Assessing mechanisms for microbial taxa and community dynamics using process models

Authors: Linwei Wu[1,2,*], Yunfeng Yang[3], Daliang Ning[2], Qun Gao[3], Huaqun Yin[4], Naija Xiao[2], Benjamin Y. Zhou[5], Si Chen[6,7], Qiang He[6,7], and Jizhong Zhou[2,3,8,9]

Affiliation: [1]Institute of Ecology, Key Laboratory for Earth Surface Processes of the Ministry of Education, College of Urban and Environmental Sciences, Peking University, Beijing, China;

[2]Institute for Environmental Genomics and Department of Microbiology and Plant Biology, University of Oklahoma, Norman, OK, USA.

[3]State Key Joint Laboratory of Environment Simulation and Pollution Control, School of Environment, Tsinghua University, Beijing, China.

[4]School of Minerals Processing and Bioengineering, Central South University, Changsha, China.

[5]Department of Mathematics, Lunt Hall, Northwestern University, Evanston, IL, USA

[6]Department of Civil and Environmental Engineering, The University of Tennessee, Knoxville, TN, USA.

[7]Institute for a Secure and Sustainable Environment, The University of Tennessee, Knoxville, TN, USA.

[8]School of Civil Engineering and Environmental Sciences, University of Oklahoma, Norman, OK, USA.

[9]Earth and Environmental Sciences, Lawrence Berkeley National Laboratory, Berkeley, CA, USA.

Running title: Process models on community dynamics

***Correspondence:** Linwei Wu linwei.wu@pku.edu.cn

Abstract

Disentangling the assembly mechanisms controlling community composition, structure, distribution, functions, and dynamics is a central issue in ecology. Although various approaches have been proposed to examine community assembly mechanisms, quantitative characterization is challenging, particularly in microbial ecology. Here, we present a novel approach for quantitatively delineating community assembly mechanisms by combining the consumer-resource model with a neutral model in stochastic differential equations (SDEs). Using time-series data from anaerobic bioreactors that target microbial 16S rRNA genes, we tested the applicability of three ecological models, the consumer-resources model, the neutral model, and the combined model. Our results revealed that model performances varied substantially as a function of population abundance and/or process conditions. The combined model performed best for abundant taxa in the treatment bioreactors where process conditions were manipulated. In contrast, the neutral model exhibited the best performance for rare taxa. Our analysis further indicated that immigration rates decreased with taxa abundance and competitions between taxa were strongly correlated with phylogeny but within a certain phylogenetic distance only. The determinism underlying taxa and community dynamics were quantitatively assessed, showing greater determinism in the treatment bioreactors which aligned with the subsequent abnormal system functioning. Given its mechanistic basis, the framework developed here is expected to be potentially applicable beyond microbial ecology.

Impact Statement

One fundamental goal in microbial ecology is to predict how microbial diversity is changed across space and time. Although spatial patterns of microbial communities have been recently intensively examined, our understanding of microbial temporal dynamics is rudimentary, primarily due to the lack of appropriate experimental data and theoretical framework. By reconciling niche and neutral perspectives, this study developed a novel process models-based framework to effectively encapsulate microbial species temporal dynamics, which is powerful for quantitatively assessing the assembly mechanisms underlying microbial community dynamics. This study represents a significant advance in explaining microbial temporal dynamics toward predictive microbial community ecology.

Keywords: neutral model; consumer-resource model; species dynamics; community assembly mechanisms.

Introduction

Microorganisms are the most diverse group of life on Earth, and play critical roles in global biogeochemical cycling of carbon, nitrogen, phosphorus, sulfur and various other elements. It is well known that microbial diversity is extremely high across various habitats (1-3). One of the fundamental goals in microbial ecology is to determine how such extremely high microbial biodiversity is generated and maintained across space and time (4). Two types of ecological processes (deterministic vs. stochastic) are influential for explaining the processes of assembling individual taxa into a local community. Niche-based theory assumes that deterministic processes, such as differences in taxonomic and functional traits, interspecies interactions (e.g., competition, predation, and mutualisms), and abiotic filtering (e.g. temperature, pH), are responsible for local community compositions (5, 6). In contrast, neutral theory proposes that all species are ecologically equivalent, thus immigration and ecological drift of stochastic birth and death shape the diversity and composition of local communities regardless of species traits (7). Although both deterministic and stochastic processes are believed to play key roles in shaping community diversity, their relative importance is still hotly debated (6-11), particularly in microbial ecology (4, 12-14). It is thus critical to quantify the extent to which deterministic or stochastic processes influence community assembly in order to influence or even manipulate microbial communities for designed functions (4, 14).

Several major approaches have been used to infer community assembly mechanisms, such as multivariate analysis, null modeling analysis, and ecological theory-based process models (i.e. niche and neutral models) (4, 15). Comparing to the multivariate and null model-based statistical approaches, the ecological theory (niche vs neutral)-based process model approach is more attractive because it allows mechanistic predictions of community dynamic behavior. One of the most widely used niche models is Lotka-Volterra competition (16, 17), which describes the dynamics of individual taxa as a function of growth rate and inter-species interaction. However, such direct effect is rarely analyzed in nature, and the competition coefficients are challenging to measure experimentally (17, 18). Such parameter-rich models are particularly intractable for studying microbial communities that typically exhibit high diversity (19-22). An alternative to the generalized Lotka-Volterra model is the consumer-resource model, which describes the dynamics of individual taxa as a function of the availability of resources. This model assumes that species interact only through competition for a few limiting resources (23, 24), which greatly reduces the number of required parameter from the square of the taxon number (pairwise species interactions) to the number of resources, and hence it is parsimonious for complex systems such as microbial communities (25). Recently, resource-related models have been successfully used for modeling microbial community diversity dynamics (26, 27).

Neutral models have also been successful in explaining some of the most widely studied patterns in community ecology, such as abundance distribution (28), rank-abundance distribution (13) and frequency-abundance distribution of individual taxa (12). However, most studies have focused on community-level predictions at one time point (29-33), but rarely examined the dynamic behavior of individual populations from neutral perspectives (13). This is an important knowledge gap to fill because temporal dynamic behavior is critical for understanding multispecies coexistence (6) and functional stability (34). Also, because both niche and neutral mechanisms play key roles in community assembly (35), several studies attempted to develop unified models to reconcile both mechanisms (6, 8-10). But such theoretical models are rarely applied to actual ecological data owing to mathematical challenges (36, 37). Recently, a stochastic differential equation (SDE)-based model that consolidates niche and neutral processes has been developed to simulate the dynamics of individual microbial taxa (13, 36). Rooted on the framework of neutral model, this SDE model considers the niche effect by incorporating an advantage term as a linear function of various environmental variables (13). However, this SDE model does not account for biotic interactions such as competition.

In this study, we developed a novel process models-based framework to quantitatively infer assembly mechanisms by integrating niche and neutral theory-based models for community dynamics. Specifically, we first developed an SDE-based combined model by incorporating consumer-resource interactions, immigration, and drift. We then compared this new model with the consumer-resource model and neutral model, for the ability to capture the temporal dynamics of individual taxa in anaerobic bioreactors. We estimated ecologically relevant model parameters such as the immigration rate and competition strength, and inferred the relative importance of stochastic vs deterministic processes in driving community dynamics. We applied this framework to analyze time-series data from anaerobic bioreactors under stable or disturbed process conditions. Our results indicated that it provides a robust, reliable process models-based tool for assessing assembly mechanisms controlling taxa and community dynamics.

