

# Deterministic and stochastic processes generating alternative states of microbiomes

Ibuki Hayashi<sup>1,\*</sup>, Hiroaki Fujita<sup>1</sup>, Hirokazu Toju<sup>1,2,3,\*</sup>

<sup>1</sup>Center for Ecological Research, Kyoto University, Otsu, Shiga 520-2133, Japan

<sup>2</sup>Center for Living Systems Information Science (CeLiSIS), Graduate School of Biostudies, Kyoto University, Kyoto 606-8501, Japan

<sup>3</sup>Laboratory of Ecosystems and Coevolution, Graduate School of Biostudies, Kyoto University, Kyoto 606-8501, Japan

\*Corresponding authors: Ibuki Hayashi, Center for Ecological Research, Kyoto University, Otsu, Shiga 520-2133, Japan. Email: hayashi.ibuki.62z@st.kyoto-u.ac.jp and Hirokazu Toju, Laboratory of Ecosystems and Coevolution, Graduate School of Biostudies, Kyoto University, Yoshida-konoe-cho, Sakyo, Kyoto 606-8501, Japan. Email: toju.hirokazu.4c@kyoto-u.ac.jp

## Abstract

The structure of microbiomes is often classified into discrete or semi-discrete types potentially differing in community-scale functional profiles. Elucidating the mechanisms that generate such “alternative states” of microbiome compositions has been one of the major challenges in ecology and microbiology. In a time-series analysis of experimental microbiomes, we here show that both deterministic and stochastic ecological processes drive divergence of alternative microbiome states. We introduced species-rich soil-derived microbiomes into eight types of culture media with 48 replicates, monitoring shifts in community compositions at six time points (8 media × 48 replicates × 6 time points = 2304 community samples). We then confirmed that microbial community structure diverged into a few state types in each of the eight medium conditions as predicted in the presence of both deterministic and stochastic community processes. In other words, microbiome structure was differentiated into a small number of reproducible compositions under the same environment. This fact indicates not only the presence of selective forces leading to specific equilibria of community-scale resource use but also the influence of demographic drift (fluctuations) on the microbiome assembly. A reference-genome-based analysis further suggested that the observed alternative states differed in ecosystem-level functions. These findings will help us examine how microbiome structure and functions can be controlled by changing the “stability landscapes” of ecological community compositions.

**Keywords:** alternative stable states, biodiversity, community assembly, dysbiosis, enterotypes, microbiome dynamics, multiple stability, transient dynamics

## Introduction

Understanding the mechanisms by which microbial community structure is organized is a major challenge in ecology and microbiology [1–4]. In a general framework of community dynamics, selection, diversification (speciation), dispersal, and drift have been considered as fundamental components of community processes [5, 6] (Fig. 1). Among the four component processes [5, 6], selection and drift can be regarded, respectively, as purely deterministic and stochastic components [7, 8], while dispersal and diversification are considered to include both deterministic and stochastic components [8]. Elucidating how those deterministic and stochastic processes collectively organize ecological community assembly is the key to predict and manage microbiome dynamics in diverse fields of applications such as human-gut microbiome therapies [9–13] and agroecosystem microbiome control [14–16].

In empirical studies of ecological communities, consequences of deterministic and stochastic ecological processes are observable as “alternative states” of community structure [3, 17–19]. Pioneering studies on human-gut microbiomes have shown that microbiome structure can be classified into semi-discrete (partially continuous) clusters despite potential numerous combinations of species or taxonomic compositions [20–23]. Although

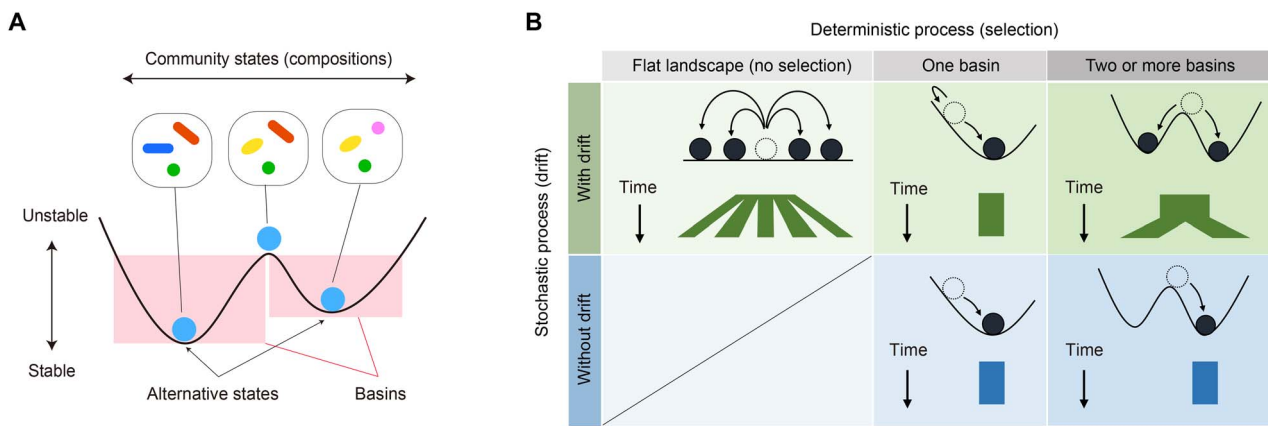
clear classification of microbiome compositions is impossible [23–26], such observations of community structure highlight the presence of strong deterministic processes organizing community structure as represented by the “stability landscape” concept of species assembly [27]. Meanwhile, the presence of alternative community compositions suggests that stochastic processes play some roles in the community differentiation. In other words, because no variation in community structure is expected under the strict assumption of deterministic processes, both deterministic and stochastic processes are necessary for the existence of categorizable community compositions [18, 28] (Fig. 1). Thus, empirical studies on community structure give essential insights into ecological community assembly. Nonetheless, it is basically difficult to develop detailed discussion on ecological community processes based on existing microbiome datasets because potential influence of environmental conditions on community structural patterns cannot be fully understood in observational studies.

In this respect, experimental studies with fully controlled environmental conditions are expected to provide ideal opportunities for examining ecological community processes in light of theories on alternative states [6, 8, 19]. By making a number of experimental microbial communities with defined

Received 11 January 2024. Revised: 14 January 2024. Accepted: 19 January 2024

© The Author(s) 2024. Published by Oxford University Press on behalf of the International Society for Microbial Ecology.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.



**Figure 1.** Conceptual framework of community assembly; (A) stability landscape of community structure; community assembly is often discussed based on the schema of “stability landscapes,” which represent the stability/instability of community states (i.e. species or taxonomic compositions); alternative states are defined as bottom positions within basins on the stability landscapes; (B) deterministic and stochastic processes in community assembly; our experiment was designed to test the contributions of deterministic and stochastic ecological processes; specifically, the results of our multireplicate microbiome experiments are expected to depend on the presence/absence of selection and drift; without drift (demographic fluctuations), all replicate communities will converge to a single community state (bottom panels); with drift, differentiation of community states can occur depending on the structure of stability landscapes; in particular, if there are two or more basins, differentiation into a small number of alternative community states will be observed as a consequence of both selection and drift (top right).

environmental conditions, we can perform strict tests of the emergence of alternative community states. In other words, the presence of multiple reproducible states in the same experimental treatment is interpreted in the framework of selection, diversification, dispersal, and drift. Despite the potential contributions of the experimental approaches to our fundamental knowledge of community assembly, few attempts (but see Estrela et al. [29]) have been made to explore alternative states of microbiomes with tens of replications.

In this study, we examined the presence of alternative community states by performing microbiome experiments under eight nutritional (medium) conditions with 48 replicates. We constructed experimental microbiomes using a forest-soil-derived community of prokaryotes as a source microbiome and then kept the 384 microbial communities (8 treatments  $\times$  48 replicates) under a fully controlled temperature condition. The experimental system was designed to examine the roles of selection and drift in microbial community assembly. With the aid of an automated pipetting system equipped in a clean laboratory environment, we monitored changes in microbiome community compositions every 2 days for 12 days based on the DNA metabarcoding of 16S rRNA gene sequences. The analysis of more than 2000 community samples then allowed us to understand how deterministic and stochastic processes could generate alternative states of ecological community structure. By exploring statistical ways for quantitatively evaluating the distributions of community compositions within state space, we extended discussions on alternative microbiome states in terms of the interplay of deterministic and stochastic ecological processes.

## Materials and Methods

### Terminology

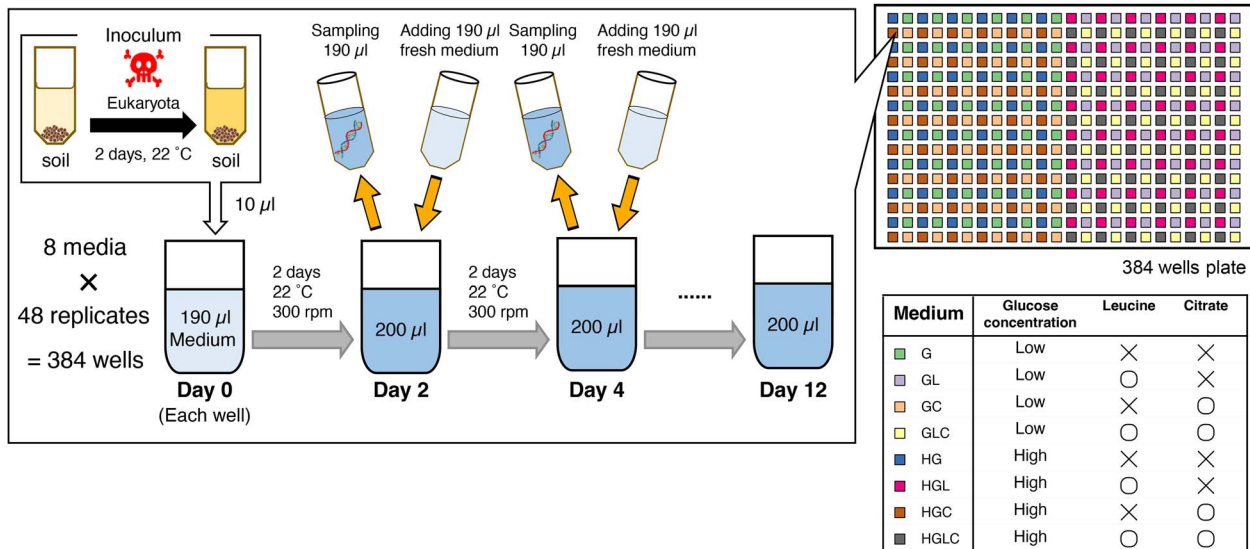
In analyzing the data obtained in this study, we need to use consistent terminology to minimize the risk of confusion and misunderstanding. It is necessary to confirm the definitions of the ecological processes whose meaning can change depending on contexts. Specifically, we use the terms “selection,” “drift,” “dispersal,” and “diversification (speciation)” in ecological

processes as conceptualized by Vellend [5]. Among the terms, selection represents expected changes in local community compositions resulting from differences in mean fitness between species, constituting purely deterministic processes [7]. Drift refers to demographic fluctuations that occur regardless of among-species difference in mean fitness, forming purely stochastic processes [7]. On the other hand, dispersal itself is not a term representing stochasticity or determinism, but it merely describes processes by which organismal individuals or their propagules move between local communities. Diversification is defined as evolutionary differentiation of genetic variants and hence it represents both stochastic (e.g. nucleotide mutation) and deterministic (i.e. natural selection) phenomena [3, 8].

