

## ARTICLE

## OPEN



# The neglected role of micronutrients in predicting soil microbial structure

Ziheng Peng <sup>1</sup>, Chunling Liang <sup>1</sup>, Min Gao <sup>1</sup>, Yu Qiu <sup>1</sup>, Yanjing Pan <sup>1</sup>, Hang Gao <sup>1</sup>, Yu Liu <sup>1</sup>, Xiaomeng Li <sup>1</sup>, Gehong Wei <sup>1</sup>✉ and Shuo Jiao <sup>1</sup>✉

Predicting the distribution patterns of soil microbial communities requires consideration of more environmental drivers. The effects of soil micronutrients on composition of microbial communities are largely unknown despite micronutrients closely relating to soil fertility and plant communities. Here we used data from 228 agricultural fields to identify the importance of micronutrients (iron, zinc, copper and manganese) in shaping structure of soil microbial communities (bacteria, fungi and protist) along latitudinal gradient over 3400 km, across diverse edaphic conditions and climatic gradients. We found that micronutrients explained more variations in the structure of microbial communities than macronutrients in maize soils. Moreover, micronutrients, particularly iron and copper, explained a unique percentage of the variation in structure of microbial communities in maize soils even after controlling for climate, soil physicochemical properties and macronutrients, but these effects were stronger for fungi and protist than for bacteria. The ability of micronutrients to predict the structure of soil microbial communities declined greatly in paddy soils. Machine learning approach showed that the addition of micronutrients substantially increased the predictive power by 9–17% in predicting the structure of soil microbial communities with up to 69–78% accuracy. These results highlighted the considerable contributions of soil micronutrients to microbial community structure, and advocated that soil micronutrients should be considered when predicting the structure of microbial communities in a changing world.

*npj Biofilms and Microbiomes* (2022)8:103 ; <https://doi.org/10.1038/s41522-022-00363-3>

## INTRODUCTION

One of the urgent transitions in microbial ecology is progressing from the descriptive records of patterns in microbial composition towards a predictive understanding of the shifts in microbial communities in response to environmental changes<sup>1,2</sup>. The composition of these communities is sensitive to surrounding environmental changes, and such shifts would affect their functioning<sup>3,4</sup>. Understanding how the composition of soil communities respond to environment changes is critical for predicting future changes in ecosystem functioning. Unfortunately, the enormous diversity of soil organisms and the complexity of environmental properties create obstacles for our understanding of the underlying dynamics in microbial communities<sup>5</sup>. Over the past decades, a growing number of survey and experimental studies have focused on the effects of climate, plant, soil properties on soil microbial communities across continental and global scale<sup>6–10</sup>. However, a large fraction of variation in soil microbial community's composition remains unexplained. Unraveling the complexities of the relationship between soil microbial communities and environment requires consideration of more environmental drivers.

In soils, most research has focused on the role of soil pH and macronutrients (e.g., nitrogen, phosphorus and potassium) in explaining the structure of soil microbial communities<sup>11–14</sup>. In contrast, soil metal micronutrients (e.g., Fe, Cu, Mn, Zn) supporting essential biological functions remain poorly investigated in terms of their effects on the structure of soil microbes. Although micronutrients concentrations are as low as milligrams per kilogram or even lower, they are indispensable to plants and play a vital role in cell growth and redox homeostasis<sup>15,16</sup>. They engage in the formation of enzyme-substrate complexes and act as

enzyme cofactors in common biochemical pathways, such as the synthesis of protein and biomolecules<sup>17,18</sup>. This points to a potentially significant role of micronutrients in affecting microbial communities' structure, which had been overlooked in the past. However, comprehensively examining the role of micronutrients in explaining the structure of microbial communities is challenging, not only because the concentrations of many micronutrients have not been systematically investigated, but also because the availability of micronutrients are affected by climate and soil physicochemical properties such as aridity, temperature, pH, organic matter and soil cation exchange capacity<sup>16</sup>. In other words, we have not determined whether micronutrients itself directly shape community structure, or whether other environmental factors that co-vary with soil micronutrients might be indirectly linked with the change in community structure via micronutrients<sup>15</sup>. Therefore, unraveling the unique fraction of the effects of soil micronutrients on microbial communities' structure will advance our understanding of the complex relationships between microbial communities and the environment.

Different groups of soil organisms respond differently to environmental factors. Climatic variables are recently reported to influence the diversity and composition of fungal communities on global scale<sup>10</sup>. The diversity patterns in fungi are found to be decoupled from bacterial patterns in forest ecosystems, with temperature and primary productivity as the major factors in predicting fungal communities but soil properties such as pH and N:P ratio for bacterial communities<sup>19</sup>. Moreover, the largest proportion of the variance in soil bacterial communities reported in recent study is explained by differences in soil chemistry, particularly soil pH, which is confirmed robust across different spatial scales and land-use types<sup>11,12</sup>. Most bacterial taxa exhibit

<sup>1</sup>State Key Laboratory of Crop Stress Biology in Arid Areas, Shaanxi Key Laboratory of Agricultural and Environmental Microbiology, College of Life Sciences, Northwest A&F University, Yangling, Shaanxi 72100, P. R. China. ✉email: [weigehong@nwafu.edu.cn](mailto:weigehong@nwafu.edu.cn); [shuojiao@nwafu.edu.cn](mailto:shuojiao@nwafu.edu.cn)

relatively narrow growth tolerances in soil pH perhaps due to the evolutionary history and life history traits<sup>12</sup>. Protists, which are dominated by consumers, are primarily affected by precipitation<sup>20</sup>. In addition, protistan communities are also associated with bacterial and fungal communities, potentially driven by interactions between them<sup>21,22</sup>. Identifying the relative importance of micronutrients to different groups of soil organisms (such as protists, bacteria and fungi) is crucial to advancing our understanding of soil communities and associated ecosystem functions.

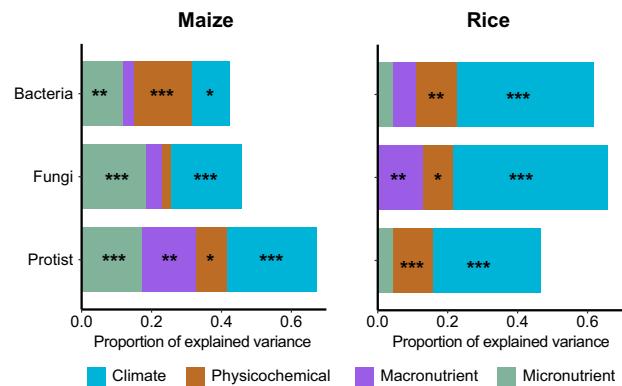
Here, we evaluated the relative contribution of soil micronutrients (total and available iron, zinc, copper, manganese) to the composition (relative abundance of phylotypes) of soil bacterial, fungal and protistan communities across latitudinal gradient over 3400 km. The importance of micronutrients was estimated after accounting for climate, soil properties and macronutrients. We used data from 114 parallel agricultural fields (maize and rice) in eastern China, covering diverse climatic and edaphic gradients. This data contained a set of 30 environmental variables-climate (annual mean temperature and precipitation), soil properties (soil pH, cation exchange capacity, sand, silt, clay, dissolved and total organic matter), macronutrients (total and available nitrogen, phosphorus, potassium and sulfur, nitrate and ammonia nitrogen, C/N, C/P and N/P) and micronutrients (total and available iron, zinc, copper, manganese). These variables were major predictors of soil organisms reported in previous study<sup>9,10,20</sup>. Microbial information on the structure of bacterial, fungal and protistan communities were assessed by 16 S rRNA, ITS and 18 S rRNA gene, respectively. We first tested whether micronutrients could explain a unique portion of the variation in the composition of soil bacterial, fungal and protistan communities after controlling for climate, soil physicochemical properties and macronutrients. Then, we identified the microbial taxa that were driven by micronutrients. Finally, we used machine learning approach to examine whether environmental factors provide markedly better prediction of microbial communities after micronutrients were added.

