Title

Scale-dependent effects of climate and geographic distance on bacterial diversity patterns across northern China's grasslands

Journal

FEMS Microbiology Ecology, 91(12)

ISSN

0168-6496

Authors

Wang, Xiaobo

Van Nostrand, Joy D

Deng, Ye

et al.

Publication Date

2015-12-01

DOI

10.1093/femsec/fiv133

Peer reviewed

Scale-dependent effects of climate and geographic distance on bacterial diversity patterns across northern China's grasslands

Xiaobo Wang[1,2,3], Joy D. Van Nostrand[3], Ye Deng[4], Xiaotao Lu[1], Chao Wang[1], Jizhong Zhou[3,5,6] and Xingguo Han[1,*]

[1]State Key Laboratory of Forest and Soil Ecology, Institute of Applied Ecology, Chinese Academy of Sciences, Shenyang 110164, PR China, [2]University of Chinese Academy of Sciences, Beijing 100049, PR China, [3]Department of Microbiology and Plant Biology, Institute for Environmental Genomics, University of Oklahoma, Norman, OK 73019, USA, [4]Research Center for Eco-Environmental Sciences, Chinese Academy of Sciences, Beijing 100085, PR China, [5]Earth Science Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA and [6]State Key Joint Laboratory of Environment Simulation and Pollution Control, School of Environment, Tsinghua University, Beijing 100084, PR China

∗Corresponding author: State Key Laboratory of Forest and Soil Ecology, Institute of Applied Ecology, Chinese Academy of Sciences, Shenyang 110164, PR China. Tel: +86 24 83970301; Fax: +86 24 83970300; E-mail: xghan@ibcas.ac.cn

Abstract

Patterns of variation in plant and animal diversity along precipitation gradients have been extensively studied, but much less is known about how and to what extent precipitation affects the biogeographic distribution of microbial diversity in arid areas across large spatial scales. Here we collected soils from 54 sites along a 3700 km transect covering a wide range of grassland ecosystems with distinct aridity gradients. We quantified the bacterial community diversity and the effects of climate, edaphic parameter and geographic distance on the bacterial community structure using high-throughput 16S rRNA gene sequencing. Of the 35 phyla detected, 6 were dominant: *Actinobacteria*, *Acidobacteria*, *Alphaproteobacteria*, *Deltaproteobacteria*, *Bacteroidetes* and *Planctomycetes*. Aridity was a major factor influencing bacterial diversity, community composition and taxon abundance. Although the pattern of bacterial species richness is markedly different from that of plant species richness, most soil bacteria were endemic to particular bioregions like macro-organisms. Community similarity significantly declined with environmental distance and geographic distance ($r = -0.579$ and -0.773, respectively). Geographic distance (historical contingencies) contributed more to bacterial community variation (36.02%) than combined environmental factors (24.06%). Overall, our results showed that geographic distance and climatic factors concurrently govern bacterial biogeographic patterns in arid and semi-arid grassland.

Keywords: microbial biogeography; dispersal limitation; spatial patterns; metagenomic sequencing; soil microbial ecology; aridity; Inner Mongolia

INTRODUCTION

Soil microorganisms play vital roles in ecosystems and mediate a number of critical processes such as biogeochemical cycling (Hodge, Campbell and Fitter 2001; Balser and Firestone 2005) and soil formation (Rillig and Mummey 2006). Despite their ubiquity and importance, our understanding of the spatial patterns of microbial diversity across large spatial scales is limited (Ward *et al.*1998; Horner-Devine, Carney and Bohannan 2004; Fierer and Jackson 2006; Bardgett and Van der Putten 2014).

There is currently debate regarding which factors are the primary drivers of microbial spatial distribution at larger spatial scales (Lauber *et al.*2009), although growing evidence indicates that soil microorganisms have a restricted global distribution due to variations in climatic, soil and plant conditions (Cho and Tiedje 2000; Zhou *et al.*2002). There is still question as to whether processes and factors that control spatial patterns for macro-organism communities, such as contemporary environmental conditions and historical contingencies, also drive the spatial variation of microbial diversity and what the relative importance of these factors is (Martiny *et al.*2006; Ramette and Tiedje 2007; Ge *et al.*2008). The traditional view is that a cosmopolitan dispersal of microbes occurs because of their minimum size and high local abundance (Finlay and Clarke 1999; Finlay 2002), and that spatial diversification in microbial communities is mainly driven by contemporary environmental conditions (McArthur, Kovacic and Smith 1988; Lauber *et al.*2008), or, as Becking put it, ‘everything is everywhere, but, the environment selects’ (Bass Becking 1934). For example, a considerable body of research has shown that soil pH is a key factor accounting for the majority of the variation in soil bacterial diversity and biogeographic distribution across regional and continental scales (Fierer and Jackson 2006; Rousk *et al.*2010; Chu *et al.*2010; Griffiths *et al.*2011). However, in different ecosystems at different spatial scales, we are likely to get completely different results as other environmental factors, such as nutrient availability (Broughton and Gross 2000; Liu *et al.*2010), salinity (Crump *et al.*2004; Lozupone and Knight 2007), plant diversity and community composition (Stephan, Meyer and Schmid 2000; Marschner *et al.*2001; Kuske *et al.*2002; Johnson *et al.*2003; Wardle *et al.*2004) and profile depth (Fierer, Schimel and Holden 2003), have been found to influence the composition and diversity of soil bacterial communities. Studies examining different biomes or ecosystem types have clearly demonstrated the importance of environmental or habitat filtering in relation to soil microorganisms, indicating that microbial spatial patterns are strongly driven by contemporary biotic and abiotic factors (Ramette and Tiedje 2007; Green, Bohannan and Whitaker 2008).

Historical contingencies, such as dispersal limitation, are also key factors affecting spatial variation of microbial diversity. Dispersal is a key process shaping biogeographic patterns in aboveground organisms and soil biota, including bacteria, fungi and nematodes (Fierer 2008; Peay, Garbelotto and Bruns 2010; Nielsen *et al.*2014). Geographic isolation of free-living

microorganisms is relatively common at local, regional and global scales (Horner-Devine, Carney and Bohannan 2004; Fierer 2008). Dispersal is likely a key limiting factor of community assemblages over space and becomes an important driver of microbial evolution, particularly at larger spatial scales (Hubbell 2001; Talbot *et al.*2014). Studies have shown a 'distance-decay relationship', in which community composition between sites becomes less similar with increasing geographic distance. Studies of *Sulfobolus* assemblages in hot spring (Papke *et al.*2003; Whitaker, Grogan and Taylor 2003) and soil *Pseudomonas* (Cho and Tiedje 2000) found that geographic distance better explained microbial community variation than environmental heterogeneity. Historical separation and geographic isolation appear to be strong key drivers in determining microbial assemblage differences at large spatial scales. As such, geographic distance may be the best predictor of diversified variation among communities, but this assumption needs to be tested further. A systematic, broad scale survey along a continuous geographic and climatic gradient would be a more effective way to reveal relationships between microbial community composition and geographic distance or environmental variables.

