Alternative stable states of soil microbiomes and ecosystem functions

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

[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

Short title: Alternative stable states of microbiomes

Subject Areas: ecology, environmental science, microbiology

Keywords: community assembly, critical transitions, hysteresis, keystone species, multiple stability, tipping points

Author for correspondence:

e-mail: toju.hirokazu.4c@kyoto-u.ac.jp

Tel, +81-77-549-8234; Fax +81-77-549-8201.

Abstract

Theory predicts that biological communities can have multiple stable equilibria of species/taxonomic compositions. The presence of such alternative stable states has been demonstrated in classic empirical studies on forest–savanna transitions and those on eutrophication in freshwater lakes. Nonetheless, it remains a major challenge to extend the ecological research 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 > 2,000 agroecosystem soil samples suggested that prokaryotic communities could have numerous local compositional equilibria. In contrast, fungal community compositions were classified into a small number of alternative stable states, suggesting fundamental difference in community assembly between prokaryotic and fungal communities. We also found that the stable states differed greatly in their associations with crop disease prevalence in agroecosystems. A further analysis highlighted microbial taxa potentially playing key roles in transitions between alternative stable states 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.

1. Introduction

The idea that biological communities can show alternative stable states have attracted both empirical and theoretical ecologists since the late 1960s [1–4]. The concept of multistability has been examined in aquatic and terrestrial ecosystems [5–7]. 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 [3,8,9]. Likewise, worldwide inventories of tree cover have shown that forests and savanna can be recognized as alternative stable states of vegetation [10–12]. Those studies have further indicated that ecosystem functions (e.g., fishery, agricultural, and forestry production) can differ greatly between alternative states of biological communities [3,5,13,14]. 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 demonstrated the presence of alternative stable states in nature, recent technical advances in microbial community (microbiome) research have come to provide an ideal opportunity for deepening our knowledge of multistability in ecosystems [15–20]. 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 [18,21,22]. 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 [23,24] (see also [25,26]). Intriguingly, these alternative gut microbiomes ("enterotypes") differ in their associations with human disease such as type II diabetes and Crohn's disease [17]. In addition to those studies on animal-associated microbiomes [23,27,28], 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 functions [16,29]. 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 firm statistical signs of alternative stable states.

In theoretical ecology, stability of community states (taxonomic or species compositions) is often discussed in the framework of stability landscapes [4,5,30]. On the landscape representing stability/instability of community structure (figure 1), stable community states are split by

unstable community states [4,5,13,30]. If those alternative stable states [4,13] differ in their community-level biological functions, 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 [31,32], it is now possible to infer "energy landscapes", which represent structure of stability landscapes, from empirical datasets of ecological communities [30]. The statistical framework allows us to explore stable states of community compositions within the "assembly graphs" [33,34] representing paths of possible community assembly [21,30] (figure 1). Although hundreds or thousands of community compositional data points are required to apply the statistical approach [21,30], 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 alternative stable states based on massive community datasets, the energy landscape analysis has been applied only to a few microbial community datasets [21,30].

We here extend the analysis of alternative stable states to soil microbiomes, which often show highest levels of compositional and structural diversity in nature. We compile a cropland soil microbiome dataset consisting of > 2,000 sampling positions across the Japan Archipelago [35] With the massive dataset, we infer the compositional stability of prokaryotic and fungal communities based on maximum entropy models of an energy landscape analysis [30]. The number of alternative stable states are then compared between prokaryotic and fungal communities. We also examine whether alternative stable states of soil microbiomes can differ in ecosystem-scale functions by focusing on potential relationship between soil microbial compositions and the prevalence of crop plant disease. We further explore key microbial taxa whose abundance critically divide alternative stable states with favorable and unfavorable ecosystem functions. 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.

2. Methods

(a) Dataset compilation

We compiled a publicly available dataset of cropland soil microbiomes (DDBJ Bioproject: PSUB018361) with its metadata of the samples [35]. In a previous study [35], 2,903 bulk soil samples collected 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 [35]. 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 [36]) was available for 1,472 samples [35]. 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 [35]. 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 [35] (figures 2 and S1).

(b) Community structure along environmental gradients

We first inspected how prokaryotic and fungal community structure varied along environmental gradients. Specifically, the scores of the principal coordinate analysis (PCoA) performed based on Bray-Curtis β-diversity in the previous study (i.e., scores defined based on the ASV-level compositions of prokaryotes or fungi [35]) were shown along each axis of edaphic factors (soil pH, electrical conductivity, C/N ratio, and available phosphorous concentration). For each plot representing relationship between environmental conditions and community structure, the density of data points was visualized with the ggplot2 3.3.6 package [37] of R v.4.1.2 [38].

(c) Energy landscape analysis

We estimated the alternative stable states of microbial community structure based on the framework of an energy landscape analysis [21,30,31] (tutorials of energy landscape analyses are available at https://community.wolfram.com/groups/-/m/t/2358581). In this study, we use the

term "energy" as defined in the following equations based on statistical physics [30,31]. Within the "assembly graphs" representing paths of community dynamics [33,34], globally stable states of community compositions can be explored as detailed previously [30]. In brief, probabilities of community states $\mathrm{p}\left(\vec{\sigma}^{(k)}\right)$ are given by

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

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

where $\vec{\sigma}^{(k)} = (\sigma_1^{(k)}, \sigma_2^{(k)}, \dots, \sigma_S^{(k)})$ is a community state vector of *k*-th community state and *S* is the total number of taxa (e.g., ASVs, species, genera, or families) examined. $\varepsilon = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_M)$ is an array of continuous values representing environmental factors (e.g., soil pH and electrical conductivity) and *M* is the total number of environmental parameters. $\sigma_{\mathrm{i}}^{(k)}$ is a binary variable 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, families that represented 10 % or more of the total sequencing reads of a target sample were coded as 1, while those representing less than 10 % of total reads were coded as 0. Then, 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 [30] we identified the stable state communities as the energy minima of the weighted network (nodes having the lowest energy compared to all its neighbors), and determined their basins of attraction [1,30]. 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.

(d) Number of alternative stable states

The number of stable states within energy landscapes of community structure was inferred, respectively, for three types of datasets, namely, the prokaryotic community matrix, the fungal matrix, and the matrix including both prokaryotes and fungi. 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. Likewise, additional analyses based on another threshold value (5 %) in the binarization of community states were conducted. Consequently, we performed an energy landscape analyses in each of the combinations of three types of data (prokaryotes, fungi, and prokaryotes + fungi), two taxonomic levels (family or order), and two threshold levels (5 % and 10 %). For each combination (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 [30]. The parameters were regularized by a logistic prior with location 0 and scale 2.0 (for environmental responses) or 0.5 (for pairwise relationships) [39]. Hyperparameters for the algorithm, including a maximum number of iteration steps T = 100000, were set according to the preliminary analysis that confirmed convergence of the model parameters. The number of stable states was then estimated by using the gradient descent algorithm for 20,000 randomly generated community states within the 2^S possible community states. In the heuristic search, sets of values of the environmental variables (soil pH, electrical conductivity, C/N ratio, and available phosphorous concentration) were drawn from those of sampling points. Then, with these environmental conditions, we determined their basins of attraction [1] based on a steepest descent procedure [30].