## RESULTS

### The contribution of micronutrients on microbial communities

We first used partial mantel analysis to quantify the relative contribution of climate, soil physicochemical properties, macronutrients and micronutrients on the structure of bacterial, fungal, and protistan communities. This approach allowed us to estimate the unique contribution of micronutrients on the structure of soil communities by controlling other predictors. Most microbial communities showed strong responses to climate and soil physicochemical properties, explaining 27%, 23%, 34% variance of bacteria, fungi and protists in maize soils, and 51%, 53%, 42% variance in paddy soils, respectively (Fig. 1 and Supplementary Table 1). Most importantly, we found that micronutrients predicted a unique part of the variation in the structure of soil bacterial, fungal and protistan communities in maize soils, accounting for 11%, 18%, and 17% of the variation, respectively (Fig. 1). However, micronutrients had little contribution to predicting the structure of microbial communities in paddy soils, suggesting that land-use change would influence the effects of micronutrients on microbial communities' composition. Macronutrients, with only significant influence on protist in maize soils and fungi in paddy soils, had a lower ability to predict the structure of microbial communities than micronutrients (Fig. 1). This suggested that micronutrients were better predictor of soil microbial structure than macronutrients. Consistent results were observed when we used multiple regression on matrices (Supplementary Fig. 2 and Supplementary Table 1).

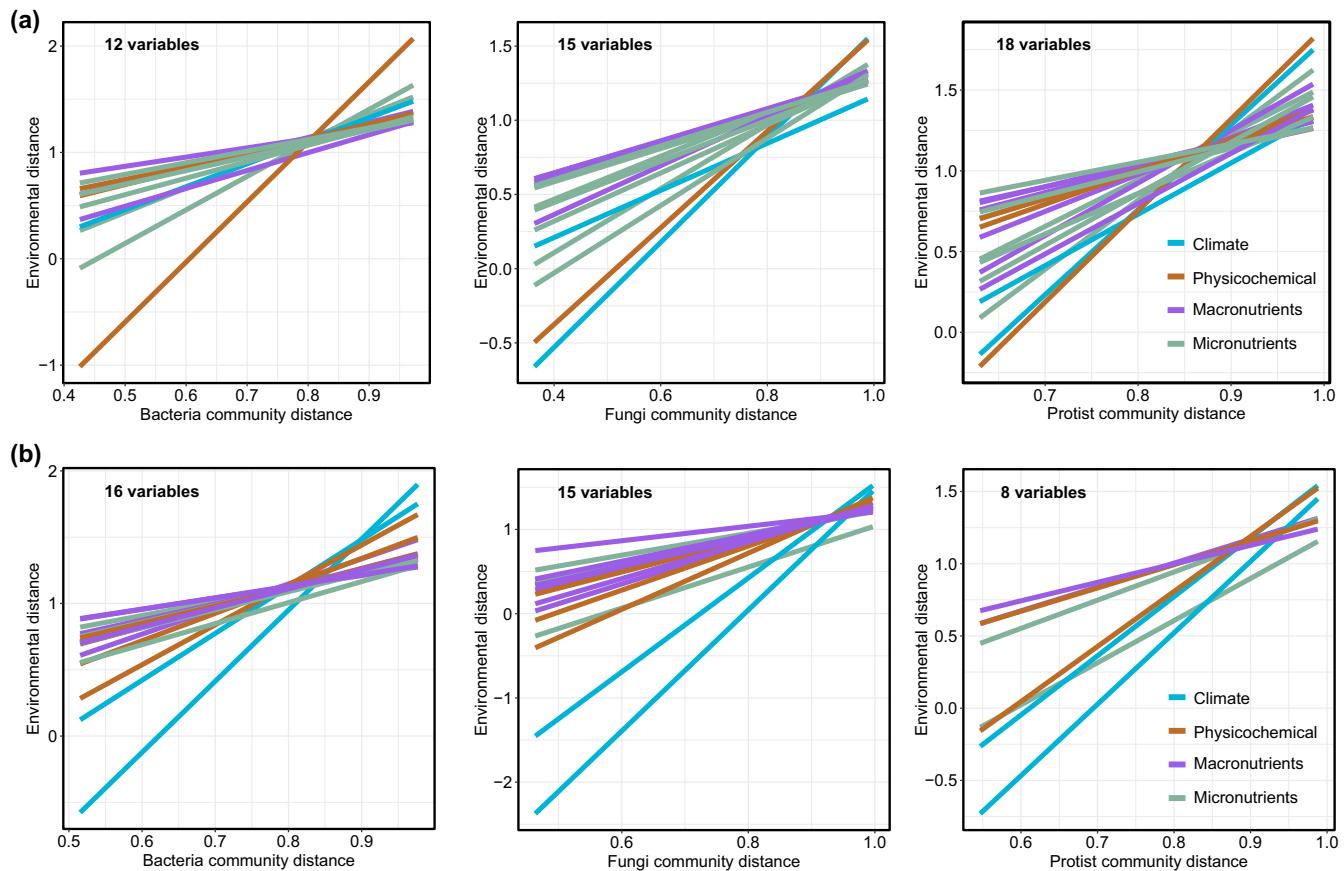
The relationship between the distances in each environmental predictor and microbial communities was then assessed to identify the major environmental factors in structuring bacterial,



**Fig. 1 Explained variance of bacterial, fungal and protistan communities in maize and paddy soils by environmental parameters using partial mantel test analysis.** Input matrices with variables: climate (mean annual temperature, mean annual precipitation), soil physicochemical properties (soil pH, CEC, sand, silt, clay, dissolved and total organic matter), macronutrients (total and available nitrogen, phosphorus, potassium and sulfur, nitrate and ammonia nitrogen) and micronutrients (total and available iron, zinc, copper, manganese). The significance of statistical test was conducted by partial mantel test analysis based on pearson correlation. Similar results were obtained using multiple regression on matrices (Supplementary Fig. 1). \*\*\* $p < 0.001$ ; \*\* $p < 0.01$ ; \* $p < 0.05$ .

fungal and protistan communities. The concentrations of several different micronutrients were found to be correlated with the structure of microbial communities, together with climate and soil properties such as mean annual temperature and precipitation, and soil pH (Fig. 2 and Supplementary Figs. 3–8). Specially, bacterial communities were associated with available iron (Pearson:  $R = 0.258$ ,  $P < 0.001$ ), total copper (Pearson:  $R = 0.101$ ,  $P < 0.05$ ), total zinc (Pearson:  $R = 0.091$ ,  $P < 0.05$ ), total manganese (Pearson:  $R = 0.120$ ,  $P < 0.01$ ), and available manganese (Pearson:  $R = 0.197$ ,  $P < 0.001$ ) in maize soils. Fungal communities were associated with total iron (Pearson:  $R = 0.159$ ,  $P < 0.01$ ), available iron (Pearson:  $R = 0.168$ ,  $P < 0.001$ ), total copper (Pearson:  $R = 0.105$ ,  $P < 0.05$ ), available copper (Pearson:  $R = 0.086$ ,  $P < 0.05$ ), total zinc (Pearson:  $R = 0.090$ ,  $P < 0.05$ ), total manganese (Pearson:  $R = 0.120$ ,  $P < 0.01$ ), and available manganese (Pearson:  $R = 0.118$ ,  $P < 0.01$ ) in maize soils. Protistan communities were associated with total iron (Pearson:  $R = 0.141$ ,  $P < 0.01$ ), available iron (Pearson:  $R = 0.269$ ,  $P < 0.001$ ), total copper (Pearson:  $R = 0.089$ ,  $P < 0.05$ ), available copper (Pearson:  $R = 0.104$ ,  $P < 0.05$ ), total manganese (Pearson:  $R = 0.190$ ,  $P < 0.01$ ), and available manganese (Pearson:  $R = 0.189$ ,  $P < 0.001$ ) in maize soils. These results suggested the considerable role of these micronutrients in explaining microbial structure in maize soils. Moreover, available iron showed the strongest relationship with microbial communities among all micronutrients in maize soils (Supplementary Table 2), suggesting the potential importance of iron to influence microbial community structure. For paddy soils, available zinc was correlated with the structure of bacterial, fungal and protistan communities (Supplementary Table 3).

We further used structural equation modeling to disentangle the direct effect of micronutrients and indirect role of other environmental predictors via micronutrients on the changes in microbial communities. To rigorously determine the unique importance of micronutrients, SEM model was constructed representing the effects of different environmental variables in three steps (Supplementary Fig. 9). Soil physicochemical and climatic properties were reported as main predictors for the structure of bacterial, fungal and protistan communities, and thus their effects were tested first as SEM model I. In addition, the



**Fig. 2 Relationships between soil bacterial, fungal, protistan communities and environmental distance in maize and paddy soils.** **a** maize soils, **(b)** paddy soils. Differences among environment and communities were estimated using mantel test base on Bray–Curtis distances. Lines represent fitted linear regressions between environmental distance and associated bacterial, fungal or protistan communities. Only environmental factors with significant relationship with the communities are shown in figure (the number in the upper left). A full version of this figure showing each environmental factor can be found in Supplementary Figs. 2–7 and Supplementary Tables 2–4.

influence of macronutrients (e.g., N, P, K, S) on microbial communities have also been widely studied and considered to be important predictors, and their effects were therefore added as SEM model II. Finally, micronutrients were added into SEM model III. To match the SEM model, we used non-metric multidimensional scaling (NMDS) to transform ASVs abundance data into low-dimensional space and kept the first two axes for bacterial, fungal and protistan communities, respectively. The micronutrients were correlated with the NMDS axis scores of bacterial, fungal, and protistan communities (Supplementary Fig. 10).

