Metagenomic analysis of ecological niche overlap and community collapse in microbiome dynamics

Hiroaki Fujita[1†], Masayuki Ushio[1,2], Kenta Suzuki[3], Masato S. Abe[4], Masato Yamamichi[5,6], Yusuke Okazaki[7], Alberto Canarini[1], Ibuki Hayashi[1], Keitaro Fukushima[8], Shinji Fukuda[9-12], E. Toby Kiers[13], and Hirokazu Toju[1†]

[1]Center for Ecological Research, Kyoto University, Otsu, Shiga 520-2133, Japan

[2]Department of Ocean Science (OCES), The Hong Kong University of Science and Technology (HKUST), Clear Water Bay, Kowloon, Hong Kong SAR, China

[3]Integrated Bioresource Information Division, BioResource Research Center, RIKEN, Tsukuba, Ibaraki 305-0074, Japan

[4]Faculty of Culture and Information Science, Doshisha University, Kyotanabe, Kyoto 610-0321, Japan

[5]School of Biological Sciences, The University of Queensland, St. Lucia, Brisbane, QLD 4072, Australia

[6]Department of International Health and Medical Anthropology, Institute of Tropical Medicine, Nagasaki University, Nagasaki 852-8523, Japan

[7]Institute for Chemical Research, Kyoto University, Gokasho, Uji, Kyoto 611-0011, Japan

[8]Faculty of Food and Agricultural Sciences, Fukushima University, Kanayagawa 1, Fukushima, Fukushima 960-1296, Japan.

[9]Institute for Advanced Biosciences, Keio University, Tsuruoka, Yamagata 997-0052, Japan.

[10]Gut Environmental Design Group, Kanagawa Institute of Industrial Science and Technology, Kawasaki, Kanagawa 210-0821, Japan.

[11]Transborder Medical Research Center, University of Tsukuba, Tsukuba, Ibaraki 305-8575,

Japan.

[12]Laboratory for Regenerative Microbiology, Juntendo University Graduate School of Medicine, Tokyo 113-8421, Japan.

[13]Department of Ecological Science, Vrije Universiteit Amsterdam, Amsterdam, the Netherlands

[†]**Correspondence:** Hiroaki Fujita (fujita.h@ecology.kyoto-u.ac.jp) or Hirokazu Toju (toju.hirokazu.4c@kyoto-u.ac.jp).

Abstract

Species utilizing the same resources often fail to coexist for extended periods of time. Such competitive exclusion mechanisms potentially underly microbiome dynamics, causing breakdowns of communities composed of species with similar genetic backgrounds of resource utilization. Although genes responsible for competitive exclusion among a small number of species have been investigated in pioneering studies, it remains a major challenge to integrate genomics and ecology for understanding stable coexistence in species-rich communities. Here, we show that community-scale analyses of functional gene redundancy can provide a useful platform for interpreting and predicting collapse of bacterial communities. Through 110-day time-series of experimental microbiome dynamics, we analyzed the metagenome-assembled genomes of co-occurring bacterial species. We then inferred ecological niche space based on the multivariate analysis of the genome compositions. The analysis allowed us to evaluate potential shifts in the level of niche overlap between species through time. We hypothesized that community-scale pressure of competitive exclusion could be evaluated by quantifying overlap of genetically determined resource-use profiles (metabolic pathway profiles) among coexisting species. We found that the degree of community compositional changes observed in the experimental microbiome was explained by the magnitude of gene-repertoire overlaps among bacterial species. The metagenome-based analysis of genetic potential for competitive exclusion will help us forecast major events in microbiome dynamics such as sudden community collapse (i.e., dysbiosis).

INTRODUCTION

Classic niche theory predicts that coexistence of species requires interspecific difference in resource use patterns [1–6]. Although some specific mechanisms can promote stable coexistence even with complete resource overlap (e.g., spatial structure of habitats and temporal variability in resource availability), similarity/dissimilarity in basic resource dependency among species is the basic factor determining the occurrence of competitive exclusion [7–9]. Therefore, evaluating the overlap of “fundamental niches”, which are defined by species’ fundamental resource requirements and resource-use capabilities [10, 11], is an essential step for understanding and predicting community-level dynamics.

Insights into fundamental niches are encrypted in species' genomes [12–14]: as species' traits are encoded in their DNA, genomic information provides the ultimate basis for evaluating target species' fundamental niches [15, 16]. Thus, potential strength of competitive interactions within ecological guilds or communities could be evaluated based on the distribution of species' gene repertoires within ecological niche space inferred with metagenomic data [12, 15, 16], also referred to as "metagenomic niche space". Although overlap of niches does not always cause competitive exclusion [7–9], higher levels of gene repertoire overlap within a community may impose greater impacts on population dynamics of constituent species.

In considering coexistence of microbial species, it is essential to examine whether such competition-driven population-level phenomena underly drastic ecological events observed at the community level. Microbial communities sometimes show sudden and substantial changes in species and/or taxonomic compositions [17–20]. Human gut microbiomes, for example, have been reported to show drastic shifts from species-rich states to "imbalanced" states with low α-diversity and overrepresentation of pathogenic species [21–24] (e.g., *Clostridium difficile*). Elucidating the ecological mechanisms causing such drastic community-level events provide fundamental insights into microbiome dynamics [24–26]. In this respect, an important challenge is to test the hypothesis that high levels of gene-repertoire overlap are observable prior to drastic community compositional changes. However, this hypothesis, to our knowledge, has not yet been tested presumably due to the paucity of time-series observations of microbiomes with substantial compositional changes. Even if such microbiome time-series data are available, analyses of potential niche (gene repertoire) overlap require another line of information. Specifically, we need data of respective species' genomes at multiple time points. Therefore, developing research systems that can overcome these constrains will deepen our understanding of microbiome ecological processes.

In this study, we test the degree to which gene-repertoire overlap changes through dynamics of species-rich microbial communities. By targeting an experimental microbial system showing rapid and substantial changes in taxonomic compositions [19], we infer niche space depicting species' gene repertoires. Work in this system using a metabolic modeling analysis demonstrated that interactions between species were keys to understand the drastic microbiome dynamics [27]. Now, by compiling the shotgun metagenomic data collected at 13 time points across the 110-day time-series of the experiment, we reveal temporal shifts in the magnitude of gene repertoire overlap among microbial species. We then examine whether a

high level of fundamental-niche overlap is observed prior to drastic changes in community structure. Overall, we explore how signs of drastic shifts in community structure are detected by inferring community-scale degree of fundamental niche overlap with the aid of genomic information.

