Phylogenetic Inference using RevBayes

Historical biogeography

Michael Landis

Introduction

How did species come to live where they're found today? To answer this, we can leverage phylogenetic, molecular, and geographical information to model species distributions as the outcome of biogeographic processes. How to best model these processes requires special consideration, such as how ranges are inherited following speciation events, how geological events might influence dispersal rates, and what factors affect rates of dispersal and extirpation. A major technical challenge of modeling range evolution is how to translate these natural processes into stochastic processes that remain tractable for inference. This tutorial provides a brief background in some of these models, then describes how to perform Bayesian inference of historical biogeography using `RevBayes`.

1 Overview of the Dispersal-Extinction-Cladogenesis model

Discrete range characters

The Dispersal-Extinction-Cladogenesis (DEC) models range evolution as a discrete-valued process (Ree et al. 2005; Ree and Smith 2008). DEC interprets taxon ranges as presence-absence data, that is, where a species is observed or not observed across multiple discrete areas. For example, say there are three areas, A, B, and C. If a species is present in areas A and C, then its range equals AC, which can also be encoded into the length-3 bit vector, 101. Bit vectors may also be transformed into (decimal) integers, *e.g.,* the binary number 101 equals the decimal number 5.

Range	Bits	Size	State
∅	000	0	0
A	100	1	1
B	010	1	2
C	001	1	3
AB	110	2	4
AC	101	2	5
BC	011	2	6
ABC	111	3	7

Table 1: Example of discrete range representations for an analysis with areas A, B, and C.

Decimal representation is rarely used in discussion, but it is useful to keep in mind when considering the total number of possible ranges for a species and when processing output. Note, `RevBayes` assigns integers to ranges in order according to size.

Anagenetic range evolution

Anagenesis refers to range evolution that occurs between speciation events within lineages. Because DEC uses discrete-valued ranges, anagenesis is modeled using a continuous-time Markov chain. This, in turn, allows us to compute transition probability of a character changing from i to j in time t through matrix exponentiation

$$\mathbf{P}_{ij}(t) = \left[\exp\left\{\mathbf{Q}t\right\}\right]_{ij},$$

where $\mathbf{Q}$ is the instantaneous rate matrix defining the rates of change between all pairs of characters, and $\mathbf{P}$ is the transition probability rate matrix. The indices i and j represent different ranges, each of which is encoded as the set of areas occupied by the species. The probability has integrated over all possible scenarios of character transitions that could occur during t so long as the chain begins in state i and ends in state j.

We can then encode $\mathbf{Q}$ to reflect the allowable classes of range evolution events with biologically meaningful parameters. We'll take a simple model of range expansion (e.g. $BC \rightarrow ABC$) and range contraction (e.g. $BC \rightarrow C$). Range expansion may also be referred to as dispersal or area gain and range contraction as extirpation, (local) extinction, or area loss. The rates in the transition matrix for three areas might appear as

$$\mathbf{Q} = \begin{array}{c|cccccccc} & \emptyset & A & B & C & AB & AC & BC & ABC \\ \hline \emptyset & - & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ A & e_A & - & 0 & 0 & d_{AB} & d_{AC} & 0 & 0 \\ B & e_B & 0 & - & 0 & d_{BA} & 0 & d_{BC} & 0 \\ C & e_C & 0 & 0 & - & 0 & d_{CA} & d_{CB} & 0 \\ AB & 0 & e_A & e_B & 0 & - & 0 & 0 & d_{AC} + d_{BC} \\ AC & 0 & e_C & 0 & e_A & 0 & - & 0 & d_{AB} + d_{CB} \\ BC & 0 & 0 & e_C & e_B & 0 & 0 & - & d_{BA} + d_{CA} \\ ABC & 0 & 0 & 0 & 0 & e_C & e_B & e_A & - \end{array},$$

where $e = (e_A, e_B, e_C)$ are the (local) extinction rates per area, and $d = (d_{AB}, d_{AC}, d_{BC}, d_{BA}, d_{CA}, d_{CB})$ are the dispersal rates between areas. Notice that the sum of rates leaving state $\emptyset$ is zero, meaning any species that loses all areas in its range remains permanently extinct.

To build our intuition, let's build a DEC rate matrix. Assume you have three areas

```
n_areas <- 3
```

First, create a matrix of dispersal rates between area pairs, with rates $d_{AB} = d_{AC} = \ldots = d_{CB} = 1$

```
for (i in 1:n_areas) {
    for (j in 1:n_areas) {
        dr[i][j] <- abs(1)
    }
}
```

Next, let's create the extirpation rates with values $e_A = e_B = e_C = 1$

```
for (i in 1:n_areas) {
    for (j in 1:n_areas) {
        er[i][j] <- abs(0)
    }
    er[i][i] <- abs(1)
}
```

When the only non-zero extirpation rates are on the diagonal of the matrix, extirpation rates are independent of what other areas the taxon is occupies. More complex models that penalize widespread ranges spanning disconnected areas are explored in later sections.

Now, create the DEC rate matrix

```
Q_DEC := fnDECRateMatrix(dispersalRates=dr, extirpationRates=er)
Q_DEC
[ [ 0.0000, 0.0000, 0.0000, 0.0000, 0.0000, 0.0000, 0.0000, 0.0000 ] ,
    1.0000, -3.0000, 0.0000, 0.0000, 1.0000, 1.0000, 0.0000, 0.0000 ] ,
    1.0000, 0.0000, -3.0000, 0.0000, 1.0000, 0.0000, 1.0000, 0.0000 ] ,
    1.0000, 0.0000, 0.0000, -3.0000, 0.0000, 1.0000, 1.0000, 0.0000 ] ,
    0.0000, 1.0000, 1.0000, 0.0000, -4.0000, 0.0000, 0.0000, 2.0000 ] ,
    0.0000, 1.0000, 0.0000, 1.0000, 0.0000, -4.0000, 0.0000, 2.0000 ] ,
    0.0000, 0.0000, 1.0000, 1.0000, 0.0000, 0.0000, -4.0000, 2.0000 ] ,
    0.0000, 0.0000, 0.0000, 0.0000, 1.0000, 1.0000, 1.0000, -3.0000 ] ]
```

Show the anagenetic transition probabilities for a branch of length `0.2`

```
tp_DEC <- Q_DEC.getTransitionProbabilities(rate=0.2)
tp_DEC
[ [ 1.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000],
  [ 0.000, 0.673, 0.013, 0.013, 0.123, 0.123, 0.005, 0.050],
  [ 0.000, 0.013, 0.673, 0.013, 0.123, 0.005, 0.123, 0.050],
  [ 0.000, 0.013, 0.013, 0.673, 0.005, 0.123, 0.123, 0.050],
  [ 0.000, 0.107, 0.107, 0.004, 0.502, 0.031, 0.031, 0.218],
  [ 0.000, 0.107, 0.004, 0.107, 0.031, 0.502, 0.031, 0.218],
  [ 0.000, 0.004, 0.107, 0.107, 0.031, 0.031, 0.502, 0.218],
  [ 0.000, 0.021, 0.021, 0.021, 0.107, 0.107, 0.107, 0.616]]
```

Substructures from the rate matrix are reflected in the transition probability matrix. Notice that ranges that are separated by multiple dispersal and extirpation events are the most improbable – e.g. going from A to BC takes three events and has probability 0.005.

By default, the `RevBayes` conditions the anagenetic range evolution process on never entering the null range when computing the transition probabilities. This is crucial so the model can both simulate and infer using identical (correct) probabilities. Massana et al. (2015) first noted that the null range results in abnormal extirpation rate and range size estimates. Their proposed solution to eliminate the null range from the state space is done with the `condition=Exclude` setting. The `nullRange=Include` setting provides no special handling of the null range, and produces the raw probabilities of Ree et al. (2005).

Cladogenetic range evolution

The cladogenetic component of the DEC model describes evolutionary change accompanying speciation events. In the context of range evolution, daughter species do not necessarily inherit their ancestral range in an identical manner. For each internal node in the reconstructed tree, one of several cladogenetic events can occur, some of which are described below.

