# 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[1,4,5†]

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

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

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

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

[5]Laboratory of Ecosystems and Coevolution, Graduate School of Biostudies, Kyoto University, Kyoto 606-8501, 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 632 archaeal, 26,868 bacterial, and 4,889 fungal operational taxonomic units detected across the fields of 19 crop plant species allowed us to conduct statistical analyses (permutational analyses of variance, generalized linear mixed models, and randomization analyses) on relationship among edaphic factors, microbiome compositions, and crop disease prevalence. We then examined whether the diverse microbes form species sets varying in potential ecological impacts on crop plants. A network analysis suggested that the observed prokaryotes and fungi were actually classified into several species sets (network modules), which differed substantially in associations with crop disease prevalence. Within the network of microbe-to-microbe coexistence, ecologically diverse microbes, such as an ammonium-oxidizing archaeum, an antibiotics-producing bacterium, and a potentially mycoparasitic fungus, were inferred to play key 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 and managing agroecosystems with high 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), although there are some exceptionally comprehensive studies comparing multiple research sites (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). In fact, ecological network studies have inferred how sets of microbial species could respond to the outbreaks or experimental introductions of crop plant pathogens (36–38). Although various statistical platforms for deciphering the architecture of such microbial interaction networks have been proposed (32, 39), hundreds or more of microbial community samples are required to gain reliable inferences on interactions that reproducibly occur in real ecosystems (40). 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 coexistence network consisting of diverse archaea, bacteria, and fungi. Specifically, we examined whether the

network architecture was partitioned into compartments (modules) of closely interacting microbial species. In addition, we tested the hypothesis that such network modules could differ in their positive/negative associations with crop plant disease/health status. To explore prokaryotic and fungal species keys to manage agroecosystem structure and functions, we further explored "core" or "hub" species that were placed at the central positions within the inferred microbial interaction network. 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 identifying sets of microbes to be controlled in sustainable crop production.

## 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 2,676 and 2,477 samples, respectively. In total, 632 archaeal operational taxonomic units (OTUs) representing 22 genera (24 families), 26,868 bacterial OTUs representing 1,120 genera (447 families), and 4,889 fungal OTUs representing 1,190 genera (495 families) were detected (Fig. 1B; Fig. S1).

The prokaryotic communities lacked apparently dominant taxa at the genus and family levels (Fig. 1B). In contrast, the fungal communities were dominated by fungi in the families Mortierellaceae, Chetomiaceae, and Nectriaceae, depending on localities (Fig. 1B). A reference database profiling of fungal functional groups suggested that the fungal communities were dominated by soil saprotrophic and plant pathogenic fungi (Fig. 1C) as characterized by the dominance of *Mortierella* and *Fusarium* 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).

### Microbiome structure and crop disease prevalence

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, and C/N ratio, (Fig. 2A-B; Figs. S2-3; Table 1). In addition, the ratio of prokaryotic abundance to fungal abundance (see

Materials and Methods for details) was associated with both prokaryotic and fungal community structure (Table 1). Nonetheless, the explanatory powers of these variable were all small as indicated by the low $R^2$ values (Table 1).

Both prokaryotic and fungal community structure was significantly associated with the severity of crop disease (Fig. 2C; Table 2). Specifically, the crop plants' disease/health status (disease level 1 vs. disease levels 2-5; see Materials and Methods) was explained by some of the principal components (PCs) defined based on prokaryotic/fungal community structure (Fig. 2B).

## Microbes associated with crop disease/health status

We explored microbial OTUs whose prevalence are associated with crop plant disease/health status. Based on a randomization analysis, prokaryotic/fungal OTUs whose distribution is biased in samples representing the minimal crop disease level (disease level 1) were screened (Fig. S4).

To examine whether the OTUs highlighted in the across-Japan spatial scale could actually show tight associations with crop disease status at local scales, the randomization analysis was performed as well in each of the six sub-datasets representing unique combinations of research sites, crop plant species, and experimental/research purposes (Data S2). Statistically significant specificity for crop disease level (FDR < 0.025; two-tailed test) was observed for at least one OTU in five of the six sub-datasets (Data S2). Among them, exceptionally strong specificity to the minimal crop disease level (standardized specificity score ≥ 6.0; FDR < 0.0001) was detected in two sub-datasets (Table 3). The relative abundance of these OTUs tightly associated with crop disease level across samples within each sub-dataset (Fig. 3).

## Microbe-to-microbe network

We then examined the network architecture of potential microbe-to-microbe interactions within the soil microbiomes. The inferred network of coexistence was subdivided into several modules, in which archaeal, bacterial, and fungal OTUs sharing environmental preferences and/or those in positive interactions were linked with each other (Figs. 3A and S5-8). The network modules differed considerably in their associations with crop-plant disease level (Fig. 4B; Fig. S5; Data S3). Modules 2, 6, and 8, for example, were characterized by microbes associated with least disease level. Module 6, which showed the highest mean specificity to the minimal crop disease level (Fig. 4B), included a bacterium allied to the genus *Gemmatimonas* (Bac_00025), that allied to the genus *Thermanaerothrix* (Bac_00258), and a *Plectosphaerella* fungus (Fun_4447) (Table 4). In contrast to these modules, Modules 1 and 7 were constituted by microbes negatively associated with crop plant health (Fig. 4B). Module 1 included a bacterium distantly allied to the genus *Ureibacillus* (Bac_00165), a *Nonomuraea* bacterium (Bac_00004), and a *Streptomyces* bacterium (Bac_00010), while Module 7 involved a *Fusarium* fungus (Fun_4028) and a *Nitrososphaera* archaeum (Arc_006) (Table 4).

## Core species within the microbial network

We next explored microbial OTUs that potentially have great impacts on community- or ecosystem-scale processes based on an analysis of the microbe-to-microbe network architecture

(Data S3). Among the microbes disproportionately found from the samples with the minimal crop disease level, a Pyrinomonadaceae bacterium allied to the genus *Brevitalea* (Bac_00182 in Module 6; Table 4), for example, showed a high betweenness centrality score (Fig. 5). Meanwhile, among the microbes negatively associated with crop health status, a bacterium distantly allied to the genus *Ureibacillus* (Bac_00165 in Module 1; Table 4) was inferred to be located at a central position within the network (Fig. 5).

We further ranked microbial OTUs in terms of their topological roles in interlinking multiple network modules. We then found that OTUs linked with many other OTUs within modules were not necessarily placed at the topological positions interconnecting different modules (Fig. 6). In Module 6, which showed high specificity to the minimal crop disease level (Fig. 4), a bacterium distantly allied to the genus *Thermanaerothrix* (Bac_00258) was designated as a "within-module hub", while a *Plectosphaerella* fungus (Fun_4447) showed a high "among-module connectivity" score (Table 4). Likewise, in Module 1, which consisted of many OTUs with negative associations with crop plant health (Fig. 4), a bacterium allied to the genus *Gemmatimonas* (Bac_00258) had the highest numbers of within-module links, while a *Curvularia* fungus (Fun_0043) was inferred to be an among-module hub (Table 5). The list of microbial OTUs placed at the interface of modules (OTUs with high among-module connectivity scores) involved a *Nitrosotenuis* archaeum, *Arenimonas*, *Arthrobacter*, and *Streptomyces* bacteria, and *Mortierella*, *Curvularia*, and *Trichoderma* fungi (Table 5).

## DISCUSSION

We here profiled the diversity of agroecosystem microbiome structure across a latitudinal gradient from cool-temperate to subtropical regions based on the analysis of > 2,000 soil samples. As partially reported in previous studies comparing microbiome compositions across broad geographic ranges (15, 22), prokaryotic and fungal community structure varied depending on season, crop plant species, former crop identity, and background soil categories (Fig. 2A; Table 1). In addition, soil chemical properties such as pH, electrical conductivity, and C/N ratio as well as the prokaryote/fungus abundance ratio significantly explained variation in microbiome structure (Table 1). In contrast, available phosphorus concentrations had significant effects on neither prokaryotic nor fungal communities in the multivariate model (Table 1), suggesting that nitrogen cycles rather than phosphorous ones are more tightly linked with microbiome structure. The integration of the microbiome datasets with agricultural field metadata allowed us to perform statistical tests of potential relationship between microbiome structure and agroecosystem performance (Fig. 2B; Table 2). A series of OTU-level analyses further highlighted taxonomically diverse prokaryotes and fungi showing strong positive or negative associations with crop health status (Figs. 3 and S4; Table 3).