RESULTS

Functional dynamics of microbiomes

We focused on the experimental microbiome showing drastic shifts in taxonomic compositions [19]. In a previous study [19], a 110-day monitoring of microbiomes was performed with six experimental settings. To set up experimental microbiomes with high diversity of bacterial species/taxa, we used natural microbial communities derived from soil or pond-water ecosystems as source inocula, rather than "synthetic" communities with pre-defined diversity. Specifically, microbiomes were set up with combinations of two source inoculum types (soil- or pond-water-derived inoculum microbiomes) and three medium types (oatmeal, oatmeal-peptone, or peptone broth medium) with eight replications ($2 \times 3 \times 8 = 48$ microbiomes; see Materials and Methods for details). From each of the 48 microbiomes, a fraction of each replicate community was sampled every 24 hours. The collected samples were subjected to the amplicon sequencing of the 16S rRNA region and the temporal changes in community compositions were monitored throughout the time-series [19]. By calculating the magnitude of time-series changes in community compositions [19] (Fig. 1A), we focused on a water-inoculum/oatmeal-medium replicate community showing the most abrupt (rapid and substantial) changes in community compositions among the 48 microbiomes examined as described in a study on metabolic interactions between species [27] (Fig. S1).

By targeting the replicate community mentioned above, we compiled shotgun metagenomic data at 13 time points across the time-series [27]. In total, 32 high-quality (> 80 % completeness and < 5 % contamination) metagenome-assembled genomes (MAGs) belonging to 20 genera (16 families; 12 orders) were detected [27] (Figs. 1B-C and 2; Fig. S2; Table S1). As indicated in the amplicon sequencing analysis [19] (Fig. 1A), drastic shifts from taxon-rich community states to oligopolistic states was observed around Day 20 in the shotgun sequencing analysis (Fig. 1B).

After the drastic community compositional change, the system reached a quasi-stable state represented by the dominance of a *Hydrotalea* (Chitinophagaceae) bacterium (Fig. 1B).

131 The MAG of the *Hydrotalea* was characterized by relatively low GC content (38 %) and
132 relatively small genome size within the community (ca. 3.1 Mb; Fig. 2A). In contrast, the two
133 bacterial MAGs consistently coexisted with the dominant *Hydrotalea* through the time-series
134 (i.e., *Terracidiphilus* and *Mangrovibacter*) had larger genome size (4.2 and 5.4 Mb,
135 respectively; Fig. 1C), characterized by various genes absent from the *Hydrotalea* genome
136 (Fig. 2; Fig. S3). Specifically, the *Terracidiphilus* MAG showed metabolic
137 pathways/processes for degrading plant-derived biopolymers (e.g., cellulose; Fig. 2),
138 potentially surviving as a primary user of polymer compounds within the plant-derived
139 (oatmeal) medium. Meanwhile, the *Mangrovibacter* MAG had pathways/processes related to
140 starch degradation (e.g., amylase) and vitamin-B_{12} transportation, which were absent from the
141 genomes of *Hydrotalea, Terracidiphilus*, and the other MAG (*Rhizomicrobium*) detected on
142 Day 40-60 (Fig. 2).

143

144 **Multivariate analysis of gene repertoires**

145 Next, we used the shotgun metagenomic data to evaluate how the level of gene repertoire
146 overlap among microbes shifted through time. We anticipated that microbial species with
147 similar resource-use abilities or restrictions have similar genomic structure. Therefore, it is
148 expected that species competing for the same resource tend to form clusters within the space
149 defined based on the principal coordinate analysis (PCoA) of dissimilarity in gene repertoires.
150 For each pair of the 32 MAGs, dissimilarity (Jaccard distance) of gene repertoires was
151 calculated based on the matrix representing the presence/absence of the 6,999 genes annotated
152 with the program Prokka [28]. A PCoA was then performed using the β-diversity information
153 (Fig. 3A). At each of the 13 time points, detected MAGs were plotted on the PCoA space.
154 Since we did not have *a priori* knowledge of specific metabolic pathways keys to the
155 microbe-to-microbe competition within the experimental microbiome, all datasets were
156 included in this multivariate analysis. Given general characteristics of multivariate analysis
157 based on β-diversity metrics, the multivariate reconstruction of ecological niche space
158 depends greatly on the genes whose presence/absence profiles vary among species, while
159 housekeeping genes possessed by most species are expected to contribute little to the
160 multivariate analysis.

161 We found that alphaproteobacterial and gammaproteobacterial MAGs respectively
162 constituted some clusters within the niche space reconstructed based on the multivariate

analysis early in the microbiome dynamics (Day 1-20; Fig. 3B). This state with high niche overlap and potential within-guild competition for resources then collapsed into a simpler community state represented by *Hydrotalea*, *Mangrovibacter*, *Terracidiphilus*, and *Rhizomicrobium* as detailed above (Fig. 3B). The space once occupied by many alphaproteobacterial and gammaproteobacterial MAGs remained unoccupied or sparsely occupied after the community compositional collapse. Even when the number of MAGs detectable with our shotgun-metagenomic sequencing increased again late in the time-series, dense aggregations of microbes with similar genomic compositions remained unobserved (Fig. 3B).

Metagenomic niche overlap

We next quantitatively evaluated dynamics in the magnitude of community-scale niche overlap within the multivariate space (Fig. 3). We developed two types of simple indices for evaluating community-scale niche overlap. The one is defined as the overall mean of gene-repertoire similarity between pairs of MAGs within a community. For a time point, the niche overlap index is calculated as:

$$\text{niche overlap score (overall mean)} = 1 - \frac{\sum_{i \in T, j \in T, i \neq j} D_{ij}}{N_T(N_T - 1)},$$

where T is the set of MAGs detected on a focal day (relative abundance > 0.1 %), D_{ij} is the Jaccard metric of dissimilarity [29] in gene compositions, and N_T is the number of MAGs detected on the day. By definition, this niche overlap value based on Jaccard dissimilarity varies from 0 (completely different repertoires of genes in all pairs of MAGs) and 1 (completely identical gene repertoires in all pairs of MAGs), allowing us to evaluate niche overlap levels of target communities within the standardized ranges. The other index is defined as mean value of gene-repertoire similarity with nearest neighbors. The alternative is calculated as:

$$\text{niche overlap score (nearest mean)} = 1 - \frac{\sum_{i \in T, i \neq j} \min_{j \in T}(D_{ij})}{N_T}.$$

This index can be modified by incorporating the information of the relative abundance of MAGs (p_i) as follows:

$$\text{niche overlap score (weighted nearest mean)} = 1 - \sum_{i \in T, i \neq j} p_i \min_{j \in T}(D_{ij}).$$