(e) Energy landscape structure

We then visualized the energy landscapes by showing the energy of each soil sample on the two-dimensional surface of the community state space defined with the abovementioned PCoA scores. Minor stable states whose shallowest “energy barrier” (i.e., deviation of energy from adjacent “tipping points”) was less than 1/10 of that of the deepest alternative stables states were

196 incorporated into the basins of deeper stable states (i.e., pruning of alternative stable states [30]).
197 Spline smoothing of the energy landscape was performed with optimized penalty scores using the
198 mgcv v.1.8-40 package [40] of R.

199

200 (f) Ecosystem functions and key taxa

201 For the inferred alternative stable states of microbial community compositions, associations with
202 crop disease prevalence were examined. We first constructed the list of soil samples whose
203 community structure was located within the basin of each alternative stable state. We then
204 characterized each alternative stable state based on the metadata of crop disease symptoms [35].
205 Specifically, for each alternative stable state, we calculated the proportion of constituent soil
206 samples with the minimum level of crop disease symptoms (the percentage of diseased plants <
207 20 or disease severity index < 20; [35]). The alternative stable states representing different levels
208 of crop disease prevalence were then compared in terms of taxonomic compositions in order to
209 explore microbial taxa that were keys to distinguish potentially disease-suppressive and disease-
210 promotive soil ecosystems.

211

212 **3. Results**

213 (a) Community structure along environmental gradients

214 On each plot showing community compositions (PCoA 1 or PCoA 2 scores) along the soil
215 environmental gradient (figure 3), multiple clusters of data points were observed for both
216 prokaryotes and fungi (figure S2). In other words, community states could be classified into some
217 clusters even under the same edaphic conditions [i.e., under the same soil pH, electrical
218 conductivity, carbon/nitrogen (C/N) ratio, and available phosphorous concentration].

219

220 (b) Number of alternative stable states

221 The results of energy landscape reconstruction fundamentally differed between prokaryotes and
222 fungi. Even if community states were defined in a binary format at the family or order level to
223 reduce the complexity of the input data (i.e., 2^S possible community states, where S denotes the
224 number of families; see Methods), more than 10^3 stable states were inferred to exist for the

225 prokaryotic dataset (figure 4). In contrast, the energy landscape of fungal community structure
226 was represented by much smaller number of stable states (figure 4). While the estimated number
227 of stable states varied depending on threshold values in the binarization of community structural
228 data, fungal communities consistently showed lower number of stable states than prokaryotic
229 communities (figure S3).

230

231 (c) Energy landscape structure

232 When the inferred alternative stable states of fungal communities were filtered based on the
233 "depth of basins" (figure 1) and the number of samples belonging to the basins, ten major stable
234 states existed within the energy landscape (figure 5a; see also figure S4a). A spline smoothing
235 suggested that positions of these alternative stable states were divided into mainly two regions
236 within the energy landscape (regions I and II in figure 5a). A supplemental analysis indicated that
237 the two regions of stable community structure differed greatly in their associations with crop
238 disease prevalence (region I, least diseased; region II, severely diseased; figure 5b).

239

240 (d) Ecosystem functions and key taxa

241 The ten alternative stable states detected were distinguished in terms of balance among 22 fungal
242 families with cosmopolitan distributions (figure 6). Hypocreaceae appeared at high proportions in
243 the stable states associated with low disease prevalence. In contrast, Aspergillaceae were
244 relatively abundant in the stable states with high disease potential. In addition to these basic
245 patterns, several families were keys to distinguish least diseased and severely diseased
246 agroecosystems. For example, high abundance of Didymellaceae, Lasiosphaeriaceae,
247 Mrakiaceae, Orbilliaceae, Piptocephalidaceae, and Plectosphaerellaceae characterized the two
248 stable states with the least disease prevalence (figure 6).

249

250 **4. DISCUSSION**

251 We here showed that the existence of alternative stable states could be examined based on a
252 statistical framework commonly applicable to diverse biological communities. The energy
253 landscape analysis allows systematic analyses of complex (species-rich) community datasets by

incorporating the information of multiple environmental factors [30]. While classic studies on alternative stable states discussed ecological processes spanning a few easily distinguishable community states (high/low tree cover in forest-savanna transitions [10–12] or macrophyte-/phytoplankton-dominated state in shallow lakes [5,9]), it is now made possible to define alternative stable states based on high-dimensional community datasets involving hundreds of species/taxa [17,21–23]. Application of the general statistical platform will enhance our understanding of how stability landscape properties differ among diverse microbial and non-microbial systems.

The energy landscape analysis of agroecosystem soil microbiomes suggested that prokaryotic communities had numerous local compositional equilibria (figure 4), presumably due to their rapid and continuous responses to changes in environmental conditions. In contrast, ecological processes of fungal communities could be interpreted as shifts between a small number of alternative stable states (figures 4 and 5). Basically, fungi can form complex hyphal structures and their individual lifespan is much longer than most archaea and bacteria. Because of these biological properties, structure and functions of fungal communities are more likely to be influenced by legacy effects of past ecological processes [41–44] than those of prokaryotes. Thus, difference in the number of alternative stable states may illuminate fundamental difference in community assembly between soil prokaryotic and fungal communities [45] (e.g., difference in the strength of priority effects [41]). Meanwhile, it should be kept in mind that the contrasting patterns may be explained by difference in basic taxonomic diversity (343 prokaryotic vs. 240 fungal families). A meta-analysis across diverse microbiomes will help us decipher confounding effects of microbiome type and α-diversity on stability landscape features.

The inference of stability landscape structure provided an opportunity for evaluating relationship between community stability and ecosystem-scale functions. The ten alternative stable states of 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 [46–48] 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 [16,49–51]. In particular, experimental studies on “synthetic” communities have reorganized our knowledge of microbiome functions [49–51]. Nonetheless, such microbial functions cannot be realized in real agroecosystems if the synthesized or designed microbiome compositions are vulnerable to biotic and abiotic

environmental changes in the wild [52]. Thus, in addition to functional properties, compositional stability is the key to manage microbiomes in agroecosystems [49,53,54]. The finding that the number of alternative stable states is much fewer in fungi than in prokaryotes (figure 4) highlights the possibility that agroecosystem functions are managed by keeping fungal community structure at targeted stable states with positive impacts on crop production. The fact that prokaryotic communities could not be classified into a small number of alternative stable states does not indicate the lack of their roles in crop production. Bacteria and archaea play fundamental roles in carbon- and nitrogen-cycles of soil ecosystems [55,56]. The point is that information of fungal taxonomic compositions is indispensable for adaptively managing microbiome structure in agricultural practices.