Our final SEM analyses showed that micronutrients had direct effects on the first and second axis of bacteria (iron and copper), the first and second axis of fungi (iron and copper), and the first axis of protists communities (iron) in maize soils after accounting for other predicting factors simultaneously, with the sole exception of protistan second axis (Fig. 3). These results indicated that micronutrients explained a unique proportion of the variation in the structure of microbial communities, although climate and soil physicochemical properties showed the largest total effects on microbial structure. Moreover, climate and soil physicochemical properties also indirectly affected the microbial communities through their effects on the composition and concentration of micronutrients, suggesting the mediating role of micronutrients between climate, soil and microbial structure (Fig. 3). For example, indirect effects of soil pH on microbial structure were largely driven by micronutrients such as iron (negative and positive effect of pH in bacterial and fungal axis one, respectively). Climate such as MAP indirectly drove microbial communities' structure through its effect on soil micronutrient concentrations (positive effects on

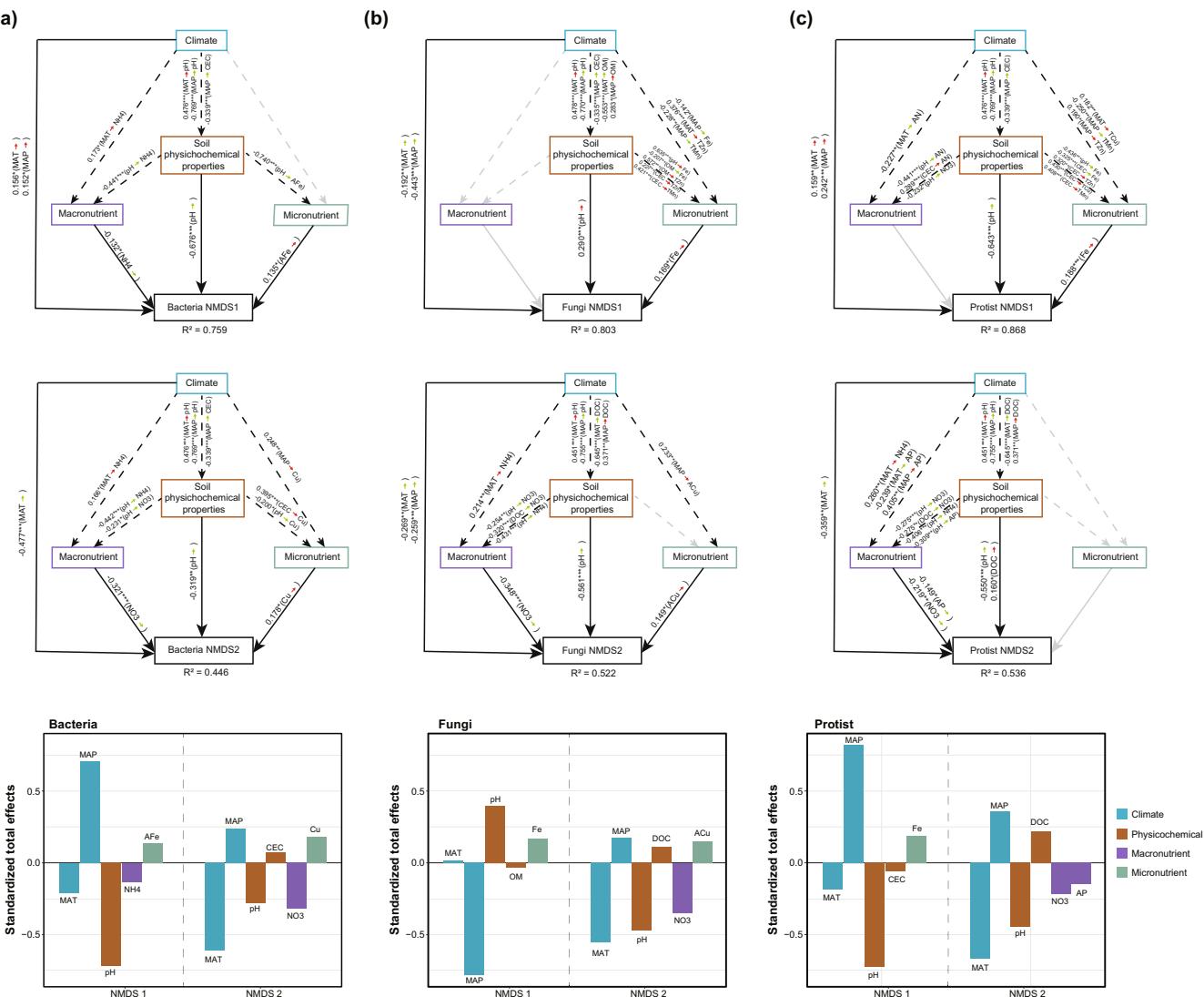
copper in bacterial and fungal second axis). When SEM analyses were repeated using data from paddy soils (Supplementary Fig. 11), we found that a unique part of the variance of microbial communities was still predicted by micronutrients. However, micronutrients had a lower ability to predict the structure of microbial communities in paddy soils than that in maize soils (two of six axis scores), with iron for fungal axis one and manganese for protistan second axis.

#### Microbial phylotypes influenced by micronutrients

Random forest analyses were conducted to identify microbial phylotypes at ASV (Supplementary Figs. 12 and 13) and genera (Fig. 4 and Supplementary Figs. 14–20) levels that were primarily affected by micronutrients. For example, genus *Gibberella* as plant pathogens was found to be an indicator of changes in micronutrients and the relative abundance decreased with the concentrations of available iron, zinc, copper and manganese (Supplementary Fig. 21). Of the environmental factors, micronutrients accounted for more than 25% of variability over half of the phylotypes (Fig. 4). Across bacterial, fungal and protistan communities, the random forest models were sensitive to multiple micronutrients abundance, such as available manganese for bacteria and protist in maize soils and available copper for fungi in paddy soils (Supplementary Fig. 22).

#### The importance of micronutrients by machine learning

To validate importance of micronutrients, we further used machine learning approach to evaluate whether the inclusion of



