University of Nottingham

UK | CHINA | MALAYSIA

MRes Environmental Science and Engineering

A meta-analysis on the impact of long-term fertilization on soil microbial communities

Shule Li

20513674

Thesis submitted to the University of Nottingham for the degree of Master by Research

March 2023

Abstract

Soils are relevant to our human life and the microbial communities that use them as habitats can actively participate in biogeochemical cycles. Fertilizer application, one of the most common agronomic management practices, is diverse and long-term in nature. However, the effects of long-term fertilization with different types of fertilizers on microbial microorganisms in soils are not fully understood. In this study, we collected bulk soil samples based on 16S rRNA sequencing from 103 publications of 10308 long-term fertilization experiments from various locations worldwide and environmental metadata corresponding to each sample. To explore the importance of different environmental variables as well as the interaction effects between variables, we evaluated three tree-based machine learning models, RandomForest, XGBoost, and LightGBM, and used the state-of-the-art interpretation method SHAP to interpret the models, whose hyperparameters were optimized by Bayesian optimization algorithm. Ultimately, 20 randomized experiments showed that soil organic carbon, inorganic fertilizer application amount, and sampling depth were the three most essential predictors of soil microbial Shannon diversity. The local SHAP imputation values revealed the robustness of the importance of soil organic carbon, as its SHAP value increased almost monotonically with its value. Furthermore, SHAP analysis for fertilization treatment duration demonstrated that the soil microbial community had reached a steady state under long-term fertilization. In addition, the interaction between the use of N fertilizer and soil organic carbon and soil pH, respectively, was revealed by SHAP interaction analysis. This work demonstrates that the tree-based machine learning algorithm combined with the interpretable machine learning algorithm SHAP has the potential to predict soil microbial Shannon diversity and to analyze global and local attribution. This is critical for capturing the level of environmental factors and directing agricultural operations in a way that preserves soil stability.

KEYWORDS: long-term fertilization; soil; 16S rRNA sequencing; microbial diversity; machine learning; SHAP.

Contents

Chapter 1 Introduction

More than one-third of the Earth's land surface is covered by the agroecosystem, which provides a wide range of services to ecological and anthropogenic networks (Tilman, Cassman et al. 2002, Smith, Martino et al. 2008). To supply the global food production needs of the growing population, crop productivity has been enhanced by the increasing use of organic fertilization and synthetic chemical fertilization in agricultural practice (Hartmann and Six 2023). However, the overuse and misuse of fertilization may bring negative impacts on the ecosystem, including soil acidification, soil degradation, eutrophication of surface water, and greenhouse gas emission (Laborde, Mamun et al. 2021).

The soil microorganisms play vital roles in biogeochemical cycling, bioremediation, climate regulation, and mediating plant growth (Bender, Wagg et al. 2016, Hartmann and Six 2023). The diversity, composition, and stability of the soil microbial community are vital to maintaining soil ecosystem sustainability and function (Dan, Sadler et al. 2020). Soil bacterial communities could be disturbed easily and interact with the surrounding environment sensitively in response to environmental change.

Long-term nutrient addition can adversely affect soil physical structure and chemical properties, which in turn affect soil microbial communities. There were extensive studies that focused on the biochemical properties and enzyme activities of soil under different fertilization strategies, whereas the studies focusing on the interaction of soil microbial communities in response to nutrient addition are still fragmented and inconsistent. For example, some studies found that long-term chemical fertilization reduced soil microbial diversity due to the decrease in soil pH, while Hou et al. showed that the microbial biomass carbon and microbial alpha diversity were higher at a proper nitrogen fertilization addition rate than the treatments without fertilization (Hou, Ren et al. 2023). In terms of organic fertilization addition, some research indicated that manure addition increased soil bacterial diversity, while Guo et

al. found manure addition had no significant effect on soil bacterial diversity (Guo, Wan et al. 2020).

There were meta-analyses that integrated studies under different fertilization managements to explore the general trends in soil biochemical properties change and the main effect factors. Recently, some studies demonstrated that soil biological data such as alpha diversity might be more precise to predict soil health status under fertilization, and several meta-analyses focused on the change of the dominant taxa and the diversity index were conducted. A global meta-analysis integrating 105 papers showed that the application of organic fertilizers significantly elevated soil organic carbon (SOC), total nitrogen (TN), and microbial biomass carbon (MBC) contents and significantly increased soil bacterial alpha diversity compared to the non-fertilized group (Dang, Li et al. 2022). Meanwhile, they found inorganic fertilizer application decreased soil pH and thus had a negative effect on soil microbial alpha diversity and community composition, such as the significant reduction in the abundance of taxa such as Verrucomicrobia, Planctomycetes, and Nitrospirae. However, these studies mainly revealed the regulation by the information of papers which were already done in bioinformatic analysis, for example, the alpha diversity indexes were calculated through different rarefaction methods among studies, as it might bring some disturbing in the results.

Thus, we used a downstream analysis framework started with the raw sequencing data, attempting to infer the composition of bacterial communities in the bulk soil under different fertilization practices, as well as to determine the bacterial community interactions, dynamics, and threshold of the living condition under a broad list of environmental metadata, and to identify the most important and relevant environmental variables to construct a robust model to predict the microbial community diversity. It will provide general principles for conducting proper fertilization practices in the agricultural ecosystem to maintain both soil fertility and productivity by adjusting the soil microbial community.

Here, we collected 16S rRNA amplicon sequencing raw data of bulk soil under different fertilization managements from various gene banks, and processed and

merged these data through bioinformatic methods. We also collected a wide range of environmental factors for each sample, considering to fully understand the relationship between soil microbial community and surrounding habitat. (1) We hypothesized that there are different effects of microbial community composition, diversity, and function under different fertilization. After long-term nutrient addition, the co-occurrence network patterns would evolve separately, with different keystone taxa merging in the microbial community. To this end, we calculated different matrixes and determined the keystone taxa of the microbial group under different fertilization practices. (2) Since considering microbial diversity is explicitly shaped by various environmental factors, we hypothesized that there would be some main effects to determine the soil microbial diversity under long-term fertilization, and it is also necessary to explore the interactive effects among environmental variables. For example, the influence of different nutrition addition rates on soil microbial diversity would be entirely different under various pH conditions. To this end, we constructed several optimized machine learning models to focus on the effect of different environmental factors on soil microbial alpha diversity, and we used state-of-art interpretable artificial methods to figure out the regulation of the main effect and the interactive effect. After figuring out the various regulation way driven by various environmental conditions in relation to the diversity index, we could find a tipping point of maintaining the diversity of the soil bacterial community. Overall, this work provided an overview of the microbial community traits under long-term fertilization, gained an understanding of how the environmental variables would influence the soil bacteria directly and interactively, and offered a robust model to predict the bacteria alpha diversity index under various fertilization management, which would identify the promising prospect.

Chapter 2 Literature review

2.1 Soil microbiome

2.1.1 The vital significance of soil microbiome within the ecosystem

Microorganisms are an essential component of biodiversity and constitute a complex ecosystem with their environment (Dan, Sadler et al. 2020). A microbiome is a collection of all microorganisms and their genetic information in a particular environment or ecosystem, which includes the interactions of microorganisms with their environment and hosts (Hartmann and Six 2023). Soil, as a link between the atmosphere, hydrosphere, lithosphere, and biosphere, contains organisms, minerals, organic matter, water, and air in the soil sphere. And the soil is essential to the survival of plant systems on Earth, as well as animal and human activity. Because of its unique physical structure and complicated chemical composition, the soil is an ideal habitat for microbial growth and reproduction. According to statistics, the total number of microorganisms in soil worldwide exceeds 10^{30}, and each gram of soil contains hundreds of millions of microbial cells (Bardgett and van der Putten 2014). Being a key link between the above- and below-ground parts of the soil ecosystem, the soil microbiome also plays an irreplaceable and critical role in soil nutrient cycling, energy flow, ecosystem structure maintenance, and ecosystem function regulation (Doran and Zeiss 2000, Bardgett and van der Putten 2014). The soil microbiome is a significant driver of elemental biogeochemical cycling. As a result, the soil microbiome is a crucial resource for maintaining normal human productive life and the earth's ecosystem equilibrium.

In terms of the decomposition and transformation of soil materials, soil microorganisms ensure the recycling of many macronutrients by participating in complex and diverse metabolic processes such as organic matter decomposition and biosynthesis, including carbon fixation, conversion of monosaccharide polysaccharides,

nitrogen fixation, and nitrification (Amundson, Berhe et al. 2015, Banerjee and van der Heijden 2023). These processes provide nutrients for plant growth and make an important contribution to the formation of soil aggregates and the accumulation of fertility. In carbon cycling, for example, bacteria mainly mineralize and break down small molecule carbohydrates, whereas fungi primarily break down refractory carbon and destroy plant leftovers. In terms of nitrogen cycling, soil microorganisms also influence nitrogen transformation efficiency (Vitousek and Howarth 1991). They regulate nitrogen fixation and loss at various phases of nitrogen transformation by using their metabolism or releasing enzymes that impact nutrient utilization efficiency. When soil microbes change nutrients in the soil, they are also engaged in the process of nutrient acquisition by plants. Some legumes, for instance, have symbiotic nitrogen-fixing bacteria in their rhizosphere that enhance nitrogen uptake and utilization. In addition, certain mycorrhizal fungi and root-promoting beneficial bacteria also improve plant uptake of nutritional components in the soil. At the same time, soil microorganisms can play a purification role for organic pollutants entering the soil from exogenous sources, and specific microbes can transform or completely degrade mineralized organic contaminants via metabolic or co-metabolic processes (Amundson, Berhe et al. 2015). Microbial metabolism can change the toxicity and efficacy of metal and metalloid pollutants under altered redox conditions in the soil environment. In conclusion, soil microorganisms are frequently involved in the majority of biophysical-chemical processes in soils, including the cyclic transformation of nutrients on the one hand and the transport and transformation of pollutants in soils on the other.

More and more studies show that soil microorganisms are to some extent related to other organisms such as plants and humans. This "One Health" status is closely related to other ecosystem services and functions such as plant diversity, decomposition of apoplastic plants, and global climate change (Banerjee and van der Heijden 2023).

Some genetic influences in soil microbial systems can ensure the relative stability of composition and structure. They, on the other hand, preserve community diversity through functional redundancy and variation, which can be classified as community structural diversity, species diversity, genetic diversity, and functional diversity. This

resistance and resilience of the soil microbial system allow the soil microbial community to respond and adjust to external dynamic changes, thus maintaining the function of the soil ecosystem (Saleem, Hu et al. 2019). When plant systems are exposed to plant pathogens, for example, inter-rooted beneficial bacteria in the soil can boost inter-rooted immunity and thereby prevent plant pathogen proliferation and dissemination. In this way, the effect of improving plant resistance, reducing plant morbidity, and reducing the risk of transmission of some human pathogenic bacteria to humans through plants is achieved (Hannula, Ma et al. 2020). Simultaneously, the dynamics of the soil microbiome act as a regulator of global climate change. The soil system develops a sequence of responses in response to changes in the external environment (Jansson and Hofmockel 2020). A temperature rise, for example, modifies microbial activity and hence accelerates the breakdown of organic carbon by soil microbes (Romero-Olivares, Allison et al. 2017). On the other hand, it indirectly affects soil carbon balance by shifting plant system functioning, such as the release of apoplastic matter, and hence contributes to CO_2 emissions. Soil microorganisms can also produce secondary metabolically active compounds like antibiotics, which are a significant source of biological resources.

Therefore, the maintenance of soil microbial community health, as reflected in soil microbial biomass, activity, and diversity, is critical for ecosystem stability, environmental protection, and rational development at regional and global scales.

2.1.2 Soil microbiome ecology research methods

Soil has a huge number of microorganisms of various taxa and functions. 1g of soil includes approximately 10^9 microbial cells and more than 10^6 microbial species, according to estimates (Bardgett and van der Putten 2014). The numerous bacteria, fungi, actinomycetes, and protozoa that live in soil are essential components of a healthy soil ecosystem. Furthermore, they serve an important role in guaranteeing the buildup and mineralization of soil organic matter, resulting in healthy plant growth, as

well as drivers of biogeochemical cycling of critical elements such as carbon, nitrogen, and phosphorus (Calderón, Spor et al. 2017). The structure and diversity of soil microbial communities, however, vary greatly due to factors such as soil-forming parent material, the great degree of geospatial variability in geographic location, the differentiation of climatic conditions such as rainfall and temperature, the type of vegetation cover, and the influence of anthropogenic activities. Soil microbial ecology has been increasingly investigated by domestic and international research scholars to examine the interaction and patterns of soil microorganisms and soil environment (Jansson and Hofmockel 2018).

Traditional methods for determining soil microbial diversity include direct assay, plate scribing dilution pure culture method, and physiological and biochemical assay. The direct assay methodology employs light and electron microscopes to directly observe and count microorganisms in a sample (Thompson, Sanders et al. 2017). The plate scribing dilution pure culture method refers to the method of isolating or screening microorganisms by inoculating a plate with a suitable dilution concentration by coating it with a gradient dilution after the suspension containing single-celled microorganisms is well mixed and homogenized. Physiological and biochemical assay techniques include the Biolog method, phospholipid fatty acid (PLFA) assay, and ergocalciferol method, etc. The Biolog method is an approach for exploring microbial diversity based on microorganisms' ability to utilize carbon sources differently. It enables the identification of pure microbe species as well as the comparative examination of overall differences in diverse environmental microbial communities. The phospholipid fatty acid (PLFA) analysis method is based on the fact that the composition and content of phospholipid fatty acids on living cell membranes are relatively stable and have high biological specificity, allowing quantification of biomass and ecological structure of large taxa in soil ecosystems (Frostegård, Tunlid et al. 1991). Ergocalciferol is a significant component of fungal and protozoan cell membranes. And the ergocalciferol method could estimate fungal biomass by detecting ergocalciferol levels. But this method cannot detect diversity.

The first generation of sequencing technology, Sanger sequencing, emerged in the

late twentieth and early twenty-first centuries, allowing for in-depth analysis of microbial communities directly at the molecular level without relying on traditional methods such as complex and laborious pure culture and determination of microbial markers (Simon, Lalonde et al. 1992). The usage of dideoxy triphosphate nucleotides with fluorescent labeling groups, which can be utilized to abort DNA strand expansions and collected by imaging equipment, is at the core of first-generation sequencing technology. However, first-generation sequencing is confined to large-scale utilization at the experimental level because of high sequencing costs and low throughput. Second-generation high-throughput sequencing technology, also known as "next-generation sequencing technology," enables rapid sequencing while synthesizing, lowering sequencing costs and increasing sequencing speed while maintaining high accuracy, as demonstrated by Roche's 454 sequencing system, Illumina's Hiseq sequencing system, and ABI's Solid sequencing system (Lane, Pace et al. 1985, Sanchez-Cid, Tignat-Perrier et al. 2022). The composition of microorganisms in environmental samples and their potential ecological functions can be studied using high-throughput sequencing techniques. Among the more widely utilized technologies in soil ecology are amplicon sequencing of marker genes and metagenomics. The method of PCR amplification of DNA collected from environmental materials using primers for specific genes and analysis of the amplification products is known as amplicon sequencing of marker genes. The marker genes include 16S rRNA, 18S rRNA, and functional genes relevant to the carbon and nitrogen cycles, etc (Callahan, Wong et al. 2019). The process of extracting all DNA from environmental samples without PCR amplification of target fragments and randomly interrupting them into fragments for sequencing, which can reflect all genetic information of the entire community in environmental samples and can deeply explore the potential functions of microbial communities, is known as metagenomics. This study is mainly based on the results of 16S rRNA sequencing of bacteria, which is described in detail here. 16S rRNA is a segment of DNA, which is the gene used by the bacterial genome to encode the small subunit of the ribosome. S is the sedimentation coefficient of ribosomes; for example, the prokaryotic ribosome is the 70S and is made up of two subunits, 50S, and 30S, with 16S being a portion of the

small subunit. The 16S gene has the character of a "molecular clock" during the evolution and proliferation of bacteria, in that it has certain structural and functional conservation. Specifically, this segment includes 9 variable regions and 10 conserved regions. The conserved regions reflect the kinship between species and can be used to design primers with generality to complete the amplification of variable regions, and the gene sequences of variable regions can reflect the differences between species, and the species can be specifically identified by amplifying, and sequencing the genes of variable regions.

Currently, 16S rRNA-based DNA sequencing of environmental samples has been applied to almost all natural environments, including common agricultural soils, natural water bodies, sewage and sludge from wastewater treatment plants, permafrost layers in polar regions, marine sediments, and plant inter-root and inter-leaf.

In soil microbial ecology, the 16S rRNA sequencing of environmental samples allows for diversity analysis of soil microbial communities. Diversity analysis includes calculating and analyzing the microbial composition of different samples, calculating α diversity to compare the diversity of species within samples, calculating β diversity to compare the diversity of community structure between samples, and exploring the diversity of community functions by functional annotation and prediction.

Data mining of 16S rRNA sequencing results can efficiently elucidate the properties of environmental samples and the associations between samples by selecting appropriate statistical tests, data processing methods, and software. For example, STAMP software (Parks, Tyson et al. 2014) can perform statistical tests for differences in species abundance and functional differences between groups and visualize the results; LefSe analysis (Segata, Izard et al. 2011) can find biomarkers for differences in abundance between different treatments; Co-occurrence analysis is used to compare the interactions between different samples and different bacteria by constructing a correlation coefficient matrix; Mantel test finds the association between samples, microbial communities, and environmental factors by comparing the correlation between the environmental factor matrix and the microbial community matrix, etc. The Mantel test is used to find the association between samples, microbial communities,

and environmental factors by comparing the correlation between environmental factors and microbial community matrices.

Various functional annotation software, such as PICRUSt (Douglas, Maffei et al. 2020), Tax4Fun (Wemheuer, Taylor et al. 2020), and FAPROTAX (Louca, Parfrey et al. 2016), is constantly being upgraded. The prediction of community function using 16S rRNA gene sequences, however, is not a replacement for sequencing and analysis of specific functional genes and metagenomics techniques, because the presence of a certain class of microorganisms does not necessarily imply that they have specific functional genes or perform related functions. However, the function gene abundance results predicted by 16S rRNA sequences can be used as a basis for subsequent experimental design of metagenomics, metatranscriptomics, and so on.

2.2 Impact of fertilizer application on soil microbiome

The soil microbiome is a significant indication of soil quality. Not only does the microbial community play a crucial role in maintaining soil ecosystem balance, but soil microbial diversity is also directly related to soil ecosystem stability and the ability to recover from disturbance stress. As a popular agronomic management practice, fertilizer application has direct or indirect impacts on soil physicochemical parameters, as well as on the biomass, composition, and function of soil microbial communities (Thiele-Bruhn, Bloem et al. 2012). Various fertilizer management approaches, as exogenous inputs of varying nature, have varying effects on soil ecosystems. The effects of inorganic fertilizer application, organic fertilizer application, and organic fertilizer mix on soil microorganisms are described below.

2.2.1 Impact of inorganic fertilizer application on soil microbiome

Inorganic fertilizers are chemically manufactured and processed fertilizers that do not contain organic matter. They include nitrogen, phosphorus, potash, and compound

fertilizers. Numerous research studies have shown that the use of inorganic fertilizers increases crop yield. However, the low utilization rate of inorganic fertilizers leads to their overuse by humanity, allowing a huge amount of inorganic fertilizer into the soil environment. Many studies are looking into how additional inorganic fertilizer stimulation in the soil environment may affect the soil microbial community. Several studies have shown that applying inorganic fertilizer to the soil enhances the organic matter content and promotes the growth and metabolism of the soil microbial community. Geisseler et al., for example, found that adding inorganic fertilizer increased microbial biomass carbon (MBC) in rice soils by 26% and soil organic carbon content by 13% in a meta-analysis of rice soils (Geisseler, Linquist et al. 2017). A global study of long-term inorganic fertilizer application revealed through metagenomic sequencing that nitrogen fertilization boosted bacterial abundance while also improving denitrification and nitrate reduction by bacteria (Li, Tremblay et al. 2020). Some studies, however, have shown that the use of inorganic fertilizers might lower the biomass of soil microbes, such as a meta-analysis by Jian et al., which found that the application of N fertilizer increased soil organic carbon and total N by 7.6% and 15.3%, respectively, but reduced MBC by 9.5% (Jian, Li et al. 2016). Liu et al. discovered that the addition of inorganic fertilizer decreased MBC by 20%, which was followed by an 8% drop in soil microbial respiration (Liu and Greaver 2010). A meta-analysis by Yang et al. of soil microbial diversity found that nitrogen fertilizer application resulted in a drop in soil microbial diversity, which was found to be dominated by a fall in pH (Yang, Cheng et al. 2020). Dang et al.'s global meta-analysis also revealed that inorganic fertilizer application reduced soil pH and had a detrimental influence on soil microbial diversity and community composition, such as a significant decrease in soil Verrucomicrobia, Planctomycetes, Nitrospirae, and other taxa abundance (Dang, Li et al. 2022).