Another important term we need to make clear before describing our microbiome experiment is “alternative states.” In ecology, the term “alternative stable states” is often used to discuss the processes by which divergence of community compositions is caused [17, 19, 30, 31]. Meanwhile, it is generally difficult to know whether the observed community compositions are at stable states (i.e. equilibria) or they are in transient processes toward stable states [32, 33]. Stability of communities has been a central topic in the history of community ecology [34, 35]. In this study, however, we investigate the divergence of community compositions irrespective of the concept of community stability. In other words, our community dataset can include information of both transient and stable states.

### Culture experiments of microbiomes

In the microbiome experiment, the source microbiome derived from the soil of the A layer (0–10 cm in depth) in the research forest of the Center for Ecological Research, Kyoto University, Shiga, Japan (34.972°N; 135.958°E). After sampling, the soil was sieved with a 4-mm stainless mesh and then 5 g of the sieved soil was mixed in 100 ml phosphate-buffered saline with cycloheximide (137 mM NaCl, 8.1 mM Na<sub>2</sub>HPO<sub>4</sub>, 2.68 mM KCl, 1.47 mM KH<sub>2</sub>PO<sub>4</sub>, and 200  $\mu$ g/ml cycloheximide). In this process, we added cycloheximide in order to exclude eukaryotes from the source microbiome. The source prokaryote microbiome was cultured at 22°C for 48 h.



**Figure 2.** Experimental design; laboratory culture system; a source microbiome deriving from forest soil was precultured with cycloheximide under room temperature for 2 days in order to remove eukaryotes; the microbiome inoculum was then introduced into eight types of media (Table S1) with 48 replicates; a fraction of the culture fluid was sampled every 2 days, and equivalent volume of fresh medium was added to the continual culture system throughout the 12-day experiment (8 media  $\times$  48 replicates  $\times$  6 time points = 2304 community samples).

We introduced the inoculum microbiome into eight types of media with different constitutions of carbon sources with 48 replicate communities per medium type (in total, 8 media  $\times$  48 replicates = 384 experimental communities; Fig. 2). To make the compositions of the media as simple as possible, we used M9 medium with minimal inorganic additives and combinations of three types of the carbon resources, specifically, glucose, leucine, and citrate as detailed in Table S1. We selected the medium systems with different combinations of glucose, leucine, and citrate because all the three carbohydrates have six carbons in their molecules, facilitating the adjustments of molarity in the experimental treatments. In the microbiome culture experiment, each of the eight media was designed based on the concentrations or presence/absence of the three carbon sources: i.e. low or high concentration of glucose (G/HG), with or without leucine (–/L), and with or without citrate (–/C). Hereafter, these medium types were designated as Medium-G (low glucose, without leucine, without citrate), GL (low glucose, with leucine, without citrate), GC (low glucose, without leucine, with citrate), GLC (low glucose, with leucine, with citrate), HG (high glucose, without leucine, without citrate), HGL (high glucose, with leucine, without citrate), HGC (high glucose, without leucine, with citrate), and HGLC (high glucose, with leucine, with citrate), respectively (Table S1).

In each well of a 240- $\mu$ l deep-well plate, 10  $\mu$ l of the diluted source microbiome solution and 190  $\mu$ l of medium were installed. Based on the quantitative amplicon sequencing detailed below, the source microbiome solution was estimated to contain  $3.12 \times 10^6$  DNA copies of the 16S rRNA gene ( $SD = 8.37 \times 10^5$ ) (i.e. the density of the 16S rRNA gene in the diluted source inoculum was  $3.12 \times 10^5$  copies/ $\mu$ l; Fig. S1). Note that a previous research estimated that a single cell of *Enterobacteriaceae* bacteria has an average of 7.3 DNA copies of 16S rRNA genes ( $SD = 0.9$ ; calculated with rrnDB; [36]), although variation in 16S rRNA gene copy numbers among prokaryote taxa has been known [36]. The deep-well plate was kept shaken at 200 rpm using a plate thermo-shaker BSR-MB100-4A (Bio Medical Sciences Co. Ltd, Tokyo) at 30°C for 2 days. After 2 days of incubation, 190  $\mu$ l out of the 200- $\mu$ l culture medium was sampled from each of the 48 wells after mixing (pipetting) every 2 days for 12 days. All pipetting manipulations were performed with high precision

using an automatic pipetting machine (EDR-384SR; BIOTEC Co. Ltd, Tokyo) placed in a laminar flow cabinet. In each sampling event, 190  $\mu$ l of fresh medium was added to each well so that the total culture volume was kept constant. In total, 2304 samples (384 communities/day  $\times$  6 time points) were collected.

Note that our study was not designed to examine historical contingency (priority effects) because the microbial species were simultaneously introduced into the experimental media. Dispersal between replicate communities was prohibited, and diversification leading to speciation events was unlikely to occur in the 12-day experiment. Thus, our aim in this study was to examine the contributions of selection and drift to microbiome dynamics (Fig. 1).

## DNA extraction

To extract DNA from each culture sample, 5  $\mu$ l of the collected aliquot was mixed with 1  $\mu$ l lysozyme solution (50 mg/ml lysozyme [Sigma], 20 mM Tris-HCl [pH 8.0], 2 mM EDTA), and the mixed solution was incubated at 37°C for 2 h. After adding proteinase K solution (1/30 [v/v] Proteinase K [Takara], 20 mM Tris-HCl [pH 8.0], 2 mM ethylenediaminetetraacetic acid (EDTA)), the aliquot was incubated at 55°C for 3 h and 95°C for 10 min. The solution was then vortexed for 10 min to increase DNA yield.

We also extracted DNA from the inoculum aliquot to reveal community structure of the source microbiome. Because the source inoculum was expected to include high concentrations of soil-derived compounds, which can inhibit polymerase chain reactions (PCRs), a commercial DNA extraction kit optimized for soil samples was used. Specifically, after 50  $\mu$ l of the inoculum sample was incubated with 350  $\mu$ l SDS buffer with proteinase K (1:30 [v/v] Proteinase K [Takara], 0.5% sodium dodecyl sulfate (SDS), 2 mM Tris-HCl [pH 8.0], 2 mM EDTA) based on the temperature profile of 55°C for 180 min and 95°C for 10 min. The aliquot (400  $\mu$ l) was then subjected to DNA extraction with DNeasy PowerSoil Kit (Qiagen).

## PCR and DNA sequencing

For the samples of the experimental microbiomes, prokaryote 16S rRNA V4 region was PCR-amplified with the forward primer 515f fused with 3–6-mer Ns for improved Illumina sequencing

quality [37] and the forward Illumina sequencing primer (5'-TCG TCG GCA GCG TCA GAT GTG TAT AAG AGA CAG-[3-6-mer Ns]-[515f]-3') and the reverse primer 806rB fused with 3-6-mer Ns and the reverse sequencing primer (5'-GTC TCG TGG GCT CGG AGA TGT GTA TAA GAG ACA G [3-6-mer Ns]-[806rB]-3') (0.2  $\mu$ M each). The buffer and polymerase system of KOD One (Toyobo) were used with the temperature profile of 35 cycles at 98°C for 10 s, 55°C for 5 s, and 68°C for 1 s. To prevent generation of chimeric sequences, the ramp rate through the thermal cycles was set to 1°C/s [38]. Illumina sequencing adaptors were then added to respective samples in the supplemental PCR using the forward fusion primers consisting of the P5 Illumina adaptor, 8-mer indexes for sample identification [39], and a partial sequence of the sequencing primer (5'-AAT GAT ACG GCG ACC ACC GAG ATC TAC AC-[8-mer index]-TCG TCG GCA GCG TC-3') and the reverse fusion primers consisting of the P7 adaptor, 8-mer indexes, and a partial sequence of the sequencing primer (5'-CAA GCA GAA GAC GGC ATA CGA GAT-[8-mer index]-GTC TCG TGG GCT CGG-3'). KOD One was used with a temperature profile of 8 cycles at 98°C for 10 s, 55°C for 5 s, and 68°C for 5 s (ramp rate = 1°C/s). The PCR amplicons of the samples were then pooled after a purification process with the AMPureXP Kit (Beckman Coulter). Primer dimers, which were shorter than 200 bp, were removed from the pooled library by supplemental purification with AMPureXP: the ratio of AMPureXP reagent to the pooled library was set to 1 (v/v) in this process. This library was further purified with E-gel SizeSelect 2 (Invitrogen), and then ca. 440-bp DNA fragments were selectively obtained. The sequencing libraries were processed in an Illumina Miseq sequencer (271 forward [R1] and 31 reverse [R4] cycles; 20% PhiX spike-in).

