⬤ISME

Predicting disease occurrence with high accuracy based on soil macroecological patterns of Fusarium wilt

Jun Yuan[1] · Tao Wen[1] · He Zhang[1] · Mengli Zhao[1] · C. Ryan Penton ⬤[2] · Linda S. Thomashow[3] · Qirong Shen ⬤[1]

Abstract

Soil-borne plant diseases are increasingly causing devastating losses in agricultural production. The development of a more refined model for disease prediction can aid in reducing crop losses through the use of preventative control measures or soil fallowing for a planting season. The emergence of high-throughput DNA sequencing technology has provided unprecedented insight into the microbial composition of diseased versus healthy soils. However, a single independent case study rarely yields a general conclusion predictive of the disease in a particular soil. Here, we attempt to account for the differences among various studies and plant varieties using a machine-learning approach based on 24 independent bacterial data sets comprising 758 samples and 22 independent fungal data sets comprising 279 samples of healthy or *Fusarium* wilt-diseased soils from eight different countries. We found that soil bacterial and fungal communities were both clearly separated between diseased and healthy soil samples that originated from six crops across nine countries or regions. *Alpha* diversity was consistently greater in the fungal community of healthy soils. While diseased soil microbiomes harbored higher abundances of *Xanthomonadaceae*, *Bacillaceae*, *Gibberella*, and *Fusarium oxysporum*, the healthy soil microbiome contained more *Streptomyces Mirabilis*, *Bradyrhizobiaceae*, *Comamonadaceae*, *Mortierella*, and nonpathogenic fungi of *Fusarium*. Furthermore, a random forest method identified 45 bacterial OTUs and 40 fungal OTUs that categorized the health status of the soil with an accuracy >80%. We conclude that these models can be applied to predict the potential for occurrence of *F. oxysporum* wilt by revealing key biological indicators and features common to the wilt-diseased soil microbiome.

These authors contributed equally: Jun Yuan, Tao Wen

✉ Qirong Shen
shenqirong@njau.edu.cn

1   The Key Laboratory of Plant Immunity, Jiangsu Provincial Key Lab for Organic Solid Waste Utilization, National Engineering Research Center for Organic-based Fertilizers, Jiangsu Collaborative Innovation Center for Solid Organic Waste Resource Utilization, Nanjing Agricultural University, Nanjing 210095, China

2   Center for Fundamental and Applied Microbiomics, Biodesign Institute, Arizona State University, Tempe, AZ, USA

3   US Department of Agriculture, Agricultural Research Service, Wheat Health, Genetics and Quality Research Unit, Pullman, WA, USA

Introduction

Microbes are of paramount importance for the sustainability of human health and the environment [1, 2]. Human gut microbes directly influence human health [3], oceanic microbes produce novel antibiotics [1, 4], while soil microbes influence food quality, plant productivity [5, 6], and the global climate through the cycling of carbon and nitrogen [7–9]. The composition and functional capacity of the soil microbiome is known to directly influence agricultural productivity by shaping critical ecosystem functions such as nutrient cycling and resistance to plant pathogens [8, 10, 11]. In addition, agricultural soil amendments, such as compost or compost-manure, have been shown to alter the native soil microbiome with potential impacts on plant disease or pathogen abundance [12]. Therefore, understanding the relationships between the soil microbiome and soil-borne plant disease is a current grand challenge in microbial ecology with applied consequences for agricultural food production worldwide [13, 14].

Caused by the fungal pathogen *Fusarium oxysporum*, Fusarium wilt is one of the most destructive soil-borne plant diseases [15], with broad host range in crop production systems [16]. *Fusarium oxysporum* can be isolated and identified from numerous native soils [17, 18]. Plant infection initiates in root tips followed by longitudinal growth into immature xylems, causing symptoms including damping-off, root rot, and vascular wilt [19]. Over the past decade, Fusarium wilt disease-suppressive soils, together with their microbiomes, have been characterized by using high-throughput sequencing (HTS) methods [20–23] and hundreds of beneficial microbes with suppressive activity against *F. oxysporum* have been isolated and tested [24]. In one study, a suppressive soil against strawberry wilt was investigated and beneficial microbial groups, as well as their active secondary metabolites, were identified [16]. Another study on Fusarium wilt of vanilla found that healthy soil harbored a more diverse fungal community with a greater abundance of *Mortierella* [22], as compared to a high disease incidence soil. Soils suppressive to Fusarium wilt of banana have been shown to harbor greater soil bacterial diversity [25], with an enrichment of *Pseudomonas* and *Tumebacillus* genera [26]. However, contrasting conditions have also been identified, with higher diversity and richness in both the fungal and bacterial communities within the diseased soil [27, 28]. These contradictions may arise due to site-specific characteristics, such as soil type, host plant, crop rotation, fertilizer management, climate, and other confounding factors [29–31], that result in a lack of a consensus concerning global microbiome properties within *Fusarium* infested soils.

Case studies have contributed an abundance of open-access data which can be utilized to evaluate the microbial world. Effectively merging and analyzing those data may provide a means by which to identify global properties of Fusarium wilt-diseased soils and their healthy counterparts such as key microbial taxa or whether crop disease can be predicted based on microbiome composition. However, the integration of microbial community data garnered from HTS in different studies is challenging [27]. "Microbial community records and information can be incomplete, processing and naming varies greatly between studies, over time, data storage is inconsistent over time, and there are few curated databases with high-quality data" [32]. As such, it may be difficult to identify underlying signals that correspond to soil disease status across studies, especially when site-specific characteristics influence the composition of the soil microbiome.

Higher Fusarium wilt disease incidence is correlated to a large increase in the population abundance of *F. oxysporum* in bulk soils as studied with under the cultivation of banana, cucumber, and other crops [33, 34]. While there is no threshold of bulk soil pathogen abundances that can be used to predict disease occurrence, it has been demonstrated that disease incidence remained low in disease-suppressive soils even after the addition of the pathogen in high abundance [35, 36]. Analyses of disease-suppressive soils have also indicated that members of the soil microbiome appear to play important roles in the suppression of *F. oxysporum* [27].

We hypothesize that there are common microbial characteristics shared among Fusarium wilt diseased and soils with on occurrence of disease. In this study we characterized the bacterial and fungal communities of bulk soils with a history of disease due to *F. oxysporum* infestation and compared them with nondiseased soils by collecting open-access microbial community data from independent studies, taking methodological and technical biases into account, and using a machine-learning approach. In this way, we attempt to reveal the underlying microbial community patterns of Fusarium wilt-diseased soils and thus, to constructs a model to predict Fusarium wilt disease under field conditions.

Materials and methods

Data collection and description

Metadata related to Fusarium wilt disease occurrence in healthy and diseased bulk soils from 9 countries or regions were collected from 37 studies with 1105 bacterial samples and 26 studies with 444 fungal samples by searching the keywords "Fusarium wilt microbiome," "Fusarium wilt community," and "Fusarium wilt structure" in Google Scholar and the National Center for Biotechnology Information (NCBI) SRA database. Here, a healthy soil is defined as on that exhibited <15% Fusarium wilt plant disease incidence while a diseased soil refers to a disease incidence >15%. A total of 24 independent bacterial data sets comprising 758 samples and 20 independent fungal data sets comprising 279 samples were randomly selected, and then separately merged to build a microbial classifier. Another 13 independent bacterial data sets (347 samples) and 6 independent fungal data sets (165 samples) were used for prediction. Sequencing data were produced either with Roche 454 technology (19.4%) or one of the Illumina platforms (80.6%). For the bacterial community studies, 11 different primer pairs (515F:806R; 515F:907R; 338F:806R; 520F:802R; 799F:1193R; 27F:533R; 563F:802R; 341F:785R; 341F:805R; 27F:533R; and 27F:518R) were identified from the metadata, accounting for 66% of samples, with the majority (72.7%) using the V4 or V3–V4 regions to produce amplicons. For fungal community studies, six different primer pairs were identified (ITS5F: ITS4R; ITS1F:ITS2R; ITS1F:ITS1R; ITS7F:ITS4R; ITS2F: ITS5R; and ITS3F:ITS4R), accounting for 66% of the samples with clear primer information (34% of the samples

with unclear primer information). All raw data were obtained from two open-access nucleic acid databases: the DNA Data Bank of Japan (5%) or the NCBI (95%). Details of metadata are provided in Supplementary Files 1 and 2, including sequence information such as crop type, location, reference, database, sequencing platform, primers, etc.

Data processing

Processing data from databases

Sequence read quality was checked by FastQC v.0.11.5 [37] and paired-end reads were joined by using the join_-paired_ends.py script in QIIME 1.9.0 [38], and then trimmed to a Phred score of at least 20 by using the split_libraries_fastq.py script in QIIME. For each data set, all reads were mapped to the full-length 16S rRNA or ITS gene sequences using the usearch global algorithm implemented in VSEARCH 2.1.3 [39] The closed-reference workflow was a database-dependent approach employing a predefined set of reference sequences with known taxonomy (the manually curated Greengenes database 13.8 or the Unite database 7.2) to cluster sequences into OTUs and assign taxonomy to reference sequences. This approach is advantageous for comparing studies that target different 16S rRNA or ITS gene regions since the underlying database is comprised of full-length sequences.

