

Using mbd

Giovanni Laudanno and Richel J.C. Bilderbeek

2018-11-06

1. Introduction

`mbd` is a package to support multiple birth model. This model accounts for a diversification process in which multiple birth (at different nodes) are allowed to occur at the same time.

The basic framework is described in Etienne et al. 2012¹. The main purpose of the package is to provide a tool to estimate, given a phylogenetic tree, the likelihood for a set of four parameters: λ , μ , ν and q . Here:

- λ is the sympatric speciation rate;
- μ is the extinction rate;
- ν is the multiple allopatric speciation trigger rate;
- q is the single-lineage speciation probability.

Those parameters may be inferred through a maximum likelihood approach using the function `mbd_ml`.

2. Setup

We will need to load the package:

```
library(mbd)
```

Also, we will set the random number generator seed to a value, so that this vignette always produces the same results:

```
set.seed(2)
```

3. Simulating an MBD tree

First we set the parameters of the MBD speciation model:

```
lambda <- 0.2 # sympatric speciation rate  
mu <- 0.15 # extinction rate;  
nu <- 2.0 # multiple allopatric speciation trigger rate  
q <- 0.1 # single-lineage speciation probability
```

Then we set some basic parameters to create a simulated tree:

¹Etienne, R.S., Haegeman, B., Stadler, T., Aze, T., Pearson, P.N., Purvis, A., Phillimore, A.B.: Diversity-dependence brings molecular phylogenies closer to agreement with the fossil record. *Proceedings of the Royal Society of London B: Biological Sciences* 279(1732), 1300–1309 (2012)

```

crown_age <- 1
sim_pars <- c(lambda, mu, nu, q)
sim <- mbd_sim(
  pars = sim_pars,
  n_0 = 2, # Use a crown age
  age = crown_age,
  cond = 1 # Condition on non-extinction
)

```

These are additional auxiliary parameters such as `n_0` (equal to 1 if we want a stem, or 2 if we want a crown), `cond` (let us condition the tree on the survival of crown species) and `crown_age` (being the total age of the tree).

