bioRxiv preprint doi: https://doi.org/10.1101/2022.08.23.505048; this version posted April 3, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Basins of attraction of microbiome structure and soil ecosystem functions

Hiroaki Fujita[1], Shigenobu Yoshida[2], Kenta Suzuki[3,4] and Hirokazu Toju[1,5,6†]

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

[2]Institute for Plant Protection, National Agriculture and Food Research Organization, Tsukuba, Ibaraki 305-8666, Japan

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

[4]Institute of Multidisciplinary Sciences, Yokohama National University. Yokohama, Kanagawa 240-8501, Japan.

[5]Laboratory of Ecosystems and Coevolution, Graduate School of Biostudies, Kyoto University, Kyoto 606-8501, Japan

[6]Center for Living Systems Information Science (CeLiSIS), Graduate School of Biostudies, Kyoto University, Kyoto 606-8501, Japan

For correspondence:

toju.hirokazu.4c@kyoto-u.ac.jp

Competing interest:

HT is the founder and director of Sunlit Seedlings Ltd. 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.

bioRxiv preprint doi: https://doi.org/10.1101/2022.08.23.505048; this version posted April 3, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Funding

This work was financially supported by JST PRESTO (JPMJPR16Q6), JST FOREST (JPMJFR2048), Human Frontier Science Program (RGP0029/2019), JSPS Grant-in-Aid for Scientific Research (20K20586), NEDO Moonshot Research and Development Program (JPNP18016), and JST CREST (JPMJCR23N5) to H.T., JSPS Grant-in-Aid for Scientific Research (20K06820 and 20H03010) to K.S., and JSPS Fellowship to H.F.

bioRxiv preprint doi: https://doi.org/10.1101/2022.08.23.505048; this version posted April 3, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Abstract

Theory predicts that biological communities can have multiple basins of attraction in terms of their species/taxonomic compositions. The presence of such basins of community structure has been examined in classic empirical studies on forest–savanna transitions and those on eutrophication in freshwater lakes. Nonetheless, it remains a major challenge to extend the investigations of multistability to species-rich microbial communities. By targeting soil microbiomes, we infer the stability landscapes of community structure based on the concepts of statistical physics. Our analysis on the compiled dataset involving 11 archaeal, 332 bacterial, and 240 fungal families detected from > 1,500 agroecosystem soil samples suggested that both prokaryotic and fungal community compositions could be classified into several basins of attraction. We also found that the basins differed greatly in their associations with crop disease prevalence in agroecosystems. A further analysis highlighted microbial taxa potentially playing key roles in transitions between basins with different ecosystem-scale functions. The statistical framework commonly applicable to diverse microbial and non-microbial communities will reorganize our understanding of relationship among community structure, stability, and functions.

bioRxiv preprint doi: https://doi.org/10.1101/2022.08.23.505048; this version posted April 3, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Introduction

The idea that the state space of biological community structure can comprise multiple basins of attraction have inspired both empirical and theoretical ecologists since the late 1960s (Beisner et al., 2003; May, 1977; Scheffer et al., 1993). The concept of multistability has been examined in aquatic and terrestrial ecosystems (Scheffer and Carpenter, 2003; Schröder et al., 2005; Suding and Hobbs, 2009). Community structure in shallow lakes, for example, is known to show two discrete states depending on nutrient (phosphorus) concentrations as represented by the bistability of charophyte densities (Ibelings et al., 2007; Scheffer et al., 1993; Smith and Schindler, 2009). Likewise, worldwide inventories of tree cover have shown that forest and savanna vegetation types possibly represent different basins of attraction (Hirota et al., 2011; Staver et al., 2011b, 2011a). Those studies have further shown that ecosystem functions (e.g., fishery, agricultural, and forestry production) can differ greatly between such basins of biological community structure (Gunderson, 2000; Scheffer et al., 2001, 1993; Scheffer and Carpenter, 2003). Consequently, understanding how structure, stability, and biological functions are organized in real communities and ecosystems has been one of the major goals in ecology.

While classic studies targeting freshwater and terrestrial biomes have explored basins of attraction based on simple characterization of community states (e.g., tree cover percentages), recent technical advances in microbial community (microbiome) research have come to provide opportunities for deepening our knowledge of biological community stability (Amor et al., 2020; Costea et al., 2017; Faust et al., 2015; Shaw et al., 2019; Toju et al., 2018; Zaneveld et al., 2017). Based on amplicon and shotgun sequencing technologies, large datasets of microbial species/taxonomic compositions have been made available, providing a basis for exploring reproducible states in microbiome community structure (Amor et al., 2020; H Fujita et al., 2023; Hayashi et al., 2024). Such high-throughput DNA sequencing studies in medicine, for example, have shown that human individuals can be classified into three or four semi-discrete clusters in terms of their intestinal microbiome compositions (Arumugam et al., 2011; Wu et al., 2011) [see also (Jeffery et al., 2012; Knights et al., 2014)]. Intriguingly, these alternative gut microbiomes ("enterotypes") differ in their associations with human disease such as type II diabetes and Crohn's disease (Costea et al., 2017). In addition to those studies on animal-associated microbiomes (Arumugam et al., 2011; Moeller et al., 2012; Yajima et al., 2023), studies on plant-associated microbiomes have started to reorganize our recognition of how multistability of phyllosphere/rhizosphere microbiome structure is associated with ecosystem-scale processes and

bioRxiv preprint doi: https://doi.org/10.1101/2022.08.23.505048; this version posted April 3, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

functions (Toju et al., 2018, 2016). Because hundreds or thousands of replicate community samples are available in such microbiome studies, it is now possible to discuss potential relationship between community structure and ecosystem functions based on statistical signs of the presence of multiple basins (and background attractors).

In theoretical ecology, stability of community states (taxonomic or species compositions) is often discussed in the framework of stability landscapes (Beisner et al., 2003; Hastings et al., 2018; Lewontin, 1969; Scheffer and Carpenter, 2003; Suzuki et al., 2021). On the landscape representing stability/instability of community structure, basins of attraction are split by "tipping points" representing unstable equilibria (Beisner et al., 2003; Scheffer et al., 2001; Scheffer and Carpenter, 2003; Suzuki et al., 2021) (Figure 1). As these basins differ in the biological functions of constituent communities, stable and highly functional community states can be explored within the stability landscapes. With the application of a recently proposed mathematical approach developed based on statistical physics (Becker and Karplus, 1997; Watanabe et al., 2014), it is now possible to infer "energy landscapes", which represent structure of stability landscapes, from empirical datasets of ecological communities (Dakos and Kéfi, 2022; Sánchez-Pinillos et al., 2024; Suzuki et al., 2021). The statistical framework allows us to explore the probabilities of community compositions within the "assembly graphs" (Coyte et al., 2021; Serván and Allesina, 2021) representing paths of possible community assembly (H Fujita et al., 2023; Suzuki et al., 2021) (Figure 1). Although hundreds or thousands of community compositional data points are required to apply the statistical approach (H Fujita et al., 2023; Suzuki et al., 2021), such energy landscape analyses will allow us to define key features of stable and highly functional microbiome states out of numerous possible combinations of microbial species or taxa. Despite the potential for systematically profiling the relationship among community structure, stability, and functions based on massive community datasets, the energy landscape analysis has been applied only to a few microbial community datasets (H Fujita et al., 2023; Suzuki et al., 2021).

We here apply the emerging statistical framework to soil microbiomes, which often show highest levels of structural diversity in nature. We compile a cropland soil microbiome dataset consisting of > 1,500 sampling positions across the Japan Archipelago (Fujita et al., 2024). With the massive dataset, we infer the compositional stability of prokaryotic and fungal communities based on maximum entropy models of the energy landscape analysis (Suzuki et al., 2021). We then examine whether the basins of attraction of soil microbiomes can differ in ecosystem-scale functions by focusing on potential relationship between soil microbial compositions and the

bioRxiv preprint doi: https://doi.org/10.1101/2022.08.23.505048; this version posted April 3, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

prevalence of crop plant disease. We also explore key microbial taxa whose abundance critically divide the basins representing favorable and unfavorable ecosystem functions. The results of the energy landscape analysis are further used to infer tipping points splitting the inferred basins. Overall, this study illustrates how we can integrate the information of community structure, stability, and functions based on a statistical platform commonly applicable to diverse microbial and non-microbial communities.

Methods

Dataset compilation

We compiled a publicly available dataset of cropland soil microbiomes (DDBJ accession: DRA015491; Figure 2) with its metadata of the samples (Fujita et al., 2024). In the previous study reporting the data (Fujita et al., 2024), 2,903 bulk soil samples collected from the field of 19 crop plant species (apple, broccoli, cabbage, celery, Chinese cabbage, eggplant, ginger, komatsuna, lettuce, onion, potato, radish, rice, satsuma mandarin, soybean, spinach, strawberry, sweet corn, tomato) across the Japan Archipelago from January 23, 2006 to July 28, 2014 (latitudes of the sampling positions: 26.1–42.8 °N) were subjected to the amplicon sequencing analysis of the prokaryotic 16S rRNA region and the fungal internal transcribed spacer 1 (ITS1) region (Fujita et al., 2024). The information of dry soil pH, electrical conductivity, carbon/nitrogen (C/N) ratio, and available phosphorous concentration was available for 2,830, 2,610, 2,346, and 2,249 samples, respectively. Likewise, the information of crop plant disease [the percentage of diseased plants or disease severity index (Chiang et al., 2017)] was available for 1,471 samples (Fujita et al., 2024). The plant pathogens surveyed were *Colletotrichum gloeosporioides* on the strawberry, *Fusarium oxysporum* on the celery, the lettuce, the strawberry, and the tomato, *Phytophthora sojae* on the soybean, *Plasmodiophora brassicae* on Cruciferae plants, *Pyrenochaeta lycopersici* on the tomato, *Pythium myriotylum* on the ginger, *Ralstonia solanacearum* on the eggplant and the tomato, and *Verticillium* spp. on Chinese cabbage (Fujita et al., 2024). After a series of quality filtering, prokaryotic and fungal community data were available for 2,318 and 2,186 samples, respectively. In total, 579 archaeal amplicon sequence variants (ASVs) representing 11 families, 26,640 bacterial ASVs representing 332 families, and 6,306 fungal ASVs representing 240 families were detected (Fujita et al., 2024) (Figures 2; Figure 2–figure supplement 1).