To investigate comprehensively the influences of environmental heterogeneity and geographic distance on the spatial variation of microbial diversity, we carried out a field survey, sampling 54 sites along a 3700 km transect across the arid and semi-arid ecosystems in northern China (Supplementary Fig. S1). The region is at the eastern end of the contiguous Eurasian steppe, which is characterized by typical continental climate with limited precipitation occurring mainly in the summer. We assessed soil bacterial diversity, community composition and biogeographic distribution along this transect using 16S rRNA gene sequencing, and collected extensive environmental data. This study aims to address three questions: (i) How does the diversity of bacterial communities vary spatially along this environmental gradient? (ii) Does either environmental heterogeneity or geographic distance exert a stronger influence on the spatial variation of bacterial diversity and community composition? And (iii) what are the relative contributions of each to the spatial patterning of bacterial communities?

Table 1. Pearson correlation coefficients between alpha diversity (OTU richness, Faith's PD and Shannon index) and environmental variables.

r	AI	Altitude	TOC	TN	TP	C/N	N/P	pH	SM	PR
Richness	0.601**	−0.118	0.537**	0.544**	−0.031	0.317*	0.318*	−0.255	0.223	0.561**
PD	0.523**	−0.037	0.448**	0.463**	−0.043	0.234	0.316*	−0.201	0.173	0.478**
Shannon	0.301*	0.024	0.236	0.285*	0.047	−0.021	0.513**	0.023	0.126	0.308*

Asterisk represent significance of correlation (*$P < 0.05$; **$P < 0.01$). PD, phylogenetic diversity; AI, aridity index; TOC, total organic carbon (%); TN, total nitrogen (%); TP, total phosphorus (%); C/N, carbon/nitrogen; N/P, nitrogen/phosphorus; SM, soil moisture (%), PR, plant richness.

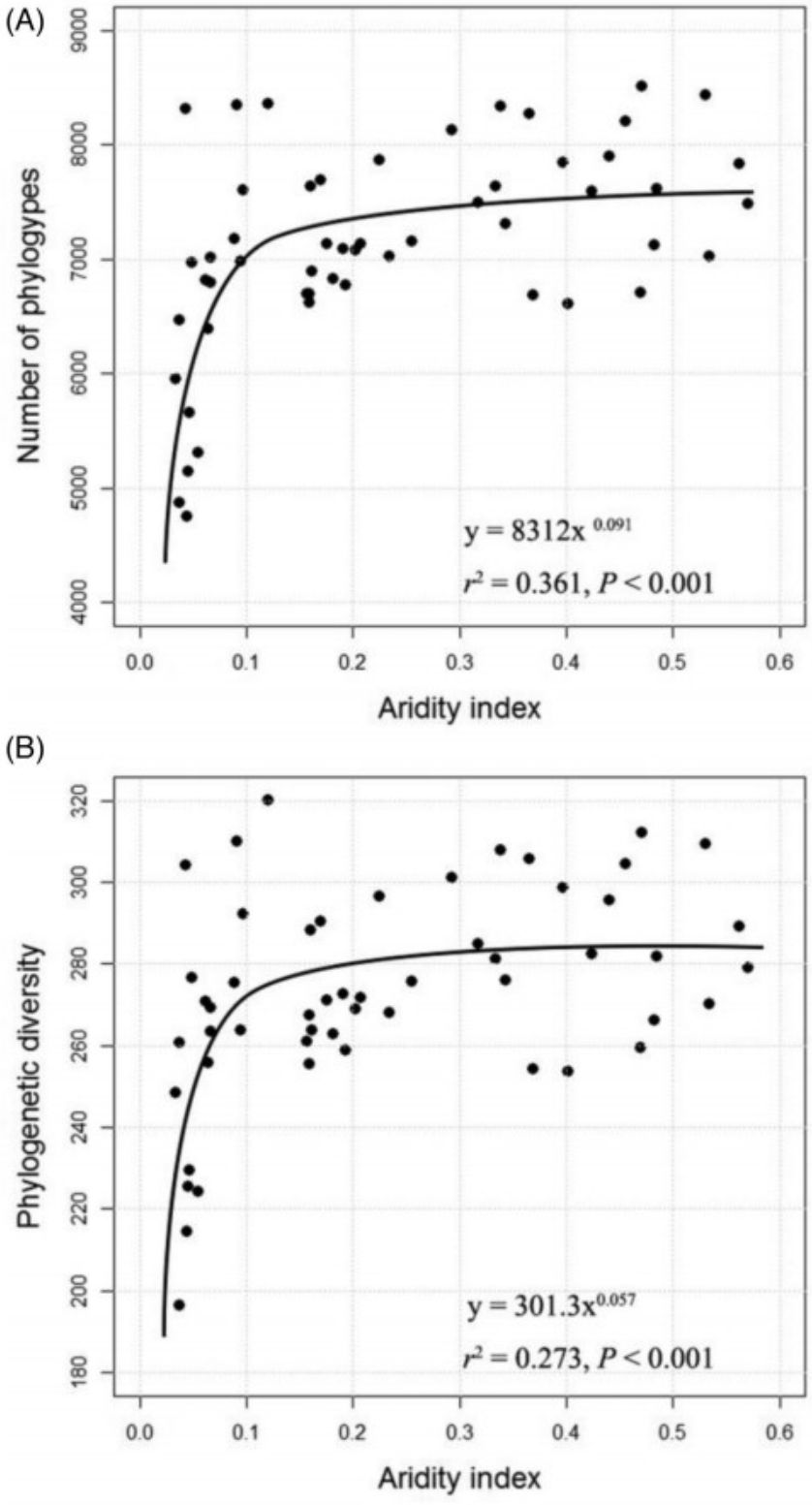


Figure 1. The relationships of bacterial operational taxonomic unit richness (**A**) and phylogenetic diversity (**B**) with aridity index based on 97% sequence identity. The communities were randomly sampled at the 173,260 sequences level.

MATERIALS AND METHODS

Experimental design and site description

The study was conducted along a 3700 km transect of arid and semi-arid grasslands from the Xinjiang Uygur Autonomous Region to eastern Inner Mongolia in northern China (87.38° E to 120.36° E, 42.22° N to 49.19° N) (Supplementary Fig. S1). The mean annual precipitation (MAP) ranges from 38 to 436 mm and the mean annual temperature (MAT) ranges from −2.9 to 9.4ºC. The sampling sites were at elevations of 530 to 1620 m (average, 1052 m) (Supplementary Table S1). From west to east, the main vegetation types are *Calligonum* spp., *Alhagi* ssp. and *Ephedra* spp. (desert steppe), *Stipa* spp., *Leymus* spp. and *Agropyron* spp. (typical steppe) and *Leymus* spp., *Stipa* spp. and *Achnatherum* spp. (meadow steppe). Dominant soil types are arid, sandy and brown loess rich in calcium. A total of 54 sites along the transect were selected based on two criteria: (i) sites had nearly intact natural plant communities only with light animal grazing; and (ii) sites represented the local flora and soil types covering an area of no less than 10,000 m^2. At each site, two large plots (50 m × 50 m) at a distance of <2 km across the transect were designed. Within each of the large plots, five quadrats (1 m × 1 m) were selected based on the five-spot-sampling method (Yang *et al.*2010).