We then examined how these microbes differing in associations with crop disease/health status form a network of coexistence. The architecture of the network involving diverse archaeal, bacterial, and fungal OTUs was highly structured, being partitioned into 11 modules (Fig. 4A). Intriguingly, the network modules varied considerably in constituent microbes' associations with crop disease levels (Fig. 4B). This result suggests that sets of microbes can be used to design soil microbiomes with crop-disease-suppressive functions. Among the detected modules, Modules 2, 6, and 8 were of particular interest with regard to the assembly of microbial OTUs positively

associated with crop health status (Figs. 4 and 5). In contrast, Modules 1 and 7 were constituted mainly by microbial OTUs negatively associated with plant health (Fig. 4B). In particular, Module 7 was characterized by the presence of a notorious plant pathogenic fungus, *Fusarium oxysporum* [(41, 42); but see (43) for diversity of their impacts on plants]. All these modules included both prokaryotes and fungi (Fig. S8; Data S3), illuminating the importance of inter-kingdom interactions (3, 33). The presence of microbial species sets differing in plant-associated ecological properties suggests that keeping specific sets of compatible prokaryotes and fungi is essential for maximizing the stability of agricultural production (3).

The analysis of network architecture further allowed us to explore core or hub species within the microbial network (Fig. 6). Because the microbes highlighted with the examined network indices occupy key positions interconnecting many other microbes (44), 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 (40) (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). The candidate list of such core species involved an ammonium-oxidizing archaeum (*Nitrosotenuis*) (45), an antibiotics-producing bacterium (*Streptomyces*) (46), a prevalent soil fungus (*Mortierella*) (47, 48), a potentially mycoparasitic fungus (*Trichoderma*) (49, 50), and fungi allied to plant pathogenic clades [*Curvularia* and *Plectosphaerella* (anamorph = *Fusarium*)] (51, 52) (Table 5). Given that many of the bacterial and fungal taxa listed above are culturable, experimental studies examining their ecological roles are awaited. Specifically, it would be intriguing to test whether substantial shifts in soil microbiome structure and functions can be caused by the introduction of those among-module hub microbes.

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 (53–55) 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 (56–59), simultaneous analyses of all prokaryotic and eukaryotic taxa (60, 61) will help us infer whole webs of biological processes. Third, meta-analyses of agroecosystem performance across diverse crop fields require utmost care because there is no firm criterion commonly applicable to different crop plant species or different pest/pathogen species. As implemented in this study, effects of such difference may be partially controlled by including them as random variables in GLMMs (Table 2). Nonetheless, local-scale analyses targeting specific crop plant species and disease symptoms (Fig. 3; Table 3; Data S2) are necessary to gain reliable inferences of potential microbial functions. Fourth, along with the potential pitfall discussed above, network modules can differ not only in properties related to crop disease/health status but also in those associated with crop plant identity or cropland management (Figs. S6-7). Again, findings in broad-geographic-scale analyses need to be supplemented by insights from local-scale observations (Fig. 3). Fifth, 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 (62–64). Sixth, 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.

Expanding the comparative microbiome analysis to different geographic regions and agroecosystem management practices will contribute to a more comprehensive understanding of microbiome structure and function. 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 (65–68) will reorganize our understanding of relationship between microbiome compositions and functions. In conclusion, this data-driven research lays the groundwork for understanding fundamental mechanisms in soil ecosystems, offering innovative strategies for the design of 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 (69) 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 (5'- GTG YCA GCM GCC GCG GTA A -3') and the reverse primer 806rB (5'- GGA CTA CNV GGG TWT CTA AT -3') were used as described elsewhere (54). 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; $1.0 \times 10^{-4}$, $5.0 \times 10^{-5}$, $2.0 \times 10^{-5}$, $1.0 \times 10^{-5}$, and $5.0 \times 10^{-6}$ nM; Table S1) to the PCR master mix solution as detailed elsewhere (54). 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 (54).

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 (5'- CTH GGT CAT TTA GAG GAA STA A -3') and the reverse primer ITS2_KYO2 (5' – TTY RCT RCG TTC TTC ATC - 3') (70). 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 ($1.0 \times 10^{-5}$, $7.0 \times 10^{-6}$, $5.0 \times 10^{-6}$, $2.0 \times 10^{-6}$, and $1.0 \times 10^{-6}$ 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 (71). The sequencing data were deposited to DNA Data Bank of Japan (DDBJ) (Bioproject accession no.: PSUB018361). The removal of low-quality sequences and OTU inferences were done using DADA2 (72) 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. The amplicon sequence variants (ASVs) obtained from the DADA2 pipeline were clustered using the vsearch v2.21.1 program (73) with the 98% and 97% cutoff sequence similarity for prokaryotes and fungi, respectively. Taxonomic annotation of the obtained prokaryotic and fungal OTUs were conducted based on the SILVA 138 SSU (74) and the UNITE all_ 25.07.2023 (75) databases, respectively, with the assignTaxonomy function of DADA2. The OTUs 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 × OTU matrix, in which a cell entry depicted the number of sequencing reads of an OTU in a sample. The samples with less than 1,000 reads were discarded from the matrices. 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 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. For the sample × OTU matrix, centered log-ratio (CLR) transformation (76–78) was performed using the ALDEx2 v1.35.0 package (79) of R.

In total, prokaryotic and fungal community data were obtained for 2,676 and 2,477 samples, respectively. For fungal OTUs, putative functional groups (e.g., “plant pathogen”) were inferred using the program FungalTraits (80). The estimation of DNA concentrations of the prokaryotic 16S rRNA and fungal ITS regions was performed, respectively, based on the

406 calibration with the standard DNA gradients (artificial DNA variants introduced to the PCR
407 master mix solutions) using the bioinformatic pipeline detailed elsewhere (54).

408

409 **Calculation of prokaryote/fungus ratio**

410 Based on the estimated concentrations of prokaryotic 16S rRNA and fungal ITS sequences in
411 template DNA solutions, we calculated the ratio of prokaryotic DNA concentrations to fungal
412 DNA concentrations in respective samples (prokaryote/fungus ratio) as follows:

413 $$\log\left(\frac{\text{prokaryotic 16S rRNA gene concentration(DNA copies/}\mu\text{L)}}{\text{fungal ITS gene concentration (DNA copies/}\mu\text{L)}}\right).$$

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

422

423 **Microbiome structure and crop disease prevalence**

424 For each of the prokaryotic and fungal datasets, permutational analyses of variance
425 (PERMANOVA) (81) were performed to examine associations between family-level community
426 compositions and variables in the metadata. Two types of PERMANOVA models were
427 constructed based on the Euclid distance ($\beta$-diversity) calculated for the CLR-transformed
428 datasets (1,000 iterations). Specifically, one is constituted by categorical explanatory variables
429 (crop plant, former crop plant, soil taxonomy, research site, and sampling month), while the other
430 included continuous explanatory variables (soil pH, electrical conductivity, C/N ratio, and
431 available phosphorous concentration, prokaryote/fungus ratio, latitude, and longitude).

432 To reduce the dimensions of the community compositional data, a principal component
433 analysis (PCA) was performed based on the Euclid distance data mentioned above. For each PCA
434 axis (axes 1 to 5) in each of the prokaryotic and fungal analyses, Pearson's correlation with each
435 chemical environmental factor (soil pH, electrical conductivity, C/N ratio, and available
436 phosphorous concentration) was calculated.

437 We then evaluated how community structure of prokaryotes and fungi were associated with
438 crop disease. For each of the prokaryotic and fungal datasets, a generalized linear mixed model
439 (GLMM) of crop-disease level (disease level 1 vs. disease levels 2-5) was constructed by
440 including the PCoA axes 1 to 5 as fixed effects. Sampling month and the identity of crop plant
441 species and experimental/research purposes in the metadata were set as random effects. A logit-
442 link function and binomial errors was assumed in the GLMM after converting the response
443 variable into a binary format [disease level 1 (= 1) vs. disease levels 2–5 (= 0)]. The analysis was
444 performed with the "glmer" function of the R lme4 package (82).

445

446 **Microbes associated with crop disease/health status**

447 For each microbial OTU constituting the modules, we evaluated specificity of occurrences in
448 samples differing in crop disease levels based on a randomization analysis. For the calculation,
449 the original sample × OTU matrices of prokaryotes and fungi were respectively rarefied to 1,000
450 reads per sample, being merged into an input data matrix. Within the combined sample × OTU
451 matrix, samples were categorized into the two crop disease levels (disease level 1 vs. disease
452 levels 2-5). Mean read counts across samples displaying each of the two disease levels were then
453 calculated for each OTU. Meanwhile, mean read counts for respective disease levels were
454 calculated as well for randomized matrices, in which disease labels of the samples were shuffled
455 (10,000 permutations). For *i*-th OTU, standardized specificity to disease level *j* ($s_{ij}$) was obtained
456 as follows:

457 $$s_{ij} = \frac{O_{ij} - Mean(R_{ij})}{SD(R_{ij})},$$

458 where $O_{ij}$ and $R_{ij}$ is the mean read counts of *i*-th OTU across disease-level-*j* samples in the
459 observed and randomized matrices, respectively, and $Mean(R_{ij})$ and $SD(R_{ij})$ indicate mean and
460 standard deviation across the randomized matrices. The *P* values obtained based on the
461 randomization analysis were adjusted with the Benjamini-Hochberg method [i.e., false discovery
462 rate (FDR)]. The relationship between the standardized specificity index and FDR is shown in
463 Figure S4. This randomization approach was also applied to the analyses of each OTU's
464 specificity to crop plant identity and that to experimental/research purpose identity (Figs. S6-7).

