Title

The spatial scale dependence of diazotrophic and bacterial community assembly in paddy soil

Journal

Global Ecology and Biogeography, 28(8)

ISSN

0960-7447

Authors

Gao, Qun

Yang, Yunfeng

Feng, Jiajie

et al.

Publication Date

2019-08-01

DOI

10.1111/geb.12917

Peer reviewed

The spatial scale dependence of diazotrophic and bacterial community assembly in paddy soil

Qun Gao[1] | Yunfeng Yang[1] | Jiajie Feng[2] | Renmao Tian[2] | Xue Guo[1] | Daliang Ning[2] | Lauren Hale[2,3] | Mengmeng Wang[1] | Jingmin Cheng[1] | Linwei Wu[2] | Mengxin Zhao[1] | Jianshu Zhao[1] | Liyou Wu[2] | Yujia Qin[2] | Qi Qi[1] | Yuting Liang[4] | Bo Sun[4] | Haiyan Chu[4] | Jizhong Zhou[1,2,5]

[1] State Key Joint Laboratory of Environment Simulation and Pollution Control, School of Environment, Tsinghua University, Beijing, China [2] Department of Microbiology and Plant Biology, Institute for Environmental Genomics, University of Oklahoma, Norman, Oklahoma [3] Agricultural Research Service, San Joaquin Valley Agricultural Sciences Center, USDA, Parlier, California [4] State Key Laboratory of Soil and Sustainable Agriculture, Institute of Soil Science, Chinese Academy of Sciences, Nanjing, China [5] Earth and Environmental Sciences, Lawrence Berkeley National Laboratory, Berkeley, California

Correspondence Jizhong Zhou, State Key Joint Laboratory of Environment Simulation and Pollution Control, School of Environment, Tsinghua University, Beijing 100084, China. Email:jzhou@ou.edu

Abstract

Aim

The factors driving microbial community β-diversity (variation in composition) at different spatial scales yield fundamental insights into the mechanisms that maintain ecosystem biodiversity, which as yet are uncertain. Here, we explore whether spatial scale-dependent patterns of β-diversity vary between microbial functional groups and bacterial taxa (i.e., diazotrophic and bacterial communities) across local to regional scales (from metres to hundreds of kilometres).

Location

Eastern China.

Time period

October and November 2015.

Major taxa studied

Diazotrophic and bacterial communities.

Methods

We use two complementary statistical tools to unveil biotic mechanisms (i.e., species association) underlying variation in β-diversity of diazotrophic and bacterial communities. We examined distance–decay slopes of both communities at the local (1–113 m), meso- (3.4–39 km) and regional (103–668 km) scales. We used an environmentally constrained checkerboard score

and topological features of association networks as indices of species association. We then calculated contributions of species association, abiotic factors and geographical distance to explain community β-diversity. The scale-dependent distance–decay relationships were also examined in ubiquitous (high occupancy across samples) and endemic communities of diazotrophs and bacteria.

Results

Diazotrophs displayed steeper distance–decay slopes than bacteria, suggesting that the β-diversity of diazotrophic communities was more variable. The distance–decay slopes were dependent on spatial scales in both communities, owing to different contributions of geographical distance, abiotic factors and species association at three spatial scales. Intriguingly, species association was greater and contributed more to community β-diversity than other forces at the local scale, implying that species association could greatly alter community structures.

Main conclusions

Drivers of diazotrophic and bacterial community β-diversity depended on spatial scales, resulting in different distance–decay patterns. Moreover, this was the first study to use two methods to demonstrate that species association played important, but as yet unrecognized, roles in driving spatial scale-dependent β-diversity.

KEYWORDS: bacterial community, β-diversity, community assembly, diazotrophic community, paddy soil, spatial scale dependence, species association

1 INTRODUCTION

The long-standing paradigm in microbial biogeography, ‘everything is everywhere, but the environment selects’, alludes to the remarkable potential for dispersal of microorganisms (Fuhrman, 2009). Most microorganisms are not cosmopolitans, thus discernible biogeographical patterns and the distance–decay rule (i.e., microbial community similarity decreases as geographical distance increases) are commonly discovered in all branches of microbial life (Green & Bohannan, 2006; Zhang et al., 2016). It is likely that there is a steady distance–decay slope across different spatial scales when measuring microbial β-diversity (i.e., variation in community composition), because community similarity is a probability rule for spatial distribution of taxa abundance (Harte, Kinzig, & Green, 1999). However, it has been argued that microbial β-diversity varies over distance owing to the fact that microorganisms are fundamentally different in their longevity, niche preferences and dispersal abilities (Green & Bohannan, 2006; Meyer et al., 2018). This argument was supported by recent studies showing that the distance–decay slopes of ammonia-oxidizing bacteria and sulfate-reducing bacteria in salt marsh sediments varied by spatial scales (Angermeyer, Crosby, & Huber, 2015; Martiny, Eisen, Penn, Allison, & Horner-Devine,

2011). In contrast, distance–decay curves should be flat where dispersal is high. This prediction has been verified in the marine environment, because ocean currents facilitate microbial dispersal (Hewson, Steele, Capone, & Fuhrman, 2006). As a consequence, it still remains elusive whether there exists spatial scale dependence of microbial β-diversity. Moreover, the underlying drivers that shape spatial scaling of microbial community assembly are largely unexplored.

Microorganisms possess a large number of functional traits and thus play an essential role in mediating the Earth's biogeochemical processes. Understanding the spatial pattern of microbial functional β-diversity is crucial for predicting and elucidating ecosystem functions. Meanwhile, incomplete sampling of microbial communities is particularly pronounced for endemic taxa, which are defined as microorganisms that have low occupancies across different samples (Meyer et al., 2018). Given that endemic taxa are restricted in ranges (Woodcock, Curtis, Head, Lunn, & Sloan, 2006), they could be vital in determining the scale-dependent biogeographical patterns. In contrast, ubiquitous taxa might exhibit similar community composition across different spatial scales (Saunders, Albertsen, Vollertsen, & Nielsen, 2016). However, the spatial scale dependence of microbial β-diversity has seldom been assessed for endemic and ubiquitous taxa.

Investigation of environmental heterogeneity has allowed microbial ecologists to describe how abiotic factors affect microbial community assembly across a wide range of natural habitats (Barberán & Casamayor, 2011). In contrast, there have been far fewer studies to explore biotic interactions between microbial taxa, which are believed to have a substantial influence on the functions or niche occupancy of microbial communities (Chaffron, Rehrauer, Pernthaler, & Mering, 2010; Freilich et al., 2010). It was found that interspecies connection affected habitat affinities or shared physiologies of microbial community members (Barberán, Bates, Casamayor, & Fierer, 2011). Niche specialization of microorganinisms was also highly relevant to their ecological relationships (Faust & Raes, 2012). Those studies have provided important implications that biotic interaction is important in structuring microbial communities.

Species association is regularly used in ecology and biogeography as a proxy for biotic interaction in the community (Cazelles, Araújo, Mouquet, & Gravel, 2016; Li, Poisot, Waller, & Baiser, 2018). One way to infer species association is based on species co-occurrence. Since Diamond (1975) pioneered the analysis of species co-occurrence in geographical space, null models have been used to infer the role of associations/interactions between pairs of species in their distributions (Diamond, 1975; Peterson, 2011; Poisot, Canard, Mouillot, Mouquet, & Gravel, 2012). Another way to approach similar questions is through network-based analysis, which offers insights into topological properties of community members (Ma et al., 2016; Newman, 2003) and is regarded as a valuable tool to identify species associations within a community (Berry & Widder, 2014; Ma et al., 2016; Proulx,

Promislow, & Phillips, 2005; Shi et al., 2016). To date, species co-occurrence patterns and topological features in microbial networks are largely unknown at different spatial scales, resulting in highly unexplained variation in microbial β-diversity in previous studies (Fierer, 2017; Shi et al., 2018; Zhou, Kang, Schadt, & Garten, 2008).

In the present study, we collected 188 paddy soil samples from 18 long-term rice fields located in Hubei, Anhui, Jiangsu and Zhejiang provinces of Eastern China. We compared the slopes of distance–decay curves of bacterial communities and diazotrophic communities responsible for biological nitrogen (N) fixation. Considering that N limitation is widespread in the natural environment (Elser et al., 2007), diazotrophic communities are crucial for soil N bioavailability, especially in paddy soil owing to N loss when the soil is flooded. However, the biogeographical patterns of diazotrophic communities have not yet been examined in paddy soil ecosystems. To tackle this problem, we engaged a nested design of soil sampling to achieve a balanced distribution of pairwise distance across different spatial scales [i.e., local scale (1–113 m), meso-scale (3.4–39 km) and regional scale (103–668 km); Supporting Information Appendix S1, Figure S1]. We then examined diazotrophic communities by sequencing N-fixing *nifH* gene amplicons and examined bacterial communities by sequencing 16S rRNA gene amplicons.

