Vignette ecophylo

E. Barthelemy, M. Jaunatre & F. Munoz

09/03/2021

Introduction

We introduce the Python package ecophylo dedicated to the coalescent-based simulation of neutral eco-evolutionary dynamics. Species assemblages and their shared ancestry can be simulated by jointly taking into account the influence of past demographic fluctuations and extinctions along with how divergent genotypes have introduced new species over time through speciation.

The model rests on two main components: (i) a demographic component driving stochastic changes in population sizes, structure and extinctions due to habitat availability possibly linked migration events, (ii) a mutation and protracted speciation component representing how divergent genotypes emerge and define new species over time.

Here we show how to simulate communities and their phylogenetic relationships for a wide range of past demographic scenarios having affected whole assemblages of species. We also provide ways to produce many simulated datasets from prior distributions so as to allow Approximate Bayesian Computation (ABC) methods to retrieve likely parameter values from the comparison of these simulations to observed diversity patterns.

Installation

explain that ecophylo is a python package, that the current vignette shows us here how to use in r blablbla

The packages can be installed from the public repository Pypi with the `pip` tool.

You can also download the *tar.gz* file from the repository and install it with pip. You need to use python3 and install the dependencies before. Note that it is a development stage.

impossible at this moment as the repository is private

```
python3 -m pip install ecophylo
# NOT RUN
# wget -L https://github.com/thegreatlizzyator/ecophylo/blob/master/dist/ecophylo-0.0.5.tar.gz
# python3 -m pip install ecophylo-0.0.5.tar.gz
```

Installation on Windows (the painfull way)

For François : First create a new project in rstudio with **version control** from the github.

You will need to install `{reticulate}` package and python dependencies.

```
if(!"reticulate"%in%installed.packages()[,1]) install.packages('reticulate') # answer yes
library(reticulate)
# python dependencies
conda_install('r-reticulate', c('tskit', 'msprime','ete3','pandas',"Numpy"))
```

```
## Warning in normalizePath(path.expand(path), winslash, mustWork):
## path[1]="C:\ProgramData\Miniconda3\envs\msprime-env/python.exe": Le
## fichier spécifié est introuvable
```

```r
py_install("ecophylo", envname = "r-reticulate", method = "conda", pip = TRUE)
```

```
## Warning in normalizePath(path.expand(path), winslash, mustWork):
## path[1]="C:\ProgramData\Miniconda3\envs\msprime-env/python.exe": Le
## fichier spécifié est introuvable
```

Simulating phylogenies

Constant size model

Our model rests on the fundamental hypothesis that fluctuations in the relative species abundances in a given habitat are driven by neutral drift dynamics, depending on the size of the assemblage (Hubbell, 2001). Thus, the dynamics of the assemblage they form can be represented by coalescence, i.e. by tracing the shared co-ancestry of extant individuals backwards in time until a single common ancestor is found (Kingman 1982).

We start by simulating the shared co-ancestry of n sampled individuals observed at present time in each assemblage using a backward simulation of the coalescent tree. Variant individuals are then sprinkled over the simulated genealogy conditionally to its topology and branch lengths.

EXPLAIN PROTRACTEDNESS - The probability that a mutation occurred on a specific branch of the genealogy, leading to a new variant individual, follows a Poisson distribution with parameter $\mu \cdot B$ where μ is the point mutation rate and B is the length of the branch – change to explain protracted parameters

Let us first simulate a case in which there are no past fluctuations nor migration events linking populations. This implies assuming that Jm the metacommunity effective size has been at equilibrium from present backward, until all lineages coalesced into their Most Recent Common Ancestor (MRCA).

```r
ecophylo <- import('ecophylo')

n <- as.integer(10) # the sample size
Jm <- as.integer(500) # the size of the metacommunity (sensu Hubbell, 2001)
mu <- as.numeric(0.01) # the point mutation rate

tree <- ecophylo$simulate(n, Jm, mu, seed = 42)
```

We can then compute summary statistics on the resulting phylogeny

```r
library(ape)
library(picante)
library(vegan)

abund <- ecophylo$getAbund(tree, n)
phylo <- read.tree(text = tree$write())
```

