Soil Prokaryotic and Fungal Biome Structures Associated with Crop Disease Status across the Japan Archipelago

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

[†]**Correspondence:** Hirokazu Toju (toju.hirokazu.4c@kyoto-u.ac.jp).

ABSTRACT

Archaea, bacteria, and fungi in the soil are increasingly recognized as determinants of agricultural productivity and sustainability. A crucial step for exploring soil microbiomes with high ecosystem functions is to perform statistical analyses on potential relationship between microbiome structure and functions based on comparisons of hundreds or thousands of environmental samples collected across broad geographic ranges. In this study, we integrated agricultural field metadata with microbial community analyses by targeting > 2,000 soil samples collected along a latitudinal gradient from cool-temperate to subtropical regions in Japan (26.1–42.8 °N). The data involving 579 archaeal, 26,640 bacterial, and 6,306 fungal amplicon sequencing variants detected across the fields of 19 crop plant species allowed us to conduct statistical analyses on relationship among edaphic factors, microbiome compositions, and crop disease prevalence. We then found that not only compositions of prokaryotic and fungal communities but also balance between prokaryotic and fungal abundance had statistically significant impacts on crop disease status. A network analysis suggested that the prokaryotes and fungi could be classified into several species sets (network modules), which differed substantially in associations with crop disease prevalence. Within the network of microbe-to-microbe coexistence, ammonium-oxidizing archaea and nitrite-oxidizing bacteria were inferred to play some roles in shifts between crop-disease-promotive and crop-disease-suppressive states of soil microbiomes. The bird's-eye view of soil microbiome structure will provide a basis for designing agroecosystems with high nutrient-use efficiency and disease-suppressive functions.

IMPORTANCE

Understanding how microbiome structure and functions are organized in soil ecosystems is one of the major challenges in both basic ecology and applied microbiology. Given the ongoing worldwide degradation of agroecosystems, building frameworks for exploring structural diversity and functional profiles of soil microbiomes is an essential task. Our study provides an overview of cropland microbiome states in light of potential crop-disease-suppressive functions. The large dataset allowed us to explore highly functional species sets that may be stably managed in agroecosystems. Furthermore, an analysis of network architecture highlighted species that are potentially used to cause shifts from disease-prevalent states of agroecosystems to disease-suppressive states. By extending the approach of comparative analyses towards broader geographic ranges and diverse agricultural practices, agroecosystem with maximized biological functions will be further explored.

KEYWORDS

alternative stable states, community stability, ecosystem functions, keystone species, meta-analysis, microbial interactions, network hubs, plant pathogens

RUNNING TITLE

Soil microbiome structure and crop disease prevalence

The ongoing global-scale degradation of agroecosystems is threatening food production (1, 2). Maximizing the functions of microbial communities (microbiomes) is a prerequisite for building bases of sustainable agriculture (3–7). Archaea, bacteria, and fungi in the soil drive cycles of carbon, nitrogen, and phosphorus within agroecosystems (8–12). Many of those microbes also work to promote crop plant's tolerance to drought and high temperature stresses as well as resistance to pests and pathogens (13–18). Importantly, those microbes vary greatly in their physiological impacts on crop plants (19–21). Therefore, gaining insights into soil microbiome compositions is an essential starting point for managing resource-use efficient and disease-tolerant agroecosystems.

Since the emergence of high-throughput DNA sequencing, a number of studies have revealed taxonomic compositions of prokaryotes and/or fungi in agroecosystem soil (22–24). Those studies have explored microbial species that potentially support crop plant growth and/or prevent crop plant disease (9, 16, 25, 26). Meanwhile, each of the previous studies has tended to focus on specific crop plant species in specific farm fields (27) [but see notable exceptions (15, 22)]. Therefore, generality in relationship between microbiome structure and functions remain to be examined in broader contexts [cf. global-scale analyses of soil microbiomes in natural ecosystems (28–31)]. In other words, we still have limited knowledge of general patterns and features common to soil microbiomes with high crop yield or those with least crop disease risk. Thus, statistical analyses comparing microbiome structure among diverse crop plants across broad geographic ranges (15, 22) are expected to deepen our understanding of microbial functions in agroecosystems. In particular, comparative studies of thousands of soil samples covering a wide range of latitudes will provide opportunities for finding general properties common to microbial communities with plant-growth-promoting or crop-disease-suppressive functions across diverse climatic conditions.

Large datasets of soil microbiomes will also allow us to estimate interspecific interactions between microbial species (3, 32, 33). Archaea, bacteria, and fungi in soil ecosystems potentially form entangled webs of facilitative or competitive interactions, collectively determining ecosystem-level functions such as the efficiency of nutrient cycles and the prevalence of plant pathogens (34, 35). Although statistical platforms for inferring such microbial interaction networks have been proposed (32, 36), hundreds or more of microbial community samples are required to gain reliable inferences on interactions that reproducibly occur in real ecosystems (37). Thus, datasets consisting of thousands of soil samples collected across a number of local ecosystems will provide fundamental insights into how soil ecological processes are driven by cross-kingdom interactions involving archaea, bacteria, and fungi.

In this study, we conducted a comparative analysis of agroecosystem soil microbiomes based on > 2,000 soil samples collected from subtropical to cool-temperate regions across the Japan Archipelago. Based on the amplicon sequencing dataset representing farm fields of 19 crop plant species, we profiled prokaryotic and fungal community compositions in conventional agricultural fields in Japan. By compiling the metadata of the soil samples, we then examined biotic and abiotic factors explaining diversity in the prevalence of crop disease. The soil microbiome dataset was further used to infer the structure of a microbe-to-microbe interaction network consisting of diverse archaea, bacteria, and fungi. We then identified clusters (modules)

of closely interacting microbial species within the network and examined whether such network modules differ in their positive/negative impacts on crop plants. We further explored "hub" or "core" species that were placed at the central positions within the inferred microbial interaction networks. Overall, this study provides an overview of soil microbial diversity of cropland soil across a latitudinal gradient, setting a basis for diagnosing soil ecosystem status and exploring microbial species keys to manage agroecosystem functions.

RESULTS

Diversity of agroecosystem microbiomes

We compiled the field metadata of 2,903 soil samples collected in the research projects of National Agricultural and Food Research Organization (NARO), Japan. The bulk soil of farmlands was sampled from subtropical to cool-temperate regions (26.1–42.8 °N) across the Japan Archipelago from 2006 to 2014, targeting 19 crop plant species (Fig. 1A; Data S1). 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 (Data S1) 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 a platform for profiling 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, prokaryotic and fungal community data were obtained from 2,318 and 2,186 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; Fig. S1).

The prokaryotic 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 (Fig. S1). Meanwhile, mycoparasitic fungi had exceptionally high proportions at some research sites, as represented by the dominance of *Trichoderma* (Hypocreaceae) at those sites (Fig. 1B; Fig. S1).

The prokaryotic and fungal communities differed in basic biodiversity patterns. In terms of α-diversity at the family level, prokaryotic communities showed higher Shannon's entropy than fungal communities in all but one of the research sites (Fig. 2A). 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. 2B). Likewise, fungi showed higher levels of community

139 structural difference between research sites than prokaryotes (Welch's test; t = -6.6, df = 168.9, P
140 < 0.0001; Fig. 2C).
141
142 **Microbiome structure and crop disease prevalence**
143 Compiling the metadata of edaphic factors, we found that variation in the community structure of
144 prokaryotes and fungi was significantly explained by crop plant identity and soil taxonomy as
145 well as by soil chemical properties such as pH, electrical conductivity, C/N ratio, and available
146 phosphorous concentration (Fig. 3A-B; Figs. S2-3; Table 1). Meanwhile, both prokaryotic and
147 fungal community structure was significantly associated with the severity of crop disease (Fig.
148 4A; Table 2). Thus, the dataset as a whole offers an opportunity for understanding mechanisms
149 spanning physical, chemical, and biological processes in agroecosystems.
150 We then examined potential factors determining the prevalence of crop disease in
151 agroecosystems. In addition to chemical environmental factors (e.g., soil pH and electrical
152 conductivity), whose associations with agricultural productivity have been routinely examined
153 (38, 39), we explored biological factors critical to crop-plant disease level. The α-diversity of
154 prokaryotes and/or fungi, for example, is an index that can be estimated with DNA-based
155 profiling of soil biological communities. In further reconstructing biological factors in a high-
156 throughput manner, we focused on the ratio of prokaryotic abundance to fungal abundance.
157 Specifically, by applying a calibrated DNA metabarcoding method using artificial standard DNA
158 sequences with pre-defined concentrations (40), the relative concentrations of prokaryotic 16S
159 rRNA gene sequences to fungal ITS gene sequences were calculated (Fig. 4B; see Materials and
160 Methods for details). We then constructed statistical models comparing effects on crop-plant
161 disease level among various types of chemical/biological factors.
162 A series of analyses indicated that prokaryote/fungus ratios had strong impacts on crop-
163 plant disease level. While soil chemistry (soil pH) had marginally significant effects on the
164 proportion of diseased crop plants, the prokaryote/fungus ratios showed much stronger
165 associations with crop plant disease level (Model 1 in Table 3). Specifically, crop plant disease
166 was more likely to occur when prokaryotes were abundant relative to fungi (Fig. 4C). The effects
167 of prokaryote/fungus ratio on plant disease level were conspicuous even when they were
168 compared with effects of prokaryotic and fungal α-diversity (Model 2 in Table 3). The threshold
169 value over which more than 50 % of plants showed disease levels 2–5 was estimated to be around
170 0.7 in a log-ratio unit (Fig. 4C), although the values potentially vary depending on various factors
171 such as crop plants, geography, and climatic conditions.
172
173 **Microbe-to-microbe network**
174 We then examined the network architecture of potential microbe-to-microbe interactions within
175 the soil microbiomes. As the number of possible interactions inflates with that of examined
176 microbes (i.e., X^2 pairs, where X denotes the number of microbial species/ASVs), we targeted
177 microbial ASVs that appeared in ≥ 10 % of samples in the network statistical analysis. In total,
178 288 ASVs (17 archaea, 183 bacteria, and 88 fungi) were analyzed based on the algorithm of

179 sparse inverse covariance estimation for ecological association inference (SPIEC-EASI) with the
180 Meinshausen-Bühlmann (MB) method (36).

181 The inferred network of coexistence was subdivided into several "modules", in which
182 archaeal, bacterial, and fungal ASVs sharing environmental preferences and/or those in positive
183 interactions were linked with each other (Fig. 5A). We found that the network modules differed
184 considerably in their associations with crop-plant disease level (Fig. 5B; Fig. S4; Data S2).
185 Modules 3 and 7, for example, were constituted by microbes with potentially negative impacts on
186 crop plants, while Modules 1, 2, 4, and 5 were characterized by microbes associated with least
187 disease level (Fig. 5B).

188

189 **Core species within the microbial network**

190 We next explored microbial ASVs that potentially have great impacts on community- or
191 ecosystem-scale processes based on an analysis of the microbe-to-microbe network architecture
192 (Data S2). We found that some ASVs allied to a clade of ammonium-oxidizing archaea (41) [a
193 *Nitrocosmicus* ASV (A_0004) and other Nitrososphaeraceae ASVs (A_0002, A_0013, and
194 A_0016)] were associated with low disease level of plants and they were placed at core (hub)
195 positions as indicated by the betweenness centrality (42) index (Fig. 6A). Among microbes
196 associated with severely damaged plants, an *Chaetomium* fungus (F_00022) had the highest
197 network centrality (Fig. 6A).

198 When the microbial ASVs were ranked in terms of their topological roles in interlinking
199 multiple network modules (Fig. 5), potentially ammonium-oxidizing archaea in the family
200 Nitrososphaeraceae (41) and a nitrite-oxidizing bacterium in the genus *Nitrospira* (B_00048) (43)
201 were highlighted (Fig. 6B). The analysis also indicated that a *Fusarium* fungus (F_09978) was
202 placed at a hub position interlinking network modules (Fig. 6B). Although the genus *Fusarium* is
203 known as a predominantly plant-pathogenic taxon (44, 45), this particular fungal ASV (F_09978)
204 was strongly associated (correlated) with less-diseased crop plants (Fig. 6A): note that some
205 *Fusarium* species/strains are known to work as beneficial plant symbionts (46).

206 In terms of topological roles within respective modules, a *Chaetomium* fungus associated
207 with crop plant disease (F_00022) served as a hub within Module 7 (Fig. 6B), which consisted of
208 microbes associated with crop disease (Fig. 5B). A PLTA13 bacterium in the class
209 Gammaproteobacteria (B_00179) and a C0119 bacterium in the class Ktedonobacteria (B_00160)
210 were hubs within Modules 2 and 6, respectively (Fig. 6B), which showed specificity to less-
211 diseased crop plants (Fig. 5B).

212

213 **DISCUSSION**

214 We here profiled the diversity of agroecosystem microbiome structure across a latitudinal
215 gradient from cool-temperate to subtropical regions based on the analysis of > 2,000 soil samples.
216 The integration of the microbiome datasets with agricultural field metadata allowed us to perform
217 statistical tests of possible factors determining crop disease prevalence/severity. Moreover, an
218 analysis of microbe-to-microbe interaction networks inferred sets of microbes associated with

219 crop plant health. The network analysis further highlighted core archaea, bacteria, and fungi that
220 potentially play key roles in community- or ecosystem-level processes.