Field sampling

Sampling was carried out in July and August 2012, near the period of highest plant biomass production and species richness, from the west to the east. At each site, spatial geographic coordinates and elevation were recorded by GPS (eTrex Venture, Garmin, USA). In each quadrat, litter was removed. Samples of live, aboveground plants were clipped, sorted by species and stored in paper bags for biomass measurement, nutrient analysis, aboveground net primary production (ANPP) determination, and calculation of plant species richness (PR). Soil samples per quadrat were collected randomly from 5–10 soil cores (2.5 cm diameter × 10 cm depth) of the upper 10 cm of soil and mixed thoroughly. Composite soils were sieved through a 2.0 mm mesh to remove roots and rocks, homogenized by hand and separated into two parts: one was preserved for subsequent characterization of soil chemistry and the other was placed into a sterile plastic bag and immediately stored at −40ºC for later DNA extraction.

Distance and climate data

The pairwise geographic distance was calculated using the Imap package in R v.3.1.0 according to the GPS coordinates of each site. MAP and MAT of each sampling site were obtained from the WorldClim global climate dataset (Hijmans *et al.*2005). Extracted data were processed in ArcGIS version 9.3 using Spatial Analysis tool (ESRI, Redlands, CA, USA). An aridity index (AI) of each site was estimated by the ratio of precipitation to potential evapotranspiration (PET) (AI = MAP/PET).

Soil physicochemical analysis

Total organic carbon (TOC) and total nitrogen (TN) were determined using wet oxidation and a modified Kjeldahl procedure as previously described (Wang *et al.*2014). Total P was measured by colorimetric analysis with ammonium molybdate and persulfate oxidation (Kuo 1996). Soil pH was measured after creating a 1:2.5 (volume) fresh soil to water slurry. Soil moisture was determined gravimetrically after drying in an oven at 105ºC for 12 h. All replicates were analysed separately and values were averaged to obtain site-level estimates using the mean of each quadrat.

Soil DNA extraction

DNA was extracted from 0.5 g soil using the MoBio PowerSoil® DNA isolation kit (MoBio Laboratories, Carlsbad, CA, USA) according to the manufacturer's protocol. With samples from the desert steppe, which had low biomass, DNA was extracted using the method of freeze-grinding and SDS-based lysis (Zhou, Bruns and Tiedje 1996) to obtain sufficient high-quality DNA. The quality of the purified DNA was assessed based on 260/280 nm and 260/230 nm absorbance ratios obtained using a NanoDrop ND-1000 spectrophotometer (NanoDrop Technologies Inc., Wilmington, DE, USA). The final DNA concentration was quantified using a PicoGreen (Life Technologies, Grand Island, NY, USA) assay (Ahn, Costa and Emanuel 1996) with a FLUOstar Optima (BMG Labtech, Jena, Germany), stored at −20ºC until use.

16S rRNA gene amplification and sequencing

To determine the soil bacterial community composition and diversity in each soil sample, an amplicon survey of a portion of the 16S rRNA gene was performed. We used a barcoded primer set (515F, 5′-GTGCCAGCMGCCGCGGTAA-3′ and 806R, 5′-GGACTACHVGGGTWTCTAAT-3′) targeting the V4 region of both bacterial and archaeal 16S rRNA genes, which has high sequence coverage for both bacteria and archaea and produces an appropriately sized amplicon (253 bp by excluding primers) for Illumina sequencing. Both primers contained Illumina adapters and the reverse primer contained a 12 bp barcode sequence unique to each sample. Purified DNA (10 ng) from each sample was used as a template for PCR amplification in a 25 µl reaction volume under the following conditions: initial denaturation at 94ºC for 1 min followed by 30 cycles of 94ºC for 20 s, 53ºC for 25 s and 68ºC for 45 s, ending with an final extension step of 10 min at 68ºC. All samples were amplified in triplicate and combined. PCR amplicons from all samples were then pooled in equimolar concentrations. Primer and primer dimers were separated out by electrophoresis on a 1% agarose gel, and final PCR products were recovered using a QIAquick gel extraction kit (Qiagen, Valencia, CA, USA). Sequencing was conducted on an Illumina Miseq sequencer at the Institute for Environmental Genomics, University of Oklahoma.

Processing of sequencing data

Raw sequences with an average quality score >30 and with no ambiguous base calls were processed using an in-house pipeline that was built on the Galaxy platform and incorporates various software tools (http://rccc.ou.edu). In brief, the sequences were quality trimmed using Btrim (Kong 2011) and were assigned to their respective samples based on the unique barcodes. Paired end reads were merged into full length sequences by FLASH v1.2.5 (Magoc and Salzberg 2011). Chimeric sequences were discarded based on prediction by Uchime (usearch v5.2.3) (Edgar *et al.*2011) using the reference database mode. Sequences were clustered into operational taxonomic units (OTUs) at the ≥97% identity threshold with UPARSE (Edgar 2013) and singleton OTUs were removed. Taxonomic assignment was carried out with the Ribosomal Database Project (RDP) classifier. To correct for sampling effort (number of analysed sequences per sample), we used a randomly selected subset of 173,260 sequences per sample for subsequent community analysis. Alpha diversity was determined using both taxonomic (number of OTUs and Shannon index, H′) and phylogenetic (Faith's phylogenetic diversity) metrics.

Statistical analysis

All statistical tests and graphics were done using the program R, version 3.1.0. Phylogenetic diversity was estimated using Faith's index, which incorporates the phylogenetic breadth across taxonomic levels (Faith 1992; Faith *et al.*2009). Matrices of the pairwise taxonomic distance between communities (Bray–Curits) and the Euclidean distance between geographic coordinates and standardized environmental variables were constructed within the package vegan (Oksanen *et al.*2013). Non-metric multidimensional scaling (NMDS) ordination was used to specifically assess changes in bacterial composition along the transect. NMDS analyses were performed using the MetaMDS function based on dissimilarities calculated using the Bray–Curtis index, and environmental vectors were fitted using the envfit and ordisurf routines. BioEnv and canonical correspondence analysis (CCA) were also performed to determine the most significant environmental variables shaping the microbial community composition. These variables were then used to construct the environmental variables matrix for variation partitioning analysis (VPA) with the vegan package. To determine the influences of geographic distance on microbial community structure, we first transformed the spatial geographic distance to rectangular data points that were suitable for constrained ordination with the package PCNM (Borcard and Legendre 2002), and then conducted CCA to select the most significant vectors for variation partitioning analysis. Detrended correspondence analysis (DCA) was also employed to determine community changes using the decorana function. In addition, cluster analysis of the bacterial communities based on the Bray–Curtis abundance-based dissimilarity matrix was conducted with the package vegan, and classification boxes were added using the ordicluster function. Best of fit modeling of the regression between environmental factors and diversity and the relative abundance of bacterial

taxa were performed in SigmaPlot (version 10.0) using ordinary least squares linear, quadratic, piecewise and power law functions. We used Mantel tests with 9999 permutations to examine the correlation between community similarity (1 minus Bray–Curtis dissimilarity index) and geographic (Pearson correlation) or environmental distance (Spearman correlation) within the vegan package.

Nucleotide sequence accession number

DNA sequences in this study have been deposited in SRA of NCBI database under accession number SRP063935.

