



## RESEARCH ARTICLE

# *ecolottery*: Simulating and assessing community assembly with environmental filtering and neutral dynamics in R

François Munoz<sup>1</sup>  | Matthias Grenié<sup>2</sup> | Pierre Denelle<sup>2</sup> | Adrien Taudière<sup>2</sup> | Fabien Laroche<sup>2,3</sup> | Caroline Tucker<sup>2,4</sup> | Cyrille Violle<sup>2</sup> 

<sup>1</sup>University Grenoble-Alpes, LECA, Grenoble Cedex 9, France

<sup>2</sup>CEFE UMR 5175, CNRS-Université de Montpellier-Université Paul-Valéry Montpellier-EPHE, Montpellier Cedex 5, France

<sup>3</sup>Irstea, UR EFNO, Centre de Nogent-sur-Vernisson, Nogent-sur-Vernisson, France

<sup>4</sup>Department of Biology, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

**Correspondence**

François Munoz

Email: francois.munoz@cirad.fr

**Funding information**

European Research Council, Grant/Award Number: ERC-StG-2014-639706-CONSTRAINTS; Marie Curie IIF, Grant/Award Number: H2020-MSCA-IF-2014-657951; French Ecole Normale Supérieure

Handling Editor: Ryan Chisholm

**Abstract**

1. We introduce the R package *ecolottery* dedicated to quick and efficient simulation of communities undergoing local neutral dynamics with environmentally filtered immigration from a reference species pool (spatially implicit model). The package includes an Approximate Bayesian Computation (ABC) tool to estimate the parameters of these processes. We present the rationale of the approach and show examples of simulations and ABC analysis.
2. The species in the reference pool differ in their abundances and trait values. Environmental filtering weights the probability of immigration success depending on trait values, while the descendants of established immigrants undergo neutral stochastic drift. The reference pool can be defined in a flexible way as representing, e.g. the composition of a broad biogeographical region, or available dispersers around local communities. The package provides a process-based alternative to the use of randomization-based null models.
3. The package proposes a coalescent-based simulation algorithm that presents significant advantages over alternative algorithms. It does not require simulating community dynamics from an initial state forward in time but does still allow measurement of the influence of environmental filtering. Because of its high calculation speed, this approach allows simulating many communities within a reasonable amount of time.
4. Diverse patterns of taxonomic, functional and phylogenetic compositions can be generated. The package can be used to explore the outcome of ecological and evolutionary processes playing at local and regional scales, and to estimate the parameters of these processes based on observed patterns.

**KEYWORDS**

Approximate Bayesian Computation, coalescent, community assembly, dispersal limitation, environmental filtering, species pool, stochastic and deterministic processes

## 1 | INTRODUCTION

A major aim of community ecology is to identify the influence of ecological processes on species dynamics and coexistence from descriptive features of realized biodiversity patterns (McGill, Enquist, Weiher,

& Westoby, 2006; Weiher & Keddy, 1995). Heuristic assembly rules have been proposed to relate basic ecological drivers to typical features (Weiher & Keddy, 1995), but the interplay of multiple drivers at multiple scales requires more mechanistic and comprehensive understanding of their combined influence (Denny & Benedetti-Cecchi,

2012). Experiments are useful to address the influence of a given driver while controlling for others. However, practical constraints prevent devising experiments at the large spatial and temporal scales at which ecological, biogeographical and evolutionary dynamics are entangled. In this context, a simulation-based approach represents a sound alternative to assess the extent to which a given pattern can be related to a given process (Peck, 2004; Zurell et al., 2010).

Simulating ecological communities according to different scenarios can help to establish a benchmark against which to infer the signatures of assembly processes from resulting patterns of taxonomic, functional or phylogenetic diversity (Botta-Dukát & Czúcz, 2016; Pearse et al., 2015). In silico experiments are developing quickly in ecology and allow researchers to address the possible outcome of (meta)community models (Gravel, Canham, Beaudet, & Messier, 2006; Kembel, 2009; May, Giladi, Ristow, Ziv, & Jeltsch, 2013). With increasing computational power, it becomes possible to simulate many communities across broad ranges of parameter values. Approximate Bayesian Computation (ABC; Csillery, Blum, Gaggiotti, & François, 2010) and Monte Carlo Markov Chain (MCMC) methods allow exploring the outcome of complex models given prior distributions of their parameters, and then determining posterior distributions of the parameters consistent with observed patterns (Hartig, Calabrese, Reineking, Wiegand, & Huth, 2011).

Although simulation-based inference of parameter values from observed patterns of biodiversity is a major opportunity for ecological studies (Gimenez et al., 2014), complex algorithms for community simulation still require intensive and time-consuming computation. Here we introduce a novel simulation toolkit for community and functional ecologists who wish to investigate the signatures of ecological processes in community composition. Each community is assembled by combining immigration from outside the community with stochastic variations in the abundance of the descendants of successful immigrants (ecological drift, Vellend, 2016). The immigrants are drawn from a reference external pool that can represent either a large-scale biogeographic or habitat context (Lessard, Belmaker, Myers, Chase, & Rahbek, 2012), or dispersers available nearby the community (Lessard et al., 2016). Environmental filtering during immigration determines varying success of immigrants depending on the way their ecological attributes allow establishment and survival in a local environmental context (Jabot, Etienne & Chave, 2008; Keddy, 1992). This spatially-implicit framework follows a hierarchical conception of community dynamics embedded in a regional context (Lortie et al., 2004; Ricklefs, 2008), and provides a process-based alternative to randomization-based null models (Gotelli & Graves, 1996; Munoz & Huneman, 2016). It is intended to be quicker and more efficient than alternative algorithms, and is thus suited to intensive simulation schemes. We provide a complementary ABC tool to assess parameters of neutral dynamics and environmental filtering from the composition of real communities. The tool allows addressing whether we can unambiguously relate a given pattern (either taxonomic, functional or phylogenetic) to some hypothesized drivers.

We propose a coalescent-based algorithm of community simulation, which rebuilds the shared ancestry of coexisting individuals in