Materials and Methods

Mathematical framework

Consumer-resource model

Because of its mathematically tractable form, MacArthur's consumer–resource model (38) has a strong impact on the theory of exploitative competition (39). In this study, we use the following equation (27, 40) for its simplicity to describe the consumer-resource interaction of Taxon i:

$$\frac{dN_i}{dt} = \left(\sum_j b_{ij} C_{ij} R_j - m_i\right) N_i \qquad [1]$$

Where N_i is the absolute abundance (i.e. population density, population per unit area) of Taxon i and R_j is the availability of Resource j. C_{ij} is the rate at which Taxon i consumes Resource j, while the quality factor, b_{ij}, represents Taxon i's ability to convert the consumed resource to its biomass. Thus, their product, $b_{ij} C_{ij}$, can represent the competition strength of Taxon i over Resource j. m_i represent the minimum maintenance cost.

The community size, $N_T = \sum_{i=1} N_i$, is implicitly a function of time. For typical microbial community data, N_T is not available. Rather, the relative abundances and the ratios between taxa abundances can be inferred from the compositional datasets (41). We can choose a reference taxon r, and take the ratio of focal taxon and the reference taxon. Let $Z_i = \log \frac{N_i}{N_r}$ be the log-ratio of Taxon i to the reference taxon r. Based on eq [1], we have:

$$\frac{d\log(N_i)}{dt} = \frac{dN_i}{N_i dt} = \sum_j b_{ij} C_{ij} R_j - m_i \qquad [2]$$

Thus:

$$\frac{dZ_i}{dt} = \frac{d\log\frac{N_i}{N_r}}{dt} = \frac{d\log(N_i)}{dt} - \frac{d\log(N_r)}{dt} = \sum_j (b_{ij} C_{ij} - b_{rj} C_{rj}) R_j - (m_i - m_r) \quad [3]$$

In this study of bioreactor dataset, the availability of resource R_j is represented by a single variable, the volatile solids (VS), in the bioreactors. R_j could be represented by other resources when applying this model in other systems. These variables are known at discrete time points. Further, Eq [3] can be expressed as a simple linear model,

$$\frac{dZ_i}{dt} = k_0 + \sum_j k_{1,j} Y_{1,j} \quad [4]$$

Where $k_0 = -(m_i - m_r)$ representing the relative maintenance cost of taxon i as compared to the reference taxon, $k_{1,j} = b_{ij}C_{ij} - b_{rj}C_{rj}$ representing the relative competition strength of taxon i over resource R_j, and $Y_{1,j} = R_j$. We can then estimate the parameters through a least-squares regression analysis based on the measured variables at discrete time points.

The neutral model

In a neutral local community, when an individual dies, it is replaced by an immigrant of Taxon *i* from a source community (i.e., regional species pool) with the probability m_i, or by regeneration from the local community with probability $1 - m_i$. Under the neutral assumption, $m_1 = m_2 = \cdots = m$. We set the mean time for replacement of an individual to be a and define a scaled time $\tau = t/a$. In a short time period $\Delta\tau \to 0$, we can expect only one replacement in the community. The species relative abundances $\boldsymbol{X}$ in a neutral model follows a Wright-Fisher Process (WFP) (42-44), which is defined by the Ito stochastic differential equation (SDE):

$$d\boldsymbol{X} = \lambda(\boldsymbol{p} - \boldsymbol{X})\, d\tau + \sigma(\boldsymbol{X}) d\boldsymbol{W} \quad [5]$$

Where $\boldsymbol{p}$ is the relative abundance of taxa in the metacommunity, $\lambda = N_T m$ is the product of local community size and taxon immigration probability, representing the relative rate of migration from the metacommunity into the local community. $\sigma(\boldsymbol{X})$ is the instantaneous standard deviation of changes in X per unit time. dW is a standard Wiener process term. The quadratic covariation between taxa is given by $\Sigma = \frac{1}{2}\sigma\sigma^T$ where (42-44)

$$\Sigma_{ij} = \begin{cases} X_i(1-X_i) & i = j \\ -X_iX_j & i \neq j \end{cases}$$

The SDE for the focal taxon i is then defined as:

$$dX_i = \lambda(p_i - X_i)\,d\tau + \sigma(X_i)dW_i = \underbrace{\lambda(p_i - X_i)\,d\tau}_{deterministic} + \underbrace{\sqrt{2X_i(1-X_i)}dW_i}_{stochastic} \qquad [6]$$

Where X_i is the relative abundance of taxon i, i.e., $X_i = \frac{N_i}{N_T}$. dW_i is a standard Wiener process term following the standard normal distribution $N(0,1)$. The first term on the right of Eq[6] represents the expect change of X_i, which is a deterministic term; the second term represents the variation of change, which is a stochastic term.

The covariation between taxon i and taxon j $(i \neq j)$ is $E\,[(dX_i - E(dX_i))\,(dX_j - E(dX_j))] = E\,(\sqrt{2X_i(1-X_i)}dW_i \times \sqrt{2X_j(1-X_j)}dW_j)$, which equals to $-2X_iX_j$. This gives us the covariance between the two Wiener processes dW_i and dW_j:

$$\rho = E(dW_i dW_j) = -\sqrt{\frac{X_iX_j}{(1-X_i)(1-X_j)}}. \qquad [7]$$

We can take the log-ratio transformation as $Z_i = \log\frac{N_i}{N_r} = \log\frac{N_i/N_T}{N_r/N_T} = \log\frac{X_i}{X_r}$. Since both X_i and X_r follow the SDE (Eq[6]), the SDE of Z_i is derived based on Ito's lemma:

$$dZ_i = \left[\frac{\partial Z_i}{\partial X_i}\lambda_i(p_i - X_i) + \frac{\partial Z_i}{\partial X_r}\lambda_r(p_r - X_r) + \frac{\partial Z_i}{\partial t}\right] d\tau$$

$$+ \left[\frac{1}{2}\frac{\partial^2 Z_i}{\partial X_i^2}\sigma^2(X_i) + \frac{\partial^2 Z_i}{\partial X_i \partial X_r}\sigma(X_i)\sigma(X_r)\rho + \frac{1}{2}\frac{\partial^2 Z_i}{\partial X_r^2}\sigma^2(X_r)\right] d\tau + \frac{\partial Z_i}{\partial X_i}\sigma(X_i)dW_i$$

$$+ \frac{\partial Z_i}{\partial X_r}\sigma(X_r)dW_r$$

That is,

$$dZ_i = \left[\frac{\lambda_i p_i - 1}{X_i} - \frac{\lambda_r p_r - 1}{X_r} + \lambda_r - \lambda_i\right] d\tau + \sqrt{\frac{2(1-X_i)}{X_i}}\, dW_i - \sqrt{\frac{2(1-X_r)}{X_r}}\, dW_r \qquad [8]$$

Given that $\tau=t/a$, and the covariance between dW_i and dW_r (Eq[7]), the above equation (eq [8]) can be written as a SDE:

$$dZ_i = \underbrace{\frac{1}{a}\left[\frac{\lambda_i p_i - 1}{X_i} - \frac{\lambda_r p_r - 1}{X_r} + \lambda_r - \lambda_i\right] dt}_{deterministic} + \underbrace{\sqrt{\frac{2}{aX_i} + \frac{2}{aX_r}}\, dW_t}_{stochastic} \qquad [9]$$