221 The prokaryotic communities of the agroecosystem soil involved diverse archaeal and
222 bacterial taxa (Fig. 1B). At the family level, an ammonium-oxidizing archaeal taxon,
223 Nitrososphaeraceae (47), was the most commonly observed, suggesting their ecosystem functions
224 related to nitrogen cycling from fertilizers (e.g., ammonium sulfate). Meanwhile, the relative
225 abundance of the archaeal taxon was small (5.5 % on average) as indicated by the lack of
226 dominant taxa within the prokaryotic communities (Fig. 1A). In contrast, the fungal communities
227 were constituted predominantly by several taxa such as Mortierellaceae, Chaetomiaceae,
228 Nectriaceae, Aspergillaceae, and Hypocreaceae (Fig. 1B). Our data also indicated that fungi
229 exhibited higher β-diversity among sampling positions or local communities than prokaryotes,
230 suggesting high spatial heterogeneity of their community structure (Fig. 2B-C). Indeed, fungal
231 communities were dominated by mycoparasitic fungi [e.g., *Trichoderma* (48) in Hypocreaceae]
232 (Fig. S1A) in some research sites, while they consisted of high proportions of plant pathogenic
233 taxa [e.g., *Fusarium* (44, 45) in Nectriaceae] (Fig. S1A) in others (Fig. 1C). These results suggest
234 that agroecosystem soil microbiomes can differ greatly in their crop-disease-suppressive and
235 crop-disease-promotive properties depending on their fungal community compositions. Indeed, a
236 statistical analysis indicated significant impacts of soil fungal community structure on crop
237 disease level, although the effects of community structure on disease prevalence was significant
238 as well for prokaryotes (Table 2).

239 We further explored ecological mechanisms determining crop disease prevalence based on
240 multiple regression models. Among the biotic and abiotic factors examined, balance between
241 prokaryotic and fungal abundance showed the strongest contributions to variation in crop disease
242 level (Table 3; Fig. 4A-B). The finding that crop disease is more prevalent under higher ratios of
243 prokaryotes to fungi (Fig. 4C) is of particular ecological and agricultural interest. Due to the
244 presence of plant-pathogenic fungi that have devastating impacts on crop production (49, 50),
245 fungicide chemicals have been used to suppress the entire fungal communities (51, 52).
246 Nonetheless, a number of studies have shown that the presence of mycoparasitic fungi (e.g.,
247 *Trichoderma*) is essential for suppressing plant pathogenic fungi within agroecosystems (48, 53).
248 Moreover, the presence of nematophagous fungal taxa [e.g., Orbiliales (54)] is the key to prevent
249 crop damage by cyst and root-knot nematodes (55). Therefore, positive sides of fungus-
250 dominated microbiome states need to be further investigated in light of taxonomic and functional
251 profiles of fungal communities.

252 A network analysis then suggested that diverse soil microbes could be classified into
253 several groups differing in associations with crop disease prevalence (Fig. 5). Such information
254 of microbial network modules helps us find optimal sets of microbes that should be inspected to
255 maintain crop-disease-suppressive soil microbiomes. Because interspecific interactions
256 (involving inter-kingdom interactions) are key drivers of microbial community dynamics (3, 33),
257 insights into stably coexisting species sets provide valuable platforms for managing
258 agroecosystem microbiome functions.

259 The analysis of network architecture further allowed us to explore core species within the
260 microbial network (Fig. 6). Because the microbes highlighted with the examined network indices

occupy key positions interconnecting many other microbes (56), their increase/decrease is expected to have profound impacts on whole community processes (3, 32, 33). In particular, control or manipulation of microbes located at the central positions interlinking different network modules (37) (i.e., microbes with high among-module connectivity; Fig. 6B) may trigger drastic shifts in microbial community structure between disease-promotive and disease-suppressive states (3). Interestingly, the candidate list of such core species involved ammonium-oxidizing archaea (Nitrosospharaeraceae) (41, 47) and a nitrite-oxidizing bacterium (*Nitrospira*) (43), which potentially play crucial roles in soil nitrogen cycling (10, 11) (Fig. 6B). These microbes are prioritized targets of research because they can be used not only to manage crop disease but also to optimize the pace of nitrate supply to crop plants.

Although the dataset across a latitudinal gradient provided an opportunity for gaining bird's-eye insights into the structure and potential functions of soil microbiomes, the results should be interpreted carefully with the recognition of potential methodological shortcomings and pitfalls. First, the approach of geographic comparison *per se* does not give a firm basis for deciphering microbial community dynamics. To gain fundamental insights into microbiome dynamics, we need to perform time-series monitoring (40, 57, 58) of soil prokaryotic and fungal community compositions. Second, information of microbial communities alone does not provide comprehensive insights into agroecosystem soil states. Given that soil ecosystem processes are driven not only by microbes but also by nematodes, arthropods, earthworms, and protists (59–62), simultaneous analyses of all prokaryotic and eukaryotic taxa (63, 64) will help us infer whole webs of biological processes. Third, amplicon sequencing approaches provide only indirect inference of biological functions. With the current capacity of sequencing and bioinformatic technologies, it is hard to assemble tens of thousands of microbial genomes based on the analysis of thousands of environmental samples. Furthermore, due to the paucity of the information of fungal ecology and physiology, it remains difficult to annotate high proportions of genes within fungal genomic data. Nonetheless, with the accumulation of methodological breakthroughs, shotgun sequencing of soil microbiomes will deepen our understanding of agroecosystem processes (65–67). Fourth, in this study, full sets of metadata were not available for all the sequenced samples, inevitably decreasing the number of samples examined in some statistical modeling. Although substantial efforts had been made to profile cropland soils in the national projects in which the soil samples were collected, continuous efforts are required to gain further comprehensive insights into agroecosystem structure and functions.

The approach of the comparative microbiome analysis can be extended to some directions. For example, comparison with soil agroecosystems in lower-latitudinal or higher-latitudinal regions or meta-analyses covering multiple continents will provide further comprehensive knowledge of the diversity of microbiome structure. In addition to extensions towards broader geographic ranges, those towards diverse agroecosystem management are of particular importance. Given that our samples were collected mainly from croplands managed with conventional agricultural practices, involvement of soil samples from regenerative or conservation agricultural fields (68–71) will reorganize our understanding of relationship between microbiome compositions and functions. Further data-driven research will uncover fundamental mechanisms driving soil ecosystems, providing novel strategies for designing sustainable agriculture.

MATERIALS AND METHODS

Soil samples and metadata

Over research projects of National Agricultural and Food Research Organization (NARO), which were carried out through five national research programs funded by Ministry of Agriculture, Forestry and Fisheries, 2,903 rhizosphere/bulk soil samples were collected from conventional agricultural fields across the Japan Archipelago from January 23, 2006, to July 28, 2014 (Data S1). When the latitude and longitude of the sampling positions were round to one decimal place, 42 research sites were distinguished. Across the metadata of the 2,903 samples, the information of 19 crop plants, 34 former crop plants (including "no crop"), 13 soil taxonomic groups (e.g., "Andosol"), 60 experimental/research purposes (e.g., "soil comparison between organic and conventional management") was described. Likewise, the metadata included the information of dry soil pH, electrical conductivity, carbon/nitrogen (C/N) ratio, and available phosphorous concentration from, 2,830, 2,610, 2,346, and 2,249 samples, respectively. In addition, the information of the severity of crop plant disease was available for 1,472 samples (tomato, 637 samples; Chinese cabbage, 336 samples; eggplant, 202 samples; celery, 97 samples; Broccoli, 96 samples, etc.). The values of the proportion of diseased plants or disease severity index (72) was normalized within the ranges from 0 to 100, and they were then categorized into five levels (level 1, 0-20; level 2, 20-40; level3; 40-60; level 4, 60-80; level 5, 80-100). The plant pathogens examined in the disease-level evaluation 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. For continuous variables within the metadata, emergent outliers (mean + 5 SD) were converted into "NA" in the data matrix used in the following statistical analyses as potential measurement/recording errors. Unrealistic electrical conductivity records (> 20) were converted into "NA" as well.

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

DNA amplification and sequencing

Profiling of soil microbial biodiversity was performed by targeting archaea, bacteria, and fungi. For the amplification of the 16S rRNA V4 region of archaea and bacteria (prokaryotes), the set of the forward primer 515f and the reverse primer 806rB were used as described elsewhere (40). The primers were fused with 3–6-mer Ns for improved Illumina sequencing quality and Illumina sequencing primers. PCR was performed using KOD ONE PCR Master Mix (TOYOBO, Osaka) with the temperature profile of 35 cycles at 98 °C for 10 seconds (denaturation), 55 °C for 5 seconds (annealing of primers), and 68 °C for 30 seconds (extension), and a final extension at 68

°C for 2 minutes. The ramp rate through the thermal cycles was set to 1 °C/sec to prevent the generation of chimeric sequences. In the PCR, we added five artificial DNA sequence variants with different concentrations (i.e., standard DNA gradients; 1.0×10^{-4}, 5.0×10^{-5}, 2.0×10^{-5}, 1.0×10^{-5}, and 5.0×10^{-6} nM; Table S1) to the PCR master mix solution as detailed elsewhere (40). By comparing the number of sequencing reads between the artificial standard DNA and real prokaryotic DNA, the concentration of prokaryotic 16S rRNA genes in template DNA samples were calibrated (40).

In addition to the prokaryotic 16S rRNA region, the internal transcribed spacer 1 (ITS1) region of fungi was amplified using the set of the forward primer ITS1F_KYO1 and the reverse primer ITS2_KYO2 (73). PCR was performed using the Illumina-sequencing fusion primer design mentioned above with the temperature profile of 35 cycles at 98 °C for 10 seconds, 53 °C for 5 s seconds, and 68 °C for 5 seconds, and a final extension at 68 °C for 2 minutes (ramp rate = 1 °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} nM; Table S1) were added to the PCR master mix as standard DNA gradients for the calibration of the ITS sequence concentrations in the template DNA samples.

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

Bioinformatics

In total, 23,573,405 sequencing reads were obtained in the Illumina sequencing (16S rRNA, 11,647,166 sequencing reads; ITS, 11,926,239 sequencing reads). The raw sequencing data were converted into FASTQ files using the program bcl2fastq 1.8.4 distributed by Illumina. For each of the 16S rRNA and fungal ITS1 regions, the output FASTQ files were demultiplexed using Claident v0.9.2022.01.26 (74). The sequencing data were deposited to DNA Data Bank of Japan (DDBJ) (Bioproject accession no.: PSUB018361). The removal of low-quality sequences and ASV inferences were done using DADA2 (75) v1.17.5 of R v3.6.3. The mean number of filtered sequencing reads obtained per sample was 3,949 and 4,075 for the prokaryotic and fungal datasets, respectively. Taxonomic annotation of prokaryotes was conducted based on the SILVA

138 SSU (76) the assignTaxonomy function of DADA2. For the taxonomic annotation of fungi, the QCauto method (74) was applied to the NCBI nucleotide sequence database bundled with Claident v0.9.2022.01.26. ASVs that were not assigned to the domain Archaea/Bacteria and the kingdom Fungi were removed from the 16S rRNA and ITS1 datasets, respectively. For each target organismal group (prokaryotes and fungi), we then obtained a sample × ASV matrix, in which a cell entry depicted the number of sequencing reads of an ASV in a sample. The samples with less than 1,000 reads were discarded from the matrices. The matrix was rarefied to 1,000 reads per sample for the downstream statistical pipeline. The number of reads was insufficient for comprehensively profiling rare microbial species, which are often targets of soil microbiome studies. However, because data matrices including numerous rare ASVs or OTUs could not be subjected to the computationally intensive ecological analyses detailed below even if we used supercomputers, we focused on major components of soil prokaryotic and fungal biomes. In other words, our purpose here was to extract major components of agroecosystem soil microbiomes across the Japan Archipelago, thereby finding core microbiome properties associated with disease-suppressive and disease-susceptible agroecosystems.

In total, prokaryotic and fungal community data were obtained for 2,318 and 2,186 samples, respectively. The number of ASVs included in the rarefied matrices were 27,219 and 6,306 for prokaryotes and fungi, respectively. For fungal ASVs, putative functional groups (e.g., "plant pathogen") were inferred using the program FungalTraits (77). The estimation of DNA concentrations of the prokaryotic 16S rRNA and fungal ITS regions was performed, respectively, based on the calibration with the standard DNA gradients (artificial DNA variants introduced to the PCR master mix solutions) using the bioinformatic pipeline detailed elsewhere (40).

Diversity of microbiome structure

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

For each of the prokaryotic and fungal datasets, permutational analyses of variance (PERMANOVA) (78) were performed to examine associations between family-level community compositions and variables in the metadata. Two types of PERMANOVA models were constructed based on Bray-Curtis β-diversity (10,000 iterations). Specifically, one is constituted by categorical explanatory variables (crop plant, former crop plant, soil taxonomy, research site, and sampling month), while the other included continuous explanatory variables (soil pH,

426 electrical conductivity, C/N ratio, and available phosphorous concentration, latitude, and
427 longitude).