bioRxiv preprint doi: https://doi.org/10.1101/2022.08.23.505048; this version posted April 3, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

146

147 **Community structure along environmental gradients**

148 We first inspected how prokaryotic and fungal community structure varied along environmental
149 gradients. For each data matrix representing the family-level compositions of prokaryotes or
150 fungi, a principal coordinate analysis (PCoA) was performed based on Bray-Curtis β-diversity.
151 The PCoA1 and PCoA2 scores were then plotted, respectively, along the axes of soil
152 environmental factors. Specifically, the axes of the environmental factors were defined based on
153 a principal component analysis (PCA) of soil pH, electrical conductivity, C/N ratio, and available
154 phosphorous concentration. In total, 1,771 and 1,664 samples for which the information of both
155 community structure and all the four environmental variables was available were included in the
156 analyses of prokaryotes and fungi, respectively. For each plot representing relationship between
157 environmental conditions and community structure, the density of data points was visualized with
158 the ggplot2 3.3.6 package (Wickham, 2011) of R v.4.1.2 (R Core Team, 2020).

159

160 **Energy landscape analysis**

161 We examined the stability landscape of soil microbiome structure based on the framework of an
162 energy landscape analysis (H Fujita et al., 2023; Suzuki et al., 2021; Watanabe et al., 2014)
163 (tutorials of energy landscape analyses are available at https://github.com/kecosz/rELA). In the
164 framework, the term "energy" is defined by the following equations based on statistical physics
165 (Suzuki et al., 2021; Watanabe et al., 2014). Within the "assembly graphs" representing paths of
166 community dynamics (Coyte et al., 2021; Serván and Allesina, 2021), probabilities of observing
167 specific community compositions can be explored as detailed previously (Suzuki et al., 2021). In
168 brief, probabilities of community states $\mathrm{p}(\vec{\sigma}^{(k)})$ are given by

169 $$P(\vec{\sigma}^{(k)}|\varepsilon) = e^{-E(\vec{\sigma}^{(k)},\varepsilon)}/Z \qquad \text{[eq. 1]}$$

170 $$Z = \sum_{i=1}^{2^S} e^{-E(\vec{\sigma}^{(i)},\varepsilon)} \qquad \text{[eq. 2]},$$

171 where $\vec{\sigma}^{(k)} = (\sigma_1^{(k)}, \sigma_2^{(k)}, \ldots, \sigma_S^{(k)})$ is a community state vector of k-th community state and S is
172 the total number of taxa (e.g., ASVs, species, genera, or families) examined. $\varepsilon = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_M)$
173 is an array of continuous values representing environmental factors (e.g., soil pH and electrical
174 conductivity) and M is the total number of environmental parameters. $\sigma_{\mathrm{i}}^{(k)}$ is a binary variable

bioRxiv preprint doi: https://doi.org/10.1101/2022.08.23.505048; this version posted April 3, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

that represents presence (1) or absence (0) of taxon *i*: i.e., there are a total of 2^S community states. As the exploration of the 2^S community states were computationally intensive, we coded community states at the family-level taxonomic compositions. Specifically, for each sample, families whose relative abundance exceeds a certain threshold value (threshold for binarization) were coded as 1, while the remaining minor families were coded as 0. Subsequently, families whose occurrence ratios (i.e., the proportions of samples in which target families were coded as 1) were less than a certain threshold (occurrence threshold) were excluded from the dataset. Likewise, families that appeared in almost all samples (1 – occurrence threshold) were excluded. Note that without such thinning of input data, the dimensions of community states are too high to be explored even using supercomputers. Therefore, exclusion of the taxa that contribute little to the classification of community states (i.e., taxa appearing only in a small fraction of samples or those appearing in most samples) is inevitable in the energy landscape analysis. Through intensive preliminary computational runs with various combinations of binarization and occurrence thresholds, we found that the number of taxa (*S*) should be kept less than 65 as detailed in the next subsection.

When input community matrix is defined, the energy of the community state $\vec{\sigma}^{(k)}$ is given by the extended pairwise maximum entropy model:

$$E(\vec{\sigma}^{(k)}, \varepsilon) = -\sum_{i=1}^{S} h_i \vec{\sigma}_i^{(k)} - \sum_{j=1}^{S} \sum_{i=1}^{M} g_{ij}\, \varepsilon_i^{(k)} \sigma_j^{(k)} - \sum_{i=1}^{S} \sum_{j=1}^{S} J_{ij} \vec{\sigma}_i^{(k)}\, \vec{\sigma}_j^{(k)}/2 \qquad \text{[eq. 3]},$$

where h_i represents the net effect of implicit abiotic factors, by which *i*-th taxon is more likely to present ($h_i > 0$) or not ($h_i < 0$), g_{ij} represents the effect of the *i*-th observed environmental factor, and J_{ij} represents a co-occurrence pattern of *i*-th and *j*-th taxa. Since the logarithm of the probability of a community state is inversely proportional to $E(\vec{\sigma}^{(k)})$, a community state having lower *E* is more frequently observed. To consider dynamics on an assembly graph defined as a network whose 2^S nodes represent possible community states and the edges represents transition path between them (two community states are adjacent only if they have the opposite presence/absence status for just one species), we assigned energy to nodes with the above equation, and so imposed the directionality in state transitions. Then, by using the steepest descent algorithm (Suzuki et al., 2021), we identified nodes having the lowest energy compared to all its neighbors within the weighted network, and determined their basins of attraction (Lewontin, 1969; Suzuki et al., 2021). These community states whose energy was lower than that of all adjacent community states represent estimated point equilibria (attractors), around which

bioRxiv preprint doi: https://doi.org/10.1101/2022.08.23.505048; this version posted April 3, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

community states are expected to show transient fluctuations due to demographic stochasticity as considered in the statistical framework (H Fujita et al., 2023; Suzuki et al., 2021) (Figure 1). Soil pH, electrical conductivity, C/N ratio, and available phosphorous concentration were included as environmental variables in the model after normalization within the ranges from 0 to 1.

Energy landscape structure

The energy landscapes of community structure were inferred, respectively, for three types of datasets, namely, the prokaryotic community matrix, the fungal matrix, and the matrix including both prokaryotes and fungi. As mentioned above, various combinations of binarization and occurrence thresholds were examined to check the reproducibility of the results. In addition to the energy landscape analysis based on the above-mentioned family-level delineation of community states, analyses based on community-state delineation at the order-level were performed. In the main body and supplementary figures of this study, we show the results at the following settings: prokaryotes (family), binarization = 0.020, occurrence = 0.10; prokaryotes (order), binarization = 0.020, occurrence = 0.10; fungi (family), binarization = 0.001, occurrence = 0.05; fungi (family), binarization = 0.001, occurrence = 0.10; prokaryotes + fungi (family), binarization = 0.030, occurrence = 0.10; prokaryotes + fungi (order), binarization = 0.030, occurrence = 0.10. Note that these thresholds were selected to make the state space (2^S) neither too simplified (e.g., $S < 30$) nor too complex ($S < 65$).

For each setting, the parameters of the extended pairwise maximum entropy model [eq. 3] were adjusted to the empirical data. More precisely, the maximum likelihood estimates of h_i, g_{ij}, and J_{ij} was obtained by a stochastic approximation method as detailed elsewhere (Suzuki et al., 2021). The parameters were regularized by a logistic prior with location 0 and scale 2.0 (for environmental responses) or 0.5 (for pairwise relationships) (Harris, 2016). Hyperparameters for the algorithm, criterion value for judging the convergence of parameters qth = 10^{-5}, were set according to a series of preliminary analyses. Based on the inferred maximum entropy model, we determined basins of attraction (Lewontin, 1969) within the energy landscape based on a steepest descent procedure (Suzuki et al., 2021). The structure of the energy landscape was visualized by showing the energy of each soil sample on the two-dimensional surface of the community state space defined with the abovementioned PCoA scores. The default setting of environmental variables (the mean value for each of soil pH, electrical conductivity, C/N ratio, and available

bioRxiv preprint doi: https://doi.org/10.1101/2022.08.23.505048; this version posted April 3, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

phosphorous concentration) was used in the energy calculation. Spline smoothing of the energy landscape was performed with optimized penalty scores using the mgcv v.1.8-40 package (Wood, 2022) of R. For each analysis of the prokaryote, fungi, and prokaryote + fungi datasets, 1,771, 1,664, and 1,474 samples for which the information of both community structure and all the four environmental variables was available were subjected to the analysis, respectively.

Ecosystem functions and key taxa

For the inferred basins of microbial community compositions, associations with crop disease prevalence were examined. We first constructed the list of soil samples whose community structure was located within each basin of attraction. We then evaluated the ecosystem-scale properties of the basins in light of the metadata of crop disease symptoms (Fujita et al., 2024). Specifically, for each basin, we calculated the proportion of constituent soil samples with the minimal level of crop disease symptoms (the percentage of diseased plants < 20 or disease severity index < 20; (Fujita et al., 2024)). The bottoms of basins representing different levels of crop disease prevalence were then compared in terms of taxonomic compositions in order to explore microbial taxa that were keys to distinguish potentially disease-suppressive and disease-promotive soil ecosystems.