The aligned results in usearch table format were directly converted to Biological Observation Matrix (BIOM) format using BIOM 2.1.5 [40]. The metadata were added to BIOM-Format 1.3.1 using biom convert for downstream analysis with QIIME. Summary information on the representation of taxonomic groups was obtained by using summarize_taxa.py script. Finally, all BIOM files of each data set were merged using QIIME. The workflow of this study is shown in Supplementary Fig. 1.

To address PCR biases (Supplementary Files 1 and 2) and biases associated with rare taxa and some groups which could be over amplified, we created a filtered subset of the data incorporating two strategies: (1) deletion of OTUs that only appeared in one-third of all samples; and (2) deletion of OTUs that only appeared in less than half of the independent data sets.

Machine-learning approach for model building

Comparison of machine-learning methods

In order to better distinguish the microbial communities of Fusarium wilt-diseased and healthy soils, we applied three well-established machine-learning algorithms; random forests (RF) [41], support vector machines (SVM) [42], and logistic regression (LR) [43], to construct models using the relative abundance of OTUs as a candidate feature. A performance evaluation of the classifier was established through fivefold cross-validation within the training set. In detail, the training set was randomly divided into five equal portions; then, during each of the five iterations, the least absolute shrinkage and selector operator [44] were chosen as the feature selection method on 80% of the training data to train the classifiers (1000 trees for RF, radial kernel for SVM, other parameters set by default) with the selected features. In addition, the trained classifiers were used for prediction or validation with the remaining 20% of the training data. After that, the predictions of all five iterations were combined and compared to the actual health condition according to the receiver operating characteristic curve (ROC) and the area under curve (AUC) score [45]. Finally, we built up the classifiers from the entire training set using the best-performing algorithm (with the highest AUC) identified through cross-validation and applied the best classifiers to the test set in order to independently validate the predictive power. Cross-validation was performed by the rfcv() function for the selection of appropriate features. The varImpPlot function was used to illustrate the importance of features in the classification, and the curves were visualized by using the *ggplot2* package in R v.3.6.1.

Modeling of bacterial metadata

We created a classification model using RF, a robust machine-learning algorithm for classification and regression that is suitable for microbial population data [46, 47]. Then, to acquire the best discriminated performance with taxa across wilt-diseased soils and soils free of Fusarium wilt, we classified the relative abundances of bacterial taxa to the phylum, class, order, family, genus, and OTU levels using the RF package v.4.6–14 in R with default parameters [41]. For each taxon, two-thirds of all samples were randomly selected for model building and the remaining one-third were used for prediction. We carried out this process five times to evaluate the accuracy of each taxon. Modeling of bacterial taxa at the OTU level most clearly discriminated the samples of diseased and healthy soils (Supplementary Table 1), and therefore, further analysis was performed at that level. The training set of metadata from 24 studies (758 samples) using the random forest (importance = TRUE, proximity = TRUE) function was utilized to generate the classification model for diseased and healthy soils. A high degree of accuracy of 98.9% was achieved.

Modeling of fungal metadata

In order to acquire the highest discriminated separation of fungal taxa among soils that did or did not give rise to Fusarium wilted plants, we classified the relative abundances

of taxa to the phylum, class, order, family, genus, and OTU levels by using the R package v.4.6–14 in R with default parameters. Model building and prediction were carried five times in the same fashion as performed with the bacterial metadata. As with bacterial taxa, modeling to the OTU level most clearly distinguished between diseased and healthy soils (Supplementary Table 1), and further analysis therefore was performed at that level. The training set of metadata from 20 studies (279 samples) using the random forest (importance = TRUE, proximity = TRUE) function was carried out to generate the classification model for diseased and healthy soils, yielding an accuracy of 96.4%.

Characterization of healthy and diseased soil microbial communities

To exclude the effects of variable sample sizes from each case study, we randomly selected 80 samples if the total number of samples was >80. For *alpha* diversity analysis, the minimum number of sequences was extracted randomly from each sample to calculate Chao1, Shannon, and Pielou evenness indices in QIIME using the alpha_diversity.py script. The relative abundance was used to standardize the OTU profiles by the normalize_table.py script before beta diversity calculation and Bray–Curtis similarity matrices were prepared with the beta_diversity.py script. The principal coordinate analysis (PCoA) plots were generated from Bray–Curtis similarity matrices by using R package ggplot2.

Significant correlations between the relative abundance of bacterial or fungal OTUs were made using the sparse correlations for compositional data algorithm implemented in the *SparCC* python module [48] and co-occurrence networks were plotted using the R package igraph [49]. Only the correlations with "R-corr" absolute value > 0.4 and p value < 0.05 were plotted. In order to describe the topology of the resulting network, a set of measures (that is, average node connectivity, average path length, diameter, cumulative degree distribution, clustering coefficient, and modularity) were calculated [50]. All statistical analyses were carried out in the R environment (http://www.r-project.org) using the vegan [50] and igraph [49] packages. To assess nonrandom patterns in the resulting network, we compared our network against its randomized version using the igraph package. Structural attributes of this network, such as the clustering coefficient and characteristic path length, were compared with those in the random network with equal nodes and edges.

Trained model validation

Soil sample collection

To further model validation, we collected soils (Supplementary File 3) of banana, cucumber, watermelon, and lily

from Hainan, Guangdong, Beijing, and Jiangsu (China), respectively, in late April of 2019. For diseased soil collection, we selected plots under continuous cropping whose plants suffered from wilt disease for at least five seasons. Newly planted plots lacking wilt disease were sampled for healthy soils. Healthy plots were selected near to the diseased plots to avoid differences induced by geographical and edaphic factors. For each soil sampling, twenty random soil cores (5 cm diameter × 20 cm length) were obtained from the 0- to 20-cm layer in an S-pattern in each plot and combined as one soil sample. The freshly collected soils were sieved through a 2-mm mesh to remove plant debris and stored at −80 °C prior to DNA extraction.

DNA extraction, gene amplification, and sequencing

Total DNA from 0.5 g of soil was extracted with the Power Lyzer PowerSoil DNA Isolation Kit (Qiagen, Germany) following the manufacturer's protocol. DNA quality and quantity were evaluated with a 1% agarose gel and a NanoDrop 2000 spectrophotometer (Thermo Scientific, Waltham, MA, USA). For taxonomic profiling, PCR products that targeted the V4 region of the bacterial 16S rRNA gene were amplified with the primers 515F: GTGYCAGC MGCCGCGGTAA and 806R: GGACTACNVGGGTWT CTAAT) [51] to yield an amplicon of 292 bp. For the fungal communities, PCR targeting the ITS2 region was carried out with the primers ITS3-F: GCATCGATGAAG AACGCAGC and ITS4-R: TCCTCCGCTTATTGATAT GC) [52]. The 50 μL reaction mixtures contained 25 μL 2× Premix Taq (Takara Biotechnology, Dalian Co. Ltd, China), 1 μL each primer (10 μM), 3 μL DNA (20 ng/μL), and 25 μL of sterilized ultrapure water. PCR amplification was performed by using a BioRad S1000 (BioRad Laboratory, CA) with the following cycles: 95 °C for 5 min, then 30 cycles of 94 °C for 30 s, 52 °C for 30 s, 72 °C for 30 s with a final extension at 72 °C for 10 min. Products were run on a 1% agarose gel and those with clear bands between 290 and 310 bp were combined for sequencing. PCR products were mixed in equidensity ratios according to the GeneTools Analysis Software (Version 4.03.05.0, SynGene) and the mixture was purified with an EZNA Gel Extraction Kit (Omega, USA). Sequencing libraries were generated using the NEBNext® Ultra™ DNA Library Prep Kit for Illumina® (New England Biolabs, USA) following the manufacturer's recommendations. Index codes were added and library quality was assessed with a Qubit® 2.0 Fluorometer (Thermo Scientific) and an Agilent Bioanalyzer 2100 system. Finally, the library was sequenced on an Illumina Hiseq 2500 platform and 250 bp paired-end reads were generated, quality filtered to obtain high-quality clean reads according to Trimmomatic (V0.33, http://www.usa dellab.org/cms/?page=trimmomatic), and sequences were

assigned to each sample based on its unique barcode. All of the clean reads were processed using the pipelines detailed above.

Statistical methods

For subsequent analyses, minimum numbers of sequences were extracted at random from each sample to calculate an alpha diversity index estimated in Qiime by the alpha_diversity.py script. A nonparametric t-test was used to determine if Shannon indices differed between Fusarium wilt-diseased and healthy soils. Before calculation of beta diversity, relative abundances were used to standardize the OTU profiles and Bray–Curtis similarity matrices were prepared using the beta_diversity.py script. Permutational multivariate analysis of variance (PERMANOVA) (Adonis, transformed data by Bray–Curtis, permutation = 999) was used to determine if beta diversity differed between the two treatments and PCoA plots were generated from Bray–Curtis similarity matrices created using R package ggplot2.

Results and discussion

Diversity and taxonomic difference between diseased and healthy soil

We randomly selected sequencing data from 24 individual bacterial studies, 19 of which were published and 5 that were available by open-access. (Fig. 1a and Supplementary File 1). The resulting merged bacterial OTU table contained data from 758 individual soil samples from six countries with over 40,000 taxa (Supplementary Fig. 2).

