



# Balance between community assembly processes mediates species coexistence in agricultural soil microbiomes across eastern China

Shuo Jiao<sup>1</sup> · Yunfeng Yang<sup>2</sup> · Yiqin Xu<sup>1</sup> · Jie Zhang<sup>1</sup> · Yahai Lu<sup>1</sup>

Received: 12 February 2019 / Revised: 17 August 2019 / Accepted: 20 September 2019 / Published online: 14 October 2019  
© The Author(s), under exclusive licence to International Society for Microbial Ecology 2019

## Abstract

Revealing the linkages between community assembly and species coexistence, which is crucial for the understanding of ecosystem diversity and functioning, is a fundamental but rarely investigated subject in microbial ecology. Here we examined archaeal, bacterial, and fungal community assembly in adjacent pairs of maize (water-unsaturated) and rice (water-saturated) fields across different habitats and regions throughout Eastern China. The high-throughput sequencing dataset was analyzed by variation partitioning, null model, and neutral community model analyses. We demonstrated that microbial community assembly was governed more by species sorting than by dispersal limitation in maize fields, and to a lesser extent in rice fields. The relative importance of species sorting in maize soils was greater at low latitudes than at high latitudes, while rice soils exhibited an opposite trend. Microbial co-occurrence associations tended to be higher when communities were primarily driven by dispersal limitation relative to species sorting. There were greater community dissimilarities between maize and rice soils in low-latitude regions, which was consistent with the higher proportion of negative edges in the correlation networks. The results indicate that a balance between species sorting and dispersal limitation mediates species coexistence in soil microbiomes. This study enhances our understanding of contemporary coexistence theory in microbial ecosystems.

## Introduction

Understanding the mechanisms that underpin microbial community diversity and biogeographic patterns is a key topic in ecology [1, 2]. Over the past few decades, the biogeographic patterns of soil microbial communities have been reported for a wide range of habitats [3–5] at regional [3, 4, 6], continental [7], and global scales [8–10]. Currently, microbial biogeography studies are turning their attention toward the processes that drive microbial

community assembly. There is an ongoing debate regarding the contributions of deterministic and stochastic processes in the assembly of microbial communities [11, 12]. Deterministic processes mainly involve ecological selection, with abiotic and biotic factors representing distinct niches that influence community assembly [13]. Conversely, stochastic processes demonstrate that species can co-occur with considerably overlapping niches when their competitive capacities are closely matched, and when random changes are not associated with environmentally derived fitness [14, 15]. Stochastic processes, which involve random birth, death, and dispersal events, can greatly influence the distribution of microbial communities and result in species composition patterns that are indistinguishable from patterns arising randomly [16–18]. However, characterizing the processes underpinning spatial variation in microbial communities across ecosystems is challenging [15, 19, 20].

One of the most common biogeographic patterns, distance-decay relationship (DDR), refers to decreasing community similarity with increasing geographic distance. DDR provides a directional model for variations in  $\beta$ -diversity across spatial scales [21, 22]. It is widely acknowledged that both deterministic and stochastic

**Supplementary information** The online version of this article (<https://doi.org/10.1038/s41396-019-0522-9>) contains supplementary material, which is available to authorized users.

✉ Yahai Lu  
luyh@pku.edu.cn

<sup>1</sup> College of Urban and Environmental Sciences, Peking University, Beijing, PR China

<sup>2</sup> State Key Joint Laboratory of Environment Simulation and Pollution Control, School of Environment, Tsinghua University, Beijing, PR China

processes influence DDR [4, 16]; however, the relative importance of the two processes varies across habitats [20]. Collectively, species sorting, dispersal limitation (alongside drift), homogeneous dispersal (also known as mass effect), and drift (acting alone) simultaneously influence distance-decay patterns [1, 13, 23–25]. Homogeneous dispersal and drift may have a minor role in the assembly of microbial communities at large spatial scales [23, 24]. Consequently, species sorting and dispersal limitation are the two major processes that influence DDRs when communities are surveyed at large spatial scales [24, 25]. The slopes of DDRs could vary based on habitat, which reflects different rates of species turnover in various locations [4].

Mechanisms of microbial community assembly have been investigated in numerous terrestrial habitats [10], including hypolithic soils [26], the hyporheic zone [27], successional soils [28], a salt marsh [11], incipient basaltic soil [29], subsurface sediment [23], a grassland [30], and oil-contaminated soils [3, 31], and aquatic ecosystems [12, 32]. Agricultural fields are typical human-managed terrestrial ecosystems that play a key role in global food security for an increasing global population [33]. Recent studies have demonstrated that deterministic processes play a dominant role in bacterial community assembly in wheat [34] and soybean fields [35]. However, contrasting water management practices are applied to different crop systems, forming water-unsaturated (e.g., maize) and water-saturated (e.g., rice) fields, which may lead to distinct microbial diversity patterns. Such habitat differentiation can influence microbial dispersal because saturated water flows potentially facilitate greater dispersal than isolated habitats or solid substrates [36]. Therefore, the contributions of species sorting and dispersal limitation to microbial assembly might vary between terrestrial and aquatic ecosystems [31, 32]. It remains unclear whether there are correlations between the microbial diversity patterns and the assembly processes across unsaturated and saturated agro-ecosystems.

Uncovering high species diversity, which has been extensively explored in macroorganism communities, is an enduring challenge for ecologists and has inspired many theories of species coexistence [15, 37–39]. Coexistence theories can be divided into two categories. One is based on the differences in ecological niches of co-occurring species (niche-based theory) [40] and the other relies on dispersal and stochastic demographic processes (neutral theory) [15]. Examination of the contributions of deterministic and stochastic processes in microbial community assembly could elucidate the ecological strategies of co-occurring species [37, 41]. Microorganisms, which have microscopic sizes and high dispersal capacity, could display complex interaction webs within an ecological niche [42]. Considering the critical roles of microorganisms in soil ecosystems [43, 44], revealing the intrinsic linkages between assembly

processes and species coexistence could facilitate the management of microbial communities for enhanced ecosystem service provisioning such as agricultural production and habitat restoration in response to environmental changes [19, 45]. However, despite increased interest in microbial community assembly processes [1, 2, 20, 23, 24, 28], we still lack an adequate understanding of how community assembly processes is related with the ecological strategies of co-occurring species in microbial ecosystems.

The aims of the present study were to (I) evaluate the relative importance of species sorting and dispersal limitation in shaping soil microbial communities across different habitats and regions in agro-ecosystems; and (II) to uncover associations between assembly processes of microbial communities and ecological strategies of co-occurring species across agro-ecosystems. We carried out a large-scale soil survey on the composition of soil archaeal, bacterial, and fungal communities, and 27 edaphic variables in adjacent pairs of maize (water-unsaturated) and rice (water-saturated) fields across Eastern China. Maize and rice are typical crops that are cultivated extensively across China, and they offer model systems for adjacent paired sampling strategies that control for the influence of spatial scale and climatic factors. Species pairs or assemblages that co-occur may exhibit similar ecological characteristics and life strategies [46, 47]; thus, co-occurring species with the low levels of environmental filtering could be more adapted in homogeneous habitats. Considering the contrasting water management practices and cultivation history, we hypothesize that (I) the relative importance of community assembly processes varies among habitats and regions; and (II) microbial communities less influenced by species sorting exhibit higher species coexistence.

## Materials and methods

### Soil sampling and data collection

One hundred and twenty-seven sites were selected in agricultural fields that had been under at least 10 years of maize or rice cultivation. There were 115 paired sites, five maize-only sites, and seven rice-only sites, yielding 120 maize and 122 rice soil samples (Supplementary Fig. S1). The sampling sites extend from 18.30°N to 48.35°N and 87.61°E to 99.91°E across Eastern China. Each of the paired sites is located in adjacent maize or rice field <5 km apart. In July–September 2017 (at the milk stage of maize and early grain-filling stage of rice), three 100 m<sup>2</sup> plots were sampled at each site, and five soil cores per plot obtained at a depth of 0–15 cm were combined.

Standard test methods were adopted to measure soil pH, moisture, cation exchange capacity, organic matter, dissolved

organic carbon, total nitrogen, available nitrogen, nitrate-nitrogen ( $\text{NO}_3$ ), ammonium-nitrogen ( $\text{NH}_4$ ), total phosphorus, available phosphorus, total potassium, available potassium, microbial biomass carbon, microbial biomass nitrogen, total iron, available iron, total sulfur, available sulfur (AS), total manganese (TMn), available manganese, total copper, available copper, total zinc, available zinc, and particle size distribution (clay, silt, and sand contents), as previously described [34, 48]. We obtained climatic data including mean annual temperature, mean diurnal temperature range, temperature annual range, mean annual precipitation, and precipitation seasonality for all sampling sites from the Worldclim database ([www.worldclim.org](http://www.worldclim.org)). In addition, we estimated the aridity index (AI, precipitation/evapotranspiration) at each site using the Global Potential Evapotranspiration database [49], which is based on interpolations provided by WorldClim [50].

Soil archaeal, bacterial, and fungal 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 a region of the 16S rRNA gene for archaea, 16S rRNA gene for bacteria, and a region of the internal transcribed spacer 1 gene for fungi. Corresponding polymerase chain reaction assays were performed with the Arch519F/Arch915R, 515F/907R, and ITS5-1737F/ITS2-2043R primer pairs, respectively [51]. Sequencing was performed on the Illumina HiSeq2500 platform (Illumina Inc., San Diego, CA, USA). The acquired sequences were filtered for quality control as previously described [52, 53]. Any chimeric sequences were removed using the USEARCH tool based on the UCHIME algorithm [54]. Sequences were then split into operational taxonomic units (OTUs) at a 3% dissimilarity level using the UPARSE pipeline [54]. OTUs with fewer than two sequences were removed, and their representative sequences were assigned to taxonomic lineages using the RDP classifier within the SILVA database (release 128) for bacteria and archaea and UNITE + INSD (UNITE and the International Nucleotide Sequence Databases) for fungi.