465 The specificity of microbial OTUs to crop disease levels was performed as well at the local
466 scale. Specifically, in each of the six sub-datasets representing unique combinations of research
467 sites, crop plant species, and experimental/research purposes, the abovementioned randomization
468 analysis was performed: each sub-dataset included 69 to 198 soil samples (Data S2). For the
469 OTUs showing exceptionally strong specificity to the minimal crop disease level (standardized
470 specificity score ≥ 6.0; FDR < 0.0001), supplemental analyses of generalized linear models
471 (GLMs) were conducted. In each GLM of crop disease/health status (disease level 1 vs. disease
472 levels 2-5) with a logit-link function with binomial errors, the relative abundance of a target OTU
473 was included as an explanatory variable.

474

475 **Microbe-to-microbe network**

476 To infer potential interactions between microbial OTUs, the algorithm of sparse inverse
477 covariance estimation for ecological association inference (SPIEC-EASI) was applied based on
478 the Meinshausen-Bühlmann (MB) method as implemented in the SpiecEasi package (39) of R. In
479 total, 2,305 soil samples from which both prokaryotic and fungal community data were available
480 were subjected to the analysis. Note that CLR-transformation was performed internally with the
481 "spiec.easi" function. The network inference based on co-occurrence patterns allowed us to
482 detect pairs of microbial OTUs that potentially interact with each other in facilitative ways and/or
483 those that might share ecological niches (e.g., preference for edaphic factors). Because estimation
484 of co-occurrence patterns was not feasible for rare nodes, the prokaryotic and fungal OTUs that
485 appeared in more than 10 % of the sequenced samples were included in the input matrix of the
486 network analysis. Network modules, within which closely associated OTUs were interlinked with

each other, were identified with the algorithm based on edge betweenness (83) using the igraph package (84) of R. For each module in the inferred co-occurrence network, mean standardized specificity to disease level 1 were calculated across constituent OTUs.

To explore potential keystone microbes within the network, we scored respective OTUs on the basis of their topological positions. Among the indices used for evaluating OTUs, betweenness centrality (85), which measures the extent to which a given nodes (OTU) 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 OTUs based on their within-module degree and among-module connectivity (86). 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 (87) was calculated for respective nodes.

## Data availability

The 16S rRNA and ITS sequencing data are available from the DNA Data Bank of Japan (DDBJ accession: DRA015491 and DRA015506). 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.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

20. Hardoim PR, van Overbeek LS, Elsas JD van. 2008. Properties of bacterial endophytes and 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.

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

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

617 35. Durán P, Thiergart T, Garrido-Oter R, Agler M, Kemen E, Schulze-Lefert P, Hacquard S.
618 2018. Microbial Interkingdom Interactions in Roots Promote Arabidopsis Survival. Cell
619 175:973-983.e14.

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

622 37. Legrand F, Chen W, Cobo-Díaz JF, Picot A, Le Floch G. 2019. Co-occurrence analysis
623 reveal that biotic and abiotic factors influence soil fungistasis against *Fusarium*
624 *graminearum*. FEMS Microbiol Ecol 95:fiz056.

625 38. Kerdraon L, Barret M, Laval V, Suffert F. 2019. Differential dynamics of microbial
626 community networks help identify microorganisms interacting with residue-borne
627 pathogens: The case of *Zymoseptoria tritici* in wheat. Microbiome 7:125.

628 39. Kurtz ZD, Müller CL, Miraldi ER, Littman DR, Blaser MJ, Bonneau RA. 2015. Sparse
629 and compositionally robust inference of microbial ecological networks. PLoS Comput Biol
630 11:e1004226.

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

633 41. Summerell BA. 2019. Resolving *Fusarium*: Current Status of the Genus. Annu Rev
634 Phytopathol 57:323–339.

635 42. Gordon TR. 2017. *Fusarium oxysporum* and the *Fusarium* wilt syndrome. Annu Rev
636 Phytopathol 55:23–39.

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

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

642 45. Lebedeva E V., Hatzenpichler R, Pelletier E, Schuster N, Hauzmayer S, Bulaev A,
643 Grigor'eva N V., Galushko A, Schmid M, Palatinszky M, Le Paslier D, Daims H, Wagner
644 M. 2013. Enrichment and genome sequence of the group I.1a ammonia-oxidizing archaeon
645 "Ca. *Nitrosotenuis uzonensis*" representing a clade globally distributed in thermal habitats.
646 PLoS One 8:e80835.

647 46. Olanrewaju OS, Babalola OO. 2019. *Streptomyces*: implications and interactions in plant
648 growth promotion. Appl Microbiol Biotechnol https://doi.org/10.1007/s00253-018-09577-
649 y.

650 47. Li F, Chen L, Redmile-Gordon M, Zhang J, Zhang C, Ning Q, Li W. 2018. *Mortierella*
651 *elongata*'s roles in organic agriculture and crop growth promotion in a mineral soil. Land
652 Degrad Dev 29:1642–1651.

653 48. Ozimek E, Hanaka A. 2021. *Mortierella* species as the plant growth-promoting fungi
654 present in the agricultural soils. Agriculture (Switzerland)
655 https://doi.org/10.3390/agriculture11010007.

656 49. Poveda J. 2021. *Trichoderma* as biocontrol agent against pests: New uses for a
657 mycoparasite. Biological Control 159:104634.

658 50. Sood M, Kapoor D, Kumar V, Sheteiwy MS, Ramakrishnan M, Landi M, Araniti F,
659 Sharma A. 2020. *Trichoderma*: The "secrets" of a multitalented biocontrol agent. Plants
660 9:762.

661 51. AbdElfatah HAS, Sallam NMA, Mohamed MS, Bagy HMMK. 2021. *Curvularia lunata* as
662 new causal pathogen of tomato early blight disease in Egypt. Mol Biol Rep 48:3001–3006.

663 52. Pétriacq P, Stassen JHM, Ton J. 2016. Spore density determines infection strategy by the
664 plant pathogenic fungus *Plectosphaerella cucumerina*. Plant Physiol 170:2325–2339.

665 53. Ushio M. 2022. Interaction capacity as a potential driver of community diversity.
666 Proceedings of the Royal Society B: Biological Sciences 289:20212690.

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

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

672 56. Menta C, Remelli S. 2020. Soil health and arthropods: From complex system to
673 worthwhile investigation. Insects 11:54.

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

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

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

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

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

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

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

693 64. Wu X, Rensing C, Han D, Xiao K-Q, Dai Y, Tang Z, Liesack W, Peng J, Cui Z, Zhang F.
694 2022. Genome-Resolved Metagenomics Reveals Distinct Phosphorus Acquisition
695 Strategies between Soil Microbiomes. mSystems 7:e01107-21.

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

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

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

68. 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.

69. 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.

70. 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.

71. 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.

72. 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.

73. Rognes T, Flouri T, Nichols B, Quince C, Mahé F. 2016. VSEARCH: a versatile open source tool for metagenomics. PeerJ 4:e2584.

74. 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.

75. Nilsson RH, Larsson KH, Taylor AFS, Bengtsson-Palme J, Jeppesen TS, Schigel D, Kennedy P, Picard K, Glöckner FO, Tedersoo L, Saar I, Kõljalg U, Abarenkov K. 2019. The UNITE database for molecular identification of fungi: Handling dark taxa and parallel taxonomic classifications. Nucleic Acids Res 47:D259–D264.

76. Quinn TP, Erb I, Gloor G, Notredame C, Richardson MF, Crowley TM. 2019. A field guide for the compositional analysis of any-omics data. Gigascience 8:1–14.

77. Gloor GB, Macklaim JM, Pawlowsky-Glahn V, Egozcue JJ. 2017. Microbiome datasets are compositional: And this is not optional. Front Microbiol https://doi.org/10.3389/fmicb.2017.02224.

78. Aitchison J, Barceló-Vidal C, Martín-Fernández JA, Pawlowsky-Glahn V. 2000. Logratio analysis and compositional distance. Math Geol 32:271–275.

79. Fernandes AD, Reid JNS, Macklaim JM, McMurrough TA, Edgell DR, Gloor GB. 2014. Unifying the analysis of high-throughput sequencing datasets: Characterizing RNA-seq, 16S rRNA gene sequencing and selective growth experiments by compositional data analysis. Microbiome 2:15.

80. 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.

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

82. Bates D, Mächler M, Bolker BM, Walker SC. 2015. Fitting linear mixed-effects models using lme4. J Stat Softw 67:1–48.