Given that different driving forces may vary by spatial scales, our first hypothesis is that distance–decay slopes of both diazotrophic and bacterial communities are spatial scale dependent. Our second hypothesis is that distance–decay slopes of the functional gene (*nifH*) are steeper than those of 16S rRNA genes, because 16S rRNA genes are more phylogenetically conserved than *nifH* (Goberna et al., 2014). Given that ubiquitous taxa are the ‘steadfast’ community members across a large number of samples and are less impacted by dispersal limitation (Martiny et al., 2006), we test a third hypothesis that ubiquitous microbes have shallower distance–decay slopes than endemic microbes.

2 METHODS

2.1 Site description

We selected 18 paddy fields in four provinces (Hubei, Anhui, Jiangsu and Zhejiang) of Eastern China, which are among the major rice production regions of China. All paddy fields are in the subtropical monsoon climate type. All fields were > 0.02 km^2, and we collected soil samples in the centre of each paddy field. Soil samples were collected from late October to early November in 2015, immediately after rice harvesting. We collected 11 soil samples (biological replicates) in each field. Specifically, six soil samples were collected along a 75-m-long transect, and another five samples were collected along a vertical 75-m-long transect (Supporting Information Appendix S1, Figure S1). Distances between two adjacent samples along both transects were 1, 5, 10, 20 and 40 m (Supporting Information Appendix S1, Figure S1). For each sample, three soil cores were taken and fully mixed

to generate a composite sample. Visible roots and rocks were discarded before samples were packed into polyethylene bags and stored in a portable 4 °C refrigerator. After immediate transportion to the laboratory, each composite sample was divided into two parts: one was stored at 4 °C for analyses of soil physicochemical factors, and the other one was stored at −80 °C for DNA extraction.

2.2 Measurements of soil physicochemical factors

In situ soil temperature (in degrees Celsius) was measured three times by thermometer to obtain an average value. In situ soil water content (as a percentage) was measured three times by hygrometer to obtain an average value. Soil pH was determined by a pH meter (E20-FiveEasy pH; Mettler Toledo, Greifensee, Switzerland) in soil water suspension (1:5, fresh soil/de-ionized water) after shaking for 30 min. Ten grams of sieved soil was weighed to a 250 mL plastic bottle with 100 mL of 2 M KCl solution. The solution in the bottle was then shaken at 250 rpm and kept at room temperature for 1 h, after which part of the soil solution was filtered into a 15 mL centrifuge tube for ammonium, nitrate and dissolved total N measurements. The other part of the soil solution was filtered again with 0.45 μm filter membranes for measurement of dissolved organic carbon by a Skalar autoanalyser using an ultraviolet digestion technique and colorimetric detection. Ammonium and nitrate concentrations were also measured colorimetrically using a spectrophotometer. Dissolved total N and dissolved organic carbon in the soil samples were measured using a Multi N/C 2100 analyser (Analytik Jena, Thuringia, Germany). The concentrations of ammonium, nitrate, dissolved total N and dissolved organic carbon were all expressed in the same units (milligrams per kilogram). The validity of the analytical data generated by the laboratory was monitored by participation in a regular interlaboratory proficiency scheme.

2.3 Soil DNA extraction

Soil DNA was extracted using a PowerMax Soil DNA Isolation Kit (MO BIO Laboratories, Inc., Carlsbad, CA, USA) after freeze-grinding mechanical lysis as previously described (Zhou, Bruns, & Tiedje, 1996). The DNA concentration was quantified by PicoGreen with a FLUOstar OPTIMA fluorescence plate reader (BMG LabTech, Ortenberg, Germany). The DNA quality was assessed based on spectrometry absorbance at wavelengths of 230, 26 and 280 nm by a NanoDrop ND-1000 Spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA).

2.4 Sequencing of *nifH* and 16S rRNA genes, amplifications and raw data processing

Extracted DNA for PCR amplification was diluted to 5 ng/μL. Primers of PolF/PolR (5′-TGCGAYCCSAARGCBGACTC-3′/5′-ATSGCCATCATYTCRCCGGA-3′) were used for amplifying *nifH* genes (Poly, Monrozier, & Bally, 2001), and primers of 515F/806R (5′-GTGCCAGCMGCCGCGGTAA-3′/5′-

GGACTACHVGGGTWTCTAAT-3′) were used for amplifying 16S rRNA genes (Caporaso et al., 2012). Two-step PCR experiments were used to prepare amplicon libraries of both *nifH* genes and 16S rRNA genes as described previously (Wu et al., 2016). Specifically, 10 cycles were used in the first step and 20 cycles in the second step for 16S rRNA genes; 12 cycles were used in the first step and 23 cycles in the second step for *nifH* genes. To increase the base diversity in the library, phasing primers were used in the second step. The PCR products were separated on a 1.5% agarose gel at 90 V for 45 min. The bands were then purified with a QIAquick Gel Extraction Kit (QIAGEN Inc., Valencia, CA, USA). Libraries were sequenced on a desktop MiSeq system (Illumina, San Diego, CA, USA; 2 × 250 bp paired ends), following the manufacturer's protocols, at the Institute for Environmental Genomics (University of Oklahoma, Norman, OK, USA).

Sequencing reads of poor quality were removed by Btrim (Kong, 2011). Chimeras were removed by Uchime (Edgar, Haas, Clemente, Quince, & Knight, 2011). For *nifH*, frame shifts were screened and corrected by Framebot software (Wang et al., 2013). Remaining *nifH* sequences were then clustered into operational taxonomic units (OTUs) with complete linkage clustering (Loewenstein, Portugaly, Fromer, & Linial, 2008) on a Galaxy platform (Afgan et al., 2016) pipeline at 95% amino acid identity (Penton et al., 2016). Taxonomic assignment was conducted by tBLASTx against the Zehr labortory's *nifH* gene database (June 2017 version; https://wwwzehr.pmc.ucsc.edu/nifH_Database_Public/), with parameters maximum target of 10 and E-value cut-off of 1×10^{-10}. The hits with amino acid identity < 95% and alignment coverage < 80% were filtered out. Only hits with known taxonomic assignment deeper than Class level were retained, and lineages of hits with the highest BLAST score were used as the taxonomic assignments of queries.

The 16S rRNA gene sequences were clustered into OTUs with UPARSE (Edgar, 2013) on the Galaxy platform at 97% nucleotide identity. Taxonomic assignment was conducted through the RDP classifier with a confidence cut-off of 0.5 (Wang, Garrity, Tiedje, & Cole, 2007). All sequences were randomly resampled to the depth of 23,000 sequences per sample for the 16S rRNA gene and 10,000 sequences per sample for the *nifH* gene. Phylogenetic trees were constructed and analysed using PyNAST alignment (v.1.0.0), FastTree (v.1.0.0) and MEGA (v.5.10, BETA2).

2.5 Species association analysis

Our first method to examine species association was based on the null model to study species co-occurrence patterns (Li et al., 2018). Within each spatial scale, we constructed a species matrix with rows for each sample and columns for each species. Values in the matrix reflect the presence or absence (1/0) of the species. We then used this species matrix to predict species distributions under environmental constraints with a species distribution model (SDM) in the R package '*biomod2*' (Thuiller et al., 2016).

After this, we calculated an environmentally constrained checkerboard score (C-score) with the function '*ecospat.cons_Cscore*' in the R package '*ecospat*' (Di Cola et al., 2017). The C-score was the mean number of checkerboard units (CUs) between all possible pairs of species (or OTUs) in a matrix. The number of CUs for any pair of species was calculated using the equation:

$$CU_{ij} = (R_i - D)(R_j - D)$$

where CU_{ij} is the C-score for species pair i and j, R_i is the total sites (the number of species occurrences) for species i, R_j is the total sites for species j, and D is the number of shared sites in which both species are present (Di Cola et al., 2017). Environmentally constrained C-scores were expected to maximize the chance of distinguishing species interactions that might shape species distribution and community assembly, because environment was factored out as a possible explanation for the species distribution patterns encountered (Di Cola et al., 2017).

The function '*ecospat.cons_Cscore*' returned the C-score index for the observed community (ObsCscoreTot), the mean of the C-score for the simulated communities by the null model (SimCscoreTot, n = 10,000), the p-values to evaluate the significance of the difference between the former two indices, and returned the standardized effect size (SES) for the whole community (SES.Tot). If $p < 0.050$, we regarded the co-occurrence of microbial community as a non-random pattern resulting from biotic interactions, because the influence of environmental variables was partitioned. Moreover, SESs that were greater than two or less than minus two were statistically significant with a probability of < 0.050. The function also returned the observed and simulated C-scores and the SES for each species pair. Species pairs with SES < −2 reflected aggregation, because they co-occurred more than expected by chance. Species pairs with SES > 2 reflected segregation of species, because they co-occurred less often than expected by chance.