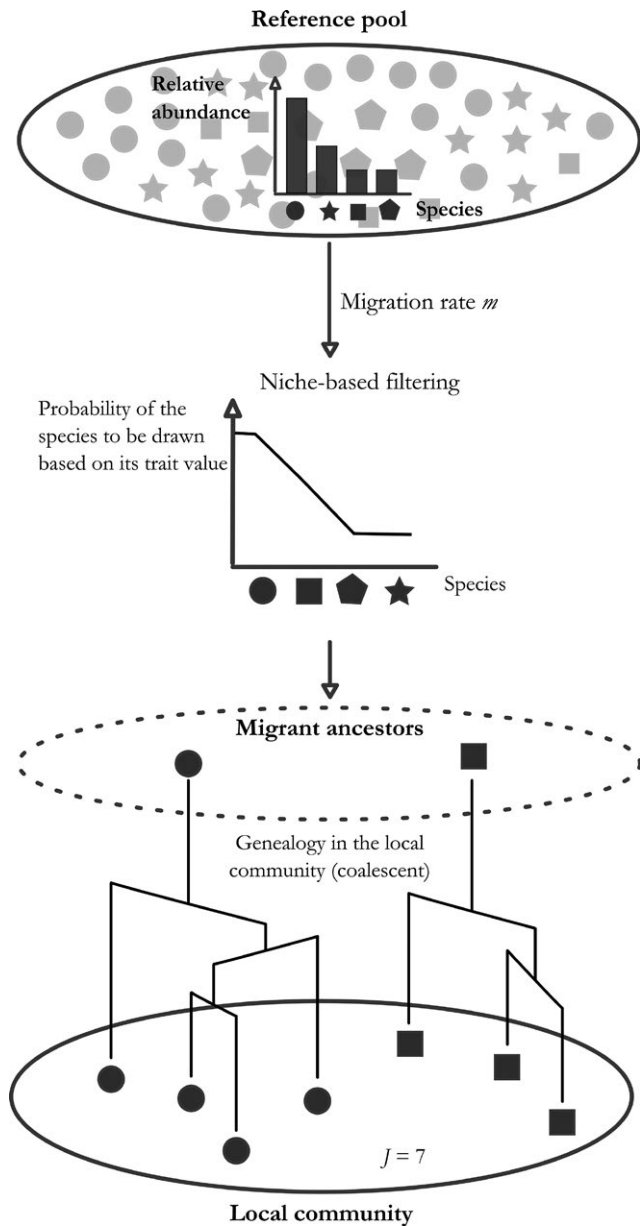
a community. Whether individuals share (or do not share) ancestors backward in time depends on immigration and local stochastic drift (Etienne & Olff, 2004). The genealogy of individuals is also called the *coalescent*. While coalescent-based modelling is popular in population genetics (Wakeley, 2004), it is seldom used in ecology (Etienne & Olff, 2004; Munoz, Couteron, Ramesh, & Etienne, 2007; Munoz, Couteron, & Ramesh, 2008; Munoz, Ramesh, & Couteron, 2014; Rosindell, Wong, & Etienne, 2008). We expose here the logic and advantages of the approach to examine community assembly. It is suited to simulation of neutral communities (Etienne & Olff, 2004; Hubbell, 2001; Munoz et al., 2007) but also allows deviations from neutrality based on niche-dependent immigration biases (Jabot, 2010; Janzen, Haegeman, & Etienne, 2015; Liu & Zhou, 2011; Munoz et al., 2014). Therefore, it allows simulation of ecological communities over a continuum of relative importance of neutral dynamics and environmental filtering.

This sampling scheme was previously implemented in ecology in MATLAB and PARI-GP softwares (Etienne & Olff, 2004; Munoz et al., 2007, 2014), or in C language (Rosindell et al., 2008), and the lack of easy-to-use tools may have limited the application for community simulation. Thus we here introduce the method in R language with the *ecolottery* package. The package also includes an alternative forward-in-time option of community simulation, which generalizes the algorithm of neutral dynamics formerly proposed in the *untb* package (Hankin, 2007) to the context of non-neutral dynamics. Although slower than the coalescent-based algorithm, the forward-in-time option allows exploring more diverse options of niche-based dynamics, such as limiting similarity related to competitive interactions (Abrams, 1983).

## 2 | GENEALOGY OF A LOCAL COMMUNITY: THE FRAMEWORK

The coalescent-based simulation rebuilds the genealogy of individuals in a community by sampling their immigrant ancestors from a reference pool of species while conditioning establishment and survival of their descendants to environmental filtering and neutral drift dynamics (Figure 1). In an initial neutral version of the model akin to Wright-Fisher/Moran models of population genetics, Etienne and Olff (2004) investigated the influence of a single immigration parameter, *I*. This parameter represents the role of successful dispersal, establishment of immigrants, and survival of descendants in determining the composition of the simulated community. Subsequent use of “immigration” will encompass all these components. Etienne and Olff (2004) further considered neutral evolutionary dynamics determining the species abundance distribution in a regional reference pool (or metacommunity) (Munoz et al., 2007, 2008). This approach was later generalized for any user-defined reference pool (Munoz et al., 2008).

In a version with environment filtering, the probability of successful immigration and subsequent survival of descendants also depends on the correspondence between species niche



**FIGURE 1** Analyzing assembly dynamics through the genealogy of individuals in a local community. The reference pool (top) includes individuals with varying ecological niche requirements (different symbols) and varying regional abundances (numbers of each symbol type). The individuals found in a local community (bottom) are the descendants of ancestors that immigrated and established in the past. The coalescent tree represents the genealogy of local individuals back to these migrant ancestors. A pool of ancestors (dashed ellipses) is drawn from the reference pool either randomly (neutral case), or depending on their ability to disperse, establish and survive in the local environment (niche-based filtering)

preferences (or some related functional trait values) and local environmental conditions (Figure 1) (Jabot et al., 2008; Munoz et al., 2014). The probability of drawing an ancestor immigrant of the genealogy from the pool is then weighted according to how well its functional trait values allow successful establishment (symbols in Figure 1). The relative species weights capture fitness differences

due to variation in functional traits among immigrants (Munoz et al., 2014).

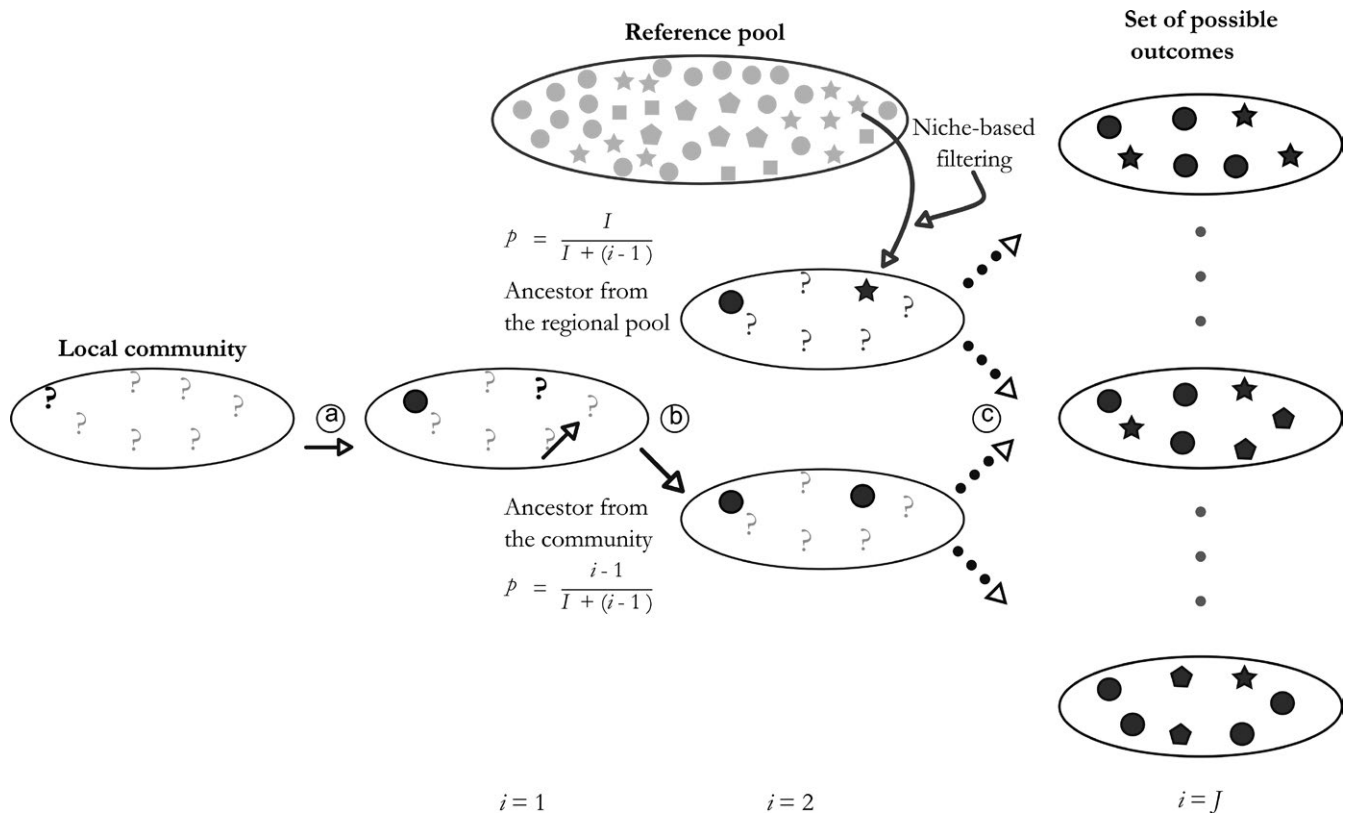
The framework thereby integrates basic dispersal and environmental filtering of immigrants from a reference pool to a local community (Lortie et al., 2004), and stochastic drift dynamics of the descendants. As such, it represents a synthetic approach to jointly investigate environmental filtering and neutral dynamics in community assembly (Gravel et al., 2006; Leibold & McPeck, 2006; Vellend, 2016). An important aspect of the approach is that ecological and evolutionary dynamics can be decoupled in the community and in a regional pool of immigrants, respectively (Munoz et al., 2014). Community dynamics are assumed to be fast enough to neglect speciation events and trait evolution at the local scale; conversely, evolutionary and biogeographical dynamics (e.g. extinction and speciation events) can occur and determine biodiversity patterns at the larger spatial and temporal scales represented by a reference pool. The immigration process is here analogous to a mainland-island model in island biogeography (Hubbell, 2001; MacArthur & Wilson, 1967), where the regional pool undergoes biogeographical and evolutionary dynamics, while the community islands follow migration-drift dynamics with possible environmental filtering.