The results indicated that the level of niche overlap was the highest on Day 1 or Day 10 and that it decreased until Day 30 (Fig. 4A). Although the niche overlap score remained low between Day 40 and 60, it increased again late in the microbiome time-series (Fig. 4B). Note that α-diversity of the community showed similar temporal shifts and it was significantly associated with the niche overlap indices (Fig. 4B-C). Through the time-series, the estimated niche overlap level was significantly associated with the magnitude of the observed community compositional changes (Fig. 5)

DISCUSSION

By developing simple metrics of among-species overlap of gene repertoires, we examined potential relationship between community-scale niche overlap and drastic changes in community structure. Early in the experimental microbiome dynamics, alphaproteobacterial and gammaproteobacterial species were present, resulting in relatively high niche-overlap scores at the community level (Figs. 3 and 4). The quasi-equilibrium state of microbial compositions then collapsed into another quasi-equilibrium represented by a small number of bacteria varying in genome size and metabolic capabilities. Throughout the time-series, higher niche overlap levels entailed greater changes in microbial community compositions (Fig. 5). These findings lead to the working hypothesis that collapse of microbiome structure is predicted by the level of potential niche overlap within multivariate metagenomic space. In light of the "limiting similarity" rule of ecological niches [30], microbial species that exceed a critical limit of genome compositional similarity are expected to compete for the same resources intensively, eventually driving competitive exclusion processes. Thus, as examined in this study, similarity/dissimilarity in genetically determined resource-use properties (i.e., fundamental niches) sets baselines for consequences of interspecific interactions.

The results also indicated that niche overlap level does not necessarily show monotonic decrease through microbial community processes. Although gene-repertoire overlap level and detectable species richness sharply declined early in the microbiome dynamics, both variables gradually increased again around Day 80 (Figs. 1A and 4B). In the resurgence process, however, the dense clusters of alphaproteobacterial or gammaproteobacterial species detected until Day 20 did not appear again within the niche space (Fig. 3B). These observations

222 suggest that once collapsed, microbial communities may not return to previous states with
223 highest levels of niche overlap, but refilling of poorly-used niches can occur under the
224 constraint of limiting similarity within niche space. Although these insights are useful, our
225 present analysis is based only on 13 time points of a microbiome experiment. Due to the
226 limitation, it remained difficult to separate effects of α-diversity from those of gene-repertoire
227 overlap (Fig. 4B-C). Thus, the statistical analysis proposed in this study need to be expanded
228 by reducing the cost of metagenomic sequencing as well as by developing more efficient
229 pipelines for the computationally intensive analyses of metagenomic datasets.

230 The approach of systematically evaluating potential overlap of ecological niches have
231 been previously explored in "community phylogenetics", in which phylogenetic
232 overdispersion/clustering is evaluated based on null model analysis of random assembly from
233 species pools [5, 31, 32]. In those studies based on phylogenetic analyses, similarity of niches
234 has been inferred based on the assumption that phylogenetically similar species have similar
235 ecological properties (e.g., resource requirements). Nonetheless, given that convergent
236 evolution of ecologically important traits is ubiquitous in the history of life [33–35], the
237 assumption of phylogenetic niche conservatism is not always met [36]. Therefore, because
238 gene repertoires are more direct proxies of species traits than phylogeny, metagenome-based
239 analyses will deepen our understanding of community processes driven by competitive
240 exclusion. Meanwhile, in the present analyses of gene repertoire overlap, we included whole
241 metagenomic datasets of the examined microbes due to the lack of *a priori* insights into the
242 metabolic pathways/processes playing essential roles in interspecific competition for
243 resources. In this respect, our analysis is a preliminary conceptual step for evaluating potential
244 overlap of fundamental niches at the community level. In future studies, analyses excluding
245 housekeeping genes [37, 38] or those focusing on specific functional groups of genes (e.g.,
246 carbohydrate degrading genes [39]) may provide more reliable inference of niche overlap.
247 Because such selection of genes can critically influence threshold niche-overlap values for
248 anticipating abrupt community compositional changes, setting a commonly applicable
249 criterion of choosing target gene sets will help us perform comparative analyses across a wide
250 range of microbial communities.

251 The simple framework for evaluating overlap of fundamental niches is applicable to
252 diverse types of microbiomes. Given that our Jaccard-dissimilarity-based indices are
253 standardized within the range from 0 to 1, the next crucial step is to examine how threshold
254 niche overlap values for anticipating microbial community collapse vary among different

types of ecosystems. Such threshold values can vary among ecosystems depending on their basic levels of sustainable functional redundancy. In our laboratory microbiome, for example, the lack of spatial structure (e.g., refuges for inferior species) and environmental fluctuations (e.g., temperature fluctuations) might have severely limited coexistence of functionally similar species (species with similar metabolic capabilities). In contrast, in human gut microbiomes, spatial complexity [40, 41] and temporally fluctuating environmental conditions [22] may reduce the risk of competitive exclusion, allowing higher levels of niche overlap within communities. Thus, extension of time-series metagenomic analyses to diverse types of ecosystems [42–45] will enhance our knowledge of relationship among ecosystem properties, functional redundancy, and community stability.

While genomic information provides an ultimate platform for inferring fundamental niches [12–14], overlap of gene repertoires may not always result in competitive exclusion of species within communities. Even in a pair of species with similar gene repertoires, differentiation in gene expression patterns may occur to avoid overlap of resource-use patterns between species, allowing coexistence of the two species in an environment. Such differentiation of "realized niches [10]" through phenotypic plasticity is potentially evaluated by transcriptomic or metabolomic analyses [46, 47]. Consequently, integration of (meta)transcriptome and (meta)metabolome analyses [48–50] with metagenome-based analyses will reorganize our understanding of deterministic processes in microbiome dynamics.