- taxonomic diversity metrics:

```r
# species richness
specnumber(abund)
```

```
## [1] 4
```

```r
# shannon's diversity
diversity(abund)
```

```
## [1] 1.168282
```

```r
# simpson's diversity
diversity(abund, "simpson")
```

```
## [1] 0.64
```

- phylogenetic diversity metrics:

```r
# Faith's phylogenetic diversity (PD)
pd(matrix(abund, nrow =1, dimnames = list("", phylo$tip.label)), phylo)[[1]]
```

```
## [1] 4937.208
```

```r
# Mean pairwise distance (MPD)
mpd(matrix(abund, nrow =1, dimnames = list("", phylo$tip.label)), cophenetic(phylo))
```

```
## [1] 3067.306
```

```r
# Mean nearest taxon distance (MNTD)
mntd(matrix(abund, nrow =1, dimnames = list("", phylo$tip.label)), cophenetic(phylo))
```

```
## [1] 672.499
```

Simulating past community-wide demographic fluctuations

However, past environmental changes such as alternating periods of contraction and expansion of suitable environmental conditions, may have shaped the evolutionary and ecological diversity of extant organisms. Specifically, climatic fluctuations have made environmental conditions alternatively common or rare over time. For instance, taxa may have undergone restriction of their distribution to local refugia during habitat contraction, from which they could expand when suitable habitat became more common. Yet, in the constant size model, Jm can only be approximately considered as the harmonic mean of the per-generation Jm dynamics. Thus by allowing Jm to vary at specific ages of the past we can relax this equilibrium assumption and account for the discordance of demographic histories between distinct meta-communities.

Let us assume that the Jm has fluctuated in the past.

```r
pastsizes <- list(as.integer(1000), as.integer(5000)) # the size of the metacommunity in the past
changetime <- list(as.integer(100),as.integer(200)) # the time (in generations) at which metacommunity
```

```r
verb = py_capture_output( # capture the verbose
  tree1 <- ecophylo$simulate(n, Jm, mu,
                             past_sizes = pastsizes,
                             changetime = changetime,
                             seed = 42, verbose = TRUE),
  type = "stderr") # verbose is eported as an error
cat(verb) # read this precious verbose
```

```
## Model =  hudson(reference_size=500)
## ==============================
## Epoch: 0 -- 100.0 generations
## ==============================
##       start     end      growth_rate |     0
##      -------- --------       -------- | --------
## 0 |   500     500                  0 |     0
##
## Events @ generation 100.0
##    - Population parameter change for -1: initial_size -> 1000
## ==================================
## Epoch: 100.0 -- 200.0 generations
## ==================================
##       start     end      growth_rate |     0
##      -------- --------       -------- | --------
## 0 |   1e+03   1e+03                0 |     0
```

```
##
## Events @ generation 200.0
##    - Population parameter change for -1: initial_size -> 5000
## ===============================
## Epoch: 200.0 -- inf generations
## ===============================
##      start      end      growth_rate |      0
##    --------  --------      -------- | --------
## 0 |  5e+03    5e+03                0 |      0
abund1 <- ecophylo$getAbund(tree1, n)
phylo1 <- read.tree(text = tree1$write())

# chunk to check if Maxime don't break the function !
pastsizes <- list(as.integer(1000)) # the size of the metacommunity in the past
changetime <- list(100) # the time (in generations) at which metacommunity size has changed


cat(py_capture_output( # capture the verbose
  tree1b <- ecophylo$simulate_dolly(n, Jm, mu,
                         past_sizes = pastsizes,
                         changetime = changetime,
                         seed = 42, verbose = TRUE),
  type = "stderr")) # verbose is eported as an error
# cat(verb) # read this precious verbose

abund1 <- ecophylo$getAbund(tree1, n)
phylo1 <- read.tree(text = tree1$write())

# TO DO: print demmography debuger here
```

We can compute the same set of statistics on the resulting phylogeny for which Jm has changed in the past an thus compare extant diversity patterns for different past demographic scenarios

- taxonomic diversity metrics:

```
# species richness
specnumber(abund1)
```

```
## [1] 5
```
```
# shannon's diversity
diversity(abund1)
```

```
## [1] 1.504788
```
```
# simpson's diversity
diversity(abund1, "simpson")
```

```
## [1] 0.76
```

Simulating sub-populations linked by vicariance and migration events

Intensive simulations along priors