RESULTS

Climate and geochemical characteristics along the transect

The major geographic and physicochemical characteristics of the 54 sampling sites are summarized in Supplementary Table S1. AI, TOC, and PR tended to increase with longitude, while pH tended to decrease. Soil pH varied from 6.35 to 9.24. The soil organic C (TOC), total N (TN), and total P (TP) varied from 0.05 to 4.46%, 0.01 to 0.38% and 0.01 to 0.08%, respectively. Plant species richness (PR) per square meter and aboveground net primary production (ANPP) at the site level varied from 0 to 23 and from 0.72 to 286.51 g m^{-2} $year^{-1}$, respectively. The aridity gradient can be considered as a surrogate of MAP and MAT gradient (Supplementary Fig. S2). TOC was positively correlated with TN ($r = 0.991$, $P < 0.01$), and also PR was positively correlated with ANPP ($r = 0.812$, $P < 0.01$). Pairwise distances between sampling sites ranged from 12 to 3700 km.

The distribution of taxa and alpha diversity

A total of 14,649,589 high-quality sequences were identified from all soil samples examined before resampling, averaging 384,823 sequences per sample (ranging from 173,260 to 596,386). The sequences were grouped into 31,248 operational taxonomic units (OTUs) using an arbitrary 97% sequence similarity cutoff. Of these sequences, 90.6% could be classified at the phylum level. All samples were compared at an equivalent sequencing depth of 173,260 randomly selected 16S rRNA gene amplicons per sample. The dominant phyla across all soils were *Actinobacteria*, *Acidobacteria*, *Bacteroidetes*, *Alphaproteobacteria*, *Deltaproteobacteria* and *Planctomycetes* (>5% average relative abundance across all soils), accounting for more than 70% of the bacterial sequences. Additional phyla including *Cyanobacteria*, *Nitrospira*, *Chlamydiae*, *OD1*, *BRC1* and *Chlorobi* were also present in nearly all soils at relatively low abundances, and 18 other rarer phyla were identified. Archaea were relatively rare in all soils (only 0.8% of the detected sequences) but were most abundant in the more arid areas. *Halobacteria* and *Thermoprotei* were the dominant archaeal groups (59% of the detected archaeal sequences).

Bacterial alpha diversity, as estimated by the number of OTUs, phylogenetic diversity (PD) and Shannon index, varied across the transect. The richness of the bacterial communities ranged from 4752 to 8518 phylotypes per sample with all samples compared at an identical sequencing depth. Of all environmental variables examined, AI was most correlated with OTU richness ($r = 0.601$, $P < 0.001$), PD ($r = 0.523$, $P < 0.001$) and Shannon index ($r = 0.301$, $P = 0.027$). In addition, the TOC, TN, N/P and PR were also correlated with alpha diversity, while soil pH and other variables such as altitude, TP, C/N and SM were weakly or not correlated with alpha diversity (Table 1). Lower bacterial diversity was observed in the more arid areas (lower AI values), as indicated by the changes in OTU richness, PD and Shannon index. The diversity (OTU richness and PD) increased steeply as AI increased in the more arid areas, but no significant variation was observed with further increases of AI (Fig. 1). Interestingly, the OTU richness of Archaea was highly negatively correlated with AI ($r = -0.740$, $P < 0.001$) and positively correlated with soil pH ($r = 0.491$, $P = 0.001$) (Supplementary Fig. S3). Unlike the pattern for bacteria, higher Archaeal diversity was found at the more xeric area.

Bacterial community structure and beta diversity

The NMDS plot of the Bray–Curtis distance ordination clearly indicated significant variability in soil bacterial community composition across the transect that was primarily related to aridity and soil pH (Fig. 2). This is further supported by correlation analyses between community Bray–Curtis distance and aridity and soil pH (Mantel test: $r = 0.669$ and 0.480, respectively, $P = 0.001$). There were also significant correlations between bacterial community composition and other measured physical and geochemical factors, including C/N, TN, TOC, TP, soil moisture (SM) and altitude, but these composite factors did not significantly improve correlation (Supplementary Table S2). The structure of bacterial communities varied greatly across the AI gradient, as indicated by detrended correspondence analysis (DCA) (Supplementary Fig. S4). Results of canonical correspondence analysis (CCA) also showed that AI had the strongest effects on bacterial community assembly (Fig. 3).

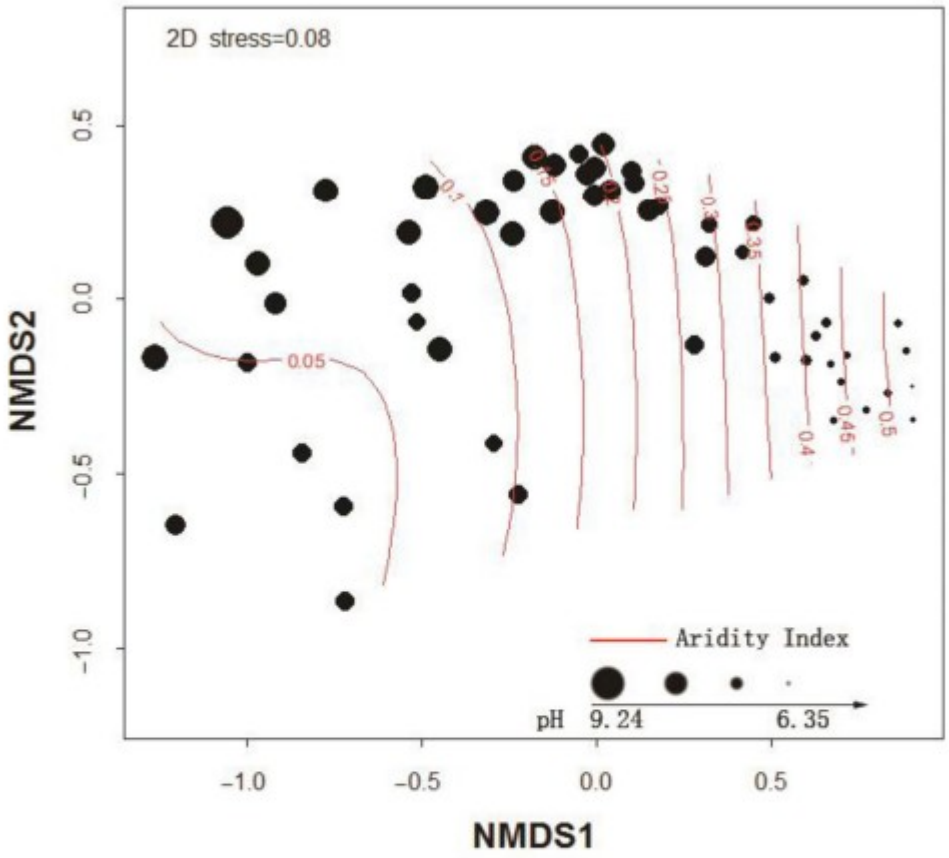


Figure 2. Non-metric multidimensional scaling (NMDS) ordination of microbial community composition from 54 sites across the transect. The compositional variation is represented with Bray–Curtis distance matrix based on the abundance of OTUs. The two dominant environmental gradients associated with ordination are represented: red lines represent the gradient of aridity index; the sizes of circles indicate the values of pH.