For the source microbiome sample, the prokaryote 16S rRNA V4 region was amplified as well. To estimate concentrations of 16S rRNA genes included in the inoculum, a quantitative amplicon sequencing platform was applied by introducing five "standard DNA" fragments with controlled concentrations to the PCR master mix solution of the first PCR process as detailed elsewhere [40, 41]. The standard DNAs were used for the *in silico* calibration of 16S rRNA gene concentrations in the target sample after sequencing as detailed in the previous study [40-42].

## Bioinformatics

In total, 25 284 304 sequencing reads were obtained with the Illumina sequencing. The raw sequencing data obtained in the Illumina sequencing were converted into FASTQ files using the program bcl2fastq 1.8.4 distributed by Illumina. The output FASTQ files were then demultiplexed with the program Claident v0.2. 2018.05.29. The sequencing reads were subsequently processed with the program DADA2 [43] v.1.18.0 of R 3.6.3 to remove low-quality data. In this process, the filterAndTrim function of DADA2 was used with the following criteria: minLen = 200, minQ = 10, PercLQ = 0.1, truncWindow = 5, maxEE = 3, truncR = 240, and avgqual = 15. Potentially chimeric sequences were removed by the removeBimeraDenovo function of DADA2. The molecular identification of the obtained amplicon sequence variants (ASVs) was performed based on the naive Bayesian classifier method [44] with the SILVA v.138.1 database [45]. Based on the *in silico* calibration of the quantitative amplicon sequencing data [40, 41], the number of DNA copies in 10  $\mu$ l of source microbiome (inoculum) was estimated to be  $3.12 \times 10^6$  copies as mentioned above (Fig. S1).

## Community-level diversity

The rarefaction curves representing relationship between the number of sequencing reads and the number of ASVs were drawn

using the vegan 2.6.4 package [46] of R. In the sequencing of experimental culture samples, the diversity of microbial ASVs reached plateaus along the axis of the number of sequencing reads (Fig. S2). Given the rarefaction curves, the dataset was rarefied to 5000 reads per sample with the rrarefy function of the R vegan package. Of the 2304 samples (8 treatments  $\times$  48 replicates  $\times$  6 time points), 2250 samples with >5000 reads were used in the following pipeline. After screening for the replicate communities for which sequencing data were available for all the six time points, 2094 samples were subjected to the following statistical analyses. In total, 718 prokaryote ASVs belonging to 19 phyla, 32 classes, 74 orders, 89 families, and 115 genera were detected.

For each sample, two types of  $\alpha$ -diversity indices, ASV richness and Shannon  $H'$ , were calculated. We then compared  $\alpha$ -diversity among each medium using Student's t-test corrected by false discovery rate (FDR) by Benjamini-Hochberg method (Table S2).

## Overview of the community structure

To visualize the diversity of the prokaryote community structure, an analysis of nonmetric multidimensional scaling (NMDS) was performed based on the Bray-Curtis metric of  $\beta$ -diversity. Likewise, to examine the dependence of community structure on medium conditions, a series of permutational multivariate analysis of variance (PERMANOVA) [47] was performed with 50 000 permutations. In each PERMANOVA model, glucose concentration (high or low;  $df = 1$ ), the presence/absence of leucine ( $df = 1$ ), or the presence/absence of citrate ( $df = 1$ ) was included as the explanatory variable. An additional model including all the medium conditions and interactions between them ( $df = 7$ ) was examined as well. The coefficients of determination ( $R^2$ ), which indicated the proportion to which given explanatory variables explained variation in dependent variables (community structure), were shown for each of the above PERMANOVA models ("glucose concentration," "leucine," "citrate," or "all" model). Each PERMANOVA model was supplemented by permutational analysis of dispersion (PERMDISP; [48]) for potential effects of the medium condition on dispersion in community structure. The NMDS, PERMANOVA, and PERMDISP were performed for each dataset of the ASV-, genus-, and family-level compositions of prokaryote communities using the vegan 2.6.4 package [46] of R.

## Observed community structural differentiation versus null-model expectation

To examine the differentiation of community compositions among replicate samples, the Bray-Curtis metric of  $\beta$ -diversity was calculated for respective pairs of samples collected on the same day in each of the same treatments (medium conditions). Note that the Bray-Curtis metric of  $\beta$ -diversity could range from 0 (identical community structure) to 1 (no overlap of microbes). A histogram of the community structural difference was shown for each day in each experimental treatment. If a small number of alternative states of community structure exist under a medium condition, there can be some peaks of community structural difference within a histogram.

To evaluate whether the obtained histograms of community dissimilarity ( $\beta$ -diversity) represented unimodal distributions, we performed an analysis comparing observed histograms to those expected under purely stochastic models. We obtained average community structure for each time point in each medium condition for each of the ASV-, genus-, and family-level datasets. Stochasticity in initial community compositions was simulated by randomly sampling microbes from the average community structure based on multinomial distributions: the size of sampling was set to 5000 (i.e. the number of sequencing reads per sample in



our empirical data) in each of the 48 simulated replicate communities. The subsequent stochastic fluctuations of the community structure were then introduced based on random-walk simulations. Each run of the random-walk simulations was performed through the following three steps. First, the Bray–Curtis metric of  $\beta$ -diversity was calculated for respective pairs of the simulated communities, and the median of  $\beta$ -diversity values was obtained (Step 1). Second, if this median was lower than the median of the empirical data, then one of the 5000 entries (ASVs, genera, or families) in each of the 48 simulated replicate communities was randomly selected and this entry was duplicated (Step 2). Third, one of the 5001 entries was randomly selected and deleted (Step 3). If the median of  $\beta$ -diversity values exceeded that of the empirical data, this random-walk process was completed. Otherwise, the random-walk simulation was repeated from the Step 1. Throughout the null-model analyses, the simulated distributions of community dissimilarity with the same degree of variance as the empirical data were obtained.

We then compared the histograms of the simulated among-replicate dissimilarity with observed histograms. If all the microbiome compositions within the state space are equally possible (i.e. lack of deterministic processes) and only stochastic events (initial stochasticity and subsequent random-walk dynamics) operate, unimodal distributions would be observed in the empirical histograms as in the simulated histograms. Alternatively, if the presence of both deterministic and stochastic processes results in the divergence of community states, multimodal distributions would be present in the empirical histograms of community dissimilarity among replicate samples. We quantified the deviation of the observed community dissimilarity distributions from the simulated ones based on the Kullback–Leibler metric of divergence as follows:

$$D_{KL}(P || Q) = \sum_x P(x) \log \left( \frac{P(x)}{Q(x)} \right)$$

where  $x$  denotes the Bray–Curtis  $\beta$ -diversity between replicate samples,  $P(x)$  was the observed distributions of Bray–Curtis  $\beta$ -diversity (i.e. empirical data), and  $Q(x)$  was the simulated distributions of Bray–Curtis  $\beta$ -diversity. The Kullback–Leibler divergences were calculated by using the `philentropy 0.7.0` package [49] of R. The larger the Kullback–Leibler divergence is, the more it expresses that the two distributions are different. The analysis was performed for each day in each experimental treatment (medium condition) for each of the ASV-, genus-, and family-level community compositional datasets.

### Likelihood ratio tests of dissimilarity distributions

We statistically examined whether the observed histograms of community dissimilarity could be explained by unimodal or multimodal distributions. In this analysis, we applied Gaussian mixture modeling [50]. For a given number of normal distributions to be combined (number of components;  $G$ ), the variance, mean, and relative weights to the other components of normal distributions are estimated. Let each overlapping normal distribution be  $f(x, \theta)$  ( $\theta$  includes mean and variance), and let the distribution of  $n$  samples (here, a sample is each dissimilarity of the pairwise communities) be  $X = \{x_1, x_2, \dots, x_i, \dots, x_n\}$ , we can estimate the parameter that maximizes the following equation:

$$L(\Psi; x_1, \dots, x_n) = \sum_{i=1}^n \log \left( \sum_{k=1}^G \pi_k f_k(x_i; \theta_k) \right)$$

where  $\Psi = \{\pi_1, \dots, \pi_{G-1}, \theta_1, \dots, \theta_G\}$  are the parameters of the mixture model and  $(\pi_1, \dots, \pi_{G-1})$  are the mixing weights. Note that  $\pi_k$  is defined for each pair of mixture components: when  $G = 1$ , the total number of parameters is 2 ( $\theta_1$  including mean and variance); when  $G = 2$ , the total number of parameters is 5 (each of  $\theta_1$  and  $\theta_2$  includes mean and variance and mixing weight between components 1 and 2 is denoted by  $\pi_1$ ;  $\pi_2$  is defined as  $\pi_2 = 1 - \pi_1$ ).

In our analysis of the empirical data, statistical models assuming unimodal distributions of community dissimilarity and those assuming two or more modes were compared based on the Bayesian information criterion (BIC). The parameter fitting was performed by the `mclust 6.0.0` package [50] of R, and the parameters were obtained via the expectation-maximization (EM) algorithm in this package. We then performed likelihood ratio tests by assuming that deviance (i.e.  $2 \times \log$ -likelihood-ratio) followed chi-square distributions. The tests were performed for each of the ASV-, genus-, and family-level datasets.

### Silhouette method for inferring the number of clusters

We investigated whether the communities belonging to each time point and medium can be divided into a number of clusters using  $k$ -medoids clustering based on the Bray–Curtis distance between communities and cluster number selection using the average silhouette coefficient [51]. In  $k$ -medoids clustering, a standard unsupervised learning approach, replicate microbiomes in each experimental treatment were divided into  $k$  clusters according to a given number of clusters  $k$ . For each  $k$ , the average silhouette coefficient were calculated, and the  $k$  that exhibits the highest average silhouette coefficient was selected as the optimal number of clusters. The average silhouette coefficient was calculated by using the `cluster 2.1.4` package [52].

