Alternative stable states in soil microbiomes of agroecosystems

Hiroaki Fujita[a], Shigenobu Yoshida[b], Kenta Suzuki[c] and Hirokazu Toju[a*]

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

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

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

*Hirokazu Toju

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

Author Contributions: H.T. conceived the project; 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 Interest Statement: The authors declare that they have no competing interests.

Classification: Biological Sciences: Environmental Sciences and Ecology

Keywords: biodiversity; community stability; ecosystem functions; multistability; species coexistence

This PDF file includes:

Main Text
Figures 1 to 5

Abstract

Archaea, bacteria, and fungi in the soil are increasingly recognized as major determinants of agricultural productivity and sustainability. Although our knowledge of soil biodiversity has been greatly enhanced for the last decade, defining key characteristics of stable and highly functional microbiomes remains a major challenge. We here show that "stable states" of soil microbiomes can be defined with their disease-suppressive effects on crop plants. We integrated agricultural field metadata with microbial community analyses by targeting > 2,000 soil samples collected across a latitudinal gradient from cool-temperate to subtropical regions, profiling microbiomes constituted, in total, by 11 archaeal, 332 bacterial, and 240 fungal families. A DNA-metabarcoding-based evaluation of balance between prokaryote and fungal abundance indicated that fungus-dominated soil is more resistant to crop plant disease. We then examined whether disease-suppressive soil can be defined based on prokaryote and/or fungal taxonomic compositions. A statistical-physics-based approach for reconstructing "energy landscapes" of community structure indicated that disease-suppressive and disease-susceptible soils could be distinguished as alternative stable states of fungal communities. On the other hand, bacterial community structure could not be classified into a small number of alternative stable states, making it difficult to build simplified criteria for diagnosing disease-suppressive functions of agroecosystems with prokaryote taxonomic compositions alone. As the 22 fungal families keys to define the alternative stable states have cosmopolitan distributions, their roles and compositional balance are potentially keys to maintain agroecosystem functions and stability in diverse types of cropland soils.

Significance Statement

Managing "healthy" soil microbial communities is a prerequisite of sustainable crop production on the Earth. Based on intensive profiling of soil archaea, bacteria, and fungi in cool-temperate to subtropical regions, we found general community-level characteristics common to disease-resistant agroecosystems. Coexistence of multiple fungal lineages that can suppress devastating pests and pathogens is the key to stabilize and maximize ecosystem functions in agriculture.

Main Text

Introduction

The ongoing global-scale degradation of agroecosystems is threatening food production (1, 2). Maximizing the functions of microbial communities is a prerequisite for building bases of sustainable agriculture (3–6). Archaea, bacteria, and fungi not only promote soil nutrient cycles but also protect crop plants from pests/pathogens and abiotic stresses (7–11). Therefore, diagnosing soil microbiome states is of central interest when we try to design and manage resource-efficient and disease-tolerant agroecosystems.

A number of studies have explored specific microbes and specific combinations of microbes promoting crop plants' growth (3, 12–14). However, such microbial functions can be lost due to unexpected changes in community structure in soil microbiomes (15). In general, microbiome compositions in the wild are consequences of multiple ecological processes (13, 16–19). Selection by environmental conditions, for example, is a primary mechanism determining establishment of microbial populations (18, 20, 21).

Furthermore, competitive and facilitative interactions between microbial species are essential factors governing local community structure (22–24). Continually influenced by these mechanisms and factors, local microbial communities differ in their functions and stability (14, 25–27). Thus, understanding of not only functions but also community-level stability is essential for maximizing the benefits provided by microbiomes (12, 22, 28). Stability in soil or plant-associated microbiomes have started to be examined in studies based on laboratory experiments (29, 30). Nonetheless, we still have limited knowledge of soil microbiome stability in wild ecosystems.

In theoretical ecology, stability of community states (taxonomic or species compositions) is often discussed in the framework of stability landscapes (31–33). On the landscape representing stability/instability of community structure (*SI Appendix*; Fig. S1), stable community states are split by unstable community states (31–34). If those "alternative stable states" (31, 34, 35) [see also (36, 37)] differ in their community-level biological functions, stable and highly functional community states can be explored within the stability landscapes. Based on a mathematical approach developed in statistical physics (38, 39), it is now possible to reconstruct "energy landscapes", which represent structure of stability landscapes, from empirical datasets of ecological communities (32). Although hundreds or thousands of community compositional data points are required to apply the statistical approach (32), such energy landscape analyses will allow us to define key features of stable and highly functional microbiome states out of numerous possible combinations of microbes.

We here perform systematic profiling of cropland microbiome structure across the Japanese Archipelago in order to define general characteristics of disease-suppressive and disease-susceptible soils (40–43) based on theoretical ecology. We integrate DNA metabarcoding results of prokaryote and fungal communities with environmental meta-datasets including > 2,000 soil samples collected from the croplands of 19 plant species. With the bird's-eye view of possible microbiome structure, we map alternative stable states within the reconstructed energy landscape of microbiome structure. We then test the hypothesis that disease-suppressive and disease-susceptible states of soil microbiomes are defined with compositions of cosmopolitan microbial taxa. Overall, this study shows how we can reduce the complexity of microbiome data to build simplified diagnostic criteria for exploring optimal soil microbiomes with high controllability and agricultural functionality.

Results

Diversity of microbiome structure

We compiled the field metadata of 2,903 soil samples collected in the research projects of National Agricultural and Food Research Organization (NARO). The bulk soil of farmlands was sampled from subtropical to cool-temperate regions (26.1–42.8 °N) across the Japanese Archipelago, targeting 19 crop plant species (Fig. 1a) from 2006 to 2014. Most of the croplands were managed with conventional agricultural practices (characterized by intensive tillage and chemical fertilizer/pesticide application), while some were experimentally controlled as organic agricultural fields. The metadata included the information of chemical [e.g., pH, electrical conductivity, carbon/nitrogen (C/N) ratio, and available phosphorous concentration], physical (e.g., soil taxonomy), and biological (e.g., crop disease level) properties, providing an invaluable platform for profiling and categorizing ecosystem states of cropland soil.

To integrate the metadata with the information of microbial community structure, we performed DNA metabarcoding analyses of both prokaryotes (archaea and bacteria) and fungi. After a series of quality filtering, prokaryote and fungal community data were obtained from 2,534 and 2,384 samples, respectively. In total, 579 archaeal amplicon sequence variants (ASVs) representing 14 genera (11 families), 26,640 bacterial ASVs representing 875 genera (332 families), and 6,306 fungal ASVs representing 573 genera (240 families) were detected (Fig. 1B; *SI Appendix*: Fig. S2).

The prokaryote communities did not have dominant taxa at the genus and family levels, although the archaeal family Nitrososphaeraceae had the highest abundance at some research sites (Fig. 1B). In contrast, the fungal communities were dominated by fungi in the families Mortierellaceae, Chetomiaceae, and Nectriaceae (Fig. 1B). A reference database profiling of fungal functional groups suggested that the fungal communities were dominated by plant pathogenic and soil saprotrophic fungi (Fig. 1C) as characterized by the dominance of *Fusarium* and *Mortierella* at the genus level (*SI Appendix*: Fig. S2). Meanwhile, mycoparasitic fungi had exceptionally high proportions at some research sites, as represented by the dominance of *Trichoderma* (Hypocreaceae) at those sites (Fig. 1B; *SI Appendix*: Fig. S2).

The prokaryote and fungal communities differed in basic biodiversity patterns. In terms of α-diversity at the family level, prokaryote communities showed higher Shannon's entropy than fungal communities in all but one of the research sites (Fig. 1D). In contrast, variability of community structure among samples, i.e., Bray-Curtis β-diversity, was higher in fungi than in prokaryotes in most study sites (Fig. 1E). Likewise, fungi showed higher levels of community structural difference between research sites than prokaryotes (Welch's test; $t = -6.6$, df = 168.9, $P < 0.0001$; Fig. 1F). These results indicate that diverse taxonomic groups of archaea and bacteria are present without dominance by particular taxa, while fungal community structure is spatially dynamic both within and across agroecosystems.

Compiling the metadata of edaphic factors, we found that variation in the community structure of prokaryotes and fungi was significantly explained by crop plant identity and soil taxonomy as well as by soil chemical properties such as pH, electrical conductivity, C/N ratio, and available phosphorous concentration (Fig. 2A-B; *SI Appendix*: Figs. S3-4; Table S1). Meanwhile, both prokaryote and fungal community structure was significantly associated with the severity of crop disease (Fig. 2C-D). Thus, the dataset as a whole offers an opportunity for understanding mechanisms spanning physical, chemical, and biological processes in agroecosystems.