## Statistical analysis

To compare the assembly processes of microbial communities across different regions, we divided the samples into two groups based on latitude (Supplementary Fig. S1): high-latitude group (latitude  $> 32^\circ$ ) and low-latitude group (latitude  $< 32^\circ$ ). This separation generally corresponds to the Qinling Mountains–Huaihe River Line (latitude  $\approx 32^\circ$ ), an important geographical (landform, climate, and soil) boundary in China [55]. Beta-diversities of archaeal, bacterial, and fungal communities were estimated based on

Bray–Curtis distances between samples, and the distances were calculated using OTU tables resampled to a minimum number of sequences from each sample (36,880 for archaea, 27,712 for bacteria, and 30,369 for fungi). To determine the significance in difference of community composition between two geographical groups, a similarity analysis (ANOSIM) was performed using the `anosim` function in “vegan” package in R [56]. The Bray–Curtis dissimilarities for archaeal, bacterial, and fungal communities between maize and rice fields or between the two geographical groups were examined using Wilcoxon rank-sum tests, performed using the `wilcox.test` function in “stats” package in R [57]. DDRs were calculated as the slopes of ordinary least-squares regressions for the relationships between geographic distances and community similarities (1—dissimilarity of the Bray–Curtis metric). All maps were created using the `geom_polygon` function in “ggplot2” package in R [58].

To determine the relative importance of species sorting and dispersal limitation, we performed a variation-partitioning analysis to separate community variation (using pairwise Bray–Curtis dissimilarity) into environmental effect and spatial effect [59]. Spatial variables were obtained from geographic distances using Moran’s eigenvector maps [60], also known as principal coordinates analysis of neighbor matrices (PCNM) [61]. The derived PCNM variables (e.g., PCNM1 and PCNM2) could capture all the detectable spatial scales in the dataset [60, 61], which were obtained using the `pcnm` function in “vegan” package in R [56]. Forward selection procedures, based on a constrained analysis of the principal coordinates model, were used to select environmental and spatial variables and performed using the `ordiR2step` function in “vegan” package in R [56]. The forward selection was halted if the significance level ( $P < 0.05$ ) of the model was reached or if no improvement was observed in the selection criterion ( $R^2$ ) of the model when adding variables. Subsequently, a two-way permutational multivariate analysis of variance (PERMANOVA) was performed using the selected variables. Pure environmental variation without a spatial component represents the effect of species sorting, while pure spatial variation without an environmental component represents the effect of dispersal limitation. To evaluate the relative importance of species sorting versus dispersal limitation between maize and rice fields and across different regions, the ratio of species sorting to dispersal limitation effect (SDER) was calculated.

Null model analysis was carried out using the framework described by Stegen et al. [23] to classify community pairs into underlying drivers of species sorting (or selection), dispersal limitation, homogeneous dispersal, and drift. The null model expectation was generated using 999 randomizations. The variation of both phylogenetic diversity and taxonomic

diversity was measured using null model-based phylogenetic and taxonomic  $\beta$ -diversity metrics, namely  $\beta$ -nearest taxon index ( $\beta$ NTI) and Bray–Curtis-based Raup–Crick ( $RC_{Bray}$ ). A significant deviation (i.e.,  $|\beta$ NTI  $> 2$ ) indicates the dominance of selection processes.  $\beta$ NTI  $< -2$  indicates significantly less phylogenetic turnover than expected (i.e., homogeneous selection) while  $\beta$ NTI  $> +2$  indicates significantly more phylogenetic turnover than expected (i.e., variable selection). Subsequently,  $RC_{Bray}$  was used to further partition the pairwise comparisons that were not assigned to selection (i.e.,  $|\beta$ NTI  $< 2$ ). The relative influence of homogenizing dispersal was quantified as the fraction of pairwise comparisons with  $|\beta$ NTI  $< 2$  and  $RC_{Bray} < -0.95$ . Dispersal limitation was quantified as the fraction of pairwise comparisons with  $|\beta$ NTI  $< 2$  and  $RC_{Bray} > 0.95$ . The fractions of all pairwise comparisons with  $|\beta$ NTI  $< 2$  and  $|RC_{Bray}| < 0.95$  were used to estimate influence of “undominated” assembly, which mostly consists of weak selection, weak dispersal, diversification, and/or drift [2, 28, 62]. To make variation partitioning and null model analysis comparable, we also estimated the SDER using the percentage of pairwise communities governed by species sorting divided by the percentage of pairwise communities governed by dispersal limitation. Here, the percentage of species sorting was calculated as the sum of homogeneous and variable selection.

A neutral community model was used to determine the contribution of stochastic processes to microbial community assembly by predicting the relationship between the frequency with which taxa occur in a set of local communities and their abundance across the wider metacommunity [63]. The model predicts that abundant taxa are more likely to be dispersed by chance and widespread across metacommunity, while rare taxa would be lost in different local communities due to ecological drift. In the model, the estimated migration rate is a parameter for evaluating the probability that a random loss of an individual in a local community would be replaced by dispersal from the metacommunity, and, therefore, is a measure of dispersal limitation. Higher  $m$  values indicate that microbial communities are less dispersal limited [63, 64]. The formula is as follows:

$$Freq_i = 1 - I(1/N | N * m * p_i, N * m * (1 - p_i))$$

where  $Freq_i$  is the occurrence frequency of taxon  $i$  across communities;  $N$  is the number of individuals per community;  $m$  is the estimated migration rate;  $p_i$  is the average relative abundance of taxon  $i$  across communities; and  $I()$  is the probability density function of beta distribution.  $R^2$  indicates the fit of the parameter based on nonlinear least-squares fitting. The overall fit of the model to the observed data was assessed by comparing the sum of squares of residuals,  $SS_{err}$ , with the total sum of squares,  $SS_{total}$ : model fit =  $1 - SS_{err}/SS_{total}$  (generalized  $R$ -squared; [64, 65]). To

determine whether the model was based on only the random sampling of the source metacommunity, we compared the fit of the neutral model with the fit of a binomial distribution model [63, 64]. Sampling from a binomial distribution represents the case where local communities are random subsets of the metacommunity in the absence of drift and dispersal limitation processes [64]. The Akaike information criterion of each model was calculated based on 1000 bootstrap replicates. R code used to fit the model and calculate goodness-of-fit statistics was according to Burns et al. [64]. In all the habitats, the neutral model outperformed the binomial distribution model, suggesting that passive dispersal and ecological drift had an impact beyond the random sampling of the source community (Supplementary Fig. S2).

To help reveal the patterns of species sorting and dispersal limitation and their influence on microbial communities based on habitats and regions, we estimated Levins' niche breadth ( $B$ ) index [66] for the microbial group's members according to the formula:

$$B_j = 1 / \sum_{i=1}^N P_{ij}^2$$

where  $B_j$  represents the habitat niche breadth of OTU  $j$  in a metacommunity;  $N$  is the total number of communities in each metacommunity;  $P_{ij}$  is the proportion of OTU  $j$  in community  $i$  [24, 67]. A high  $B$ -value for a given OTU indicates its wide habitat niche breadth. The community-level  $B$ -value ( $B_{com}$ ) was calculated as the average of  $B$ -values from all taxa occurring in one community [24]. We expect a microbial group with a wider niche breadth to be more metabolically flexible at the community level [24, 67]. The analysis was conducted using the “niche.width” function in “spaa” package in R [68].

To identify significantly enriched taxa of all three microbial groups, we applied Wilcoxon rank-sum test (`wilcox.test` function in “stats” package in R [57]) of the relative abundance of each OTU ( $> 0.01\%$ ) between the maize and rice fields. OTUs with significantly higher ( $P < 0.05$ ) abundance in maize samples were grouped as maize-enriched OTUs, whereas those with significantly ( $P < 0.05$ ) higher abundance in rice samples were categorized as rice-enriched OTUs, and the remaining OTUs with no significant differences in relative abundance were categorized as ‘Others’. In addition, we identified significantly ( $P < 0.05$ ) enriched OTUs in the high- and low-latitude regions in both maize and rice fields. OTUs with significantly higher abundance in high-latitude regions were grouped as high latitude-enriched OTUs, whereas those with significantly higher abundance in low-latitude regions were categorized as low latitude-enriched OTUs, and the remaining OTUs with no significant differences in relative abundance were categorized as ‘Others’.

To estimate species coexistence across different habitats and regions, metacommunity co-occurrence networks consisting of all the members of the three microbial groups were constructed. To reduce rare OTUs in the dataset, we removed OTUs with a relative abundance <0.01%. Robust correlations with Spearman's correlation coefficients ( $\rho$ ) > 0.6 and false discovery rate-corrected  $p$ -values < 0.01 were used to construct networks. Each node represents one OTU, and each edge represents a strong and significant correlation between two nodes. To describe the topology of the networks, we calculated a set of metrics: average path length, network diameter, average degree, clustering coefficient, and graph density. Average path length refers to the average network distance between all pairs of nodes; network diameter refers to the greatest distance between the nodes that exist in the network; average degree refers to the average connections of each node with another unique node in the network; clustering coefficient represents the degree to which the nodes tend to cluster together; and graph density refers to the intensity of connections among nodes [69, 70]. Therefore, higher average degree, clustering coefficient, and graph density suggest a more connected network. In addition, lower average path lengths and diameters indicate closer associations in the network [47, 48]. The node-level topological features, including degree, betweenness, and closeness centrality, were calculated for each node. Degree refers to the number of connections for a particular node; betweenness centrality refers to the potential influence of a particular node on the connections of other nodes; and the closeness centrality of a node refers to the average distance of the node to any other node [71, 72]. High values of these topological features suggest a core position of a node in the network, while low values suggest a peripheral position [31, 48]. Sub-network analyses of archaeal, bacterial, and fungal communities were performed separately using the induced\_subgraph function in "igraph" package in R [73]. Networks were visualized using the interactive Gephi platform [74].

Microbial community dissimilarities between maize and rice fields were calculated for each pairwise comparison of soil samples and were mapped using ordinary kriging interpolation. The analysis was performed using the "automap" package in R [75], which performs the interpolation process by automatically estimating a semivariogram and performing kriging. We cross-validated the maps using the autoKrige.cv function in the "automap" package in R [75], which automatically isolates the model structure dataset. We extracted the predicted values for the selected soil samples and tested their correlation with the observed values based on a Pearson correlation analysis (cor.test function in "stats" package in R [57]). We also calculated the Spearman correlations between differences in the relative abundance of the major phyla versus differences in soil properties, which were estimated between maize and rice soils. We applied the multiple

regression model (lm function in "stats" package in R [57]) with variance decomposition analysis (calc.relimp function in the "relaimpo" package in R [76]) to estimate the importance of the differences in soil physicochemical properties in explaining the dissimilarities in microbial communities and differences in the relative abundance of selected phyla. The relationships among microbial taxa were estimated by establishing correlation networks for soil samples from the high- or low-latitude regions, which considered both positive (Spearman's  $\rho$  > 0.6) and negative (Spearman's  $\rho$  < -0.6) edges. Other parameters for the network construction are as described above.

