

1 *ecophylo*: Simulating and assessing eco-evolutionary dynamics  
2 under past environmental changes in Python and R

3  
4 Elizabeth Barthelemy<sup>1\*</sup>, Maxime Jaunatre<sup>1</sup>, and Francois Munoz<sup>1</sup>

5 <sup>1</sup> Universite Grenoble Alpes, LIPhy, 140 Rue de la Physique, 38402 Saint-Martin-d Heres, France

6 Running title: *ecophylo*: simulating eco-evolutionary dynamics

7 Research article

8 Word count: XXXXX

9 (\*) Corresponding author, elizabeth.barthelemy@univ-grenoble-alpes.fr

## Abstract

- We introduce the Python package *ecophylo* dedicated to coalescent-based simulation of 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 shared co-ancestry of present individuals is simulated backward in time using coalescent theory. Speciation events are then sprinkled over the simulated genealogy conditionally to its topology and branch lengths. The phylogenetic relationships amongst individuals and their abundances are finally obtained by merging paraphyletic clades into single species. Coalescent reconstruction of the genealogy of individuals can be simulated to represent past demographic fluctuations due to varying habitat availability, or include multiple communities linked by migration events.
- The package includes tools to simulate large numbers of datasets and associated summary statistics, so that Approximate Bayesian Computation methods can be used to estimate parameter values of these processes.. Diverse patterns of taxonomic and phylogenetic compositions can be generated. The first version of the package allows simulating neutral coalescent genealogies, and will incorporate further non-neutral eco-evolutionary scenarios in future. The package can be used to explore how past demographic fluctuations have affected species abundances and phylogenetic relationships, and to estimate the parameters of these processes based on observed patterns. We provide step by step examples in both Python and R languages.

Key-words: eco-evolutionary modelling; community phylogeny; coalescent; demographic stochasticity, ecological drift, extinction-speciation dynamics

## Introduction

Observed species distributions and biodiversity patterns are shaped by current ecological processes but also reflect the influence of past evolutionary and biogeographic dynamics @svenning\_influence\_2015. For instance, alternating periods of contraction and expansion of suitable environmental conditions should affect both demographic and speciation-extinction dynamics in a given habitat over time (Barthelemy et al. Frontiers 2021, ref ?). Hence, a fundamental goal of biogeography is to better understand how changes in suitable environmental conditions due to past climatic and geomorphologic history have shaped current biodiversity patterns @bennett\_milankovitch\_1990. Historical biogeography approaches typically investigate species appearance, extinction and migration events without considering the role of demography and dispersal

events in community assembly over time (e.g., Yu, Harris & He, 2010). Meanwhile, population genetics and phylogeographic studies emphasize the joint role of migration, mutation and drift in driving patterns of genetic diversity in space and time (Avice, 2009). Specifically, comparative phylogeographic approaches aim to grasp congruent or differing influence of past historical events led by environmental fluctuations on several co-occurring taxa (Arbogast and Kenagy 2001, Swenson 2019, Overcast et al. 2019, 2020). Concurrently, the neutral theory of biogeography has underlined the role of migration, speciation and drift in shaping patterns of taxonomic diversity in space and time (Hubbell, 2001). Thus, comprehensive and integrative modelling approaches are still needed to address biodiversity dynamics at multiple spatial and temporal scales, specifically regarding how elementary mechanisms driving species coexistence and spatial dynamics in an ecological perspective have influenced over time evolutionary trajectories of species assemblages.

However, most studies investigating the intertwined influence of ecological and evolutive forces on the relative abundances of species and associated eco-evolutive modelling approaches have assumed near-equilibrium dynamics in the past (Hubbell, 2001). The neutral theory of biodiversity itself (Hubbell 2001) defines the fundamental biodiversity number  $\theta$  as equal to  $2 \cdot J_m \cdot \nu$  (with  $J_m$  the metacommunity effective size and  $\nu$  the speciation rate per lineage per generation) which has been estimated for various biomes across the globe (ref?). However, this number assumes that  $J_m$  has been at equilibrium from present backward, until all lineages coalesced into their Most Recent Common Ancestor (MRCA). However,  $J_m$  can only be approximately considered as the harmonic mean of the per-generation  $J_m$  dynamics. Thus by allowing  $J_m$  to vary at specific ages of the past we can relax this equilibrium assumption and account for the discordance of demographic histories between distinct metacommunities. Yet, in the case where past environmental variations (especially climate) have caused the sizes of communities to fluctuate over time, these should have affected ecological drift dynamics. Thus, in the case where these fluctuations occur rapidly compared to the expected time-to-equilibrium of speciation, migration and drift dynamics (for instance with long-lived organisms with slow population dynamics), we expect that current biodiversity patterns should retain the signature of past environmental fluctuations.

