

# Phylogenetic Inference using RevBayes

## *State-Specific Diversification Rate Estimation*

Sebastian Höhna and Will Freyman

## 1 Overview: Diversification Rate Estimation

Models of speciation and extinction are fundamental to any phylogenetic analysis of macroevolutionary processes (*e.g.*, divergence time estimation, diversification rate estimation, continuous and discrete trait evolution, and historical biogeography). First, a prior model describing the distribution of speciation events over time is critical to estimating phylogenies with branch lengths proportional to time. Second, stochastic branching models allow for inference of speciation and extinction rates. These inferences allow us to investigate key questions in evolutionary biology.

Diversification-rate parameters may be included as nuisance parameters of other phylogenetic models—*i.e.*, where these diversification-rate parameters are not of direct interest. For example, many methods for estimating species divergence times—such as BEAST (Drummond et al. 2012), MrBayes (Ronquist et al. 2012), and RevBayes (Höhna et al. 2016)—implement ‘relaxed-clock models’ that include a constant-rate birth-death branching process as a prior model on the distribution of tree topologies and node ages. Although the parameters of these ‘tree priors’ are not typically of direct interest, they are nevertheless estimated as part of the joint posterior probability distribution of the relaxed-clock model, and so can be estimated simply by querying the corresponding marginal posterior probability densities. In fact, this may provide more robust estimates of the diversification-rate parameters, as they accommodate uncertainty in the other phylogenetic-model parameters (including the tree topology, divergence-time estimates, and the other relaxed-clock model parameters). More recent work, *e.g.*, Heath et al. (2014), uses macroevolutionary models (the fossilized birth-death process) to calibrate phylogenies and thus to infer dated trees.

In these tutorials we focus on the different types of macroevolutionary models to study diversification processes and thus the diversification-rate parameters themselves. Nevertheless, these macroevolutionary models should be used for other evolutionary questions, when an appropriate prior distribution on the tree and divergence times is needed.

### 1.1 Types of Hypotheses for Estimating Diversification Rates

Many evolutionary phenomena entail differential rates of diversification (speciation – extinction); *e.g.*, adaptive radiation, diversity-dependent diversification, key innovations, and mass extinction. The specific study questions regarding lineage diversification may be classified within three fundamental categories of inference problems. Admittedly, this classification scheme is somewhat arbitrary, but it is nevertheless useful, as it allows users to navigate the ever-increasing number of available phylogenetic methods. Below, we describe each of the fundamental questions regarding diversification rates.

**(1) Diversification-rate through time estimation** *What is the (constant) rate of diversification in my study group?* The most basic models estimate parameters of the stochastic-branching process (*i.e.*, rates of speciation and extinction, or composite parameters such as net-diversification and relative-extinction

rates) under the assumption that rates have remained constant across lineages and through time; *i.e.*, under a constant-rate birth-death stochastic-branching process model (Nee et al. 1994). Extensions to the (basic) constant-rate models include diversification-rate variation through time (Stadler 2011; Höhna 2015). First, we might ask whether there is evidence of an episodic, tree-wide increase in diversification rates (associated with a sudden increase in speciation rate and/or decrease in extinction rate), as might occur during an episode of adaptive radiation. A second question asks whether there is evidence of a continuous/gradual decrease in diversification rates through time (associated with decreasing speciation rates and/or increasing extinction rates), as might occur because of diversity-dependent diversification (*i.e.*, where competitive ecological interactions among the species of a growing tree decrease the opportunities for speciation and/or increase the probability of extinction, *e.g.*, Höhna (2014)). Third, we can ask whether changes in diversification rates are correlated with environmental factors, such as environmental CO<sub>2</sub> or temperature (Condamine et al. 2013). A final question in this category asks whether our study tree was impacted by a mass-extinction event (where a large fraction of the standing species diversity is suddenly lost, *e.g.*, May et al. (2016)). The common theme of these studies is that the diversification process is tree-wide, that is, all lineages of the study group have the exact same rates at a given time.

**(2) Diversification-rate variation across branches estimation** *Is there evidence that diversification rates have varied significantly across the branches of my study group?* Models have been developed to detect departures from rate constancy across lineages; these tests are analogous to methods that test for departures from a molecular clock—*i.e.*, to assess whether substitution rates vary significantly across lineages (Alfaro et al. 2009; Rabosky 2014). These models are important for assessing whether a given tree violates the assumptions of rate homogeneity among lineages. Furthermore, these models are important to answer questions such as: *What are the branch-specific diversification rates?*; and *Have there been significant diversification-rate shifts along branches in my study group, and if so, how many shifts, what magnitude of rate-shifts and along which branches?*

**(3) Character-dependent diversification-rate estimation** *Are diversification rates correlated with some variable in my study group?* Character-dependent diversification-rate models aim to identify overall correlations between diversification rates and organismal features (binary and multi-state discrete morphological traits, continuous morphological traits, geographic range, etc.). For example, one can hypothesize that a binary character, say if an organism is herbivorous/carnivorous or self-compatible/self-incompatible, impact the diversification rates. Then, if the organism is in state 0 (*e.g.*, is herbivorous) it has a lower (or higher) diversification rate than if the organism is in state 1 (*e.g.*, carnivorous) (Maddison et al. 2007).

## 2 Models

We begin this section with a general introduction to the stochastic birth-death branching process that underlies inference of diversification rates in RevBayes. This primer will provide some details on the relevant theory of stochastic-branching process models. We appreciate that some readers may want to skip this somewhat technical primer; however, we believe that a better understanding of the relevant theory provides a foundation for performing better inferences. We then discuss a variety of specific birth-death models, but emphasize that these examples represent only a tiny fraction of the possible diversification-rate models that can be specified in RevBayes.

### 2.1 The birth-death branching process

Our approach is based on the *reconstructed evolutionary process* described by Nee et al. (1994); a birth-death process in which only sampled, extant lineages are observed. Let  $N(t)$  denote the number of species at time

$t$ . Assume the process starts at time  $t_1$  (the ‘crown’ age of the most recent common ancestor of the study group,  $t_{\text{MRCA}}$ ) when there are two species. Thus, the process is initiated with two species,  $N(t_1) = 2$ . We condition the process on sampling at least one descendant from each of these initial two lineages; otherwise  $t_1$  would not correspond to the  $t_{\text{MRCA}}$  of our study group. Each lineage evolves independently of all other lineages, giving rise to exactly one new lineage with rate  $b(t)$  and losing one existing lineage with rate  $d(t)$  (Figure 1 and Figure 2). Note that although each lineage evolves independently, all lineages share both a common (tree-wide) speciation rate  $b(t)$  and a common extinction rate  $d(t)$  (Nee et al. 1994; Höhna 2015). Additionally, at certain times,  $t_{\text{M}}$ , a mass-extinction event occurs and each species existing at that time has the same probability,  $\rho$ , of survival. Finally, all extinct lineages are pruned and only the reconstructed tree remains (Figure 1).

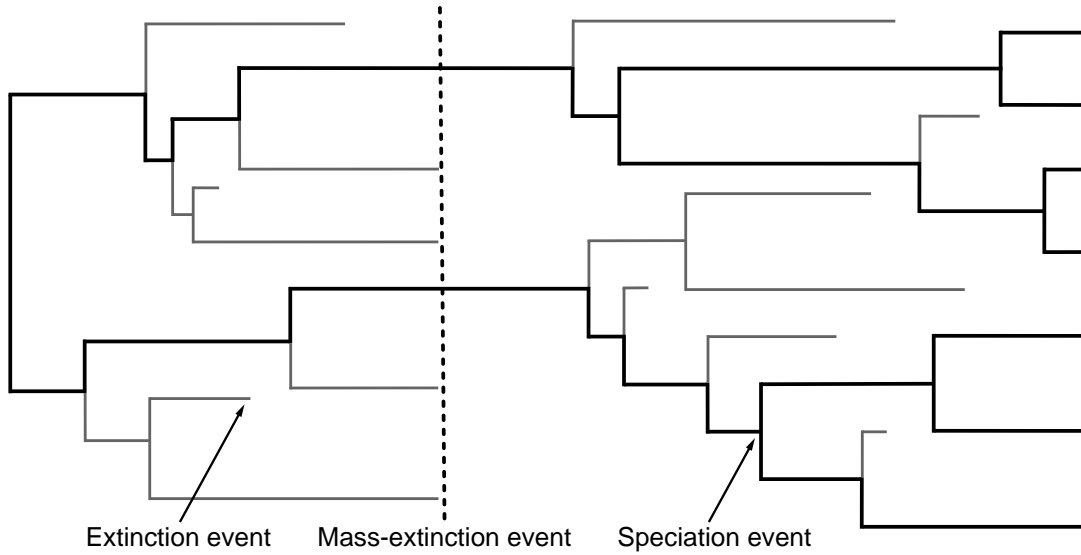


Figure 1: A realization of the birth-death process with mass extinction. Lineages that have no extant or sampled descendant are shown in gray and surviving lineages are shown in a thicker black line.