All statistical analyses were performed in the R environment (v3.5.1; <http://www.r-project.org/>), using "vegan" [56], "stats" [57], "fdrtool" [77], "igraph" [73], "hmisc" [78], "ggplot2" [58], "spaa" [68], "picante" [79], "automap" [75], "relaimpo" [76], and "gplots" [80] packages.

## Results

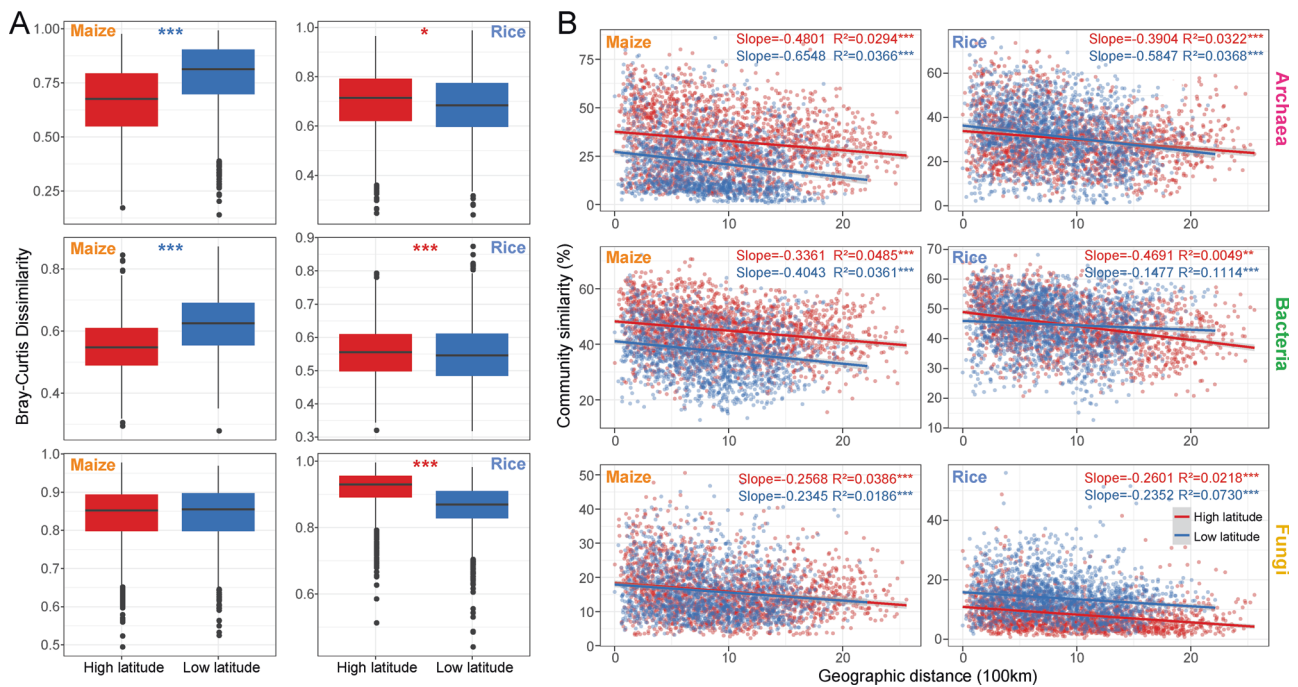
### Distance-decay of community similarity

We estimated the DDRs for archaeal, bacterial, and fungal communities in both rice and maize soils across eastern China spanning a geographic distance of 3679 km (Supplementary Fig. S3). Although the DDRs were significant ( $P < 0.001$ ), the fitness values were relatively low ( $R^2 < 0.1$ ), indicating weak decay of community similarity with geographic distance in agricultural soils. The slope of rice soils (-0.3819 for archaea; -0.3575 for bacteria; and -0.3277 for fungi) was steeper than that of the maize soils (-0.3520 for archaea; -0.2113 for bacteria; and -0.2467 for fungi). Maize soils exhibited significantly higher  $\beta$ -diversity in archaeal and bacterial communities, but lower  $\beta$ -diversity in fungal communities than rice soils (Supplementary Fig. S3).

Only minor differences in microbial community composition were observed between the two geographical groups based on the ANOSIM test ( $R_{\text{ANOSIM}} = 0.1679$ ,  $P < 0.001$  for archaea;  $R_{\text{ANOSIM}} = 0.1345$ ,  $P < 0.001$  for bacteria; and  $R_{\text{ANOSIM}} = 0.1978$ ,  $P < 0.001$  for fungi). Microbial  $\beta$ -diversity was significantly ( $P < 0.001$ ) higher in low-latitude maize soils than in high-latitude soils, while rice fields exhibited an opposite trend (Fig. 1a). Furthermore, we observed weak DDRs for each of the microbial groups across all regions (Fig. 1b).

### Assembly processes and coexistence in soil microbiome

Variation partitioning and null model analyses revealed that the SDERs of archaeal, bacterial, and fungal communities were consistently higher in maize than in rice fields; this



**Fig. 1** General patterns of microbial beta-diversity in soil samples from maize and rice fields in different regions. **a** Compositional variation between high-latitude and low-latitude regions in maize or rice fields based on Bray–Curtis distance. The boxplot shows community dissimilarities between high-latitude and low-latitude regions. Red asterisks indicate significantly higher values in high-latitude regions

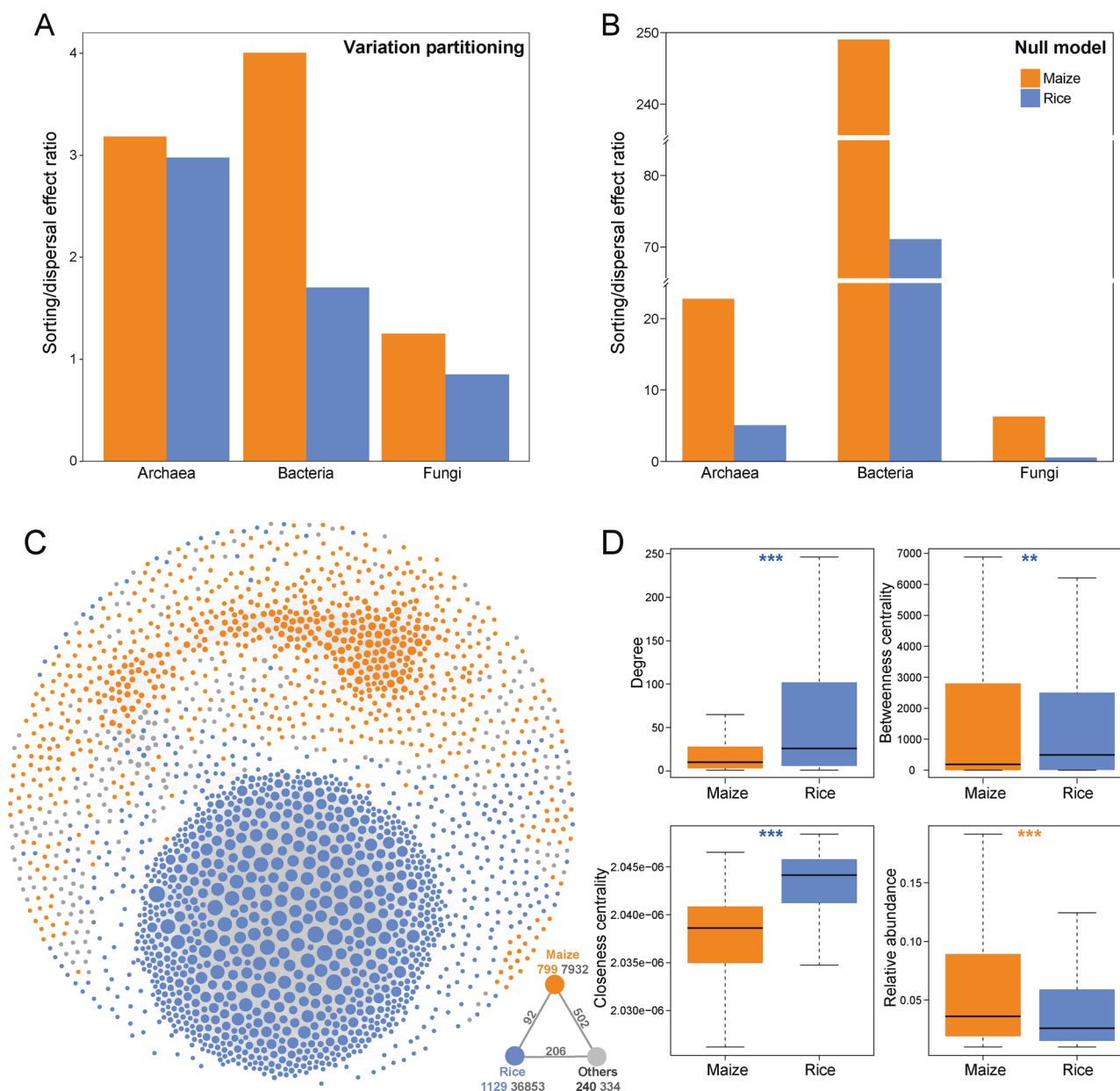
(\* $p < 0.05$ ; \*\*\* $p < 0.001$ ; Wilcoxon rank-sum test); blue asterisks indicate the significantly higher values in low-latitude regions. **b** Distance-decay curves showing Bray–Curtis similarity against geographic distances between sampling sites. Solid lines denote the ordinary least-squares linear regressions. Asterisks denote significant correlation (\*\*\* $p < 0.0001$ )

result suggests that the relative contribution of species sorting was higher in maize fields than in rice fields (Fig. 2a, b, Supplementary Figs. S4, S5, and Table S1). Null model analysis revealed a negligible degree of homogeneous dispersal (Supplementary Fig. S5), confirming its minor role in the assembly of microbial communities at large-scale distances. In addition, the archaeal, bacterial, and fungal communities fit the neutral community model, and the degree of fit was higher in rice than in maize fields. The migration rates tended to be higher in rice than in maize fields (Supplementary Table S2), suggesting that microbial communities were less limited by dispersal in rice fields.

Subsequently, we inferred a metacommunity co-occurrence network based on correlation relationships, capturing 45,919 associations among 2168 microbial OTUs (Fig. 2c). We identified 1265 and 1502 significantly enriched OTUs in maize and rice soils, respectively. The rice- and maize-enriched OTUs formed independent modules, and rice-enriched OTUs exhibited much closer interconnections than maize-enriched OTUs. In addition, we examined the node-level topological features of different groups of OTUs (Fig. 2d). Values of the topological features including degree, betweenness centrality, and closeness centrality were significantly higher ( $p < 0.01$ ) in rice-enriched OTUs than in maize-enriched OTUs.