In our analysis across the Japan Archipelago, 22 out of the 240 fungal families detected were keys to distinguish least diseased and severely diseased states of soil microbiomes (figure 6). Among them, Hypocreaceae, Orbiliaceae, Piptocephalidaceae, and Plectosphaerellaceae are of particular interest because they include many species utilized as biological control agents for suppressing plant pathogenic fungi and/or plant parasitic nematodes [57–60]. Fungal pathogens and cyst/root-knot nematodes have continued to have devastating impacts on global crop production [48,61,62]. Because those fungi and nematodes cannot be eradicated with chemical fungicide/pesticides, understanding of the biological processes in which their outbreaks are prevented in nature is an essential step for implementing sustainable agriculture. Our results suggested that plant disease could be suppressed under the coexistence of multiple nematophagous and mycoparasitic fungal taxa [63] (figure 6). Thus, statistical analyses of stability landscapes allow the exploration of key species or taxa [64,65], whose management could result in transitions from unfavorable ecosystem states to favorable ones [3,5,13,14]. Given that all the 22 fungal families highlighted in our analysis have cosmopolitan distributions, a next crucial step is to test whether the alternative stable states defined across the Japan Archipelago can be used to categorize disease-suppressive and disease-susceptible microbiomes in other regions on the globe.

Although the energy landscape analysis enhances our understanding of community stability and functions, its results should be interpreted with caution. First, given that classic empirical studies identified alternative stable states with system-specific simple criteria (e.g., high/low tree cover [10–12]), special care should be taken when we extend the approach to species-rich (high-dimensional) community datasets. In other words, unambiguous and broadly applicable criteria based on statistical evaluation are the prerequisite for comparative analyses of community

multistability. Although we applied a straightforward definition of stables states (nodes having the lowest energy compared to all its neighbors within an assembly graph [30]; figure 1) in light of classic theoretical studies [1–4], continuous methodological improvements should be explored towards further comprehensive analyses. Second, our analysis on hyper-diverse soil microbiomes incurred substantial computational costs, forcing us to limit the energy landscape analysis to family-level input data. Further improvements of codes are necessary for inferring stability landscapes at genus-, species-, or strain-level analyses. Third, it should be acknowledged that detailed discussion on ecological processes require time-series datasets [66–68]. Because our present data lacked the information of temporal changes in community structure, we could not discuss potential routes of microbiome transitions within the state space. Thus, monitoring of microbiome compositions [20,28] is necessary for distinguishing transient states [1,69,70] from stable states [21] or gaining deeper insights into shifts between alternative stable states [66,67].

The energy landscape framework of multistability analysis is readily applicable to a wide range of microbiome datasets. Application to human microbiome data is of particular interest in terms of the confirmation of the existence of discrete enterotypes as alternative stable states [26]. In addition, insights into the key microbial species/taxa that would play key roles in the transitions from disease-associated microbiome states to healthy ones will open new directions of microbiome therapy. Furthermore, time-series analyses of community dynamics on stability landscapes will allow us to forecast and prevent transitions into unfavorable community states (e.g., dysbiosis [21,71]). Along with such extensions of observational research, experimental studies controlling key species/taxa or environmental parameters [7] will promote both basic and applied sciences of ecosystem functions.

Data accessibility. The accession number of the DDBJ Sequence Read Archive: PSUB018361.

The microbial community data are provided with codes at our GitHub repository (https://github.com/hiro-toju/Soil_EnergyLandscape_NARO3000).

Authors' contributions. H.F., S.Y. and H.T. designed the work. H.F. performed molecular experiments. H.F. and H.T. analyzed the data. H.T. wrote the paper with H.F., S.Y., and K.S.

Competing interests. We declare no competing interests.

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 FOREST (JPMJFR2048) to H.T., JSPS Grant-in-Aid for Scientific Research (20K06820 and 20H03010) to K.S., and JSPS Fellowship to H.F.

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

References

1. Lewontin RC. 1969 The meaning of stability. *Brookhaven Symp Biol* **22**, 13–23. (doi:10.2307/j.ctv14161w4.26)

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

3. 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)

4. 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)

5. 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)

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

7. 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)

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

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

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

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

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

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

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

15. 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)

16. Toju H *et al.* 2018 Core microbiomes for sustainable agroecosystems. *Nat Plants* **4**, 247–257. (doi:10.1038/s41477-018-0139-4)

17. Costea PI *et al.* 2017 Enterotypes in the landscape of gut microbial community composition. *Nat Microbiol* **3**, 8–16. (doi:10.1038/s41564-017-0072-8)

18. 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)

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

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

21. Fujita H *et al.* 2023 Alternative stable states, nonlinear behavior, and predictability of microbiome dynamics. *Microbiome* **11**, 63.

22. Hayashi I, Fujita H, Toju H. 2023 Deterministic and stochastic processes generating alternative states of microbiomes. *bioRxiv* , https://doi.org/10.1101/2023.04.03.535343.

23. Arumugam M *et al.* 2011 Enterotypes of the human gut microbiome. *Nature* **473**, 174–180. (doi:10.1038/nature09944)

24. Wu GD *et al.* 2011 Linking long-term dietary patterns with gut microbial enterotypes. *Science (1979)* **334**, 105–108. (doi:10.1126/science.1208344)

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

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

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

28. 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)

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

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

31. 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)

32. 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)

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

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

35. Fujita H, Yoshida S, Suzuki K, Toju H. 2023 Soil Prokaryotic and Fungal Biome Structures Associated with Crop Disease Status across the Japan Archipelago. *bioRxiv* , https://doi.org/10.1101/2023.08.31.555816.

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

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

38. 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://www.R-project.org/.*

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

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

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

42. 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)

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

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

45. Ives AR, Carpenter SR. 2007 Stability and diversity of ecosystems. *Science (1979)* **317**, 58–62. (doi:10.1126/science.1133258)

46. 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)

47. Mendes R *et al.* 2011 Deciphering the rhizosphere microbiome for disease-suppressive bacteria. *Science* **332**, 1097–100. (doi:10.1126/science.1203980)

48. 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)

49. Vorholt JA, Vogel C, Carlström CI, Müller DB. 2017 Establishing Causality: Opportunities of Synthetic Communities for Plant Microbiome Research. *Cell Host Microbe*. **22**. (doi:10.1016/j.chom.2017.07.004)

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

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

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

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

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

55. Kuypers MMM, Marchant HK, Kartal B. 2018 The microbial nitrogen-cycling network. *Nat Rev Microbiol* **16**, 263–276. (doi:10.1038/nrmicro.2018.9)

56. Jansson JK, Hofmockel KS. 2020 Soil microbiomes and climate change. *Nat Rev Microbiol* **18**, 35–46. (doi:10.1038/s41579-019-0265-7)

503 57. Yang Y, Yang E, An Z, Liu X. 2007 Evolution of nematode-trapping cells of predatory
504 fungi of the Orbiliaceae based on evidence from rRNA-encoding DNA and multiprotein
505 sequences. *Proc Natl Acad Sci U S A* **104**, 8379–8384. (doi:10.1073/pnas.0702770104)

506 58. Poveda J. 2021 *Trichoderma* as biocontrol agent against pests: New uses for a
507 mycoparasite. *Biological Control* **159**, 104634. (doi:10.1016/j.biocontrol.2021.104634)

