Development of a machine-learning model to identify the impacts of pesticides characteristics on soil microbial communities from high-throughput sequencing data

Ke, M.; Xu, N.; Zhang, Z.; Qiu, D.; Kang, J.; Lu, T.; ... ; Qian, H.

Citation

Ke, M., Xu, N., Zhang, Z., Qiu, D., Kang, J., Lu, T., ... Qian, H. (2022). Development of a machine-learning model to identify the impacts of pesticides characteristics on soil microbial communities from high-throughput sequencing data. *Environmental Microbiology*, *24*(11), 5561-5573. doi:10.1111/1462-2920.16175

Received: 30 March 2022 | Accepted: 11 August 2022
DOI: 10.1111/1462-2920.16175

RESEARCH ARTICLE

Development of a machine-learning model to identify the impacts of pesticides characteristics on soil microbial communities from high-throughput sequencing data

Mingjing Ke[1] | **Nuohan Xu**[1] | **Zhenyan Zhang**[1] | **Danyan Qiu**[1] | **Jian Kang**[1] | **Tao Lu**[1] | **Tingzhang Wang**[2] | **W. J. G. M. Peijnenburg**[3,4] | **Liwei Sun**[1] | **Baolan Hu**[5] | **Haifeng Qian**[1]

[1]College of Environment, Zhejiang University of Technology, Hangzhou, People's Republic of China

[2]Key Laboratory of Microbial Technology and Bioinformatics of Zhejiang Province, Hangzhou, People's Republic of China

[3]Institute of Environmental Sciences (CML), Leiden University, Leiden, The Netherlands

[4]National Institute of Public Health and the Environment (RIVM), Center for Safety of Substances and Products, Bilthoven, The Netherlands

[5]Department of Environmental Science, Zhejiang University, Hangzhou, People's Republic of China

Correspondence
Haifeng Qian, College of Environment, Zhejiang University of Technology, Hangzhou 310032, People's Republic of China.
Email: hfqian@zjut.edu.cn

Funding information
Key R&D Program of Zhejiang, Grant/Award Number: 2022C02029; National Natural Science Foundation of China, Grant/Award Numbers: 21577128, 21777144, 21777145, 21976161

Abstract

High-throughput sequencing (HTS) of soil environmental DNA provides an advanced insight into the effects of pesticides on soil microbial systems. However, the association between the properties of the pesticide and its ecological impact remains methodically challenging. Risks associated with pesticide use can be minimized if pesticides with optimal structural traits were applied. For this purpose, we merged the 20 independent HTS studies, to reveal that pesticides significantly reduced beneficial bacteria associated with soil and plant immunity, enhanced the human pathogen and weaken the soil's ecological stability. Through the machine-learning approach, correlating these impacts with the physicochemical properties of the pesticides yielded a random forest model with good predictive capabilities. The models revealed that physical pesticide properties such as the dissociation constant (pKa), the molecular weight and water solubility, determined the ecological impact of pesticides to a large extent. Moreover, this study identified that eco-friendly pesticides should possess a value of $pKa > 5$ and a molecular weight in the range of 200–300 g/mol, which were found to be conducive to bacteria related to plant immunity promotion and exerted the lowest fluctuation of human opportunistic pathogen and keystone species. This guides the design of pesticides for which the impacts on soil biota are minimized.

INTRODUCTION

Soil is a critical ecosystem at the interface of the atmosphere, lithosphere, hydrosphere and biosphere and is also the keystone to maintaining the health of plants, animals and humans by providing food and other ecosystem services (Chaparro et al., 2012; Gentile & Weir, 2018; Ley et al., 2006). Pesticides are extensively applied in agriculture to increase crop yields. This has resulted in serious contamination of soil ecosystems (Zhang, 2018). Thereupon it is a great challenge to balance the ever-increasing demands of the growing world population and acceptable environmental impacts, as affected by increasingly variable and increasingly extreme climate conditions (Pimentel & Burgess, 2013; Singh et al., 2022; Vörösmarty et al., 2000). It is therefore of paramount imperative for soil ecosystems to optimize inputs (mineral and organic fertilizers, pesticides) with minimal impact on humans, animals and the ecosystem. This requires the establishment of a well-functioning and well-validated assessment system of the impacts of pesticides on soil ecosystems.

Mingjing Ke and Nuohan Xu contributed equally to this work.

As integral components of microbial ecosystems, bacteria operate at the heart of many critical terrestrial, aquatic and host-associated processes, providing critical ecosystem functions, while playing a key role in various biogeochemical processes and in soil health (Howe et al., 2014; Li et al., 2022). High-throughput DNA sequencing methods have allowed for the unprecedented promotion of our understanding of the impacts of pesticide contamination on the soil bacterial biodiversity (Caporaso et al., 2012; Davison et al., 2015). Many empirical studies reveal the enormous biological diversity and heterogeneity of soil communities after pesticide application (Monard et al., 2011; Xu et al., 2019). Historically, the studies that evaluated the factors modifying the different effects of pesticides on soil microbial communities, were mostly focused on soil properties at varying pesticide concentrations, different exposure times and so on (Deng et al., 2022; Zheng et al., 2022). However, little attention is paid to the characteristic properties of the pesticides (Feng et al., 2019; Liu et al., 2019; Wang et al., 2020). Furthermore, it is a great challenge to draw consistent and convincing rules with regard to the impacts of pesticides on soil microbial communities. The complexity of predicting the impacts of pesticides could be significantly reduced if it was possible to attribute the impact of pesticides on microbial ecological systems to a limited number of pesticides characteristics on the basis of existing literature reports. Rumschlag et al. (2020) demonstrated consistent effects of pesticides on aquatic microbial communities and ecosystem functions, which could be attributed to some specific pesticide classes or types of pesticides, and this considerably simplified the prediction of the complex effects of pesticides on aquatic ecosystems. Currently, we are far from a clear understanding of soil communities and ecosystems. In particular, the specific linkages between pesticide characteristics and soil community biodiversity and ecosystem functions are not well established. Solving these problems would allow optimizing of environmental inputs (such as optimizing pesticide design, application of microbial fertilizer, etc.), to reduce environmental pollution, as well as to enhance environmental and agricultural sustainability.

However, high heterogeneity in technical (e.g., sequencing platforms, sequencing depth, PCR bias, different clustering methods, etc.) and soil environmental (e.g., soil pH, soil properties, sampling location, etc.) factors between sequencing studies, caused great difficulties in generalized assessment for the impact of pesticide characteristics on the soil community structure (Mcdonald et al., 2012; Pawluczyk et al., 2015). Machine learning is a collection of data-analytical techniques, each of which has the potential to handle biological networks and large heterogeneous multi-dimensional datasets to construct models that simulate complex relationships (Camacho et al., 2018; Zhang, Zhang, Wang, et al., 2022). Furthermore, the integration of microbial sequence data from independent studies has been a favourable option. For example, Ramirez et al. (2017) revealed macroecological patterns in bacterial communities across global soils for addressing global-scale biogeography questions, and Yuan et al. (2020) answered ecological questions by identifying biological indicators and featured commons of the wilt-diseased soil microbiome. The objectives of this study focused on: (1) identifying the effect of pesticides on the soil microbial community composition as well as the variation in beneficial and pathogenic bacteria; (2) determining how pesticides affect microbial stability and keystones species using modularity and network cohesion; (3) establishing a predictive model to correlate pesticide characteristics with soil microbiota variation. All of these targets can be addressed based on merging independent taxonomy-based data sets, using powerful machine learning approaches. The accurate prediction of the contribution of pesticides characteristics on soil ecosystem function variation could guide the design of eco-friendly pesticides.

EXPERIMENTAL PROCEDURES

Data processing (per study)

The microbial sequence data were processed using the QIIME 22020.8 core distribution, following the standard operating procedures (Bolyen et al., 2019). Before using Cutadapt to remove adaptors and primers (Martin, 2011), an initial visualization of reading quality using the packages FastQC was performed. Then, the raw read files of independent studies were imported to QIIME 2. We used Dada2 to join paired-end reads (Callahan et al., 2016), denoise sequences as well as resolve amplicon sequence variants (ASVs). All low-quality reads were then filtered using default quality thresholds.