We generated subnetworks for rice- and maize- enriched communities and calculated a set of network-level topological features (Supplementary Table S3). The average degree, clustering coefficient, and graph density were higher in the rice subnetwork than the maize subnetwork, suggesting that rice-enriched OTUs were more interconnected. The average path length and diameter were lower in the rice subnetwork, revealing closer relationships among rice-enriched communities. We also performed separate network analyses for the archaeal, bacterial, and fungal sub-communities (Supplementary Fig. S6). Rice-enriched OTUs consistently exhibited much closer interconnections than maize-enriched OTUs, which was confirmed by higher values of the topological features in rice-enriched OTUs. These results indicated that species co-occurred more frequently within rice-field microbial communities than in maize-field microbial communities.

Considering the significant differences in microbial  $\beta$ -diversity, we compared the SDERs of microbial communities between the high- and low-latitude regions in maize and rice fields. Both variation partitioning (Fig. 3, Supplementary Fig. S4, Tables S4 and S5) and null model analysis (Supplementary Figs. S5 and S7) showed that the ratio was higher in low-latitude regions than in high-latitude regions in maize soils, while rice soils exhibited an opposite trend.



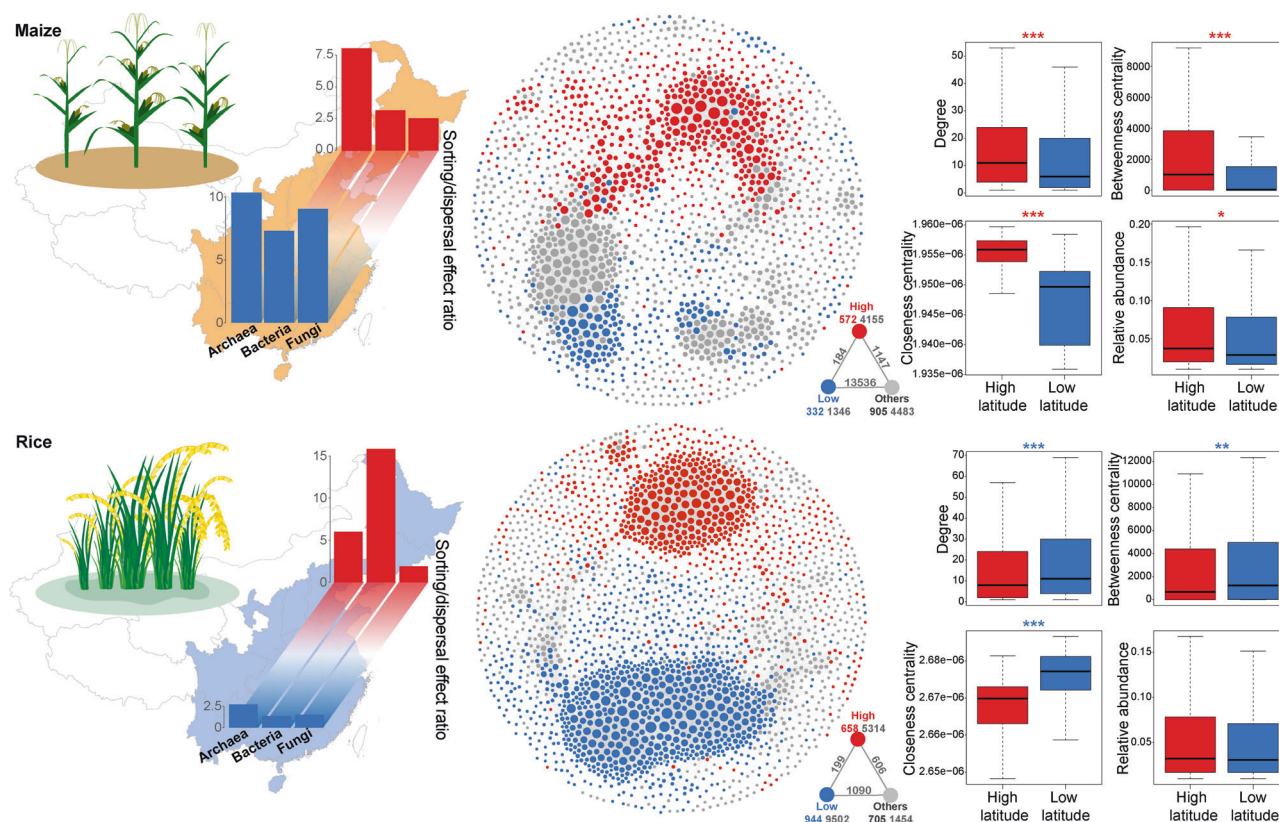
**Fig. 2** Community assembly and co-occurrence patterns in soil microbiomes in maize and rice fields. The sorting/dispersal effect ratio obtained based on variation partitioning (**a**) and null model (**b**) analysis. Metacommunity co-occurrence network of microbial taxa (**c**). A colored co-occurrence network is shown for maize- or rice-enriched taxa. A connection indicates a strong (Spearman's  $\rho > 0.6$ ) and significant (FDR-corrected  $p < 0.01$ ) correlation. The size of each node is proportional to the degree of the OTUs; the thickness of a connection between two nodes (i.e., an edge) is proportional to the value of the Spearman's correlation coefficient. The network figure on the bottom right is a summary of node-edge statistics. Colored numbers represent the number of nodes belonging to the corresponding category (i.e.,

there were 1129 nodes in the "Rice" category). Dark gray numbers indicate the number of inner-connections. The numbers adjacent to edge connections represent cross-group interactions. **d** Unique node-level topological features of different taxa categories, specifically the degree, betweenness, and closeness centrality, and the relative abundance of maize- and rice-enriched taxa. Blue asterisks indicate the significantly high values in rice-enriched taxa (\*\* $p < 0.01$ ; \*\*\* $p < 0.001$ ; Wilcoxon rank-sum test); orange asterisks indicate the significantly high properties in maize-enriched taxa. Maize, maize-enriched OTUs; Rice, rice-enriched OTUs; Others, OTUs not specific to a group

The microbial communities fit the neutral community model, and the degree of fit was higher in maize soils from the high-latitude region than the low-latitude region, and rice soils exhibited an opposite trend again. Migration rates exhibited a similar trend (Table 1), which suggested that microbial

communities were less limited by dispersal at high latitudes for maize soils and at low latitudes for rice soils.

To estimate microbial coexistence in different regions, metacommunity co-occurrence networks were established for maize and rice soils (Fig. 3). We also identified



**Fig. 3** Community assembly and co-occurrence patterns of soil microbiomes in high- and low-latitude regions in maize and rice fields. The sorting/dispersal effect ratio obtained based on variation partitioning and null model analysis (left panels). Metacommunity co-occurrence networks of microbial taxa in maize and rice soils (middle panels). The networks are colored based on high latitude- or low latitude-enriched taxa. A connection indicates a strong (Spearman's  $\rho > 0.6$ ) and significant (FDR-corrected  $p < 0.01$ ) correlation. The size of each node is proportional to the degree of the OTUs; the thickness of a connection between two nodes (i.e., an edge) is proportional to the value of Spearman's correlation coefficient. The network figure on the bottom right is a summary of node-edge statistics. Colored numbers represent the number of nodes belonging to the corresponding

category (i.e., there were 332 nodes in the "Low" latitude category). Dark gray numbers indicate the number of inner connections. The numbers adjacent to edge connections represent cross-group interactions. Unique node-level topological features of different taxa categories, specifically the degree, betweenness, and closeness centrality, and the relative abundance of high latitude- or low latitude-enriched taxa (right panels). Red asterisks indicate the significantly higher values significantly in high latitude-enriched taxa ( $*p < 0.05$ ;  $**p < 0.01$ ;  $***p < 0.001$ ; Wilcoxon rank-sum test); blue asterisks indicate the significantly higher properties in low latitude-enriched taxa. High, high latitude-enriched OTUs; Low, low latitude-enriched OTUs; Others, OTUs not specific to a group

significantly enriched OTUs in the high and low-latitude regions in maize and rice fields. In maize fields, 809 and 509 significantly enriched OTUs were selected in high-latitude and low-latitude regions, respectively. In rice fields, 911 and 1106 significantly enriched OTUs were selected in the high-latitude and low-latitude regions, respectively. OTUs enriched in different regions formed independent modules. High latitude-enriched OTUs in maize fields and low latitude-enriched OTUs in rice fields exhibited much more interconnections, as confirmed by the node-level and network-level topological features (Fig. 3, Supplementary Table S6). The subnetworks for different microbial groups exhibited similar results (Supplementary Figs. S8 and S9). Overall, these results indicate that microbial co-occurrence is more frequent in high-latitude maize soils and low-latitude rice soils.

The community-level habitat niche breadths ( $B_{com}$ ) were estimated to reveal the contributions of species sorting and dispersal limitation to microbial community assembly. Significantly higher mean  $B_{com}$  values were observed in archaeal, bacterial, and fungal subcommunities of rice fields than in maize fields (Fig. 4). In maize fields, the mean  $B_{com}$  values were significantly higher in the high-latitude regions compared with the low-latitude regions; however, in rice fields, an opposite trend was observed for the mean  $B_{com}$  values, except for the archaeal community.