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

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

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

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

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

774 **ACKNOWLEDGEMENTS**

775 We are grateful to the reviewers whose comments improved the manuscript. We thank the
776 SuperComputer System, Institute for Chemical Research, Kyoto University for the use of super
777 computers. This work was financially supported by JST PRESTO (JPMJPR16Q6), JST FOREST
778 (JPMJFR2048), JST PRESTO (JPMJCR23N5), Human Frontier Science Program
779 (RGP0029/2019), JSPS Grant-in-Aid for Scientific Research (20K20586) and NEDO Moonshot
780 Research and Development Program (JPNP18016) to H.T., JSPS Grant-in-Aid for Scientific
781 Research (20K06820 and 20H03010) to K.S., and JSPS Fellowship to H.F.
782

783 **Table 1 |** Effects of environmental variables on prokaryotic/fungal community structure. For each
784 set of categorical/continuous environmental variables, a PERMANOVA was performed for each
785 of the prokaryotic and fungal community datasets.

786

| Model | Dataset | Variable | df | $R^2$ | *F* | *P* |
|---|---|---|---|---|---|---|
| Categorical variables | Prokaryotes | Research site | 34 | 0.051 | 5.13 | < 0.001 |
| | | Month | 11 | 0.014 | 4.36 | < 0.001 |
| | | Crop | 16 | 0.006 | 1.21 | 0.030 |
| | | Former crop | 22 | 0.024 | 3.78 | < 0.001 |
| | | Soil category | 10 | 0.006 | 2.12 | < 0.001 |
| | | Residual | 2301 | 0.673 | | |
| | | Total | 2400 | 1.000 | | |
| | Fungi | Research site | 33 | 0.064 | 6.91 | < 0.001 |
| | | Month | 11 | 0.012 | 3.96 | < 0.001 |
| | | Crop | 15 | 0.006 | 1.54 | < 0.001 |
| | | Former crop | 20 | 0.023 | 4.14 | < 0.001 |
| | | Soil category | 10 | 0.005 | 1.75 | < 0.001 |
| | | Residual | 2109 | 0.591 | | |
| | | Total | 2206 | 1.000 | | |
| Continuous variables | Prokaryotes | pH | 1 | 0.011 | 16.52 | < 0.001 |
| | | Electrical conductivity | 1 | 0.009 | 14.09 | < 0.001 |
| | | Available P | 1 | 0.009 | 13.81 | 0.626 |
| | | C/N ratio | 1 | 0.004 | 6.19 | < 0.001 |
| | | Latitude | 1 | 0.006 | 9.00 | < 0.001 |
| | | Longitude | 1 | 0.008 | 11.41 | < 0.001 |
| | | Prokaryote/fungus ratio | 1 | 0.004 | 5.98 | < 0.001 |
| | | Residual | 1408 | 0.936 | | |
| | | Total | 1415 | 1.000 | | |
| | Fungi | | | | | |
| | | pH | 1 | 0.013 | 19.64 | < 0.001 |
| | | Electrical conductivity | 1 | 0.011 | 17.49 | < 0.001 |
| | | Available P | 1 | 0.009 | 13.38 | 0.477 |
| | | C/N ratio | 1 | 0.008 | 12.74 | < 0.001 |
| | | Latitude | 1 | 0.016 | 25.58 | < 0.001 |
| | | Longitude | 1 | 0.017 | 26.37 | 0.230 |
| | | Prokaryote/fungus ratio | 1 | 0.009 | 13.61 | < 0.001 |
| | | Residual | 1408 | 0.904 | | |
| | | Total | 1415 | 1 | | |

787

788

789 **Table 2 |** Relationship between prokaryotic/fungal community structure on the disease level of
790 crop plants. A GLMM of crop plants' disease level (disease level 1 vs. disease levels 2-5) with a
791 logit-link function and binomial errors was constructed by setting principal components of
792 prokaryotic/fungal community structure (Fig. 2) as explanatory variables (fixed effects). The
793 identity of experimental/research purposes, sampling month, and crop plant species were
794 included as random effects in the GLMM.

795

| Dataset | Variable | *z* | *P* |
|---|---|---|---|
| Prokaryotes | PC1 | 1.59 | 0.1111 |
| (*N* = 1,379) | PC2 | -1.65 | 0.1000 |
| | PC3 | 1.82 | 0.0684 |
| | PC4 | -2.32 | 0.0205 |
| | PC5 | 3.98 | 0.0001 |
| Fungi | PC1 | 1.52 | 0.1281 |
| (*N* = 1,320) | PC2 | 1.39 | 0.1656 |
| | PC3 | -2.11 | 0.0348 |
| | PC4 | 2.62 | 0.0089 |
| | PC5 | -0.84 | 0.4002 |

796

**Table 3** | Prokaryotic and fungal OTUs showing highest associations with crop health status within local croplands. Among the six sub-datasets representing unique combinations of research sites, crop plant species, and research experimental/research purposes, OTUs showing strongest specificity to the minimal crop disease level (*z*-standardized specificity to disease level 1 ≥ 6.0) were observed in two sub-datasets ("eggplant in Kuki City" and "tomato in Kashihara City"). The OTUs are shown with the NCBI BLAST top-hit results. See Data S2 for the full results.

| | | | | Specificity to disease level 1 | | BLAST top-hit results | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Site | Crop | Experiment/research identity | ID | Score | FDR | Scientific Name | Query cover (%) | E value | Identity (%) | Accession |
| Kuki | Eggplant | Control of bacterial wilt | Bac_00034 | 6.53 | < 0.0001 | *Denitratisoma oestradiolicum* | 100 | 9.0E-103 | 95.1 | KF810120.1 |
| | | | Bac_00044 | 6.67 | < 0.0001 | *Nocardioides cynanchi* | 100 | 4.0E-121 | 99.6 | CP044344.1 |
| | | | Bac_00061 | 7.12 | < 0.0001 | *Piscinibacter aquaticus* | 100 | 4.0E-121 | 99.6 | KY284087.1 |
| | | | Bac_00224 | 7.52 | < 0.0001 | *Dongia* sp. | 100 | 9.0E-123 | 100.0 | AB835804.1 |
| | | | Bac_00237 | 8.76 | < 0.0001 | *Chondromyces robustus* | 100 | 2.0E-89 | 91.8 | AJ233942.2 |
| | | | Fun_0059 | 6.83 | < 0.0001 | *Moesziomyces aphidis* | 100 | 9.0E-123 | 100.0 | MH777069.1 |
| | | | Fun_1871 | 6.27 | < 0.0001 | *Pseudeurotium bakeri* | 100 | 3.0E-122 | 100.0 | MK911621.1 |
| | | | Fun_3676 | 7.72 | < 0.0001 | *Cladosporium proteacearum* | 100 | 8.0E-118 | 100.0 | OR857360.1 |
| | | | Fun_3688 | 6.38 | < 0.0001 | *Nigrospora sphaerica* | 100 | 8.0E-118 | 100.0 | OP113684.1 |
| | | | Fun_3993 | 7.01 | < 0.0001 | *Fusarium equiseti* | 100 | 3.0E-112 | 99.6 | MT588081.1 |
| | | | Fun_4311 | 8.75 | < 0.0001 | *Gibellulopsis nigrescens* | 100 | 5.0E-110 | 100.0 | OP498056.1 |
| Kashihara | Tomato | Control of *Fusarium* wilt | Bac_00031 | 6.74 | < 0.0001 | *Ramlibacter algicola* | 100 | 9.0E-123 | 100.0 | NR_175506.1 |
| | | | Bac_00861 | 6.98 | < 0.0001 | *Rhizomicrobium* sp. | 100 | 4.0E-116 | 98.4 | LN876448.1 |
| | | | Fun_0056 | 6.28 | < 0.0001 | *Corynascus sepedonium* | 100 | 9.0E-123 | 100.0 | OW986289.1 |

**Table 4** | Representative prokaryotic and fungal OTUs in network modules with highly positive/negative associations with crop plant health. In each of the modules 2, 6, and 8 (Fig. 4), the top-five OTUs with the highest specificity to the minimal crop disease level (specificity to disease level 1; see Fig. S4 for the relationship between the specificity score and FDR). For each OTU, network degree, betweenness centrality, within-module degree (*z*-standardized), and among-module connectivity (Fig. 6) are presented with the NCBI BLAST top-hit results. Likewise, in each of the modules 1 and 7 (Fig. 4), the top-five OTUs negatively associated with the minimal crop disease level are shown. See Data S3 for the full results.