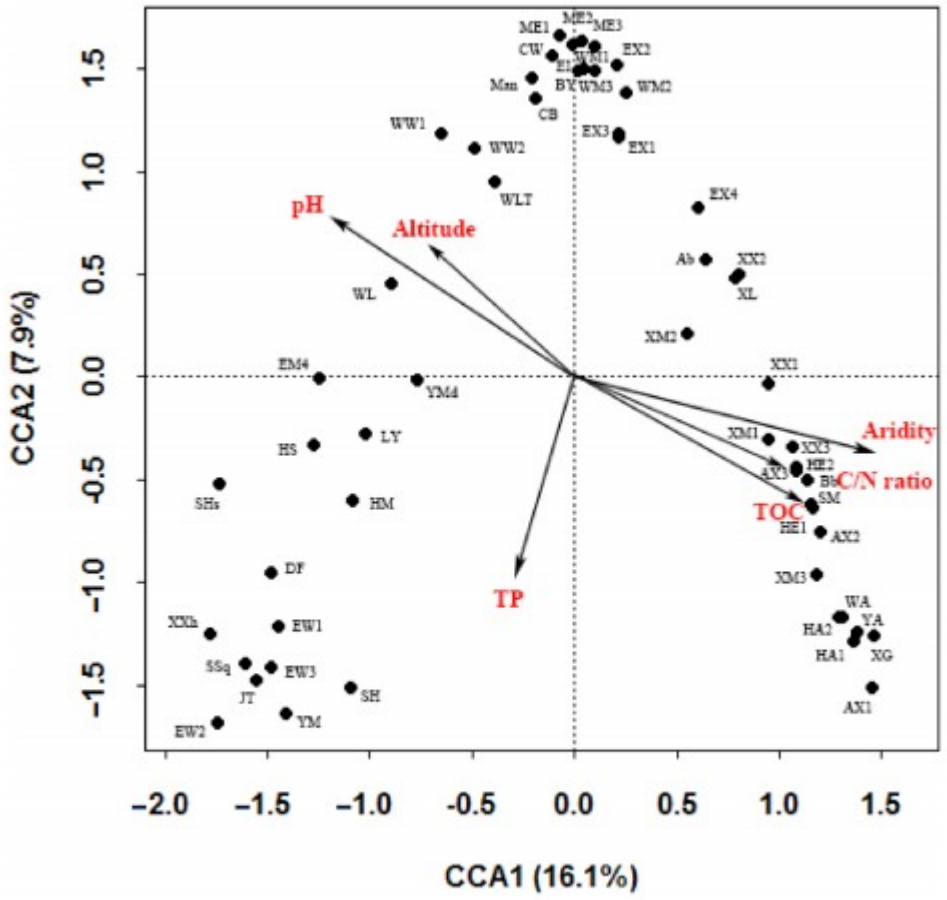


Figure 3. Canonical correspondence analysis (CCA) of the bacterial communities with site abbreviation. Model ANOVA test: $^{*}P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$.

Pairwise community similarity between all soil samples was calculated using the Bray–Curtis abundance-based dissimilarity index, which was highly correlated with the incidence-based Jaccard index (Mantel test: $r = 0.993$, $P = 0.001$). Community similarity was negatively correlated with both geographic distance (Mantel test: $r = -0.773$, $P < 0.001$) and environmental distance (Mantel test: $r = -0.579$, $P < 0.001$) for each pairwise set of samples (Fig. 4). Thus, both environmental factors and geographic distance are key determinants in shaping bacterial community structures. Variation partition analysis was performed to quantify the relative contributions of environmental parameters and geographic distance to bacterial community structure. A subset of environmental variables (aridity, TP, pH, TOC, C/N ratio, altitude) was selected by the BioEnv procedure and together explained 24.06% of the bacterial community variation, whereas geographic distance alone explained 36.02% of the variation, leaving 39.92% unexplained (Fig. 5). Of the selected environmental variables, aridity, TP, pH, TOC, C/N ratio, and altitude individually explained 5.63%, 5.33%, 5.12%, 3.10%, 2.53%, 2.35% of observed variation, respectively (Fig. 5).

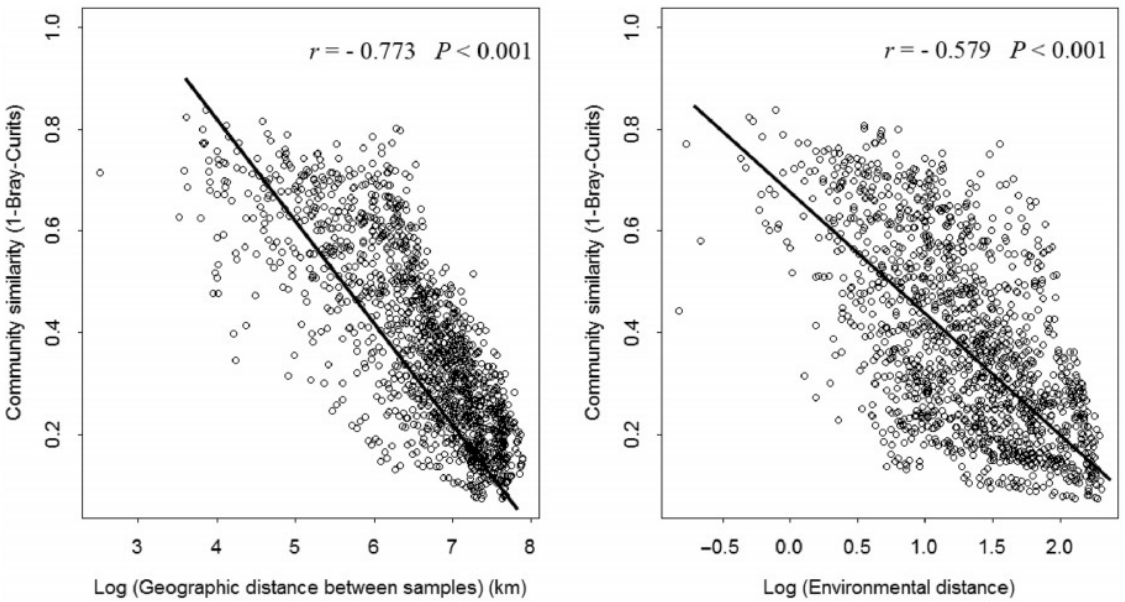
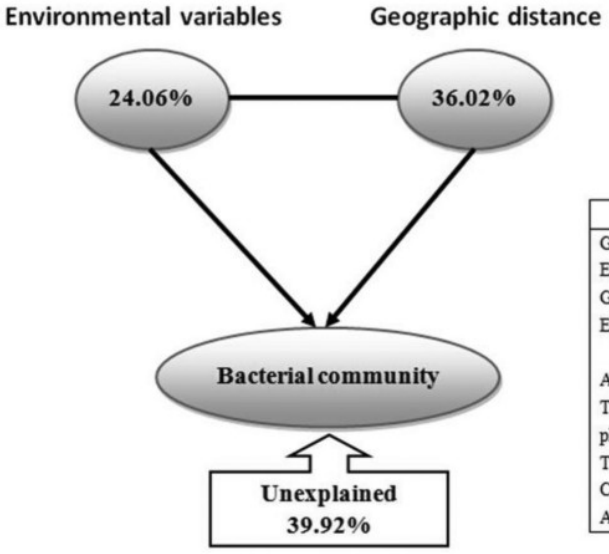


Figure 4. Relationships between the log of geographic distance, log of environmental distance and bacterial community similarity estimated by the Bray–Curtis index.