### Drivers of microbial community dissimilarities between maize and rice fields

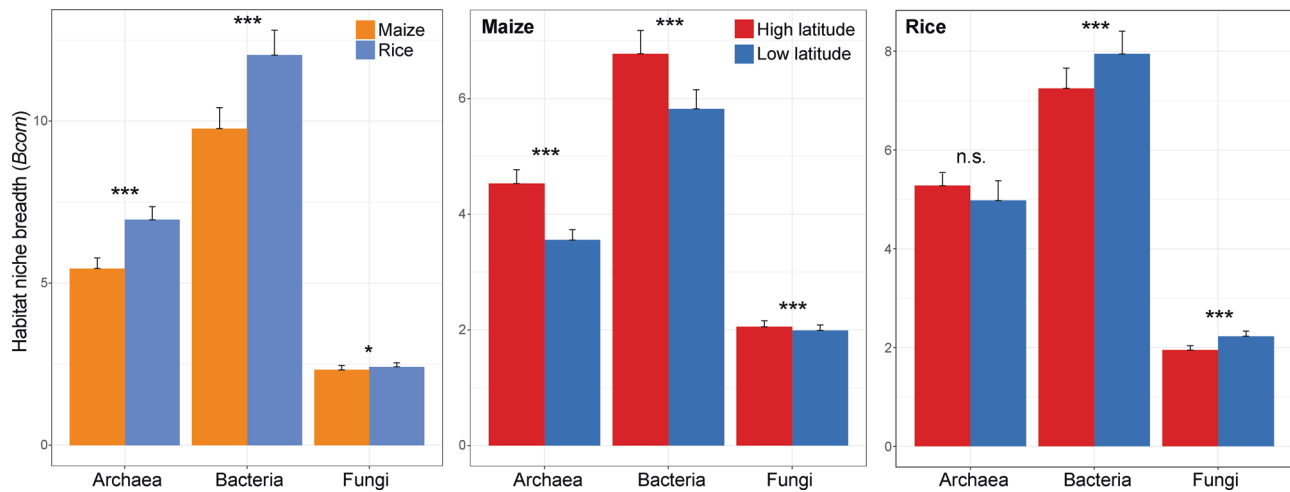
Since distinct assembly processes in microbial communities were observed in different regions in both maize and rice



**Table 1** Fit of the neutral model in microbial communities across different regions in maize and rice fields

Habitats	Region	Archaea		Bacteria		Fungi	
		$m$	$R^2$	$m$	$R^2$	$m$	$R^2$
Maize	High latitude	0.0498	0.8300	0.4928	0.8644	0.0171	0.4511
	Low latitude	0.0257	0.7219	0.3623	0.8517	0.0169	0.4364
Rice	High latitude	0.0666	0.8074	0.5784	0.8661	0.0126	0.3273
	Low latitude	0.0848	0.8287	0.6459	0.8707	0.0306	0.5422

$m$  indicates the estimated migration rate and  $R^2$  indicates the fit to the neutral model

**Fig. 4** Comparison of mean habitat niche breadths ( $B_{com}$ ) in all taxa in archaeal, bacterial, and fungal community samples between

maize and rice soils, and high- and low-latitude regions of maize and rice fields (n.s.  $p > 0.05$ ; \*\*\* $p < 0.001$ ; Wilcoxon rank-sum test)

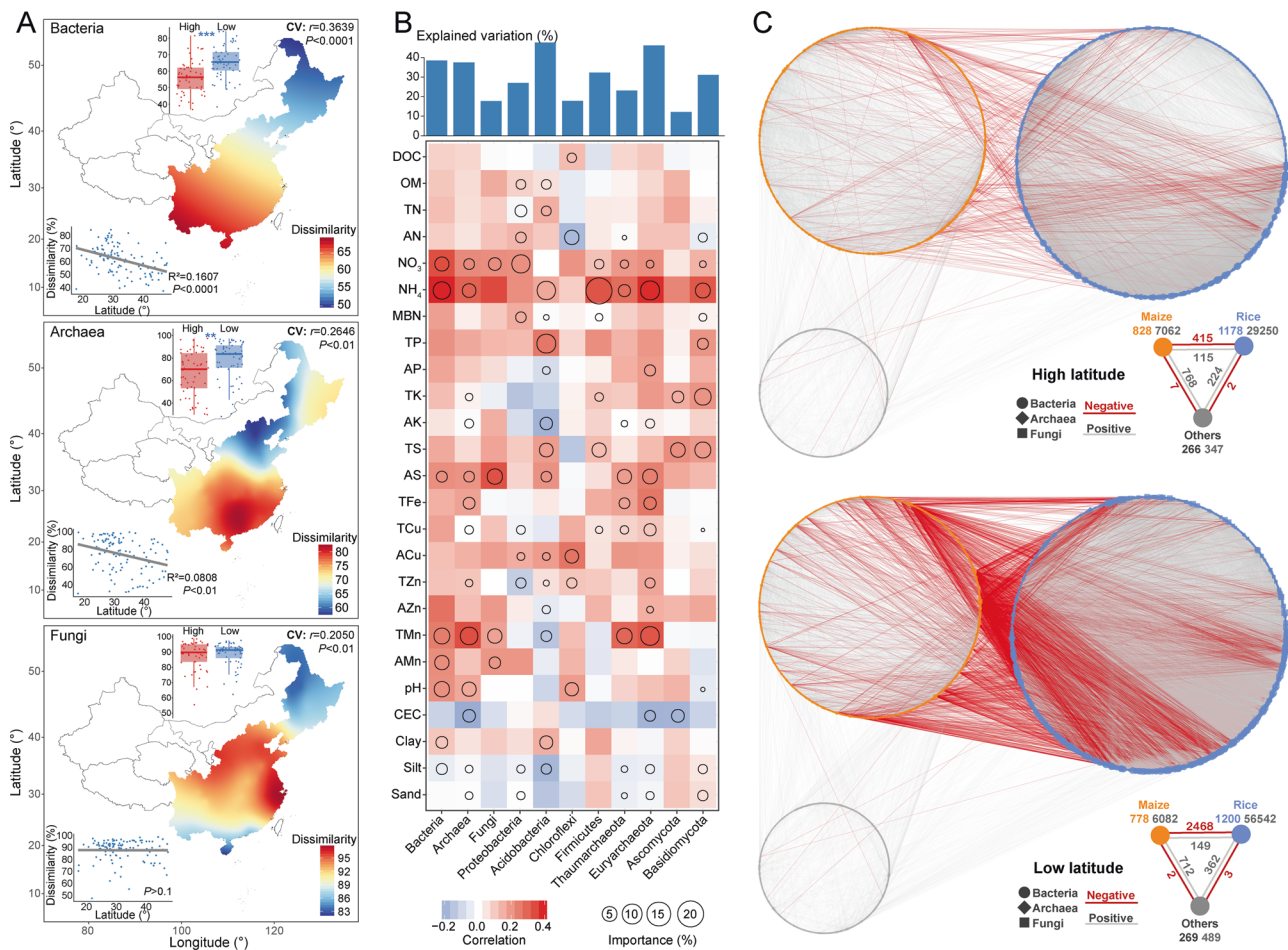
fields, we explored the intrinsic factors driving the community variation. The community dissimilarities between maize and rice soils for each pairwise set of soil samples were estimated, and their spatial distributions were mapped (Fig. 5a). Dissimilarities in bacterial and archaeal communities, but not fungal communities, were higher at low latitudes (Fig. 5a). Although the  $R^2$  of the linear fitting models were not high, the trends could be obviously observed in the predicated maps. In addition, significant and negative linear regression relationships against latitude were observed for some abundant microbial phyla, confirming the higher dissimilarities in low-latitude regions (Supplementary Fig. S10). To identify soil properties that affect dissimilarities of microbial communities and differences in relative abundances of microbial phyla, we correlated these values to the differences in soil properties between maize and rice fields for each pairwise set of soil samples.  $\text{NH}_4$ ,  $\text{NO}_3$ , TMn, and AS were strong, positive predictors for dissimilarities of microbial communities, and differences in relative abundances of most phyla. For example,  $\text{NH}_4$  was significantly related to bacterial community dissimilarity, and phyla *Acidobacteria*, *Firmicutes*, and *Euryarchaeota*; TMn for archaeal and bacterial community dissimilarity, and phyla *Thaumarchaeota*, and *Euryarchaeota*. The

significance of these observations was confirmed by linear least-squares regression (Supplementary Figs. S11 and S12).

Relationships among microbial taxa were estimated by constructing correlation networks for soil samples from high- or low-latitude regions (Fig. 5c). The high-latitude network consisted of 2272 nodes with 37,761 positive edges and 429 negative edges. There were 415 negative edges between maize- and rice-enriched nodes. The low-latitude network contained 2247 nodes linked by 64,334 positive edges and 2473 negative edges, and 2468 negative edges were between maize- and rice-enriched nodes. The low-latitude network constituted a higher proportion of negative edges than the high-latitude network, which correlated with the higher dissimilarities in microbial communities in the low-latitude regions.

## Discussion

The mechanisms underpinning community diversity have been explored extensively in microbial ecology [2]. Here we examined archaeal, bacterial, and fungal community assembly in agro-ecosystems across different habitats and



**Fig. 5** Spatial distribution and drivers of dissimilarities in bacterial, archaeal, and fungal communities between maize and rice soils for each pairwise set of soil samples. **a** The predicted spatial distributions of community dissimilarities between pairwise maize and rice soils are mapped and their associations with latitude for each pairwise set of soil samples are estimated via linear least-squares regression analysis. First-order polynomial fits are shown in gray solid lines. The cross-validation (“CV”) of the maps based on Pearson correlation between the predicted and observed values in each sampling site. The boxplot shows differences in community dissimilarities between high- and low-latitude regions (\*\* $p < 0.01$ ; \*\*\* $p < 0.001$ ; Wilcoxon rank-sum test). **b** Contributions of soil properties to the dissimilarities of microbial communities and differences in relative abundances of microbial phyla based on correlation and best multiple regression model. We examined the correlation of the values with the differences in soil properties between maize and rice fields for each pairwise set of

regions throughout eastern China. We demonstrated that: (i) microbial community assembly was governed more by species sorting than by dispersal limitation in maize fields, and to a lesser extent in rice fields; (ii) the relative importance of species sorting in maize soils was greater at low latitudes than at high latitudes, while rice soils exhibited an opposite trend; and (iii) microbial co-occurrence associations tended to be higher when communities were primarily driven by dispersal limitation relative to species sorting.

soil samples and identified the major predictors. Circle size represents the variable importance (that is, proportion of explained variability calculated via multiple regression modeling and variance decomposition analysis). Colors represent Spearman correlations. The abbreviations of soil properties accorded to the *Method*. **c** Correlation networks for microbial taxa in soils in high- and low-latitude regions. The networks are colored based on maize or rice-enriched taxa. A connection indicates a strong and significant (FDR-corrected  $p < 0.01$ ) correlation, divided into positive (Spearman’s  $\rho > 0.6$ ; dark gray) or negative (Spearman’s  $\rho < -0.6$ ; red) edges. The size of each node is proportional to the relative abundance of the OTUs; the thickness of a connection between two nodes (i.e., an edge) is proportional to the value of Spearman’s correlation coefficient. Maize, maize-enriched OTUs; Rice, rice-enriched OTUs; Others, non-significantly different OTUs

In the present study, we only observed weak DDRs (fitness  $R^2 < 0.1$ ) in microbial communities in maize and rice fields (Supplementary Fig. S2). The fitness values of the relationships were substantially lower than the values reported in natural terrestrial ecosystems [3, 4, 6], suggesting that the spatial structures of soil microbial communities were less apparent in human-managed agricultural ecosystems. We further examined the community assembly processes driving the relationships. Lower sorting/dispersal

effect ratios for microbial communities were observed in rice fields, indicating that the microbial communities in the rice fields were less constrained by environment than the microbial communities in the maize fields (Fig. 2). This is supported by the results of the neutral community model (Supplementary Table S2). In addition, microbial groups in the rice fields demonstrated wider niche breadths, which were quantified using community-level habitat niche breadth analysis (Fig. 4), than the microbial groups in the maize fields. The finding is consistent with the results of other studies that report that habitat generalists with wider niche breadths are influenced less by environmental factors [24, 67]. A potential explanation is that paddy soils could be unique habitats due to constant flooding management activities and long-term regular dry–wet cycles [81], resulting in less environment filtering even at large spatial scales. In systems with less environmental variation or with species pools characterized by environmental generalists, stochastic processes may overwhelm deterministic processes [20]. In particular, species sorting, defined as the ecological forces that alter community structure due to fitness differences among organisms and environmental heterogeneity among regional pools, is a deterministic process in the absolute sense [2, 14]. Conversely, dispersal limitation may not be used as the sole evidence for stochastic processes since dispersal can either be deterministic, stochastic, or both [2, 25].