277 # MATERIALS AND METHODS

278 **Time-series data of experimental microbiomes**

279 We used the experimental system of the microbiome time-series monitoring described in a
280 previous study [19]. In the experiment, microbiomes differing in the magnitude of community
281 compositional shifts were constructed across the six treatments defined by the combinations
282 of two inoculum sources and three types of media. One of the source microbiomes derived
283 from the soil collected from the A layer (0-10 cm in depth) in the research forest of Center for
284 Ecological Research, Kyoto University, Otsu, Japan (34.972 ºN; 135.958 ºE). The other
285 source inoculum was prepared by collecting water from a pond ("Shoubuike") near Center for
286 Ecological Research (34.974 ºN, 135.966 ºE). Each of the source inocula was introduced into
287 oatmeal (Medium-A), oatmeal-peptone (Medium-B), or peptone (Medium-C) broth media
288 with eight replicates. Thus, in total, 48 experimental microcosms (two source microbiomes ×
289 three media × eight replicates) were constructed in a deep-well plate (1000-μL-scale culture in
290 each well). The plate was kept shaken at 1,000 rpm at 23 ºC. After five-day pre-incubation,
291 200 μL out of the 1,000-μL culture medium was sampled from each well every 24 hours for
292 110 days. In each sampling event, 200 μL of fresh medium was added to each well so that the
293 total culture volume was kept constant. In total, 5,280 samples (48 communities/day × 110
294 days) were collected through the time-series experiment. After DNA extraction, the samples
295 were subjected to the amplicon sequencing analysis of the 16S rRNA region [19]. To quantify
296 the speed and magnitude of community shifts through time, the "abruptness" index was
297 calculated through the time-series of each replicate microcosm in each experimental treatment
298 [19]. Specifically, an estimate of the abruptness index for time point t was obtained as the
299 Bray-Curtis β-diversity between average community compositions from time points $t - 4$ to t
300 and those from $t + 1$ to $t + 5$ (i.e., dissimilarity between 5-day time-windows). The Bray-
301 Curtis β-diversity [51] was calculated as $\frac{\sum_{i=1}^{n}|X_{ij}-X_{ik}|}{\sum_{i=1}^{n}(X_{ij}+X_{ik})}$, where X_{ij} and X_{ik} denoted relative
302 abundance of microbial amplicon sequence variant (ASV) i in the compared time windows (j,
303 from $t - 4$ to t; k, from $t + 1$ to $t + 5$). An abruptness score larger than 0.5 indicates that
304 turnover of more than 50 % of community compositions occurred between the time-windows
305 [19].

306

307 **Shotgun metagenomics**

308 Focusing on a replicate microcosm in which the most rapid and substantial turnover of
309 community compositions was observed (replicate no. 5 of Water/Medium-A treatment; Fig.
310 S1), shotgun metagenomic sequencing was conducted by targeting 13 samples (Day 1, 10, 20,
311 24, 30, 40, 50, 60, 70, 80, 90, 100, and 110) as described elsewhere [27]. Specifically, each
312 DNA sample was processed with Nextera XT DNA Library Preparation Kit (Illumina) and
313 sequenced with the DNBSEQ-G400 (BGI; 200-bp paired-end sequencing). From the output
314 data, sequencing adaptors were removed using Cutadapt [52] 2.5 and quality filtering was
315 performed with Fastp [53] 0.21.0: ca. 10 Gb/sample was subjected to the analysis [in total,
316 159.96 Gb (1000.301 M reads)]. The sequences of each sample were assembled with
317 metaSPAdes [54] 3.15.2. Binning was then performed with MetaWRAP [55] 1.3.2, followed
318 by quality assessing with CheckM [56] 1.1.3. The identity between MAGs were calculated
319 using FastANI [57] 1.33 and MAGs with > 99 % identity were dereplicated through the time-
320 series (Table S1). In the dereplication, the MAGs with the highest completeness and N50
321 statistics were selected as representative MAGs. Read-coverage was then calculated with
322 CoverM [58] 0.6.0, followed by taxonomic annotation was performed using GTDB-Tk [59,
323 60] 1.6. Only the MAGs with > 80 % completeness and < 5 % contamination were used in the
324 downstream analyses. Gene annotation was performed with Prokka [28] 1.14.6, yielding
325 6,999 annotated genes (Data S1). To conduct additional functional annotation of genes, the
326 orthology numbers of Kyoto Encyclopedia of Genomes (KEGG) were retrieved using
327 GhostKOALA [61] 2.2. For respective microbial MAGs (bins), completeness of metabolic
328 pathways was estimated with KEGG decoder [62] 1.3. Based on the matrix representing
329 KEGG metabolic pathway/process profiles of respective MAGs (Data S2), a heatmap
330 showing pathway/process completeness was drawn (Fig. S3).

331

332 **Background environmental conditions**

333 For the 13 samples subjected to the shotgun metagenomic analysis, concentrations of
334 ammonium (NH_4^+) and nitrate (NO_3^-) were measured to obtain supplementary information of
335 background environmental conditions. Colorimetric methods with a modified indophenol
336 reaction [63, 64] and the VCl3/Griess assay were applied for the measurements of NH_4^+ and
337 NO_3^-, respectively. Samples were run in triplicates via a standard addition method to account
338 for individual matrix effects [65].

339

340 **Multivariate analysis of the metagenomic space**

341 Based on the whole matrix representing the profiles of the 6,999 genes (Data S1), the Jaccard
342 metric of distance was calculated for each pair of the 32 microbial MAGs (D_{ij}, where i and j
343 represent MAGs). The Jaccard-distance estimates were then used to perform a principal
344 coordinate analysis (PCoA). Using the obtained principal coordinate scores, all the microbial
345 MAGs detected through the time-series were plotted on a multivariate space consisting of the
346 first three PCoA axes (PCoA 1, PCoA 2, and PCoA 3). At each time point, the MAGs detected
347 with the shotgun metagenomic sequencing (defined as the MAGs whose relative abundance is
348 greater than 0.1 %) was plotted on the three-dimensional space defined with the PCoA axes.

349

350 **Evaluation of niche overlap level**

351 The community-scale magnitude of potential niche overlap among species was evaluated
352 based on the shogun metagenomic sequencing dataset. We developed two types of simple
353 indices for evaluating community-scale niche overlap as detailed in the Results section. To
354 test whether a high level of fundamental-niche overlap is observed prior to drastic changes in
355 microbial community structure, we examined relationship between the above niche overlap
356 index and time-series shifts in community structure (Bray-Curtis β-diversity between present
357 and next time points through the time-series of the shotgun metagenomic data).

358

359 # REFERENCES

360 1. Gause GF. The Struggle for Coexistence. 1934. Williams & Wilkins, Baltimore.

361 2. Hardin G. The competitive exclusion principle. *Science (1979)* 1960; **131**: 1292–1297.

362 3. Volterra V. Variations and fluctuations of the number of individuals in animal species
363 living together. *ICES Journal of Marine Science* 1928; **3**: 3–51.

364 4. Zaret TM, Rand AS. Competition in tropical stream fishes: support for the competitive
365 exclusion principle. *Ecology* 1971; **52**: 336–342.

366 5. Mayfield MM, Levine JM. Opposing effects of competitive exclusion on the
367 phylogenetic structure of communities. *Ecol Lett* 2010; **13**: 1085–1093.

6. Grime JP. Competitive exclusion in herbaceous vegetation. *Nature* 1973; **242**: 344–347.

7. Letten AD, Ke PJ, Fukami T. Linking modern coexistence theory and contemporary niche theory. *Ecol Monogr* 2017; **87**: 161–177.

8. Chesson P. Updates on mechanisms of maintenance of species diversity. *Journal of Ecology* 2018; **106**: 1773–1794.

9. Chesson P. Mechanisms of maintenance of species diversity. *Annu Rev Ecol Syst* 2000; **31**: 343–366.

10. Chase JM, Leibold MA. Ecological niches: linking classical and contemporary approaches. *Biodiversity and Conservation* . 2004. University of Chicago Press, Chicago.