The reference pool can thus represent a biogeographic context at large spatial scale, but it can also be restricted to species occupying a specific habitat (Kraft, Valencia, & Ackerly, 2008), or to a spatially delimited neighbourhood surrounding the community (Lessard et al., 2016). In the latter case, the coalescent-based framework represents combined influence of species dispersal from a neighbourhood, environmental filtering and drift dynamics in the community. Alternative definitions of the pool can thus be used to address specific hypotheses on the respective roles of biogeography, environmental filtering, and neutral dynamics.

### 3 | SIMULATION ALGORITHM IN COALESC

*coalesec* is the core function in *ecolottery* performing coalescent-based simulation of ecological communities. Let us consider a given pool of species  $s$  with relative abundances  $f_s$ , and a community of size  $J$ . The migration rate is denoted as  $m$ . When an individual dies in the community, it can be replaced by the offspring of one of the  $J - 1$  remaining individuals with probability  $1 - m$ , or by a migrant with probability  $m$ . The immigration parameter  $I = m(J - 1)/(1 - m)$  represents the number of immigrants competing with local offspring to replace a dead individual in the community at each mortality event (Etienne & Olff, 2004).

The simulation algorithm determines which individuals in the sample share the same immigrant ancestor. The simulation begins by drawing a first individual from the reference pool (Figure 2): it is the ancestor of the first individual of the community. The taxonomic identity of this first individual is the identity of its ancestor. The second individual drawn from the community can be (i) either a descendant of the same ancestor as the first individual (in which case the second and the first individuals are conspecific), or (ii) a

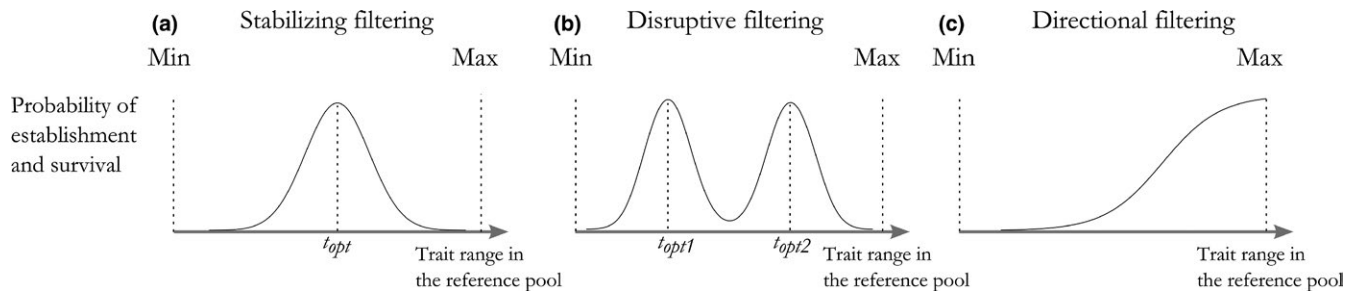


**FIGURE 2** Coalescent-based sampling of a simulated community. Individuals with different trait values have different symbols. The top large ellipse represents the reference pool; the smaller bottom ellipses represent successive stages of coalescent-based sampling to generate a simulated community. In (a), the first individual sampled in the community is the descendant of an immigrant ancestor drawn from the reference pool. In (b), the second sampled individual is either the descendant of the same ancestor, or of another ancestor also drawn from the reference pool. Subsequent draws of individuals are performed with the same two choices of sharing an ancestor with previously sampled individuals, or being the descendant of a new immigrant ancestor, until sampling of  $J$  individuals in (c). Sampling of immigrants from the reference pool can depend on niche-based filtering, i.e., the probability of immigration success is weighted by the fit of species niche requirements to local environment

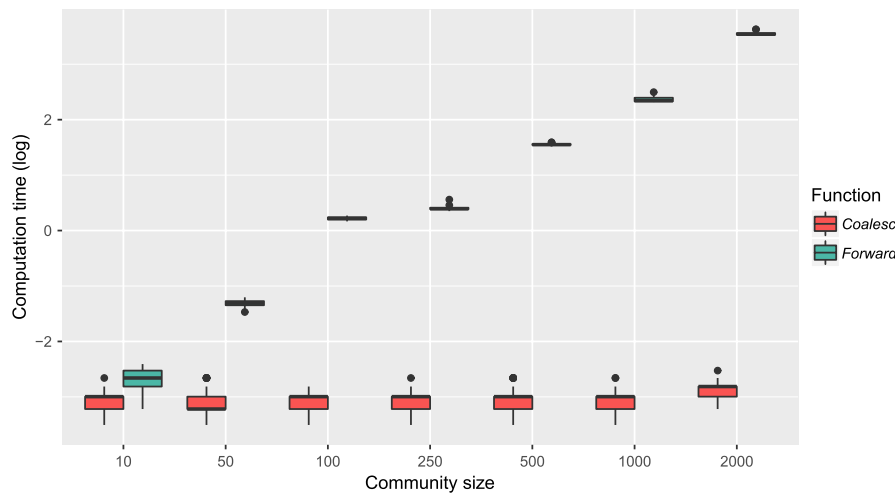
descendant of another immigrant ancestor. The probabilities of (i) and (ii) depend on the immigration parameter,  $I$ . If the descendants of immigrant ancestors have the same probability of survival until present, the probability of (i) is  $1/(I+1)$  and the probability of (ii) is  $I/(I+1)$ . A binomial lottery is used to calculate whether the second individual descends from a new immigrant ancestor. If the ancestor is a new immigrant, it can be drawn from the reference pool and the identity of the second individual can be assigned accordingly. Note that the new immigrant can be, but need not to be, from a different taxon from the first immigrant. The third individual drawn from the community can be the descendant of one of the ancestors of the two first individuals, with probability  $2/(I+2)$ , or the descendant of a new immigrant ancestor, with probability  $I/(I+2)$ . In the former case, we randomly select one of the two first individuals as being conspecific of the third individual. Otherwise, we select a new immigrant ancestor of the pool as before. This process is repeated until reaching the desired number of individuals,  $J$ . Thus, individual  $i$  can be the descendant of one of the ancestors of the  $(i-1)$  first individuals, with probability  $(i-1)/(I+i-1)$ , or of a new immigrant ancestor, with probability  $I/(I+i-1)$ .

