rates ($\lambda_0 < \lambda_1$; dashed curves). The histograms display posterior probabilities over the last 7500 points from 3 independent MCMC chains, discarding the first 2500 points of each chain. The vertical lines indicate the true parameter values used in simulating the trees. The y-axes differ between plots but are scaled so the area under each curve integrates to 1. The horizontal bars indicate the 95% credibility intervals for the equal rate (upper bar) and unequal rate (lower bar) tree.

The difference between the diversification rates for the two character states (relative diversification rate; $r_{\text{rel}} = r_1 - r_0$) gives a summary of the strength of differential diversification. The relative diversification rate

for equal rate trees was well estimated, centered around the true value and with a narrow credibility interval. For unequal rate trees, the posterior probability distribution was flatter but still centered around the true value. For
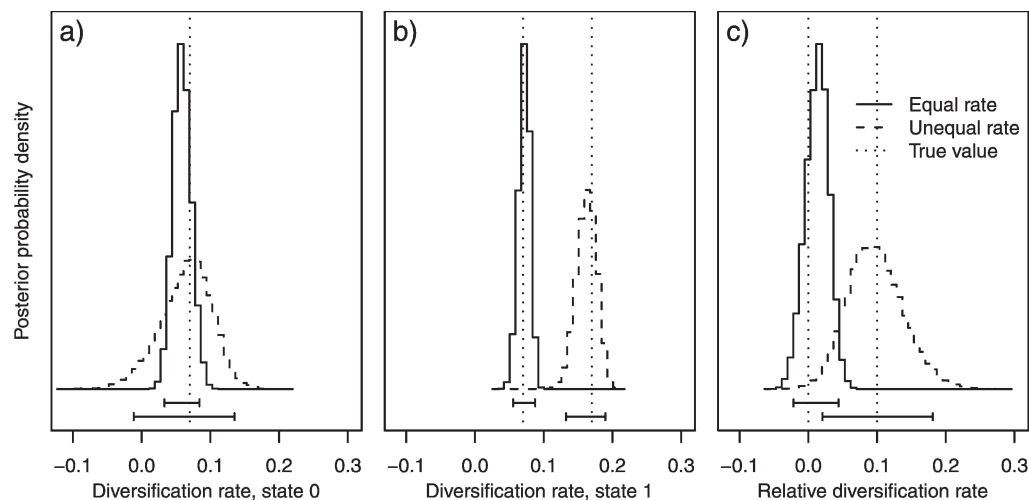

FIGURE 4.   Posterior probability distribution for the diversification rates (a) in state 0 (r_0) and (b) in state 1 (r_1) and (c) the relative diversification rate ($r_{\text{rel}} = r_1 - r_0$) for an equal rate tree ($\lambda_0 = \lambda_1$, solid curves) and an unequal rate tree ($\lambda_0 < \lambda_1$, dashed curves). The 95% credibility intervals are indicated by the horizontal bars, and the vertical lines indicate the true parameter values used in simulating the trees. Parameters are as indicated in text.

the example shown in Fig. 4, the posterior probability of $r_{rel} \leq 0$ for the unequal rate tree was 0.004, so we would correctly conclude that character state 1 increased diversification rate in this case.

Effect of Decreasing Phylogenetic Knowledge

As fewer species were included in a phylogeny or as more species fell within unresolved clades, parameters were less accurately and precisely estimated (Fig. 5). Accuracy and precision were essentially unaffected for nearly completely resolved phylogenies (75–100% complete) and for most parameters precision did not deteriorate substantially until trees contained fewer than $\approx 50\%$ of the total possible tips. For all parameters, the mean parameter estimate increased when phylogenetic

resolution became very low. This reflects the skew in the posterior probability distribution (Fig. 3), which increased with reduced phylogenetic resolution as the prior distribution increasingly dominates the posterior distribution. In addition, the prior means were higher than any of the simulated rates (approximately 0.35 for unequal rate trees and 0.16 for equal rate trees). Because medians are less sensitive to skew, the median parameter estimates were less affected by decreasing phylogenetic information than the mean, but they still tended to increase at very low phylogenetic resolution (not shown). The decrease in accuracy and precision was most pronounced for the rate parameters for the rare state on unequal rate trees (λ_0, μ_0, and q_{01}). Decreasing phylogenetic resolution increased the uncertainty of the parameter estimates, with the widths of the credibility intervals growing as the proportion of
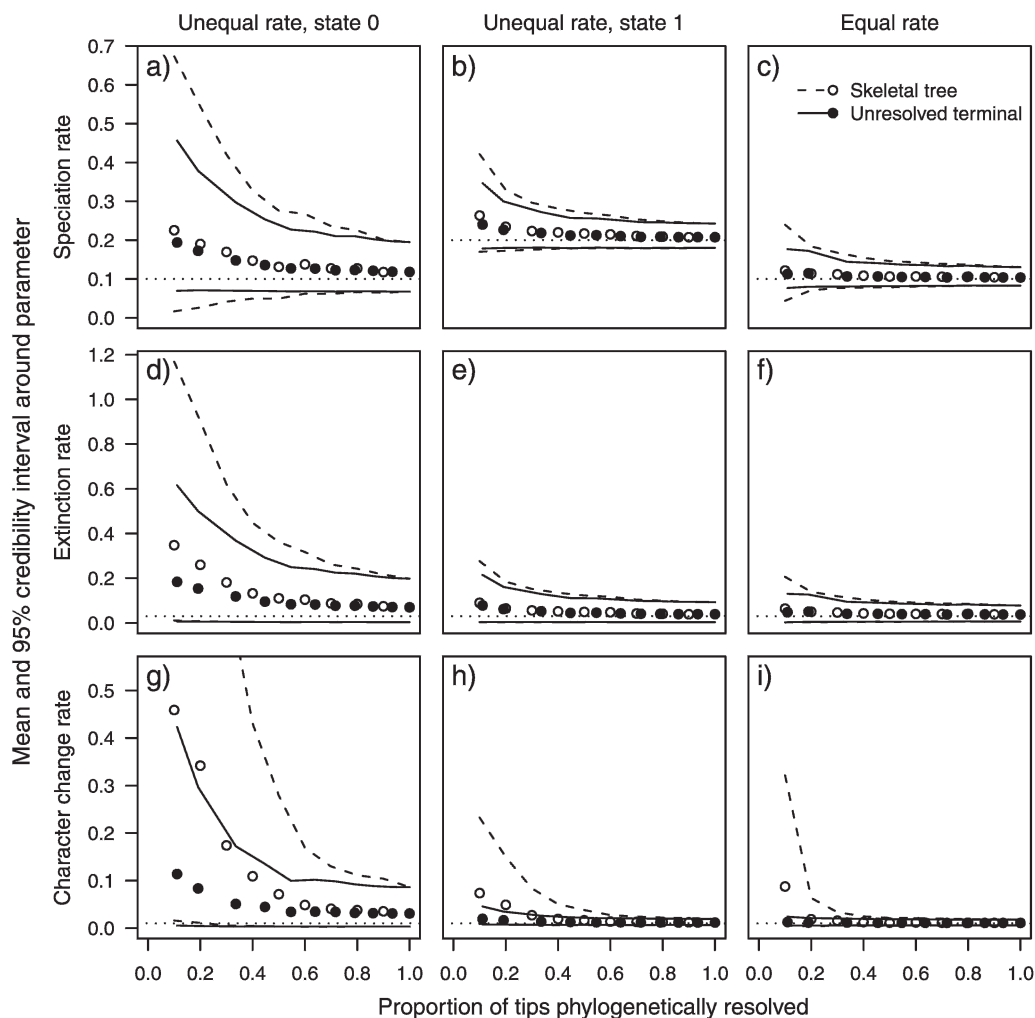

