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 and requests for materials should be addressed to H.F. (email: fujita.h@ecology.kyoto-u.ac.jp) or H.T. (email: toju.hirokazu.4c@kyoto-u.ac.jp).

Abstract

Species utilizing the same resources ultimately do not coexist for long periods of time. Such competitive exclusion mechanisms potentially underly dynamics of microbiomes, causing breakdowns of communities constituted by species with similar genetic backgrounds of resource use. Nonetheless, it remains a major challenge to integrate genomics and ecology for understanding deterministic processes of species coexistence in species-rich communities. We here show that community-scale analyses of functional gene redundancy provide statistical platforms for interpreting and predicting collapse of bacterial communities. Through 110-day time-series of experimental microbiome dynamics, we analyzed the metagenome-assembled genomes of coexisting bacterial species. We then reconstructed ecological niche space based on the multivariate analysis of the genome compositions in order to evaluate potential shifts in the level of niche overlap between species through time. Specifically, 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 metabolic pathway (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).

Classic niche theory predicts that coexistence of species requires interspecific difference in resource ranges[1–6]. Although some specific mechanisms can guarantee coexistence even in the presence of niche overlap (e.g., spatial structure of habitats and temporal variability in resource availability), similarity/dissimilarity in basic resource dependency among species is the key 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]. In other words, as species' traits are encoded in their DNA, reconstructed genomes provide 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 reconstructed with metagenomic data[12,15,16] ("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.

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[17–19]. 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[20–23] (e.g., *Clostridium difficile*). Elucidating the ecological mechanisms by which such drastic community-level events are caused provide fundamental insights into microbiome dynamics[23–25]. In this respect, an important challenge is to test the hypothesis that higher levels of gene-repertoire overlap are observable prior to drastic community compositional changes than after such 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, data of respective species' genomes at multiple time points. Therefore, developing research systems that overcome these current constrains is expected to deepen our understanding of microbiome ecological processes.

We here show how degree of 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 reconstruct niche space depicting species' gene repertoires. Based on a whole-genome shotgun metagenomic analysis at 13 time points within the 110-day time-series of the microbiome 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 reconstructing community-scale degree of fundamental niche overlap with the aid of genomic information.

Results

Target microbiome. 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, natural microbial communities derived from soil or pond-water ecosystems, rather than "synthetic" communities with pre-defined diversity, were used as source inocula. 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 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] ("abruptness" index; Fig. 1a; Supplementary Fig. 1), 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 (Fig. 1a; Supplementary Fig. 1).

Functional dynamics of microbiomes. By targeting the replicate community mentioned above, we performed whole-genome shotgun sequencing at 13 time points across the time-series. In total, 32 high-quality (> 80 % completeness and < 5 % contamination) metagenome-assembled genomes (MAGs) belonging to 20 genera (16 families; 12 orders) were detected (Figs. 1b-c and 2; Supplementary Fig. 2; Supplementary Table 1). 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). The MAG of the *Hydrotalea* was characterized by relatively low GC content (38 %) and relatively small genome size within the community (ca. 3.1 Mb; Fig. 2a). In contrast, the two bacterial MAGs consistently coexisted with the dominant *Hydrotalea* through the time-series (i.e., *Terracidiphilus* and *Mangrovibacter*) had larger genome size (4.2 and 5.4 Mb, respectively; Fig. 1c), characterized by various genes absent from the *Hydrotalea* genome (Fig. 2; Supplementary Fig. 3). Specifically, the *Terracidiphilus* MAG showed metabolic

132 pathways/processes for degrading plant-derived biopolymers (e.g., cellulose; Fig. 2),
133 potentially surviving as a primary user of polymer compounds within the plant-derived
134 (oatmeal) medium. Meanwhile, the *Mangrovibacter* MAG had pathways/processes related to
135 starch degradation (e.g., amylase) and vitamin-B_{12} transportation, which were absent from the
136 genomes of *Hydrotalea, Terracidiphilus*, and the other MAG (*Rhizomicrobium*) detected on
137 Day 40-60 (Fig. 2).

138

139 **Multivariate analysis of gene repertoires.** The whole-genome shotgun metagenomic data
140 were used to evaluate how the level of gene repertoire overlap among microbes shifted
141 through time. We anticipated that microbial species with similar resource-use abilities or
142 restrictions have similar genomic structure. Therefore, it is expected that species competing
143 for the same resource tend to form clusters within the space defined based on the principal
144 coordinate analysis (PCoA) of dissimilarity in gene repertoires. For each pair of the 32
145 MAGs, dissimilarity (Jaccard β-diversity) of gene repertoires was calculated based on the
146 matrix representing the presence/absence of the 6,999 genes annotated with the program
147 Prokka[26]. A PCoA was then performed using the β-diversity information (Fig. 3a). At each of
148 the 13 time point, detected MAGs were plotted on the PCoA space. Since we did not have *a*
149 *priori* knowledge of specific metabolic pathways keys to the microbe-to-microbe competition
150 within the experimental microbiome, the entire datasets were used in this multivariate
151 analysis. Given general characteristics of multivariate analysis based on β-diversity metrics,
152 the multivariate reconstruction of ecological niche space depends greatly on the genes whose
153 presence/absence profiles vary among species, while housekeeping genes possessed by most
154 species are expected to contribute little to the multivariate analysis.