With environmental filtering, the probabilities of sampling immigrant ancestors are affected by the match of their niche preferences with local environmental conditions. Let us consider that an immigrant bears a trait value  $t$ , which influences immigration success in the community depending on a filtering function  $filt(t)$ . The same sampling procedure as above is applicable, except that the immigrants are drawn from the pool with relative probability  $filt(t)$ . The higher this probability, the better adapted the immigrant to the local environment, and the greater chance it has to successfully establish in the community (Jabot, 2010; Munoz et al., 2014). Figure 3 illustrates different kinds of environmental filtering depending on the definition of  $filt(t)$ .

The function *coalesc* implements this basic simulation scheme. In the example below, a community of size  $J = 400$  is simulated.  $m = 1$  means that any dead individual in the community is replaced by an immigrant from the pool, while  $m < 1$  means that both immigrants and local offspring can replace a dead individual. The user must also provide information on the composition of the reference pool, by setting either a  $\theta$  value, to simulate a pool of size  $J_{pool}$  with log-series abundance distribution, or explicit species abundances in the pool for *pool*.



**FIGURE 3** Several types of environmental filtering. The probability that an individual immigrates and establishes in a community can depend on its trait value(s) when environmental filtering occurs. Different kinds of filtering are illustrated: (a) a stabilizing environmental filtering around an optimal trait value called  $t_{opt}$ ; (b) a disruptive environmental filtering generating two modes around  $t_{opt1}$  and  $t_{opt2}$ ; (c) a directional environmental filtering with increasing probability of success along a gradient of trait values. Other types of niche-based filtering can be defined using the argument *filt*



**FIGURE 4** Comparison of computation time of alternative coalescent-based (red) and forward-in-time (blue) algorithms in *ecolottery* package. The simulation time in seconds is log-transformed. It is given for a single simulated community related to a given reference pool with log-series distribution of abundances and uniform distribution of trait values between 0 and 1 across species. In forward-in-time simulation, one individual dies at each time step and is replaced by either local offspring or by an immigrant. The migration rate is fixed to  $m = 0.5$  and the immigrants undergo Gaussian environmental filtering around  $t_{opt} = 0.5$ , with standard deviation  $\sigma = 0.1$ . We simulated varying community size (abscissa). Simulation time quickly increased with community size in forward-in-time simulation, while it remained very low in coalescent-based simulation. See Appendix S2 for more details and the complete code

The *pool* object contains at least three columns; an individual index, a species index and the values of one or multiple traits. If trait values are not provided in *pool*, the user can provide them separately, using the *traits* object. In this case, it is assumed that *traits* includes species trait values, without intraspecific variation. If no trait information is provided, species trait values are drawn from a uniform distribution between 0 and 1.

```
res <- coalesc(J = 400, m = 0.5, theta = 50,
             filt = NULL, pool = NULL, traits = NULL)
summary(res)

##      Length Class      Mode
## com   3      data.frame list
## pool  3      data.frame list
```

The resulting object *res* has two components: community composition in *com* and reference pool composition in *pool*. A tutorial in Appendix S1 shows how to simulate neutral dynamics and environmental filtering in diverse situations with *coalesc*, and how to characterize the taxonomic, functional and phylogenetic compositions of simulated communities.

We also provide in *ecolottery* a sister function *forward* to simulate communities using a classical forward-in-time approach. From a given initial community composition, *forward* simulates at each time step a number of mortality events and replacement through local reproduction and immigration from an external pool. The *forward* function thus simulates communities with neutral dynamics and environmental filtering following the spatially-implicit framework of Figure 1. It can also condition the mortality events to trait



differences among the individuals present at each time step, and thus simulate limiting similarity (Abrams, 1983). It extends the scope of the *untb* function of Hankin (2007) to the context of non-neutral dynamics.

In Appendix S2, we compare the computation times of *forward* and *coalesc* for comparable parameterization of immigration and environmental filtering, and for varying community sizes.

Figure 4 shows a comparison of computation times between the two approaches. While calculation time of *coalesc* remains low and increases slowly with community size, calculation time of *forward* increases dramatically and is almost 3,000 times greater than *coalesc* for a community with 2,000 individuals. In addition, checking the convergence to stationarity of *forward* can be challenging, while there is no such issue with *coalesc*, since it simulates community composition at equilibrium (see Appendix S2).

## 4 | INFERENCE OF PARAMETERS USING APPROXIMATE BAYESIAN COMPUTATION

A long-standing objective of community ecology is to infer the ecological processes yielding observed patterns in community composition. A challenging issue is to disentangle the respective influences of neutral and non-neutral dynamics from their combined signatures in local communities. Because it is quick and efficient, *coalesc* can be used to simulate a large number of communities over a broad range of parameters of environmental filtering and neutral dynamics. By performing Approximate Bayesian Computation (ABC, Csillery et al., 2010), we can then compare the structure of an observed community to the simulated communities and determine the plausible parameter values corresponding to the observed community structure. We provide the *coalesc\_abc* function in *ecolottery* to perform such ABC inference in a flexible way. Parallel computing is proposed to improve calculation speed in multi-core computers.

In the following example, a reference community is generated with known parameter values of immigration and environmental filtering. Stabilizing environmental filtering along a Gaussian distribution provides an establishment advantage to immigrants displaying trait values close to an optimum  $t_{opt}$ ; this advantage decreases from this optimum value according to the standard deviation,  $\sigma$ , of the Gaussian curve (see Figure 3).