508 59. Lazarus KL, Benny GL, Ho HM, Smith ME. 2017 Phylogenetic systematics of syncephalis
509 (zoopagales, zoopagomycotina), a genus of ubiquitous mycoparasites. *Mycologia* **109**, 333–
510 349. (doi:10.1080/00275514.2017.1307005)

511 60. Bourne JM, Kerry BR, De Leij FAAM. 1996 The importance of the host plant on the
512 interaction between root-knot nematodes (*Meloidogyne* spp.) and the Nematophagous
513 fungus, *Verticillium chlamydosporium* Goddard. *Biocontrol Sci Technol* **6**, 539–548.
514 (doi:10.1080/09583159631172)

515 61. Onkendi EM, Kariuki GM, Marais M, Moleleki LN. 2014 The threat of root-knot
516 nematodes (Meloidogyne spp.) in Africa: A review. *Plant Pathol.* **63**, 727–737.
517 (doi:10.1111/ppa.12202)

518 62. Dita M, Barquero M, Heck D, Mizubuti ESG, Staver CP. 2018 Fusarium wilt of banana:
519 Current knowledge on epidemiology and research needs toward sustainable disease
520 management. *Front Plant Sci* **9**, 1468. (doi:10.3389/fpls.2018.01468)

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

524 64. Paine RT. 1966 Food Web Complexity and Species Diversity. *Am Nat* **100**, 65–75.
525 (doi:10.1086/282400)

526 65. Power ME *et al.* 1996 Challenges in the quest for keystones. *Bioscience* **46**, 609–620.
527 (doi:10.2307/1312990)

528 66. Scheffer M *et al.* 2012 Anticipating critical transitions. *Science (1979)* **338**, 344–348.
529 (doi:10.1126/science.1225244)

530 67. Scheffer M *et al.* 2009 Early-warning signals for critical transitions. *Nature* **461**, 53–59.
531 (doi:10.1038/nature08227)

532 68. Davidson TA, Sayer CD, Jeppesen E, Søndergaard M, Lauridsen TL, Johansson LS, Baker
533 A, Graeber D. 2023 Bimodality and alternative equilibria do not help explain long-term

patterns in shallow lake chlorophyll-a. *Nat Commun* **14**, 398. (doi:10.1038/s41467-023-36043-9)

69. Fukami T, Nakajima M. 2011 Community assembly: alternative stable states or alternative transient states? *Ecol Lett* **14**, 973–84. (doi:10.1111/j.1461-0248.2011.01663.x)

70. Hastings A *et al.* 2018 Transient phenomena in ecology. *Science (1979)* **361**, eaat6412. (doi:10.1126/science.aat6412)

71. 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)

Electric supplementary materials

Figure S1. Community structure of the source data.

Figure S2. Community structure along environmental gradients (density plots).

Figure S3. Number of the stable states of community structure (threshold for binarization = 5 %).

Figure S4. Alternative stable states defined for order-level fungal data (threshold for binarization = 5 %).

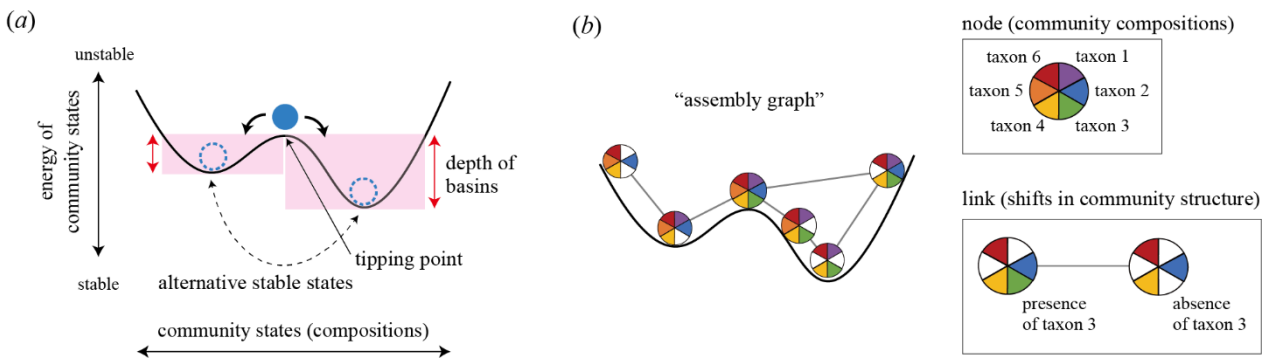

Figure 1. Schema of alternative stable states of ecological communities. (*a*) Stability landscapes. The structure of "stability landscapes" showing relationship between community states (species or taxonomic compositions) and their stability is inferred based on the energy landscape analysis. The "energy" of each community state is calculated with maximum entropy models as detailed in Methods. Lower energy represents a more stable community state on an stability landscape. Multiple stable states (bottoms) on an stability landscape are often called "alternative stable states" in theoretical ecology. (*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.

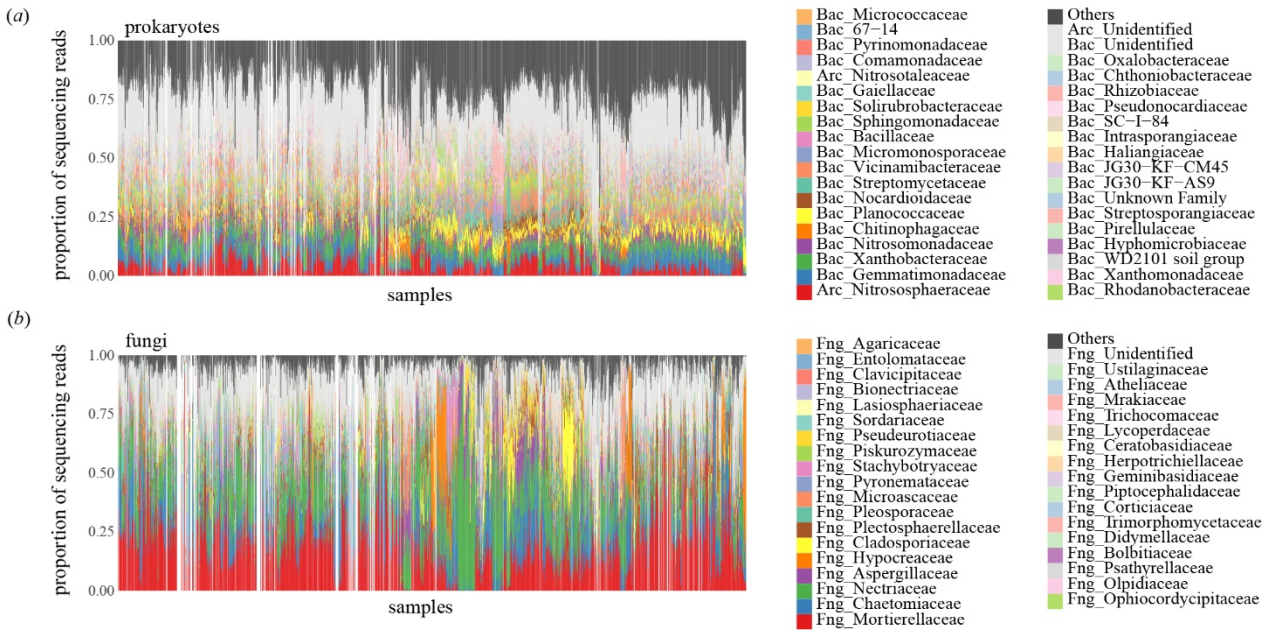

Figure 2. Community structure of the source data. The family-level compositions of prokaryotes (*a*) and fungi (*b*) are shown. The soil samples from which DNA sequence data were unavailable for either prokaryotic 16S rRNA or fungal ITS regions are indicated as blanks. See figure S1 for community compositions at the order and genus levels.

