

OPEN

Soil bacterial diversity correlates with precipitation and soil pH in long-term maize cropping systems

Wenjun Tan^{1,2}, Junman Wang^{1,2}, Wenqing Bai¹, Jiejun Qi¹ & Weimin Chen^{1*}

Unraveling the key drivers of bacterial community assembly in agricultural soils is pivotal for soil nutrient management and crop productivity. Presently, the drivers of microbial community structure remain unexplored in maize cropping systems under complex and variable environmental scenarios across large spatial scales. In this study, we conducted high-throughput 16S rRNA gene sequencing and network analysis to identify the major environmental factors driving bacterial community diversity and co-occurrence patterns in 21 maize field soils across China. The results show that mean annual precipitation and soil pH are the major environmental factors that shape soil bacterial communities in maize soils. The similarities of bacterial communities significantly decreased with increasing geographic distance between different sites. The differences in spatial turnover rates across bacterial phyla indicate the distinct dispersal capabilities of bacterial groups, and some abundant phyla exhibited high dispersal capabilities. *Aeromicrobium*, *Friedmanniella*, *Saccharothrix*, *Lamia*, *Rhodococcus*, *Skermanella*, and *Pedobacter* were identified as keystone taxa. Based on the node-level and network-level topological features, members of the core microbiome were more frequently found in the center of the ecosystem network compared with other taxa. This study highlights the major environmental factors driving bacterial community assembly in agro-ecosystems and the central ecological role of the core microbiome in maintaining the web of complex bacterial interactions.

As an important component of terrestrial ecosystems, agricultural soils are the productive engine of the Earth and account for over one third of the global land area¹. Microbes have an immense amount of diversity, play important roles in global biogeochemical processes, and drive the functioning of diverse ecosystems². In agricultural soils, microbes are critical to crop productivity by participating in organic matter decomposition, humus formation, and nutrient transformation^{3,4}. Understanding how biodiversity is generated and maintained, as well as determining the drivers of microbial distribution patterns across different scales is the central goal of microbial ecology⁵. There has been a growing number of studies on microbial biogeography in diverse environments^{6–9}. Microbes are distributed worldwide; however, in a given habitat, varying environmental factors can lead to different niches and select for adapted microbes. For example, the microbial distribution can be substantially influenced by soil pH, nutrient availability, and climate conditions^{10–13}.

The spatial patterns of microbial diversity can reveal the key drivers of biodiversity and help predict the risk of biodiversity loss. The distance-decay relationship (DDR) has been well documented for the spatial distributions of plant and animal communities, as the community similarity decreases with increasing geographic distance⁵. Recent studies have shown that the non-random distribution of terrestrial and aquatic microbial communities is also affected by geographic distance and exhibits distance-decay patterns across different habitats^{6,7,14}. Thus, spatial distance may induce dispersal limitation, influencing the biogeography of microbes⁵. However, it is still challenging to characterize and quantify microbial communities distributed in agro-ecosystems under complex and variable environmental scenarios on a large scale. Maize (*Zea mays* L.) is one of the major cereal crops worldwide and it is widely cultivated across China, making a suitable model for assessing such broad-scale questions. In addition, maintaining a stable state of the plant community and ecosystem management would allow for observation of the influence of edaphic and other abiotic factors on soil microbial communities.

In a particular habitat type, a suite of members broadly distributed amongst microbial communities at different locations is defined as the core microbiome^{7,15}. Given its ubiquity in a specific environment, discovering

¹Shaanxi Key Laboratory of Agricultural and Environmental Microbiology, College of Life Sciences, Northwest A&F University, Yangling, Shaanxi, 712100, P. R. China. ²These authors contributed equally: Wenjun Tan and Junman Wang. *email: chenwm029@nwsuaf.edu.cn

Site	Edaphic						Geographic		Climatic		Soil type
	pH	AK ^a (mg/kg)	OC (mg/kg)	TN (mg/kg)	AN (mg/kg)	AP (mg/kg)	Latitude (°)	Longitude (°)	PRE (mm)	TEM (°C)	
1	7.68	117.10	18.26	0.78	43.75	61.98	18.65	109.67	2162.8	24.8	sandy clay loam
2	5.46	167.70	26.40	1.06	103.25	15.83	25.00	101.51	937.2	16.2	clay loam
3	7.93	125.20	14.87	0.68	35.00	17.44	27.92	112.75	1377.0	17.4	loamy clay
4	8.00	116.10	10.42	0.44	19.25	8.30	29.23	120.05	1386.6	17.7	sandy loam
5	7.96	145.40	29.72	0.96	98.00	30.48	29.52	106.66	1108.0	18.4	sandy loam
6	8.39	55.10	3.16	0.44	3.50	3.39	31.88	106.16	1013.5	16.6	loamy clay
7	5.23	135.30	13.62	0.56	94.50	21.82	33.28	114.30	863.1	14.8	clay loam
8	8.33	198.20	16.49	0.84	31.50	14.17	34.36	109.12	569.4	13.8	clay loam
9	8.65	234.70	16.75	0.50	35.00	17.03	35.15	111.23	519.2	13.2	silty clay loam
10	7.87	86.50	14.86	0.56	87.50	9.45	35.95	116.57	614.3	14.0	clay loam
11	8.72	208.30	11.67	0.44	49.00	4.92	36.25	106.39	425.5	6.9	clay loam
12	8.74	102.80	5.86	0.16	5.25	4.45	36.59	107.27	409.5	9.2	silty clay loam
13	6.37	288.20	15.25	0.62	31.50	87.36	36.64	119.41	546.5	12.8	loamy clay
14	5.83	135.40	37.52	1.24	63.00	175.37	37.36	114.79	478.2	13.2	sandy loam
15	8.76	207.30	13.73	0.50	5.25	38.06	38.45	77.25	61.3	12.0	sandy loam
16	8.50	80.70	23.17	0.68	29.75	127.64	39.52	112.53	407.3	6.1	sandy loam
17	8.86	248.90	18.94	0.78	87.50	52.70	39.88	117.73	623.8	11.9	clay loam
18	7.04	105.00	10.06	0.56	21.00	14.10	39.94	121.52	617.5	9.9	sandy clay loam
19	8.45	143.40	8.81	0.40	175.00	57.88	40.75	111.97	396.5	7.3	loamy clay
20	8.04	323.00	34.16	1.00	47.25	56.16	43.87	126.75	580.0	5.0	loamy clay
21	8.18	144.40	26.59	1.00	66.50	28.88	46.51	124.96	434.6	3.8	sandy loam

Table 1. Environmental parameters of the 21 soil samples collected from maize cropping systems across China. ^aAK, available potassium; OC, organic carbon; TN, total nitrogen; AN, available nitrogen; AP, available phosphorus; PRE, mean annual precipitation; TEM, mean annual temperature.

a core microbiome is critical to understand the assembly and stability of microbial communities¹⁵. Determining the factors driving assembly of the core microbiome in agricultural soils has implications for understanding the microbial functional groups in agro-ecosystems. Moreover, the interactions among microbes form a complex network in specific ecosystem niches. Microbial co-occurrence patterns can help to uncover the complex interactions of microbial communities and delineate the underlying ecological processes¹⁶. Network analysis can identify the keystone microbial taxa that have the greatest influence on community structure and determine the ecological roles of specific microbial groups of concern.

Here we investigated the drivers of bacterial community assembly in maize cropping systems across a large spatial scale. The main objectives of the study were: (i) to examine the bacterial community diversity and structure in maize soils by high-throughput 16S rRNA gene sequencing; ii) to identify the major environmental factors that influence bacterial community diversity, and; (iii) to identify the keystone taxa in maize cropping systems via network analysis.