11. Hutchinson GE. Concluding Remarks. *Cold Spring Harb Symp Quant Biol* 1957; **22**: 415–427.

12. Régimbeau A, Budinich M, Larhlimi A, Pierella Karlusich JJ, Aumont O, Memery L, et al. Contribution of genome-scale metabolic modelling to niche theory. *Ecol Lett* 2022; **25**: 1352–1364.

13. Smith SR, Dupont CL, McCarthy JK, Broddrick JT, Oborník M, Horák A, et al. Evolution and regulation of nitrogen flux through compartmentalized metabolic networks in a marine diatom. *Nat Commun* 2019; **10**: 4552.

14. Palomo A, Pedersen AG, Fowler SJ, Dechesne A, Sicheritz-Pontén T, Smets BF. Comparative genomics sheds light on niche differentiation and the evolutionary history of comammox *Nitrospira*. *ISME Journal* 2018; **12**: 1779–1793.

15. Alneberg J, Bennke C, Beier S, Bunse C, Quince C, Ininbergs K, et al. Ecosystem-wide metagenomic binning enables prediction of ecological niches from genomes. *Commun Biol* 2020; **3**: 1–10.

16. Fahimipour AK, Gross T. Mapping the bacterial metabolic niche space. *Nat Commun* 2020; **11**: 1–8.

17. Carding S, Verbeke K, Vipond DT, Corfe BM, Owen LJ. Dysbiosis of the gut microbiota in disease. *Microb Ecol Health Dis* 2015; **26**: 26191.

397 18. Ravel J, Brotman RM, Gajer P, Ma B, Nandy M, Fadrosh DW, et al. Daily temporal
398 dynamics of vaginal microbiota before, during and after episodes of bacterial
399 vaginosis. *Microbiome* 2013; **1**: 29.

400 19. Fujita H, Ushio M, Suzuki K, Abe M, Yamamichi M, Iwayama K, et al. Alternative
401 stable states, nonlinear behavior, and predictability of microbiome dynamics.
402 *Microbiome* 2023; **11**: 63.

403 20. Yajima D, Fujita H, Hayashi I, Shima G, Suzuki K, Toju H. Core species and
404 interactions prominent in fish-associated microbiome dynamics. *Microbiome* 2023; **11**:
405 53.

406 21. Lahti L, Salojärvi J, Salonen A, Scheffer M, de Vos WM. Tipping elements in the
407 human intestinal ecosystem. *Nat Commun* 2014; **5**: 4344.

408 22. David LA, Maurice CF, Carmody RN, Gootenberg DB, Button JE, Wolfe BE, et al.
409 Diet rapidly and reproducibly alters the human gut microbiome. *Nature* 2014; **505**:
410 559–563.

411 23. Kho ZY, Lal SK. The human gut microbiome - A potential controller of wellness and
412 disease. *Front Microbiol* 2018; **9**: 1835.

413 24. Kriss M, Hazleton KZ, Nusbacher NM, Martin CG, Lozupone CA. Low diversity gut
414 microbiota dysbiosis: drivers, functional implications and recovery. *Curr Opin*
415 *Microbiol* 2018; **44**: 34–40.

416 25. Costello EK, Stagaman K, Dethlefsen L, Bohannan BJM, Relman DA. The application
417 of ecological theory toward an understanding of the human microbiome. *Science*
418 *(1979)* 2012; **336**: 1255–1262.

419 26. Huttenhower C, Gevers D, Knight R, Abubucker S, Badger JH, Chinwalla AT, et al.
420 Structure, function and diversity of the healthy human microbiome. *Nature* 2012; **486**:
421 207–214.

422 27. Fujita H, Ushio M, Suzuki K, Abe MS, Yamamichi M, Okazaki Y, et al. Facilitative
423 interaction networks in experimental microbial community dynamics. *Front*
424 *Microbiol* . 2023. , **14**: 1153952

425 28. Seemann T. Prokka: Rapid prokaryotic genome annotation. *Bioinformatics* 2014; **30**:

426 2068–2069.

427 29. Anderson MJ, Crist TO, Chase JM, Vellend M, Inouye BD, Freestone AL, et al.
428 Navigating the multiple meanings of β diversity: A roadmap for the practicing
429 ecologist. *Ecol Lett* 2011; **14**: 19–28.

430 30. Macarthur R, Levins R. The limiting similarity, convergence, and divergence of
431 coexisting species. *American Naturalist* 1967; **101**: 377–385.

432 31. Webb CO, Ackerly DD, McPeek MA, Donoghue MJ. Phylogenies and community
433 ecology. *Annu Rev Ecol Syst* 2002; **33**: 475–505.

434 32. Cavender-Bares J, Kozak KH, Fine PVA, Kembel SW. The merging of community
435 ecology and phylogenetic biology. *Ecol Lett* 2009; **12**: 693–715.

436 33. Marvig RL, Sommer LM, Molin S, Johansen HK. Convergent evolution and adaptation
437 of Pseudomonas aeruginosa within patients with cystic fibrosis. *Nat Genet* 2015; **47**:
438 57–64.

439 34. McCutcheon JP, McDonald BR, Moran NA. Convergent evolution of metabolic roles
440 in bacterial co-symbionts of insects. *Proc Natl Acad Sci U S A* 2009; **106**: 15394–
441 15399.

442 35. Merhej V, Royer-Carenzi M, Pontarotti P, Raoult D. Massive comparative genomic
443 analysis reveals convergent evolution of specialized bacteria. *Biol Direct* 2009; **4**: 13.

444 36. Losos JB. Phylogenetic niche conservatism, phylogenetic signal and the relationship
445 between phylogenetic relatedness and ecological similarity among species. *Ecol Lett* .
446 2008. , **11**: 995–1007

447 37. Gibson DG, Glass JI, Lartigue C, Noskov VN, Chuang R-Y, Algire M a, et al. Creation
448 of a bacterial cell controlled by a chemically synthesized genome. *Science* 2010; **329**:
449 52–6.

450 38. Maiden MCJ, van Rensburg MJJ, Bray JE, Earle SG, Ford SA, Jolley KA, et al. MLST
451 revisited: The gene-by-gene approach to bacterial genomics. *Nat Rev Microbiol* 2013;
452 **11**: 728–736.

453 39. Flint HJ, Scott KP, Duncan SH, Louis P, Forano E. Microbial degradation of complex
454 carbohydrates in the gut. *Gut Microbes* 2012; **3**: 289–306.