In particular, the impacts of inorganic fertilizers on soil microorganisms can be examined from the following aspects. The first is the influence of added inorganic fertilizer on soil structure, such as the application of massive quantities of potassium fertilizer, which causes the exchange and soaking of salt-based ions such as Ca^{2+} and

Mg^{2+} in the soil, subsequently causing the soil to become slabbed and less permeable. Hence, the aforementioned processes influence the distribution and survival of soil microbial communities in soil aggregates. According to Shun et al., the addition of inorganic fertilizer reduces the versatility of soil agglomerates and alters soil enzyme activity (Han, Delgado-Baquerizo et al. 2021). The second is the effect of inorganic fertilizer addition on soil physicochemical properties. The addition of inorganic fertilizer alters the pH and organic matter content of the soil, influencing the makeup and function of the microbial community indirectly. The effect of inorganic fertilizer application varies depending on the soil type because of variations in pH values. Hou et al., for example, demonstrated that at an applied N fertilizer rate of 200 kg N $ha^{-1}yr^{-1}$, soil pH was significantly lower while MBC was significantly higher in black soil (Hou, Ren et al. 2023). Furthermore, microbial alpha diversity was higher at lower planting densities than in the treatment without N fertilization. However, when 400 kg N $ha^{-1}yr^{-1}$ N fertilizer was applied, soil microbial diversity was much lower than in the control and the treatment with moderate N fertilizer application. According to Hui et al., N fertilizer application reduced microbial diversity in black soils by 13.2% to 48.5%, with pH being the most important determinant (Wang, Xu et al. 2018). Meanwhile, the fall in soil pH caused by long-term N fertilizer promoted the proliferation of acidophilic bacteria. Shi et al. discovered that long-term inorganic fertilizer application to red soils substantially lowered soil microbial biomass and functional bacterial activity compared to unfertilized controls (Shi, Zhao et al. 2021). According to Xun et al., pH correction with moderate amounts of lime applied to acidic red soils could minimize the negative impacts of long-term inorganic fertilizer treatment and increase network stability (Xun, Huang et al. 2017).

2.2.2 Impact of organic fertilizer application on soil microbiome

Organic fertilizers are fertilizers composed of organic substances, including manure and urine from human life; animal manure and manure compost from livestock and poultry

farming green manure, cake manure, and biogas fertilizer from some agricultural production. Organic fertilizers provide nutrients such as nitrogen, phosphorus, potassium, amino acids, and trace elements, which are essential in agricultural production. They can greatly improve soil structure by stimulating the development of soil aggregates and altering the distribution and activity of microbes, in addition to providing macronutrients or micronutrients. Shun et al., for example, demonstrated that applying organic fertilizer enhanced the multifunctionality of soil aggregates. when compared to treatments without and with inorganic fertilizer, the activity of C, N, P, and S cycling functional enzymes in soil micro agglomerates with particle size less than 53 um was dramatically increased (Han, Delgado-Baquerizo et al. 2021). Tian's research found that long-term organic fertilizer application enhanced the fraction of soil macro agglomerates and the nutritional content of C, N, and P in soil macro agglomerates. Also, long-term organic fertilizer application enhanced the biomass of soil bacteria in each particle size agglomeration. Meanwhile, the long-term application of organic fertilizer enhanced crop yield compared to the long-term application of inorganic fertilizer (Tian, Zhu et al. 2022).

Simultaneously, a number of long-term localization investigations have revealed that applying organic fertilizer greatly enhances soil MBC, soil microbial respiration, and enzyme activity, supporting the creation of more stable microbial contact networks. A global meta-analysis integrating 105 papers showed that organic fertilizer application significantly increased SOC, TN, and MBC in soil (Dang, Li et al. 2022). Moreover, when compared to the control group that did not receive fertilizer, the administration of organic fertilizer dramatically enhanced the alpha diversity of soil bacteria. On the one hand, long-term field experiments conducted in black soil in northeastern China by Xiaojing et al. revealed that the addition of manure as organic fertilizer directly enhanced the available phosphorus (AP) concentration in the soil (Hu, Gu et al. 2023). Organic fertilizer, on the other hand, altered the number of relevant functional genes involved in the soil phosphorus cycle, expedited microbial phosphorus transformation, and decreased microbial phosphorus uptake as well as enhanced plant phosphorus bioavailability. Semenov's research also revealed that exogenous microbes from

organic fertilizer application die out quickly, whereas other native taxa proliferate over time as a result of changes in organic matter, nutrition, and physicochemical qualities caused by organic fertilizer application (Semenov, Krasnov et al. 2021). Andera et al. collected soil samples for analysis after applying fertilizer to Italian farms where cow manure, chicken manure, and pig manure were applied individually (Laconi, Mughini-Gras et al. 2021). The results reveal that bacteria from manure had no effect on the soil's native microbiome thirty days after application and that bacteria from manure did not survive in the soil. It can be seen that the effect of organic fertilizers on soil microorganisms needs to be explored in terms of the nature of the organic fertilizer itself and the duration of application. Various organic fertilizers introduce different types of carbon sources and carbon-to-nitrogen ratios into the soil, resulting in variations in the abundance of microbes with various functions. As when plant straw is added, the stimulation of high cellulose exogenous sources can make cellulolytic microorganisms more competitive. A 47-year field trial showed that continuous application of green manure increased bacterial biomass in the soil compared to treatments without manure and that the ratio of fungi to bacteria was higher in soils treated with green manure compared to treatments with manure application. The abundance of microorganisms with a predominantly polymer-based carbon source utilization increased when manure was added (Elfstrand, Hedlund et al. 2007). However, the degradation rates of different manures differed, as did the impacts on soil microbes. Peng et al. discovered that the alpha diversity of bacteria in soil samples treated with fresh swine manure was substantially larger than that of the unfertilized treatment group, whereas the alpha diversity of bacteria in soil samples treated with chicken manure was not significantly different from that of the control group(Li, Wu et al. 2020). However, chicken manure application considerably affected the community structure of soil microorganisms, with Plancomycetaceae and Thauera having significantly higher relative abundance than the other treatment groups. In terms of application timing, long-term application of organic fertilizers is generally considered to improve soil structure and increase the accumulation of organic matter in the soil. For the short-term application of organic fertilizers, some studies have shown that application of organic

fertilizers during the crop growing season causes soil microorganisms to respond in a short-term time frame, as reflected by an increase in soil microbial load and activity. However, the response varies depending on the type of organic fertilizer used. Chunmei et al. discovered that short-term application of composted pig manure fertilizer disturbed soil bacteria more intensely than fresh chicken manure and pig manure, resulting in a lower abundance of bacteria Actinobacteria and nitrifying bacteria Nitrospirae, both of which are involved in organic matter synthesis (Ye, Huang et al. 2022). Several research, however, found that using organic manure as a slow-release fertilizer had little influence on soil microbial community structure in the short term, owing to the fact that organic manure does not immediately offer fast-acting nutrients (Pimentel, Hepperly et al. 2005).

2.2.3 Impact of combined inorganic and organic fertilizer application on soil microbiome

Organic fertilizer application can provide organic carbon and common nutrients like N, P, and K, as well as medium and micronutrients or active compounds like amino acids, humic acid, and so on. With the physical action of macromolecules, it can also improve soil structure and physicochemical qualities, and it is a more prevalent nutrient addition measure in agricultural management. However, the actual nutrient content of organic fertilizers is low, and because it is a slow-release fertilizer, it takes time for microorganisms to break down into plant-available nutrients when applied to the soil as an exogenous substance. As a result, an increasing number of research are being conducted to investigate the appropriate application of organic and inorganic fertilizers in conjunction.

A combination of organic and inorganic fertilizers assures increased crop yields and enhances soil fertility without causing acidification, which probably occurs when inorganic fertilizers are over-applied. It has been demonstrated that a combination of organic and inorganic fertilizers can boost the carbon sequestration capacity of surface

tillage soils, which is important for sustainable agriculture (Chaudhary, Dheri et al. 2017). This is because when there is sufficient and stable organic matter in the soil, the soil's microbial community becomes more stable and healthier. And this technique of application may be able to generate a consistent improvement in agricultural food production (Shahid, Nayak et al. 2017). A comprehensive trial was also conducted by Enke et al. They evaluated soil chemical and biological indicators for a control group without fertilizer application, application of inorganic fertilizer, application of organic fertilizer, and farmyard manure mixed with inorganic fertilizer under their farming practices for 30 years (Liu, Yan et al. 2010). The results showed that soil SOC, TN, MBC, and microbial biomass nitrogen (MBN) increased significantly when organic and inorganic fertilizers were mixed, as did the activity of functional enzymes in the soil, and crop yields were significantly higher than in the organic fertilizer only group, the inorganic fertilizer only group and the control group with no fertilizer.

There are different views on whether the buffering effect of organic and inorganic fertilizer blends on soil pH plays a major role in regulating the soil microbial community. As one study found, a continuous manure and inorganic fertilizer application strategy for 33 years increased soil pH from 5.7 to 6.5, and this increase in pH buffering capacity increased the activity of cellulases and convertases in the soil (Saha, Prakash et al. 2008). Nonetheless, another study indicated that the increase in soil enzyme activity caused by the combined application of manure and inorganic fertilizer was caused by the increase in nutrients from the manure treatment rather than the buffering impact of pH (Zhang, Sun et al. 2019). At the same time, the effectiveness of organic application combined with chemical fertilizers on soil microorganisms is highly dependent on local meteorological circumstances and soil properties. In the tropics, for example, inorganic fertilizers are easily lost through surface runoff, leaching, and volatilization, whereas some organic fertilizers are mineralized at a faster rate by microorganisms. As a consequence, while performing relevant investigations, these environmental elements needy to be considered.

In terms of the response of soil microorganisms to combined organic and inorganic fertilizer, some studies have focused on the effect of enzyme activity in soil as an

indicator of a fertility evaluation. Zhang et al. applied pig manure and inorganic fertilizers in different proportions to a subtropical red soil region in southern China, and they discovered that combined fertilizer applications significantly increased the activities of β-1,4-glucosidase(βG), β-1,4-N-acetylglucosaminidase (NAG), and leucine aminopeptidase (LAP) enzyme activities in soil (Zhang, Dong et al. 2015). Moreover, Zhang proposed that the amount of inorganic phosphorus fertilizer applied should not exceed 44 kg ha^{-1} yr^{-1}. Aside from investigating the indicator of microbial response: and enzyme activity, another part of the study focused on the dynamic response of the microbial community to fertilizer application, DNA extraction of the soil bacterial community, and comprehensive analysis in conjunction with soil physicochemical properties. According to Hui et al., bacteria are more susceptible to manure, and the combined application of nitrogen fertilizer and manure in black soils can greatly enhance the abundance and diversity of bacteria, with the abundance of bacteria increasing twice as much as the diversity increasing 46.6% (Wang, Xu et al. 2018). Shi et al. reported that combining organic and inorganic fertilizers in acidic red soils could adjust the pH of the soil, increase the diversity of nitrogen-fixing bacteria, create a more complex and stable co-occurrence network, and increase the potential efficiency of biological nitrogen fixation in soil (Shi, Zhao et al. 2021). The results of Dali et al. showed that the combination of organic and inorganic fertilizers increased the content of total organic carbon (TOC), TN, available potassium (AK), total dissolved nitrogen (TDN), available phosphorus (AP), as well as increasing crop yields and improving the activity of soil microorganisms and the diversity of metabolic pathways for decomposing organic matter (Song, Dai et al. 2022). Furthermore, metagenomic sequencing studies indicated that long-term application of organic-inorganic blended fertilizers enhanced genes for organophosphorus mineralization by bacteria while decreasing genes for phosphorus assimilation by microorganisms, consequently controlling crop nutrient uptake (Hu, Gu et al. 2022). Long-term use of organic-inorganic mixed fertilizers was also found to enhance the soil microbiome's nitrogen fixation and nitrification (Li, Wang et al. 2020). Much of the research indicates that combined organic and inorganic fertilizers alter the structure and function of soil

microbial communities and have an impact on the biogeochemical cycling of important elements.

Chapter 3 Data collection and processing

3.1 Data collection

A comprehensive literature survey was performed through the Web of Science Core Collection, Google Scholar, and Scopus up to October 2022 using keywords “long-term” or “decades” or “years” and “fertiliz*” and “microbial community” and “soil”. The articles were selected based on the following criteria: (1) studies containing high-throughput sequencing data of the 16S rRNA gene in bulk soil were included; (2) long-term fertilization experiments and field surveys with explicit agricultural practice history were included; (3) the field trail with less than one-year duration was excluded.

We recorded the accession number of raw sequencing data and then excluded the studies with absence or inaccurate data. Based on these steps, we establish the bacterial sequencing dataset from bulk soil under long-term fertilization, which contains a total of 10308 samples from 103 individual publications. For each sample, we also collected a wide range of parameters for accurate analysis including environmental factors, soil properties, agricultural practice factors, and sequencing conditions.

Environmental factors, as background information for each sampling point, include geographical location, altitude, mean annual temperature (MAT), mean annual precipitation (MAP), ecosystem type (grassland or cropland), and sampling depth.

Soil properties include soil pH, soil organic carbon (SOC), total nitrogen (TN), ammonium nitrogen (NH_4^+-N), nitrate nitrogen (NO_3^--N), and soil texture (classified based on the United States Department of Agriculture, USDA, source: http://www.nrcs.usda.gov/). To obtain climate condition data that were not provided in the publication, we extracted MAP and MAT from the WorldClimate (https://www.world climate.com) based on the latitude and longitude of sampling sites.

Agricultural practice factors include four fertilization types (control group without any fertilization, chemical fertilization, organic fertilization, and chemical fertilization combined with organic fertilization), fertilization practice duration, the annual application rate of each fertilization, crop rotation patterns (rotation or monoculture), soil status when sampling (fallow or planting), whether the sampling site plant cover crop or apply straw mulching management.

And sequencing conditions include primer pairs, sequencing region for 16S rRNA, and sequencing instruments.

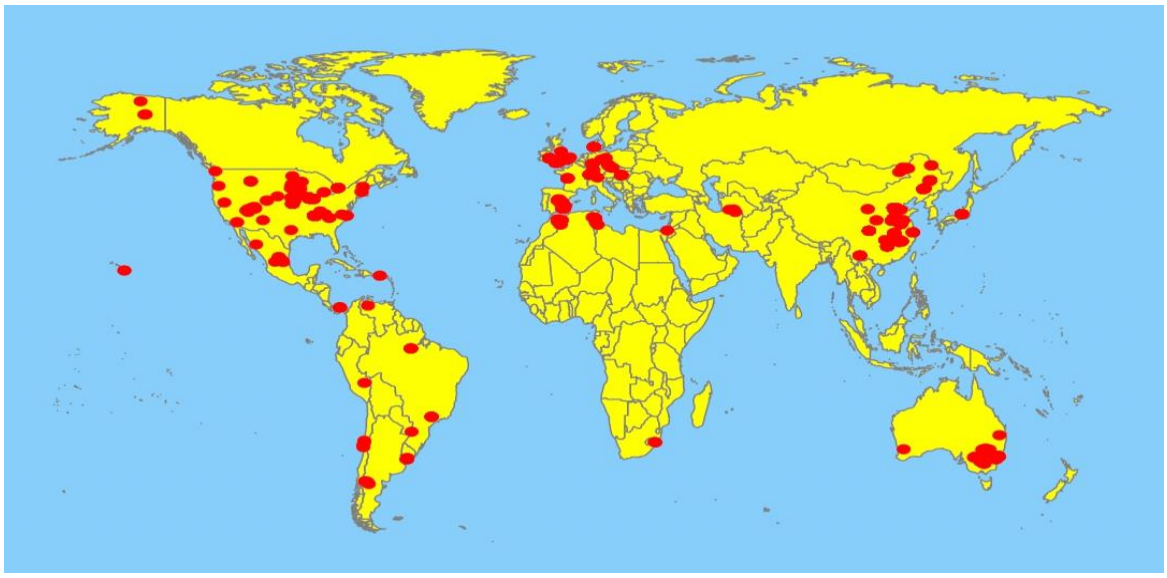

Figure 1. The map of the samples' distribution

3.2 Bioinformatics analysis

According to the accession number, we downloaded the raw data as FASTQ files from NCBI by using the Sratoolkit tool. Bioinformatic processing was performed using QIIME2 (version 2021.11.0) following our previous study. For each individual study, we used Cutadapt to remove primers that were recorded in the literature and then used VSEARCH to join the paired end reads. As for samples of which primers were already removed or reads were already joined before uploading to NCBI, these steps were skipped.

We used default quality thresholds by QIIME2's quality-filtered command to filter the low-quality reads and used Deblur to denoise sequences, then the amplicon sequence variants (ASVs) for each individual study were obtained. Then we merged ASVs from all the studies by merge-seqs command and made taxonomical annotation by the feature-classifier command based on the full-length 16S rRNA gene SILVA v138 database, which could integrate studies sequenced in different 16S rRNA gene regions by phylogenetic placement. ASVs annotated to mitochondria, chloroplasts, and those that could not be classified at the kingdom level there were removed. ASVs that were present in ten or fewer samples were also removed for downstream analysis. There were eventually 2176 samples remaining for further analysis after removing those below 2000 reads and rarefying samples to 2000 reads. Shannon's diversity index and Chao1 index were calculated to evaluate the alpha diversity of each sample.

In addition, to prove the chosen rarefaction level did not bring bias to our results, we performed correlation analysis between the diversity indexes under different rarefaction levels (rarefied to 2000 versus 10000 reads per sample). We found that there were highly significant correlations between the Chao1 indexes under the rarefaction level of 2000 reads per sample and of 10000 reads per sample ($R^2 = 0.941$; $P < 0.0001$), and the same results showed in Shannon's diversity indexes ($R^2 = 0.985$; $P < 0.0001$), which provided evidence that our rarefaction option did not influence the global analysis pattern of soil microbial community.

Finally, we used a conservative algorithm, the Functional Annotation of Prokaryotic Taxa (FAPROTAX), to analyze the potential functional groups of bacteria, which is widely used in functional annotations of prokaryotic taxa in environmental samples.

3.3 Model environmental variables processing

As mentioned in the data collection part, we collected various environmental data including environmental conditions, soil properties, agricultural practice factors, and

sequencing conditions. Firstly, we grouped fertilization application rate, duration of fertilization use, and some of soil properties into several classes according to the data properties (e.g., soil can be classified as acidic, neutral, or alkaline based on soil pH.) and data distributions: (1) annual application rate of organic fertilization < 4000 kg ha^{-1} , 4000 kg ha^{-1} ≤ annual application rate of organic fertilization ≤ 13000 kg ha^{-1} , annual application rate of organic fertilization > 13000 kg ha^{-1} ; (2) annual application rate of chemical fertilization < 200 kg ha^{-1} , 200 kg ha^{-1} ≤ annual application rate of chemical fertilization ≤ 400 kg ha^{-1} , annual application rate of chemical fertilization > 400 kg ha^{-1} ; (3) annual application rate of nitrogen fertilization < 100 kg ha^{-1} , 100 kg ha^{-1} ≤ annual application rate of nitrogen fertilization ≤ 250 kg ha^{-1} , annual application rate of nitrogen fertilization > 250 kg ha^{-1} (we also divided the annual application rate of potassium fertilization and phosphorus fertilization into three categories including low, medium, high application amount based on the data distribution, respectively.); (4) duration of fertilization use < 5 years, 5 years ≤ duration of fertilization use ≤ 15 years, 15 years < duration of fertilization < 25 years, 25 years ≤ duration of fertilization ≤ 30 years, duration of fertilization >30 years; (5) one rotation period ≤ 1 year, one rotation period > 1 year; (6) soil pH < 6.5, 6.5 ≤ soil pH ≤ 7.5, soil pH > 7.5; (7) SOC < 10 mg kg^{-1}, 10 mg kg^{-1} ≤ SOC ≤ 20 mg kg^{-1}, 20 mg kg^{-1} < SOC < 30 mg kg^{-1}, SOC ≥ 30 mg kg^{-1} ; (8) TN < 1.5 mg kg^{-1}, 1.5 mg kg^{-1} ≤ SOC ≤ 3 mg kg^{-1}, SOC > 3 mg kg^{-1}. After this step, we got continuous and discrete variables.

To gain more necessary variables to develop the model, we investigated one-hot encoding to transform data with discrete features. One-hot coding is mainly used to encode data with discrete features. This method maps the categorical values to integer values, and each integer value is represented as a binary vector, then the discrete features could be analyzed as continuous features without considering the underlying numerical relationship among features.

Chapter 4 Methodology

As we hypothesized that there would be different effects on microbial communities under different fertilization managements, we focused on the composition and diversity of microbial communities first. To better understand the interactions of taxa in microbial communities, we constructed co-occurrence networks of microbial species for different treatments. The co-occurrence network would show the evolution patterns by calculating the correlation matrixes of microbial taxa, which could determine the closeness and association among taxa. We could also estimate the stability of the networks and determine the keystone taxa of different treatments by calculating and comparing related metrics of co-occurrence networks.

Because soil microbial diversity would be disturbed by various environmental factors, we used different tree-based machine-learning models to determine the importance of environmental variables to the Shannon index. And the environmental variables with high importance indicate a stronger positive or negative influence on soil microbial diversity. We hypothesized that these important variables would bring main effects on the Shannon index of the soil microbial community, whereas at the same time, there would be interactions among these variables. So here we used a SHAP value-based method to explore the main effects and interactive effects.

The work can be almost divided into seven parts, which are shown in Figure 2.

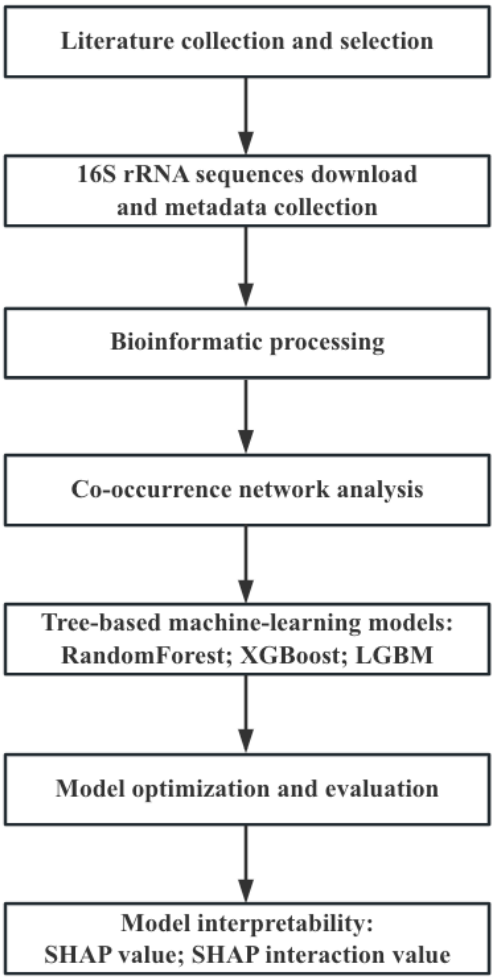

Figure 2. The work-flow chart of the entire work

4.1 Co-occurrence network analysis

Studies have demonstrated that the evaluation of microbial co-occurrence networks could investigate the complexity and the interactions among taxa.