In order to prevent a large number of samples within a single study from dominating the analysis of microbial community diversity, we randomly filtered the samples when such case studies contained more than 80 samples. Then, we rarefied (without replacement) the sequencing reads of each sample to 3000 to calculate the Chao1, Shannon, and Pielou evenness indices. The results of these analyses revealed no clear differences in bacterial alpha diversity between healthy and diseased soils (Fig. 2a), but principal coordinates analysis (PCoA) with Bray–Curtis distance showed that the bacterial communities differed significantly between the two soil types ($p = 0.001$, PERMANOVA by Adonis) (Fig. 2b). The diseased soils harbored a higher relative abundance of *Firmicutes*, *Bacteroidetes*, *Choroflexi*, and *Gemmatimonadetes*, while the healthy soils harbored higher relative abundances of *Proteobacteria*, *Actinobacteria*, and *Acidobacteria* (Fig. 2c). We also found that there were more unique bacterial OTUs in the diseased soils (806) than in the healthy soils (120) (Fig. 2d).

We also incorporated the sequencing data from 20 individual fungal studies related to Fusarium wilt into our analyses (Fig. 1b and Supplementary file 1). All of these soil samples had been amplicon-sequenced across the ITS region (Supplementary file 1). Among the 20 studies, sequencing data from 16 studies were published and 4 of them were available by open access. The merged fungal OTU table contained 279 individual soil samples and over 8000 taxa. We then rarefied (without replacement) the sequencing reads of each sample to 3000 for alpha diversity calculations of the Chao1, Shannon, and Pielou evenness indices, which together indicated significantly greater diversity in the healthy than the diseased soils (Fig. 2e). Similarly, PCoA analysis with Bray–Curtis distance also showed a clear ($p = 0.001$, PERMANOVA by Adonis) difference between the fungal communities of healthy and diseased soils (Fig. 2f). The relative abundance of *Ascomycota* was higher in diseased soils, while that of *Mortierellomycota* was greater in healthy soils (Fig. 2g and Supplementary File 4; Wilcoxon test; adjusted p values < 0.05). We also found that there were more unique OTUs in the fungal communities of the diseased soil (37) as compared to the healthy soils (31) (Fig. 2h).

To highlight common features among the microbial communities, we also filtered the bacterial and fungal OTUs as above by deleting those that appeared in less than one-third of all of the samples or in less than half of the independent data sets. After filtering, the tables contained 1339 bacterial OTUs and 177 fungal OTUs for downstream analysis, accounting for 37.7% of the total bacterial reads and 49.9% of the total fungal reads (Supplementary Fig. 2). The resulting alpha diversity of the bacterial community was significantly higher in diseased than in healthy soil samples (Supplementary Fig. 3), while no significant difference was found between diseased and healthy soil fungal communities (Supplementary Fig. 4).

Both bacterial and fungal communities can be biomarkers to distinguish diseased and healthy soils

To determine whether properties of the soil bacterial or fungal communities could be used as biomarkers to distinguish between diseased and healthy soils, we built three machine-learning models by random forest, SVM, and LR. Consideration of the AUC and ROC curves derived from the models revealed that the RF model exhibited the best accuracy rate with which to predict the classification of samples (Fig. 3c, d). Next, we established with the bacterial and fungal data that the models trained at the OTU level showed the highest levels of classification accuracy, 95% and 92%, respectively, for both models.

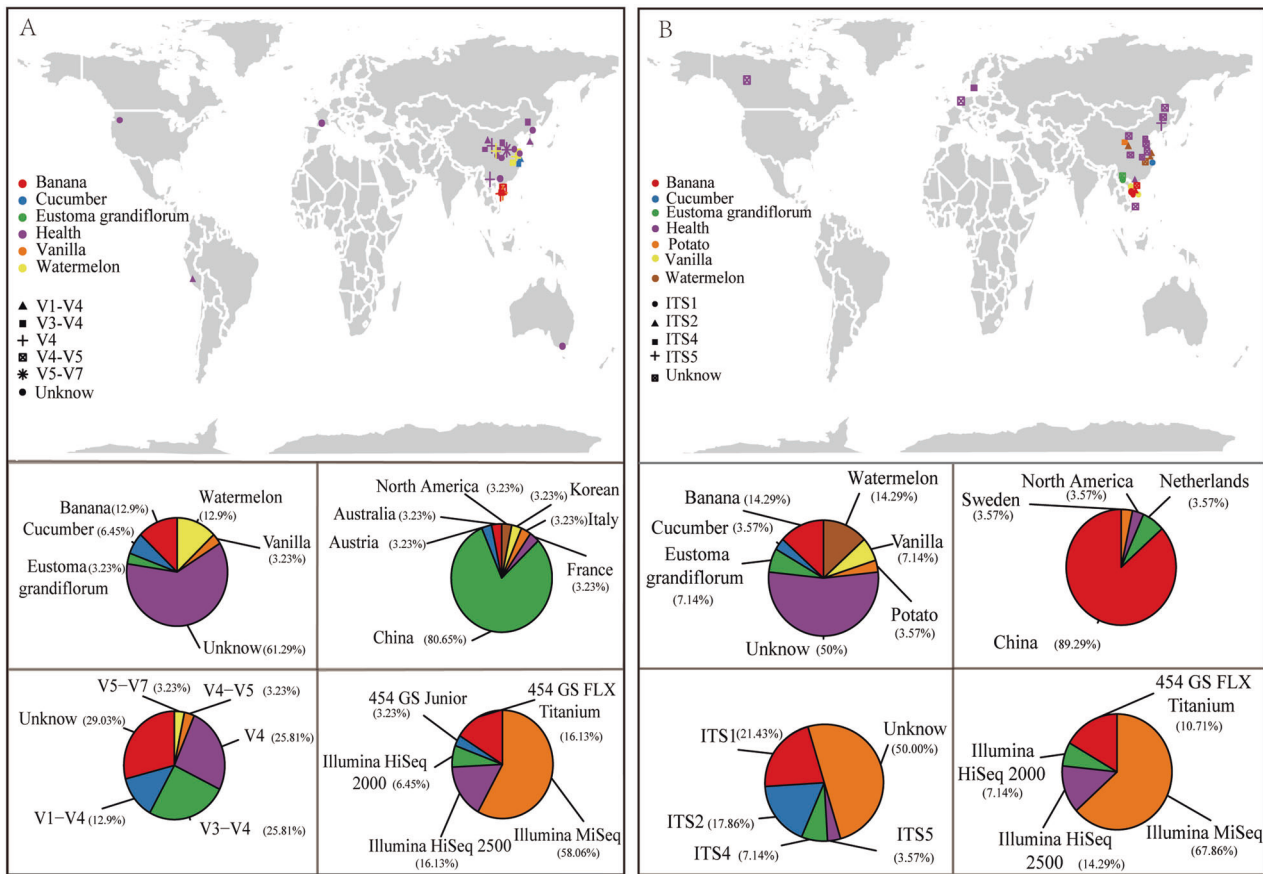For bacterial model validation, we analyzed 13 independent data sets from the NCBI SRA database including

Fig. 1 Merging of data from independent studies. The geographic region, information about country and zones, primers, amplicon region, and sequencing platforms used in this study are displayed. **a** and **b** show the broad geographical distribution of the data sourced from 24 individual bacterial studies and 22 individual fungal studies, respectively. The color of points refers to the identity of the sampled plants. The shape of point indicates locations of amplicons sequenced. The four subplots below **a** and **b** show the proportion of samples among crops, the proportion of samples among countries and zones, the proportion of samples sequenced with different primers, and the proportion of samples sequenced with different platforms.

26 diseased soil samples and 321 healthy soil samples. The bacterial model showed 93.4% average accuracy for all samples collected, with 89.6% accuracy for diseased samples and 95.4% for healthy samples (Supplementary Table 2). At the same time, we used 6 independent fungal data sets from the NCBI SRA database, including 21 diseased soil samples and 144 healthy soil samples for fungal model validation. The fungal model exhibited an 87.8% average accuracy for all samples collected, with 95 % accuracy for diseased samples and 78.9% for healthy samples (Supplementary Table 2).

As an additional test of the universality of the model, field soil samples of 4 different crops, including 20 soil samples, were collected for further verification. Our bacterial model showed 80% average accuracy for all samples, with 90% accuracy for diseased samples and 70% for healthy samples (Fig. 3i). The fungal model showed 87.2% average accuracy for all samples, with 90% accuracy for diseased soil samples and 84.2% for healthy soil samples (Fig. 3i). Collectively, our results indicate that both the bacterial and the fungal models in this study can serve to reasonably predict whether Fusarium wilt disease will occur (>15% incidence) or not in a particular soil.