Microbial diversity often exhibits strong habitat-specific patterns [4]. In the present study, higher dissimilarities in microbial communities were observed between maize and rice soils in low-latitude regions, which was consistent with the higher proportion of negative edges in the correlation networks (Fig. 5). The observation could be because higher temperatures in low-latitude regions caused higher metabolic rates and more rapid succession, resulting in greater dissimilarity tendency between maize and rice soil microbial communities [82]. In addition, soil microbiota in warmer regions (e.g., low latitudes) are more sensitive to environmental changes [5]. As opposed to high-latitude maize soils, we demonstrated that low-latitude soils were driven more strongly by species sorting than by dispersal limitation (Fig. 3). In contrast, rice paddy soils exhibited an opposite trend (i.e., comparatively more species sorting in higher latitudes). The dynamics of soil microbial communities we observed could be associated with the cultivation histories of the different crops grown in water-saturated and unsaturated agricultural fields. Rice agriculture was originally developed in the middle and lower Yangtze River Basin, while the origin of dry-land agriculture is along the Yellow River Basin in northern China [83], which correspond to the sampling sites in the low- and high-latitude regions, respectively, in the present study. A long-term history of cultivation could positively select

microorganisms with fitness advantages under relatively constant environmental conditions, resulting in less environment filtering. This is in line with a previous study [84], which demonstrated that stochastic processes played greater roles in driving microbial community assembly under long-term warming in a tall-grass prairie ecosystem than deterministic processes [84]. Given that maize soils are unsaturated, microorganisms in such habitats could have limited dispersal capacity and hence would potentially experience higher dispersal limitation. Therefore, our results suggest that microbial community assembly is less environmentally filtered in agricultural fields under long-term cultivation. The findings could facilitate the understanding of the mechanism of generation and maintenance of the ubiquitous microbial taxa, which are dominant in terrestrial ecosystems at the global scale [8].

A few potential limitations in the interpretation of our findings should be considered. First, one may argue that a large proportion of variation remains unexplained in the variation partitioning. The unexplained variation could be largely attributed to the unmeasured environmental variables and large-scale sampling. However, results from the null and neutral theory-based process models collectively support the robustness of our conclusions. Variation partitioning analysis focuses on partitioning beta-diversity of metacommunity in a region [85], while null model analysis examines the pairwise metacommunities [23]. The neutral model predicts the relationship between the frequency with which taxa occur in a set of local communities and their abundance across the wider metacommunity [63, 64]. Such frameworks provide complementary information and should be simultaneously applied when considering their strengths and weaknesses [2]. Second, we used the SDERs rather than their absolute values when comparing microbial communities across different habitats and regions. It was difficult to evaluate the relative importance of species sorting versus dispersal limitation using their absolute values since the total variation explained by the two processes varied among the regional metacommunities. For instance, the absolute value of dispersal limitation for bacterial communities was higher in rice fields than in maize fields (Supplementary Figs. S4 and S5). However, bacteria in the rice fields had wider niche breadths (Fig. 4) and higher migration rates (Supplementary Table S1) than the bacteria in the maize fields. In addition, the microorganisms in the rice fields exhibited higher site occupation than the microorganisms in the maize fields (Supplementary Fig. 13). The relative importance of species sorting and dispersal limitation seemingly influences the differences in microbiota assembly across systems, as demonstrated by a recent study [24].

Determining the linkages between community assembly and species coexistence is fundamental for the

understanding of mechanisms underpinning community diversity [2, 37, 86]. The contemporary coexistence theory emphasizes that coexistence depends on both niche differences and fitness differences [39]. Kraft et al. applied a null model approach to examine the co-occurrence patterns of more than 1100 tree species in an Amazonian forest, and they demonstrated that co-occurring trees are often less ecologically similar than predicted by a niche-free (neutral) model. In addition, niche-based processes drive convergence and divergence in key aspects of plant ecological strategy [37]. While the ecological strategies of co-occurring species are well documented in plant communities [37–39, 41], few studies have examined the co-occurrence strategies in microbial communities. Microorganisms occupying a specific ecological niche can form complex interaction networks [42]. Co-occurrence networks have provided novel insight into potential interactions via revealing the niche spaces shared by community members [42, 46, 87]. Since complex soil processes are driven by interactions among soil bacteria, archaea, and fungi, research on soil microbiomes has increasingly focused on all the microbial members as a whole [88–90]. Species sorting became a less dominant assembly process in rice field communities as microbial co-occurrence increased (Fig. 2). Similar trends were observed in the high-latitude and the low-latitude regions in maize and rice fields, respectively (Fig. 3).

Environmental filtering refers to how abiotic factors prevent species without certain physiological traits from occurring in local communities [41]. Different to macroorganisms, microorganisms exhibit a relatively high degree of metabolic flexibility, physiological tolerance, dispersal, rapid growth, and evolutionary adaptation via horizontal gene transfer [91, 92]. Such features may enable the occupation of wider niches, resulting in greater coexistence of species within metacommunities [12, 93, 94]. In addition, microorganisms are less environmentally filtered than macroorganisms [24, 95]. In tropical forests, niche-based assembly of tree communities (e.g., habitat specialization and strategy differentiation) contributes to species coexistence [37].

In the microbial ecosystems investigated in the present study, we showed that a balance between species sorting and dispersal limitation mediates species coexistence in agricultural soil microbiomes across eastern China. This conclusion is based on the linkages between community assembly and species coexistence estimated according to the relative importance of the two processes. However, a previous study demonstrated that more interconnected microbial communities experienced lower turnover and susceptibility to homogenizing selection in unperturbed aquifers [96]. Such discrepancies may be attributed to the following reasons: (i) compared with

aquatic ecosystems, selection could be more intense in soil ecosystems with lower fluidity, resulting in lower degrees of microbial co-occurrence due to competition for limited resources; (ii) in our agricultural fields, human management (e.g., cultivation or irrigation) could have resulted in distinct linkages between community assembly and species coexistence, relative to the unperturbed environments.

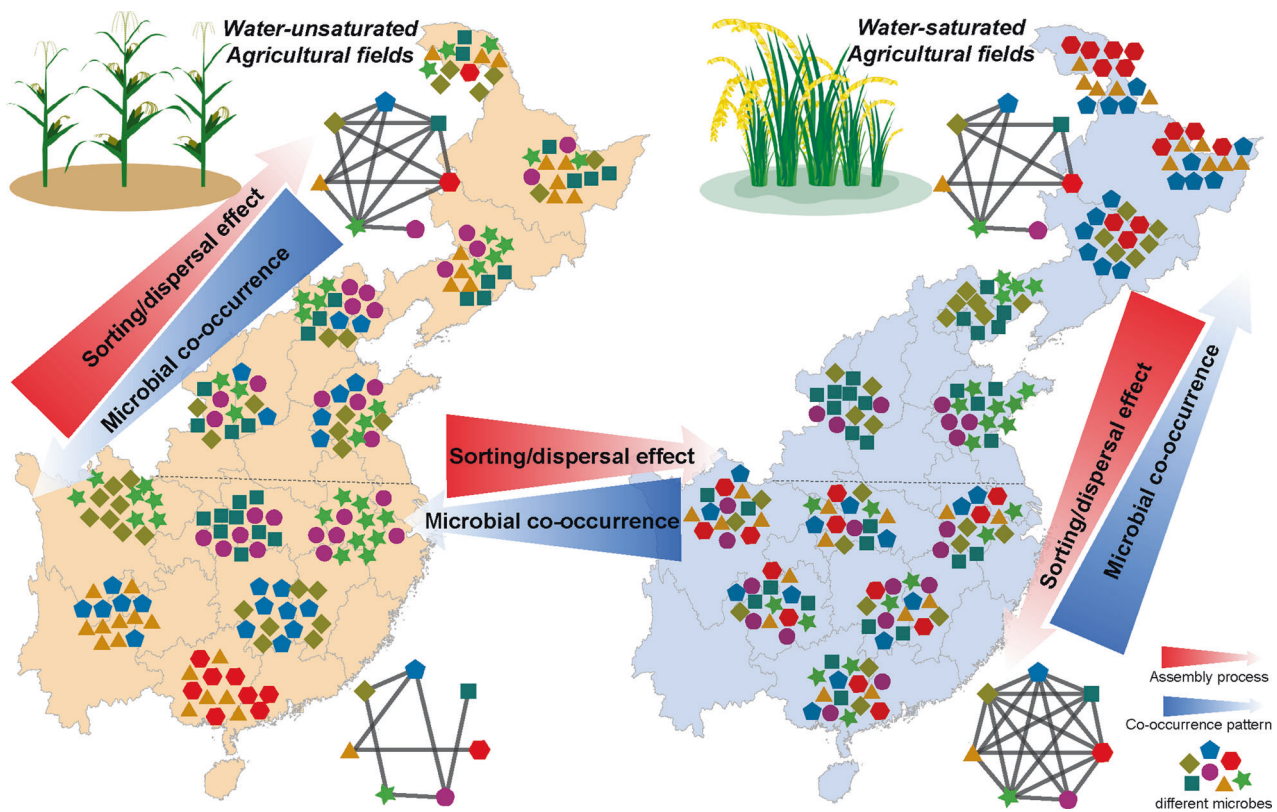
## Data synthesis

We propose a conceptual paradigm (Fig. 6) to describe the different microbial community assembly processes in two distinct agro-ecosystems across a large geographical distance. First, dispersal limitation is more dominant in the assembly of microbial communities in water-saturated (e.g., rice) fields than unsaturated (e.g., maize) fields. Second, distinct processes drive microbial community assembly across different regions in saturated and unsaturated agricultural fields. That is, microbial communities in high-latitude maize soils or low-latitude rice soils tend to be less driven by species sorting relative to dispersal limitation. In addition, microbial co-occurrence is more widespread under weaker environmental filtering. Considering the importance of microorganisms for overall ecosystem functioning [43, 44], it is crucial to understand the mechanisms underpinning the generation and maintenance of microbial diversity in response to current and future environmental changes. By uncovering the intrinsic linkages between assembly processes and species coexistence, the present study could facilitate the management of microbial communities in agro-ecosystems for better adaptation to habitat fragmentation and climate change, in addition to informing restoration strategies based on reassembly and co-occurrence processes.