Combined processing

All studies were merged by QIIME2's merge-seqsc commands. To normalize the difference of 16S rRNA variable regions, all feature sequences were annotated with the full-length 16S rRNA gene SILVA database for subsequent analysis. Besides, to address PCR biases and biases associated with rare taxa and some groups which could be over-amplified, data processing was limited to the following strategies: (1) deletion of samples with a total number of sequences less than 2000; (2) deletion of ASVs for which the absolute abundance was less than 10; (3) Removed ASVs with mitochondria and chloroplast; Random filtering is needed in case of studies containing more than 90 samples as well as less than 10 samples (Buda et al., 2018; Thabtah et al., 2020; Xu et al., 2022; Yuan et al., 2020). We finally acquired 297 controls and 430 pesticide-treated samples (Data Set S1) that encompassed 8 herbicides, 2 insecticides and 1 agricultural fungicide.

Co-occurrence network analysis

We constructed two types of co-occurrence networks. For all networks, the significance of the correlations (the Spearman's correlation coefficient >0.5 at $p < 0.05$) between taxa abundances were calculated using the R psych package (version 4.0.3).

To explore the effect of pesticides on the stability of the soil microbial network, we performed Spearman rank correlations between all pairs of bacteria at the genus level within pesticides-treated and pesticides-free groups separately (as shown in Figure 2A, B). To illustrate the complexity and stability of the microbial community network, topological parameters including modularity, nodes, as well as the number of negative: positive cohesion were calculated by Gephi (version 0.9.2) (Assenov et al., 2007; Jacomy et al., 2014). We identified keystone OTUs separately for the pesticides and the control interaction networks and defined them as those nodes within the top 1% of node degree values of each network (Banerjee et al., 2016; Hartman et al., 2018).

We then constructed a co-occurrence network to visualize correlations between 184 genera that were present both in pesticides-treated and pesticides-free groups (Figure S2). Seven subnetworks (modules) were extracted from the merged network using the Louvain algorithm (Blondel et al., 2008; Csardi & Nepusz, 2006). Louvain starts from the community structure that separates all vertices. Next, it tries to move each vertex from its community to another, picks the move that increases the modularity most, and iterates until no change increases the modularity anymore. It then replaces the vertices with the detected communities and performs the same operations on the newly obtained graph, until the modularity cannot be increased anymore (Csardi & Nepusz, 2006; Didier et al., 2018).

Biomarker analysis

To acquire the biomarkers with pesticide treatment, we constructed a classification model that the relative abundances of bacterial taxa at the class level were calculated by the RandomForest package in R (version 4.0.3), and all parameters were default. Microbes were ranked by feature importance based on the parameter of mean decrease accuracy. The number of biomarkers was determined using a 10-fold cross-validation implemented with the function 'rfcv' in the RandomForest package with five repeats. The stabilized cross-validation error was obtained when using 10 microbes, therefore, the 10 most important microbes as the biomarkers related to pesticides treatment.

Construction of predictive RF models

We used an RF regression model to construct the relationship between pesticides characteristics and microbial abundance. We constructed seven separate RF regression models for seven microbial modules. For every model, the physical and chemical properties of the pesticides served as independent variables, and the total relative abundances of microbes in each module served as dependent variables. To avoid overfitting, the original dataset was partitioned into 10 folds, in which 90% of the training data were used to train the classifiers, and the remaining 10% was used for validation of the trained classifiers. Furthermore, the parameter of R^2 (coefficient) was used to ensure the result's reliability and applicability, and the parameter of RMSE (root mean square error) was used to ensure the accuracy of the result. We selected the model with an R^2 value greater than 0.8 for subsequent functional analysis from seven modules. To estimate the importance of different properties, the increase of mean-square error (% IncMSE) of each module was calculated by the R package 'RandomForest'. The construction and prediction performance of the RF regression model (n_estimators $= 8$) and the verification of the 10-fold cross-validation method was both used by the scikit-learn (Version 0.24.0) library of the python language (Python 3.9).

Statistical methods

The α-diversity index of the Shannon and Richness indices was calculated using the R vegan and picante package (version 4.0.3). Principal coordinates analysis (PCoA) plots were generated from Bray–Curtise distance created using R package ggplot2. Permutational multivariate analysis of variance (PERMANOVA by ADONIS, transformed data by Bray–Curtise, permutation $= 999$) was used to determine if beta diversity differed between the pesticides treated and the control groups. Opportunistic human pathogens were searched from an online database (https://www.hartmann-science-center.com/en/hygiene-knowledge/pathogens-a-z). Beneficial bacteria were collected from publications that were listed in Data Set S2. To avoid false-positive results, the p-value of Spearman's correlation in network analysis was amended using the Benjamini–Hochberg's FDR (false discovery rate) method (Chen et al., 2018). The Spearman correlations (positive correlation: $R > 0$; positive correlation: $R < 0$; $p < 0.05$) of the microbial trait and drivers were calculated using the R psych package. We considered '$R^2 > 0.8$' to be a model with strong applicability and more reliable prediction results. The remaining results are presented as the means ± standard deviations (SDs). Significant differences ($p < 0.05$) were evaluated by the Kruskal–Wallis test using the R function kruskal. test (version 4.0.3).

RESULTS

Data collection and description

Microbial high-throughput sequencing (HTS) data in pesticide-treated and reference soils were collected by means of a literature search using the keywords 'Pesticides and soil microbiome,' 'Pesticides and soil community,' 'Herbicides and 16S,' 'Insecticide and 16S,' and 'Fungicides and 16S' in Google Scholar and the National Center for Biotechnology Information (NCBI) Sequence Read Archive (SRA) database. Unfortunately, the collection of microbial community HTS data was heavily constrained. Only parts of the assays in the most eligible publications have accession numbers of raw sequence data sets that could be downloaded from public databases (we obtained 30 studies with accession numbers from 600 studies approximately). Besides, some HTS data with incomplete microbial community records and with a sample name that was not clearly marked, could not be used for analysis. Given the restrictions indicated, we finally acquired 20 independent studies from 10 countries (Figure 1A). Details of the metadata of these 20 independent studies are provided in Data Set S3, which includes sequence information such as the description of the study, accession number, location, assay, primers, collection time and so on.

Diversity and taxonomic profile after pesticides treatment

To assess how microbial community properties change after pesticides exposure, we compared α-diversity indices (Richness and Shannon). No significant differences between the bacterial Shannon index of pesticide-treated and pesticide-free soils were found, while a significant increase was observed in Richness in pesticide-treated groups (Figure 1B, $p = 0.0268$). PCoA with Bray–Curtise distance showed that the