155 We then found that alphaproteobacterial and gammaproteobacterial MAGs respectively
156 constituted some clusters within the niche space reconstructed based on the multivariate
157 analysis early in the microbiome dynamics (Day 1-20; Fig. 3b). This state with high niche
158 overlap and potential within-guild competition for resources then collapsed into a simpler
159 community state represented by *Hydrotalea*, *Mangrovibacter*, *Terracidiphilus*, *and*
160 *Rhizomicrobium* as detailed above (Fig. 3b). The space once occupied by many
161 alphaproteobacterial and gammaproteobacterial MAGs remained unoccupied or sparsely
162 occupied after the community compositional collapse. Even when the number of MAGs
163 detectable with our shotgun-metagenomic sequencing increased again late in the time-series,
164 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). In our analysis, the niche overlap index was defined as:

$$1 - \frac{\sum_{i \in D, j \in D} \beta_{ij}}{N_D},$$

where D is the set of MAGs detected on a focal day (time point), β_{ij} is the Jaccard metric of dissimilarity in gene compositions, and N_D is the number of MAGs detected on the day. By definition, this niche overlap value 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 results indicated that the level of niche overlap was the highest on Day 1 and that it gradually decreased until Day 20 (Fig. 4a-b). A slight increase in the niche overlap index was observed on Day 24, but it dropped sharply by Day 30 (Fig. 4a-b). Although the niche overlap score remained low between Day 40 and 60, it increased again on Day 70 (Fig. 4b).

We then found that the estimated niche overlap level significantly explained the magnitude of the observed community compositional changes (t = 5.525, df = 10, P = 0.00025; Fig. 5). In other words, higher levels of gene-repertoire overlap within a community were followed by larger shifts in community compositions at subsequent time points.

Discussion

We here showed that among-species overlap of gene repertoires are observable prior to 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 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[27], pairs of microbial species that exceed a

critical limit of genome compositional similarity are expected to compete for the same resources, 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 suggest that once collapsed, microbial communities may not return to previous states with highest levels of niche overlap, but refilling of poorly-used niches can occur under the constraint of limiting similarity within niche space.

The simple framework for evaluating overlap of fundamental niches is applicable to diverse types of microbiomes. Given that our β-diversity-based index is standardized within the range from 0 to 1, next crucial step is to examine how threshold 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 environmental fluctuations (e.g., temperature fluctuations) and spatial structure (e.g., refuges for inferior species) might severely limited coexistence of functionally similar species (species with similar metabolic capabilities). In contrast, in human gut microbiomes, spatially complexity[28,29] and fluctuating environmental conditions[21] 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[30–33] will enhance our knowledge of relationship among ecosystem properties, functional redundancy, and community stability.

In this study, we included whole metagenomic datasets of the examined microbes due to the lack of *a priori* insights into the metabolic pathways/processes playing essential roles in interspecific competition for resources. In this respect, our analysis is a preliminary conceptual step for evaluating potential overlap of fundamental niches at the community level. In future studies, analyses excluding housekeeping genes[34,35] or those focusing on specific functional groups of genes (e.g., carbohydrate degrading genes[36]) may provide more

228 reliable inference of niche overlap. Because such selection of genes can critically influence
229 threshold niche-overlap values for anticipating abrupt community compositional changes,
230 setting a commonly applicable criterion of choosing target gene sets will help us perform
231 comparative analyses across a wide range of microbial communities.

232 While genomic information provides an ultimate platform for inferring fundamental
233 niches[12–14], overlap of gene repertoires may not always result in competitive exclusion of
234 species within communities. Even in a pair of species with similar gene repertoires,
235 differentiation in gene expression patterns may occur to avoid overlap of resource-use
236 patterns between species, allowing coexistence of the two species in an environment. Such
237 differentiation of "realized niches[10]" through phenotypic plasticity is potentially evaluated by
238 transcriptomic or metabolomic analyses[37,38]. Consequently, integration of (meta)transcriptome
239 and (meta)metabolome analyses[39–41] with metagenome-based analyses will reorganize our
240 understanding of deterministic processes in microbiome dynamics.