Microbial features recognition of diseased and healthy soil

We carried out tenfold cross-validation with five repeats to evaluate the importance of bacterial OTUs as potential indicators. The cross-validation error curve stabilized (Fig. 3g) at the 45 most relevant OTUs, which were defined as biomarker taxa with relative abundance ranging from 0.16 to 4.84% (Fig. 4c). The most important OTU belonged to the *Gemmatimonadetes* (Gemm-1, Fig. 4c) and was enriched in diseased soils. Among all of the important OTUs, 33 showed higher relative abundance in diseased soils while 12 showed higher relative abundance in healthy soils (FDR adjusted $p < 0.05$, Wilcoxon rank sum test; Fig. 4b). The OTUs more abundant in diseased

Fig. 2 General descriptions of the Fusarium wilt-diseased and healthy soil microbial (bacterial and fungal) communities. a and **e** illustrate the alpha diversity of soil bacterial and fungal communities, respectively. The Chao1, Shannon, and Pielou indices were calculated with all OTUs (48521 bacterial and 8840 fungal) merged from 244 diseased (122 bacterial and 122 fungal) and 793 healthy (636 bacterial and 157 fungal) samples. The horizontal bars within boxes represent the median. The tops and bottoms of boxes represent 75th and 25th quartiles, respectively. All outliers were plotted as individual points. Asterisk "*" represents a significant difference between groups (nonparametric t-test), and "ns" indicates no significant difference between groups. **b** and **f** Show the principal coordinates analysis (PCoA) with Bray–Curtis dissimilarity performed on the taxonomic profile (at the OTU level) for diseased and healthy microbial (**b**: bacterial and **f**: fungal) communities. *R*- and *p* values were evaluated via Adonis test. **c** and **g** Show the relative abundance (%) of the major phyla present in the microbial (**c**: bacterial and **g**: fungal) communities in diseased and healthy soil samples. **d** and **h**: Venn diagram of shared and unique OTUs numbers observed in diseased and healthy soil microbial (**d**: bacterial; **h** fungal) communities.
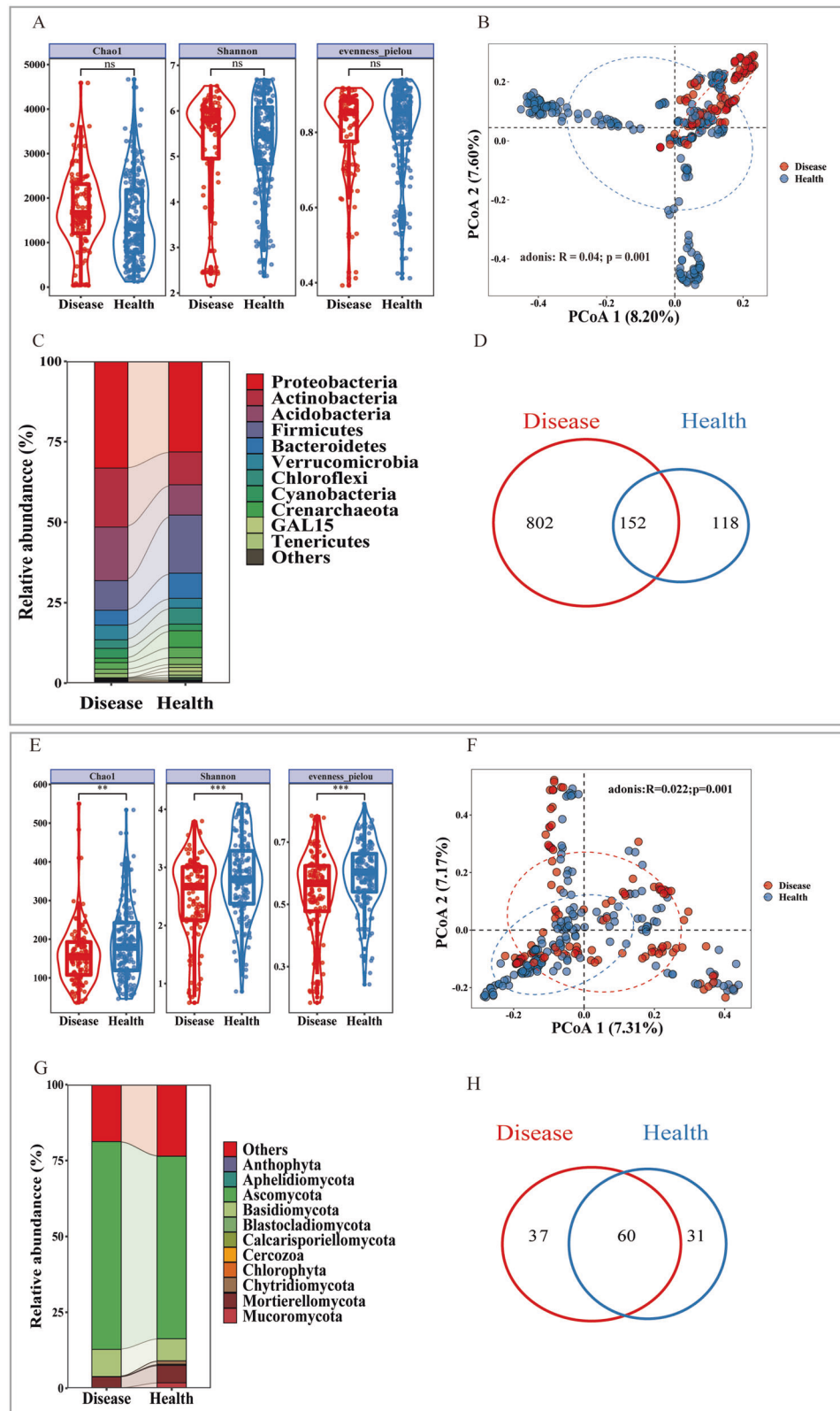
soils were classified as *Proteobacteria* (*Alpha*- and *Beta*-) and *Acidobacteria*, followed by *Sphingomonadaceae*, *Koribacteraceae*, *Kaistobacter*, and *Xanthomonadaceae* (Supplementary Fig. 5, Supplementary File 5, and Fig. 4c). The OTUs more abundant in healthy soils were classified mainly as *Actinobacteria* and *Proteobacteria* (*Alpha*- and
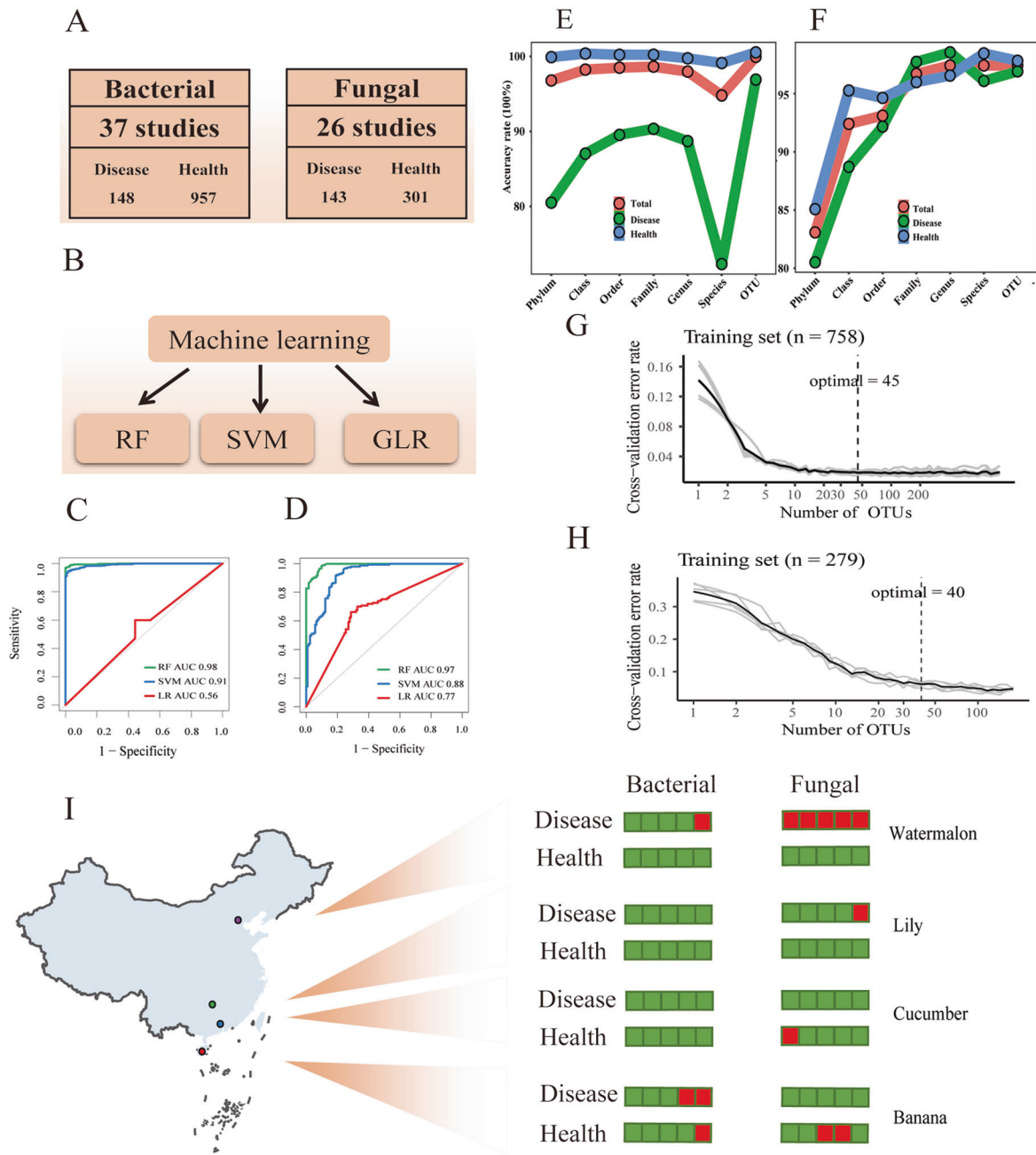
Fig. 3 The workflow of the machine-learning method for model building and testing. a The samples (bacterial 758; fungal 279) used in machine-learning process. **b** Three machine-learning methods (random forest (RF), support vector machines (SVM), and logistic regression (LR)) used to build the models. **c, d** The AUC and ROC curves of models built with the three machine-learning methods at the OTU level of microbial (c: Bacterial, d: Fungal) communities, with the RF model showing the highest accuracy rate predicting the classification of samples (AUC = 0.98). The RF method was used to build the models with both soil bacterial (**e**) and fungal (**f**) samples at phylum, class, order, family, genus, and OTU level, respectively. Green, blue, and red lines represented the accuracy of healthy samples, diseased samples, and total samples, respectively. The models built at the OTU level show the highest classification accuracy for both bacterial and fungal samples. The top 45 bacterial OTUs (**g**) and top 40 fungal OTUs (**h**) identified by applying RF classification of the relative abundance of diseased and healthy samples. **i** Prediction with samples collected from China to the best classifier. Green indicates correct, and red indicates incorrect, with five replicates for each of the diseased and healthy soil samples.