### Potential differentiation in community-level functions

To examine the potential differentiation of ecosystem functions among the replicate communities, we inferred the community-level functional profiles based on the phylogenetic estimation of gene repertoires with `PICRUSt2` [53]. The gene repertoire data were then subjected to NMDS, PERMANOVA, and `PERMDISP` based on the Bray–Curtis  $\beta$ -diversity. In total, the relative functional compositions of 392 metabolic pathways/processes were examined as input data. Based on the inferred profiles of metabolic pathways/processes, histograms of the Bray–Curtis  $\beta$ -diversity among replicate samples were drawn at respective time points within respective treatments.

The metabolic pathway/process information was also used to compare potential functional properties among the clusters of microbial compositions. For each of the ASV-, genus-, and family-level data of community compositions, mean pathway abundances were calculated for each of the clusters inferred with the average silhouette coefficient. The metabolic pathways whose pathway abundances varied considerably among those community compositional clusters were explored based on the Shannon entropy metric: 30 metabolic pathways/processes with the lowest Shannon entropy scores were selected from the 392 metabolic pathways/processes.

## Results

### Overview of the community structure

In each of the experimental treatments (medium conditions), taxonomic richness and Shannon's diversity index at the ASV, genus,

and family levels decreased through time (Figs 3A, S3, and S4). The  $\alpha$ -diversity indices varied depending also on medium conditions (Table S2). The addition of leucine or citrate, for example, significantly increased  $\alpha$ -diversity of the experimental microbiomes (FDR:  $q < 0.01$ ) (Fig. 3A; see Table S2). In all the experimental treatments, the decline of the number of ASVs reached plateaus until Day 8 (Fig. S4).

At the genus level, the microbiomes were dominated by the four genera, *Klebsiella*, *Raoultella*, *Pseudomonas*, and *Cedecea*, although there was substantial variation in the balance of these taxa among the medium conditions examined (Fig. 4). For example, in the media containing citrate (Medium-GC, GLC, HGC, and HGLC), the relative abundance of *Citrobacter* was higher than in other medium conditions. In the media containing leucine (Medium-GL, GLC, HGL, and HGLC), *Cedecea* were more abundant than in other medium conditions. Moreover, within each of the medium conditions, *Serratia* became dominant only in some replicate communities (Fig. 4).

At the family level, the communities were characterized by *Enterobacteriaceae*, *Pseudomonadaceae*, and *Comamodaceae*, with *Enterobacteriaceae* being particularly dominant (Fig. S6). *Comamodaceae* tended to appear at high proportions in the media with high concentration of glucose. The substantial among-replicate variation in community compositions observed at the ASV- and genus-level analyses was also evident in the family-level analysis. Specifically, in the media containing leucine (medium-GL, GLC, HGL, and HGLC), *Yersiniaceae*, which includes *Serratia*, dominated only in some replicate communities within each experimental treatment.

At all-time points, differences in community compositions were significantly explained by differences in medium (Figs 3B, S7–S8 and Table S3). The results also indicated that the overall variance explained by medium conditions decreased from Day 4 to 12 at the ASV level (Fig. 3B). Among the components of the medium conditions, concentrations of glucose had the largest contributions to community structure at the family level (Fig. S7). Meanwhile, the presence/absence of leucine had the highest impacts on the ASV- and genus-level community structure until Day 6 (Figs 3B and S7). The community structure within the NMDS plot distributed depending on both time points and medium conditions (Fig. 3C). The PERMDISP indicated that the concentration of glucose and the presence/absence of citrate greatly influenced dispersion of family-level community structure among samples (Fig. S8 and Table S4). In contrast, the presence/absence of leucine was the major determinant of community structural dispersion at the genus- and ASV-level (Fig. S8).

## Community structural differentiation

In some medium conditions (experimental treatments), large community structural differences among replicate samples were observed. For example, in the Medium-GLC treatment, substantial differences in community compositions were observed among replicate samples at the ASV, genus, and family levels, and these differences seemingly increased until Day 10 (Figs 5, S9, and S10). The community compositions seemed to be classified into some categories within the NMDS plot (Figs 5, S9, and S10) due to the dominance of *Serratia* in some, but not all, replicate samples (Fig. 4).

In fact, the histogram of community structural difference between samples (i.e. Bray–Curtis  $\beta$ -diversity between pairs of replicate samples) showed seemingly bimodal or multimodal distributions (Figs 6, S11, and S12). Such bimodal or multimodal

distributions of  $\beta$ -diversity were observed in all the eight medium conditions (Fig. 6), although among-replicate differentiation in community structure on the NMDS surface was conspicuous in some treatments (Medium-G, GL, GLC, HGL, HGC, and HGLC) but not in others (Medium-GC and HG) (Fig. 5).

Within the deep-well plate used in the experiment, substantially different community compositions were observed between adjacent replicate wells in each experimental treatment (Figs S13 and S14). This fact suggests that potential fine-scale heterogeneity of temperature or humidity within the small culture plate does not fully explain the observed among-replicate divergence of community structure.

## Observed community structural differentiation versus null-model expectation

The deviations of the observed distributions of community compositional dissimilarity from simulated distributions were then quantified based on the Kullback–Leibler divergence metric. The empirical data with clear multimodal distributions were highlighted by high Kullback–Leibler divergence (e.g.  $1 \leq$  Kullback–Leibler divergence; Figs 6, S11 and S12, and S15).

## Likelihood ratio tests of dissimilarity distributions

The BIC-based model selection in the Gaussian mixture modeling showed that the observed distributions of community compositional dissimilarity ( $\beta$ -diversity) were better explained by models assuming two or more modes than those assuming unimodal distributions. The subsequent likelihood ratio tests further indicated the presence of multiple peaks within the distributions of  $\beta$ -diversity among replicate samples (FDR  $< 10^{-9}$  for all of the 48 media  $\times$  day combinations; Table S5).

## Silhouette method for inferring the number of clusters

On average, 2.56 (SD=0.85), 2.56 (SD=1.11), and 2.81 (SD=1.20) clusters were identified for the community compositional data at the ASV, genus, and family levels, respectively (Figs S16–S18; Table S5). Replicate samples belonging to respective clusters are shown on the 2D surfaces of NMDS (Figs S16 and S17 and S18).

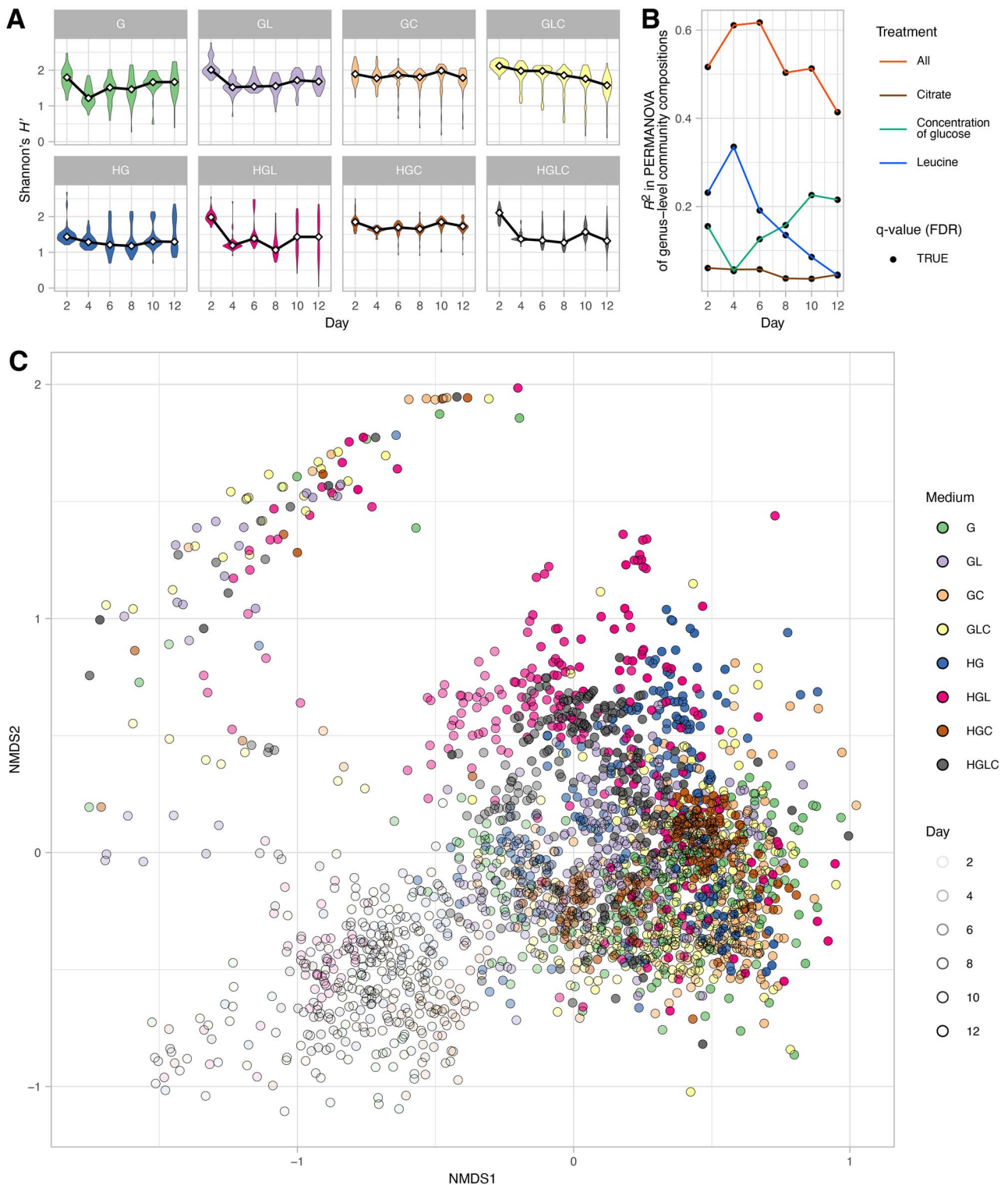
## Potential differentiation in community-level functions

As observed in the above analyses on the community compositions, functional profiles of the community samples significantly differed depending on the medium conditions (Fig. 7A and B). Moreover, among-replicate differentiation was evident in the functional profile dataset (Fig. 7C and D), although the extent of such differentiation varied among medium conditions (Figs S19 and S20).