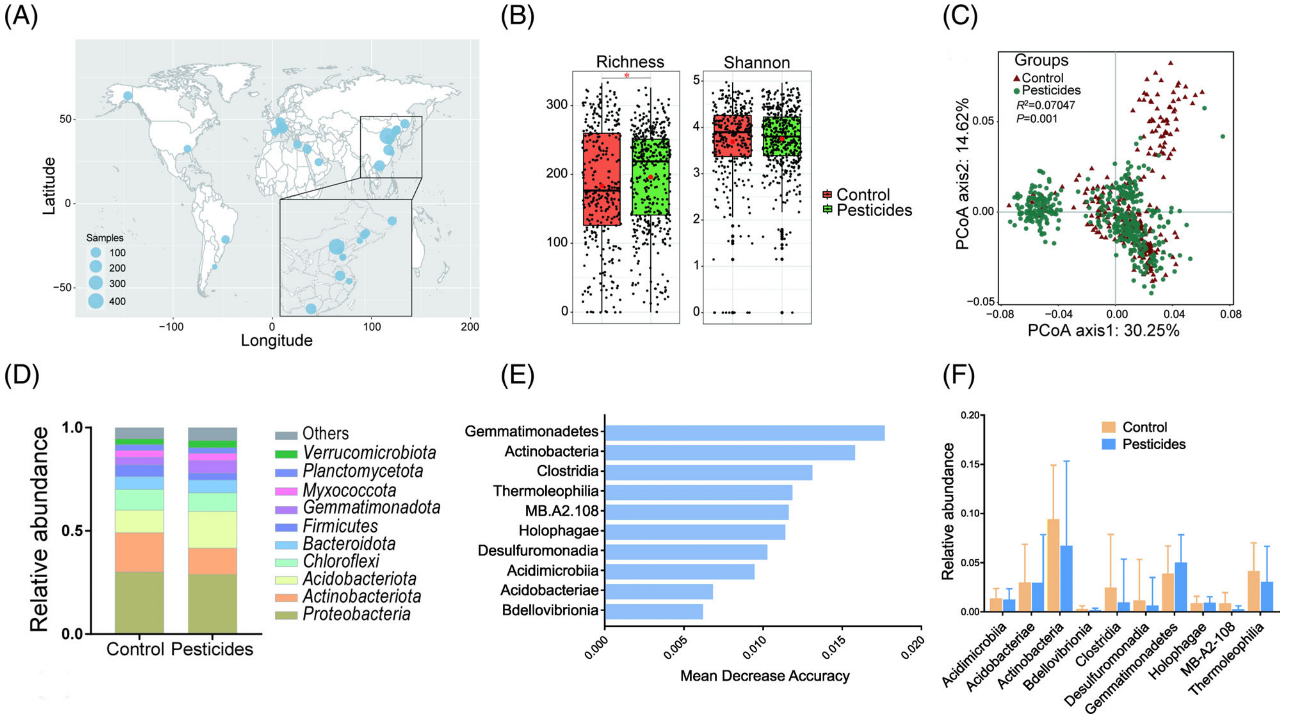

FIGURE 1 Overview of sample collection and effects of the pesticides on diversity and structure of microbial communities. (A) Sample profile. The geographic region includes information about the country. Also, the number of independent studies and samples were displayed. A node represents a single independent experiment. (B) Diversity index (Shannon and richness). *Represents statistically significant differences at $p < 0.05$ from a one-way ANOVA. (C) Principal coordinate analysis (PCoA) score plots of bacteria community profiles in different treatments. (D) Taxonomic comparison of the 10 most abundant phyla in the pesticide-treated and control groups. (E) The importance of the top 10 biomarker taxa at the class level. Line charts represent the changes of tenfold cross-validation error with the increasing number of classes. (F) Relative abundance (RA) of top 10 biomarkers in pesticide-treated and control groups.

bacterial communities changed significantly after pesticide treatment (Figure 1C, $R^2 = 0.07047$, $p = 0.001$, ANOSIM). At the phylum level, the soil bacterial communities in all samples after pesticide treatment were predominantly composed of *Proteobacteria*, *Actinobacteriota*, *Acidobacteriota* and *Chloroflexi* (Figure 1D). Several bacteria phyla, such as *Acidobacteriota* ($p < 0.0001$), *Gemmatimonadota* ($p < 0.0001$) and *Verrucomicrobiota* ($p < 0.001$) presented higher abundance (increased significantly by 25–64%) than in the control, while *Actinobacteriota* ($p < 0.0001$) and *Firmicutes* ($p < 0.001$) were at lower relative abundance (significant decreases by 33% and 41%, respectively) (Figure 1D). The Venn diagram revealed a similar microbial community composition as most of the bacterial taxa were shared in control and pesticide-treated soils, but there were still more unique bacterial taxa in the pesticide-treated soils (317) than in the control (230) (Figure S1). Generally, our results suggested that pesticide contamination disturbed predominated phyla bacterial taxa, and also exhibited a significant alteration in bacterial diversity.

A classification RF model to correlate soil bacterial taxonomic biomarkers with pesticides treatment

The random forest classification model which predicts whether or not pesticides have been treated, was constructed based on the microbial community composition. Biomarkers were chosen based on the feature importance of the classification model. We performed a 10-fold cross-validation with five repeats to evaluate the importance of indicator bacterial class. The cross-validation error was achieved minimum when 10 important classes were used, which were regarded as biomarkers in the model (Figure 1E). Of these, 5 classes (*Actinobacteria*, *Bdellovibrionia*, *Clostridia*, *MB-A2-108*, *Thermoleophilia*) showed lower relative abundance in pesticide treatment than the control, and only *Gemmatimonadetes* increased in pesticide treatment (Figure 1F). The biomarker for soil microbiota identification can be served as a novel approach for evaluating the effects of pesticides on soil health.

Pesticide treatment destabilizes microbial networks

Our results showed that pesticides could alter soil microbial community diversity and structure, but it was still uncertain whether and how pesticides impact the network stability of microbial communities, which decides their function and sustainability. Microbial network complexity was displayed by various network topological parameters including nodes, centrality, the ratio of *negative*: *positive* cohesion and network modularity (Figure 2A, B). The resulting size (total nodes) of the networks without isolated nodes was decreased by 14% upon pesticide treatment compared with the control group (Figure 2C). Besides, the ratio of negative: positive cohesion also decreased shapely from 0.0391 to 0.0118 after pesticide treatment (Figure 2D). The centrality and modularity of the microbial community networks decreased slightly after pesticides-treatment, indicating that microbial communities after pesticides treatment are less compartmentalized than in the control group (Figure 2E, F). We also identified the keystone species separately for the unpolluted and pesticides-polluted microbial interaction network (Data Set 4). Ten keystone genera were found in the control micro-network with high node degree values, including *Ellin6055*, *SJA-15*, *Syntrophorhabdus*, *BBMC-4*, *BSV26 Bacteroidetes_vadinHA17* and so on. In the pesticides network, the 9 genera identified as keystones were *Crossiella*, *Marmoricola*, OLB13, *Blastocatella*, *Stenotrophobacter*, *Actinomadura*, *AT-s3-28*, uncultured group, *Microcoleus-Es-Yyy1400*. These genera belonged to the phyla *Actinobacteriota*, *Chloroflex* and *Cyanobacteria*. All of the keystone species did not overlap between the unpolluted and pesticide-polluted microbial interaction networks.

The profile of beneficial and pathogenic soil bacteria after pesticides treatment

The compositional alteration of beneficial and pathogenic bacteria was a critical indicator of soil health, which was closely associated with the ability of plant defences and crop production (Lehmann et al., 2020). To disentangle the consistent impacts of pesticides on beneficial and pathogenic bacteria taxa, we detected 35 plant growth-promoting and soil disease suppression bacteria, as well as 40 pathogenic bacteria related to human and soil-borne bacteria pathogens by comparing our data with the online website (Data Set S2). Although several bacteria (*Chthoniobacter*, *Steroidobacter*, *Acidibacter* and *Latescibacterota*) were increased, 6 out of 43 beneficial bacteria showed a markable reduction as compared to unpolluted soil (Figure 3), such as *Nocardioides*, *Streptomyces*, *Bacillus*, *JG30-KF-CM45*, *Agromyces* and *Terrabacter*. Besides, pesticide treatment induced a significant increase of 8 out of 40 pathogenic bacteria, including *Blastocatella*, *Ferruginibacter*, *Brevundimonas*, *Legionella*, *Flavobacterium* and so on (Figure 3), while it decreased the abundances of *Roseomonas*, *Clostridium_sensu_stricto_1*, *Escherichia-Shigella* and *Aeromonas*, compared with unpolluted soil.