Crop disease and microbiome structure

We then focused on factors determining disease-suppressive and disease-susceptible properties of agroecosystem soil. In addition to chemical environmental factors (e.g., soil pH and electrical conductivity), whose associations with agricultural productivity have been routinely examined (44, 45), we explored biological factors critical to crop-plant disease level. The α-diversity of prokaryotes and/or fungi, for example, is an index that can be estimated with DNA-based profiling of soil biological communities. In further reconstructing biological factors in a high-throughput manner, we focused on the ratio of prokaryote abundance to fungal abundance. As the classic chromatography-based method for quantifying phospholipid fatty acids specific to bacteria or fungi (46) could require impractically long time for processing > 2,000 samples, we developed an alternative approach for estimating the ratio of prokaryote abundance to fungal abundance (hereafter, prokaryote/fungus ratio) based on DNA metabarcoding. Specifically, the relative concentrations of prokaryote 16S rRNA gene sequences to fungal ITS gene sequences were calculated based

150 on a quantitative amplicon sequencing analysis (Fig. 3A; see Materials and Methods for details). We then
151 constructed statistical models comparing effects on crop-plant disease level among various types of
152 chemical/biological factors.

153 A series of analysis indicated prokaryote/fungus ratio had strong impacts on crop-plant disease level.
154 While soil chemistry (soil pH) had marginally significant effects on the proportion of diseased crop plants, the
155 prokaryote/fungus ratio showed much stronger associations with crop plant disease level (Model 1 in Fig.
156 3B). Specifically, crop plant disease was more likely to occur when prokaryotes were abundant relative to
157 fungi (Fig. 3C). The effects of prokaryote/fungus ratio on plant disease level were conspicuous even when
158 they were compared with effects of prokaryote and fungal α-diversity (Model 2 in Fig. 3B). The threshold
159 values over which more than 50 % of plants showed disease levels 2–5 were estimated to be around 0.7 in a
160 log-ratio unit (Fig. 3C), although the values potentially vary depending on various factors such as crop plants,
161 geography, and climatic conditions.

162

163 **Alternative stables states**

164 We next reconstructed alternative stable states (31, 34, 37) of microbiomes in order to define community
165 structure of disease-suppressive and disease-susceptible soils. In an energy landscape analysis of statistical
166 physics (32), we modeled how microbial community compositions (community states) depend on soil
167 chemical factors and interactions between microbes. The modeling approach allowed us to calculate the
168 "energy" of each community state. We then estimated the shape of an energy landscape, on which lower
169 energy represents more stable community states (*SI Appendix*; Fig. S1). Stable states were then defined as
170 the community compositions whose energy is lower than all possible adjacent community compositions (32).

171 The results of energy landscape reconstruction fundamentally differed between prokaryotes and
172 fungi. Even if community states were defined in a binary format at the family or order level to reduce the
173 complexity of the input data (i.e., 2^S possible community states, where S denotes the number of families; see
174 Materials and Methods), more than 10^3 stable states were inferred to exist for the prokaryote dataset (Fig.
175 4). This result suggests that ecological processes of soil prokaryote communities are described as shifts
176 between many equally possible compositions and that categorization of ecosystem functions based on
177 prokaryote community structure is challenging. In contrast, the energy landscape of fungal community
178 structure was represented by much smaller number of stable states (Fig. 4). While the estimated number of
179 stable states varied depending on threshold values in the binarization of community structural data, fungal
180 communities consistently showed lower number of stable states than prokaryote communities (*SI Appendix*:
181 Fig. S5).

182 We then filtered the alternative stable states of fungal communities based on the "depth of basins" (*SI*
183 *Appendix*: Fig. S1) and the number of samples belonging to the basins of the stable states (32), obtaining 10
184 major stable states representing the energy landscape (Fig. 5A; see also *SI Appendix*: Fig. S6). By integrating
185 the information of the energy landscape (Fig. 5A) with that of crop-plant disease level (Fig. 5B), we ranked
186 the stable states in terms of possible disease-suppressive functions (Fig. 5C). We then found that all soil
187 samples belonging to the top-three stable states had associations with the lowest crop-disease level, while
188 92 % of samples belonging to the lowest-rank stable state showed higher disease levels (Fig. 5C).

189 The 10 stable states were distinguished in light of balance among 22 fungal families with
190 cosmopolitan distributions (Fig. 5C). Hypocreaceae, which embraced well-known mycoparasitic taxa [e.g.,

191 *Trichoderma* (47)], appeared at high proportions in stable states with potential disease-suppressive
192 functions. In contrast, Aspergillaceae, which included both plant pathogens and saprotrophs (48), were
193 relatively abundant in stable states with high disease potential. In addition to these basic patterns, several
194 families were keys to define disease-suppressive or disease-susceptible soils. For example, high abundance
195 of Didymellaceae, Lasiosphaeriaceae, Mrakiaceae, Orbilliaceae, Piptocephalidaceae, and
196 Plectosphaerellaceae characterized the two stable states with the highest performance (Fig. 3D). Among the
197 families, Orbiliaceae, Piptocephalidaceae, and Plectosphaerellaceae have been known to include
198 nematophagous and mycoparasitic taxa (49–51), presumably working as suppressors of plant-parasitic
199 nematodes and fungi in the soil.

200

201 **Discussion**

202 By integrating large datasets of agricultural field properties and DNA metabarcoding, we found that
203 seemingly diverse microbiome structure could be categorized into several stable states varying in
204 ecosystem-level properties. The analysis of energy landscapes suggested that prokaryote communities had
205 numerous local compositional equilibria or quasi-equilibria (Fig. 4), presumably due to their rapid and
206 continuous responses to changes in environmental conditions. In contrast, ecological processes of fungal
207 communities could be interpreted as shifts between a small number of alternative stable states (Fig. 5).
208 Basically, fungi can form complex hyphal structures and their individual lifespan is much longer than most
209 archaea and bacteria. Because of these biological properties, structure and functions of fungal communities
210 are more likely to be influenced by legacy effects of past ecological processes (7, 37, 52, 53) than those of
211 prokaryotes. Consequently, fungal taxonomic compositions can give essential insights into current status of
212 agroecosystem functions, while prokaryote community structure is too flexible to be used for building
213 simplified diagnostic criteria for soil microbiome functions.

214 These insights into alternative stable states add an important dimension of discussion on the use of
215 microbiomes in agriculture. Beyond investigations on single species/strains of microbes, microbiome studies
216 have explored sets of microbes that collectively maximize biological functions (3, 12–14). In particular,
217 experimental studies on "synthetic" communities have reorganized our understanding of microbiome
218 functions (12–14). Nonetheless, such microbial functions cannot be realized in real agroecosystems if the
219 synthesized or designed microbiome compositions are vulnerable to biotic and abiotic environmental
220 changes in the wild (15). Thus, in addition to functional properties, compositional stability is the key to
221 manage microbiomes in agroecosystems (12, 22, 28). The finding that the number of alternative stable
222 states is much fewer in fungi than in prokaryotes (Fig. 4) highlights the possibility that agroecosystem
223 functions are managed by keeping fungal community structure at targeted stable states with positive
224 impacts on crop production. The fact that prokaryote communities could not be classified into a small
225 number of alternative stable states does not indicate lack of their roles in crop production. Bacteria and
226 archaea play fundamental roles in carbon- and nitrogen-cycles of soil ecosystems (54, 55). The point is that
227 information of fungal taxonomic compositions is indispensable for adaptively managing microbiome
228 structure in agricultural practices.

229 In our analysis across the Japanese Archipelago, 22 out of the 240 fungal families detected were keys
230 to distinguish disease-suppressive and disease-susceptible soil microbiomes (Fig. 5C). Among them,
231 Hypocreaceae, Orbiliaceae, Piptocephalidaceae, and Plectosphaerellaceae are of particular interest because
232 they include many species utilized as biological control agents for suppressing plant pathogenic fungi and/or
233 plant parasitic nematodes (47, 49–51). Fungal pathogens and cyst/root-knot nematodes have continued to
234 have devastating impacts on global crop production (43, 56, 57). 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 showed that plant disease could be suppressed under coexistence of multiple nematophagous and mycoparasitic fungal taxa (Fig. 5). Therefore, shifting fungal community structure towards the stable states represented by the nematophagous/mycoparasitic families will provide more comprehensive solutions than choosing single biological control agents with best disease-suppressive functions.

Although we focused on macroscopic structure of microbial communities by performing energy landscape analysis at the family and order levels, analyses at lower taxonomic levels will enhance our understanding of microbiome stability. Positive or adverse impacts on plants often differ substantially among microbial species or strains (58). Thus, analyses at the level of ASVs or operational taxonomic units (OTUs) are necessary if we try to find "keystone" microbial species (59, 60) at each local ecosystem. Meanwhile, such knowledge of specific species or strains are not utilized in other climatic or geographic regions in which the focal species/strains are absent (19). Thus, for the purpose of building broadly applicable criteria of stable and functional microbiomes, energy landscape analyses of taxonomic compositions, rather than those of species- or strain-level compositions, are advocated. 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 Japanese Archipelago can be used to categorize disease-suppressive and disease-susceptible microbiomes in other regions on the globe.