FIGURE 5. Uncertainty around BISSE parameter estimates as a function of phylogenetic knowledge. Points represent the mean for the estimate of each parameters, and the curves above and below indicate the mean 95% credibility interval, averaged over 30 different phylogenies. Dashed curves/open circles represent skeletal trees and solid curves/filled circles represent terminally unresolved trees. The horizontal dotted line indicates the true rate from the simulations. For skeletal trees, the proportion of tips reflects phylogenetic completeness, whereas for terminally unresolved trees, it represents the level of phylogenetic resolution. Trees were evolved with unequal speciation rates ($\lambda_1 > \lambda_0$, first two columns) or equal speciation rates (final column) and contained 500 species before sampling. Credibility intervals were calculated over the last 7500 points of three independent MCMC chains per tree, discarding the first 2500 points.

tips sampled decreased (Fig. 5). On unequal rate trees at very low phylogenetic resolution, the posterior probability distribution for q_{01} became very similar to the prior distribution.

The decrease in accuracy and precision with decreasing phylogenetic knowledge was more pronounced for skeletal trees than for terminally unresolved trees. This is because of the additional information that the unresolved clades contain in addition to the branching structure (i.e., the number of species and their states). In general, for a given number of tips present in a phylogeny (i.e., sampled species in skeletal trees, resolved species plus unresolved clades in terminally unresolved trees), terminally unresolved trees had lower bias in the mean parameter estimates and narrower credibility intervals than did skeletal trees. However, for well-estimated parameters (e.g., λ_1 and μ_1 on the unequal rate tree; Fig. 5b,e), the difference in uncertainty between these methods was small. The difference in precision was particularly pronounced for the character transition rates, which were typically well estimated on terminally unresolved trees, even with low phylogenetic resolution (Fig. 5g–i).

Rates of net diversification were well estimated as phylogenetic resolution decreased, despite increasing uncertainty in extinction rates (Fig. 6). For equal rate trees, the net diversification rate for each trait ($r_i = \lambda_i - \mu_i$) and the relative diversification rate r_{rel} were fairly insensitive to phylogenetic resolution, with no bias in the mean parameter estimates and little increase in the width of the credibility intervals, especially for terminally unresolved trees (Fig. 6a,c,e). Where speciation rates differed (unequal rate trees), the estimated diversification rates were sensitive to decreasing phylogenetic resolution, but less so than for the individual parameters. Particularly for terminally unresolved trees, the net diversification rate was well estimated even with low phylogenetic resolution. With the parameters used here, differential diversification was detectable on unequal rate trees at the 5% significance level until fewer than 30% of taxa were explicitly included in terminally unresolved trees and until 50% of taxa were included using skeletal trees (Fig. 6f).

APPLICATION TO SHOREBIRD DATA

Sexual dimorphism in body size or other traits in birds is thought be driven by sexual selection (Darwin 1871). Larger males might be favored by females or fare better in intrasexual conflict over mates, whereas small males (reversed sexual dimorphism) might be favored when sexual displays are acrobatic (Figuerola 1999). Sexual differences in any trait may indicate different optima for the two sexes, and therefore, intersexual conflict, which may increase speciation rates (Parker and Partridge 1998; Gavrilets 2000; Jablonski 2008). Comparative evidence linking sexual dimorphism with speciation is mixed. Several studies have found that sexual dimorphism in plumage or other display traits might promote increased diversification (Barraclough

et al. 1995; Parker and Partridge 1998; Owens et al. 1999), whereas other studies failed to find correlations between measures of sexual selection and diversification rates (e.g., Gage et al. 2002; Morrow and Pitcher 2003; Morrow et al. 2003).

Here, we use a recent supertree of shorebirds (Charadriiformes) to investigate the correlation between speciation rate and sexual dimorphism. Thomas et al. (2004) constructed a complete supertree of all 350 shorebird species. Although complete, this tree lacks resolution among many of the terminal clades, with large polytomies including up to 50 species. For each polytomy, we collapsed all species descended from any lineage within the polytomy into a terminal unresolved clade (Fig. 7). The resulting tree had 134 tips (with the 215 unresolved species included in 14 unresolved terminal clades). Many of the branch lengths in this tree are not strictly proportional to time, which reduces the information about extinction rates available in the tree.