241

242

Methods

Time-series data of experimental microbiomes. We used the experimental system of the microbiome time-series monitoring described in a previous study[19]. In the experiment, microbiomes differing in the magnitude of community compositional shifts were constructed across the six treatments defined by the combinations of two inoculum sources and three types of media. One of the source microbiomes derived from the soil collected from the A layer (0-10 cm in depth) in the research forest of Center for Ecological Research, Kyoto University, Kyoto, Japan (34.972 ºN; 135.958 ºE). The other source inoculum was prepared by collecting water from a pond ("Shoubuike") near Center for Ecological Research (34.974 ºN, 135.966 ºE). Each of the source inocula was introduced into oatmeal (Medium-A), oatmeal-peptone (Medium-B), or peptone (Medium-C) broth media with eight replicates. Thus, in total, 48 experimental microcosms (two source microbiomes × three media × eight replicates) were constructed in a deep-well plate (1000-μL-scale culture in each well). The plate was kept shaken at 1,000 rpm at 23 ºC. After five-day pre-incubation, 200 μL out of the 1,000-μL culture medium was sampled from each well every 24 hours for 110 days. In each sampling event, 200 μL of fresh medium was added to each well so that the total culture volume was kept constant. In total, 5,280 samples (48 communities/day × 110 days) were collected through the time-series experiment. After DNA extraction, the samples were subjected to the amplicon sequencing analysis of the 16S rRNA region[19]. To quantify the speed and magnitude of community shifts through time, the "abruptness" index was calculated through the time-series of each replicate microcosm in each experimental treatment[19]. Specifically, an estimate of the abruptness index for time point t was obtained as the Bray-Curtis β-diversity between average community compositions from time points $t - 4$ to t and those from $t + 1$ to $t + 5$ (i.e., dissimilarity between 5-day time-windows). The Bray-Curtis β-diversity[42] 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 abundance of microbial amplicon sequence variant (ASV) i in the compared time windows (j, from $t - 4$ to t; k, from $t + 1$ to $t + 5$). An abruptness score larger than 0.5 indicates that turnover of more than 50 % of community compositions occurred between the time-windows[19].

Whole-genome shotgun metagenomics. Focusing on a replicate microcosm in which the most rapid and substantial turnover of community compositions was observed (replicate no. 5

275 of Water/Medium-A treatment; Supplementary Fig. 1), whole-genome shotgun metagenomics
276 was conducted by targeting 13 samples (Day 1, 10, 20, 24, 30, 40, 50, 60, 70, 80, 90, 100,
277 110). Each DNA sample was processed with Nextera XT DNA Library Preparation Kit
278 (Illumina) and sequenced with the DNBSEQ-G400 (BGI; 200-bp paired-end sequencing).
279 From the output data, sequencing adaptors were removed using Cutadapt[43] 2.5 and quality
280 filtering was performed with Fastp[44] 0.21.0: ca. 10 Gb/sample was subjected to the analysis
281 [in total, 159.96 Gb (1000.301 M reads)].

282 The sequences of each sample were assembled with metaSPAdes[45] 3.15.2. Binning was
283 then performed with MetaWRAP[46] 1.3.2, followed by quality assessing with CheckM[47] 1.1.3.
284 Only the MAGs with > 80 % completeness and < 5 % contamination were used in the
285 downstream analyses. The identity between MAGs were calculated using FastANI[48] 1.33 and
286 MAGs with ≥ 98 % identity were dereplicated through the time-series (Supplementary Table
287 1). Read-coverage was then calculated with CoverM[49] 0.6.0, followed by taxonomic
288 annotation was performed using GTDB-Tk[50,51] 1.6. Gene annotation was performed with
289 Prokka[26] 1.14.6, yielding 6,999 annotated genes (Supplementary Data 1). To conduct
290 additional functional annotation of genes, the orthology numbers of Kyoto Encyclopedia of
291 Genomes (KEGG) were retrieved using GhostKOALA[52] 2.2. For respective microbial MAGs
292 (bins), completeness of metabolic pathways was estimated with KEGG decoder[53] 1.3. Based
293 on the matrix representing KEGG metabolic pathway/process profiles of respective MAGs
294 (Supplementary Data 2), a heatmap showing pathway/process completeness was drawn
295 (Supplementary Fig. 3).

296

297 **Background environmental conditions.** For the 13 samples subjected to the shotgun
298 metagenomic analysis, concentrations of ammonium (NH_4^+) and nitrate (NO_3^-) were
299 measured to obtain supplementary information of background environmental conditions.
300 Colorimetric methods with a modified indophenol reaction[54,55] and the VCl3/Griess assay
301 were applied for the measurements of NH_4^+ and NO_3^-, respectively. Samples were run in
302 triplicates via a standard addition method to account for individual matrix effects[56].

303

304 **Multivariate analysis of the metagenomic space.** Based on the whole matrix representing
305 the profiles of the 6,999 genes (Supplementary Data 1), the Jaccard metric of β-diversity was
306 calculated for each pair of the 32 microbial MAGs (β_{ij}, where i and j represent MAGs). The

307 β-diversity estimates were then used to perform a principal coordinate analysis. Using the
308 obtained principal coordinate scores, all the microbial MAGs detected through the time-series
309 were plotted on a multivariate space consisting of the first three PCoA axes (PCoA 1, PCoA 2,
310 and PCoA 3). At each time point, the MAGs detected with the shotgun metagenomic
311 sequencing (defined as the MAGs whose relative abundance is greater than 0.1 %) was
312 plotted on the three-dimensional space defined with the PCoA axes.

