

APPLICATION

miaSim: an R/Bioconductor package to easily simulate microbial community dynamics

Yu Gao^{1,2}  | Yağmur Şimşek³ | Emma Gheysen² | Tuomas Borman³ | Yi Li¹ |
Leo Lahti³  | Karoline Faust²  | Daniel Rios Garza² 

¹Key Laboratory of Integrated Regulation and Resource Development on Shallow Lakes, Ministry of Education, College of Environment, Hohai University, Nanjing, People's Republic of China

²Laboratory of Molecular Bacteriology (Rega Institute), Department of Microbiology, Immunology and Transplantation, KU Leuven, Leuven, Belgium

³Department of Computing, University of Turku, Turku, Finland

Correspondence

Karoline Faust
Email: karoline.faust@kuleuven.be

Daniel Rios Garza
Email: danielriosgarza@gmail.com

Funding information

Academy of Finland, Grant/Award Number: 330887; China Scholarship Council, Grant/Award Number: 202006710066; H2020 European Research Council, Grant/Award Number: 801747 and 952914

Handling Editor: Steven Kembel

Abstract

1. Microbiomes never stop changing. Their compositions and functions are shaped by the complex interplay of intrinsic and extrinsic drivers, such as growth and migration rates, species interactions, available nutrients and environmental conditions. Mathematical models help us make sense of these complex drivers and intuitively explain how, why and when specific microbiome states are reached while others are not. To make simulations of microbiome dynamics intuitive and accessible, we present miaSim.
2. miaSim provides users with a wide range of possibilities to match their specific assumptions and scenarios, starting from a core implementation of four widely used models (namely the stochastic logistic model, MacArthur's consumer-resource model, Hubbell's neutral model and the generalized Lotka-Volterra model) and several of their derivations. The diverse model implementations share the same data structures and, whenever possible, share state variables, which significantly facilitates cross-model combinations and comparisons.
3. We combined and simulated some published examples of microbiome models in miaSim and performed cross-model comparisons and tested diverse model assumptions. Our examples illustrate the reliability, robustness and user-friendliness of the package. In addition, miaSim is accompanied by miaSimShiny, which allows users to explore the parameter space of their models in real-time in an intuitive graphical interface. miaSim is fully compatible with the 'miaverse', an R/Bioconductor framework for microbiome data science, allowing users to combine and compare model simulations with microbiome datasets. The stable version of miaSim is available through Bioconductor 3.15, and the version for future development is available at <https://github.com/microbiome/miaSim>. miaSimShiny is available at <https://github.com/gaoyu19920914/miaSimShiny>.
4. We anticipate that miaSim will significantly facilitate the task of simulating microbiome dynamics, highlighting the role of ecological simulations as important tools in microbiome data science.

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2023 The Authors. *Methods in Ecology and Evolution* published by John Wiley & Sons Ltd on behalf of British Ecological Society.

KEYWORDS

community dynamics, ecological interactions, microbial community, microbial ecology, modelling tool

1 | INTRODUCTION

Microorganisms are key players in the biosphere and greatly impact human life, environment and society (Arumugam et al., 2011; Logan & Rabaey, 2012; Natarajan & Bhatt, 2020; Patowary et al., 2016; Toju et al., 2018). The rapid development of sequencing and computing technologies contributed to an increase in our understanding of the structure and function of microbiomes in diverse environments, including laboratory incubations, human body sites, activated sludge in wastewater treatment plants, and river beds (Huttenhower et al., 2012; Khazaei et al., 2020; Li et al., 2020; Peces et al., 2022). However, microbiomes are inherently complex and dynamic systems whose trajectories depend on many interactions between microbes and their environments (Kehe et al., 2021; Palmer & Foster, 2022).

Mathematical models help us make sense of these complex systems, by providing a simplified explanatory framework for their underlying dynamics (Bengtsson-Palme, 2020; Coenen et al., 2020; Song et al., 2014). Over the years, a variety of microbiome models have been proposed that account for particular characteristics of the system (briefly summarized in Table 1). Of note, here we focus on the 'well-mixed' models. For microbiome models that account for spatial structure, see the works of Bauer et al. (2017), Dukovski et al. (2021), Goldschmidt et al. (2021) and Hellweger et al. (2016).

With the help of these diverse models, a number of ecological insights have been obtained by fitting and simulating experimental data. For example, macroecological laws of variation and diversity (Gamma abundance fluctuation distribution, Taylor's law and the mean abundance distribution of different species in microbial communities) were reproduced with a stochastic logistic model (Grilli, 2020), while the dynamics of bacteria that inhibit *Clostridium difficile* infection in the human gut microbiome were successfully predicted with a generalized Lotka-Volterra model (Buffie et al., 2015), and properties of large-scale microbiome data sets, such as the relationship between richness and environmental harshness, as well as abundance distributions, were reproduced with a consumer-resource model (Marsland, Cui, & Mehta, 2020). Recent simulation studies have also explored the effect of ecological memory on microbial communities (Khalighi et al., 2022).

Modelling formalisms focus on different factors shaping microbial community dynamics, but the patterns that these models intend to explain generally reflect diverse eco-evolutionary processes. In some cases, combining and comparing different models is helpful for understanding community dynamics (Greenblum et al., 2013) or for investigating the possible origin of particular properties, for example memory in time-series data (Faust et al., 2018). Furthermore, different model formalisms imply different generative processes. For example, simulations of generalized Lotka-Volterra models with random interaction matrices were used to illustrate that communities

with numerous interacting species or communities with a few strongly cooperative species tend to be unstable (Coyte et al., 2015; May, 1972), while a consumer-resource model revealed that large communities may be stable when resources are explicitly considered (Butler & O'Dwyer, 2018). These conflicting observations emerge from the simulation of community dynamics with various formalisms, illustrating the potential benefit of interpreting microbiome data in the light of diverse models. Currently, cross-model comparison is limited since the different published models depend on knowledge of their specific interfaces and programming languages.

To fill in this gap, we developed miaSim, an R/Bioconductor package that facilitates the task of modelling and simulation of microbiome time series. Here, we provide detailed explanations of miaSim's models, functions, parameters, and data structures. Moreover, we present case studies using miaSim and its R-Shiny extension miaSim-Shiny, which illustrate how these tools allow researchers to computationally explore a variety of questions in microbial ecology.

2 | miaSim R PACKAGE

2.1 | An overview of the package

The workflow of miaSim is described in Figure 1. It consists of preprocessing, simulating and postprocessing steps. Compatibility between different models is facilitated by sharing common names, parameters and data structures. The common data structures also facilitate the downstream comparison and visualization of simulations.

A minimal number of user inputs are required to start simulating the models (Figure 1). Most of the parameters have default values. For instance, to generate a simulation of the generalized Lotka-Volterra model, we only need to specify the number of species (using the `ExampleGLV <- simulateGLV(n_species = 4)` command). miaSim will automatically generate a random interaction matrix and simulate the model. The time series may then be visualized with downstream tools in R/Bioconductor, for instance using `miaViz::plotSeries(ExampleGLV)`. While most models require only the number of species or resources to be specified, the user has control over all the parameters listed in Table 2.

Different community models (e.g. the generalized Lotka-Volterra model, the consumer-resource model, etc.) are implemented in the package, and examples of their practical application are provided in separate reproducible vignette notebooks that are available through the package's website (Şimşek et al., 2023). Pre-processing functions with specific controlling variables (e.g. generator of inter-species interaction matrices based on different types, ratios and strengths of interactions) and post-processing functions with different objectives (e.g. converting function and other utilities) are also written separately

TABLE 1 Microbiome models. Exponential growth and logistic growth models describe population dynamics, based on simple phenomenological parameters such as growth rates and carrying capacities. Models based on the generalized Lotka-Volterra model go further by explicitly considering species interactions (Lotka, 1925; May, 1972; Volterra, 1926). Kinetic models generally use differential equations to explicitly describe microorganism growth in terms of metabolite consumption and production (D'hoë et al., 2018; Monod, 1949) and changes in environmental conditions such as pH (Kettle et al., 2018). Consumer-resource models are particular examples of kinetic models that use simplified assumptions for resource dynamics (Chesson, 1990). Hubbell's unified neutral theory of biodiversity describes the dynamics at the per capita level where individuals have equal rates of birth, death, speciation and migration (Hubbell, 2001). Hubbell's neutral model is equivalent to the hierarchical Dirichlet process (HDP) when population sizes are large (Harris et al., 2017). Other important models such as the Genome-scale metabolic models and agent-based models are beyond the scope of this study.

Model formalisms	Features	Implementations
Exponential growth model	Growth of non-interacting single species	growthrates (Petzoldt, 2020)
Stochastic logistic model	Growth with maximum capacity/rate limit optional stochasticity	growthcurver (Sprouffske & Wagner, 2016), stochastic logistic model (Descheemaeker & de Buyl, 2020)
Generalized Lotka-Volterra model (GLV)	Inter-species interactions	web-gLV (Kuntal et al., 2019) MDSINE (Bucci et al., 2016)
Kinetic model	Changes in resources (nutrients/metabolites)	Community Simulator (Marsland, Cui, Goldford, et al., 2020), rescomp (Letten, 2022) microPop (Kettle et al., 2018), Micodymora (Delattre et al., 2019)
Hubbell's neutral model	Identical birth/death probabilities for individuals Migration from metacommunity	untb (Hankin, 2019) Hierarchical Dirichlet process (HDP) (Harris et al., 2017)

to keep the structure of the package simple and extensible. All functions contain detailed explanations and practical examples.