We used a database of bird traits with separate measurements for males and females of body mass, wing length, tarsus length, bill length, and tail length (Lislevand et al. 2007). For each trait, we computed a standardized measure of dimorphism as $(x_m - x_f)/\bar{x}$, where x_m and x_f are the trait values in the males and females, respectively, and $\bar{x}$ is the mean of the male and female values. We regarded species as dimorphic if the absolute value of this dimorphism measure was greater than some threshold value for at least one of the five traits. This data set did not include state information for 77 species (22%). These were treated using the methods described in the Incomplete Character State Knowledge section.

Although the general form of the marginal posterior distributions was well characterized after 10,000 steps of the MCMC algorithm, it was difficult to characterize some of the peaks in the multimodal posterior distributions. To improve resolution, we ran eight independent MCMC chains for 100,000 iterations. The precise credibility intervals changed slightly, but not our general conclusions.

The relationship between sexual dimorphism and speciation and diversification rates depended on the threshold difference in body size used. For low to medium thresholds of sexual dimorphism ($\leq 15\%$), the maximum likelihood and mean posterior probability speciation and diversification rates were higher for sexually dimorphic lineages than monomorphic lineages (Fig. 8). In contrast, for very high thresholds (20%), the diversification rate for dimorphic species was lower than that of monomorphic species. The difference was supported by high posterior probability values only for the 15% threshold. The maximum likelihood extinction rate and the mode of the posterior probability distribution was generally zero for both character states across all threshold values examined. We found that character transition rates from sexual dimorphism to monomorphism were higher than the reverse across most thresholds used (Fig. 8), with this difference being most pronounced at 15% (significant at the 5% level for
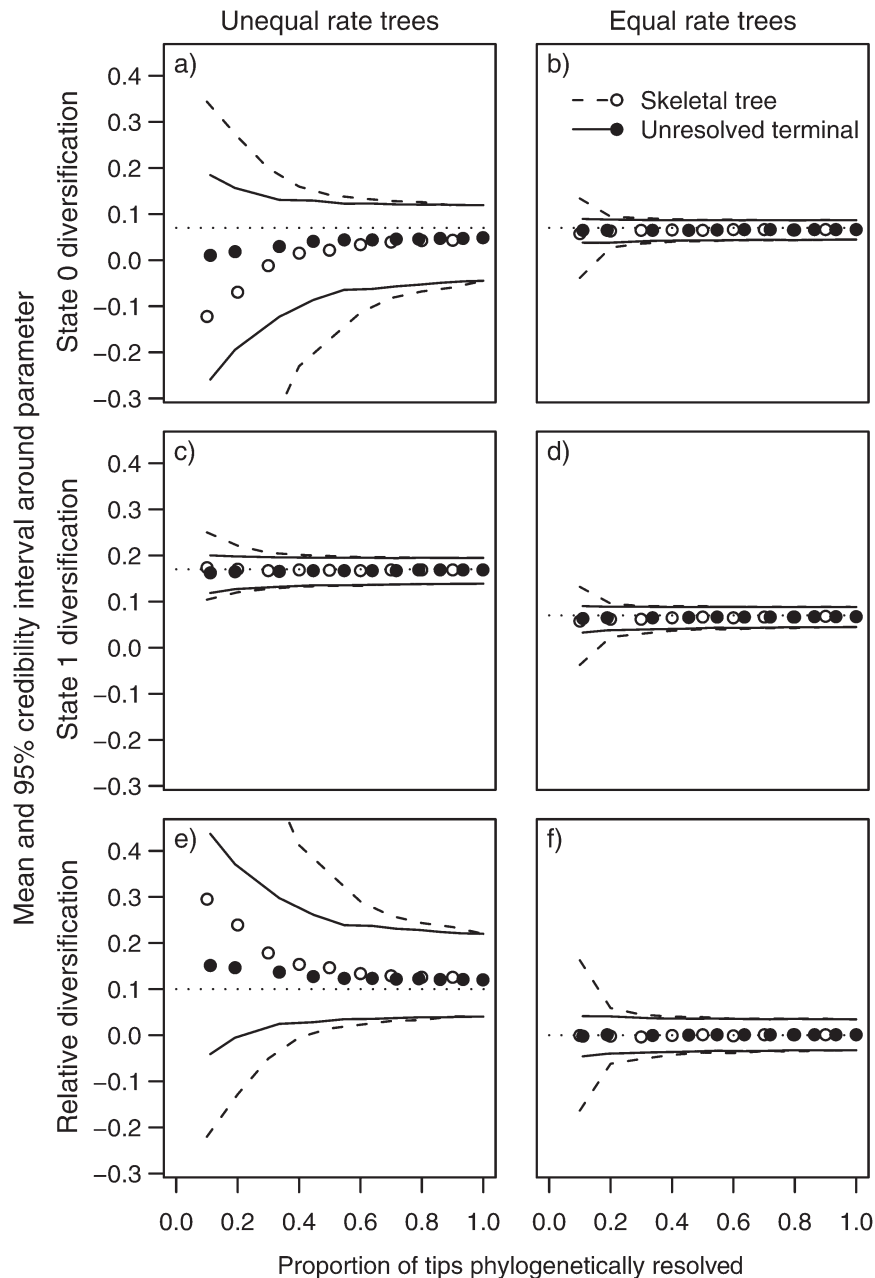
FIGURE 6. Uncertainty around diversification rate estimates as a function of phylogenetic knowledge. Panels a and b show the net diversification rate in state 0 ($\lambda_0 - \mu_0$), panels c and d show the net diversification rate in state 1 ($\lambda_1 - \mu_1$), and panels e and f show the relative diversification rate, r_{rel}. See Fig. 5 for details.

the 10% and 15% thresholds). It is perhaps not surprising that the choice of threshold has such an effect, as the dimorphic state becomes rare as the threshold is raised, and the rarer a state the more likely its diversification rate would be biased downward.