Where dW_t is a Wiener process term, which follows a normal distribution $N(0, dt)$. Further, Eq [9] can be expressed as a simple linear model,

$$\frac{dZ_i}{dt} = k_0 + k_2 Y_2 + k_3 Y_3 + \varepsilon \qquad [10]$$

where $k_0 = \frac{\lambda_r - \lambda_i}{a}$, $k_2 = \frac{\lambda_i p_i - 1}{a}$, $Y_2 = \frac{1}{X_i}$, $k_3 = -\frac{\lambda_r p_r - 1}{a}$, $Y_3 = \frac{1}{X_r}$ and ε is an error term given by $\varepsilon = \sqrt{\frac{2}{aX_i} + \frac{2}{aX_r}}\frac{dW_t}{dt}$. The parameters can be estimated through a weighted least-squares regression analysis, in which the weights are $\frac{dt}{\frac{2}{X_i} + \frac{2}{X_r}}$. The weighted errors should be normally distributed and the standard residual error of the linear regression model should be $\sqrt{\frac{1}{a}}$. We then estimate the parameter product, $\lambda_i p_i$, based on the coefficient of variable Y_2. Further, p_i can be estimated as the mean relative abundance of taxon i, and λ_i can be derived by dividing the estimated $\lambda_i p_i$ to p_i.

The combined model

A combined model of taxon dynamics was further developed to include both exploitative competition and neutral factors. The term of ‘relative growth’ (can be positive or negative) caused by the resource consuming (eq [3]) is added to the deterministic part of the SDE (eq [9]) without change, since it is purely deterministic which wouldn’t bring in any uncertainty. The combined model is thus given by:

$$dZ_i = \underbrace{\left[\frac{\lambda_i p_i - 1}{aX_i} - \frac{\lambda_r p_r - 1}{aX_r} + \frac{\lambda_r}{a} - \frac{\lambda_i}{a} + \sum_j (b_{ij}C_{ij} - b_{rj}C_{rj})R_j - (m_i - m_r)\right] dt}_{deterministic} + \underbrace{\sqrt{\frac{2}{aX_i} + \frac{2}{aX_r}}\, dW_t}_{stochastic} \qquad [11]$$

Further, Eq [11] can be expressed as a simple linear model,

$$\frac{dZ_i}{dt} = k_0 + \sum_j k_{1,j} Y_{1,j} + k_2 Y_2 + k_3 Y_3 + \varepsilon \qquad [12]$$

Where $Z_i = \log \frac{X_i}{X_r}$ is the log ratio of the relative abundance of taxon i to the reference taxon r. $k_0 = \frac{\lambda_r}{a} - \frac{\lambda_i}{a} + m_r - m_i$, $k_{1,j} = b_{ij}C_{ij} - b_{rj}C_{rj}$ representing the relative competition strength of taxon i on resource R_j, and $Y_{1,j} = R_j$. $k_2 = \frac{\lambda_i p_i - 1}{a}$, $Y_2 = \frac{1}{X_i}$, $k_3 = -\frac{\lambda_r p_r - 1}{a}$, $Y_3 = \frac{1}{X_r}$ and ε is an error term given by $\varepsilon = \sqrt{\frac{2}{aX_i} + \frac{2}{aX_r}}\frac{dW_t}{dt}$. The parameters can be estimated through a weighted least-squares regression analysis, in which the weights are $\frac{dt}{\frac{2}{X_i} + \frac{2}{X_r}}$. The weighted errors should be normally distributed and the standard residual error of the linear regression model should be $\sqrt{\frac{1}{a}}$.

p_i can be estimated as the mean relative abundance of taxon i. We can estimate the parameters k_0, k_1, k_2 and k_3 in the linear model, by which the model parameters $b_{ij}C_{ij} - b_{rj}C_{rj}$, λ_i and p_i can be further derived.

Determinism

The SDE of the combined model (eq [11]) can be written as

$$dZ = \underbrace{\mu d\tau}_{deterministic} + \underbrace{\sigma dW}_{stochastic}$$

Where μ is the expected change of Z per unit time and σ is the instantaneous standard deviation of changes in Z per unit time. dW is a standard Wiener process term. We define taxa determinism as the inverse of the variation coefficient, that is,

$$determinism = \frac{\mu}{\sigma} \qquad [13]$$

After parameter estimation using weighted least-squares regression analysis, the taxa determinism can be calculated for each taxon at each time point based on eq [13]. For the combined model, the determinism of taxon i can be calculated based on parameters of the linear model eq [12]:

$$determinism = \frac{(k_0 + \sum_j k_{1,j} R_j + \frac{k_2}{X_i} + \frac{k_3}{X_r}) \times a}{\sqrt{\frac{2}{X_i} + \frac{2}{X_r}}} \qquad [14]$$

Note that the stochasticity is calculated on the scaled time unit τ. Then, the community-level determinism is calculated as the mean determinism among taxa, either weighted by the relative abundance of each taxon (weighted determinism) or not (unweighted determinism).

Anaerobic bioreactor operation and 16S rRNA gene sequencing

The operation of anaerobic bioreactors, biomass sampling and chemical analyses were processed as previously described (45). In brief, two sets of triplicated, continuous anaerobic bioreactors (i.e., the control bioreactors C1, C2 and C3, and the treatment bioreactors D1, D2 and D3) were operated at 35 ºC and fed at 4-hr intervals, each with a working volume of 3.6 L. The control bioreactors were fed with dairy manure at a constant rate and continuously operated for 501 days, which showed a stable organic matter level (Fig. S1a). The treatment bioreactors were operated for 100 days before they collapsed by supplementing incremental poultry waste, resulting in higher ammonia toxicity (Fig. S1b). Sludge samples were generally taken every 3 to 10 days from each bioreactor, resulting in 53 time points for control and 11 time points for treatment bioreactors.

DNA extraction and 16S rRNA gene sequencing were processed as previously described (45). In brief, biomass samples were subjected to suspension in 630 µL of DNA-extraction buffer, subsequently undergoing treatment with a lysozyme mixture (60 µL, 37 °C, 60 min), a protease mixture (60 µL, 37 °C, 30 min), and 20% sodium dodecyl sulfate (80 µL, 37 °C, 90 min). The treated sample suspension was then extracted using phenol-chloroform-isoamyl alcohol (25:24:1) at 65 °C for 20 min, followed by extraction with chloroform-isoamyl alcohol (24:1) to obtain a supernatant. Further, DNA extract was combined with 0.6 volume of isopropanol and stored overnight at 4 °C; DNA was obtained through centrifugation followed by washing with 70% cold ethanol, drying, and resuspension in nuclease-free water. The purity and concentration of DNA were subsequently assessed utilizing a NanoDrop spectrophotometer (NanoDrop Technologies Inc., Wilmington, DE, USA). The V4 region of microbial 16S rRNA gene was amplified by primer pairs of 515F and 806R (46). PCR amplicon sequencing was conducted on the MiSeq Illumina platform at the Institute for Environmental Genomics (IEG), University of Oklahoma. Sequences were processed to generate exact sequence variants (ESVs) by UNOISE3 (47) at the 100% sequence similarity threshold. ESVs with fewer than eight reads were removed using the default '-minsize' values. Taxonomy was assigned with a confidence cutoff of 50% using the RDP classifier (48). The reference taxon was then chosen as the one with the top frequency and relative abundance, which was ESV1 that were detected at all time points.