Although we herein profiled alternative stable states of microbiomes by focusing mainly on conventional agricultural fields, soil microbiomes with higher functions and stability are potentially explored in the framework of conservation agriculture (61) or regenerative agriculture (62). Minimum tillage (63), for example, can decrease physical damage to fungal hyphae (64), thereby increasing plants' nutrient uptake through mycorrhizal hyphal networks and suppressing pests/pathogens by the functions of entomophagous, nematophagous, and mycoparasitic fungi. In this respect, the fact that crop disease frequency decreased towards the range of dominance by fungi over prokaryotes (Fig. 3) is of particular interest. We also found that the prokaryote/fungus ratio varied considerably among soil samples with similar prokaryote/fungal community compositions (Fig. 3A). Thus, analyses of balance between prokaryote and fungal abundance, in line with energy landscape analyses of community states, will help us monitor and control fungal biomes for further exploration of sustainable crop production.

We herein explored stable microbial community structure of disease-suppressive soil based on an intensive profiling of farmland soil across the Japanese Archipelago. Although fungi are often dismissed in studies on soil microbiomes, analyses of fungal biomes give essential insights into agroecosystem states (7, 43). For further understanding relationship among structure, stability, and functions of soil microbiomes, it is necessary to incorporate additional measures of microbiome performance (e.g., crop yield and nitrogen utilization efficiency) into community ecological analyses. Moreover, developing ways to shift microbiome structure from unfavorable to favorable stable states is a crucial challenge for controlling agroecosystem functions with minimum anthropogenic intervention. Bird's-eye views of stable microbiome compositions will provide platforms for integrating ecology, genomics, and microbiology towards the realization of "nature-positive" world (65).

Materials and Methods

278 **Soil samples and metadata**

279 Over research projects of National Agricultural and Food Research Organization (NARO), which were carried
280 out through five national research programs funded by Ministry of Agriculture, Forestry and Fisheries, 2,903
281 rhizosphere/bulk soil samples were collected from conventional agricultural fields across the Japanese
282 Archipelago from January 23, 2006, to July 28, 2014 (*SI Appendix*: Dataset S1). When the latitude and
283 longitude of the sampling positions were round to one decimal place, 42 research sites were distinguished.
284 Across the metadata of the 2,903 samples, the information of 19 crop plants, 34 former crop plants
285 (including "no crop"), 13 soil taxonomic groups (e.g., "Andosol"), 60 experimental/research purposes (e.g.,
286 "soil comparison between organic and conventional management") was described. Likewise, the metadata
287 included the information of dry soil pH, electrical conductivity, carbon/nitrogen (C/N) ratio, and available
288 phosphorous concentration from, 2,830, 2,610, 2,346, and 2,249 samples, respectively. In addition, the
289 information of the severity of crop plant disease was available for 1,472 samples (tomato, 637 samples;
290 Chinese cabbage, 336 samples; eggplant, 202 samples; celery, 97 samples; Broccoli, 96 samples, etc.). The
291 values of the proportion of diseased plants or disease severity index was normalized within the ranges from
292 0 to 100, and they were then categorized into five levels (level 1, 0-20; level 2, 20-40; level3; 40-60; level 4,
293 60-80; level 5, 80-100). The plant pathogens examined in the disease-level evaluation were *Colletotrichum*
294 *gloeosporioides* on the strawberry, *Fusarium oxysporum* on the celery, the lettuce, the strawberry, and the
295 tomato, *Phytophthora sojae* on the soybean, *Plasmodiophora brassicae* on Cruciferae plants, *Pyrenochaeta*
296 *lycopersici* on the tomato, *Pythium myriotylum* on the ginger, *Ralstonia solanacearum* on the eggplant and
297 the tomato, and *Verticillium* spp. on Chinese cabbage. For continuous variables within the metadata,
298 emergent outliers (mean + 5 SD) were converted into "NA" in the data matrix used in the following statistical
299 analyses as potential measurement/recording errors. Unrealistic electrical conductivity records (> 20) were
300 converted into "NA" as well.

301 At each sampling position, five soil sub-samples collected from the upper layer (0-10 cm in depth) at
302 five points (ca. 100 g each) were mixed. The mixed soil sample (ca. 500 g) was then sieved with 2-mm mesh
303 in the field. The samples were stored at -20 °C until DNA extraction. In laboratory conditions, 0.4 g of soil
304 (fresh weight) was subjected to DNA extraction with FastDNA SPIN Kit for Soil (Q-BioGene).

305

306 **DNA amplification and sequencing**

307 Profiling of soil microbial biodiversity was performed by targeting archaea, bacteria, and fungi. For the
308 amplification of the 16S rRNA V4 region of archaea and bacteria (prokaryotes), the set of the forward primer
309 515f (66) and the reverse primer 806rB (67) were used. The primers were fused with 3–6-mer Ns for
310 improved Illumina sequencing quality (68) and Illumina sequencing primers as detailed in a previous study
311 (69). PCR was performed using KOD ONE PCR Master Mix (TOYOBO, Osaka) with the temperature profile of
312 35 cycles at 98 °C for 10 seconds (denaturation), 55 °C for 5 seconds (annealing of primers), and 68 °C for 30
313 seconds (extension), and a final extension at 68 °C for 2 minutes. The ramp rate through the thermal cycles
314 was set to 1 °C/sec to prevent the generation of chimeric sequences (70). In the PCR, we added five artificial
315 DNA sequence variants with different concentrations (i.e., standard DNA gradients; 1.0×10^{-4}, 5.0×10^{-5}, 2.0
316 $\times 10^{-5}$, 1.0×10^{-5}, and 5.0×10^{-6} nM; *SI Appendix*: Table S2) to the PCR master mix solution in order to calibrate
317 16S rRNA gene copy numbers in the template DNA samples as detailed elsewhere (30, 71, 72). Likewise, the
318 internal transcribed spacer 1 (ITS1) region of fungi was amplified using the set of the forward primer
319 ITS1F_KYO1 and the reverse primer ITS2_KYO2 (73). PCR was performed using the Illumina-sequencing

320 fusion primer design mentioned above with the temperature profile of 35 cycles at 98 °C for 10 seconds, 53
321 °C for 5 s seconds, and 68 °C for 5 seconds, and a final extension at 68 °C for 2 minutes (ramp rate = 1
322 °C/sec). Newly designed artificial sequence variants (1.0×10^{-5}, 7.0×10^{-6}, 5.0×10^{-6}, 2.0×10^{-6}, and 1.0×10^{-6}
323 nM; *SI Appendix*: Table S2) were added to the PCR master mix as standard DNA gradients for the calibration
324 of the ITS sequence concentrations in the template DNA samples.

325 The PCR products of the prokaryote 16S rRNA and fungal ITS1 regions were respectively subjected to
326 the additional PCR step for linking Illumina sequencing adaptors and 8-mer sample identifier indexes (74)
327 with the amplicons as detailed elsewhere (69). The temperature profile in the PCR was 8 cycles at 98 °C for
328 10 seconds, 55 °C for 5 seconds, and 68 °C for 5 seconds, and a final extension at 68 °C for 2 minutes. The
329 PCR products were then pooled for each of the 16S rRNA and fungal ITS1 regions after a
330 purification/equalization process with the AMPureXP Kit (Beckman Coulter, Inc., Brea). Primer dimers, which
331 were shorter than 200 bp, were removed from the pooled library by supplemental purification with
332 AMpureXP: the ratio of AMPureXP reagent to the pooled library was set to 0.8 (v/v) in this process. The
333 sequencing libraries of the two regions were processed in an Illumina MiSeq sequencer (10% PhiX spike-in).
334 Because the quality of forward sequences is generally higher than that of reverse sequences in Illumina
335 sequencing, we optimized the MiSeq run setting in order to use only forward sequences. Specifically, the run
336 length was set 271 forward (R1) and 31 reverse (R4) cycles to enhance forward sequencing data: the reverse
337 sequences were used only for discriminating between prokaryote 16S and fungal ITS1 sequences in the
338 following bioinformatic pipeline.

339

340 **Bioinformatics**

341 In total, 23,573,405 sequencing reads were obtained in the Illumina sequencing (16S rRNA, 11,647,166
342 sequencing reads; ITS, 11,926,239 sequencing reads). The raw sequencing data were converted into FASTQ
343 files using the program bcl2fastq 1.8.4 distributed by Illumina. For each of the 16S rRNA and fungal ITS1
344 regions, the output FASTQ files were demultiplexed using Claident v0.9.2022.01.26 (75). The sequencing
345 data were deposited to DNA Data Bank of Japan (DDBJ) (Bioproject accession no.: PSUB018361). The
346 removal of low-quality sequences and ASV inferences were done using DADA2 (76) v1.17.5 of R v3.6.3 (77).
347 The mean number of filtered sequencing reads obtained per sample was 3,949 and 4,075 for the prokaryote
348 and fungal datasets, respectively. Taxonomic annotation of prokaryotes was conducted based on the SILVA
349 138 SSU (Quast et al., 2013) the assignTaxonomy function of DADA2. For the taxonomic annotation of fungi,
350 the QCauto method (75) was applied to the NCBI nucleotide sequence database bundled with Claident
351 v0.9.2022.01.26. ASVs that were not assigned to the domain Archaea/Bacteria and the kingdom Fungi were
352 removed from the 16S rRNA and ITS1 datasets, respectively. For each target organismal group (prokaryotes
353 and fungi), we then obtained a sample × ASV matrix, in which a cell entry depicted the number of
354 sequencing reads of an ASV in a sample. The samples with less than 1,000 reads were discarded from the
355 matrices. The matrix was rarefied to 1,000 reads per sample for the downstream statistical pipeline. The
356 number of reads was insufficient for comprehensively profiling rare microbial species, which are often
357 targets of soil microbiome studies. However, because data matrices including numerous rare ASVs or OTUs
358 could not be subjected to the computationally intensive ecological analyses detailed below even if we used
359 supercomputers, we focused on major components of soil prokaryote and fungal biomes. In other words,
360 our purpose here was to extract major components of agroecosystem soil microbiomes across the Japanese
361 Archipelago, thereby developing simple diagnostic criteria for distinguishing disease-suppressive and
362 disease-susceptible soils.