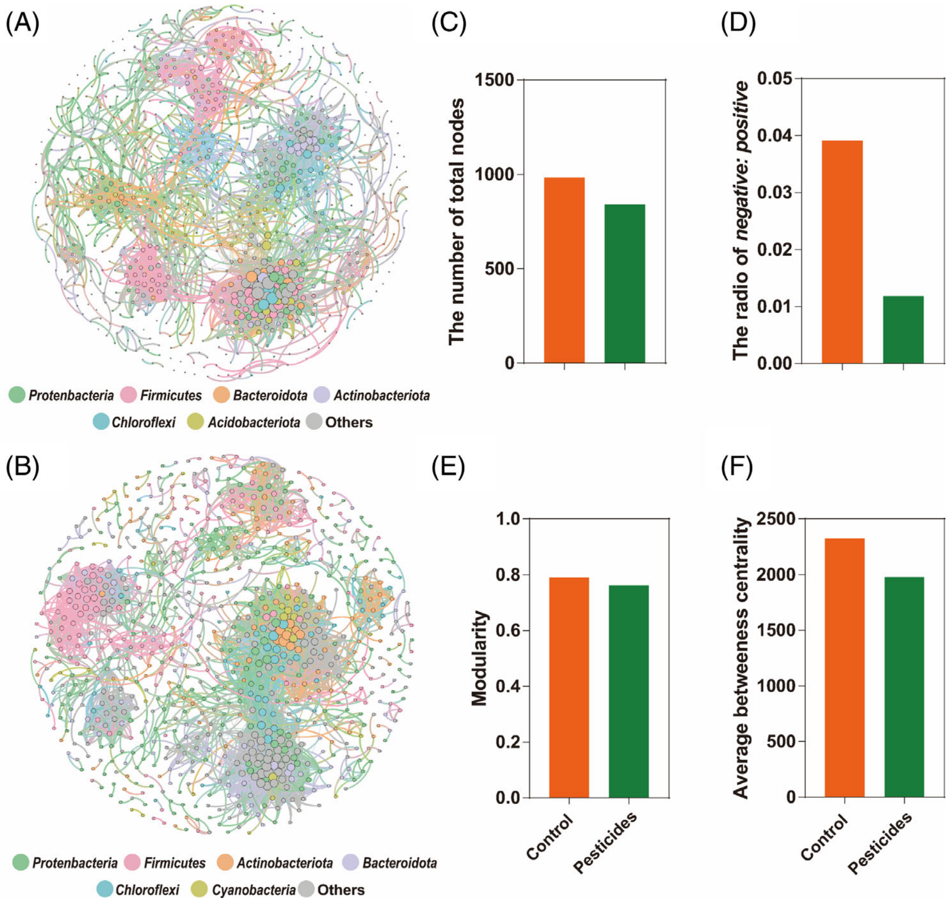

FIGURE 2 Effects of pesticides on microbial networks. (A) Visualization of constructed soil ecological networks treated with pesticides. (B) Visualization of constructed soil ecological networks in the control group. Various network topological parameters: (C) nodes, (D) the ratio of *negative: Positive*, (E) modularity and (F) centrality.

Development of a model to connect pesticide characteristics and microbial community traits

Given the significant discrepancy in the diversity and the interactions of microbial communities caused by pesticide treatments, it is imperative to figure out which pesticide characteristics played a decisive role in this interaction. We established a regression model using an RF to reveal the attributive mode of pesticides characteristics to microbial community traits (Figure 4A). A total of 24 physicochemical properties of pesticides were collected to serve as independent variables in the model (Data Set S5). Seven microbial modules that are expected to share some homologous function features among the community and to act collectively in the microbial community mechanisms (Chun et al., 2019), were identified based on the Louvain algorithm from an association network which visualizes co-occurrence patterns between pesticide treatments and the control group (Figure S2). The total relative abundance of microbes in each module served as the model-dependent variable (Data Set S6). To pledge suitable fitness, we used 10-fold cross-validation for evaluating the model performance. The model for 10-fold cross prediction displayed a good performance with values of R^2 mostly ranging from 65% to 96%. On the basis of this result, the three modules of the prediction models (modules 1, 3 and 7), with a good value of R^2 (>0.8) and low RSEM (<0.05) were chosen to analyse the contribution of pesticide characteristics on microbial traits (Figure 4B, Figure S3). For module 1, the dissociation constant (pKa) explained 21.7% of the variation in microbial communities associated with pesticide treatment, and amido-groups (9.9%) dominated the contributions of the chemical properties of the pesticides applied. The molecular weight, explaining 25% of the microbial alteration induced by pesticides was the most important physical factor for module 3 after pesticides addition, while chemical properties related to the ester functionality accounted for 3.9% of the microbial alteration. The microbial traits in Module 7 were mostly related to the water solubility of the pesticides (21.8%), while heterocyclic chemical groups (12.3%) were the next most important contributor to microbial community diversity in this module (Figure 4C).

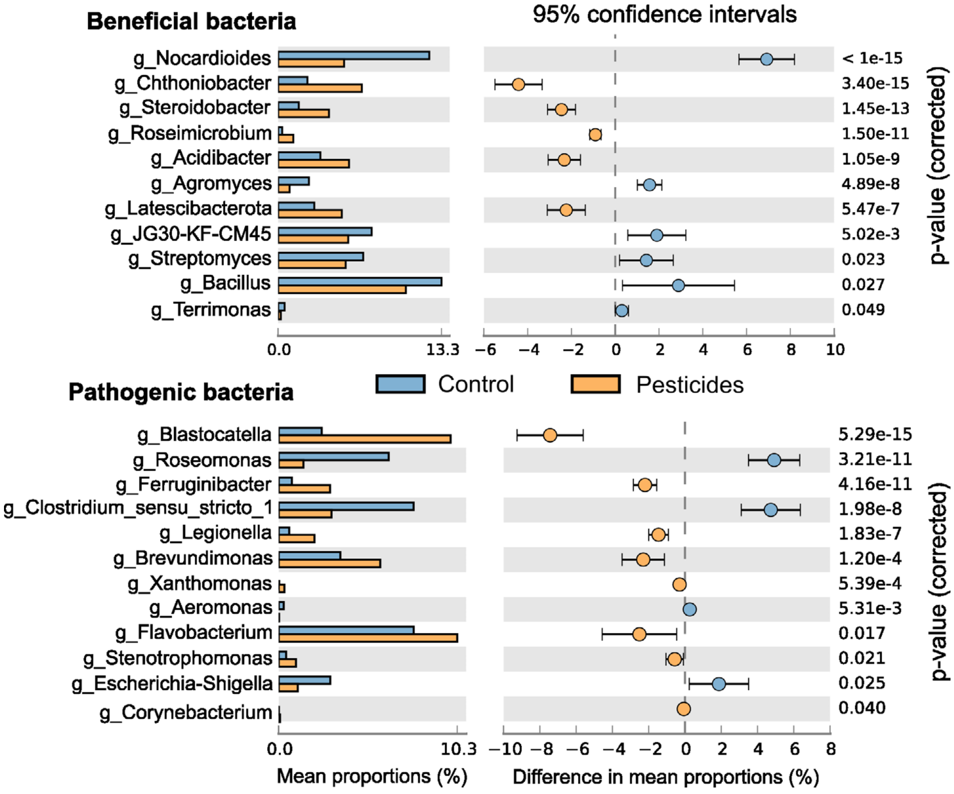

FIGURE 3 Different beneficial and pathogenic soil bacteria at the genera level in pesticide-treated and control groups. The *p*-values were amended by Benjamini–Hochberg FDR. The significant differences were evaluated by the two-tailed Welch's *t*-test ($p < 0.05$)