Since there were only 11 time points for each treatment bioreactor, we combined the time series of the triplicate bioreactors together to improve the liability of model fitting. For example, if the dependent variable (as for eq [4], eq [10] & eq [12]) of one taxon in treatment bioreactor D1 is $\left(\frac{dZ_i}{dt}\right)_{D1} = [\left(\frac{z_{i,t2}-z_{i,t1}}{t2-t1}\right)_{D1,1}, \dots, \left(\frac{z_{i,t11}-z_{i,t10}}{t11-t10}\right)_{D1,10}]$; the dependent variable of this taxon in D2 is $\left(\frac{dZ_i}{dt}\right)_{D2} = [\left(\frac{z_{i,t2}-z_{i,t1}}{t2-t1}\right)_{D2,1}, \dots, \left(\frac{z_{i,t11}-z_{i,t10}}{t11-t10}\right)_{D2,10}]$ and that in D3 is $\left(\frac{dZ_i}{dt}\right)_{D3} = [\left(\frac{z_{i,t2}-z_{i,t1}}{t2-t1}\right)_{D3,1}, \dots, \left(\frac{z_{i,t11}-z_{i,t10}}{t11-t10}\right)_{D3,10}]$, then the dependent variable for the combined time-series is $\left(\frac{dZ_i}{dt}\right)_{D} = [\left(\frac{dZ_i}{dt}\right)_{D1}, \left(\frac{dZ_i}{dt}\right)_{D2}, \left(\frac{dZ_i}{dt}\right)_{D3}] = [\left(\frac{z_{i,t2}-z_{i,t1}}{t2-t1}\right)_{D1,1}, \dots, \left(\frac{z_{i,t11}-z_{i,t10}}{t11-t10}\right)_{D1,10}, \left(\frac{z_{i,t2}-z_{i,t1}}{t2-t1}\right)_{D2,1}, \dots, \left(\frac{z_{i,t11}-z_{i,t10}}{t11-t10}\right)_{D2,10}, \left(\frac{z_{i,t2}-z_{i,t1}}{t2-t1}\right)_{D3,1}, \dots, \left(\frac{z_{i,t11}-z_{i,t10}}{t11-t10}\right)_{D3,10}]$. Similarly, the independent variables can be combined in the same way. The combined dependent and independent variables for the treatment bioreactors were then used for the linear regression analyses based on the least-squares method. We note that this is not a standard way to apply the model fitting for common time-series data. Yet, this combination method may provide an option for replicated time-series. In fact, fluctuations in microbial community compositions were highly consistent for the three replicated treatment bioreactors (Fig. S1c), which enabled us to test the dynamical pattern of microbial taxa based on the combined time-series.

Results

Overview of modeling framework

To assess the mechanisms controlling community dynamics, raw time-series sequence data are first processed to generate relative abundances of individual taxa represented as exact sequence variants (ESVs) (Fig. 1, i). The reference taxon is chosen as the one with the top frequency and relative abundance, and the ratio of taxa abundance to the abundance of the reference taxon is then calculated for each taxon. The observed time-series data of each taxon are then fitted with the neutral, consumer-resource, and combined models, respectively (Fig. 1, ii). The performances of the three models are compared according to the Akaike information criteria (AIC) values, aiming to reveal potential mechanisms driving the dynamics of individual taxa. Ecologically important parameters, such as λ_i (the rate of migration from the metacommunity into the local community) and $b_i C_i - b_r C_r$ (relative competition strength to the resource), are estimated using the least-square method for model fitting (Fig. 1, iii). Finally, the determinism for taxa and community dynamics are assessed based on the SDEs of the combined model (Fig. 1, iv), as the SDEs comprise the deterministic and stochastic part. It is noted that, while the immigration is generally considered as a stochastic process (15), it is included in the deterministic part of the SDEs (eq [6], eq [9] & eq [11]). In fact, the immigration process acts as a restoring force which makes the relative abundance return to its mean value when there is a deviation between the current relative abundance and the mean relative abundance.

Model performances on taxon dynamics

To illustrate how the process model-based framework (Fig. 1) is applied to microbial time-series data, we collected longitudinal data from two contrasting sets of anaerobic bioreactors, each with three replicates: There were a total of 53 time points from the control bioreactors in which stable process conditions were maintained over 500 days, and 11 time points from the treatment bioreactors over 100 days during which the resource levels were incrementally raised until process conditions deteriorated to an ultimate collapse. A total of 6,799 microbial taxa, represented by ESVs, were detected, which were present in at least one sample in control or treatment bioreactors.

Further, models were fitted using the least-squares method for each taxon under control or treatment conditions, requiring the taxon to present in at least six time points (for example, a taxon present in at least 6 out of 53 time points in bioreactor C1). Specifically, we combined the time series of the triplicate treatment bioreactors together to improve the liability of model fitting (see Methods for details), and fitted the models on taxa which were present in at least 6 out of 33 time points in treatment bioreactors. In addition, the mean relative abundance of each taxon in control or treatment bioreactors was calculated, based on which taxa were classified into three groups: the abundant taxa (mean relative abundance > 0.1%), the moderate taxa (mean relative abundance between 0.01% and 0.1%), and the rare taxa (mean relative abundance < 0.01%) (Table S1).

To identify the mechanisms driving the dynamics of individual taxon, the relative performances of the three models were compared based on AIC values. In the treatment bioreactors, the combined model had the best fit for 58% of the abundant taxa (Fig. 2a), suggesting that most abundant taxa were driven by both stochastic drift and deterministic immigration and competition. In contrast, the neutral model had best fit for 38% of the abundant taxa, and the consumer-resource model had best fit for only 4% of the abundant taxa. For rare taxa, 58% of them in the treatment bioreactors found best fit with the neutral model, suggesting that rare taxa were mainly shaped by immigration and drift. The importance of neutral processes was even more conspicuous in the control bioreactors, since the neutral model had the best fit for 79% of all abundant taxa and 74% of rare taxa. Therefore, neutral processes of immigration and drift were identified to drive the dynamics of the majority of rare taxa, particularly in the control bioreactors. When examining the model performance for the entire community, the neutral model had the best fit for most taxa in both the control (75% of all taxa) and treatment bioreactors (57% of all taxa) (Fig. S2a), which was expected as the rare taxa contributed to the majority of the taxa number (Table S1).

Model performance was further examined across the major phyla. The neutral model was the best for most rare taxa in both treatment and control bioreactors regardless of phylogenetic relationships (Fig. S3), while the combined model performed better than the other two models for the abundant taxa in treatment bioreactors for 5 out of the top 7 phyla such as *Firmicutes* and *Bacteroidetes* (Fig. S3). These results suggested that model performance was largely unrelated to microbial phylogeny.

Because the combined model includes both the neutral and consumer-resource interaction terms, the R^2 values from the least square-squares fitting are almost always the largest for the combined model (Fig. 2b). On average, the combined model can explain 36% ± 20% (mean ± s.d.) of the variations in taxon dynamics based on the R^2 values, while the neutral model can explain 31% ± 19% and the consumer-resource model can only explain 4% ± 8% of the variations (Fig. S2b). Regarding the ability to represent taxon dynamics under different treatment conditions, the neutral model could explain more variations of the abundant taxa in the control than the treatment bioreactors (mean R^2 value: 22% v.s. 16%; $P < 0.0001$ by two-tailed *t*-test) (Fig. 2b). It also performed better on the rare taxa in the control than the treatment bioreactors (mean R^2 value: 36% v.s. 32%; $P < 0.0001$ by two-tailed *t*-test). In contrast, the consumer-resource model or the combined model was able to represent taxon dynamics in the treatment bioreactors better than those in the control bioreactors, as the mean R^2 values were significantly higher in the treatment than the control bioreactors for abundant, moderate, as well as rare taxa ($P < 0.02$ by two-tailed *t*-test). Therefore, the relative performance of these three models are dependent on taxa abundance and process conditions in the ecosystem of interest.