363 In total, prokaryote and fungal community data were obtained for 2,534 and 2,384 samples,
364 respectively. The number of ASVs included in the rarefied matrices were 27,219 and 6,306 for prokaryotes
365 and fungi, respectively. For fungal ASVs, putative functional groups (e.g., "plant pathogen") were inferred
366 using the program FungalTraits (78). The estimation of DNA concentrations of the prokaryote 16S rRNA and
367 fungal ITS regions was performed, respectively, based on the calibration with the standard DNA gradients
368 (artificial DNA variants introduced to the PCR master mix solutions) as detailed elsewhere (30).

369

370 **Diversity of microbiome structure**

371 For each sample, Shannon's α-diversity estimates of prokaryotes and fungi were respectively calculated.
372 Because biological entities of ASVs or operational taxonomic units (OTUs) differ considerably between
373 prokaryotes and fungi, we compared α-diversity between the two organismal groups at the family level.
374 Likewise, mean Bray-Curtis dissimilarity (β-diversity) in family-level compositions within each of the 42
375 research sites was calculated as a measure of within-site variability of prokaryote/fungal community
376 structure. In addition, the degree of among-site community differentiation was evaluated for prokaryotes
377 and fungi. Specifically, after calculating mean family-level compositions at each research site, Bray-Curtis β-
378 diversity was calculated in all pairs of research sites. The α-diversity, within-site β-diversity, and among-site
379 β-diversity estimates were respectively compared between prokaryotes and fungi.

380 For each of the prokaryote and fungal datasets, permutational analyses of variance (PERMANOVA)
381 were performed to examine associations between family-level community compositions and variables in the
382 metadata. Two types of PERMANOVA models were constructed based on Bray-Curtis β-diversity (10,000
383 iterations). Specifically, one is constituted by categorical explanatory variables (crop plant, former crop plant,
384 soil taxonomy, research site, and sampling month), while the other included continuous explanatory
385 variables (soil pH, electrical conductivity, C/N ratio, and available phosphorous concentration, latitude, and
386 longitude).

387 To reduce the dimensions of the community composition data, a principal coordinate analysis (PCoA)
388 was performed based on the Bray-Curtis β-diversity calculated with the rarefied sample × ASV matrix of
389 prokaryotes or fungi. For each PCoA axis (axes 1 to 5) in each of the prokaryote and fungal analyses,
390 Pearson's correlation with each chemical environmental factor (soil pH, electrical conductivity, C/N ratio,
391 and available phosphorous concentration) was calculated.

392

393 **Crop disease and microbiome structure**

394 We then evaluated how community structure of prokaryotes and fungi were associated with crop disease.
395 For each of the prokaryote and fungal datasets, a generalized linear mixed model (GLMM) of crop-disease
396 level was constructed by including the PCoA axes 1 to 5 as fixed effects. Sampling month and
397 experimental/research purposes in the metadata were set as random effects. A logit-link function and
398 binomial error was assumed in the GLMM after converting the response variable into a binary format
399 [disease level 1 (= 1) vs. disease levels 2–5 (= 0)].

400 We also examined how balance between prokaryote and fungal abundance could be associated with
401 crop disease levels. Based on the estimated concentrations of prokaryote 16S rRNA and fungal ITS sequences

402 in template DNA solutions, we calculated the ratio of prokaryote DNA concentrations to fungal DNA
403 concentrations in respective samples (prokaryote/fungus ratio) as follows:

404 $$\log\left(\frac{\text{prokaryote 16S rRNA gene concentration(DNA copies/}\mu\text{L)}}{\text{fungal ITS gene concentration (DNA copies/}\mu\text{L)}}\right).$$

405 Although potential variation in DNA extraction skills of researchers might affect absolute DNA
406 concentrations in the template DNA solutions, balance between prokaryote and fungal DNA in each
407 template DNA sample could be used as a reliable measure. The DNA-metabarcoding-based approach of
408 estimating prokaryote/fungus ratio has advantage over quantitative-PCR-based approaches. Specifically, the
409 former approach allows us to eliminate effects of nonspecific PCR amplification based on DNA sequencing
410 data, while the latter is affected by "contamination" of nontarget amplicons (e.g., plastid DNA in 16S rRNA
411 sequencing and plant DNA in ITS sequencing).

412 Potential effects of the prokaryote/fungus ratio on crop-disease level were examined in a GLMM with
413 a logit-link function and binomial errors. In the GLMM, soil pH and electrical conductivity were included as
414 fixed effects in addition to the prokaryote/fungus ratio. In the GLMM, C/N ratio and available phosphorous
415 concentration were not included due to the small number of samples with both crop-disease information
416 and the chemical variables. Sampling month and experimental/research purpose information in the
417 metadata were included as random effects in the GLMM. To examine the effects of other aspects of
418 biological factors, we constructed another GLMM including the ASV-level Shannon diversity of prokaryotes,
419 that of fungi, and the prokaryote/fungus ratio. All the explanatory variables were *z*-standardized (zero-mean
420 and unit-variance) before the GLMM analysis. To visualize how the proportion of least-diseased (disease-
421 level 1) samples varied along the axis of the prokaryote/fungus ratio, we performed a supplementary
422 analysis with generalized linear models (GLMs). Specifically, for the five experimental/research purpose
423 categories from which 10 or more disease-level 1 (least-diseased) samples and 10 or more disease-level 2–5
424 (severely diseased) samples were obtained, a GLM of disease level was constructed by including the
425 prokaryote/fungus ratio as an explanatory variable.

426

427 **Alternative stables states**

428 We estimated the alternative stable states of microbial community structure based on the framework of an
429 energy landscape analysis(32, 38) (tutorials of energy landscape analyses are available at
430 https://community.wolfram.com/groups/-/m/t/2358581). Probabilities of community states $\mathrm{p}\left(\vec{\boldsymbol{\sigma}}^{(\boldsymbol{k})}\right)$ are
431 given by

432 $$P\left(\vec{\boldsymbol{\sigma}}^{(\boldsymbol{k})}|\boldsymbol{\varepsilon}\right) = \boldsymbol{e}^{-\boldsymbol{H}(\vec{\boldsymbol{\sigma}}^{(\boldsymbol{k})},\boldsymbol{\varepsilon})}/Z$$

433 $$Z = \sum_{\boldsymbol{i}=\boldsymbol{1}}^{\boldsymbol{2^S}} \boldsymbol{e}^{-\boldsymbol{H}(\vec{\boldsymbol{\sigma}}^{(\boldsymbol{i})},\boldsymbol{\varepsilon})},$$

434 where $\vec{\boldsymbol{\sigma}}^{(\boldsymbol{k})} = (\sigma_1^{(k)}, \sigma_2^{(k)}, \dots, \sigma_S^{(k)})$ is a community state vector of *k*-th community state and *S* is the total
435 number of taxa (e.g., ASVs, species, genera, or families) examined. $\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_M)$ is an array of
436 continuous values representing environmental factors (e.g., soil pH and electrical conductivity) and *M* is the
437 total number of environmental parameters. $\sigma_{\mathrm{i}}^{(k)}$ is a binary variable that represents presence (1) or absence
438 (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 $\vec{\boldsymbol{\sigma}}^{(\boldsymbol{k})}$ is given by the extended pairwise maximum entropy model:

$$H(\vec{\boldsymbol{\sigma}}^{(\boldsymbol{k})}, \boldsymbol{\varepsilon}) = -\sum_{\boldsymbol{i}=\mathbf{1}}^{\boldsymbol{S}} h_{\boldsymbol{i}} \vec{\boldsymbol{\sigma}}_{\boldsymbol{i}}^{(\boldsymbol{k})} - \sum_{\boldsymbol{j}=\mathbf{1}}^{\boldsymbol{S}} \sum_{\boldsymbol{i}=\mathbf{1}}^{\boldsymbol{M}} \boldsymbol{g}_{\boldsymbol{ij}}\, \boldsymbol{\varepsilon}_{\boldsymbol{i}}^{(\boldsymbol{k})} \boldsymbol{\sigma}_{\boldsymbol{j}}^{(\boldsymbol{k})} - \sum_{\boldsymbol{i}=\mathbf{1}}^{\boldsymbol{S}} \sum_{\boldsymbol{j}=\mathbf{1}}^{\boldsymbol{S}} \boldsymbol{J}_{\boldsymbol{ij}} \vec{\boldsymbol{\sigma}}_{\boldsymbol{i}}^{(\boldsymbol{k})} \vec{\boldsymbol{\sigma}}_{\boldsymbol{j}}^{(\boldsymbol{k})} / \mathbf{2},$$

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 $H(\vec{\boldsymbol{\sigma}}^{(\boldsymbol{k})})$, a community state having lower *H* 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 (32) 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 based on a steepest descent procedure starting from each node. As environmental variables, soil pH, electrical conductivity, C/N ratio, and available phosphorous concentration were included after normalization within the ranges from 0 to 1.