An additional analysis further suggested that the clusters defined based on taxonomic compositions differed substantially in the inferred functional profiles. The metabolic pathways/processes whose abundances showed the greatest variation among community compositional clusters involved teichoic acid (poly-glycerol) biosynthesis, vanillin/vanillate degradation, and bacteriochlorophyll-*a* biosynthesis (Fig. S21).

## Discussion

Based on the experimental design with many replicates in each of the multiple treatments, we examined how microbial community assembly was driven by deterministic and stochastic



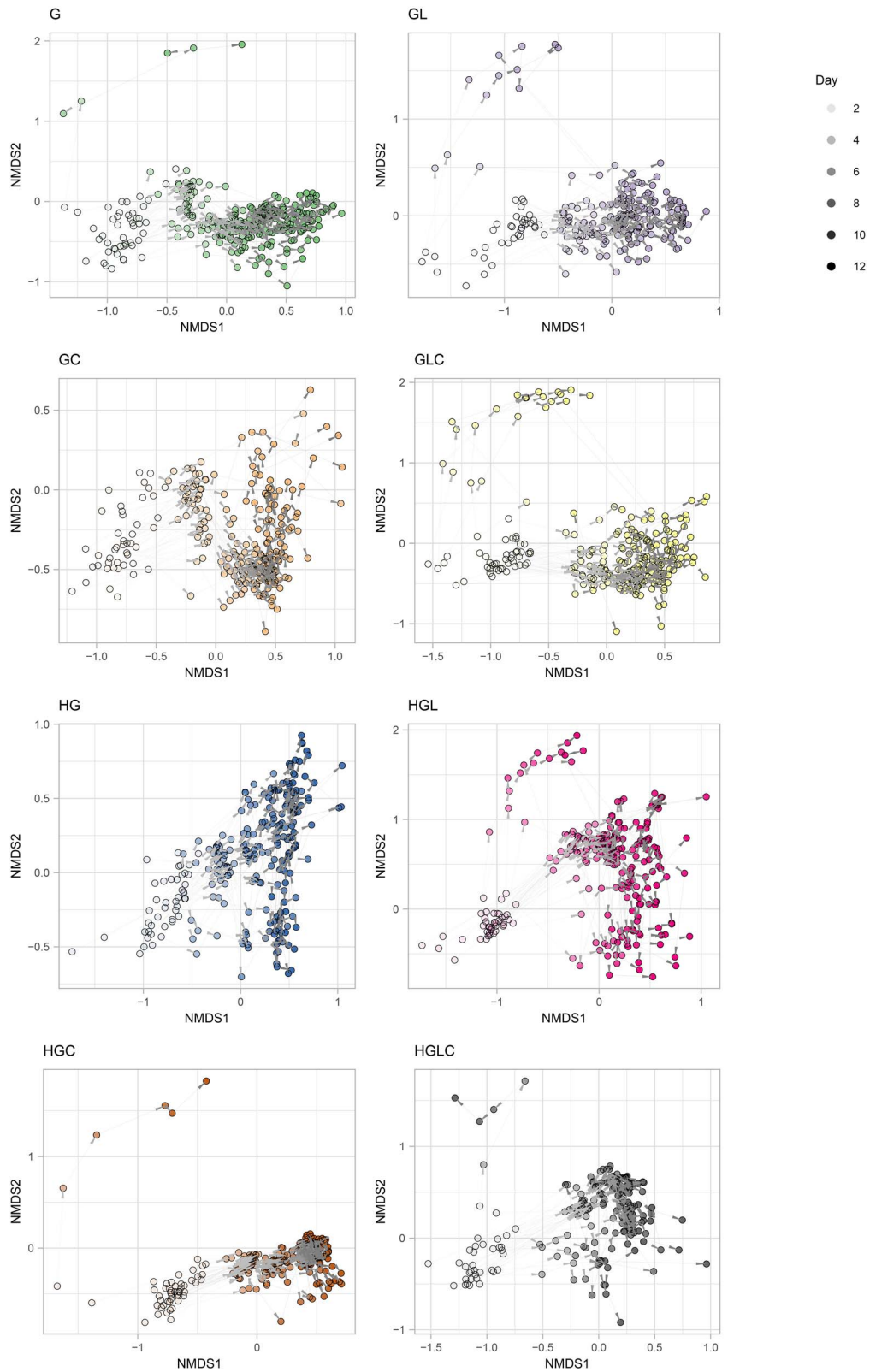
**Figure 3.** Overview of the community structure; (A)  $\alpha$ -diversity of the communities; in each experimental treatment (medium condition), Shannon's diversity index for ASV-level community compositions is shown for each day; the results of Student's t-test are shown in Table S2; (B) dependence of community structure on medium conditions; in each PERMANOVA model of ASV-level community compositions, glucose concentration (high or low;  $df = 1$ ), the presence/absence of leucine ( $df = 1$ ), or the presence/absence of citrate ( $df = 1$ ) was included as the explanatory variable; an additional model including all the medium conditions and interactions between them ( $df = 7$ ) was examined as well; the coefficients of determination ( $R^2$ ), which indicate the proportion to which given explanatory variables explain variation in dependent variables (community structure), are shown for each of the above PERMANOVA models ("glucose concentration," "leucine," "citrate," or "all" model); (C) overview of the community structure; the ASV-level community compositions are shown along the axes of NMDS (stress = 0.157).



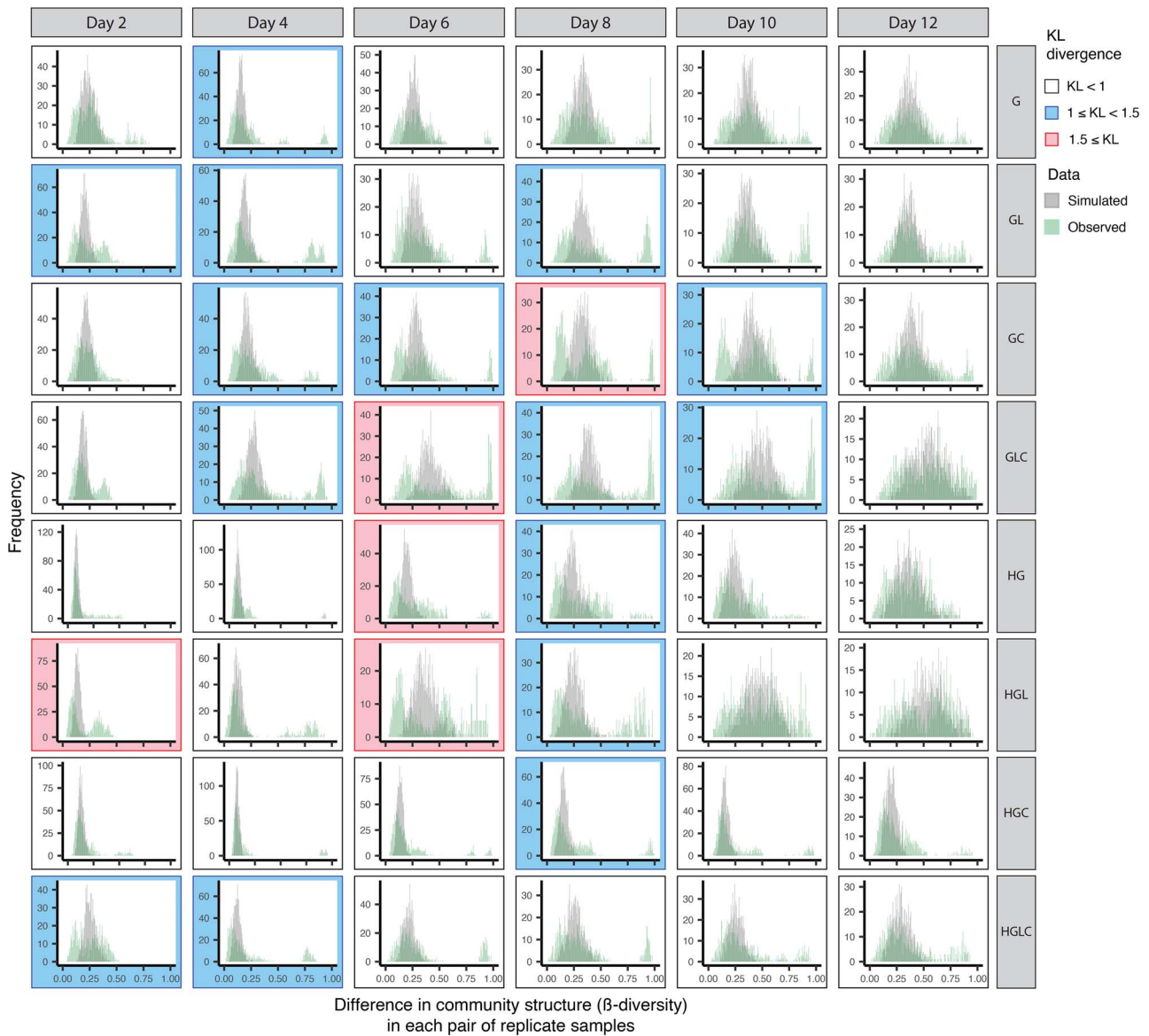


**Figure 4.** Variation in community structure among replicate samples; for each replicate community in each experimental treatment, changes in genus-level community compositions (relative abundance) are shown; the numbers shown at the top of the bar plots refer to time points (days); the replicate samples were ordered based on unweighted pair group method with arithmetic mean (UPGMA) analyses performed on Day 12 for respective experimental treatments; the order of replicate communities on successive days is the same as that on Day 12.





**Figure 5.** Time-series changes in community structure; for each replicate community in each experimental treatment, time-series changes in ASV-level community structure are shown with arrows on the NMDS surface defined in Fig. 3C.