| | | Specificity to disease level 1 | | Network scores | | | | BLAST top-hit results | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Module | OTU ID | Score | FDR | Degree | Betweenness | Within-module degree | Module connectivity | Scientific Name | Query cover (%) | E value | Identity (%) | Accession |
| Module 2 | Bac_00147 | 7.526 | < 0.0001 | 4 | 0.001 | -0.693 | 0.625 | *Botrimarina hoheduenensis* | 99 | 2.0E-90 | 92.2 | NR_173585.1 |
| (positive) | Bac_00469 | 6.307 | < 0.0001 | 5 | 0.001 | -1.214 | 0.720 | *Luteitalea pratensis* | 100 | 9.0E-108 | 96.3 | NR_156918.1 |
| | Bac_00061 | 4.657 | < 0.0001 | 3 | 0.001 | -1.214 | 0.667 | *Piscinibacter aquaticus* | 100 | 4.0E-121 | 99.6 | KY284087.1 |
| | Bac_00002 | 4.652 | < 0.0001 | 10 | 0.019 | 0.347 | 0.720 | *Arthrobacter globiformis* | 100 | 9.0E-123 | 100.0 | OR780585.1 |
| | Bac_00463 | 4.166 | < 0.0001 | 5 | 0.002 | 0.347 | 0.320 | *Flavobacterium sufflavum* | 100 | 9.0E-123 | 100.0 | NR_171469.1 |
| Module 6 | Bac_00025 | 9.477 | < 0.0001 | 5 | 0.001 | 0.189 | 0.000 | *Gemmatimonas aurantiaca* | 100 | 4.0E-86 | 91.0 | KF228166.1 |
| (positive) | Bac_00258 | 7.205 | < 0.0001 | 14 | 0.011 | 2.587 | 0.357 | *Thermanaerothrix daxensis* | 100 | 4.0E-76 | 88.6 | NR_117865.1 |
| | Fun_4447 | 7.151 | < 0.0001 | 10 | 0.017 | -0.610 | 0.780 | *Plectosphaerella cucumerina* | 100 | 2.0E-108 | 100.0 | MT529301.1 |
| | Fun_0071 | 7.093 | < 0.0001 | 6 | 0.003 | -1.009 | 0.722 | *Schizothecium miniglutinans* | 100 | 2.0E-104 | 95.5 | MW472119.1 |
| | Bac_00182 | 6.787 | < 0.0001 | 21 | 0.037 | 1.788 | 0.649 | *Brevitalea aridisoli* | 100 | 9.0E-93 | 92.7 | NR_151987.1 |
| Module 8 | Bac_00294 | 9.679 | < 0.0001 | 7 | 0.002 | 0.000 | 0.571 | *Luteitalea pratensis* | 100 | 4.0E-81 | 89.8 | NR_156918.1 |
| (positive) | Bac_00171 | 9.499 | < 0.0001 | 11 | 0.012 | 1.063 | 0.628 | *Membranihabitans marinus* | 100 | 9.0E-88 | 91.4 | OL441066.1 |
| | Fun_4070 | 8.282 | < 0.0001 | 10 | 0.006 | -0.532 | 0.700 | *Mortierella* sp. | 80 | 3.0E-87 | 99.5 | OP799297.1 |
| | Bac_00138 | 7.904 | < 0.0001 | 20 | 0.023 | 1.595 | 0.715 | *Ectothiorhodospira mobilis* | 100 | 4.0E-86 | 91.0 | MG264243.1 |
| | Bac_00339 | 7.297 | < 0.0001 | 5 | 0.003 | -1.063 | 0.480 | *Vicinamibacter silvestris* | 100 | 2.0E-74 | 88.2 | NR_151905.1 |
| Module 1 | Bac_00165 | -9.211 | < 0.0001 | 19 | 0.035 | 0.903 | 0.582 | *Ureibacillus suwonensis* | 96 | 1.0E-76 | 89.5 | JX914499.1 |
| (negative) | Bac_00004 | -7.370 | < 0.0001 | 17 | 0.021 | 1.134 | 0.512 | *Nonomuraea phyllanthi* | 100 | 9.0E-123 | 100.0 | CP045572.1 |
| | Bac_00010 | -7.370 | < 0.0001 | 4 | 0.001 | -0.714 | 0.375 | *Streptomyces ardesiacus* | 100 | 9.0E-123 | 100.0 | OR873433.1 |
| | Bac_00282 | -7.289 | < 0.0001 | 9 | 0.009 | 0.672 | 0.000 | *Hyphomicrobium aestuarii* | 100 | 4.0E-111 | 97.1 | NR_104954.1 |
| | Bac_00036 | -6.531 | < 0.0001 | 16 | 0.011 | 1.596 | 0.305 | *Nitrolancea hollandica* | 100 | 9.0E-88 | 91.5 | MW367426.1 |
| Module 7 | Fun_4028 | -8.585 | < 0.0001 | 5 | 0.000 | -0.490 | 0.320 | *Fusarium oxysporum* | 100 | 6.0E-114 | 100.0 | MF281350.2 |
| (negative) | Arc_006 | -8.544 | < 0.0001 | 8 | 0.006 | -0.490 | 0.594 | *Nitrososphaera viennensis* | 100 | 9.0E-123 | 100.0 | NR_134097.1 |
| | Bac_00195 | -8.079 | < 0.0001 | 14 | 0.019 | 1.218 | 0.449 | *Chromobacterium amazonense* | 100 | 3.0E-88 | 91.5 | OQ061977.1 |
| | Bac_00062 | -7.281 | < 0.0001 | 13 | 0.011 | 1.218 | 0.379 | *Sphingomonas segetis* | 100 | 4.0E-116 | 98.4 | NR_175421.1 |
| | Bac_00098 | -6.557 | < 0.0001 | 18 | 0.026 | 1.218 | 0.537 | *Geodermatophilus normandii* | 100 | 9.0E-118 | 98.8 | MT214187.1 |

**Table 5** | Within- and among-module hubs in the network. In each of the modules highlighted in Table 4 and Figure 6, the top-three OTUs with the highest within-module degree (*z*-standardized) or among-module connectivity are shown. See Data S3 for the full results.

| Focused | | | Specificity to disease level 1 | | Network scores | | | | BLAST top-hit results | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| index | Module | OTU ID | Score | FDR | Degree | Betweenness | Within-module degree | Module connectivity | Scientific Name | Query cover (%) | E value | Identity (%) | Accession |
| Within | Module 2 | Bac_00011 | 3.104 | 0.0013 | 13 | 0.018 | 1.907 | 0.663 | *Pseudomonas izuensis* | 100 | 9.0E-123 | 100.0 | OR841525.1 |
| | (positive) | Bac_00057 | 2.741 | 0.0063 | 6 | 0.002 | 1.387 | 0.000 | *Massilia violaceinigra* | 100 | 9.0E-123 | 100.0 | MT373681.1 |
| | | Bac_00002 | 4.652 | < 0.0001 | 10 | 0.019 | 0.347 | 0.720 | *Arthrobacter globiformis* | 100 | 9.0E-123 | 100.0 | OR780585.1 |
| | Module 6 | Bac_00258 | 7.205 | < 0.0001 | 14 | 0.011 | 2.587 | 0.357 | *Thermanaerothrix daxensis* | 100 | 4.0E-76 | 88.6 | NR_117865.1 |
| | (positive) | Bac_00182 | 6.787 | < 0.0001 | 21 | 0.037 | 1.788 | 0.649 | *Brevitalea aridisoli* | 100 | 9.0E-93 | 92.7 | NR_151987.1 |
| | | Arc_004 | 6.089 | < 0.0001 | 19 | 0.027 | 1.388 | 0.742 | *Nitrosotenuis chungbukensis* | 100 | 9.0E-123 | 100.0 | CP130341.1 |
| | Module 8 | Bac_00138 | 7.904 | < 0.0001 | 20 | 0.023 | 1.595 | 0.715 | *Ectothiorhodospira mobilis* | 100 | 4.0E-86 | 91.0 | MG264243.1 |
| | (positive) | Bac_00260 | 5.320 | < 0.0001 | 9 | 0.006 | 1.595 | 0.346 | *Crenobacter cavernae* | 100 | 4.0E-96 | 93.5 | CP031337.1 |
| | | Bac_00171 | 9.499 | < 0.0001 | 11 | 0.012 | 1.063 | 0.628 | *Membranihabitans marinus* | 100 | 9.0E-88 | 91.4 | OL441066.1 |
| | Module 1 | Bac_00014 | -4.481 | < 0.0001 | 22 | 0.025 | 3.676 | 0.000 | *Gemmatimonas phototrophica* | 100 | 2.0E-89 | 91.8 | CP011454.1 |
| | (negative) | Bac_00055 | -5.245 | < 0.0001 | 18 | 0.015 | 2.752 | 0.000 | *Microbispora rosea* | 100 | 9.0E-123 | 100.0 | MN826183.1 |
| | | Fun_0033 | -1.088 | 0.1752 | 17 | 0.026 | 2.521 | 0.000 | *Trichoderma atroviride* | 100 | 9.0E-123 | 100.0 | MN429074.1 |
| | Module 7 | Bac_00045 | -1.380 | 0.1103 | 17 | 0.017 | 2.356 | 0.304 | *Pseudolabrys taiwanensis* | 100 | 9.0E-113 | 97.6 | CP031417.1 |
| | (negative) | Bac_00116 | -3.636 | 0.0003 | 13 | 0.006 | 2.071 | 0.000 | *Streptomyces spinosirectus* | 100 | 9.0E-123 | 100.0 | CP090447.1 |
| | | Bac_00218 | -5.182 | < 0.0001 | 13 | 0.010 | 1.502 | 0.260 | *Dyella ginsengisoli* | 100 | 9.0E-123 | 100.0 | KY228986.1 |
| Among | Module 2 | Bac_00039 | -1.782 | 0.0537 | 11 | 0.027 | 0.347 | 0.744 | *Arenimonas daechungensis* | 100 | 4.0E-116 | 98.4 | NR_109442.1 |
| | (positive) | Bac_00002 | 4.652 | < 0.0001 | 10 | 0.019 | 0.347 | 0.720 | *Arthrobacter globiformis* | 100 | 9.0E-123 | 100.0 | OR780585.1 |
| | | Bac_00469 | 6.307 | < 0.0001 | 5 | 0.001 | -1.214 | 0.720 | *Luteitalea pratensis* | 100 | 9.0E-108 | 96.3 | NR_156918.1 |
| | Module 6 | Fun_4447 | 7.151 | < 0.0001 | 10 | 0.017 | -0.610 | 0.780 | *Plectosphaerella cucumerina* | 100 | 2.0E-108 | 100.0 | MT529301.1 |
| | (positive) | Fun_3979 | 5.139 | < 0.0001 | 6 | 0.008 | -1.409 | 0.778 | *Enterocarpus grenotii* | 100 | 8.0E-113 | 99.6 | OU989357.1 |
| | | Arc_004 | 6.089 | < 0.0001 | 19 | 0.027 | 1.388 | 0.742 | *Nitrosotenuis chungbukensis* | 100 | 9.0E-123 | 100.0 | CP130341.1 |
| | Module 8 | Bac_00138 | 7.904 | < 0.0001 | 20 | 0.023 | 1.595 | 0.715 | *Ectothiorhodospira mobilis* | 100 | 4.0E-86 | 91.0 | MG264243.1 |
| | (positive) | Bac_00180 | 2.213 | 0.0224 | 10 | 0.008 | 0.000 | 0.700 | *Rubrobacter spartanus* | 100 | 4.0E-66 | 86.3 | NR_158052.1 |
| | | Fun_4070 | 8.282 | < 0.0001 | 10 | 0.006 | -0.532 | 0.700 | *Mortierella kuhlmanii* | 100 | 9.0E-53 | 84.5 | MH860115.1 |
| | Module 1 | Fun_0043 | 0.485 | 0.3420 | 7 | 0.008 | -0.714 | 0.735 | *Curvularia senegalensis* | 100 | 9.0E-123 | 100.0 | MT476857.1 |
| | (negative) | Fun_3610 | 1.694 | 0.0649 | 12 | 0.014 | -0.714 | 0.708 | *Acremonium alternatum* | 100 | 2.0E-118 | 100.0 | MT529342.1 |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Bac_00042 | 1.601 | 0.0774 | 12 | 0.020 | -0.252 | 0.625 | *Fulvivirgaceae bacterium* | 100 | 9.0E-108 | 96.3 | OQ733332.1 |
| Module 7 | Fun_0027 | 2.861 | 0.0040 | 5 | 0.004 | -1.059 | 0.640 | *Trichoderma virens* | 100 | 9.0E-123 | 100.0 | MT530036.1 |
| (negative) | Bac_00032 | -1.253 | 0.1331 | 4 | 0.001 | -1.059 | 0.625 | *Streptomyces cyaneus* | 100 | 9.0E-123 | 100.0 | OR807486.1 |
| | Bac_00070 | 2.191 | 0.0242 | 4 | 0.005 | -1.059 | 0.625 | *Luteimonas aestuarii* | 100 | 9.0E-123 | 100.0 | OQ255277.1 |