313

314 **Evaluation of niche overlap level.** The community-scale magnitude of potential niche
315 overlap among species was evaluated based on the whole-genome shogun sequencing dataset.
316 Specifically, the niche overlap index was defined as:

317 $$1 - \frac{\sum_{i \in D, j \in D} \beta_{ij}}{N_D},$$

318 where D is the set of MAGs detected on a focal day (relative abundance > 0.1 %), β_{ij} is the
319 Jaccard metric of dissimilarity in gene compositions (defined in the previous section), and N_D
320 is the number of MAGs detected on the day.

321 The scores of the niche overlap index were shown on a two-dimensional space
322 representing metabolic pathway/process compositions of the whole community at respective
323 time points. At each time point, the gene repertoires of the detected MAGs (the MAGs whose
324 relative abundance is greater than 0.1 %) were summed, yielding a matrix representing 13
325 time points × 6,999 genes. The matrix was used to calculate dissimilarity (Jaccard β-diversity)
326 in microbiome-scale gene repertoires among time points to perform a PCoA analysis.

327 To test whether a high level of fundamental-niche overlap is observed prior to drastic
328 changes in microbial community structure, we examined relationship between the above niche
329 overlap index and time-series shifts in community structure (Bray-Curtis β-diversity between
330 present and next time points through the time-series of the shotgun metagenomic data).

331

332 Data availability

333 The 16S rRNA sequencing data reported in a previous study[19] are available from the DNA
334 Data Bank of Japan (DDBJ) with the accession number DRA013352, DRA013353,
335 DRA013354, DRA013355, DRA013356, DRA013368 and DRA013379. The whole-genome

shotgun metagenomics data 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 Supplementary Data 1 and 2.

Code availability

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].

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.

Additional information

Supplementary information is available for this paper at [URL to be supplied by the publisher]

Correspondence and requests for materials should be addressed to fujita.h@ecology.kyoto-u.ac.jp or toju.hirokazu.4c@kyoto-u.ac.jp.

References

1. Gause, G. F. *The Struggle for Coexistence*. (Williams & Wilkins, 1934).

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

3. Volterra, V. Variations and fluctuations of the number of individuals in animal species living together. *ICES Journal of Marine Science* **3**, 3–51 (1928).

4. Zaret, T. M. & Rand, A. S. Competition in tropical stream fishes: support for the competitive exclusion principle. *Ecology* **52**, 336–342 (1971).

5. Mayfield, M. M. & Levine, J. M. Opposing effects of competitive exclusion on the phylogenetic structure of communities. *Ecol Lett* **13**, 1085–1093 (2010).

6. Grime, J. P. Competitive exclusion in herbaceous vegetation. *Nature* **242**, 344–347 (1973).

7. Letten, A. D., Ke, P. J. & Fukami, T. Linking modern coexistence theory and contemporary niche theory. *Ecol Monogr* **87**, 161–177 (2017).

8. Chesson, P. Updates on mechanisms of maintenance of species diversity. *Journal of Ecology* **106**, 1773–1794 (2018).

9. Chesson, P. Mechanisms of maintenance of species diversity. *Annu Rev Ecol Syst* **31**, 343–366 (2000).

10. Chase, J. M. & Leibold, M. A. *Ecological niches: linking classical and contemporary approaches*. *Biodiversity and Conservation* (University of Chicago Press, 2004). doi:10.1023/b:bioc.0000029366.24837.fc.

11. Hutchinson, G. E. Concluding Remarks. *Cold Spring Harb Symp Quant Biol* **22**, 415–427 (1957).

12. Régimbeau, A. *et al.* Contribution of genome-scale metabolic modelling to niche

391 theory. *Ecol Lett* **25**, 1352–1364 (2022).

392 13. Smith, S. R. *et al.* Evolution and regulation of nitrogen flux through
393 compartmentalized metabolic networks in a marine diatom. *Nat Commun* **10**, 4552
394 (2019).

395 14. Palomo, A. *et al.* Comparative genomics sheds light on niche differentiation and the
396 evolutionary history of comammox *Nitrospira*. *ISME Journal* **12**, 1779–1793 (2018).

397 15. Alneberg, J. *et al.* Ecosystem-wide metagenomic binning enables prediction of
398 ecological niches from genomes. *Commun Biol* **3**, 119 (2020).

399 16. Fahimipour, A. K. & Gross, T. Mapping the bacterial metabolic niche space. *Nat*
400 *Commun* **11**, 4887 (2020).

401 17. Carding, S., Verbeke, K., Vipond, D. T., Corfe, B. M. & Owen, L. J. Dysbiosis of the
402 gut microbiota in disease. *Microb Ecol Health Dis* **26**, 26191 (2015).