Most models return the simulation as a *TreeSummarizedExperiment* data object (Huang et al., 2021) containing several elements, including the simulation results, initial states and parameters (Figure 2). As detailed in the online tutorial, there are also functions available to extract desired time points from a simulation or to combine results across multiple simulations. Furthermore, there are cases where a large number of repeated simulations are needed. For example, when performing parameter sweeps, *miaSim*'s execution time may be improved by using one of R's parallel looping constructs to execute independent simulations in parallel (see Supporting Information File S1). Execution times are only bottlenecks in simulations with a very large number of species, as illustrated by the execution times of *miaSim*'s Hubbell and GLV models compared to a MATLAB implementation (Supporting Information File S1). In general, the shared data container facilitates the combination of simulations and opens up the possibility to treat simulated data in batches. Further examples and functionality are provided in the online vignette at <https://microbiome.github.io/miaSim> (Şimşek et al., 2023).

2.2 | Mathematical descriptions of the *miaSim* models

2.2.1 | Logistic model

The logistic model describes the growth of populations without interaction terms. We based the model on the Verhulst equation (Bacaër, 2011), adding terms for migration rates (m_i) and stochastic noise (σ):

$$\frac{dN_i}{dt} = r_i N_i \left(1 - \frac{N_i}{K_i} \right) + m_i + \sigma,$$

where N_i , r_i , and K_i are, respectively, the density, specific growth rate and carrying capacity of the i th population. The user can also simulate a logistic model that contains a nonrenewable resource by specifying death rates. In this case, similar to common setups in experimental microbiology, the community will behave like a batch culture where all species die when resources run out.

2.2.2 | Generalized Lotka-Volterra model

The Generalized Lotka-Volterra (GLV) model describes the growth of interacting populations. Interactions are encoded by adding a matrix (A) with positive and negative interaction coefficients. The *randomA* function generates random interaction matrices with a variety of possible distributions and assumptions, including the choice for a defined proportion of interaction types (e.g. competition, mutualism, commensalism, parasitism and amensalism), random or power-law distributions for the interaction strengths, and the possibility to structure the matrix into subgroups of strongly interacting species. Following the formulation of Fisher and Mehta (Fisher & Mehta, 2014), we define the GLV as

$$\frac{dN_i}{dt} = r_i N_i + \sum_j a_{ij} N_i N_j + m_i + \sigma,$$

where a_{ij} is the interaction coefficient of the j th species on the i th species. By default, the diagonal values of A are fixed to -0.5 , indicating intraspecific competition for habitat, nutrients, etc. A discrete version is also available based on the Ricker model as described by Fisher and Mehta (Fisher & Mehta, 2014).

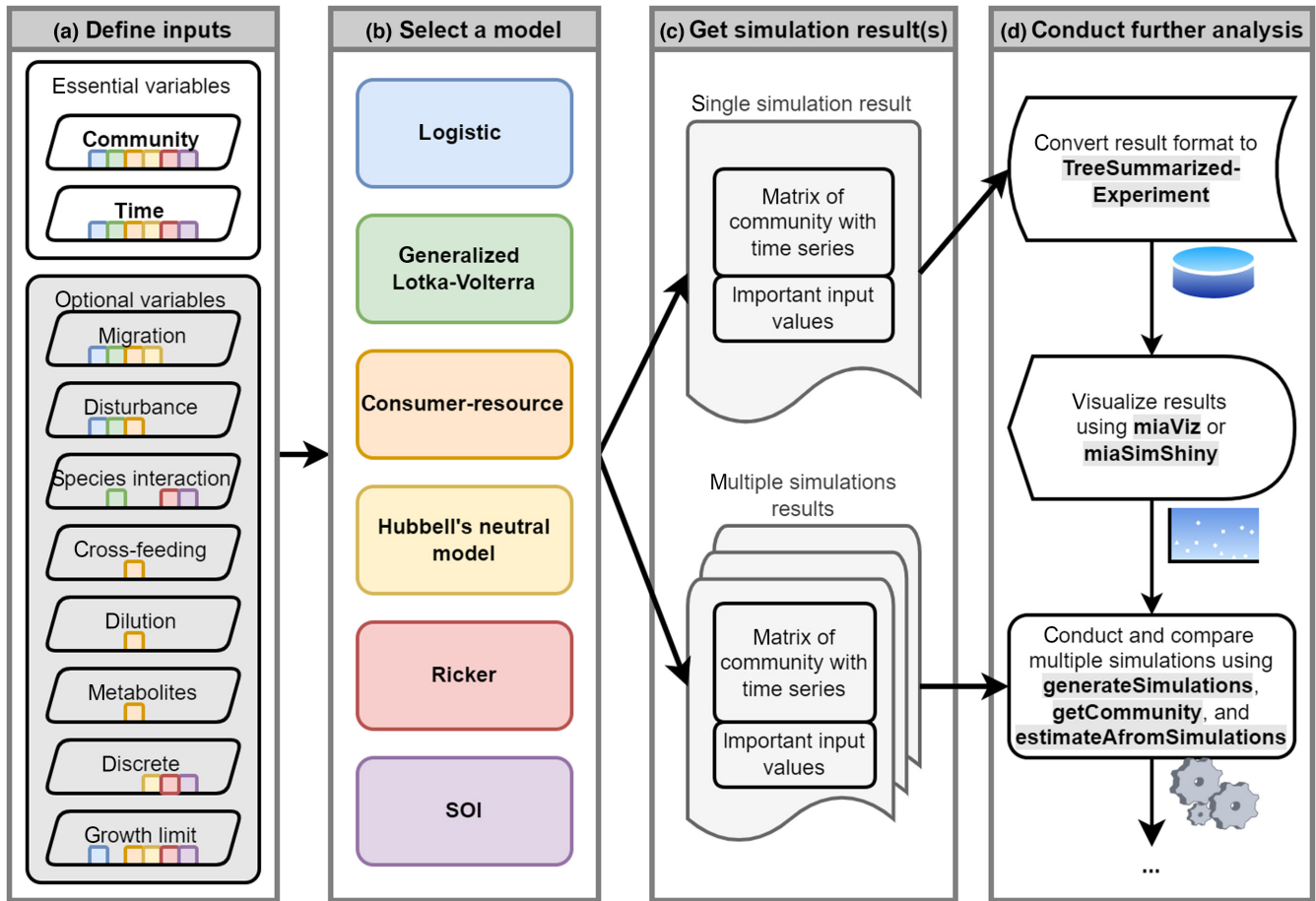


FIGURE 1 Flowchart of modelling in miaSim. Arrows are possible flows of data, and blocks are available data/models/modules. (a) To start using miaSim, essential variables should be defined (either the number of species or resources and the total time to simulate), while optional variables (e.g. a list of metabolites and their quantities for consumer-resource model, a matrix describing interspecies interactions for generalized Lotka-Volterra model, etc.) will be generated if not given. Coloured blocks near each input indicate their applicability to different models in the next step. (b) A model should be selected among all available models (including the logistic model, the generalized Lotka-Volterra model, the consumer-resource model, Hubbell's neutral model, the Ricker model and the self-organized instability model). (c) The results containing simulation data from different models will be stored in an identical format for one single simulation or a series of different simulations. (d) Finally, data are returned in the TreeSummarizedExperiment format, which can be visualized using miaViz or miaSimShiny. In addition, with the evaluation of modelling results, some parameters can be adjusted, generating a new single simulation or a series of multiple simulations using alternative models. A series of multiple simulations can be generated by a single call to the “generateSimulations” function.

2.2.3 | Consumer-resource model

Based on the original formulation of MacArthur (MacArthur, 1970), the consumer-resource model describes the growth rates of each species i as a linear combination of their consumed metabolites. This model explicitly accounts for metabolite concentrations. Since growth rates depend on resource intake rate, we model metabolite uptake with a Monod functional form. Additionally, the model allows microbes to secrete metabolic by-products that may be used by other species.

The resource stoichiometries are defined by a matrix E , which can be generated using the function `randomE`. This function allows users to control several parameters, such as the number of metabolites to consume or produce, the ratio between maintenance and consumption, and trophic preferences for species or subgroups of species (e.g. Wang et al., 2019).

The following differential equation defines the growth rates of species i :

$$\frac{dN_i}{dt} = N_i \left(r_i \sum_{j=1}^k e_{ij} f_{ij} - \phi \right),$$

where r_i is the maximum growth rate of species i , while N_i is its abundance at time t . e_{ij} is the yield of substrate j for species i , f_{ij} is the feeding form, which constrains the quantity of substrate that can be consumed at each time step, and ϕ is a dilution term.

The function uses the Monod equation as the feeding form $f_{ij} = \frac{S_j}{k_{ij} + S_j}$, where k_{ij} is the Monod constant for species i of substrate j and S_j is the concentration of substrate j at time t .

Resources change according to the following differential equation:

$$\frac{dS_j}{dt} = \phi(S_{j0} - S_j) - \sum_{i=1}^n e_{ij} f_{ij} N_i + \sum_{i=1}^n p_{ij} N_i \left(\sum_{j=1}^k e_{ij} f_{ij} \right),$$

TABLE 2 Important input parameters (parameter types corresponding to the classification of inputs in Figure 1a). A detailed table of parameters can be found in Supporting Information Table S1 and in the documentation of each function.