Beginning with the simplest case first, suppose the range of a species is A the moment before speciation occurs at an internal phylogenetic node. Since the species range is size one, both daughter lineages necessarily inherit the ancestral species range (A). In DEC parlance, this is called a *narrow sympatry* event. Now, suppose the ancestral range is ABC. Under *subset sympatric cladogenesis*, one lineage identically inherits the ancestral species range, ABC, while the other lineage inherits only a single area, i.e. only A or B or C. For *widspread sympatric cladogenesis*, both lineages inherit the ancestral range, ABC. Under *allopatric cladogenesis*, the ancestral range is split evenly among daughter lineages, e.g. one lineage may inherit AB and the other inherits C. Finally, supposing the ancestral range is A, *jump dispersal cladogenesis* results in one daughter lineage inheriting the ancestral range A, and the other daughter lineage inheriting a previously uninhabited area, B or C. For an excellent overview of described state transitions for cladogenetic events, see Matzke (2012).

Make the cladogenetic probability event matrix

```
clado_event_types = [ "s", "a" ]
clado_event_probs <- simplex( 1, 1 )
P_DEC := fnDECCladoProbs(eventProbs=clado_event_probs,
                         eventTypes=clado_event_types,
                         numCharacters=n_areas)
```

`clado_event_types` defines what cladogenetic event types are used. `"a"` and `"s"` indicate allopatry and subset sympatry, as described in (Ree et al. 2005). Other cladogenetic events include jump dispersal (`"j"`; Matzke 2012) and full sympatry (`"f"`; Landis et al. 2013). The cladogenetic event probability matrix will assume that `eventProbs` and `eventTypes` share the same order.

Print the cladogenetic transition probabilities

```
P_DEC
   [
     ( 1 -> 1, 1 ) = 1.0000,
     ( 2 -> 2, 2 ) = 1.0000,
     ( 3 -> 3, 3 ) = 1.0000,
     ...
```

```
    ( 7 -> 7, 1 ) = 0.0833,
    ( 7 -> 7, 2 ) = 0.0833,
    ( 7 -> 7, 3 ) = 0.0833
  ]
```

The cladogenetic probability matrix becomes very sparse for large numbers of areas, so only non-zero values are shown. Each row reports a triplet of states—the ancestral state and the two daughter states—with the probability associated with that event. Since these are proper probabilities, the sum of probabilities for a given ancestral state over all possible cladogenetic outcomes equals one.

Things to consider

The probabilities of anagenetic change along lineages must account for all combinations of starting states and ending states. For 3 areas, there are 8 states, and thus $8 \times 8 = 64$ probability terms for pairs of states. For cladogenetic change, we need transition probabilities for all combinations of states before cladogenesis, after cladogenesis for the left lineage, and after cladogenesis for the right lineage. Like above, for three areas, there are 8 states, and $8 \times 8 \times 8 = 512$ cladogenetic probability terms.

Of course, this model can be specified for more than three areas. Let's consider what happens to the size of $\mathbf{Q}$ when the number of areas, N, becomes large. For three areas, $\mathbf{Q}$ is size 8×8. For ten areas, $\mathbf{Q}$ is size $2^{10} \times 2^{10} = 1024 \times 1024$, which approaches the largest size matrices that can be exponentiated in a practical amount of time. For twenty areas, $\mathbf{Q}$ is size $2^{20} \times 2^{20} \approx 10^6 \times 10^6$ and exponentiation is not viable. Thus, selecting the discrete areas for a DEC analysis should be done with regard to what one hopes to learn through the analysis itself.

Some questions

? For the three-area DEC rate matrix above, what is the rate of leaving state AC in terms of dispersal and extinction parameters?

? What series of transition events might explain a lineage evolving from range ABC to range A? From range AB to range C?

? Imagine a DEC rate matrix with four areas, $ABCD$. What would be the dispersal rate for $Q_{BC,BCD}$? How many states does a DEC rate matrix with four areas have? What is the relationship between the number of areas and the number of states under the DEC model?

? Given the state is AB before cladogenesis, and allowing subset sympatry, widespread sympatry, and allopatry, what are the 7 possible states in the daughter lineages after cladogenesis?

? For three areas, there are three narrow, four widespread, 18 subset sympatric events and 12 allopatric cladogenesis events. What proportion of terms in the cladogenesis matrix are zero?

Recommended tutorials

The **Historical Biogeography** tutorial will assume the reader has completed the following `RevBayes` tutorials

- **`Rev` Basics**
- **Molecular Models of Character Evolution**
- **Running and Diagnosing an MCMC Analysis**
- **Divergence Time Estimation and Node Calibrations**

2 Simple DEC analysis

The following series of tutorials will reconstruct the ancestral ranges of the silversword alliance (Tribe *Madiinae*), a young and diverse clade of about 50 species (Baldwin et al. 1991). Although silverswords are endemic to Hawaii, they are nested within a larger clade alongside tarweeds, which are native to the California coastline. The size and age of the silversword clade, combined with our knowledge of Hawaiian island formation, makes it an ideal system to explore concepts in historical biogeography and phylogeny. For further reading, consult: Carlquist (1959); Baldwin and Sanderson (1998).

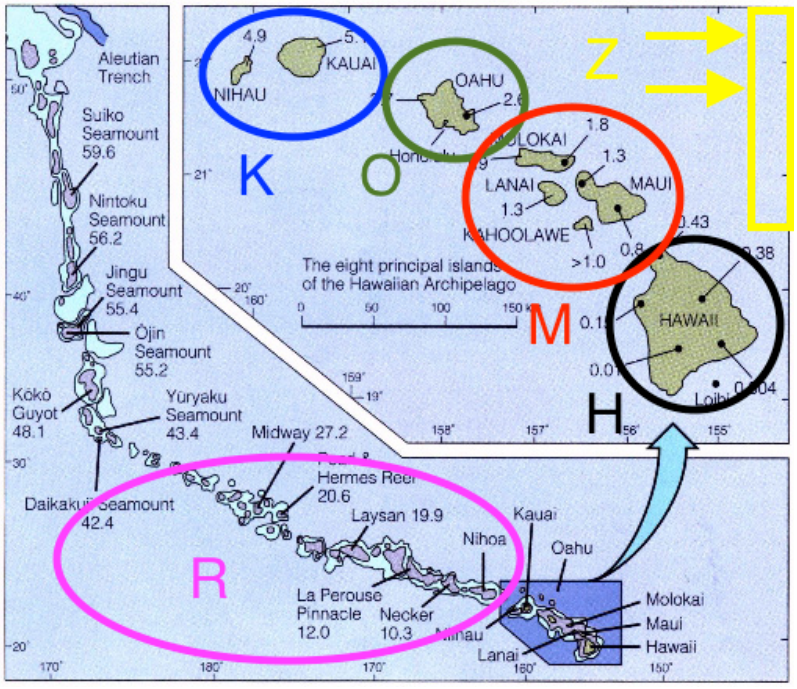

Figure 1: A beautiful figure of the discrete areas for tutorial. Six areas are shown: Kauai and Niihau (K); Oahu (O); Maui-Nui, Lanai, and Molokai (M); Hawaii (H); the remaining Hawaiian islands (R); and the North American mainland (Z).

For this tutorial we'll focus entirely on the silversword alliance and the modern Hawaiian archipelago. To begin, we'll use just four areas, K, O, M, and H, and include areas R and Z in later analyses (Fig 1).

Analysis

First, we'll create some variables to manage files

```
range_fn = "data/n4/silversword.n4.range.nex"
tree_fn = "data/n4/silversword.tre"
out_fn = "output/simple"
```

then read in our character data as binary presence-absence characters

Range	Areas	Size	State
∅	0000	0	0
K	1000	1	1
O	0100	1	2
M	0010	1	3
H	0001	1	4
KO	1100	2	5
KM	1010	2	6
OM	0110	2	7
KH	1001	2	8
OH	0101	2	9
MH	0011	2	10
KOM	1110	3	11
KOH	1101	3	12
KMH	1011	3	13
OMH	0111	3	14
KOMH	1111	4	15

Table 2: Area coding used for four areas: K is Kauai and Nihoa; O is Oahu; M is Maui Nui, Lanai, and Molokai; H is Hawaii island.

```
dat_range_01 = readDiscreteCharacterData(range_fn)
```

then encode the species ranges into natural numbers

```
dat_range_n   = formatDiscreteCharacterData(dat_range_01, "DEC")
```

Record the number of areas

```
n_areas = dat_range_01.nchar()
```

You can view the taxon data to see how characters are coded

```
dat_range_01[1]
 Argyroxiphium_grayanum_East_Maui:
    0010
dat_range_n[1]
 Argyroxiphium_grayanum_East_Maui:
    3
```

For now, we'll assume we know the dated species phylogeny without error.

```
tree <- readTrees(tree_fn)[1]
```

Next, we'll build the anagenetic rate matrix for the DEC model. In its simplest form, this requires a dispersal rate and an extirpation rate. For this analysis, we'll assume that all pairs of areas share the same dispersal rate and all areas share the same extirpation rate. To gain greater control to observe and manage prior sensitivity, we'll reparameterize the DEC rate matrix to report the *relative* rates of dispersal versus extirpation events. One we have the relative rate matrix, we'll scale the *absolute* rate of anagenesis in geological time units with a biogeographic clock parameter, similar to the molecular clock parameter used in dating analyses.