Disconnectivity graphs

For the reconstructed energy landscape, we inferred “disconnectivity graphs” (Suzuki et al., 2021) representing how basins of attraction were split by tipping points (Figure 1A). Within a disconnectivity graph, community states whose energy is much lower than the energy of connected tipping points are expected to be resistant to perturbations (demographic stochasticity). In contrast, community states with small energy gaps to tipping points may be shifted from current basins to adjacent basins with minimal perturbations.

Results

Community structure along environmental gradients

bioRxiv preprint doi: https://doi.org/10.1101/2022.08.23.505048; this version posted April 3, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

On each plot showing community compositions (PCoA1 or PCoA2 scores) along the soil environmental gradient (Figure 3), multiple clusters of data points were observed for both prokaryotes and fungi (Figure 3–figure supplements 1-2). In other words, community states are expected to be classified into some clusters even under equivalent edaphic conditions.

Energy landscape structure

The energy landscape of the family-level prokaryotic data included several major basins differing remarkably in associations with the prevalence of crop plant disease (Figure 4). Specifically, 59.6% of soil samples located within a basin (basin ID = 0IK1G2) were associated with the minimal plant-disease level, while the proportion was only 10.7% for another basin (LQWZ02) (Figure 4C-D). The presence of basins differing greatly in their associations with plant-disease levels was inferred as well at the order-level analysis of the prokaryotic data (Figure 4–figure supplement 1). Such variation in crop disease prevalence among inferred basins was observed also for the family-level analysis of fungal community structure (Figure 5). Specifically, while 57.9% of samples belonging to the basin 7QH9moTf8Xa, but none of the samples belonging to the basin 68C0849W020, were associated with the minimal plant-disease level (Figure 5D). Meanwhile, such difference in associations with disease prevalence was moderate in an analysis in which a smaller number of fungal families were examined to define community states (Figure 5–figure supplement 1). The presence of multiple basins, which differed in associations with crop-disease prevalence, was suggested even when prokaryotic and fungal community data were simultaneously analyzed (Figure 4–figure supplements 2-3).

Ecosystem functions and key taxa

In an analysis of the prokaryotic community structure, 19 families were keys to distinguish basins differing in associations with crop-disease prevalence (Figure 4D). The presence of Pyrinomonadaceae and Vicinamibacteraceae, for example, was unique to the basin with the highest proportion of samples showing the minimal plant-disease level (Figure 4D). Likewise, in an analysis of the fungal community structure, the basin associated closely with the minimal plant-disease prevalence (7QH9moTf8Xa) was defined by the presence of several families such as Basidiobolaceae, Cordycipitaceae, and Gelatinodiscaceae (Figure 5D). The exploration of

bioRxiv preprint doi: https://doi.org/10.1101/2022.08.23.505048; this version posted April 3, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

295 microbial taxa keys to distinguish basins with different ecosystem-level functions can be
296 performed at other taxonomic levels (e.g., order-level; Figures 4–figure supplements 1 and 3).

297

298 **Disconnectivity graphs**

299 Within the energy landscape of the family-level analysis of prokaryotes (Figure 4), both the
300 basins associated with the least-diseased (OIK1G2) and most-diseased (N21H04) crop status
301 were the deepest among the inferred basins (i.e., showing the largest energy gaps from the bottom
302 to tipping points; Figure 6A-B). In the family-level analysis of fungi, the basin associated with
303 the least-diseased status (7QH9moTf8Xa) was the deepest, while the other basin representing the
304 most-diseased status (68C0849W020) was the shallowest (Figure 6C).

305

306 # Discussion

307 We have estimated the stability landscape structure of complex microbiomes based on a
308 statistical framework commonly applicable to diverse types of biological communities. The
309 energy landscape analysis allows systematic analyses of taxon-rich community datasets by
310 incorporating the information of multiple environmental factors (Dakos and Kéfi, 2022; Sánchez-
311 Pinillos et al., 2024; Suzuki et al., 2021). While classic studies on community multistability have
312 discussed ecological processes spanning a few intuitively distinguishable community states
313 [high/low tree cover in forest-savanna transitions (Hirota et al., 2011; Staver et al., 2011a, 2011b)
314 or macrophyte-/phytoplankton-dominated state in shallow lakes (Ibelings et al., 2007; Scheffer
315 and Carpenter, 2003)], it is now made possible to define basins of attraction based on high-
316 dimensional community datasets involving hundreds of species/taxa (Arumugam et al., 2011;
317 Costea et al., 2017; H Fujita et al., 2023; Guim Aguadé-Gorgorió et al., 2023; Hayashi et al.,
318 2024). Application of the general statistical platform will enhance our understanding of how
319 stability landscape properties differ among diverse microbial and non-microbial systems.

320 Despite numerous potential compositions (2^S community states; S is the number of
321 considered species/taxa), the prokaryotic and fungal community states were grouped into small
322 numbers of basins within energy landscapes (Figures 4-5). This result suggests that soil
323 microbiome structure remain within certain regions even after demographic perturbations. In

bioRxiv preprint doi: https://doi.org/10.1101/2022.08.23.505048; this version posted April 3, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

other words, once trapped in a basin of attraction, large shifts in community structure would not occur without perturbations whose strength exceed certain thresholds (Beisner et al., 2003; Lewontin, 1969; May, 1977; Scheffer et al., 1993). Importantly, the threshold strength of perturbations is estimated as the energy gap between bottoms of basins and tipping points (Suzuki et al., 2021) (Figure 6A). Furthermore, potential paths of community structural transitions can be quantitatively inferred as illustrated in disconnectivity graphs (Suzuki et al., 2021) (Figures 6B-C). Such statistical framework of quantitative science will entail novel opportunities for testing theories on biological community processes in the era of massive datasets.

Among potential processes or mechanisms underlying the multistability of community structure, historical contingency is of particular interest (Fukami, 2015). In the local assembly of microbial communities, early colonizers or residents can prevent the settlement of followers by constructing physical barriers (e.g., biofilms and mycelia) (Baümler and Sperandio, 2016; Fukami, 2015; Leopold et al., 2017; Verbruggen et al., 2013; Werner and Kiers, 2015) or emitting antibiotics (Mendes et al., 2013; Raaijmakers et al., 2002). In addition to those antagonistic effects on late colonizers, webs of mutualistic or commensalistic interactions within the microbiomes of early colonizers (Elias and Banin, 2012; Hiroaki Fujita et al., 2023; Zelezniak et al., 2015) would influence community dynamics. Due to such "priority effects" (Fukami, 2015), bacterial and fungal community compositions may persist within limited ranges of community states without substantial perturbations. Given that abilities to form physical or chemical barriers can differ greatly among microbial species/taxa (Mendes et al., 2013; Raaijmakers et al., 2002; Werner and Kiers, 2015), such variation in constituent species' priority effects may underly the observed variation in the depth of basins (Figure 6B-C).

The inference of stability landscape structure provided an opportunity for evaluating relationship between community stability and ecosystem-scale functions. The basins of attraction of prokaryotic/fungal community structure differed considerably in associations with crop disease prevalence (Figure 5), suggesting the presence of "stable and favorable" and "stable but unfavorable" states of microbiomes (Mendes et al., 2011; Schlatter et al., 2017; Yuan et al., 2020) in terms of agricultural productivity. This finding adds an important dimension of discussion on the use of microbes in agriculture. Beyond investigations on single species/strains of microbes, microbiome studies have explored sets of microbes that collectively maximize biological functions (Jansson and Hofmockel, 2018; Toju et al., 2018; Trivedi et al., 2020; Vorholt et al., 2017). In particular, experimental studies on "synthetic" communities have reorganized our

bioRxiv preprint doi: https://doi.org/10.1101/2022.08.23.505048; this version posted April 3, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

356 knowledge of microbiome functions (Jansson and Hofmockel, 2018; Trivedi et al., 2020; Vorholt
357 et al., 2017). Nonetheless, such microbial functions cannot be realized in real agroecosystems if
358 the synthesized or designed microbiome compositions are vulnerable to biotic and abiotic
359 environmental changes in the wild (Mazzola and Freilich, 2017). Thus, in addition to functional
360 properties, compositional stability is the key to manage microbiomes in agroecosystems (Faust
361 and Raes, 2012; Toju et al., 2020; Vorholt et al., 2017).

362 In our analysis across the Japan Archipelago, prokaryotic and fungal taxa keys to
363 distinguish least-diseased and severely-diseased states of soil microbiomes were highlighted
364 (Figures 4-5). Among them, Basidiobolaceae and Cordycipitaceae are of particular interest
365 because they include many species potentially utilized as biological control agents for
366 suppressing pest insects (Meyling and Eilenberg, 2007; Möckel et al., 2022). Gelatinodiscaceae is
367 another fungal taxon playing potentially important roles as symbionts of plants (Johnston et al.,
368 2019). These results illuminate the hypothesis that plant disease could be suppressed under the
369 coexistence of multiple prokaryotic and fungal taxa with favorable ecosystem functions (Toju et
370 al., 2018; Toju and Tanaka, 2019). Thus, statistical analyses of stability landscapes allow the
371 exploration of key species or taxa (Paine, 1966; Power et al., 1996), whose management could
372 result in transitions from unfavorable ecosystem states to favorable ones (Gunderson, 2000;
373 Scheffer et al., 2001, 1993; Scheffer and Carpenter, 2003). Given that most prokaryotic and
374 fungal families highlighted in our analysis have cosmopolitan distributions, a next crucial step is
375 to test whether the basins defined across the Japan Archipelago can be used to categorize disease-
376 suppressive and disease-susceptible microbiomes in other regions on the globe.