The ABC analysis is based on simulating a large number of communities using *coalesc*, for varying values of migration rate  $m$ , of filtering parameters  $t_{opt}$  and  $\sigma$ , and for a user-defined reference pool. Summary statistics (calculated by the function *f.sumstats*) are used to characterize resulting patterns of taxonomic and functional composition in communities. Posterior distributions of parameter values for a given community are then derived according to a comparison of the summary statistics from this community with the summary statistics from simulated communities.

```
# Trait-dependent filtering function
filt_gaussian <- function(t, params)
  exp(-(t-params[1])^2/(2*params[2]^2))

# Definition of parameters of environmental filtering
and their range
params <- data.frame(rbind(c(0, 1), c(0.05, 1)))
row.names(params) <- c("topt", "sigmaopt")

# Number of values to sample in prior distributions
nb.samp <- 1000000 # Should be large

# Basic summary statistics
f.sumstats <- function(com) array(dimnames=
  list(c("cwm", "cww", "cws", "cwk", "S", "Es")),
  c(mean(com[,3]), var(com[,3]),
  e1071::skewness(com[,3]),
  e1071::kurtosis(com[,3]),
  vegan::specnumber(table(com[,2])),
  vegan::diversity(table(com[,2]))))

# A reference community is here simulated
(known parameters)
comm <- coalesc(J = 400, m = 0.5, theta = 50,
  filt = function(x) filt_gaussian(x, c(0.2, 0.1)))

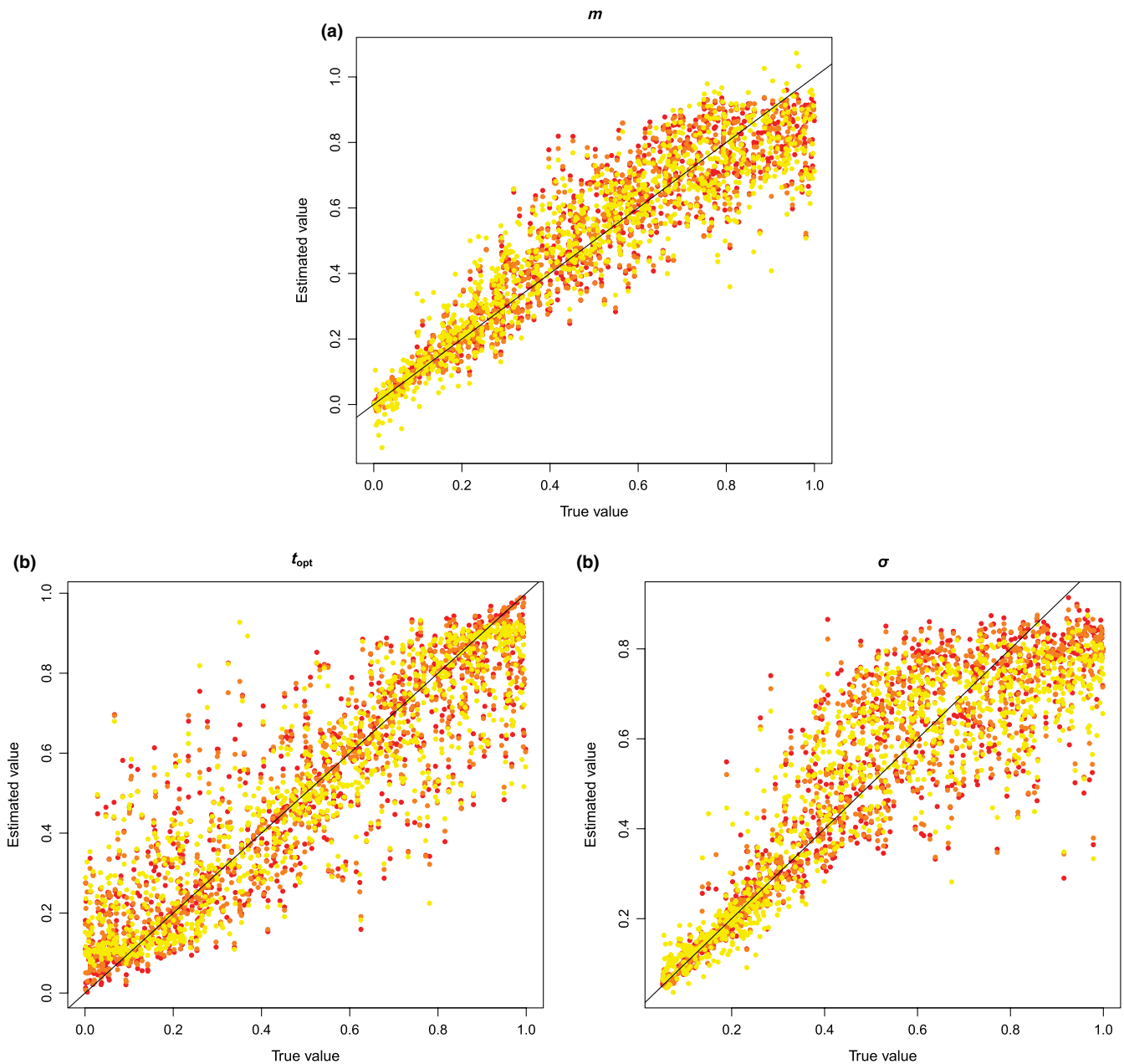
# ABC estimation of the parameters based on summary
statistics of the observed community

# The function makes vary the migration rate, m,
and the parameters of environmental
filtering defined in params
res <- coalesc_abc(comm$com, comm$pool,
  f.sumstats = f.sumstats,
  filt.abc = filt_gaussian,
  params = params, nb.samp = nb.samp,
  parallel = T, tol = 1,
  pkg = c("e1071", "vegan"),
  method = "neuralnet")

plot(res$abc, param=res$par)

The posterior distributions of the parameters can be compared to
the expected values used to simulate the reference community. In ad-
dition, cross-validation can be performed to evaluate the accuracy of
estimation for a broad range of parameter values, e.g. using the cv4abc
function of package abc.

pos.na <- which(rowSums(is.na(res$ss))!=0)
res$cv <- cv4abc(param = res$par[-pos.na,],
  sumstat = res$ss[-pos.na,], nval = 1000,
  tols = c(0.01, 0.1, 1), method = "neuralnet")
```



**FIGURE 5** Cross-validation analysis of simulated communities with limited migration and stabilizing environmental filtering, using the function `plot.cv4abc` function of package `abc`. Neutral (migration rate  $m$  in a) and non-neutral (Gaussian stabilizing filtering with mean  $t_{opt}$  in b and standard deviation  $\sigma$  in c) parameters are estimated using `coalesc_abc`, based on the four first moments of the local trait composition, species richness and Shannon taxonomic diversity. 1,000 validation simulations are considered, expected parameter values are on abscissa and estimated values are on ordinates. Cross-validation is performed for three tolerance levels with increasing value from red to yellow (0.01, 0.1 and 1)

Figure 5 shows the result of cross-validation analysis. All the parameters are estimated reasonably accurately.  $m$  and  $t_{opt}$  are well estimated over the whole range of values between 0 and 1.  $\sigma$  is less well estimated when it becomes large, i.e. in this case a wide environmental filter selects all trait values almost equivalently and may not be discernible from pure neutral dynamics.

In Appendix S3, we also provide an example of ABC analysis to estimate neutral parameters in the Barro Colorado Island rainforest of Panama, with step-by-step explanation of the calculations.

## 5 | DISCUSSION

We propose a coalescent-based framework for quick and efficient simulation of ecological communities assembled through migration and environmental filtering from a regional species pool, and neutral stochastic drift (Vellend, 2016). The `coalesc_abc` function combines the simplicity and speed of coalescent-based simulation with powerful estimation of parameters using Approximate Bayesian Computation (ABC, Csillery et al., 2010). In this regard, it should be useful for a

broad audience of ecologists interested in exploring the linkage of ecological processes to biodiversity patterns and the role of processes in producing observed community composition.