Results

Geochemistry of maize soils. The samples included a diverse array of soil types, including sandy clay loam, clay loam, loamy clay, sandy loam, silty clay loam, and loam (Table 1). The soils were slightly acidic to alkaline, with the pH varying from 5.23 to 8.86 (pH > 7 in 15 of 21 samples). There was a wide variation in the mean annual precipitation across different sites (61.3–2,162.8 mm). The pairwise distances between the sites ranged from 87 to 4,130 km (Table S1, Supporting Information). There were no correlations between edaphic factors and geographic distances ($P > 0.1$; Mantel test). However, the two climatic factors (mean annual temperature and precipitation) were negatively correlated with latitude ($P < 0.05$; Spearman correlation).

Distribution of bacterial taxa. We obtained 613,504 high-quality sequences in total, with 18,844 to 36,891 sequences per sample (mean = 29,214). There were 5,318 operational taxonomic units (OTUs), 5,307 of which were affiliated with the bacterial domain, and the remaining 11 with the archaeal domain. Five dominant bacterial phyla (relative abundance > 5%) accounted for 89.7% of the total sequences: Proteobacteria (43.1%), Actinobacteria (24.4%), Acidobacteria (10.7%), Chloroflexi (6.0%), and Gemmatimonadetes (5.5%). At the class level, Proteobacteria was dominated by Betaproteobacteria, Alphaproteobacteria, Gammaproteobacteria, and Deltaproteobacteria; Actinobacteria was dominated by Actinobacteria and Thermoleophilia (overall relative abundance, 21.88%). The dominant archaea was *Candidatus Nitrososphaera* (Crenarchaeota phylum), which accounted for 0.11% of all sequences. Additionally, Bacteroidetes, Planctomycetes, Cyanobacteria, Firmicutes, Nitrospirae, Armatimonadetes, and Verrucomicrobia were detected in most samples at low relative abundances (< 5%) (Fig. 1).



Figure 1. Locations of the 21 sampling sites with their phylum-level distributions of soil bacterial communities in maize cropping systems across China. The map was generated using GenGIS software (Version II, <http://kiwi.cs.dal.ca/GenGIS>).

Alpha diversity, measured by OTU richness ($r = -0.398$, $P < 0.1$, marginally significant) and Shannon–Wiener index ($r = -0.66$, $P < 0.01$, significant), was negatively correlated with mean annual precipitation (Table S2, Supporting Information). The Shannon–Wiener index was also negatively correlated with mean annual temperature ($r = -0.449$, $P < 0.05$, significant). Meanwhile, there was a marginally significant positive correlation between OTU richness and soil pH ($r = 0.409$, $P < 0.1$, marginally significant). These significant correlations were confirmed by the linear regression relationships, as shown in Figs. 2 and 3.

Environmental factors influencing bacterial community structure. Major environmental factors that shaped the microbial communities were determined by Constrained analysis of principal coordinates (CAP) based on Weighted Unifrac distance (Fig. S1, Supporting Information). Mean annual precipitation and soil pH were significantly linked with bacterial community structure (PERMANOVA; Supplementary Table S3). Additionally, mean annual precipitation, which explained higher degree of variance ($R^2 = 0.277$), was a major driver of bacterial community structure. The effect of mean annual precipitation on the bacterial community structure was evident at a low taxonomic resolution, as the relative abundances of dominant phyla were significantly changed along with the annual precipitation gradient, as determined by linear regression relationships (Fig. 4). With the increasing of mean annual precipitation, the relative abundance of Proteobacteria was significantly increased and the relative abundance of Gemmatimonadetes, and Nitrospirae significantly decreased.

The influence of spatial distance on the variation in community structure was estimated by DDR, a distance-decay similarity model. Significant negative DDRs were found in a linear regression using beta diversity (Weighted Unifrac distance; Fig. S2, Supporting Information). At different taxonomic levels, many dominant phyla also exhibited significant DDRs. Interestingly, Planctomycetes and Verrucomicrobia showed steeper slopes than other phyla (Fig. 5).

Given the important influence of soil types on the bacterial community structure based on prior knowledge, we estimated the effect of soil texture on bacterial community structure. Soil samples were divided into three groups, according to the different soil types, including clay loam, loamy clay and sandy loam. Based on PCoA analysis, we did not observe clear clustering of these samples, according to soil types (Fig. S3, Supporting Information). This was confirmed by ANOSIM and permutational multivariate analysis of variance (ADONIS) tests confirmed that no significant difference was observed in bacterial community structure among soil types ($P > 0.1$). Furthermore, bacterial alpha diversity showed no significant difference among soil types ($P > 0.1$), estimated via one-way analysis of variance. These results indicated that soil types did not influence the diversity and structure of bacterial community in Maize soils of present work.

Bacterial community assembly of the core microbiome. We identified 421 core taxa in seven phyla and seventy genera. In general, *Janthinobacterium*, *Ramlibacter*, *Kaistobacter*, *Lysobacter*, and *Thermomonas* were the predominant members of the core microbiome from the Proteobacteria phylum. *Nocardiooides*, *Streptomyces*, and *Rhodococcus* were the predominant members of the core microbiome from the Actinobacteria phylum (Fig. S4, Supporting Information). Based on Bray–Curtis dissimilarity, the core microbiome contributed to a substantial fraction (mean = 64.45%, range = 39.32–81.83%) of community dissimilarity despite accounting for only 7.92% of the total OTUs (Fig. S5, Supporting Information).

To explore the influences of environmental factors on the assembly of the core microbiome, we estimated the associations between the core microbiome and environmental factors by correlation analysis. There were

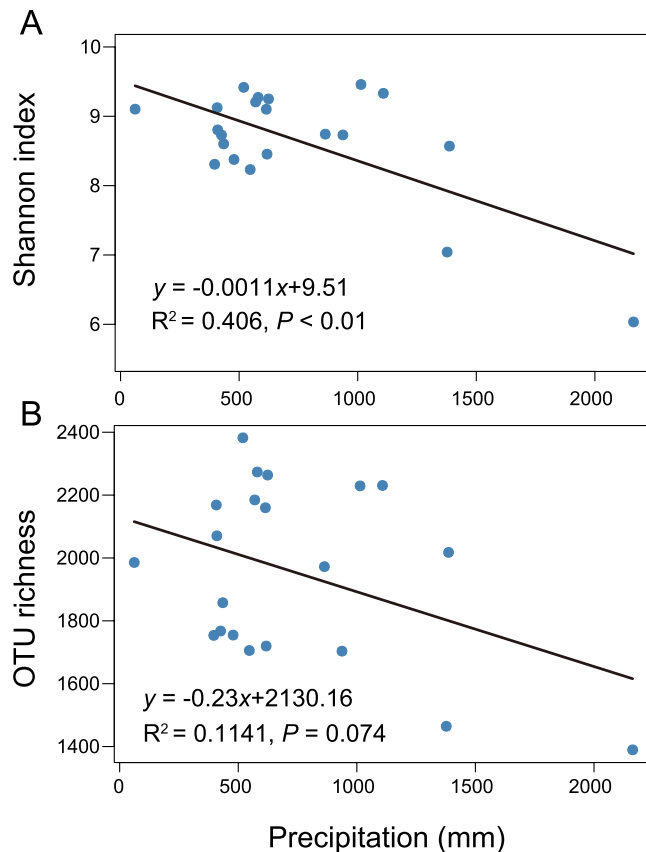


Figure 2. The linear regression relationship between mean annual precipitation and either Shannon-Wiener index (A) or operational taxonomic unit (OTU) richness (B) of soil bacterial communities in maize cropping systems ($n = 21$).

significant [false discovery rate (FDR)-corrected $P < 0.05$] and strong ($r > 0.6$ or < -0.6) correlations with four environmental factors (Fig. 6). For example, *Nitrospira* and *Skermanella* were positively correlated with soil pH; *Nannocys* was positively correlated with available nitrogen, and; *Pilimelia*, *Nonomuraea*, *Rubellimicrobium*, and *Sinorhizobium* were negatively correlated with mean annual precipitation; *Aeromicrobium* was negatively correlated with mean annual temperature.