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

434

435 **Crop disease and microbiome structure**

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

442 We also examined how balance between prokaryotic and fungal abundance could be
443 associated with crop disease levels. Based on the estimated concentrations of prokaryotic 16S
444 rRNA and fungal ITS sequences in template DNA solutions, we calculated the ratio of
445 prokaryotic DNA concentrations to fungal DNA concentrations in respective samples
446 (prokaryote/fungus ratio) as follows:

447 $$\log\left(\frac{\text{prokaryotic 16S rRNA gene concentration(DNA copies/}\mu\text{L)}}{\text{fungal ITS gene concentration (DNA copies/}\mu\text{L)}}\right).$$

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

456 Potential effects of the prokaryote/fungus ratio on crop-disease level were examined in a
457 GLMM with a logit-link function and binomial errors. In the GLMM, soil pH and electrical
458 conductivity were included as fixed effects in addition to the prokaryote/fungus ratio. In the
459 GLMM, C/N ratio and available phosphorous concentration were not included due to the small
460 number of samples with both crop-disease information and the chemical variables. Sampling
461 month and experimental/research purpose information in the metadata were included as random
462 effects in the GLMM. To examine the effects of other aspects of biological factors, we
463 constructed another GLMM including the ASV-level Shannon diversity of prokaryotes, that of
464 fungi, and the prokaryote/fungus ratio. All the explanatory variables were *z*-standardized (zero-
465 mean and unit-variance) before the GLMM analysis. To visualize how the proportion of least-

466 diseased (disease-level 1) samples varied along the axis of the prokaryote/fungus ratio, we
467 performed a supplementary analysis with generalized linear models (GLMs). Specifically, for the
468 five experimental/research purpose categories from which 10 or more disease-level 1 (least-
469 diseased) samples and 10 or more disease-level 2–5 (severely diseased) samples were obtained, a
470 GLM of disease level was constructed by including the prokaryote/fungus ratio as an explanatory
471 variable.

472

473 **Microbiome network modules**

474 To infer potential interactions between microbial ASVs, the algorithm of sparse inverse
475 covariance estimation for ecological association inference (SPIEC-EASI) was applied based on
476 the Meinshausen-Bühlmann (MB) method as implemented in the SpiecEasi package (36) of R. In
477 total, 2,117 soil samples from which both prokaryotic and fungal community data were available
478 were subjected to the analysis. The network inference based on co-occurrence patterns allowed us
479 to detect pairs of microbial ASVs that potentially interact with each other in facilitative ways
480 and/or those that might share ecological niches (e.g., preference for edaphic factors). Because
481 estimation of co-occurrence patterns was not feasible for rare nodes, the prokaryotic and fungal
482 ASVs that appeared in more than 10 % of the sequenced samples were included in the input
483 matrix of the network analysis. Network modules, within which closely associated ASVs were
484 interlinked with each other, were identified with the algorithm based on edge betweenness (79)
485 using the igraph package (80) or R.

486 We then characterized the inferred network modules based on their potential impacts on
487 crop-plant disease level. For each microbial ASV constituting the modules, we evaluated
488 specificity of occurrences in samples differing in crop disease levels based on a randomization
489 analysis. For the calculation, the rarefied matrices of prokaryotes and fungi were combined.
490 Within the combined sample × ASV matrix, samples were categorized into the five disease
491 levels. Mean read counts across samples displaying each of the five disease levels were then
492 calculated for each ASV. Meanwhile, mean read counts for respective disease levels were
493 calculated as well for randomized matrices, in which disease labels of the samples were shuffled
494 (10,000 permutations). For *i*-th ASV, standardized specificity to disease level *j* (s_{ij}) was obtained
495 as follows:

496 $$s_{ij} = \frac{O_{ij} - Mean(R_{ij})}{SD(R_{ij})},$$

497 where O_{ij} and R_{ij} is the mean read counts of *i*-th ASV across disease-level-*j* samples in the
498 observed and randomized matrices, respectively, and $Mean(R_{ij})$ and $SD(R_{ij})$ indicate mean and
499 standard deviation across the randomized matrices. An s_{ij} value higher than 4 or that lower than
500 – 4 indicates statistically strong specificity of the *i*-th ASV to the disease level *j* (Fig. S4). For
501 each module in the inferred co-occurrence network, we then calculated mean standardized
502 specificity to disease level 1 across constituent ASVs.

503

504 **Network hubs**

To explore potential keystone microbes within the network, we scored respective ASVs on the basis of their topological positions. Among the indices used for evaluating ASVs, betweenness centrality (42), which measures the extent to which a given nodes (ASV) is located within the shortest paths connecting pairs of other nodes in a network, is commonly used to find hubs mediating flow of effects in a network. The network centrality scores were normalized as implemented in the igraph packages of R. In addition, by focusing on the above-mentioned network modules, we ranked ASVs based on their within-module degree and among-module connectivity (81). The former index is obtained as the number of nodes linked with a target node within a target network module, suggesting the topological importance of a node within the module it belongs to. The latter index represents the extent to which a node is linked with other nodes belonging to different network modules. Within-module degree was *z*-standardized (i.e., zero-mean and unit-variance) within each module, while among-module connectivity was defined to vary between 0 to 1. In addition to those indices for evaluating topological roles within a network, eigenvector centrality (82) was calculated for respective nodes.

Data availability

The 16S rRNA and ITS sequencing data are available from the DNA Data Bank of Japan (DDBJ: Bioproject: PSUB018361). The microbial community data are deposited at our GitHub repository (https://github.com/hiro-toju/Soil_Microbiome_NARO3000).

Code availability

All the R scripts used to analyze the data are available at the GitHub repository (https://github.com/hiro-toju/Soil_Microbiome_NARO3000).

REFERENCES

1. Springmann M, Clark M, Mason-D'Croz D, Wiebe K, Bodirsky BL, Lassaletta L, de Vries W, Vermeulen SJ, Herrero M, Carlson KM, Jonell M, Troell M, DeClerck F, Gordon LJ, Zurayk R, Scarborough P, Rayner M, Loken B, Fanzo J, Godfray HCJ, Tilman D, Rockström J, Willett W. 2018. Options for keeping the food system within environmental limits. Nature 562:519–525.
2. Tilman D, Balzer C, Hill J, Befort BL. 2011. Global food demand and the sustainable intensification of agriculture. Proceedings of the National Academy of Sciences 108:20260–20264.
3. Toju H, Peay KG, Yamamichi M, Narisawa K, Hiruma K, Naito K, Fukuda S, Ushio M, Nakaoka S, Onoda Y, Yoshida K, Schlaeppi K, Bai Y, Sugiura R, Ichihashi Y, Minamisawa K, Kiers ET. 2018. Core microbiomes for sustainable agroecosystems. Nat Plants 4:247–257.

542 4. Busby PE, Soman C, Wagner MR, Friesen ML, Kremer J, Bennett A, Morsy M, Eisen JA,
543 Leach JE, Dangl JL. 2017. Research priorities for harnessing plant microbiomes in
544 sustainable agriculture. PLoS Biol 15:e2001793.

545 5. Berendsen RL, Pieterse CMJ, Bakker PAHM. 2012. The rhizosphere microbiome and
546 plant health. Trends Plant Sci 17:478–486.

547 6. Bommarco R, Kleijn D, Potts SG. 2013. Ecological intensification: Harnessing ecosystem
548 services for food security. Trends Ecol Evol 28:230–238.

549 7. French E, Kaplan I, Iyer-Pascuzzi A, Nakatsu CH, Enders L. 2021. Emerging strategies for
550 precision microbiome management in diverse agroecosystems. Nat Plants 7:256–267.

551 8. Busby PE, Ridout M, Newcombe G. 2016. Fungal endophytes: modifiers of plant disease.
552 Plant Mol Biol 90:645–655.

553 9. Santoyo G, Moreno-hagelsieb G, Orozco-mosqueda C, Glick BR. 2016. Plant growth-
554 promoting bacterial endophytes. Microbiol Res 183:92–99.

555 10. Carey CJ, Dove NC, Beman JM, Hart SC, Aronson EL. 2016. Meta-analysis reveals
556 ammonia-oxidizing bacteria respond more strongly to nitrogen addition than ammonia-
557 oxidizing archaea. Soil Biol Biochem 99:158–166.

558 11. Schauss K, Focks A, Leininger S, Kotzerke A, Heuer H, Thiele-Bruhn S, Sharma S, Wilke
559 BM, Matthies M, Smalla K, Munch JC, Amelung W, Kaupenjohann M, Schloter M,
560 Schleper C. 2009. Dynamics and functional relevance of ammonia-oxidizing archaea in
561 two agricultural soils. Environ Microbiol 11:446–456.

562 12. Chen M, Arato M, Borghi L, Nouri E, Reinhardt D. 2018. Beneficial services of arbuscular
563 mycorrhizal fungi – from ecology to application. Front Plant Sci 9:1270.

564 13. Baker BJ, de Anda V, Seitz KW, Dombrowski N, Santoro AE, Lloyd KG. 2020. Diversity,
565 ecology and evolution of Archaea. Nat Microbiol 5:887–900.

566 14. van der Heijden MGA, Klironomos JN, Ursic M, Moutoglis P, Streitwolf-Engel R, Boller
567 T, Sanders IR, Wiemken A. 1998. Mycorrhizal fungal diversity determines plant
568 biodiversity, ecosystem variability and productivity. Nature 396:69–72.

569 15. Yuan J, Wen T, Zhang H, Zhao M, Penton CR, Thomashow LS, Shen Q. 2020. Predicting
570 disease occurrence with high accuracy based on soil macroecological patterns of Fusarium
571 wilt. ISME Journal 14:2936–2950.

572 16. Mendes R, Kruijt M, de Bruijn I, Dekkers E, van der Voort M, Schneider JHM, Piceno
573 YM, DeSantis TZ, Andersen GL, Bakker P a HM, Raaijmakers JM. 2011. Deciphering the
574 rhizosphere microbiome for disease-suppressive bacteria. Science 332:1097–100.

575 17. Schlatter D, Kinkel L, Thomashow L, Weller D, Paulitz T. 2017. Disease suppressive
576 soils: New insights from the soil microbiome. Phytopathology 107:1284–1297.

577 18. Wei Z, Gu Y, Friman V-PP, Kowalchuk GA, Xu Y, Shen Q, Jousset A. 2019. Initial soil
578 microbiome composition and functioning predetermine future plant health. Sci Adv 5:1–
579 11.

580 19. Hori Y, Fujita H, Hiruma K, Narisawa K, Toju H. 2021. Synergistic and Offset Effects of
581 Fungal Species Combinations on Plant Performance. Front Microbiol 12:713180.

582 20. Hardoim PR, van Overbeek LS, Elsas JD van. 2008. Properties of bacterial endophytes and
583 their proposed role in plant growth. Trends Microbiol 16:463–471.

21. Kiers ET, Duhamel M, Beesetty Y, Mensah JA, Franken O, Verbruggen E, Fellbaum CR, Kowalchuk GA, Hart MM, Bago A, Palmer TM, West SA, Vandenkoornhuyse P, Jansa J, Bücking H. 2011. Reciprocal rewards stabilize cooperation in the mycorrhizal symbiosis. Science (1979) 333:880–882.

22. Jiao S, Yang Y, Xu Y, Zhang J, Lu Y. 2020. Balance between community assembly processes mediates species coexistence in agricultural soil microbiomes across eastern China. ISME Journal 14:202–216.

23. Ray P, Lakshmanan V, Labbé JL, Craven KD. 2020. Microbe to Microbiome: A Paradigm Shift in the Application of Microorganisms for Sustainable Agriculture. Front Microbiol 11:622926.

24. Hartman K, van der Heijden MGA, Wittwer RA, Banerjee S, Walser JC, Schlaeppi K. 2018. Cropping practices manipulate abundance patterns of root and soil microbiome members paving the way to smart farming. Microbiome 6:14.

25. Shen Z, Xue C, Penton CR, Thomashow LS, Zhang N, Wang B, Ruan Y, Li R, Shen Q. 2019. Suppression of banana Panama disease induced by soil microbiome reconstruction through an integrated agricultural strategy. Soil Biol Biochem 128:164–174.

26. Carrión VJ, Perez-Jaramillo J, Cordovez V, Tracanna V, De Hollander M, Ruiz-Buck D, Mendes LW, van Ijcken WFJ, Gomez-Exposito R, Elsayed SS, Mohanraju P, Arifah A, van der Oost J, Paulson JN, Mendes R, van Wezel GP, Medema MH, Raaijmakers JM. 2019. Pathogen-induced activation of disease-suppressive functions in the endophytic root microbiome. Science (1979) 366:606–612.

27. Santos LF, Olivares FL. 2021. Plant microbiome structure and benefits for sustainable agriculture. Curr Plant Biol 26:100198.