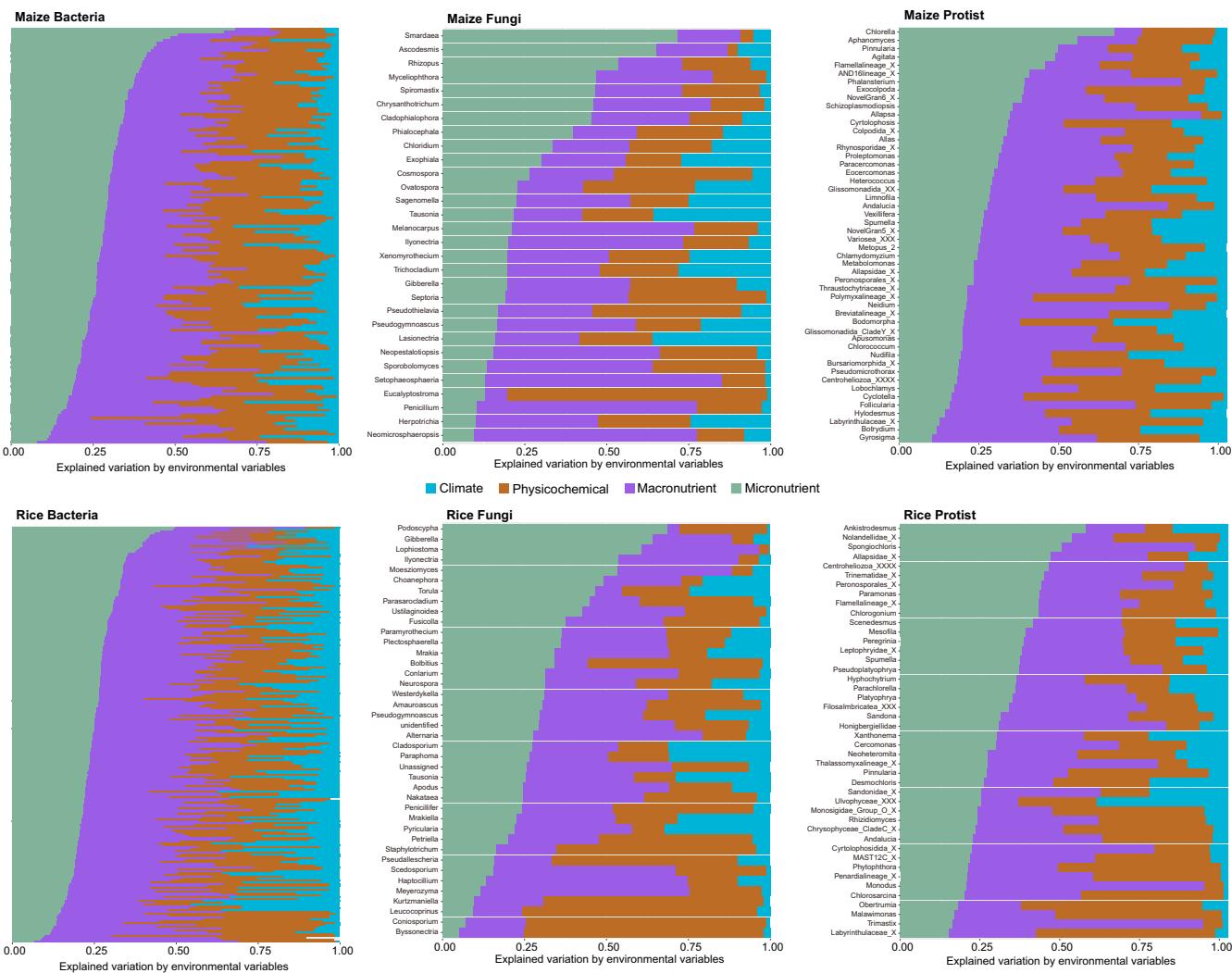
**Fig. 3 Structural equation models identifying the direct (full lines) and indirect (dotted lines) influence of different predictors on the structure of bacteria, fungi and protistan communities in maize soils. a** bacteria, **(b)** fungi, **(c)**, protist. Black lines indicate significant and numbers adjacent to lines were indicative of the effect size of the relationship. We grouped the climatic and edaphic properties into the same box in the model for graphical simplicity, which did not represent latent variables.  $R^2$  denotes the proportion of variance explained. Red arrows represented positive paths, and green arrows represented negative paths. Significance levels were as follows:  $*p < 0.05$ ,  $**p < 0.01$ , and  $***p < 0.001$ . The total standardized effects on SEM on the structure of microbial communities were calculated as sum of the direct and indirect effects. Information about our a priori model was provided in Supplementary Fig. 8. TZN total zinc, AZn available zinc, TFe total iron, AFe available iron, TCu total copper, ACu available copper, TMn total manganese, CEC cation exchange capacity.

micronutrients could provide a better, higher accuracy in predicting the structure of microbial communities in agricultural fields (Fig. 5 and Supplementary Fig. 23). Machine learning (ML) algorithms provided powerful tools for explaining multivariate, nonlinear, and non-monotonic relationships<sup>23,24</sup>. Ten classic ML algorithms - Bagged Regression Tree (BaRT), Cubist, Fast Nearest Neighbor (FNN), Gradient Boosting Machines (GBM), Weighted k-Nearest Neighbor (KKNN), Kernel Support Vector Machine (KSVM), Random Forest (RF), Ranger, Rpart and Support Vector Machine (SVM) were employed to assess the prediction performance of microbial communities' models. As expected, we found that the models including micronutrients had a lower mean absolute error (median of 0.483), mean squared error (median of 0.415) and root mean squared error (median of 0.633) than those obtained from the model excluding micronutrients in predicting the structure of bacterial, fungal and protistan communities (Fig. 5 and Supplementary Tables 4–6), suggesting that the addition of

micronutrients substantially increased the predictive power of environmental variables. Moreover, ensemble machine learning methods such as BaRT and RF showed a lower mean absolute error (BaRT: 0.445; RF: 0.446) and root mean squared error (BaRT: 0.593; RF: 0.584), yielding better prediction accuracy compared to other methods. The observed vs. predicted the structure of bacterial, fungal and protistan communities in test dataset matched well for RF and BaRT with values aligned close to the 1:1 line in models including micronutrients than those values in models excluding micronutrients (Fig. 5).

## DISCUSSION

Together, our results provided empirical evidence that micronutrients could leave a considerable effect on the structure of soil bacterial, fungal and protistan communities. Our work highlighted that micronutrients explained a unique and significant portion in

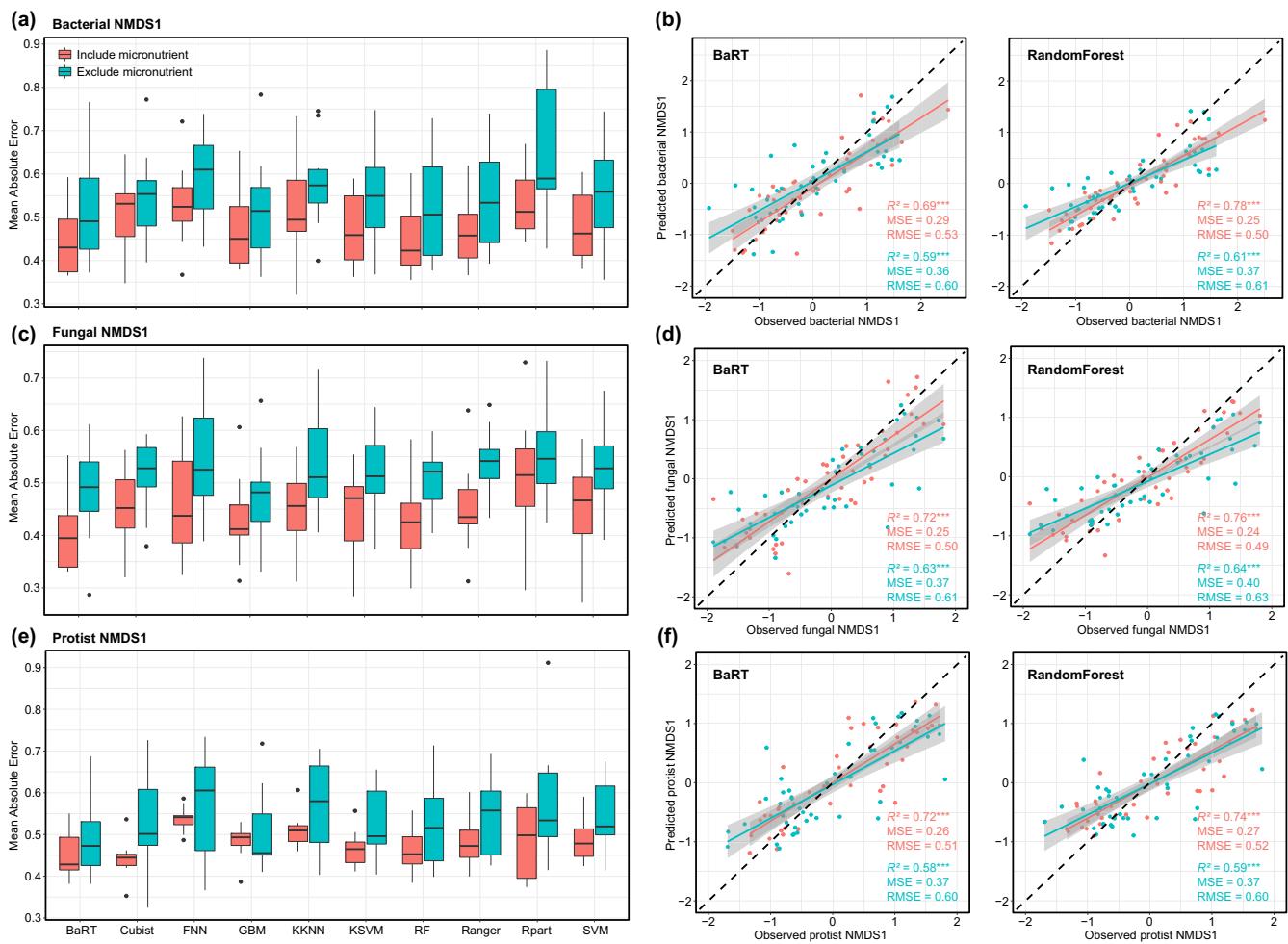


**Fig. 4 Environmental variables explaining the structure of bacterial, fungal and protistan communities.** Contribution of climate, soil physicochemical properties, macronutrients and micronutrients to the variation explained by the complete random forest model for each microbial genus where out-of-bag  $R^2 > 0$ .