Association networks of diazotrophs and bacteria were constructed as previously described (Deng et al., 2012; Zhou et al., 2010). Only OTUs detected in ≥ 141 of the 188 biological replicates were kept for bacterial community network construction, and 138 of 185 biological replicates were kept for diazotrophic community network construction, following a random matrix theory (RMT) algorithm (Deng et al., 2012). The association network examined the pairwise correlation coefficients of species based on OTU abundance data. The network construction and network topology characterization were processed by the network analysis pipeline at http://ieg2.ou.edu/MENA. We also constructed sub-networks for each sample from the global network using the R package '*igraph*' (Ma et al., 2016). The network topological properties, including average degree, average clustering coefficient and modularity, were calculated for each sample by the R functions '*knn*', '*transitivity*' and '*modularity*', respectively, in the '*igraph*' package.

The average degree referred to species connectivity in the community (Zhou et al., 2010). The average clustering coefficient was used to measure how well nodes were connected with their neighbours (Deng et al., 2016). Modularity was used to demonstrate a network that could be divided naturally into communities or modules (Deng et al., 2012). Higher modularity indicated a higher number of within-cluster associations than between-cluster associations compared with random expectation (Clauset, Newman, & Moore, 2004). These topological properties were regarded as biotic factors in examining their contribution to the variation in microbial β-diversity at different spatial scales using multiple regression on matrices (MRM). We were not able to include environmentally constrained C-scores in MRM because C-scores could not be calculated with one sample. A general framework of species co-occurrence analysis and RMT-based network analysis is illustrated in the Supporting Information (Appendix S1, Figure S2).

2.6 Statistical analyses

Variation of environmental factors across sampling sites was determined by one-way ANOVA followed by the least significant difference (LSD) test. We calculated the rates of distance decay (i.e., slopes of distance–decay curves) of microbial communities at three spatial scales: the local scale (0–113 m), the meso-scale (3.4–39 km) and the regional scale (103–668 km). The slope at each spatial scale was calculated based on the following equation:

$$\ln(S) = \ln(a) + z\ln(G)$$

where S is the microbial community similarity, G is the geographical distance, a is an intercept parameter and z is the slope coefficient of the distance–decay curve (Martiny et al., 2011). The microbial phylogenetic β-diversity (phylogenetic distance) was calculated based on the matrix of abundance-weighted UniFrac distance using the R package '*GUifrac*', because abundance-weighted microbial phylogenetic distance involves both taxonomic and phylogenetic information in communities (Burns et al., 2015). Microbial community similarity was calculated as one minus the phylogenetic distance.

To explore whether the sequencing depth accounted for differences in the distance–decay slopes of diazotrophic and bacterial communities, we resampled the sequence of both communities to obtain communities with different sequence depth. Specifically, bacterial communities with 11,500 and 18,400 sequences were generated by resampling 50 and 80% of overall 16S rRNA sequences, respectively. Likewise, diazotrophic communities with 5,000 and 8,000 sequences were generated by resampling 50 and 80% of overall *nifH* sequences, respectively. We performed the resampling steps in an in-house Galaxy pipeline (http://zhoulab5.rccc.ou.edu:8080/) using both OTU table and corresponding representative sequences. We then calculated the distance–decay slopes for each community at each of the three spatial scales.

We tested whether the distance–decay slopes were significantly different from zero or whether the distance–decay slopes at three spatial scales differed substantially from each other. To this end, we used matrix permutations to compare the observed slopes within the three spatial scales with the distribution of slopes observed in those ranges over 999 permutations, following the method in a previous study (Martiny et al., 2011).

We used the MRM analyses to explore the significance of geographical distance, abiotic factors (dissolved organic carbon, dissolved total N, soil water content, soil temperature and soil pH) and biotic factors (average degree, average clustering coefficient and modularity) in relationship to microbial β-diversity. We applied a ln-transformation on geographical distance because our samples ranged over many orders of magnitude, which made the data points skewed. Abiotic and biotic factors were ln-transformed to achieve normalization except for soil pH, which is a logarithm format of hydrogen ion concentration (Bates, 1964). We then calculated the dissimilarity matrix (by the Euclidean method) for each of the geographical distance, abiotic and biotic factors. The microbial community dissimilarity matrix was calculated based on the abundance-weighted Unifrac distance. We performed modified MRM code based on the R package '*ecodist*' (Goslee, 2007) to disentangle the potential relationships between microbial β-diversity and factors at each spatial scale. The R^2 value of the MRM model represents the total explanatory power of all factors involved in the model, and the partial regression coefficient, *b*, represents the relative contribution of each factor. To remove covariant factors (Harrell, 2001), we performed the MRM model and removed the non-significant factors. Then we performed the MRM model again, following the steps shown previously (Martiny et al., 2011).

To examine the relative contribution of each factor at the three spatial scales, we performed MRM models according to spatial scales. For example, at the local scale, we calculated R^2 and the partial regression coefficient, *b*, for only those pairwise comparisons between 1 and 113 m.

3 RESULTS

3.1 Soil bacterial and diazotrophic communities

We identified > 15,000 bacterial OTUs in 188 paddy soil samples and > 6,000 diazotrophic OTUs in 183 samples, discarding five samples with extremely low diazotrophic OTU richness. α-*Proteobacteria* were the most abundant in diazotrophic communities, followed by δ-*Proteobacteria* and β-*Proteobacteria* (Supporting Information Appendix S1, Figure S3a). Only 42% of diazotrophic sequences were classified, suggesting that our knowledge about the taxonomic information on diazotrophic communities in paddy soil was limited. In contrast, < 20% of overall bacterial sequences were taxonomically unclassified (Supporting Information Appendix S1, Figure S3b). δ-*Proteobacteria* were the most abundant in bacterial communities, followed

by *Acidobacteria*, α-*Proteobacteria*, β-*Proteobacteria*, *Chloroflexi* and γ-*Proteobacteria*.

Almost all bacterial OTUs (96.1%) and diazotrophic OTUs (91.6%) were detected in more than one sample. Abundant OTUs tended to have higher occupancy in both diazotrophic and bacterial communities (Supporting Information Appendix 1, Figure S4). Therefore, communities were divided into sub-groups according to the occupancy of OTUs [i.e., OTUs detected in 0–25% (endemic taxa) and in 75–100% (ubiquitous taxa) of all samples; Nekola & White, 2004]. α-*Proteobacteria* were still the most abundant in ubiquitous diazotrophic communities, accounting for 24.3% of total OTUs (Supporting Information Appendix S1, Figure S3a). δ-*Proteobacteria* were still the most abundant in ubiquitous bacterial communities, accounting for 13.2% of total OTUs (Supporting Information Appendix S1, Figure S3b).

3.2 Species co-occurrence patterns

Overall species co-occurrence patterns of bacterial and diazotrophic communities were non-random across spatial scales, as indicated by C-scores (Supporting Information Appendix S1, Tables S1 and S2). Both communities showed the highest C-scores at the local scale, the medium C-scores at the meso-scale and the lowest C-scores at the regional scale (Figure 1a,b). The result suggested that species interacted more frequently at the local scale. Species interactions became weaker as spatial scale enlarged.

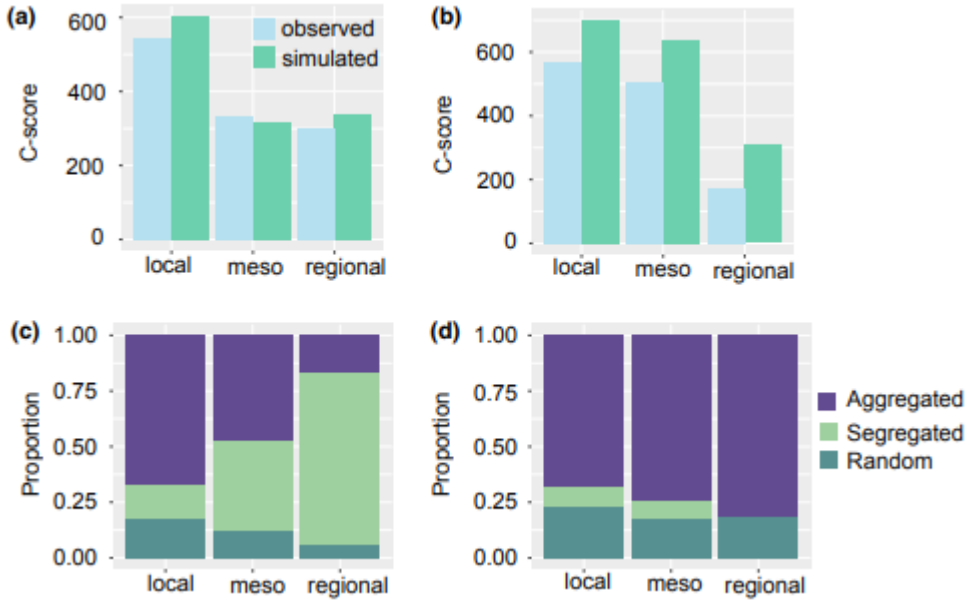

Figure 1. Mean of observed and simulated environmentally-constrained C-scores of (a) bacterial community and (b) diazotrophic community at three spatial scales. Only observed C-scores that are significantly ($P < .050$) different with simulated C-scores are averaged. Relative proportion of aggregated, segregated and random species pairs of (c) bacterial community and (d) diazotrophic community at three spatial scales

Across spatial scales, most of the species pairs significantly co-occurred for both bacterial and diazotrophic communities (Figure 1c,d). Among bacterial species pairs, more species pairs were shown to be aggregated at the local and meso-scales (Figure 1c), suggesting that species co-occurred more than expected at smaller scales. The proportion of segregated species pairs of the bacterial community increased with enlarged spatial scales, suggesting that species associations of bacterial communities have simplified (Figure 1c).