Compared to other available simulation tools, a coalescent-based approach does not simulate the dynamics forward in time, but determines the shared ancestry of individuals sampled in a community at equilibrium. This offers several appealing properties for community simulation: first, we do not need to simulate the dynamics of all individuals that were present through time, but only to determine the lineages that have provided descendants in the observed community. Therefore, the coalescent-based approach is expected to be quicker by several orders of magnitude compared to a forward-in-time approach, as illustrated in Figure 4. Second, the sampling algorithm directly produces a community at a stationary state, so that there is no difficulty related to the choice of the number of generations to simulate or checks for convergence (see Appendix S2). Despite the significant gain of computation time, ABC analysis can still be computationally intensive when increasing the number of parameters and thus the number of simulations required. Therefore, reasonable complexity must be considered in these models.

The approach is in line with the current conceptual synthesis combining neutral and niche-based processes in the study of community dynamics (Munoz & Huneman, 2016; Vellend, 2016). The reference pool considered in simulations can result from neutral speciation-extinction dynamics (default case, Hubbell, 2001), be defined following specific scenarios of trait evolution (e.g. Appendix S1.6), or be based on some predefined flora or fauna information. We can then address how ecological processes driving local community assembly, and biogeographical and evolutionary processes underlying the composition of the reference pool, jointly influence patterns of taxonomic, functional and phylogenetic diversity. We can simulate one or several communities related to the same reference pool, and address patterns of *beta* diversity among communities depending on environmental filtering and neutral dynamics (Munoz et al., 2008, e.g. Appendix S3).

The spatial structure of communities can influence metacommunity dynamics, and this influence cannot be addressed based on a spatially implicit framework (Economio & Keitt, 2007). Recent advances in the modelling of spatially explicit coalescent should allow extending *coalest* to model community assembly in spatially explicit networks (Kelleher, Etheridge, & Barton, 2014). Comparing the outcome of dynamics through the spatially implicit scheme of *ecolottery* to alternative spatially-explicit schemes will help better understand the emergent properties of spatial biodiversity dynamics. Currently it is not clear how reliably a spatially implicit approach can represent the outcome of spatially explicit dynamics and so how local to regional biodiversity dynamics can scale (Morozov & Poggiale, 2012; Munoz, Beeravolu, Pélissier, & Couteron, 2013).

Although adopting a spatially implicit approach, some recent works have proposed defining “process-based species pools” restricted to potential immigrants of a given habitat (Kraft et al., 2008), or to immigrants available from the neighbourhood of communities (Lessard et al., 2016). These alternative pools are considered in randomization schemes to define more or less restrictive null models (Gotelli & Graves,

1996), in order to address the influence of specific processes on simulated patterns. *coalest* can also be used with alternate species pools and alternate parameterizations of environmental filtering and neutral dynamics, to address and test the relative influence of hypothesized drivers. It differs from randomization approaches that shuffle species trait values, site environmental characteristics, or community composition, but do not consider actual parameters of the ecological processes.

Addressing the influence of environmental filtering on species abundances and coexistence is a basic objective of community and functional ecologists (Cornwell & Ackerly, 2009; McGill et al., 2006; Shipley, Vile, & Garnier, 2006). The functions *coalest* and *forward* allow custom definition of environmental filtering, through the input function *filt*. The filtering function can be designed to represent, for instance, stabilizing, directional, or disruptive filtering (Figure 3 and Appendix S1.4), in analogy with the basic types of selective pressures in evolutionary biology (Shipley, 2013). The filtering process represents a weighted lottery drawing immigrants from the reference pool (Lortie et al., 2004), where assigned weights depend on the properties of species (here based on functional traits defined in the *pool* or *traits* input arguments) and local environment (in the example of stabilizing filtering used in ABC estimation above, through the definition of  $t_{opt}$  and  $\sigma$ ) and so can be visualized as the fit of each species' niche to local environmental conditions. The parameters *m* and *filt* of the weighted lottery then represent the respective roles of neutral dispersal limitation and trait-dependent environmental filtering during immigration and establishment.

The definition of environmental filtering is flexible and can integrate the influence of many biological traits (Cadotte & Tucker, 2017; Kraft et al., 2015). Therefore, we must not consider our definition of “trait” in the examples as a specific attribute, but as a more integrative measure of species fitness. Likewise, relating this species' “trait” to a local environment optimum  $t_{opt}$  in the example of environmental filtering represents a synthetic response of species to a set of possibly numerous environmental factors. The user can decompose the basic functional dimensions representing different components of fitness by addressing, for instance, the influence of dispersal traits versus the influence of traits related to local reproduction and competition. In the multi-trait example of Appendix S1.5, we show how to integrate the combined influence of filtering on several traits. The user may consider contrasted filtering operating on different traits, the outcome of this filtering on phylogenetic composition when niche conservatism differs among traits, etc.

Another major objective in community ecology is to address the extent to which limiting similarity and niche differentiation determine species coexistence (Abrams, 1983). It is possible to use *coalest* to model disruptive filtering, i.e. only groups of species with contrasted traits can coexist in the community (see Appendix S1.4, in which we define such a filtering function). This would represent differentiation among ecological guilds, which may reflect, for instance, a vertical stratification of plants in vegetation, and where each guild represents a group of functionally similar species (Vergnon, Dulvy, & Freckleton, 2009). Another specific case of competitive interactions concerns the priority effect (Fukami, 2015). In this case, the order of species



arrival determines community composition, as first arrived species have a competitive advantage and prevent subsequent establishment of other species. Because the logic of priority effect is basically forward-in-time, it represents a case where coalescent-based approach cannot be applied. For this reason, we provide the sister function *forward* to simulate communities forward in time: with input argument *limit.sim* = *T*, community assembly is conditioned to the trait dissimilarity of coexisting species at each time step (Appendix S1.7). A forward-in-time simulation approach is also preferable to address the influence of density dependence, i.e. when species fitness varies over time depending on species abundance variations. Nevertheless, coalescent-based approaches can be designed to address the influence of fitness differences in local community dynamics as, e.g. by applying a rejection algorithm on the output of the coalescent process (Donnelly, Nordborg, & Joyce, 2001).

The user can analyse a number of metrics of taxonomic, functional and phylogenetic diversity in simulated communities (Munoz et al., 2007, 2008, 2014) to test the influence of entangled environmental filtering and neutral dynamics. In the examples of Appendix S1.4, we calculate average trait values (classically known as Community Weighted Mean, CWM, Garnier et al., 2004) in local communities undergoing environmental filtering along an environmental gradient. We also illustrate in Appendix S1.6 how niche conservatism and environmental filtering can produce phylogenetic clustering in the community (Mouquet et al., 2012). However, these approaches often consider *ad hoc* hypotheses about the influence of niche-based processes on these metrics, as, e.g. environmental filtering should decrease the range and variance of local trait variation, while shift of CWM should occur along environmental gradients (Cornwell & Ackerly, 2009). With the interplay of multiple processes driving local composition, how trait-based metrics reflect underlying drivers is expected to be less straightforward. Specifically, we can disentangle the influence of these drivers by performing cross-validation of parameter estimation in ABC estimation with *coalesc\_abc*. As shown in the basic example of ABC estimation with neutral dynamics and Gaussian environmental filtering (Figure 5), cross-validation indicates that we can disentangle the influence of these drivers from summary statistics representing functional composition (the four first moments, CWM, CWV, CWS, CWK; Enquist et al., 2015) and taxonomic composition (here richness and Shannon diversity).