We constructed the co-occurrence networks of the soil microbial community under different fertilization treatments based on a Spearman correlation method. Firstly, we remove ASVs of which the cumulative occurrence frequency was presenting below 0.01%. P values were corrected by a multiple test way based on the Benjamini-Hochberg false discovery rate (FDR). A random matrix theory (RMT) approach was employed to determine the correlation similarity threshold, then the pairwise

association between ASVs with an adjusted P-value below 0.05 and a score higher than the threshold were retained. Networks were visualized in Gephi.

We use natural connectivity to determine the discrimination of structural stability from the complex network by removing nodes in the network and evaluating how rapidly the stability decrease. Then we calculated the modularity of each network by the greedy modularity optimization method. Nodes in the same module are tightly connected with each other and less connected with the nodes outside the module.

Then we estimated the connectivity of each node by calculating within module connectivity (Zi) and among module connectivity (Pi), which were used to classify the nodes into four categories based on their topological roles in the whole network, including module hubs ($Zi > 2.5$ and $Pi < 0.62$), network hubs ($Zi > 2.5$ and $Pi > 0.62$), connectors ($Zi < 2.5$ and $Pi > 0.62$) and peripherals ($Zi < 2.5$ and $Pi < 0.62$). Among them, the module hubs and connectors mean the nodes which are highly connected with other nodes within a module and among the modules, respectively, which may be an important part to interact and mediate in one module (module hubs) and among modules (connectors). Network hubs are highly connected both within and among modules. Taxa of network hubs are expected to be vital in functioning and mediating interactions for the whole community. Peripherals are nodes with fewer connections than other nodes.

To describe the topological features of the networks comprehensively, we also calculated a set of metrics including average degree, graph density, average path length, network diameter, clustering coefficient, betweenness centrality, and closeness centrality. The degrees refer to the number of connections for one node; the average degree refers to the average connections of all the nodes in the network; the average path length refers to the average value of the distance between any pairs of the nodes; graph density refers to the ratio of the number of edges that occur to the number of possible edges, reflecting the cohesive nature of the network; the network diameter refers to the maximum value of the distance between any two nodes in the network; cluster coefficient refers to the closeness of the nodes in the network, also known as transferability. Among these parameters, average degree, graph density, and cluster coefficient indicate the associations of the network, the higher parameters suggest a

more connected network. As for each node of the network, we used the betweenness centrality and closeness centrality to describe the node-level features. Betweenness centrality refers to the number of shortest paths through a particular node and closeness centrality refers to the average distance of one node to other nodes, with higher values indicating the shorter distance from other nodes. Higher values of betweenness centrality and closeness centrality indicate a more important core position of a node in the network.

4.2 Tree-based machine-learning models

Tree-based machine learning models are increasingly common nonlinear models for the prediction and attribution study of biotic and abiotic dynamics within ecosystems. RandomForest, XGBoost, and LightGBM are tree-based machine-learning models that perform exceptionally well on regression applications. These three models are utilized in this study to assess the performance impacts on microbial Shannon diversity regression prediction and to achieve the optimal model.

This study utilizes the RandomForest, LightGBM, and XGBoost regression approach built in the Python version of sklearn, lightgbm, and xgboost library. In this work, the 5-fold cross-validation procedures are applied to test the accuracy of the model. In 5-fold cross-validation, the original data are randomly separated into 5 groups, four subsets of which are utilized as training data, while the remaining one subset is preserved for validation.

4.2.1 RandomForest

RandomForest, a machine learning method introduced by Breiman in 2001. It is an expanded variation of bagging and varies from bagging chiefly in that it involves randomized feature selection. RandomForest offers high prediction accuracy, is more tolerant to outliers and noise, and is less prone to overfitting. The technique is simple and quick, as well as straightforward to implement. It is widely used in data mining and

modeling. The general workflow of the RandomForest algorithm is as follows.

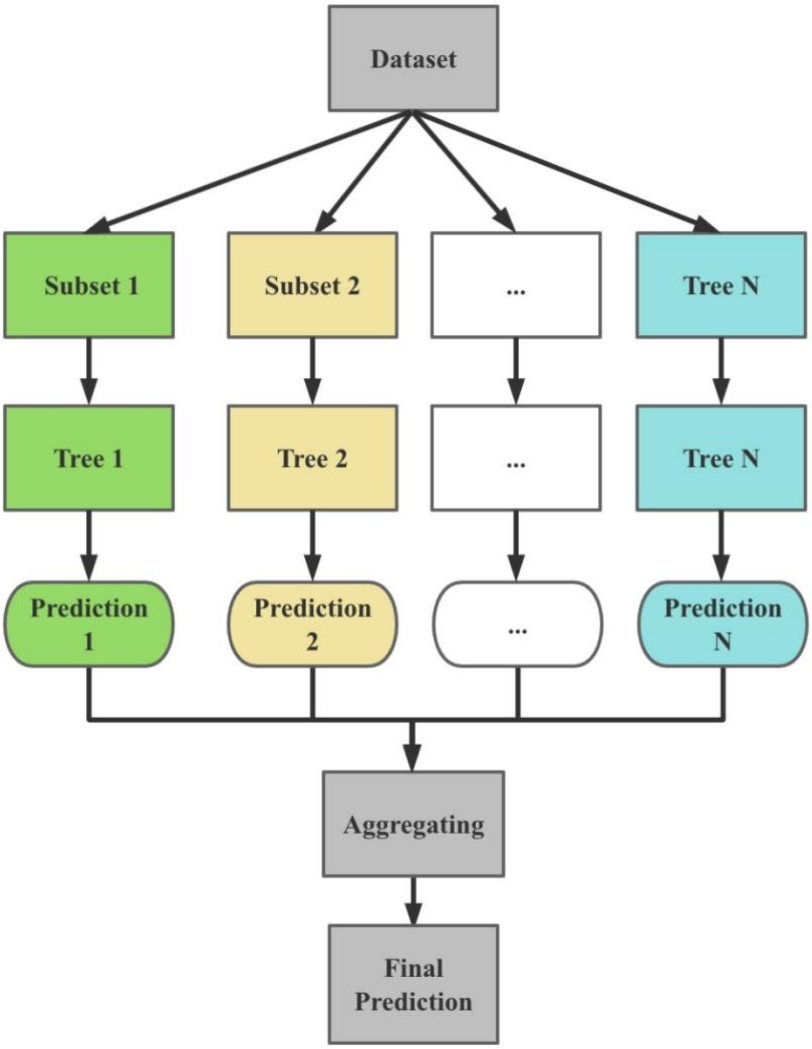

Figure 3. The flow chart of the RandomForest algorithm

4.2.2 XGBoost

XGBoost is one of the boosting algorithms, which has been the major option of many researchers in various issues due to its great computing efficiency and strong prediction outcomes. As the boosting approach is dependent on the gradient direction of the loss function to identify the weak prediction model for each step, the algorithm is termed gradient boosting. Shrinkage and column subsampling algorithms are used to minimize model bias and variation in XGBoost. The general workflow of the XGBoost algorithm is as follows.

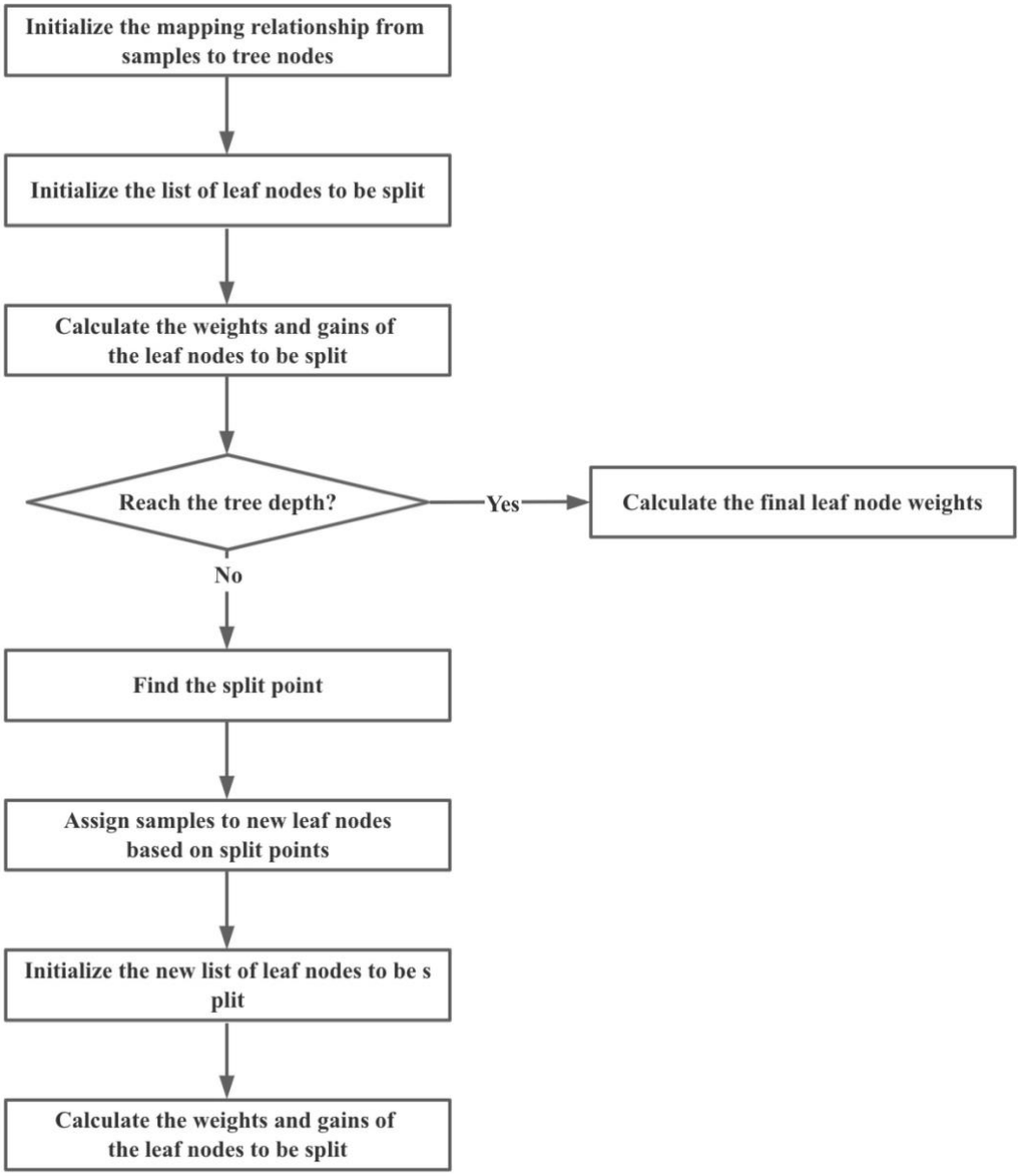

Figure 4. The flow chart of the XGBoost algorithm

4.2.3 LightGBM

LightGBM, newly created by Microsoft, is a highly efficient algorithm based on GBDT for handling issues with high-dimensional features and large-size data. LightGBM is a cutting-edge gradient-boosting framework that leverages tree-based learning approaches. It is designed with quicker training speed, less memory use, and greater accuracy. LightGBM contains two innovative features: gradient-based one-

sided sample (GOSS) and exclusive feature bundling (EFB). These two characteristics assist lower the data sample size and feature size during model training, without compromising accuracy or efficiency considerably. In addition, LightGBM is adjusted to overcome the limitations of other models, such as huge memory consumption, sluggish training speed, and extended running duration, by employing the histogram technique to simplify the data representation and minimize memory use.

4.2.4 Tree-based machine-learning model hyperparameters

It is very crucial to understand and clarify the meaning and tuning range of hyperparameters of machine learning model algorithms before tuning. As all three algorithms are tree-based, they have many hyperparameter settings that are essentially the same or similar. For the convenience of illustration, parameters having the same or comparable meanings are discussed together.

boosting: the type of base evaluator to be used. Usually, the default "gbdt" is used. num_iterations: the number of weak classifiers in integrated learning, or the number of trees if the base evaluator is a tree model. In general, a larger value is good for improving the model's precision but also increases the complexity and computation time of the model. When the num of iterations reaches the threshold, the accuracy will no longer increase but fluctuate in a small range.

learning_rate: the learning rate of the modeling process, which can also be interpreted as the step length of the learning process. A large value will speed up the iterations and reach the limit of the algorithm very quickly. However, it may not converge to the true optimum. Conversely, the smaller the value, the more space is left for the tree to be built later, which may lead to better model accuracy. However, the iteration speed will be slower.

n_estimators: the number of training trees in tree-based models.

max_depth(or num_leaves): The maximum depth of the tree and the number of leaves per tree. The theoretical connection is num_leaves = 2^(max_depth). However, this simple conversion is not good in practice. The reason is that for a fixed number of

leaves, leaf-type trees are usually much deeper than depth-type trees. Unconstrained depths can lead to overfitting. They can be set too large to make the tree more complex and more accurate, but this leads to overfitting. Therefore, it is not good to set its value too high.

reg_alpha: The weight coefficient of the regularization term, which is set to prevent overfitting. The larger the value the more conservative the model is.

gamma: The reduction of the minimization loss function requires the leaf nodes of the tree to make a division, subtracting some leaf nodes whose presence makes the loss function larger, so this value acts as a pruning function. The larger the value, the more conservative the algorithm is.

min_child_weight: The least sample weight sum among the child nodes. If the total sample weights of a leaf node is less than this value, then the splitting process is finished. The larger the value of this parameter the more conservative the algorithm is.

bagging_fraction: the proportion of the sample that is sampled, taking a value between 0 and 1. Feeding the model a little fewer data at a time makes it less overfitting and allows for better generalization, but the model will train more slowly.

The above parameters are largely meant to improve the training speed of the model, the model accuracy, and the generalization ability of the model. In many circumstances, the three are in conflict. High model training speed tends to lead to some loss of accuracy. Model accuracy increase may lead to overfitting. The purpose of modeling is to obtain a model with as high a precision as feasible without taking a lot of time and space and at the same time with generalization capabilities. Consequently, how determining the suitable hyperparameter values is of major relevance for the comprehensive performance of the model.

4.3 Bayesian optimization algorithm

The parameters of a machine learning model directly determine the model's efficacy and efficiency. Some of these parameters can be approximated via optimization

methods, but others, known as hyperparameters, cannot be learned from the input and therefore have to be supplied before the model is trained. The RandomForest, XGBoost, and LightGBM models all include numerous hyperparameters, and the values of the hyperparameters have a large influence on the model's regression prediction, hence hyperparameter optimization is a critical stage in modeling. Random search, grid search, genetic algorithm, and Bayesian tuning are among the tuning strategies. Relying on grid search to obtain model hyperparameters is highly challenging since it frequently takes an inordinate amount of time and server processing resources.

Bayesian optimization is an excellent method for hyperparameter optimization. Its algorithm design makes it feasible to obtain accurate models with both accuracy and efficacy. And it is essentially a robust algorithm that is efficient for the stochastic, nonconvex, and even discontinuous basis objective functions. Thus, this study applies the Bayesian optimization algorithm to optimize the parameters of the tree-based machine learning models.

The Bayesian optimization algorithm finds the optimal value of a function by constructing the posterior probability of the output of a black box function with a known finite number of sample points. Unlike grid search and random search, the Bayesian optimization algorithm framework is sequential, which means the current optimal value search is based on the outcomes of previous searches and takes full use of the known data.

The Bayesian optimization algorithm is based on Bayes' theorem. The principle of Bayesian optimization is to find the X^* in the hyperparameter space that makes the model generalization performance optimal in the number of dimensions d, which can be expressed as:

$$x^* = argmaxf(x)$$

Bayesian optimization employs a probabilistic surrogate model to simulate the present black-box objective function and an acquisition function to estimate the most probable position of the best advantage based on the current data. To prevent falling into local optima, Bayesian optimization algorithms frequently integrate a degree of randomness, making a tradeoff between random exploration and taking values based

on the posterior distribution.

According to Bayes' theorem, the model parameters are updated in the following equation:

$$p(f|D_{1:t}) = \frac{p(D_{1:t}|f)p(f)}{p(D_{1:t})}$$

$$D_{1:t} = \{(x_1, y_1), (x_2, y_2), \dots, (x_t, y_t)\}$$

$$y_t = f(x_t) + \varepsilon_t$$

Where f is the unknown objective function; $D_{1:t}$ is the observed training set; x_t is the hyperparameter vector; y_t is the observed value; ε_t is the observed error; $p(f)$ is the prior probability model of f; $p(D_{1:t}|f)$ is the likelihood distribution of y; $p(D_{1:t})$ is the marginal likelihood distribution, and $p(f|D_{1:t})$ is the posterior probability model of f, indicating the confidence level of the unknown objective function rectified a priori by the observed data set.

The sampling function is the referenced qualification for the Bayesian optimization algorithm to acquire the next sample point in the hyperparameter space. In this study, expected improvement method is employed as the sample function, which can be represented as:

$$\begin{cases} x_{t+1} = argmax f(x) \\ \alpha(x) = [\mu(x) - q^+]\phi(Z) + \sigma(x)\varphi(z) \end{cases}$$

where x_{t+1} is the hyperparameter of the evaluation; $\alpha(x)$ the objective function; $\mu(x)$ is the mean value; $\sigma(x)$ is the standard deviation value; q^+ is the maximum value of the current objective function; $\phi(Z)$ is the cumulative distribution function of the Gaussian distribution, and $\varphi(z)$ is the probability density function of the distribution.

4.4 Interpretable method: SHAP

The state-of-the-art TreeExplainer-based SHAP is based on the Shapley value from the cooperative game theory, namely, TreeExplainer-based SHapley Additive exPlanations. In this paper this approach is used to find important environmental variables for soil microbial diversity and to investigate local attribution and interactions between

variables in the Shannon diversity index predicted by the tree-based machine-learning model

The goal of the variable attribution approach is to explain the model by assigning an importance value to each feature used by the model. A linear model can be described as:

$$f(x) = \beta_0 + \beta_1 x_1 + \cdots + \beta_d x_d$$

Linear models are considered to be inherently interpretable because they summarize the importance of each feature in terms of scalar values. The importance of variable i for a linear model f is $\phi(f) = \beta_i$. $\phi(f)$ is called global variable attribution, which explains the model in a holistic way. Because the variables in a linear model are constrained to have a linear relationship with the model predictions, the relationship between that variable and the model output can be described in terms of the slope of the variable under sufficient statistics. Nevertheless, in many circumstances, we may choose to give a more personalized interpretation, especially for sample-specific predictions $f(x^s)$, rather than for the model as a whole. As a result, $\phi_i(f,\ x^s)$ is local variable attribution, which meets $\phi_i(f,\ x^s) = \beta_i x_i^s$, explaining the variable of a particular sample s.

In practical application scenarios, the relationship between x^s and $f(x^s)$ constructed by machine learning models is generally a nonlinear model that allows for complex interactions between variables. Therefore, we no longer have the slope in the linear model to summarize the global variables ascribed to $\phi(f)$. Instead, it is a very good way to utilize the Shapley method to attribute $\phi_i(f,\ x^s)$ to local variables to explain the model's predictions for a specific sample.

Shapley value is a method of assigning credits to each player in a cooperative game. We proceed via Shapley to acquire the variable contribution to the prediction of individual models in the three tree-based machine learning models presented in this study. The process can be described as follows.

$$\phi_i(v) = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!\,(|N| - |S| - 1)!}{|N|!} (v(S \cup \{i\}) - v(S))$$

Where i is the instance to be interpreted; $N = \{1, \ldots, d\}$ and S is the set and

subset of variables used in the model, respectively; $|N|$ is the number of the variables and $v(S)$ is the prediction of the subset S;

SHAP interprets the predicted output of the model as the sum of the attribute values of each individual variable as follows.

$$f(x) = \phi_0 + \sum_{i=1}^{|N|} \phi_i x_i$$

Where f is the interpretable model; ϕ_0 denotes the mean value of the prediction values and $\phi_i x_i$ is the SHAP value for variable i in sample x.

In the tree-based SHAP approach, the imputation of variables can be described below.

$$\phi_i = \frac{1}{M} \sum_{S \subseteq N \backslash \{i\}} \frac{1}{\binom{M-1}{|S|}} C(i|S)$$

where $N = \{1, \dots, d\}$ is the set of variables used in the model; S is the subset of $N\backslash\{i\}$; $|S|$ is the number of the variables; $C(i|S) = f(S \cup \{i\}) - f(S)$ is the contribution of variable i to the interpretable model f; We use an approximate method to compute the hard-to-compute $f(S)$, namely, path-dependent feature perturbation algorithm. This perturbation causes the samples to end up dropping to different leaf nodes. Thus, passing each path from the root to a leaf node of the tree is equivalent to perturbing a subset of the input features. Each leaf node will contain a proper proportion of all possible subsets in the set N. Then the SHAP value calculating process can be rewritten as:

$$\phi_i = \sum_{j \in N} \sum_{P \in S_j} \frac{\omega(|P|, j)}{L_j \binom{L_j - 1}{|P|}} \left(P_0^{i,j} - P_z^{i,j}\right) v_j$$

Where S_j is the subset of variables that appear at leaf node j; P is the subset of S_j; L_j is the path length from the root node to the leaf node i. $\omega(|P|, j)$ is the proportion of all subsets of P; $P_0^{i,j}$ and $P_z^{i,j}$ denote the proportion of subsets containing and not containing variable i in all subsets, respectively.