Gamma-), followed by *Streptomyces Mirabilis Bradyrhizobiaceae*, *Comamonadaceae*, and *Chthoniobacteraceae* (Supplementary Fig. 5 and Supplementary File 5).

An identical method was used to evaluate the importance of fungal OTU features. The cross-validation error curve stabilized at 40 OTU features, which were
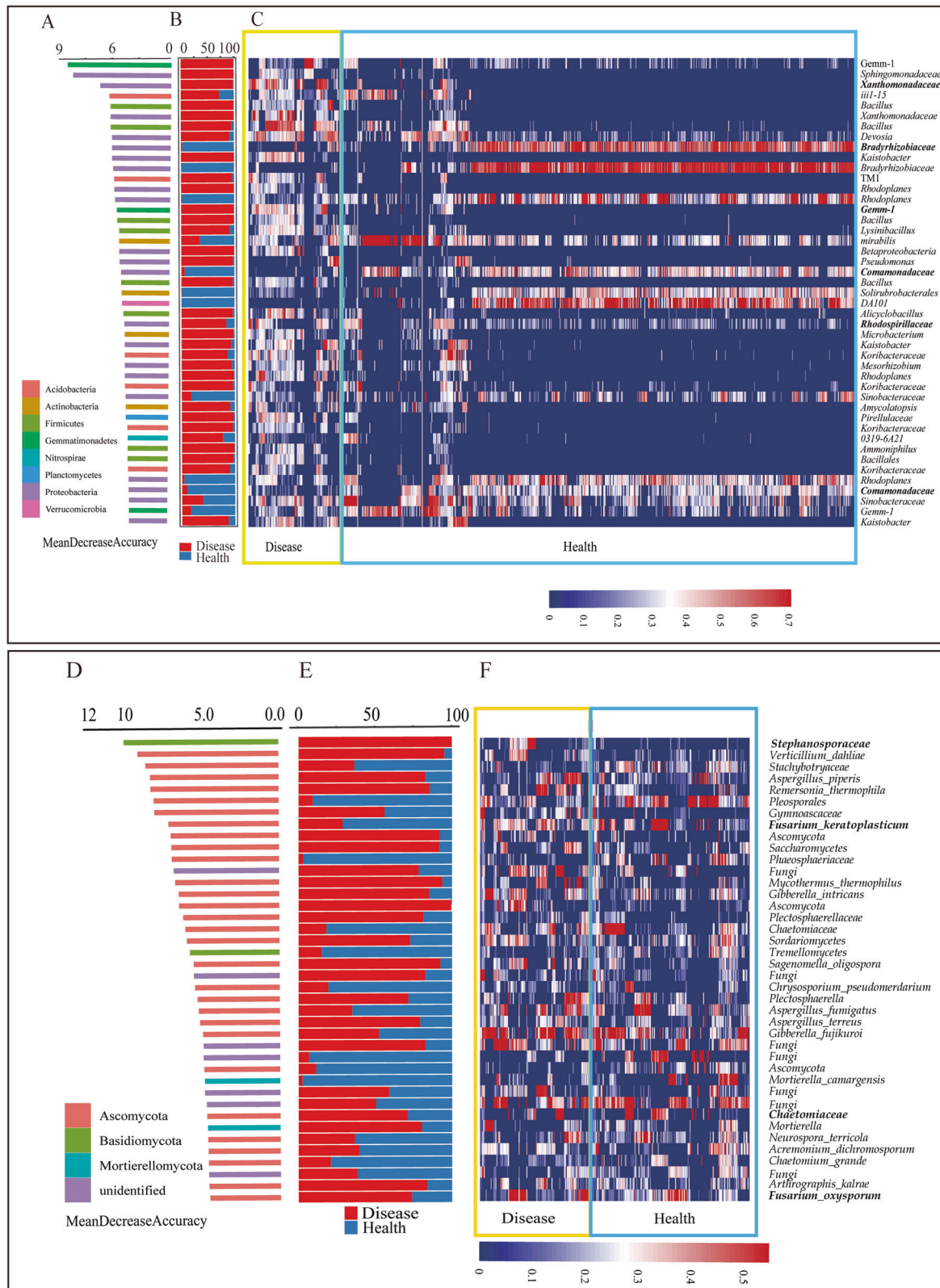
defined as biomarker taxa (Fig. 3h). Twenty-six of those showed higher relative abundance in diseased soils, and fourteen had higher relative abundance in healthy

soils. The *Fusarium oxysporum* and *Mycothermus thermophilus* were proportionally enriched in diseased soils, while *Aspergillus fumigatus*, *Mortierella*

◀ **Fig. 4 RF model to detect bacterial taxa that accurately distinguish diseased and healthy samples.** Top feature taxa (45 bacterial OTUs and 40 fungal OTUs) are ranked in descending order of importance to the accuracy of the bacterial (**a**) or fungal (**d**) model. Color represents the phylum of OTUs classified. **b, e** The relative proportion of the feature OTUs (**b** for bacterial, **e** for fungal) in healthy and diseased samples are displayed. **c, f** Heatmap analysis of the abundance of the top 45 bacterial OTUs (1339 bacterial OTUs merged from 122 bacterial and 636 bacterial samples) and 40 fungal OTUs (177 fungal OTUs merged from 122 fungal and 157 fungal samples).

camargensis, and *Neurospora terricola* were proportionally enriched in healthy soils (Supplementary Fig. 6, Supplementary File 6, and Fig. 4f). Not surprisingly, the relative abundance of *F. oxysporum* was greater in diseased soils (7.8%) than in healthy soils (4.%). Other *Fusarium*-OTUs were distributed between diseased and healthy soils. For example, the OTU *Fusarium keratoplasticum* had a higher relative abundance in healthy soils (Supplementary Fig. 7).

Co-occurrence network analysis of the features

Co-occurrence network analysis was used to assess interactions across features, and only the significant correlations ($|R| > 0.4$, $p < 0.05$) are shown in this analysis. The results revealed a greater number of nodes and links in the healthy microbial (both bacterial and fungal) feature networks (Supplementary Table 3). Further insight into the bacterial feature network illustrated higher average degree, centralization-closeness, and clustering coefficient values in healthy than in diseased soil (Supplementary Table 4). Some OTUs, such as those classified as *Bacillaceae*, *Hyphomicrobiaceae*, *Bradyrhizobiaceae*, *Comamonadaceae*, *Streptomycetaceae*, and *Chthoniobacteraceae* (red in the healthy feature network) contained more links and higher relative abundances in the healthy bacterial network. Other OTUs, such as *Alicyclobacillaceae* and *Sphingomonadaceae* (red in the diseased feature network), contained more links and higher relative abundances in the diseased bacterial networks. OTUs 1111883 and 549433, classified as *Gemmatimonadetes* that exhibited a high importance through model recognition, contained more connections and higher relative abundances in the diseased (Fig. 5a), compared to the healthy bacterial network (Supplementary Files 7, 8).

Unlike the bacterial feature networks, there were more nodes and links, as well as higher average degree, centralization-closeness, and clustering coefficient values, in the diseased than the healthy fungal feature network (Supplementary Table 4). Three OTUs belonging to the genera *Aspergillus* and *Remersonia* and the family *Stephanosporaceae* showed greater closeness, center degree, and betweenness in the diseased (Fig. 5c) than the healthy fungal feature network (Fig. 5d and Supplementary Files 9 and 10) while OTUs belonging to *Chaetomiaceae* showed higher closeness, center degree, and betweenness values in the healthy fungal feature network. Unlike in the diseased feature network, no connection with *F. oxysporum* was found in the healthy feature network (Supplementary Files 9 and 10). Notably, fewer nodes and connections were included in the healthy feature network than in the diseased one, perhaps because the OTUs used there were derived from a machine-learning process that focused mainly on diseased soil microbial features.