455 40. Tropini C, Earle KA, Huang KC, Sonnenburg JL. The Gut Microbiome: Connecting
456 Spatial Organization to Function. *Cell Host Microbe* 2017; **21**: 433–442.

457 41. Earle KA, Billings G, Sigal M, Lichtman JS, Hansson GC, Elias JE, et al. Quantitative
458 Imaging of Gut Microbiota Spatial Organization. *Cell Host Microbe* 2015; **18**: 478–
459 488.

460 42. Jansson JK, Hofmockel KS. Soil microbiomes and climate change. *Nat Rev Microbiol* .
461 2020. , **18**: 35–46

462 43. Fierer N. Embracing the unknown: Disentangling the complexities of the soil
463 microbiome. *Nat Rev Microbiol* 2017; **15**: 579–590.

464 44. Trivedi P, Leach JE, Tringe SG, Sa T, Singh BK. Plant–microbiome interactions: from
465 community assembly to plant health. *Nat Rev Microbiol* 2020; **18**: 607–621.

466 45. Venter JC, Remington K, Heidelberg JF, Halpern AL, Rusch D, Eisen JA, et al.
467 Environmental Genome Shotgun Sequencing of the Sargasso Sea. *Science (1979)*
468 2004; **304**: 66–74.

469 46. Nowinski B, Moran MA. Niche dimensions of a marine bacterium are identified using
470 invasion studies in coastal seawater. *Nat Microbiol* 2021; **6**: 524–532.

471 47. Pereira FC, Berry D. Microbial nutrient niches in the gut. *Environ Microbiol* 2017; **19**:
472 1366–1378.

473 48. Heintz-Buschart A, Wilmes P. Human gut microbiome: function matters. *Trends*
474 *Microbiol* 2018; **26**: 563–574.

475 49. Schirmer M, Franzosa EA, Lloyd-Price J, McIver LJ, Schwager R, Poon TW, et al.
476 Dynamics of metatranscription in the inflammatory bowel disease gut microbiome. *Nat*
477 *Microbiol* 2018; **3**: 337–346.

478 50. Turner TR, Ramakrishnan K, Walshaw J, Heavens D, Alston M, Swarbreck D, et al.
479 Comparative metatranscriptomics reveals kingdom level changes in the rhizosphere
480 microbiome of plants. *ISME Journal* 2013; **7**: 2248–2258.

481 51. Legendre P, de Cáceres M. Beta diversity as the variance of community data:
482 Dissimilarity coefficients and partitioning. *Ecol Lett* 2013; **16**: 951–963.

483 52. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing
484 reads. *EMBnet J* 2011; **17**: https://doi.org/10.14806/ej.17.1.200.

485 53. Chen S, Zhou Y, Chen Y, Gu J. Fastp: An ultra-fast all-in-one FASTQ preprocessor.
486 *Bioinformatics* 2018; **34**: i884–i890.

487 54. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al.
488 SPAdes: A new genome assembly algorithm and its applications to single-cell
489 sequencing. *Journal of Computational Biology* 2012; **19**: 455–477.

490 55. Uritskiy G v., DiRuggiero J, Taylor J. MetaWRAP—a flexible pipeline for genome-
491 resolved metagenomic data analysis. *Microbiome* 2018; **6**: 158.

492 56. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. CheckM: Assessing
493 the quality of microbial genomes recovered from isolates, single cells, and
494 metagenomes. *Genome Res* 2015; **25**: 1043–1055.

495 57. Jain C, Rodriguez-R LM, Phillippy AM, Konstantinidis KT, Aluru S. High throughput
496 ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nat*
497 *Commun* 2018; **9**: 5114.

498 58. Woodcroft B. CoverM: program available at https://github.com/wwood/CoverM. 2021.

499 59. Chaumeil PA, Mussig AJ, Hugenholtz P, Parks DH. GTDB-Tk: A toolkit to classify
500 genomes with the genome taxonomy database. *Bioinformatics* 2020; **36**: 1925–1927.

501 60. Parks DH, Chuvochina M, Rinke C, Mussig AJ, Chaumeil P-A, Hugenholtz P. GTDB:
502 an ongoing census of bacterial and archaeal diversity through a phylogenetically
503 consistent, rank normalized and complete genome-based taxonomy. *Nucleic Acids Res*
504 2022; **50**: D785–D794.

505 61. Kanehisa M, Sato Y, Morishima K. BlastKOALA and GhostKOALA: KEGG Tools for
506 Functional Characterization of Genome and Metagenome Sequences. *J Mol Biol* 2016;
507 **428**: 726–731.

508 62. Graham ED, Heidelberg JF, Tully BJ. Potential for primary productivity in a globally-
509 distributed bacterial phototroph. *ISME Journal* 2018; **12**: 1861–1866.

510 63. Kandeler E, Gerber H. Short-term assay of soil urease activity using colorimetric
511 determination of ammonium. *Biol Fertil Soils* 1988; **6**: 68–72.

64. Hood-Nowotny R, Umana NH-N, Inselbacher E, Oswald- Lachouani P, Wanek W. Alternative methods for measuring inorganic, organic, and total dissolved nitrogen in soil. *Soil Science Society of America Journal* 2010; **74**: 1018–1027.

65. Taylor BW, Keep CF, Hall RO, Koch BJ, Tronstad LM, Flecker AS, et al. Improving the fluorometric ammonium method: Matrix effects, background fluorescence, and standard additions. *J North Am Benthol Soc* 2007; **26**: 167–177.

ACKNOWLEDGEMENTS

Computation time was provided by the SuperComputer System, Institute for Chemical Research, Kyoto University. This work was financially supported by JST PRESTO (JPMJPR16Q6), Human Frontier Science Program (RGP0029/2019), JSPS Grant-in-Aid for Scientific Research (20K20586), NEDO Moonshot Research and Development Program (JPNP18016), and JST FOREST (JPMJFR2048) to H.T., JSPS Grant-in-Aid for Scientific Research (20K06820 and 20H03010) to K.S., and JSPS Fellowship to H.F. and A.C.

AUTHOR CONTRIBUTIONS

H.T. designed the work with H.F. H.F. and A.C. performed experiments. H.F. analyzed the data with Y.O., and H.T. H.F. and H.T. wrote the paper with all the authors.

COMPETING INTERESTS

The authors declare no competing interests.

DATA AVAILABILITY STATEMENT

The 16S rRNA sequencing data reported in a previous study [19] are available from the DNA Data Bank of Japan (DDBJ) with the accession number DRA013352, DRA013353, DRA013354, DRA013355, DRA013356, DRA013368 and DRA013379. The shotgun metagenomic data reported previous [27] are available with the DDBJ accession number DRA013382. The microbial community data are deposited at our GitHub repository (https://github.com/hiroakif93/MTS_nicheSpace) [to be publicly available after acceptance of the paper]. The matrices of the shotgun metagenomic data are available as Data S1 and 2. All the scripts used to analyze the data are available at the GitHub repository (https://github.com/hiroakif93/MTS_nicheSpace) [to be publicly available after acceptance of the paper].