403 18. Ravel, J. *et al.* Daily temporal dynamics of vaginal microbiota before, during and after
404 episodes of bacterial vaginosis. *Microbiome* **1**, 29 (2013).

405 19. Fujita, H. *et al.* Alternative stable states, nonlinear behavior, and predictability of
406 microbiome dynamics. *bioRxiv* https://doi.org/10.1101/2022.08.23.505041 (2022).

407 20. Lahti, L., Salojärvi, J., Salonen, A., Scheffer, M. & de Vos, W. M. Tipping elements in
408 the human intestinal ecosystem. *Nat Commun* **5**, 4344 (2014).

409 21. David, L. A. *et al.* Diet rapidly and reproducibly alters the human gut microbiome.
410 *Nature* **505**, 559–563 (2014).

411 22. Kho, Z. Y. & Lal, S. K. The human gut microbiome - A potential controller of wellness
412 and disease. *Front Microbiol* **9**, 1835 (2018).

413 23. Kriss, M., Hazleton, K. Z., Nusbacher, N. M., Martin, C. G. & Lozupone, C. A. Low
414 diversity gut microbiota dysbiosis: drivers, functional implications and recovery. *Curr*
415 *Opin Microbiol* **44**, 34–40 (2018).

416 24. Costello, E. K., Stagaman, K., Dethlefsen, L., Bohannan, B. J. M. & Relman, D. A. The
417 application of ecological theory toward an understanding of the human microbiome.
418 *Science (1979)* **336**, 1255–1262 (2012).

419 25. Huttenhower, C. *et al.* Structure, function and diversity of the healthy human
420 microbiome. *Nature* **486**, 207–214 (2012).

421 26. Seemann, T. Prokka: Rapid prokaryotic genome annotation. *Bioinformatics* **30**, 2068–
422 2069 (2014).

423 27. Macarthur, R. & Levins, R. The limiting similarity, convergence, and divergence of
424 coexisting species. *American Naturalist* **101**, 377–385 (1967).

425 28. Tropini, C., Earle, K. A., Huang, K. C. & Sonnenburg, J. L. The Gut Microbiome:
426 Connecting Spatial Organization to Function. *Cell Host Microbe* **21**, 433–442 (2017).

427 29. Earle, K. A. *et al.* Quantitative Imaging of Gut Microbiota Spatial Organization. *Cell*
428 *Host Microbe* **18**, 478–488 (2015).

429 30. Jansson, J. K. & Hofmockel, K. S. Soil microbiomes and climate change. *Nat Rev*
430 *Microbiol* **18**, 35–46 (2020).

431 31. Fierer, N. Embracing the unknown: Disentangling the complexities of the soil
432 microbiome. *Nat Rev Microbiol* **15**, 579–590 (2017).

433 32. Trivedi, P., Leach, J. E., Tringe, S. G., Sa, T. & Singh, B. K. Plant–microbiome
434 interactions: from community assembly to plant health. *Nat Rev Microbiol* **18**, 607–621
435 (2020).

436 33. Venter, J. C. *et al.* Environmental Genome Shotgun Sequencing of the Sargasso Sea.
437 *Science (1979)* **304**, 66–74 (2004).

438 34. Gibson, D. G. *et al.* Creation of a bacterial cell controlled by a chemically synthesized
439 genome. *Science* **329**, 52–6 (2010).

440 35. Maiden, M. C. J. *et al.* MLST revisited: The gene-by-gene approach to bacterial
441 genomics. *Nat Rev Microbiol* **11**, 728–736 (2013).

442 36. Flint, H. J., Scott, K. P., Duncan, S. H., Louis, P. & Forano, E. Microbial degradation of
443 complex carbohydrates in the gut. *Gut Microbes* **3**, 289–306 (2012).

444 37. Nowinski, B. & Moran, M. A. Niche dimensions of a marine bacterium are identified
445 using invasion studies in coastal seawater. *Nat Microbiol* **6**, 524–532 (2021).

446 38. Pereira, F. C. & Berry, D. Microbial nutrient niches in the gut. *Environ Microbiol* **19**,

1366–1378 (2017).

39. Heintz-Buschart, A. & Wilmes, P. Human gut microbiome: function matters. *Trends Microbiol* **26**, 563–574 (2018).

40. Schirmer, M. *et al.* Dynamics of metatranscription in the inflammatory bowel disease gut microbiome. *Nat Microbiol* **3**, 337–346 (2018).

41. Turner, T. R. *et al.* Comparative metatranscriptomics reveals kingdom level changes in the rhizosphere microbiome of plants. *ISME Journal* **7**, 2248–2258 (2013).

42. Legendre, P. & de Cáceres, M. Beta diversity as the variance of community data: Dissimilarity coefficients and partitioning. *Ecol Lett* **16**, 951–963 (2013).

43. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J* **17**, https://doi.org/10.14806/ej.17.1.200 (2011).