Competition strengths among different taxa

Ecologically important parameters such as $b_iC_i - b_rC_r$ reflecting the relative competition strength can be estimated with relative taxon abundance data at discrete time points, based on the consumer-resource model or the combined model. Considering the better performance of the combined model than the consumer-resource model, here the parameters were estimated based on the combined model to enable the comparison across taxa, which were summarized in Table S2. The top three most competitive taxa in the treatment bioreactors were identified to be associated with the genera *Ornithinicoccus*, unclassified *Ruminococcaceae* and *Gottschalkia*, suggesting them as strong competitors for the organic substrates.

It is curious whether phylogenetically closely related taxa are more likely to have similar competition strengths. Thus, we examined the relationship between taxa phylogeny and the estimated relative competition strength. When the sequence similarity between taxa was larger than 70%, the difference in b_iC_i had a significant negative correlation with sequence similarity in the treatment bioreactors (Spearman's *rho* = -0.04, $P < 0.0001$) (Fig. 3a), suggesting that closely related microbial taxa had similar competition strengths (i.e., phylogenetic signal) when resource levels were altered. The negative correlation between competition strength difference and sequence similarity robustly held under even higher sequence similarity (Spearman's *rho* = -0.04, $P < 0.0001$ for sequence similarity > 80% and Spearman's *rho* = -0.07, $P = 0.003$ for sequence similarity > 90%). However, such negative correlation did not hold when sequence similarity of the 16S rRNA gene was less than 70% (Spearman's *rho* = 0.03 for treatment bioreactors). For control bioreactors, the negative correlation between sequence similarity and the difference in b_iC_i was observed when sequence dissimilarity was larger than 85% (Spearman's *rho* = -0.06, $P < 0.0001$) but not below that threshold (Fig. 3a). Therefore, the phylogenetic signal of resource competition strengths is relevant only within certain phylogenetic distances. It is also noted that, although significant, the correlations were weak (absolute Spearman's *rho* < 0.1), suggesting that phylogeny could only explain a minor proportion of variations in taxa resource competition strengths.

Since the mean $b_i C_i$ difference of microbial taxa was substantially larger in control bioreactors (0.21± 0.19, mean ± s.d.) than that in treatment bioreactors (0.16± 0.14, mean ± s.d.), microbial responses to resource levels were more predictable in the treatment bioreactors, where changes in resource levels could lead to greater environmental selection. As a result, temporal dynamics of closely related ESVs was more similar in the treatment bioreactors than the control bioreactors. For example, ESV4 and ESV 221, which are 98.82% similar in sequence, belong to the same genus T78 of family *Anaerolineaceae*. The temporal dynamics of their relative abundance were not correlated (Pearson's $r = 0.17$, $P = 0.36$) in the control bioreactors (Fig. 3b, 3d) but significantly correlated (Pearson's $r = 0.54$, $P = 0.001$) in the treatment bioreactors (Fig. 3c, 3e).

Negative correlation between immigration rates and taxa abundances

The neutral model presented the best fit for most taxa in the control bioreactors (Fig. 2a). We further examined how the estimated λ_i, which represented the immigration rates, varied across all taxa. The estimated relative immigration rates were similar for the same ESVs across triplicate bioreactors but highly different among various taxa, ranging in 10^4 folds. The estimated values of λ_i were negatively and significantly (Spearman's *rho* = -0.95 ~ -0.92, $P < 0.0001$) correlated with the average relative abundances of ESVs (Fig. 4a). Furthermore, the estimated λ_i values were highly variable within each phylum because they were negatively dependent on taxa abundance

(Fig. S4), suggesting that the estimated immigration rates were related to abundance but not phylogeny.

The probability density distribution of individual taxon abundance under equilibrium can be derived for the neutral model (12). Such abundance distribution is not possible for the consumer-resource or the combined model because taxon dynamics is dependent on the resource variable in these models. The probability density distributions of the relative abundances of an ESV can be predicted by λ_i and p_i (the relative abundance of that ESV in the source community) in the neutral model, which were shown to follow a beta distribution (12). Exemplified by the distributions of relative abundances for several representative ESVs ranging from abundant to rare ones in the control bioreactors, the beta distributions predicted the dynamics of ESVs well, with much higher λ_i values for the rarer taxa (Figure 4b & Fig. S5). These results suggested that the neutral model could be used to predict the range of fluctuation for each microbial taxon under equilibrium, which may be valuable for assessing the boundaries of population abundance in a stable microbial community.

Higher determinism in the treatment bioreactors

The determinism of taxa at certain time points was calculated based on the parameters estimated of the combined model using the above-mentioned approach (Fig. 1). Interestingly, taxa determinism showed significant negative correlation with the mean relative abundance of taxa in both control (Spearman's *rho* = -0.53, $P < 0.0001$) and treatment bioreactors (Spearman's *rho* = -0.55, $P < 0.0001$), suggesting that rare taxa tended to be more predictable than abundant taxa. Further, the mean taxa determinism was higher in treatment than control bioreactors for abundant (mean determinism: 16 v.s. 13; $P < 0.0001$ by two-tailed *t*-test), moderate (mean determinism: 57

v.s. 54; $P = 0.01$ by two-tailed t-test) and rare taxa (mean determinism: 196 v.s. 152; $P < 0.0001$ by two-tailed t-test) (Fig. 5a).

The community-level determinism was further derived by aggregating the determinism of co-occurring taxa within the community. The abundance-weighted and unweighted community determinism were not different between the control and treatment bioreactors before Day 90 (P =0.06 ~ 0.94 by two-tailed t-test on each time point) (Fig. 5b). On Day 90, the mean weighted community determinism of treatment bioreactors was significantly higher than that of controls (P =0.02 by two-tailed t-test). On Day 97 which was prior to the collapse of treatment bioreactors, both the weighted and unweighted community determinism were substantially higher in the treatment bioreactors than controls (P =0.004 for weighted community determinism and P =0.04 for unweighted community determinism by two-tailed t-test) (Fig. 5b), indicating stronger selection in the treatment bioreactors.

Discussion

Untangling ecological processes controlling community dynamics is a major challenge in microbial ecology, primarily due to the lack of appropriate theoretical framework and long-term time-series datasets (13, 49). With recent advances of genomics technology, massive longitudinal data can be rapidly obtained across different environmental conditions (50), which offer great opportunities for testing microbial ecological theories (15, 51). Here, we described a novel process models-based framework, to quantitatively assess the assembly mechanisms controlling community dynamics. Different from statistical approaches such as VPA (52, 53) and null model-based methods (15, 51, 54, 55), the process models are mechanistically developed to enable the prediction of community dynamics and their underlying mechanisms. Our analyses demonstrate that this framework could discern the relative importance of deterministic processes (immigration, resource competition) and stochastic process of drift in driving taxa and community dynamics. The developed framework represents a significant advance in reconciling both niche and neutral theories for predicting community dynamics and underlying mechanisms toward predictive microbial ecology, the ultimate goal in this field.