variation in soil microbial communities beyond what could be predicted from climate, soil physicochemical properties or macronutrients. These results emphasized the fact that micronutrients could be used to gain a better and more accurate understanding and predicting in the composition of soil communities, which was helpful to understanding mechanisms of microbial communities to the surrounding environment, including micronutrients gain and loss, and fertilization strategies. For example, our results suggested that concentrations of iron and copper might directly affect bacterial, fungal and protistan communities independent of the climate, soil physicochemical properties and macronutrients in maize soils. This result was consistent with previous study demonstrating that iron and its complexes affected microbial community structure, with increasing bacterial interactions and the number of mutually beneficial taxa in soils<sup>25</sup>. Moreover, experimental studies dedicating to the study of organic fertilizers mainly focused on the effects of organic carbon and macronutrients (N, P and K) on microbial communities while ignoring micronutrients<sup>26,27</sup>. However, the applications of organic fertilizers were also linked with soil micronutrients status in soils<sup>28</sup>. A recent global grassland study showed that nitrogen and phosphorus fertilization favored fungal pathogens<sup>29</sup>, in contrast to our study that Gibberella decreased with abundance of micronutrients. These results suggest that the effects of organic

fertilizers on microbial communities are complex, and the combined effects of both macro- and micronutrients should be considered.

In the present study, the structure of fungal and protistan communities showed stronger relationships with micronutrients than the bacterial communities, which might attribute to physiological and ecological differences among these groups. Bacterial communities composition was reported to be primarily determined by soil physicochemical properties such as soil pH<sup>11</sup>, consistent with our results. The effects of other environmental variables on bacterial communities were far weaker than those of soil properties. In addition, the concentration of available iron was closely linked with bacterial communities, which might be associated with ubiquitous iron-reducing and oxidizing bacteria that determined bioavailability of iron<sup>30</sup>. Fungal communities are key decomposers of plant necromass and symbionts of plant growth involving soil micronutrient cycles, and tend to be largely driven by plant communities and the climate associated with plant distribution<sup>10,31</sup>. Arbuscular mycorrhizal fungi increases ferric chelate reductase activity as well as Fe, Zn, S and P in plant under Fe-deficiency, which is related to the bioavailability of Fe at rhizosphere zone<sup>32</sup>. Fungi are more effective than acidophilic autotrophic and heterotrophic bacteria for bioremediation of heavy metals (Zn and Cu) in sediments<sup>33</sup>. Unlike bacteria and



**Fig. 5 Comparison of prediction errors of different machine learning algorithms for the structure of bacterial, fungal and protistan communities.** Machine learning algorithms include Bagged Regression Tree (BaRT), Cubist, Fast Nearest Neighbor (FNN), Gradient Boosting Machines (GBM), Weighted k-Nearest Neighbor (KKNN), Kernel Support Vector Machine (KSVM), Random Forest (RF), Ranger, Rpart and Support Vector Machine (SVM), showing bacterial (a), fungal (c), and protistan communities (e). The center line of boxplot represents the median of data. The boxplot bounds the interquartile range (IQR) divided by the median, and whiskers extend to a maximum of 1.5 times the IQR beyond the box. Other observed data points outside the boundary of the whiskers are plotted as outliers, shown as dot. Predicted vs. observed the structure of bacterial (b), fungal (d) and protistan (f) communities of test dataset derived from bagged regression tree (BaRT) and random forest (RF).  $R^2$ , coefficient of determination; MSE, mean squared error; RMSE, root mean squared error. The significance of statistical test was conducted by BaRT and RandomForest regression models. \* $p < 0.05$ , \*\* $p < 0.01$ , and \*\*\* $p < 0.001$ .

fungi, the composition of protists is dominated by consumers and is found to be jointly structured by climate, soil physicochemical properties, macronutrients, micronutrients and its prey<sup>20,34</sup>.

As reported, palaeoclimate, current climate, soil development, soil chemistry, topography, and vegetation account for a large variation in the structure of microbial communities<sup>9,10,35–37</sup>. However, micronutrients were demonstrated to explain a unique and significant part of variation in our study. In addition, climate and soil properties could also indirectly alter microbial communities' structure by regulating micronutrients, suggesting that climate and soil properties cannot account solely (via direct effects) for soil microbial communities. For example, the SEM models showed that, besides the direct effect, soil properties had an indirect effect on fungal and protistan communities through their negative effect on soil iron concentrations. Linkages of micronutrients to microbial communities advance our understanding of the pathways by which climatic and soil properties affect microbial communities and how microbial communities respond to ongoing global changes. Nevertheless, there was still a large proportion of unexplained variations in microbial communities structure, which might be attributed to stochastic processes

such as homogenizing dispersal and ecological drift and unconsidered factors such as crop genotypes<sup>38</sup>.

Moreover, micronutrients had a lower ability to predict the structure of microbial communities in paddy soils than in maize soils. This suggested that different agricultural practices and crop types had distinct effects on the unique contribution of micronutrients as a predictor of microbial structure. Compared to maize soils, paddy fields are subject to more agricultural disturbances, such as experiencing more human-driven physical practices and suffering larger fluctuations in soil moisture. Soil moisture is one of major drivers on microbial communities structure<sup>39,40</sup>, and drought-rewetting events might attenuate the direct contribution of micronutrients on soil microbial communities by directionally promoting moisture-insensitive taxa<sup>41,42</sup>.

Overall, our work provides fundamental insights into the complexities of relationships between environment and soil microbial communities, and quantifies the relative contribution of micronutrients in predicting the structure of bacterial, fungal and protistan communities. We show that soil micronutrients explain a unique proportion of the variation in microbial communities' structure even after controlling for climate, soil

physicochemical properties and macronutrients. Our results from machine learning indicate that the inclusion of micronutrients in the model increases our ability to predict the structure of soil microbial communities. Thus, given the important role of micronutrients in predicting microbial communities, information on micronutrients is essential to fully unravel the complexity relationships between environment and microbial communities and the impact of global changes on belowground communities.

## METHODS

### Data collection and soil analysis

We used data from agricultural fields under long-term cultivation with maize and paddy soils across eastern China<sup>43,44</sup>. The dataset included 114 paired sites, amounting to 114 maize and 114 paddy soil samples along a wide latitudinal gradient over 3400 km, across diverse edaphic conditions and climate gradient. Collected paddy soils are around maize soils (within 5 km) in each paired site (Supplementary Fig. 1). The soil samples were collected during the growing season (July–September 2017). In each site, a composite soil sample were composed of 15 soil cores taken at a depth of 0–15 cm from three plots (each 100 m<sup>2</sup>) and was separated into two portions. One portion was air-dried and used for soil physicochemical properties, macronutrients and micronutrients analyses, and the other was stored at –80 °C until DNA extraction.