The global interpretation of a variable is the mean value of the sum of the absolute values of its local interpretations for all samples, which can be considered feature

importance as well.

Additionally, the SHAP values of a variable can be deconstructed into its main effects plus its SHAP interaction values with all other variables.

Similar to the SHAP value, the interactive effects can be further analyzed based on the Shapley interaction index from game theory, which enables the separate assessment of main and interaction local effects for each instance. Mathematically, the SHAP interaction value between i and j is described as follows.

$$\phi_{i,j}(f,x) = \sum_{S \subseteq N \setminus \{i,j\}} \frac{|S|!\,(|N| - |S| - 2)!}{2(|N| - 1)!} \nabla_{i,j}(f,x,S)$$

When $i \neq j$, and

$$\nabla_{i,j}(f,x,S) = f_x(S \cup \{i,j\}) - f_x(S \cup \{i\}) - f_x(S \cup \{j\}) + f_x(S)$$

Where $N = \{1, \ldots, d\}$ is the set of variables used in the model.

Specifically, considering pairwise variables i and j, the SHAP value of variable i for an instance can be decomposed into three parts. The first part is the main effect, representing an individual contribution without any interaction of other input variables. In addition, the second part is the interactive effect between the two variables i and j. And the third is the interactive effects between variable i and all other variables except j, namely, the residual.

4.5 Model performance metrics

The prediction of soil bacterial microbial diversity for long-term fertilization treatment is a regression problem. As a result, the RandomForest, XGBoost, and LightGBM regression model performances were assessed using the coefficient of determination (R Squared, R^2), mean squared error (MSE), root mean squared error (RMSE), and mean absolute error (MAE), defined as follows.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|$$

$$SSE = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

$$SSR = \sum_{i=1}^{n}(\hat{y}_i - \overline{y_i})^2$$

$$SSE = \sum_{i=1}^{n}(y_i - \overline{y_i})^2$$

$$R^2 = 1 - \frac{SSE}{SST} = \frac{SSR}{SST} = \frac{\sum_{i=1}^{n}(\hat{y}_i - \overline{y_i})^2}{\sum_{i=1}^{n}(y_i - \overline{y_i})^2}$$

Where y_i is the observed value, and $\hat{y}_i$ is the predicted value.

MSE is the residual squared mean of the observed value and the predicted value, and its smaller value indicates better model performance. RMSE is the arithmetic square root of the mean of the squared residuals of the observed and predicted values, which is sensitive to extreme and minimal values. And this metric can show the degree of dispersion of the sample, with smaller values suggesting better model performance. MAE is the average of the absolute errors of the observed and predicted values, which can better reflect the error of the predicted values. R^2 is the ratio of SSR to SST, which value ranges from [0,1]. The R^2 value closer to 1 means the better the model performance.

Chapter 5 Results

5.1 Microbial diversity and community composition

Table 1 shows the top ten taxa in terms of relative abundance at the phylum level under different long-term fertilization treatments. It is noteworthy that in terms of the frequency of phylum, the top ten ranked phylum in the control without fertilizer and in the treatment with inorganic fertilizer were higher than 90% in all samples. The top three in all samples were Proteobacteria, Actinobacteriota, and Myxococcota, with frequencies of 0.998, 0.995, and 0.974, respectively.

Specifically, the relative abundance of Crenarchaeota was 1.68% in the non-fertilized treatment group and 2.25% in the inorganic fertilized treatment group. These results showed that fertilization had a beneficial effect on the growth of Crenarchaeota, with the greatest increase in abundance with organic fertilizer alone, followed by organic-inorganic mixed fertilizer, and the least increase in abundance with inorganic fertilizer.

At the same time, it was discovered that Chloroflexi was also the most abundant taxon after fertilization, with the largest increase of 30.54% with organic-inorganic combined fertilization, followed by 13.53% with inorganic fertilization and 5.765% with organic fertilization. The taxa whose abundance decreased after fertilization were Verrucomicrobiota, Planctomycetota, and Myxococcota, among which Verrucomicrobiota was the most sensitive to organic fertilization, with 44.96% reduction in abundance by organic fertilization alone and 38.3% reduction by combined organic-inorganic fertilization; Myxococcota was also sensitive to the application of organic fertilizer, with a reduction of 18.91% (organic fertilizer only), 21.81% (organic-inorganic mixture) and 8.29% (inorganic fertilizer only), respectively.

Table 1. The top ten taxa in terms of relative abundance of soil bacteria at the phylum level under different long-term fertilization treatments.: CK (the control group without

fertilization), IF (the inorganic fertilization treatment group), OF (the organic fertilization treatment group), and IFOF (the inorganic and organic fertilization combined treatment group).

	Phylum	Relative abundance		Phylum	Relative abundance
CK	Proteobacteria	23.66%	IF	Proteobacteria	25.14%
	Acidobacteriota	17.93%		Acidobacteriota	17.14%
	Actinobacteriota	16.33%		Actinobacteriota	15.96%
	Chloroflexi	7.53%		Chloroflexi	8.70%
	Verrucomicrobiota	5.80%		Verrucomicrobiota	5.76%
	Bacteroidota	5.42%		Bacteroidota	4.65%
	Planctomycetota	4.83%		Planctomycetota	4.39%
	Gemmatimonadota	3.87%		Gemmatimonadota	4.22%
	Myxococcota	3.15%		Myxococcota	2.89%
	Firmicutes	2.10%		Firmicutes	2.57%
OF	Proteobacteria	23.13%	IFOF	Proteobacteria	22.49%
	Acidobacteriota	16.74%		Acidobacteriota	18.78%
	Actinobacteriota	12.11%		Actinobacteriota	16.72%
	Bacteroidota	9.95%		Chloroflexi	10.84%
	Chloroflexi	7.99%		Bacteroidota	4.63%
	Planctomycetota	4.37%		Gemmatimonadota	4.23%
	Gemmatimonadota	4.83%		Verrucomicrobiota	3.58%
	Crenarchaeota	3.54%		Planctomycetota	3.16%
	Firmicutes	3.34%		Crenarchaeota	2.62%
	Verrucomicrobiota	3.19%		Myxococcota	2.46%

The top ten taxa in terms of relative abundance at the order level under different long-term fertilization treatments are shown in Table 2. After fertilization, the relative abundance of Rhizobiales, Gaiellales, Chthoniobacterales, Pyrinomonadales, and

Solirubrobacterales dropped. Specifically, the relative abundance of Rhizobiales decreased by 25.94% with mixed fertilizers, 21.67% with organic fertilizers, and the lowest rate of 9.06% with inorganic fertilizers. Gaiellales' response to fertilization is also notable. Gaiellales had the biggest decline in relative abundance when only organic fertilizer was used, with a decrease ratio of 39.10%, but this taxon had a smaller decrease ratio when only inorganic fertilizer and organic-inorganic mixed fertilizer were used, with 8.52% and 9.66%, respectively.

Also, we found that the taxa that significantly increased in relative abundance after fertilization were Xanthomonadales and Nitrososphaerales. After the addition of inorganic fertilizer, mixed fertilizer, and organic fertilizer, the relative abundance of Xanthomonadales increased by 106.31%, 71.15%, and 62.46%, respectively. Nitrososphaerales increased in relative abundance by 20.77%, 45.82%, and 96.52%, respectively, with inorganic, mixed, and organic fertilizers.

Relative abundance of Acidobacteriales increased by 23.19% and 29.54% with inorganic and mixed fertilizers, respectively, but dropped by 28.69% with organic fertilizers alone. Similarly, Gemmatimonadales showed an increase in relative abundance when inorganic fertilizers only and mixed fertilizers were applied, with an increase of 10.97% and 10.48%, respectively; but a decrease in the relative abundance of 11.85% when only organic fertilizers were applied. Frankiales grew in relative abundance by 36.37% and 32.58% when inorganic fertilizer was supplied, including inorganic fertilizer only and mixed groups, respectively, but decreased by 39.84% when just organic fertilizer was applied. Interestingly, Vicinamibacterales was the taxon that responded positively to organic fertilization, with relative abundance increasing by 8.42% and 19.45% when mixed fertilizer and organic fertilizer were used alone, respectively, while relative abundance decreased slightly when inorganic fertilizer was used alone, at a rate of 0.12%.

We also identified the taxon that responded positively only to organic fertilizer application only, Chitinophagales, which increased its relative abundance by 54.61% with organic fertilizer application only but decreased its relative abundance by 6.93% and 27.81% with the addition of inorganic fertilizer and mixed fertilizer, respectively.

Further, we found that the effect of fertilization on the high abundance taxon Burkholderiales was not significant, and the relative abundance of Burkholderiales all decreased slightly when different fertilizers were applied, but the decrease ratios were 0.52% (inorganic fertilizer application), 1.87% (organic-inorganic mixture fertilizer application), and 1.49% (organic fertilizer application).

Table 2. The top ten taxa in terms of relative abundance of soil bacteria at the order level under different long-term fertilization treatments.

	Order	Relative abundance		Order	Relative abundance
CK	Rhizobiales	7.002%	IF	Burkholderiales	6.946%
	Burkholderiales	6.983%		Rhizobiales	6.368%
	Vicinamibacterales	5.172%		Vicinamibacterales	5.166%
	Gaiellales	4.548%		Gaiellales	4.161%
	Chthoniobacterales	3.759%		Gemmatimonadales	3.865%
	Chitinophagales	3.707%		Chitinophagales	3.450%
	Pyrinomonadales	3.564%		Pyrinomonadales	3.101%
	Gemmatimonadales	3.482%		Acidobacteriales	2.970%
	Acidobacteriales	2.411%		Chthoniobacterales	2.835%
	Solirubrobacterales	2.193%		Xanthomonadales	2.575%
OF	Burkholderiales	6.879%	IFOF	Burkholderiales	6.852%
	Vicinamibacterales	6.178%		Vicinamibacterales	5.608%
	Chitinophagales	5.731%		Rhizobiales	5.185%
	Rhizobiales	5.485%		Gaiellales	4.109%
	Nitrososphaerales	3.238%		Gemmatimonadales	3.847%
	Gemmatimonadales	3.070%		Acidobacteriales	3.124%
	Gaiellales	2.770%		Pyrinomonadales	2.723%
	Bacillales	2.292%		Chitinophagales	2.676%
	Sphingomonadales	2.267%		Chthoniobacterales	2.488%
	Pyrinomonadales	2.039%		Nitrososphaerales	2.402%

In addition, we classified the soil samples into three categories based on pH: acidic, neutral, and alkaline soils, with the goal of determining whether there are substantial changes in microbial composition in distinct acid and alkaline states, generated following long-term fertilization. We found that the top ten phyla of soil microbial composition under all three classifications were Proteobacteria, Acidobacteriota, Actinobacteriota, Chloroflexi, Bacteroidota, Verrucomicrobiota, Planctomycetota, Gemmatimonadota, Myxococcota, and Firmicutes. Among them, Proteobacteria, Acidobacteriota, and Actinobacteriota were the top three taxa in terms of relative abundance.

Acidobacteriota had a relative abundance of 17.52% in acidic soils, making it the second most common taxon in the total abundance of microbial communities in acidic soils. But Acidobacteriota placed third in overall abundance for alkaline and neutral soils, with relative abundances of 14.76% and 16.89%, respectively. Meanwhile, the total abundance of Proteobacteria taxa in acidic soils was 25.47%, greater than in neutral (21.59%) and alkaline (22.43%) soils. However, the relative abundance of Actinobacteriota in alkaline soils was highest (20.03%). The relative abundance of Actinobacteriota in acidic and neutral soils was 15% and 17.73%, respectively.

At the same time, we found that Gemmatimonadota was more abundant in alkaline soils with a relative abundance of 4.8%, in contrast to 3.73% and 3.82% in acidic and neutral soils, respectively. Bacteroidota taxa preferred acidic and neutral soils with a relative abundance of 5.8% and 6.19%, respectively, but in alkaline soils, the relative abundance was 3.9%. We further analyzed the microbial composition of different pH soils at the order level.

Table 3. The top ten taxa in terms of relative abundance of acid, Neutral, and alkaline soil bacteria at the order level.

	Order	Relative abundance		Order	Relative abundance
Acid	Burkholderiales	7.284%	Neutral	Burkholderiales	6.346%
	Rhizobiales	6.655%		Vicinamibacterales	6.119%
	Vicinamibacterales	4.346%		Rhizobiales	5.591%
	Gaiellales	4.025%		Gaiellales	4.347%
	Acidobacteriales	3.691%		Chitinophagales	3.662%
	Gemmatimonadales	3.579%		Gemmatimonadales	3.360%
	Chitinophagales	3.512%		Pyrinomonadales	3.228%
	Chthoniobacterales	3.319%		Chthoniobacterales	2.805%
	Pyrinomonadales	2.945%		Bacillales	2.242%
	Sphingomonadales	2.570%		Solirubrobacterales	2.064%
Alkaline	Nitrososphaerales	8.180%			
	Vicinamibacterales	7.230%			
	Burkholderiales	6.353%			
	Gaiellales	5.229%			
	Rhizobiales	5.122%			
	Gemmatimonadales	3.365%			
	Solirubrobacterales	2.894%			
	Pyrinomonadales	2.337%			
	Micrococcales	2.277%			
	Chitinophagales	2.267%			

The top ten taxa in terms of relative abundance of acid, Neutral, and alkaline soil bacteria at the order level are shown in Table 3. Interestingly, we discovered that the percentage of Burkholderiales and Rhizobiales in acidic soils was higher than in neutral and alkaline soils. Moreover, Acidobacteriales taxa were abundant in acidic soils with a relative abundance of 3.691%, but only 0.7% and 1% in alkaline and neutral soils, respectively. Similarly, the Ktedonobacterales taxon was 1.8% in acidic soils but only

0.53% and 0.45% in alkaline and neutral soils, accordingly. Chthoniobacterales were the favored taxon in acidic and neutral soils, with relative abundances of 3.332% and 2.8%, respectively, while alkaline soils had a relative abundance of only 0.93%.

It is worth noting that some taxa prefer to grow in alkaline soils, such as Nitrososphaerales, which has the highest relative abundance in alkaline soils (8.2%), but only 0.98% and 1.996% in acidic and neutral soils. The relative abundance of Gaiellales was 5.2% in alkaline soils and 4.02% and 4.35% in acidic and neutral soils, respectively.

Micrococcales, Pirellulales, and Thermomicrobiales were also taxa that preferred alkaline soils and their relative abundance in alkaline soils was more than double that of neutral and acidic soils. The relative abundance of Cytophagales taxa in neutral and alkaline soils was 1.12% and 1.16%, respectively, whereas it was just 0.07% in acidic soils.

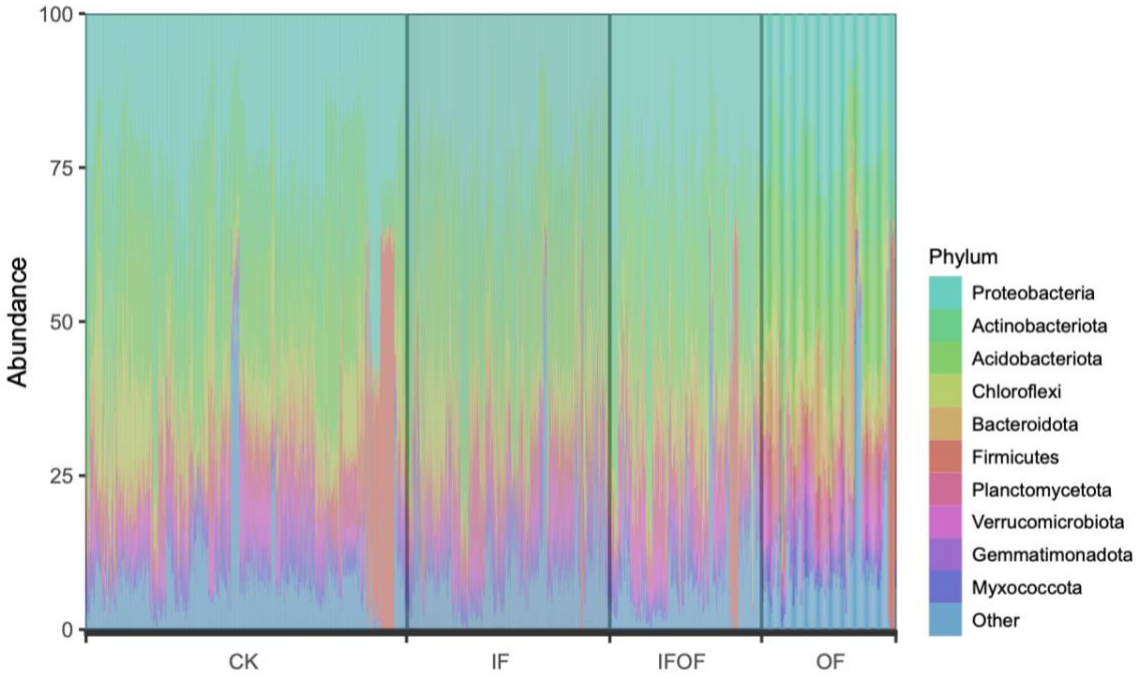

Figure 5. Stacked histograms of the relative abundance of taxa at the phylum level for each soil sample. The composition of each sample forms a vertical bar and is presented in grouped clusters based on the type of fertilizer applied.

In addition to the exploration of the changes in soil microbial species abundance under different fertilization treatments, we also evaluated microbial alpha diversity for each sample. Obviously, Figure 6 indicates that the four groups show almost identical patterns for the different microbial alpha diversity indexes. Both the Shannon index and the inverse Simpson index of soil microbial alpha diversity were considerably decreased after the application of inorganic fertilizers. The Shannon diversity index did not change significantly for soils with just organic fertilizer treatment, but the inverse Simpson diversity index did. In addition, microbial alpha diversity was reduced to varying degrees in soils with a combination of organic and inorganic fertilizers. In general, fertilizer application to the soil had an effect on bacterial alpha diversity, but it is crucial to highlight that the levels of the effect differed depending on the fertilizer.

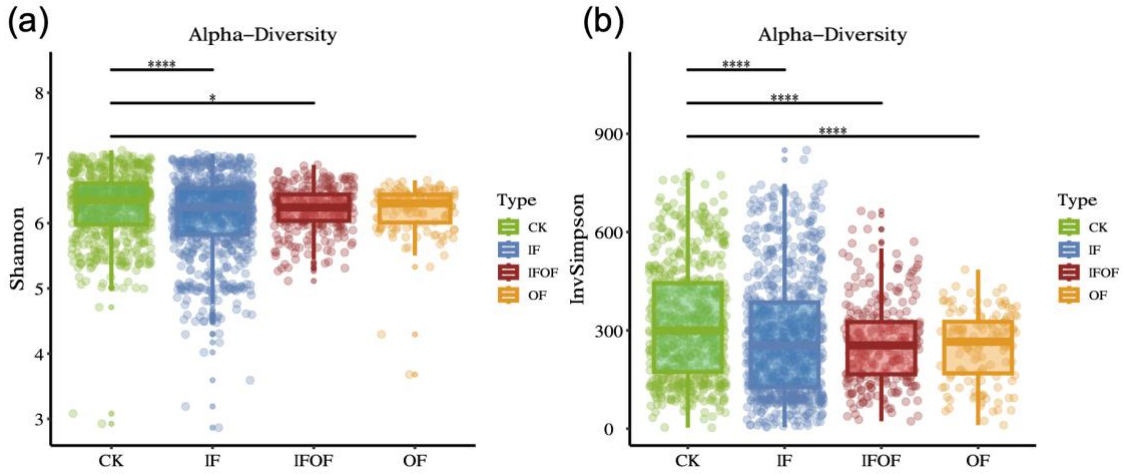

Figure 6. The boxplots of soil microbial alpha diversity for three different fertilizer treatment groups and the control group, containing Shannon diversity index (a) and inverse Simpson diversity index (b). Significance tests were performed separately for each fertilization group and control group. P values are indicated by asterisks: *P≤0.05; **P≤0.01; ***P≤0.001; ****P≤0.0001.

Generally, fertilization practices would change soil microbial composition and diversity. Inorganic fertilization decreased soil microbial diversity significantly, while the addition of organic fertilization did not show increases in the diversity indexes.

5.2 Comparison of microbial co-occurrence networks under different fertilization treatments

Based on the correlation and significance of the ASVs, microbial co-occurrence networks were constructed for the four groups of inorganic fertilizer, organic fertilizer, organic-inorganic combined fertilizer, and no fertilizer. Figure 7 depicts the different soil microbial co-occurrence networks under these four treatments. Each point represents a bacterial microorganism, and the edges between points indicate that there is a strong correlation between pairs of points that exceeds the RMT method's threshold, and the P-value after the FDR method is less than 0.05, denoting that the relationship between pairs of points is significantly strong. The color of the points is used to identify which modules they belong to. It is worth noting that the colors of the points reflect different modules in different co-occurrence networks, implying we need to analyze the module composition of each co-occurrence network independently.