Microbial rarity can result from both stochastic and deterministic processes (56). For instance, low local abundance can emerge by stochastic population fluctuation. A recently immigrated taxon might also be rare when it first enters a new community. Niche processes, including abiotic and biotic factors, can have crucial roles in driving taxon rarity. Rare biosphere members can be ascribed to narrow niche breadth, thus remaining generally inactive and at low density in most conditions but becoming dominant when favorable conditions arise (57, 58), which is best illustrated by the extreme case of microbial dormancy. An alternative is the competition-colonization trade-off hypothesis, which is rooted in the classic niche-based ecology predicting that taxa with low competitive ability may remain rare rather than going extinct due to the advantage in immigration and colonization (59, 60). Since microbial dynamics are very fast, competitive exclusion may not have sufficient time to play out (61). Our study suggested that immigration played important roles in driving community dynamics, especially for rare taxa (Fig. 4). Rare microbial populations were shown to have the best fit to the neutral model in both control and treatment bioreactors (Fig. 2a), indicating a dominant role of immigration and drift in shaping rare taxa dynamics, consistent with the observation that ecological drift was pronounced for rare planktonic eukaryotes (62). Further, the estimated relative immigration rate was higher for rare taxa than abundant taxa (Fig. 4a). This also supports the competition-colonization trade-off hypothesis that rare taxa are recruited mainly through immigration (58, 63). It was noted that the determinism of rare taxa was higher than abundant taxa (Fig. 5a), which could be explained by their immigration rate. Higher immigration rate of a taxon would result in less variations in its relative abundances, as the taxon tend to return to its correspondent relative abundance in the metacommunity (12), i.e., higher determinism of taxa dynamics. In contrast, taxa with low immigration rate are less affected by the metacommunity, which may be subject to larger effects of local drift and result in more variations in their relative abundances.

Deterministic processes of resource competition might play an important role in shaping the dynamics of abundant taxa in treatment bioreactors, consistent with the resource-related theory. The resource ratio-theory successfully explained the 'paradox of enrichment' in sludge bioreactors, i.e., higher resource level of nitrogen and oxygen initially increased and then decreased the diversity of the ammonia oxidizing bacteria (26), as a result of competition among multiple taxa with different resource-ratio requirements. A modified consumer-resource model to include nonspecific cross-feeding interactions explained experimental results that many microbial taxa could co-exist in a single-resource environment (27). Exploitative competition, rooted in the consumer-resource model, significantly contributed to abundant taxa dynamics in the disturbed environment (Fig. 2a), possibly because increases in resources stimulated the competition among abundant microbial populations. As a result, the determinism at the community level was significantly higher in the treatment bioreactors as compared to the controls (Fig. 5b).

The estimated competition strengths showed stronger phylogenetic signal in the treatment than control bioreactors (Fig. 3a). Temporal dynamics patterns of closely related ESVs were more similar in treatment bioreactors than controls (Fig. 3b), resonating with the physics-based theory that views microbial community as a fully disordered background with unstructured individuals (i.e., behaviors of individuals are not clustered by their taxonomic identity) (64), and that imposing disturbance will order the disordered individuals based on traits, resulting in ecological clusters that are disturbance-dependent.

Understanding the mechanisms underlying community assembly is important not only to ecologists but also to practitioners. The relative importance of deterministic vs stochastic processes in controlling microbial community assembly has attracted increasingly interests in the last several years (4). Since the treatment reactors were operated under fluctuated resource levels (45), the microbial communities in treatment reactors appear more filtered compared to the control reactors under stable operating conditions, resulting in higher determinism. Our findings that deterministic processes are more important for controlling the taxa and community dynamics in the treatment reactors (Fig. 5) are highly consistent with this expectation. In addition, the knowledge learned in this study could help environmental engineers maintain microbial systems for desired functions. For example, the neutral model could predict how taxa fluctuate in the control bioreactors (exemplified in Fig. 4b). Given its simplicity, the neutral model could be useful in long-term monitoring of stable systems such as wastewater treatment plants and human guts. The deviation of certain taxa from the predicated range may signify abnormal conditions of the system. Also, the increase of community determinism could provide early warnings for the system functional instability, as exemplified by the treatment bioreactors prior to system collapse (Fig. 5b). The relative competition strengths inferred from the consumer-resource model or the combined model can be used to identify functionally important taxa. Since abundant microbial populations play significant roles in biogeochemical cycling in ecosystems (65), it is interesting to examine how changes in such functionally important taxa would affect resources such as the carbon pool by considering the coupled dynamics of resource and consumer under the framework of ecological stoichiometry (66).

In this study, we demonstrated the applicability of the novel modeling framework in representing the bacterial community dynamics of anaerobic bioreactors. Given its mechanistic basis, the framework developed in this study is expected to be potentially applicable in other ecosystems such as soils, oceans and guts, and also to other organisms such as eukaryotic microorganisms and plants. We expect the neutral model to be an appropriate tool for modeling taxa dynamics in relatively stable environments such as human guts, while the combined model might be better for the abundant taxa in ecosystems with fluctuate resource levels such as soils. However, the performance of different models, as well as the driving forces governing taxa dynamics in different ecosystems remain to be tested. It is also noted that these models possess certain limitations. For example, the resource level is assumed to linearly affect the taxa growth in the consumer-resource model and the combined model, which may not capture the complicated interaction between consumer and resource in nature. In addition, to achieve reliable parameter estimation for the SDE-based models, extensive time-series data of high frequency and duration must be collected, which often entails significant time and effort.

Acknowledgements

The initial development of the theoretical framework for modeling species dynamics (Phase I) was supported by National Science Foundation of China to Y.Y. (41825016). The further improvement of the theoretical framework for assessing community-level determinism (Phase II) was supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research (DOE-BER) (DE-SC0014079, DE-SC0016247, and DE-SC0020163) to J.Z.; also part of ENIGMA- Ecosystems and Networks Integrated with Genes and Molecular Assemblies (http://enigma.lbl.gov), a Scientific Focus Area Program at Lawrence Berkeley National Laboratory, supported by DOE-BER under contract number DE-AC02-05CH11231; and by the U.S. National Science Foundation (NSF) (EF-1065844, EF-2025558). The maintenance of the bioreactors was partly supported by a U.S. Environmental Protection Agency Grant XA-83539201, and the Science Alliance—Tennessee Center of Excellence, to Q.H.. The China Scholarship Council (CSC) provided support for L.W., Q.G., and H.Y..

Author contributions

All authors contributed the intellectual input and assistance to this study and manuscript preparation. L.W., Y.Y. and J.Z. conceived the research questions. L.W., D.N. and J.Z. developed the mathematical framework. Q.G., H.Y., Q. H., and S.C. contributed the experimental data. L.W. performed statistical analysis with help from N.X and B.Y.Z.. L.W. and J.Z. wrote the paper with inputs from Y.Y. and Q.H..

Ethics Statement

No animals and human were involved in this study.

Competing financial interests

The authors declare no competing interests.

Data and code availability

Sequence data are accessible in the GenBank database under the accession number SRP070491. R codes on the modeling and statistical analyses are available at https://github.com/Linwei-Wu/species_dynamics_models.