Although the paradigm offers potentially novel insight, we acknowledge some limitations in the study. For example, the paradigm is based on model agro-ecosystems with an adjacent-pair sampling strategy that controls for the influence of spatial scale and climatic factors. Therefore, considering the importance of soil processes in overall ecosystem functioning, simulation studies should be evaluated both theoretically and experimentally across a broad range of systems.

## Data availability

The raw sequence data reported in this paper are available in the NCBI Sequence Read Archive under BioProject PRJNA544819. R codes on the statistical analyses are available at <https://github.com/shuojiao/Community-assembly-processes-and-species-coexistences-of-agricultural-soils-microbiome.git>.



**Fig. 6** A conceptual paradigm showing soil microbial community assembly processes in low- and high-latitude regions in water-unsaturated (e.g., maize) and water-saturated (e.g., rice) fields, and

their associations with microbial co-occurrence patterns. The geometric symbols represent microbial OTUs

**Acknowledgements** This work was supported by the National Key Research and Development Program of China (2016YFD0200306), the National Science Foundation of China (41807030), the National Postdoctoral Program for Innovative Talents (BX201700005), and the China Postdoctoral Science Foundation (2018M630041).

**Author contributions** SJ and YL conceived and designed the experiments; SJ collected soil samples with the help of YX and JZ; SJ performed the experiments, analyzed the data; SJ, YY, and YL contributed to manuscript development and revisions.

## Compliance with ethical standards

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

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

## References

- Nemergut DR, Schmidt SK, Tadashi F, O'Neill SP, Bilinski TM, Stanish LF, et al. Patterns and processes of microbial community assembly. *Microbiol Mol Biol Rev.* 2013;77:42–356.
- Zhou J, Ning D. Stochastic community assembly: does it matter in microbial ecology? *Microbiol Mol Biol Rev.* 2017;81:e00002–00017.
- Jiao S, Liu ZS, Lin YB, Yang J, Chen WM, Wei GH. Bacterial communities in oil contaminated soils: Biogeography and co-occurrence patterns. *Soil Biol Biochem.* 2016;98:64–73.
- Wang XB, Lu XT, Yao J, Wang ZW, Deng Y, Cheng WX, et al. Habitat-specific patterns and drivers of bacterial beta-diversity in China's drylands. *ISME J.* 2017;11:1345–58.
- Jiao S, Chen W, Wang J, Zhang L, Yang F, Lin Y, et al. Plant growth and oil contamination alter the diversity and composition of bacterial communities in agricultural soils across china. *Land Degrad Dev.* 2018;29:1660–71.
- Griffiths RI, Thomson BC, James P, Bell T, Bailey M, Whiteley AS. The bacterial biogeography of British soils. *Environ Microbiol.* 2011;13:1642–54.
- Fierer N, Jackson RB. The diversity and biogeography of soil bacterial communities. *Proc Natl Acad Sci USA.* 2006;103:626–31.
- Delgado-Baquerizo M, Oliverio AM, Brewer TE, Benavent-Gonzalez A, Eldridge DJ, Bardgett RD, et al. A global atlas of the dominant bacteria found in soil. *Science.* 2018;359:320–5.
- Fierer N, Strickland MS, Liptzin D, Bradford MA, Cleveland CC. Global patterns in below ground communities. *Ecol Lett.* 2009;12:1238–49.
- Nemergut DR, Costello EK, Hamady M, Lozupone C, Jiang L, Schmidt SK, et al. Global patterns in the biogeography of bacterial taxa. *Environ Microbiol.* 2011;13:135–44.
- Dini-Andreote F, Stegen JC, van Elsas JD, Salles JF. Disentangling mechanisms that mediate the balance between stochastic and deterministic processes in microbial succession. *Proc Natl Acad Sci USA.* 2015;112:E1326–32.

12. Stegen JC, Lin X, Konopka AE, Fredrickson JK. Stochastic and deterministic assembly processes in subsurface microbial communities. *ISME J.* 2012;6:1653–64.
13. Vellend M. Conceptual synthesis in community ecology. *Q Rev Biol.* 2010;85:183–206.
14. Chase JM, Myers JA. Disentangling the importance of ecological niches from stochastic processes across scales. *Philos T R Soc B.* 2011;366:2351–63.
15. Hubbell SP. The unified neutral theory of biodiversity and biogeography (MPB-32), vol. 32. Princeton University Press; 2001.
16. Martiny JB, Eisen JA, Penn K, Allison SD, Horner-Devine MC. Drivers of bacterial  $\beta$ -diversity depend on spatial scale. *Proc Natl Acad Sci USA.* 2011;108:7850–4.
17. Whitaker RJ, Grogan DW, Taylor JW. Geographic barriers isolate endemic populations of hyperthermophilic archaea. *Science.* 2003;301:976–8.
18. Ramette A, Tiedje JM. Multiscale responses of microbial life to spatial distance and environmental heterogeneity in a patchy ecosystem. *Proc Natl Acad Sci USA.* 2007;104:2761–6.
19. Nemergut DR, Schmidt SK, Fukami T, O'Neill SP, Bilinski TM, Stanish LF, et al. Patterns and Processes of Microbial Community Assembly. *Microbiol Mol Biol R.* 2013;77:342–56.
20. Jianjun W, Ji S, Yucheng W, Chen T, Janne S, James CS, et al. Phylogenetic beta diversity in bacterial assemblages across ecosystems: deterministic versus stochastic processes. *ISME J.* 2013;7:1310–21.
21. Morlon H, Chuyong G, Condit R, Hubbell S, Kenfack D, Thomas D, et al. A general framework for the distance-decay of similarity in ecological communities. *Ecol Lett.* 2008;11:904–17.
22. Anderson MJ, Crist TO, Chase JM, Vellend M, Inouye BD, Free-stone AL, et al. Navigating the multiple meanings of beta diversity: a roadmap for the practicing ecologist. *Ecol Lett.* 2011;14:19–28.
23. Stegen JC, Lin XJ, Fredrickson JK, Chen XY, Kennedy DW, Murray CJ, et al. Quantifying community assembly processes and identifying features that impose them. *ISME J.* 2013;7:2069–79.
24. Wu WX, Lu HP, Sastri A, Yeh YC, Gong GC, Chou WC, et al. Contrasting the relative importance of species sorting and dispersal limitation in shaping marine bacterial versus protist communities. *ISME J.* 2018;12:485–94.
25. Hanson CA, Fuhrman JA, Horner-Devine MC, Martiny JB. Beyond biogeographic patterns: processes shaping the microbial landscape. *Nat Rev Microbiol.* 2012;10:497–506.
26. Stomeo F, Valverde A, Pointing SB, McKay CP, Warren-Rhodes KA, Tuffin MI, et al. Hypolithic and soil microbial community assembly along an aridity gradient in the Namib Desert. *Extremophiles.* 2013;17:329–37.
27. Graham EB, Crump AR, Resch CT, Fansler S, Arntzen E, Kennedy DW, et al. Coupling spatiotemporal community assembly processes to changes in microbial metabolism. *Front Microbiol.* 2016;7:1949.
28. Tripathi BM, Stegen JC, Kim M, Dong K, Adams JM, Lee YK. Soil pH mediates the balance between stochastic and deterministic assembly of bacteria. *ISME J.* 2018;12:1072–83.
29. Sengupta A, Stegen JC, Meira Neto AA, Wang Y, Neilson JW, Chorover J, et al. Assessing microbial community patterns during incipient soil formation from basalt. *J Geophys Res Biogeosci.* 2019;124:941–58.
30. Li XL, Zhang JL, Gai JP, Cai XB, Christie P, Li XL. Contribution of arbuscular mycorrhizal fungi of sedges to soil aggregation along an altitudinal alpine grassland gradient on the Tibetan Plateau. *Environ Microbiol.* 2015;17:2841–57.
31. Jiao S, Chen WM, Wei GH. Biogeography and ecological diversity patterns of rare and abundant bacteria in oil-contaminated soils. *Mol Ecol.* 2017;26:5305–17.
32. Liu L, Yang J, Yu Z, Wilkinson DM. The biogeography of abundant and rare bacterioplankton in the lakes and reservoirs of China. *ISME J.* 2015;9:2068–77.
33. Mondiale B. World Development report: agriculture for development. Washington, DC: The International Bank for Reconstruction and Development/The World Bank; 2008.
34. Shi Y, Li Y, Xiang X, Sun R, Yang T, He D, 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* 2018;6:27.
35. Zhang BG, Zhang J, Liu Y, Guo YQ, Shi P, Wei GH. Biogeography and ecological processes affecting root-associated bacterial communities in soybean fields across China. *Sci Total Environ.* 2018;627:20–7.
36. Cermeno P, Falkowski PG. Controls on Diatom Biogeography in the Ocean. *Science.* 2009;325:1539–41.
37. Kraft NJB, Valencia R, Ackerly DD. Functional traits and niche-based tree community assembly in an Amazonian forest. *Sci.* 2008;322:580.
38. Adler PB, Alex F, Kleinhesselink AR, Kraft NJB. Trait-based tests of coexistence mechanisms. *Ecol Lett.* 2014;16:1294–306.
39. HilleRisLambers J, Adler PB, Harpole WS, Levine JM, Mayfield MM. Rethinking community assembly through the lens of coexistence theory. *Annu Rev Ecol Evol Syst.* 2012;43:227–48.
40. Silvertown J. Plant coexistence and the niche. *Trends Ecol Evol.* 2004;19:605–11.
41. Kraft NJB, Adler PB, Godoy O, James EC, Fuller S, Levine JM. Community assembly, coexistence and the environmental filtering metaphor. *Funct Ecol.* 2015;29:592–9.
42. Faust K, Raes J. Microbial interactions: from networks to models. *Nat Rev Microbiol.* 2012;10:538–50.
43. Falkowski PG, Fenchel T, Delong EF. The microbial engines that drive Earth's biogeochemical cycles. *Science.* 2008;320:1034–9.
44. Delgado-Baquerizo M, Maestre FT, Reich PB, Jeffries TC, Gaitan JJ, Encinar D, et al. Microbial diversity drives multifunctionality in terrestrial ecosystems. *Nat Commun.* 2016;7:10541.
45. Silva MCP, Dias ACF, van Elsas JD, Salles JF. Spatial and temporal variation of archaeal, bacterial and fungal communities in agricultural soils. *Plos One.* 2012;7:e51554.
46. Eiler A, Heinrich F, Bertilsson S. Coherent dynamics and association networks among lake bacterioplankton taxa. *ISME J.* 2012;6:330–42.
47. Barberán A, Bates ST, Casamayor EO, Fierer N. Using network analysis to explore co-occurrence patterns in soil microbial communities. *ISME J.* 2012;6:343–51.
48. Ma B, Wang H, Dsouza M, Lou J, He Y, Dai Z, et al. Geographic patterns of co-occurrence network topological features for soil microbiota at continental scale in eastern China. *ISME J.* 2016;10:1891–901.
49. Zorner RJ, Trabucco A, Bossio DA, Verchot LV. Climate change mitigation: a spatial analysis of global land suitability for clean development mechanism afforestation and reforestation. *Agriculture Ecosyst Environ.* 2008;126:67–80.
50. Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A. Very high resolution interpolated climate surfaces for global land areas. *Int J Climatol.* 2005;25:1965–78.
51. Jiao S, Chen W, Wang J, Du N, Li Q, Wei G. Soil microbiomes with distinct assemblies through vertical soil profiles drive the cycling of multiple nutrients in reforested ecosystems. *Microbiome.* 2018;6:146.
52. Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Lozupone CA, Turnbaugh PJ, et al. Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proc Natl Acad Sci USA.* 2011;108:4516–22.
53. Dissanayake AJ, Purahong W, Wubet T, Hyde KD, Zhang W, Xu H, et al. Direct comparison of culture-dependent and culture-independent molecular approaches reveal the diversity of fungal endophytic communities in stems of grapevine (*Vitis vinifera*). *Fungal Divers.* 2018;90:85–107.