For all soil samples, we measured soil physicochemical properties (soil pH, CEC, sand, silt, clay, dissolved and total organic matter), macronutrients (total and available nitrogen, phosphorus, potassium and sulfur, nitrate and ammonia nitrogen, C/N, C/P and N/P) and micronutrients (total and available iron, zinc, copper, manganese). Soil properties were determined using standardized protocols<sup>45,46</sup>. Briefly, soil texture was determined using a simple sifting and sedimentation-based approach<sup>47</sup>, and the pH of the soil was assessed in a 1:5 suspension (soil to distilled water). Organic matter was determined colorimetrically following oxidation with a combination of potassium dichromate and sulfuric acid<sup>48</sup>. Dissolved organic carbon (DOC) was measured by high-temperature combustion, the high temperature (110 °C) conversion of inorganic carbon to dissolved CO<sub>2</sub>, and purging this from the sample, the remaining (organic) carbon is then oxidized at a high temperature to CO<sub>2</sub> which can be detected by the instrument's nondispersive infrared (NDIR) sensor<sup>49</sup>. Cation exchange capacity (CEC) of the soil was measured by a modified ammonium-acetate compulsory displacement<sup>50</sup>. Total nitrogen (TN) was determined using a Flash 2000 NC Analyzers (Thermo Scientific, MA, USA). Ammonium (NH<sub>4</sub><sup>+</sup>) and nitrate (NO<sub>3</sub><sup>–</sup>) concentrations in extracts were determined colorimetrically by automated segmented flow analysis (AAIII; Bran and Luebbe, Germany). Determination of total phosphorus (TP) and total potassium (TK) were used potassium dichromate-sulfuric acid digestion and ammonium acetate-flame photometry method, respectively<sup>51</sup>. Available phosphorus (AP) was extracted by 0.5 M NaHCO<sub>3</sub> and determined using the molybdenum blue method. Available potassium (AK) was determined in 1 M ammonium acetate extracts by flame photometry (FP640, INASA, China). After digestion with HNO<sub>3</sub>/H<sub>2</sub>O<sub>2</sub> mixture and extraction with DTPA<sup>52</sup>, we measured total and available metal concentrations using inductively coupled plasma atomic emission spectroscopy (Iris Intrepid II XDL; Thermo Fisher Scientific). Because the DTPA method is widely utilized in neutral to basic soils, as is the case for most of the soils investigated, and the DTPA extraction for copper, iron, manganese, and zinc is standardized, so we used it as an indication of the accessible pool of metals in soils and limited our analysis to them. We obtained climatic data including mean annual temperature (MAT) and mean annual precipitation (MAP) based on the sampling sites from the Worldclim database ([www.worldclim.org](http://www.worldclim.org)).

### Molecular analysis

Soil bacterial, fungal and protistan communities were analyzed using high-throughput sequencing. Total genomic DNA was extracted from soil samples using a FastDNA SPIN Kit for Soil (MP Biochemicals, Solon, OH, USA). Microbial communities were profiled by targeting the V4-V5 region of 16S rRNA gene for bacteria, a region of the internal transcribed spacer 1 gene for fungi, and the V4 region of 18S rRNA in protists. Corresponding polymerase chain reaction assays were performed with the 515 F/907 R (~450 bp)<sup>53</sup>, ITS5-1737F/ITS2-2043R (~300 bp)<sup>54</sup> and TAREuk454FWD1/TAREuk-REV3 (~380 bp)<sup>55</sup> primer pairs, respectively<sup>44</sup>. PCR amplification was performed in a 50 µl volume: 25 µl 2x Premix Taq (Takara Biotechnology, Dalian Co. Ltd, China), 1 µl each primer (10 µM) and 3 µl DNA (20 ng/µl) template. The PCR thermal cycling conditions were performed by thermocycling: 5 min at 94 °C for initialization, followed by 30 cycles of 30 s denaturation at 94 °C, 30 s annealing at 52 °C, 30 s extension at 72 °C, and 10 min final elongation at 72 °C. The length and concentration of the PCR product were detected by 1% agarose gel electrophoresis. Sequencing libraries were generated using NEBNext® Ultra™ II DNA Library Prep Kit for Illumina® (New England Biolabs, MA, USA) following manufacturer's recommendations and index codes were added. Sequencing was performed on the Illumina HiSeq2500 platform (Illumina Inc, San Diego, CA, USA). Bioinformatics processing, including filtering, dereplication, sample inference, chimera identification, and merging of paired-end reads, was performed using the Divisive Amplicon Denoising Algorithm 2 (DADA2)<sup>56</sup>, a model-based approach for correcting Illumina amplicon errors without constructing OTUs. Compared to other methods, DADA2 identified more real variants and output fewer spurious sequences. Taxonomical annotation of the representative sequences of amplicon sequence variants (ASVs) was performed with a naïve Bayesian classifier using the Silva v. 138 for bacteria<sup>57</sup>, UNITE + INSD v. 8.3 for fungi<sup>58</sup>, and Protist Ribosomal Reference database v. 4.13.0 for protist<sup>59</sup>. Bacteria and protists were defined as all prokaryotic and eukaryotic taxa, except Archaea, chloroplasts, mitochondria in 16S gene and Rhodophyta, Streptophyta, Metazoa and fungi in 18S gene for subsequent analyses<sup>20,60</sup>.

### Correlation and regression analyses

We first examined the relationship between environmental variables and the bacterial, fungal and protistan communities through partial Mantel test and multiple regression on distance matrices using the "ecodist" package in R<sup>61,62</sup>, as shown Supplementary Table 1. To quantify dissimilarity in bacterial, fungal and protistan communities across sampling sites, dissimilarity matrices based on Bray–Curtis dissimilarities of ASVs were calculated. Here we used four environmental explanatory matrices: climate variables (annual mean temperature and precipitation), (2) soil physicochemical properties (soil pH, cation exchange capacity, sand, silt, clay, dissolved and total organic matter), (3) macronutrients (total and available nitrogen, phosphorus, potassium and sulfur, nitrate and ammonia nitrogen) and (4) micronutrients (total and available iron, zinc, copper, manganese). To quantify dissimilarity in climate, soil properties, macronutrients and micronutrients across sampling sites, we calculated Euclidean dissimilarities. Four distance matrices were calculated that represent climate, soil properties, macronutrients and micronutrients. Each environmental matrix was individually associated with the microbial communities based on Pearson's correlations, leaving three environmental variables as controlling distance matrix in partial Mantel test, as shown in Fig. 1. Multiple regression analysis was performed on the four distance matrices and microbial communities, as shown Supplementary Fig. 2. Mantel test was further used to assess the significance of the association between each environmental predictor of 30 variables and bacterial, fungal and protistan communities, respectively, as shown in (Fig. 2).

## SEM analyses

We used structural equation modeling that synthesized prior knowledge in model to disentangle the direct and indirect role of the complex relationships among climate, soil physicochemical properties, macronutrients, micronutrients and microbial community. Structural equation modeling allowed us to partition causal influences among multiple predictors and separated the direct and indirect effects of model predictors on the structure of microbial community. Variables were log-transformed and standardized prior to improve normality and linearity.

To rigorously evaluate the relative effects of micronutrients on the structure of microbial community, three SEMs representing the effects of different variables were constructed stepwise (Supplementary Fig. 9), as shown in previous study that micronutrients have important role on global biomass production<sup>15</sup>. First, climatic and soil physicochemical properties were incorporated in model and tested, as temperature, precipitation and soil properties were expected to be the important drivers of the composition of microbial communities<sup>11,63</sup>. Second, the effects of macronutrients (such as N, P, and K) on microbial communities had also been extensively studied and were thus incorporated into the model<sup>64,65</sup>. Finally, we tested whether micronutrients explained additional variation in microbial communities due to their important role in life chemistry and ecosystem functioning<sup>16,66</sup>. Specifically, MAT and MAP were firstly used to build the model together with soil physicochemical properties SOM, DOC, CEC, pH and texture. Each of the soil physicochemical properties was added separately to the model including climatic factors. All those that significantly contributed ( $p < 0.05$ ) to explaining additional variation were retained, as following a similar approach in<sup>15,67</sup>. In the following steps, nitrogen, phosphorus, potassium and sulfur were separately added to the previous model and those that had a significant contribution were retained. The same procedure was applied in the next step for micronutrients (iron, zinc, copper, and manganese). When both total and available metal were retained, the metal was grouped into one composite variable by summing the product of total and available metal with their coefficient in the full SEM model. The model was then reconstructed substituting the individual metal with the composite variable. We grouped the different categories of predictors (climate, soil physicochemical properties, macronutrients and micronutrients) in the same box in the model for graphical simplicity, but these boxes did not represent latent variables.

Three metrics were used to quantify the goodness of fit of SEM models: the  $\chi^2$  test, the Root Mean Square Error of Approximation (RMSEA), and Comparative Fit Index (CFI). Specially, the closer to 1 CFI value, closer to 0 RMSEA values, and higher  $\chi^2$  and RMSEA P values, the better model perform. With a good model fit, we were able to interpret the path coefficients of the model and their associated P values. A path coefficient was analogous to the partial correlation coefficient, and described the strength and sign of the relationship between two variables. Structural equation models were constructed using the "lavaan" package in R<sup>68</sup>.