As with previous studies, our results suggest that evidence for a correlation between sexual dimorphism and diversification rates is mixed at best (Barraclough et al. 1995; Parker and Partridge 1998; Owens et al. 1999; Gage et al. 2002; Morrow and Pitcher 2003; Morrow et al. 2003). However, rather than dividing groups arbitrarily

into clades that have just one character state (e.g., Barraclough et al. 1995), our approach allowed us to make use of all of the available phylogenetic and character state information.

DISCUSSION

In this paper, we have developed two methods for estimating the effect of a trait on speciation and extinction rates from incomplete and incompletely resolved phylogenies. Testing these methods with simulations, it
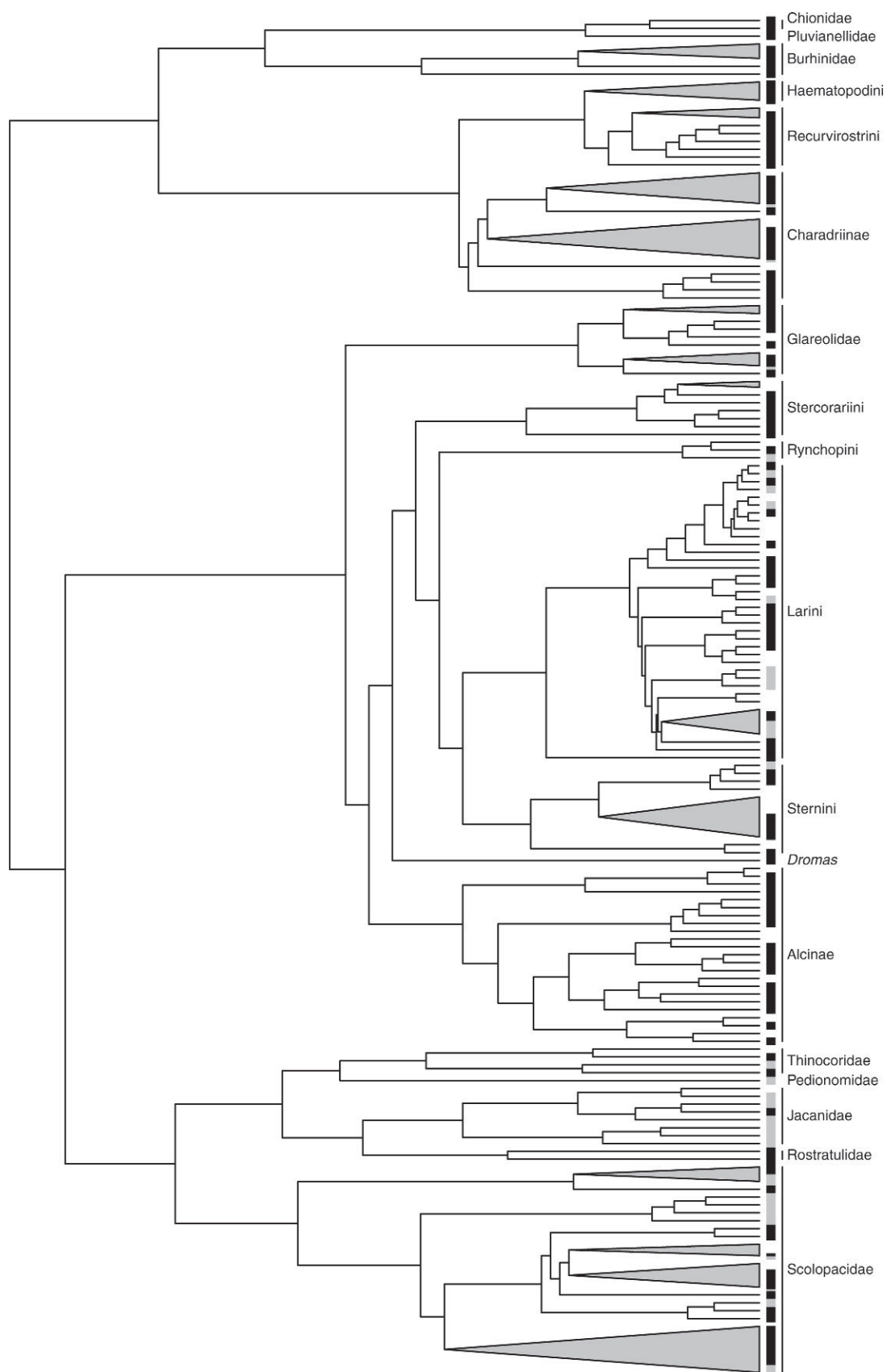
FIGURE 7. (Continued)

was possible to estimate diversification rates from even poorly sampled phylogenies (Fig. 6). Where trees were simulated with equal speciation rates, there was little increase in uncertainty in the estimates of differential diversification with decreasing phylogenetic information, even when as few as 20% of species were phylogenetically placed. This is surprising because terminally unresolved trees lack much of the fine branching structure present at the tips of a completely resolved phylogeny (Fig. 1). However, the power to detect differences in individual parameters depended more strongly on phylogenetic structure.

Because the terminally unresolved clade method uses the branching structure available to the skeletal method (for a given number of tips in a tree), differences between these two methods are due to the additional information about the placement of the missing taxa in the terminal unresolved clades. In cases where a given species sample can be reasonably assumed to be a random draw from all extant species, the skeletal tree method provides a simple way of estimating speciation and extinction. In particular, the phylogenetic relationships and character states of nonsampled species do not need to be incorporated. Where phylogenies are almost complete, the loss of power using this method is fairly low. For poorly sampled phylogenies (fewer than 25% species included in our 500 species phylogenies), the uncertainty around parameters became very large to the point where inference was not possible (Figs. 5–6).