377 Although the energy landscape analysis enhances our understanding of community stability
378 and functions, its results should be interpreted with caution. First, given that classic empirical
379 studies examined community multistability with system-specific simple criteria [e.g., high/low
380 tree cover (Hirota et al., 2011; Staver et al., 2011a, 2011b)], special care should be taken when
381 we extend the approach to species-rich (high-dimensional) community datasets (Guim Aguadé-
382 Gorgorió et al., 2023). In other words, unambiguous and broadly applicable criteria based on
383 statistical evaluation are the prerequisite for comparative analyses of community multistability.
384 Although we applied a straightforward statistical definition of basins of attraction (Suzuki et al.,
385 2021) (Figure 1) in light of classic theoretical studies (Beisner et al., 2003; Lewontin, 1969; May,
386 1977; Scheffer et al., 1993), continuous methodological improvements should be explored
387 towards further comprehensive analyses. Second, our analysis on hyper-diverse soil microbiomes

bioRxiv preprint doi: https://doi.org/10.1101/2022.08.23.505048; this version posted April 3, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

388 incurred substantial computational costs, forcing us to limit the energy landscape analysis to
389 family-level input data. Further improvements of codes are necessary for inferring stability
390 landscapes at genus-, species-, or strain-level analyses. Third, it should be acknowledged that
391 detailed discussion on ecological processes require time-series datasets (Davidson et al., 2023;
392 Scheffer et al., 2012, 2009). Because our present data lacked the information of temporal changes
393 in community structure, we are unable to discuss the frequency and pace of community structural
394 transitions between basins of attraction. Monitoring of microbiome compositions (Faust et al.,
395 2015; Hayashi et al., 2024; Yajima et al., 2023) is necessary for filling the gap between
396 theoretical and empirical studies (Long et al., 2024).

397 The energy landscape framework of multistability analysis is readily applicable to a wide
398 range of microbiome datasets. Application to human microbiome data is of particular interest in
399 terms of the confirmation of the existence of multiple basins of attraction (Jeffery et al., 2012). In
400 addition, insights into the key microbial species/taxa that would play key roles in the transitions
401 from disease-associated microbiome states to healthy ones will open new directions of
402 microbiome therapy. Furthermore, time-series analyses of community dynamics on stability
403 landscapes will allow us to forecast and prevent transitions into unfavorable community states
404 [e.g., dysbiosis (Carding et al., 2015; H Fujita et al., 2023; Long et al., 2024)]. Along with such
405 extensions of observational research, experimental studies controlling key species/taxa or
406 environmental parameters (Schröder et al., 2005) will promote both basic and applied sciences of
407 ecosystem functions.

408

409 **Acknowledgements.** We thank the SuperComputer System, Institute for Chemical Research,
410 Kyoto University for the use of super computers.

411

412 **Author contributions**

413 H.F., S.Y. and H.T. designed the work. H.F. performed molecular experiments. H.F. and H.T.
414 analyzed the data. H.T. wrote the paper with H.F., S.Y., and K.S.

415

416 **Author ORCIDs**

bioRxiv preprint doi: https://doi.org/10.1101/2022.08.23.505048; this version posted April 3, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Hiroaki Fujita: https://orcid.org/0009-0003-5422-9775

Shigenobu Yoshida: https://orcid.org/0000-0001-5105-0807

Kenta Suzuki: https://orcid.org/0000-0002-0639-2303

Hirokazu Toju: https://orcid.org/0000-0002-3362-3285

Data accessibility

The accession number of the DDBJ Sequence Read Archive: DRA015491. The microbial community matrices are provided with codes at our GitHub repository (https://github.com/hiro-toju/Soil_EnergyLandscape_NARO3000) [to be released after the acceptance of the manuscript].

References

Amor DR, Ratzke C, Gore J. 2020. Transient invaders can induce shifts between alternative stable states of microbial communities. *Sci Adv* **6**:eaay8676. doi:10.1126/sciadv.aay8676

Arumugam M, Raes J, Pelletier E, Le Paslier D, Yamada T, Mende DR, Fernandes GR, Tap J, Bruls T, Batto JM, Bertalan M, Borruel N, Casellas F, Fernandez L, Gautier L, Hansen T, Hattori M, Hayashi T, Kleerebezem M, Kurokawa K, Leclerc M, Levenez F, Manichanh C, Nielsen HB, Nielsen T, Pons N, Poulain J, Qin J, Sicheritz-Ponten T, Tims S, Torrents D, Ugarte E, Zoetendal EG, Wang J, Guarner F, Pedersen O, De Vos WM, Brunak S, Doré J, Weissenbach J, Ehrlich SD, Bork P. 2011. Enterotypes of the human gut microbiome. *Nature* **473**:174–180. doi:10.1038/nature09944

Baümler AJ, Sperandio V. 2016. Interactions between the microbiota and pathogenic bacteria in the gut. *Nature*. doi:10.1038/nature18849

Becker OM, Karplus M. 1997. The topology of multidimensional potential energy surfaces: Theory and application to peptide structure and kinetics. *Journal of Chemical Physics* **106**:1495. doi:10.1063/1.473299

Beisner BE, Haydon DT, Cuddington K. 2003. Alternative stable states in ecology. *Front Ecol Environ* **1**:376–382. doi:10.1890/1540-9295(2003)001[0376:ASSIE]2.0.CO;2

Carding S, Verbeke K, Vipond DT, Corfe BM, Owen LJ. 2015. Dysbiosis of the gut microbiota in disease. *Microb Ecol Health Dis* **26**:26191. doi:10.3402/mehd.v26.26191

bioRxiv preprint doi: https://doi.org/10.1101/2022.08.23.505048; this version posted April 3, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

446 Chiang KS, Liu HI, Bock CH. 2017. A discussion on disease severity index values. Part I: warning
447 on inherent errors and suggestions to maximise accuracy. *Annals of Applied Biology*
448 **171**:139–154. doi:10.1111/aab.12362

449 Costea PI, Hildebrand F, Manimozhiyan A, Bäckhed F, Blaser MJ, Bushman FD, De Vos WM,
450 Ehrlich SD, Fraser CM, Hattori M, Huttenhower C, Jeffery IB, Knights D, Lewis JD, Ley
451 RE, Ochman H, O'Toole PW, Quince C, Relman DA, Shanahan F, Sunagawa S, Wang J,
452 Weinstock GM, Wu GD, Zeller G, Zhao L, Raes J, Knight R, Bork P. 2017. Enterotypes in
453 the landscape of gut microbial community composition. *Nat Microbiol* **3**:8–16.
454 doi:10.1038/s41564-017-0072-8

455 Coyte KZ, Rao C, Rakoff-Nahoum S, Foster KR. 2021. Ecological rules for the assembly of
456 microbiome communities. *PLoS Biol* **19**:e3001116. doi:10.1371/JOURNAL.PBIO.3001116

457 Dakos V, Kéfi S. 2022. Ecological resilience: What to measure and how. *Environmental Research*
458 *Letters* **17**:043003. doi:10.1088/1748-9326/ac5767

459 Davidson TA, Sayer CD, Jeppesen E, Søndergaard M, Lauridsen TL, Johansson LS, Baker A,
460 Graeber D. 2023. Bimodality and alternative equilibria do not help explain long-term
461 patterns in shallow lake chlorophyll-a. *Nat Commun* **14**:398. doi:10.1038/s41467-023-
462 36043-9

463 Elias S, Banin E. 2012. Multi-species biofilms: Living with friendly neighbors. *FEMS Microbiol*
464 *Rev* **36**:990–1004. doi:10.1111/j.1574-6976.2012.00325.x

465 Faust K, Lahti L, Gonze D, de Vos WM, Raes J. 2015. Metagenomics meets time series analysis:
466 Unraveling microbial community dynamics. *Curr Opin Microbiol* **25**:56–66.
467 doi:10.1016/j.mib.2015.04.004

468 Faust K, Raes J. 2012. Microbial interactions: From networks to models. *Nat Rev Microbiol*
469 **10**:538–550. doi:10.1038/nrmicro2832

470 Fujita H, Ushio M, Suzuki K, Abe M, Yamamichi M, Iwayama K, Canarini A, Hayashi I,
471 Fukushima K, Fukuda S, Kiers E, Toju H. 2023. Alternative stable states, nonlinear behavior,
472 and predictability of microbiome dynamics. *Microbiome* **11**:63. doi:10.1186/s40168-023-
473 01474-5

474 Fujita Hiroaki, Ushio M, Suzuki K, Abe MS, Yamamichi M, Okazaki Y, Canarini A, Hayashi I,
475 Fukushima K, Fukuda S, Kiers ET, Toju H. 2023. Facilitative interaction networks in
476 experimental microbial community dynamics. *Front Microbiol* **14**:1153952.
477 doi:10.3389/fmicb.2023.1153952

bioRxiv preprint doi: https://doi.org/10.1101/2022.08.23.505048; this version posted April 3, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

478 Fujita H, Yoshida S, Suzuki K, Toju H. 2024. Soil prokaryotic and fungal biome structures
479 associated with crop disease status across the Japan Archipelago. *mSphere* in press.

480 Fukami T. 2015. Historical contingency in community assembly: integrating niches, species
481 pools, and priority effects. *Annu Rev Ecol Evol Syst* **46**:1–23. doi:10.1146/annurev-ecolsys-
482 110411-160340

483 Guim Aguadé-Gorgorió, Jean-François Arnoldi, Matthieu Barbier, Sonia Kéfi. 2023. A taxonomy
484 of multiple stable states in complex ecological communities. *bioRxiv*
485 https://doi.org/10.1101/2023.08.30.555051.