## Randomforest modeling

To explore the environmental predictors of microbial phylotypes and genera, we modeled distribution of each phylotype or genus by Random Forest<sup>69</sup>. Many of the predictor variables used in ecological studies were either naturally (e.g., the availability of metal with total metal) or functionally (e.g., soil moisture was calculated as a function of precipitation and evapotranspiration) correlated. While some of these predictors may determine species distribution or abundance other collinear predictors may not. The random subset approach for fitting predictor variables at each node could result in a correlated but less influential predictor standing in for more highly influential predictors in the early splits of an individual tree depending upon which predictor is

selected in the subset. The overall predictive ability of the forest for each phylotype or genus was calculated as the average proportion of out-of-bag data variance explained by the fitted forest. In model, the accuracy importance of predictor was quantified as the decrease in performance when each predictor was randomly permuted but other predictors are not modified. To show variable importance across all modeled phylotypes or genera, the relative importance of each predictor was calculated as a sum of predictor relative importances of all Random Forests for individual phylotype or genus weighted by Random Forest predictive ability (out-of-bag  $R^2$ )<sup>10,70,71</sup>. Monotonic function of each predictor that delineates where the compositional change points occurred along the gradient of the predictor was obtained.

## Modeling techniques

Machine learning (ML) algorithms were powerful tools for explaining multivariate, nonlinear, and non-monotonic relationships. We further used machine learning approach to test whether environmental factors provide markedly better prediction accuracies of microbial communities when micronutrients were added. Ten classic ML algorithms (i.e., Bagged Regression Tree (BaRT), Cubist, Fast Nearest Neighbor (FNN), Gradient Boosting Machines (GBM), Weighted k-Nearest Neighbor (KKNN), Kernel Support Vector Machine (KSVM), Random Forest (RF), Ranger, Rpart and Support Vector Machine (SVM)) were selected for regression prediction to assess the prediction performance of microbial communities' models. Our aim was to assess whether environmental factors including micronutrients provided better prediction accuracies than environmental factors excluding micronutrients, and thus environment factors was divided into two group: including micronutrients (climate, soil physicochemical properties, macronutrients and micronutrients) and excluding micronutrients. Five-fold cross-validation was calculated to estimate test errors of the models. For each model, 80% of the dataset was randomly separated into a training set, with the remaining 20% used as a testing set. Three performance parameters were calculated simultaneously to evaluate the machine learning models, including mean absolute error (MAE), mean squared error (MSE), and root mean squared error (RMSE) obtained from the test data set. Lower value represented better the prediction performance of models. Machine learning methods were conducted using the "mlr3" package in R<sup>72</sup>.

## Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this article.

## DATA AVAILABILITY

Environmental data are archived in <https://doi.org/10.6084/m9.figshare.19765978.v1>. Sequencing data are accessible on GenBank via BioProject accession PRJNA544819.

## CODE AVAILABILITY

R code used for data processing, graph creation, and statistical analyses related to this study are available on GitHub at <https://github.com/Pong2021/The-neglected-role-of-micronutrients-in-predicting-soil-microbial-structure>.

Received: 27 June 2022; Accepted: 29 November 2022;

Published online: 27 December 2022

## REFERENCES

1. Antwis, R. E. et al. Fifty important research questions in microbial ecology. *FEMS Microbiology Ecology* **93**, fix044 (2017).
2. Tecon, R. et al. Bridging the Holistic-Reductionist Divide in Microbial Ecology. *mSystems* **4**, e00265–00218 (2019).