The terminal unresolved clade approach can avoid most of this loss of power, provided all species not included in the phylogeny can be grouped into terminally unresolved clades. The effect of including the terminally unresolved clades was strongest for the character transition rates (q_{01} and q_{10}, Fig. 5), and it allowed detection of differential diversification on poorly sampled trees (Fig. 6). However, the terminally unresolved tree method can only be used where every species can be assigned to a terminal unresolved clade. Deeper phylogenetic uncertainties such as unresolved paraphyletic groups have not yet been incorporated, and some known phylogenetic information may need to be discarded to use the current methods by including only terminal unresolved clades (see Fig. 1d). This method is also substantially more computationally demanding than the skeletal tree approach and is limited at present to unresolved clades that contain fewer than approximately 200 species. With 200 species, there are more than 20,000 possible clade compositions (numbers of species in each state), and even with modern matrix exponentiation techniques, the calculations become both very slow and prone to numerical underflow (Sidje 1998).

Missing data and incompleteness are generally unavoidable in comparative macroevolutionary analyses. Frequently, phylogenetic trees will contain species that are less related than expected by chance to maximize coverage over the true phylogeny (e.g., Moyle et al. 2009). In these cases, the terminally unresolved tree method will be appropriate. If a phylogeny is almost complete, missing only a few taxa, but for which the placement is uncertain (e.g., the cases considered by Bokma 2008), the skeletal tree method should be satisfactory (Fig. 5). It may not always be possible to know with complete certainty where taxa that are not included in a tree should be placed within a terminally unresolved tree. In this case, one could run an analysis over possible placements of missing taxa, integrating over this uncertainty (Lutzoni et al. 2001). It is probably not possible to know in general for cases such as Fig. 1d whether collapsing known structure into an terminal unresolved clade or treating the tree as a skeleton tree will suffer the least reduction in power, as these approaches both lose data. A final caution about the pattern of missing taxa: outgroups are generally poorly sampled relative to the ingroup and should be removed from the tree prior to calculation of likelihoods with BISSE. Poorly sampled outgroups would certainly violate assumptions of random taxon sampling.

Maddison et al. (2007) tested whether simpler models may fit the data better by comparing models where rates were character dependent to simpler models where rates were character independent (e.g., a model where $\lambda_0 \neq \lambda_1$ to a model where $\lambda_0 = \lambda_1$) using likelihood ratio tests. Model selection between the full model and reduced models could also be done in a Bayesian framework using reversible jump Markov chain Monte Carlo (RJMCMC; Green 1995). RJMCMC alters the Markov chain to propose different models at some steps, for example, changing from the full six parameter model to one of the three simpler five parameter models. The posterior probability distribution of different models can then be directly compared. An RJMCMC approach would provide a natural way of removing parameters from the analysis where there is little to no phylogenetic signal (e.g., q_{01} in the poorly resolved unequal rate tree, Fig. 5g). This approach has been used successfully elsewhere in phylogenetic inference (e.g., Pagel and Meade 2006).

Using either of the sampling methods explored here, BISSE likelihoods may be computed for partially complete phylogenies. Future work is needed to handle much larger unresolved clades (> 200 species) and to handle orphan taxa, whose phylogenetic position is deep in the tree and uncertain. Such extensions are needed to analyze "higher level" phylogenies that are complete at a taxonomic level above species but contain

FIGURE 7. Phylogenetic tree of the 350 species of shorebirds (Charadriiformes) and measures of sexual dimorphism are based on Thomas et al. (2004). Gray triangles indicate unresolved clades, with the height of the triangle being proportional to the square root of the number of species. Character states at the 15% threshold level are indicated at the tips; gray = sexually dimorphic, black = sexually monomorphic, and white = no data. For clarity, only family or subfamily names are shown.
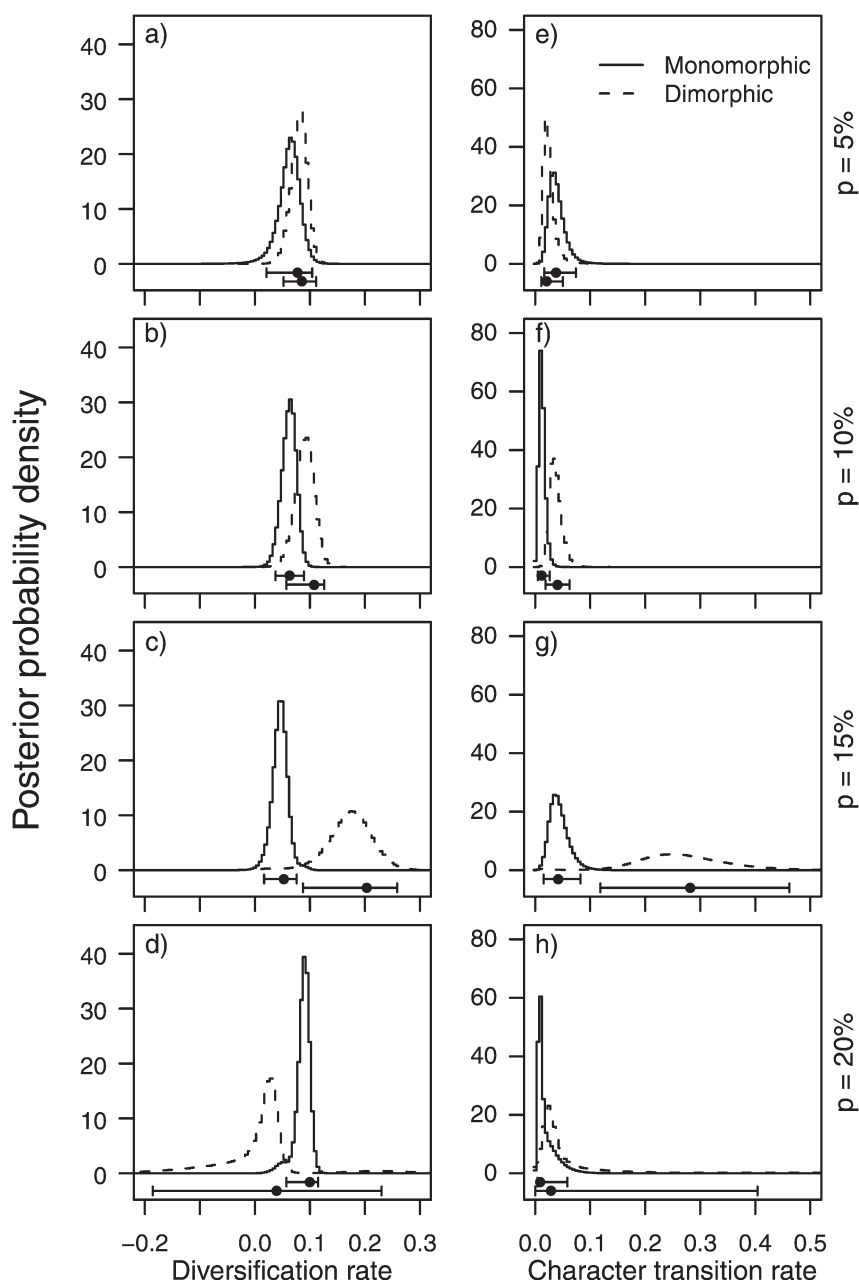
FIGURE 8. Marginal posterior probability distributions for the sexual dimorphism-dependent diversification rates (a–d) and character transition rates (e–h) inferred from a supertree of shorebirds (Thomas et al. 2004). Panels in different rows use a different threshold level of sexual dimporphism to classify species as monomorphic and dimorphic. Solid curves show the distribution for sexually monomorphic species and dashed curves for dimorphic species. The horizontal bar and point indicate the 95% credibility interval and maximum likelihood estimate.