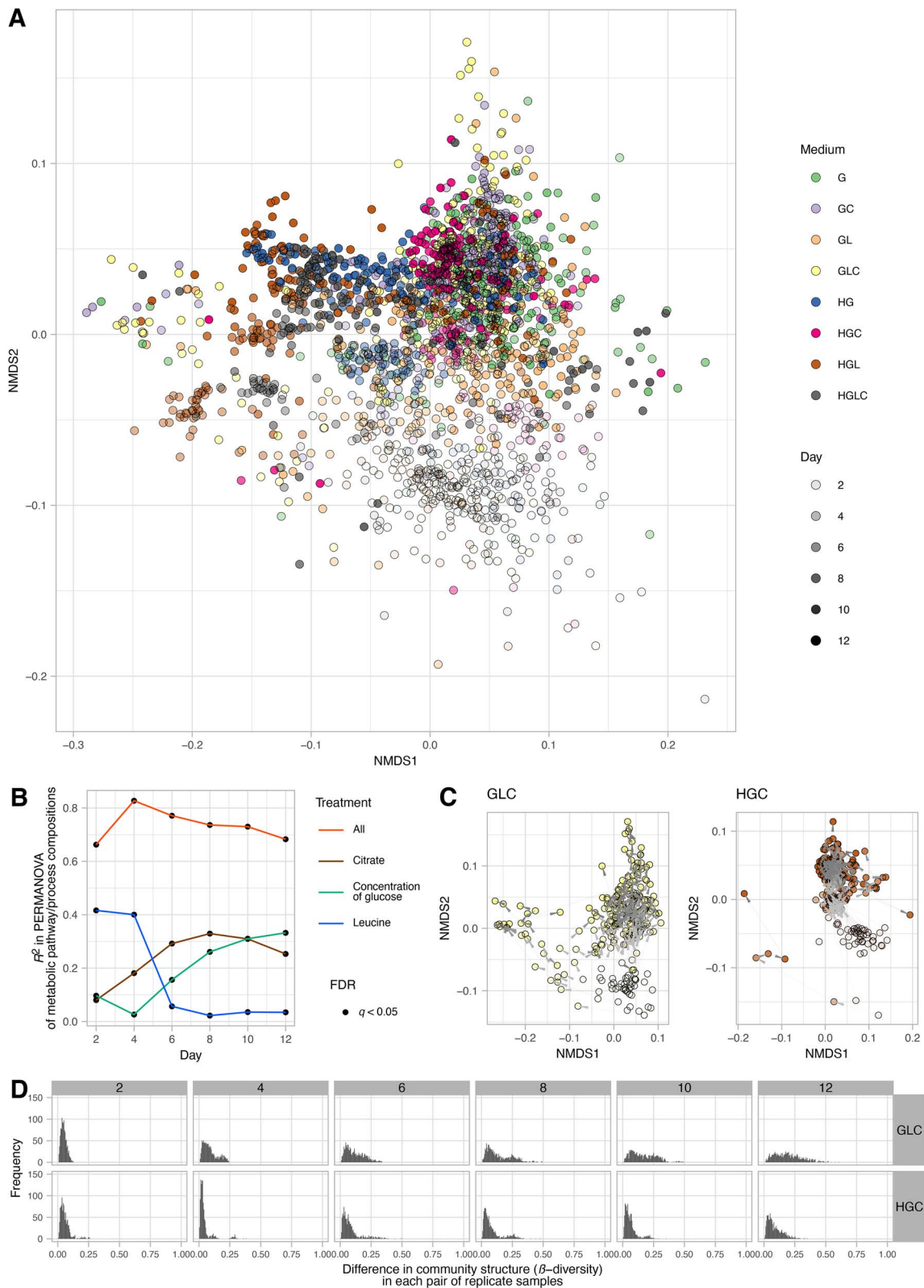


**Figure 6.** Histograms of community structural differentiation; for each experimental treatment, difference in ASV-level community structure (Bray–Curtis  $\beta$ -diversity) between replicate communities is shown as a histogram for each day; the distributions simulated by assuming purely stochastic processes (stochasticity at colonization events and subsequent random-walk processes) and the distributions observed in the empirical data are shown; each panel is color coded according to the score of Kullback–Leibler divergence (KL divergence) between the simulated and observed distributions (Fig. S15); a higher KL divergence score represents greater difference between the simulated and observed distributions; see Figs S11 and S12 for results based on the genus- and family-level community structure.

processes. As reported in previous studies [29, 54, 55], the alpha diversity and community structure of the experimental microbiomes have changed drastically within the first few days (Figs 4 and S5 and S6), but the pace of the change has decreased dramatically late in the time series (Figs 5 and S9 and S10). We then observed community compositional variation both among and within experimental treatments (Figs 3–5). Although our experiment is based on a single source microbiome and hence the generality of the findings should be strictly examined in future studies expanding the experimental approach, the results provide an opportunity for discussing how deterministic and stochastic processes operate in microbiome assembly. Hereafter, we discuss the ecological processes potentially underlying the microbiome variation in the framework outlined in Fig. 1.

In terms of deterministic ecological processes, the importance of selection has been discussed for decades in microbiology [7, 8, 56–60]. Indeed, in our study, the community structure differed remarkably depending on the medium conditions (Fig. 3), indicating that selection [8, 59] operated in the community processes of the experimental microbiomes. For example, *Citrobacter*, which has the ability to metabolize citrate, occurred almost exclusively in the media containing citrate (Fig. 4). Moreover, the observed distributions of the community compositional dissimilarity among replicate samples deviated quantitatively and qualitatively from what was expected under purely stochastic processes (Figs 6 and S11 and S12). Thus, selection can be regarded as a fundamental mechanism determining community structure.

In addition to selection, drift was shown to organize the community structure of the experimental microbiomes. We found



**Figure 7.** Functional profiles of the microbiome; (A) overview of functional compositions; based on a phylogenetic inference of metabolic pathway compositions (a reference-genome analysis of constituent bacteria based on PICRUSTS2), community samples are plotted on a NMDS surface (stress = 0.125); (B) dependence of community functional profiles on medium conditions; in each PERMANOVA model of metabolic pathway/process compositions, glucose concentration (high or low;  $df = 1$ ), the presence/absence of leucine ( $df = 1$ ), or the presence/absence of citrate ( $df = 1$ ) was included as the explanatory variable; an additional model including all the medium conditions and interactions between them ( $df = 7$ ) was examined as well; the coefficient of determination ( $R^2$ ) is shown for each day; (C) time-series changes in community functional profiles; the results on two experimental treatments (Medium-GLC and HGC) are shown; see Fig. S19 for full results; (D) histograms of community functional differentiation; the results on two experimental treatments (Medium-GLC and HGC) are shown; see Fig. S20 for full results.



that community compositions could diverge into a small number of reproducible alternative states even under the same environmental (medium) conditions (Figs 3–6). This finding indicates that selection and drift could collectively generate alternative states of community structure in microbiome dynamics (Fig. 1). On a stability landscape of community structure, divergence into a small number of reproducible states never occur without selection (Fig. 1). In addition, stochastic processes like drift (demographic fluctuations) are necessary for such divergence because without stochasticity, identical consequences are always expected. Meanwhile, careful interpretation is required when we discuss the stochastic processes caused by drift. In our experiment, drift might influence the community assembly not exclusively at the (pre-)colonization stage (i.e. stochastic sampling effects in the pipetting of inoculum microbiomes) but also at the postcolonization stage (i.e. “random walk” fluctuation of community compositions). The relative contributions of precolonization and postcolonization stochastic events need to be examined in future studies. In our present experiment,  $3.12 \times 10^6$  DNA copies of the 16S rRNA gene were introduced into each replicate community at the precolonization stage, while 10  $\mu$ l out of 200  $\mu$ l of culture media was left for successive time points (i.e. 5% dilution rate in each transition between time points) at the postcolonization stage. Experiments comparing multiple source microbiome density and multiple post-colonization bottleneck levels will provide further crucial knowledge of how drift can drive community dynamics leading to alternative states.

In our results, community structural differences, which were evident at the ASV or genus levels (Figs 3–6), were less conspicuous at the family level (Figs S6, S9, and S11) partly due to the dominance of bacteria belonging to *Enterobacteriaceae*. Such convergence of community structure at higher taxonomic levels may stem from redundancy in functional compositions of microbial communities as has been reported in previous studies [9, 61, 62]. However, in some experimental treatments (e.g. Medium-GLC), substantial divergence of taxonomic compositions among replicate communities was evident even at the family level (Fig S6). Thus, substantial functional differentiation could occur through the emergence of alternative states of microbiome structure.

In fact, a reference-genome-based analysis suggested that divergence into alternative community structure could entail functional differentiation of microbial communities. The inferred community-level gene repertoires were differentiated into some clusters within each experimental treatment (Fig 7C), resulting in bimodal or multimodal distributions of pairwise community dissimilarity (Fig 7D). We further found that the clusters identified in the community compositional analyses differed substantially in their metabolic pathway/process profiles (Fig. S21). This observation is of particular interest because knowledge of the processes driving divergence into functionally different microbiome states is essential in diverse fields of applied sciences. In medicine, for example, the structure of human gut microbiomes varies considerably among host individuals, exhibiting complex associations with host human health [20, 22]. Furthermore, recent studies have suggested that fish-associated and plant-associated microbiomes can be classified into several community compositional clusters potentially differing in physiological impacts on host organisms [14, 63]. To extend the discussion on functional divergence of microbiomes, the genomic information of the microbes constituting microbiomes need to be enriched with shotgun metagenomic analyses [64–66].

From a detailed inspection of the experimental results, we found that shifts between alternative states could occur in microbial community dynamics. Although differences in community structure were expanded from Day 2 to Day 10, transitions between alternative states seemed to occur within the NMDS plots in some experimental treatments (Medium-GL, GLC, and HGL; Fig. 5). This observation illuminates the ecological theory that transitions between alternative stable states are possible if demographic fluctuations of microbial populations within the communities are large enough to cross the “boundaries” splitting basins of stability landscapes [30, 31]. Alternatively, the observed dynamics may be interpreted as transient dynamics toward a large basin within a stability landscape [19, 33]. Although it is notoriously difficult to distinguish alternative transient states from alternative stable states based on current frameworks of empirical datasets, further feedback between theoretical and empirical investigations will promote our understanding of large shifts in community structure [42]. Besides, quantitative evaluation of the degree to which community compositions were distributed within state space (Figs S16–S18) will help us establish a comprehensive conceptual framework beyond dichotomy between discrete and continuous community states.