First, create a parameter for the arrival rate for anagenetic range evolution events. We'll apply an uninformative prior to the rate's magnitude by first assigning a uniform distribution to the $\log_{10}$ rate.

```
log10_rate_bg ~ dnUniform(-4,2)
log10_rate_bg.setValue(-2)
moves[1] = mvSlide(log10_rate_bg, weight=4)
```

Create a deterministic function to convert the rate from log-scale to linear-scale.

```
rate_bg := 10^log10_rate_bg
```

This gives us a uniform prior over orders of magnitude, ranging from 10^{-4} to 10^2 events per million years.

Because the rate matrix will describe the relative anagenetic event rates, we will assume that dispersal occurs at the relative (fixed) rate of one.

```
dispersal_rate <- abs(1)
```

then create the dispersal rate matrix

```
for (i in 1:n_areas) {
  for (j in 1:n_areas) {
    dr[i][j] <- dispersal_rate
  }
}
```

Next, assign the prior distribution to the relative extirpation rate and assign it a move. The prior distribution of extirpation rates uses `log_sd` and `log_mean` values to give it the expected value of one – i.e. ranges are expected to lose and gain areas at the same rate under the prior.

```
log_sd <- 0.5
log_mean <- ln(1) - 0.5*log_sd^2
extirpation_rate ~ dnLognormal(mean=log_mean, sd=log_sd)
moves[2] = mvScale(extirpation_rate, weight=2)
```

then create a matrix of extirpation rates

```
for (i in 1:n_areas) {
  for (j in 1:n_areas) {
    er[i][j] <- abs(0)
  }
  er[i][i] := extirpation_rate
}
```

This diagonal matrix results in per-area extirpation rates that are mutually independent. All non-diagonal extirpation rates equal zero. (More on penalized ranges and off-diagonal rates later.) We can now create our relative rate matrix, `Q_DEC`, with the `fnDECRateMatrix` function.

```
Q_DEC := fnDECRateMatrix(dispersalRates=dr, extirpationRates=er)
```

Note, `fnDECRateMatrix` does not rescale its elements in any way, so transition rates share the same time scale as the underlying tree (typically, millions of years). This is in contrast to the standard molecular substitution processes (e.g. `fnGTR`) whose rates are rescaled such that the process is expected to produce one event per site per unit time.

Next, we need to create the cladogenetic probability matrix. Cladogenetic event probabilities are given by a transition probability matrix and do not require a rate matrix. First, we will create a vector of prior weights on cladogenesis events. Here, we assume all cladogenetic events are equiprobable.

```
clado_event_types <- [ "s", "a" ]
clado_event_probs <- simplex(1, 1)
P_DEC := fnDECCladoProbs(eventProbs=clado_event_probs,
                         eventTypes=clado_event_types,
                         numCharacters=n_areas)
```

Finally, all our DEC model components are encapsulated in the `dnPhyloCTMCClado` distribution, which is similar to `dnPhyloCTMC` except specialized to integrate over cladogenetic events. Although this dataset has four areas, it is recognized single character with states valued from 1 to 2^4, hence `nSites=1`.

```
m_bg ~ dnPhyloCTMCClado(tree=tree,
                        Q=Q_DEC,
                        cladoProbs=P_DEC,
                        branchRates=rate_bg,
                        nSites=1,
                        type="NaturalNumbers")
```

The remaining tasks should be familiar from previous tutorials, so we can proceed briskly. Attach the observed ranges to the model.

```
m_bg.clamp(dat_range_n)
```

Compose the model.

```
mymodel = model(m_bg)
```

Add the monitors. (The `mnJointConditionalAncestralState` monitor will be described later.)

```
monitors[1] = mnScreen(rate_bg, extirpation_rate, printgen=100)
monitors[2] = mnModel(file=out_fn+".params.log", printgen=10)
monitors[3] = mnFile(tree, file=out_fn+".tre", printgen=10)
monitors[4] = mnJointConditionalAncestralState(tree=tree,
                                               ctmc=m_bg,
                                               filename=out_fn+".states.log",
                                               type="NaturalNumbers",
                                               printgen=10,
                                               withTips=true,
                                               withStartStates=true)
```

Create the MCMC object, and run the chain after burn-in.

```
mymcmc = mcmc(moves, monitors, mymodel)
mymcmc.run(3000)
```

Results

Example results have around found in `output_example/epoch.*`

The `mnJointConditionalAncestralState` monitor above created a states file. Each row in the states file lists the joint sample of ancestral states conditioned on the tip values for the whole tree. Each column

corresponds to the phylogenetic node index for that particular MCMC sample. The index is used to later correspond the state samples with the tree samples when the topology is a random variable. (More on this in the ancestral state tutorial.)

The script located at `scripts/make_anc_states.Rev` contains code to construct an ancestral state tree. To use it for other analyses, just modify the `out_str` variable below.

Open a new RevBayes session. Set up the files we'll work with.

```
out_str = "output/simple"
out_state_fn = out_str + ".states.log"
out_tree_fn = out_str + ".tre"
out_mcc_fn = out_str + ".mcc.tre"
```

Get the ancestral state trace

```
state_trace = readAncestralStateTrace(file=out_state_fn)
```

Get the ancestral state tree trace. It is important to use `readAncestralTreeTrace` and not `readTreeTrace` to properly annotate the tree with ancestral states.

```
tree_trace = readAncestralStateTreeTrace(file=out_tree_fn, treetype="clock")
```

Read the maximum clade credibility tree and write it to file

```
mcc_tree = mccTree(tree_trace, file=out_mcc_fn)
```

Build the ancestral state tree

```
anc_tree = ancestralStateTree(tree=mcc_tree,
                              ancestral_state_trace_vector=state_trace,
                              tree_trace=tree_trace,
                              include_start_states=true,
                              file=out_str+".ase.tre",
                              burnin=0,
                              summary_statistic="MAP",
                              site=0)
```

We can review the output from `ancestralStateTree` in FigTree (Fig 2).

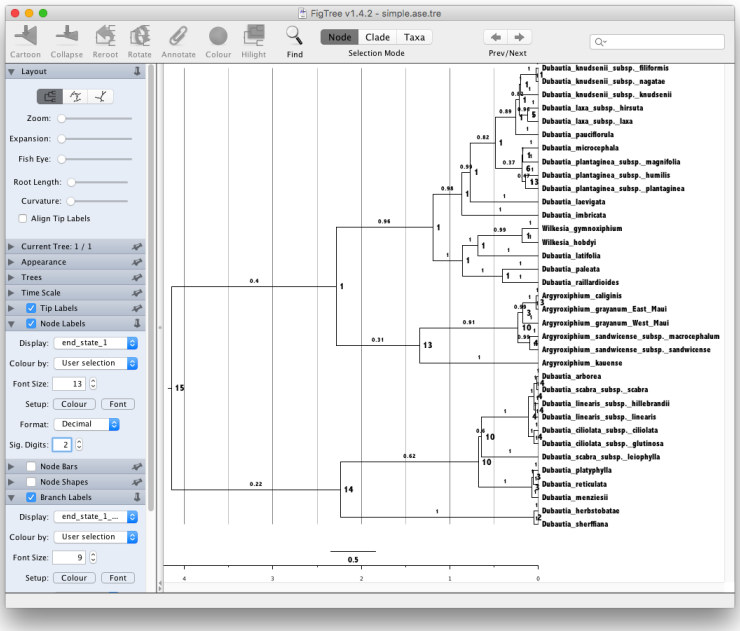

Figure 2: Tree with ancestral state estimates. The most probable end state of each branch (before cladogenesis) is shown at each node. Branches are labeled with the posterior probability for the ancestral state on the tipwards end of the branch.