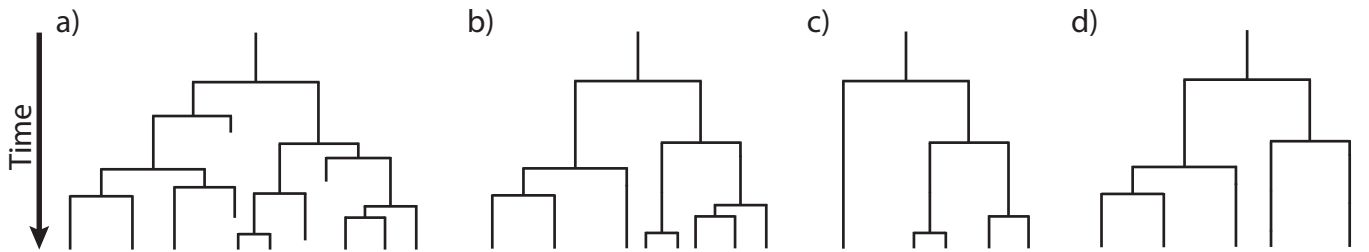


Figure 2: **Examples of trees produced under a birth-death process.** The process is initiated at the first speciation event (the ‘crown-age’ of the MRCA) when there are two initial lineages. At each speciation event the ancestral lineage is replaced by two descendant lineages. At an extinction event one lineage simply terminates. (A) A complete tree including extinct lineages. (B) The reconstructed tree of tree from A with extinct lineages pruned away. (C) A *uniform* subsample of the tree from B, where each species was sampled with equal probability,  $\rho$ . (D) A *diversified* subsample of the tree from B, where the species were selected so as to maximize diversity.

To condition the probability of observing the branching times on the survival of both lineages that descend from the root, we divide by  $P(N(T) > 0 | N(0) = 1)^2$ . Then, the probability density of the branching times,

$\mathbb{T}$ , becomes

$$P(\mathbb{T}) = \frac{\overbrace{P(N(T) = 1 \mid N(0) = 1)^2}^{\text{both initial lineages have one descendant}}}{\underbrace{P(N(T) > 0 \mid N(0) = 1)^2}_{\text{both initial lineages survive}}} \times \prod_{i=2}^{n-1} \overbrace{i \times b(t_i)}^{\text{speciation rate}} \times \overbrace{P(N(T) = 1 \mid N(t_i) = 1)}^{\text{lineage has one descendant}},$$

and the probability density of the reconstructed tree (topology and branching times) is then

$$P(\Psi) = \frac{2^{n-1}}{n!(n-1)!} \times \left( \frac{P(N(T) = 1 \mid N(0) = 1)}{P(N(T) > 0 \mid N(0) = 1)} \right)^2 \times \prod_{i=2}^{n-1} i \times b(t_i) \times P(N(T) = 1 \mid N(t_i) = 1) \quad (1)$$

We can expand Equation (1) by substituting  $P(N(T) > 0 \mid N(t) = 1)^2 \exp(r(t, T))$  for  $P(N(T) = 1 \mid N(t) = 1)$ , where  $r(u, v) = \int_u^v d(t) - b(t)dt$ ; the above equation becomes

$$\begin{aligned} P(\Psi) &= \frac{2^{n-1}}{n!(n-1)!} \times \left( \frac{P(N(T) > 0 \mid N(0) = 1)^2 \exp(r(0, T))}{P(N(T) > 0 \mid N(0) = 1)} \right)^2 \\ &\quad \times \prod_{i=2}^{n-1} i \times b(t_i) \times P(N(T) > 0 \mid N(t_i) = 1)^2 \exp(r(t_i, T)) \\ &= \frac{2^{n-1}}{n!} \times \left( P(N(T) > 0 \mid N(0) = 1) \exp(r(0, T)) \right)^2 \\ &\quad \times \prod_{i=2}^{n-1} b(t_i) \times P(N(T) > 0 \mid N(t_i) = 1)^2 \exp(r(t_i, T)). \end{aligned} \quad (2)$$

For a detailed description of this substitution, see [Höhna \(2015\)](#). Additional information regarding the underlying birth-death process can be found in ([Thompson 1975](#); Equation 3.4.6) and [Nee et al. \(1994\)](#) for constant rates and [Höhna \(2013; 2014; 2015\)](#) for arbitrary rate functions.

To compute the equation above we need to know the rate function,  $r(t, s) = \int_t^s d(x) - b(x)dx$ , and the probability of survival,  $P(N(T) > 0 \mid N(t) = 1)$ . [Yule \(1925\)](#) and later [Kendall \(1948\)](#) derived the probability that a process survives ( $N(T) > 0$ ) and the probability of obtaining exactly  $n$  species at time  $T$  ( $N(T) = n$ ) when the process started at time  $t$  with one species. Kendall's results were summarized in Equation (3) and Equation (24) in [Nee et al. \(1994\)](#)

$$P(N(T) > 0 \mid N(t) = 1) = \left( 1 + \int_t^T \left( \mu(s) \exp(r(t, s)) \right) ds \right)^{-1} \quad (3)$$

$$\begin{aligned} P(N(T) = n \mid N(t) = 1) &= (1 - P(N(T) > 0 \mid N(t) = 1) \exp(r(t, T)))^{n-1} \\ &\quad \times P(N(T) > 0 \mid N(t) = 1)^2 \exp(r(t, T)) \end{aligned} \quad (4)$$

An overview for different diversification models is given in [Höhna \(2015\)](#).

*Sidebar: Phylogenetic trees as observations*

The branching processes used here describe probability distributions on phylogenetic trees. This probability distribution can be used to infer diversification rates given an “observed” phylogenetic tree. In reality we never observe a phylogenetic tree itself. Instead, phylogenetic trees themselves are estimated from actual observations, such as DNA sequences. These phylogenetic tree estimates, especially the divergence times, can have considerable uncertainty associated with them. Thus, the correct approach for estimating diversification rates is to include the uncertainty in the phylogeny by, for example, jointly estimating the phylogeny and diversification rates. For the simplicity of the following tutorials, we take a shortcut and assume that we know the phylogeny without error. For publication quality analysis you should always estimate the diversification rates jointly with the phylogeny and divergence times.

### 3 Estimating Character-Dependent Speciation & Extinction Rates

#### 3.1 Outline

This tutorial describes how to specify a character-dependent branching-process models in **RevBayes**; a birth-death process where diversification rates are dependent on the state of a discrete character (Maddison et al. 2007; FitzJohn 2012; Beaulieu and O’Meara 2016). The probabilistic graphical model is given for this tutorial. Finally, you will estimate character-dependent speciation and extinction rates using Markov chain Monte Carlo (MCMC).

#### 3.2 Requirements

We assume that you have read and hopefully completed the following tutorials:

- [Getting started](#)
- [Rev basics](#)
- [Basic Diversification Rate Estimation](#)

Note that the [Rev basics tutorial](#) introduces the basic syntax of **Rev** but does not cover any phylogenetic models. You may skip the [Rev basics tutorial](#) if you have some familiarity with **R**. We tried to keep this tutorial very basic and introduce all the language concepts and theory on the way. You may only need the [Rev basics tutorial](#) for a more in-depth discussion of concepts in **Rev**.

### 4 Data and files

We provide the data file(s) which we will use in this tutorial. You may want to use your own data instead. In the **data** folder, you will find the following files

- [primates\\_tree.nex](#): Dated primates phylogeny including 233 out of 367 species from [Magnuson-Ford and Otto \(2012\)](#).
- [primates\\_habitat.nex](#): Habitat data from [Magnuson-Ford and Otto \(2012\)](#). There is also a larger set of [discrete morphological characters](#). The type of characters are described in the file [primates\\_morph\\_description](#)

- Open the tree `data/primates_tree.nex` in FigTree.
- Open the character data file `data/primates_morph.nex` in a text editor.

## 5 The Theory behind Character-dependent diversification rate models

RevBayes implements the multi-state extension of BiSSE, just as implemented in `diversitree`. The main differences are just which priors you use on the parameters and the inference procedure, *i.e.*, the specifics of the MCMC algorithm. Here we will first describe the general theory about the model, borrowing heavily from the supplementary material of [Moore et al. \(2016\)](#). You may want to skip over this section if you are not interested in math behind. Then we will show how to run this analysis in RevBayes.