The number of stable states within energy landscapes of community structure was inferred, respectively, for three types of datasets, namely, the prokaryote community matrix, the fungal matrix, and the matrix including both prokaryotes and fungi. In addition to the energy landscape analyses based on the abovementioned 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 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 were adjusted to fit the model to the empirical data. 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.

We then reconstructed energy landscapes for the settings with which community structure could be characterized by relatively small numbers of stable states (i.e., the family-level fungal data matrix defined at the 10-% threshold and the order-level fungal data matrix defined at the 10-% threshold; Fig. 4; *SI Appendix*: Fig. S5). 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 incorporated into the basins of deeper stable states [i.e., pruning of alternative stable states (32)]. The "energy" estimates were then plotted against the PCoA axes mentioned above. Spline smoothing of the landscape was performed with optimized penalty scores using the mgcv v.1.8-40 package (79) of R.

We then showed the energy of each sample on the two-dimensional surface of the state space inferred with the above-mentioned PCoA: the horizontal axes indicated PCoA 1 and 2 scores and the vertical axis indicated the energy of target samples' community states defined at the ASV level. On the energy

landscape, samples belonging to the same alternative stable states were indicated by colors. For the alternative stable states that represented 10 or more samples, the proportions of disease-level-1 samples were calculated.

Availability of data and materials

The 16S rRNA and ITS sequencing data are available from the DNA Data Bank of Japan (DDBJ: Bioproject: PRJDB14264). The microbial community data are deposited at our GitHub repository (https://github.com/hiro-toju/Soil_Microbiome_NARO3000) [to be released after acceptance of the paper]. All the R scripts used to analyze the data are available at the GitHub repository (https://github.com/hiro-toju/Soil_Microbiome_NARO3000) [to be released after acceptance of the paper].

Acknowledgments

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

References

1. M. Springmann, *et al.*, Options for keeping the food system within environmental limits. *Nature* **562**, 519–525 (2018).

2. D. Tilman, C. Balzer, J. Hill, B. L. Befort, Global food demand and the sustainable intensification of agriculture. *Proceedings of the National Academy of Sciences* **108**, 20260–20264 (2011).

3. H. Toju, *et al.*, Core microbiomes for sustainable agroecosystems. *Nat Plants* **4**, 247–257 (2018).

4. P. E. Busby, *et al.*, Research priorities for harnessing plant microbiomes in sustainable agriculture. *PLoS Biol* **15**, e2001793 (2017).

5. R. L. Berendsen, C. M. J. Pieterse, P. A. H. M. Bakker, The rhizosphere microbiome and plant health. *Trends Plant Sci* **17**, 478–486 (2012).

6. R. Bommarco, D. Kleijn, S. G. Potts, Ecological intensification: Harnessing ecosystem services for food security. *Trends Ecol Evol* **28**, 230–238 (2013).

7. K. G. Peay, P. G. Kennedy, J. M. Talbot, Dimensions of biodiversity in the Earth mycobiome. *Nat Rev Microbiol* **14**, 434–447 (2016).

8. B. J. Baker, *et al.*, Diversity, ecology and evolution of Archaea. *Nat Microbiol* **5**, 887–900 (2020).

9. R. Hayat, S. Ali, U. Amara, R. Khalid, I. Ahmed, Soil beneficial bacteria and their role in plant growth promotion: A review. *Ann Microbiol* **60**, 579–598 (2010).

518 10. K. K. Newsham, A meta-analysis of plant responses to dark septate root endophytes. *New Phytol* **190**,
519 783–93 (2011).

520 11. M. G. A. van der Heijden, *et al.*, Mycorrhizal fungal diversity determines plant biodiversity, ecosystem
521 variability and productivity. *Nature* **396**, 69–72 (1998).

522 12. J. A. Vorholt, C. Vogel, C. I. Carlström, D. B. Müller, Establishing Causality: Opportunities of Synthetic
523 Communities for Plant Microbiome Research. *Cell Host Microbe* **22** (2017).

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

526 14. J. K. Jansson, K. S. Hofmockel, The soil microbiome — from metagenomics to metaphenomics. *Curr*
527 *Opin Microbiol* **43**, 162–168 (2018).

528 15. M. Mazzola, S. Freilich, Prospects for biological soilborne disease control: Application of indigenous
529 versus synthetic microbiomes. *Phytopathology* **107**, 256–263 (2017).

530 16. S. Jiao, Y. Yang, Y. Xu, J. Zhang, Y. Lu, Balance between community assembly processes mediates
531 species coexistence in agricultural soil microbiomes across eastern China. *ISME Journal* **14**, 202–216
532 (2020).

533 17. K. Faust, L. Lahti, D. Gonze, W. M. de Vos, J. Raes, Metagenomics meets time series analysis:
534 Unraveling microbial community dynamics. *Curr Opin Microbiol* **25**, 56–66 (2015).

535 18. L. Tedersoo, *et al.*, Global diversity and geography of soil fungi. *Science (1979)* **346**, 1256688 (2014).

536 19. H. Toju, A. S. Tanabe, H. Sato, Network hubs in root-associated fungal metacommunities. *Microbiome*
537 **6**, 116 (2018).

538 20. J. M. Talbot, *et al.*, Endemism and functional convergence across the North American soil
539 mycobiome. *Proceedings of the National Academy of Sciences* **111**, 6341–6346 (2014).

540 21. M. Bahram, *et al.*, Structure and function of the global topsoil microbiome. *Nature* **560**, 233–237
541 (2018).

542 22. K. Faust, J. Raes, Microbial interactions: From networks to models. *Nat Rev Microbiol* **10**, 538–550
543 (2012).

544 23. H. Toju, S. Yamamoto, A. S. A. S. Tanabe, T. Hayakawa, H. S. H. S. Ishii, Network modules and hubs in
545 plant-root fungal biomes. *J R Soc Interface* **13**, 20151097–20151097 (2016).

546 24. M. G. A. van der Heijden, M. Hartmann, Networking in the plant microbiome. *PLoS Biol* **14**, 1–9
547 (2016).

548 25. K. Hartman, *et al.*, Deciphering composition and function of the root microbiome of a legume plant.
549 *Microbiome* **5**, 1–13 (2017).

550 26. J. A. Vorholt, C. Vogel, C. I. Carlström, D. B. Müller, Establishing causality: opportunities of dynthetic
551 vommunities for plant microbiome tesearch. *Cell Host Microbe* **22**, 142–155 (2017).

27. L. W. Mendes, *et al.*, Soil-borne microbiome: linking diversity to function. *Microb Ecol* **70**, 255–265 (2015).

28. H. Toju, *et al.*, Scoring species for synthetic community design: network analyses of functional core microbiomes. *Front Microbiol* **11**, 1361–1361 (2020).

29. J. E. Goldford, *et al.*, Emergent simplicity in microbial community assembly. *Science (1979)* **361**, 469–474 (2018).

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

31. B. E. Beisner, D. T. Haydon, K. Cuddington, Alternative stable states in ecology. *Front Ecol Environ* **1**, 376–382 (2003).

32. K. Suzuki, S. Nakaoka, S. Fukuda, H. Masuya, Energy landscape analysis elucidates the multistability of ecological communities across environmental gradients. *Ecol Monogr* **91**, 1–21 (2021).

33. M. Scheffer, S. R. Carpenter, Catastrophic regime shifts in ecosystems: Linking theory to observation. *Trends Ecol Evol* **18**, 648–656 (2003).

34. M. Scheffer, S. Carpenter, J. a Foley, C. Folke, B. Walker, Catastrophic shifts in ecosystems. *Nature* **413**, 591–596 (2001).

35. P. J. Mumby, A. Hastings, H. J. Edwards, Thresholds and the resilience of Caribbean coral reefs. *Nature* **450**, 98–101 (2007).

36. A. Hastings, *et al.*, Transient phenomena in ecology. *Science (1979)* **361**, eaat6412 (2018).

37. T. Fukami, Historical contingency in community assembly: integrating niches, species pools, and priority effects. *Annu Rev Ecol Evol Syst* **46**, 1–23 (2015).