	Explained	P
Geographic distance	59.69%	0.001
Environmental Variables	33.58%	0.001
Geographic distance | Environmental Variables	36.02%	0.008
Environmental Variables | Geographic distance	24.06%	0.001
Aridity | TP+pH+TOC+ C/N ratio+Altitude	5.63%	0.005
TP | Aridity+pH+TOC+C/N ratio+Altitude	5.33%	0.005
pH | Aridity+TP+TOC+C/N ratio+Altitude	5.12%	0.005
TOC | Aridity+TP+pH+C/N ratio+Altitude	3.10%	0.010
C/N ratio | Aridity+TP+pH+TOC+Altitude	2.53%	0.015
Altitude | Aridity+TP+pH+TOC+C/N ratio	2.35%	0.010

Figure 5. Variation partition analysis of the effects of geographic distance and environmental variables on bacterial community structure. Significance (P) was determined by an ANOVA-like permutation test in CCA.

The bacterial communities from all soils were clustered into five groups based on Bray–Curtis distance (Fig. 6). Group 1 consisted of eight sites located in Xinjiang and Gansu province, from longitude 87° 23′ E to 97° 16′ E and had AIs lower than 0.1. Group 2 contained only one site near a Yardang Landform. Group 3 consisted of six sites from Gansu province and western Inner Mongolia, from longitude 99° 52′ E to 103° 45′ E and had AIs lower than 0.05. Group 4 consisted of 18 sites with 0.1 < AI < 0.3 from the middle of Inner Mongolia from longitude 104° 53′ E to 113° 28′ E. Group 5 contained 21 sites in the wetter areas with 0.3 < AI < 0.53 from northeast of Inner

Mongolia from longitude 114° 05′ E to 120° 21′ E. Soils collected from proximal locations harbored similar bacterial communities.

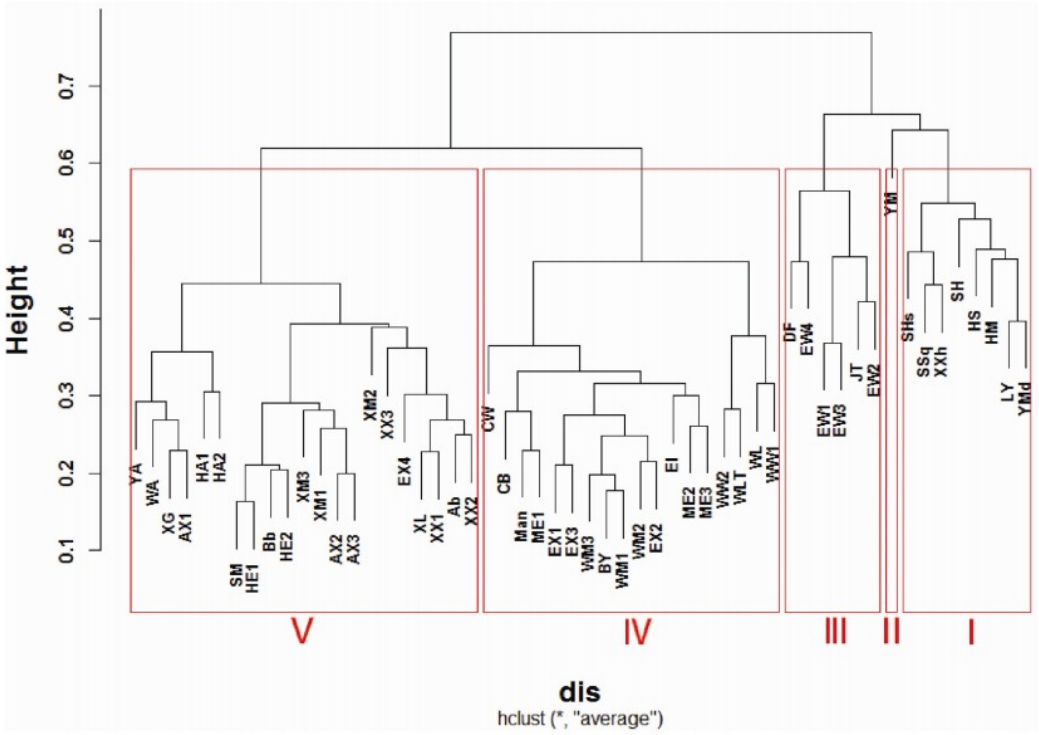


Figure 6. Cluster analysis of bacterial communities based on Bray–Curtis distance matrix. The analysis was performed using vegan package in R.

The taxonomic abundance of dominant phyla

The relative abundance of dominant bacterial groups in each site changed across the aridity gradient. The relative abundance of *Alphaproteobacteria*, *Betaproteobacteria*, *Gammaproteobacteria* and *Bacteroidetes* showed non-linear relationships with AI, with decreased abundance in the more arid areas and increased abundance in areas with higher AI (Supplementary Fig. S5). In contrast, the relative abundance of *Acidobacteria* and *Planctomycetes* increased with increasing AI in the more arid areas but showed no further variation as AI increased more (Supplementary Fig. S5). Significant negative correlation was observed between the AI and the relative abundance of *Actinobacteria* and *Chloroflexi* ($r = -0.845$ and -0.510, respectively, $P < 0.0001$), while the opposite trend was observed for the phyla *Deltaproteobacteria* and *Verrucomicrobia* ($r = 0.850$ and 0.901, respectively, $P < 0.0001$) (Supplementary Fig. S5). Soil pH revealed a negative correlation with the relative abundance of most of the dominant phyla, but showed positive correlation with a few phyla such as *Actinobacteria* ($r = 0.729$, $P < 0.0001$), *Firmicutes* ($r = 0.419$, $P = 0.002$) and *Chloroflexi* ($r = 0.652$, $P < 0.0001$).

DISCUSSION

A large number of studies have shown that water availability is positively correlated with richness, diversity and abundance of macroorganims (e.g. Hawkins *et al*. 2003). Interestingly, in our study bacterial alpha diversity and abundance did not strictly follow this water-availability paradigm. Although more arid soils harbored far lower diversity than other relative mesic soils, the variation in bacterial community alpha diversity was larger at more arid areas than at the more mesic areas (Fig. 1). Abrupt changes in bacterial diversity within extremely dry areas (AI < 0.1) indicate that soil microbes living in areas of lower water availability may be easily activated by even small rainfall events at levels even lower than needed for plants in similar areas (Collins *et al.*2008; Dijkstra *et al.*2012). In fact, our data clearly showed that plant species richness and biomass significantly increased with increasing water availability along the aridity gradient (Supplementary Fig. S6). It is apparent that the pattern of plant species richness followed a different trend compared with that of bacterial OTU richness, suggesting that the spatial patterns of macroscopic plants are governed by different mechanisms, possibly because soil microorganisms are more stress-tolerant than macroorganisms (Gaston 2000; Wardle 2002).