28. Tedersoo L, Bahram M, Põlme S, Kõljalg U, Yorou NS, Wijesundera R, Ruiz LV, Vasco-Palacios AM, Thu PQ, Suija A, Smith ME, Sharp C, Saluveer E, Saitta A, Rosas M, Riit T, Ratkowsky D, Pritsch K, Põldmaa K, Piepenbring M, Phosri C, Peterson M, Parts K, Pärtel K, Otsing E, Nouhra E, Njouonkou AL, Nilsson RH, Morgado LN, Mayor J, May TW, Majuakim L, Lodge DJ, Lee S, Larsson KH, Kohout P, Hosaka K, Hiiesalu I, Henkel TW, Harend H, Guo LD, Greslebin A, Grelet G, Geml J, Gates G, Dunstan W, Dunk C, Drenkhan R, Dearnaley J, de Kesel A, Dang T, Chen X, Buegger F, Brearley FQ, Bonito G, Anslan S, Abell S, Abarenkov K. 2014. Global diversity and geography of soil fungi. Science (1979) 346:1256688.

29. Delgado-Baquerizo M, Oliverio AM, Brewer TE, Benavent-González A, Eldridge DJ, Bardgett RD, Maestre FT, Singh BK, Fierer N. 2018. A global atlas of the dominant bacteria found in soil. Science (1979) 359:320–325.

30. Egidi E, Delgado-Baquerizo M, Plett JM, Wang J, Eldridge DJ, Bardgett RD, Maestre FT, Singh BK. 2019. A few Ascomycota taxa dominate soil fungal communities worldwide. Nat Commun 10:2369.

31. Bahram M, Hildebrand F, Forslund SK, Anderson JL, Soudzilovskaia NA, Bodegom PM, Bengtsson-Palme J, Anslan S, Coelho LP, Harend H, Huerta-Cepas J, Medema MH, Maltz MR, Mundra S, Olsson PA, Pent M, Põlme S, Sunagawa S, Ryberg M, Tedersoo L, Bork P. 2018. Structure and function of the global topsoil microbiome. Nature 560:233–237.

32. Faust K, Raes J. 2012. Microbial interactions: From networks to models. Nat Rev Microbiol 10:538–550.

628 33. van der Heijden MGA, Hartmann M. 2016. Networking in the plant microbiome. PLoS
629 Biol 14:1–9.

630 34. Saleem M, Hu J, Jousset A. 2019. More than the sum of its parts: microbiome biodiversity
631 as a driver of plant growth and soil health. Annu Rev Ecol Evol Syst 50:145–168.

632 35. Durán P, Thiergart T, Garrido-Oter R, Agler M, Kemen E, Schulze-Lefert P, Hacquard S.
633 2018. Microbial Interkingdom Interactions in Roots Promote Arabidopsis Survival. Cell
634 175:973-983.e14.

635 36. Kurtz ZD, Müller CL, Miraldi ER, Littman DR, Blaser MJ, Bonneau RA. 2015. Sparse
636 and compositionally robust inference of microbial ecological networks. PLoS Comput Biol
637 11:e1004226.

638 37. Toju H, Yamamoto S, Tanabe AS, Hayakawa T, Ishii HS. 2016. Network modules and
639 hubs in plant-root fungal biomes. J R Soc Interface 13:20151097–20151097.

640 38. Goulding KWT. 2016. Soil acidification and the importance of liming agricultural soils
641 with particular reference to the United Kingdom. Soil Use Manag 32:390–399.

642 39. Corwin DL, Lesch SM. 2005. Apparent soil electrical conductivity measurements in
643 agriculture. Comput Electron Agric 46:11–43.

644 40. Fujita H, Ushio M, Suzuki K, Abe M, Yamamichi M, Iwayama K, Canarini A, Hayashi I,
645 Fukushima K, Fukuda S, Kiers E, Toju H. 2023. Alternative stable states, nonlinear
646 behavior, and predictability of microbiome dynamics. Microbiome 11:63.

647 41. Zhalnina K v., Dias R, Leonard MT, de Quadros PD, Camargo FAO, Drew JC, Farmerie
648 WG, Daroub SH, Triplett EW. 2014. Genome sequence of Candidatus nitrososphaera
649 evergladensis from group I.1b enriched from everglades soil reveals novel genomic
650 features of the ammonia-oxidizing archaea. PLoS One 9:e101648.

651 42. Freeman LC. 1977. A set of measures of centrality based on betweenness. Sociometry
652 40:35–41.

653 43. Lücker S, Wagner M, Maixner F, Pelletier E, Koch H, Vacherie B, Rattei T, Damsté JSS,
654 Spieck E, le Paslier D, Daims H. 2010. A Nitrospira metagenome illuminates the
655 physiology and evolution of globally important nitrite-oxidizing bacteria. Proc Natl Acad
656 Sci U S A 107:13479–84.

657 44. Summerell BA. 2019. Resolving *Fusarium*: Current Status of the Genus. Annu Rev
658 Phytopathol 57:323–339.

659 45. Gordon TR. 2017. *Fusarium oxysporum* and the *Fusarium* wilt syndrome. Annu Rev
660 Phytopathol 55:23–39.

661 46. de Lamo FJ, Takken FLW. 2020. Biocontrol by *Fusarium oxysporum* Using Endophyte-
662 Mediated Resistance. Front Plant Sci 11:37.

663 47. Kerou M, Offre P, Valledor L, Abby SS, Melcher M, Nagler M, Weckwerth W, Schleper
664 C. 2016. Proteomics and comparative genomics of *Nitrososphaera viennensis* reveal the
665 core genome and adaptations of archaeal ammonia oxidizers. Proc Natl Acad Sci U S A
666 113:E7937–E7946.

667 48. Poveda J. 2021. *Trichoderma* as biocontrol agent against pests: New uses for a
668 mycoparasite. Biological Control 159:104634.

669 49. Fisher MC, Gurr SJ, Cuomo CA, Blehert DS, Jin H, Stukenbrock EH, Stajich JE,
670 Kahmann R, Boone C, Denning DW, Gow NAR, Klein BS, Kronstad JW, Sheppard DC,

671 Taylor JW, Wright GD, Heitman J, Casadevall A, Cowen LE. 2020. Threats posed by the
672 fungal kingdom to humans, wildlife, and agriculture. mBio 11:e00449-20.

673 50. Dita M, Barquero M, Heck D, Mizubuti ESG, Staver CP. 2018. Fusarium wilt of banana:
674 Current knowledge on epidemiology and research needs toward sustainable disease
675 management. Front Plant Sci https://doi.org/10.3389/fpls.2018.01468.

676 51. Feng Y, Huang Y, Zhan H, Bhatt P, Chen S. 2020. An Overview of Strobilurin Fungicide
677 Degradation:Current Status and Future Perspective. Front Microbiol 11:389.

678 52. Yang C, Hamel C, Vujanovic V, Gan Y. 2011. Fungicide: Modes of Action and Possible
679 Impact on Nontarget Microorganisms. ISRN Ecol 2011:130289.

680 53. Sood M, Kapoor D, Kumar V, Sheteiwy MS, Ramakrishnan M, Landi M, Araniti F,
681 Sharma A. 2020. *Trichoderma*: The "secrets" of a multitalented biocontrol agent. Plants
682 9:762.

683 54. Yang Y, Yang E, An Z, Liu X. 2007. Evolution of nematode-trapping cells of predatory
684 fungi of the Orbiliaceae based on evidence from rRNA-encoding DNA and multiprotein
685 sequences. Proc Natl Acad Sci U S A 104:8379–8384.

686 55. Toju H, Tanaka Y. 2019. Consortia of anti-nematode fungi and bacteria in the rhizosphere
687 of soybean plants attacked by root-knot nematodes. R Soc Open Sci 6:181693.

688 56. Toju H, Yamamichi M, Jr PRG, Olesen JM, Mougi A, Yoshida T, Thompson JN. 2017.
689 Species-rich networks and eco-evolutionary synthesis at the metacommunity level. Nat
690 Ecol Evol 1:0024.

691 57. Ushio M. 2022. Interaction capacity as a potential driver of community diversity.
692 Proceedings of the Royal Society B: Biological Sciences 289:20212690.

693 58. Faust K, Lahti L, Gonze D, de Vos WM, Raes J. 2015. Metagenomics meets time series
694 analysis: Unraveling microbial community dynamics. Curr Opin Microbiol 25:56–66.

695 59. Menta C, Remelli S. 2020. Soil health and arthropods: From complex system to
696 worthwhile investigation. Insects 11:54.

697 60. Liu T, Chen X, Gong X, Lubbers IM, Jiang Y, Feng W, Li X, Whalen JK, Bonkowski M,
698 Griffiths BS, Hu F, Liu M. 2019. Earthworms Coordinate Soil Biota to Improve Multiple
699 Ecosystem Functions. Current Biology 29:3420–3429.

700 61. Neher DA. 2010. Ecology of plant and free-living nematodes in natural and agricultural
701 soil. Annu Rev Phytopathol 48:371–394.

702 62. Geisen S, Mitchell EAD, Adl S, Bonkowski M, Dunthorn M, Ekelund F, Fernández LD,
703 Jousset A, Krashevska V, Singer D, Spiegel FW, Walochnik J, Lara E. 2018. Soil protists:
704 A fertile frontier in soil biology research. FEMS Microbiol Rev 42:293–323.

705 63. Zhao ZB, He JZ, Geisen S, Han LL, Wang JT, Shen JP, Wei WX, Fang YT, Li PP, Zhang
706 LM. 2019. Protist communities are more sensitive to nitrogen fertilization than other
707 microorganisms in diverse agricultural soils. Microbiome 7:33.

708 64. Kageyama T, Toju H. 2022. Effects of source sample amount on biodiversity surveys of
709 bacteria, fungi, and nematodes in soil ecosystems. Front Ecol Evol 10:959945.

710 65. Quince C, Walker AW, Simpson JT, Loman NJ, Segata N. 2017. Shotgun metagenomics,
711 from sampling to analysis. Nat Biotechnol 35:833–844.

66. Neal AL, Barrat HA, Bacq-Lebreuil A, Qin Y, Zhang X, Takahashi T, Rubio V, Hughes D, Clark IM, Cárdenas LM, Gardiner LJ, Krishna R, Glendining ML, Ritz K, Mooney SJ, Crawford JW. 2023. Arable soil nitrogen dynamics reflect organic inputs via the extended composite phenotype. Nat Food 4:51–60.

67. Wu X, Rensing C, Han D, Xiao K-Q, Dai Y, Tang Z, Liesack W, Peng J, Cui Z, Zhang F. 2022. Genome-Resolved Metagenomics Reveals Distinct Phosphorus Acquisition Strategies between Soil Microbiomes. mSystems 7:e01107-21.

68. Luo G, Li L, Friman VP, Guo J, Guo S, Shen Q, Ling N. 2018. Organic amendments increase crop yields by improving microbe-mediated soil functioning of agroecosystems: A meta-analysis. Soil Biol Biochem 124:105–115.

69. Kassam A, Friedrich T, Derpsch R. 2019. Global spread of conservation agriculture. International Journal of Environmental Studies 76:29–51.

70. Schreefel L, Schulte RPO, de Boer IJM, Schrijver AP, van Zanten HHE. 2020. Regenerative agriculture – the soil is the base. Glob Food Sec 26:100404.

71. Li Y, Zhang Q, Cai Y, Yang Q, Chang SX. 2020. Minimum tillage and residue retention increase soil microbial population size and diversity: Implications for conservation tillage. Science of the Total Environment 716:137164.

72. 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.

73. Toju H, Tanabe AS, Yamamoto S, Sato H. 2012. High-coverage ITS primers for the DNA-based identification of ascomycetes and basidiomycetes in environmental samples. PLoS One 7:e40863.

74. Tanabe AS, Toju H. 2013. Two new computational methods for universal DNA barcoding: a benchmark using barcode sequences of bacteria, archaea, animals, fungi, and land plants. PLoS One 8:e76910.

75. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. 2016. DADA2: High-resolution sample inference from Illumina amplicon data. Nat Methods 13:581–583.

76. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glöckner FO. 2013. The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools. Nucleic Acids Res 41:D590–D596.

77. Põlme S, Abarenkov K, Henrik Nilsson R, Lindahl BD, Clemmensen KE, Kauserud H, Nguyen N, Kjøller R, Bates ST, Baldrian P, Frøslev TG, Adojaan K, Vizzini A, Suija A, Pfister D, Baral HO, Järv H, Madrid H, Nordén J, Liu JK, Pawlowska J, Põldmaa K, Pärtel K, Runnel K, Hansen K, Larsson KH, Hyde KD, Sandoval-Denis M, Smith ME, Toome-Heller M, Wijayawardene NN, Menolli N, Reynolds NK, Drenkhan R, Maharachchikumbura SSN, Gibertoni TB, Læssøe T, Davis W, Tokarev Y, Corrales A, Soares AM, Agan A, Machado AR, Argüelles-Moyao A, Detheridge A, de Meiras-Ottoni A, Verbeken A, Dutta AK, Cui BK, Pradeep CK, Marín C, Stanton D, Gohar D, Wanasinghe DN, Otsing E, Aslani F, Griffith GW, Lumbsch TH, Grossart HP, Masigol H, Timling I, Hiiesalu I, Oja J, Kupagme JY, Geml J, Alvarez-Manjarrez J, Ilves K, Loit K, Adamson K, Nara K, Küngas K, Rojas-Jimenez K, Bitenieks K, Irinyi L, Nagy LL, Soonvald L, Zhou LW, Wagner L, Aime MC, Öpik M, Mujica MI, Metsoja M, Ryberg M, Vasar M, Murata M, Nelsen MP, Cleary M, Samarakoon MC, Doilom M, Bahram M,

Hagh-Doust N, Dulya O, Johnston P, Kohout P, Chen Q, Tian Q, Nandi R, Amiri R, Perera RH, dos Santos Chikowski R, Mendes-Alvarenga RL, Garibay-Orijel R, Gielen R, Phookamsak R, Jayawardena RS, Rahimlou S, Karunarathna SC, Tibpromma S, Brown SP, Sepp SK, Mundra S, Luo ZH, Bose T, Vahter T, Netherway T, Yang T, May T, Varga T, Li W, Coimbra VRM, de Oliveira VRT, de Lima VX, Mikryukov VS, Lu Y, Matsuda Y, Miyamoto Y, Kõljalg U, Tedersoo L. 2020. FungalTraits: a user-friendly traits database of fungi and fungus-like stramenopiles. Fungal Divers 105:1–16.