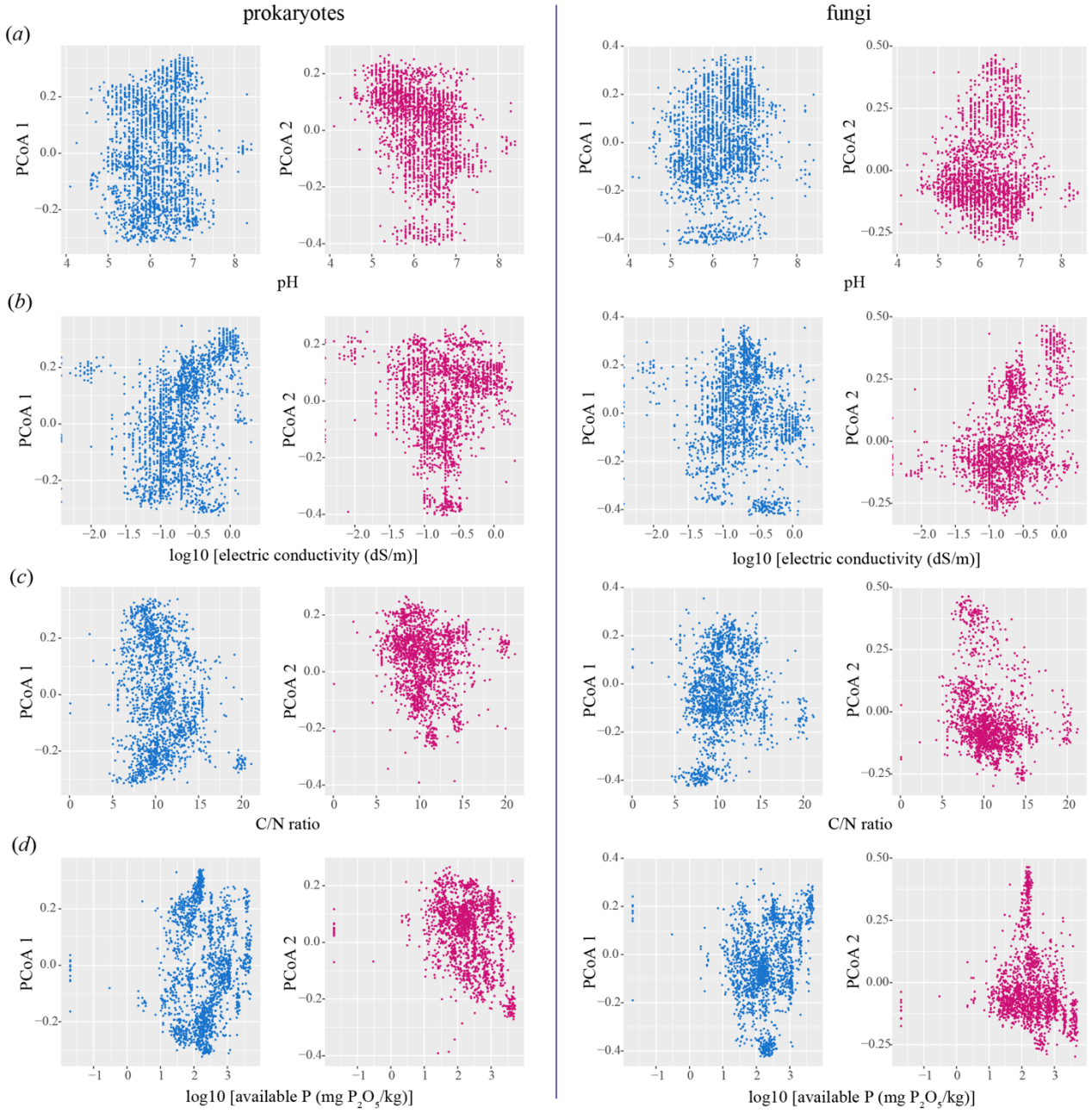

Figure 3. Community structure along environmental gradients. The scores representing prokaryotic/fungal community compositions (PCoA 1 and 2 scores) are shown along each axis of soil environmental factors. (*a*) pH. (*b*) electrical conductivity. (*c*) C/N ratio. (*d*) available phosphorous concentration. See figure S2 for density plots.

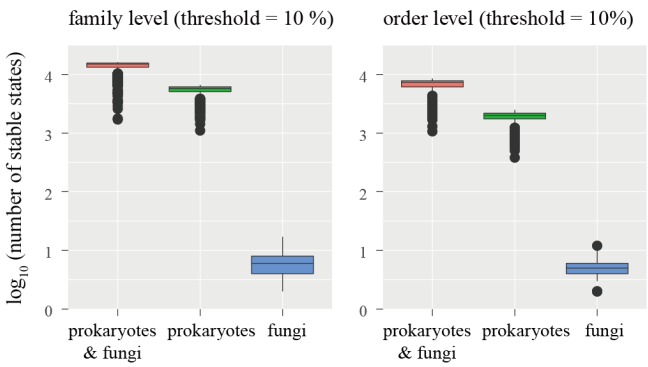

Figure 4. Number of the stable states of community structure. The parameters of the extended pairwise maximum entropy model were adjusted to fit the model to the empirical data (see Methods). The number of stable states in the energy landscape was then estimated by using the gradient descent algorithm for 20,000 randomly generated community states within the 2^S possible community states, where S denotes the number of families or orders. In the heuristic search, sets of values of the environmental variables (soil pH, electrical conductivity, C/N ratio, and available phosphorous concentration) were drawn from those of sampling points. The boxplots show variation in the number of stable states across different sets of environmental variables. A box represents the first quartile, median and the third quartile, while bars indicate the largest/smallest values within 1.5 × inter-quartile ranges from the box: outliers were indicated as circles. The analysis was performed, respectively, for the dataset representing both prokaryotes and fungi, the prokaryote dataset, and the fungus dataset. The threshold value in the binarization of the community data was set at 10 %: see figure S3 for results based on another threshold value (5 %).