Name	Description	Type
n_species	Number of species	Community
names_species	Names of species	Community
x0	Initial abundances of simulated species	Community
growth_rates	Growth rates of simulated species	Community
t_start	Starting time point of the simulation	Time
t_end	Ending time point of the simulation	Time
t_step	Time interval of the simulation	Time
t_store	Number of time points to return	Time
sigma_migration	Intensity of migration at each time point	Migration
migration_p	Probability of migration from metacommunity	Migration
metacommunity_probability	Probabilities of each species migrating from the metacommunity	Migration
stochastic	Whether to introduce noise/stochasticity in the simulation	Disturbances
sigma_drift	Noise strength at each time point	Disturbances
epoch_p	Probability of random periodic changes	Disturbances
sigma_epoch	Strength of random periodic changes	Disturbances
sigma_external	Strength of external disturbances	Disturbances
t_external_events	Starting time points of external disturbances	Disturbances
t_external_durations	Durations of external disturbances	Disturbances
A	A matrix defining the inter-species interactions	Inter-species interaction
E	A matrix defining relations between species and resources	Cross-feeding
resources_dilution	Quantities of resources in diluting resources	Dilution
inflow_rate	Inflow rate of continuous cultivation	Dilution
outflow_rate	Outflow rate of continuous cultivation	Dilution
volume	Volume of continuous cultivation	Dilution
n_resources	Number of resources	Metabolites
names_resources	Names of resources	Metabolites
resources	Initial quantities of resources	Metabolites

TABLE 2 Continued

Name	Description	Type
k_events	Number of events to simulate before updating the sampling distributions	Discrete
monod_constant	Monod growth constant for n_species with n_resources	Growth limit
carrying_capacities	Max populations of species in logistic model	Growth limit
death_rates	Death rates of each species	Growth limit
error_variance	Variance of measurement error	Miscellaneous
norm	Whether to normalize the simulated results to relative abundances	Miscellaneous

where ϕ is a dilution term, S_{j0} is the concentration of substrate j in the feed, and p_{ij} is the yield of the production of substrate j that results from the growth of species i .

In practice, a species either consumes or produces a metabolite (or is indifferent to its presence) allowing us to summarize e_{ij} and p_{ij} in a single matrix (E , containing the same number of rows as species in the community and the same number of columns as the number of metabolites that are used and produced). To distinguish them, the entries of E are, respectively, positive and negative for the consumed and produced substrates.

2.2.4 | Hubbell neutral model

Hubbell's neutral model treats individuals as functionally indistinguishable units. Community dynamics are simulated by sampling from per capita probabilities of birth, death and migration (Hubbell, 2001). Our implementation is optimized by a chemical stochastic simulation algorithm (Gillespie, 1976), where multiple events may be simulated at a single time point (Cai & Xu, 2007). In the model, the total number of individuals is fixed, so the number of new births or immigrations sum up to the number of deaths during simulation steps. Additionally, users may define species-specific growth rates, which result in nonneutral dynamics but retain the demographic noise derived from using a per capita-based approach (Constable et al., 2016).

2.2.5 | Self-organized instability model

The self-organized instability (SOI) model combines the species interactions of the generalized Lotka-Volterra model with the immigration and extinction properties of Hubbell's neutral model (Solé et al., 2002). At a particular time step in the dynamics of the community, the following events occur in order: (i) immigration. The colonization of an empty site by species i according to the probability of immigration μ_i , (ii) death. A site occupied by species

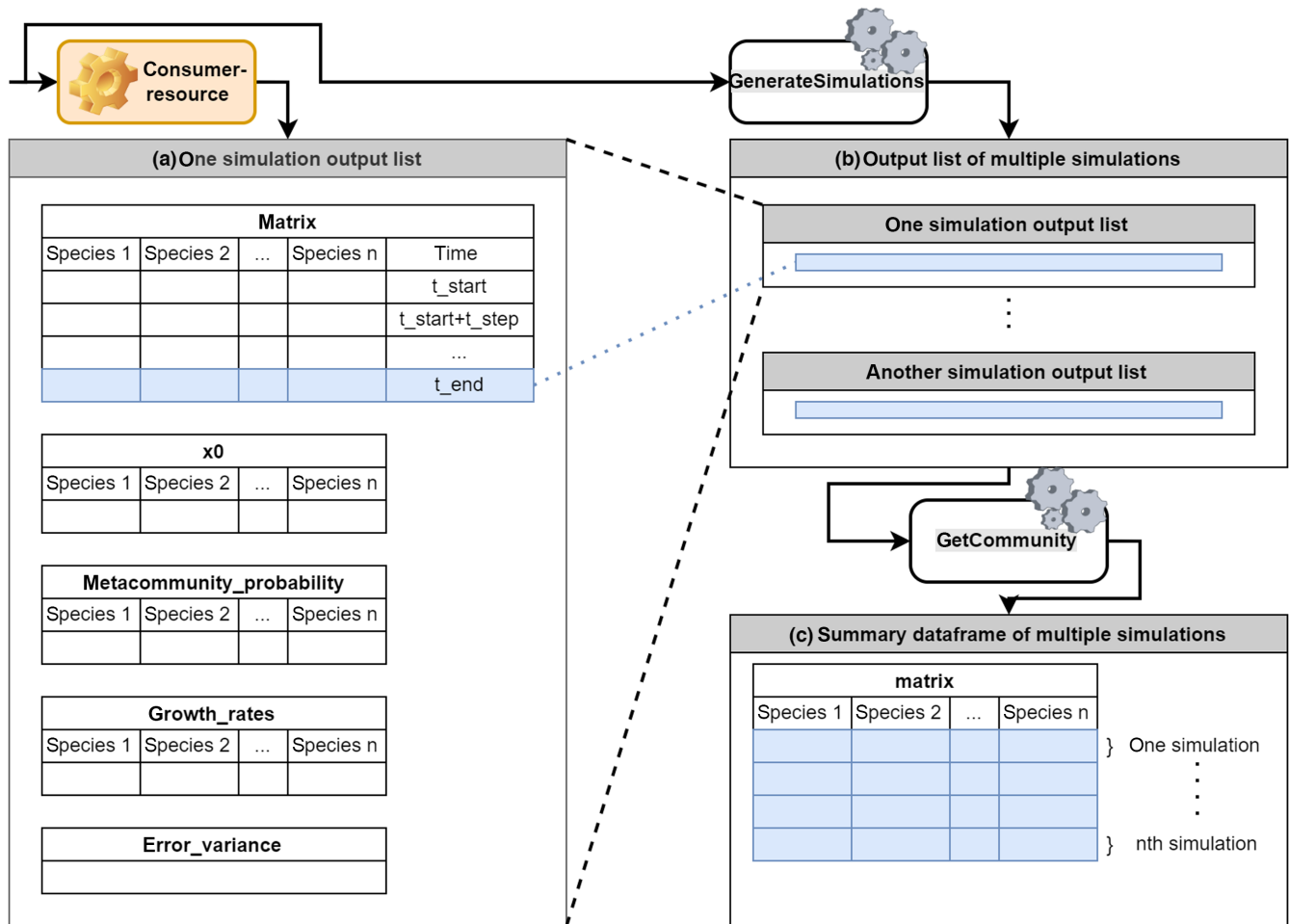


FIGURE 2 Data structures of the outputs in miaSim. (a) A result of the consumer-resource model simulation is a list acquired by running the `simulateConsumerResource` function once and contains a matrix of community compositions at different time points in the simulation, initial abundances, probabilities of dispersal from the metacommunity, growth rates and the magnitude of the measurement error in the model. (b) A list of multiple simulations generated using a series of inputs by running `generateSimulations` function. (c) A summary data frame of multiple simulations acquired from a series of simulating results by applying the `getCommunity` function to a list of multiple simulations. In this data frame, each row represents one independent simulation.

i becomes an empty site according to the probability of extinction e_i . (iii) interaction. Similarly to the GLV model, between two species i and j , the interaction is defined by α_{ij} . Similarly to the Hubbell model, the SOI model is implemented with an optimized version of Gillespie's chemical stochastic simulation algorithm (Cai & Xu, 2007; Gillespie, 1976).

3 | CASE STUDIES

Next, we illustrate the utility of miaSim, by qualitatively reproducing studies from literature using functions from miaSim. Additionally, we used miaSim to introduce some simple yet inspiring simulations that were not in the original studies but reflect interesting extensions to their ecological hypotheses and patterns. We did not perform an exact comparison of miaSim's results with the replicated studies due to the stochastic parameters in the published examples, whose random seed and/or exact values were not available in their original

publications. However, the qualitative comparisons with the original publications leave no doubt that the results obtained with miaSim have the same interpretation as those of their original authors. Scripts to reproduce the simulations are available in the package `vi-gnettes` (Şimşek et al., 2023).

3.1 | Origin of different community types with strongly interacting species

Different human gut microbial community types (also known as “enterotypes”) were reported in healthy individuals (Arumugam et al., 2011). Their correlations with dietary habits and health status motivated the investigation of their differences in taxonomic, functional, and ecological properties (Costea et al., 2018). However, the origin of these different community types is still under debate. Multistability, host heterogeneity, and interspecific interactions could all be alternative explanations for the genesis of enterotypes

(Falony et al., 2016; Gonze et al., 2017). In a series of simulations using the generalized Lotka-Volterra model, Gibson et al. showed that different community types could emerge by combining subsampling from a metacommunity with a heterogeneous distribution of species interaction strength, and suggested that the presence of species that interact strongly with many other species, referred to as strongly interacting species (SIS), could give rise to different stable community types (Gibson et al., 2016).