78. Anderson MJ. 2001. A new method for non-parametric multivariate analysis of variance. Austral Ecol 26:32–46.

79. Newman M, Girvan M. 2004. Finding and evaluating community structure in networks. Phys Rev E 69:026113.

80. Csardi G, Nepusz T. 2006. The igraph software package for complex network research. InterJournal Complex Systems 1695:1–9.

81. Olesen JM, Bascompte J, Dupont YL, Jordano P. 2007. The modularity of pollination networks. Proc Natl Acad Sci U S A 104:19891–19869.

82. Bonacich P. 1972. Factoring and weighting approaches to status scores and clique identification. J Math Sociol 2:113–120.

ACKNOWLEDGEMENTS

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), 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.

Table 1 | Effects of environmental variables on prokaryotic/fungal community structure. For each set of categorical/continuous environmental variables, a PERMANOVA was performed for prokaryotic or fungal community dataset. The numbers of samples for which both community and environmental variable data were obtained are shown parentheses. The number of samples from which both sample metadata and DNA metabarcoding data were obtained is shown in a parenthesis in each model.

Model	Dataset	Variable	df	R^2	F	P
Categorical variables	Prokaryotes	Research site	33	0.0818	12.03	< 0.0001
		Month	11	0.0203	8.96	< 0.0001
		Crop	16	0.0064	1.94	< 0.0001
		Former crop	22	0.0382	8.42	< 0.0001
		Soil category	9	0.0060	3.22	< 0.0001
		Residual	2169	0.4473		
		Total	2267	1.0000		
	Fungi	Research site	33	0.0733	8.80	< 0.0001
		Month	11	0.0279	10.06	< 0.0001
		Crop	15	0.0071	1.87	< 0.0001
		Former crop	20	0.0372	7.38	< 0.0001
		Soil category	10	0.0060	2.37	< 0.0001
		Residual	2034	0.5132		
		Total	2131	1.0000		
Continuous variables	Prokaryotes	pH	1	0.0182	32.43	< 0.0001
		Electrical conductivity	1	0.0205	36.66	< 0.0001
		Available P	1	0.0125	22.25	< 0.0001
		C/N ratio	1	0.0078	13.93	< 0.0001
		Latitude	1	0.0144	25.75	< 0.0001
		Longitude	1	0.0163	29.09	< 0.0001
		Residual	1600	0.8957		
		Total	1606	1.0000		
	Fungi					
		pH	1	0.0117	19.98	< 0.0001
		Electrical conductivity	1	0.0221	37.68	0.0184
		Available P	1	0.0172	29.30	0.0098
		C/N ratio	1	0.0101	17.17	0.0038

Latitude	1	0.0171	29.14	< 0.0001
Longitude	1	0.0150	25.59	< 0.0001
Residual	1507	0.8833		
Total	1513	1.0000		

793 **Table 2 |** Relationship between prokaryotic/fungal community structure on the disease level of
794 crop plants. A GLMM of crop plants' disease level (disease level 1 vs. disease levels 2-5) with a
795 logit-link function and binomial errors was constructed by setting principal coordinates of
796 prokaryotic/fungal community structure (Fig. 2) as explanatory variables (fixed effects). The
797 identity of research projects and sampling month were included as random effects in the GLMM.

798

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

799

800

801 **Table 3** | Factors associated with the disease level of crop plants. In each GLMM with a logit-
802 link function and binomial errors, crop plants disease level [least-diseased (disease level 1) = 1,
803 severely-diseased (disease level 2-5) = 0] was included as response variable. In addition to
804 bacterium/fungal abundance ratio, soil environmental variables (pH and electrical conductivity;
805 Model 1) and α-diversity estimates of prokaryotes and fungi (Model 2) were included as
806 explanatory variables (fixed effects). The identity of research projects and sampling month were
807 included as random effects in the GLMM. A negative estimate indicates that the proportion of
808 least-diseased plants decreased with increasing values of a target explanatory variable.

809

Variable	Estimate	*z*	*P*
Model 1 (*N* = 1,069)			
pH	-0.268	-1.94	0.0524
Electrical conductivity	-0.181	-0.88	0.3765
Prokaryote/fungus ratio	-0.520	-4.46	< 0.0001
Model 2 (*N* = 1,111)			
Prokaryote diversity	-0.162	-1.23	0.2189
Fungal diversity	-0.331	-2.10	0.0355
Prokaryote/fungus ratio	-0.500	-4.19	< 0.0001

810
811

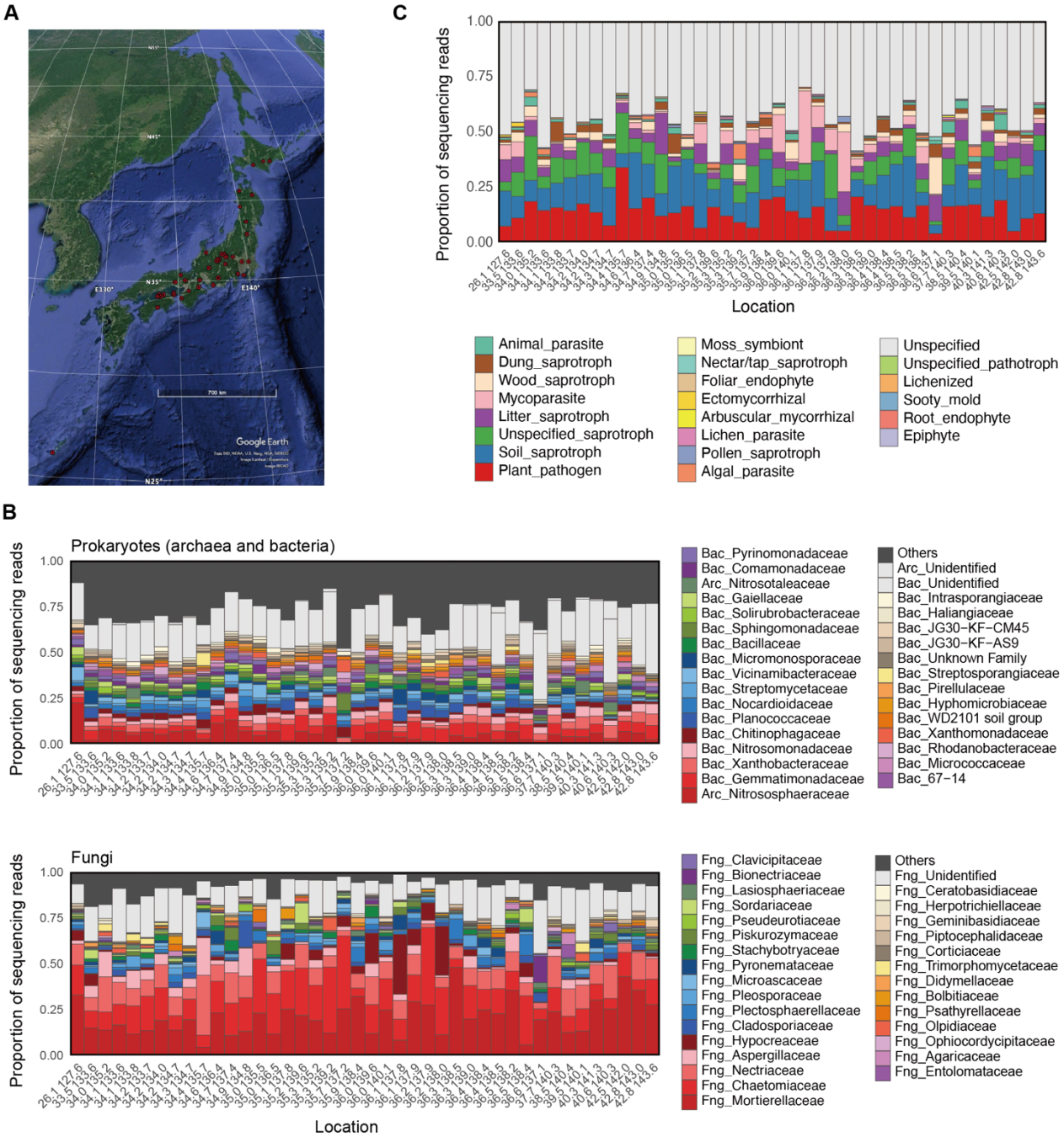


Fig. 1 | Comparison of soil microbiome structure across the Japan Archipelago. (A) Map of research sites across the Japan Archipelago. The 2,903 soil samples were grouped into 42 research sits were when their latitude and longitude profiles were round to one decimal place. (B) Taxonomic compositions of prokaryotes (archaea and bacteria; top) and fungi (bottom) at the family level. See Fig. S1 for results at the genus, order, and class levels. (C) Compositions of functional groups of fungi.

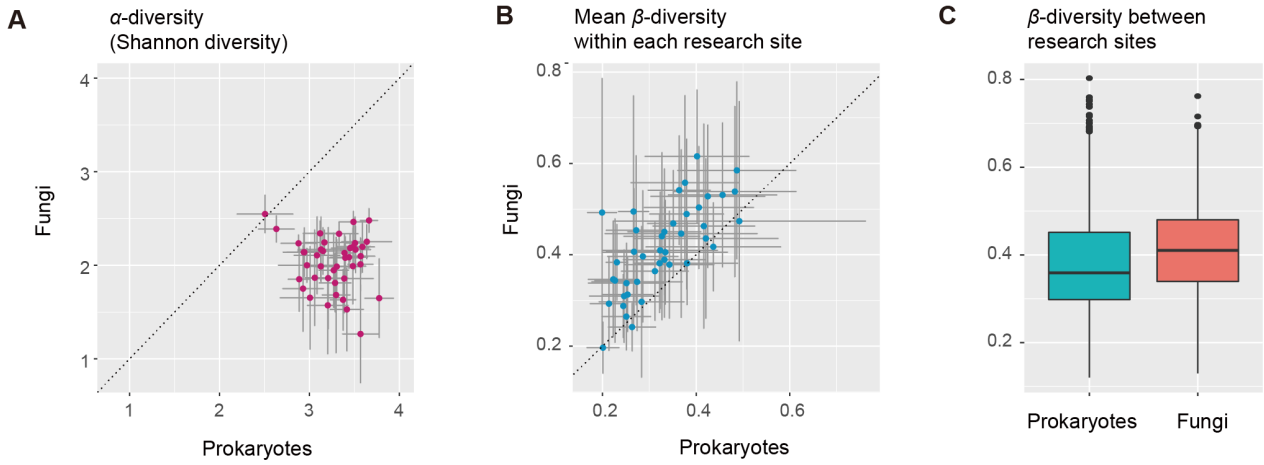


Fig. 2 | Diversity indices of agroecosystem microbiomes. (A) Comparison of α-diversity between prokaryotes and fungi. Mean Shannon diversity at the family level was shown for each research site. Bars represent standard deviations. (B) Mean community dissimilarity (β-diversity) between pairs of samples within each research site. (C) Community dissimilarity (β-diversity) in mean community compositions between pairs of research sites.

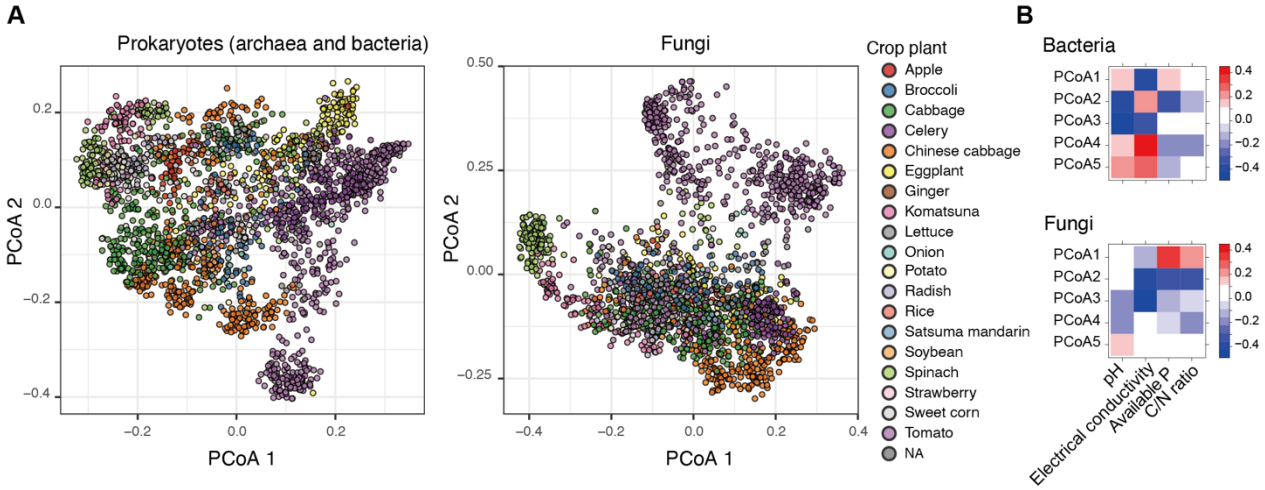


Fig. 3 | Dimensions of soil microbiome structure. (A) Prokaryote and fungal community structure. Principal co-ordinate analyses (PCoA) were performed based on ASV-level compositional matrices respectively for the prokaryotic and fungal communities. The identify of crop plants is shown by colors. See Figures S2-3 for relationship between community structure and environmental factors. (B) Correlation between PCoA scores and soil environmental factors.