44. Chen, S., Zhou, Y., Chen, Y. & Gu, J. Fastp: An ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**, i884–i890 (2018).

45. Bankevich, A. *et al.* SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology* **19**, 455–477 (2012).

46. Uritskiy, G. v., DiRuggiero, J. & Taylor, J. MetaWRAP—a flexible pipeline for genome-resolved metagenomic data analysis. *Microbiome* **6**, 158 (2018).

47. Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P. & Tyson, G. W. CheckM: Assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res* **25**, 1043–1055 (2015).

48. Jain, C., Rodriguez-R, L. M., Phillippy, A. M., Konstantinidis, K. T. & Aluru, S. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nat Commun* **9**, 5114 (2018).

49. Woodcroft B. CoverM: program available at https://github.com/wwood/CoverM. Preprint at (2021).

50. Chaumeil, P. A., Mussig, A. J., Hugenholtz, P. & Parks, D. H. GTDB-Tk: A toolkit to classify genomes with the genome taxonomy database. *Bioinformatics* **36**, 1925–1927 (2020).

51. Parks, D. H. *et al.* GTDB: an ongoing census of bacterial and archaeal diversity through a phylogenetically consistent, rank normalized and complete genome-based taxonomy. *Nucleic Acids Res* **50**, D785–D794 (2022).

52. Kanehisa, M., Sato, Y. & Morishima, K. BlastKOALA and GhostKOALA: KEGG Tools for Functional Characterization of Genome and Metagenome Sequences. *J Mol Biol* **428**, 726–731 (2016).

53. Graham, E. D., Heidelberg, J. F. & Tully, B. J. Potential for primary productivity in a globally-distributed bacterial phototroph. *ISME Journal* **12**, 1861–1866 (2018).

54. Kandeler, E. & Gerber, H. Short-term assay of soil urease activity using colorimetric determination of ammonium. *Biol Fertil Soils* **6**, 68–72 (1988).

55. Hood-Nowotny, R., Umana, N. H.-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* **74**, 1018–1027 (2010).

56. Taylor, B. W. *et al.* Improving the fluorometric ammonium method: Matrix effects, background fluorescence, and standard additions. *J North Am Benthol Soc* **26**, 167–177 (2007).

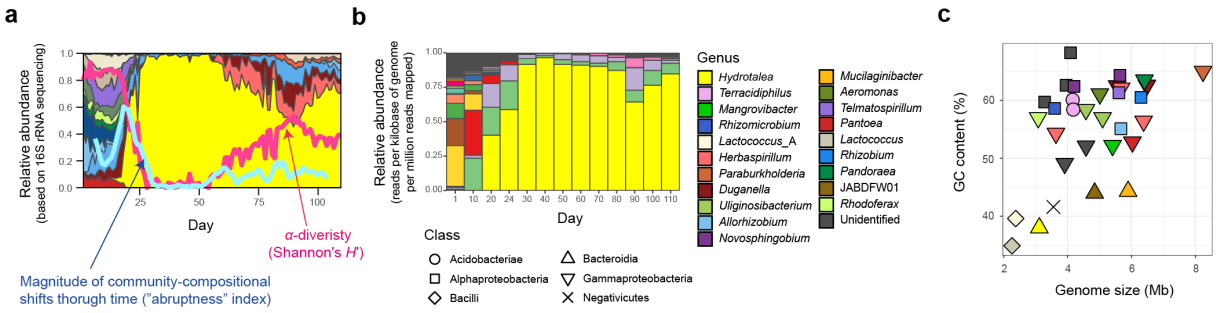

Fig. 1 | Community and ecosystem dynamics. a, Time-series data of community structure. For the replicate microcosm that showed the most abrupt community compositional changes through the 110-day microbiome experiment[19] (Supplementary Fig. 1), family-level taxonomic compositions inferred with 16S rRNA sequencing are shown. The blue line represents the speed and magnitude of community compositional changes around each time point ("abruptness" index[19]; see 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 Supplementary Figure 1 for color profiles of bacterial families. Reproduced from the data of a previous study[19]. **b**, Taxonomic compositions inferred with whole-genome shotgun 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 **b** for colors and symbols.

508

509 **Fig. 2 | Metabolic pathway/process profiles of the MAGs.** KEGG metabolic
510 pathways/profiles of the reconstructed bacterial genomes (MAGs) are shown. The detection
511 (relative abundance > 0.1 %) of each microbial MAG on each day within the whole-genome
512 shotgun data is indicated in the panel below. Only the microbial MAGs with > 80 %
513 completeness and < 5 % contamination were included (Supplementary Table 1). The five
514 MAGs that co-occurred from Day 40 to 60 and metabolic pathways/processes mentioned in
515 the main text are highlighted. Only the metabolic pathways/processes with highly
516 heterogeneous patterns across microbial MAGs are shown. See Supplementary Figure 3 for

517 detailed profiles of the metabolic pathways/processes.