Nodes are annotated with the first three most probable ancestral states along with their posterior probabilities. When the tree is a random variable, as it is in later exercises, additional information about phylogenetic uncertainty is reported.

Finally, we can also generate a figure with ancestral states in R using RevGadgets that is suitable for publication.

```
# RevGadgets requires development tools for installation
install.packages("devtools")
library(devtools)

# Install RevGadgets
install_github("revbayes/RevGadgets")

# Note about ggtree dependency:
# RevGadgets requires ggtree version 1.5.14 or greater. This can be installed directly
    from GitHub:
install_github("GuangchuangYu/ggtree")
```

Once this is installed you can generate a figure by executing `source("plot_anc_state.n4.R")` from within an R session (Fig 3). Modifying the source file allows you to use the script with different datasets.

Notice that the model infers a widespread ancestral range for the clade (KOMH) approximately four million

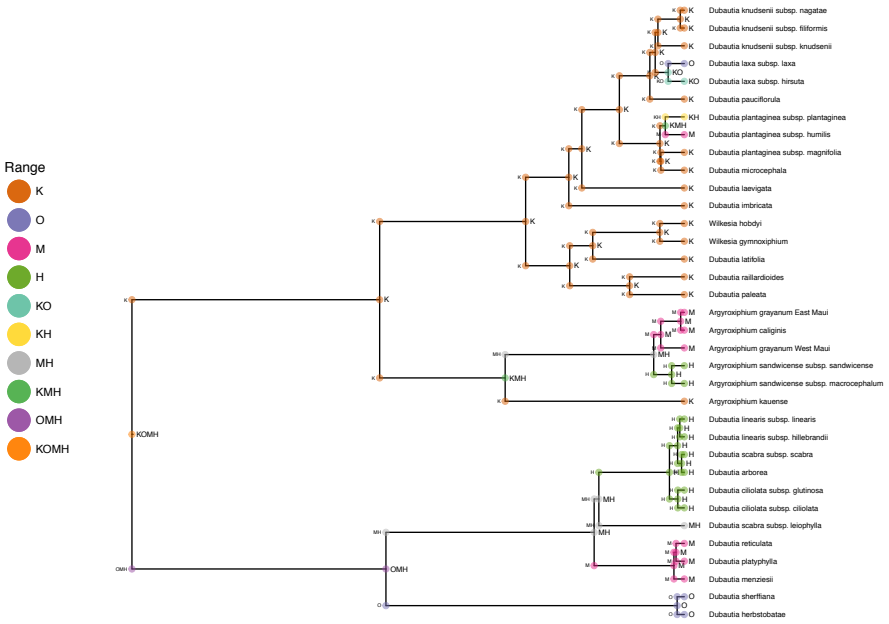

Figure 3: Tree with ancestral state estimates for the "simple" analysis. Nodes are annotated with ancestral states before and after cladogenetic events. Most probable states are shown. Colors of markers indicate the range state. Sizes of markers indicate the posterior probability of that state.

years ago when only Kauai existed. Similar geologically unrealistic widespread ranges are estimated for the Agyroxiphium clade (KMH) and the Dubautia sheriffiana+arborea clade (OMH). The remaining tutorials will focus on improvements to the simple DEC model presented here.

3 An improved DEC analysis

In this section, we'll focus on techniques that allow permit more realistic biogeographic analyses. Topics include applying range size constraints, area connectivity, stratified/epoch models, function-valued dispersal rates, and incorporating uncertainty in paleogeographic event time estimates. These modifications should produce more realistic ancestral range estimates, e.g. that a volcanic island may only be colonized once it has formed, and that distance should have some bearing on dispersal rate.

The remaining tutorials will make use of geographical and paleogeographical measurements for the Hawaiian archipelago, summarized in Table 3. Even though we will continue to use four areas (K, O, M, H) for Section 3, we will use all six areas (R, K, O, M, H, Z) in future sections, hence the full table is given for reference. The paleogeographical information is encoded in the files named `hawaii.n4.times.txt`, `hawaii.n4.distances.txt`, and `hawaii.n4.connectivity.*.txt`.

area	code	a_{max}	a_{min}	$g_{\bullet R}$	$g_{\bullet K}$	$g_{\bullet O}$	$g_{\bullet M}$	$g_{\bullet H}$	$g_{\bullet Z}$
Older islands	R	-	-	-	261	406	500	680	3900
Kauai	K	5.15	5.05	-	-	145	239	419	3900
Oahu	O	3.7	2.2	-	-	-	059	239	3900
Maui Nui	M	1.8	1.3	-	-	-	-	082	3900
Hawaii	H	0.7	0.3	-	-	-	-	-	3900
Mainland	Z	-	-	-	-	-	-	-	-

Table 3: Hawaiian paleogeography model. The six areas are given in Figure 1. Ages a_{max} and a_{min} report the maximum and minimum origination times for the given island. Distances g_{ij} report the shortest geographical distance from the coast of the row's area to the column's area (at present).

Analysis

Start by creating our filename variables

```
range_fn = "data/n4/silversword.n4.range.nex"
tree_fn = "data/n4/silversword.tre"
out_fn = "output/epoch"
geo_fn = "data/n4/hawaii.n4"
times_fn = geo_fn + ".times.txt"
dist_fn = geo_fn + ".distances.txt"
```

Create some helper analysis variables

```
mvi = 1
mni = 1
n_gen = 5000
```

Read in the presence-absence range characters and record the number of areas in the dataset

```
dat_range_01 = readDiscreteCharacterData(range_fn)
n_areas <- dat_range_01.nchar()
```

Often, biogeographers wish to limit to the maximum allowable range size. This prohibits widespread species ranges and to reduce the total number of range states in the analysis, thus benefitting computational efficiency. Suppose we disallowed ranges from including more than two areas. The total number of ranges equals $\sum_{k=0}^{m} \binom{n}{k}$ where n is the total number of areas, m is the maximum number of permissible areas, and $\binom{n}{k}$ is the number of ways to sample k unordered areas from a pool of n areas.

```
max_areas <- 2
n_states <- 0
for (k in 0:max_areas) n_states += choose(n_areas, k)
```

Then format the dataset for the reduced state space

```
dat_range_n = formatDiscreteCharacterData(dat_range_01, "DEC", n_states)
```

Our state space now includes only 11 states (∅, K, O, M, H, KO, KM, OM, KH, OH, MH).

Next, we'll set up the paleogeographic model. Read in the list of minimum and maximum ages of island formation

```
time_bounds <- readDataDelimitedFile(file=times_fn, delimiter=" ")
n_epochs <- time_bounds.size()
```

Read in a vector of matrices that describe the connectivity between areas over time. Note, there is one connectivity matrix per epoch, ordered from oldest to youngest.

```
for (i in 1:n_epochs) {
  epoch_fn[i] = geo_fn + ".connectivity." + i + ".txt"
  connectivity[i] <- readDataDelimitedFile(file=epoch_fn[i], delimiter=" ")
}
```

Read in the matrix of distances between all pairs of areas (km). For simplicity, we will assume that distances remain constant over time, even though they certainly vary.

```
distances <- readDataDelimitedFile(file=dist_fn, delimiter=" ")
```

For now, we'll assume we know the dated species phylogeny without error.

```
tree <- readTrees(tree_fn)[1]
```

Dispersal rates might make use of some extrinsic information, such as geographical distances between areas (MacArthur and Wilson 1967; Webb and Ree 2012). We model this as $d_{ij} = ae^{-bg_{ij}}$ where g_{ij} is the geographical distance between areas i and j and a and b are parameters that scale distance on linear and exponential scales, respectively. Note that all dispersal rates equal a when $b = 0$. The variables a, b, d_{ij}, and g_{ij} are stored in memory as `dispersal_rate`, `distance_scale`, `dr[i][j]`, and `distances[i][j]`.

```
log10_rate_bg ~ dnUniform(-4,2)
log10_rate_bg.setValue(-2)
rate_bg := 10^log10_rate_bg
moves[mvi++] = mvSlide(log10_rate_bg, weight=4)
```