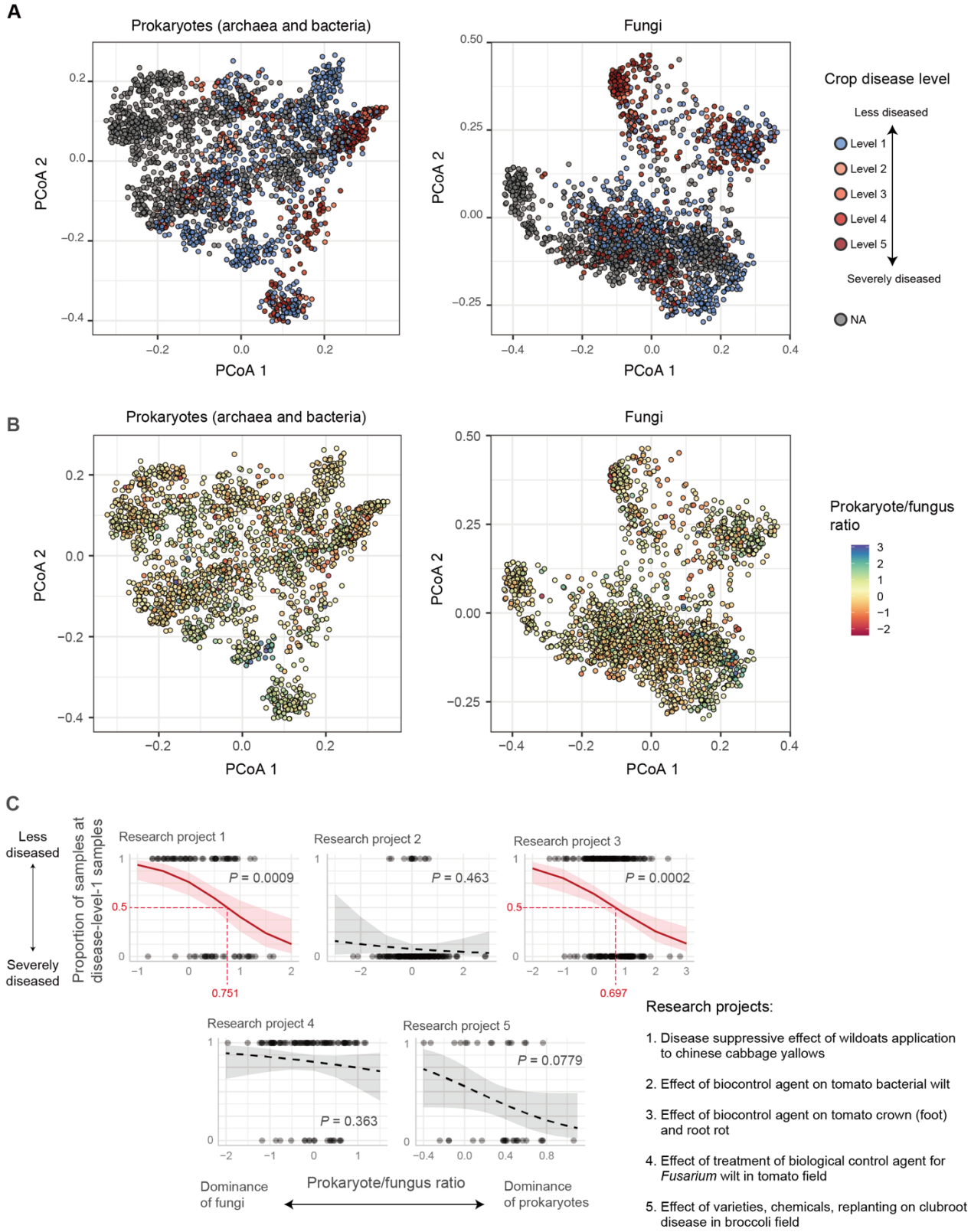
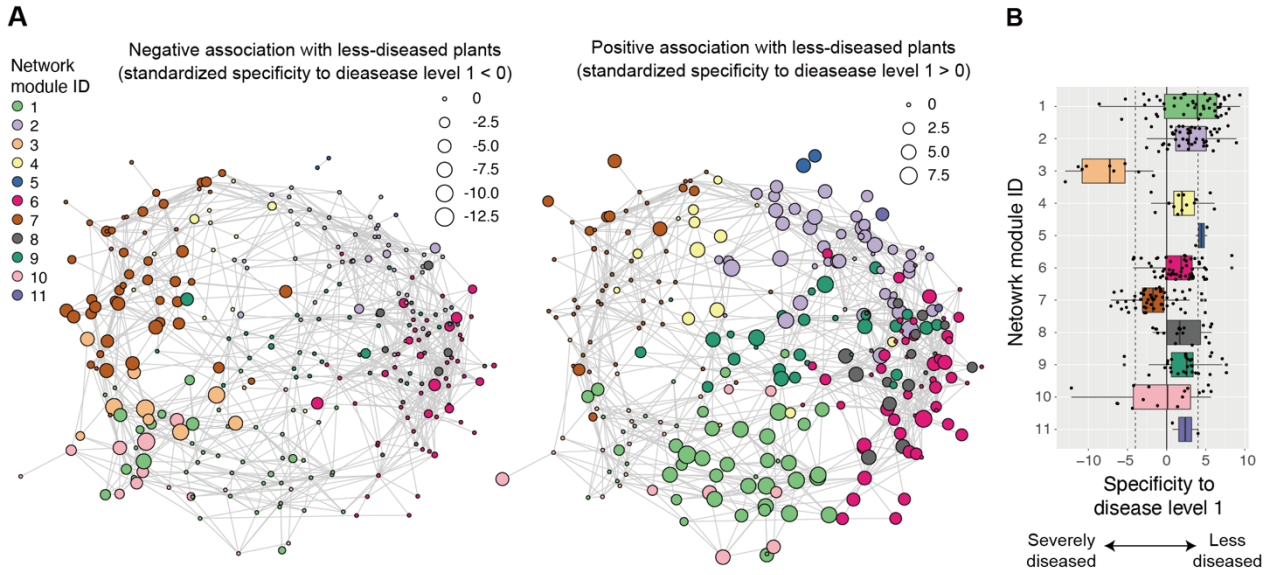


Fig. 4 | Crop disease and soil microbiome properties. (A) Crop disease level and microbial community structure. On the PCoA surface defined in Figure 3, crop disease level (see Materials and Methods) is indicated. (B) Balance between prokaryotic and fungal abundance. The ratio of prokaryotic abundance to fungal abundance was evaluated based on estimated DNA

concentrations [log $\left(\frac{\text{prokaryotic 16S rRNA gene concentration(DNA copies/}\mu\text{L)}}{\text{fungal ITS gene concentration (DNA copies/}\mu\text{L)}}\right)$]. The prokaryote/fungus ratio is shown for each sample on the PCoA surfaces of prokaryotic and fungal community compositions. (C) Representative relationship between prokaryote/fungus ratio on crop disease. For each of the five research projects with enough number of disease-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.



854
855

856 **Fig. 5 |** Architecture of microbe-to-microbe network. (A) Co-occurrence networks of archaea,
857 bacteria, and fungi. Specificity of occurrences to disease-level-1 (the lowest disease level)
858 samples (Fig. S4) is shown for each ASV within the network. The specificity is shown as node
859 size separately for negative (left) and positive (right) associations with least-diseased states of
860 crop plants. Colors indicate network modules, in which microbial ASVs in
861 commensalistic/mutualistic interactions and/or those sharing environmental preferences are
862 densely linked with each other. See Figure S5 for taxonomy (archaea, bacteria, or fungi) of
863 respective nodes. (B) Characteristics of network modules. Mean specificity to disease-level-1
864 samples is shown for each network module.
865

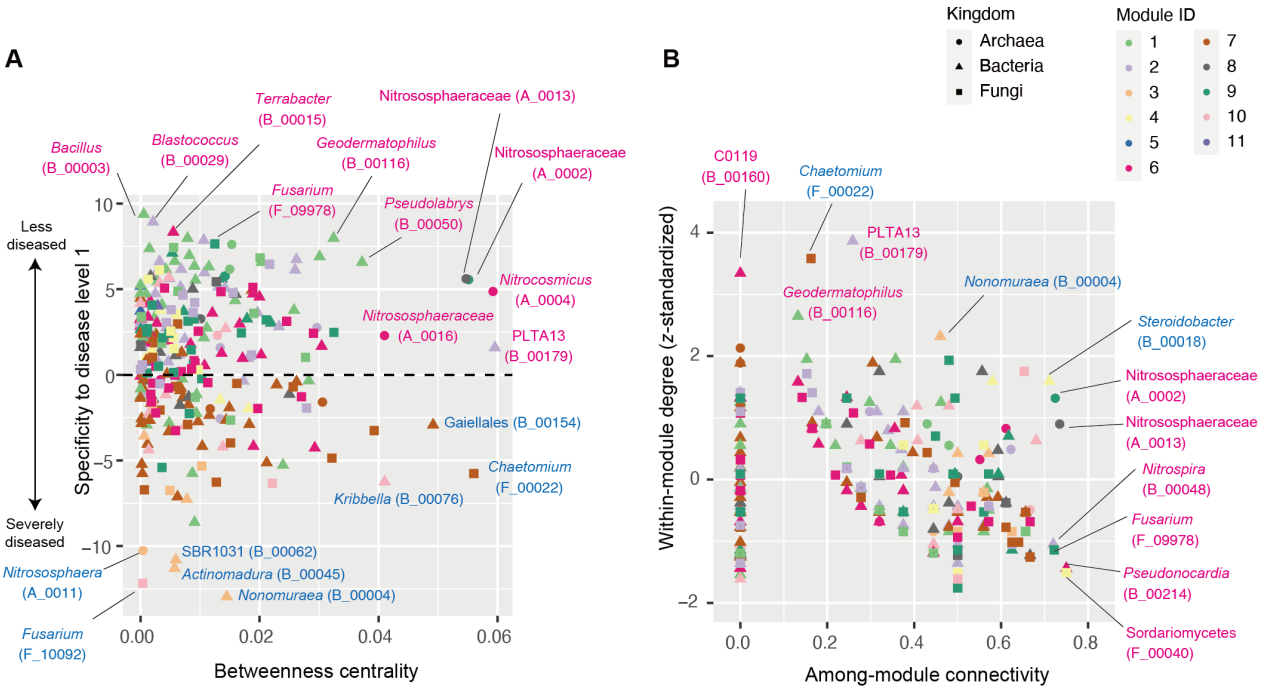
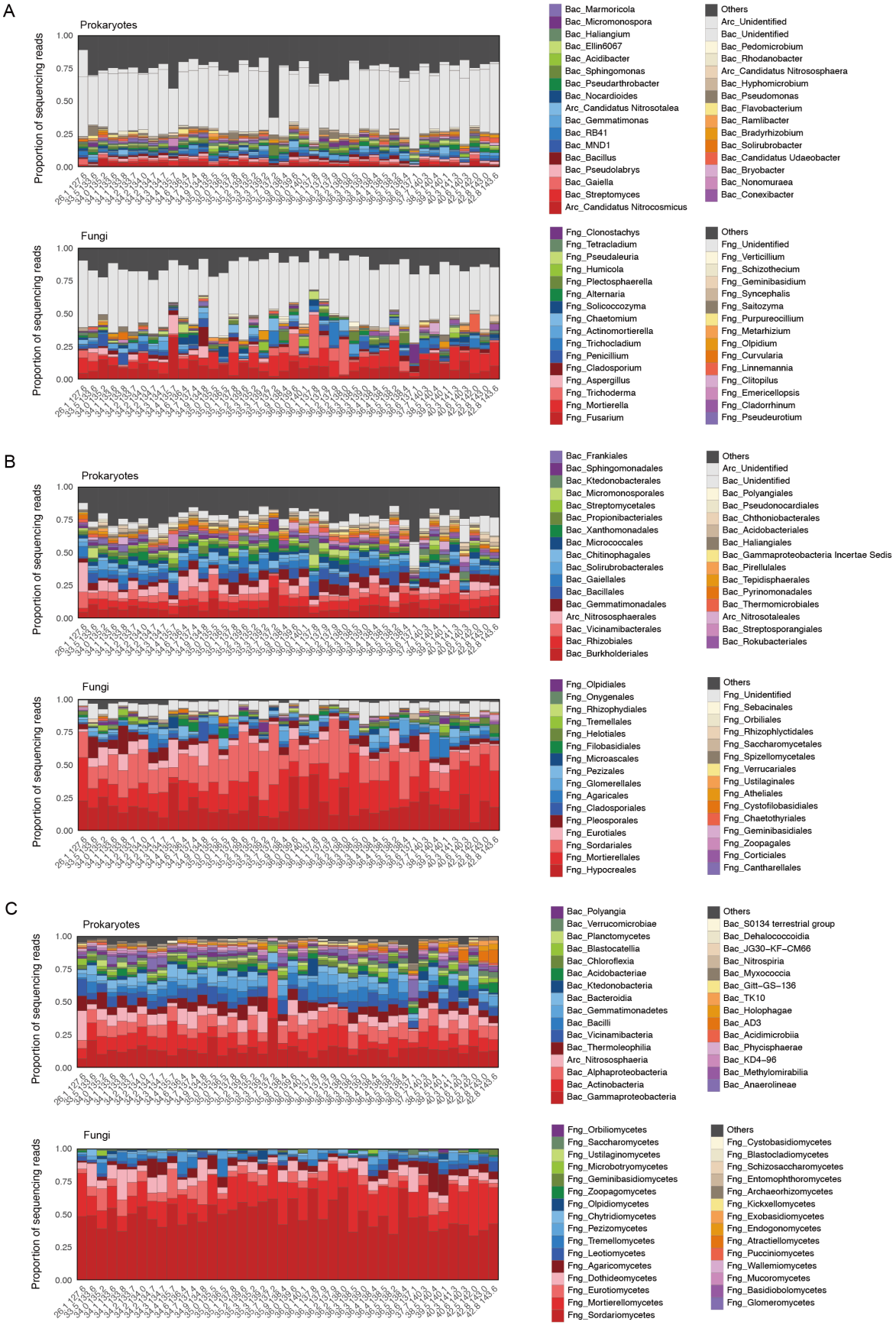