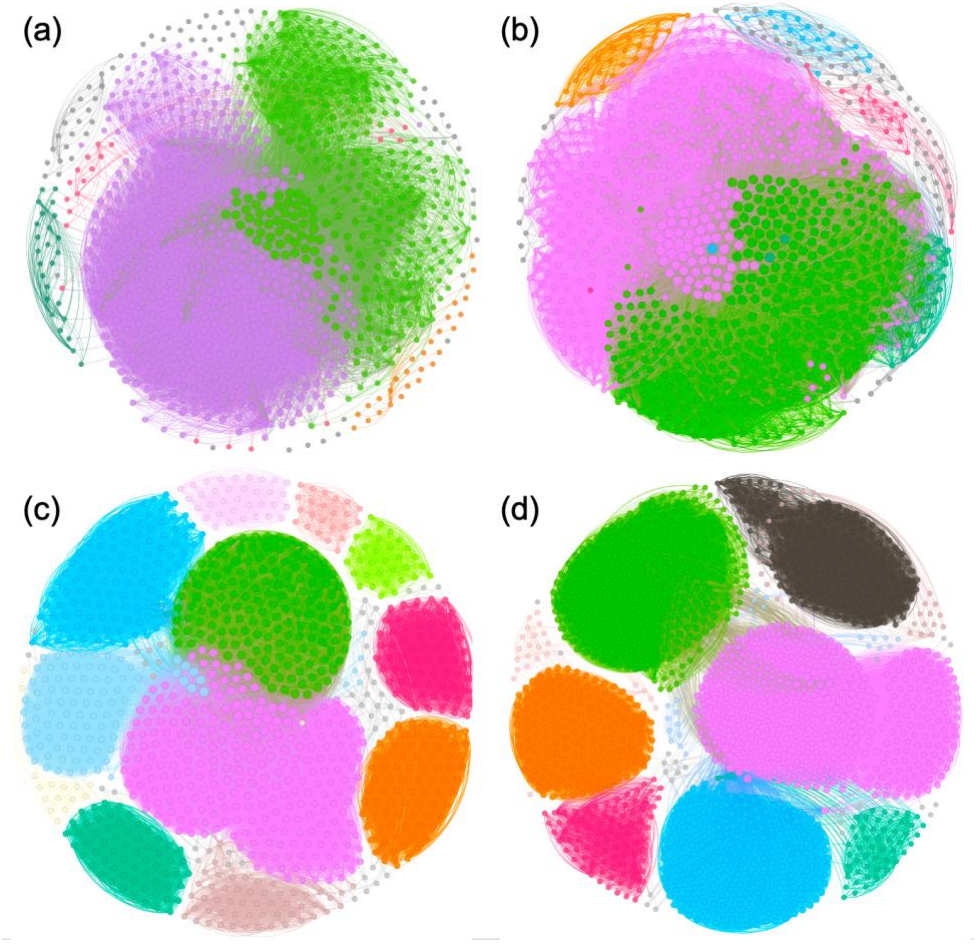

Figure 7. Soil microbial co-occurrence networks, which were constructed based on different fertilization treatment samples, (a) denotes the control group, (b) denotes the inorganic fertilization group, (c) denotes the inorganic-organic combined fertilization group, and (d) denotes the organic fertilization group.

The co-occurrence network with organic fertilizer application and organic-inorganic blended fertilizer application is obviously more complex than the co-occurrence network without fertilizer application and with inorganic fertilizer only, as evidenced by the co-occurrence network with organic fertilizer and organic-inorganic blended fertilizer application having a greater number of modules. Numerically, the degree of modularity of the co-occurrence network was 0.208 for the control group that did not receive fertilization, 0.204 for the treatment group that received only inorganic fertilization, 0.607 for the treatment group that received inorganic-organic mixed fertilization, and 0.738 for the treatment group that received only organic fertilization. Furthermore, the number of points in the co-occurrence network of both the organic-inorganic mixed fertilizer treatment group and the organic fertilizer treatment group was greater than that of the control group and the inorganic fertilizer treatment group, particularly the number of points in the co-occurrence network of the organic fertilizer treatment group (1509), which was 62.4% and 79.0% greater than that of the control group (929) and the inorganic fertilizer treatment group (843), respectively. This supports recent findings that organic fertilizer application might increase the complexity of soil networks.

Figure 8 depicts the degrees, eigen centrality parameters, and network destruction resistance of the microbial co-occurrence network for each of the four treatments. Furthermore, the average degree of the cooccurrence network in the inorganic fertilizer and non-fertilizer treatment groups was 248.9 and 278.2, respectively, which were greater than the average degree of the cooccurrence network in the organic fertilizer and inorganic-organic mixed fertilizer treatment groups, 196.6 and 106.8, respectively. Also, the graph density of the inorganic fertilizer (0.296) and no fertilizer (0.3) treatment groups was higher than that of the organic fertilizer (0.13) and combined

inorganic-organic fertilizer (0.106) groups. Furthermore, Eigenvector centrality considers both the weight magnitude of the corresponding connections of the points and the centrality of the points' neighboring nodes, i.e., nodes connected to the central node have greater centrality than those connected to the secondary nodes. Figure 8 (b) clearly shows the distribution of Eigenvector centrality in the four co-occurrence networks, indicating that the control and inorganic fertilizer treatment groups have higher values than the organic fertilizer treatment group, and are significantly higher than the inorganic-organic mixed fertilizer treatment group. The trends of Eigenvector centrality, mean degree size and plot density could indicate that whether organic fertilizer was applied or not had a great influence on the connectivity of the soil microbial co-occurrence network.

The natural connectivity index depicts the degree of connectedness. The trend of the network's natural connectivity may be fitted linearly during the random removal of 500 points one by one from the four individual co-occurring networks, as shown in Figure 8 (a). However, the slope of the network's natural connectedness for the no-fertilizer and inorganic fertilizer-only groups, both approximately 0.45, is greater than that of the organic-inorganic combined fertilizer and organic fertilizer groups, both about 0.2. In the case of random point deletion, the natural connectivity of the network is naturally affected. The magnitude of this effect, however, may be described by the slope, and the assumption that networks with a lower slope have greater resilience to destruction and stability is credible. In conclusion, investigations on the natural connection of microbial co-occurrence networks in terms of resistance to destruction illustrate that networks with organic or organic-inorganic blended fertilizers are more stable.

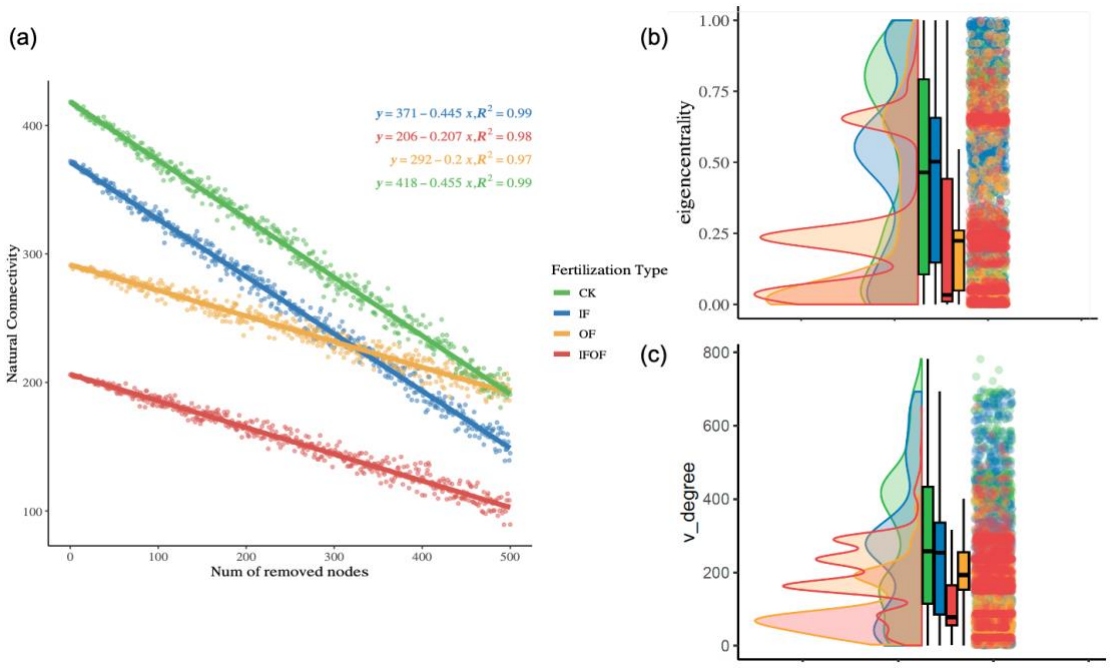

Figure 8. eigenvector centrality (b) and degree (c) for the three treatment groups and the control group. The point represents the point that appears in the co-occurrence network, and the color of the point indicates which group of co-occurrence networks it is in. (a) shows the process of natural connectivity resistance detection for the four co-occurrence networks.

We found that different long-term fertilization induces the soil microbial community to evolve separately, with more highly modular networks after the addition of organic fertilization. At the same time, the networks of organic fertilization were more stable, which means that the long-term use of organic fertilization would make the soil microbial community more connective and stable.

5.3 Soil microbial community function

The FAPROTAX method allows functional annotation of amplicon sequencing data. And the FAPROTAX functional annotation of soil bacterial communities yielded 92 functional community groupings. We classified these functional groups into four categories in order to investigate the functional composition of soil microbial

communities under varied fertilization treatments and climatic circumstances. The four functional groups are those associated with the carbon cycle, those associated with the nitrogen cycle, those associated with the sulfur cycle, and other functions. Here we focus on the first three functions related to the material cycle, and the functional annotation results are shown in Figure 9.

Functions related to the carbon cycle include chemoheterotrophy, aerobic chemoheterotrophy, fermentation, hydrocarbon degradation, methylotrophy, aromatic compound degradation, phototrophy, aromatic hydrocarbon degradation, methanol oxidation, photoheterotrophy, oxygenic photoautotrophy.

In the carbon cycle, chemoheterotrophy, aerobic chemoheterotrophy and photoheterotrophy were the three functions that appeared with the highest frequency. It is worth noting that soil microbial samples treated with inorganic fertilizer had a high abundance in all three functions. In particular, when inorganic fertilizer was applied to the soil, the chemoheterotrophy function increased by roughly 11.8% as compared to the soil without fertilizer. Yet, for both organic and organic-inorganic fertilizer combinations, the function of their chemoheterotrophy did not differ significantly from the control group. Interestingly, for the aerobic chemoheterotrophy function, these soils showed the same pattern across fertilization treatment groups, i.e., a 10.2% increase in functional abundance with inorganic fertilization, with little difference between the other two fertilizers and no fertilization. Yet, soils treated solely with organic fertilizers exhibited distinct functions, including fermentation, hydrocarbon degradation, and methylotrophy. These three carbon cycle functions were significantly higher in soil samples fed with only organic fertilizers than in soil samples fertilized with no fertilizer, inorganic fertilizer, or organic-inorganic blends. The organic-inorganic fertilizer combination appears to be distinguished in the carbon cycle function as aromatic compound degradation and phototrophy function that is significantly lower than the other three cases, but the enhancement of methanol oxidation function brought by the organic and inorganic fertilizer combination is not available in the other cases. Specifically, it was 40.0%, 12.5%, and 6.6% greater, respectively, than the control, inorganic fertilizer application, and organic fertilizer application groups.

For the nitrogen cycle, the functions associated with the nitrogen cycle include nitrification, denitrification, nitrite ammonification, nitrite respiration, aerobic ammonia oxidation, aerobic nitrite oxidation, nitrate reduction, nitrate respiration, nitrogen respiration, and nitrogen fixation.

As seen in Figure 9, the four most frequent functions in the nitrogen cycle were nitrification, denitrification, nitrite respiration, and aerobic ammonia oxidation. It is noteworthy that denitrification, nitrite respiration, and nitrogen fixation decreased after fertilization, regardless of the type of fertilizer applied. After applying only inorganic fertilizers to the soil, nitrate respiration increased by 54.9% compared to the non-fertilizer treatment group, while this value increased to 155.5% and 101.6% when the comparison was made with organic-inorganic fertilizer mix and organic fertilizer. Nitrogen respiration followed the same pattern as nitrate respiration. Also, when compared to the other three treatment groups, the treatment group with solely organic fertilizers had the highest levels of nitrification, nitrite ammonification, aerobic ammonia oxidation, and aerobic nitrite oxidation. Surprisingly, there was a similar but lower level of variation in the soils mixed with organic and inorganic fertilizers for these four nitrogen cycle functions, i.e., for these four functions, the organic-inorganic fertilizer mix showed an increase compared to the no fertilizer group, but the increase was less than that of the soils with only organic fertilizers.

Finally, sulfur cycle-related functions include sulfate respiration, respiration of sulfur compounds, sulfur respiration, thiosulfate respiration, and dark oxidation of sulfur compounds. It is worth noting that, with the exception of dark oxidation of sulfur compounds, the organic fertilizer-only group had the highest values of sulfur-related functions among the four treatment groups. However, the organic fertilizer soil had the second highest functional level of dark oxidation of sulfur compounds and was only 24.1% lower than the highest value holder, organic-inorganic mixed fertilizer soil, despite the fact that the organic fertilizer soil was about 78.7% higher than the control group.

Consequently, these results suggested that applying different fertilizers to the soil alters not just the soil microbial community, but also the potential functions associated

with soil nutrient cycling.

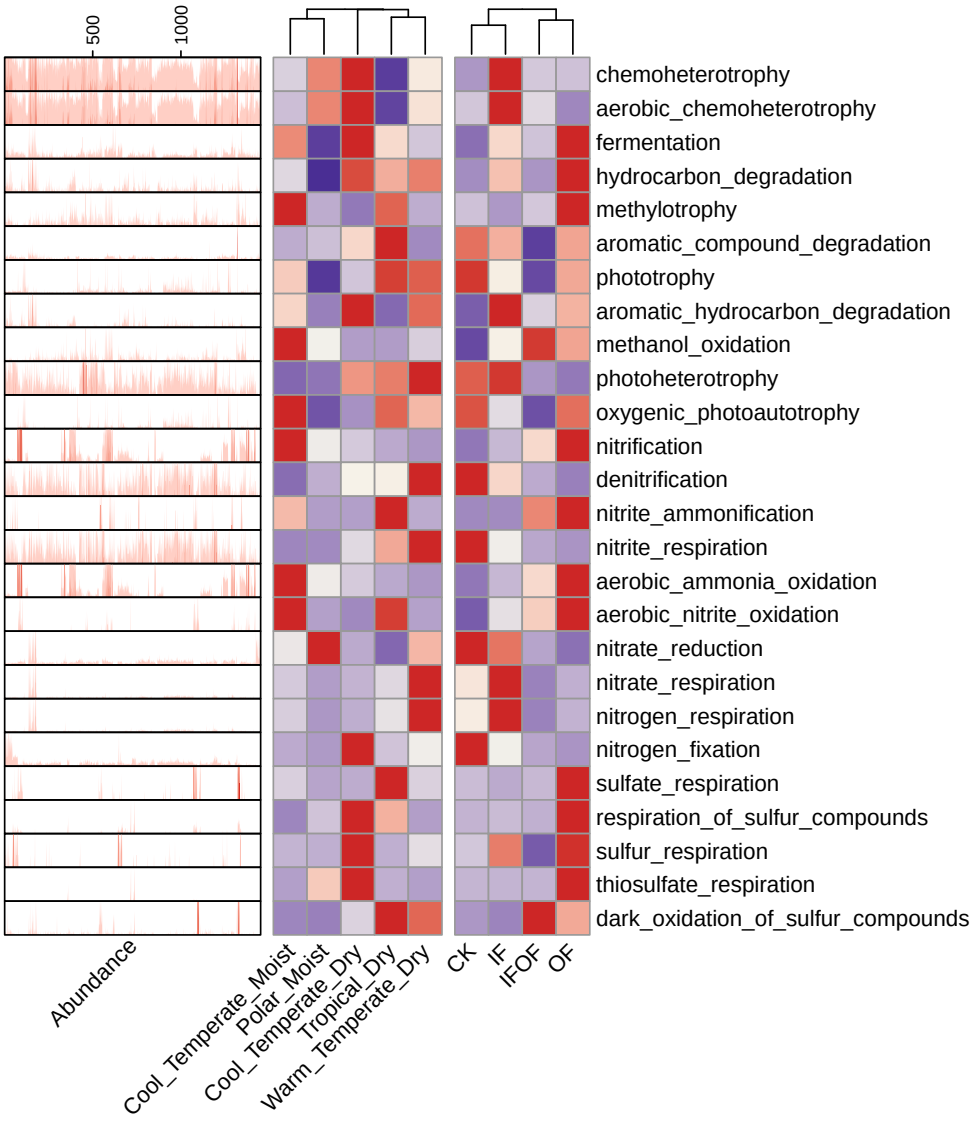

Figure 9. The heatmap of soil microbial functions under different fertilization treatments and different climatic zones, and the abundance of each function occurring in soil microbial samples.

5.4 Prediction of soil microbial diversity under different fertilization treatments

Prediction of soil bacterial microbial diversity under different fertilization treatments is a regression problem. In this study, we adopted five efficient tree-based machine-

learning models: RandomForest, XGBoost, and LightGBM. We used a Bayesian optimization algorithm to perform 500 iterations of each model hyperparameter at five different starting positions. This approach allows the optimal combination of model hyperparameters to be obtained over an estimable time horizon, allowing each model to compare the accuracy of each model for test machine detection while maximizing the model's benefits. The optimal hyperparameters for each model are shown in Table 4 below.

Table 4. The optimized hyperparameters of tree-based models

RandomForest	n_estimators	911
	min_samples_split	5
	min_samples_leaf	8
	max_depth	5
XGBoost	booster	'gbtree'
	gamma	3.36
	learning_rate	0.01
	max_depth	3
	min_child_weight	18
	n_estimators	969
	reg_alpha	0
LightGBM	boosting_type	'gbdt'
	n_estimators	727
	max_depth	3
	num_leaves	6
	learning_rate	0.01
	reg_alpha	18.41
	reg_lambda	14.18

We evaluate model performance using four scores: R^2, MSE, RMSE, and MAE. These four scores are obtained using distinct data calculation formulas and have different performance in different situations, each with its own set of advantages and limitations, which can help us understand the model's prediction performance in many aspects and more accurately. For example, RMSE is very sensitive for extreme data, which is sometimes noisy data. Nevertheless, MAE, defined as the average absolute value of the difference between the observed and predicted values, is much more robust to outliers. Therefore, we believe that when we compare the prediction results of a machine learning model trained on the training data set to the actual results on the test data set, the prediction values that are closer to the real values are better. That is, the machine learning model that predicts more accurately is better. It is vital to note that the model performs better when the R^2 value is close to 1. Smaller MSE, RMSE, and MAE values, on the other hand, indicate that the model performs better. The model metrics are shown in Table 5 below.

Table 5. The model metrics of tree-based models

	R^2	MSE	RMSE	MAE
RandomForest	0.501	0.172	0.414	0.259
XGBoost	0.542	0.146	0.382	0.251
LightGBM	0.534	0.149	0.386	0.252

RandomForest does not perform as well as XGBoost and LightGBM in general, and its scores are lower than those of the other two GBDT-based models invented later than it. The regression model is trained by minimizing the root mean square error (RMSE) between observed and predicted microbial soil microbial diversity. Furthermore, the MAE scores of these two exact gradients boosting decision tree models are nearly identical. As a result, XGBoost was chosen as the final model for the following variable imputation effect analysis as the best-performing model.

5.5 Revealing important environmental variables for microbial diversity in long-term fertilization soils

The SHAP approach is employed to interpret the contribution of model variables to a tree-based machine learning model. SHAP is a novel and effective machine learning interpretation method for analyzing the contribution of machine learning model predictions. SHAP can quantitatively indicate the local contribution of predictors for each sample. Variables having high absolute SHAP values are critical for the model's prediction accuracy. The models were randomly reconstructed using twenty distinct random seeds to eliminate bias due to randomness, and the SHAP list for each model was counted. We achieved a robust global variable importance result.

Figure 10 shows the importance of the SHAP global variables for XGBoost previously trained for microbial diversity regression prediction. The results are sorted from greatest to smallest by the median. Clearly, the most crucial variable is soil organic carbon. Soil organic carbon (SOC), total soil nitrogen (TN), and soil pH value were important in terms of soil physicochemical properties from various perspectives; as for fertilization, the type of fertilizer applied and the amount of inorganic fertilizer applied were very important in determining diversity; from the climatic perspective, mean annual temperature and mean annual precipitation were also important to a large extent; I n terms of sampling conditions, sample depth and elevation of sampling sites are also critical for the model to effectively predict diversity.

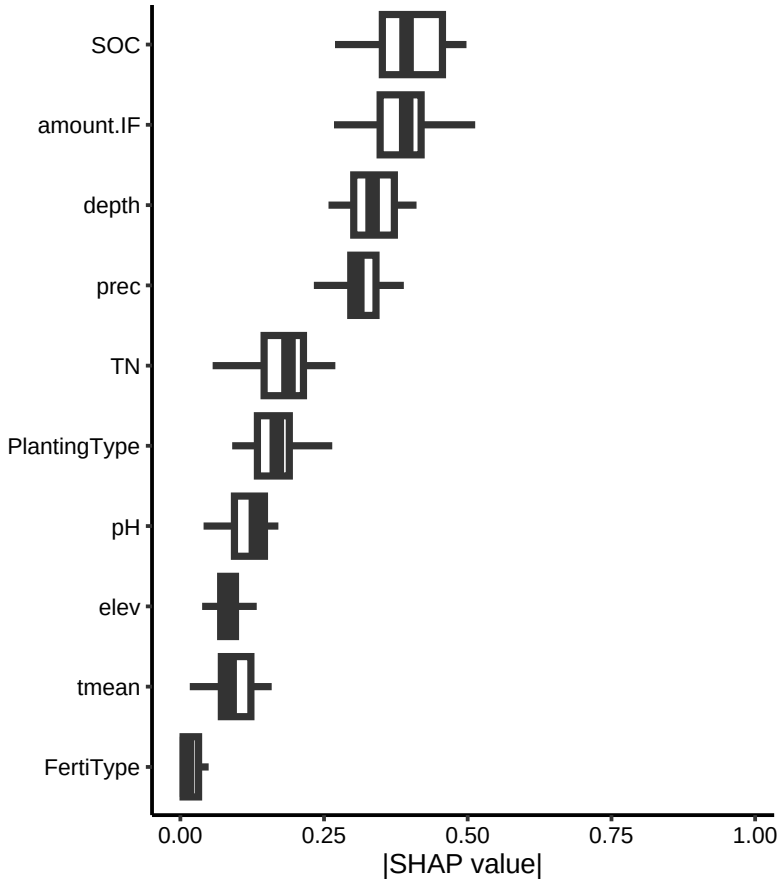

Figure 10. The importance of model variables sorted in descending order. The value was calculated using the absolute SHAP values of the distinct models constructed randomly for the full sample of 20 times.