Fix the base dispersal rate to 1

```
dispersal_rate <- abs(1)
```

Add a distance scale parameter

```
distance_scale ~ dnUnif(0,20)
distance_scale.setValue(0.01)
moves[mvi++] = mvScale(distance_scale, weight=3)
```

Now we can assign rates that are functions of distance between all pairs of areas

```
for (i in 1:n_epochs) {
  for (j in 1:n_areas) {
    for (k in 1:n_areas) {
      dr[i][j][k] <- abs(0)
      if (connectivity[i][j][k] > 0) {
        dr[i][j][k] := dispersal_rate * exp(-distance_scale * distances[j][k])
      }
    }
  }
}
```

It is unlikely that widespread ranges persist across disjunct areas for long periods of time. Extirpation is more likely to occur in fragmented ranges than well-connected ranges, where peripheral populations are continuously reinforced from the center.

```
log_sd <- 0.5
log_mean <- ln(1) - 0.5*log_sd^2
extirpation_rate ~ dnLognormal(mean=log_mean, sd=log_sd)
moves[mvi++] = mvScale(extirpation_rate, weight=2)
```

```
for (i in 1:n_epochs) {
  for (j in 1:n_areas) {
    for (k in 1:n_areas) {
      er[i][j][k] <- abs(0.0)
    }
    er[i][j][j] := extirpation_rate
  }
}
```

Build a rate matrix for each time interval

```
for (i in 1:n_epochs) {
  Q_DEC[i] := fnDECRateMatrix(dispersalRates=dr[i],
                              extirpationRates=er[i],
                              maxRangeSize=max_areas)
}
```

Treat epoch times as random variables. The present is always the present.

```
for (i in 1:n_epochs) {
  time_max[i] <- time_bounds[i][1]
  time_min[i] <- time_bounds[i][2]
  if (i != n_epochs) {
      epoch_times[i] ~ dnUniform(time_min[i], time_max[i])
      moves[mvi++] = mvSlide(epoch_times[i], delta=(time_max[i]-time_min[i])/2)
  } else {
      epoch_times[i] <- 0.0
  }
}
```

Create the epoch rate generator object

```
Q_DEC_epoch := fnEpoch(Q=Q_DEC, times=epoch_times, rates=rep(1,n_epochs))
```

Here, we treat the probability of different types of cladogenetic events as a random variable to be estimate.

```
clado_event_types <- [ "s", "a" ]
p_sympatry ~ dnUniform(0,1)
p_allopatry := abs(1.0 - p_sympatry)
clado_type_probs := simplex(p_sympatry, p_allopatry)
moves[mvi++] = mvSlide(p_sympatry, weight=2)
P_DEC := fnDECCladoProbs(eventProbs=clado_type_probs,
                         eventTypes=clado_event_types,
                         numCharacters=n_areas,
                         maxRangeSize=max_areas)
```

For this dataset, we assume cladogenetic probabilities are constant with respect to geological time.

Only Kauai exists 4 Ma at the assumed time of origin of the clade.

```
rf_DEC <- rep(0, n_states)
rf_DEC[2] <- 1
rf_DEC <- simplex(rf_DEC)
```

Create the phylogenetic model

```
m_bg ~ dnPhyloCTMCClado(tree=tree,
                        Q=Q_DEC_epoch,
                        cladoProbs=P_DEC,
                        branchRates=rate_bg,
                        rootFrequencies=rf_DEC,
                        type="NaturalNumbers",
                        nSites=1)
```

Attach the dataset

```
m_bg.clamp(dat_range_n)
```

And the rest we've done before...

```
monitors[mni++] = mnScreen(printgen=100, rate_bg, extirpation_rate, distance_scale)
monitors[mni++] = mnModel(file=out_fn+".model.log", printgen=10)
monitors[mni++] = mnFile(tree, filename=out_fn+".tre", printgen=10)
monitors[mni++] = mnJointConditionalAncestralState(tree=tree,
                                                   ctmc=m_bg,
                                                   type="NaturalNumbers",
```

```
                                                    withTips=true,
                                                    withStartStates=true,
                                                    filename=out_fn+".states.log",
                                                    printgen=10)
```

the model

```
mymodel = model(m_bg)
```

build and run MCMC

```
mymcmc = mcmc(mymodel, moves, monitors)
mymcmc.run(n_gen)
```

Results

Example results have around found in `output_example/epoch.*`

When compared to the ancestral state estimates from the "simple" analysis (Figure 3), these results are more consonant with what we know about the origination times of the islands and lineages (Figure XX). First, this reconstruction asserts that the clade originated in the modern Hawaiian islands at a time when only Kauai was above sea level. The ancestral range for Clade I is Maui (M) with probability 0.41 and Maui+Hawaii (MH) with probability 0.33. Similarly, the Clade II does not reconstruct OMH as an ancestral range, since Maui and Hawaii had not yet formed 2.4 Ma. It may be that these are correct biogeographic reconstructions, or they may be artifacts of assuming a fixed and errorless phylogeny. The next tutorials discuss how to jointly estimate phylogeny and biogeography, which potentially improves the estimation of divergence times, tree topology, and ancestral ranges.

[Look at posterior estimates for distance.]

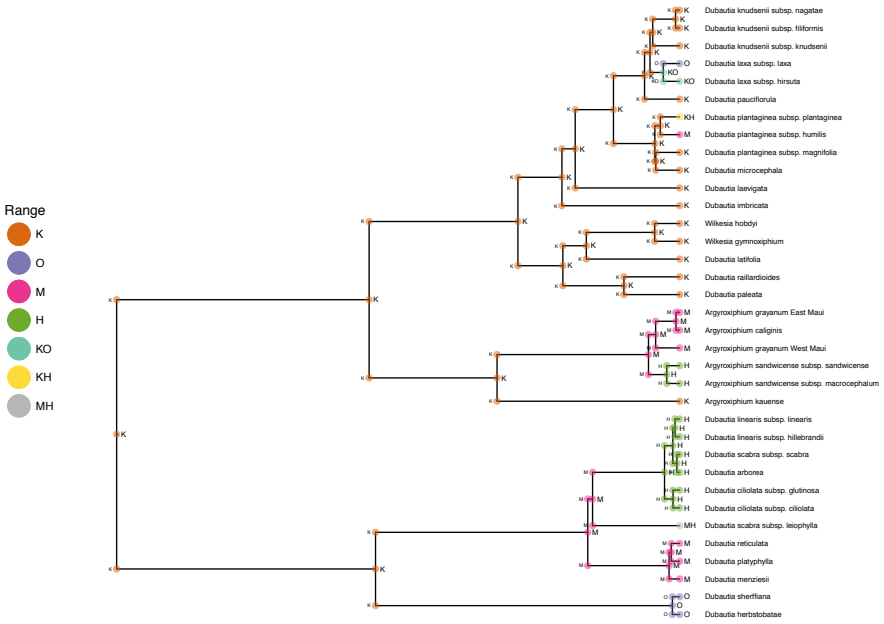

Figure 4: Tree with ancestral state estimates. Nodes are annotated with ancestral states before and after cladogenetic events. Most probable states are shown. Colors of markers indicate the range state. Sizes of markers indicate the posterior probability of that state.

4 Biogeographic dating using DEC

This analysis will jointly estimate phylogeny and biogeography. One benefit is that the biogeographic analysis will intrinsically accommodate phylogenetic uncertainty, both in terms of topology and branch lengths. Another is that paleogeographic evidence has the potential provide information about the geological timing of speciation events in the phylogeny (Landis 2016). Finally, biogeographic data may lend support to certain phylogenetic relationships that have poor resolution otherwise.

As mentioned in Section 2, Hawaiian silverswords are nested in a larger group of plants, the tarweeds. Fossil pollen evidence indicates that tarweeds diversified during a period of aridification from 15–5 Ma in the western regions of North America (Baldwin et al. 1991). Although the oldest Hawaiian island that silverswords inhabit is Kauai, it is possible that silverswords first colonized older islands in the Emperor Island chain that predate the formation of Kauai at about 5.1 Ma.

This makes traditional node-based biogeographic calibrations challenging, because it would require a strong assumption about when and how many times the oldest silversword lineages colonized Kauai. Did silverswords colonize Kauai once directly from the California coast? Or did the colonize the younger islands multiple times from older islands in the chain? And did the event occur immediately after Kauai surfaced or much later? Because we cannot observe the timing and nature of this event directly, this process-based biogeographic dating approach does so through probabilistic inference.