518

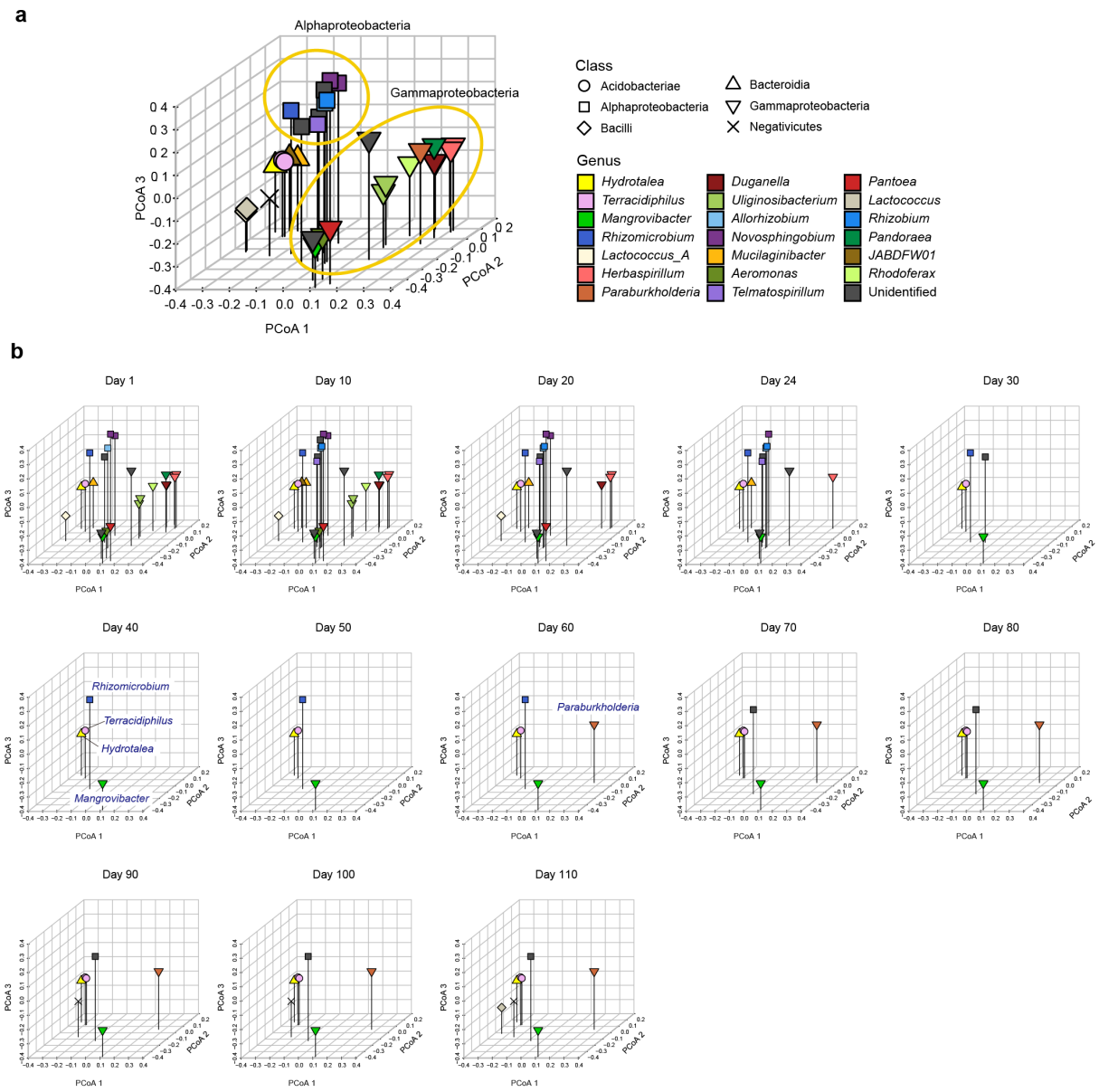

Fig. 3 | Metagenomic niche space. **a**, Distributions of MAGs within metagenomic niche space. Based on dissimilarity in gene repertoires, microbial MAGs that appeared in the time-series of the target microcosm were plotted on the three-dimensional space defined by the principal coordinate analysis (PCoA) 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**.