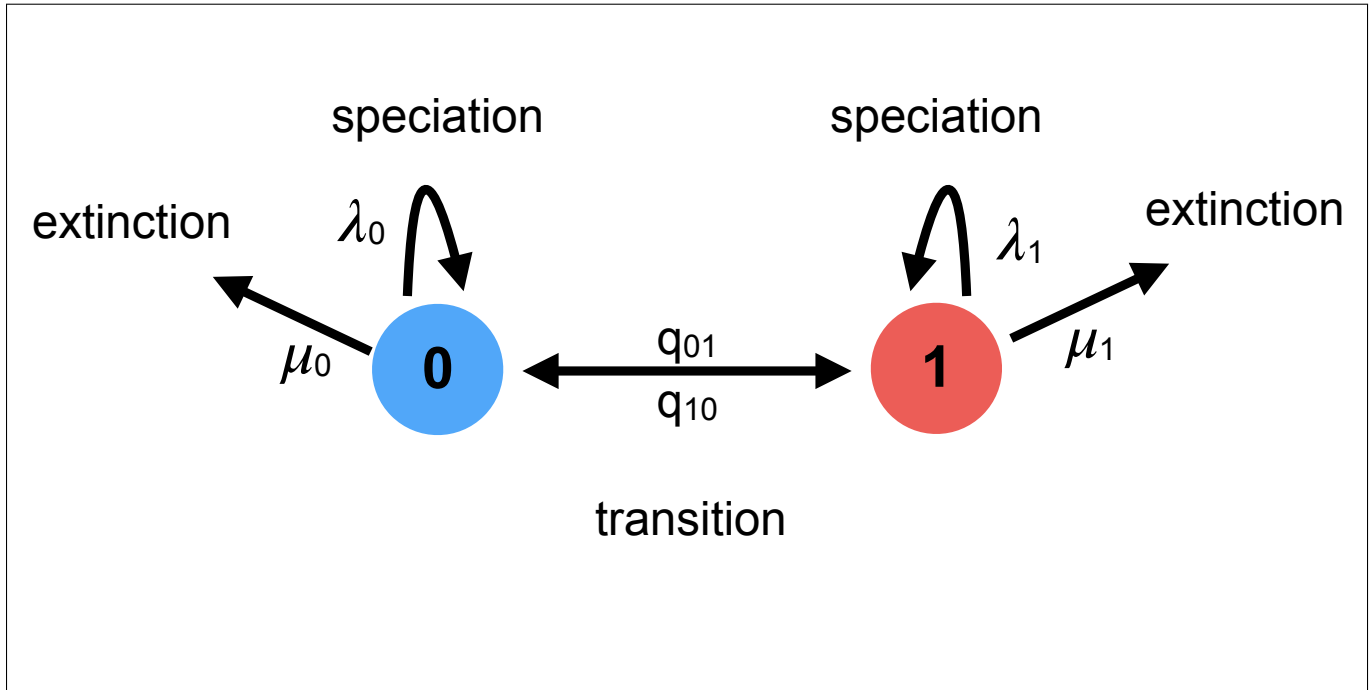


Figure 3: A schematic overview of the BiSSE model. Each lineage has a binary state associated to it and thus can either be in state 0 (blue) or state 1 (red). When a lineage is in state 0 (blue), it can either (a) speciate with rate  $\lambda_0$  which results into two descendant lineage both being in state 0 (blue); (b) it can go extinct with rate  $\mu_0$ ; or (c) it can switch the state to 1 (red) with rate  $q_{01}$ . The same type of events are possible when a lineage is in state 1 (red) but with rate  $\lambda_1$ ,  $\mu_1$ , and  $q_{10}$  respectively.

### 5.1 Likelihood function

We introduce the theory and likelihood function of the BiSSE (Binary State Speciation and Extinction) model developed in [Maddison et al. \(2007\)](#). The BiSSE model assumes a finite number of states (two; one for each of the binary states), where the states of each extant species is known (indeed, it is an observed discrete trait). The general approach adopted by BiSSE models is to derive a set of ordinary differential equations (ODEs) that describe how the probability of observing a descendant clade changes along a branch in the observed phylogeny. Each equation in this set describes how the probability of observing the clade changes through time if it is in a particular state over that time period; collectively, these equations are called  $\frac{dD_{N,i}(t)}{dt}$ , where the subscript  $N$  refers to the descendant clade and the subscript  $i$  refers to  $i^{\text{th}}$  state.

Computing the likelihood proceeds by establishing an initial value problem: in principle, if we know the probabilities of observing a lineage at some specific time (*e.g.*, the present), and know how those probabilities change over time (described by the ODEs), then we can compute the probabilities of observing those lineages at some earlier time (*e.g.*, the root). Assuming that there are exactly  $k$  possible states, we initialize  $k$  probabilities at each tip in the phylogeny; we then compute how each of those  $k$  probabilities changes down each branch in the tree using the above set of  $k$  ODEs. At each node in the tree, we take the product of each of the  $k$  probabilities for the descendants of that node (multiplied by the instantaneous speciation rate for each of the  $k$  states to account for the observed speciation event at the node) as the initial values for the ancestral branch subtending that node. Proceeding in this way down the tree results in a set of  $k$  probabilities at the root; these  $k$  probabilities represent the probability of observing the phylogeny conditional on the root being in each of the states (*i.e.*, the  $i^{\text{th}}$  conditional probability is the probability of observing the tree given that the root is in state  $i$ ). The overall likelihood of the tree is a weighted average of the  $k$  probabilities at the root, where the weighting scheme represents the assumed probability that the root was in each of the  $k$  states.

As with all birth-death process models, special care must be taken to account for the possibility of extinction. Specifically, the above ODEs must accommodate lineages that may arise along each branch in the tree that subsequently go extinct before the present (and so are unobserved). This requires a second set of  $k$  ODEs,  $\frac{dE_i(t)}{dt}$ , which define how the probability of extinction in process  $i$  changes over time. These ODEs must be solved to compute the differential equations  $\frac{dD_{N,i}(t)}{dt}$ , as we will demonstrate when we derive both sets of equations in the following sections.

This framework therefore requires four distinct pieces of information to compute the likelihood of the data:

In the following sections we detail how each of these components is determined for increasingly complex birth-death process models.

### 5.1.1 Binary-state space of diversification processes

Consider a (time-independent) birth-death process with two possible states (*e.g.*, a binary state), with diversification rates  $\{\lambda_0, \mu_0\}$  and  $\{\lambda_1, \mu_1\}$ . We define  $D_{N,0}(t)$  as the probability of observing lineage  $N$  descending from a particular branch at time  $t$ , given that the lineage at that point is in state 0 (with rate parameters  $\lambda_0$ , and  $\mu_0$ ). To compute the probability of observing the lineage at some earlier point,  $D_{N,0}(t + \Delta t)$ , we enumerate all possible events that could occur within the interval  $\Delta t$ . Assuming that  $\Delta t$  is small—so that the probability of any two events occurring in the interval is negligible—there are four possible scenarios within the interval:

1. nothing happens in the interval;
2. the process changes  $0 \rightarrow 1$ ;
3. a speciation event occurs and the left descendant subsequently goes extinct before the present, or;
4. a speciation event occurs and the right descendant subsequently goes extinct before the present.



We can thus compute  $D_{N,0}(t + \Delta t)$  as (see [Maddison et al. \(2007\)](#) and [FitzJohn et al. \(2009\)](#) for a more complete elucidation):

$$\begin{aligned}
 D_{N,0}(t + \Delta t) = & (1 - \mu_0 \Delta t) \times && \text{in all cases, no extinction of the observed lineage} \quad (5) \\
 & [(1 - q \Delta t)(1 - \lambda_0 \Delta t) D_{N,0}(t) && \text{Case (1) nothing happens} \\
 & + q_{01} \Delta t (1 - \lambda_0 \Delta t) D_{N,1}(t) && \text{Case (2) process change but no speciation} \\
 & + (1 - q_{01} \Delta t) \lambda_0 \Delta t E_0(t) D_{N,0}(t) && \text{Case (3) no process change, speciation, extinction} \\
 & + (1 - q \Delta t) \lambda_0 \Delta t E_0(t) D_{N,0}(t)] && \text{Case (4) no process change, speciation, extinction}
 \end{aligned}$$

A matching equation can be written down for  $D_{N,1}(t + \Delta t)$ .

Define  $E_0(t)$  as the probability that a lineage in state 0 at time  $t$  goes extinct before the present. To determine the extinction probability at an earlier point,  $E_0(t + \Delta t)$ , we can again enumerate all the possible events in the interval  $\Delta t$ :

1. the lineage goes extinct within the interval;
2. the lineage neither goes extinct nor speciates, resulting in a single lineage that must eventually go extinct before the present;
3. the lineage neither goes extinct nor speciates, but there is a state change, resulting in a single lineage that must go extinct before the present, or;
4. the lineage speciates in the interval, resulting in *two* lineages that must eventually go extinct before the present.

$$\begin{aligned}
 E_0(t + \Delta t) = & \mu_0 \Delta t + && \text{Case (1) extinction in the interval} \quad (6) \\
 & (1 - \mu_0 \Delta t) \times && \text{no extinction in the interval and ...} \\
 & [(1 - q_{01} \Delta t)(1 - \lambda_0 \Delta t) E_0(t) && \text{Case (2) nothing happens, but subsequent extinction} \\
 & + q_{01} \Delta t (1 - \lambda_0 \Delta t) E_1(t) && \text{Case (3) process change and subsequent extinction} \\
 & + (1 - q \Delta t) \lambda_0 \Delta t E_0(t)^2] && \text{Case (4) speciation and subsequent extinctions}
 \end{aligned}$$

Again, a matching equation  $E_1(t + \Delta t)$  can be written down.