To address these issues, it is necessary that we transition from our simpler 4-area model to a richer 6-area model (see Figure 1). The mainland area (Z) is necessary to force the silversword and tarweed clade

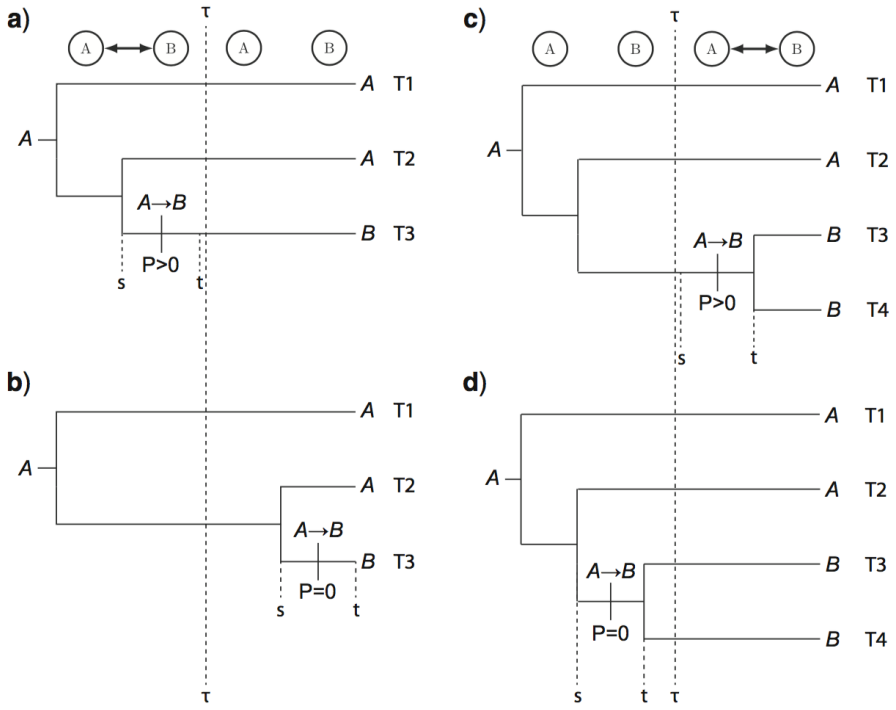

Figure 5: Cartoon of biogeographic transition probabilities as functions of geological time, and how that relates to speciation times. (a) Areas split, dispersal before split, positive probability; (b) Areas split, dispersal after split, zero probability; (c) Areas merge, dispersal after merge, positive probability; (d) Areas merge, dispersal before merge, zero probabilty. Figure from Landis (2016).

to originate apart from the islands. The area corresponding to the older island chain (R) is necessary because we do not know *a priori* whether silverswords colonized the modern islands directly from Z, or first colonized R and only later dispersed into the younger islands any number of times. This permits the silversword origin time to precede the formation of Kauai when the dispersal rate is large.

Additionally, we will add three tarweed taxa to our dataset, increasing the total number of taxa to 38. We'll use internal transcribed spacer (ITS) to estimate the phylogeny, which is a 600 bp non-coding locus that is historically important for plant systematics. Because the locus is relatively short, it will also leave us with a fair amount of phylogenetic uncertainty in branch length and topology estimates. However, because we're estimating phylogeny and biogeography, it will be correctly incorporated into our ancestral range estimates.

Note: like all dating methods, including node calibration methods, tip dating methods, and fossilized birth death dating methods, process-based biogeographic dating estimates are prior sensitive and dataset dependent. Applying this model to alternative data sets should be done with care!

Much of this tutorial will be similar to the previous sections, except we are adding a birth-death process and a molecular substitution process to the model graph.

Analysis

Create the necessary input/output variables.

```
range_fn = "data/n6/silversword.n6.range.nex"
mol_fn = "data/n6/silversword.mol.nex"
tree_fn = "data/n6/silversword.tre"
out_fn = "output/test_epoch_phy"
geo_fn = "data/n6/hawaii.n6"
times_fn = geo_fn + ".times.txt"
dist_fn = geo_fn + ".distances.txt"
```

Add the analysis helper variables

```
mvi = 1
mni = 1
n_gen = 1e5   # more parameters, longer run!
```

Read in the molecular alignment

```
dat_mol = readDiscreteCharacterData(mol_fn)
```

Impose limits to the maximum range size, both to prohibit widespread species ranges and to improve the computational efficiency of the method.

First, get the number of areas

```
dat_range_01 = readDiscreteCharacterData(range_fn)
n_areas <- dat_range_01.nchar()
```

Suppose we wanted to forbid ranges from being three or more areas in size. The total number of ranges is $\sum_{k=0}^{m} \binom{n}{k}$ where n is the total number of areas, m is the maximum number of permissible areas, and $\binom{n}{k}$ is the number of ways to sample k unordered areas from a pool of n areas.

```
max_areas <- 2
n_states <- 0
for (k in 0:max_areas) n_states += choose(n_areas, k)
```

Then format the dataset for the reduced state space

```
dat_range_n = formatDiscreteCharacterData(dat_range_01, "DEC", n_states)
```

Read in the list of minimum and maximum ages of island formation

```
time_bounds <- readDataDelimitedFile(file=times_fn, delimiter=" ")
n_epochs <- time_bounds.size()
```

Read in a vector of matrices that describe the connectivity between areas over time. Note, there is one connectivity matrix per epoch, ordered from oldest to youngest.

```
for (i in 1:n_epochs) {
  epoch_fn[i] = geo_fn + ".connectivity." + i + ".txt"
  connectivity[i] <- readDataDelimitedFile(file=epoch_fn[i], delimiter=" ")
}
```

Read in the matrix of distances between all pairs of areas (km). For simplicity, we will assume that distances remain constant over time, even though they certainly vary.

```
distances <- readDataDelimitedFile(file=dist_fn, delimiter=" ")
```

For this analysis, all starting trees should have a non-zero probability. However, in general, it is possible that certain combinations of phylogeny, biogeography, and paleogeography have zero-valued likelihoods should the epoch model introduce reducible rate matrix structures. In addition, the initial MCMC state must have a non-zero probability for it to work properly. For MCMC to operate correctly, it must be initialized with a state with non-zero probability. See the supplemental of Buerki et al. (2011) for an explanation.

```
tree_init = readTrees(tree_fn)[1]
```

Store some basic information about the taxon set, the number of taxa, and the number of branches in the tree

```
taxa = tree_init.taxa()
n_taxa = taxa.size()
n_branches = 2 * n_taxa - 2
```

4.0.1 The tree model

In this exercise we will also be estimating the phylogeny (topology and branch lengths), meaning our tree will be a stochastic node with a prior distribution. For this, we'll use a constant rate birth-death process.

Assign root age with a maximum age of 15Ma to reflect the fossil pollen record for Californian tarweeds [cite].

```
root_age ~ dnUniform(0, 15)
moves[mvi++] = mvScale(root_age, weight=2)
```

Assign the proportion of sampled taxa (we have a non-uniform sampling scheme, but this should suffice).

```
rho <- 35/50
```

Assign the birth and death priors. It is important to note that the birth and death priors induce a root age distribution through the birth-death process. These priors generate a relatively uniform root age distribution between 2.5–15 Ma in the absence of data (i.e. running MCMC with the `underPrior=true` option).

```
birth ~ dnExp(1)
moves[mvi++] = mvScale(birth)
death ~ dnExp(1)
moves[mvi++] = mvScale(death)
```

Instantiate a tree variable generated by a birth-death process

```
tree ~ dnBDP(lambda=birth, mu=death, rho=rho, rootAge=root_age, taxa=taxa)
```

Add topology and branch length moves

```
moves[mvi++] = mvNNI(tree, weight=n_branches/2)
moves[mvi++] = mvFNPR(tree, weight=n_branches/8)
moves[mvi++] = mvNodeTimeSlideUniform(tree, weight=n_branches/2)
```

Provide a starting tree (improves mixing, not essential)

```
tree.setValue(tree_init)
root_age.setValue(tree_init.rootAge())
```

4.0.2 The molecular model

In addition, to inform our branch lengths (in relative time units) and our topology, we will specify a simple HKY+Γ4+UCLN model of molecular substitution.