Here we reproduced their most important simulations using miaSim. Local communities with different species interaction strengths were generated using the “randomA” function with a specific custom “interactions” argument. As performed in the original study, we generated interaction matrices by drawing random values from a power

law probability distribution $P(\alpha)$. Different values of α {7, 3, 2, 1.6, 1.01} were explored. As shown in Figure 3a, decreasing α leads to the formation of unevenly distributed species interaction strengths. As a result, the interactions between species changed from approximately even and weak interactions to stronger interactions centered on a few species, leading to the formation of SIS (Figure 3b). Five types of local communities were assembled separately using the five different values of α . For each type, 80 species were randomly drawn from their local species pool to form 100 random initial communities. Then, they were set to the same initial abundances and simulated with the generalized Lotka-Volterra model with their corresponding interaction subnetwork (where only the 80 local species are included). Dimensionality reduction of the final states of these

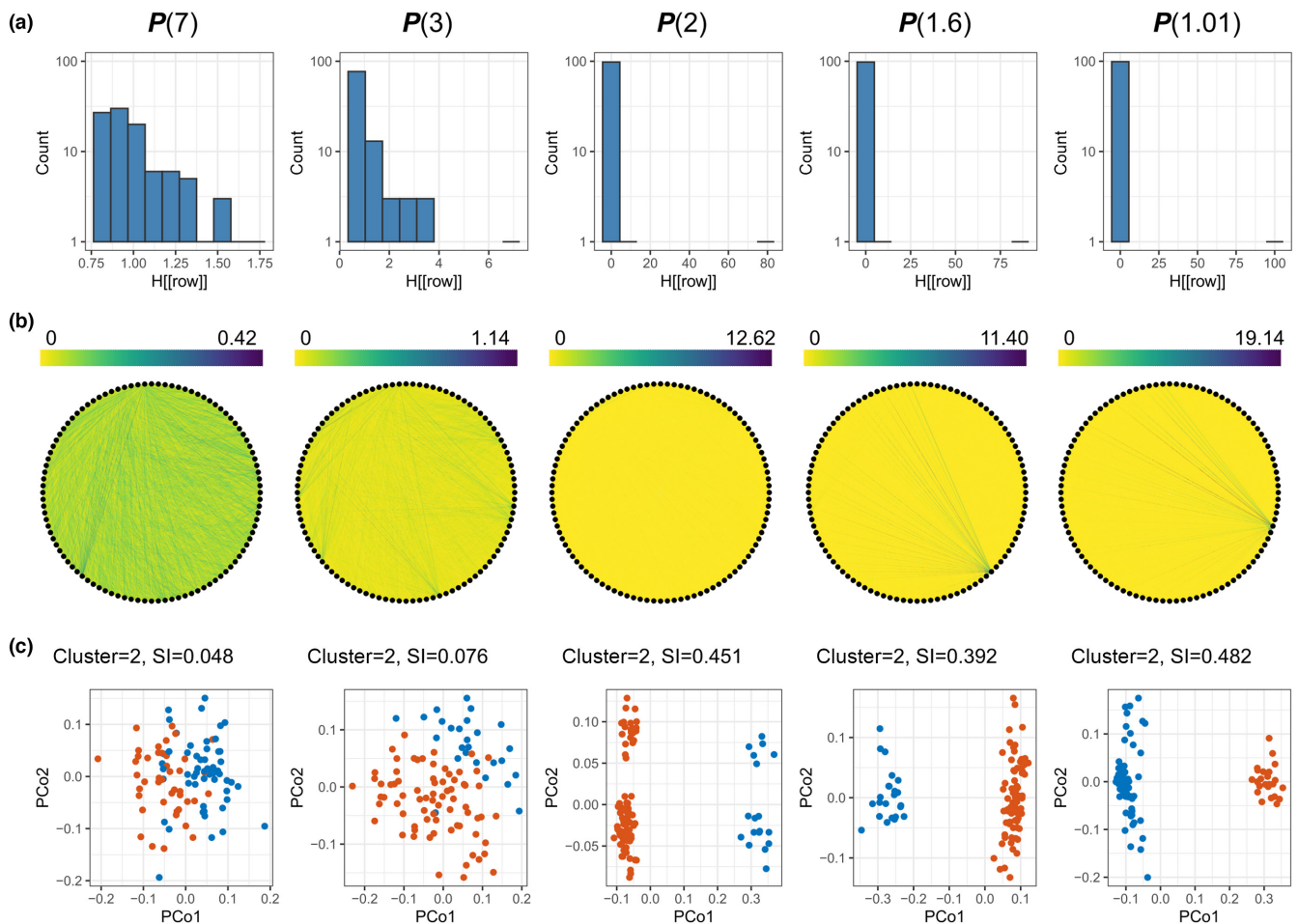


FIGURE 3 Impact of heterogeneity of interaction-strength on the distinctness of community types. This figure is a recomputation of the model from (Gibson et al., 2016) using miaSim but with different random seeds and specific numerical values. With different interaction strength heterogeneities, we generated a series of inter-species interaction matrices, and subsequently, simulated the corresponding community changes using the generalized Lotka-Volterra model. (a) Histograms of the interaction heterogeneity matrix H with different α {7, 3, 2, 1.6, 1.01} in a power-law distribution. The highest H display an increasing value with α approaching 1 (b) Randomly generated networks of interspecies interactions. Each network contains 100 nodes, indicating 100 species in the local species pool, from which 80 species were chosen randomly 100 times to form 100 initial communities in every interaction strength pattern. From light yellow to dark blue, the edge colours indicate different strengths of interactions between species. (c) Principal coordinate analyses (PCoA) of the final state of communities using the Jensen-Shannon distance. In every plot, each point represents a final community, and number of clusters is determined by the Silhouette Index (SI) acquired using Partitioning-Around-Medoid (PAM) algorithm. SI represents the goodness of clustering. It increases with the decrease of α , implying that the Strongly Interacting Species (SIS) in the last interaction patterns promotes the formation of different community types.

simulated communities displayed a gradient of separation into enterotypes (Figure 3c).

In addition to reproducing the results from (Gibson et al., 2016) in *miaSim*, we explored the simulation results from *miaSim*, which showed that the two clusters seen in the principal coordinate analysis (PCoA) for simulations with the last interaction network were separated by the presence/absence of the SIS (Supporting Information Figure S1), which illustrates the mechanisms by which the SIS could contribute to the formation of community types.

3.2 | A nutrient concentration threshold in microbial community diversity

An important challenge of microbial ecology is to unravel the relative contributions of abiotic and biotic factors to community diversity (New & Brito, 2020). In different environments (e.g. marine, soil, fresh water, air, human gut, etc.) microbial community diversity is driven by different factors, including the availability of nutrients (Chen et al., 2019; Crowther et al., 2019; Oliphant et al., 2019; Palmer & Ruhi, 2019; Shade et al., 2013). We used *miaSim* to explore the hypothetical relationship between nutrient concentration and community diversity, mimicking a gradient from poor (oligotrophic) to nutrient-rich (eutrophic) environments. We explored the hypothesis suggested by Marsland et al. that microbiomes exhibit a threshold for nutrient concentration in the relationship between community diversity and the availability of nutrients or, more generally, “energy fluxes” (Marsland III et al., 2019). Below this threshold, according to Marsland et al, microbiome diversity is externally modulated by the total influx of nutrients, and once the threshold is exceeded, the diversity is controlled by ecological interactions (Marsland III et al., 2019).

To test this hypothesis, we used *miaSim* to simulate consumer-resource models with combinations of 10 initial communities, 10 nutrient types and 6 quantities of nutrients. This simulation does not specifically replicate Marsland III et al.'s (2019) simulation, but only broadly tests the same nutrient threshold hypothesis. In our simulations, each community contained at most 10 species, and to control the effect of the initial composition, the beta-diversity between the first initial community and other randomly generated initial communities were chosen to be in an increasing gradient (the Bray–Curtis dissimilarity index increased from 0.05 to 0.85); the nutrient compositions, each containing five varieties of nutrients, were also structured into a gradient of increasing dissimilarities; and the quantities of nutrients were set as a geometric progression from 1 to 10^5 with a common ratio of 10.

Plotting the results of simulations into a uniform manifold approximation and Projection (UMAP) (Supporting Information Figure S2) or PCoA (Supporting Information Figure S3) (UMAP is recommended for microbiome data; Armstrong et al., 2021), we can observe that an increase in the concentration of nutrients (from 1 to 10^2) leads to an increase in the community beta diversity, that is communities spread more in the UMAP projection, but

beyond 10^2 the clusters do not change (Supporting Information Figures S4 and S5). To illustrate this trend, we plotted the average distances between simulated communities in the UMAP and PCoA projections against the nutrient concentrations (Supporting Information Figures S6 and S7). In our simulations, the nutrient saturation threshold was near 10^2 .