Fig. 6 | Network cores (hubs) within the microbial network. (A) Betweenness centrality of ASVs. For each ASV, betweenness centrality within the co-occurrence network is shown along the horizontal axis. Specificity to disease-level-1 samples is shown along the vertical axis. (B) Topological roles of ASVs within and across network modules. For each ASV, within-module degree represents the number of the ASVs linked with the target ASV within a module (*z*-standardized). Among-module connectivity represents the extent to which an ASV interlinks ASVs belonging to different network modules.


A
Prokaryotes
Proportion of sequencing reads
1.00
0.75
0.50
0.25
0.00
Bac_Marmoricola
Bac_Micromonospora
Bac_Haliangium
Bac_Ellin6067
Bac_Acidibacter
Bac_Sphingomonas
Bac_Pseudarthrobacter
Bac_Nocardioides
Arc_Candidatus Nitrosotalea
Bac_Gemmatimonas
Bac_RB41
Bac_MND1
Bac_Bacillus
Bac_Pseudolabrys
Bac_Gaiella
Bac_Streptomyces
Arc_Candidatus Nitrocosmicus
Others
Arc_Unidentified
Bac_Unidentified
Bac_Pedomicrobium
Bac_Rhodanobacter
Arc_Candidatus Nitrososphaera
Bac_Hyphomicrobium
Bac_Pseudomonas
Bac_Flavobacterium
Bac_Ramlibacter
Bac_Bradyrhizobium
Bac_Solirubrobacter
Bac_Candidatus Udaeobacter
Bac_Bryobacter
Bac_Nonomuraea
Bac_Conexibacter
Fungi
Fng_Clonostachys
Fng_Tetracladium
Fng_Pseudaleuria
Fng_Humicola
Fng_Plectosphaerella
Fng_Alternaria
Fng_Solicoccozyma
Fng_Chaetomium
Fng_Actinomortierella
Fng_Trichocladium
Fng_Penicillium
Fng_Cladosporium
Fng_Aspergillus
Fng_Trichoderma
Fng_Mortierella
Fng_Fusarium
Others
Fng_Unidentified
Fng_Verticillium
Fng_Schizothecium
Fng_Geminibasidium
Fng_Syncephalis
Fng_Saitozyma
Fng_Purpureocillium
Fng_Metarhizium
Fng_Olpidium
Fng_Curvularia
Fng_Linnemannia
Fng_Clitopilus
Fng_Emericellopsis
Fng_Cladorrhinum
Fng_Pseudeurotium
B
Prokaryotes
Proportion of sequencing reads
1.00
0.75
0.50
0.25
0.00
Bac_Frankiales
Bac_Sphingomonadales
Bac_Ktedonobacterales
Bac_Micromonosporales
Bac_Streptomycetales
Bac_Propionibacteriales
Bac_Xanthomonadales
Bac_Micrococcales
Bac_Chitinophagales
Bac_Solirubrobacterales
Bac_Gaiellales
Bac_Bacillales
Bac_Gemmatimonadales
Arc_Nitrosophaerales
Bac_Vicinamibacterales
Bac_Rhizobiales
Bac_Burkholderiales
Others
Arc_Unidentified
Bac_Unidentified
Bac_Polyangiales
Bac_Pseudonocardiales
Bac_Chthoniobacterales
Bac_Acidobacteriales
Bac_Haliangiales
Bac_Gammaproteobacteria Incertae Sedis
Bac_Pirellulales
Bac_Tepidisphaerales
Bac_Pyrinomonadales
Bac_Thermomicrobiales
Arc_Nitrosotaleales
Bac_Streptosporangiales
Bac_Rokubacteriales
Fungi
Fng_Olpidiales
Fng_Onygenales
Fng_Rhizophydiales
Fng_Tremellales
Fng_Helotiales
Fng_Filobasidiales
Fng_Microascales
Fng_Pezizales
Fng_Glomerellales
Fng_Agaricales
Fng_Cladosporiales
Fng_Pleosporales
Fng_Eurotiales
Fng_Sordariales
Fng_Mortierellales
Fng_Hypocreales
Others
Fng_Unidentified
Fng_Sebacinales
Fng_Orbiliales
Fng_Rhizophlyctidales
Fng_Saccharomycetales
Fng_Spizellomycetales
Fng_Verrucariales
Fng_Ustilaginales
Fng_Atheliales
Fng_Cystofilobasidiales
Fng_Chaetothyriales
Fng_Geminibasidiales
Fng_Zoopagales
Fng_Corticiales
Fng_Cantharellales
C
Prokaryotes
Proportion of sequencing reads
1.00
0.75
0.50
0.25
0.00
Bac_Polyangia
Bac_Verrucomicrobiae
Bac_Planctomycetes
Bac_Blastocatellia
Bac_Chloroflexia
Bac_Acidobacteriae
Bac_Ktedonobacteria
Bac_Bacteroidia
Bac_Gemmatimonadetes
Bac_Bacilli
Bac_Vicinamibacteria
Bac_Thermoleophilia
Arc_Nitrososphaeria
Bac_Alphaproteobacteria
Bac_Actinobacteria
Bac_Gammaproteobacteria
Others
Bac_S0134 terrestrial group
Bac_Dehalococcoidia
Bac_JG30–KF–CM66
Bac_Nitrospiria
Bac_Myxococcia
Bac_Gitt–GS–136
Bac_TK10
Bac_Holophagae
Bac_AD3
Bac_Acidimicrobiia
Bac_Phycisphaerae
Bac_KD4–96
Bac_Methylomirabilia
Bac_Anaerolineae
Fungi
Proportion of sequencing reads
1.00
0.75
0.50
0.25
0.00
Fng_Orbiliomycetes
Fng_Saccharomycetes
Fng_Ustilaginomycetes
Fng_Microbotryomycetes
Fng_Geminibasidiomycetes
Fng_Zoopagomycetes
Fng_Olpidiomycetes
Fng_Chytridiomycetes
Fng_Pezizomycetes
Fng_Tremellomycetes
Fng_Leotiomycetes
Fng_Agaricomycetes
Fng_Dothideomycetes
Fng_Eurotiomycetes
Fng_Mortierellomycetes
Fng_Sordariomycetes
Others
Fng_Cystobasidiomycetes
Fng_Blastocladiomycetes
Fng_Schizosaccharomycetes
Fng_Entomophthoromycetes
Fng_Archaeorhizomycetes
Fng_Kickxellomycetes
Fng_Exobasidiomycetes
Fng_Endogonomycetes
Fng_Atractiellomycetes
Fng_Pucciniomycetes
Fng_Wallemiomycetes
Fng_Mucoromycetes
Fng_Basidiobolomycetes
Fng_Glomeromycetes

878 **Fig. S1 |** Microbial taxonomic compositions at the genus, order, and class levels. (A) Genus-level
879 taxonomic compositions of prokaryotes and fungi. See Figure 1a for the map of the research sites. (B)
880 Order-level taxonomic compositions. (C) Class-level taxonomic compositions.
881

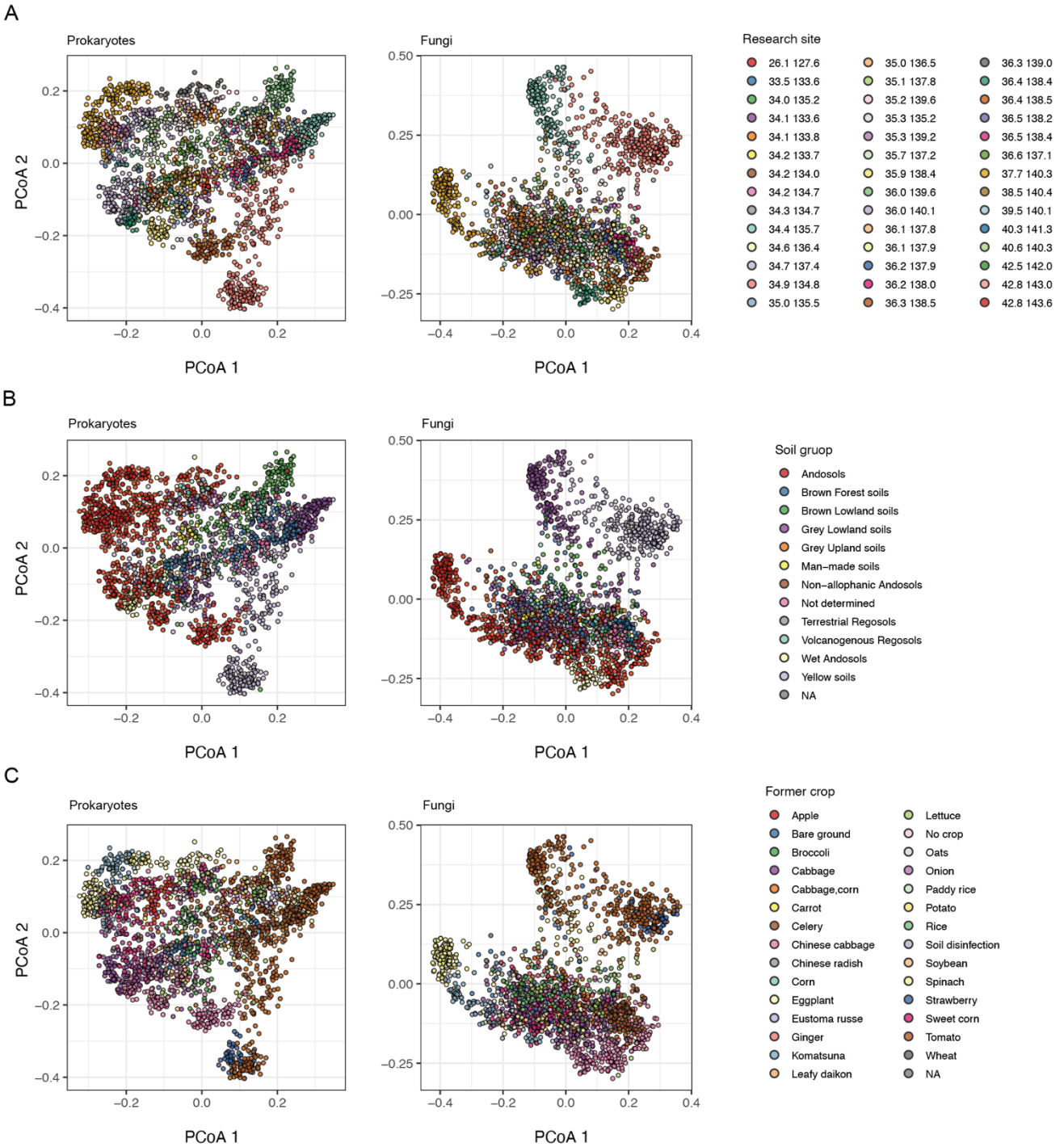


Fig. S2 | Community structure and metadata properties of the samples. (A) Prokaryote/fungal community structure and research-site profiles. Research sites are indicated by colors on the PCoA surface of prokaryotic/fungal community structure. (B) Soil taxonomy profile. c, Former crop plant.