814

815

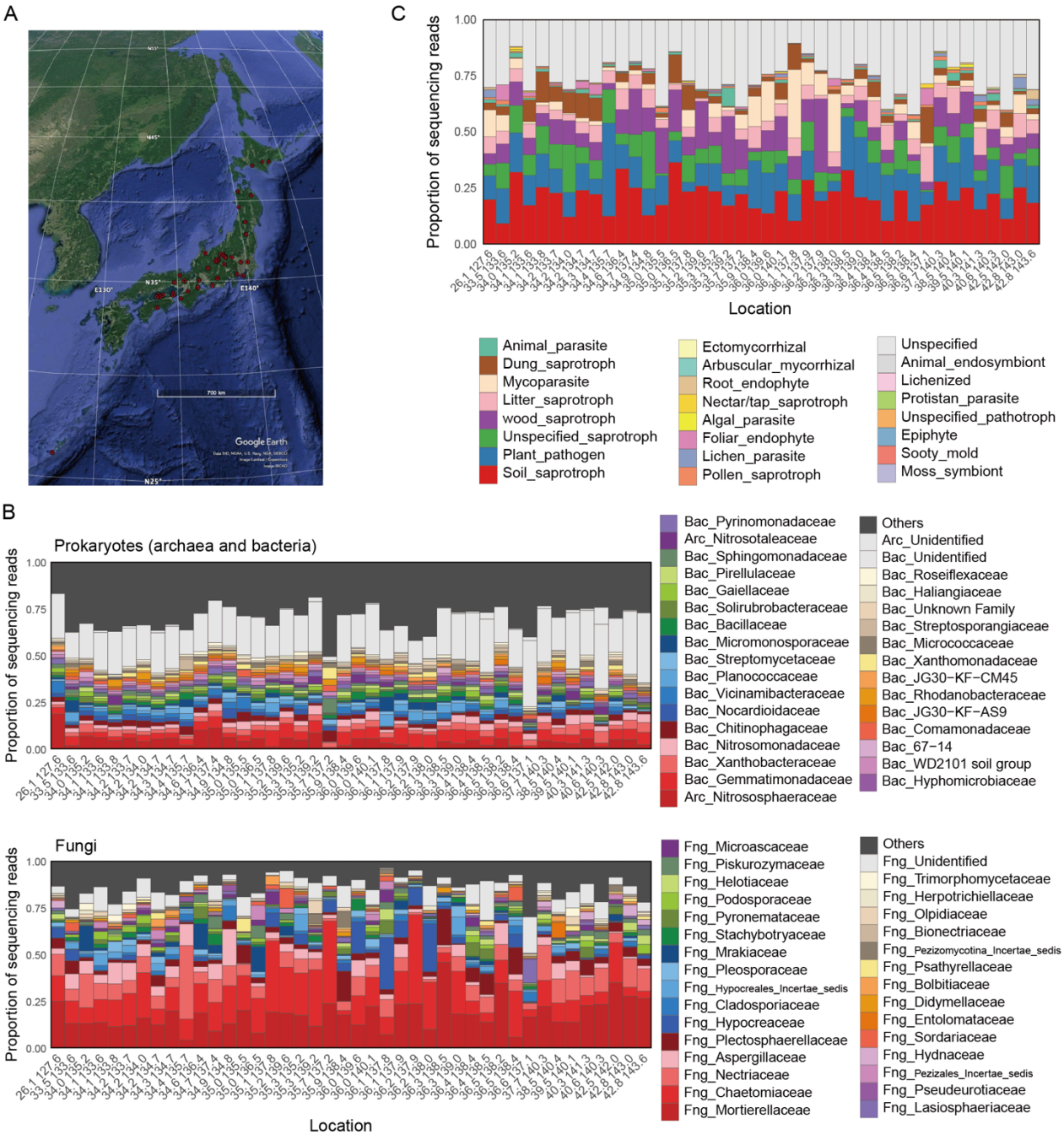


**Fig. 1 |** Comparison of soil microbiome structure across the Japan Archipelago. (A) Map of research sites across the Japan Archipelago. The 2,903 soil samples were grouped into 42 research sits were when their latitude and longitude profiles were round to one decimal place. (B) Taxonomic compositions 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.