Next, we focus on the clustering patterns observed in the UMAP projections. Similar to the effect that SIS have on the formation of clusters from metacommunities simulated with the GLV model (Figure 3), both nutrient concentration and diversity lead to the formation of enterotypes in metacommunities simulated with the consumer-resource model (Figure 4). Clusters are frequently observed when there is an increase in resource concentration in different growing environments (cultivating medium; i.e. communities 1 and 2; communities 3, 4 and 7; communities 5, 8 and 9; communities 6 and 10 in Figure 4). This suggests that, according to the consumer-resource model, enterotypes are expected if local communities encounter an environment where resources are abundant, similar to the environment that microbes encounter in the human gut. It is also, to our knowledge, the first time that enterotypes emerge in consumer-resource model simulations that mimic a move from a resource-rich to a resource-poor environment.

3.3 | Nutrient complexity impacts the composition and diversity of the microbiome

In addition to the total concentration of nutrients, another factor that could externally modulate the composition of the microbiome is the number of different types of resources that microbes encounter in their environment (Madigan et al., 2018). In contrast to the previous section, where we explored nutrient concentration, here we fixed the total concentrations and varied the number of resource types. Qualitatively following the simulations from a previous study (Pacheco et al., 2021), we investigated the impacts of environmental complexity (numbers of available nutrients) for different community sizes.

First, we used a consumer-resource model in *miaSim* that follows the same parameters used by Pacheco et al. (Pacheco et al., 2021), smaller and larger communities (consisting, respectively 3, 4 and 13 species) were compared in different environments containing different combinations of carbon sources (the number of sources were 1, 2, 4, 8, 16 and 32). Minor model adaptations are commented on in the source code (Şimşek et al., 2023).

Similarly to (Pacheco et al., 2021), we also observed a nonadditive relationship between environment complexity and growth yields (Figure 5). Next, we explored a wider range of parameters than the ones used by (Pacheco et al., 2021) to better understand the key factors that impact the composition and diversity of the microbiome when increasing the number of alternative carbon sources. First, we simulated a wider range of niche overlaps (ρ) by extending the average fraction of resources usable by each species (θ) to 0.25, 0.1, and 0.05. Interestingly, we found that in larger communities, the growth yield increased with the number of resources only when θ

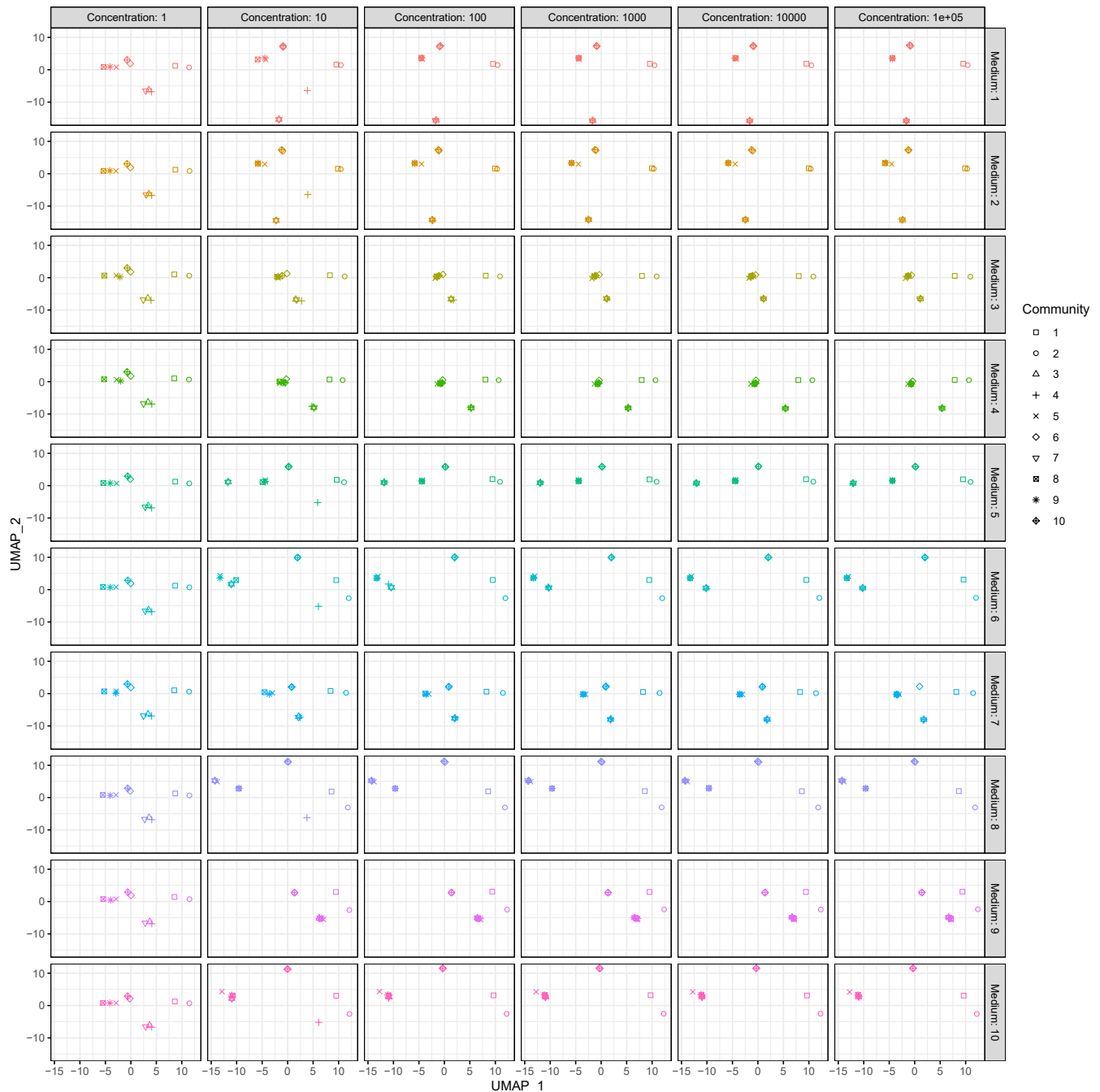


FIGURE 4 Uniform Manifold Approximation and Projection (UMAP) of simulated communities in different media (both components and concentrations). In each row, the relative composition of nutrients in the medium is the same, but from left to right, the quantities of resources for microbial communities increase in a geometric sequence. In each subfigure, different symbols represent their different initial community compositions. Each subfigure in the last column presents a common clustering result: four community types (communities 1 and 2; communities 3, 4 and 7; communities 5, 8 and 9; communities 6 and 10) are consistently observed in different environments.

is relatively small, and the increase is negatively correlated with the value of θ (Supporting Information Figure S8). When θ is large, then the niche overlap is large, too, and most species can simultaneously use most of the resources; in contrast, when θ is relatively small, the resources are only available for their specialists and could be left unused in smaller communities.

To further explore the impact of θ on the diversity of microbial communities, we measured the average Euclidean

distance between different initial communities in UMAP and found that larger communities (with 13 species) are less likely to be influenced by increasing the number of resources (Supporting Information Figure S9). The increase in the number of alternative carbon sources also leads to a decrease in the beta diversity, as the average amount of each resource is decreased, causing their consumption and growth to be more evenly distributed between populations of different species.

4 | DISCUSSION

Unified theories or general concepts are important pursuits in community ecology. A number of patterns initially found in macroecology also apply in microbial ecology, including the species-area relationship, species-abundance distribution, distance-dissimilarity correlation, etc. (Lennon & Locey, 2017). Due to the challenges in handling and analysing large amounts of data acquired from microbiomes, much effort is devoted to reducing the gaps between

ecological theories, observed patterns, available tools and accumulating data (Mascarenhas et al., 2020).

For example, the conceptual synthesis of different ecological processes (selection, dispersal, drift and speciation) in community ecology (Vellend, 2010) inspired the development of tools to quantify these processes (Stegen et al., 2013); the discussion of stochasticity in microbial ecology (Zhou & Ning, 2017) paved the way towards quantifying ecological stochasticity using null model algorithms (Ning et al., 2019). We carried out this work

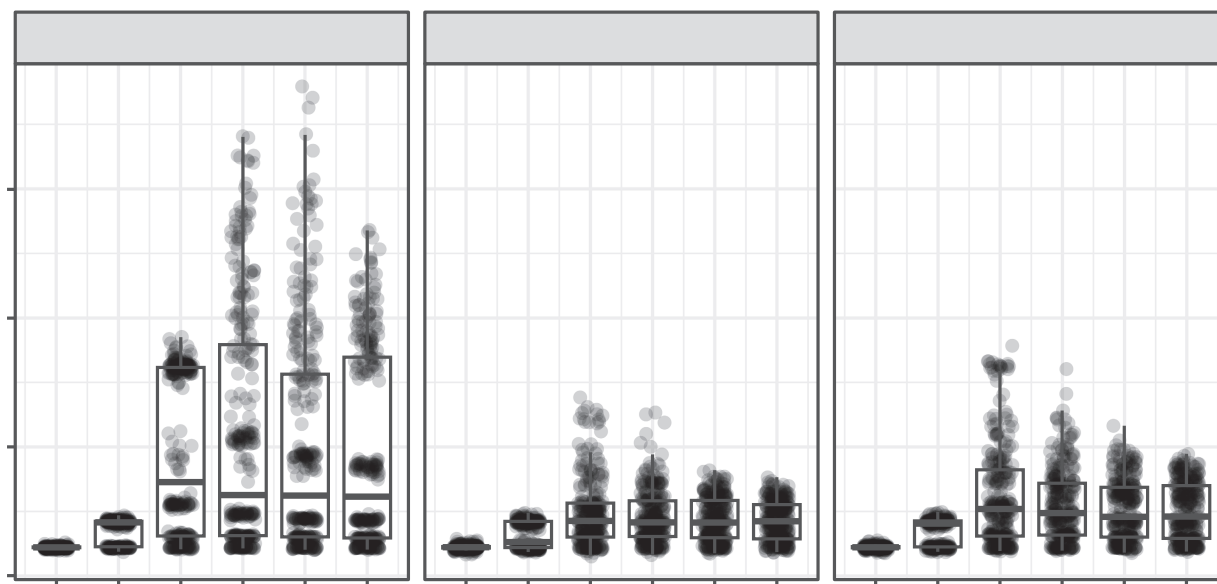


FIGURE 5 Changes in community growth yield when more alternative carbon sources are added. This figure is a recomputation of the model from (Pacheco et al., 2021) using miaSim with a different random seed and slight model differences from the original publication, but qualitatively similar. Box plots contain the distribution of growth yields at the end of simulations of a complex community with 13 species (left) and simple communities with 3 species (middle) and 4 species (right) growing at the same concentration but different types of carbon sources. The central lines indicate the median values, and the top and bottom box edges are the 25th and 75th percentiles, respectively.