Sequencing data used in this study varied with regard to the primers used, the amplified gene regions and sequence lengths. We aligned the sequencing reads to full-length 16S rRNA gene or ITS gene sequences in Greengene or Unite for merging disparate sequence data as a reasonable way to integrate all sequence data before further analyses. Similar methods were previously implemented by Adams et al. [53] where 16S rRNA gene data sets from 23 studies were combined by using a closed-reference OTU method and Cornejo-Granados et al. [54] where environmental and host factors shaping the structure and function of shrimp microbiota were identified by using the same method. Recently, it has been shown that disparate amplicon sequence data could be combined at the taxonomy-based level [32], but only with bacterial community amplicon sequences. Another method, the de novo approach, was reported to include more OTUs than the reference-based method by retaining all of the sequences generated [55]. However, in studies that used the de novo approach [56, 57], it did not work well for sequencing metadata with multiple primers (Supplementary File 1 and 2) across different case studies. In this study, we succeeded in analyzing both bacterial and fungal metadata related to Fusarium wilt disease. However, the percentage of shared OTUs among all studies we included in our analysis were only about 15%, indicating that the majority of OTUs were specific to individual studies. This may be due to sampling bias, DNA extraction bias, and sequencing biases [58], but even so, we were still able to obtain common features of the bacterial and fungal communities in Fusarium wilt-diseased soils. The RF algorithm we utilized has been widely applied in microbial ecology studies [56, 59, 60], such as to model the difference between the microbiota of mice fed normal and high-salt diets [61]. We also found that the RF model showed excellent performance in distinguishing between the microbial community characteristics of diseased and healthy soils (Fig. 3c, d). The model can be built at any taxonomic level, but in our study we found that the OTU level provided the best results for both bacterial and fungal communities (Fig. 3e, f). The model also can perform well at other taxonomic levels, as when family-level random
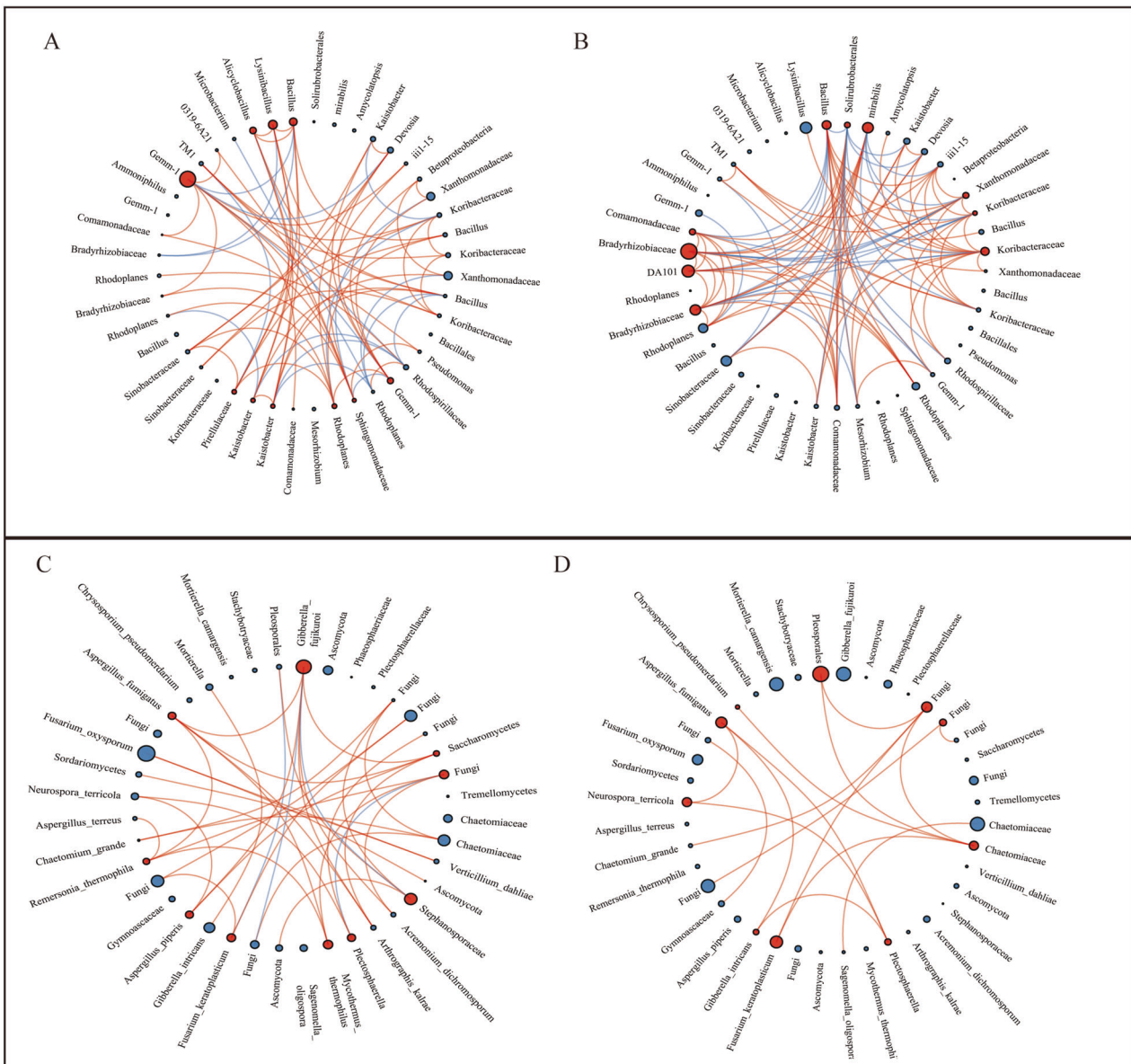
Fig. 5 Co-occurrence networks of the top microbial features. a and **b** show the diseased and healthy networks with 45 bacterial OTUs, respectively. **c** and **d** Show the diseased and healthy networks with 40 fungal OTUs, respectively. Edges represent significant Spearman correlations ($\rho > |0.4|$, $p < 0.05$). Light blue lines represent a significant negative correlation and light red lines represent a significant positive correlation. The size of the points represents the relative abundance of OTUs in each microbial community. The red nodes represent the top ten node values in each network.

forest classifiers were used to distinguish the root microbiomes of indica and japonica rice [47]. Collectively, the RF algorithm appears to be a suitable choice for extracting microbial features especially with metadata.

By merging global metadata, we found that the alpha diversity of the fungal community was significantly greater in healthy soils than in diseased soils, whereas there was no difference in the alpha diversity of bacterial microbial communities. Case studies have reported similar results with vanilla and potato field soil samples [22, 62]. It is thought that a large increase in the population of a fungal

pathogen within a soil will disrupt the inherent balance of the soil fungal community, leading to a decrease in diversity [22]. This is also reflected in the difference in the overall composition of the microbial communities of our healthy and diseased soils. The OTU classified as *F. oxysporum* was the most important fungal community feature by which to distinguish between diseased and healthy soils (Fig. 4d–f). These results revealed that while specific features of the healthy soil microbial community might be difficult to distinguish, the principal feature of the diseased soil microbial community was the high population of *F*.

oxysporum [63]. Although the relative abundance of *F. oxysporum* was high in the diseased soils, it has been shown to vary greatly among studies. Thus, in similar studies concerning the banana wilt soil microbiome, Shen et al. identified 10^6 copies/g soil, while only 10^4 copies/g soil was found by Wang et al. [64, 65]. As another example, the relative abundance of *Fusarium oxysporum* in vanilla Fusarium wilt soil was 0.1–0.45%, but 12–32% was found in diseased banana soils [64, 66]. As such, it does not appear that we can solely utilize the relative abundances of specific members such as *F. oxysporum* alone to identify whether a soil harbors the potential to cause Fusarium wilt disease (Supplementary Fig. 7). Furthermore, within the healthy soils there was greater diversity among other members of the genus *Fusarium* and increased relative abundance of other *Fusarium* species (nonpathogenic fungi of Fusarium wilt disease) which do not cause Fusarium wilt disease (Fig. 5c). Unexpectedly, the *F. keratoplasticum*, nonpathogenic fungi of Fusarium wilt disease was proportionally higher in healthy than in diseased soils and were recognized to be an important characteristic of the fungal communities by which to distinguish between them (Fig. 5d, Supplementary Fig. 7, and Supplementary file 11). Co-occurrence analysis also showed more connections with *F. keratoplasticum* than with *F. oxysporum* in the healthy feature network. Nonpathogenic *Fusarium* has been used as a biocontrol agent of *Fusarium* wilt disease of many crops [67, 68].

Among the bacteria, *Xanthomonadaceae* exhibited a higher relative abundance in the diseased than in the healthy soils (Supplementary Fig. 5). Coincidentally, *Xanthomonadaceae* have been reported as phytopathogens [69, 70] and recently were shown to be positively correlated with Fusarium wilt disease [71]. In another report, monoculture of *Pseudostellaria heterophylla* increased the population size of *Xanthomonadaceae* in soil [72], suggesting that the bacteria might function to aid Fusarium wilt disease. Moreover, we also found *Gemmatimonadetes* in greater relative abundance and with more connections in the diseased network. These results implicate *Gemmatimonadetes* as a diseased-related microbial group, as also indicated in an earlier study of rhizosphere samples from Fusarium wilt-diseased lily [73].

We also identified some potential beneficial bacteria among the microbes with higher relative abundance and position in the healthy soil co-occurrence network. Among these were *Streptomyces* (*micabili*, Fig. 5), of which many strains are known to suppress *F. oxysporum* and promote plant growth [74–77]. Overall, the features that emerged from the models built to separate healthy and diseased soils provided convincing evidence of conserved features among the microbial communities associated with diverse Fusarium wilt-diseased crops from various soil types.