First specify a base rate for the molecular clock. This prior is uniform over orders of magnitude, between 10^-6 and 10^3

```
log10_rate_mol ~ dnUniform(-6, 3)
log10_rate_mol.setValue(-1)
moves[mvi++] = mvSlide(log10_rate_mol, weight=5, delta=0.2)
rate_mol := 10^log10_rate_mol
```

Assign log-normal relaxed clock rate multipliers to each branch in the tree. These priors have a mean of 1 so each branch prefers a strict clock model in the absence of data.

```
branch_sd <- 1.0
branch_mean <- 0.0 - 0.5 * branch_sd^2
for (i in 1:n_branches) {
    branch_rate_multiplier[i] ~ dnLognormal(mean=branch_mean, sd=branch_sd)
    moves[mvi++] = mvScale(branch_rate_multiplier[i])
    branch_rates[i] := rate_mol * branch_rate_multiplier[i]
}
```

Now we'll create an HKY rate matrix. First the transition-transversion rate ratio (with prior with mean=1)

```
kappa ~ dnGamma(2,2)
moves[mvi++] = mvScale(kappa)
```

the base frequencies over A, C, G, and T

```
bf ~ dnDirichlet([1,1,1,1])
moves[mvi++] = mvSimplexElementScale(bf, alpha=10, weight=2)
```

then using the base frequencies and TsTv rate ratio to build the matrix

```
Q_mol := fnHKY(kappa, bf)
```

Next, we'll create a +Γ4 across site rate variation model. This requires a parameter to control how much site rate heterogeneity there is.

```
alpha ~ dnUniform(0,50)
moves[mvi++] = mvScale(alpha)
```

and a discretized Gamma distribution with 4 categories

```
site_rates := fnDiscretizeGamma(alpha, alpha, 4)
```

When `alpha` is large, then the Gamma distribution centers its density around the rate multiplier of 1, meaning that all sites evolve at similar rates. When `alpha` is small, the Gamma distribution presents more site rate heterogeneity.

Finally, we'll create our molecular model of substitution

```
m_mol ~ dnPhyloCTMC(Q=Q_mol, tree=tree, branchRates=branch_rates, siteRates=site_rates,
    type="DNA", nSites=dat_mol.nchar())
```

and attach the ETS dataset

```
m_mol.clamp(dat_mol)
```

4.0.3 The biogeographic model

The biogeographic model is identical to that described in Section 3.

First, create biogeographic rate parameter.

```
log10_rate_bg ~ dnUniform(-4,2)
log10_rate_bg.setValue(-2)
rate_bg := 10^log10_rate_bg
moves[mvi++] = mvSlide(log10_rate_bg, weight=4)
```

The relative dispersal rate is fixed to 1

```
dispersal_rate <- abs(1)
```

the distance scale parameter

```
distance_scale ~ dnUnif(0,20)
distance_scale.setValue(0.001)
moves[mvi++] = mvScale(distance_scale, weight=3)
```

Now we can assign rates that are functions of distance between all pairs of areas

```
for (i in 1:n_epochs) {
  for (j in 1:n_areas) {
    for (k in 1:n_areas) {
     dr[i][j][k] <- abs(0)
     if (connectivity[i][j][k] > 0) {
       dr[i][j][k] := dispersal_rate * exp(-distance_scale * distances[j][k])
     }
    }
  }
}
```

It is unlikely that widespread ranges persist across disjunct areas for long periods of time. Extirpation is more likely to occur in fragmented ranges than well-connected ranges, where peripheral populations are continuously reinforced from the center.

```
log_sd <- 0.5
log_mean <- ln(1) - 0.5*log_sd^2
extirpation_rate ~ dnLognormal(mean=log_mean, sd=log_sd)
moves[mvi++] = mvScale(extirpation_rate, weight=2)

for (i in 1:n_epochs) {
  for (j in 1:n_areas) {
    for (k in 1:n_areas) {
      er[i][j][k] <- abs(0.0)
    }
    er[i][j][j] := extirpation_rate
  }
}
```

Build a rate matrix for each time interval

```
for (i in 1:n_epochs) {
  Q_DEC[i] := fnDECRateMatrix(dispersalRates=dr[i],
                        extirpationRates=er[i],
                        maxRangeSize=max_areas)
}
```

Treat epoch times as random variables. The present is always the present.

```
for (i in 1:n_epochs) {
  time_max[i] <- time_bounds[i][1]
```

```
  time_min[i] <- time_bounds[i][2]
  if (i != n_epochs) {
    epoch_times[i] ~ dnUniform(time_min[i], time_max[i])
    moves[mvi++] = mvSlide(epoch_times[i], delta=(time_bounds[i][1]-time_bounds[i][2])
        /2)
  } else {
    epoch_times[i] <- 0.0
  }
}
```

Create the epoch rate generator object

```
Q_DEC_epoch := fnEpoch(Q=Q_DEC, times=epoch_times, rates=rep(1, n_epochs))
```

Here, we treat the probability of different types of cladogenetic events as a random variable to be estimate.

```
clado_event_types <- [ "s", "a" ]
p_sympatry ~ dnUniform(0,1)
p_allopatry := abs(1.0 - p_sympatry)
moves[mvi++] = mvSlide(p_sympatry, delta=0.1, weight=2)
clado_event_probs := simplex(p_sympatry, p_allopatry)
P_DEC := fnDECCladoProbs(eventProbs=clado_event_probs,
                         eventTypes=clado_event_types,
                         numCharacters=n_areas,
                         maxRangeSize=max_areas)
```

For this dataset, we assume cladogenetic probabilities are constant with respect to geological time.

```
rf_DEC <- rep(0, n_states)
rf_DEC[n_areas+1] <- 1 # Mainland (Z) is the only possible starting state
rf_DEC <- simplex(rf_DEC)
```

Create the phylogenetic model

```
m_bg ~ dnPhyloCTMCClado(tree=tree,
                        Q=Q_DEC_epoch,
                        cladoProbs=P_DEC,
                        branchRates=rate_bg,
                        rootFrequencies=rf_DEC,
                        type="NaturalNumbers",
                        nSites=1)
```

Attach the dataset

```
m_bg.clamp(dat_range_n)
```

Add a deterministic node to monitor the crown age of the silversword radiation

```
ingroup_clade <- clade("Wilkesia_hobdyi",
                       "Dubautia_reticulata",
                       "Dubautia_microcephala",
                       "Argyroxiphium_caliginis")

ingroup_age := tmrca(tree, ingroup_clade)
```

Add add nodes that report the posterior probability that the clade originates before or after a given island. When the first argument in of the `ifelse` function returns `true`, the node has value `1` and `0` otherwise. Thus, the mean of this variable gives the posterior probability that the inequality is satisfied.

```
for (i in 1:n_epochs) {
  ingroup_older_island[i] := ifelse(ingroup_age > epoch_times[i], 1, 0)
}
```

```
monitors[mni++] = mnScreen(printgen=100, ingroup_age)
monitors[mni++] = mnModel(file=out_fn+".model.log", printgen=100)
monitors[mni++] = mnFile(tree, filename=out_fn+".tre", printgen=100)
monitors[mni++] = mnJointConditionalAncestralState(tree=tree,
                                                   ctmc=m_bg,
                                                   type="NaturalNumbers",
                                                   withTips=true,
                                                   withStartStates=true,
                                                   filename=out_fn+".states.log",
                                                   printgen=100)
```

Compose the model object. Because `ingroup_older_island` does not contribute to the model likelihood, it must be manually introduced to the model object.

```
mymodel = model(m_bg, ingroup_older_island)
```

```
mymcmc = mcmc(mymodel, moves, monitors)
mymcmc.run(n_gen)
```

Results

Example results are found in `output_example/epoch_phy.*` *and* `output_example/simple_phy.*`

To understand the influence of the epoch model on ancestral range and divergence time estimation, it is important to run addition analyses with alternative settings. Scripts to jointly estimate molecular evolution, historical biogeographic, and phylogenetic parameters are available as `scripts/run_simple_phy.Rev` and `scripts/run_epoch_phy.Rev`. The "epoch" analysis is identical to the analysis just described. The "simple" analysis is similar to the "epoch" analysis, except it substitutes the paleogeography-aware model of range evolution (see Section 3) for a paleogeography-naive model (see Section 2).