## 5.2 Extension to a multi-state birth-death process

We can expand the BiSSE model to accommodate an arbitrary number of processes,  $k$ , by writing a set of  $k$  difference equations  $D_{N,0}(t + \Delta t), D_{N,1}(t + \Delta t), \dots, D_{N,k}(t + \Delta t)$ . The resulting differential equations are (see [Maddison et al. \(2007\)](#) for the two-process case and [FitzJohn \(2010\)](#) for the  $k$ -process case):

$$\begin{aligned}
 \frac{dD_{N,i}(t)}{dt} = & - \left( \lambda_i + \mu_i + \sum_{j \neq i}^k q \right) D_{N,i}(t) + 2\lambda_i E_i(t) D_{N,i}(t) + \sum_{i \neq j}^k q D_{N,j}(t) \\
 \frac{dE_i(t)}{dt} = & - \left( \lambda_i + \mu_i + \sum_{j \neq i}^k q \right) E_i(t) + \lambda_i E_i(t)^2 + \mu_i + \sum_{i \neq j}^k q E_j(t)
 \end{aligned}$$



Table 1: **BiSSE model parameters and their interpretation**

Parameter	Interpretation
$\Psi$	Phylogenetic tree with divergence times.
$T$	The root age.
$q_{01}$	The rate of shifts from 0 to 1.
$q_{10}$	The rate of shifts from 1 to 0.
$\lambda_0$	Speciation rate for state 0.
$\mu_0$	Extinction rate for state 0.
$\lambda_1$	Speciation rate for state 1.
$\mu_1$	Extinction rate for state 1.

Initial probabilities are assigned according to the observed discrete states: if species  $i$  has state  $j$ , then  $D_{i,j}(0) = 1$  for the observed state (or  $D_{i,j}(0) = \rho_j$  when species sampling is incomplete), and  $D_{i,j}(0) = 0$  for all other ( $\neq j$ ) states. Initial extinction probabilities are set to 0 (since there is no time for extinction to occur at the present) or to  $1 - \rho_i$  if incomplete taxon sampling is used. Root probabilities are either weighted using equal probabilities (uniformly), by a vector of pre-defined root stationary probabilities (informative), or by the stationary distribution of the model, to compute the overall likelihood of the data.

## 6 Character-dependent diversification rates

Now let's start to analyze an example in **RevBayes** using the BiSSE model.

### 6.1 Read the tree

Begin by reading in the observed tree and the morphological data. We have both stored in separate nexus files.

```
observed_phylogeny <- readTrees("data/primates_tree.nex")[1]
data <- readCharacterData("data/primates_solitariness.nex")
```

Note, the character-dependent birth-death process currently uses always the first character/site in the alignment file. We have therefore split the morphological dataset into several small files that include only one morphological character each.

From the tree, we can get some helpful variables:

```
taxa <- observed_phylogeny.taxa()
```

Additionally, we can initialize an iterator variable for our vector of moves and monitors:

```
mvi = 0
mni = 0
```

Finally, we create a helper variable that specifies the number of states that the morphological character has.

```
NUM_STATES = 2
```

Using this variable we can easily change our script to use a different morphological character with a different number of states. We will also use this variable in our second example on hidden-state speciation and extinction model.

## 6.2 Specifying the model

The basic idea behind the model in this example is that speciation and extinction rates are dependent on a binary character (Maddison et al. 2007).

### 6.2.1 Priors on rates

We start by specifying prior distributions on the diversification rates. We will assume here an identical prior distribution on the speciation and extinction rate. Furthermore, we will use a normal distribution as the prior distribution on the log of the speciation and extinction rate. Hence, we will use a mean of  $\frac{\ln(\frac{\#Taxa}{2})}{tree-age}$  which is the expected net-diversification rate.

```
rate_mean <- ln( ln(367.0/2.0) / observed_phylogeny.rootAge() )
rate_sd <- 2.0
```

Now we can specify our character-specific speciation and extinction rate parameters. As we just said before, we are going to use normal distributions for the prior on the log-speciation and log-extinction rate. Here we will use a **for**-loop to specify speciation and extinction parameters for each character, *e.g.*, two in a binary state case.

```
for (i in 1:NUM_STATES) {

  ### Create a lognormal distributed variable for the diversification rate
  log_speciation[i] ~ dnNormal(mean=rate_mean,sd=rate_sd)
  speciation[i] := exp( log_speciation[i] )
  moves[++mvi] = mvSlide(log_speciation[i],delta=0.20,tune=true,weight=3.0)

  ### Create a lognormal distributed variable for the turnover rate
  log_extinction[i] ~ dnNormal(mean=rate_mean,sd=rate_sd)
  extinction[i] := exp( log_extinction[i] )
  moves[++mvi] = mvSlide(log_extinction[i],delta=0.20,tune=true,weight=3.0)

}
```

Great! Now we have create our main variables of interest, the state-specific speciation and extinction rate variables with some meaningful prior distributions. Additionally, we have already created moves on these variables. So we can now turn our focus on the remaining variables of the model.

Next, we specify the transition rates  $q_{01}$  and  $q_{10}$  between the two states 0 and 1. Each transition rate between observed states is drawn from an exponential distribution with a mean of 10 character state transitions over the entire tree. We hope that this is a reasonable prior but also leaves enough uncertainty in our prior knowledge. (You may want to compare the posterior to the prior and/or check the resulting posterior estimates for different choices of the prior!)

```
rate_pr := observed_phylogeny.treeLength() / 10
rate_12 ~ dnExponential(rate_pr)
rate_21 ~ dnExponential(rate_pr)
```

For both rate variable we specify a scaling move.

```
moves[++mvi] = mvScale( rate_12, weight=2 )
moves[++mvi] = mvScale( rate_21, weight=2 )
```

Finally, we build a rate matrix for the relative-rate of change between categories. This is because we need a rate matrix in our state-dependent birth-death process.

```
rate_matrix := fnFreeBinary( [rate_12, rate_21 ], rescaled=false)
```

A specific note here is that we do not rescale the rate matrix. This is very important because otherwise rate matrices, as used for molecular evolution, are always rescaled to have an average rate of 1.0. If such a rescaled rate matrix was used, then you need to provide an overall rate scalar  $\delta$ .

### 6.2.2 Prior on the root state

Create a variable with the prior probabilities of each rate category at the root. We are using a flat Dirichlet distribution as the prior on each state. In this case we are actually estimating the prior frequencies of the root states. There has been some discussion about this in [FitzJohn et al. \(2009\)](#). You could also fix the prior probabilities for the root states to be equal (generally not recommended), or use empirical state frequencies.

```
rate_category_prior ~ dnDirichlet( rep(1,NUM_STATES) )
moves[++mvi] = mvDirichletSimplex(rate_category_prior,tune=true,weight=2)
```

### 6.2.3 Incomplete Taxon Sampling

We know that we have sampled 233 out of 367 living primate species. To account for this we can set the sampling parameter as a constant node with a value of 233/367

```
rho <- observed_phylogeny.ntips()/367
```

### 6.2.4 Root age

The birth-death process requires a parameter for the root age. In this exercise we use a fix tree and thus we know the age of the tree. Hence, we can get the value for the root from the [Magnuson-Ford and Otto \(2012\)](#) tree.

```
root <- observed_phylogeny.rootAge()
```

### 6.2.5 The time tree

Now we have all of the parameters we need to specify the full episodic birth-death model. We initialize the stochastic node representing the time tree.

```
timetree ~ dnCDBDP( rootAge      = root,
                    speciationRates = speciation,
                    extinctionRates = extinction,
                    Q              = rate_matrix,
                    pi              = rate_category_prior,
                    delta           = 1.0,
                    rho             = rho,
                    condition       = "survival",
                    taxa            = taxa )
```

And then we attach data to it.

```
timetree.clamp( observed_phylogeny )
timetree.clampCharData( data )
```

Finally, we create a workspace object of our whole model using the `model()` function.

```
mymodel = model(rate_matrix)
```

The `model()` function traversed all of the connections and found all of the nodes we specified.

## 6.3 Running an MCMC analysis

### 6.3.1 Specifying Monitors

For our MCMC analysis, we set up a vector of *monitors* to record the states of our Markov chain. The first monitor will model all numerical variables, specifically we are interested in speciation rates  $\lambda_0$  and  $\lambda_1$  as well as the extinction rates  $\mu_0$  and  $\mu_1$ .

```
monitors[++mni] = mnModel(filename="output/primates_BiSSE.log", printgen=1)
```

The second monitor is a new type of monitor: an joint-ancestral-states monitor. This monitor takes a draw from joint posterior distribution of the ancestral states. Thus, with this output file we will be able to make a nice plot with ancestral states.

```
monitors[++mni] = mnJointConditionalAncestralState(tree=timetree, cdbdp=timetree, type
  ="Standard", printgen=1, withTips=true, withStartStates=false, filename="output/
  anc_states_primates_BiSSE.log")
```

Finally, we add a screen monitor showing some updates during the MCMC run.

```
monitors[++mni] = mnScreen(printgen=10, rate_12, rate_21, speciation, extinction)
```

### 6.3.2 Initializing and Running the MCMC Simulation

With a fully specified model, a set of monitors, and a set of moves, we can now set up the MCMC algorithm that will sample parameter values in proportion to their posterior probability. The `mcmc()` function will create our MCMC object:

```
mymcmc = mcmc(myModel, monitors, moves)
```

First, we will run a pre-burnin to tune the moves and to obtain starting values from the posterior distribution.

```
mymcmc.burnin(generations=5000,tuningInterval=200)
```

Now, run the MCMC:

```
mymcmc.run(generations=20000)
```

### 6.3.3 Summarizing ancestral states

After our MCMC run has finished, we read-in again our samples from the joint-ancestral-state posterior distribution.

```
anc_states = readAncestralStateTrace("output/anc_states_primates_BiSSE.log")
```

Then we can use this trace and our fixed tree to compute the posterior probabilities of the ancestral states and prepare the output for plotting. We will use the function called **ancestralStateTree** which stores the tree with ancestral states automatically in a file.

```
anc_tree = ancestralStateTree(tree=observed_phylogeny, ancestral_state_trace_vector=
  anc_states, include_start_states=false, file="output/
  anc_states_primates_BiSSE_results.tree", burnin=0, summary_statistic="MAP", site=0)
```

### 6.3.4 Plotting ancestral states

Let us first plot the ancestral states mapped on the phylogeny. We will use R and the package **RevGadgets**. Execute the following code in R.

```
library(RevGadgets)

tree_file = "output/anc_states_primates_BiSSE_results.tree"

plot_ancestral_states(tree_file, summary_statistic="MAP",
  tip_label_size=0,
  xlim_visible=NULL,
  node_label_size=0,
  show_posterior_legend=TRUE,
  node_size_range=c(2, 6),
  alpha=0.75)

output_file = "RevBayes_Anc_States_BiSSE.pdf"
ggsave(output_file, width = 11, height = 9)
```

The resulting plot is shown in Figure 7. We see both the maximum *a posteriori* (MAP) estimate for each node as well as the posterior probability of the states represented by the size of the dots.