TABLE 3 Comparison of community modelling tools (✓: Supported).

Features\models		miaSim	growthrates	growthcurver	microPop	web-gLV	Micodymora	rescomp	untb
Available models	Exponential		✓						
	Logistic	✓	✓	✓					
	Monod				✓				
	Lotka-Volterra	✓				✓			
	Consumer-resource	✓			✓		✓	✓	
	Hubbell's neutral	✓							✓
	User-defined		✓	✓					
Variables in simulations	Initial states	✓	✓	✓	✓	✓	✓	✓	✓
	Time points	✓	✓	✓	✓	✓	✓	✓	✓
	Cultivation modes	✓	✓		✓		✓	✓	
	Stochasticity	✓							
Miscellaneous functions	Batch simulation	✓		✓	✓				
	Easy visualization	✓	✓	✓	✓	✓		✓	✓

in the same spirit to facilitate the comparison of classical community models.

To further emphasize the versatility of miaSim, the features in various microbial community modelling tools are summarized in Table 3.

Different models can generate the same ecological patterns. miaSim is therefore developed to simulate microbiomes and compare different results coming from various models, to tell which model better fits specific data under a certain scenario or whether several models and their hypotheses can explain the observed ecological pattern. For instance, in the first two case studies, the GLV and consumer-resource model both reproduced the formation of different community types but relied on different mechanisms.

In addition, miaSim supports the *TreeSummarizedExperiment* data container and the associated package ecosystem in R/Bioconductor, allowing scientists to easily combine and compare their model simulations with microbiome datasets. In the future, functions that specifically fit sets of parameters to experimental time series can be added, allowing users to better compare models and experimental data. This can also be the result of a community effort, since miaSim is openly available as a part of the R/Bioconductor project.

To conclude, miaSim is designed to gather different community models into one place to facilitate the comparison between models, and to help the validation/discovery of ecological insights. In addition to miaSim, with the help of an intuitive graphical user interface and explanations provided in miaSimShiny, both novice and proficient users are empowered to effectively simulate the community dynamics and explore ecological patterns of microorganisms, and to contribute to the study of microbial ecology in various scenarios.

AUTHOR CONTRIBUTIONS

Karoline Faust, Daniel Rios Garza and Leo Lahti conceived the ideas; all authors contributed critically to the design and development of the software package miaSim; Yu Gao and Daniel Rios Garza gathered and analysed the data, and drafted the manuscript. All authors contributed to the writing and reviewing of the draft and gave final approval for publication.

ACKNOWLEDGEMENTS

We thank Didier Gonze for providing the MATLAB implementations of the GLV and Hubbell models that were compared to miaSim (Supporting Information File S1).

FUNDING INFORMATION

This work is supported by funding from the European Research Council (ERC) under the European Union's Horizon 2020 Research and Innovation Programme (grant agreement no. 801747), the China Scholarship Council (grant no. 202006710066), the Academy of Finland (grant no. 330887 to L.L.) and the Horizon 2020 Programme of the European Union (FindingPheno grant 952914 to T.B.).

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

PEER REVIEW

The peer review history for this article is available at <https://www.webofscience.com/api/gateway/wos/peer-review/10.1111/2041-210X.14129>.

DATA AVAILABILITY STATEMENT

The version of miaSim referenced by the manuscript as well as the code used to generate the simulations are available and archived in Zenodo (Şimşek et al., 2023), which includes the vignettes to replicate the case studies described in Section 3.

ORCID

Yu Gao  <https://orcid.org/0000-0003-4406-2978>

Leo Lahti  <https://orcid.org/0000-0001-5537-637X>

Karoline Faust  <https://orcid.org/0000-0001-7129-2803>

Daniel Rios Garza  <https://orcid.org/0000-0003-3865-2146>

REFERENCES

- Armstrong, G., Martino, C., Rahman, G., Gonzalez, A., Vázquez-Baeza, Y., Mishne, G., & Knight, R. (2021). Uniform manifold approximation and projection (UMAP) reveals composite patterns and resolves visualization artifacts in microbiome data. *mSystems*, 6(5), e00691-21. <https://doi.org/10.1128/mSystems.00691-21>
- Arumugam, M., Raes, J., Pelletier, E., Le Paslier, D., Yamada, T., Mende, D. R., Fernandes, G. R., Tap, J., Bruls, T., Batto, J.-M., Bertalan, M., Borruel, N., Casellas, F., Fernandez, L., Gautier, L., Hansen, T., Hattori, M., Hayashi, T., Kleerebezem, M., ... Bork, P. (2011). Enterotypes of the human gut microbiome. *Nature*, 473(7346), 174–180. <https://doi.org/10.1038/nature09944>
- Bacaër, N. (2011). Verhulst and the logistic equation (1838). In N. Bacaër (Ed.), *A short history of mathematical population dynamics* (pp. 35–39). Springer. https://doi.org/10.1007/978-0-85729-115-8_6
- Bauer, E., Zimmermann, J., Baldini, F., Thiele, I., & Kaleta, C. (2017). BacArena: Individual-based metabolic modeling of heterogeneous microbes in complex communities. *PLoS Computational Biology*, 13(5), e1005544. <https://doi.org/10.1371/journal.pcbi.1005544>
- Bengtsson-Palme, J. (2020). Microbial model communities: To understand complexity, harness the power of simplicity. *Computational and Structural Biotechnology Journal*, 18, 3987–4001. <https://doi.org/10.1016/j.csbj.2020.11.043>
- Bucci, V., Tzen, B., Li, N., Simmons, M., Tanoue, T., Bogart, E., Deng, L., Yeliseyev, V., Delaney, M. L., Liu, Q., Olle, B., Stein, R. R., Honda, K., Bry, L., & Gerber, G. K. (2016). MDSINE: Microbial dynamical systems inference engine for microbiome time-series analyses. *Genome Biology*, 17(1), 121. <https://doi.org/10.1186/s13059-016-0980-6>
- Buffie, C. G., Bucci, V., Stein, R. R., McKenney, P. T., Ling, L., Gobourne, A., No, D., Liu, H., Kinnebrew, M., Viale, A., Littmann, E., van den Brink, M. R. M., Jenq, R. R., Taur, Y., Sander, C., Cross, J. R., Toussaint, N. C., Xavier, J. B., & Pamer, E. G. (2015). Precision microbiome reconstitution restores bile acid mediated resistance to *Clostridium difficile*. *Nature*, 517(7533), Article 7533. <https://doi.org/10.1038/nature13828>
- Butler, S., & O'Dwyer, J. P. (2018). Stability criteria for complex microbial communities. *Nature Communications*, 9(1), 2970. <https://doi.org/10.1038/s41467-018-05308-z>