Bacterial co-occurrence patterns and keystone taxa. First, we structured a soil bacterial network based on correlations between OTUs to explore the overall distribution of co-occurrence patterns and the ecological role of the core microbiome. The resulting network consisted of 5,312 nodes (i.e., OTUs) and 181,157 edges (mean = 34 edges per node). Topological analysis showed that the APL was 3.237 edges with a ND of 4.401 edges, and the CC was 0.162 (Table S4, Supporting Information). The structural properties of the real-world network were greater than those of an identically sized random Erdős–Rényi network (APL = 2.398, CC = 0.0128, and ND = 3). These suggest that the real bacterial network was non-randomly distributed and had a highly connected topological structure.

The betweenness centrality values were significantly higher in the core microbiome nodes than in other nodes (Fig. 7). This suggests that core taxa were more likely to be located in central position within the network. To verify this observation, a sub-network was generated for the core microbiome and network-level topological features were analyzed (Table 2). The values of topological features were higher in the core microbiome sub-network than in the other sub-network, indicating that the former sub-network was more connected and complex than the latter sub-network.

The co-occurrence patterns of soil microbes were examined at the genus level and keystone genera were identified. The genus-level network consisted of 184 nodes and 639 edges (Fig. 7). The nodes primarily belonged to four bacterial phyla: *Proteobacteria*, *Actinobacteria*, *Firmicutes*, and *Bacteroidetes*. Seven genera were considered keystone based on betweenness centrality scores. *Aeromicrobium*, *Friedmanniella*, *Saccharothrix*, *Lamia*, and *Rhodococcus* were keystone genera from the Actinobacteria phylum; *Skermanella* is from the Proteobacteria phylum, and *Pedobacter* was a keystone genus from the Bacteroidetes phylum. Five of the keystone genera were members of the core microbiome (Fig. S6, Supporting Information).

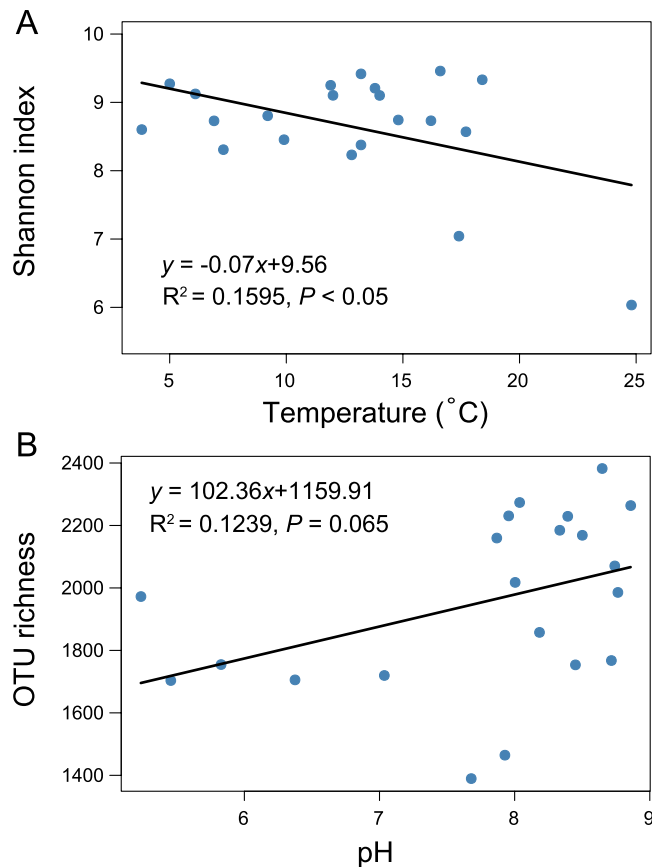


Figure 3. The linear regression relationships between: mean annual temperature and Shannon–Wiener index of soil bacterial communities (A), and; soil pH and operational taxonomic unit (OTU) richness of soil bacterial communities (B) in maize cropping systems ($n = 21$).

Discussion

Given the global role of microbes in the environment, determining the drivers of microbial community assembly is an important issue in microbial ecology¹⁷. Microbes inhabiting agricultural soils are closely linked to crop productivity through complex biogeochemical processes¹⁸. In this study, we investigated the major environmental factors that drive bacterial community assembly in maize soils across a large spatial scale with a latitude gradient.

Maize, which possesses exceptional phenotypic and molecular diversity, can grow in diverse soils with a pH range of 5.7–8.0¹⁹. In the present study, the maize soils were slightly acidic to alkaline, within the suitable pH range for growth of maize. The variation of soil pH might be related with the wide variety of soil textures. In addition, most of our soil samples (71%) were circumneutral to alkaline ($\text{pH} > 7$). Bacterial communities can be shaped by complex environmental factors. Significant correlations have been frequently found between the pH and bacterial alpha-diversity in soils^{20,21} and lake sediments²², indicating that pH is a predictor for soil microbial diversity. In the present study, OTU richness showed a marginally significant and positive correlation with soil pH in the range of 5.23–8.86, in agreement with previous observation in British soils over a broader pH range from 3 to 9²⁰. Soil pH could affect the bioavailability of carbon and nitrogen substrate as well as toxic metals, indirectly influencing soil microbes²³. The richness and diversity of bacterial communities are key maintainers for the productivity and stability of agricultural soil ecosystems³.

Two climatic factors, mean annual precipitation and temperature, also predicted bacterial diversity in the maize soils. These two variables were negatively correlated with the Shannon–Wiener index of soil bacterial communities in the current study, although we also acknowledged that the significant relationship between the Shannon index and the average annual precipitation may be driven by two outliers. This result could be supported by a continental-scale study across eastern China²⁴, which reported higher bacterial diversity at high latitudes in maize fields. However, conflicting results were reported in natural terrestrial ecosystems that soil bacterial diversity was positively or not correlated with temperature^{10,13}. The discrepancy may be associated with the complex and heterogeneous environments across different habitats across multiple scales.