486 Gunderson LH. 2000. Ecological resilience - In theory and application. *Annu Rev Ecol Syst*
487 **31**:425–439. doi:10.1146/annurev.ecolsys.31.1.425

488 Harris DJ. 2016. Inferring species interactions from co-occurrence data with Markov networks.
489 *Ecology* **97**:3308–3314. doi:10.1002/ecy.1605

490 Hastings A, Abbott KC, Cuddington K, Francis T, Gellner G, Lai YC, Morozov A, Petrovskii S,
491 Scranton K, Zeeman M Lou. 2018. Transient phenomena in ecology. *Science (1979)*
492 **361**:eaat6412. doi:10.1126/science.aat6412

493 Hayashi I, Fujita H, Toju H. 2024. Deterministic and stochastic processes generating alternative
494 states of microbiomes. *ISME Communications* **4**:ycae007. doi:10.1093/ismeco/ycae007

495 Hirota M, Holmgren M, Van Nes EH, Scheffer M. 2011. Global resilience of tropical forest and
496 savanna to critical transitions. *Science (1979)* **334**:232–235. doi:10.1126/science.1210657

497 Ibelings BW, Portielje R, Lammens EHRR, Noordhuis R, Van Den Berg MS, Joosse W, Meijer
498 ML. 2007. Resilience of alternative stable states during the recovery of shallow lakes from
499 eutrophication: Lake Veluwe as a case study. *Ecosystems* **10**:4–16. doi:10.1007/s10021-006-
500 9009-4

501 Jansson JK, Hofmockel KS. 2018. The soil microbiome — from metagenomics to metaphenomics.
502 *Curr Opin Microbiol* **43**:162–168. doi:10.1016/j.mib.2018.01.013

503 Jeffery IB, Claesson MJ, O'Toole PW, Shanahan F. 2012. Categorization of the gut microbiota:
504 Enterotypes or gradients? *Nat Rev Microbiol* **10**:591–592. doi:10.1038/nrmicro2859

505 Johnston PR, Quijada L, Smith CA, Baral HO, Hosoya T, Baschien C, Pärtel K, Zhuang WY,
506 Haelewaters D, Park D, Carl S, López-Giráldez F, Wang Z, Townsend JP. 2019. A
507 multigene phylogeny toward a new phylogenetic classification of Leotiomycetes. *IMA*
508 *Fungus* **10**:1–22. doi:10.1186/s43008-019-0002-x

bioRxiv preprint doi: https://doi.org/10.1101/2022.08.23.505048; this version posted April 3, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

509 Knights D, Ward TL, McKinlay CE, Miller H, Gonzalez A, McDonald D, Knight R. 2014.
510 Rethinking enterotypes. *Cell Host Microbe* **16**:433–437. doi:10.1016/j.chom.2014.09.013

511 Leopold DR, Wilkie JP, Dickie IA, Allen RB, Buchanan PK, Fukami T. 2017. Priority effects are
512 interactively regulated by top-down and bottom-up forces: evidence from wood decomposer
513 communities. *Ecol Lett* **20**:1054–1063. doi:10.1111/ele.12803

514 Lewontin RC. 1969. The meaning of stability. *Brookhaven Symp Biol* **22**:13–23.
515 doi:10.2307/j.ctv14161w4.26

516 Long C, Deng J, Nguyen J, Liu Y-Y, Alm EJ, Solé R, Saavedra S. 2024. Structured community
517 transitions explain the switching capacity of microbial systems. *Proceedings of the National*
518 *Academy of Sciences* **121**:e2312521121. doi:10.1073/pnas.2312521121

519 May RM. 1977. Thresholds and breakpoints in ecosystems with a multiplicity of stable states.
520 *Nature* **269**:471–477. doi:10.1038/269471a0

521 Mazzola M, Freilich S. 2017. Prospects for biological soilborne disease control: Application of
522 indigenous versus synthetic microbiomes. *Phytopathology* **107**:256–263.
523 doi:10.1094/PHYTO-09-16-0330-RVW

524 Mendes R, Garbeva P, Raaijmakers JM. 2013. The rhizosphere microbiome: Significance of plant
525 beneficial, plant pathogenic, and human pathogenic microorganisms. *FEMS Microbiol Rev*
526 **37**:634–663. doi:10.1111/1574-6976.12028

527 Mendes R, Kruijt M, de Bruijn I, Dekkers E, van der Voort M, Schneider JHM, Piceno YM,
528 DeSantis TZ, Andersen GL, Bakker P a HM, Raaijmakers JM. 2011. Deciphering the
529 rhizosphere microbiome for disease-suppressive bacteria. *Science* **332**:1097–100.
530 doi:10.1126/science.1203980

531 Meyling N V., Eilenberg J. 2007. Ecology of the entomopathogenic fungi Beauveria bassiana and
532 Metarhizium anisopliae in temperate agroecosystems: Potential for conservation biological
533 control. *Biological Control* **43**:145–155. doi:10.1016/j.biocontrol.2007.07.007

534 Möckel L, Meusemann K, Misof B, Schwartze VU, De Fine Licht HH, Voigt K, Stielow B, de
535 Hoog S, Beutel RG, Buellesbach J. 2022. Phylogenetic Revision and Patterns of Host
536 Specificity in the Fungal Subphylum Entomophthoromycotina. *Microorganisms* **10**:256.
537 doi:10.3390/microorganisms10020256

538 Moeller AH, Degnan PH, Pusey AE, Wilson ML, Hahn BH, Ochman H. 2012. Chimpanzees and
539 humans harbour compositionally similar gut enterotypes. *Nat Commun* **3**:1179.
540 doi:10.1038/ncomms2159

bioRxiv preprint doi: https://doi.org/10.1101/2022.08.23.505048; this version posted April 3, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Paine RT. 1966. Food Web Complexity and Species Diversity. *Am Nat* **100**:65–75. doi:10.1086/282400

Power ME, Tilman D, Estes JA, Menge BA, Bond WJ, Mills LS, Daily G, Castilla JC, Lubchenco J, Paine RT. 1996. Challenges in the quest for keystones. *Bioscience* **46**:609–620. doi:10.2307/1312990

R Core Team. 2020. R: A language and environment for statistical computing. *R: A language and environment for statistical computing R Foundation for Statistical Computing, Vienna, Austria URL https://wwwR-project.org/*.

Raaijmakers JM, Vlami M, de Souza JT. 2002. Antibiotic production by bacterial biocontrol agents. *Antonie Van Leeuwenhoek* **81**:537–547. doi:10.1023/A:1020501420831

Sánchez-Pinillos M, Dakos V, Kéfi S. 2024. Ecological dynamic regimes: A key concept for assessing ecological resilience. *Biol Conserv* **289**:110409. doi:10.1016/j.biocon.2023.110409

Scheffer M, Bascompte J, Brock WA, Brovkin V, Carpenter SR, Dakos V, Held H, Van Nes EH, Rietkerk M, Sugihara G. 2009. Early-warning signals for critical transitions. *Nature* **461**:53–59. doi:10.1038/nature08227

Scheffer M, Carpenter S, Foley J a, Folke C, Walker B. 2001. Catastrophic shifts in ecosystems. *Nature* **413**:591–596. doi:10.1038/35098000

Scheffer M, Carpenter SR. 2003. Catastrophic regime shifts in ecosystems: Linking theory to observation. *Trends Ecol Evol* **18**:648–656. doi:10.1016/j.tree.2003.09.002

Scheffer M, Carpenter SR, Lenton TM, Bascompte J, Brock W, Dakos V, Van De Koppel J, Van De Leemput IA, Levin SA, Van Nes EH, Pascual M, Vandermeer J. 2012. Anticipating critical transitions. *Science (1979)* **338**:344–348. doi:10.1126/science.1225244

Scheffer M, Hosper SH, Meijer ML, Moss B, Jeppesen E. 1993. Alternative equilibria in shallow lakes. *Trends Ecol Evol* **8**:275–279. doi:10.1016/0169-5347(93)90254-M

Schlatter D, Kinkel L, Thomashow L, Weller D, Paulitz T. 2017. Disease suppressive soils: New insights from the soil microbiome. *Phytopathology* **107**:1284–1297. doi:10.1094/PHYTO-03-17-0111-RVW

Schröder A, Persson L, De Roos AM. 2005. Direct experimental evidence for alternative stable states: A review. *Oikos* **110**:3–19. doi:10.1111/j.0030-1299.2005.13962.x

Serván CA, Allesina S. 2021. Tractable models of ecological assembly. *Ecol Lett* **24**:1029–1037. doi:10.1111/ele.13702

bioRxiv preprint doi: https://doi.org/10.1101/2022.08.23.505048; this version posted April 3, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

572 Shaw LP, Bassam H, Barnes CP, Walker AS, Klein N, Balloux F. 2019. Modelling microbiome
573 recovery after antibiotics using a stability landscape framework. *ISME Journal* **13**:1845–
574 1856. doi:10.1038/s41396-019-0392-1

575 Smith VH, Schindler DW. 2009. Eutrophication science: where do we go from here? *Trends Ecol*
576 *Evol* **24**:201–207. doi:10.1016/j.tree.2008.11.009

577 Staver AC, Archibald S, Levin S. 2011a. Tree cover in sub-Saharan Africa: Rainfall and fire
578 constrain forest and savanna as alternative stable states. *Ecology* **92**:1063–1072.
579 doi:10.1890/10-1684.1

580 Staver AC, Archibald S, Levin SA. 2011b. The global extent and determinants of savanna and
581 forest as alternative biome states. *Science (1979)* **334**:230–232.
582 doi:10.1126/science.1210465

583 Suding KN, Hobbs RJ. 2009. Threshold models in restoration and conservation: a developing
584 framework. *Trends Ecol Evol* **24**:271–279. doi:10.1016/j.tree.2008.11.012

585 Suzuki K, Nakaoka S, Fukuda S, Masuya H. 2021. Energy landscape analysis elucidates the
586 multistability of ecological communities across environmental gradients. *Ecol Monogr* **91**:1–
587 21. doi:10.1002/ecm.1469

588 Toju H, Abe MS, Ishii C, Hori Y, Fujita H, Fukuda S. 2020. Scoring species for synthetic
589 community design: network analyses of functional core microbiomes. *Front Microbiol*
590 **11**:1361–1361. doi:10.3389/fmicb.2020.01361