Another crucial issue in community and functional ecology is how intraspecific variation of ecological properties influences community dynamics (Bolnick et al., 2011; Violle et al., 2012). The niche breadth and overlap of coexisting species are basic components of competition theory (MacArthur & Levins, 1967). Coalescent-based simulation of community assembly in *ecolottery* represents individual dynamics based on individual properties, and is therefore suited to address the role of intraspecific trait variation among immigrants, when assuming perfect heritability and no mutation during community dynamics. In Appendix S1.3, an example of community simulation with intraspecific variation is provided. Changes in inter- and intra-specific trait variations are expected between regional and local scales depending on community assembly processes and their comparison allows inferring the influence of these processes (Violle et al., 2012). As the

descendants of immigrants display the same trait values than their ancestor in this approach, we can assess the diversity of phenotypes related to distinct genotypes, which is also appropriate to address the relationship between genetic and specific diversity in communities (Laroche, Jarne, David, Lamy, & Massol, 2015; Vellend & Geber, 2005). Further refinements should also allow addressing the influence of phenotypic plasticity.

## 6 | CONCLUSION

Even though the coalescent-based simulation scheme is a simplified approach to community assembly, it allows straightforward and quick simulation of community composition in diverse situations involving environmental filtering and neutral dynamics. It is suited for intensive simulation schemes, and thus allows ABC estimation of the parameters of assembly dynamics based on observed community composition. Thanks to great flexibility in parameterization of simulations and in the choice of summary statistics for ABC analyses, *ecolottery* will allow in-depth investigation of how the nature and quality of functional and phylogenetic information (e.g. types of phylogenetic and functional distances, structural versus ultrametric type of tree, resolution of the tree, continuous and categorical traits), and how the scale and the composition of reference pools (nature and uncertainty of regional abundance data, inclusion or not of rare species) can affect correct inference of underlying ecological processes. Therefore, *ecolottery* offers options to test theoretical hypotheses on the role of entangled processes, as well as a benchmark for estimating uncertainty in hypothesis testing depending on methodological choices (Perronne, Munoz, Borgy, Reboud, & Gaba, 2017).

## ACKNOWLEDGEMENTS

P.D. and C.V. were supported by the European Research Council (ERC) Starting Grant Project "Ecophysiological and biophysical constraints on domestication of crop plants" (Grant ERC-StG-2014-639706-CONSTRAINTS). C.T. acknowledges the European Commission for the Marie Curie IIF (H2020-MSCA-IF-2014-657951). M.G. received PhD funding from the French Ecole Normale Supérieure.

## AUTHORS' CONTRIBUTIONS

F.M. conceived the study and built the basic architecture of the *ecolottery* package. M.G. and P.D. did substantial work on adding functionalities, testing and cleaning code. All the authors contributed substantially to setting up the framework and to writing the manuscript.

## DATA ACCESSIBILITY

The version 1.0.0 of *ecolottery* is available on CRAN, <https://cran.r-project.org/web/packages/ecolottery/index.html>. The development version is available on Github (<https://github.com/frmunoz/>

ecolottery/tree/master/pkg), and can be installed by using the following function from package *devtools*,

```
devtools::install_github("frmunoz/ecolottery/pkg")
```

## ORCID

François Munoz  <http://orcid.org/0000-0001-8776-4705>

Cyrille Violle  <http://orcid.org/0000-0002-2471-9226>

## REFERENCES

- Abrams, P. (1983). The theory of limiting similarity. *Annual Review of Ecology and Systematics*, 14, 359–376.
- Bolnick, D. I., Amarasekare, P., Araujo, M. S., Burger, R., Levine, J. M., Novak, M., ... Vasseur, D. A. (2011). Why intraspecific trait variation matters in community ecology. *Trends in Ecology and Evolution*, 26, 183–192.
- Botta-Dukát, Z., & Czúcz, B. (2016). Testing the ability of functional diversity indices to detect trait convergence and divergence using individual-based simulation. *Methods in Ecology and Evolution*, 7, 114–126.
- Cadotte, M. W., & Tucker, C. M. (2017). Should environmental filtering be abandoned? *Trends in Ecology & Evolution*, 32, 429–437.
- Cornwell, W. K., & Ackerly, D. D. (2009). Community assembly and shifts in plant trait distributions across an environmental gradient in coastal California. *Ecological Monographs*, 79, 109–126.
- Csillery, K., Blum, M. G. B., Gaggiotti, O. E., & François, O. (2010). Approximate bayesian computation (abc) in practice. *Trends in Ecology & Evolution*, 25, 410–418.
- Denny, M., & Benedetti-Cecchi, L. (2012). Scaling up in ecology: Mechanistic approaches. *Annual Review of Ecology, Evolution, and Systematics*, 43, 1–22.
- Donnelly, P., Nordborg, M., & Joyce, P. (2001). Likelihoods and simulation methods for a class of nonneutral population genetics models. *Genetics*, 159, 853–867.
- Economo, E. P., & Keitt, T. H. (2007). Species diversity in neutral metacommunities: A network approach. *Ecology Letters*, 11, 52–62.
- Enquist, B., Norberg, J., Bonsor, S., Violle, C., Webb, C., Henderson, A., ... Savage, V. (2015). Scaling from traits to ecosystems: Developing a general trait driver theory via integrating trait-based and metabolic scaling theories. *Advances in Ecological Research*, 52, 249–318.
- Etienne, R. S., & Olff, H. (2004). A novel genealogical approach to neutral biodiversity theory. *Ecology Letters*, 7, 170–175.
- Fukami, T. (2015). Historical contingency in community assembly: Integrating niches, species pools, and priority effects. *Annual Review of Ecology, Evolution, and Systematics*, 46, 1–23.
- Garnier, E., Cortez, J., Billes, G., Navas, M. L., Roumet, C., Debussche, M., ... Toussaint, J. P. (2004). Plant functional markers capture ecosystem properties during secondary succession. *Ecology*, 85, 2630–2637.
- Gimenez, O., Buckland, S. T., Morgan, B. J. T., Bez, N., Bertrand, S., Choquet, R., ... de Rexstad, E. (2014). Statistical ecology comes of age. *Biology Letters*, 10, 20140698.
- Gotelli, N. J., & Graves, G. R. (1996). *Null models in ecology*. Washington, DC: Smithsonian Institution.
- Gravel, D., Canham, C. D., Beaudet, M., & Messier, C. (2006). Reconciling niche and neutrality: The continuum hypothesis. *Ecology Letters*, 9, 399–409.
- Hankin, R. (2007). Introducing *untb*, an R package for simulating ecological drift under the unified neutral theory of biodiversity. *Journal of Statistical Software*, 22, 1–15.
- Hartig, F., Calabrese, J. M., Reineking, B., Wiegand, T., & Huth, A. (2011). Statistical inference for stochastic simulation models – Theory and application. *Ecology Letters*, 14, 816–827.
- Hubbell, S. (2001). *The unified neutral theory of biodiversity and biogeography*. Princeton, NJ and Oxford, UK: Princeton University Press.
- Jabot, F. (2010). A stochastic dispersal-limited trait-based model of community dynamics. *Journal of Theoretical Biology*, 262, 650–661.
- Jabot, F., Etienne, R. S., & Chave, J. (2008). Reconciling neutral community models and environmental filtering: Theory and an empirical test. *Oikos*, 117, 1308–1320.
- Janzen, T., Haegeman, B., & Etienne, R. S. (2015). A sampling formula for ecological communities with multiple dispersal syndromes. *Journal of Theoretical Biology*, 374, 94–106.
- Keddy, P. A. (1992). Assembly and response rules – 2 goals for predictive community ecology. *Journal of Vegetation Science*, 3, 157–164.
- Kelleher, J., Etheridge, A. M., & Barton, N. H. (2014). Coalescent simulation in continuous space: Algorithms for large neighbourhood size. *Theoretical Population Biology*, 95, 13–23.
- Kembel, S. W. (2009). Disentangling niche and neutral influences on community assembly: Assessing the performance of community phylogenetic structure tests. *Ecology Letters*, 12, 949–960.
- Kraft, N. J. B., Adler, P. B., Godoy, O., James, E., Fuller, S., & Levine, J. M. (2015). Community assembly, coexistence, and the environmental filtering metaphor. *Functional Ecology*, 29, 592–599.
- Kraft, N. J. B., Valencia, R., & Ackerly, D. D. (2008). Functional traits and niche-based tree community assembly in an Amazonian forest. *Science*, 322, 580–582.
- Laroche, F., Jarne, P., David, P., Lamy, T., & Massol, F. (2015). A neutral theory for interpreting correlations between species and genetic diversity in communities. *The American Naturalist*, 185, 59–69.
- Leibold, M. A., & McPeck, M. A. (2006). Coexistence of the niche and neutral perspectives in community ecology. *Ecology*, 87, 1399–1410.
- Lessard, J.-P., Belmaker, J., Myers, J. A., Chase, J. M., & Rahbek, C. (2012). Inferring local ecological processes amid species pool influences. *Trends in Ecology & Evolution*, 27, 600–607.
- Lessard, J.-P., Weinstein, B. G., Borregaard, M. K., Marske, K. A., Martin, D. R., McGuire, J. A., ... Graham, C. H. (2016). Process-based species pools reveal the hidden signature of biotic interactions amid the influence of temperature filtering. *The American Naturalist*, 187, 75–88.
- Liu, J., & Zhou, S. (2011). Asymmetry in species regional dispersal ability and the neutral theory. *PLoS ONE*, 6, e24128.
- Lortie, C. J., Brooker, R. W., Choler, P., Kikvidze, Z., Michalet, R., Pugnaire, F. I., & Callaway, R. M. (2004). Rethinking plant community theory. *Oikos*, 107, 433–438.
- MacArthur, R., & Levins, R. (1967). The limiting similarity, convergence, and divergence of coexisting species. *The American Naturalist*, 101, 377–385.
- MacArthur, R., & Wilson, E. (1967). *The theory of island biogeography*. Princeton, NJ: Princeton University Press.
- May, F., Giladi, I., Ristow, M., Ziv, Y., & Jeltsch, F. (2013). Metacommunity, mainland-island system or island communities? Assessing the regional dynamics of plant communities in a fragmented landscape. *Ecography*, 36, 842–853.
- McGill, B. J., Enquist, B. J., Weiher, E., & Westoby, M. (2006). Rebuilding community ecology from functional traits. *Trends in Ecology & Evolution*, 21, 178–185.
- Morozov, A., & Poggiale, J.-C. (2012). From spatially explicit ecological models to mean-field dynamics: The state of the art and perspectives. *Ecological Complexity*, 10, 1–11.
- Mouquet, N., Devictor, V., Meynard, C. N., Munoz, F., Bersier, L.-F., Chave, J., ... Thuiller, W. (2012). Ecophylogenetics: Advances and perspectives. *Biological Reviews*, 87, 769–785.
- Munoz, F., & Huneman, P. (2016). From the neutral theory to a comprehensive and multiscale theory of ecological equivalence. *The Quarterly Review of Biology*, 91, 321–342.
- Munoz, F., Beeravolu, C. R., Pélissier, R., & Couteron, P. (2013). Do spatially-implicit estimates of neutral migration comply with seed dispersal data in tropical forests? *PLoS ONE*, 8, e72497.