The local SHAP values of the regression model can quantify whether a variable has a positive or negative effect on the predicted outcome, as well as the magnitude and volatility of the effect value across different ranges of variable values. Three significant factors were considered here: fertilizer application duration, soil organic carbon (SOC), and soil total nitrogen (TN). The predicted attributed SHAP values based on the variables for all samples are shown in the scatter plot Figure 11. As shown in Figure 11 (a), when the soil was fertilized for less than about 20 years, the SHAP values for the duration of fertilization were negative, indicating that the effect of fertilization application duration was negative. Yet, when the fertilizer application period exceeded 20 years, the fertilizer treatment had a positive effect on microbial diversity. Noticeably, for the same duration of fertilizer application, the SHAP values were taken differently,

and intuitively, the distribution of SHAP values at the same horizontal coordinate on the image was wide rather than narrow. This is an effect of the interaction of other variables for fertilization duration. For example, the same duration of fertilizer application in acidic and alkaline soils can lead to different effects of fertilizer duration, or there can be differences in the magnitude of the effect of the same duration of fertilizer application under different soil physicochemical conditions. Figure 11 (b) shows that Soil organic carbon content has mostly positive effects on soil microbial diversity. Specifically, there is a tipping point in the SHAP value of SOC around its SOC value of 10, implying that at smaller values than the tipping point, the soil is deficient in organic matter nutrients, which inhibits the enhancement of soil microbial diversity. However, if the SOC content of the soil reaches a healthy level, i.e., above the tipping point, the organic matter content of the soil has a positive effect on the maintenance of microbial diversity in the soil. In the same way, Figure 11 (c) shows that soil total nitrogen content is mostly positive for soil microbial diversity, which is consistent with the understanding that soil total nitrogen content represents, to some extent, the nutrient level of the soil. Consistent with the results for SOC, there was a tipping point in the trend of SHAP values for soil total nitrogen. This means that when the total nitrogen concentration of the soil falls below about 1, the soil is nitrogen-depleted and not conducive to the growth of soil microbial diversity. When the soil total nitrogen content surpasses the tipping point, the soil total nitrogen has a positive effect on the regression model's predictions, regardless of the amount beyond which it is exceeded.

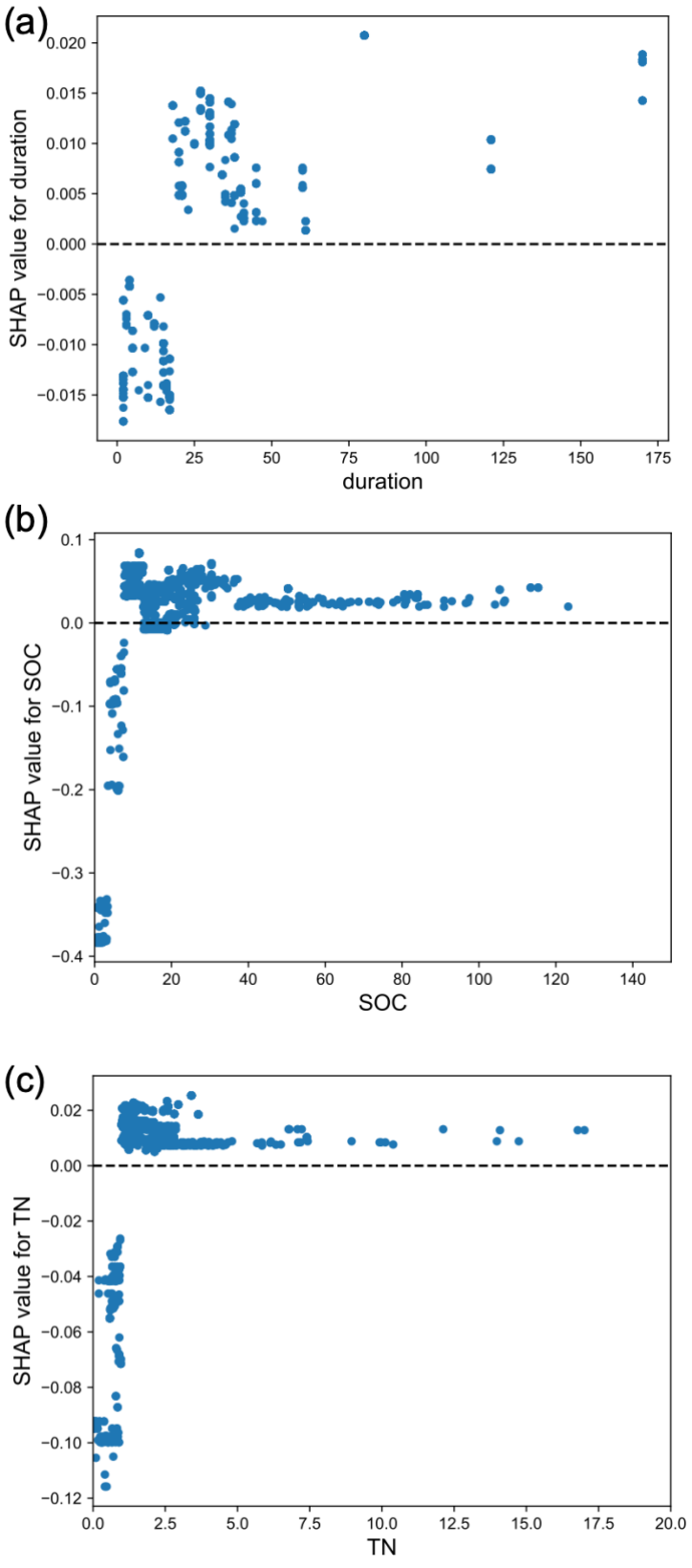

Figure 11. Scatter plots of the values of soil organic carbon and soil total nitrogen and their corresponding SHAP values for all samples at the time of fertilization treatment.

5.6 Environmental variables' main effect and pairwise interaction effects of soil microbial diversity.

One of the most important advantages of the game-theoretic-based SHAP algorithm is that it outputs local SHAP effects for each individual sample of the variables for each of the three model variables we have chosen. Furthermore, the effect of a single variable can be decomposed into the variable's direct effect and an indirect effect based on the variable-variable interaction. When studying the effects of pairs of variables, the SHAP effect of the specified variable can be split into a sum of three components. These three components are SHAP main effect (i.e., the effect of the specified variable alone); SHAP interaction effect (i.e., the quantification of the interaction effect between the specified variable and another variable, equally, the change in the effect of the presence or absence of the other variable on the effect of this variable); and SHAP residual effect (i.e., the sum of the effects of all variables other than these two variables on the specified variable).

Figure 12 depicts the effect of pH on soil microbial diversity, as well as the interaction effect of nitrogen fertilizer application on the SHAP value. of pH. The four subplots in Figure 12 represent the SHAP effect of pH; the SHAP main effect of pH; the interaction effect of pH and nitrogen fertilizer application; and the residual effect of pH, respectively. It can be seen from Figure 12 (a) that acidic soil has an inhibitory effect on microbial diversity, while alkaline soil has a promoting effect on that. Figure 12 (b) shows the main effect of pH on soil microbial diversity. It is clear that this main effect is much narrower than the SHAP of pH, which is also consistent with the principle. Figure 12 (c) illustrates that a low pH (< 5) nitrogen fertilizer application is advantageous to soil microbial diversity. Figure 12 (d) supports the results in (c) because the residual effect of pH is unstable, i.e., the SHAP residuals are distributed in different positions over the range of pH values. This proves that the conclusions we obtained earlier are robust.

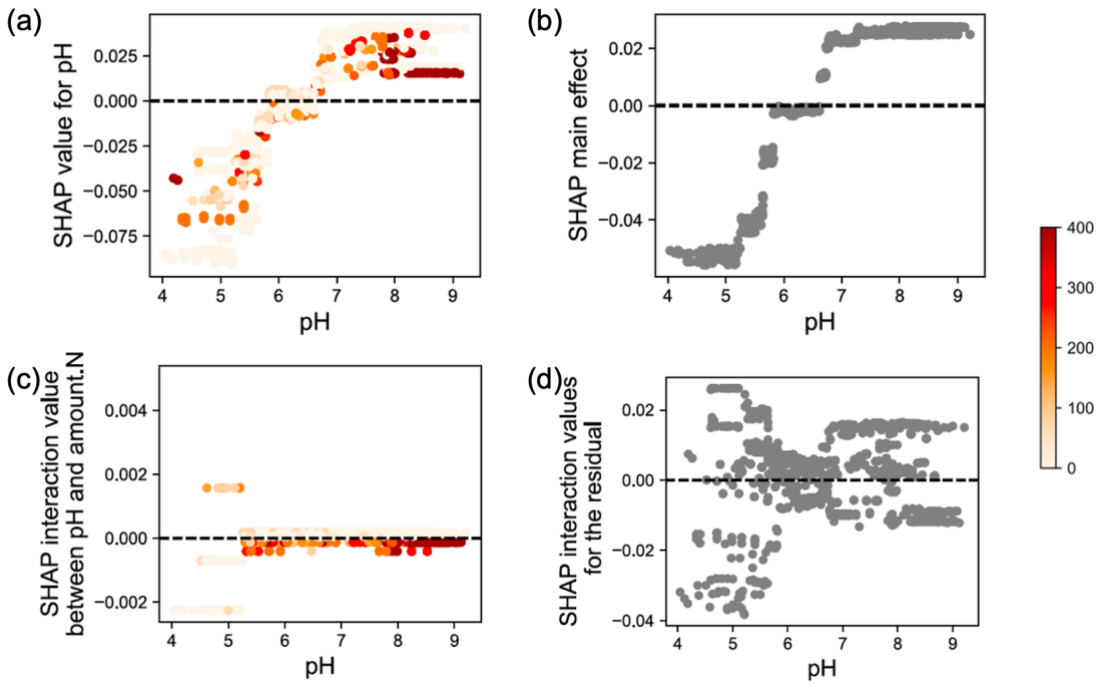

Figure 12. The effect of pH on soil microbial diversity and the interaction effect between nitrogen fertilizer application and pH.

Figure 13 depicts the interaction effect of soil organic carbon and the amount of nitrogen fertilizer placed on soil organic carbon on the SHAP value of soil microbial diversity. Figures 13 (a) and (b) show that, like pH, increasing soil organic carbon concentration has a positive influence on soil microbial diversity. From Figure 13 (c), the negative effect of the interaction between nitrogen fertilizer application and soil organic carbon was due to the excessive application of nitrogen fertilizer. In addition, as shown in Figure 13, a moderate application of nitrogen fertilizer can have a favorable influence on microbial diversity. It should be highlighted, however, that moderate nitrogen fertilizer application was a positive effect on pH and soil organic carbon, which interacted with diverse variables.

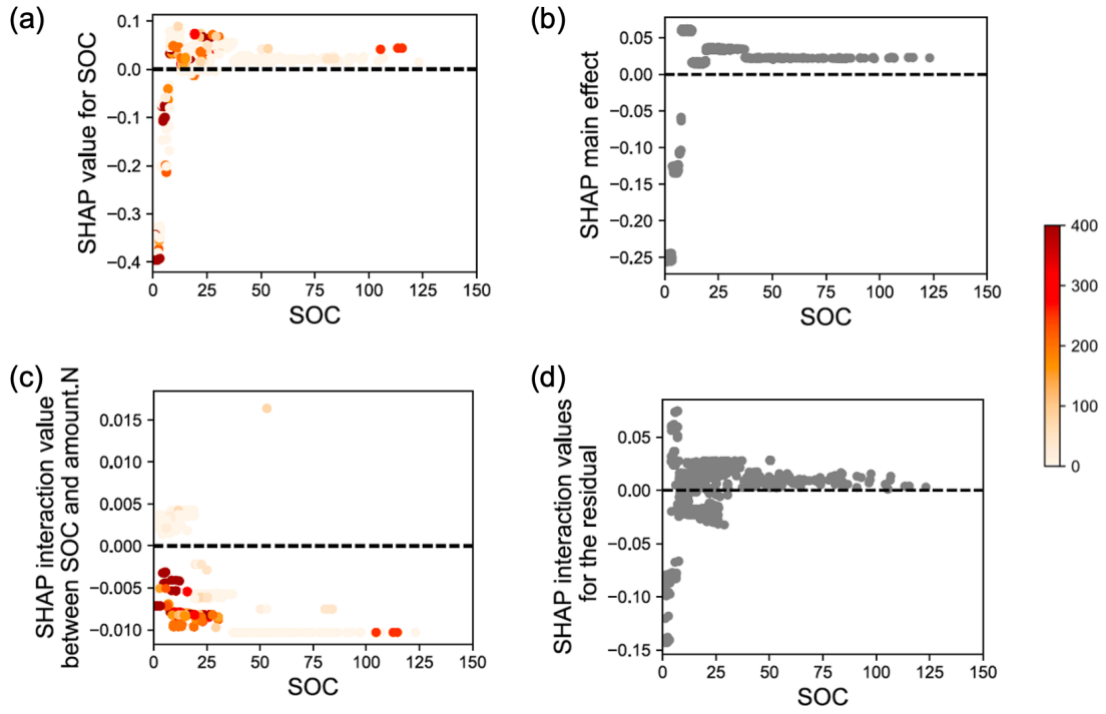

Figure 13. The effect of soil organic carbon on soil microbial diversity and the interaction effect between nitrogen fertilizer application and soil organic carbon.

Figure 14 shows the SHAP values of the annual mean temperature and the SHAP interaction values between pH and annual mean temperature on the XGBoost model. The effect of annual mean temperature on soil bacterial microbial diversity was divided by 14°C, as shown in Figure 14 (a), with environments less than 14°C having a negative effect on soil microbial diversity and environments greater than 14°C having a positive effect on soil microbial diversity. Also, at a cut-off of about 14 degrees Celsius, more acidic soils in soils less than 14 degrees Celsius have a beneficial effect on the SHAP values of the annual mean temperature. This beneficial effect is offset by a more pH-rich environment in soils above 14°C. Additionally, Figure 14 (d) reveals that for the SHAP values of annual mean temperature, the residual variables other than annual mean temperature and pH are still somewhat unstable, indicating the validity of our findings. This may be due to the fact that the ambient temperature affects the soil temperature to some extent. Different microorganisms have their own temperatures adapted for growth and reproduction, which directly affects microbial diversity. Furthermore, variations in soil temperature alter the activity of numerous enzymes in

the soil, resulting in changes in the function of the soil microbial community and thus indirectly influencing soil microbial diversity.

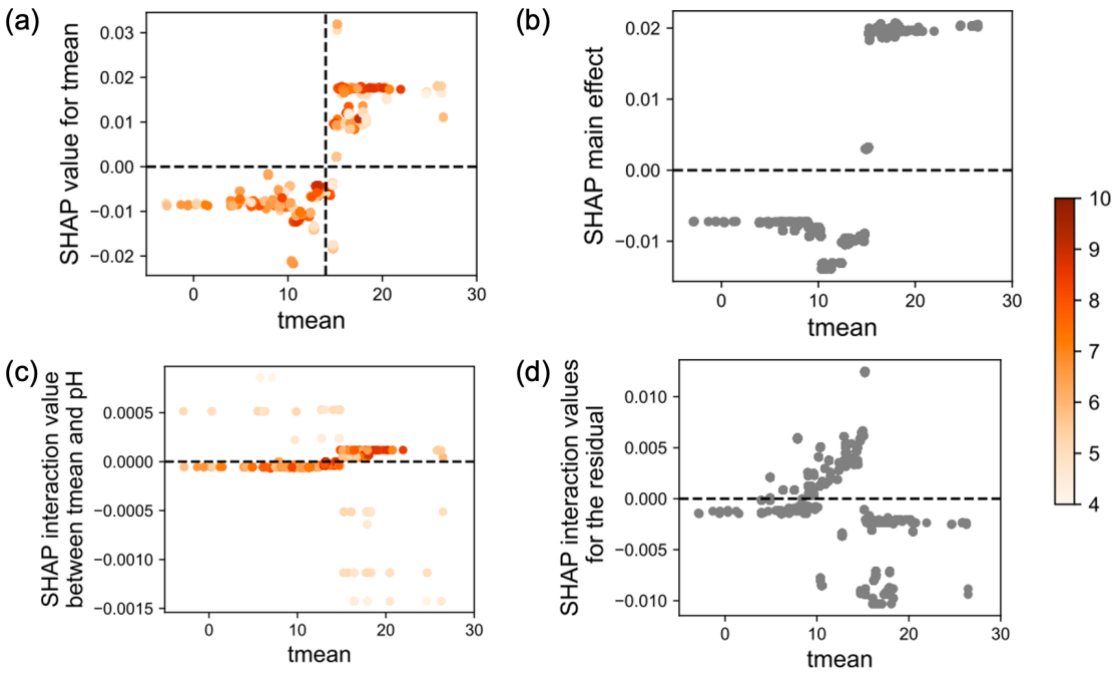

Figure 14. The effect of annual mean temperature on soil microbial diversity and the interaction effect between soil pH and annual mean temperature.

Chapter 6 Discussion

6.1 Soil microbial community composition and functional composition under different fertilization treatments

Here, we used a comprehensive meta-analysis approach to explore the microbial community composition and diversity under different long-term fertilization managements. We also studied the microbial community pattern under different soil pH conditions to investigate the potential response of typical taxa. Large differences in the soil microbial community composition were determined among three commonly employed fertilization practices. This is not surprising, since the response of the relative

abundance and diversity of soil microbial communities is determined by their own biochemical characteristics, metabolic adaptations, and ecological selection, which are closely linked to the physical and chemical properties of the soil under different fertilization practices (Geisseler and Scow 2014). We provided a macroscopic view of integrating global studies for a more universal conclusion.

After the long-term application of inorganic fertilization, the diversity of the soil bacterial community decreased, which is consistent with the findings of previous studies (Larkin, Griffin et al. 2010). There may be several reasons for this phenomenon. Firstly, the addition of NH_4^+ may promote the procession of nitrification, which leads to the release of H^+, thus increasing soil acidification with the decrease of soil pH(Geisseler, Linquist et al. 2017). When soil pH changed to acid status, it would bring negative effects on the growth of soil microorganisms. Secondly, it has been shown that the application of inorganic fertilization might increase the availability of nitrogen in the soil, which would decrease the diversity of the soil microbial community indirectly (Zeng, Liu et al. 2016). Thirdly, the application of inorganic fertilization might stimulate the growth of some specific taxa, then the evenness would get lower, which finally causes lower bacterial diversity (Hartmann, Frey et al. 2015). Besides, the results of SHAP analysis showed that the amount of inorganic fertilization was the third most important influencing factor affecting the soil microbial diversity, which means that inorganic fertilization could strongly influence the diversity index with a negative effect. However, we found that the relative abundance of some taxa was elevated in soils with long-term inorganic fertilizer application, such as Xanthomonadales and Acidobacteriales. The positive response of Xanthomonadales to inorganic fertilizer application was reported in many studies, for example, a significant decrease in soil bacterial diversity and a significant change in bacterial composition at the phyla level was reported in a field with long-term inorganic fertilizer application in the USA, while among all the dynamic taxa, Xanthomonadales was one of the most abundant taxa (Campbell, Polson et al. 2010). The long-term inorganic fertilization practices were found to result in significantly different bacterial community composition with a higher relative abundance of some acidotolerant taxa, for example, Acidobacteriales, which

could respond positively to the decrease of soil pH (Megyes, Borsodi et al. 2021). Meanwhile, members of Xanthomonadales and Acidobacteriales were demonstrated to play key roles in exchanging nutrients among bacterial species by the ability of degrading hydrocarbons, they could utilize various carbon substrates and survive in nutrient-prone or low-nutrient areas, so the higher abundance of Xanthomonadales and Acidobacteriales under organic fertilization also agreed well with the results of our analysis in the functional annotation of the species (Figure 9), which showed that the functions related to carbon cycling, especially chemoheterotroph, were higher under inorganic fertilization than the un-fertilized treatments and organic fertilization treatments.

Microbial diversity under long-term organic fertilization and combined fertilization did not show trends of increase, which were in accordance with previous studies. While the addition of organic fertilization significantly increased the proportion of some taxa, including Chitinophagales, Vicinamibacterales. Chitinophagales were identified as bioindicators of the environment with high available phosphorous and nutrients (Mason, Eagar et al. 2021). Some studies also demonstrated that the relative abundance of Chitinophagales was positively correlated with the increasing pH, so it could be concluded to the facts that the addition of organic fertilization (including the organic fertilization treatments and organically combined with inorganic fertilization treatments) improves the buffering capacity of soil pH. Meanwhile, it is worth noting that Chitinophagales were reported to survive in the high concentration of aminoglycosides antibiotics, which means that the addition of manure might bring some antibiotic-resistant bacteria into the soil (Deng, Mao et al. 2022). However, some studies demonstrated that the microorganisms from manure could not live in the soil for a long time. Our finding provided a view that the existence of antibiotic resistance bacteria might be stimulated by the antibiotic pressure under lasting organic fertilization. Some members of Vicinamibacterales have been proven to solubilize phosphate in the soil, which could increase the available phosphorous in the soil and promote plant growth (Wu, Rensing et al. 2022). In future work, we may further focus on the isolation and identification of phosphorus-solubilizing bacteria in soil by long-term fertilization

in order to explore a more effective transformation of phosphorus. The results in FAPROTAX also indicated that the functional microbial group of sulfate respiration and respiration of sulfur compounds were significantly higher than the un-fertilized group, which showed great potential in the functional microbial communities to be further identified and utilized under long-term organic fertilization and co-fertilization.