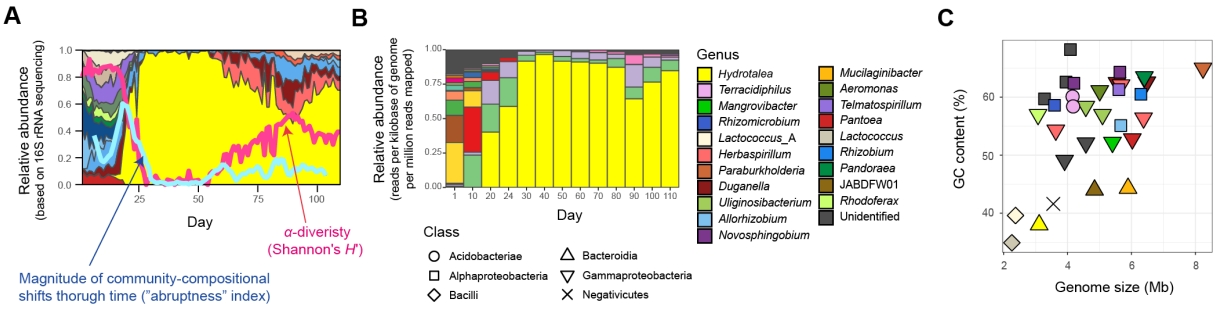

Fig. 1 Community and ecosystem dynamics. A Time-series data of community structure. For the replicate microcosm that showed the most abrupt community compositional changes through the 110-day microbiome experiment [19] (Fig. S1), family-level taxonomic compositions inferred with 16S rRNA sequencing are shown. The blue line represents the speed and magnitude of community compositional changes around each time point ("abruptness" index [19]; see Materials and Methods). The red line indicates α-diversity (Shannon's *H'*) of microbial ASVs [19]. Note that a value larger than 0.5 represents turnover of more than 50 % of microbial ASV compositions. See Fig. S1 for color profiles of bacterial families. Reproduced from the data of a previous study [19]. **B** Taxonomic compositions inferred with shotgun metagenomic sequencing. At each of the 13 time points through the time-series of the target microcosm, the relative abundance of each MAG was estimated based on the normalized read coverage value (reads per kilobase of genome per million reads mapped). **C** Genome size and GC nucleotide content of the MAGs detected in the target microcosm. See panel **A** for colors and symbols.

Fig. 2 Metabolic pathway/process profiles of the MAGs. KEGG metabolic pathways/profiles of the reconstructed bacterial genomes (MAGs) are shown. The detection (relative abundance > 0.1 %) of each microbial MAG on each day within the shotgun metagenomic data is indicated in the panel below. Only the microbial MAGs with > 80 % completeness and < 5 % contamination were included (Table S1). The five MAGs that co-occurred from Day 40 to 60 and metabolic pathways/processes mentioned in the main text are highlighted. Only the metabolic pathways/processes with highly heterogeneous patterns across microbial MAGs are shown. See Fig. S3 for detailed profiles of the metabolic

pathways/processes.

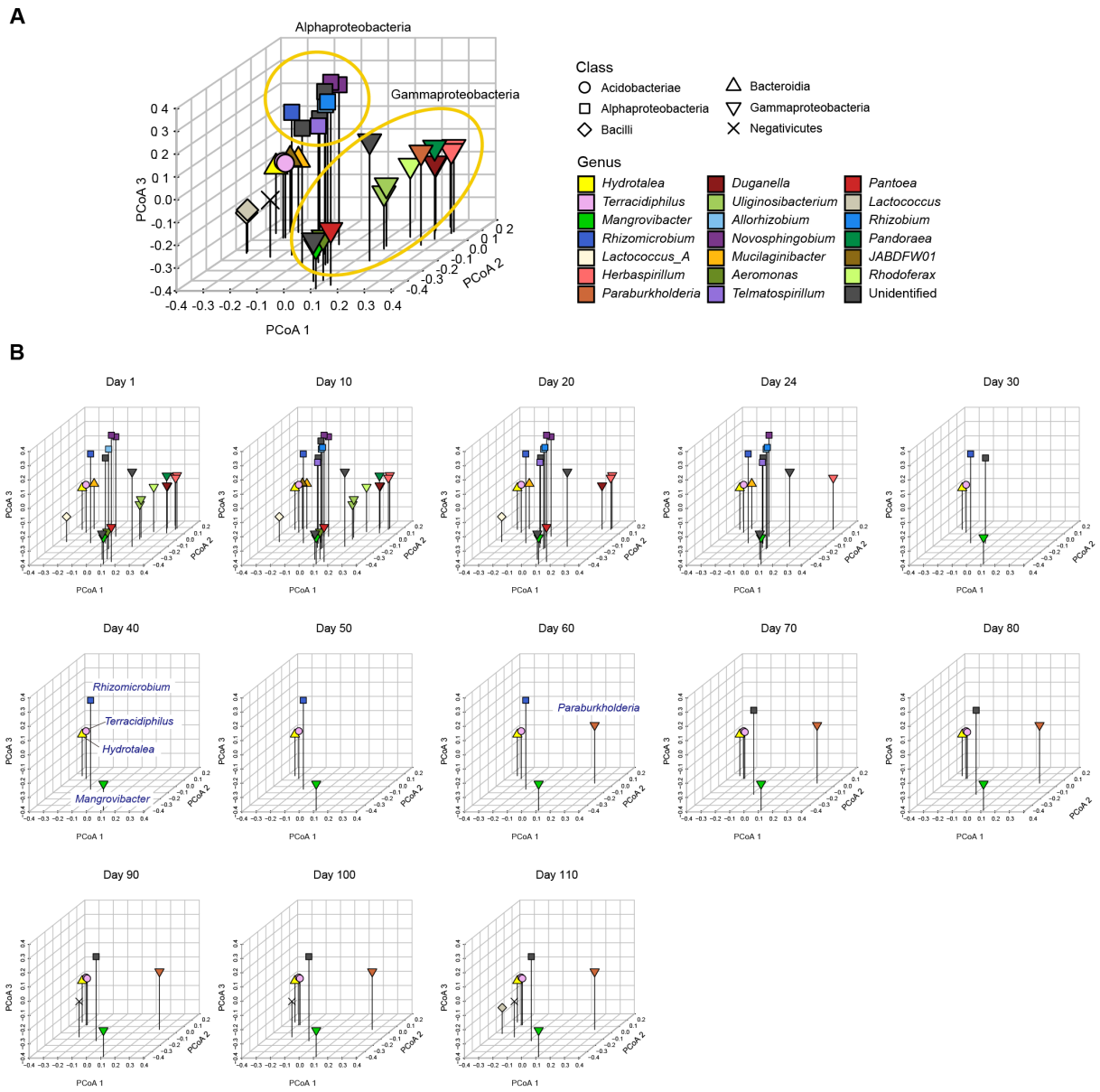

Fig. 3 Metagenomic niche space. **A** Distributions of MAGs within metagenomic niche space. Based on dissimilarity in gene repertoires, microbial MAGs that appeared in the time-series of the target microcosm were plotted on the three-dimensional space defined by the principal coordinate analysis (PCoA) of 6,999 genes. **B** Changes in the distributions of microbial MAGs within niche space. At each time point, detected MAGs (relative abundance > 0.1 %) were plotted on the space defined in the multivariate analysis in the in the panel **A**.