### 6.3.5 Plotting diversification rates

Now let us plot the diversification rate estimates. Again, we are going to use R for our plotting. Specifically, we will use the package **ggplot2** but you can also use any other package that you prefer. We are only taking advantage of reading in the tab-delimited file as a table and plot the different diversification rate parameters. Note that we also rely on another provided R script for plotting multiply plots in one file.

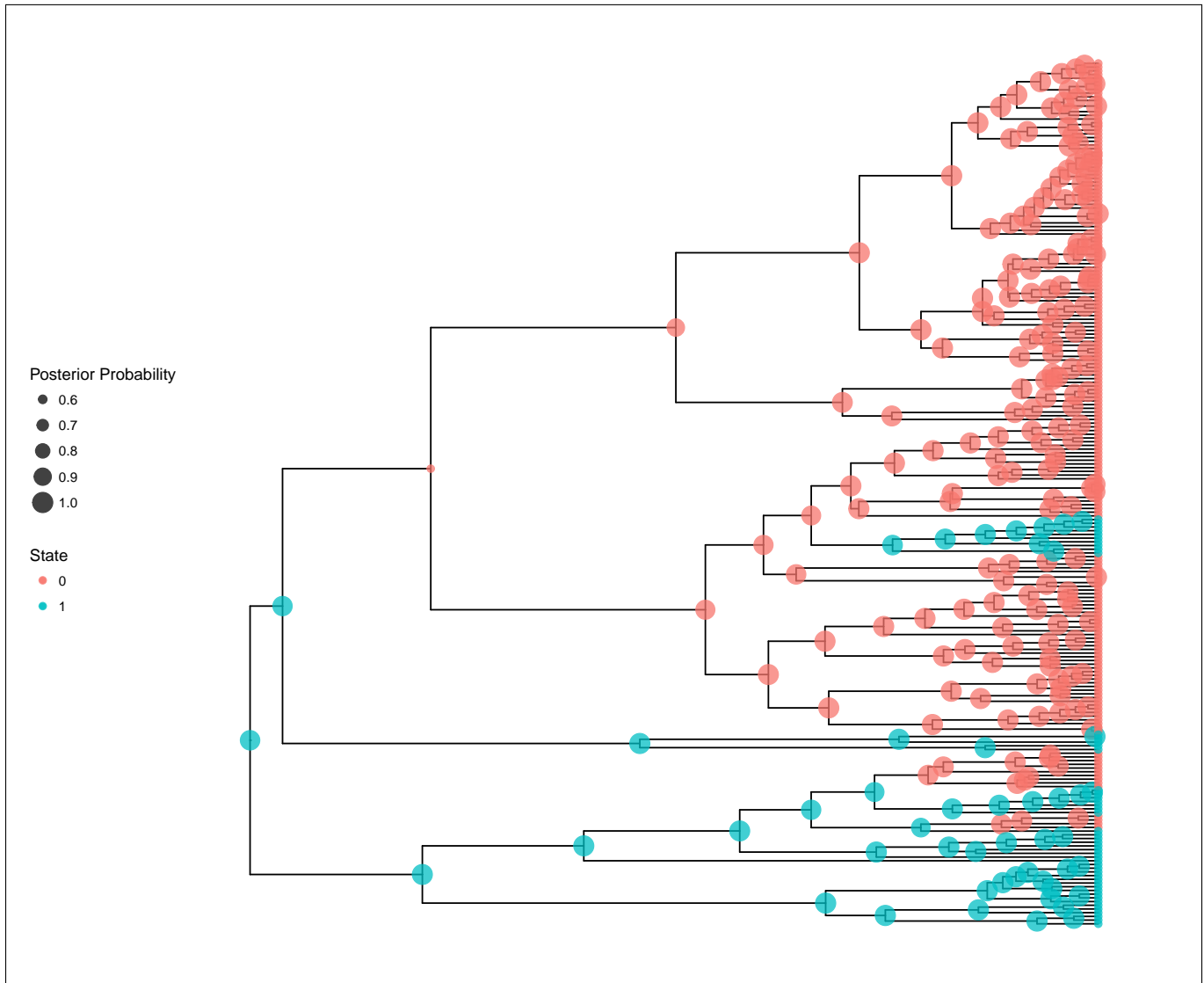


Figure 4: Estimated ancestral states for the activity period of primates.

```
library(ggplot2)
source("scripts/multiplot.R")

data <- read.table("output/primates_BiSSE.log",header=TRUE)

dat_ext <- data.frame(dens = c(data$extinction.1, data$extinction.2), Type = rep(c(
  "1", "2"), each = length(data$extinction.1)))
dat_spec <- data.frame(dens = c(data$speciation.1, data$speciation.2), Type = rep(c(
  "1", "2"), each = length(data$extinction.1)))
dat_div <- data.frame(dens = c(data$speciation.1-data$extinction.1, data$speciation.2-
  data$extinction.2), Type = rep(c("1", "2"), each = length(data$extinction.1)))
dat_rel <- data.frame(dens = c(data$extinction.1/data$speciation.1, data$extinction.2/
  data$speciation.2), Type = rep(c("1", "2"), each = length(data$extinction.1)))
```



Table 2: HiSSE model parameters and their interpretation

Parameter	Interpretation
$\Psi$	Phylogenetic tree with divergence times.
$T$	The root age.
$q_{01}$	The rate of shifts from 0 to 1.
$q_{10}$	The rate of shifts from 1 to 0.
$\lambda_0$	Speciation rate for state 0.
$\mu_0$	Extinction rate for state 0.
$\lambda_1$	Speciation rate for state 1.
$\mu_1$	Extinction rate for state 1.

```
pdf("RevBayes_BiSSE_Results.pdf")

p1 <- ggplot(dat_spec, aes(x = dens, fill = Type)) + labs(title = "Speciation", x="
  Rate", y="Posterior Density") + geom_density(alpha = 0.5)
p2 <- ggplot(dat_ext, aes(x = dens, fill = Type)) + labs(title = "Extinction", x="Rate
  ", y="Posterior Density") + geom_density(alpha = 0.5)
p3 <- ggplot(dat_div, aes(x = dens, fill = Type)) + labs(title = "Net-Diversification
  ", x="Rate", y="Posterior Density") + geom_density(alpha = 0.5)
p4 <- ggplot(dat_rel, aes(x = dens, fill = Type)) + labs(title = "Relative Extinction
  ", x="Rate", y="Posterior Density") + geom_density(alpha = 0.5)

multiplot(p1, p2, p3, p4)
dev.off()
```

## 6.4 Exercise

1. Run an MCMC simulation to estimate the posterior distribution of the speciation rate and extinction rate.
2. Visualize the state-specific diversification rates using R.
3. Do you see evidence for rate differences between the two states?
4. Repeat this analysis for a different binary morphological character.

## 7 Accommodating uncorrelated diversification rate changes

RevBayes implements the multi-state extension of BiSSE, just as implemented in `diversitree`. The main differences are just which priors you use on the parameters and the inference procedure, *i.e.*, the specifics of the MCMC algorithm. Here we will first describe the general theory about the model, borrowing heavily from the supplementary material of [Moore et al. \(2016\)](#). You may want to skip over this section if you are not interested in math behind. Then we will show how to run this analysis in RevBayes.