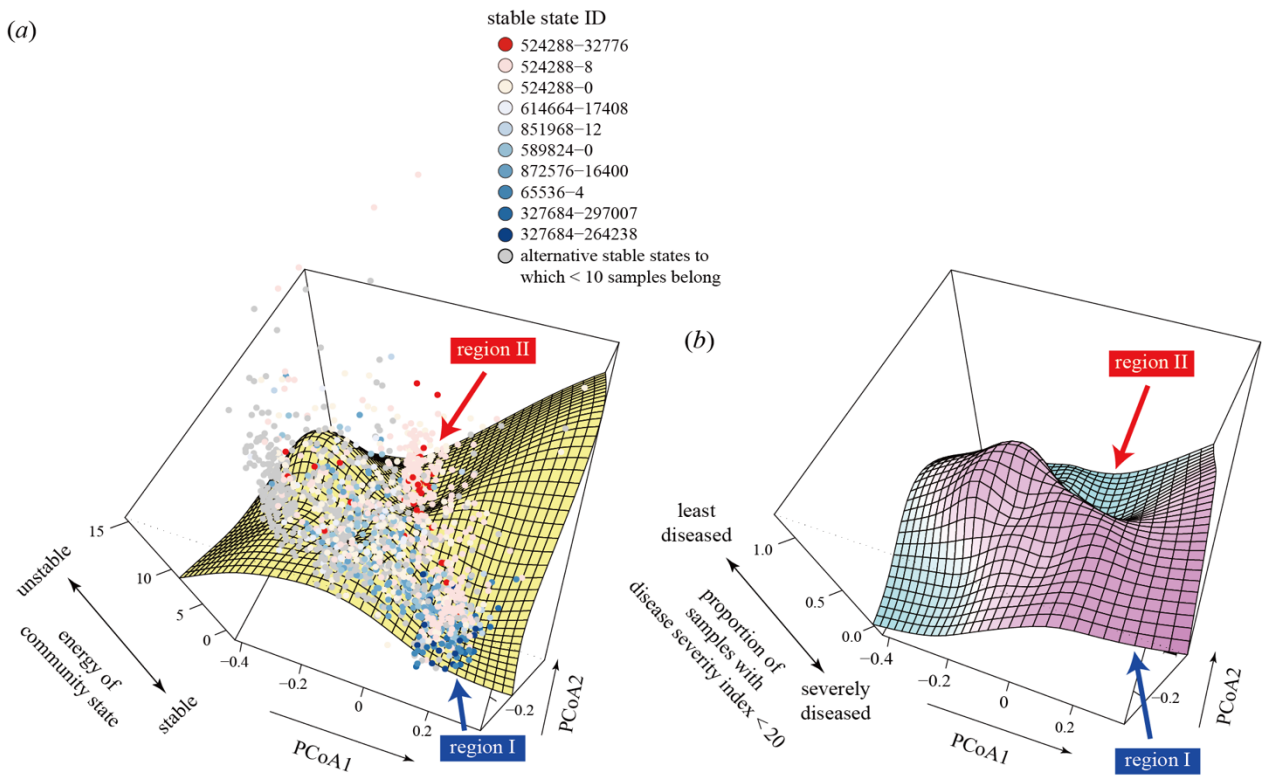

Figure 5. Alternative stable states of microbiomes. (*a*) Energy landscape of fungal community structure. The surface of energy levels was reconstructed on the PCoA space of fungal community structure (PCoA and PCoA 2 scores in figure 3) based on spline smoothing. Community states with lower energy are inferred to be more stable. Data points (samples) indicated by the same color belong to the basins of the same alternative stable states (defined at the family-level taxonomic compositions), whose energy is lower than any adjacent community structure. Two major regions with low energy (i.e., high community stability) are indicated by arrows (regions I and II). (*b*) Relationship between community structure and crop disease prevalence. The two major regions with low community state energy (regions I and II in panel *a*) are indicated by arrows.

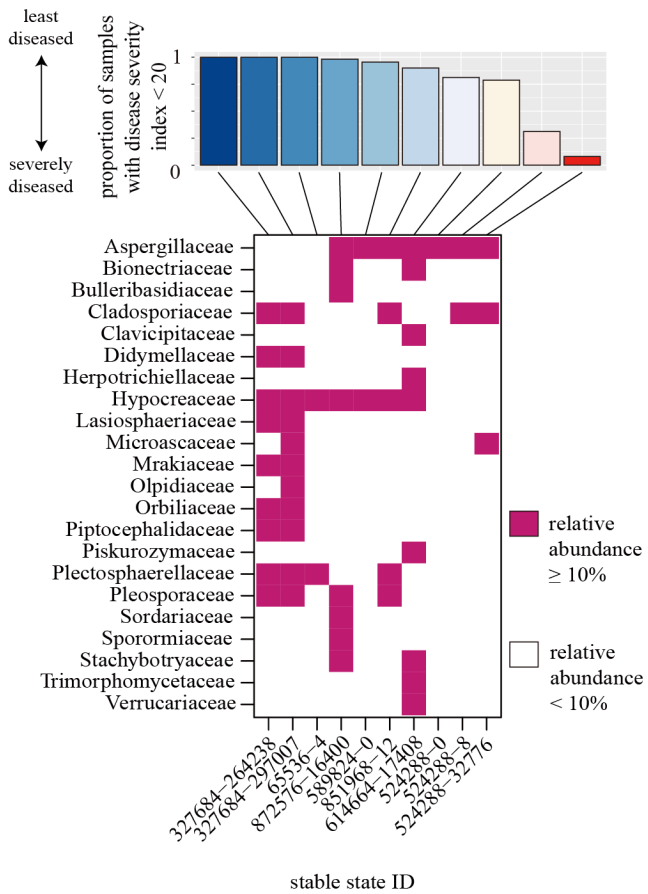

Figure 6. Key taxa whose abundance divide alternative stable states. The 10 alternative stable states representing ≥ 10 samples were compared in terms of their associations with crop disease conditions. The mean proportion of soil samples with the minimum level of crop disease symptoms (the percentage of diseased plants < 20 or disease severity index < 20) is shown for each alternative stable states. To reduce the complexity of input data to a computationally feasible range, alternative stable states were explored based on family-level community compositions (threshold for binarization = 10 %; see Methods for details).