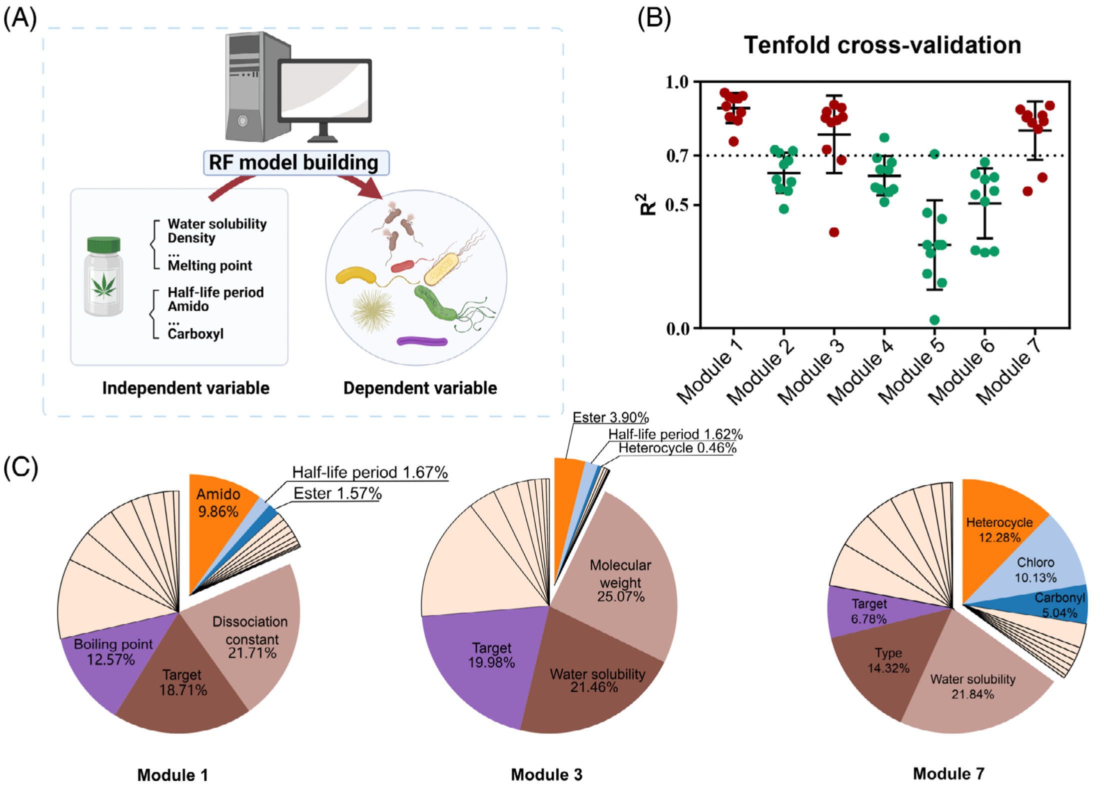

FIGURE 4 Model building by RF for connection between pesticide characteristics and microbial community trait and function. (A) RF model building (B) the values of R^2 in the validation model evaluated by tenfold cross-validation. (C) The contribution of pesticides characteristic on microbial traits.

Linking crucial pesticides physical properties to soil ecosystem functioning

Our RF model indicated that the physical properties of pesticides played important roles in the impact of pesticides on microbial ecosystems. The ecological risk of pesticides can be minimized in targeted ways, as achieved by only applying pesticides because of their specific physical properties, which have a minimal impact on soil biota. Most of the differentiated pathogenic bacteria belonged to module 1, while most of the differentiated beneficial bacteria belonged to module 3. Among keystone species in the pesticides network, there are 2 species belonging to module 1, and 4 species belonging to module 3. We, therefore, constructed a relationship between specific physical properties of the pesticides in model 1 and in model 3 and the beneficial and pathogenic bacteria as well as the keystone species (Figure 5, Figure S4). The relative abundance of pathogenic bacteria and keystone species in module 1 was increased significantly by pesticides for which the pKa ranged from −5 to 5, while they exerted no alteration or a markable decrease when the pesticides were equipped with a value of the pKa in the range of (< −5) and (≥5) (Figure 5A, Table S1). Pesticides with a molecular weight in the range of (200–300) and (≥350) did not alter or increase the beneficial bacteria significantly in module 3 (Figure 5B, Table S2). Furthermore, the keystone species were sensitive to the application of pesticides for which the molecular weight is greater than 300 (Figure 5B, Table S2).

DISCUSSION

Using amplicon sequencing approaches, researchers gained important insights into the biodiversity and potential functions of microbial communities from individual pesticides-polluted soils in natural or laboratory conditions (Bhuiyan et al., 2011; Delmont et al., 2012; Tringe et al., 2005). Various heterogeneous conditions in individual mesocosm experiments have made it difficult to clearly and uniformly understand the impact of pesticides on soil microorganisms. In this study, we performed a large-scale analysis of the soil microbial traits and the functional features by merging microbial community HTS data. However, although the urgent need for standardization of data deposition and management was emphasized in earlier reports (Ramirez et al., 2017; Santos & Branco, 2012; Tedersoo et al., 2015), the acquisition process of microbial community HTS data was still constrained by many challenges. On the one hand, the poor accessibility of HTS data and environmental metadata remains unresolved. On the other hand, data analysis was restricted to publicly available data with low quality as well as incomplete information (e.g., barcoded primers, sampling location, incompatible sample names, etc.). Thus, it should be strengthened for the standardization and management of raw data submission in public databases.

Identifying the general characteristics of microbial biodiversity in agricultural ecosystems and their response to environmental disturbances is essential for protecting ecosystem processes and regulating the management and common practice of agricultural production. Soil biodiversity is known to be associated intimately with soil quality and ecosystem functioning and services, such as soil decomposition and the global carbon, nitrogen and other nutrients cycles, especially in agricultural ecosystems (de Zelicourt et al., 2013; Naylor et al., 2020; Toju et al., 2018; Xun et al., 2019). We noticed that pesticides contamination induced a markable increase of α-diversity as well as a considerable alteration of the composition of dominant microbial communities. It is likely that pesticides may increase the ecological niche width by increasing the proportion of source rare taxa, leading to increased α-diversity (Okie et al., 2015). Besides, Coyte et al. (2015) predicted that high species diversity destabilized microbiome communities via a model that was built on ecological network theory. The most dominant bacterial taxa accounted for a high proportion of network hubs and connectors and were proven to be strongly associated with network reorganization (Shi et al., 2016; Williams et al., 2014). The strong response of soil microbial communities induced by pesticides suggests a disturbance of the soil ecological network.

In our study, pesticide contamination induced a dramatic fluctuation of the different network properties, with lower centrality and modularity. We also found a reduced dominance of negative interactions after pesticide treatment. Negative links stabilize an ecological network after environmental disturbance (Coyte et al., 2015; Hernandez et al., 2021; Yuan et al., 2021). Our findings thus indicate that pesticides weakened the microbial network stability (Rooney et al., 2006). This can be explained by the increase in bacterial diversity and markable alteration of dominant taxa. Besides, considerable amounts of keystone species that play an ecologically important role by determining community dynamics and microbiome functioning, were varied dramatically after pesticides exposure (Banerjee et al., 2016; Hartman et al., 2018; Xun et al., 2021). These keystone species constitute a complex ecological network where they are interconnected through exchanges of materials, energy, and information and withstand a diverse array of abiotic and biotic factors (e.g., temperature, climate warming, drought, etc.) (de Vries et al., 2021; Xun et al., 2019; Yuan et al., 2021). Therefore, our findings indicated that pesticides destabilized the soil ecology, thus potentially weakening the soil biotic competition.

Scientists have paid more attention to how pesticides manage pests and diseases and enhance crop yield. Unfortunately, really little attention was paid to