6.2 Microbial co-occurrence networks of different types of fertilization treatments

In microbial co-occurrence networks, species grouped in the same module are highly interconnected with each other, and species in the same module interact with each other more frequently than those with microorganisms located in other modules. The modularity of the soil microbial network under organic fertilization (both only organic fertilization treatments and co-fertilization treatments) is higher than that of the un-fertilized and chemical fertilization treatments (Figure 7 (a) (b) (c) (d)). Since modules can be explained as ecological niches for microorganisms, we found that the application of organic fertilizers leads to a more pronounced differentiation of microbial ecological niches, perhaps due to the use of complex nutrients selectively enriching microbes with specific functions, which is in good agreement with our findings in the functional annotation part (Figure 9). Meanwhile, we found that the stability of co-occurrence networks performed well under organic fertilization. To this end, we can deduce that the soil networks under organic fertilization assign more modules for various functions, and the networks owned stable interaction after long-term continuous fertilization, which is consistent with previous studies (Ling, Zhu et al. 2016).

6.3 Feasible soil microbial diversity prediction and contribution analysis by tree-based machine-learning models and SHAP approach

As our meta-analysis was based on various studies, it may bring cascading effects

on the analysis of soil microbial patterns. To understand comprehensively how environmental variables could influence soil microbial diversity directly and indirectly, we performed a contribution analysis by tree-based machine-learning models and used an interpretable SHAP analysis. We ranked and analyzed the potential influence of a large number of factors including climate conditions, soil physical and chemical properties of samples, and sequencing conditions.

We found that the amount of nitrogen addition should be determined by soil pH and soil organic carbon, as these environmental variables were various under long-term fertilization. For example, the addition of nitrogen fertilization was believed to suppress the soil microbial diversity, however, we found that when it was considered to add into a high SOC concentration soil, the negative influence would change into a positive influence. A meta-analysis of the influence of long-term nitrogen addition to soil diazotroph community showed consistent results (Zheng, Xu et al. 2023). They found that the negative effect of nitrogen fertilization addition on the diversity of nitrogen-fixing microbial taxa diminished with increasing soil organic carbon, and explained that this may be due to the presence of more potential carbon energy driving microbial nitrogen fixation at higher organic carbon concentrations and the fact that high organic carbon concentrations would ensure the carbon and nitrogen stoichiometric balance which nitrogen-fixing microbes themselves need to maintain. Therefore, our study also needs to be further refined, and we need to explore the specific mechanisms of these interactive factors.

Chapter 7 Conclusion

In this thesis, we outline the significance of soil microbes within the ecosystem and the related research background. Data from 10308 long-term fertilization publications from 103 publications worldwide were collected, including 16S rRNA amplicon sequencing and environmental metadata. Then, we performed data downloading, data processing, and data merging for all 16SrRNA amplicon data to obtain a large species-

annotated table of soil bacteria. We also compared the effects of different fertilizer types on soil microorganisms, including species composition, microbial alpha diversity, and microbial co-occurrence networks. Finally, we proposed using three tree-based machine learning models, RandomForest, XGBoost, and LightGBM, in conjunction with the interpretable machine learning method SHAP, to predict and attribute soil microbial Shannon diversity under various fertilization treatments. The main conclusions are as follows:

(1) Through biochemical processing and statistical analysis of 10308 soil samples from long-term fertilizer application experiments, we found patterns of changes in soil microbial community composition after the long-term application of various fertilizers. And by exploring the significantly changed species, we analyzed their significance in terms of their microbial preferences and functions. Furthermore, we evaluated the alpha diversity of soil bacteria and discovered that it was significantly lower in samples with inorganic fertilizer treatment to the control group without fertilizer application. FAPROTAX functional annotations were also performed for soil microorganisms in different fertilization treatments. Moreover, the functional intensity of the biogenic elemental cycle, including carbon, nitrogen, and sulfur, was estimated under different fertilization treatments.

(2) By calculating the correlation between microbes and the P-value after FDR adjustment, and after filtering with a random matrix theory threshold, this paper obtained the microbial co-occurrence network after three different long-term fertilizer treatments and the control group without fertilizer. The study discovered that using only organic fertilizer or organic-inorganic combined fertilizer resulted in a more modular microbial network with more diversified modules. In this article, 500 random node deletions were performed on each of the four constructed networks, with the changes in natural connectedness recorded and fitted. The module destruction resistance test results were found to be close to linear for all three treatments, with the destruction resistance of the network with organic fertilizer and organic-inorganic combined fertilizer application is higher.

(3) In addition, we innovatively build three tree-based machine learning models in

combination with the interpretable algorithm SHAP and optimize each of these models with hyperparameters through plain Bayesian optimization. After obtaining the best-fit models, we performed local and global quantitative attribution of variables for model prediction using the game-theoretic-based Shapley value and used 20 randomized construction experiments to identify the three most important predictors of soil bacterial Shannon diversity: soil organic carbon, inorganic fertilizer application amount, and sampling depth. To investigate the extent of the interaction between variables in the local contribution, we calculated the SHAP interaction value. As a result, the interaction between the amount of nitrogen fertilizer applied and the soil organic carbon and soil pH, respectively, was revealed.

Reference

Amundson, R., A. A. Berhe, J. W. Hopmans, C. Olson, A. E. Sztein and D. L. Sparks (2015). "Soil and human security in the 21st century." Science **348**(6235): 1261071.

Banerjee, S. and M. G. A. van der Heijden (2023). "Soil microbiomes and one health." Nature Reviews Microbiology **21**(1): 6-20.

Bardgett, R. D. and W. H. van der Putten (2014). "Belowground biodiversity and ecosystem functioning." Nature **515**(7528): 505-511.

Bender, S. F., C. Wagg and M. G. A. van der Heijden (2016). "An Underground Revolution: Biodiversity and Soil Ecological Engineering for Agricultural Sustainability." Trends in Ecology & Evolution **31**(6): 440-452.

Calderón, K., A. Spor, M.-C. Breuil, D. Bru, F. Bizouard, C. Violle, R. L. Barnard and L. Philippot (2017). "Effectiveness of ecological rescue for altered soil microbial communities and functions." The ISME Journal **11**(1): 272-283.

Callahan, B. J., J. Wong, C. Heiner, S. Oh, C. M. Theriot, A. S. Gulati, S. K. McGill and M. K. Dougherty (2019). "High-throughput amplicon sequencing of the full-length 16S rRNA gene with single-nucleotide resolution." Nucleic Acids Res **47**(18): e103.

Campbell, B. J., S. W. Polson, T. E. Hanson, M. C. Mack and E. A. G. Schuur (2010). "The effect of nutrient deposition on bacterial communities in Arctic tundra soil." Environmental Microbiology **12**(7): 1842-1854.

Chaudhary, S., G. S. Dheri and B. S. Brar (2017). "Long-term effects of NPK fertilizers and organic manures on carbon stabilization and management index under rice-wheat cropping system." Soil and Tillage Research **166**: 59-66.

Dan, N., N. Sadler, A. Bhattacharjee, E. B. Graham and J. K. Jansson (2020). "Soil Microbiomes Under Climate Change and Implications for Carbon Cycling." Annual Review of Environment and Resources **45**(1).

Dang, P., C. Li, C. Lu, M. Zhang, T. Huang, C. Wan, H. Wang, Y. Chen, X. Qin, Y. Liao and K. H. M. Siddique (2022). "Effect of fertilizer management on the soil bacterial community in agroecosystems across the globe." Agriculture, Ecosystems & Environment **326**: 107795.

Deng, Y., C. Mao, Z. Lin, W. Su, C. Cheng, Y. Li, Q. Gu, R. Gao, Y. Su and J. Feng (2022). "Nutrients, temperature, and oxygen mediate microbial antibiotic resistance in sea bass (Lateolabrax maculatus) ponds." Science of The Total Environment **819**: 153120.

Doran, J. W. and M. R. Zeiss (2000). "Soil health and sustainability: managing the biotic component of soil quality." Applied Soil Ecology **15**(1): 3-11.

Douglas, G. M., V. J. Maffei, J. R. Zaneveld, S. N. Yurgel, J. R. Brown, C. M. Taylor, C. Huttenhower and M. G. I. Langille (2020). "PICRUSt2 for prediction of metagenome functions." Nature Biotechnology **38**(6): 685-688.

Elfstrand, S., K. Hedlund and A. Mårtensson (2007). "Soil enzyme activities, microbial community composition and function after 47 years of continuous green manuring." Applied Soil Ecology **35**(3): 610-621.

Frostegård, Å., A. Tunlid and E. Bååth (1991). "Microbial biomass measured as total lipid phosphate in soils of different organic content." Journal of Microbiological Methods **14**(3): 151-163.

Geisseler, D., B. A. Linquist and P. A. Lazicki (2017). "Effect of fertilization on soil microorganisms in paddy rice systems – A meta-analysis." Soil Biology and Biochemistry **115**: 452-460.

Geisseler, D. and K. M. Scow (2014). "Long-term effects of mineral fertilizers on soil microorganisms – A review." Soil Biology and Biochemistry **75**: 54-63.

Guo, Z., S. Wan, K. Hua, Y. Yin, H. Chu, D. Wang and X. Guo (2020). "Fertilization regime has a greater effect on soil microbial community structure than crop rotation and growth stage in an agroecosystem." Applied Soil Ecology **149**: 103510.

Han, S., M. Delgado-Baquerizo, X. Luo, Y. Liu, J. D. Van Nostrand, W. Chen, J. Zhou and Q. Huang (2021). "Soil aggregate size-dependent relationships between microbial functional diversity and multifunctionality." Soil Biology and Biochemistry **154**: 108143.

Hannula, S. E., H. K. Ma, J. E. Pérez-Jaramillo, A. Pineda and T. M. Bezemer (2020). "Structure

and ecological function of the soil microbiome affecting plant-soil feedbacks in the presence of a soil-borne pathogen." Environ Microbiol **22**(2): 660-676.

Hartmann, M., B. Frey, J. Mayer, P. Mäder and F. Widmer (2015). "Distinct soil microbial diversity under long-term organic and conventional farming." The ISME Journal **9**(5): 1177-1194.

Hartmann, M. and J. Six (2023). "Soil structure and microbiome functions in agroecosystems." Nature Reviews Earth & Environment **4**(1): 4-18.

Hou, S., H. Ren, F. Fan, M. Zhao, W. Zhou, B. Zhou and C. Li (2023). "The effects of plant density and nitrogen fertilization on maize yield and soil microbial communities in the black soil region of Northeast China." Geoderma **430**: 116325.

Hu, X., H. Gu, J. Liu, D. Wei, P. Zhu, X. a. Cui, B. Zhou, X. Chen, J. Jin, X. Liu and G. Wang (2022). "Metagenomics reveals divergent functional profiles of soil carbon and nitrogen cycling under long-term addition of chemical and organic fertilizers in the black soil region." Geoderma **418**: 115846.

Hu, X., H. Gu, J. Liu, D. Wei, P. Zhu, X. a. Cui, B. Zhou, X. Chen, J. Jin, X. Liu and G. Wang (2023). "Metagenomic strategies uncover the soil bioavailable phosphorus improved by organic fertilization in Mollisols." Agriculture, Ecosystems & Environment **349**: 108462.

Jansson, J. K. and K. S. Hofmockel (2018). "The soil microbiome—from metagenomics to metaphenomics." Current Opinion in Microbiology **43**: 162-168.

Jansson, J. K. and K. S. Hofmockel (2020). "Soil microbiomes and climate change." Nature Reviews Microbiology **18**(1): 35-46.

Jian, S., J. Li, J. Chen, G. Wang, M. A. Mayes, K. E. Dzantor, D. Hui and Y. Luo (2016). "Soil extracellular enzyme activities, soil carbon and nitrogen storage under nitrogen fertilization: A meta-analysis." Soil Biology and Biochemistry **101**: 32-43.

Laborde, D., A. Mamun, W. Martin, V. Piñeiro and R. Vos (2021). "Agricultural subsidies and global greenhouse gas emissions." Nature Communications **12**(1): 2601.

Laconi, A., L. Mughini-Gras, R. Tolosi, G. Grilli, A. Trocino, L. Carraro, F. Di Cesare, P. Cagnardi and A. Piccirillo (2021). "Microbial community composition and antimicrobial resistance in agricultural soils fertilized with livestock manure from conventional farming in Northern Italy." Science of The Total Environment **760**: 143404.

Lane, D. J., B. Pace, G. J. Olsen, D. A. Stahl and N. R. Pace (1985). "Lane D, Pace B, Olsen G, Stahl D, Sogin M, Pace N.. Rapid determination of 16S ribosomal sequences for phylogenetic analyses. Proc Natl Acad Sci USA 82: 6955-6959." Proceedings of the National Academy of Sciences **82**(20): 6955-6959.

Larkin, R. P., T. S. Griffin and C. W. Honeycutt (2010). "Rotation and Cover Crop Effects on

Soilborne Potato Diseases, Tuber Yield, and Soil Microbial Communities." Plant Disease **94**(12): 1491-1502.

Li, P., J. Q. Wu, C. Y. Sha, C. M. Ye and S. F. Huang (2020). "Effects of Manure and Organic Fertilizer Application on Soil Microbial Community Diversity in Paddy Fields." Huan jing ke xue= Huanjing kexue / [bian ji, Zhongguo ke xue yuan huan jing ke xue wei yuan hui "Huan jing ke xue" bian ji wei yuan hui.] **41**(9): 4262-4272.

Li, Y., J. Tremblay, L. D. Bainard, B. Cade-Menun and C. Hamel (2020). "Long-term effects of nitrogen and phosphorus fertilization on soil microbial community structure and function under continuous wheat production." Environ Microbiol **22**(3): 1066-1088.

Li, Y., C. Wang, T. Wang, Y. Liu, S. Jia, Y. Gao and S. Liu (2020) "Effects of Different Fertilizer Treatments on Rhizosphere Soil Microbiome Composition and Functions." Land **9** DOI: 10.3390/land9090329.

Ling, N., C. Zhu, C. Xue, H. Chen, Y. Duan, C. Peng, S. Guo and Q. Shen (2016). "Insight into how organic amendments can shape the soil microbiome in long-term field experiments as revealed by network analysis." Soil Biology and Biochemistry **99**: 137-149.

Liu, E., C. Yan, X. Mei, W. He, S. H. Bing, L. Ding, Q. Liu, S. Liu and T. Fan (2010). "Long-term effect of chemical fertilizer, straw, and manure on soil chemical and biological properties in northwest China." Geoderma **158**(3): 173-180.

Liu, L. and T. L. Greaver (2010). "A global perspective on belowground carbon dynamics under nitrogen enrichment." Ecology Letters **13**(7): 819-828.

Louca, S., L. W. Parfrey and M. Doebeli (2016). "Decoupling function and taxonomy in the global ocean microbiome." Science **353**(6305): 1272-1277.

Mason, L. M., A. Eagar, P. Patel, C. B. Blackwood and J. L. DeForest (2021). "Potential microbial bioindicators of phosphorus mining in a temperate deciduous forest." Journal of Applied Microbiology **130**(1): 109-122.

Megyes, M., A. K. Borsodi, T. Árendás and K. Márialigeti (2021). "Variations in the diversity of soil bacterial and archaeal communities in response to different long-term fertilization regimes in maize fields." Applied Soil Ecology **168**: 104120.

Parks, D. H., G. W. Tyson, P. Hugenholtz and R. G. Beiko (2014). "STAMP: statistical analysis of taxonomic and functional profiles." Bioinformatics **30**(21): 3123-3124.

Pimentel, D., P. Hepperly, J. Hanson, D. Douds and R. Seidel (2005). "Environmental, Energetic, and Economic Comparisons of Organic and Conventional Farming Systems." BioScience **55**(7): 573-582.

Romero-Olivares, A. L., S. D. Allison and K. K. Treseder (2017). "Soil microbes and their response to experimental warming over time: A meta-analysis of field studies." Soil Biology

and Biochemistry **107**: 32-40.

Saha, S., V. Prakash, S. Kundu, N. Kumar and B. L. Mina (2008). "Soil enzymatic activity as affected by long term application of farm yard manure and mineral fertilizer under a rainfed soybean–wheat system in N-W Himalaya." European Journal of Soil Biology **44**(3): 309-315.

Saleem, M., J. Hu and A. Jousset (2019). "More Than the Sum of Its Parts: Microbiome Biodiversity as a Driver of Plant Growth and Soil Health." Annual Review of Ecology, Evolution, and Systematics **50**(1): 145-168.

Sanchez-Cid, C., R. Tignat-Perrier, L. Franqueville, L. Delaurière, T. Schagat and T. M. Vogel (2022). "Sequencing Depth Has a Stronger Effect than DNA Extraction on Soil Bacterial Richness Discovery." Biomolecules **12**(3).

Segata, N., J. Izard, L. Waldron, D. Gevers, L. Miropolsky, W. S. Garrett and C. Huttenhower (2011). "Metagenomic biomarker discovery and explanation." Genome Biology **12**(6): R60.

Semenov, M. V., G. S. Krasnov, V. M. Semenov, N. Ksenofontova, N. B. Zinyakova and A. H. C. van Bruggen (2021). "Does fresh farmyard manure introduce surviving microbes into soil or activate soil-borne microbiota?" Journal of Environmental Management **294**: 113018.

Shahid, M., A. K. Nayak, C. Puree, R. Tripathi, B. Lal, P. Gautam, P. Bhattacharyya, S. Mohanty, A. Kumar, B. B. Panda, U. Kumar and A. K. Shukla (2017). "Carbon and nitrogen fractions and stocks under 41 years of chemical and organic fertilization in a sub-humid tropical rice soil." Soil and Tillage Research **170**: 136-146.

Shi, W., H.-Y. Zhao, Y. Chen, J.-S. Wang, B. Han, C.-P. Li, J.-Y. Lu and L.-M. Zhang (2021). "Organic manure rather than phosphorus fertilization primarily determined asymbiotic nitrogen fixation rate and the stability of diazotrophic community in an upland red soil." Agriculture, Ecosystems & Environment **319**: 107535.

Simon, L., M. Lalonde and T. D. Bruns (1992). "Specific amplification of 18S fungal ribosomal genes from vesicular-arbuscular endomycorrhizal fungi colonizing roots." Applied & Environmental Microbiology **58**(1): 291-295.

Smith, P., D. Martino, Z. Cai, D. Gwary, H. Janzen, P. Kumar, B. McCarl, S. Ogle, F. O'Mara, C. Rice, B. Scholes, O. Sirotenko, M. Howden, T. McAllister, G. Pan, V. Romanenkov, U. Schneider, S. Towprayoon, M. Wattenbach and J. Smith (2008). "Greenhouse gas mitigation in agriculture." Philos Trans R Soc Lond B Biol Sci **363**(1492): 789-813.

Song, D., X. Dai, T. Guo, J. Cui, W. Zhou, S. Huang, J. Shen, G. Liang, P. He, X. Wang and S. Zhang (2022). "Organic amendment regulates soil microbial biomass and activity in wheat-maize and wheat-soybean rotation systems." Agriculture, Ecosystems & Environment **333**: 107974.

Thiele-Bruhn, S., J. Bloem, F. T. de Vries, K. Kalbitz and C. Wagg (2012). "Linking soil biodiversity and agricultural soil management." Current Opinion in Environmental

Sustainability **4**(5): 523-528.