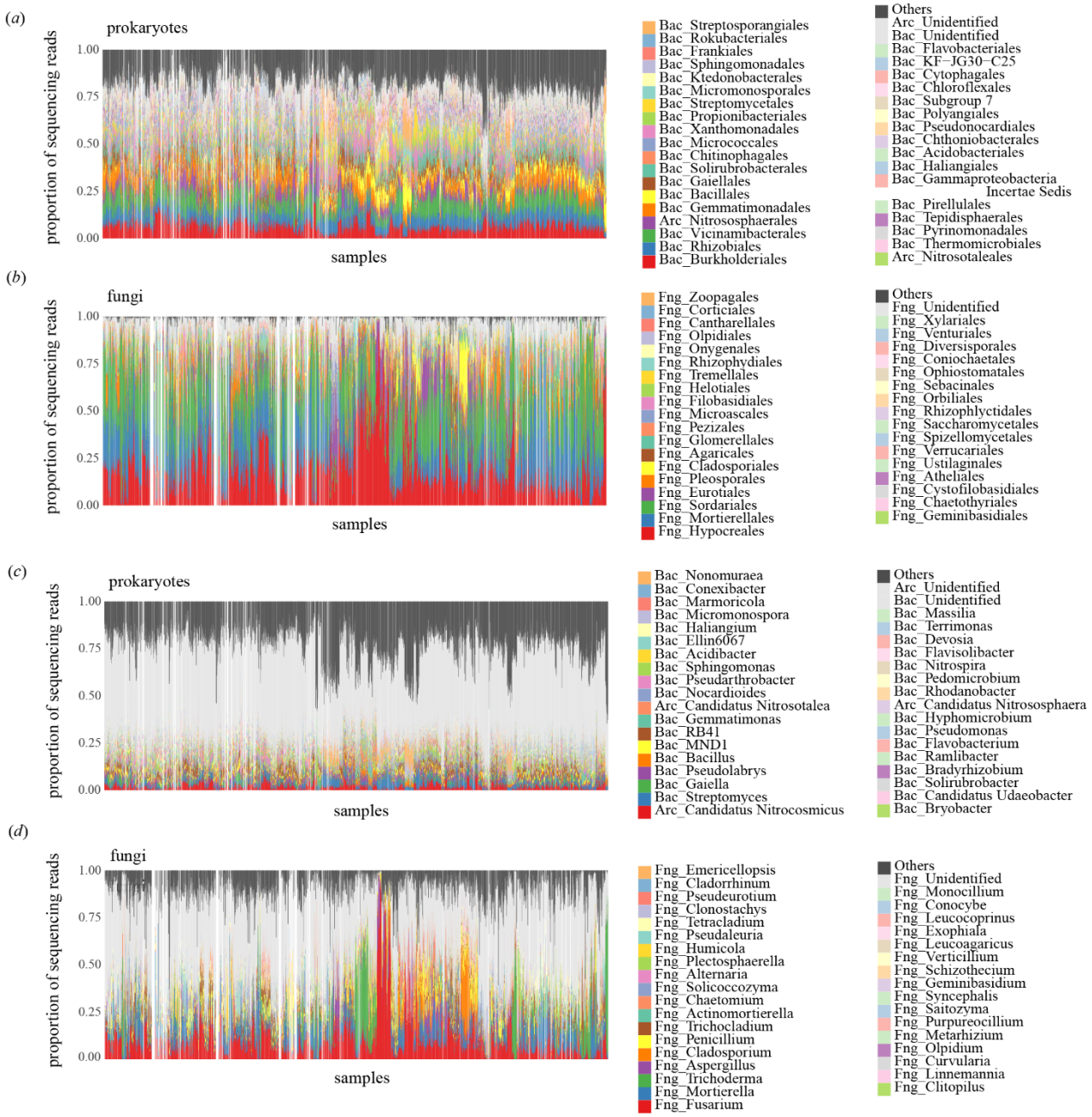

Figure S1. Community structure of the source data. (*a*) Order-level compositions of prokaryotic communities. (*b*) Order-level compositions of fungal communities. (*c*) Genus-level compositions of prokaryotic communities. (*d*) Genus-level compositions of fungal communities. The soil samples from which DNA sequence data were unavailable for either prokaryotic 16S rRNA or fungal ITS regions are indicated as blanks.

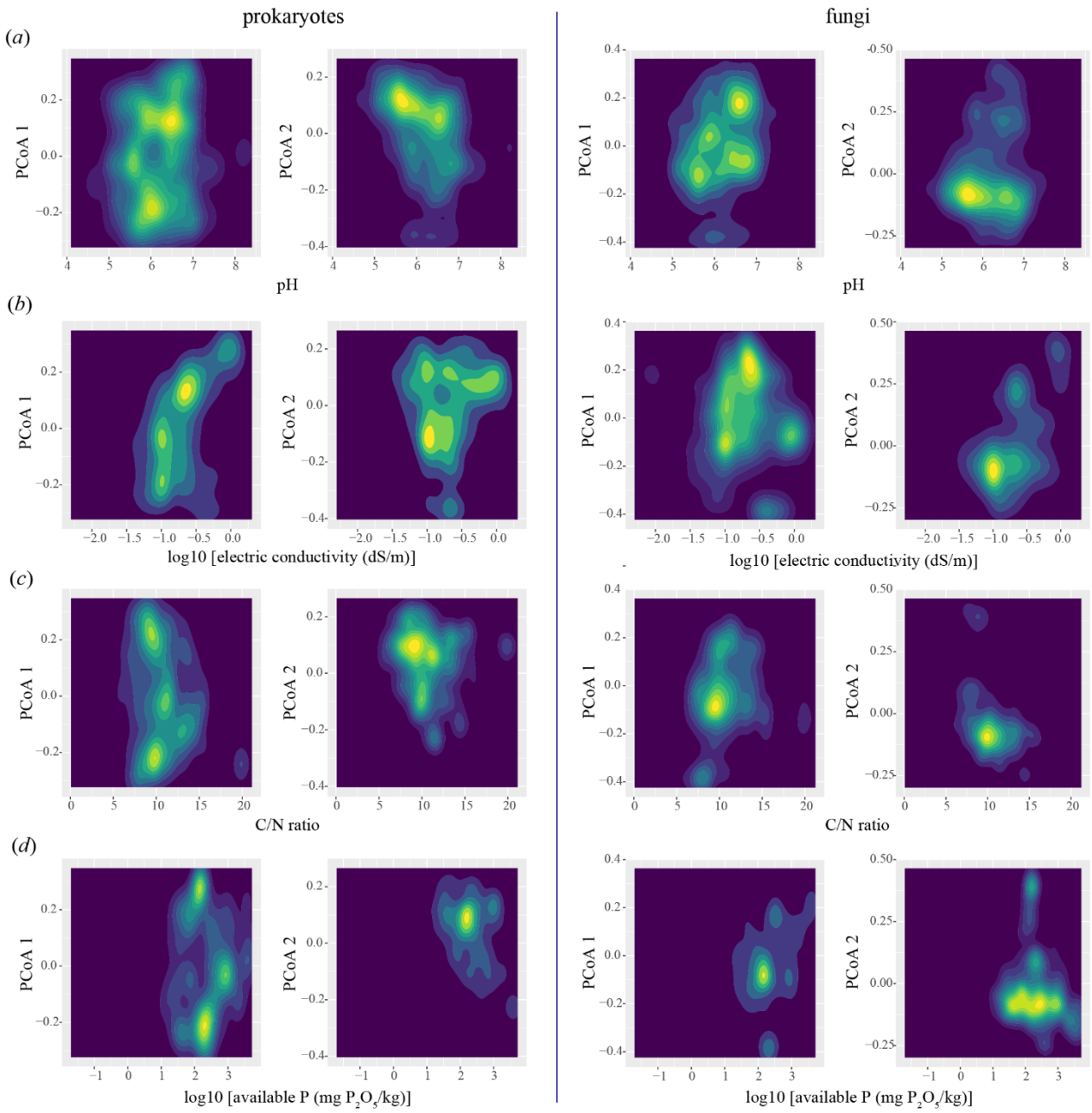

Figure S2. Community structure along environmental gradients (density plots). The scores representing prokaryotic/fungal community compositions (PCoA 1 and 2 scores) are shown along each axis of soil environmental factors. Based on the distributions of data points in figure 3, the contour surfaces were inferred. (*a*) pH. (*b*) electrical conductivity. (*c*) C/N ratio. (*d*) available phosphorous concentration.