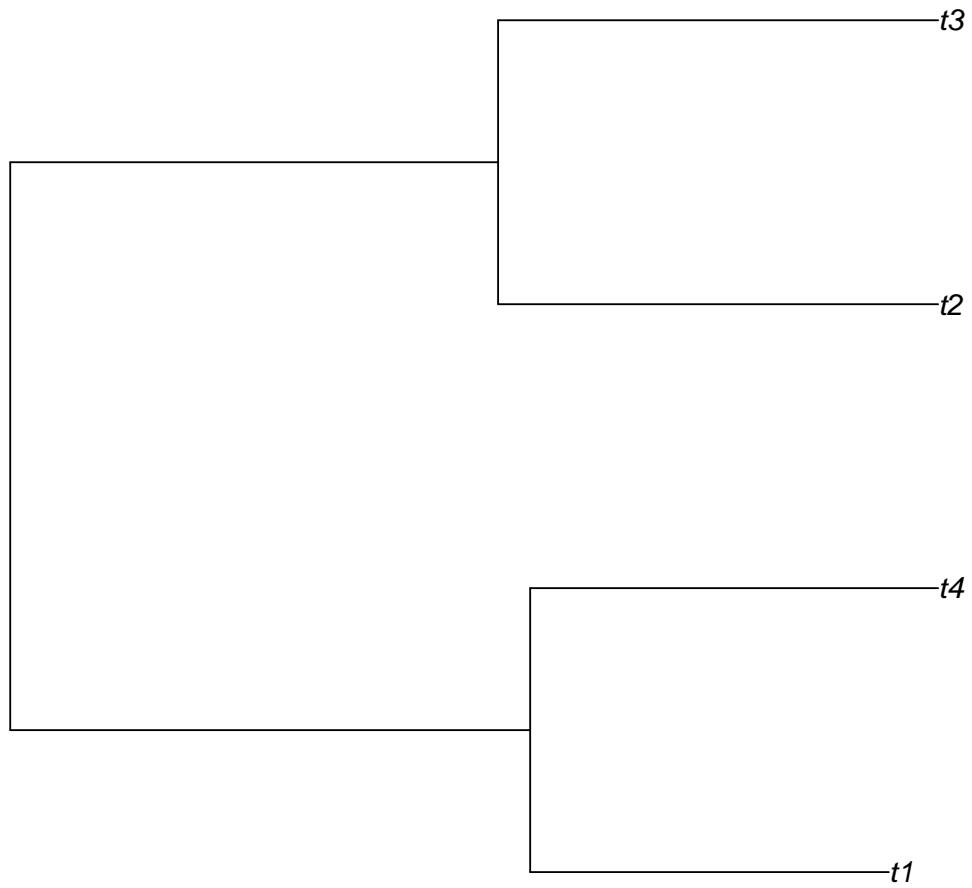
4. Showing the results

The `sim` list contain different objects. You can visualize the full tree and keep track of species that die before present time (missing species):

```

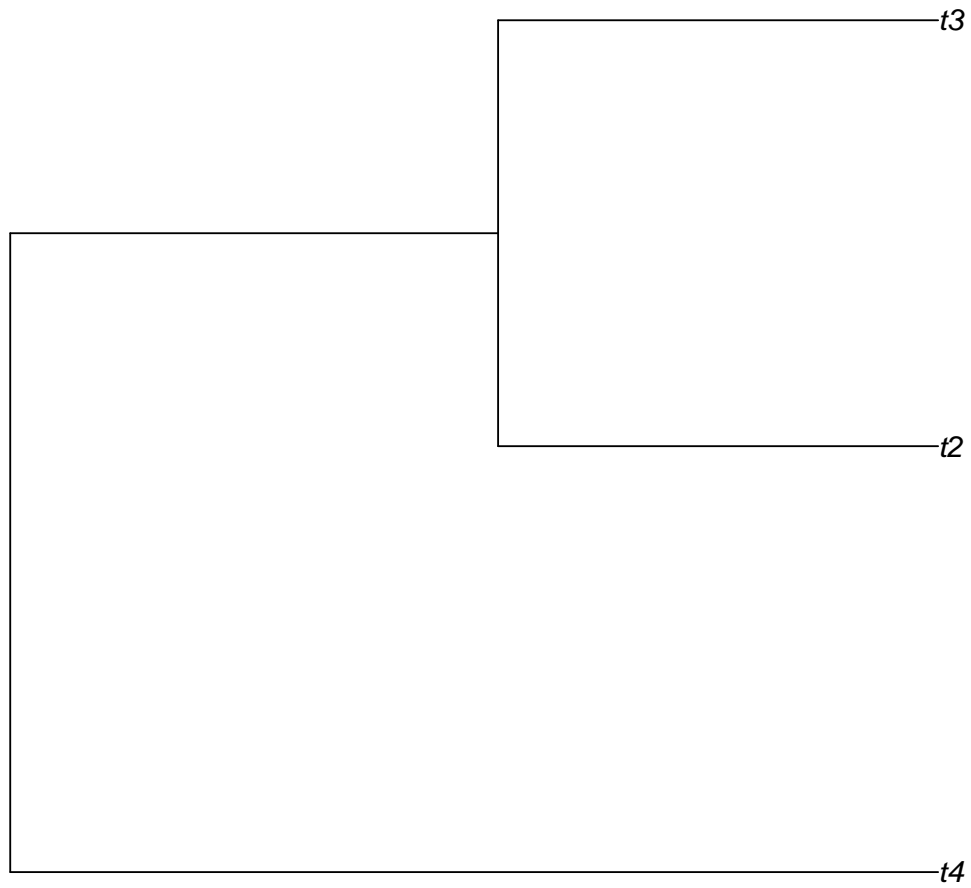
graphics::plot(sim$full_tree)

```



If you want you can also show the reconstructed tree, obtained cutting the kinks from the previous tree:

```
graphics::plot(sim$reconstructed_tree)
```



The dataframe containing all the information related to this tree is labeled as L (or L-table). You can access it by typing:

```
knitr::kable(head(sim$l_matrix))
```

birth_time	parent	id	death_time
1.0000000	0	-1	0.0531426
1.0000000	-1	2	-1.0000000
0.4742348	2	3	-1.0000000
0.4396944	-1	-4	-1.0000000

In this table:

- the first column are the times at which species are born;
- the second contains labels of the species' parents; positive and negative values only indicate whether the species belongs to the left or right crown lineage;
- the third contains the labels of the daughter species themselves; positive and negative values only indicate whether the species belongs to the left or right crown lineage;
- the fourth contains the extinction times of the species. If this is equal to -1, then the species is still extant.

To recover the vector of branching times use:

```
knitr::kable(head(sim$brts))
```

	x
	1.0000000
	0.4742348

In this vector values repeated more than once indicate that a multiple birth process has taken place.