Our analysis, the first of its kind, yielded results that potentially can be used to distinguish if a soil is (or not) conductive to Fusarium wilt-diseased of a susceptible plant is grown in it. We set up the threshold value of disease incidence at 15%, to indicate that soils in which plants are grown with less than 15% disease incidence may be recognized as "health soils." Interestingly, our models were built excluding data from the soils collected from lily plots in China, but they were highly accurate in predicting the nature of the microbial communities present (Fig. 3i). We also recognize that our analysis did not include all possible studies, and that our future studies should be extended to include additional reports with different crops. In this regard we have built a website tool (https://wentaomicro.shinyapps.io/foc_16s/) to enable, from sequence data, the prediction of Fusarium wilt disease in susceptible plants and to improve the models by including new sequences. Furthermore, combining the bacterial and fungal sequencing data to construct a single model ultimately will be more convenient for researchers to compare with their own data, though we currently support the use of the separate bacterial and fungal models. It should be noted that our models are based on the sequencing data from soils those without drastic factitious disturbance, such as antimicrobial compounds application, that would likely affect the composition of the microbial community and impact the \predictive accuracy of the model. For example, the treatment of a soil with fungicides may result in a lack of disease incidence with a concomitant alteration in the composition of the fungal community. The loss of a proportion of the functional fungal community would likely result in aberrant model predictions. In this study, we combined global bacterial and fungal HTS data of Fusarium wilt disease soil samples from multiple independent sources for the detection of microbial patterns of disease and the identification of microbial community characteristics that can be used to predict soil health in relation to Fusarium wilt disease. Common microbial characteristics of Fusarium wilt disease were revealed among various crops. The 45 bacterial and 40 fungal OTUs selected through the RF method as models for the prediction can be considered as Fusarium wilt disease-associated core microorganisms. The abundance and diversity of *Fusarium* species and *Streptomyces* species are important for soil health [67, 78] and undisputedly, the higher relative abundance of *F. oxysporum* is clearly evident in the diseased soil. Our study provides a list of microbes that potentially take a pivotal role in the process of Fusarium wilt disease. Further studies on the interactions among these core microbes are needed to better understand the Fusarium wilt disease-associated microbial community and to control crop disease.

Data availability

The authors declare that the data supporting the findings of this study are available within the paper and its supplementary information files. Raw sequence data obtained in this study have been deposited in Genome Sequence Archive in the BIG Data Center, Chinese Academy of Sciences under accession codes CRA002340. All scripts for computational analysis and corresponding raw data are available at https://github.com/taowenmicro/Wen-etal-200214-paper-code.

Author Contributions JY and TW conducted all experiments, conceived the study, and wrote the paper; HZ and MZ collected sequencing data; QS conceived the study, supervised the study, and wrote the paper; CRP and LST provided critical comments on the study, and helped write the paper.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

1. Ley RE, Peterson DA, Gordon JI. Ecological and evolutionary forces shaping microbial diversity in the human intestine. Cell. 2006;124:837–48.
2. Whitman WB, Coleman DC, Wiebe WJ. Prokaryotes: the unseen majority. Proc Natl Acad Sci USA. 1998;95 6578–83.
3. Gentile CL, Weir TL. The gut microbiota at the intersection of diet and human health. Science. 2018;362:776–80.
4. Penesyan A, Kjelleberg S, Egan S. Development of novel drugs from marine surface associated microorganisms. Mar drugs. 2010; 8:438–59.
5. Chaparro JM, Sheflin AM, Manter DK, Vivanco JM. Manipulating the soil microbiome to increase soil health and plant fertility. Biol Fertil Soils. 2012;48:489–99.
6. Mäder P, Fliessbach A, Dubois D, Gunst L, Fried P, Niggli U. Soil fertility and biodiversity in organic farming. Science. 2002; 296:1694–7.
7. Classen AT, Sundqvist MK, Henning JA, Newman GS, Moore JA, Cregger MA, et al. Direct and indirect effects of climate change on soil microbial and soil microbial-plant interactions: what lies ahead? Ecosphere. 2015;6:1–21.
8. de Vries FT, Griffiths RI, Bailey M, Craig H, Girlanda M, Gweon HS, et al. Soil bacterial networks are less stable under drought than fungal networks. Nat Commun. 2018;9:3033.
9. Van Der Heijden MG, Bardgett RD, Van Straalen NM. The unseen majority: soil microbes as drivers of plant diversity and productivity in terrestrial ecosystems. Ecol Lett. 2008;11: 296–310.
10. Hawkes CV, Wren IF, Herman DJ, Firestone MK. Plant invasion alters nitrogen cycling by modifying the soil nitrifying community. Ecol Lett. 2005;8:976–85.
11. Yuan J, Zhao J, Wen T, Zhao M, Li R, Goossens P, et al. Root exudates drive the soil-borne legacy of aboveground pathogen infection. Microbiome. 2018;6:156.
12. De Corato U, Patruno L, Avella N, Lacolla G, Cucci G. Composts from green sources show an increased suppressiveness to soil-borne plant pathogenic fungi: Relationships between physico-chemical properties, disease suppression, and the microbiome. Crop Prot. 2019;124:104870.
13. Finkel OM, Castrillo G, Paredes SH, González IS, Dangl JL. Understanding and exploiting plant beneficial microbes. Curr Opin Plant Biol. 2017;38:155–63.
14. Berendsen RL, Pieterse CM, Bakker PA. The rhizosphere microbiome and plant health. Trends plant Sci. 2012;17:478–86.
15. Ploetz RC. Fusarium wilt of banana. Phytopathology. 2015;105: 1512–21.
16. Cha J-Y, Han S, Hong H-J, Cho H, Kim D, Kwon Y, et al. Microbial and biochemical basis of a Fusarium wilt-suppressive soil. ISME J. 2016;10:119.
17. Gordon TR. Fusarium oxysporum and the Fusarium Wilt Syndrome. Annu Rev Phytopathol. 2017;55:23–39.
18. Laurence MH, Burgess LW, Summerell BA, Liew ECY. High levels of diversity in Fusarium oxysporum from non-cultivated ecosystems in Australia. Fungal Biol. 2012;116:289–97.
19. Nyvad B, Fejerskov O. An ultrastructural-study of bacterial invasion and tissue breakdown in human experimental root-surface caries. J Dent Res. 1990;69:1118–25.
20. Klein E, Ofek M, Katan J, Minz D, Gamliel A. Soil suppressiveness to Fusarium disease: shifts in root microbiome associated with reduction of pathogen root colonization. Phytopathology. 2012;103:23–33.
21. Mendes LW, Mendes R, Raaijmakers JM, Tsai SM. Breeding for soil-borne pathogen resistance impacts active rhizosphere microbiome of common bean. ISME J. 2018;12:3038–42.
22. Xiong W, Li R, Ren Y, Liu C, Zhao Q, Wu H, et al. Distinct roles for soil fungal and bacterial communities associated with the suppression of vanilla Fusarium wilt disease. Soil Biol Biochem. 2017;107:198–207.
23. Ye XF, Li ZK, Luo X, Wang WH, Li YK, Li R, et al. A predatory myxobacterium controls cucumber Fusarium wilt by regulating the soil microbial community. Microbiome. 2020; 8:17.
24. Zhang S, Raza W, Yang X, Hu J, Huang Q, Xu Y, et al. Control of Fusarium wilt disease of cucumber plants with the application of a bioorganic fertilizer. Biol Fertil Soils. 2008;44:1073.
25. Fu L, Penton CR, Ruan Y, Shen Z, Xue C, Li R, et al. Inducing the rhizosphere microbiome by biofertilizer application to suppress banana Fusarium wilt disease. Soil Biol Biochem. 2017;104: 39–48.
26. Shen Z, Ruan Y, Xue C, Zhong S, Li R, Shen Q, et al. Soils naturally suppressive to banana Fusarium wilt disease harbor unique bacterial communities. Plant. 2015;393:21–33.
27. De Corato U, Patruno L, Avella N, Salimbeni R, Lacolla G, Cucci G, et al. Soil management under tomato-wheat rotation increases the suppressive response against Fusarium wilt and tomato shoot

growth by changing the microbial composition and chemical parameters. Appl Soil Ecol. 2020;154:103601.

28. Zhou D, Jing T, Chen Y, Wang F, Qi D, Feng R, et al. Deciphering microbial diversity associated with Fusarium wilt-diseased and disease-free banana rhizosphere soil. BMC Microbiol. 2019;19:161.

29. da C Jesus E, Marsh TL, Tiedje JM, de S Moreira FM. Changes in land use alter the structure of bacterial communities in Western Amazon soils. ISME J. 2009;3:1004–11.

30. Fierer N, Jackson RB. The diversity and biogeography of soil bacterial communities. Proc Natl Acad Sci USA. 2006;103: 626–31.

31. Mercado-Blanco J, Abrantes I, Barra Caracciolo A, Bevivino A, Ciancio A, Grenni P, et al. Belowground microbiota and the health of tree crops. Front Microbiol. 2018;9:1006.

32. Ramirez KS, Knight CG, de Hollander M, Brearley FQ, Constantinides B, Cotton A, et al. Detecting macroecological patterns in bacterial communities across independent studies of global soils. Nat Microbiol. 2018;3:189–96.