591 Toju H, Peay KG, Yamamichi M, Narisawa K, Hiruma K, Naito K, Fukuda S, Ushio M, Nakaoka
592 S, Onoda Y, Yoshida K, Schlaeppi K, Bai Y, Sugiura R, Ichihashi Y, Minamisawa K, Kiers
593 ET. 2018. Core microbiomes for sustainable agroecosystems. *Nat Plants* **4**:247–257.
594 doi:10.1038/s41477-018-0139-4

595 Toju H, Tanaka Y. 2019. Consortia of anti-nematode fungi and bacteria in the rhizosphere of
596 soybean plants attacked by root-knot nematodes. *R Soc Open Sci* **6**:181693.
597 doi:10.1098/rsos.181693

598 Toju H, Yamamoto S, Tanabe AS, Hayakawa T, Ishii HS. 2016. Network modules and hubs in
599 plant-root fungal biomes. *J R Soc Interface* **13**:20151097–20151097.
600 doi:10.1098/rsif.2015.1097

601 Trivedi P, Leach JE, Tringe SG, Sa T, Singh BK. 2020. Plant–microbiome interactions: from
602 community assembly to plant health. *Nat Rev Microbiol* **18**:607–621. doi:10.1038/s41579-
603 020-0412-1

bioRxiv preprint doi: https://doi.org/10.1101/2022.08.23.505048; this version posted April 3, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Verbruggen E, van der Heijden MGA, Rillig MC, Kiers ET. 2013. Mycorrhizal fungal establishment in agricultural soils: Factors determining inoculation success. *New Phytologist* **197**:1104–1109. doi:10.1111/j.1469-8137.2012.04348.x

Vorholt JA, Vogel C, Carlström CI, Müller DB. 2017. Establishing causality: opportunities of synthetic communities for plant microbiome research. *Cell Host Microbe* **22**:142–155. doi:10.1016/j.chom.2017.07.004

Watanabe T, Masuda N, Megumi F, Kanai R, Rees G. 2014. Energy landscape and dynamics of brain activity during human bistable perception. *Nat Commun* **5**:4765. doi:10.1038/ncomms5765

Werner GDA, Kiers ET. 2015. Order of arrival structures arbuscular mycorrhizal colonization of plants. *New Phytologist* **205**:1515–1524. doi:10.1111/nph.13092

Wickham H. 2011. ggplot2. *Wiley Interdiscip Rev Comput Stat* **3**:180–185. doi:10.1002/wics.147

Wood S. 2022. mgcv: Mixed GAM computation vehicle with automatic smoothness estimation available at https://cran.r-project.org/web/packages/mgcv/index.html.

Wu GD, Chen J, Hoffmann C, Bittinger K, Chen YY, Keilbaugh SA, Bewtra M, Knights D, Walters WA, Knight R, Sinha R, Gilroy E, Gupta K, Baldassano R, Nessel L, Li H, Bushman FD, Lewis JD. 2011. Linking long-term dietary patterns with gut microbial enterotypes. *Science (1979)* **334**:105–108. doi:10.1126/science.1208344

Yajima D, Fujita H, Hayashi I, Shima G, Suzuki K, Toju H. 2023. Core species and interactions prominent in fish-associated microbiome dynamics. *Microbiome* **11**:53. doi:10.1186/s40168-023-01498-x

Yuan J, Wen T, Zhang H, Zhao M, Penton CR, Thomashow LS, Shen Q. 2020. Predicting disease occurrence with high accuracy based on soil macroecological patterns of Fusarium wilt. *ISME Journal* **14**:2936–2950. doi:10.1038/s41396-020-0720-5

Zaneveld JR, McMinds R, Thurber RV. 2017. Stress and stability: Applying the Anna Karenina principle to animal microbiomes. *Nat Microbiol* **2**:17121. doi:10.1038/nmicrobiol.2017.121

Zelezniak A, Andrejev S, Ponomarova O, Mende DR, Bork P, Patil KR. 2015. Metabolic dependencies drive species co-occurrence in diverse microbial communities. *Proc Natl Acad Sci U S A* **112**:6449–6454. doi:10.1073/pnas.1421834112

bioRxiv preprint doi: https://doi.org/10.1101/2022.08.23.505048; this version posted April 3, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

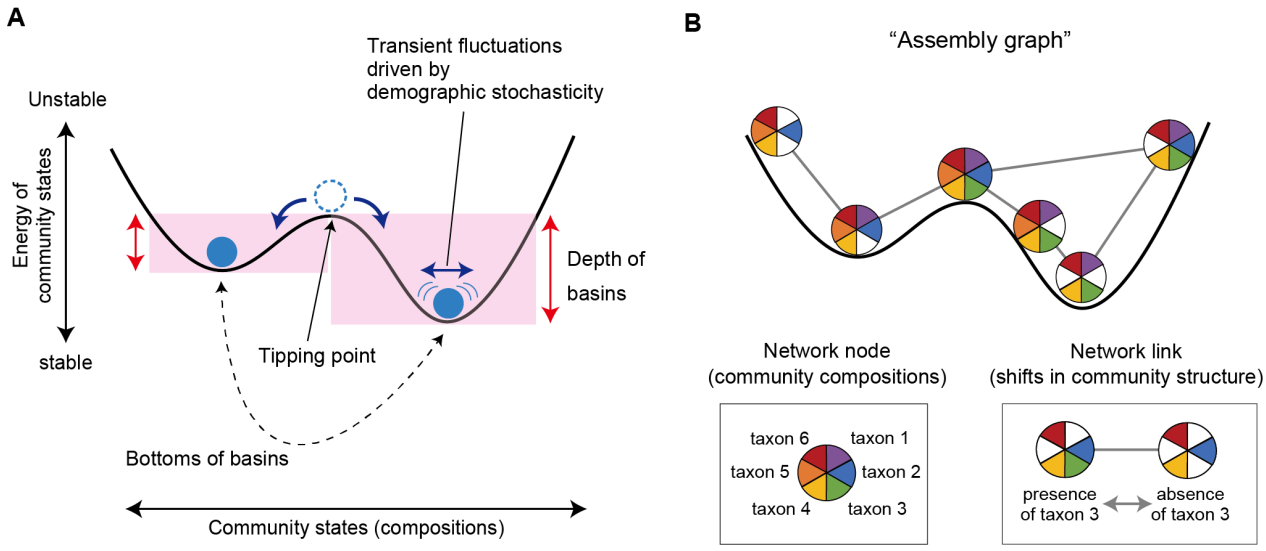


Figure 1. Schema of multistability of ecological communities. (**A**) Basins of attraction and tipping points. The structure of "stability landscapes" showing relationship between community states (species or taxonomic compositions) and their stability is inferred based on the energy landscape analysis. The "energy" of each community state is calculated with maximum entropy models as detailed in Methods. Lower energy represents a more stable community state on a stability landscape. Transient fluctuations around the bottoms of basins (i.e., point attractors) are assumed as probabilistic phenomena in the statistical approach. (**B**) Assembly graph. To explore numerous possible states of real ecological communities, input data are binarized in the energy landscape analysis. Potential transitions between community states are then considered within "assembly graphs", in which paths between different species/taxonomic compositions are treated as network links. Thus, by the assembly-graph approach, the energy landscape analysis provides a general framework for inferring the structure of stability landscapes in empirical studies of complex microbiome datasets.

bioRxiv preprint doi: https://doi.org/10.1101/2022.08.23.505048; this version posted April 3, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