54. Edgar RC, Haas BJ, Clemente JC, Quince C, Knight R. UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics*. 2011;27:2194–200.
55. Qi W, Liu S, Zhao M, Liu Z. China's different spatial patterns of population growth based on the "Hu Line". *J Geographical Sci*. 2016;26:1611–25.
56. Oksanen J, Blanchet FG, Kindt R, Legendre P, Minchin PR, O'Hara R, et al. *Vegan: community ecology package*. R package version 22-1, <http://CRAN.R-project.org/package=vegan>. 2015.
57. Field A, Miles J, Field Z. *Discovering statistics using R*. London: Sage publications; 2012.
58. Wickham H. *ggplot2: elegant graphics for data analysis*. New York: Springer; 2016.
59. Yeh YC, Peres-Neto PR, Huang SW, Lai YC, Tu CY, Shiah FK, et al. Determinism of bacterial metacommunity dynamics in the southern East China Sea varies depending on hydrography. *Ecography*. 2015;38:198–212.
60. Griffith DA, Peres-Neto PR. Spatial modeling in ecology: the flexibility of eigenfunction spatial analyses. *Ecology*. 2006;87:2603–13.
61. Dray S, Legendre P, Peres-Neto PR. Spatial modelling: a comprehensive framework for principal coordinate analysis of neighbour matrices (PCNM). *Ecol Model*. 2006;196:483–93.
62. Stegen JC, Lin XJ, Fredrickson JK, Konopka AE. Estimating and mapping ecological processes influencing microbial community assembly. *Front Microbiol*. 2015;6:370.
63. Sloan WT, Lunn M, Woodcock S, Head IM, Nee S, Curtis TP. Quantifying the roles of immigration and chance in shaping prokaryote community structure. *Environ Microbiol*. 2006;8:732–40.
64. Burns AR, Stephens WZ, Stagaman K, Wong S, Rawls JF, Guillemin K, et al. Contribution of neutral processes to the assembly of gut microbial communities in the zebrafish over host development. *ISME J*. 2016;10:655–64.
65. Ostman O, Drakare S, Kritzberg ES, Langenheder S, Logue JB, Lindstrom ES. Regional invariance among microbial communities. *Ecol Lett*. 2010;13:118–27.
66. Levins R. *Evolution in changing environments*. Princeton, New Jersey: Princeton University Press; 1968.
67. Pandit SN, Kolasa J, Cottenie K. Contrasts between habitat generalists and specialists: an empirical extension to the basic metacommunity framework. *Ecology*. 2009;90:2253.
68. Zhang J, Zhang MJ. *Spaa: Species association analysis*. R package version 0.2.1, <http://CRAN.R-project.org/package=spaa>. 2013.
69. Newman ME. The structure and function of complex networks. *Siam Rev*. 2003;45:167–256.
70. Newman ME. Modularity and community structure in networks. *Proc Natl Acad Sci USA*. 2006;103:8577–82.
71. Greenblum S, Turnbaugh PJ, Borenstein E. Metagenomic systems biology of the human gut microbiome reveals topological shifts associated with obesity and inflammatory bowel disease. *Proc Natl Acad Sci USA*. 2012;109:594–9.
72. Berry D, Widder S. Deciphering microbial interactions and detecting keystone species with co-occurrence networks. *Front Microbiol*. 2014;5:219.
73. Csardi G, Nepusz T. The igraph software package for complex network research. *Int J Complex Syst*. 2006;1695:1–9.
74. Bastian M, Heymann S, Jacomy M. *Gephi: an open source software for exploring and manipulating networks*. ICWSM. 2009;8:361–2.
75. Hiemstra P, Hiemstra MP. Package 'automap'. *Compare*. 2013; 105:10.
76. Grömping U. Relative importance for linear regression in R: the package relaimpo. *J Stat Softw*. 2006;17:1–27.
77. Klaus B, Strimmer K. *fdrtool: Estimation of (local) false discovery rates and higher criticism*. R package version 1.2.15, <https://CRAN.R-project.org/package=fdrtool>. 2015.
78. Harrell Jr FE. *Hmisc: harrell miscellaneous*. R Package Version 3.0-12, <https://CRAN.R-project.org/package=Hmisc>. 2013.
79. Kembel SW, Cowan PD, Helmus MR, Cornwell WK, Morlon H, Ackerly DD, et al. *Picante: R tools for integrating phylogenies and ecology*. *Bioinformatics*. 2010;26:1463–4.
80. Warnes GR, Bolker B, Bonebakker L, Gentleman R, Huber W, Liaw A et al. *gplots: Various R programming tools for plotting data*. R package version 2.4, <https://CRAN.R-project.org/package=gplots>. 2009.
81. Hu HW, Zhang LM, Yuan CL, He JZ. Contrasting Euryarchaeota communities between upland and paddy soils exhibited similar pH-impacted biogeographic patterns. *Soil Biol Biochem*. 2013;64:18–27.
82. Liang Y, Jiang Y, Wang F, Wen C, Deng Y, Xue K, et al. Long-term soil transplant simulating climate change with latitude significantly alters microbial temporal turnover. *ISME J*. 2015;9:2561–72.
83. Zhao Z. New data and new issues for the study of origin of rice agriculture in China. *Archaeol Anthr Sci*. 2010;2:99–105.
84. Guo X, Feng JJ, Shi Z, Zhou XS, Yuan MT, Tao XY, et al. Climate warming leads to divergent succession of grassland microbial communities. *Nat Clim Change*. 2018;8:813–8.
85. McArdle BH, Anderson MJ. Fitting multivariate models to community data: a comment on distance-based redundancy analysis. *Ecology*. 2001;82:290–7.
86. Vályi K, Mardhiah U, Rillig MC, Hempel S. Community assembly and coexistence in communities of arbuscular mycorrhizal fungi. *ISME J*. 2016;10:2341.
87. Steele JA, Countway PD, Xia L, Vigil PD, Beman JM, Kim DY, et al. Marine bacterial, archaeal and protistan association networks reveal ecological linkages. *ISME J*. 2011;5:1414–25.
88. Ma B, Dai ZM, Wang HZ, Dsouza M, Liu XM, He Y, et al. Distinct biogeographic patterns for archaea, bacteria, and fungi along the vegetation gradient at the continental scale in Eastern China. *Msystems* 2017;2:pii: e00174–16.
89. Morrien E, Hannula SE, Snoek LB, Helmsing NR, Zweers H, de Hollander M, et al. Soil networks become more connected and take up more carbon as nature restoration progresses. *Nat Commun*. 2017;8:14349.
90. Jiao S, Chen WM, Wang JL, Du NN, Li QP, Wei GH. Soil microbiomes with distinct assemblies through vertical soil profiles drive the cycling of multiple nutrients in reforested ecosystems. *Microbiome*. 2018;6:146.
91. Fuhrman JA, Cram JA, Needham DM. Marine microbial community dynamics and their ecological interpretation. *Nat Rev Microbiol*. 2015;13:133–46.
92. Allison SD, Martiny JB. Resistance, resilience, and redundancy in microbial communities. *Proc Natl Acad Sci USA*. 2008;105:11512–9.
93. Leibold MA, McPeck MA. Coexistence of the niche and neutral perspectives in community ecology. *Ecology*. 2006;87:1399–410.
94. Lange M, Eisenhauer N, Sierra CA, Bessler H, Engels C, Griffiths RI, et al. Plant diversity increases soil microbial activity and soil carbon storage. *Nat Commun*. 2015;6:6707.
95. Farjalla VF, Srivastava DS, Marino NAC, Azevedo FD, Dib V, Lopes PM, et al. Ecological determinism increases with organism size. *Ecology*. 2012;93:1752–9.
96. Danczak RE, Johnston MD, Kenah C, Slattery M, Wilkins MJ. Microbial community cohesion mediates community turnover in unperturbed aquifers. *Msystems*. 2018;3:pii: e00066–18.