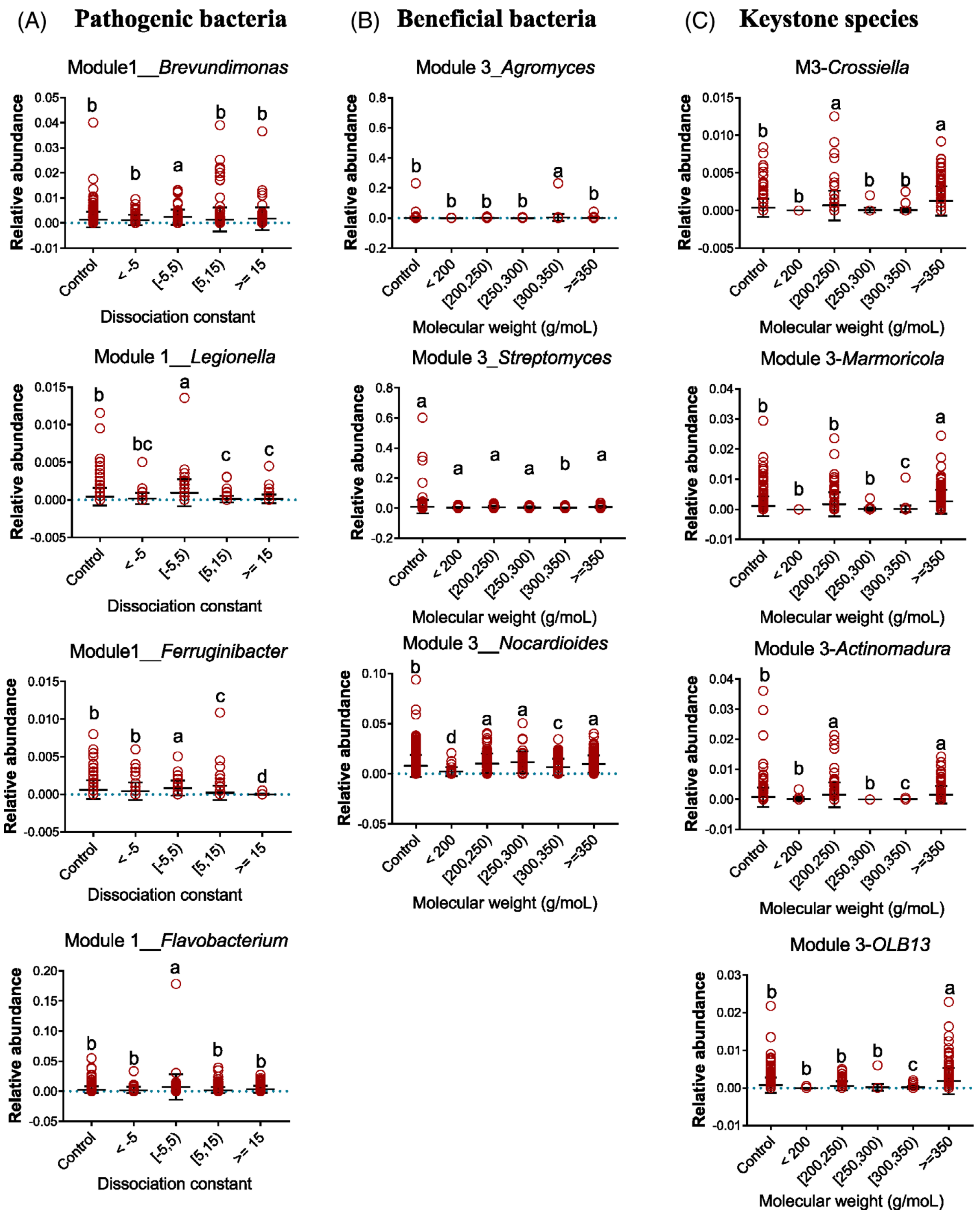

FIGURE 5 The scatterplot showed the relative abundance (RA) alteration of soil ecological parameters in different quantitative value distributions of physical properties of pesticides. (A) The RA alteration of pathogenic bacteria in modules 1 across different values of the dissociation constant; (B) the RA alteration of beneficial bacteria in modules 3 across different values of molecular weight; (C) the RA alteration of keystone species in modules 3 across different values of molecular weight. The comparison between the control and the pesticide-treated group was based on means ± standard deviations (SDs) (the midline in the figure). Different letters represent significant differences between different treatments ($p < 0.05$).

the impact of these chemicals on non-target microorganisms, especially the beneficial and pathogenic bacteria in the soil (Duke, 2018). This study revealed that pesticides had a considerable impact on bacteria associated with soil health and plant immunity as well as human health. Several genera of bacteria like *Streptomyces*, *Nocardioides* and *Bacillus*, which are well-known not only to be fungal antagonists of soilborne fungal pathogens but to also trigger plant-induced systemic resistance (ISR) and promote plant growth, were decreased significantly in pesticide treatment (Inderbitzin et al., 2017; Lazcano et al., 2021). Furthermore, the plant pathogenic bacteria *Xanthomonas* and *Blastocatella* were increased significantly in pesticides treatment (Molder et al., 2021). Therefore, although pesticides protect plants from diseases or insects, they could potentially contribute to the increase of soil pathogen invasiveness and the weakening of plant immunity after repeated application of pesticides, making soils and plants more dependent on pesticides. Besides, *JC30-KF-CM50* which has the capability of degradation of antibiotic pollutants was reduced by pesticides exposure, while the dominant hosts of antibiotics resistance genes (ARGs), such as *Ferruginibacter* and *Flavobacterium* were highly abundant in pesticide-treated soil (Wang et al., 2020). Notably, human opportunistic pathogens such as *Brevundimonas*, *Stenotrophomonas* and *Legionella* (Berg et al., 2009; Chun et al., 2019), were promoted after pesticides treatment (*Legionella*). Generally, our results reminded us that pesticide contamination might accelerate the spread of pathogens and ARGs in the soil.

Machine learning has the capability to identify meaningful and regular patterns in a complex set of data, consistently predicting various biological responses (Ahneman et al., 2018; Jiang et al., 2021; Jordan & Mitchell, 2015). By a machine learning approach, we identified biomarker taxa that can serve to discriminate the pesticides impact on soil biodiversity from a complex chemical soil system, which can be a novel quality indicator for agricultural soil remediation effectiveness of pesticides-pollution. Besides, compared with support vector machines, neural networks, and other machine learning algorithms, the RF model has been demonstrated to exhibit excellent prediction accuracy on heterogeneous metagenome data with quantitative and qualitative factors (Vujkovic-Cvijin et al., 2020; Zhang, Zhang, Cui, et al., 2022). Ban et al. (2020) clarified the complex relationships between nanoparticle properties and corona composition using RF, and successfully predicted the cellular recognition mediated by a functional corona protein. Wright et al. (2021) also revealed that environmental and methodological factors, and not plastic type were shaping dominantly the plastisphere through the construction of an RF model. We constructed a good model (with $R^2 > 0.8$) with high accuracy and reliability, that allows screening of the most important properties of pesticides that explain most of the variance in microbial soil composition after pesticide application. This provided the possibility for exploring the most important properties of pesticides that modify microbial communities. Importantly, our well-performing predictive models revealed that the physical pesticide properties determined the ecological impact of pesticides to a larger extent than the contribution of the chemical pesticide properties. The dissociation constant (pKa), the molecular weight, and water solubility were the most important pesticide properties that determine the microbial alteration of the three modules. Notably, pesticides type has always been recognized as the most important factor affecting the microbial community in many single studies. However, this factor was less important than other physical properties in our models.

The pesticides addition exhibited a strong fluctuation in keystone species, suggesting a disruption of the soil ecological system. The profile of keystone species as well as the beneficial and pathogenic bacteria across different value ranges of pesticides physical properties provide a baseline for the design of eco-friendly pesticides. Although pesticides pollution reduces the stress competitiveness and disease resistance of soil microbial systems, subsequently constraining soil ecosystem function, these negative impacts can be minimized by adjusting and restricting the properties of the pesticides applied. Pesticides equipped with a pKa value of more than 5 could decrease the abundance of the human opportunistic pathogen. Besides, pesticides with a molecular weight in the range of 200–300 g/mol and more than 350 g/mol were conducive to the growth of beneficial bacteria related to plant immune promotion. Keystone species exerted the lowest fluctuation with a molecular weight below 300 g/mol.

CONCLUSION

Our results indicated that pesticides significantly increased microbial diversity and disturbed dominant phyla bacteria. The decrease in soil microbiome stability and the increase of beneficial bacteria involved in the ability for disease suppression in soil and the immunity of the plants themselves indicated that a susceptible soil system was induced by pesticides. Importantly, we revealed through machine learning methods that pesticides characteristics, such as pKa, the molecular weight and water solubility, play a large role in shaping soil microbial communities as well as functional traits. This study provided an accurate prediction model to uncover pesticide characteristics that determine soil microbial functionality and health, as well as a guide to design and optimize pesticide molecules for which the environmental risk is minimized.

AUTHOR CONTRIBUTIONS

MJK and NHX designed and performed all practical aspects of the study with guidance from HFQ. MJK wrote the first draft of the manuscript and HFQ, BLH and WJGMP contributed substantially to revisions. JK and DYQ and searched all related literature and collected the publicly accessible databases. NHX, TZW and ZYZ were responsible for the Machine learning model construction and related data analysis. MJK and

NHX performed the visualization of all data and the artistic design of all figures. HFQ and TL acquired funding for this project. All authors read and approved the final manuscript.

CONFLICT OF INTEREST

The authors declare that they have no competing interests.

DATA AVAILABILITY STATEMENT

Raw sequence data associated with this study have been published in the National Center for Biotechnology Information (NCBI) Sequence Read Archive (SRA) database. Information for all metadata used is provided in Tables. All scripts for computational analysis and corresponding raw data are available at https://github.com/xsnhxu/Data-of-pesticides.

ORCID

Baolan Hu https://orcid.org/0000-0003-4700-3498
Haifeng Qian https://orcid.org/0000-0003-0807-9991

REFERENCES

Ahneman, D.T., Estrada, J.G., Lin, S., Dreher, S.D. & Doyle, A.G. (2018) Predicting reaction performance in C-N cross-coupling using machine learning. *Science*, 360(6389), 186–190.

Assenov, Y., Ramírez, F., Schelhorn, S.-E., Lengauer, T. & Albrecht, M. (2007) Computing topological parameters of biological networks. *Bioinformatics*, 24, 282–284.