In the present study, both mean annual precipitation and soil pH were significantly linked to bacterial community structure in maize soils based on beta-diversity metrics. Similarly, annual precipitation and temperature were found significantly linked to the functional structure of microbial communities in a soil transplant experiment designed to simulate climate change²⁵. Annual precipitation was a larger contributor to variation in bacterial community structure than soil pH in the maize cropping systems. This might be attributed to the large variability of precipitation among the sampling sites, which were across a large spatial scale and covered large landmass

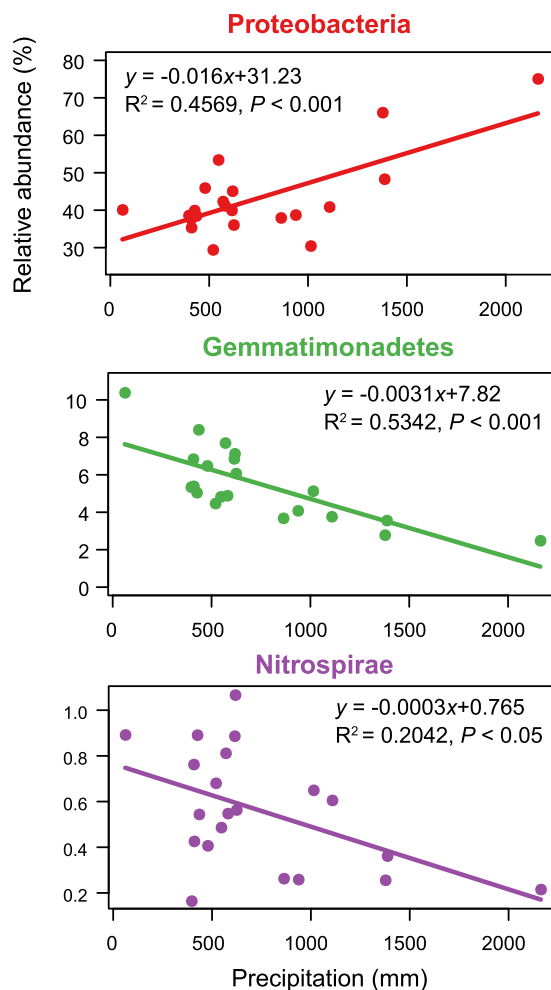


Figure 4. The linear regression relationships between the relative abundances of the dominant bacterial phyla and mean annual precipitation in maize cropping systems ($n = 21$).

and various climate regions in China. Precipitation could change soil moisture, and sudden change in moisture is stressful to microbes, as they must expend energy to regulate osmotic pressure to their microenvironment²⁶. Additionally, annual precipitation could compensate for water tables and affect redox condition seasonality²⁷. In addition, climate factors can change soil geochemistry and thereby impact microbial community structure. For example, a national soil survey of Scotland revealed that soil pH and precipitation altered the diversity of ammonia oxidizer communities and their capacity for nitrification¹¹. Moreover, the combination of warming and decreased precipitation altered bacterial community structure in a Tibetan plateau alpine grassland after a 1-year manipulation²⁸.

The above findings suggest that precipitation is a major factor impacting bacterial community assembly in soils; this is supported by our observations in the current study. A few dominant phyla were significantly correlated with mean annual precipitation in the maize cropping systems. With increasing annual precipitation, Proteobacteria increased, while Gemmatimonadetes, and Nitrospirae decreased in maize soils. Similarly, altered precipitation was found to specifically impact Gammaproteobacteria in alpine grassland²⁸. Another study found that Gemmatimonadetes was inversely correlated with soil moisture²⁹.

An important issue in biogeography is estimating whether spatial distance creates genetic variation⁵. The DDR estimates the variation in beta-diversity across spatial scales^{30,31}. In the present study, the variation in soil bacterial community decreased linearly with increasing geographic distance, indicating the spatial structure of bacterial communities in maize soils. This result could be supported by other studies demonstrating the robust DDRs of microbial communities in agricultural fields^{32,33}.

Various spatial scales could result in a difference in the distance-decay slope between habitats due to the dependence of distance-decay on the spatial scale^{34,35}, which is known to affect the assembly of soil bacterial communities in wheat fields³³. Here we found that the distance-decay slope ($w = -0.043$) in maize soils with an interval of 4129.6 km was steeper than in dryland habitats of northern China across a 4000 km transect³⁵, including alpine grassland ($w = -0.041$), desert ($w = -0.017$), desert grassland ($w = -0.014$), and typical grassland ($w = -0.017$). The slopes of these relationships across habitats can differ, reflecting varying rates of species turnover in their habitats³⁵. In particular, agricultural fields are typical human-managed terrestrial ecosystems, resulting in distinct microbial community assembly patterns compared to natural ecosystems³⁶. Agricultural

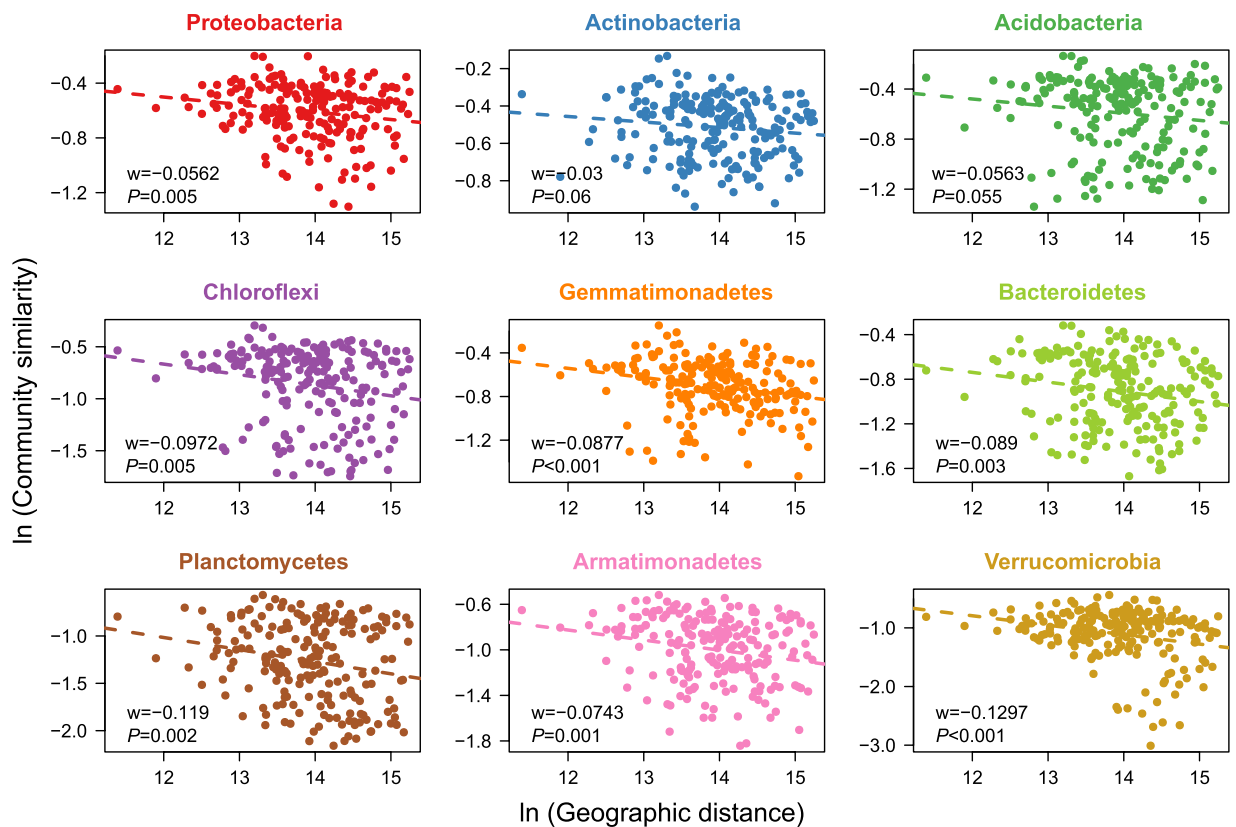


Figure 5. The distance-decay relationships of dominant bacterial phyla in maize cropping systems ($n = 21 \times 21$). A “w” represents the slopes of the linear regression curve, and a “P” represents the significance of the linear regression model.

soils can form unique, ephemeral habitats across local sites during long-term tillage³⁷, thus generating more spatially structured microbial communities across large scales. Dispersal limitations, which can be attributed to geographic distances, reflect a stochastic process that influences the beta diversity and biogeographic patterns of microbes⁵. This is in accordance with previous findings in diverse habitats that microbial distributions are not only driven by deterministic process (i.e., environmental heterogeneity), but can also be governed by stochastic process (i.e., dispersal limitations)^{6,7}.