3.3 Spatial scale dependence of microbial β-diversity

The distance–decay relationships of diazotrophic and bacterial communities were examined at the local (1–113 m), meso- (3.4–39 km) and regional (103–668 km) scales. When distance increased from 1 to 668 km, the dissimilarity of diazotrophic communities increased from 0.090 to 0.720, which was substantially higher than that of bacterial communities (from 0.060 to 0.520) (Figure 2). Significant ($p < 0.050$) distance–decay relationships of diazotrophic communities were observed at the local and regional scales but not at the meso-scale (Figure 3a), suggesting that the distance–decay slopes of diazotrophic communities were scale dependent. Scale-dependent distance–decay slopes were also observed for bacterial communities (Figure 3b), but a significant distance–decay slope of bacterial communities was observed only at the local scale (Figure 3b), suggesting that sampling at the meso- and regional scales did not increase community dissimilarity.

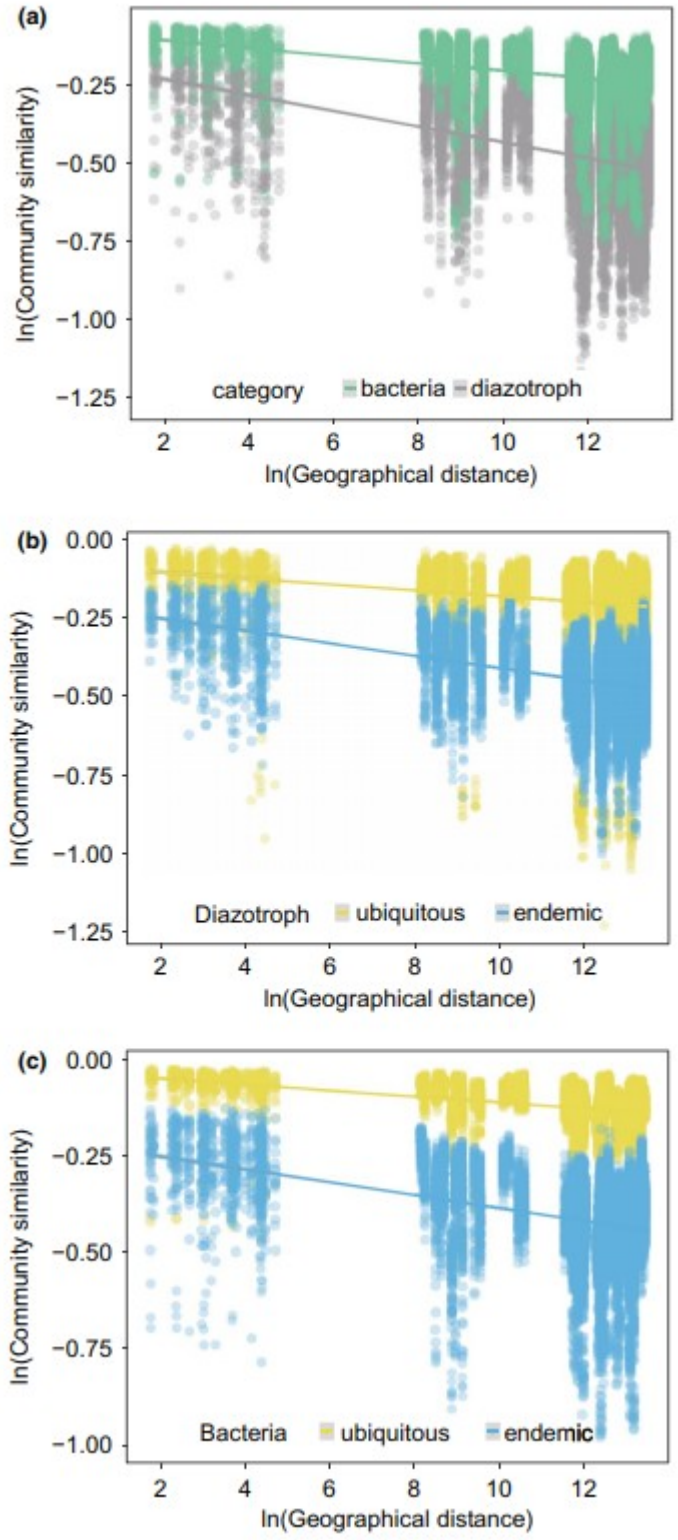

Figure 2. The distance-decay relationships for (a) overall diazotrophic and bacterial communities, (b) ubiquitous and endemic diazotrophic communities and (c) ubiquitous and endemic bacterial communities. The slopes of the distance-decay relationships are significantly ($P < .050$) lower than zero

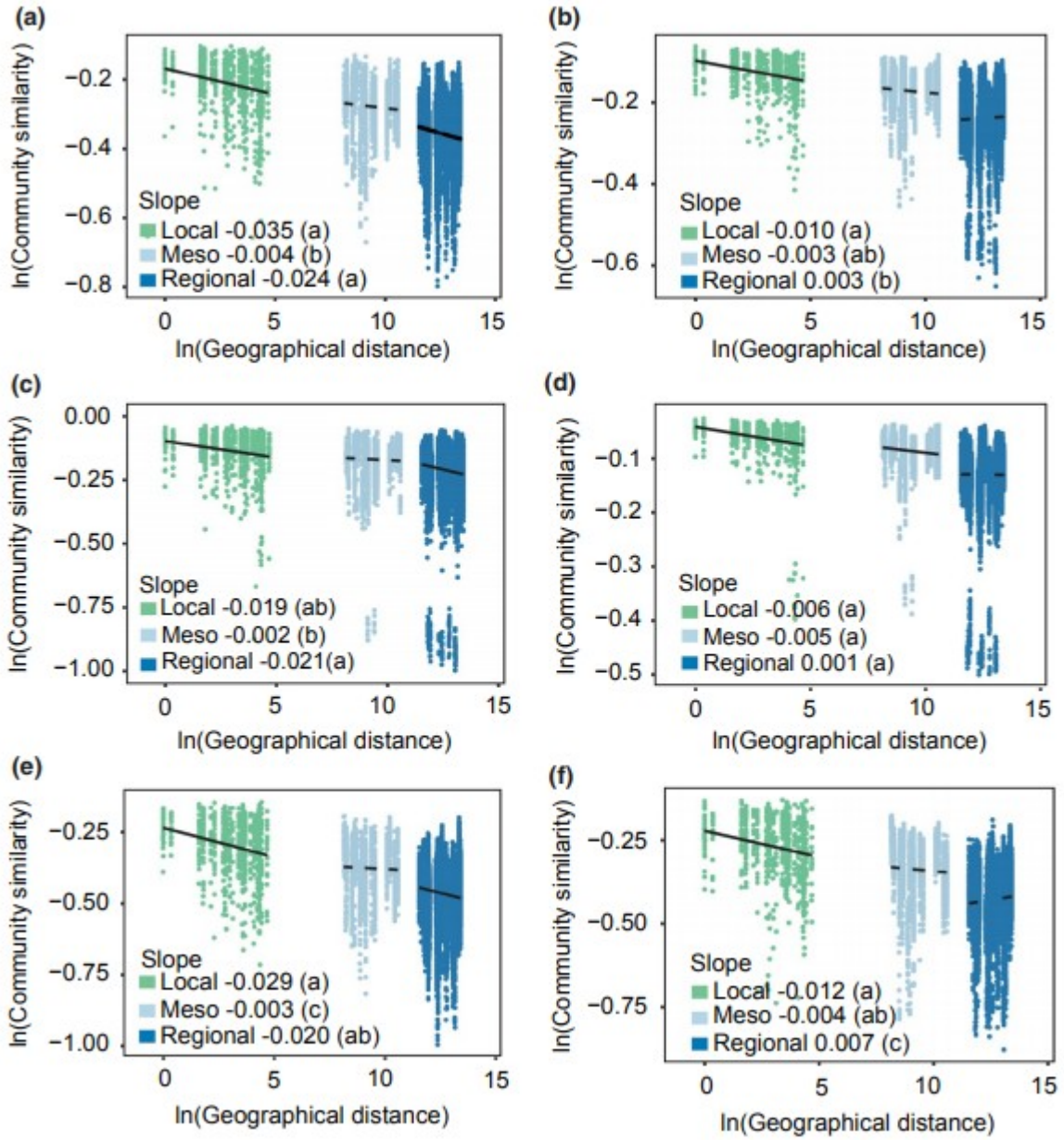

Figure 3. The distance-decay relationships for (a) overall diazotrophic communities, (b) overall bacterial communities, (c) ubiquitous diazotrophic communities, (d) ubiquitous bacterial communities, (e) endemic diazotrophic communities and (f) endemic bacterial communities at the local (green), meso (light blue) and regional (dark blue) scales. The slopes of distance-decay relationships are labeled in figures, and all lines (except the dash lines) are significantly ($P < .050$) lower than zero. Significantly ($P < .050$) different slopes at three spatial scales are represented by different alphabets in parentheses

Resampling of communities showed that sequence depth did not alter the patterns of distance-decay for both bacterial and diazotrophic communities across spatial scales (Supporting Information Appendix S1, Table S3), rejecting the possibility that differences in sampling depth underlie different scale-dependent distance–decay patterns of both communities.