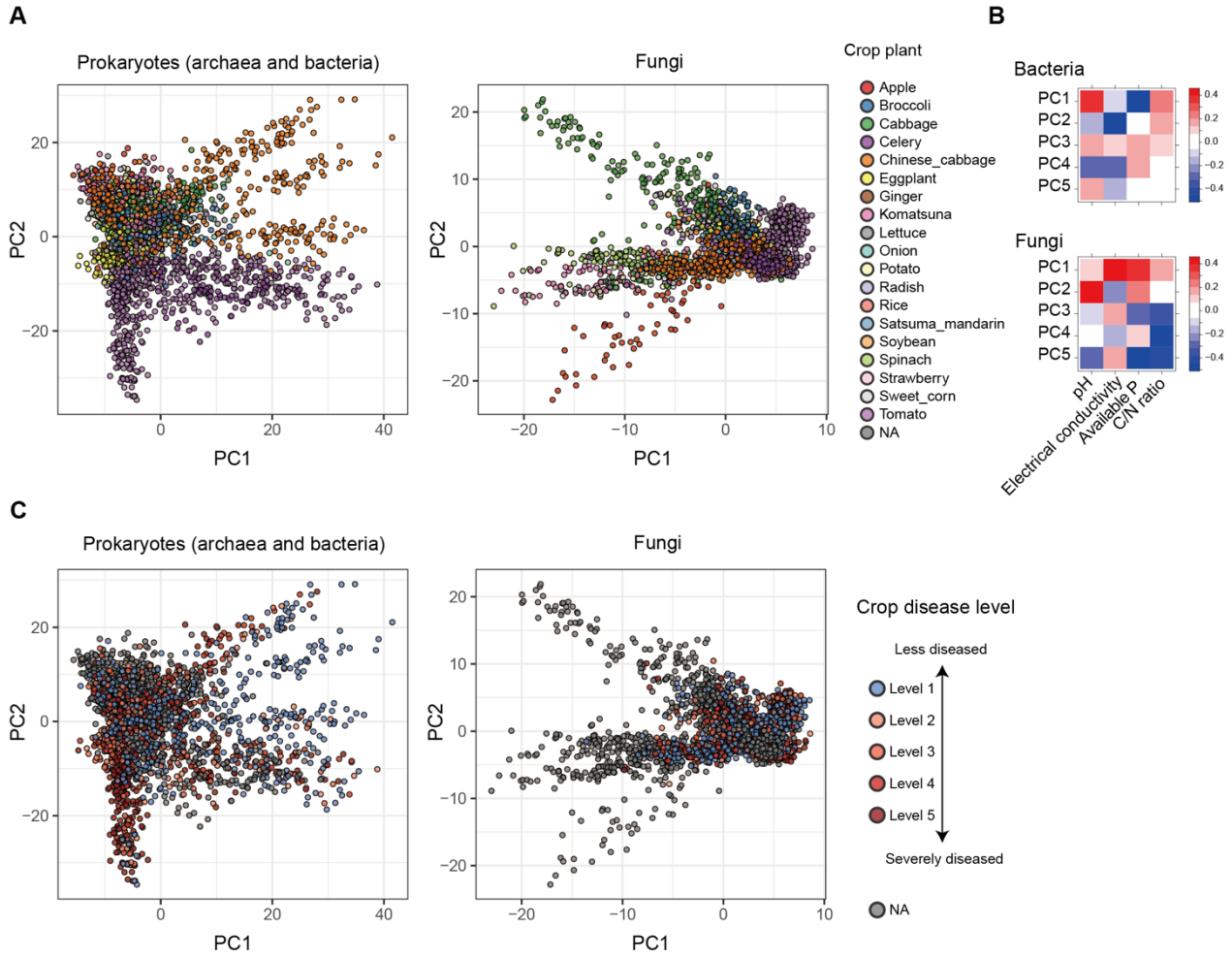


**Fig. 2 |** Dimensions of soil microbiome structure. (A) Prokaryote and fungal community structure. Principal co-ordinate analyses (PCA) were performed based on OTU-level compositional matrices respectively for the prokaryotic and fungal communities. The identify of crop plants is shown by colors. See Figures S2-3 for relationship between community structure and environmental factors. (B) Correlation between PCA scores and soil environmental factors. (C) Crop disease level and microbial community structure. On the PCA surface, crop disease level (see Materials and Methods) is indicated.

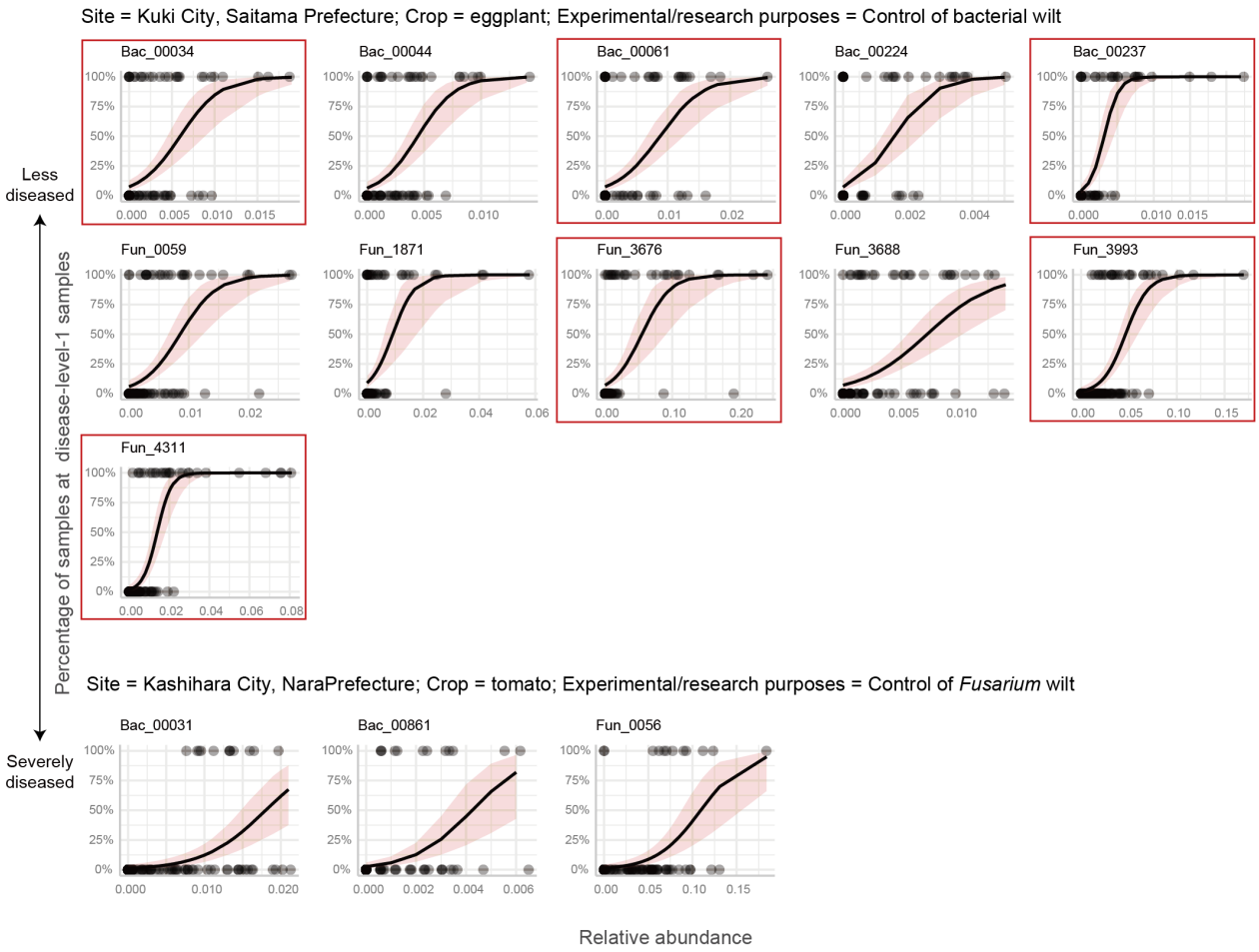


**Fig. 3** | Relationship between OTU abundance and crop plant health. Among the six sub-datasets representing unique crop plant × site combinations, OTUs showing strongest specificity to the minimal crop disease level (*z*-standardized specificity to disease level 1 ≥ 6.0) were observed in two sub-datasets ("eggplant in Kuki City" and "tomato in Kashihara City"; Table 4; see Data S2 for full results). For each OTU in each sub-dataset, generalized linear model with a logit function and binomial errors was constructed to examine relationship between OTU relative abundance and crop disease level (level 1 vs. levels 2-5). All the regression lines are statistically significant (FDR < 0.0001). The OTUs exhibiting statistically significant specificity to disease level 1 in the analysis with the entire dataset (FDR < 0.025; two-tailed test; Fig. S4-5) are highlighted with red squares.