Soil pH has been considered as the major factor determining soil bacterial diversity and composition (Fierer and Jackson 2006; Lauber *et al.*2009; Chu *et al.*2010; Griffiths *et al.*2011), yet results from this study indicated that bacterial diversity and community composition is mainly correlated with AI rather than soil pH. This may be partly because the sampling sites in this study were located in dry land where the soil pH ranged from neutral to alkaline (pH 6.35–9.24). The lack of acidic sites in our transect may have partially obscured the effect of soil pH on bacterial diversity. This effect has been well documented in a long-term N deposition experiment in the *Leymus chinensis* steppe (Yao *et al.*2014), where bacterial diversity decreased significantly only when the pH dropped below 6. Since all of our samples have pH values higher than 6, the effect of pH on diversity may not be apparent in these arid and semi-arid grassland ecosystems. Aridity may play such an important role in structuring the bacterial community diversity over space for at least two possible reasons. First, water availability plays determining roles in regulating and maintaining ecosystem functioning, especially in arid areas. Our study sites were located in arid and semi-arid grasslands where water is a key limiting factor influencing the cycling of nutrients and the diversity and productivity of plant communities, which in turn affect the quantity and quality of plant litter input to below-ground communities, i.e. microorganisms (Bai *et al.*2008). This speculation is supported by two recent studies of microbial biogeographic patterns in the Chihuahuan Desert on the Mexican Plateau (Clark *et al.*2009) and Negev Desert in southern Israel (Angel *et al.*2010), which showed the distribution pattern of soil microbial community in the arid and semi-arid areas correlated mainly with soil water content and organic matter. Secondly, the effect of environmental factors on the belowground community structure is

scale-dependent (Ettema and Wardle 2002; Fierer *et al.*2009; Bardgett and Wardle 2010). At small or local scales (centimeters to meters), spatial patterns of soil biota are often determined by variation in the physicochemical properties of soil such as soil carbon and nutrient availability (Wardle 2002). At regional and continental scales, climate factors such as precipitation or temperature become more important than soil physicochemical parameters (Tedersoo *et al.*2012; Bardgett and Van der Putten 2014).

In this study, all bacterial communities were dominated by six major groups, *Actinobacteria*, *Acidobacteria*, *Alphaproteobacteria*, *Deltaproteobacteria*, *Bacteroidetes* and *Planctomycetes*. We observed that bacterial diversity and the relative abundance of the dominant phyla were strongly correlated with aridity but also that changes in the abundance of these dominant phyla varied across the AI gradient (Supplementary Fig. S5). This suggests that the taxonomic patterns are largely driven by differences in the abundance of major taxonomic groups. The *Actinobacteria* phylum was more abundant in more arid areas, whereas *Verrucomicrobia*, in contrast, was less abundant in those areas, similar to other studies of desert microbial communities (Pointing *et al.*2009; Fierer *et al.*2012). A non-linear relationship was observed between *Alphaproteobacteria*, *Betaproteobacteria*, *Gammaproteobacteria* and *Bacteroidetes* and AI. The abundance of these phyla showed a significant increase with increasing AI above a particular threshold, which may be due to increases in soil carbon availability. For example, *Betaproteobacteria* and *Bacteroidetes*, which are categorized as copiotrophic groups (Fierer, Bradford and Jackson 2007), are considered as the initial metabolizers of labile carbon input and are thus more abundant in soils with high carbon availability (McCaig, Glover and Prosser 1999; Axelrood *et al.*2002; Padmanabhan *et al.*2003). More interestingly, below a certain AI, abundance of these four phyla decreased with increasing AI until reaching their lowest abundance at an AI tipping point. We speculate that this disparity could be caused by differences in carbon use efficiency. In contrast to the copiotrophic *Bacteroidetes*, *Acidobacteria* generally exhibit oligotrophic attributes and are negatively correlated with soil carbon availability (Fierer, Bradford and Jackson 2007; Mannisto, Tiirola and Haggblom 2007). Also, some studies reported that *Acidobacterial* abundance is regulated by pH, with the highest abundances being present in environments with the lowest pH values (Mannisto, Tiirola and Haggblom 2007; Lauber *et al.*2008). In our study, however, the relative abundance of *Acidobacteria* was influenced by AI and the lowest abundances were present in the most arid areas and with a sharp increase in abundance with increasing AI. This is likely because soils in more arid areas are more alkaline. The effect of soil pH on *Acidobacterial* abundance is probably indirect but further research is required to test this hypothesis.

Our results showed a significant correlation between bacterial community dissimilarity and geographic distance or environmental distance (Fig. 4),

which indicates that the spatial variation of biodiversity and composition in bacterial communities across a large-scale gradient may reflect concurrent influences of contemporary environmental heterogeneity and historical contingencies. This conclusion was further confirmed by results of cluster analysis in which similar bacterial communities were observed in locations or habitats in the proximity, and that the occurrence of endemism in different groups was also driven by environmental factors like aridity (Fig. 6). Previous studies claimed that dispersal limitation is less important for microorganisms than contemporary environmental selection due to their small size providing an essentially unlimited capacity for long distance dispersal (Fenchel, Esteban and Finlay 1997; Finlay 2002). However, our results clearly demonstrated that geographic distance explained more of the variation in bacterial community structure than the selected environmental variables (36.02% and 24.06%, respectively) (Fig. 5). This result suggests that local environmental conditions likely play a secondary role to large-scale geographic processes in structuring microbial communities, while historical contingencies may contribute more to shaping patterns of spatial variability in microbial communities across large spatial scales. By contrast, some surveys at smaller spatial scales found that environmental factors were primary determinants in regulating the bacterial biogeographic patterns rather than distance (Horner-Devine, Carney and Bohannan 2004; Hollister *et al.*2010). This discrepancy could be attributed to the different scales at which the studies were conducted, because the relative importance of environmental and historical factors depends on the survey scale (Ricklefs 2004). As a result, the factors operating at large spatial scales, like dispersal limitation, may become apparent only in studies of spatial structure rather than over smaller scales or site-specific habitats.

This study represents an attempt to investigate bacterial diversity and spatial distribution patterns with a unique transect survey across one of the largest geographic gradient in a single biome. We show that the diversity and composition of soil bacterial communities can be well predicated by aridity in arid and semi-arid ecosystems. Both geographic distance and climatic factors govern bacterial biogeographic patterns, suggesting that the factors operating at large spatial scales, like climate and dispersal limitation, are the determinants of bacterial community structure. Future research should focus more on the mechanisms and processes that regulate microbial biodiversity over such large scales of homogeneous landscapes and understanding the structure–function relationship of microbial communities in the ecosystem.

Acknowledgments

We thank all the members of the Shenyang Sampling Campaign Team from the Institute of Applied Ecology, Chinese Academy of Sciences for their assistance during field sampling. We also thank Daliang Ning and Chongqing Wen for their assistance in laboratory work and Xiangzhen Li and Shuijin Hu for comments on the earlier version. This work was financially supported by

the Strategic Priority Research Program of the Chinese Academy of Sciences (XDB15010401), the State Key Laboratory of Forest and Soil Ecology (LFSE2015-19), the Office of the Vice President for Research at the University of Oklahoma, the Collaborative Innovation Center for Regional Environmental Quality, and the Youth Innovation Promotion Association CAS (2014174).