3.4 Discerning drivers of microbial community assembly

To identify abiotic drivers for the distance–decay relationships of microbial communities, we measured a number of climatic and soil geochemical factors to reveal the extent of environmental heterogeneity across sampling sites (Supporting Information Appendix S1, Figure S5). Soil temperature, soil water content, soil pH, concentrations of ammonium, nitrate, dissolved organic carbon and dissolved total N were significantly ($p < 0.050$) different at the regional scale. For example, Jurong had the highest concentration of dissolved total N and nitrate but the lowest soil water content and soil temperature. Xiantao had the highest soil pH and ammonium content. Similar to the observation of high environmental heterogeneity, we observed high variation in network topological properties, with Changxing showing the highest average degree, average clustering coefficient and modularity (Supporting Information Appendix S1, Figure S5). The topological properties showed weak correlations with abiotic factors for both bacterial and diazotrophic communities (Supporting Information Appendix S1, Figure S6).

The MRM analyses showed that geographical distance explained 27% of variation for diazotrophic communities, which was similar to abiotic factors (25%; Table 1). Soil pH was the most important abiotic factor, followed by soil ammonium content and dissolved organic carbon (Table 2). Biotic factors explained only 17% of variation for diazotrophic communities, among which the average degree was the dominant factor. On the contrary, geographical distance explained 17% of variation and biotic factors explained 40% of variation for overall bacterial communities (Table 1), with modularity as the dominant biotic factor (Table 2). Abiotic factors made similar contribution (*c*. 23%) to bacterial and diazotrophic communities (Table 1).

TABLE 1 Relative contributions of geographical distance, abiotic factors and biotic factors to diazotrophic and bacterial communities with different occupancies

		Overall		Endemic		Ubiquitous	
		R^2	*p*	R^2	*p*	R^2	*p*
Diazotroph	Geographical distance	0.273	.001	0.285	.001	0.129	.001
	Abiotic selection[a]	0.254	.001	0.251	.001	0.198	.001
	Biotic selection[b]	0.174	.001	0.097	.001	0.395	.001
Bacteria	Geographical distance	0.171	.001	0.270	.001	0.244	.001
	Abiotic selection	0.230	.001	0.193	.001	0.296	.001
	Biotic selection	0.403	.001	0.180	.001	0.456	.001

[a]Abiotic factors: dissolved organic carbon, dissolved total nitrogen, ammonia, nitrate, soil water content, soil temperature and soil pH.
[b]Biotic factors: average degree, average clustering coefficient and modularity.

TABLE 2 Relative contribution of abiotic and biotic factors for diazotrophic and bacterial communities at different spatial scales

Diazotrophic communities	The overall scale	The local scale	The meso scale	The regional scale
	R^2= 0.501 *b*	R^2= 0.341 *b*	R^2= 0.176 *b*	R^2= 0.286 *b*
Geographic distance	0.392*	0.103*		0.151**
Dissolved organic carbon	0.058*		0.124*	
Ammonium	0.123**			0.141**
Soil temperature				0.102*
Soil pH	0.219**	0.182**	0.224**	0.304**
Average degree	0.295**	0.482**		0.331**
Average clustering coefficient	0.064*			0.098*
Bacterial communities	**The overall scale**	**The local scale**	**The meso scale**	**The regional scale**
	R^2= 0.440 *b*	R^2= 0.253 *b*	R^2= 0.037 *b*	R^2= 0.160 *b*
Geographic distance	0.332**	0.178*		
Dissolved organic carbon	0.055*			
Dissolved total *N*				0.159*
Ammonium	0.115*			
Soil pH	0.215**	0.173*		
Average degree		0.403**		
Average clustering coefficient			0.164*	
Modularity	0.345**			

Note: The variation (R^2) of natural-logarithm transformed community distance matrix explained by the prediction factors and the partial regression coefficient (*b*) of each prediction factor are reported. When a partial regression coefficient is shown, the significance level is labeled by * when $p \leq 0.050$ and by ** when $p \leq 0.001$.

Relative contributions of geographical distance, abiotic and biotic factors to microbial communities varied by spatial scales (Table 2). Specifically, geographical distance and average degree contributed to diazotrophic communities at the local and regional scales but not at the meso-scale. In comparison, geographical distance, average degree and soil pH contributed to bacterial communities at the local but not at the meso- and regional scales.

3.5 Ubiquitous and endemic communities

Spatial scale dependence might differ in ubiquitous and endemic communities; therefore, we examined their distance–decay slopes and possible drivers across spatial scales. Ubiquitous communities of diazotrophs and bacteria showed shallower distance–decay slopes than endemic communities (Figure 2b,c). Geographical distance made a smaller contribution to the variation of ubiquitous communities than endemic communities, whereas biotic factors made a higher contribution to the variation of ubiquitous communities than endemic communities (Supporting Information Appendix S1, Tables S4 and S5).

The distance–decay slopes of ubiquitous diazotrophic communities were similar at the local and regional scales, and were much larger than that at the meso-scale (Figure 3c). Consistently, average degree contributed to ubiquitous diazotrophic communities at the local and regional scales but not at the meso-scale (Supporting Information Appendix S1, Table S4). In contrast, ubiquitous bacterial communities showed no difference in distance–decay slopes across three spatial scales (Figure 3d), suggesting that there was no spatial scale dependence for ubiquitous bacterial communities. This might result from similar contributions of dominant drivers, including soil pH and average degree, across the three spatial scales (Supporting Information Appendix S1, Table S5). Both endemic communities of bacteria and diazotrophs had steeper distance–decay slopes at the local scale than that at the meso- and regional scales (Figure 3e,f), corresponding to the finding that geographical distance contributed to variation of endemic communities at the local but not at the meso- and regional scales (Supporting Information Appendix S1, Tables S4 and S5).

4 DISCUSSION

Few studies have mapped imprints of species associations on changes of microbial β-diversity (Ohlmann et al., 2018). Using C-scores and topological properties of RMT-based association networks to interpret species associations, we were able to examine the biotic effects underpinning spatial variation in microbial β-diversity. We found that after controlling species responses to environmental changes, species interactions (as indicated by C-scores) still persisted (Figure 1). The C-score was the highest at the local scale (Figure 1a,b) and became lower as spatial scales increased, suggesting that species interacted more frequently at smaller spatial scales. Consistently, network topological properties made a higher contribution to community β-diversity at the local scale than at the meso- and regional scales (Table 2).

Although numerous studies have reported differences in the biogeographical patterns of microbial taxa, there have been very few studies to disentangle differences between distance–decay patterns of taxonomic and functional microbial communities (Angermeyer et al., 2015). For the diazotrophic community, the slopes of distance–decay curves were significantly different at the three spatial scales, unveiling high heterogeneity of diazotrophic N-fixing capabilities in paddy soil. This corresponded to high heterogeneity of dissolved total N, nitrate and ammonium concentrations across spatial scales (Supporting Information Appendix S1, Figure S5). Long-term cultivation can promote higher biological N-fixing potential of paddy soil, which might be attributed to enlarged diazotrophic communities (Bannert et al., 2011). Hence, a high variability of diazotrophic β-diversity might have important implications for soil N bioavailability. Notably, we did not observe a significant distance–decay slope for diazotrophic communities at the meso-scale (Figure 3a), suggesting that different paddy fields might share similar functional traits to achieve stable N-fixing functions. Intriguingly, scale-

dependent distance decay was also observed for both ubiquitous and endemic diazotrophic communities (Figure 3c,e), suggesting that the scale dependence of the diazotrophic community was likely to be robust across OTUs with different occupancies.