38. T. Watanabe, N. Masuda, F. Megumi, R. Kanai, G. Rees, Energy landscape and dynamics of brain activity during human bistable perception. *Nat Commun* **5**, 4765 (2014).

39. O. M. Becker, M. Karplus, The topology of multidimensional potential energy surfaces: Theory and application to peptide structure and kinetics. *Journal of Chemical Physics* **106**, 1495 (1997).

40. D. Schlatter, L. Kinkel, L. Thomashow, D. Weller, T. Paulitz, Disease suppressive soils: New insights from the soil microbiome. *Phytopathology* **107**, 1284–1297 (2017).

41. L. L. Kinkel, M. G. Bakker, D. C. Schlatter, A coevolutionary framework for managing disease-suppressive soils. *Annu Rev Phytopathol* **49**, 47–67 (2011).

42. R. Mendes, *et al.*, Deciphering the rhizosphere microbiome for disease-suppressive bacteria. *Science* **332**, 1097–100 (2011).

43. J. Yuan, *et al.*, Predicting disease occurrence with high accuracy based on soil macroecological patterns of Fusarium wilt. *ISME Journal* **14**, 2936–2950 (2020).

585 44. K. W. T. Goulding, Soil acidification and the importance of liming agricultural soils with particular
586 reference to the United Kingdom. *Soil Use Manag* **32**, 390–399 (2016).

587 45. D. L. Corwin, S. M. Lesch, Apparent soil electrical conductivity measurements in agriculture. *Comput*
588 *Electron Agric* **46**, 11–43 (2005).

589 46. A. Frostegård, E. Bååth, The use of phospholipid fatty acid analysis to estimate bacterial and fungal
590 biomass in soil. *Biol Fertil Soils* **22**, 59–65 (1996).

591 47. J. Poveda, *Trichoderma* as biocontrol agent against pests: New uses for a mycoparasite. *Biological*
592 *Control* **159**, 104634 (2021).

593 48. F. Tekaia, J. P. Latgé, *Aspergillus fumigatus*: Saprophyte or pathogen? *Curr Opin Microbiol* **8**, 385–392
594 (2005).

595 49. Y. Yang, E. Yang, Z. An, X. Liu, Evolution of nematode-trapping cells of predatory fungi of the
596 Orbiliaceae based on evidence from rRNA-encoding DNA and multiprotein sequences. *Proc Natl Acad*
597 *Sci U S A* **104**, 8379–8384 (2007).

598 50. K. L. Lazarus, G. L. Benny, H. M. Ho, M. E. Smith, Phylogenetic systematics of syncephalis (zoopagales,
599 zoopagomycotina), a genus of ubiquitous mycoparasites. *Mycologia* **109**, 333–349 (2017).

600 51. J. M. Bourne, B. R. Kerry, F. A. A. M. de Leij, The importance of the host plant on the interaction
601 between root-knot nematodes (*Meloidogyne* spp.) and the Nematophagous fungus, *Verticillium*
602 *chlamydosporium* Goddard. *Biocontrol Sci Technol* **6**, 539–548 (1996).

603 52. P. G. Kennedy, K. G. Peay, T. D. Bruns, Root tip competition among ectomycorrhizal fungi: are priority
604 effects a rule or an exception? *Ecology* **90**, 2098–107 (2009).

605 53. G. D. A. Werner, E. T. Kiers, Order of arrival structures arbuscular mycorrhizal colonization of plants.
606 *New Phytologist* **205**, 1515–1524 (2015).

607 54. M. M. M. Kuypers, H. K. Marchant, B. Kartal, The microbial nitrogen-cycling network. *Nat Rev*
608 *Microbiol* **16**, 263–276 (2018).

609 55. J. K. Jansson, K. S. Hofmockel, Soil microbiomes and climate change. *Nat Rev Microbiol* **18**, 35–46
610 (2020).

611 56. E. M. Onkendi, G. M. Kariuki, M. Marais, L. N. Moleleki, The threat of root-knot nematodes
612 (Meloidogyne spp.) in Africa: A review. *Plant Pathol* **63**, 727–737 (2014).

613 57. M. Dita, M. Barquero, D. Heck, E. S. G. Mizubuti, C. P. Staver, Fusarium wilt of banana: Current
614 knowledge on epidemiology and research needs toward sustainable disease management. *Front*
615 *Plant Sci* **9**, 1468 (2018).

616 58. A. Hartmann, M. Schmid, D. van Tuinen, G. Berg, Plant-driven selection of microbes. *Plant Soil* **321**,
617 235–257 (2008).

618 59. S. Banerjee, K. Schlaeppi, M. G. A. van der Heijden, Keystone taxa as drivers of microbiome structure
619 and functioning. *Nat Rev Microbiol* **16**, 567–576 (2018).

620 60. H. Toju, *et al.*, Species-rich networks and eco-evolutionary synthesis at the metacommunity level. *Nat*
621 *Ecol Evol* **1**, 0024 (2017).

622 61. A. Kassam, T. Friedrich, R. Derpsch, Global spread of conservation agriculture. *International Journal of*
623 *Environmental Studies* **76**, 29–51 (2019).

624 62. L. Schreefel, R. P. O. Schulte, I. J. M. de Boer, A. P. Schrijver, H. H. E. van Zanten, Regenerative
625 agriculture – the soil is the base. *Glob Food Sec* **26**, 100404 (2020).

626 63. Y. Li, Q. Zhang, Y. Cai, Q. Yang, S. X. Chang, Minimum tillage and residue retention increase soil
627 microbial population size and diversity: Implications for conservation tillage. *Science of the Total*
628 *Environment* **716**, 137164 (2020).

629 64. E. Verbruggen, E. Toby Kiers, Evolutionary ecology of mycorrhizal functional diversity in agricultural
630 systems. *Evol Appl* **3**, 547–560 (2010).

631 65. UNDP, SCBD, UNEP-WCMC, "Creating a nature-positive future: the contribution of protected areas
632 and other effective area-based conservation measures" (2021).

633 66. J. G. Caporaso, *et al.*, Global patterns of 16S rRNA diversity at a depth of millions of sequences per
634 sample. *Proc Natl Acad Sci U S A* **108**, 4516–4522 (2011).

635 67. A. Apprill, S. Mcnally, R. Parsons, L. Weber, Minor revision to V4 region SSU rRNA 806R gene primer
636 greatly increases detection of SAR11 bacterioplankton. *Aquatic Microbial Ecology* **75**, 129–137 (2015).

637 68. D. S. Lundberg, S. Yourstone, P. Mieczkowski, C. D. Jones, J. L. Dangl, Practical innovations for high-
638 throughput amplicon sequencing. *Nat Methods* **10**, 999–1002 (2013).

639 69. H. Toju, H. Kurokawa, T. Kenta, Factors influencing leaf- and root-associated communities of bacteria
640 and fungi across 33 plant orders in a grassland. *Front Microbiol* **10**, 1–14 (2019).

641 70. J. L. Stevens, R. L. Jackson, J. B. Olson, Slowing PCR ramp speed reduces chimera formation from
642 environmental samples. *J Microbiol Methods* **93**, 203–205 (2013).

643 71. M. Ushio, Interaction capacity as a potential driver of community diversity. *Proceedings of the Royal*
644 *Society B: Biological Sciences* **289**, 20212690 (2022).

645 72. M. Ushio, *et al.*, Quantitative monitoring of multispecies fish environmental DNA using high-
646 throughput sequencing. *Metabarcoding Metagenom* **2**, 1–15 (2018).

647 73. H. Toju, A. S. A. Tanabe, S. Yamamoto, H. Sato, High-coverage ITS primers for the DNA-based
648 identification of ascomycetes and basidiomycetes in environmental samples. *PLoS One* **7**, e40863
649 (2012).

650 74. M. Hamady, J. J. Walker, J. K. Harris, N. J. Gold, R. Knight, Error-correcting barcoded primers for
651 pyrosequencing hundreds of samples in multiplex. *Nat Methods* **5**, 235–237 (2008).

652 75. A. S. Tanabe, H. Toju, Two new computational methods for universal DNA barcoding: a benchmark
653 using barcode sequences of bacteria, archaea, animals, fungi, and land plants. *PLoS One* **8**, e76910
654 (2013).

76. B. J. Callahan, *et al.*, DADA2: High-resolution sample inference from Illumina amplicon data. *Nat Methods* **13**, 581–583 (2016).

77. R Core Team, 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/* (2020).

78. S. Põlme, *et al.*, FungalTraits: a user-friendly traits database of fungi and fungus-like stramenopiles. *Fungal Divers* **105**, 1–16 (2020).