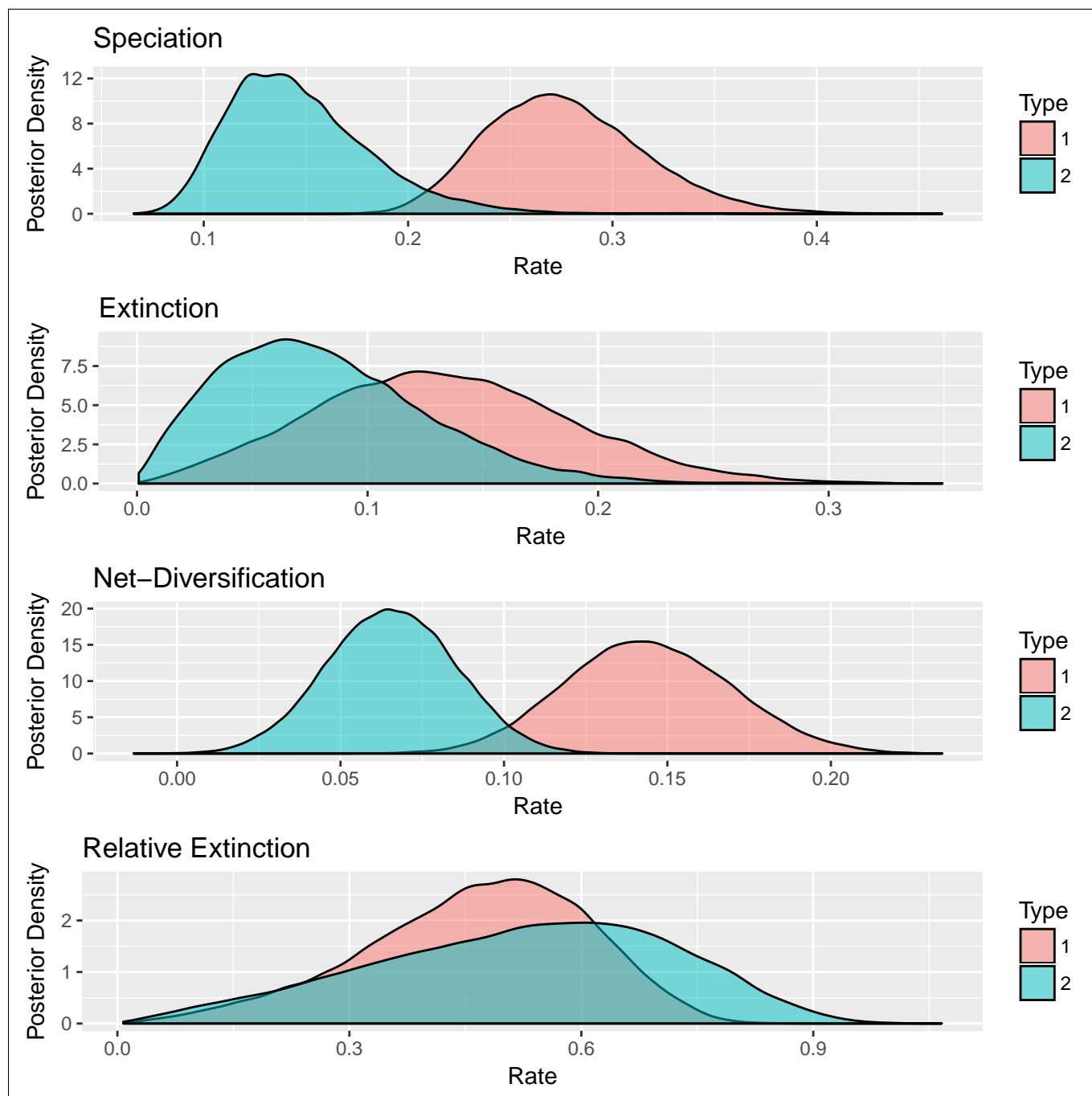


Figure 5: Estimated diversification rate for activity period (state 1 = Diurnal and state 2 = Nocturnal). We see that there is a nice difference in the estimated speciation rates but only little difference in the estimated

## 7.1 Read the tree

Begin by reading in the observed tree and the morphological data. We have both stored in separate nexus files.

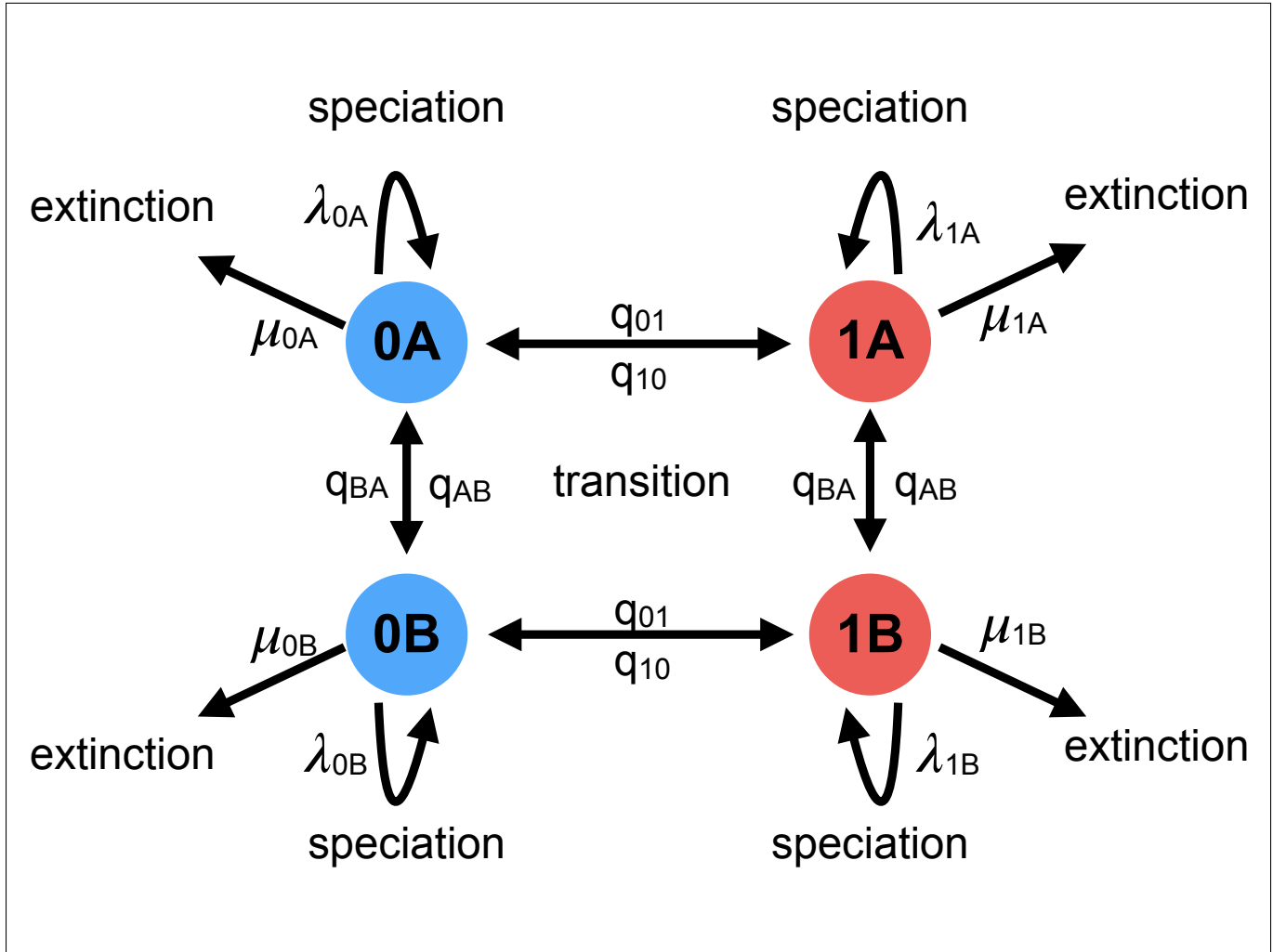


Figure 6: A schematic overview of the HiSSE model. Each lineage has a binary state associated to it and thus can either be in state 0 (blue) or state 1 (red). When a lineage is in state 0 (blue), it can either (a) speciate with rate  $\lambda_0$  which results into two descendant lineage both being in state 0 (blue); (b) it can go extinct with rate  $\mu_0$ ; or (c) it can switch the state to 1 (red) with rate  $q_{01}$ . The same type of events are possible when a lineage is in state 1 (red) but with rate  $\lambda_1$ ,  $\mu_1$ , and  $q_{10}$  respectively.

```
observed_phylogeny <- readTrees("data/primates_tree.nex")[1]
data <- readCharacterData("data/primates_solitariness.nex")
```

Note, the character-dependent birth-death process currently uses always the first character/site in the alignment file. We have therefore split the morphological dataset into several small files that include only one morphological character each.

From the tree, we can get some helpful variables:

```
taxa <- observed_phylogeny.taxa()
```

Additionally, we can initialize an iterator variable for our vector of moves and monitors:

```
mvi = 0
mni = 0
```

Finally, we create a helper variable that specifies the number of states that the morphological character has.

```
NUM_STATES = 2
NUM_HIDDEN = 2
NUM_RATES = NUM_STATES * NUM_HIDDEN
```

Using this variable we can easily change our script to use a different morphological character with a different number of states. We will also use this variable in our second example on hidden-state speciation and extinction model.