- Cai, X., & Xu, Z. (2007). K-leap method for accelerating stochastic simulation of coupled chemical reactions. *The Journal of Chemical Physics*, 126(7), 074102. <https://doi.org/10.1063/1.2436869>
- Chen, J., McIlroy, S. E., Archana, A., Baker, D. M., & Panagiotou, G. (2019). A pollution gradient contributes to the taxonomic, functional, and resistome diversity of microbial communities in marine sediments. *Microbiome*, 7(1), 104. <https://doi.org/10.1186/s40168-019-0714-6>
- Chesson, P. (1990). MacArthur's consumer-resource model. *Theoretical Population Biology*, 37(1), 26–38. [https://doi.org/10.1016/0040-5809\(90\)90025-Q](https://doi.org/10.1016/0040-5809(90)90025-Q)
- Coenen, A. R., Hu, S. K., Luo, E., Muratore, D., & Weitz, J. S. (2020). A primer for microbiome time-series analysis. *Frontiers in Genetics*, 11, 310. <https://doi.org/10.3389/fgene.2020.00310>
- Constable, G. W. A., Rogers, T., McKane, A. J., & Tarnita, C. E. (2016). Demographic noise can reverse the direction of deterministic selection. *Proceedings of the National Academy of Sciences of the United States of America*, 113(32), E4745–E4754. <https://doi.org/10.1073/pnas.1603693113>
- Costea, P. I., Hildebrand, F., Arumugam, M., Bäckhed, F., Blaser, M. J., Bushman, F. D., de Vos, W. M., Ehrlich, S. D., Fraser, C. M., Hattori, M., Huttenhower, C., Jeffery, I. B., Knights, D., Lewis, J. D., Ley, R. E., Ochman, H., O'Toole, P. W., Quince, C., Relman, D. A., ... Bork, P. (2018). Enterotypes in the landscape of gut microbial community composition. *Nature Microbiology*, 3(1), 8–16. <https://doi.org/10.1038/s41564-017-0072-8>
- Coyte, K. Z., Schluter, J., & Foster, K. R. (2015). The ecology of the microbiome: Networks, competition, and stability. *Science*, 350(6261), 663. <https://doi.org/10.1126/science.aad2602>
- Crowther, T. W., van den Hoogen, J., Wan, J., Mayes, M. A., Keiser, A. D., Mo, L., Averill, C., & Maynard, D. S. (2019). The global soil community and its influence on biogeochemistry. *Science*, 365(6455), eaav0550. <https://doi.org/10.1126/science.aav0550>
- Delattre, H., Chen, J., Wade, M., & Soyer, O. S. (2019). *Thermodynamic modelling of synthetic communities predicts minimum free energy requirements for sulfate reduction and methanogenesis* (p. 857276). bioRxiv <https://doi.org/10.1101/857276>
- Descheemaeker, L., & de Buyl, S. (2020). Stochastic logistic models reproduce experimental time series of microbial communities. *eLife*, 9, e55650. <https://doi.org/10.7554/eLife.55650>
- D'hoel, K., Vet, S., Faust, K., Moens, F., Falony, G., Gonze, D., Lloréns-Rico, V., Gelens, L., Danckaert, J., De Vuyst, L., & Raes, J. (2018). Integrated culturing, modeling and transcriptomics uncovers complex interactions and emergent behavior in a three-species synthetic gut community. *eLife*, 7, e37090. <https://doi.org/10.7554/eLife.37090>
- Dukovski, I., Bajić, D., Chacón, J. M., Quintin, M., Vila, J. C. C., Sulheim, S., Pacheco, A. R., Bernstein, D. B., Riehl, W. J., Korolev, K. S., Sanchez, A., Harcombe, W. R., & Segrè, D. (2021). A metabolic modeling platform for the computation of microbial ecosystems in time and space (COMETS). *Nature Protocols*, 16(11), 5030–5082. <https://doi.org/10.1038/s41596-021-00593-3>
- Falony, G., Joossens, M., Vieira-Silva, S., Wang, J., Darzi, Y., Faust, K., Kurilshikov, A., Bonder, M. J., Valles-Colomer, M., Vandeputte, D., Tito, R. Y., Chaffron, S., Rymenans, L., Verspecht, C., De Sutter, L., Lima-Mendez, G., D'hoel, K., Jonckheere, K., Homola, D., ... Raes, J. (2016). Population-level analysis of gut microbiome variation. *Science*, 352(6285), 560–564. <https://doi.org/10.1126/science.aad3503>
- Faust, K., Bauchinger, F., Laroche, B., de Buyl, S., Lahti, L., Washburne, A. D., Gonze, D., & Widder, S. (2018). Signatures of ecological processes in microbial community time series. *Microbiome*, 6(1), 120. <https://doi.org/10.1186/s40168-018-0496-2>
- Fisher, C. K., & Mehta, P. (2014). The transition between the niche and neutral regimes in ecology. *Proceedings of the National Academy of Sciences of the United States of America*, 111, 13111–13116. <https://doi.org/10.1073/pnas.1405637111>
- Gibson, T. E., Bashan, A., Cao, H.-T., Weiss, S. T., & Liu, Y.-Y. (2016). On the origins and control of community types in the human microbiome. *PLoS Computational Biology*, 12(2), e1004688. <https://doi.org/10.1371/journal.pcbi.1004688>
- Gillespie, D. T. (1976). A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *Journal of Computational Physics*, 22(4), 403–434. [https://doi.org/10.1016/0021-9991\(76\)90041-3](https://doi.org/10.1016/0021-9991(76)90041-3)
- Goldschmidt, F., Caduff, L., & Johnson, D. R. (2021). Causes and consequences of pattern diversification in a spatially self-organizing microbial community. *The ISME Journal*, 15(8), 2415–2426. <https://doi.org/10.1038/s41396-021-00942-w>
- Gonze, D., Lahti, L., Raes, J., & Faust, K. (2017). Multi-stability and the origin of microbial community types. *The ISME Journal*, 11(10), Article 10. <https://doi.org/10.1038/ismej.2017.60>
- Greenblum, S., Chiu, H.-C., Levy, R., Carr, R., & Borenstein, E. (2013). Towards a predictive systems-level model of the human microbiome: Progress, challenges, and opportunities. *Current Opinion in Biotechnology*, 24(4), 810–820. <https://doi.org/10.1016/j.copbio.2013.04.001>
- Grilli, J. (2020). Macroecological laws describe variation and diversity in microbial communities. *Nature Communications*, 11(1), 4743. <https://doi.org/10.1038/s41467-020-18529-y>
- Hankin, R. K. S. (2019). *untb: Ecological Drift under the UNTB* (1.7-4). <https://CRAN.R-project.org/package=untb>
- Harris, K., Parsons, T. L., Ijaz, U. Z., Lahti, L., Holmes, I., & Quince, C. (2017). Linking statistical and ecological theory: Hubbell's unified neutral theory of biodiversity as a hierarchical dirichlet process. *Proceedings of the IEEE*, 105(3), 516–529. <https://doi.org/10.1109/JPROC.2015.2428213>
- Hellweger, F. L., Clegg, R. J., Clark, J. R., Plugge, C. M., & Kreft, J.-U. (2016). Advancing microbial sciences by individual-based modelling. *Nature Reviews Microbiology*, 14(7), 461–471. <https://doi.org/10.1038/nrmicro.2016.62>
- Huang, R., Sonesson, C., Ernst, F. G. M., Rue-Albrecht, K. C., Yu, G., Hicks, S. C., & Robinson, M. D. (2021). TreeSummarizedExperiment: A S4 class for data with hierarchical structure. *F1000Research*, 9, 1246. <https://doi.org/10.12688/f1000research.26669.2>
- Hubbell, S. P. (2001). *The unified neutral theory of biodiversity and biogeography*. Princeton University Press.
- Huttenhower, C., Gevers, D., Knight, R., Abubucker, S., Badger, J. H., Chinwalla, A. T., Creasy, H. H., Earl, A. M., FitzGerald, M. G., Fulton, R. S., Giglio, M. G., Hallsworth-Pepin, K., Lobos, E. A., Madupu, R., Magrini, V., Martin, J. C., Mitreva, M., Muzny, D. M., Sodergren, E. J., ... The Human Microbiome Project Consortium. (2012). Structure, function and diversity of the healthy human microbiome. *Nature*, 486(7402), 207–214. <https://doi.org/10.1038/nature11234>
- Kehe, J., Ortiz, A., Kulesa, A., Gore, J., Blainey, P. C., & Friedman, J. (2021). Positive interactions are common among culturable bacteria. *Science Advances*, 7(45), eabi7159. <https://doi.org/10.1126/sciadv.abi7159>
- Kettle, H., Holtrop, G., Louis, P., & Flint, H. J. (2018). microPop: Modelling microbial populations and communities in R. *Methods in Ecology and Evolution*, 9(2), 399–409. <https://doi.org/10.1111/2041-210X.12873>
- Khalighi, M., Sommeria-Klein, G., Gonze, D., Faust, K., & Lahti, L. (2022). Quantifying the impact of ecological memory on the dynamics of interacting communities. *PLoS Computational Biology*, 18(6), e1009396. <https://doi.org/10.1371/journal.pcbi.1009396>
- Khazaei, T., Williams, R. L., Bogatyrev, S. R., Doyle, J. C., Henry, C. S., & Ismagilov, R. F. (2020). Metabolic multistability and hysteresis in a model aerobic-anaerobic microbiome community. *Science Advances*, 6(33), eaba0353. <https://doi.org/10.1126/sciadv.aba0353>
- Kuntal, B. K., Gadgil, C., & Mande, S. S. (2019). Web-gLV: A web based platform for Lotka-Volterra based modeling and simulation of