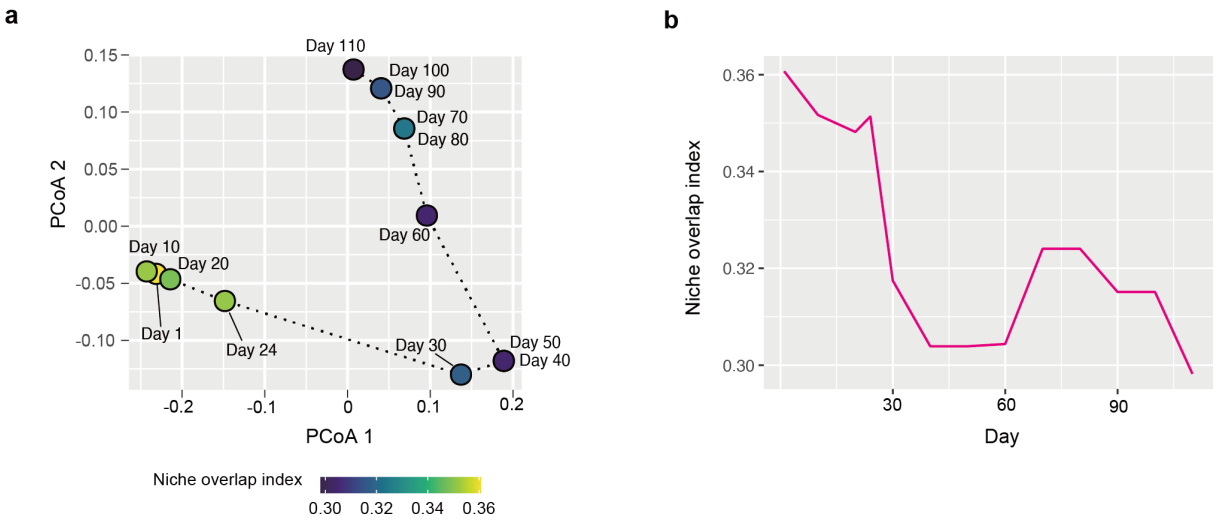

Fig. 4 | Dynamics of niche-overlap level. **a**, Community-level profiles of metabolic pathways/processes and niche overlap index. The niche overlap index was defined as $1 - \bar{\beta}$, where $\bar{\beta}$ was mean Jaccard dissimilarity (β-diversity) of gene compositions between pairs of the microbial MAGs detected at a target time point. The scores of the niche overlap index were 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.

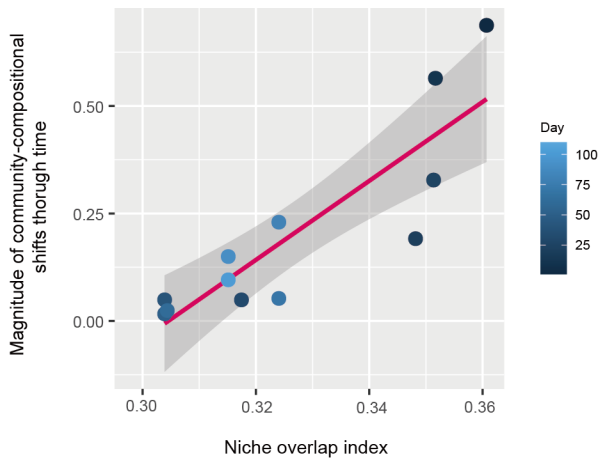

Fig. 5 | Niche overlap level and community compositional shifts. The magnitude of community compositional changes observed in the microbiome was regressed on niche overlap index obtained based on the whole-genome shotgun analysis. Niche overlap index at each time point and time-series shifts in community 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 line is shown with 95 % confidence intervals.

Supplementary Figure Captions

Supplementary Fig. 1 | Dynamics of family-level community structure. The dynamics of microbial family-level compositions were visualized based on the 16S rRNA sequencing data of the previous study[19]. The replicate microcosm (replicate no. 5 in Water/Medium-A treatment), which is subjected to the whole-genome shotgun sequencing analysis, is highlighted. The blue line represents the speed and magnitude of community compositional changes around each time point ("abruptness" index[19]; see Methods). The red line indicates α-diversity (Shannon's *H'*) of microbial ASVs[19]. Reproduced from the data of a previous study[19].

Supplementary Fig. 2 | Overview of the whole-genome shotgun sequencing data. a, Comparison of relative abundance of bacterial taxa (families) between 16S rRNA amplicon sequencing[19] (reproduced from the previous study[19]; top) and whole-genome shotgun sequencing (this study; bottom). **b**, Correlation between the family-level relative abundance of 16S rRNA and whole-genome shotgun sequencing data (Spearman's correlation; $\rho = 0.667$, df = 794, $P < 0.05$). Each point represents each family at each time point. **c**, Background chemical properties. Changes in NO_3^- and NH_4^+ concentrations in the ecosystem are shown for the time points with whole-genome shotgun metagenomic data.

Supplementary Fig. 3 | Detailed information of the metabolic pathway/process profiles of the MAGs. The KEGG metabolic pathways/processes of the reconstructed bacterial genomes (MAGs) are shown. The detection (relative abundance > 0.1 %) of each microbial MAG on each day within the whole-genome shotgun data is indicated in the panel below. Only the microbial MAGs with > 80 % completeness and < 5 % contamination were included (Supplementary Table 1). The five MAGs that co-occurred from Day 40 to 60 and metabolic pathways/profiles mentioned in the main text are highlighted. The detailed definition of the KEGG metabolic pathways/processes is available at https://github.com/bjtully/BioData/blob/master/KEGGDecoder/KOALA_definitions.txt.