33. Qiu M, Zhang R, Xue C, Zhang S, Li S, Zhang N, et al. Application of bio-organic fertilizer can control Fusarium wilt of cucumber plants by regulating microbial community of rhizosphere soil. Biol Fertil Soils. 2012;48:807–16.

34. Wang B, Li R, Ruan Y, Ou Y, Zhao Y, Shen Q. Pineapple–banana rotation reduced the amount of Fusarium oxysporum more than maize–banana rotation mainly through modulating fungal communities. Soil Biol Biochem. 2015b;86:77–86.

35. Alabouvette C. Fusarium wilt suppressive soils: an example of disease-suppressive soils. Australas Plant Pathol. 1999;28:57–64.

36. Hornby D. Suppressive soils. Australas Plant Pathol. 1983;21: 65–85.

37. Andrews S. FastQC: a quality control tool for high throughput sequence data. 2010. http://www.bioinformatics.babraham.ac.uk/projects/fastqc.

38. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, et al. QIIME allows analysis of high-throughput community sequencing data. Nat Methods. 2010;7:335.

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

40. McDonald D, Clemente JC, Kuczynski J, Rideout JR, Stombaugh J, Wendel D, et al. The Biological Observation Matrix (BIOM) format or: how I learned to stop worrying and love the ome-ome. GigaScience. 2012;1:7.

41. Liaw A, Wiener M. Classification and regression by randomForest. R N. 2002;2:18–22.

42. Cortes C, Vapnik V. Support-vector networks. Mach Learn. 1995;20:273–97.

43. Wright RE. Logistic regression. In Grimm LG, & Yarnold PR, (Eds), Reading and understanding multivariate statistics (p. 217–244). Washington, DC: American Psychological Association; 1995.

44. Tibshirani R. Regression shrinkage and selection via the lasso. J R Stat Soc. 1996;58:267–88.

45. Sing T, Sander O, Beerenwinkel N, Lengauer T. ROCR: visualizing classifier performance in R. Bioinformatics. 2005;21: 3940–1.

46. Statnikov A, Henaff M, Narendra V, Konganti K, Li Z, Yang L, et al. A comprehensive evaluation of multicategory classification methods for microbiomic data. Microbiome. 2013;1:11.

47. Zhang J, Liu Y-X, Zhang N, Hu B, Jin T, Xu H, et al. NRT1.1B is associated with root microbiota composition and nitrogen use in field-grown rice. Nat Biotechnol. 2019;37:676–84.

48. Friedman J, Alm EJ. Inferring correlation networks from genomic survey data. PLoS Comput Biol. 2012;8:e1002687.

49. Csardi G, Nepusz T. The igraph software package for complex network research. Int J Complex Syst. 2006;1695:1–9.

50. Newman ME. The structure and function of complex networks. SIAM Rev. 2003;45:167–56.

51. Walters W, Hyde ER, Berg-Lyons D, Ackermann G, Humphrey G, Parada A, et al. Improved bacterial 16S rRNA gene (V4 and V4-5) and fungal internal transcribed spacer marker gene primers for microbial community surveys. Msystems. 2016;1:e00009–15.

52. McKay G, Brown AE, Bjourson A, Mercer P. Molecular characterisation of Alternaria linicola and its detection in linseed. Eur J Plant Pathol. 1999;105:157–66.

53. Adams RI, Bateman AC, Bik HM, Meadow JF. Microbiota of the indoor environment: a meta-analysis. Microbiome. 2015;3:49.

54. Cornejo-Granados F, Gallardo-Becerra L, Leonardo-Reza M, Ochoa-Romo JP, Ochoa-Leyva A. A meta-analysis reveals the environmental and host factors shaping the structure and function of the shrimp microbiota. PeerJ. 2018;6:e5382.

55. Koren O, Knights D, Gonzalez A, Waldron L, Segata N, Knight R, et al. A guide to enterotypes across the human body: meta-analysis of microbial community structures in human microbiome datasets. PLoS Comput Biol. 2013;9:e1002863.

56. Duvallet C, Gibbons SM, Gurry T, Irizarry RA, Alm EJ. Meta-analysis of gut microbiome studies identifies disease-specific and shared responses. Nat Commun. 2017;8:1784.

57. Rocca JD, Simonin M, Blaszczak JR, Ernakovich JG, Gibbons SM, Midani FS, et al. The Microbiome Stress Project: towards a global meta-analysis of environmental stressors and their effects on microbial communities. Front Microbiol. 2018;9:3272.

58. Gonzalez A, Navas-Molina JA, Kosciolek T, McDonald D, Vázquez-Baeza Y, Ackermann G, et al. Qiita: rapid, web-enabled microbiome meta-analysis. Nat Methods. 2018;15:796–8.

59. Baxter NT, Ruffin MT, Rogers MA, Schloss PD. Microbiota-based model improves the sensitivity of fecal immunochemical test for detecting colonic lesions. Genome Med. 2016;8:37.

60. Belk A, Xu ZZ, Carter DO, Lynne A, Bucheli S, Knight R, et al. Microbiome data accurately predicts the postmortem interval using random forest regression models. Genes. 2018;9:104.

61. Wilck N, Matus MG, Kearney SM, Olesen SW, Forslund K, Bartolomaeus H, et al. Salt-responsive gut commensal modulates TH17 axis and disease. Nature. 2017;551:585–9.

62. Manici L, Caputo F. Fungal community diversity and soil health in intensive potato cropping systems of the east Po valley, northern Italy. Ann Appl Biol. 2009;155:245–58.

63. Ploetz RC. Fusarium wilt of banana is caused by several pathogens referred to as Fusarium oxysporum f. sp. cubense. Phytopathology. 2006;96:653–6.

64. Shen Z, Xue C, Taylor PWJ, Ou Y, Wang B, Zhao Y, et al. Soil pre-fumigation could effectively improve the disease suppressiveness of biofertilizer to banana Fusarium wilt disease by reshaping the soil microbiome. Biol Fertil Soils. 2018;54:793–806.

65. Wang B, Li R, Ruan Y, Ou Y, Zhao Y, Shen Q. Pineapple–banana rotation reduced the amount of Fusarium oxysporum more than maize–banana rotation mainly through modulating fungal communities. Soil Biol Biochem. 2015a;86:77–86.

66. Wu X, Guo S, Jousset A, Zhao Q, Shen Q. Bio-fertilizer application induces soil suppressiveness against Fusarium wilt disease by reshaping the soil microbiome. Soil Biol Biochem. 2017;114: 238–47.

67. Forsyth LM, Smith LJ, Aitken EA. Identification and characterization of non-pathogenic Fusarium oxysporum capable of increasing and decreasing Fusarium wilt severity. Mycological Res. 2006;110: 929–35.

68. Lemanceau P, Alabouvette C. Biological control of fusarium diseases by fluorescent Pseudomonas and non-pathogenic Fusarium. Crop Prot. 1991;10:279–86.

69. Liu L, Kong J, Cui H, Zhang J, Wang F, Cai Z, et al. Relationships of decomposability and C/N ratio in different types of

organic matter with suppression of Fusarium oxysporum and microbial communities during reductive soil disinfestation. Biol Control. 2016;101:103–13.

70. Pieretti I, Royer M, Barbe V, Carrere S, Koebnik R, Cociancich S, et al. The complete genome sequence of Xanthomonas albilineans provides new insights into the reductive genome evolution of the xylem-limited Xanthomonadaceae. BMC genomics. 2009;10:616.

71. Li X, Zhang YN, Ding C, Jia Z, He Z, Zhang T, et al. Declined soil suppressiveness to Fusarium oxysporum by rhizosphere microflora of cotton in soil sickness. Biol Fertil soils. 2015;51:935–46.

72. Wu L, Yang B, Li M, Chen J, Xiao Z, Wu H, et al. Modification of rhizosphere bacterial community structure and functional potentials to control pseudostellaria heterophylla replant disease. Plant Dis. 2019;104:25–34.

73. Shang Q, Yang G, Wang Y, Wu X, Zhao X, Hao H, et al. Illumina-based analysis of the rhizosphere microbial communities associated with healthy and wilted Lanzhou lily (Lilium davidii var. unicolor) plants grown in the field. World J Microbiol Biotechnol. 2016;32:95.

74. Abbasi S, Safaie N, Sadeghi A, Shamsbakhsh M. Streptomyces strains induce resistance to Fusarium oxysporum f. sp. lycopersici race 3 in tomato through different molecular mechanisms. Front Microbiol. 2019;10:1505.

75. Liotti RG, da Silva Figueiredo MI, Soares MA. Streptomyces griseocarneus R132 controls phytopathogens and promotes growth of pepper (Capsicum annuum). Biol Control. 2019;138:104065.

76. Tahvonen R. Microbial control of plant diseases with Streptomyces spp. 1. Eppo Bull. 2010;18:55–9.

77. Yang Y, Zhang S-W, Li K-t. Antagonistic activity and mechanism of an isolated Streptomyces corchorusii stain AUH-1 against phytopathogenic fungi. World J Microbiol Biotechnol. 2019;35:145.

78. Cha J-Y, Han S, Hong H-J, Cho H, Kim D, Kwon Y, et al. Microbial and biochemical basis of a Fusarium wilt-suppressive soil. ISME J. 2015;10:119–29.