Ban, Z., Yuan, P., Yu, F.B., Peng, T., Zhou, Q.X. & Hu, X.G. (2020) Machine learning predicts the functional composition of the protein corona and the cellular recognition of nanoparticles. *Proceedings of the National Academy of Sciences of the United States of America*, 117(19), 10492–10499.

Banerjee, S., Kirkby, C.A., Schmutter, D., Bissett, A., Kirkegaard, J. A. & Richardson, A.E. (2016) Network analysis reveals functional redundancy and keystone taxa amongst bacterial and fungal communities during organic matter decomposition in an arable soil. *Soil Biology and Biochemistry*, 97, 188–198.

Berg, K.A., Lyra, C., Sivonen, K., Paulin, L., Suomalainen, S., Tuomi, P. et al. (2009) High diversity of cultivable heterotrophic bacteria in association with cyanobacterial water blooms. *The ISME Journal*, 3(3), 314–325.

Bhuiyan, M.A., Rahman, M., Rahman, M. & Hasan, M.A. (2011) Metagenomic analysis of a permafrost microbial community reveals a rapid response to thaw. *Nature*, 480(7377), 368–371.

Blondel, V.D., Guillaume, J.L., Lambiotte, R. & Lefebvre, E. (2008) Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 10, P10008.

Bolyen, E., Rideout, J.R., Dillon, M.R., Bokulich, N.A., Abnet, C.C., Al-Ghalith, G.A. et al. (2019) Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nature Biotechnology*, 37(9), 852–857.

Buda, M., Maki, A. & Mazurowski, M.A. (2018) A systematic study of the class imbalance problem in convolutional neural networks. *Neural Networks*, 106, 249–259.

Callahan, B.J., McMurdie, P.J., Rosen, M.J., Han, A.W., Johnson, A. J.A. & Holmes, S.P. (2016) DADA2: high-resolution sample inference from Illumina amplicon data. *Nature Methods*, 13(7), 581–583.

Camacho, D.M., Collins, K.M., Powers, R.K., Costello, J.C. & Collins, J.J. (2018) Next-generation machine learning for biological networks. *Cell*, 173, 1581–1592.

Caporaso, J.G., Lauber, C.L., Walters, W.A., Berg-Lyons, D., Huntley, J., Fierer, N. et al. (2012) Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *The ISME Journal*, 6(8), 1621–1624.

Chaparro, J.M., Sheflin, A.M., Manter, D.K. & Vivanco, J.M. (2012) Manipulating the soil microbiome to increase soil health and plant fertility. *Biology and Fertility of Soils*, 48(5), 489–499.

Chen, X.Z., Doerge, R.W. & Heyse, J.F. (2018) Multiple testing with discrete data: proportion of true null hypotheses and two adaptive FDR procedures. *Biometrical Journal*, 60(4), 761–779.

Chun, S.J., Cui, Y.S., Lee, C.S., Cho, A.R., Baek, K., Choi, A. et al. (2019) Characterization of distinct cyanoHABs-related modules in microbial recurrent association network. *Frontiers in Microbiology*, 10, 1637.

Coyte, K.Z., Schluter, J. & Foster, K.R. (2015) The ecology of the microbiome: networks, competition, and stability. *Science*, 350(6261), 663–666.

Csardi, G. & Nepusz, T. (2006) The igraph software package for complex network research. *Interjournal Complex Systems*, 1695, 1–9.

Davison, J., Moora, M., Opik, M., Adholeya, A., Ainsaar, L., Ba, A. et al. (2015) Global assessment of arbuscular mycorrhizal fungus diversity reveals very low endemism. *Science*, 349(6251), 970–973.

De Vries, F.T., Griffiths, R.I., Bailey, M., Craig, H., Girlanda, M., Gweon, H.S. et al. (2021) Soil bacterial networks are less stable under drought than fungal networks. *Nature Communications*, 9(1), 3033.

de Zelicourt, A., Al-Yousif, M. & Hirt, H. (2013) Rhizosphere microbes as essential partners for plant stress tolerance. *Molecular Plant*, 6(2), 242–245.

Delmont, T.O., Prestat, E., Keegan, K.P., Faubladier, M., Robe, P., Clark, I.M. et al. (2012) Structure, fluctuation and magnitude of a natural grassland soil metagenome. *The ISME Journal*, 6(9), 1677–1687.

Deng, Y., Debognies, A., Zhang, Q., Zhang, Z.Y., Zhou, Z.G., Zhang, J.F. et al. (2022) Effects of ofloxacin on the structure and function of freshwater microbial communities. *Aquatic Toxicology*, 244, 106084.

Didier, G., Valdeolivas, A. & Baudot, A. (2018) Identifying communities from multiplex biological networks by randomized optimization of modularity. *F1000Research*, 7,1042.

Duke, S.O. (2018) Interaction of chemical pesticides and their formulation ingredients with microbes associated with plants and plant pests. *Journal of Agricultural and Food Chemistry*, 66(29), 7553–7561.

Feng, J.Y., Xu, Y., Ma, B., Tang, C.X., Brookes, P.C., He, Y. et al. (2019) Assembly of root-associated microbiomes of typical rice cultivars in response to lindane pollution. *Environment International*, 31, 104975.

Gentile, C.L. & Weir, T.L. (2018) The gut microbiota at the intersection of diet and human health. *Science*, 362(6416), 776–780.

Hartman, K., van der Heijden, G.A., Wittwer, R.A., Banerjee, S., Walser, J.C., Schlaeppi, K. et al. (2018) Cropping practices manipulate abundance patterns of root and soil microbiome members paving the way to smart farming. *Microbiome*, 6(1), 14.

Hernandez, D.J., David, A.S., Menges, E.S., Searcy, C.A. & Afkhami, M.E. (2021) Environmental stress destabilizes microbial networks. *ISME Journal*, 15(6), 1722–1734.

Howe, A.C., Jansson, J.K., Malfatti, S.A., Tringe, S.G., Tiedje, J.M. & Brown, C.T. (2014) Tackling soil diversity with the assembly of large, complex metagenomes. *Proceedings of the National Academy of Sciences of the United States of America*, 111(16), 4904–4909.

Inderbitzin, P., Ward, J., Barbella, A., Solares, N., Izyumin, D., Burman, P. et al. (2017) Soil microbiomes associated with verticillium wilt suppressive broccoli and chitin amendments are enriched with potential biocontrol agents. *Phytopathology*, 108(1), 31–43.

Jacomy, M., Venturini, T., Heymann, S. & Bastian, M. (2014) ForceAtlas2, a continuous graph layout algorithm for handy network visualization designed for the Gephi software. *PLoS One*, 29(6), e98679.

Jordan, M.I. & Mitchell, T.M. (2015) Machine learning: trends, perspectives, and prospects. *Science*, 349(6245), 255–260.

Jiang, G.F., Wang, N.Q., Zhang, Y.Y., Wang, Z., Yu, J.B., Zhang, Y. et al. (2021) The relative importance of soil moisture in predicting bacterial wilt disease occurrence. *Soil Ecology Letters*, 2021, 1–11.

Lazcano, C., Boyd, E., Holmes, G., Hewavitharana, S., Pasulka, A. & Ivors, K. (2021) The rhizosphere microbiome plays a role in the resistance to soil-borne pathogens and nutrient uptake of strawberry cultivars under field conditions. *Scientific Reports*, 11(1), 3188.

Lehmann, J., Bossio, D.A., Kögel-Knabner, I. & Rillig, M.C. (2020) The concept and future prospects of soil health. *Nature Reviews Earth & Environment*, 1(10), 544–553.

Ley, R.E., Peterson, D.A. & Gordon, J.I. (2006) Ecological and evolutionary forces shaping microbial diversity in the human intestine. *Cell*, 124(4), 837–848.

Li, M.H., Hu, J.L. & Lin, X.G. (2022) The roles and performance of arbuscular mycorrhizal fungi in intercropping systems. *Soil Ecology Letters*. https://doi.org/10.1007/s42832-021-0107-1

Liu, X., Chen, K., Chuang, S., Xu, X. & Jiang, J. (2019) Shift in bacterial vommunity structure drives different atrazine-degrading efficiencies. *Frontiers in Microbiology*, 10, 88.