Furthermore, a number of bacterial phyla showed significant DDRs with differing slopes in the maize soils, indicating their distinct turnover rates along the spatial gradient. The dominant phyla Proteobacteria, Actinobacteria, and Acidobacteria had lower slopes and stronger dispersal capabilities, while the less abundant phyla Planctomycetes and Verrucomicrobia showed lower dispersal capabilities. Dispersal capabilities may be linked to phylum abundance. Abundant bacteria are readily dispersed, as many individual cells can potentially be involved in a dispersal event; rare bacteria with lower abundances should show lower dispersal rates than abundant taxa³⁸. Verreydt, *et al.*³⁹ found that the dispersal capabilities of bacterial groups varied greatly in aquatic habitats, similar to our observations in maize soils.

In particular, we found that Proteobacteria and Gemmatimonadetes were correlated to both geographic distance and environmental factors, although no autocorrelations were found between geographic distance and edaphic factors. Thus, the change in the relative abundances of particular microbes may be ascribed to both dispersal limitations and environmental heterogeneity. Stochastic processes generate random variation in the relative abundances of species (i.e., dispersal limitation and ecological drift) and create patchiness in community composition⁴⁰. Deterministic processes indicate that environmental selection including abiotic and biotic factors could determine the community assembly⁴¹. The above results indicate the combined influence of deterministic and stochastic processes on the bacterial community assembly in maize cropping systems.

Network analysis was used to explore the bacterial co-occurrence patterns and provide insights into bacterial interactions in soils from maize cropping systems. When compared with a random network, the non-random co-occurrence of the bacterial communities was observed across the maize soils; this indicates the role of deterministic processes in bacterial community assembly of agro-ecosystems⁷. In this study, the core microbiome was broadly distributed and accounted for a high fraction of community beta diversity but a low proportion of total OTUs in the maize soils. According to the node-level (betweenness centrality) and network-level topological features, we found that members of the core microbiome were more often located in central ecological positions than other taxa. A more connected and complex sub-network was therefore generated for the core microbiome. Betweenness centrality scores can be used to determine the most influential taxa (e.g., keystone species) that are

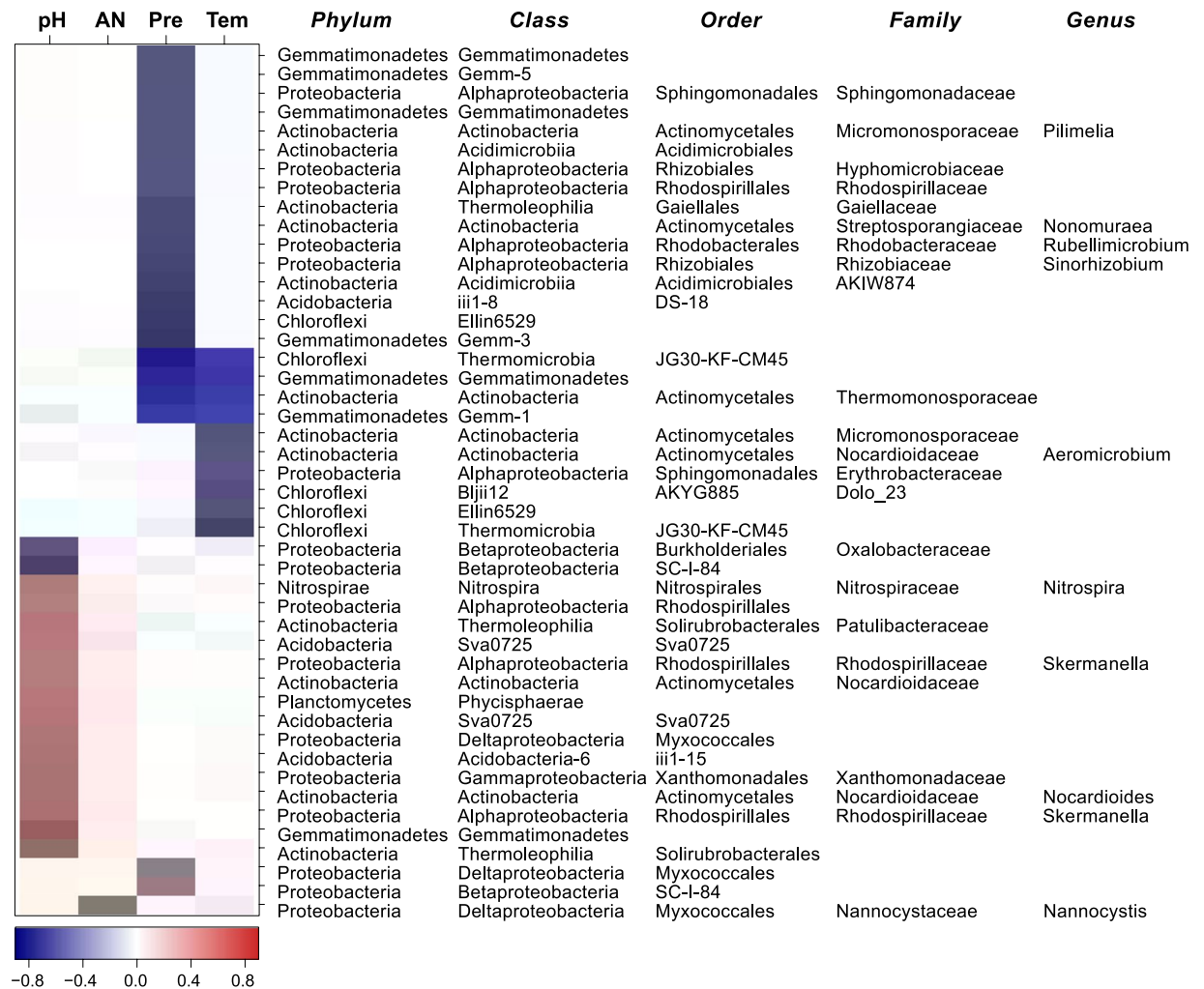


Figure 6. Spearman correlations between the core microbiome and environmental factors in maize cropping systems. The values of correlation coefficients are indicated according to the scale bar. Only significant ($P < 0.05$) and strong (correlation coefficient, $r > 0.6$ or < -0.6) correlations are shown; others are left blank. AN, available nitrogen; Pre, Mean annual precipitation, and; Tem, Mean annual temperature.

responsible for maintaining connectivity within the ecological network⁴². Thus, the core microbiome might play a vital role in maintaining the complex connections between microbes in agro-ecosystems.

In a co-occurrence network analysis, the topological features of nodes are used to determine the potential importance of microorganisms, such as keystone species^{43–45}. Different with core taxa, keystone species were determined via network analysis. Based on betweenness centrality scores, *Aeromicrobium*, *Friedmanniella*, *Saccharothrix*, *Lamia*, *Rhodococcus*, *Skermanella*, and *Pedobacter* were considered keystone taxa in the maize soils. *Aeromicrobium* and *Rhodococcus* are characteristic of herbicide metabolism⁴⁶. *Friedmanniella* has been detected in lignocellulose degradation of composted agricultural waste⁴⁷. *Saccharothrix* possesses antifungal activity⁴⁸. *Pedobacter* has been isolated from a herbicide-contaminated soil⁴⁹. These keystone taxa may play versatile roles in soil ecological processes in agro-ecosystems.