The fact that patterns in the divergence into alternative states differed among environmental conditions give significant implications for the “controllability” of microbiomes. We found that the number and structure of alternative states differed depending on medium conditions (Figs 5 and S9 and S10). This result suggests that the shapes of underlying stability landscapes, which are formed by selection, can be changed by the addition of specific chemicals to the microbial ecosystems. Such changes in stability landscape structure have been intensively discussed in theoretical ecology [31] but explored in a few empirical studies [67, 68]. Thus, this study indicates that microbiome experiments under a series of environmental conditions provide ideal opportunities for investigating how microbiome structure can be managed by changing the structure of background stability landscapes based on the manipulation of environmental conditions.

Although the experiment using field-collected source microbiomes allowed us to explore broad state space of possible community structure, experiments based on explicitly defined sets of microbial species will provide complementary insights. A more mechanism-based understanding may become possible by conducting similar experiments on multispecies systems with known genomes and metabolic pathways. In this respect, experiments on synthetic communities (SynCom) [69–71] are expected to promote reductionistic understanding of microbial community processes [72, 73]. With the aid of genome-based metabolic modeling [74–76], for example, potential consequences of competitive and facilitative interactions between microbial species may be inferred at the community level [66]. For further understanding of microbiome assembly, it is also important to apply empirical frameworks for describing complex time-series processes of ecological communities [42, 77, 78] in light of shifts between alternative states [42]. Exploring potential contributions of phages to stochastic dynamics of bacterial communities is another intriguing direction of research. Interdisciplinary studies based on microbiology, genomics, and theoretical ecology will reorganize our knowledge of the stability and dynamics of microbiomes.

## Author contributions

Ibuki Hayashi and Hirokazu Toju designed the work. Ibuki Hayashi performed the experiments. Ibuki Hayashi and Hiroaki Fujita

analyzed the data. Ibuki Hayashi and Hirokazu Toju wrote the paper with HF.

## Supplementary material

Supplementary material is available at ISME Communications online.

## Conflicts of interest

H.T. is the founder and director of Sunlit Seedlings Ltd. This had no impact on the design and implementation of the study. The other authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Funding

This work was financially supported by NEDO Moonshot Research and Development Program (JPNP18016), JST FOREST (JPMJFR2048), JST CREST (JPMJCR23N5), and JSPS Grant-in-Aid for Scientific Research (20K20586) to H.T.

## Data availability

The 16S rRNA gene sequence data are available from the DNA Data Bank of Japan (DDBJ; accession number, Bioproject PRJDB15353). The microbial community data and all the R scripts used for the statistical analyses are available at the GitHub repository (<https://github.com/Ibuki-Hayashi/deterministic-and-stochastic-processes-generating-alternative-states>).

## References

- Diamond JM. Assembly of species communities. In: Cody M.L., Diamond J.M. (eds.), *Ecology and Evolution of Communities*. Harvard University Press, Boston, 1975, 342–444.
- Chesson P. Mechanisms of maintenance of species diversity. *Annu Rev Ecol Syst* 2000;**31**:343–66. <https://doi.org/10.1146/annurev.ecolsys.31.1.343>.
- Nemergut DR, Schmidt SK, Fukami T et al. Patterns and processes of microbial community assembly. *Microbiol Mol Biol Rev* 2013;**77**:342–56. <https://doi.org/10.1128/MMBR.00051-12>.
- Coyte KZ, Schluter J, Foster KR. The ecology of the microbiome: networks, competition, and stability. *Science* 2015;**350**:663–6. <https://doi.org/10.1126/science.aad2602>.
- Vellend M. Conceptual synthesis in community ecology. *Q Rev Biol* 2010;**85**:183–206. <https://doi.org/10.1086/652373>.
- Vellend M. *The Theory of Ecological Communities*. Princeton University Press, New Jersey, 2017.
- Vellend M, Srivastava DS, Anderson KM et al. Assessing the relative importance of neutral stochasticity in ecological communities. *Oikos* 2014;**123**:1420–30. <https://doi.org/10.1111/oik.01493>.
- Zhou J, Ning D. Stochastic community assembly: does it matter in microbial ecology? *Microbiol Mol Biol Rev* 2017;**81**:10–1128. <https://doi.org/10.1128/MMBR.00002-17>.
- David LA, Maurice CF, Carmody RN et al. Diet rapidly and reproducibly alters the human gut microbiome. *Nature* 2013;**505**:559–63. <https://doi.org/10.1038/nature12820>.
- Shreiner AB, Kao JY, Young VB. The gut microbiome in health and in disease. *Curr Opin Gastroenterol* 2015;**31**:69–75. <https://doi.org/10.1097/MOG.000000000000139>.
- Palleja A, Mikkelsen KH, Forslund SK et al. Recovery of gut microbiota of healthy adults following antibiotic exposure. *Nat Microbiol* 2018;**3**:1255–65. <https://doi.org/10.1038/s41564-018-0257-9>.
- Chng KR, Ghosh TS, Tan YH et al. Metagenome-wide association analysis identifies microbial determinants of post-antibiotic ecological recovery in the gut. *Nat Ecol Evol* 2020;**4**:1256–67. <https://doi.org/10.1038/s41559-020-1236-0>.
- Fan Y, Pedersen O. Gut microbiota in human metabolic health and disease. *Nat Rev Microbiol* 2020;**19**:55–71. <https://doi.org/10.1038/s41579-020-0433-9>.
- Toju H, Peay KG, Yamamichi M et al. Core microbiomes for sustainable agroecosystems. *Nat Plants* 2018;**4**:247–57. <https://doi.org/10.1038/s41477-018-0139-4>.
- Compant S, Samad A, Faist H et al. A review on the plant microbiome: ecology, functions, and emerging trends in microbial application. *J Adv Res* 2019;**19**:29–37. <https://doi.org/10.1016/j.jare.2019.03.004>.
- Ke J, Wang B, Yoshikuni Y. Microbiome engineering: synthetic biology of plant-associated microbiomes in sustainable agriculture. *Trends Biotechnol* 2021;**39**:244–61. <https://doi.org/10.1016/j.tibtech.2020.07.008>.
- Law R, Morton RD. Alternative permanent states of ecological communities. *Ecology* 1993;**74**:1347–61. <https://doi.org/10.2307/1940065>.
- Chase JM, Myers JA. Disentangling the importance of ecological niches from stochastic processes across scales. *Philos Trans R Soc B Biol Sci* 2011;**366**:2351–63. <https://doi.org/10.1098/rstb.2011.0063>.
- Fukami T. Historical contingency in community assembly: integrating niches, species pools, and priority effects. *Annu Rev Ecol Evol Syst* 2015;**46**:1–23. <https://doi.org/10.1146/annurev-ecolsys-110411-160340>.
- Arumugam M, Raes J, Pelletier E et al. Enterotypes of the human gut microbiome. *Nature* 2011;**473**:174–80. <https://doi.org/10.1038/nature09944>.
- Zaneveld JR, McMinds R, Thurber RV. Stress and stability: applying the Anna Karenina principle to animal microbiomes. *Nat Microbiol* 2017;**2**:1–8. <https://doi.org/10.1038/nmicrobiol.2017.121>.
- Costea PI, Hildebrand F, Manimozhiyan A et al. Enterotypes in the landscape of gut microbial community composition. *Nat Microbiol* 2017;**3**:8–16. <https://doi.org/10.1038/s41564-017-0072-8>.
- Cheng M, Ning K. Stereotypes about enterotype: the old and new ideas. *Genomics Proteomics Bioinformatics* 2019;**17**:4–12. <https://doi.org/10.1016/j.gpb.2018.02.004>.
- Jeffery IB, Claesson MJ, O'Toole PW et al. Categorization of the gut microbiota: enterotypes or gradients? *Nat Rev Microbiol* 2012;**10**:591–2. <https://doi.org/10.1038/nrmicro2859>.
- Knights D, Ward TL, McKinlay CE et al. Rethinking enterotypes. *Cell Host Microbe* 2014;**16**:433–7. <https://doi.org/10.1016/j.chom.2014.09.013>.
- Liang C, Tseng HC, Chen HM et al. Diversity and enterotype in gut bacterial community of adults in Taiwan. *BMC Genomics* 2017;**18**:1–11. <https://doi.org/10.1186/s12864-016-3261-6>.
- Folke C, Carpenter S, Walker B et al. Regime shifts, resilience, and biodiversity in ecosystem management. *Annu Rev Ecol Evol Syst* 2004;**35**:557–81. <https://doi.org/10.1146/annurev.ecolsys.35.021103.105711>.
- Adler PB, HilleRisLambers J, Levine JM. A niche for neutrality. *Ecol Lett* 2007;**10**:95–104. <https://doi.org/10.1111/j.1461-0248.2006.00996.x>.