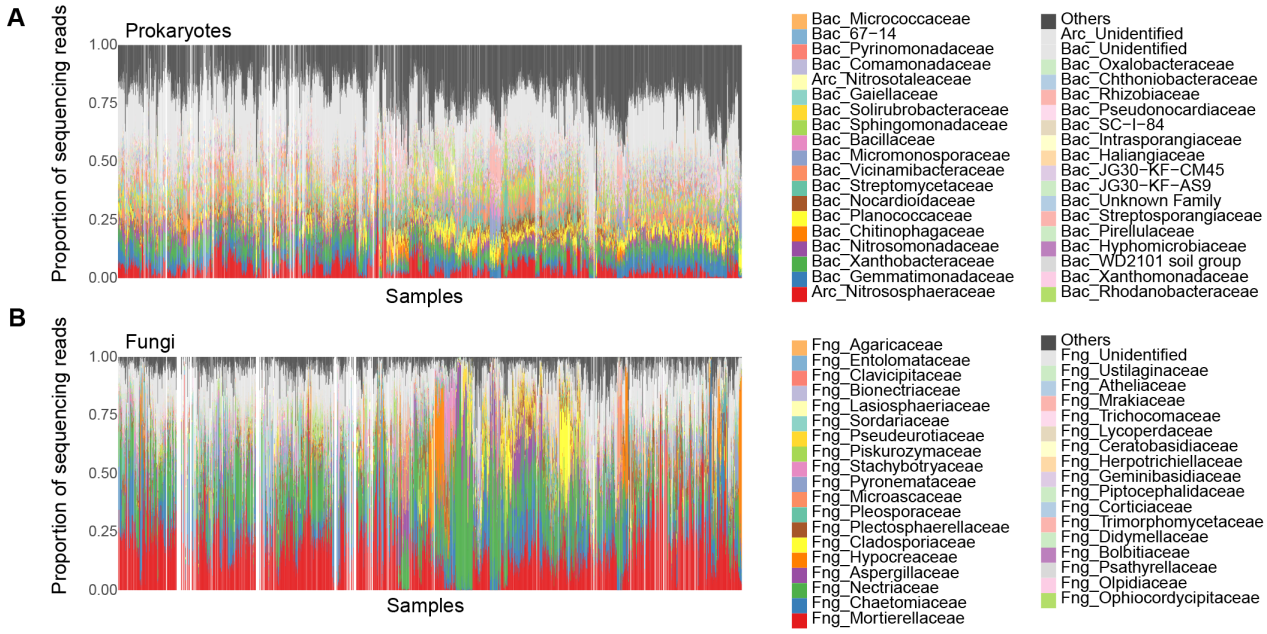
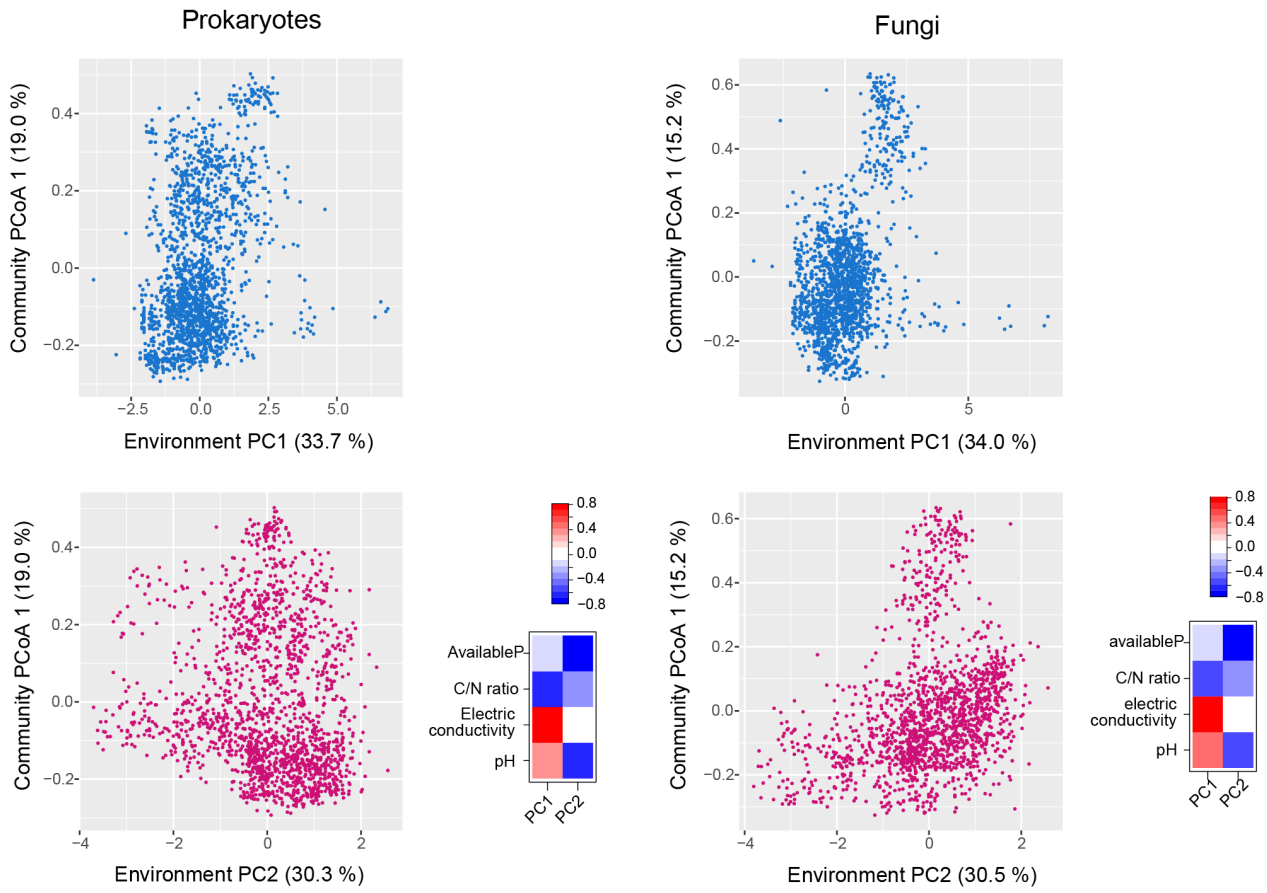


Figure 2. Community structure of the source data. The family-level compositions of prokaryotes (**A**) and fungi (**B**) are shown based on the source dataset (Fujita et al., 2024). The soil samples from which DNA sequence data were unavailable for either prokaryotic 16S rRNA or fungal ITS regions are indicated as blanks.

Figure supplement 1. Community structure of the source data (order- and genus-level compositions).

bioRxiv preprint doi: https://doi.org/10.1101/2022.08.23.505048; this version posted April 3, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.



658

659 **Figure 3.** Community structure along environmental gradients. The scores representing
660 prokaryotic/fungal community compositions (community PCoA1 scores) are shown along each
661 PCA axis of soil environmental conditions. Regarding the environmental PCA axes, factor
662 loadings of environmental variables examined (pH, electrical conductivity, C/N ratio, and
663 available phosphorous concentration) are shown separately for prokaryotic (N = 1,771) and
664 fungal (N = 1,664) datasets.

665 **Figure supplement 1.** Prokaryotic community structure along environmental gradients (detailed
666 analyses).

667 **Figure supplement 2.** Fungal community structure along environmental gradients (detailed
668 analyses).

669

bioRxiv preprint doi: https://doi.org/10.1101/2022.08.23.505048; this version posted April 3, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

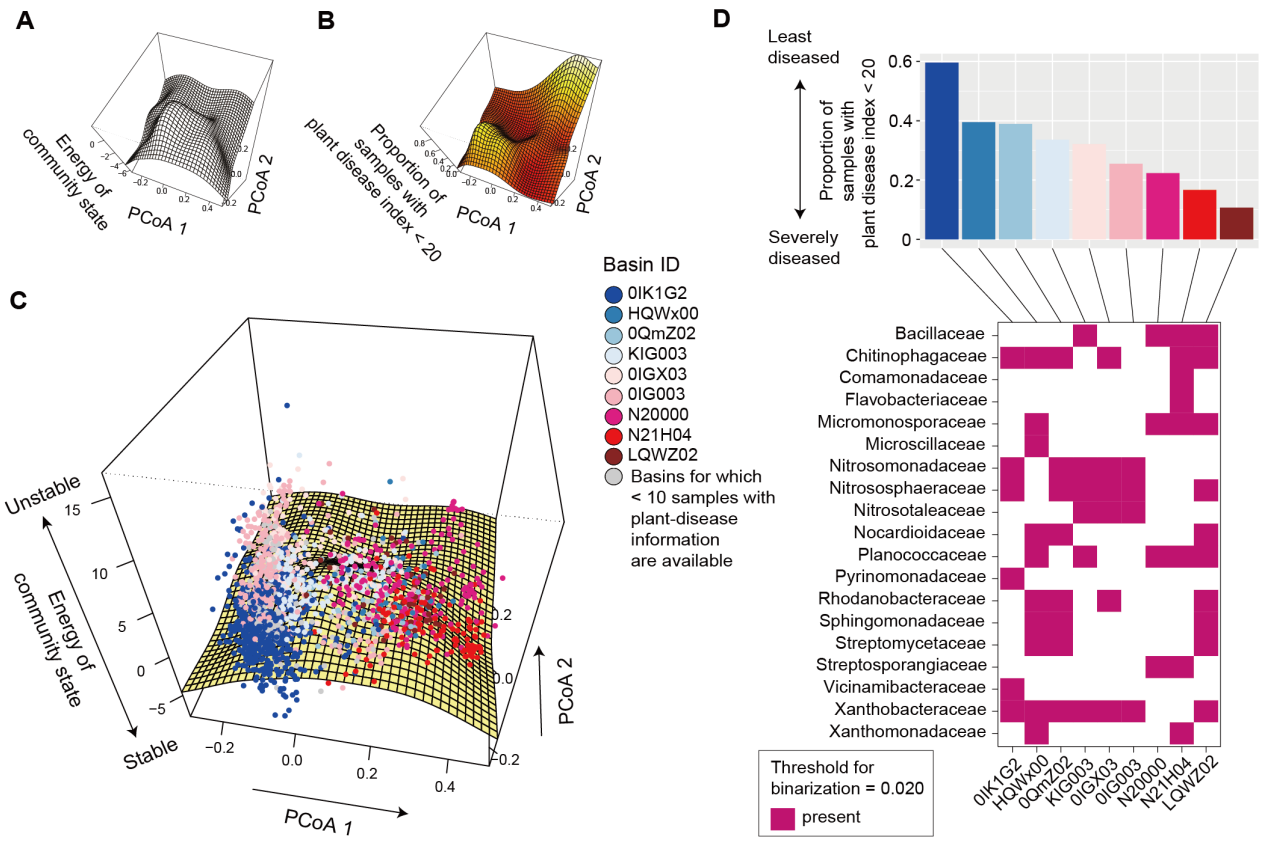


Figure 4. Energy landscape of prokaryotic communities. (**A**) Inferred energy landscape of family-level prokaryotic community structure (threshold for binarization = 0.020; occurrence threshold = 0.10; $S = 35$). The surface of energy levels was reconstructed across the PCoA space of fungal community structure (community PCoA1 and PCoA2 scores in Figures 2–figure supplement 1) based on spline smoothing. Community states with lower energy are inferred to be more stable. (**B**) Landscape of crop disease prevalence. Across the PCoA space of prokaryotic compositions, the proportion of samples with disease severity index < 20 is shown based on spline smoothing. (**C**) Community data points on the energy landscape. The axis of "energy of community state" is more expanded than that in panel **A** in order to cover the range of samples. Data points (samples) indicated by the same color belong to the same basins of attraction, which are represented by the IDs of the community states whose energy is lower than that of any adjacent community states (i.e., bottoms of basins). (**D**) Key taxa whose abundance represent basins. In the upper panel, the mean proportion of soil samples with the minimum level of plant (crop) disease symptoms (the percentage of diseased plants < 20 or disease severity index < 20) is shown for each basin. The lower panel indicates the key taxa whose abundance characterizes difference among the bottoms of the basins.

Figure supplement 1. Energy landscape of prokaryotic communities (order-level compositions;

bioRxiv preprint doi: https://doi.org/10.1101/2022.08.23.505048; this version posted April 3, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

688 threshold for binarization = 0.020; occurrence threshold = 0.10; S = 32).