Sources of uncertainties should be noted when interpreting the results of this study. Firstly, number of samples is small and there are no duplicates at each site, which may affect our results. In the further work, three composite samples per site would be taken into account. Second, although we sampled at the same time, we failed to consider the specific management system of each site, such as the variety of corn, the amount of fertilizer applied and the farming method. Therefore, in addition to natural rainfall, the source of water is then agricultural irrigation, but our study did not include artificial water supply due to large-scale sampling, which may still influence the results.

In conclusion, this is a detailed study of soil bacterial communities in maize cropping systems across a continental scale in China. We determined the major environmental factors correlating with bacterial community diversity and explored the bacterial co-occurrence patterns in 21 maize soils across China. Mean annual precipitation and soil pH were the main factors that shaped the bacterial communities. Different spatial turnover rates of microbes suggest distinct dispersal capabilities of different bacterial groups. The central ecological role of the core microbiome suggests its vital roles in promoting soil ecological processes and maintaining the complex bacterial network. The non-random co-occurrence patterns and identification of keystone taxa provide new insights

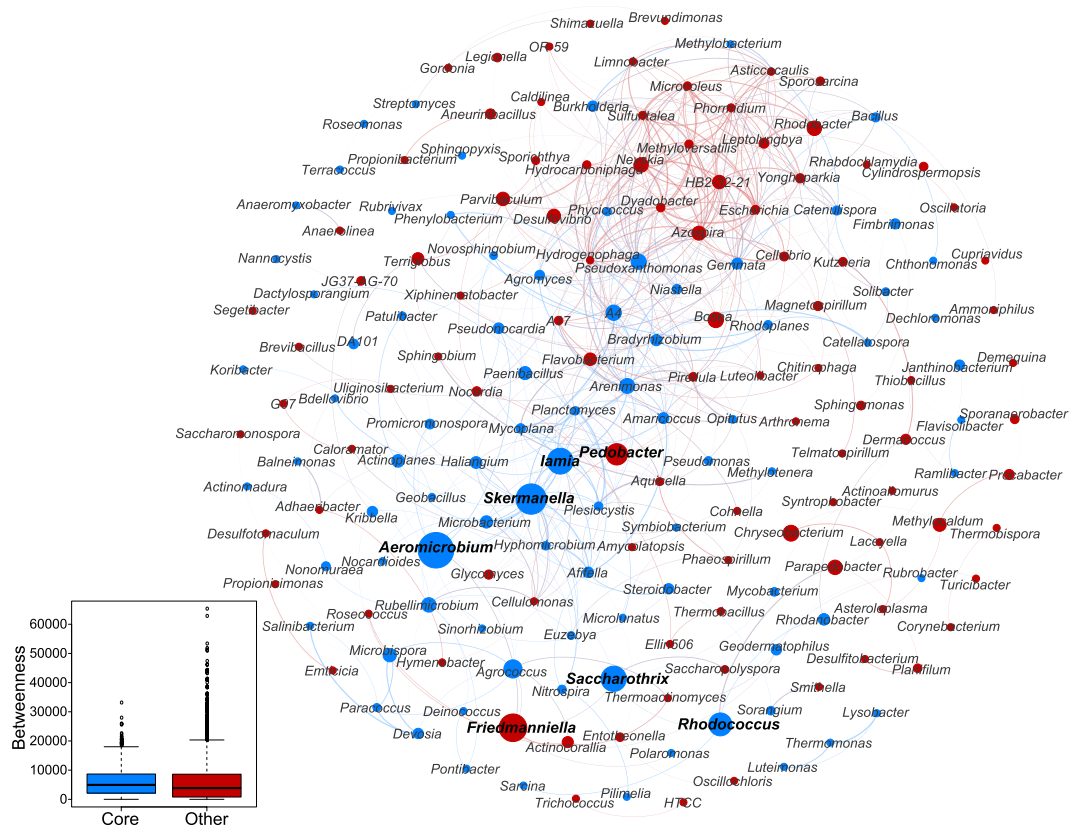


Figure 7. Network of co-occurring bacterial genera in maize cropping systems based on correlation analysis. A connection stands for a strong (Spearman $r > 0.6$) and significant ($P < 0.01$) correlation. The size of each node is proportional to the betweenness centrality value; the thickness of the connection between two nodes (edge) is proportional to the value of the Spearman correlation coefficient. The nodes colors are based on categorization as a core (“Blue”) or other (“red”) taxa. The boxplot shows the comparison of betweenness centrality values between the core and other taxa in maize cropping systems. The values were significantly different between the two groups based on a Wilcoxon Rank Sum test ($P < 0.05$).

	Clustering coefficient	Average path length	Network diameter	Graph density
Core	0.462	3.473	0.055	6.506
Others	0.126	3.247	0.013	4.410

Table 2. Network-level topological features of the sub-networks for the core and other taxa in maize cropping systems.

into microbial community assembly in agro-ecosystems. Furthermore, microbial taxonomic and functional data across temporal scales should be integrated to better characterize the patterns of microbial biogeography and dynamics in agro-ecosystems. Particularly, the investigation the specific distribution of the core microbiomes across different crops agro-ecosystems might improve our understanding of nutrient management and crop health in agro-ecosystems, which should be further explored.

Methods

Soil sampling and data collection. Twenty-one soil samples were collected from maize fields over 10 years old in 19 provinces across China (18.62°–46.51° N; 77.25°–126.75° E) during July–August 2014 (Fig. 1). Since the maize growing period varied across regions, we collected the bulk soils rather than rhizosphere soils. At each sampling site, five random soil cores were taken from a depth of 0–15 cm in a 100 × 100 m plot and combined to form a composite sample. The samples were immediately transported to the laboratory on dry ice, where they were sieved through a 2.0-mm mesh to remove plant debris and rocks³⁶. Subsamples for physicochemical analysis were air-dried. Soil pH, organic carbon, total nitrogen, available nitrogen, available phosphorus, available potassium, and texture were analyzed using standard soil testing procedures⁵⁰. Mean annual air temperature and precipitation data were obtained from the WorldClim database (<http://www.worldclim.org>). Subsamples for microbial analysis were stored at –80 °C until DNA was extracted.

DNA extraction and Illumina sequencing analysis of 16S rRNA gene amplicons. Total genomic DNA was extracted from 0.5 g of each soil sample using the MP FastDNA[®] SPIN Kit for soil (MP Biochemicals, Solon, OH, USA) following the manufacturer's instructions. The V4 and V5 hypervariable regions of the bacterial 16S rRNA gene were PCR amplified using the primer pair 515F and 907R⁷. The purified PCR amplicons were sequenced using an Illumina Miseq (300-bp paired-end reads) platform (Illumina Inc., San Diego, CA, USA). The acquired sequences were filtered by quality according to Caporaso, *et al.*⁵¹. Chimeric sequences were removed using the UCHIME algorithm⁵² in the USEARCH package v. 6.1544⁵³. Sequences were grouped by taxonomy and assigned to OTUs at the 97% sequence identity level using the UPARSE package (<http://drive5.com/uparse/>)⁵⁴. After singleton removal, representative sequences were taxonomically assigned using the Ribosomal Database Project naïve Bayesian rRNA classifier within the SILVA database (release 128) at an 80% confidence threshold⁵¹. All sequences were deposited in the NCBI Sequence Read Archive database (Biosample number: SAMN06105854-SAMN06105874).