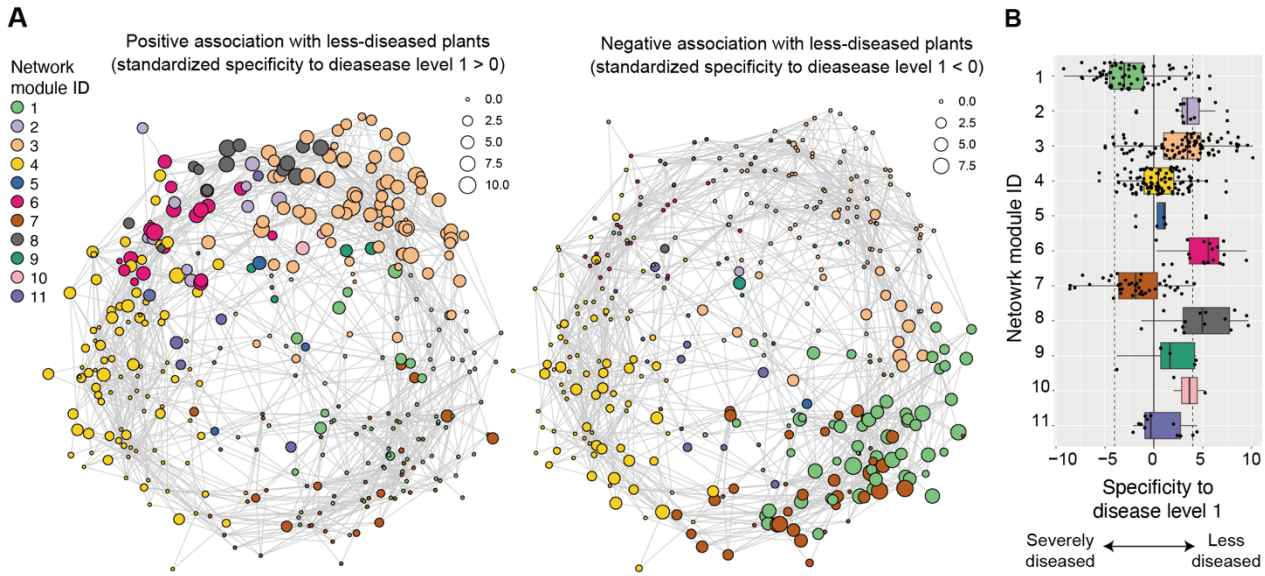


**Fig. 4 |** Architecture of microbe-to-microbe network. (A) Co-occurrence networks of archaea, bacteria, and fungi. Specificity of occurrences to disease-level-1 (the lowest disease level) samples (Figs. S4-5) is shown for each OTU within the network. The specificity is shown as node size separately for positive (left) and negative (right) associations with least-diseased states of crop plants. Colors indicate network modules, in which microbial OTUs in commensalistic/mutualistic interactions and/or those sharing environmental preferences are densely linked with each other. See Figure S8 for taxonomy (archaea, bacteria, or fungi) of respective nodes. (B) Characteristics of network modules. Mean specificity to the minimal crop disease level (disease level 1; left in the panel A) is shown for each network module.

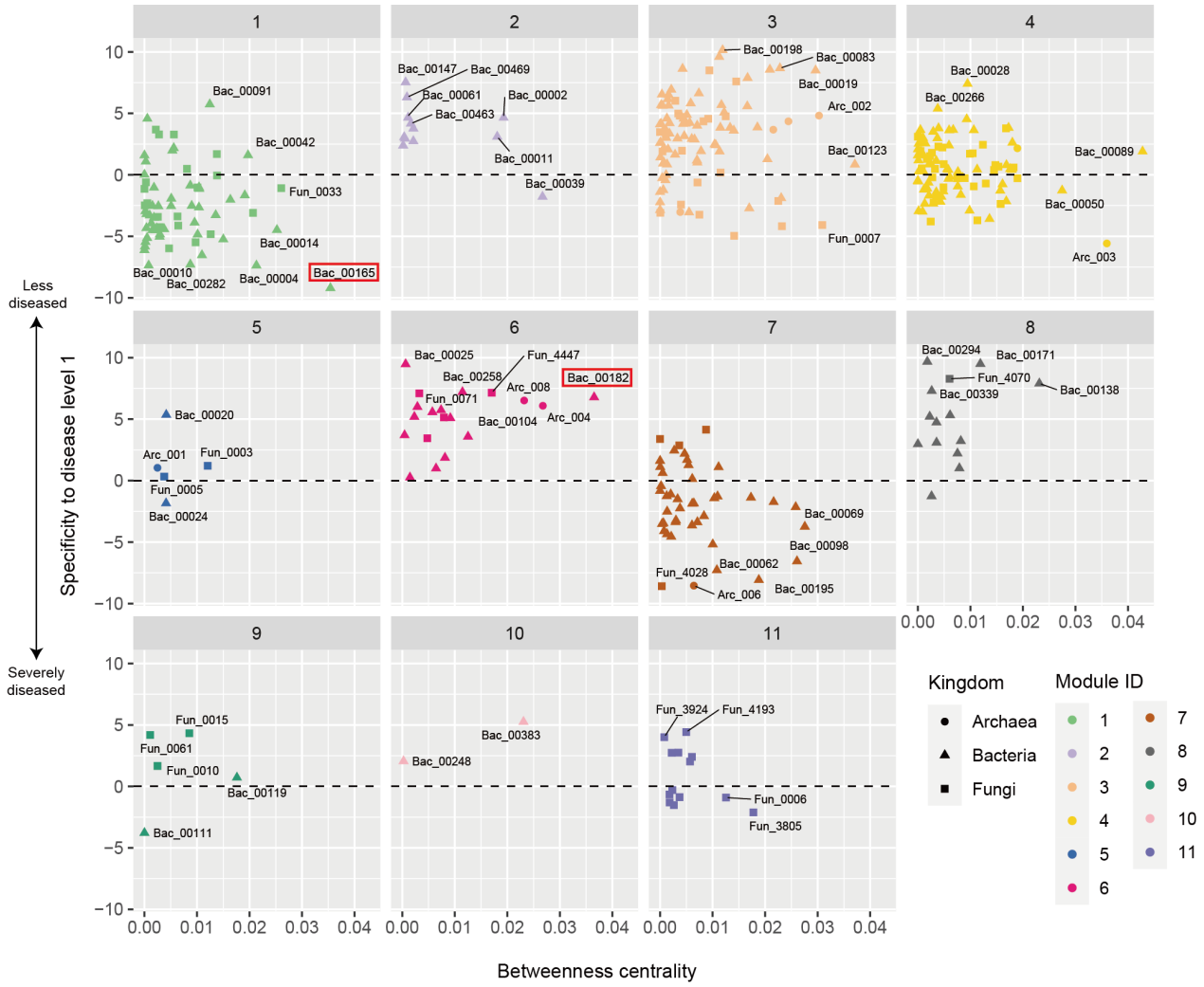


**Fig. 5 |** Properties of the microbe-to-microbe network modules. For each network module, specificity to the minimal crop disease level (disease level 1) is shown for each prokaryote/fungal OTU along the vertical axis. Betweenness centrality, which measures the extent to which an OTU is located within the shortest paths connecting pairs of other nodes in a network, is shown along the horizontal axis. The OTUs mentioned in the main text are highlighted with red squares.

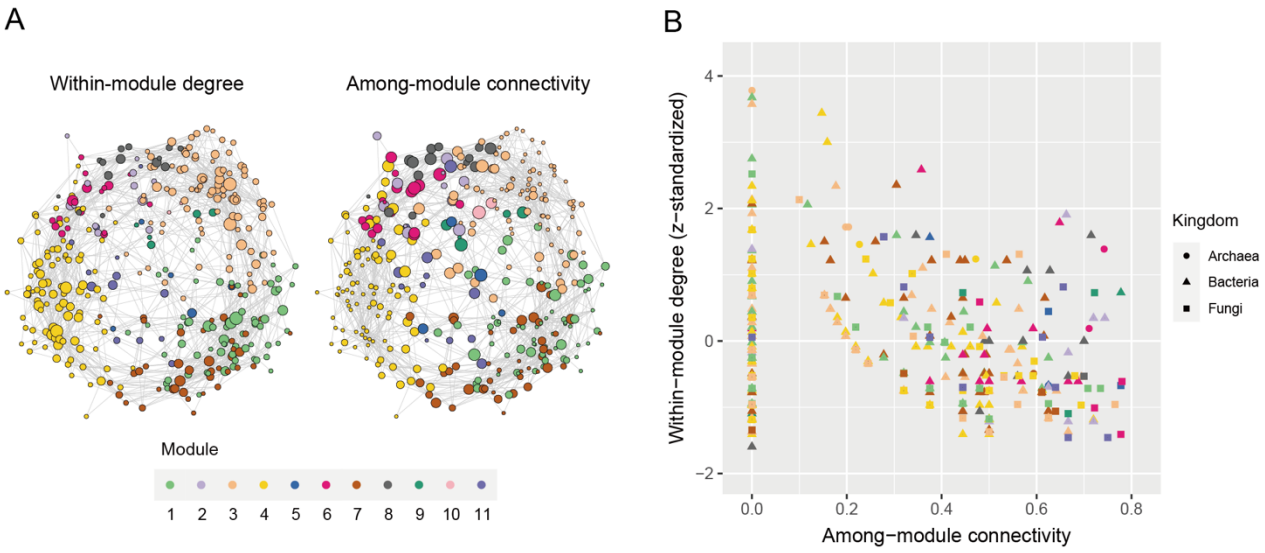


**Fig. 6** | Topological roles of OTUs within and across network modules. (A) Position of potential hubs within the network. In each graph, node size roughly represents within-module degree (left) or among-module connectivity (right). (B) Network hub indices. For each OTU, within-module degree represents the number of the OTUs linked with the target OTU within a module (*z*-standardized). Among-module connectivity represents the extent to which an OTU interlinks OTUs belonging to different network modules. The prokaryotic/fungal OTU with the highest within-module degree or among-module connectivity in each of the modules 1, 2, 6, 7, and 8 (highlighted in the main text and Table 4) is indicated with its OTU ID. See Table 5 for the taxonomic profiles of the OTUs.