Soil pH was the only factor that contributed to the variation of diazotrophic community β-diversity at three spatial scales (Table 2), suggesting a strong impact of soil pH on diversification of diazotrophic communities in paddy soil. This result was consistent with extensive studies showing soil pH to be an important abiotic factor shaping microbial community structure in diverse ecosystems (Liu et al., 2016; Shen et al., 2016).

In contrast, a significant distance–decay slope of overall bacterial communities was observed only at the local scale (Figure 3b). The result was consistent with another study of grassland soil microbial communities showing that pronounced heterogeneity of microbial β-diversity was significant only at centimetre scales (O'Brien et al., 2016). Such phenomena might be attributable to several causes. First, paddy soil may become homogenized as a result of cyclic tillage and irrigation within a field. Therefore, homogenized abiotic factors driven by those agricultural practices may contribute more to bacterial community variation compared with spatial distance, which offsets the distance effect. Second, at the local scale, we observed the highest proportion of aggregated species pairs in the bacterial communities, compared with the meso- and regional scales (Figure 1c). A previous study showed that patchy species aggregation in bulk soil might lead to divergent composition of communities (Faust & Raes, 2012). Species competition–cooperation trade-offs might also have shifted with changes in availale nutrients in the rhizosphere (Yang et al., 2018), leading to variation in community composition within the local scale (Chesson & Huntly, 1997; Kneitel & Chase, 2004; Tilman, 2000). Third, sampling across the whole region might not have contributed to significant increases in bacterial diversity above what was already observed at the local scale. This might be because the sequencing depth was not yet sufficient to capture fully the underlying genomic composition of bacterial communities, which have large species pools (Supporting Information Appendix S1, Figure S7).

The spatial scale dependence of bacterial β-diversity could be contingent on habitat types (Lozupone & Knight, 2007; Zinger, Boetius, & Ramette, 2014). For example, the sessile lifestyle of sediment bacteria caused spatial isolation, and stronger variations in environmental conditions triggered different distance–decay slopes (Zinger et al., 2014). A study on the bacterial distance–decay relationship in different habitats showed that marine bacteria had much shallower distance–decay slopes than sediment bacteria but had a similar magnitude of distance–decay slopes to soil bacteria (Ranjard et al., 2013). Furthermore, given that water flow enables the dispersal of marine bacteria, no significant spatial scale dependence was reported for marine bacterial β-diversity (Zinger et al., 2014). Consistently, we detected no distance–decay slopes of bacterial communities at the meso- and regional

scales, and a shallower distance–decay slope (≤ 0.010) at the local scale than those in sediment (0.003–0.070) (Schauer, Bienhold, Ramette, & Harder, 2010), suggesting that paddy soils were generally less limited in bacterial species dispersal.

Although the ubiquitous diazotrophic community exhibited spatial scale dependence in β-diversity, this was not observed for the ubiquitous bacterial community (Figure 3c,d). As the spatial scale increased to encompass a greater environmental gradient, ubiquitous bacteria, with broader niche adaptation than ubiquitous diazotrophs, might have been less susceptible to changes in spatial scales. Moreover, if the community dynamics (dispersal, local adaptation and colonization) were more stochastic, it is plausible that the distance–decay would deteriorate as nearby communities became as different as distant communities (Bell, 2010). We inferred that ubiquitous bacterial community dynamics might be more stochastic than that of ubiquitous diazotrophs owing to the much larger species pool.

In comparison, endemic communities of both diazotrophic and bacterial communities had steeper distance–decay slopes at the local scale than at the meso- and regional scales (Figure 3e,f). The result was consistent with a modelling framework showing that the relative contribution of endemic species to β-diversity is higher at the local scale because endemic species are more aggregated (Morlon et al., 2008).

The biotic mechanisms that accounted for the spatial scaling of microbial β-diversity were probably involved with species associations. For example, competitive exclusion caused by limited nutrient sources among species has been suggested to limit coexistence of species and result in segregation of microbes (Leibold, 1998; Macarthur & Levins, 1967). In contrast, metabolic interdependence among taxa may favour species coexistence that leads to aggregation of microbes (Zelezniak et al., 2015). Those species associations might occur simultaneously and jointly, which contributes to the observed variation in community composition (Violle, Nemergut, Pu, & Jiang, 2011). Based on the null model result, we observed 81–94% non-random species pairs (Figure 1c,d), which was much higher than that observed in plant communities (Blois et al., 2014; Li & Waller, 2016). For example, Blois et al. (2014) found only 2% of plant species pairs to be significant (Blois et al., 2014). Li and Waller (2016) found 16–31% of species pairs of plant communities to have significant associations, even at a smaller scale (1 m^2 quadrats) than ours (Li & Waller, 2016). Those results suggested that species interactions could be more important in structuring microbial communities than previously appreciated, because microbial species were more likely to be aggregated and segregated.

The MRM analysis showed that modularity of networks made the largest contribution to β-diversity of bacterial communities (Table 2). Previous studies have shown that the topology-based modules in microbial networks could be perceived as functional units in microbial communities (Luo, Zhong,

Yang, & Zhou, 2006). Another study also interpreted modules as niches for microbial communities (Chaffron et al., 2010). Accordingly, the higher impact of modularity might be linked to a greater extent of segregation among bacterial species into niches and functional units (i.e., functional differentiation). Null model-based analysis showed a higher proportion of segregated species pairs in the bacterial community than in the diazotrophic community across spatial scales (Figure 1c,d), probably owing to more diverse niches and functional units of bacterial communities than diazotrophic communities.

Despite the potential of species association in structuring microbial communities, our results should be interpreted carefully, because complex microbial associations and interactions are challenging to measure in natural environments. Therefore, using multiple methods to interpret and compare species association patterns is of great importance to demonstrate the reliability of the results. In our study, environmentally constrained species co-occurrence patterns and RMT-based association networks showed consistency in interpreting species association across spatial scales. Our study thus offers a new, reliable way to explore the biotic mechanisms underlying community spatial assemblies.

In summary, this study demonstrates the spatial scale dependence of diazotrophic and bacterial community assemblies in paddy soil and reveals the driving mechanisms. This represents the first attempt to compare spatial assembly processes of microbial functional communities with those for bacterial communities. Given the importance of species associations and their potential relationships with species composition, future empirical and theoretical research that investigates the biotic effect on changes to microbial β-diversity are needed. Moreover, given that ubiquitous species are potentially important organisms for ecosystem functions (Saunders et al., 2016; Zhang, Shao, & Ye, 2012), spatial dependence should be taken into account when examining the diversity of ubiquitous functional communities and associated ecosystem services.

DATA AVAILABILITY STATEMENT

MiSeq data of 16S rRNA gene and *nifH* gene sequencing are available online (https://www.ncbi.nlm.nih.gov/bioproject?LinkName=sra_bioproject&from_uxml:id=5312558), with the accession number PRJNA438873.

REFERENCES

Afgan, E., Baker, D., van den Beek, M., Blankenberg, D., Bouvier, D., Čech, M., … Goecks, J. (2016). The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update. *Nucleic Acids Research*, 44(W1), W3– W10. https://doi.org/10.1093/nar/gkw343.

Angermeyer, A., Crosby, S. C.., & Huber, J. A. (2015). Decoupled distance–decay patterns between dsrA and 16S rRNA genes among salt marsh sulfate-

reducing bacteria. *Environmental Microbiology*, 18, 75– 86. https://doi.org/10.1111/1462-2920.12821

Bannert, A., Kleineidam, K., Wissing, L., Mueller-Niggemann, C., Vogelsang, V., Welzl, G., … Schloter, M. (2011). Changes in diversity and functional gene abundances of microbial communities involved in nitrogen fixation, nitrification, and denitrification in a tidal wetland versus paddy soils cultivated for different time periods. *Applied and Environmental Microbiology*, 77, 6109– 6116. https://doi.org/10.1128/AEM.01751-10

Barberán, A., & Casamayor, E. O. (2011). Euxinic freshwater hypolimnia promote bacterial endemicity in continental areas. *Microbial Ecology*, 61, 465– 472. https://doi.org/10.1007/s00248-010-9775-6

Barberán, A., Bates, S. T., Casamayor, E. O., & Fierer, N. (2011). Using network analysis to explore co-occurrence patterns in soil microbial communities. *The ISME Journal*, 6, 343– 351. https://doi.org/10.1038/ismej.2011.119

Bates, R. G. (1964). *Determination of pH: Theory and practice*. Retrived from http://jes.ecsdl.org/cgi/doi/10.1149/1.2403829

Bell, T. (2010). Experimental tests of the bacterial distance–decay relationship. *The ISME Journal*, 4, 1357– 1365. https://doi.org/10.1038/ismej.2010.77

Berry, D., & Widder, S. (2014). Deciphering microbial interactions and detecting keystone species with co-occurrence networks. *Frontiers in Microbiology*, 5, 219. https://doi.org/10.3389/fmicb.2014.00219