Data analyses. Alpha diversity (OTU richness and Shannon–Wiener index) and beta diversity were calculated using QIIME (<http://qiime.org/index.html>) based on a subsample containing a minimum number of sequences (18,844)⁷. The same subsample was used to calculate the pairwise Weighted UniFrac distances between samples, i.e., bacterial community structure. A soil sampling map showing the phylum-level distribution of bacterial communities in pie charts was generated using GenGIS II⁵⁵.

To identify the major environmental attributes that shape soil bacterial communities, Constrained analysis of principal coordinates (CAP) and permutational MANOVA were performed between environmental factors and bacterial community structure (i.e., Weighted UniFrac distance), using capscale and anova procedure in the “vegan” package⁵⁶. To test the influence of spatial distance on the variation in community structure, we used DDR to explore the correlations between community similarity and geographic distance³⁵. The DDR was estimated for each pairwise set of samples as the slope of the linear least-squares regression curve between geographic distance and bacterial community similarity (based on 1 – dissimilarity of the Weighted UniFrac distance matrix). In addition, the DDRs were calculated for different bacterial phyla to estimate their turnover rates along the spatial gradient.

The partitioned Bray-Curtis dissimilarity between two samples was calculated as the fraction of beta diversity attributed to the core microbiome (defined as bacterial taxa present in all soil samples, $n = 21$). The core taxa were used to calculate the summation in the numerator of the Bray-Curtis dissimilarity expression, while all of the taxa were used to calculate the scaling summation in the denominator^{57,58}. Spearman correlations between the core microbiome and environmental factors were estimated to explore the influences of environmental factors on the assembly of the core microbiome, which were displayed by heatmap.

Network analysis was carried out to explore the co-occurrence patterns of bacterial taxa and the ecological role of the core microbiome^{32,59}. Spearman correlations between every two OTUs were estimated, and statistically robust correlations with the correlation coefficient (r) > 0.6 and false discovery rate (FDR)-corrected P -value < 0.01 were selected to form a correlation network. Each node in the network represents an OTU and each edge stands for a statistically robust correlation between OTUs. Average path length (APL), network diameter (ND), graph density (GD), and clustering coefficient (CC) were calculated to describe network topology using the IGRAPH package⁶⁰ in R environment. Meanwhile, 10,000 random Erdős–Rényi networks were generated with random probability of connections assigned to any node and compared with the real-world network in terms of topology. Furthermore, sub-network was generated for the core microbiome using the subgraph procedure in the IGRAPH package⁶⁰, and network-level topology was analyzed to verify the central position of the core microbiome in the network.

In our study, we have applied the most common normalizations, namely total sum scaling (i.e., converting counts to proportions by dividing each species count in a sample by the total sum of counts from within that sample) and subsampling⁶¹, both of which effectively convert counts into relative measures of abundance⁶². We acknowledge that the normalization is important when calculating correlations among OTUs, as mentioned in Carr, *et al.*⁶². In addition, the state-of-the-art methods such as SparCC, SPIEC-EASI and FlashWeave seem to be the most appropriate options to construct the network. In the future work, these new methods and normalization of microbial taxa should be taken into account when conducting the network analysis.

To further explore the co-occurrence patterns of soil microbes at the genus level and identify the keystone genera, the co-occurrence network at the genus level was constructed. The nodes represent genera and edges stand for strong ($r > 0.6$) and significant (FDR-corrected $P < 0.01$) Spearman correlations between genera. The networks were visualized using Gephi. The genera with the highest betweenness centrality (a measure for the relevance of a node as capable of holding together communicating nodes) were considered keystone species⁴².

Unless otherwise indicated, all statistics analyses were performed in the R environment (<http://www.r-project.org>).

Data availability

All sequences were deposited in the NCBI Sequence Read Archive database (Biosample number: SAMN06105854-SAMN06105874).

Received: 28 May 2019; Accepted: 9 March 2020;

Published online: 07 April 2020

References

1. Smith, P. *et al.* Greenhouse gas mitigation in agriculture. *Philos. T R. Soc. B* **363**, 789–813 (2008).
2. Dubey, A. *et al.* Soil microbiome: a key player for conservation of soil health under changing climate. *Biodivers. Conserv.* **28**, 2405–2429 (2019).
3. Van Der Heijden, M. G., Bardgett, R. D. & Van Straalen, N. M. The unseen majority: soil microbes as drivers of plant diversity and productivity in terrestrial ecosystems. *Ecol. Lett.* **11**, 296–310 (2008).