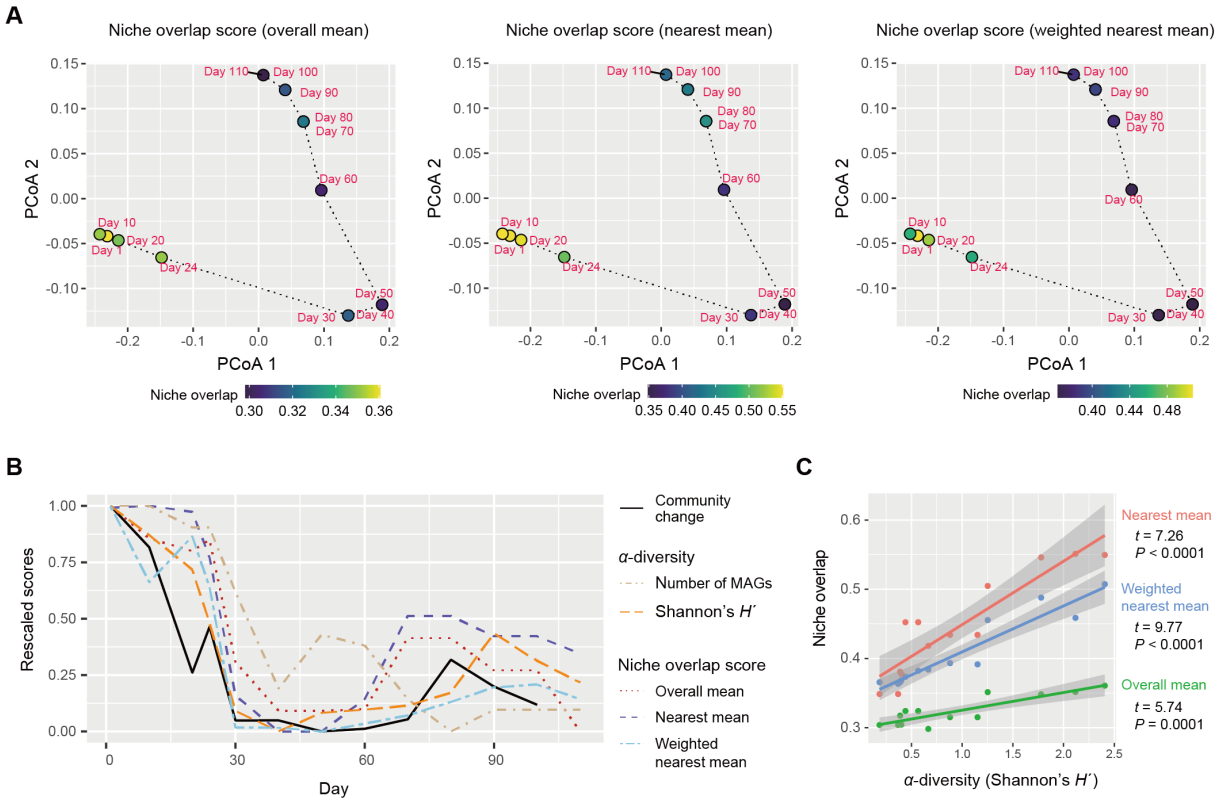

Fig. 4 Dynamics of niche-overlap level. **A** Community-level profiles of metabolic pathways/processes and niche overlap index. The niche overlap indices were defined based on the Jaccard similarity/dissimilarity of gene compositions between pairs of the microbial MAGs detected at a target time point. Three types of niche overlap indices are shown on a PCoA surface representing community-level compositions of genes. On the PCoA surface, time points are distributed based on the sum of the gene repertoires of the detected MAGs. **B** Dynamics of niche-overlap levels. Niche overlap scores are shown across the time-series. The magnitude of community compositional changes (Bray-Curtis β-diversity between present and next time points through the time-series of the shotgun metagenomic data) and α-diversity indices of the communities are shown as well. **C** Relationship between α-diversity and niche overlap scores. The lines represent linear regressions (with 95 % confidence intervals).

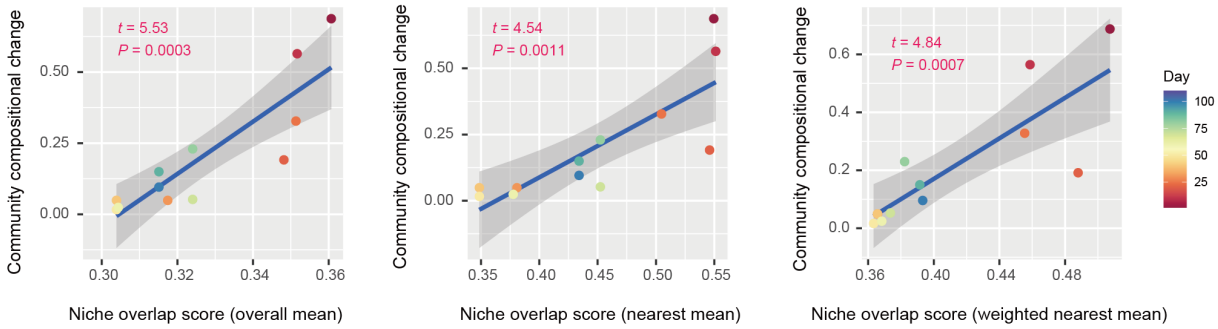

Fig. 5 Niche overlap level and community compositional shifts. The magnitude of community compositional changes observed in the microbiome was regressed on each niche overlap index obtained based on the shotgun metagenomic analysis. Niche overlap index at each time point and time-series shifts in community structure (Bray-Curtis β-diversity between present and next time points through the time-series of the shotgun metagenomic data) are shown along horizontal and vertical axes, respectively. The regression lines are shown with 95 % confidence intervals.