5. Likelihood estimation

To estimate the likelihood of a phylogeny, one needs

- parameter estimates
- the branching times (topology is irrelevant)

In this case, we will use

- the parameter estimates that we used to generate the tree
- the branching times of the tree we generated

The function `mbd_loglik` shows the likelihood of the parameters having generated the branching times:

```
mbd::mbd_loglik(
  pars = c(lambda, mu, nu, q),
  brts = sim$brts,
  n_0 = 2, # Crown age
  cond = 1 # Non-extinction
)
```

```
## [1] -1.237561
```

6. Maximum likelihood estimation

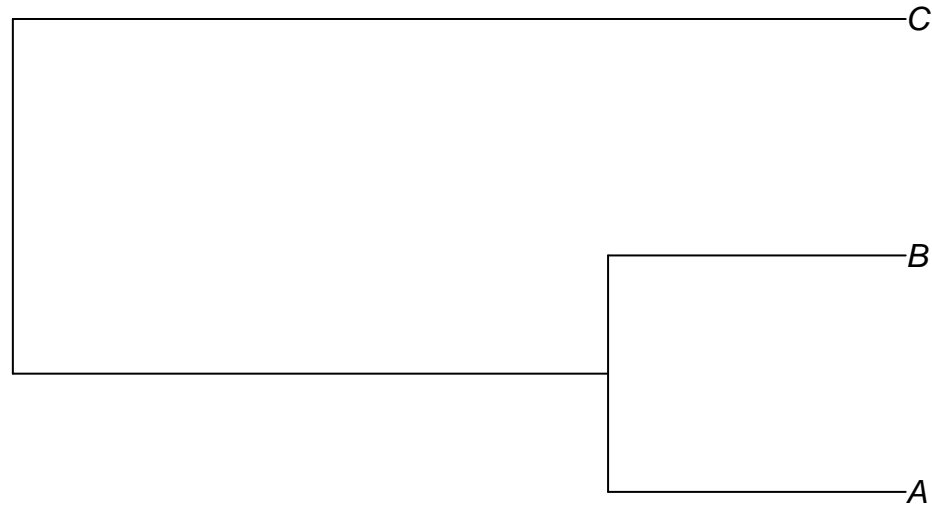
You can also ask for an optimization of the likelihood function using the function `mbd_ml` either considering a variation of all the parameters (λ, μ, ν, q) or any their subset of your choice.

Because this is a heavy calculation, we will use a simpler tree:

```

phylogeny <- ape::read.tree(text = "((A:1, B:1):2, C:3);")
ape::plot.phylo(phylogeny)

```



```
brts <- ape::branching.times(phylogeny)
```

If you want for example to maximize the likelihood only for the parameter q :

```

brts <- sim$brts
start_pars <- c(0.2, 0.15, 1, 0.15)
optim_ids <- c(FALSE, FALSE, FALSE, TRUE)
n_0 <- 2
cond <- 1
out <- mbd::mbd_ml(
  start_pars = start_pars,
  true_pars = sim_pars,
  optim_ids = optim_ids,
  brts = brts,
  cond = cond,
  n_0 = n_0,
  verbose = TRUE
)

```

```

## You are optimizing q
## You are fixing lambda mu nu

```

```
## Optimizing the likelihood - this may take a while.
## The loglikelihood for the initial parameter values is -1.24789
## 1 0.15 -1.24788982671216 initial
## 2 0.135287846481876 -1.24103363564973 expand
## 3 0.106964656964657 -1.23648686307913 expand
## 4 0.106964656964657 -1.23648686307913 contract inside
## 5 0.113912133891213 -1.23636226608897 contract inside
## 6 0.110427528675704 -1.23631219027019 contract inside
## 7 0.112167101827676 -1.23630954234395 contract inside
## 8 0.111296634489956 -1.2363038943574 contract inside
## 9 0.111296634489956 -1.2363038943574 contract inside
## 10 0.111296634489956 -1.2363038943574 contract inside
## 11 0.111405368383288 -1.23630383988072 contract inside
## 12 0.111405368383288 -1.23630383988072 contract inside
## 13 0.111378182915196 -1.23630383308519 contract inside
## 14 0.111378182915196 -1.23630383308519 contract inside
## 15 0.111378182915196 -1.23630383308519 contract inside
## 16 0.111378182915196 -1.23630383308519 contract inside
## 17 0.111378182915196 -1.23630383308519 contract inside
## 18 0.111378182915196 -1.23630383308519 contract inside
## Optimization has terminated successfully.
##
## Maximum likelihood parameter estimates: lambda: 0.2 mu: 0.15 nu: 2 q: 0.111
## Maximum loglikelihood: -1.236304
```

```
knitr::kable(out)
```

lambda	mu	nu	q	loglik	df	conv
0.2	0.15	2	0.1113782	-1.236304	1	0