large numbers of species in their unresolved clades (e.g., Davies et al. 2004; Hackett et al. 2008).

REFERENCES

Barraclough T.G., Harvey P.H., Nee S. 1995. Sexual selection and taxonomic diversity in passerine birds. Proc. R. Soc. Lond. B 259: 211–215.

Barraclough T.G., Nee S., Harvey P.H. 1998. Sister-group analysis in identifying correlates of diversification. Evol. Ecol. 12:751–754.

Bokma F. 2008. Bayesian estimation of speciation and extinction probabilities from (in)complete phylogenies. Evolution 62:2441–2445.

Cardillo M. 1999. Latitude and rates of diversification in birds and butterflies. Proc. R. Soc. Lond. B Biol. Sci. 266:1221–1225.

Churchill G.A. 2000. Inferring ancestral character states. In: Evolutionary biology. Volume 32. Chapter 6. New York: Kluwer Academic. p. 117–134.

Coyne J.A., Orr H.A. 2004. Speciation. Sunderland (MA): Sinauer Associates.

Darwin C. 1871. The descent of man, and selection in relation to sex. London: Murray.

Davies T.J., Barraclough T.G., Chase M.W., Soltis P.S., Soltis D.E. 2004. Darwin's abominable mystery: insights from a supertree of the angiosperms. Proc. Natl. Acad. Sci. U.S.A. 101:1904–1909.

Farrell B.D. 1998. "Inordinate fondness" explained: why are there so many beetles. Science 281:555–559.

Figuerola J. 1999. A comparative study on the evolution of reversed size dimorphism in monogamous waders. Biol. J. Linn. Soc. Lond. 67:1–18.

Gage M.J.G., Parker G.A., Nylin S., Wiklund C. 2002. Sexual selection and speciation in mammals, butterflies and spiders. Proc. R. Soc. Lond. B 269:2309–2316.

Gavrilets S. 2000. Rapid evolution of reproductive barriers driven by sexual conflict. Nature 403:886–889.

Gelman A., Carlin J.B., Stern H.S., Rubin D.B. 1995. Bayesian data analysis. London: Chapman & Hall.

Goldberg E.E., Igić B. 2008. On phylogenetic tests of irreversible evolution. Evolution 62:2727–2741.

Green P.J. 1995. Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. Biometrika 82:711–732.

Hackett S.J., Kimball R.T., Reddy S. Bowie R.C.K., Braun E.L., Braun M.J., Chojnowski J.L., Cox W.A., Han K.-L., Harshman J., Huddleston C.J., Marks B.D., Miglia K.J., Moore W.S., Sheldon F.H., Steadman D.W., Witt C.C., Yuri T. 2008. A phylogenomic study of birds reveals their evolutionary history. Science 320:1763–1768.

Heilbuth J.C. 2000. Lower species richness in dioecious clades. Am. Nat. 156:221–241.

Huelsenbeck J.P., Larget B., Miller R.E., Ronquist F. 2002. Potential applications and pitfalls of Bayesian inference of phylogeny. Syst. Biol. 51:673 – 688.

Jablonski D. 2008. Species selection: theory and data. Annu. Rev. Ecol. Evol. Syst. 39:501–524.

Lislevand T., Figuerola J., Székely T. 2007. Avian body sizes in relation to fecundity, mating system, display behavior, and resource sharing. Ecology 88:1605.

Lutzoni F., Pagel, M. Reeb V. 2001. Major fungal lineages are derived from lichen symbiotic ancestors. Nature 411:937–940.

MacKay D.J.C. 2003. Information theory, inference, and learning algorithms. New York: Cambridge University Press.

Maddison W.P., Maddison D.R. 2006. Mesquite: a modular system for evolutionary analysis. Version 1.1. Available from: URL http://www.mesquiteproject.org.

Maddison W.P., Maddison D.R. 2008. MESQUITE: a modular system for evolutionary analysis. Version 2.5. Available from: URL http://www.mesquiteproject.org.

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.