### 7.1.1 Priors on rates

We start by specifying prior distributions on the diversification rates. We will assume here an identical prior distribution on the speciation and extinction rate. Furthermore, we will use a normal distribution as the prior distribution on the log of the speciation and extinction rate. Hence, we will use a mean of  $\frac{\ln(\frac{\#Taxa}{2})}{tree-age}$  which is the expected net-diversification rate.

```
rate_mean <- ln( ln(367.0/2.0) / observed_phylogeny.rootAge() )
rate_sd <- 2.0
```

Now we can specify our character-specific specification and extinction rate parameters. As we just said before, we are going to use normal distributions for the prior on the log-speciation and log-extinction rate. Here we will use a **for**-loop to specify speciation and extinction parameters for each character, *e.g.*, two in a binary state case.

```
for (i in 1:NUM_STATES) {

  ### Create a lognormal distributed variable for the diversification rate
  log_speciation[i] ~ dnNormal(mean=rate_mean,sd=rate_sd)
  speciation[i] := exp( log_speciation[i] )
  moves[++mvi] = mvSlide(log_speciation[i],delta=0.20,tune=true,weight=3.0)

  ### Create a lognormal distributed variable for the turnover rate
  log_extinction[i] ~ dnNormal(mean=rate_mean,sd=rate_sd)
  extinction[i] := exp( log_extinction[i] )
  moves[++mvi] = mvSlide(log_extinction[i],delta=0.20,tune=true,weight=3.0)
```

```
}
```

Now we need to create the variable for the hidden states.

```
for (i in 1:(NUM_HIDDEN-1)) {

  ### Create an exponential distributed variable for the diversification rate
  speciation_beta[i] ~ dnExp(1.0)
  moves[++mvi] = mvScale(speciation_beta[i],lambda=0.20,tune=true,weight=2.0)

  ### Create an normal distributed variable for the turnover rate
  extinction_beta[i] ~ dnNormal(0.0,1.0)
  moves[++mvi] = mvSlide(extinction_beta[i],delta=0.20,tune=true,weight=2.0)

}
```

Finally, we match the rates to all possible —hidden and observed— states.

```
for (j in 1:NUM_HIDDEN) {
  for (i in 1:NUM_STATES) {
    if ( j == 1) {
      speciation[i] := exp( speciation_alpha[i] )
      extinction[i] := exp( extinction_alpha[i] )
    } else {
      index = i+(j*NUM_STATES)-NUM_STATES
      speciation[index] := speciation[index-NUM_STATES] * exp( speciation_beta[j-1] )
      extinction[index] := exp( extinction_alpha[i] + extinction_beta[j-1] )
    }
  }
}
```

As before, we specify a rate prior.

```
rate_pr := observed_phylogeny.treeLength() / 10
rate_12 ~ dnExponential(rate_pr)
rate_21 ~ dnExponential(rate_pr)
```

For both rate variable we specify a scaling move.

```
moves[+mvi] = mvScale( rate_12, weight=2 )
moves[+mvi] = mvScale( rate_21, weight=2 )
```

Finally, we build a rate matrix for the relative-rate of change between categories. This is because we need a rate matrix in our state-dependent birth-death process.

```
Q := [ rate_12, rate_21 ]
```

Set up the transition rate matrix for hidden states. We assume the transitions among the hidden states are all equal and drawn from an exponential distribution.

```
hidden_rate ~ dnExponential(rate_pr)
moves[+mvi] = mvScale(hidden_rate, lambda=0.2, tune=true, weight=5)

for (i in 1:(NUM_HIDDEN * (NUM_HIDDEN - 1))) {
  R[i] := hidden_rate
}
```

Create the rate matrix for the combined observed and hidden states

```
rate_matrix := fnHiddenStateRateMatrix(Q, R, rescaled=false)
```

A specific note here is that we do not rescale the rate matrix. This is very important because otherwise rate matrices, as used for molecular evolution, are always rescaled to have an average rate of 1.0. If such a rescaled rate matrix was used, then you need to provide an overall rate scalar  $\delta$ .

### 7.1.2 Prior on the root state

Create a variable with the prior probabilities of each rate category at the root. We are using a flat Dirichlet distribution as the prior on each state. In this case we are actually estimating the prior frequencies of the root states. There has been some discussion about this in [FitzJohn et al. \(2009\)](#). You could also fix the prior probabilities for the root states to be equal (generally not recommended), or use empirical state frequencies.

```
rate_category_prior ~ dnDirichlet( rep(1, NUM_STATES) )
moves[+mvi] = mvDirichletSimplex(rate_category_prior, tune=true, weight=2)
```

### 7.1.3 Incomplete Taxon Sampling

We know that we have sampled 233 out of 367 living primate species. To account for this we can set the sampling parameter as a constant node with a value of 233/367

```
rho <- observed_phylogeny.ntips()/367
```

#### 7.1.4 Root age

The birth-death process requires a parameter for the root age. In this exercise we use a fix tree and thus we know the age of the tree. Hence, we can get the value for the root from the [Magnuson-Ford and Otto \(2012\)](#) tree.

```
root <- observed_phylogeny.rootAge()
```

#### 7.1.5 The time tree

Now we have all of the parameters we need to specify the full episodic birth-death model. We initialize the stochastic node representing the time tree.

```
timetree ~ dnCDBDP( rootAge      = root,
                    speciationRates = speciation,
                    extinctionRates = extinction,
                    Q              = rate_matrix,
                    pi              = rate_category_prior,
                    delta           = 1.0,
                    rho             = rho,
                    condition       = "survival",
                    taxa            = taxa )
```

And then we attach data to it.

```
timetree.clamp( observed_phylogeny )
timetree.clampCharData( data )
```

Finally, we create a workspace object of our whole model using the `model()` function.

```
mymodel = model(rate_matrix)
```

The `model()` function traversed all of the connections and found all of the nodes we specified.

## 7.2 Running an MCMC analysis



### 7.2.1 Specifying Monitors

For our MCMC analysis, we set up a vector of *monitors* to record the states of our Markov chain. For more details

```
monitors[++mni] = mnModel(filename="output/primates_HiSSE.log", printgen=1)
monitors[++mni] = mnJointConditionalAncestralState(tree=timetree, cdbdp=timetree, type
  ="NaturalNumbers", printgen=1, withTips=true, withStartStates=false, filename="
  output/anc_states_primates_HiSSE.log")
monitors[++mni] = mnScreen(printgen=100, Q, R)
```

### 7.2.2 Initializing and Running the MCMC Simulation

With a fully specified model, a set of monitors, and a set of moves, we can now set up the MCMC algorithm that will sample parameter values in proportion to their posterior probability. The `mcmc()` function will create our MCMC object:

```
mymcmc = mcmc(myModel, monitors, moves)
```

First, we will run a pre-burnin to tune the moves and to obtain starting values from the posterior distribution.

```
mymcmc.burnin(generations=5000,tuningInterval=200)
```

Now, run the MCMC:

```
mymcmc.run(generations=20000)
```

### 7.2.3 Summarizing ancestral states

After the MCMC run we summarize and estimate the joint-ancestral-state estimates.

```
anc_states = readAncestralStateTrace("output/anc_states_primates_HiSSE.log")
anc_tree = ancestralStateTree(tree=observed_phylogeny, ancestral_state_trace_vector=
  anc_states, include_start_states=false, file="output/
  anc_states_primates_HiSSE_results.tree", burnin=0, summary_statistic="MAP", site=0)
```

### 7.2.4 Plotting diversification rates

Again, we plot the diversification rate as before.

```

library(ggplot2)
source("scripts/multiplot.R")

data <- read.table("output/primates_HiSSE.log",header=TRUE)

start <- round(0.5*length(data$extinction.1))
end <- length(data$extinction.1)

HiSSE_types <- rep(c("1A", "2A", "1B", "2B"), each = length(data$extinction.1[start:
  end]))
dat_ext <- data.frame(dens = c(data$extinction.1[start:end], data$extinction.2[start:
  end], data$extinction.3[start:end], data$extinction.4[start:end]), Type =
  HiSSE_types)
dat_spec <- data.frame(dens = c(data$speciation.1[start:end], data$speciation.2[start:
  end], data$speciation.3[start:end], data$speciation.4[start:end]), Type =
  HiSSE_types)
dat_div <- data.frame(dens = c(data$speciation.1[start:end]-data$extinction.1[start:
  end], data$speciation.2[start:end]-data$extinction.2[start:end], data$speciation.3[
  start:end]-data$extinction.3[start:end], data$speciation.4[start:end]-
  data$extinction.4[start:end]), Type = HiSSE_types)
dat_rel <- data.frame(dens = c(data$extinction.1[start:end]/data$speciation.1[start:
  end], data$extinction.2[start:end]/data$speciation.2[start:end], data$extinction.3[
  start:end]/data$speciation.3[start:end], data$extinction.4[start:end]/
  data$speciation.4[start:end]), Type = HiSSE_types)

pdf("RevBayes_HiSSE_Results.pdf")

p1 <- ggplot(dat_spec, aes(x = dens, fill = Type)) + labs(title = "Speciation", x="
  Rate", y="Posterior Density") + geom_density(alpha = 0.5) + guides(fill=
  guide_legend(ncol=2,byrow=TRUE))
p2 <- ggplot(dat_ext, aes(x = dens, fill = Type)) + labs(title = "Extinction", x="Rate
  ", y="Posterior Density") + geom_density(alpha = 0.5) + guides(fill=guide_legend(
  ncol=2,byrow=TRUE))
p3 <- ggplot(dat_div, aes(x = dens, fill = Type)) + labs(title = "Net-Diversification
  ", x="Rate", y="Posterior Density") + geom_density(alpha = 0.5) + guides(fill=
  guide_legend(ncol=2,byrow=TRUE))
p4 <- ggplot(dat_rel, aes(x = dens, fill = Type)) + labs(title = "Relative Extinction
  ", x="Rate", y="Posterior Density") + geom_density(alpha = 0.5) + guides(fill=
  guide_legend(ncol=2,byrow=TRUE))

multiplot(p1, p2, p3, p4)
dev.off()