Thompson, L. R., J. G. Sanders, D. McDonald, A. Amir, J. Ladau, K. J. Locey, R. J. Prill, A. Tripathi, S. M. Gibbons, G. Ackermann, J. A. Navas-Molina, S. Janssen, E. Kopylova, Y. Vázquez-Baeza, A. González, J. T. Morton, S. Mirarab, Z. Zech Xu, L. Jiang, M. F. Haroon, J. Kanbar, Q. Zhu, S. Jin Song, T. Kosciolek, N. A. Bokulich, J. Lefler, C. J. Brislawn, G. Humphrey, S. M. Owens, J. Hampton-Marcell, D. Berg-Lyons, V. McKenzie, N. Fierer, J. A. Fuhrman, A. Clauset, R. L. Stevens, A. Shade, K. S. Pollard, K. D. Goodwin, J. K. Jansson, J. A. Gilbert, R. Knight, J. L. A. Rivera, L. Al-Moosawi, J. Alverdy, K. R. Amato, J. Andras, L. T. Angenent, D. A. Antonopoulos, A. Apprill, D. Armitage, K. Ballantine, J. í. Bárta, J. K. Baum, A. Berry, A. Bhatnagar, M. Bhatnagar, J. F. Biddle, L. Bittner, B. Boldgiv, E. Bottos, D. M. Boyer, J. Braun, W. Brazelton, F. Q. Brearley, A. H. Campbell, J. G. Caporaso, C. Cardona, J. Carroll, S. C. Cary, B. B. Casper, T. C. Charles, H. Chu, D. C. Claar, R. G. Clark, J. B. Clayton, J. C. Clemente, A. Cochran, M. L. Coleman, G. Collins, R. R. Colwell, M. Contreras, B. B. Crary, S. Creer, D. A. Cristol, B. C. Crump, D. Cui, S. E. Daly, L. Davalos, R. D. Dawson, J. Defazio, F. Delsuc, H. M. Dionisi, M. G. Dominguez-Bello, R. Dowell, E. A. Dubinsky, P. O. Dunn, D. Ercolini, R. E. Espinoza, V. Ezenwa, N. Fenner, H. S. Findlay, I. D. Fleming, V. Fogliano, A. Forsman, C. Freeman, E. S. Friedman, G. Galindo, L. Garcia, M. A. Garcia-Amado, D. Garshelis, R. B. Gasser, G. Gerdts, M. K. Gibson, I. Gifford, R. T. Gill, T. Giray, A. Gittel, P. Golyshin, D. Gong, H.-P. Grossart, K. Guyton, S.-J. Haig, V. Hale, R. S. Hall, S. J. Hallam, K. M. Handley, N. A. Hasan, S. R. Haydon, J. E. Hickman, G. Hidalgo, K. S. Hofmockel, J. Hooker, S. Hulth, J. Hultman, E. Hyde, J. D. Ibáñez-Álamo, J. D. Jastrow, A. R. Jex, L. S. Johnson, E. R. Johnston, S. Joseph, S. D. Jurburg, D. Jurelevicius, A. Karlsson, R. Karlsson, S. Kauppinen, C. T. E. Kellogg, S. J. Kennedy, L. J. Kerkhof, G. M. King, G. W. Kling, A. V. Koehler, M. Krezalek, J. Kueneman, R. Lamendella, E. M. Landon, K. Lane-deGraaf, J. LaRoche, P. Larsen, B. Laverock, S. Lax, M. Lentino, I. I. Levin, P. Liancourt, W. Liang, A. M. Linz, D. A. Lipson, Y. Liu, M. E. Lladser, M. Lozada, C. M. Spirito, W. P. MacCormack, A. MacRae-Crerar, M. Magris, A. M. Martín-Platero, M. Martín-Vivaldi, L. M. Martínez, M. Martínez-Bueno, E. M. Marzinelli, O. U. Mason, G. D. Mayer, J. M. McDevitt-Irwin, J. E. McDonald, K. L. McGuire, K. D. McMahon, R. McMinds, M. Medina, J. R. Mendelson, J. L. Metcalf, F. Meyer, F. Michelangeli, K. Miller, D. A. Mills, J. Minich, S. Mocali, L. Moitinho-Silva, A. Moore, R. M. Morgan-Kiss, P. Munroe, D. Myrold, J. D. Neufeld, Y. Ni, G. W. Nicol, S. Nielsen, J. I. Nissimov, K. Niu, M. J. Nolan, K. Noyce, S. L. O'Brien, N. Okamoto, L. Orlando, Y. O. Castellano, O. Osuolale, W. Oswald, J. Parnell, J. M. Peralta-Sánchez, P. Petraitis, C. Pfister, E. Pilon-Smits, P. Piombino, S. B. Pointing, F. J. Pollock, C. Potter, B. Prithiviraj, C. Quince, A. Rani, R. Ranjan, S. Rao, A. P. Rees, M. Richardson, U. Riebesell, C. Robinson, K. J. Rockne, S. M. Rodriguezl, F. Rohwer, W. Roundstone, R. J. Safran, N. Sangwan, V. Sanz, M. Schrenk, M. D. Schrenzel, N. M. Scott, R. L. Seger, A. Seguin-Orlando, L. Seldin, L. M. Seyler, B. Shakhsheer, G. M. Sheets, C. Shen, Y. Shi, H. Shin, B. D. Shogan, D. Shutler, J. Siegel, S. Simmons, S. Sjöling, D. P. Smith, J. J. Soler, M. Sperling, P. D. Steinberg, B. Stephens, M. A. Stevens, S. Taghavi, V. Tai, K. Tait, C. L. Tan, N. Tas¸, D. L. Taylor, T. Thomas, I. Timling, B. L. Turner, T. Urich, L. K. Ursell, D. van der Lelie, W. Van Treuren, L. van Zwieten, D. Vargas-Robles, R. V. Thurber, P. Vitaglione, D. A. Walker, W. A. Walters, S. Wang, T. Wang, T. Weaver, N. S. Webster, B. Wehrle, P. Weisenhorn, S. Weiss, J. J. Werner, K. West, A. Whitehead, S. R. Whitehead, L. A. Whittingham, E. Willerslev, A. E.

Williams, S. A. Wood, D. C. Woodhams, Y. Yang and C. The Earth Microbiome Project (2017). "A communal catalogue reveals Earth's multiscale microbial diversity." Nature **551**(7681): 457-463.

Tian, S., B. Zhu, R. Yin, M. Wang, Y. Jiang, C. Zhang, D. Li, X. Chen, P. Kardol and M. Liu (2022). "Organic fertilization promotes crop productivity through changes in soil aggregation." Soil Biology and Biochemistry **165**: 108533.

Tilman, D., K. G. Cassman, P. A. Matson, R. Naylor and S. Polasky (2002). "Agricultural sustainability and intensive production practices." Nature **418**(6898): 671-677.

Vitousek, P. M. and R. W. Howarth (1991). "Nitrogen limitation on land and in the sea: How can it occur?" Biogeochemistry **13**(2): 87-115.

Wang, H., M. Xu, B. Zhou, X. Ma and Y. Duan (2018). "Response and driving factors of bacterial and fungal community to long-term fertilization in black soil." Scientia Agricultura Sinica **51**: 914-925.

Wemheuer, F., J. A. Taylor, R. Daniel, E. Johnston, P. Meinicke, T. Thomas and B. Wemheuer (2020). "Tax4Fun2: prediction of habitat-specific functional profiles and functional redundancy based on 16S rRNA gene sequences." Environmental Microbiome **15**(1): 11.

Wu, X., C. Rensing, D. Han, K.-Q. Xiao, Y. Dai, Z. Tang, W. Liesack, J. Peng, Z. Cui and F. Zhang (2022). "Genome-Resolved Metagenomics Reveals Distinct Phosphorus Acquisition Strategies between Soil Microbiomes." mSystems **7**(1): e01107-01121.

Xun, W., T. Huang, W. Li, Y. Ren, W. Xiong, W. Ran, D. Li, Q. Shen and R. Zhang (2017). "Alteration of soil bacterial interaction networks driven by different long-term fertilization management practices in the red soil of South China." Applied Soil Ecology **120**: 128-134.

Yang, Y., H. Cheng, H. Gao and S. An (2020). "Response and driving factors of soil microbial diversity related to global nitrogen addition." Land Degradation & Development **31**(2): 190-204.

Ye, C., S. Huang, C. Sha, J. Wu, C. Cui, J. Su, J. Ruan, J. Tan, H. Tang and J. Xue (2022). "Changes of bacterial community in arable soil after short-term application of fresh manures and organic fertilizer." Environmental Technology **43**(6): 824-834.

Zeng, J., X. Liu, L. Song, X. Lin, H. Zhang, C. Shen and H. Chu (2016). "Nitrogen fertilization directly affects soil bacterial diversity and indirectly affects bacterial community composition." Soil Biology and Biochemistry **92**: 41-49.

Zhang, X., W. Dong, X. Dai, S. Schaeffer, F. Yang, M. Radosevich, L. Xu, X. Liu and X. Sun (2015). "Responses of absolute and specific soil enzyme activities to long term additions of organic and mineral fertilizer." Science of The Total Environment **536**: 59-67.

Zhang, Y., C. Sun, Z. Chen, G. Zhang, L. Chen and Z. Wu (2019). "Stoichiometric analyses of

soil nutrients and enzymes in a Cambisol soil treated with inorganic fertilizers or manures for 26 years." Geoderma **353**: 382-390.

Zheng, M., M. Xu, D. Li, Q. Deng and J. Mo (2023). "Negative responses of terrestrial nitrogen fixation to nitrogen addition weaken across increased soil organic carbon levels." Science of The Total Environment **877**: 162965.

Appendix: 103 Publication used in the research

Number	Title	Accession Number
1	Bacterial Preferences for Specific Soil Particle Size Fractions Revealed by Community Analyses	PRJEB11366
2	Long-term nitrogen fertilization decreased the abundance of inorganic phosphate solubilizing bacteria in an alkaline soil	SRP072392
3	Rare microbial taxa as the major drivers of ecosystem multifunctionality in long-term fertilized soils	SRP140546
4	Mineral vs. Organic Amendments: Microbial Community Structure, Activity and Abundance of Agriculturally Relevant Microbes Are Driven by Long-Term Fertilization Strategies	PRJEB9307
5	Changes in Soil Microbial Activity, Bacterial Community Composition and Function in a Long-Term Continuous Soybean Cropping System After Corn Insertion and Fertilization	PRJNA658343
6	Fifteen-Year Application of Manure and Chemical Fertilizers Differently Impacts Soil ARGs and Microbial Community Structure	PRJEB29291
7	Organic farming induces changes in soil microbiota that affect agro-ecosystem functions	SRP074459
8	Response of Soil Microbes and Soil Enzymatic Activity to 20 Years of Fertilization	PRJNA644622
9	Microscale heterogeneity of soil bacterial communities under long-term fertilizations in fluvo-aquic soils	PRJNA720043
10	Organic and inorganic fertilizers respectively drive bacterial and fungal community compositions in a fluvo-aquic soil in northern China	PRJNA542408
11	Variations in soil bacterial taxonomic profiles and putative functions in response to straw incorporation combined with N fertilization during the maize growing season	SRP076696
12	Long-term organic fertilizer substitution increases rice yield by improving soil properties and regulating soil bacteria	PRJNA657529
13	Effect of 35 years inorganic fertilizer and manure amendment on structure of bacterial and archaeal communities in black soil of northeast	SRP059822

	China	
14	Distinct aggregate stratification of antibiotic resistome in farmland soil with long-term manure application	PRJNA423105
15	Composition, predicted functions, and co-occurrence networks of fungal and bacterial communities_ Links to soil organic carbon under long-term fertilization in a rice-wheat cropping system	PRJNA608072
16	Understanding the Responses of Soil Bacterial Communities to Long-Term Fertilization Regimes Using DNA and RNA Sequencing	SRP081067
17	Soil bacterial community structure and functional responses across a long-term mineral phosphorus (Pi) fertilisation gradient differ in grazed and cut grasslands	PRJEB21592
18	Long-term phosphorus deficiency decreased bacterial-fungal network complexity and efficiency across three soil types in China as revealed by network analysis	PRJNA525518
19	Organic amendment mitigates the negative impacts of mineral fertilization on bacterial communities in Shajiang black soil	PRJNA395622
20	Abundance, diversity, and structure of Geobacteraceae community in paddy soil under long-term fertilization practices	PRJNA561459
21	Soil microbial communities following 20 years of fertilization and crop rotation practices in the Czech Republic	PRJNA587449
22	The Response of the Soil Microbiota to Long-Term Mineral and Organic Nitrogen Fertilization is Stronger in the Bulk Soil than in the Rhizosphere	PRJEB27860
23	Soil Microbial Composition and phoD Gene Abundance Are Sensitive to Phosphorus Level in a Long-Term Wheat-Maize Crop System	PRJNA559597
24	Responses of microbial communities to a gradient of pig manure amendment in red paddy soils	SRP218113
25	Fate of heavy metals and bacterial community composition following biogas slurry application in a single rice cropping system	PRJNA503803
26	Organic amendments shift the phosphorus-correlated microbial co-occurrence pattern in the peanut rhizosphere network during long-term fertilization regimes	PRJEB22659
27	Bacterial diversity in soils subjected to long-term chemical fertilization can be more stably maintained with the addition of livestock manure than wheat straw	PRJEB7295
28	Long-term effects of manure and chemical fertilizers on soil antibiotic resistome	SRP111302
29	Soil Bacterial Communities Under Different Long-Term Fertilization Regimes in Three Locations Across the Black Soil Region of Northeast China	SRP080887
30	Long-Term Fertilization Shapes the Putative Electrotrophic Microbial Community in Paddy Soils Revealed by Microbial Electrosynthesis Systems	PRJNA600772
31	Long-term combined application of manure and chemical fertilizer	SRP135963

	sustained higher nutrient status and rhizospheric bacterial diversity in reddish paddy soil of Central South China	
32	Long-term application of swine manure and sewage sludge differently impacts antibiotic resistance genes in soil and phyllosphere	SRP248233
33	Contrasting assembly mechanisms and drivers of soil rare and abundant bacterial communities in 22-year continuous and non-continuous cropping systems	CRA003510
34	Fertilization Shapes Bacterial Community Structure by Alteration of Soil pH	SRP062251
35	Change of soil microbial community under long-term fertilization in a reclaimed sandy agricultural ecosystem	SRP148524
36	Effects of Manure and Chemical Fertilizer on Bacterial Community Structure and Soil Enzyme Activities in North China	PRJNA706481
37	Differential long-term fertilization alters residue-derived labile organic carbon fractions and microbial community during straw residue decomposition	PRJNA644514
38	Distinct rhizomicrobiota assemblages and plant performance in lettuce grown in soils with different agricultural management histories	PRJNA659405
39	Disentangling the impact of contrasting agricultural management practices on soil microbial communities – Importance of rare bacterial community members	PRJNA659405
40	Reduced tillage, cover crops and organic amendments affect soil microbiota and improve soil health in Uruguayan vegetable farming systems	PRJEB47825
41	Long-term fertilization and tillage regimes have limited effects on structuring bacterial and denitrifier communities in a sandy loam UK soil	PRJNA646084
42	Soil biota shift with land use change from pristine rainforest and Savannah (Cerrado) to agriculture in southern Amazonia	PRJEB49432
43	Microbial community structure is affected by cropping sequences and poultry litter under long-term no-tillage	PRJNA587110
44	Impact of long-term agricultural management practices on soil prokaryotic communities	e-mail from author
45	Long-Term Fertilization Affects Soil Microbiota, Improves Yield and Benefits Soil	SRP137050
46	Predicting soil farming system and attributes based on soil bacterial community	e-mail from author
47	Long-term stability of soil bacterial and fungal community structures revealed in their abundant and rare fractions	PRJNA686771
48	Comparison of Soil Bacterial Communities from Juvenile Maize Plants of a Long-Term Monoculture and a Natural Grassland	e-mail from author
49	Crop cover is more important than rotational diversity for soil multifunctionality and cereal yields in European cropping systems	PRJNA547528
50	High-Resolution Indicators of Soil Microbial Responses to N	PRJEB35080

	Fertilization and Cover Cropping in Corn Monocultures	
51	The Reaction of Cellulolytic and Potentially Cellulolytic Spore-Forming Bacteria to Various Types of Crop Management and Farmyard Manure Fertilization in Bulk Soil	PRJNA809390
52	Differentiation of individual clusters of comammox Nitrospira in an acidic Ultisol following long-term fertilization	PRJNA665141
53	Environmental and Anthropogenic Factors Shape Major Bacterial Community Types Across the Complex Mountain Landscape of Switzerland	DRA011948
54	Nitrogen fertilization directly affects soil bacterial diversity and indirectly affects bacterial community composition	PRJEB8961
55	Inorganic Nitrogen Application Affects Both Taxonomical and Predicted Functional Structure of Wheat Rhizosphere Bacterial Communitiess	PRJEB8961
56	Prokaryotic Community Structure of Long-Term Fertilization Field Andisols in Central Japan	PRJNA454003
57	Microbial Communities in Soils and Endosphere of Solanum tuberosum L. and their Response to Long-Term Fertilization	DRA007565
58	Composition of soil viral and bacterial communities after long-term tillage, fertilization, and cover cropping management	PRJNA645139
59	Long-term high-P fertilizer input decreased the total bacterial diversity but not phoD-harboring bacteria in wheat rhizosphere soil with available-P deficiency	PRJNA744634
60	Long-term effects of maize straw return and manure on the microbial community in cinnamon soil in Northern China using 16S rRNA sequencing	CNP0000424
61	Different impacts of manure and chemical fertilizers on bacterial community structure and antibiotic resistance genes in arable soils	PRJNA588667
62	Long-term no-till increases soil nitrogen mineralization but does not affect optimal corn nitrogen fertilization practices relative to inversion tillage	PRJNA342077
63	Impacts of switching tillage to no-tillage and vice versa on soil structure, enzyme activities and prokaryotic community profiles in Argentinean semi-arid soils	PRJNA826740
64	Long-term land use in Amazon influence the dynamic of microbial communities in soil and rhizosphere	PRJNA644644
65	Liming does not counteract the influence of long-term fertilization on soil bacterial community structure and its co-occurrence pattern	PRJNA764025
66	Long-term nitrogen fertilization, but not short-term tillage reversal, affects bacterial community structure and function in a no-till soil	PRJNA325648
67	Effects of nitrogen and phosphorus addition on microbial community composition and element cycling in a grassland soil	PRJNA325652
68	Variations in the diversity of soil bacterial and archaeal communities in response to different long-term fertilization regimes in maize fields	PRJNA596166
69	Nitrogen leaching greatly impacts bacterial community and denitrifiers	PRJNA663467

	abundance in subsoil under long-term fertilization	
70	Forest-to-agriculture conversion in Amazon drives soil microbial communities and N-cycle	PRJEB32468
71	Microbial Signatures in Fertile Soils Under Long-Term N Management	PRJNA513834
72	Differential Resilience of Soil Microbes and Ecosystem Functions Following Cessation of Long-Term Fertilization	PRJNA771382
73	Variation in Bacterial Community Structure Under Long-Term Fertilization, Tillage, and Cover Cropping in Continuous Cotton Production	PRJNA577961
74	Bacterial indicator taxa in soils under different long-term agricultural management	PRJNA747698
75	Long-term liming promotes drastic changes in the composition of the microbial community in a tropical savanna soil	PRJNA255111
76	Long-Term Nutrient Enrichment of an Oligotroph-Dominated Wetland Increases Bacterial Diversity in Bulk Soils and Plant Rhizospheres	PRJNA647807
77	Long-Term N Fertilization Decreased Diversity and Altered the Composition of Soil Bacterial and Archaeal Communities	PRJNA599142
78	Diversity of archaea and niche preferences among putative ammonia-oxidizing Nitrososphaeria dominating across European arable soils	PRJNA771382
79	Land-use intensification differentially affects bacterial, fungal and protist communities and decreases microbiome network complexity	PRJEB35080
80	Long Term Influence of Fertility and Rotation on Soil Nitrification Potential and Nitrifier Communities	PRJNA741976
81	Long-term fertilizer and crop-rotation treatments differentially affect soil bacterial community structure	e-mail from author
82	Effect of long-term organic and mineral fertilization strategies on rhizosphere microbiota assemblage and performance of lettuce	e-mail from author
83	Impact of Long-Term Manure and Sewage Sludge Application to Soil as Organic Fertilizer on the Incidence of Pathogenic Microorganisms and Antibiotic Resistance Genes	SRP133289
84	Long-Term Nitrogen Fertilization Elevates the Activity and Abundance of Nitrifying and Denitrifying Microbial Communities in an Upland Soil: Implications for Nitrogen Loss From Intensive Agricultural Systems	PRJNA681506
85	Long-term N inputs shape microbial communities more strongly than current-year inputs in soils under 10-year continuous corn cropping	PRJEB26423
86	Long-term nitrogen and phosphorus fertilization reveals that phosphorus limitation shapes the microbial community composition and functions in tropical montane forest soil	PRJNA736632
87	Long-term nitrogen fertilization alters microbial community structure and denitrifier abundance in the deep vadose zone	PRJNA646084
88	The ecological clusters of soil organisms drive the ecosystem multifunctionality under long-term fertilization	PRJEB34566
89	Organic amendments drive shifts in microbial community structure and	PRJNA643413

	keystone taxa which increase C mineralization across aggregate size classes	
90	Nitrogen-dependent bacterial community shifts in root, rhizome and rhizosphere of nutrient-efficient Miscanthus x giganteus from long-term field trials	e-mail from author
91	Long-term effects of nitrogen and phosphorus fertilization on soil microbial community structure and function under continuous wheat production	PRJNA527123
92	Long-term and legacy effects of manure application on soil microbial community composition	PRJNA448773
93	Fertilization changes soil microbiome functioning, especially phagotrophic protists	PRJNA390038
94	Long-Term Compost Amendment Changes Interactions and Specialization in the Soil Bacterial Community, Increasing the Presence of Beneficial N-Cycling Genes in the Soil	PRJNA498197
95	Assessing the long-term impact of urease and nitrification inhibitor use on microbial community composition, diversity and function in grassland soil	PRJEB38121
96	The Bacterial Composition and Diversity in a Eucalyptus pellita Plantation in South Sumatra, Indonesia	PRJNA790732
97	Site-Specific Conditions Change the Response of Bacterial Producers of Soil Structure-Stabilizing Agents Such as Exopolysaccharides and Lipopolysaccharides to Tillage Intensity	PRJNA838114
98	Different traits from the paddy soil and upland soil regulate bacterial community and molecular composition under long-term fertilization regimes	PRJNA555481
99	Chemical nature of soil organic carbon under different long-term fertilization regimes is coupled with changes in the bacterial community composition in a Calcaric Fluvisol	PRJNA591958
100	Beyond the snapshot: identification of the timeless, enduring indicator microbiome informing soil fertility and crop production in alkaline soils	SRP126437
101	Biodiversity of key-stone phylotypes determines crop production in a 4-decade fertilization experiment	PRJNA726588
102	Long-term nitrogen input alters plant and soil bacterial, but not fungal beta diversity in a semiarid grassland	SRP126794
103	Divergent responses of bacterial activity, structure, and co-occurrence patterns to long-term unbalanced fertilization without nitrogen, phosphorus, or potassium in a cultivated vertisol	PRJNA573484