29. Estrela S, Vila JCC, Lu N et al. Functional attractors in microbial community assembly. *Cell Syst* 2022;**13**:29–42.e7. <https://doi.org/10.1016/j.cels.2021.09.011>.
30. Beisner BE, Haydon DT, Cuddington K. Alternative stable states in ecology. *Front Ecol Environ* 2003;**1**:376–82.
31. Scheffer M, Carpenter S, Foley JA et al. Catastrophic shifts in ecosystems. *Nature* 2001;**413**:591–6.
32. Fukami T, Nakajima M. Community assembly: alternative stable states or alternative transient states? *Ecol Lett* 2011;**14**:973–84. <https://doi.org/10.1111/j.1461-0248.2011.01663.x>.
33. Hastings A, Abbott KC, Cuddington K et al. Transient phenomena in ecology. *Science* 2018;**361**:6406, eaat6412. <https://doi.org/10.1126/science.aat6412>.
34. McCann KS. The diversity–stability debate. *Nature* 2000;**405**:228–33. <https://doi.org/10.1038/35012234>.
35. Lozupone CA, Stombaugh JI, Gordon JI et al. Diversity, stability and resilience of the human gut microbiota. *Nature* 2012;**489**:220–30. <https://doi.org/10.1038/nature11550>.
36. Klappenbach JA, Saxman PR, Cole JR et al. rrndb: the ribosomal RNA operon copy number database. *Nucleic Acids Res* 2001;**29**:181–4. <https://doi.org/10.1093/nar/29.1.181>.
37. Lundberg DS, Yourstone S, Mieczkowski P et al. Practical innovations for high-throughput amplicon sequencing. *Nat Methods* 2013;**10**:999–1002. <https://doi.org/10.1038/nmeth.2634>.
38. Stevens JL, Jackson RL, Olson JB. Slowing PCR ramp speed reduces chimera formation from environmental samples. *J Microbiol Methods* 2013;**93**:203–5. <https://doi.org/10.1016/j.mimet.2013.03.013>.
39. Hamady M, Walker JJ, Harris JK et al. Error-correcting barcoded primers for pyrosequencing hundreds of samples in multiplex. *Nat Methods* 2008;**5**:235–7. <https://doi.org/10.1038/nmeth.1184>.
40. Ushio M, Murakami H, Masuda R et al. Quantitative monitoring of multispecies fish environmental DNA using high-throughput sequencing. *Metabarcoding Metagenom* 2018;**2**:1–7.
41. Ushio M. Interaction capacity as a potential driver of community diversity. *Proc R Soc B* 2022;**289**:289. <https://doi.org/10.1098/rspb.2021.2690>.
42. Fujita H, Ushio M, Suzuki K et al. Alternative stable states, nonlinear behavior, and predictability of microbiome dynamics. *Microbiome* 2023;**11**:1–16. <https://doi.org/10.1186/s40168-023-01474-5>.
43. Callahan BJ, McMurdie PJ, Rosen MJ et al. DADA2: High-resolution sample inference from Illumina amplicon data. *Nat Methods* 2016;**13**:581–3. <https://doi.org/10.1038/nmeth.3869>.
44. Wang Q, Garrity GM, Tiedje JM et al. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl Environ Microbiol* 2007;**73**:5261–7. <https://doi.org/10.1128/AEM.00062-07>.
45. Quast C, Pruesse E, Yilmaz P et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res* 2013;**41**:D590–6. <https://doi.org/10.1093/nar/gks1219>.
46. Oksanen J, Guillaume B, Michael F, et al. vegan: Community Ecology Package. <https://cran.r-project.org/package=vegan>. (Accessed November 14, 2023).
47. Anderson MJ. A new method for non-parametric multivariate analysis of variance. *Austral Ecol* 2001;**26**:32–46.
48. Anderson MJ, Ellingsen KE, McArdle BH. Multivariate dispersion as a measure of beta diversity. *Ecol Lett* 2006;**9**:683–93. <https://doi.org/10.1111/j.1461-0248.2006.00926.x>.
49. Drost H-G. Philentropy: information theory and distance quantification with R. *J Open Source Softw* 2018;**3**:765. <https://doi.org/10.21105/joss.00765>.
50. Scrucca L, Fop M, Murphy TB et al. mclust 5: clustering, classification and density estimation using Gaussian finite mixture models. *R J* 2016;**8**:289–317. <https://doi.org/10.32614/RJ-2016-021>.
51. Rousseeuw PJ. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *J Comput Appl Math* 1987;**20**:53–65. [https://doi.org/10.1016/0377-0427\(87\)90125-7](https://doi.org/10.1016/0377-0427(87)90125-7).
52. Maechler M, Rousseeuw P, Struyt A, Hubert M. cluster: cluster analysis basics and extensions. R package version 2.1.4 — For new features, see the ‘Changelog’ file (in the package source). <https://CRAN.R-project.org/package=cluster>. (Accessed november 14, 2023).
53. Douglas GM, Maffei VJ, Zaneveld JR et al. PICRUSt2 for prediction of metagenome functions. *Nat Biotechnol* 2020;**38**:685–8. <https://doi.org/10.1038/s41587-020-0548-6>.
54. Goldford JE, Lu N, Bajić D et al. Emergent simplicity in microbial community assembly. *Science* 2018;**361**:469–74. <https://doi.org/10.1126/science.aat1168>.
55. Dal Bello M, Lee H, Goyal A et al. Resource–diversity relationships in bacterial communities reflect the network structure of microbial metabolism. *Nat Ecol Evol* 2021;**5**:1424–34. <https://doi.org/10.1038/s41559-021-01535-8>.
56. Sloan WT, Lunn M, Woodcock S et al. Quantifying the roles of immigration and chance in shaping prokaryote community structure. *Environ Microbiol* 2006;**8**:732–40. <https://doi.org/10.1111/j.1462-2920.2005.00956.x>.
57. Zhou J, Liu W, Deng Y et al. Stochastic assembly leads to alternative communities with distinct functions in a bioreactor microbial community. *MBio* 2013;**4**:4. <https://doi.org/10.1128/mBio.00584-12>.
58. Burns AR, Stephens WZ, Stagaman K et al. Contribution of neutral processes to the assembly of gut microbial communities in the zebrafish over host development. *ISME J* 2015;**10**:655–64. <https://doi.org/10.1038/ismej.2015.142>.
59. Kraft NJB, Adler PB, Godoy O et al. Community assembly, coexistence and the environmental filtering metaphor. *Funct Ecol* 2015;**29**:592–9. <https://doi.org/10.1111/1365-2435.12345>.
60. Ning D, Deng Y, Tiedje JM et al. A general framework for quantitatively assessing ecological stochasticity. *Proc Natl Acad Sci U S A* 2019;**116**:16892–8. <https://doi.org/10.1073/pnas.1904623116>.
61. Louca S, Jacques SMS, Pires APF et al. High taxonomic variability despite stable functional structure across microbial communities. *Nat Ecol Evol* 2016;**1**:1–12. <https://doi.org/10.1038/s41559-016-0015>.
62. Louca S, Parfrey LW, Doebeli M. Decoupling function and taxonomy in the global ocean microbiome. *Science* 2016;**353**:1272–7. <https://doi.org/10.1126/science.aaf4507>.
63. Yajima D, Fujita H, Hayashi I et al. Core species and interactions prominent in fish-associated microbiome dynamics. *Microbiome* 2023;**11**:1–15. <https://doi.org/10.1186/s40168-023-01498-x>.
64. Alneberg J, Bennis C, Beier S et al. Ecosystem-wide metagenomic binning enables prediction of ecological niches from genomes. *Commun Biol* 2020;**3**:1–10. <https://doi.org/10.1038/s42003-020-0856-x>.
65. Fahimipour AK, Gross T. Mapping the bacterial metabolic niche space. *Nat Commun* 2020;**11**:1–8. <https://doi.org/10.1038/s41467-020-18695-z>.
66. Fujita H, Ushio M, Suzuki K et al. Facilitative interaction networks in experimental microbial community dynamics. *Front Microbiol* 2023;**14**:1153952. <https://doi.org/10.3389/fmicb.2023.1153952>.
67. Klausmeier CA. Regular and irregular patterns in semi-arid vegetation. *Science* 1999;**284**:1826–8. <https://doi.org/10.1126/science.284.5421.1826>.



68. Rietkerk M, Dekker SC, De Ruiter PC et al. Self-organized patchiness and catastrophic shifts in ecosystems. *Science* 2004;**305**: 1926–9. <https://doi.org/10.1126/science.1101867>.
69. De Roy K, Marzorati M, Van den Abbeele P et al. Synthetic microbial ecosystems: an exciting tool to understand and apply microbial communities. *Environ Microbiol* 2014;**16**:1472–81. <https://doi.org/10.1111/1462-2920.12343>.
70. Vrancken G, Gregory AC, Huys GRB et al. Synthetic ecology of the human gut microbiota. *Nat Rev Microbiol* 2019;**17**:754–63. <https://doi.org/10.1038/s41579-019-0264-8>.
71. Carlström CI, Field CM, Bortfeld-Miller M et al. Synthetic microbiota reveal priority effects and keystone strains in the Arabidopsis phyllosphere. *Nat Ecol Evol* 2019;**3**:1445–54. <https://doi.org/10.1038/s41559-019-0994-z>.
72. Lloréns-Rico V, Simcock JA, Huys GRB et al. Single-cell approaches in human microbiome research. *Cell* 2022;**185**: 2725–38. <https://doi.org/10.1016/j.cell.2022.06.040>.
73. Liu YX, Qin Y, Bai Y. Reductionist synthetic community approaches in root microbiome research. *Curr Opin Microbiol* 2019;**49**:97–102. <https://doi.org/10.1016/j.mib.2019.10.010>.
74. Marsland R, Cui W, Goldford J et al. Available energy fluxes drive a transition in the diversity, stability, and functional structure of microbial communities. *PLoS Comput Biol* 2019;**15**:15. <https://doi.org/10.1371/journal.pcbi.1006793>.
75. Machado D, Maistrenko OM, Andrejev S et al. Polarization of microbial communities between competitive and cooperative metabolism. *Nat Ecol Evol* 2021;**5**:195–203. <https://doi.org/10.1038/s41559-020-01353-4>.
76. Dukovski I, Bajić D, Chacón JM et al. A metabolic modeling platform for the computation of microbial ecosystems in time and space (COMETS). *Nat Protoc* 2021;**16**:5030–82. <https://doi.org/10.1038/s41596-021-00593-3>.
77. Hsieh CH, Glaser SM, Lucas AJ et al. Distinguishing random environmental fluctuations from ecological catastrophes for the North Pacific Ocean. *Nature* 2005;**435**:336–40. <https://doi.org/10.1038/nature03553>.
78. Ushio M, Hsieh C, Masuda R et al. Fluctuating interaction network and time-varying stability of a natural fish community. *Nature* 2018;**554**:360–3. <https://doi.org/10.1038/nature25504>.