References

1. Wu L, Ning D, Zhang B, Li Y, Zhang P, Shan X, et al. Global diversity and biogeography of bacterial communities in wastewater treatment plants. Nature Microbiology. 2019;4:1183–95.
2. Locey KJ, Lennon JT. Scaling laws predict global microbial diversity. Proceedings of the National Academy of Sciences. 2016;113(21):5970-5.
3. Thompson LR, Sanders JG, McDonald D, Amir A, Ladau J, Locey KJ, et al. A communal catalogue reveals Earth's multiscale microbial diversity. Nature. 2017;551(7681):457-63.
4. Zhou J, Ning D. Stochastic Community Assembly: Does It Matter in Microbial Ecology? Microbiology and Molecular Biology Reviews. 2017;81(4):e00002-17.
5. Chesson P. Mechanisms of maintenance of species diversity. Annual review of Ecology and Systematics. 2000;31(1):343-66.
6. Tilman D. Niche tradeoffs, neutrality, and community structure: a stochastic theory of resource competition, invasion, and community assembly. Proceedings of the National Academy of Sciences of the United States of America. 2004;101(30):10854-61.
7. Hubbell SP. The unified neutral theory of biodiversity and biogeography. Levin SA, Horn HS, editors. Princeton, New Jersey: Princeton University Press; 2001 2001. 375 p.
8. Leibold MA, McPeek MA. Coexistence of the niche and neutral perspectives in community ecology. Ecology. 2006;87(6):1399-410.
9. Gravel D, Canham CD, Beaudet M, Messier C. Reconciling niche and neutrality: the continuum hypothesis. Ecology letters. 2006;9(4):399-409.
10. Adler PB, HilleRisLambers J, Levine JM. A niche for neutrality. Ecology letters. 2007;10(2):95-104.
11. Chisholm RA, Pacala SW. Niche and neutral models predict asymptotically equivalent species abundance distributions in high-diversity ecological communities. Proceedings of the National Academy of Sciences. 2010;107(36):15821-5.
12. Sloan WT, Lunn M, Woodcock S, Head IM, Nee S, Curtis TP. Quantifying the roles of immigration and chance in shaping prokaryote community structure. Environmental microbiology. 2006;8(4):732-40.
13. Ofiţeru ID, Lunn M, Curtis TP, Wells GF, Criddle CS, Francis CA, et al. Combined niche and neutral effects in a microbial wastewater treatment community. Proceedings of the National Academy of Sciences. 2010;107(35):15345-50.
14. Liu Z, Cichocki N, Hübschmann T, Süring C, Ofiţeru ID, Sloan WT, et al. Neutral mechanisms and niche differentiation in steady-state insular microbial communities revealed by single cell analysis. Environmental microbiology. 2019;21(1):164-81.
15. Ning D, Yuan M, Wu L, Zhang Y, Guo X, Zhou X, et al. A quantitative framework reveals ecological drivers of grassland microbial community assembly in response to warming. Nature Communications. 2020;11(1):4717.
16. Malcai O, Biham O, Richmond P, Solomon S. Theoretical analysis and simulations of the generalized Lotka-Volterra model. Physical Review E. 2002;66(3):031102.
17. Tilman D. Interspecific competition and multispecies coexistence. In 'Theoretical Ecology: Principles and Applications'.(Eds. RM May and AR McLean.) pp. 84–97. Oxford, UK: Oxford University Press; 2007.
18. Chesson P. MacArthur's consumer-resource model. Theoretical Population Biology. 1990;37(1):26-38.
19. Bucci V, Tzen B, Li N, Simmons M, Tanoue T, Bogart E, et al. MDSINE: Microbial Dynamical Systems INference Engine for microbiome time-series analyses. Genome biology. 2016;17(1):121.
20. Marino S, Baxter NT, Huffnagle GB, Petrosino JF, Schloss PD. Mathematical modeling of primary succession of murine intestinal microbiota. Proceedings of the National Academy of Sciences. 2014;111(1):439-44.

21. Gibbons SM, Kearney SM, Smillie CS, Alm EJ. Two dynamic regimes in the human gut microbiome. PLoS computational biology. 2017;13(2):e1005364.
22. Dam P, Fonseca LL, Konstantinidis KT, Voit EO. Dynamic models of the complex microbial metapopulation of lake mendota. NPJ systems biology and applications. 2016;2(1):1-7.
23. Ramakrishna R, Ramkrishna D, Konopka AE. Microbial growth on substitutable substrates: Characterizing the consumer-resource relationship. Biotechnology and bioengineering. 1997;54(1):77-90.
24. MacArthur R, Levins R. Competition, habitat selection, and character displacement in a patchy environment. Proceedings of the National Academy of Sciences. 1964;51(6):1207-10.
25. Tilman D. Resource competition and community structure: Princeton university press; 1982.
26. Bellucci M, Ofiţeru ID, Beneduce L, Graham DW, Head IM, Curtis TP. A preliminary and qualitative study of resource ratio theory to nitrifying lab-scale bioreactors. Microbial biotechnology. 2015;8(3):590-603.
27. Goldford JE, Lu N, Bajić D, Estrela S, Tikhonov M, Sanchez-Gorostiaga A, et al. Emergent simplicity in microbial community assembly. Science. 2018;361(6401):469-74.
28. Volkov I, Banavar JR, Hubbell SP, Maritan A. Neutral theory and relative species abundance in ecology. Nature. 2003;424(6952):1035-7.
29. Keil P, Herben T, Rosindell J, Storch D. Predictions of Taylor's power law, density dependence and pink noise from a neutrally modeled time series. Journal of theoretical biology. 2010;265(1):78-86.
30. McGill BJ. A test of the unified neutral theory of biodiversity. Nature. 2003;422(6934):881-5.
31. Turnbull LA, Manley L, Rees M. Niches, rather than neutrality, structure a grassland pioneer guild. Proceedings of the Royal Society B: Biological Sciences. 2005;272(1570):1357-64.
32. Volkov I, Banavar JR, He F, Hubbell SP, Maritan A. Density dependence explains tree species abundance and diversity in tropical forests. Nature. 2005;438(7068):658-61.
33. Dornelas M, Connolly SR, Hughes TP. Coral reef diversity refutes the neutral theory of biodiversity. Nature. 2006;440(7080):80-2.
34. De Schryver P, Vadstein O. Ecological theory as a foundation to control pathogenic invasion in aquaculture. The ISME journal. 2014;8(12):2360-8.
35. Stokes C, Archer S. Niche differentiation and neutral theory: an integrated perspective on shrub assemblages in a parkland savanna. Ecology. 2010;91(4):1152-62.
36. Li J, Shen X. An improved neutral community model for temporal observations in microbial communities. Ecological Modelling. 2018;388:108-14.
37. Etienne RS, Alonso D, McKane AJ. The zero-sum assumption in neutral biodiversity theory. Journal of theoretical biology. 2007;248(3):522-36.
38. MacArthur R. Species packing and competitive equilibrium for many species. Theoretical population biology. 1970;1(1):1-11.
39. Carroll IT, Cardinale BJ, Nisbet RM. Niche and fitness differences relate the maintenance of diversity to ecosystem function. Ecology. 2011;92(5):1157-65.
40. Advani M, Bunin G, Mehta P. Statistical physics of community ecology: a cavity solution to MacArthur's consumer resource model. Journal of Statistical Mechanics: Theory and Experiment. 2018;2018(3):033406.
41. Morton JT, Marotz C, Washburne A, Silverman J, Zaramela LS, Edlund A, et al. Establishing microbial composition measurement standards with reference frames. Nature communications. 2019;10(1):1-11.
42. Washburne AD, Burby JW, Lacker D. Novel covariance-based neutrality test of time-series data reveals asymmetries in ecological and economic systems. PLoS computational biology. 2016;12(9):e1005124.
43. Fisher RA. The genetical theory of natural selection. Oxford: Clarendon Press; 1930.