*intérêt de rajouter de la vicariance, de la migration avec plusieurs communautés locales ?*

Thus, past environmental fluctuations have influenced the evolutionary trajectory of communities by affecting the local coexistence of species, their migration opportunities, extinction risks and altering speciation pathways in time. Here we propose a novel simulation-based approach in which we consider how assembly dynamics in finite and heterogeneous environments (ecological perspective) affect speciation and extinction dynamics over a long-term (evolutionary perspective), depending on environmental changes over time.

We devise an eco-evolutionary model, resting on the reconstruction of the past co-ancestry of extant individuals, and subsequent grouping of said individuals into distinct species under a model of point mutations and protracted speciation of monophyletic genetic groups. Because, only the ancestry of extant observed individuals is traced back, coalescent methods do not require simulating lineages with no descendants in the present and are thereby much faster than their forward-in-time alternatives (Munoz et al. 2018). Thus, without assuming to which species each individual belongs, we can simulate the genealogies of individuals in an assemblage which can have undergone fluctuations in the size of  $J_m$ , been split into sub-assemblages linked by migration and/or vicariance or any combination of these processes with possible differences in their relative importance (Figure 1) (Hudson 2002, Kelleher and Lohse 2020).

We see multiple advantages to this approach as it may allow for designing *in silico* experiments that have been growing in ecology to address the possible outcome of (meta)community models. Simulating ecological communities and their phylogenetic relations according to different scenarios can help establish a benchmark against which to infer the signatures of community-wide past biogeographic processes from the resulting patterns of taxonomic and phylogenetic diversity. Thus, simulation-based approaches hold great potential for inference with the increase in computation power and the emergence of likelihood-free inference, such as Approximate Bayesian Computation (ABC).

We expose here the logic and advantages of the approach to examine how past community-wide non-equilibrium dynamics have shaped patterns of taxonomic and phylogenetic diversity. The initial version of the method presented here is suited to the simulation of neutral eco-evolutionary dynamics, but future versions will allow further deviation from neutrality. We introduce the method with the *ecophylo* package in Python language which can be called in R to analyse simulated patterns of diversity. The package includes options to simulate large numbers of datasets over broad ranges of parameters and scenarios of past demographic events and fluctuations. These methods are destined to be used alongside ABC methods so as to estimate parameters of past demographic fluctuation from the observation of actual patterns of diversity.

## Simulation algorithm in *ecophylo.simulate*

### Coalescent-based simulation of genealogies

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). Following the Wright-Fisher model, we assume that all individuals shrunk to their haplotypes can reproduce freely within the assemblage, and with equal fitness (neutral assumption). 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). Generations in the model are discrete and non-overlapping, and a single coalescence event can happen at a given generation. We can thus simulate the shared co-ancestry of  $n$  sampled individuals observed at present time in each assemblage using a backward simulation of the coalescent tree. Coalescence is an event such as two lineages at generation  $t$  share the same ancestor at generation  $t-1$ , which defines a bifurcating node in the stochastic genealogy of individuals. Only the ancestry of individuals observed at present is traced back, so that coalescent methods do not require simulating lineages with no extant descendants and are thereby much faster than their forward-in-time alternatives @munoz\_ecolottery\_2018. When  $n \ll J_m(T)$ , the distribution of coalescence times can be approximated as an exponential law with parameter  $\lambda$  proportional to  $1/(2 \cdot J_m(T))$  {Wakeley, 2009}. The topology of the genealogy is thus influenced at each predefined period by the corresponding value of  $J_m(T)$ . Subpops, migration etc... Thus, without assuming to which species each individual belongs, we start by simulating the genealogies of individuals in an assemblage experiencing past demographic fluctuations and/or linked by vicariance or migration events using the ms coalescent simulator (Hudson 2002; Kelleher, 2020).