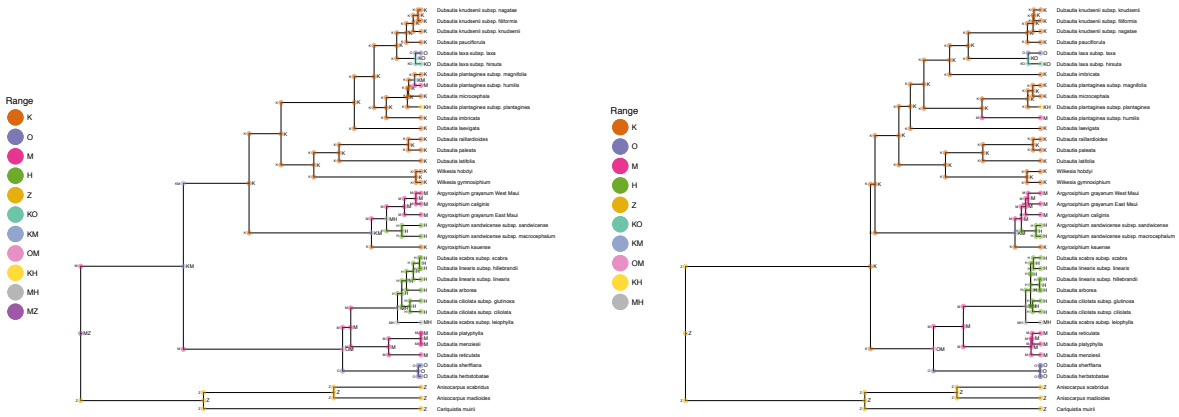

Figure 6: Joint estimate of phylogeny and biogeography. The left panel ignores paleogeography while the right panel conditions on it.

We see that simple analysis (Fig 6, Left) estimates the ancestral range at the root of the clade as Maui + Mainland (MZ). This is unrealistic, both because of the extreme distance between those areas, but also the simple analysis estimates the root age to be $10.3(HPD4.6, 15.0)Ma$, well before Maui originated.

The epoch analysis (Fig 6, Right) produces more sensible ancestral range estimates, with Kauai being colonized first, and younger islands only being colonized as they become available. When comparing the results to the earlier fixed-phylogeny epoch results 4, we recover a greater role for cladogenesis for the younger speciation events. These two analyses only differ in terms of whether the phylogeny is fixed or estimated, so it is likely a result of phylogenetic error in the fixed tree.

In Tracer, one can look at the sampled posterior of island ages in comparison the origination time of crown silverswords (Fig). The left panel shows the simple analysis, where crown silverswords often originate before the formation of Kauai. The right panel shows that crown silverswords probably originated before the formation of Maui, but after the formation of Kauai. By tabulating the results of the deterministic variable `ingroup_older_island`, we measure the posterior probability that crown silverswords originated before or after each particular epoch in the model.

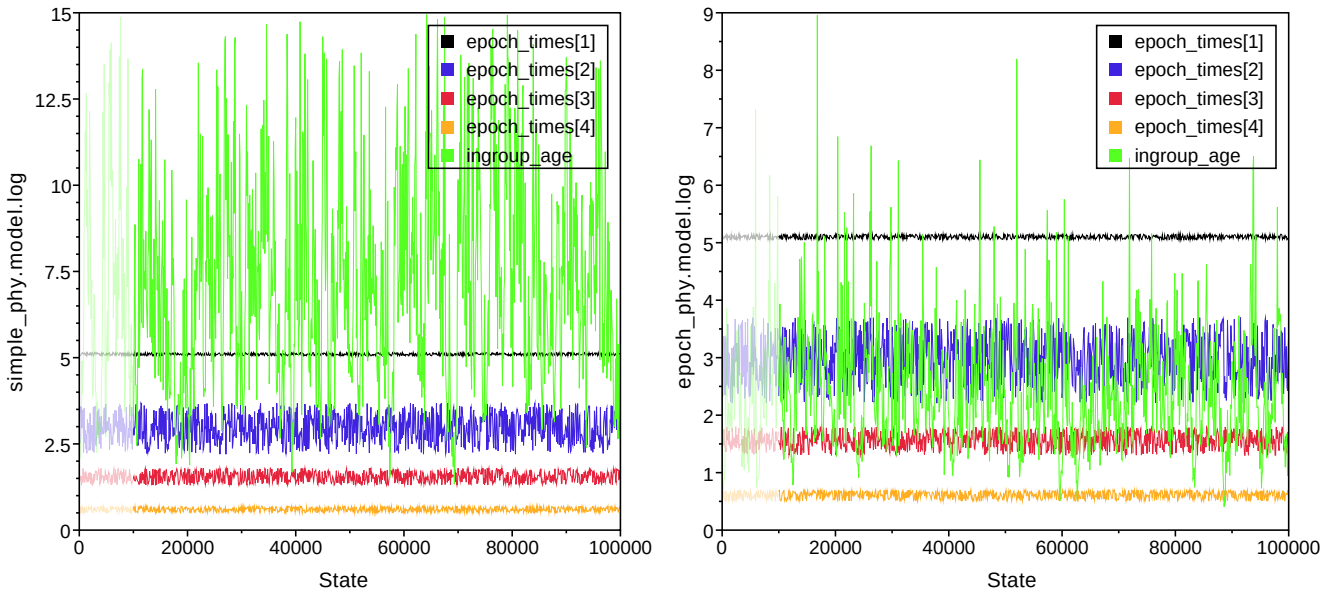

Figure 7: Plot of trace for island ages and the origin time of living silverswords. The left panel ignores paleogeography while the right panel conditions on it.

Model	$P(a_s > a_K)$	$P(a_s > a_O)$	$P(a_s > a_M)$	$P(a_s > a_H)$
simple	0.72	0.94	0.99	1.00
epoch	0.02	0.26	0.84	0.99

Table 4: Posterior probability that the age of crown silverswords (a_s) is older than the origination times of K, O, M, and H (a_K, a_O, a_M, a_H, respectively). The "simple" model (Left) ignores paleogeography while the "epoch" model (Right) conditions on it.

References

Baldwin, B. G., D. W. Kyhos, J. Dvorak, and G. D. Carr. 1991. Chloroplast dna evidence for a north american origin of the hawaiian silversword alliance (asteraceae). Proceedings of the National Academy of Sciences 88:1840–1843.

Baldwin, B. G. and M. J. Sanderson. 1998. Age and rate of diversification of the hawaiian silversword alliance (compositae). Proceedings of the National Academy of Sciences 95:9402–9406.

Buerki, S., F. Forest, N. Alvarez, J. A. Nylander, N. Arrigo, and I. Sanmartín. 2011. An evaluation of new parsimony-based versus parametric inference methods in biogeography: a case study using the globally distributed plant family sapindaceae. Journal of Biogeography 38:531–550.

Carlquist, S. 1959. Studies on madinae: anatomy, cytology, and evolutionary relationships. Aliso 4:171–236.

Landis, M. J. 2016. Biogeographic dating of speciation times using paleogeographically informed processes. Systematic biology Page syw040.

Landis, M. J., N. J. Matzke, B. R. Moore, and J. P. Huelsenbeck. 2013. Bayesian analysis of biogeography when the number of areas is large. Systematic Biology 62:789–804.

MacArthur, R. H. and E. O. Wilson. 1967. Theory of Island Biogeography. vol. 1. Princeton University Press.

Massana, K. A., J. M. Beaulieu, N. J. Matzke, and B. C. O'Meara. 2015. Non-null effects of the null range in biogeographic models: Exploring parameter estimation in the dec model. bioRxiv Page 026914.

Matzke, N. J. 2012. Founder-event speciation in biogeobears package dramatically improves likelihoods and alters parameter inference in dispersal–extinction–cladogenesis dec analyses. Frontiers of Biogeography 4:210.

Ree, R. H., B. R. Moore, C. O. Webb, M. J. Donoghue, and K. Crandall. 2005. A likelihood framework for inferring the evolution of geographic range on phylogenetic trees. Evolution 59:2299–2311.

Ree, R. H. and S. A. Smith. 2008. Maximum likelihood inference of geographic range evolution by dispersal, local extinction, and cladogenesis. Systematic Biology 57:4–14.

Webb, C. O. and R. Ree. 2012. Historical biogeography inference in malesia. Biotic evolution and environmental change in Southeast Asia Pages 191–215.

Version dated: January 4, 2017