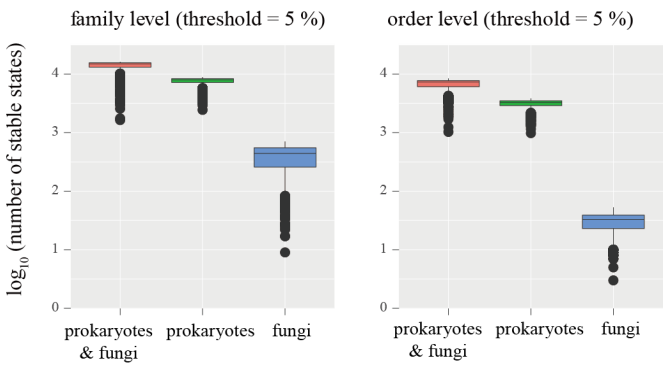

Figure S3. Number of the stable states of community structure (threshold for binarization = 5 %). The parameters of the extended pairwise maximum entropy model were adjusted to fit the model to the empirical data (see Methods). The number of stable states in the energy landscape was then estimated by using the gradient descent algorithm for 20,000 randomly generated community states within the 2^S possible community states, where S denotes the number of families or orders. In the heuristic search, sets of values of the environmental variables (soil pH, electrical conductivity, C/N ratio, and available phosphorous concentration) were drawn from those of sampling points. The boxplots show variation in the number of stable states across different sets of environmental variables. A box represents the first quartile, median and the third quartile, while bars indicate the largest/smallest values within 1.5 × inter-quartile ranges from the box: outliers were indicated as circles. The analysis was performed, respectively, for the dataset representing both prokaryotes and fungi, the prokaryote dataset, and the fungus dataset. Even for the fungus data set, the number of inferred stable states was much higher with the 5 % threshold than with the 10 % threshold at both family- and order-level data. Therefore, the discussion on the alternative stable states of microbiomes was conducted based on the 10 % threshold in the main body of this paper. Further exploration of community state space at various threshold values and taxonomic levels will enhance our understanding of alternative stable states of microbiome structure.

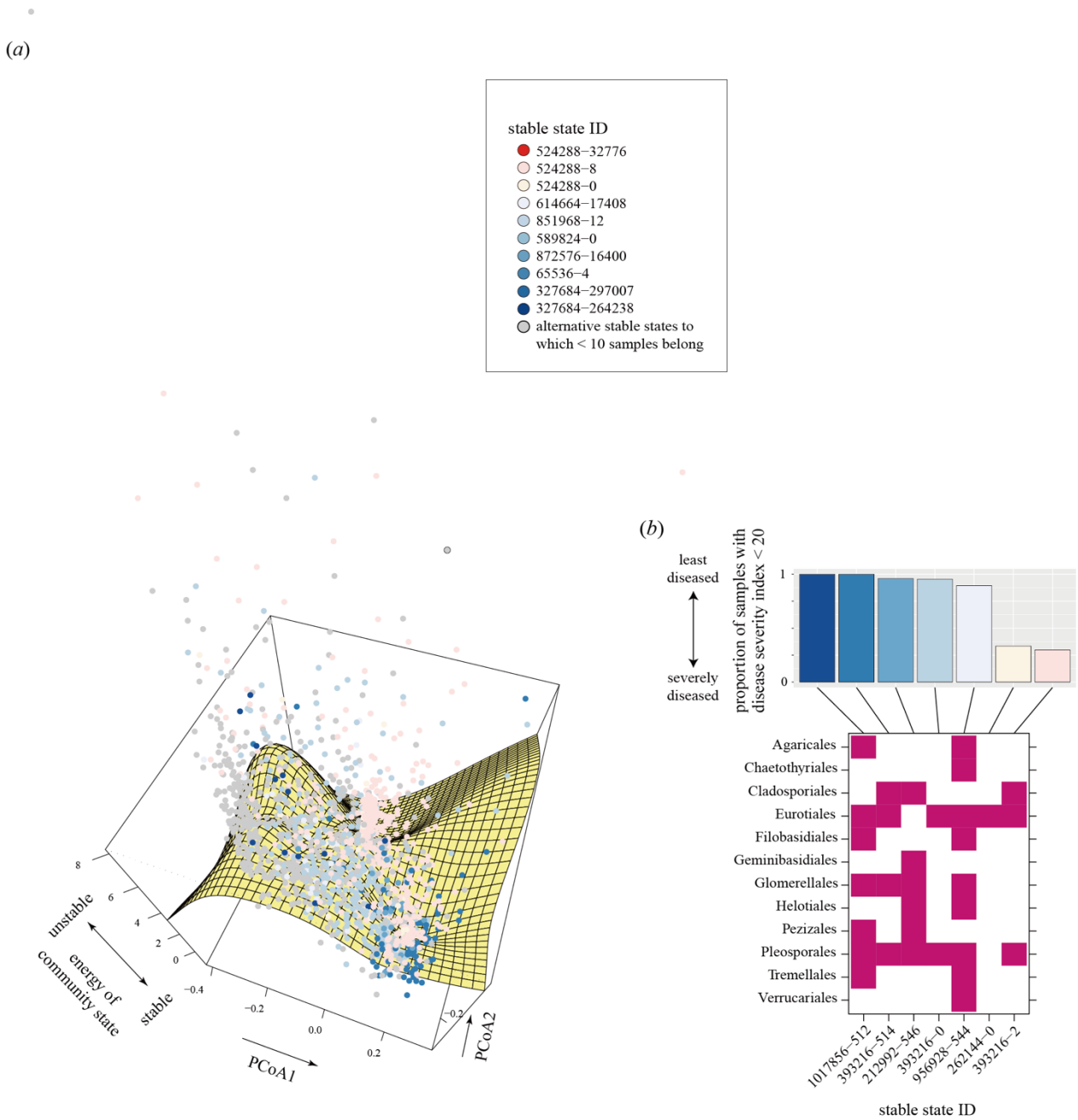

Figure S4. Alternative stable states defined for order-level fungal data (threshold for binarization = 5 %). (*a*) Energy landscape of fungal community structure. The surface of energy levels was reconstructed on the PCoA dimensions of fungal community structure based on spline smoothing. Community states with lower energy are inferred to be more stable. Data points (samples) indicated by the same color belong to the basins of the same alternative stable states, whose energy is lower than any adjacent community structure. (*b*) Definitions of the stable states. To reduce the complexity of input data to a computationally feasible range, alternative stable states were explored based on order-level community compositions (threshold for binarization = 5 %).

The seven alternative stable states representing ≥ 10 samples were compared in terms of their associations with crop disease prevalence.