## Phylogenetic reconstruction and species abundances

We sprinkled speciation events over the simulated genealogies conditionally to their topology and branch lengths. The probability that a mutation occurred on a specific branch of the genealogy, leading to a new variant descendant, followed a Poisson distribution with parameter  $\mu \cdot B$  where  $\mu$  is the point mutation rate and  $B$  is the length of the branch. The descendants stemming from this branch defined a new haplotype clade. Since an extant species should be a monophyletic clade and include haplotypes differing from other species, all paraphyletic clades of haplotypes at present were merged to form a single species. Therefore, monophyletic lineages with distinct genotypes and older than two generations were considered a distinct species (Manceau et al. 2016). We derived thusly the phylogenetic relationships among individuals as well as the number of individuals descending from a speciation event in the genealogy, which defined the species abundance in the sample at present.

## Multi-species eco-evolutionary dynamics

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 using the *ecophylo* package in Python language. We provide examples in R language making use of the functions provided in the package. 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.

## Simulating multi-species past demographic fluctuations

The *ecophylo* package essentially articulates itself around the *simulate* function. This function implements the above-mentioned simulation algorithm and allows users to simulate a phylogeny in Newick format given the desired parameter combinations accounting for the demographic history of Jm.

In the following example, an assemblage of species and their phylogenetic relationships, is simulated assuming that Jm has fluctuated in the past over 3 predefined periods. Users can input the desired time periods (in generations before present) by specifying the *changetime* parameter, yet the *timeframes* function can be used to determine the periods while accounting for loss of resolution in more ancient periods (Boitard et.al 2016 Plos Genetics). Changes in sizes can be instantaneous (by providing a list of past sizes in *past\_sizes*) or gradual if users provide a growth rate for each period of time (by providing a list of growth rates in *parameter name* #TODO: do this!).

```
library(reticulate)
ecophylo <- import('ecophylo')

n <- as.integer(10) # the number of simulated individuals
Jm <- as.integer(500) # the size of the assemblage at present
mu <- as.numeric(0.01) # the point mutation rate

epochs <- ecophylo$timeframes(I= as.integer(3), T= as.integer(10000), a=0.3) # the dates at which the a

sizes <- list(as.integer(1000), as.integer(5000), as.integer(500)) # the different past sizes

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

We can then compute summary statistics on the resulting phylogeny. The *getAbund* function allows us to

146 retrieve the number of individuals descending from a speciation event in the genealogy, thus defining the  
147 species abundance in the sample at present.

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

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

148 • taxonomic diversity metrics:

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

149 ## [1] 6

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

150 ## [1] 1.748067

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

151 ## [1] 0.82

152 • phylogenetic diversity metrics:

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

153 ## [1] 20474.73

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

154 ## [1] 8765.149

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

## [1] 3611.605

These summary statistics can then be used to compare different eco-evolutionary scenarios having yielded different patterns of extant community composition.

## Simulating multi-species population structure and history

*introduce why we're doing this etc*

## Discussion

Studies in community ecology often discount the influence of diversification events, by customarily considering the regional species pool from which communities assemble via rapid local dynamics as a snapshot in evolutionary history. However recent studies have called a fuller integration of biogeographic history into community ecology (Swenson 2019, Overcast et al. 2019, 2020) as well as for a more mechanistic modeling of key processes shaping species pools (Denelle et al. 2019). With the novel simulation algorithm proposed in the *ecophylo* package we explicitly represent how elementary ecological mechanisms such as ecological drift and migration (Vellend 2010) in a finite, heterogeneous and fluctuating environment shape evolutionary trajectories.

## Conclusion

## Data accessibility

The *ecophylo* package is available on <https://github.com/thegreatlizzyator/ecophylo> The *ecophylo* package can be installed by applying the following command in Python and can then be called into R provided the prior installation of *reticulate*

```
# INSTALL ECOPHYLO HERE LOOOOL
```

## Authors' contributions statement

EB conceived the study and built the basic architecture of the *ecophylo* package. EB and MJ did substantial work on adding functionalities, testing and cleaning code. All the authors contributed to setting up the framework and to write the manuscript.