Mitra S., Landel H., Pruett-Jones S.. 1996. Species richness covaries with mating system in birds. Auk 113:544–551.

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

Morrow E.H., Pitcher T.E. 2003. Sexual selection and the risk of extinction in birds. Proc. R. Soc. Lond. B 270:1793–1799.

Morrow E.H., Pitcher T.E., Arnqvist G. 2003. No evidence that sexual selection is an 'engine of speciation' in birds. Ecol. Lett. 6:228–234.

Moyle R.G., Filardi C.E., Smith C.E., Diamond J. 2009. Explosive Pleistocene diversification and hemispheric expansion of a "great speciator." Proc. Natl. Acad. Sci. U.S.A. 106:1863–1868.

Neal R.M. 2003. Slice sampling. Ann. Stat. 31:705–767.

Nee S. 2006. Birth-death models in macroevolution. Annu. Rev. Ecol. Evol. 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.

Owens I.P.F., Bennett P.M., Harvey P.H. 1999. Species richness among birds: body size, life histoy, sexual selection or ecology. Proc. R. Soc. Lond. B 266:933–939.

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

Pagel M., Meade A. 2006. Bayesian analysis of correlated evolution of discrete characters by reversible-jump Markov chain Monte Carlo. Am. Nat. 167:808–825.

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

Parker G.A., Partridge L. 1998. Sexual conflict and speciation. Philos. Trans. R. Soc. Lond. B Biol. Sci. 353:261–274.

Rabosky D.L., Donnellan S.C., Talaba A.L., Lovette I.J. 2007. Exceptional among-lineage variation in diversification rates during the radiation of Australia's most diverse vertebrate clade. Proc. R. Soc. Lond. B Biol. Sci. 274:2915–2923.

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

Ricklefs R.E. 2007. Estimating diversification rates from phylogenetic information. Trends Ecol. Evol. 22:601–610.

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

Sidje R.B. 1998. EXPOKIT: a software package for computing matrix exponentials. ACM Trans. Math. Softw. 24:130–156.

Stanley S.M. 1975. A theory of evolution above the species level. Proc. Natl. Acad. Sci. U.S.A. 72:646–650.

Thomas G.H., Wills M.A., Székely T. 2004. A supertree approach to shorebird phylogeny. BMC Evol. Biol. 4:28.

APPENDIX 1

Root-State Calculations

At the root, R, we have the two probabilities, D_{R0} and D_{R1}, corresponding to the possible character states at the root. The overall likelihood must sum over the probabilities that the root was in each state. In Schluter et al. (1997), the Ds at the root were weighted evenly, which assumes that the lineage arose out of a group with 50% of taxa in each state, therefore potentially being out of equilibrium with the inferred model of evolution. The model parameters provide some knowledge, however. For example, if transitions away from character state 0 are much more frequent than the reverse ($q_{01} > q_{10}$) and if speciation/extinction rates do not depend strongly on the character state, then we would expect that the system is more likely to be in state 1 than in state 0 at any point in time, including at the root. In Maddison et al. (2007), the information provided by the model was used to weight the Ds at the root by the equilibrium frequencies for the character states given by the model (following Maddison WP and Maddison DR 2006). This

implicitly assumes that a sufficient amount of time has passed prior to the root, so that the root state can be assumed to be a random draw from an equilibrium distribution. However, this assumption does not account for cases where the traits are novel or have yet to reach equilibrium.

Here, we treat the root state as a nuisance parameter (Gelman et al. 1995) and use an alternative root assignment that weights each root state according to its probability of giving rise to the extant data, given the model parameters and the tree. This probability is given by the likelihood given that the root is in state i divided by the sum of the likelihoods over both root states, $D_{Ri}/(D_{R0} + D_{R1})$. The overall likelihood is then as follows:

$$D_R = D_{R0}\frac{D_{R0}}{D_{R0} + D_{R1}} + D_{R1}\frac{D_{R1}}{D_{R0} + D_{R1}}. \quad (10)$$

As a test case, consider a tree that consists of a single branch with an infinitesimally short branch length where the single extant taxon is in state 0. Assuming that none of the transition parameters is very large, then D_{R0} is nearly one and D_{R1} is nearly zero. Assigning the root state according to the probability of the root leading to the data, as we do in Equation (10), we infer that there is nearly a 100% probability that the root was in state 0 and the overall probability of the data given the model is nearly 1. Assigning the root state uniformly, we would be assuming that there is a 50% probability that the root state was 1, even though we know this to be impossible (more precisely, it has an infinitesimally small probability of being true given the infinitesimally short branch and the fact that the single extant taxon is in state 0). Similarly, assigning the root state according to the equilibrium distribution would give some nonzero probability to the root being in state 1, unless $q_{01} = 0$.

Furthermore, if there is directional change in the character, assigning the root to the equilibrium distribution incorrectly forces the character to take on the value that it will have in the long-term future, not what it is likely to have been in the past. For example, if all organisms started in state 0 and are evolving into state 1 ($q_{01} \gg q_{10}$, with all else equal), the equilibrium distribution method will incorrectly assign the root to state 1, even if most extant species are still in state 0. By contrast, assigning the root states according to their relative likelihoods of explaining the data will assign a high probability on the root having been in state 0 when most species are in state 0. Equation (10) has the further advantage that the only quantities needed for its calculation are D_{R0} and D_{R1}, which are already known once BISSE has traversed the tree. Goldberg and Igić (2008) have recently explored the effect of root state on BISSE calculations and found that they can have a strong effect on conclusions, especially where character change is unidirectional. Our approach (Equation (10)) can approximate the ancestral root state and result in reasonable character change rate estimates for the situations described above. In practice,

sensitivity to the root state is easily detected by comparing D_{R0} and D_{R1}.

APPENDIX 2

Character-Independent Model

When the speciation and extinction rates do not depend on a character, our likelihood calculations reduce to existing models of character-independent evolution. Analytical solutions for these models are known, removing the need to use numerical approaches to calculating the likelihoods.