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

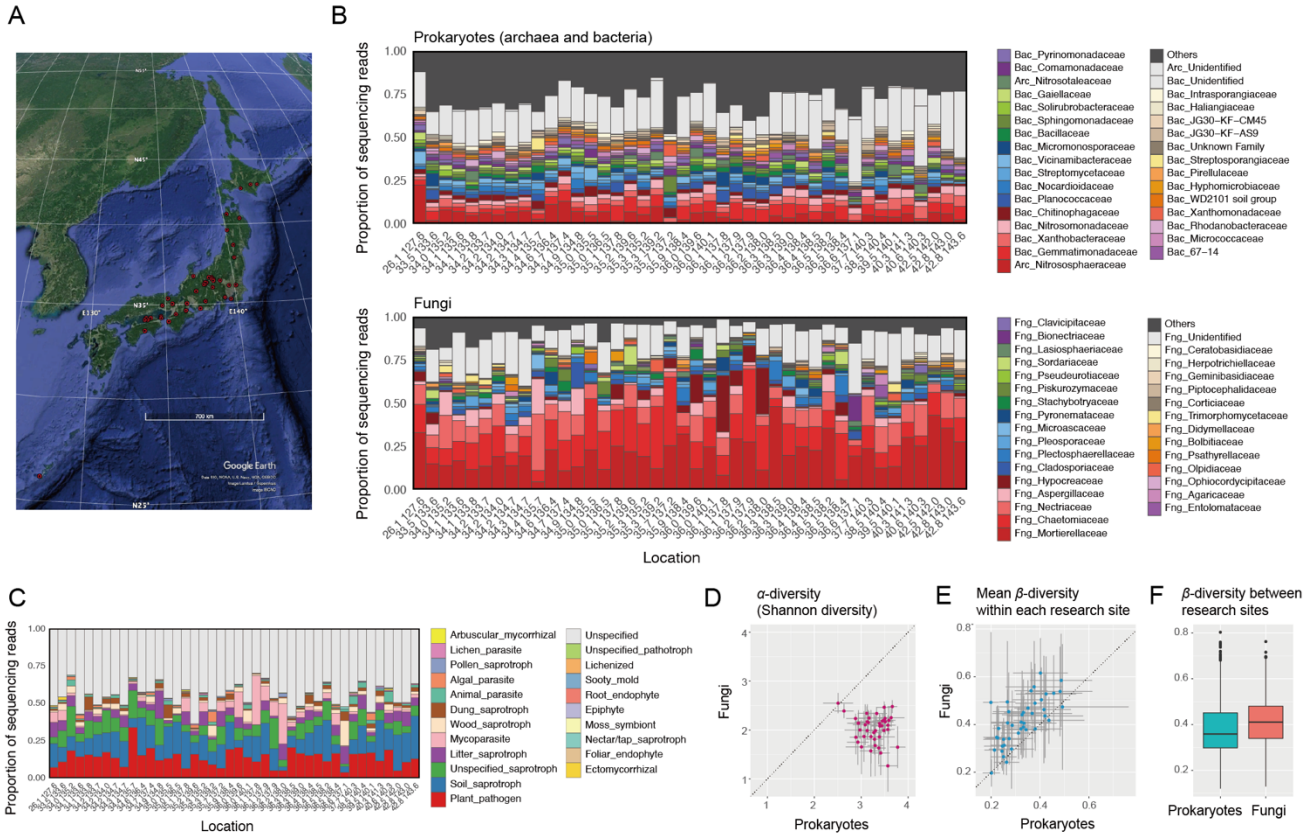
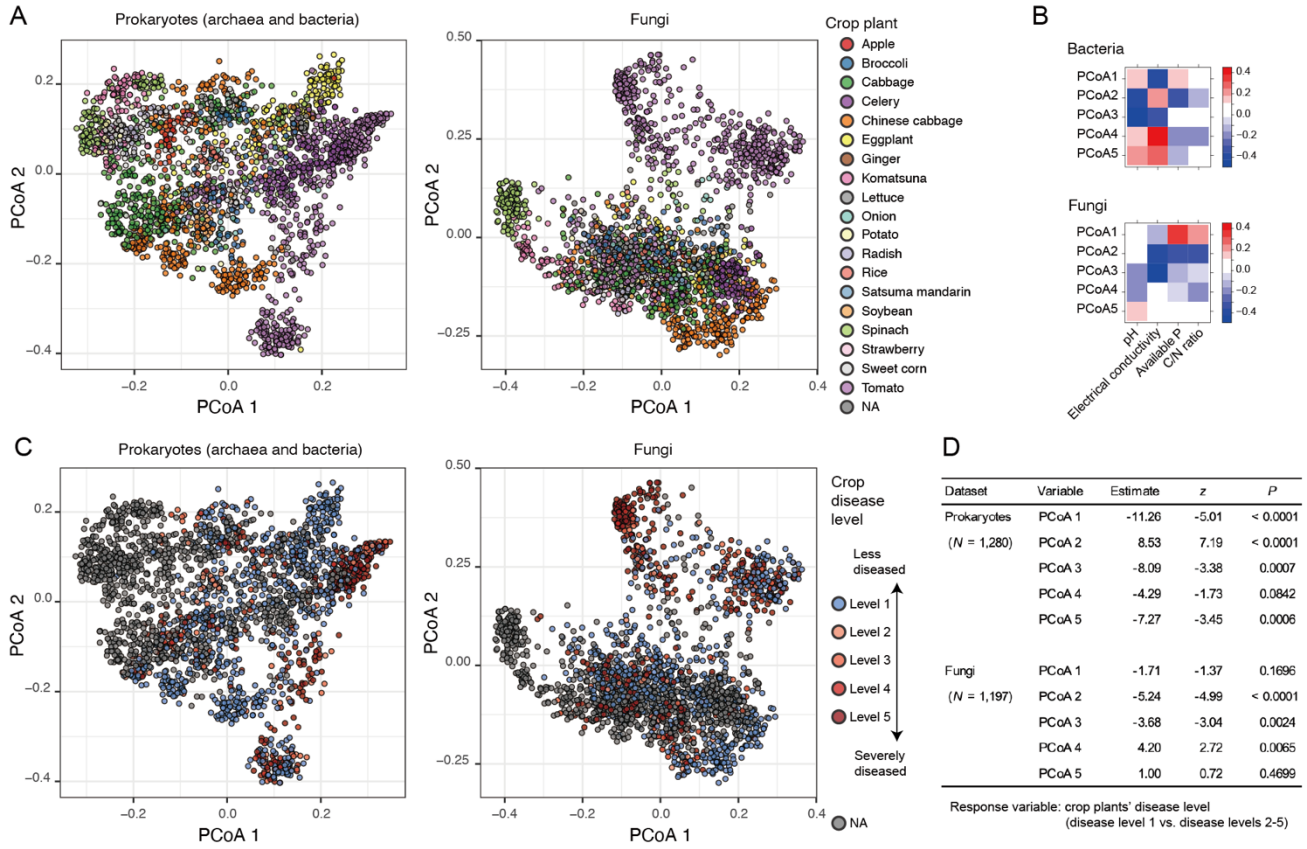


Figure 1. Diversity of agroecosystem microbiomes. (*A*) Map of research sites across the Japanese Archipelago. The 2,903 soil samples were grouped into 42 research sits were when their latitude and longitude profiles were round to one decimal place. (*B*) Taxonomic compositions of prokaryotes (archaea and bacteria; top) and fungi (bottom) at the family level. See *SI Appendix*: Fig. S2 for results at the genus, order, and class levels. (*C*) Compositions of functional groups of fungi. (*D*) Comparison of α-diversity between prokaryotes and fungi. Mean Shannon diversity at the family level was shown for each research site. Bars represent standard deviations. (*E*) Mean community dissimilarity (β-diversity) between pairs of samples within each research site. (*F*) Community dissimilarity (β-diversity) in mean community compositions between pairs of research sites.



Dataset	Variable	Estimate	z	P
Prokaryotes	PCoA 1	-11.26	-5.01	< 0.0001
(*N* = 1,280)	PCoA 2	8.53	7.19	< 0.0001
	PCoA 3	-8.09	-3.38	0.0007
	PCoA 4	-4.29	-1.73	0.0842
	PCoA 5	-7.27	-3.45	0.0006
Fungi	PCoA 1	-1.71	-1.37	0.1696
(*N* = 1,197)	PCoA 2	-5.24	-4.99	< 0.0001
	PCoA 3	-3.68	-3.04	0.0024
	PCoA 4	4.20	2.72	0.0065
	PCoA 5	1.00	0.72	0.4699

Response variable: crop plants' disease level (disease level 1 vs. disease levels 2-5)

Figure 2. Dimensions of soil microbiome structure. (*A*) Prokaryote and fungal community structure. Principal co-ordinate analyses (PCoA) were performed based on ASV-level compositional matrices respectively for the prokaryote and fungal communities. The identify of crop plants is shown by colors. See *SI Appendix*: Figs S3-4 for relationship between community structure and environmental factors. (*B*) Correlation between PCoA scores and soil environmental factors. (*C*) Microbiome structure and crop disease levels. (*D*) Relationship between prokaryote/fungal community structure on the disease level of crop plants. A GLMM of crop plants' disease level (disease level 1 vs. disease levels 2-5) with a logit-link function and binomial errors was constructed by setting principal coordinates of prokaryote/fungal community structure as explanatory variables (fixed effects). The identity of research projects and sampling month were included as random effects in the GLMM.