689 **Figure supplement 2.** Energy landscape of communities including both prokaryotes and fungi
690 (family-level compositions; threshold for binarization = 0.030; occurrence threshold = 0.10; S =
691 31).

692 **Figure supplement 3.** Energy landscape of communities including both prokaryotes and fungi
693 (order-level compositions; threshold for binarization = 0.030; occurrence threshold = 0.10; S =
694 32).

695

bioRxiv preprint doi: https://doi.org/10.1101/2022.08.23.505048; this version posted April 3, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

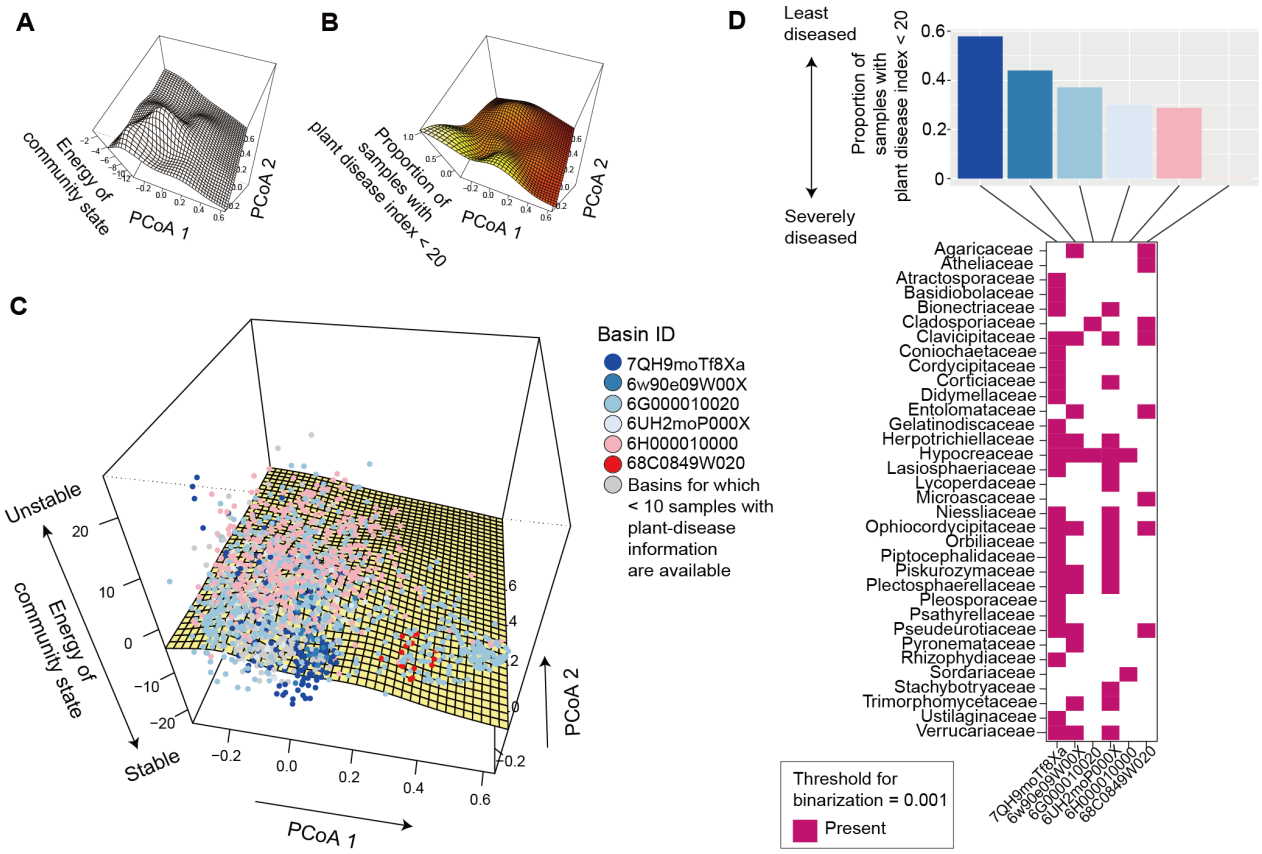


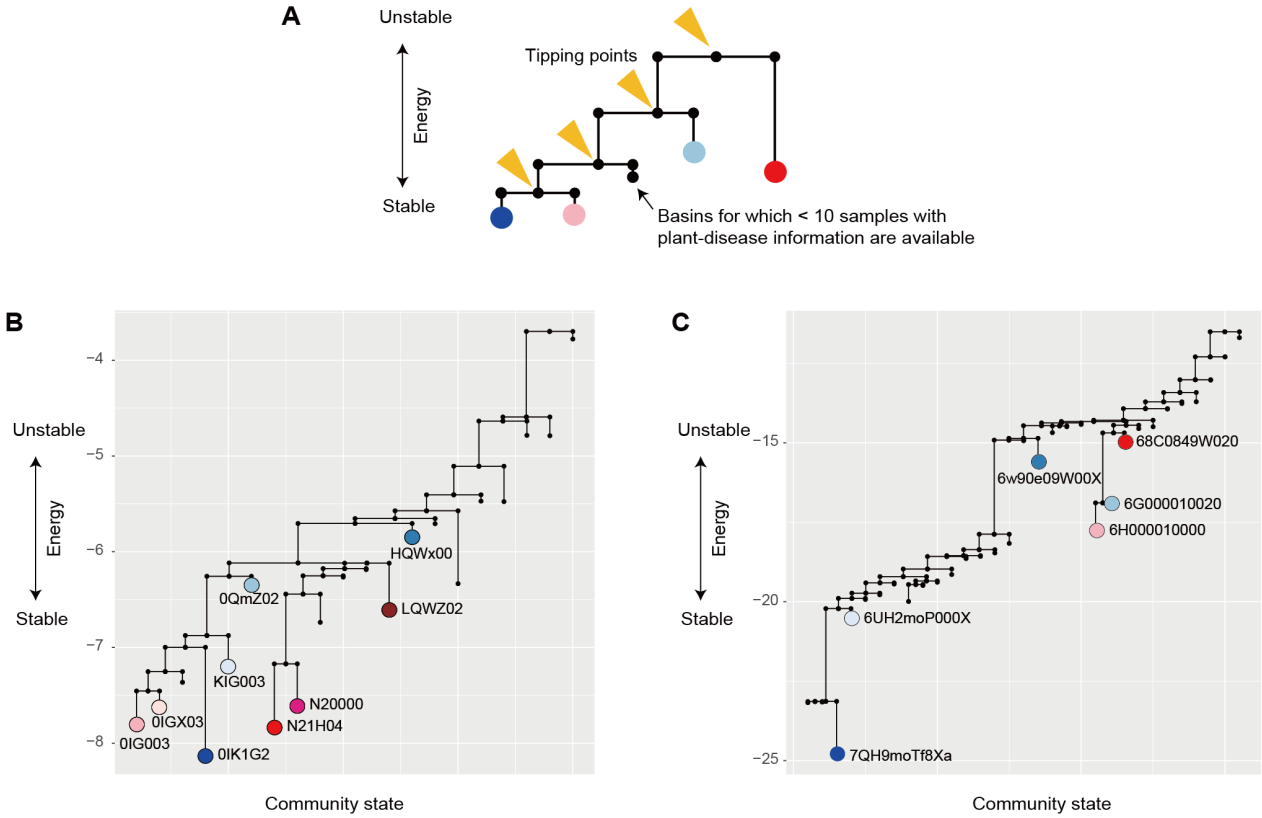
Figure 5. Energy landscape of fungal communities. (**A**) Inferred energy landscape of family-level fungal community structure (threshold for binarization = 0.001; occurrence threshold = 0.05; S = 62). The surface of energy levels was reconstructed across the PCoA space of fungal community structure (community PCoA1 and PCoA2 scores in in Figures 2–figure supplement 2) based on spline smoothing. Community states with lower energy are inferred to be more stable. (**B**) Landscape of crop disease prevalence. Across the PCoA space of prokaryotic compositions, the proportion of samples with disease severity index < 20 is shown based on spline smoothing. (**C**) Community data points on the energy landscape. The axis of "energy of community state" is more expanded than that in panel **A** in order to cover the range of samples. Data points (samples) indicated by the same color belong to the same basins of attraction, which are represented by the IDs of the community states whose energy is lower than that of any adjacent community states (i.e., bottoms of basins). (**D**) Key taxa whose abundance represent basins. In the upper panel, the mean proportion of soil samples with the minimum level of plant (crop) disease symptoms (the percentage of diseased plants < 20 or disease severity index < 20) is shown for each basin. The lower panel indicates the key taxa whose abundance characterizes difference among the bottoms of the basins.

bioRxiv preprint doi: https://doi.org/10.1101/2022.08.23.505048; this version posted April 3, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

713 **Figure supplement 1.** Energy landscape of fungal communities (family-level compositions;
714 threshold for binarization = 0.001; occurrence threshold = 0.10; $S = 42$).

715

bioRxiv preprint doi: https://doi.org/10.1101/2022.08.23.505048; this version posted April 3, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.



716

717 **Figure 6.** Disconnectivity graphs of the energy landscapes. (**A**) Schema of a disconnectivity
718 graph. The energy of the "tipping points" splitting basins of attraction are presented across the
719 axis of 2^S possible community states, where S denotes the number of the species or taxa
720 examined. The energy of the bottom of each basin is shown. (**B**) Tipping points and basins on the
721 energy landscape of prokaryotes. The major basins of attraction with ≥ 10 samples with plant-
722 disease information are highlighted with the colors defined in Figure 4. (**C**) Tipping points and
723 basins on the energy landscape of fungi. The major basins of attraction with ≥ 10 samples with
724 plant-disease information are highlighted with the colors defined in Figure 5.

725