4. Singh, J. S. & Gupta, V. K. Soil microbial biomass: A key soil driver in management of ecosystem functioning. *Sci. Total. Env.* **634**, 497–500 (2018).
5. Martiny, J. B., Eisen, J. A., Penn, K., Allison, S. D. & Horner-Devine, M. C. Drivers of bacterial β -diversity depend on spatial scale. *Proc. Natl Acad. Sci. USA* **108**, 7850–7854 (2011).
6. Wang, K. *et al.* Bacterial biogeography in the coastal waters of northern Zhejiang, East China Sea is highly controlled by spatially structured environmental gradients. *Env. Microbiol.* **17**, 3898–3913 (2015).
7. Jiao, S. *et al.* Bacterial communities in oil contaminated soils: Biogeography and co-occurrence patterns. *Soil. Biol. Biochem.* **98**, 64–73 (2016).
8. Karimi, B. *et al.* Biogeography of soil bacteria and archaea across France. *Sci. Adv.* **4**, eaat1808 (2018).
9. Delgado-Baquerizo, M. Obscure soil microbes and where to find them. *ISME J.* **13**, 2120–2124 (2019).
10. Fierer, N. & Jackson, R. B. The diversity and biogeography of soil bacterial communities. *Proc. Natl Acad. Sci. USA* **103**, 626 (2006).
11. Yao, H. *et al.* Multi-factorial drivers of ammonia oxidizer communities: evidence from a national soil survey. *Env. Microbiol.* **15**, 2545–2556 (2013).
12. Delgado-Baquerizo, M. *et al.* Microbial diversity drives multifunctionality in terrestrial ecosystems. *Nat. Commun.* **7**, 10541 (2016).
13. Zhou, J. *et al.* Temperature mediates continental-scale diversity of microbes in forest soils. *Nat. Commun.* **7**, 12083 (2016).
14. Power, J. F. *et al.* Microbial biogeography of 925 geothermal springs in New Zealand. *Nat. Commun.* **9**, 2876 (2018).
15. Shade, A. & Handelsman, J. Beyond the Venn diagram: the hunt for a core microbiome. *Env. Microbiol.* **14**, 4–12 (2012).
16. Fuhrman, J. A. Microbial community structure and its functional implications. *Nature* **459**, 193–199 (2009).
17. Hanson, C. A., Fuhrman, J. A., Horner-Devine, M. C. & Martiny, J. B. Beyond biogeographic patterns: processes shaping the microbial landscape. *Nat. Rev. Microbiol.* **10**, 497–506 (2012).
18. Toju, H. *et al.* Core microbiomes for sustainable agroecosystems. *Nat. Plants* **4**, 247–257 (2018).
19. Peiffer, J. A. *et al.* Diversity and heritability of the maize rhizosphere microbiome under field conditions. *Proc. Natl Acad. Sci. USA* **110**, 6548–6553 (2013).
20. Griffiths, R. I. *et al.* The bacterial biogeography of British soils. *Env. Microbiol.* **13**, 1642–1654 (2011).
21. Ren, B. *et al.* Soil pH and plant diversity shape soil bacterial community structure in the active layer across the latitudinal gradients in continuous permafrost region of Northeastern China. *Sci. Rep.* **8**, 5619 (2018).
22. Xiong, J. *et al.* Geographic distance and pH drive bacterial distribution in alkaline lake sediments across Tibetan Plateau. *Env. Microbiol.* **14**, 2457–2466 (2012).
23. Kemmitt, S. J., Wright, D., Goulding, K. W. & Jones, D. L. pH regulation of carbon and nitrogen dynamics in two agricultural soils. *Soil. Biol. Biochem.* **38**, 898–911 (2006).
24. Jiao, S., Xu, Y., Zhang, J., Hao, X. & Lu, Y. Core Microbiota in Agricultural Soils and Their Potential Associations with Nutrient Cycling. *Msystems* **4**, e00313–00318 (2019).
25. Zhao, M. *et al.* Microbial mediation of biogeochemical cycles revealed by simulation of global changes with soil transplant and cropping. *ISME J.* **8**, 2045–2055 (2014).
26. Evans, S. E. & Wallenstein, M. D. Soil microbial community response to drying and rewetting stress: does historical precipitation regime matter? *Biogeochemistry* **109**, 101–116 (2012).
27. Bubier, J. L. & Moore, T. R. An ecological perspective on methane emissions from northern wetlands. *Trends Ecol. Evol.* **9**, 460–464 (1994).
28. Zhang, K. *et al.* Effects of short-term warming and altered precipitation on soil microbial communities in alpine grassland of the Tibetan Plateau. *Front Microbiol* **7** (2016).
29. DeBruyn, J. M., Nixon, L. T., Fawaz, M. N., Johnson, A. M. & Radosevich, M. Global biogeography and quantitative seasonal dynamics of Gemmatimonadetes in soil. *Appl. Env. Microb.* **77**, 6295–6300 (2011).
30. Morlon, H. *et al.* A general framework for the distance-decay of similarity in ecological communities. *Ecol. Lett.* **11**, 904–917 (2008).
31. Anderson, M. J. *et al.* Navigating the multiple meanings of beta diversity: a roadmap for the practicing ecologist. *Ecol. Lett.* **14**, 19–28 (2011).
32. Jiao, S., Xu, Y., Zhang, J. & Lu, Y. Environmental filtering drives distinct continental atlases of soil archaea between dryland and wetland agricultural ecosystems. *Microbiome* **7**, 15 (2019).
33. 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**, 27 (2018).
34. Nekola, J. C. & White, P. S. The distance decay of similarity in biogeography and ecology. *J. Biogeogr.* **26**, 867–878 (1999).
35. Wang, X. B. *et al.* Habitat-specific patterns and drivers of bacterial beta-diversity in China's drylands. *ISME J.* **11**, 1345–1358 (2017).
36. Neher, D. A. Soil community composition and ecosystem processes - Comparing agricultural ecosystems with natural ecosystems. *Agroforest Syst.* **45**, 159–185 (1999).
37. Hu, H. W., Zhang, L. M., Yuan, C. L. & He, J. Z. Contrasting Euryarchaeota communities between upland and paddy soils exhibited similar pH-impacted biogeographic patterns. *Soil. Biol. Biochem.* **64**, 18–27 (2013).
38. Liu, L., Yang, J., Yu, Z. & Wilkinson, D. M. The biogeography of abundant and rare bacterioplankton in the lakes and reservoirs of China. *ISME J* (2015).
39. Verreydt, D. *et al.* Dispersal-mediated trophic interactions can generate apparent patterns of dispersal limitation in aquatic metacommunities. *Ecol. Lett.* **15**, 218–226 (2012).
40. Chase, J. M. & Myers, J. A. Disentangling the importance of ecological niches from stochastic processes across scales. *Philos. T Roy. Soc. B* **366**, 2351–2363 (2011).
41. Vellend, M. Conceptual synthesis in community ecology. *Q. Rev. Biol.* **85**, 183–206 (2010).
42. Vick-Majors, T. J., Priscu, J. C. & Amaral-Zettler, L. A. Modular community structure suggests metabolic plasticity during the transition to polar night in ice-covered Antarctic lakes. *ISME J.* **8**, 778–789 (2014).
43. Eiler, A., Heinrich, F. & Bertilsson, S. Coherent dynamics and association networks among lake bacterioplankton taxa. *ISME J.* **6**, 330–342 (2012).
44. Steele, J. A. *et al.* Marine bacterial, archaeal and protistan association networks reveal ecological linkages. *ISME J.* **5**, 1414–1425 (2011).
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. Topp, E., Mulbry, W. M., Zhu, H., Nour, S. M. & Cuppels, D. Characterization of s-triazine herbicide metabolism by a *Nocardioideis* sp. isolated from agricultural soils. *Appl. Env. Microb.* **66**, 3134–3141 (2000).
47. Yu, H. *et al.* Microbial community succession and lignocellulose degradation during agricultural waste composting. *Biodegradation* **18**, 793–802 (2007).
48. Zitouni, A. *et al.* *Nocardioideis* and *Saccharothrix* genera in Saharan soils in Algeria: isolation, biological activities and partial characterization of antibiotics. *Res. Microbiol.* **156**, 984–993 (2005).
49. Zhang, H. *et al.* *Pedobacter nanyangensis* sp. nov., isolated from herbicide-contaminated soil. *Int. J. Syst. Evol. Micr* **65**, 3517–3521 (2015).
50. Bao, S. D. *Soil and agricultural chemistry analysis* (China Agriculture Press, Beijing, 2000).
51. Caporaso, J. G. *et al.* Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proc. Natl Acad. Sci. USA* **108**, 4516–4522 (2011).

52. Edgar, R. C., Haas, B. J., Clemente, J. C., Quince, C. & Knight, R. UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* **27**, 2194–2200 (2011).
53. Edgar, R. C. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**, 2460–2461 (2010).
54. Edgar, R. C. J. N. M. UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nat. Methods* **10**, 996 (2013).
55. Parks, D. H. *et al.* GenGIS 2: Geospatial analysis of traditional and genetic biodiversity, with new gradient algorithms and an extensible plugin framework. *PLoS one* **8**, e69885 (2013).
56. Oksanen, J. *et al.* vegan: Community Ecology Package. *R package version 2.2-1* (2015).
57. Jiao, S., Chen, W. M. & Wei, G. H. Biogeography and ecological diversity patterns of rare and abundant bacteria in oil-contaminated soils. *Mol. Ecol.* **26**, 5305–5317 (2017).
58. Shade, A. *et al.* Conditionally rare taxa disproportionately contribute to temporal changes in microbial diversity. *MBio* **5**, e01371–01314 (2014).
59. Delgado-Baquerizo, M. *et al.* A global atlas of the dominant bacteria found in soil. *Science* **359**, 320–325 (2018).
60. Csardi, G. & Nepusz, T. The igraph software package for complex network research. *InterJournal, Complex. Syst.* **1695**, 1–9 (2006).
61. Weiss, S. *et al.* Normalization and microbial differential abundance strategies depend upon data characteristics. *Microbiome* **5**, 27 (2017).
62. Carr, A., Diener, C., Baliga, N. S. & Gibbons, S. M. Use and abuse of correlation analyses in microbial ecology. *ISME J*, **1** (2019).

Acknowledgements

This work was supported by the National Science Foundation of China (31570493). We thank Dr. Shuo Jiao for comments that improved this manuscript.

Author contributions

All authors contributed intellectual input and assistance to this study and the manuscript preparation. W.C. conceived the study. W.T., J.W. and W.B. collected the samples. W.C. and W.T. performed data analysis and wrote the manuscript. J.Q. contributed to manuscript revisions.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41598-020-62919-7>.

Correspondence and requests for materials should be addressed to W.C.

Reprints and permissions information is available at 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) 2020