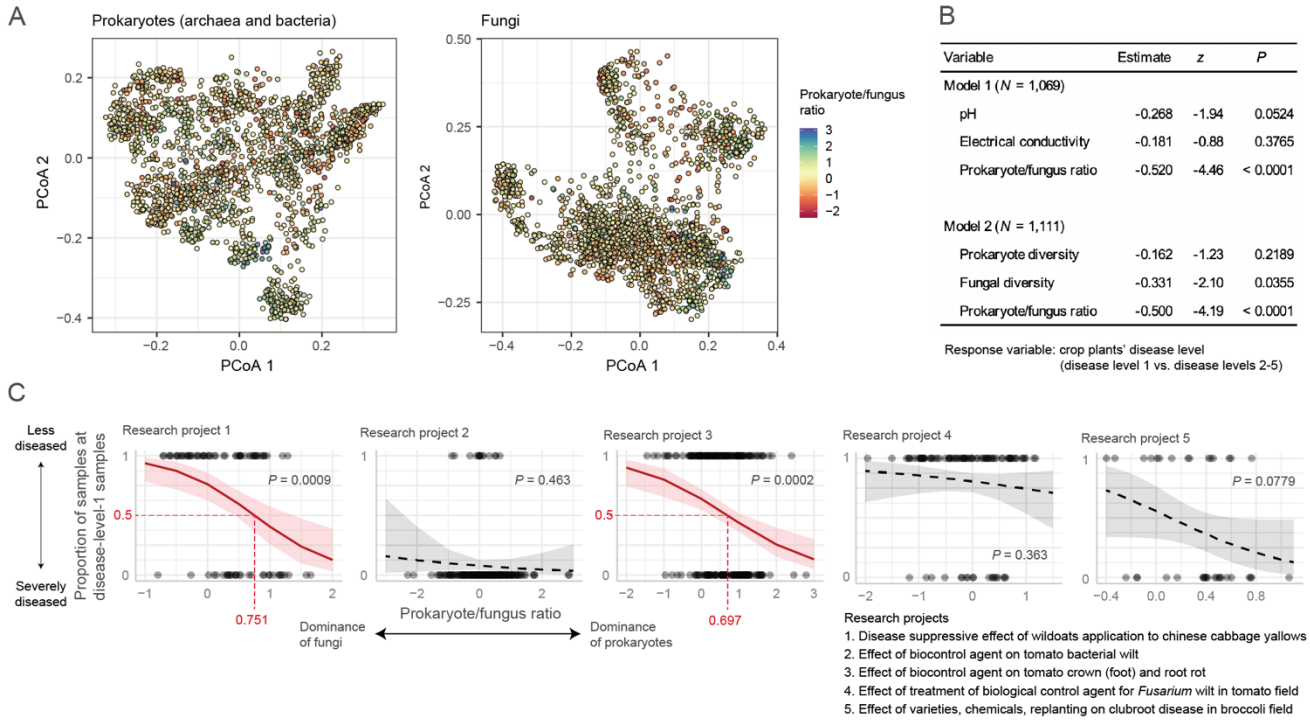


Figure 3. Balance between prokaryote and fungal abundance. (*A*) Prokaryote/fungus ratio. The ratio of prokaryote abundance to fungal abundance was evaluated based on estimated DNA concentrations [$\log\left(\frac{\text{prokaryote 16S rRNA gene concentration(DNA copies/}\mu\text{L)}}{\text{fungal ITS gene concentration (DNA copies/}\mu\text{L)}}\right)$]. The prokaryote/fungus ratio is shown for each sample on the PCoA surfaces of prokaryote and fungus community compositions. (*B*) Effects of prokaryote/fungus ratio on the disease level of crop plants. In each GLMM with a logit-link function and binomial errors, crop plants disease level [least-diseased (disease level 1) = 1, severely-diseased (disease level 2-5) = 0] was included as response variable. In addition to bacterium/fungal abundance ratio, soil environmental variables (pH and electrical conductivity; Model 1) and α-diversity estimates of prokaryotes and fungi (Model 2) were included as explanatory variables (fixed effects). The identity of research projects and sampling month were included as random effects in the GLMM. A negative estimate indicates that the proportion of least-diseased plants decreased with increasing values of a target explanatory variable. (*C*) Representative relationship between prokaryote/fungus ratio on crop disease. For each of the five research projects with enough number of disese-level-1 and disease-level-2-5 samples, a GLM was constructed. For the two research projects with significant effects of prokaryote/fungus ratio on crop disease level, the prokaryote/fungus ratios at which 50 % of samples display disease-level-1 symptoms are shown.

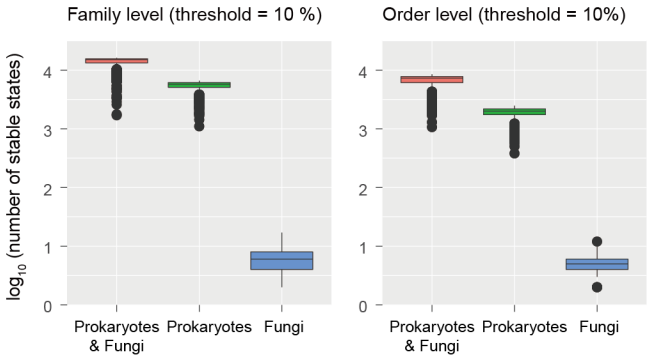


Figure 4. Number of the stable states of community structure. The parameters of the extended pairwise maximum entropy model were adjusted to fit the model to the empirical data (see Methods). The number of stable states in the energy landscape was then estimated by using the gradient descent algorithm 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 *SI Appendix*: Fig. S5 for results based on another threshold value (5 %).

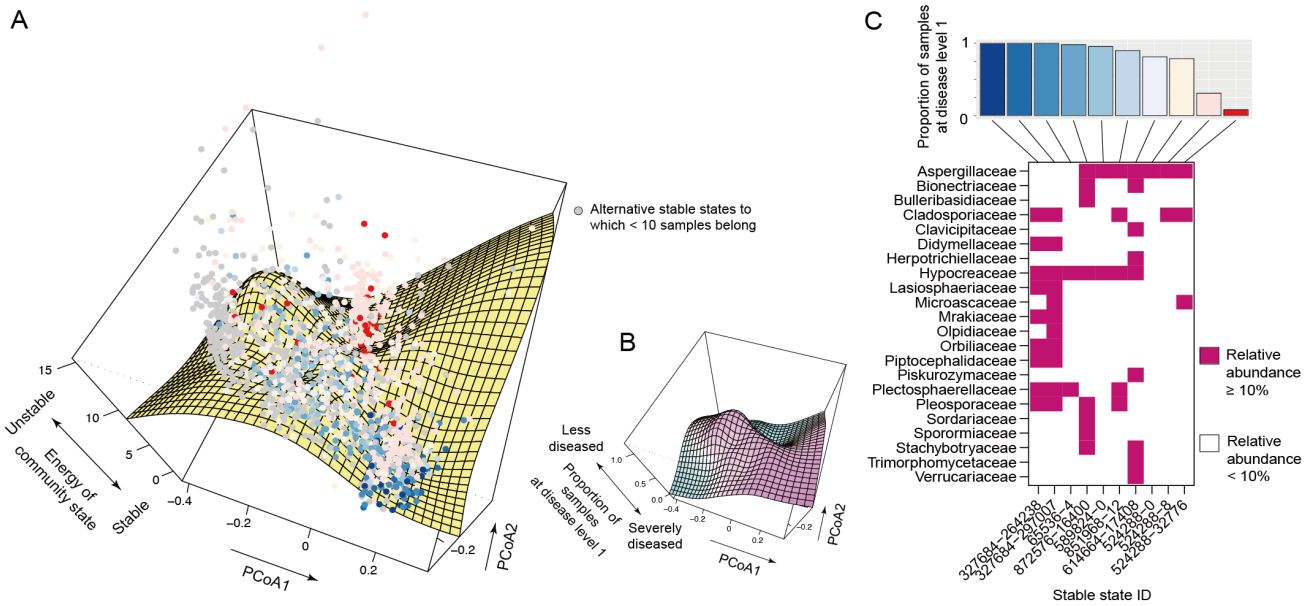


Figure 5. Alternative stable states of microbiomes. (*A*) Energy landscape of fungal community structure. The surface of energy levels was reconstructed on the PCoA dimensions (Fig. 2) of fungal community structure based on spline smoothing. Community states with lower energy are inferred to be more stable. Data points (samples) indicated by the same color belong to the basins of the same alternative stable states (defined at the family-level taxonomic compositions), whose energy is lower than any adjacent community structure. See panel *C* for details about the alternative stable states. (*B*) Surface of crop disease conditions. The area of community structure with less diseased states and lower energy (i.e., community stability; panel *A*) are expected to represent stably-manageable microbiomes with potential disease-suppressive functions. (*C*) Definitions of the 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 %). The 10 alternative stable states representing ≥ 10 samples were compared in terms of their associations with crop disease conditions.