- microbial populations. *Frontiers in Microbiology*, 10 <https://www.frontiersin.org/article/10.3389/fmicb.2019.00288>
- Lennon, J. T., & Locey, K. J. (2017). Macroecology for microbiology. *Environmental Microbiology Reports*, 9(1), 38–40. <https://doi.org/10.1111/1758-2229.12512>
- Letten, A. D. (2022). *rescomp: An R package for defining, simulating and visualizing ODE models of consumer-resource interactions* (p. 2022.01.11.475574). bioRxiv <https://doi.org/10.1101/2022.01.11.475574>
- Li, Y., Fan, L., Zhang, W., Zhu, X., Lei, M., & Niu, L. (2020). How did the bacterial community respond to the level of urbanization along the Yangtze River? *Environmental Science: Processes & Impacts*, 22(1), 161–172. <https://doi.org/10.1039/C9EM00399A>
- Logan, B. E., & Rabaey, K. (2012). Conversion of wastes into bioelectricity and chemicals by using microbial electrochemical technologies. *Science*, 337(6095), 686–690. <https://doi.org/10.1126/science.1217412>
- Lotka, A. J. (1925). *Elements of physical biology*. Williams & Wilkins. <https://archive.org/details/elementsofphysic017171mbp/mode/2up>
- MacArthur, R. (1970). Species packing and competitive equilibrium for many species. *Theoretical Population Biology*, 1(1), 1–11. [https://doi.org/10.1016/0040-5809\(70\)90039-0](https://doi.org/10.1016/0040-5809(70)90039-0)
- Madigan, M. T., Bender, K. S., Buckley, D. H., Sattley, W. M., & Stahl, D. A. (2018). *Brock biology of microorganisms* (15th ed.). Pearson.
- Marsland, R., III, Cui, W., Goldford, J., Sanchez, A., Korolev, K., & Mehta, P. (2019). Available energy fluxes drive a transition in the diversity, stability, and functional structure of microbial communities. *PLoS Computational Biology*, 15(2), e1006793. <https://doi.org/10.1371/journal.pcbi.1006793>
- Marsland, R., Cui, W., Goldford, J., & Mehta, P. (2020). The community simulator: A python package for microbial ecology. *PLoS ONE*, 15(3), e0230430. <https://doi.org/10.1371/journal.pone.0230430>
- Marsland, R., Cui, W., & Mehta, P. (2020). A minimal model for microbial biodiversity can reproduce experimentally observed ecological patterns. *Scientific Reports*, 10(1), 3308. <https://doi.org/10.1038/s41598-020-60130-2>
- Mascarenhas, R., Ruziska, F. M., Moreira, E. F., Campos, A. B., Loiola, M., Reis, K., Trindade-Silva, A. E., Barbosa, F. A. S., Salles, L., Menezes, R., Veiga, R., Coutinho, F. H., Dutilh, B. E., Guimarães, P. R., Assis, A. P. A., Ara, A., Miranda, J. G. V., Andrade, R. F. S., Vilela, B., & Meirelles, P. M. (2020). Integrating computational methods to investigate the macroecology of microbiomes. *Frontiers in Genetics*, 10, 1344. <https://doi.org/10.3389/fgene.2019.01344>
- May, R. M. (1972). Will a large complex system be stable? *Nature*, 238(5364), Article 5364. <https://doi.org/10.1038/238413a0>
- Monod, J. (1949). The growth of bacterial cultures. *Annual Review of Microbiology*, 3(1), 371–394. <https://doi.org/10.1146/annurev.mi.03.100149.002103>
- Natarajan, A., & Bhatt, A. S. (2020). Microbes and microbiomes in 2020 and beyond. *Nature Communications*, 11(1), Article 1. <https://doi.org/10.1038/s41467-020-18850-6>
- New, F. N., & Brito, I. L. (2020). What is metagenomics teaching us, and what is missed? *Annual Review of Microbiology*, 74, 117–135. <https://doi.org/10.1146/annurev-micro-012520-072314>
- Ning, D., Deng, Y., Tiedje, J. M., & Zhou, J. (2019). A general framework for quantitatively assessing ecological stochasticity. *Proceedings of the National Academy of Sciences of the United States of America*, 116(34), 16892. <https://doi.org/10.1073/pnas.1904623116>
- Oliphant, K., Parreira, V. R., Cochrane, K., & Allen-Vercoe, E. (2019). Drivers of human gut microbial community assembly: Coadaptation, determinism and stochasticity. *The ISME Journal*, 13(12), 3080–3092. <https://doi.org/10.1038/s41396-019-0498-5>
- Pacheco, A. R., Osborne, M. L., & Segrè, D. (2021). Non-additive microbial community responses to environmental complexity. *Nature Communications*, 12(1), 2365. <https://doi.org/10.1038/s41467-021-22426-3>
- Palmer, J. D., & Foster, K. R. (2022). Bacterial species rarely work together. *Science*, 376(6593), 581–582. <https://doi.org/10.1126/science.abn5093>
- Palmer, M., & Ruhli, A. (2019). Linkages between flow regime, biota, and ecosystem processes: Implications for river restoration. *Science*, 365(6459), eaaw2087. <https://doi.org/10.1126/science.aaw2087>
- Patowary, K., Patowary, R., Kalita, M. C., & Deka, S. (2016). Development of an efficient bacterial consortium for the potential remediation of hydrocarbons from contaminated sites. *Frontiers in Microbiology*, 7 <https://www.frontiersin.org/articles/10.3389/fmicb.2016.01092>
- Peces, M., Dottorini, G., Nierychlo, M., Andersen, K. S., Dueholm, M. K. D., & Nielsen, P. H. (2022). Microbial communities across activated sludge plants show recurring species-level seasonal patterns. *ISME Communications*, 2(1), 18. <https://doi.org/10.1038/s43705-022-00098-4>
- Petzoldt, T. (2020). *Growthrates: Estimate growth rates from experimental data* (0.8.2). <https://CRAN.R-project.org/package=growthrates>
- Shade, A., Gregory Caporaso, J., Handelsman, J., Knight, R., & Fierer, N. (2013). A meta-analysis of changes in bacterial and archaeal communities with time. *The ISME Journal*, 7, 1493–1506.
- Şimşek, Y., Lahti, L., Gao, Y., Garza, D., Turaga, N., Borman, T., & Jwokaty, Hallucigenia-Sparsi, & Daenarys8. (2023). *Microbiome/miaSim: V3.15-Alpha*. Zenodo <https://doi.org/10.5281/zenodo.7900177>
- Solé, R. V., Levin, S. A., Solé, R. V., Alonso, D., & McKane, A. (2002). Self-organized instability in complex ecosystems. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 357(1421), 667–681. <https://doi.org/10.1098/rstb.2001.0992>
- Song, H.-S., Cannon, W. R., Beliaev, A. S., & Konopka, A. (2014). Mathematical modeling of microbial community dynamics: A methodological review. *Processes*, 2(4), 711–752. <https://doi.org/10.3390/pr2040711>
- Sprouffske, K., & Wagner, A. (2016). Growthcurver: An R package for obtaining interpretable metrics from microbial growth curves. *BMC Bioinformatics*, 17(1), 172. <https://doi.org/10.1186/s12859-016-1016-7>
- Stegen, J. C., Konopka, A., Murray, C. J., Kennedy, D. W., Fredrickson, J. K., Rockhold, M. L., Chen, X., & Lin, X. (2013). Quantifying community assembly processes and identifying features that impose them. *The ISME Journal*, 7(11), 2069–2079. <https://doi.org/10.1038/ismej.2013.93>
- Toju, H., Peay, K. G., Yamamichi, M., Narisawa, K., Hiruma, K., Naito, K., Fukuda, S., Ushio, M., Nakaoka, S., Onoda, Y., Yoshida, K., Schlaeppli, K., Bai, Y., Sugiura, R., Ichihashi, Y., Minamisawa, K., & Kiers, E. T. (2018). Core microbiomes for sustainable agroecosystems. *Nature Plants*, 4(5), 247–257. <https://doi.org/10.1038/s41477-018-0139-4>
- Vellend, M. (2010). Conceptual synthesis in community ecology. *The Quarterly Review of Biology*, 85(2), 183–206. <https://doi.org/10.1086/652373>
- Volterra, V. (1926). *Variazioni e fluttuazioni del numero d'individui in specie animali conviventi*. Memorie della R. Accademia dei Lincei.
- Wang, T., Goyal, A., Dubinkina, V., & Maslov, S. (2019). Evidence for a multi-level trophic organization of the human gut microbiome. *PLoS Computational Biology*, 15(12), e1007524. <https://doi.org/10.1371/journal.pcbi.1007524>
- Zhou, J., & Ning, D. (2017). Stochastic community assembly: Does it matter in microbial ecology? *Microbiology and Molecular Biology Reviews*, 81(4), e00002-17. <https://doi.org/10.1128/MMBR.00002-17>

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

Supplementary Table S1: Complete list of available parameters.

Supplementary File S1. Comparison of repeated miaSim simulations

using one or multiple processors and comparison of execution times with MATLAB implementations of the Hubbell and GLV models.

Supplementary Figure S1. The left panel demonstrates the best clustering result (identified using the K-medoid clustering algorithm, see details in Supplementary File 1) of PCoA (2 clusters), and right panel shows the two clusters of communities with and without SISs. The two types of communities corresponding well the two clusters.

Supplementary Figure S2. UMAP of communities after the simulations.

Supplementary Figure S3. PCoA of communities after the simulations.

Supplementary Figure S4. UMAP of communities after the simulations. Different panels are simulations with different resource concentrations.

Supplementary Figure S5. PCoA of communities after the simulations. Different panels are simulations with different resource concentrations.

Supplementary Figure S6. Line plot of average distance between communities in UMAP with same initial community but under different resource concentrations. The value indicates the beta diversity among these communities.

Supplementary Figure S7. Line plot of average distance between

communities in PCoA with same initial community but under different resource concentrations. The value indicates the beta diversity among these communities.

Supplementary Figure S8. Increase of growth yield with the number of resources happens in larger communities (13 species, not 3 or 4 species) only when niche overlap (ρ) is relatively low. p values in each panel indicates the test result of paired one-sided t test between growth yield with 1 single resource and 32 resources.

Supplementary Figure S9. Average Euclidean distance in UMAP between different communities after the simulations with different numbers of resources under a gradient of average proportion of usable resources (θ).

How to cite this article: Gao, Y., Şimşek, Y., Gheysen, E., Borman, T., Li, Y., Lahti, L., Faust, K., & Garza, D. R. (2023). miaSim: an R/Bioconductor package to easily simulate microbial community dynamics. *Methods in Ecology and Evolution*, 00, 1–14. <https://doi.org/10.1111/2041-210X.14129>