```

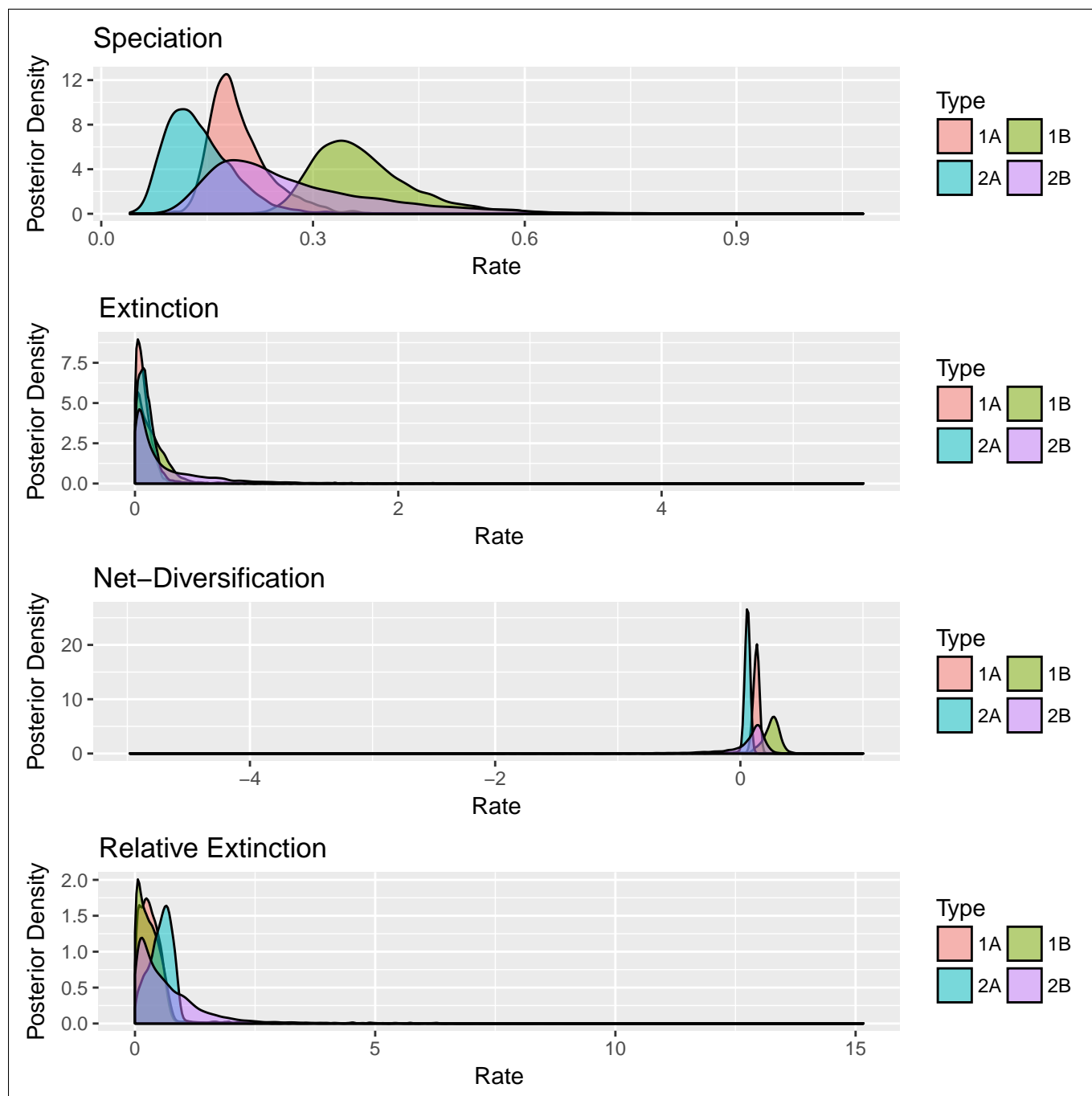


Figure 7: Estimated diversification rate for activity period (state 1 = Diurnal and state 2 = Nocturnal). We see that there is a nice difference in the estimated speciation rates but only little difference in the estimated

### 7.3 Exercise

1. Run an MCMC simulation to estimate the posterior distribution of the speciation rate and extinction rate.
2. Visualize the state-specific diversification rates using R.

3. Do you see evidence for rate differences between the two states?
4. Do you see differences to the previous BiSSE estimates?
5. Repeat this analysis for a different binary morphological character.

## References

- Alfaro, M., F. Santini, C. Brock, H. Alamillo, A. Dornburg, D. Rabosky, G. Carnevale, and L. Harmon. 2009. Nine exceptional radiations plus high turnover explain species diversity in jawed vertebrates. *Proceedings of the National Academy of Sciences* 106:13410–13414.
- Beaulieu, J. M. and B. C. O'Meara. 2016. Detecting hidden diversification shifts in models of trait-dependent speciation and extinction. *Systematic Biology* 65:583–601.
- Condamine, F. L., J. Rolland, and H. Morlon. 2013. Macroevolutionary perspectives to environmental change. *Ecology Letters* .
- Drummond, A., M. Suchard, D. Xie, and A. Rambaut. 2012. Bayesian phylogenetics with beauti and the beast 1.7. *Molecular Biology and Evolution* 29:1969–1973.
- FitzJohn, R., W. Maddison, and S. Otto. 2009. Estimating trait-dependent speciation and extinction rates from incompletely resolved phylogenies. *Systematic Biology* 58:595–611.
- FitzJohn, R. G. 2010. Quantitative traits and diversification. *Systematic Biology* 59:619–633.
- FitzJohn, R. G. 2012. Diversitree: comparative phylogenetic analyses of diversification in R. *Methods in Ecology and Evolution* 3:1084–1092.
- Heath, T. A., J. P. Huelsenbeck, and T. Stadler. 2014. The fossilized birth-death process for coherent calibration of divergence-time estimates. *Proceedings of the National Academy of Sciences* 111:E2957–E2966.
- Höhna, S. 2013. Fast simulation of reconstructed phylogenies under global time-dependent birth-death processes. *Bioinformatics* 29:1367–1374.
- Höhna, S. 2014. Likelihood Inference of Non-Constant Diversification Rates with Incomplete Taxon Sampling. *PLoS One* 9:e84184.
- Höhna, S. 2015. The time-dependent reconstructed evolutionary process with a key-role for mass-extinction events. *Journal of Theoretical Biology* 380:321–331.
- Höhna, S., M. J. Landis, T. A. Heath, B. Boussau, N. Lartillot, B. R. Moore, J. P. Huelsenbeck, and F. Ronquist. 2016. RevBayes: Bayesian phylogenetic inference using graphical models and an interactive model-specification language. *Systematic Biology* 65:726–736.
- Kendall, D. G. 1948. On the generalized "birth-and-death" process. *The Annals of Mathematical Statistics* 19:1–15.
- Maddison, W., P. Midford, and S. Otto. 2007. Estimating a binary character's effect on speciation and extinction. *Systematic Biology* 56:701.

- Magnuson-Ford, K. and S. P. Otto. 2012. Linking the investigations of character evolution and species diversification. *The American Naturalist* 180:225–245.
- May, M. R., S. Höhna, and B. R. Moore. 2016. A Bayesian Approach for Detecting the Impact of Mass-Extinction Events on Molecular Phylogenies When Rates of Lineage Diversification May Vary. *Methods in Ecology and Evolution* 7:947–959.
- Moore, B. R., S. Höhna, M. R. May, B. Rannala, and J. P. Huelsenbeck. 2016. Critically evaluating the theory and performance of Bayesian analysis of macroevolutionary mixtures. *Proceedings of the National Academy of Sciences* 113:9569–9574.
- Nee, S., R. M. May, and P. H. Harvey. 1994. The Reconstructed Evolutionary Process. *Philosophical Transactions: Biological Sciences* 344:305–311.
- Rabosky, D. L. 2014. Automatic detection of key innovations, rate shifts, and diversity-dependence on phylogenetic trees. *PLoS One* 9:e89543.
- Ronquist, F., M. Teslenko, P. van der Mark, D. L. Ayres, A. Darling, S. Höhna, B. Larget, L. Liu, M. A. Suchard, and J. P. Huelsenbeck. 2012. Mrbayes 3.2: efficient bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology* 61:539–542.
- Stadler, T. 2011. Mammalian phylogeny reveals recent diversification rate shifts. *Proceedings of the National Academy of Sciences* 108:6187–6192.
- Thompson, E. 1975. *Human evolutionary trees*. Cambridge University Press Cambridge.
- Yule, G. 1925. A mathematical theory of evolution, based on the conclusions of dr. jc willis, frs. *Philosophical Transactions of the Royal Society of London. Series B, Containing Papers of a Biological Character* 213:21–87.

Version dated: March 14, 2017