Blois, J. L., Gotelli, N. J., Behrensmeyer, A. K., Faith, J. T., Lyons, S. K., Williams, J. W., … Wing, S. (2014). A framework for evaluating the influence of climate, dispersal limitation, and biotic interactions using fossil pollen associations across the late Quaternary. *Ecography*, 37, 1095– 1108. https://doi.org/10.1111/ecog.00779

Burns, A. R., Stephens, W. Z., Stagaman, K., Wong, S., Rawls, J. F., Guillemin, K., & Bohannan, B. J. M. (2015). Contribution of neutral processes to the assembly of gut microbial communities in the zebrafish over host development. *The ISME Journal*, 10, 655– 664. https://doi.org/10.1038/ismej.2015.142

Caporaso, J. G., Lauber, C. L., Walters, W. A., Berg-Lyons, D., Huntley, J., Fierer, N., … Knight, R. (2012). Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *The ISME Journal*, 6, 1621– 1624. https://doi.org/10.1038/ismej.2012.8

Cazelles, K., Araújo, M. B., Mouquet, N., & Gravel, D. (2016). A theory for species co-occurrence in interaction networks. *Theoretical Ecology*, 9, 39– 48. https://doi.org/10.1007/s12080-015-0281-9

Chaffron, S., Rehrauer, H., Pernthaler, J., & von Mering, C. (2010). A global network of coexisting microbes from environmental and whole-genome sequence data. *Genome Research*, 20, 947– 959. https://doi.org/10.1101/gr.104521.109

Chesson, P., & Huntly, N. (1997). The roles of harsh and fluctuating conditions in the dynamics of ecological communities. *The American Naturalist*, 150, 519– 553. https://doi.org/10.1086/286080

Clauset, A., Newman, M. E., & Moore, C. (2004). Finding community structure in very large networks. *Physical Review E*, 70, 066111. https://doi.org/10.1103/PhysRevE.70.066111

Deng, Y., Jiang, Y.-H., Yang, Y., He, Z., Luo, F., & Zhou, J. (2012). Molecular ecological network analyses. *BMC Bioinformatics*, 13, 113. https://doi.org/10.1186/1471-2105-13-113

Deng, Y., Zhang, P., Qin, Y., Tu, Q., Yang, Y., He, Z., … Zhou, J. (2016). Network succession reveals the importance of competition in response to emulsified vegetable oil amendment for uranium bioremediation. *Environmental Microbiology*, 18, 205– 218. https://doi.org/10.1111/1462-2920.12981

Di Cola, V., Broennimann, O., Petitpierre, B., Breiner, F. T., D'Amen, M., Randin, C., … Guisan, A. (2017). ecospat: An R package to support spatial analyses and modeling of species niches and distributions. *Ecography*, 40, 774– 787. https://doi.org/10.1111/ecog.02671

Diamond, J. (1975). Assembly of species communities. In M. L. Cody & J. M. Diamond (Eds.), *Ecology and evolution of communities* (pp. 342–444). Cambridge, MA: Harvard University Press.

Edgar, R. C. (2013). UPARSE: Highly accurate OTU sequences from microbial amplicon reads. *Nature Methods*, 10, 996– 998. https://doi.org/10.1038/nmeth.2604

Edgar, R. C., Haas, B. J., Clemente, J. C., Quince, C., & Knight, R. (2011). UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics*, 27, 2194– 2200. https://doi.org/10.1093/bioinformatics/btr381

Elser, J. J., Bracken, M. E., Cleland, E. E., Gruner, D. S., Harpole, W. S., Hillebrand, H., … Smith, J. E. (2007). Global analysis of nitrogen and phosphorus limitation of primary producers in freshwater, marine and terrestrial ecosystems. *Ecology Letters*, 10, 1135– 1142. https://doi.org/10.1111/j.1461-0248.2007.01113.x

Faust, K., & Raes, J. (2012). Microbial interactions: From networks to models. *Nature Reviews Microbiology*, 10, 538. https://doi.org/10.1038/nrmicro2832

Fierer, N. (2017). Embracing the unknown: Disentangling the complexities of the soil microbiome. *Nature Reviews Microbiology*, 15, 579. https://doi.org/10.1038/nrmicro.2017.87

Freilich, S., Kreimer, A., Meilijson, I., Gophna, U., Sharan, R., & Ruppin, E. (2010). The large-scale organization of the bacterial network of ecological co-occurrence interactions. *Nucleic Acids Research*, 38, 3857– 3868. https://doi.org/10.1093/nar/gkq118

Fuhrman, J. A. (2009). Microbial community structure and its functional implications. *Nature*, 459, 193– 199. https://doi.org/10.1038/nature08058

Goberna, M., Navarro-Cano, J. A., Valiente-Banuet, A., Garcia, C., & Verdú, M. (2014). Abiotic stress tolerance and competitionrelated traits underlie phylogenetic clustering in soil bacterial communities[J]. *Ecology letters*, 17(10), 1191– 1201. https://doi.org/10.1111/ele.12341.

Goslee, G. C. (2007). The ecodist package for dissimilarity based analysis of ecological data. *Journal of Statistical Software*, 22, 1– 15.

Green, J., & Bohannan, B. J. (2006). Spatial scaling of microbial biodiversity. *Trends in Ecology and Evolution*, 21, 501– 507. https://doi.org/10.1016/j.tree.2006.06.012

Harrell, F. E., Jr. (2001). *Regression modeling strategies, with applications to linear models, survival analysis and logistic regression*. Springer. Retrived from https://link.springer.com/book/10.1007/978-1-4757-3462-1?page=1

Harte, J., Kinzig, A., & Green, J. (1999). Self-similarity in the distribution and abundance of species. *Science*, 284, 334– 336. https://doi.org/10.1126/science.284.5412.334

Hewson, I., Steele, J. A., Capone, D. G., & Fuhrman, J. A. (2006). Temporal and spatial scales of variation in bacterioplankton assemblages of oligotrophic surface waters. *Marine Ecology Progress Series*, 311, 67– 77. https://doi.org/10.3354/meps311067

Kneitel, J. M., & Chase, J. M. (2004). Trade-offs in community ecology: Linking spatial scales and species coexistence. *Ecology Letters*, 7, 69– 80. https://doi.org/10.1046/j.1461-0248.2003.00551.x

Kong, Y. (2011). Btrim: A fast, lightweight adapter and quality trimming program for next-generation sequencing technologies. *Genomics*, 98, 152– 153. https://doi.org/10.1016/j.ygeno.2011.05.009

Leibold, M. A. (1998). Similarity and local co-existence of species in regional biotas. *Evolutionary Ecology*, 12, 95– 110. https://doi.org/10.1023/A:1006511124428

Li, D., & Waller, D. (2016). Long-term shifts in the patterns and underlying processes of plant associations in Wisconsin forests. *Global Ecology and Biogeography*, 25, 516– 526. https://doi.org/10.1111/geb.12432

Li, D., Poisot, T., Waller, D. M., & Baiser, B. (2018). Homogenization of species composition and species association networks are decoupled. *Global Ecology and Biogeography*, 27, 1481– 1491. https://doi.org/10.1111/geb.12825

Liu, J., Sui, Y., Yu, Z., Yao, Q., Shi, Y., Chu, H., … Wang, G. (2016). Diversity and distribution patterns of acidobacterial communities in the black soil zone of northeast China. *Soil Biology and Biochemistry*, 95, 212– 222. https://doi.org/10.1016/j.soilbio.2015.12.021

Loewenstein, Y., Portugaly, E., Fromer, M., & Linial, M. (2008). Efficient algorithms for accurate hierarchical clustering of huge datasets: Tackling the entire protein space. *Bioinformatics*, 24, i41– i49. https://doi.org/10.1093/bioinformatics/btn174

Lozupone, C. A., & Knight, R. (2007). Global patterns in bacterial diversity. *Proceedings of the National Academy of Sciences USA*, 104, 11436– 11440. https://doi.org/10.1073/pnas.0611525104

Luo, F., Zhong, J., Yang, Y., & Zhou, J. (2006). Application of random matrix theory to microarray data for discovering functional gene modules. *Physical Review E*, 73, 031924. https://doi.org/10.1103/PhysRevE.73.031924

Ma, B., Wang, H., Dsouza, M., Lou, J., He, Y., Dai, Z., … Gilbert, J. A. (2016). Geographic patterns of co-occurrence network topological features for soil microbiota at continental scale in eastern China. *The ISME Journal*, 10, 1891–1901. https://doi.org/10.1038/ismej.2015.261

Macarthur, R., & Levins, R. (1967). The limiting similarity, convergence, and divergence of coexisting species. *The American Naturalist*, 101, 377– 385. https://doi.org/10.1086/282505

Martiny, J. B. H., Bohannan, B. J. M., Brown, J. H., Colwell, R. K., Fuhrman, J. A., Green, J. L., … Staley, J. T. (2006). Microbial biogeography: Putting microorganisms on the map. *Nature Reviews. Microbiology*, 4, 102– 112. https://doi.org/10.1038/nrmicro1341

Martiny, J. B. H., Eisen, J. A., Penn, K., Allison, S. D., & Horner-Devine, M. C. (2011). Drivers of bacterial β-diversity depend on spatial scale. *Proceedings of the National Academy of Sciences USA*, 108, 7850– 7854.