Martin, M. (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet Journal*, 17(1), 10–12.

Mcdonald, D., Price, M.N., Goodrich, J., Nawrocki, E.P., Desantis, T. Z., Probst, A. et al. (2012) An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. *The ISME Journal*, 6(3), 610–618.

Molder, D.T., Poncheewin, W., Schaap, P.J. & Koehorst, J.J. (2021) Machine learning approaches to predict the plant-associated phenotype of Xanthomonas strains. *BMC Genomics*, 22(1), 848.

Monard, C., Vandenkoornhuyse, P., Le Bot, B. & Binet, F. (2011) Relationship between bacterial diversity and function under biotic control: the soil pesticide degraders as a case study. *The ISME Journal*, 5(6), 1048–1056.

Naylor, D., Sadler, N., Bhattacharjee, A., Graham, E.B., Anderton, C. R., McClure, R. et al. (2020) Soil microbiomes under climate change and implications for carbon cycling. *Annual Review of Environment and Resources*, 45, 29–59.

Okie, J.G., Horn, D., Storch, D., Barrett, J.E., Gooseff, M.N., Kopsova, L. et al. (2015) Niche and metabolic principles explain patterns of diversity and distribution: theory and a case study with soil bacterial communities. *Proceedings of the Royal Society B*, 282(1809), 20142630.

Pawluczyk, M., Weiss, J., Links, M.G., Aranguren, M.E., Wilkinson, M.D. & Egea-Cortines, M. (2015) Quantitative evaluation of bias in PCR amplification and next-generation sequencing derived from metabarcoding samples. *Analytical and Bioanalytical Chemistry*, 407(7), 1841–1848.

Pimentel, D. & Burgess, M. (2013) Soil erosion threatens food production. *Agriculture*, 3(3), 443–463.

Ramirez, K.S., Knight, C.G., de Hollander, M., Brearley, F.Q., Constantinides, B., Cotton, A. et al. (2017) Detecting macroecological patterns in bacterial communities across independent studies of global soils. *Nature Microbiology*, 3(2), 189–196.

Rooney, N., McCann, K., Gellner, G. & Moore, J.C. (2006) Structural asymmetry and the stability of diverse food webs. *Nature*, 442(7100), 265–269.

Rumschlag, S.L., Mahon, M.B., Hoverman, J.T., Raffel, T.R., Carrick, H.J., Hudson, P.J. et al. (2020) Consistent effects of pesticides on community structure and ecosystem function in freshwater systems. *Nature Communications*, 11(1), 6333.

Santos, A.M. & Branco, M. (2012) The quality of name-based species records in databases. *Trends in Ecology & Evolution*, 27(1), 6–7.

Shi, S.J., Nuccio, E.E., Shi, Z.J., He, Z.L., Zhou, J.Z. & Firestone, M. K. (2016) The interconnected rhizosphere: high network complexity dominates rhizosphere assemblages. *Ecology Letters*, 19(8), 926–936.

Singh, R., Kumar, T., Verma, P., Pal Singh, B. & Raghubanshi, A.S. (2022) Compatible package-based agriculture systems: an urgent need for agro-ecological balance and climate change adaptation. *Soil Ecology Letters*, 4(3), 1–26. https://doi.org/10.1007/s42832-021-0087-1

Tedersoo, L., Ramirez, K.S., Nilsson, R.H., Kaljuvee, A., Koljalg, U. & Abarenkov, K. (2015) Standardizing metadata and taxonomic identification in metabarcoding studies. *Gigascience.*, 4(1), 34.

Thabtah, F., Hammoud, S., Kamalov, F. & Gonsalves, A. (2020) Data imbalance in classification: Experimental evaluation. *Information Sciences*, 513, 429–441.

Toju, H., Peay, K.G., Yamamichi, M., Narisawa, K., Hiruma, K., Naito, K. et al. (2018) Core microbiomes for sustainable agroecosystems. *Nature Plants*, 4(9), 247–257.

Tringe, S.G., Mering, C.V., Kobayashi, A., Salamov, A.A., Chen, K., Chang, H.W. et al. (2005) Comparative metagenomics of microbial communities. *Science*, 308(5721), 554–557.

Vörösmarty, C.J., Green, P., Salisbury, J. & Lammers, R.B. (2000) Global water resources: vulnerability from climate change and population growth. *Science*, 289(5477), 284–288.

Vujkovic-Cvijin, I., Sklar, J., Jiang, L.J., Natarajan, L., Knight, R. & Belkaid, Y. (2020) Host variables confound gut microbiota studies of human disease. *Nature*, 587(7834), 448–454.

Wang, Z., Yuan, S., Deng, Z., Wang, Y. & Wang, X. (2020) Evaluating responses of nitrification and denitrification to the co-selective pressure of divalent zinc and tetracycline based on resistance genes changes. *Bioresource Technology*, 314, 123769.

Williams, R.J., Howe, A. & Hofmockel, K.S. (2014) Demonstrating microbial cooccurrence pattern analyses within and between ecosystems. *Frontiers in Microbiology*, 5(1), 358.

Wright, R.J., Langille, M. & Walker, T.R. (2021) Food or just a free ride? A meta-analysis reveals the global diversity of the plastisphere. *The ISME Journal*, 15(3), 789–806.

Xu, N., Zhang, Z., Shen, Y., Zhang, Q., Liu, Z., Yu, Y. et al. (2022) Compare the performance of multiple binary classification models in microbial high-throughput sequencing datasets. *Science Total Environment*, 837, 155807. https://doi.org/10.1016/j.scitotenv.2022.155807

Xu, X.H., Zarecki, R., Medina, S., Ofaim, S., Liu, X.W., Chen, C. et al. (2019) Modeling microbial communities from atrazine contaminated soils promotes the development of biostimulation solutions. *The ISME Journal*, 13(2), 494–508.

Xun, W.B., Li, W., Xiong, W., Ren, Y., Liu, Y.P., Miao, Y.Z. et al. (2019) Diversity-triggered deterministic bacterial assembly constrains community functions. *Nature Communications*, 10, 3833.

Xun, W.B., Liu, Y.P., Li, W., Ren, Y., Xiong, W., Xu, Z.H. et al. (2021) Specialized metabolic functions of keystone taxa sustain soil microbiome stability. *Microbiome*, 9(1), 35.

Yuan, J., Wen, T., Zhang, H., Zhao, M.L., Penton, C.R., Thomashow, L.S. et al. (2020) Predicting disease occurrence with high accuracy based on soil macroecological patterns of Fusarium wilt. *The ISME Journal*, 14(12), 2936–2950.

Yuan, M.M., Guo, X., Wu, L.W., Zhang, Y., Xiao, N.J., Ning, D.L. et al. (2021) Climate warming enhances microbial network complexity and stability. *Nature Climate Change*, 11(4), 343–U100.

Zhang, W. (2018) Global pesticide use: profile, trend, cost/benefit and more. *Proceedings of the International Academy of Ecology and Environmental Sciences*, 8(1), 1–27.

Zhang, Z.Y., Zhang, Q., Wang, T.Z., Xu, N.H., Lu, T., Hong, W.J. et al. (2022a) Assessment of global health risk of antibiotic resistance genes. *Nature Communications*. https://doi.org/10.1038/s41467-022-29283-8

Zhang, Z.Y., Zhang, Q., Cui, H.Z., Li, Y., Xu, N.H., Lu, T. et al. (2022b) Composition identification and functional verification of bacterial community in disease-suppressive soils by machine learning. *Environmental Microbiology*. https://doi.org/10.1111/1462-2920.15902

Zheng, B.Y., Zhao, Q.Q., Feng, L., Zhang, Z.Y., Zhang, Q., Zhang, F. et al. (2022) Regulative effect of imazethapyr on *Arabidopsis thaliana* growth and rhizosphere microbial community through multiple generations of culture. *Plant and Soil*, 473, 625–637. https://doi.org/10.1007/s11104-022-05318-3

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Ke, M., Xu, N., Zhang, Z., Qiu, D., Kang, J., Lu, T. et al. (2022) Development of a machine-learning model to identify the impacts of pesticides characteristics on soil microbial communities from high-throughput sequencing data. *Environmental Microbiology*, 1–13. Available from: https://doi.org/10.1111/1462-2920.16175