Skeletal trees.—With character-independent speciation rates, the skeletal tree likelihoods reduce to the method of Nee et al. (1994). The character-independent analogues of Equation (1) are as follows:

$$\frac{dD_N}{dt} = -(\lambda + \mu)D_N(t) + 2\lambda E(t)D_N(t), \quad (11a)$$

$$\frac{dE}{dt} = \mu - (\mu + \lambda)E(t) + \lambda E(t)^2 \quad (11b)$$

(Maddison et al. 2007), where λ and μ are the character-independent speciation and extinction rates. If a fraction, f, of all species are sampled in the phylogeny, then the initial conditions are $E(0) = 1 - f$ and $D(0) = f$. It is possible to derive an analytical solution to Equations (11) describing changes along a single branch. Using the initial condition $E(0) = 1 - f$, the solution to Equation (11b) for the extinction rate is as follows:

$$E(t)=1 - \frac{f(\lambda - \mu)}{f\lambda - e^{(\lambda-\mu)t}(\mu - \lambda(1 - f))} \quad \text{if } \lambda \neq \mu, \quad (12a)$$

$$E(t)=\frac{1 - f + f\lambda t}{1 - f\lambda t} \quad \text{if } \lambda = \mu. \quad (12b)$$

Substituting Equation (12) into Equation (11a) and solving for $D_N(t)$ gives

$$D_N(t) = e^{-(\lambda-\mu)(t-t_N)}\frac{(f\lambda - e^{-(\lambda-\mu)t_N}(\mu - \lambda(1-f)))^2}{(f\lambda - e^{-(\lambda-\mu)t}(\mu - \lambda(1-f)))^2}$$
$$\times D_N(t_N) \text{ if } \lambda \neq \mu, \quad (13a)$$

$$D_N(t) = \frac{(1 + f\lambda t_N)^2}{(1 + f\lambda t)^2}D_N(t_N) \quad \text{if } \lambda = \mu, \quad (13b)$$

where t_N represents the time depth (since the present) of node N. Equations (12) and (13) reduce to Equations (9) and (10) in Maddison et al. (2007) if sampling is complete ($f = 1$).

These equations can be used as in Maddison et al. (2007) to compute the likelihood for the entire

phylogeny:

$$D_R(t_R) = \left[\prod_{k=1}^{2n} e^{-(\lambda-\mu)(t_{k,b}-t_{k,t})} \right.$$

$$\left. \times \left(\frac{f\lambda - e^{-(\lambda-\mu)t_{k,t}}(\mu - \lambda(1-f))}{f\lambda - e^{-(\lambda-\mu)t_{k,b}}(\mu - \lambda(1-f))} \right)^2 \right] \lambda^n \quad \text{if } \lambda \neq \mu,$$

(14a)

$$D_R(t_R) = \left[\prod_{k=1}^{2n} \frac{(1+f\lambda t_{k,t})^2}{(1+f\lambda t_{k,b})^2} \right] \lambda^n \quad \text{if } \lambda = \mu \quad (14b)$$

where $t_{k,b}$ and $t_{k,t}$ are the times at the base and tip of the kth branch, respectively, and the product is taken over all $2n$ branches for a tree containing n nodes.

Equation (14) is consistent with the results from Nee et al. (1994) for the character-independent case. Nee et al. (1994) does not explicitly give equations for the probability of a sampled phylogeny, given a speciation rate, extinction rate, and sampling probability, so we state them here. Using the notation of Nee et al. (1994), the probability of the data is as follows:

$$L = (N-1)! f^{N-2} \lambda^{N-2} \left(\prod_{i=3}^{N} P_s(t_i, T) \right)$$

$$\times (1 - u_s(x_2))^2 \prod_{i=3}^{N} (1 - u_s(x_i)), \quad (15)$$

where N is the number of tips in the phylogeny, x_i is the time between the present and the node that splits the phylogeny into i branches (so that x_2 is the distance to the root), $P_s(t_i, T)$ is the probability that a lineage originating at time t_i leaves at least one descendant at the present, time T (given by Nee et al. 1994, equation (34)),

and $u_s(t)$ is

$$u_s(t) = 1 - \frac{1-a}{f\, e^{rx_i} - a + 1 - f}, \quad (16)$$

where $a = \mu/\lambda$ and $r = \lambda - \mu$. Equation (16) can be derived from equations (29) and (33) in Nee et al. (1994). After some algebra, equation (14a) can be shown to be equal to equation (15), after conditioning on the existence of a root node and two surviving lineages (see Maddison et al. 2007 for similar calculations with $f = 1$).

Terminally unresolved trees.—To calculate the likelihood for terminally unresolved trees, we must first calculate the probability of unresolved clades given the speciation and extinction rates. We could rederive $\mathbf{Q}$ and $\vec{x}$, ignoring character state changes, which now do not affect diversification, and use Equation (3) to compute the probability of the clade. However, without transitions, this can be viewed as a birth–death process for which an analytical solution is available. The probability of k lineages arising and surviving to the present from a single ancestor over a period of time t is given by Nee et al. (1994):

$$P(k,t) = \frac{\lambda - \mu}{\lambda - \mu e^{-(\lambda-\mu)t}} (1 - u(t)) u(t)^{k-1}, \quad k > 0, \lambda \neq \mu,$$

(17a)

$$P(k,t) = \frac{(t\lambda)^{i-1}}{(1+t\lambda)^{i+1}}, \quad k > 0, \lambda = \mu, \quad (17b)$$

where $u(t)$ is $u_s(t)$ from Equation (16) with $f = 1$. If an unresolved clade has n species and originated at time t_N, then $P(n, t_N)$ can be used for $D_N(t_N)$ in Equation (9) of Maddison et al. (2007). This approach has been used by Rabosky et al. (2007) to estimate speciation and extinction rates from a terminally unresolved lizard phylogeny.