Meyer, K. M., Memiaghe, H., Korte, L., Kenfack, D., Alonso, A., & Bohannan, B. J. M. (2018). Why do microbes exhibit weak biogeographic patterns? *The ISME Journal*, 12, 1404– 1413.

Morlon, H., Chuyong, G., Condit, R., Hubbell, S., Kenfack, D., Thomas, D., … Green, J. L. (2008). A general framework for the distance–decay of similarity in ecological communities. *Ecology Letters*, 11, 904– 917. https://doi.org/10.1111/j.1461-0248.2008.01202.x

Nekola, J. C., & White, P. S. (2004). The distance decay of similarity in biogeography and ecology. *Journal of Biogeography*, 26, 867– 878.

Newman, M. E. J. (2003). The structure and function of complex networks. *SIAM Review*, 45, 167– 256. https://doi.org/10.1137/S003614450342480

O'Brien, S. L., Gibbons, S. M., Owens, S. M., Hampton-Marcell, J., Johnston, E. R., Jastrow, J. D., … Antonopoulos, D. A. (2016). Spatial scale drives patterns

in soil bacterial diversity. *Environmental Microbiology*, 18, 2039– 2051. https://doi.org/10.1111/1462-2920.13231

Ohlmann, M., Mazel, F., Chalmandrier, L., Bec, S., Coissac, E., Gielly, L., … Thuiller, W. (2018). Mapping the imprint of biotic interactions on β-diversity. *Ecology Letters*, 21, 1660– 1669. https://doi.org/10.1111/ele.13143

Penton, C. R., Yang, C., Wu, L., Wang, Q., Zhang, J., Liu, F., Qin, Y., Deng, Y., Hemme, C. L., Zheng, T., Schuur, E. A. G., Tiedje, J., & Zhou, J.. (2016). NifH-harboring bacterial community composition across an Alaskan permafrost thaw gradient. *Frontiers in Microbiology*, 7, 1894.

Peterson, A. T. (2011). Ecological niche conservatism: a time-structured review of evidence. *Journal of Biogeography*, 38, 817– 827. https://doi.org/10.1111/j.1365-2699.2010.02456.x

Poisot, T., Canard, E., Mouillot, D., Mouquet, N., & Gravel, D.(2012). The dissimilarity of species interaction networks. *Ecology Letters*, 15, 1353–1361. https://doi.org/10.1111/ele.12002

Poly, F., Monrozier, L. J., & Bally, R. (2001). Improvement in the RFLP procedure for studying the diversity of *nifH* genes in communities of nitrogen fixers in soil. *Research in Microbiology*, 152, 95– 103. https://doi.org/10.1016/S0923-2508(00)01172-4

Proulx, S. R., Promislow, D. E. L., & Phillips, P. C. (2005). Network thinking in ecology and evolution. *Trends in Ecology and Evolution*, 20, 345– 353. https://doi.org/10.1016/j.tree.2005.04.004

Ranjard, L., Dequiedt, S., Chemidlin Prévost-Bouré, N., Thioulouse, J., Saby, N., Lelievre, M., … Lemanceau, P. (2013). Turnover of soil bacterial diversity driven by wide-scale environmental heterogeneity. *Nature Communications*, 4, ncomms2431. https://doi.org/10.1038/ncomms2431

Saunders, A. M., Albertsen, M., Vollertsen, J., & Nielsen, P. H. (2016). The activated sludge ecosystem contains a core community of abundant organisms. *The ISME Journal*, 10, 11– 20. https://doi.org/10.1038/ismej.2015.117

Schauer, R., Bienhold, C., Ramette, A., & Harder, J. (2010). Bacterial diversity and biogeography in deep-sea surface sediments of the South Atlantic Ocean. *The ISME Journal*, 4, 159– 170. https://doi.org/10.1038/ismej.2009.106

Shen, C., Shi, Y., Ni, Y., Deng, Y., Van Nostrand, J. D., He, Z., … Chu, H. (2016). Dramatic increases of soil microbial functional gene diversity at the treeline ecotone of Changbai Mountain. *Frontiers in Microbiology*, 7, 1184.

Shi, S., Nuccio, E. E., Shi, Z. J., He, Z., Zhou, J., & Firestone, M. K. (2016). The interconnected rhizosphere: High network complexity dominates rhizosphere assemblages. *Ecology Letters*, 19, 926– 936. https://doi.org/10.1111/ele.12630

Shi, Y., Li, Y., Xiang, X., Sun, R., Yang, T., He, D., … Chu, H. (2018). 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. https://doi.org/10.1186/s40168-018-0409-4

Thuiller, W., Georges, D., Engler, R., Breiner, F., Georges, M. D., & Thuiller, C. W. (2016). Package 'biomod2' [J]. Species distribution modeling within an ensemble forecasting framework. Retrived from ftp://ftp2.de.freebsd.org/pub/misc/cran/web/packages/biomod2/biomod2.pdf

Tilman, D. (2000). Causes, consequences and ethics of biodiversity. *Nature*, 405, 208– 211. https://doi.org/10.1038/35012217

Violle, C., Nemergut, D. R., Pu, Z., & Jiang, L. (2011). Phylogenetic limiting similarity and competitive exclusion. *Ecology Letters*, 14, 782– 787. https://doi.org/10.1111/j.1461-0248.2011.01644.x

Wang, Q., Garrity, G. M., Tiedje, J. M., & Cole, J. R. (2007). Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Applied and Environmental Microbiology*, 73, 5261– 5267. https://doi.org/10.1128/AEM.00062-07

Wang, Q., Quensen, J. F., 3rd, Fish, J. A., Lee, T. K., Sun, Y., Tiedje, J. M., & Cole, J. R. (2013). Ecological patterns of *nifH* genes in four terrestrial climatic zones explored with targeted metagenomics using FrameBot, a new informatics tool. *Mbio*, 4, e00592-13. https://doi.org/10.1128/mBio.00592-13

Woodcock, S., Curtis, T. P., Head, I. M., Lunn, M., & Sloan, W. T. (2006). Taxa–area relationships for microbes: The unsampled and the unseen. *Ecology Letters*, 9, 805– 812. https://doi.org/10.1111/j.1461-0248.2006.00929.x

Wu, L., Yang, Y., Chen, S. i., Zhao, M., Zhu, Z., Yang, S., … He, Q. (2016). Long-term successional dynamics of microbial association networks in anaerobic digestion processes. *Water Research*, 104, 1– 10. https://doi.org/10.1016/j.watres.2016.07.072

Yang, T., Tedersoo, L., Soltis, P. S., Soltis, D. E., Gilbert, J. A., Sun, M., … Chu, H. (2018). Phylogenetic imprint of woody plants on the soil mycobiome in natural mountain forests of eastern China. *The ISME Journal*, 13, 686– 697.

Zelezniak, A., Andrejev, S., Ponomarova, O., Mende, D. R., Bork, P., & Patil, K. R. (2015). Metabolic dependencies drive species co-occurrence in diverse microbial communities. *Proceedings of the National Academy of Sciences USA*, 112, 6449– 6454. https://doi.org/10.1073/pnas.1421834112

Zhang, T., Shao, M.-F., & Ye, L. (2012). 454 Pyrosequencing reveals bacterial diversity of activated sludge from 14 sewage treatment plants. *The ISME Journal*, 6, 1137– 1147. https://doi.org/10.1038/ismej.2011.188

Zhang, Y., Cong, J., Lu, H., Deng, Y., Liu, X., Zhou, J., & Li, D. (2016). Soil bacterial endemism and potential functional redundancy in natural broadleaf

forest along a latitudinal gradient. *Scientific Reports*, 6, 28819. https://doi.org/10.1038/srep28819

Zhou, J., Bruns, M. A., & Tiedje, J. M. (1996). DNA recovery from soils of diverse composition. *Applied and Environmental Microbiology*, 62, 316– 322.

Zhou, J., Deng, Y., Luo, F., He, Z., Tu, Q., & Zhi, X. (2010). Functional molecular ecological networks. *MBio*, 1, e00169-10.

Zhou, J., Kang, S., Schadt, C. W., & Garten, C. T. (2008). Spatial scaling of functional gene diversity across various microbial taxa. *Proceedings of the National Academy of Sciences USA*, 105, 7768– 7773. https://doi.org/10.1073/pnas.0709016105

Zinger, L., Boetius, A., & Ramette, A. (2014). Bacterial taxa–area and distance–decay relationships in marine environments. *Molecular Ecology*, 23, 954– 964. https://doi.org/10.1111/mec.12640