44. Wright S. Evolution in Mendelian populations. Genetics. 1931;16(2):97.
45. Wu L, Yang Y, Chen S, Jason Shi Z, Zhao M, Zhu Z, et al. Microbial functional trait of rRNA operon copy numbers increases with organic levels in anaerobic digesters. The ISME journal. 2017;11(12):2874-8.
46. Wu L, Wen C, Qin Y, Yin H, Tu Q, Van Nostrand JD, et al. Phasing amplicon sequencing on Illumina Miseq for robust environmental microbial community analysis. BMC microbiology. 2015;15(1):125.
47. Edgar RC. Accuracy of taxonomy prediction for 16S rRNA and fungal ITS sequences. PeerJ. 2018;6:e4652.
48. Wang Q, Garrity GM, Tiedje JM, Cole JR. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. Applied and environmental microbiology. 2007;73(16):5261-7.
49. Faust K, Lahti L, Gonze D, De Vos WM, Raes J. Metagenomics meets time series analysis: unraveling microbial community dynamics. Current opinion in microbiology. 2015;25:56-66.
50. Zhou J, He Z, Yang Y, Deng Y, Tringe SG, Alvarez-Cohen L. High-Throughput Metagenomic Technologies for Complex Microbial Community Analysis: Open and Closed Formats. mBio. 2015;6(1):e02288-14.
51. Ning D, Deng Y, Tiedje JM, Zhou J. A general framework for quantitatively assessing ecological stochasticity. Proceedings of the National Academy of Sciences. 2019;116(34):16892-8.
52. Legendre P, Borcard D, Peres-Neto PR. Analyzing beta diversity: Partitioning the spatial variation of community composition data. Ecological Monographs. 2005;75(4):435-50.
53. Hanson CA, Fuhrman JA, Horner-Devine MC, Martiny JBH. Beyond biogeographic patterns: processes shaping the microbial landscape. Nature Reviews Microbiology. 2012;10(7):497-506.
54. Gotelli NJ, Graves GR, Gotelli NJ, Graves GR. Null models in ecology. Washington, DC: Smithsonian Institution Press; 1996. 368 p.
55. Zhou J, Deng Y, Zhang P, Xue K, Liang Y, Van Nostrand JD, et al. Stochasticity, succession, and environmental perturbations in a fluidic ecosystem. Proceedings of the National Academy of Sciences. 2014;111(9):E836-E45.
56. Jousset A, Bienhold C, Chatzinotas A, Gallien L, Gobet A, Kurm V, et al. Where less may be more: how the rare biosphere pulls ecosystems strings. The ISME journal. 2017;11(4):853.
57. Aanderud ZT, Jones SE, Fierer N, Lennon JT. Resuscitation of the rare biosphere contributes to pulses of ecosystem activity. Frontiers in microbiology. 2015;6:24.
58. Shade A, Jones SE, Caporaso JG, Handelsman J, Knight R, Fierer N, et al. Conditionally rare taxa disproportionately contribute to temporal changes in microbial diversity. MBio. 2014;5(4):e01371-14.
59. Cadotte MW, Mai DV, Jantz S, Collins MD, Keele M, Drake JA. On testing the competition-colonization trade-off in a multispecies assemblage. The American Naturalist. 2006;168(5):704-9.
60. Lowe WH, McPeek MA. Is dispersal neutral? Trends in ecology & evolution. 2014;29(8):444-50.
61. Koskella B, Hall LJ, Metcalf CJE. The microbiome beyond the horizon of ecological and evolutionary theory. Nat Ecol Evol. 2017;1(11):1606-15.
62. Xue Y, Chen H, Yang JR, Liu M, Huang B, Yang J. Distinct patterns and processes of abundant and rare eukaryotic plankton communities following a reservoir cyanobacterial bloom. The ISME journal. 2018;12(9):2263-77.
63. Pedrós-Alió C. Marine microbial diversity: can it be determined? Trends in microbiology. 2006;14(6):257-63.
64. Tikhonov M. Theoretical microbial ecology without species. Physical Review E. 2017;96(3):032410.
65. Saunders AM, Albertsen M, Vollertsen J, Nielsen PH. The activated sludge ecosystem contains a core community of abundant organisms. The ISME journal. 2016;10(1):11.

66. Moe SJ, Stelzer RS, Forman MR, Harpole WS, Daufresne T, Yoshida T. Recent advances in ecological stoichiometry: insights for population and community ecology. Oikos. 2005;109(1):29-39.

Figure legends

Figure 1. Overview of the framework. (i), The raw sequence data is processed to generate the time-series of taxa relative abundances and the abundance ratio of focal taxon to the reference taxon. (ii), The neutral, consumer-resource and combined model are fitted using the least-square methods for each taxon. (iii), Key parameters can be estimated from modelling. (iv) The taxa and community determinism are assessed based on the estimated parameters of the combined model.

Figure 2. **Model fitting on microbial taxa in control bioreactors with stable substrate feeding and treatment bioreactors with incremental substrate feeding. a,** Percentages of the neutral model (N), the consumer-resource model (CR) and the combined model (C) being the best model describing taxon dynamics. For each taxon, we fitted the three models, and the best model for that taxon was determined as the one with lowest Akaike information criteria (AIC) value. Three groups of taxa were classified by mean relative abundance, with mean relative abundance < 0.01% for rare taxa, from 0.01% to 0.1% for moderate taxa, and > 0.1% for abundant taxa. **b,** The distribution of R^2 values of the three models.

Figure 3. **The relationship between ESVs' sequence dissimilarity and the difference of estimated $\boldsymbol{b_i C_i}$ representing the competition strength for resource. a**, Smoothed lines showing the mean difference in $\boldsymbol{b_i C_i}$ at different sequence dissimilarity levels between ESVs. The shaded area represents the 95% confidence interval. **b**, The time series of two taxa in the control reactors. The two taxa, ESV4 and ESV221, were both from genus T78 of the family Anaerolineaceae, and they were 98.8% similar in 16S sequences. **c**, The time series of ESV4 and ESV221 in the treatment reactors showing consistent fluctuations of their relative abundances. **d-e**, The correlation between ESV4 and ESV221 in control (**d**) and treatment (**e**) reactors.

Figure 4. **Testing the neutral model on species time series in control bioreactors. a**, The estimated λ_i from the neutral model versus the mean relative abundance of all taxa in each reactor. **b**, Prediction of the neutral model on the distribution of relative abundances of several exemplified ESVs. When the local community size was large, the relative abundance of a specific taxon followed a beta distribution under neutral scenarios, of which the shape was determined by parameters λ_i and p_i (the relative abundance of this taxon in the source community) (12). The grey histograms represent the observed value, and the blue shadow represent the model predictions using the parameters λ_i and p_i calibrated from the time series.

Figure 5. The species-level and community-level determinism. a, The predicted determinism across taxa under control and treatment bioreactors. **b-c,** Comparisons of the predicted unweighted (**b**) and weighted (**c**) community-level determinism between the control and treatment reactors.

The lines represent the mean determinism of the three replicated control or treatment bioreactors, and the error bars represent the standard deviations.