3. Boetius, A. Global change microbiology—big questions about small life for our future. *Nat. Rev. Microbiol.* **17**, 331–332 (2019).
4. Cavicchioli, R. et al. Scientists' warning to humanity: microorganisms and climate change. *Nat. Rev. Microbiol.* **17**, 569–586 (2019).
5. Crowther, T. W. et al. The global soil community and its influence on biogeochemistry. *Science* **365**, eaav0550 (2019).
6. Tedersoo, L. et al. Global diversity and geography of soil fungi. *Science* **346**, 1078–+ (2014).
7. Davison, J. et al. Global assessment of arbuscular mycorrhizal fungus diversity reveals very low endemism. *Science* **349**, 970 (2015).
8. Bahram, M. et al. Structure and function of the global topsoil microbiome. *Nature* **560**, 233–237 (2018).
9. Delgado-Baquerizo, M. et al. Palaeoclimate explains a unique proportion of the global variation in soil bacterial communities. *Nat. Ecol. Evol.* **1**, 1339–1347 (2017).
10. Větrovský, T. et al. A meta-analysis of global fungal distribution reveals climate-driven patterns. *Nat. Commun.* **10**, 5142 (2019).
11. Lauber Christian, L., Hamady, M., Knight, R. & Fierer, N. Pyrosequencing-based assessment of soil pH as a predictor of soil bacterial community structure at the continental scale. *Appl. Environ. Microbiol.* **75**, 5111–5120 (2009).
12. Rousk, J. et al. Soil bacterial and fungal communities across a pH gradient in an arable soil. *ISME J.* **4**, 1340–1351 (2010).
13. Dai, Z. et al. Long-term nutrient inputs shift soil microbial functional profiles of phosphorus cycling in diverse agroecosystems. *ISME J.* **14**, 757–770 (2020).
14. Fan, K. et al. Microbial resistance promotes plant production in a four-decade nutrient fertilization experiment. *Soil Biol. Biochem.* **141**, 107679 (2020).
15. Radujković, D. et al. Soil properties as key predictors of global grassland production: Have we overlooked micronutrients? *Ecol. Lett.* **24**, 2713–2725 (2021).
16. Moreno-Jiménez, E. et al. Aridity and reduced soil micronutrient availability in global drylands. *Nat. Sustainability* **2**, 371–377 (2019).
17. Fageria, N., Baligar, V. & Clark, R. Micronutrients in crop production. *Adv. Agron.* **77**, 185–268 (2002).
18. Broadley, M., Brown, P., Cakmak, I., Rengel, Z. & Zhao, F. Function of nutrients: micronutrients. *Marschner's mineral nutrition of higher plants* 191–248 (Elsevier, 2012).
19. Liu, S. et al. Decoupled diversity patterns in bacteria and fungi across continental forest ecosystems. *Soil Biol. Biochem.* **144**, 107763 (2020).
20. Oliverio, A. M. et al. The global-scale distributions of soil protists and their contributions to belowground systems. *Sci. Adv.* **6**, eaax8787 (2020).
21. Crowther, T. W., Boddy, L. & Jones, T. H. Species-specific effects of soil fauna on fungal foraging and decomposition. *Oecologia* **167**, 535–545 (2011).
22. De Gruyter, J. et al. Patterns of local, intercontinental and interseasonal variation of soil bacterial and eukaryotic microbial communities. *FEMS Microbiol. Ecol.* **96**, fiaa018 (2020).
23. Li, H. et al. Decipher soil organic carbon dynamics and driving forces across China using machine learning. *Glob. Change Biol.* **28**, 3394–3410 (2022).
24. Keskin, H., Grunwald, S. & Harris, W. G. Digital mapping of soil carbon fractions with machine learning. *Geoderma* **339**, 40–58 (2019).
25. Zhang, C. et al. Fe(II) addition drives soil bacterial co-occurrence patterns and functions mediated by anaerobic and chemoautotrophic taxa. *Microorganisms* **10**, 547 (2022).
26. Liu, J. et al. Long-term organic fertilizer substitution increases rice yield by improving soil properties and regulating soil bacteria. *Geoderma* **404**, 115287 (2021).
27. Zhong, W. et al. The effects of mineral fertilizer and organic manure on soil microbial community and diversity. *Plant Soil* **326**, 511–522 (2010).
28. Li, B. Y. et al. Soil micronutrient availability to crops as affected by long-term inorganic and organic fertilizer applications. *Soil Tillage Res.* **96**, 166–173 (2007).
29. Lekberg, Y. et al. Nitrogen and phosphorus fertilization consistently favor pathogenic over mutualistic fungi in grassland soils. *Nat. Commun.* **12**, 3484 (2021).
30. Kim, S. A. & Guerinot, M. L. Mining iron: Iron uptake and transport in plants. *FEBS Lett.* **581**, 2273–2280 (2007).
31. Buzzard, V. et al. Continental scale structuring of forest and soil diversity via functional traits. *Nat. Ecol. Evol.* **3**, 1298–1308 (2019).
32. Rahman, M. A. et al. Arbuscular Mycorrhizal Symbiosis Mitigates Iron (Fe)-Deficiency Retardation in Alfalfa (*Medicago sativa* L.) Through the Enhancement of Fe Accumulation and Sulfur-Assisted Antioxidant Defense. *Int. J. Mol. Sci.* **21**, 2219 (2020).
33. Dell'Anno, F. et al. Fungi can be more effective than bacteria for the bioremediation of marine sediments highly contaminated with heavy metals. *Microorganisms* **10**, 993 (2022).
34. Chen, B. et al. Trophic interrelationships drive the biogeography of protistan community in agricultural ecosystems. *Soil Biol. Biochem.* **163**, 108445 (2021).
35. Makiola, A. et al. Environmental and plant community drivers of plant pathogen composition and richness. *New Phytologist* **233**, 496–504 (2021).
36. Delgado-Baquerizo, M. et al. The influence of soil age on ecosystem structure and function across biomes. *Nat. Commun.* **11**, 4721 (2020).
37. Barberán, A. et al. Relating belowground microbial composition to the taxonomic, phylogenetic, and functional trait distributions of trees in a tropical forest. *Ecol. Lett.* **18**, 1397–1405 (2015).
38. Zhou, J. & Ning, D. Stochastic community assembly: does it matter in microbial ecology? *Microbiol. Mol. Biol. Rev.* **81**, e00002–e00017 (2017).
39. Brockett, B. F. T., Prescott, C. E. & Grayston, S. J. Soil moisture is the major factor influencing microbial community structure and enzyme activities across seven biogeoclimatic zones in western Canada. *Soil Biol. Biochem.* **44**, 9–20 (2012).
40. Frindte, K., Pape, R., Werner, K., Löfli, J. & Knie, C. Temperature and soil moisture control microbial community composition in an arctic-alpine ecosystem along elevational and micro-topographic gradients. *ISME J.* **13**, 2031–2043 (2019).
41. Qi, J. et al. Responses of soil bacterial community structure and function to dry-wet cycles more stable in paddy than in dryland agricultural ecosystems. *Glob. Ecol. Biogeogr.* **31**, 362–377 (2022).
42. Meisner, A. et al. Soil microbial legacies differ following drying-rewetting and freezing-thawing cycles. *ISME J.* **15**, 1207–1221 (2021).
43. Jiao, S., Yang, Y., Xu, Y., Zhang, J. & Lu, Y. Balance between community assembly processes mediates species coexistence in agricultural soil microbiomes across eastern China. *ISME J.* **14**, 202–216 (2019).
44. Jiao, S., Lu, Y. & Wei, G. Soil multitrophic network complexity enhances the link between biodiversity and multifunctionality in agricultural systems. *Glob. Change Biol.* **28**, 140–153 (2022).
45. Ma, B. et al. Geographic patterns of co-occurrence network topological features for soil microbiota at continental scale in eastern China. *ISME J.* **10**, 1891–1901 (2016).
46. Shi, Y. et al. Spatial scale affects the relative role of stochasticity versus determinism in soil bacterial communities in wheat fields across the North China Plain. *Microbiome* **6**, 1–12 (2018).
47. Kettler, T. A., Doran, J. W. & Gilbert, T. L. Simplified Method for Soil Particle-Size Determination to Accompany Soil-Quality Analyses. *Soil Sci. Soc. Am. J.* **65**, 849–852 (2001).
48. Hess, T. M. *Tropical Soil Biology and Fertility: A Handbook of Methods*. Edited by J. M. Anderson and J. S. I. Ingram. Wallingford: CAB International (1989), pp. 171, £15.00. *Experimental Agriculture* **26**, 245–245 (1990).
49. Lawacz, W. Dissolved and particulate organic carbon (DOC and POC) in ecological studies. *Acta Hydrochimica et. Hydrobiologica* **5**, 363–367 (1977).
50. Sumner, M. E. & Miller, W. P. Cation Exchange Capacity and Exchange Coefficients in *Methods of soil analysis*. **5**, 1201–1229 (1996).
51. Zhang, Z., Abuduwaili, J. & Jiang, F. J. A. J. O. C. Relationship of heavy metals and soil N, P, K and total salts in Tianshan Mountains, Central Asia. *Asian J. Chem.* **25**, 8971–8975 (2013).
52. Gupta, U. C., Wu, K. & Liang, S. Micronutrients in soils, crops, and livestock. *Earth Sci. Front.* **15**, 110–125 (2008).
53. Ye, J., Song, Z., Wang, L. & Zhu, J. Metagenomic analysis of microbiota structure evolution in phytoremediation of a swine lagoon wastewater. *Bioresour. Technol.* **219**, 439–444 (2016).
54. Bellemain, E. et al. ITS as an environmental DNA barcode for fungi: an *in silico* approach reveals potential PCR biases. *BMC Microbiol.* **10**, 189 (2010).
55. Stoeck, T. et al. Multiple marker parallel tag environmental DNA sequencing reveals a highly complex eukaryotic community in marine anoxic water. *Mol. Ecol.* **19**, 21–31 (2010).
56. Callahan, B. J. et al. DADA2: High-resolution sample inference from Illumina amplicon data. *Nat. Methods* **13**, 581–583 (2016).
57. Wang, Q., Garrity George, M., Tiedje James, M. & Cole James, R. Naïve Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl. Environ. Microbiol.* **73**, 5261–5267 (2007).
58. Nilsson, R. H. et al. The UNITE database for molecular identification of fungi: handling dark taxa and parallel taxonomic classifications. *Nucleic Acids Res.* **47**, D259–D264 (2019).
59. Guillou, L. et al. The Protist Ribosomal Reference database (PR2): a catalog of unicellular eukaryote Small Sub-Unit rRNA sequences with curated taxonomy. *Nucleic Acids Res.* **41**, D597–D604 (2013).
60. Guo, S. et al. Protists as main indicators and determinants of plant performance. *Microbiome* **9**, 64 (2021).
61. Oksanen, J. et al. Package 'vegan'. *Community Ecol. Package Version* **2**, 1–295 (2013).
62. Goslee, S. C. & Urban, D. L. The ecodist package for dissimilarity-based analysis of ecological data. *J. Stat. Softw.* **22**, 1–19 (2007).
63. Delgado-Baquerizo, M. et al. Carbon content and climate variability drive global soil bacterial diversity patterns. *Ecol. Monogr.* **86**, 373–390 (2016).
64. Li, Y., Tremblay, J., Bainard, L. D., Cade-Menun, B. & Hamel, C. Long-term effects of nitrogen and phosphorus fertilization on soil microbial community structure and function under continuous wheat production. *Environ. Microbiol.* **22**, 1066–1088 (2020).

65. Zhou, Z., Wang, C., Zheng, M., Jiang, L. & Luo, Y. Patterns and mechanisms of responses by soil microbial communities to nitrogen addition. *Soil Biol. Biochem.* **115**, 433–441 (2017).
66. White, J. G. & Zasoski, R. J. Mapping soil micronutrients. *Field Crops Res.* **60**, 11–26 (1999).
67. Grace, J. B. et al. Integrative modelling reveals mechanisms linking productivity and plant species richness. *Nature* **529**, 390–393 (2016).
68. Rosseel, Y. *lavaan: An R package for structural equation modeling*. *J. Stat. Softw.* **48**, 36 (2012).
69. Breiman, L. Random Forests. *Mach. Learn.* **45**, 5–32 (2001).
70. Ellis, N., Smith, S. J. & Pitcher, C. R. Gradient forests: calculating importance gradients on physical predictors. *Ecology* **93**, 156–168 (2012).
71. Stephenson, F. et al. Using Gradient Forests to summarize patterns in species turnover across large spatial scales and inform conservation planning. *Divers. Distrib.* **24**, 1641–1656 (2018).
72. Lang, M. et al. *mlr3*: A modern object-oriented machine learning framework in R. *J. Open Source Softw.* **4**, 1903 (2019).

## ACKNOWLEDGEMENTS

This work was supported by the National key research and development program of China (grant No.: 2021YFD1900500), the Fundamental Research Funds for the Central Universities (grant No.: 2022HHZX005), National Science Foundation for Excellent Young Scholars of China (grant No.: 42122050), National Science Foundation of China (grant No.: 42077222, 41830755, and 41807030).

## AUTHOR CONTRIBUTIONS

All authors contributed intellectual input and assistance to this study and the manuscript preparation. Z.P. conducted the experiments, analyzed the data, and wrote the manuscript. G.W. and S.J. conceived and designed the experiments and revised the manuscript. C.L. conducted bioinformatics processing of sequencing data. M.G., Y.Q., Y.P., H.G., Y.L. and X.L. performed the research.

## COMPETING INTERESTS

The authors declare no competing interests.

## ADDITIONAL INFORMATION

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41522-022-00363-3>.

**Correspondence** and requests for materials should be addressed to Gehong Wei or Shuo Jiao.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

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



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022