- Munoz, F., Couteron, P., & Ramesh, B. (2008). Beta-diversity in spatially implicit neutral models: A new way to assess species migration. *The American Naturalist*, 172, 116–127.
- Munoz, F., Couteron, P., Ramesh, B., & Etienne, R. S. (2007). Estimating parameters of neutral communities: From one single large to several small samples. *Ecology*, 88, 2482–2488.
- Munoz, F., Ramesh, B. R., & Couteron, P. (2014). How do habitat filtering and niche conservatism affect community composition at different taxonomic resolutions. *Ecology*, 95, 2179–2191.
- Pearse, W. D., Cadotte, M., Cavender-Bares, J., Ives, A. R., Tucker, C., Walker, S., & Helmus, M. R. (2015). Pez: Phylogenetics for the environmental sciences. *Bioinformatics*, 31, 2888–2890.
- Peck, S. L. (2004). Simulation as experiment: A philosophical reassessment for biological modeling. *Trends in Ecology & Evolution*, 19, 530–534.
- Perronne, R., Munoz, F., Borgy, B., Reboud, X., & Gaba, S. (2017). How to design trait-based analyses of community assembly mechanisms: Insights and guidelines from a literature review. *Perspectives in Plant Ecology, Evolution and Systematics*, 25, 29–44.
- Ricklefs, R. E. (2008). Disintegration of the ecological community. *The American Naturalist*, 172, 741–750.
- Rosindell, J., Wong, Y., & Etienne, R. S. (2008). A coalescence approach to spatial neutral ecology. *Ecological Informatics*, 3, 259–271.
- Shipley, B. (2013). *From plant traits to vegetation structure: Chance and selection in the assembly of ecological communities*. Cambridge, UK: Cambridge University Press.
- Shipley, B., Vile, D., & Garnier, E. (2006). From plant traits to plant communities: A statistical mechanistic approach to biodiversity. *Science*, 314, 812–814.
- Vellend, M. (2016). *The theory of ecological communities (mpb-57)*. Princeton, NJ: Princeton University Press.
- Vellend, M., & Geber, M. A. (2005). Connections between species diversity and genetic diversity. *Ecology Letters*, 8, 767–781.
- Vernon, R., Dulvy, N. K., & Freckleton, R. P. (2009). Niches versus neutrality: Uncovering the drivers of diversity in a species-rich community. *Ecology Letters*, 12, 1079–1090.
- Vielle, C., Enquist, B. J., McGill, B. J., Jiang, L., Albert, C. H., Hulshof, C., ... Messier, J. (2012). The return of the variance: Intraspecific variability in community ecology. *Trends in Ecology & Evolution*, 27, 244–252.
- Wakeley, J. (2004). *Coalescent theory – An introduction*. Greenwood Village, CO: Roberts & Company Publishers.
- Weiher, E., & Keddy, P. A. (1995). Assembly rules, null models, and trait dispersion – New questions front old patterns. *Oikos*, 74, 159–164.
- Zurell, D., Berger, U., Cabral, J. S., Jeltsch, F., Meynard, C. N., Munkemüller, T., ... Schröder, B. (2010). The virtual ecologist approach: Simulating data and observers. *Oikos*, 119, 622–635.

## SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

**How to cite this article:** Munoz F, Grenié M, Denelle P, et al. *ecolottery*: Simulating and assessing community assembly with environmental filtering and neutral dynamics in R. *Methods Ecol Evol*. 2017;00:1–11. <https://doi.org/10.1111/2041-210X.12918>