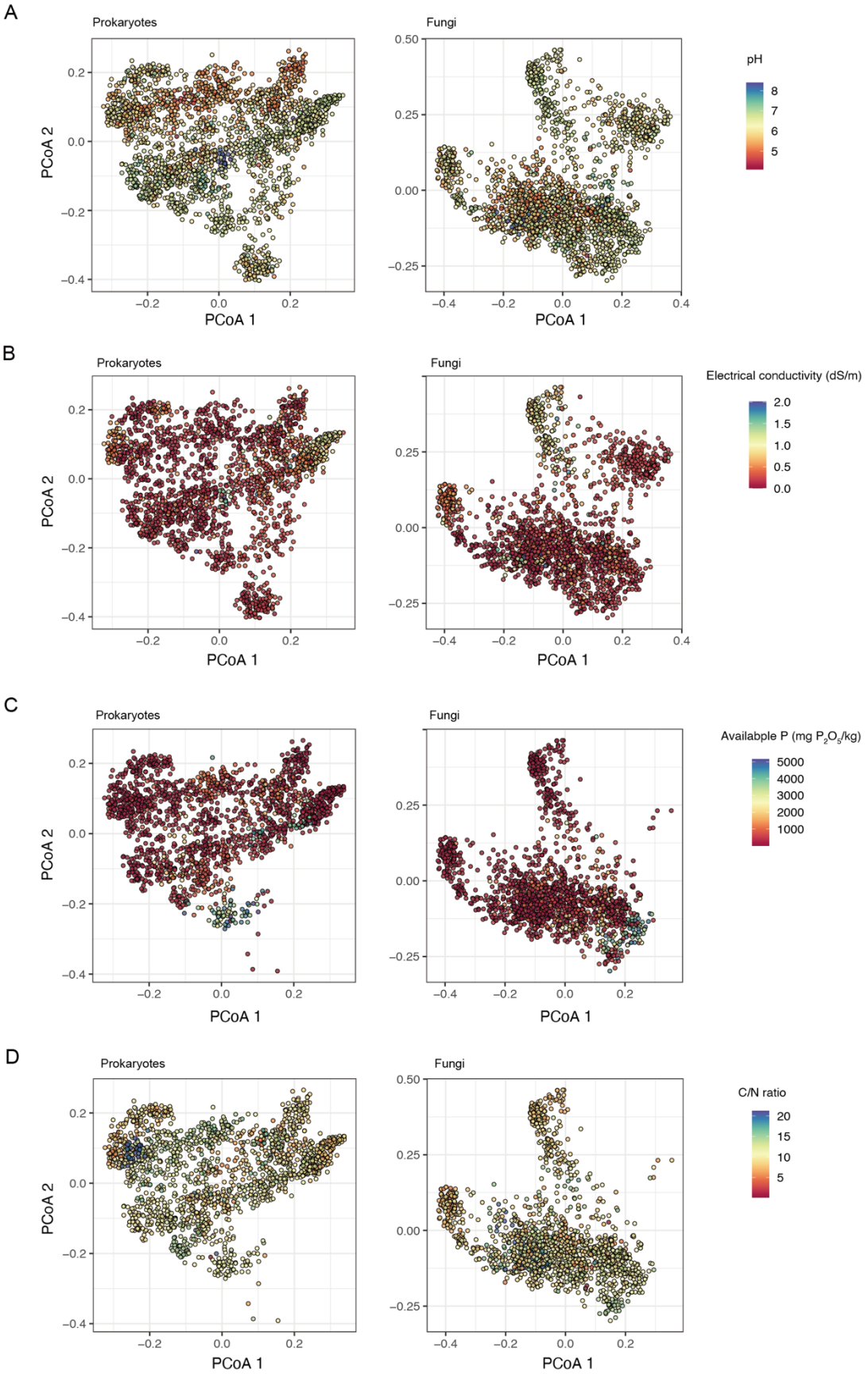

A
Prokaryotes
Fungi
pH
8
7
6
5
PCoA 2
PCoA 1
B
Prokaryotes
Fungi
Electrical conductivity (dS/m)
2.0
1.5
1.0
0.5
0.0
PCoA 2
PCoA 1
C
Prokaryotes
Fungi
Availabple P (mg P_2O_5/kg)
5000
4000
3000
2000
1000
PCoA 2
PCoA 1
D
Prokaryotes
Fungi
C/N ratio
20
15
10
5
PCoA 2
PCoA 1

Fig. S3 | Community structure and soil chemical properties. (A) Prokaryote/fungal community structure and soil pH of the samples. (B) Electrical conductivity. (C) Available phosphorous concentration. (D) Carbon to nitrogen ratio.

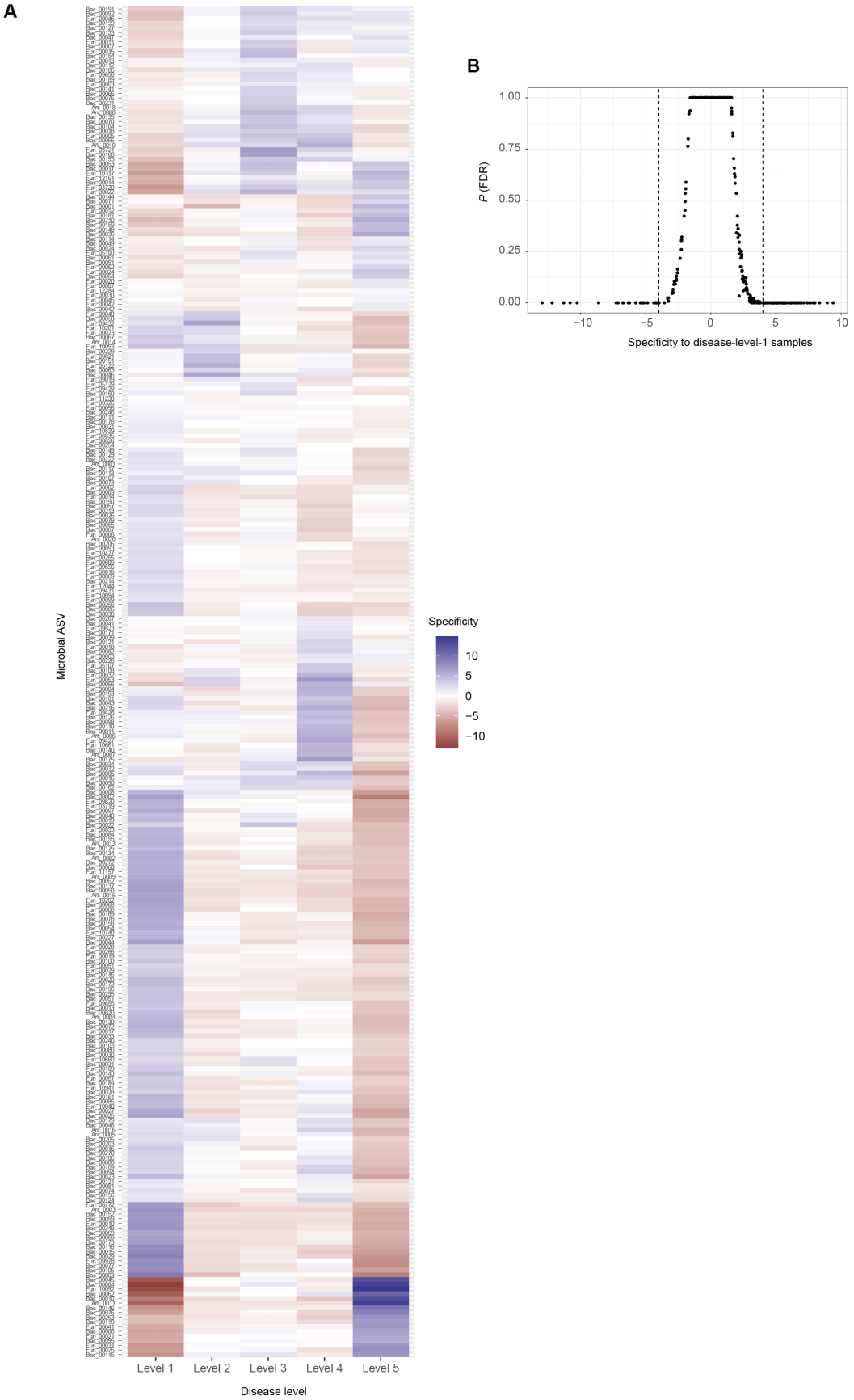
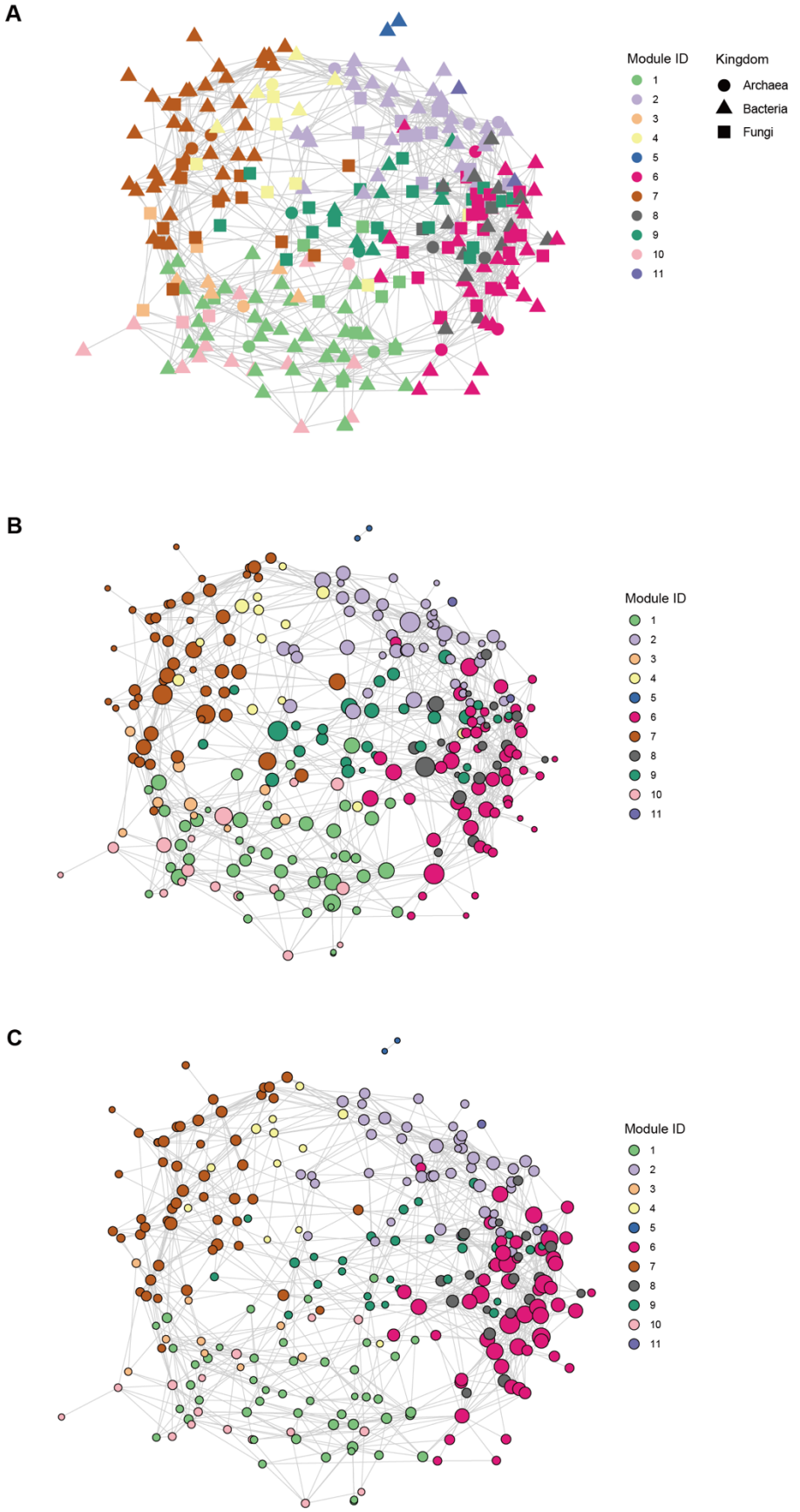


Fig. S4 | Each ASV's specificity to disease levels. (A) Standardized specificity to disease level (s_{ij}). Based on a randomization approach, each ASV's specificity to samples differing in crop

disease levels was evaluated. A higher value of the specificity index indicates that a microbial ASV displayed higher abundance in samples of the target disease level than that expected by chance. (B) Specificity scores and statistical significance. An s_{ij} value higher than 4 or that lower than -4 indicates statistically strong specificity of an ASV to a target disease level.


A
Module ID
1
2
3
4
5
6
7
8
9
10
11
Kingdom
Archaea
Bacteria
Fungi
B
Module ID
1
2
3
4
5
6
7
8
9
10
11
C
Module ID
1
2
3
4
5
6
7
8
9
10
11

902
903

904 **Fig. S5 |** Network centrality scores. (A) Distribution of archaea, bacteria, and fungi within the co-
905 occurrence network. (B) Each ASV's betweenness centrality score within the network. The node
906 size represents relative betweenness centrality scores. (C) Each ASV's eigenvector centrality
907 score within the network. The node size represents relative eigenvector centrality scores.

908