References

Ahn SJ, Costa J, Emanuel JR. PicoGreen quantitation of DNA: Effective evaluation of samples pre- or post-PCR. Nucl Acids Res 1996;24:2623–5.

Angel R, Soares MI, Ungar ED, et al. Biogeography of soil archaea and bacteria along a steep precipitation gradient. ISME J 2010;4:553–63.

Axelrood PE, Chow ML, Radomski CC, et al. Molecular characterization of bacterial diversity from British Columbia forest soils subjected to disturbance. Can J Microbiol 2002;48: 655–74.

Bai YF, Wu JG, Xing Q, et al. Primary production and rain use efficiency across a precipitation gradient on the Mongolia plateau. Ecology 2008;89:2140–53.

Balser TC, Firestone MK. Linking microbial community composition and soil processes in a California annual grassland and mixed-conifer forest. Biogeochemistry 2005;73: 395–415.

Bardgett RD, Van der Putten WH. Belowground biodiversity and ecosystem functioning. Nature 2014;515:505–11.

Bardgett RD, Wardle DA. Aboveground-belowground Linkages: Biotic Interactions, Ecosystem Processes, and Global Change. Oxford: Oxford University Press, 2010.

Bass Becking LGM. Geobiologie of Inleiding tot de Milieukunde. The Hague: Van Stockum & Zoon, 1934.

Borcard D, Legendre P. All-scale spatial analysis of ecological data by means of principal coordinates of neighbour matrices. Ecol Model 2002;153:51–68.

Broughton LC, Gross KL. Patterns of diversity in plant and soil microbial communities along a productivity gradient in a Michigan old-field. Oecologia 2000;125:420–7.

Cho JC, Tiedje JM. Biogeography and degree of endemicity of fluorescent Pseudomonas strains in soil. Appl Environ Microbiol 2000;66:5448–56.

Chu H, Fierer N, Lauber CL, et al. Soil bacterial diversity in the Arctic is not fundamentally different from that found in other biomes. Environ Microbiol 2010;12:2998–3006.

Clark JS, Campbell JH, Grizzle H, et al. Soil microbial community response to drought and precipitation variability in the Chihuahuan Desert. Microb Ecol 2009;57:248–60.

Collins SL, Sinsabaugh RL, Crenshaw C, et al. Pulse dynamics and microbial processes in aridland ecosystems. J Ecol 2008;96:413–20.

Crump BC, Hopkinson CS, Sogin ML, et al., Microbial biogeography along an estuarine salinity gradient: Combined influences of bacterial growth and residence time. Appl Environ Microbiol 2004;70:1494–505.

Dijkstra FA, Augustine DJ, Brewer P, et al. Nitrogen cycling and water pulses in semiarid grasslands: are microbial and plant processes temporally asynchronous? Oecologia 2012;170:799– 808.

Edgar RC. UPARSE: highly accurate OTU sequences from microbial amplicon reads. Nat Methods 2013;10:996–8.

Edgar RC, Haas BJ, Clemente JC, et al. UCHIME improves sensitivity and speed of chimera detection. Bioinformatics 2011;27:2194–200.

Ettema CH, Wardle DA. Spatial soil ecology. Trends Ecol Evol 2002;17:177–83.

Faith DP. Conservation evaluation and phylogenetic diversity. Biol Conserv 1992;61:1–10.

Faith DP, Lozupone CA, Nipperess D, et al. The cladistic basis for the phylogenetic diversity (PD) measure links evolutionary features to environmental gradients and supports broad applications of microbial ecology’s “phylogenetic beta diversity” framework. Int J Mol Sci 2009;10:4723–41.

Fenchel T, Esteban GF, Finlay BJ. Local versus global diversity of microorganisms: cryptic diversity of ciliated protozoa. Oikos 1997;80:220–5.

Fierer N. Microbial biogeography: patterns in microbial diversity across space and time. In: Zengler K (ed.). Accessing Uncultivated Microorganisms: From the Environment to Organisms and Genomes and Back. Washington DC: ASM Press, 2008, 95– 115.

Fierer N, Bradford MA, Jackson RB. Toward an ecological classification of soil bacteria. Ecology 2007;88:1354–64.

Fierer N, Carney KM, Horner-Devine MC, et al. The biogeography of ammonia-oxidizing bacterial communities in soil. Microb Ecol 2009;58:435–45.

Fierer N, Jackson RJ. The diversity and biogeography of soil bacterial communities. Proc Natl Acad Sci U S A 2006;103:626–31.

Fierer N, Leff JW, Adams BJ, et al. Cross-biome metagenomic analyses of soil microbial communities and their functional attributes. Proc Natl Acad Sci U S A 2012;109:21390–5.

Fierer N, Schimel J, Holden P. Variations in microbial community composition through two soil depth profiles. Soil Biol Biochem 2003;35:167–76.

Finlay BJ. Global dispersal of free-living microbial eukaryote species. Science 2002;296:1061–3.

Finlay BJ, Clarke KJ. Ubiquitous dispersal of microbial species. Nature 1999;400:828.

Gaston KJ. Global patterns in biodiversity. Nature 2000;405:220–7.

Ge Y, He J, Zhu Y, et al. Differences in soil bacterial diversity: driven by contemporary disturbances or historical contingencies? ISME J 2008;2:254–64.

Green JL, Bohannan BJM, Whitaker RJ. Microbial biogeography: From taxonomy to traits. Science 2008;320:1039–43.

Griffiths RI, Thomson BC, James P, et al. The bacterial biogeography of British soils. Environ Microbiol 2011;13:1642–54.

Hawkins BA, Field R, Cornell HV, et al. Energy, water, and broad-scale geographic patterns of species richness. Ecology 2003;84:3105–17.

Hijmans RJ, Cameron SE, Parra JL, et al. Very high resolution interpolated climate surfaces for global land areas. Int J Climatol 2005;25:1965–78.

Hodge A, Campbell CD, Fitter AH. An arbuscular mycorrhizal fungus accelerates decomposition and acquires nitrogen directly from organic material. Nature 2001;413:297–9.

Hollister EB, Engledow AS, Hammett AJM, et al. Shifts in microbial community structure along an ecological gradient of hypersaline soils and sediments. ISME J 2010;4:829–38.

Horner-Devine MC, Carney KM, Bohannan BJM. An ecological perspective on bacterial biodiversity. Proc Biol Sci 2004;271:113–22.

Hubbell SP. The Unified Neutral Theory of Biodiversity and Biogeography. Princeton: Princeton University Press, 2001.

Johnson D, Booth RE, Whiteley AS, et al. Plant community composition affects the biomass, activity and diversity of microorganisms in limestone grassland soil. Eur J Soil Sci 2003;54:671–7.

Kong Y. Btrim: A fast, lightweight adapter and quality trimming program for next-generation sequencing technologies. Genomics 2011;98:152–3.

Kuo S. Phosphorus. In: Sparks DL (ed.). Methods of Soil Analysis. Madison: SSSA and ASA, 1996, 869–919.

Kuske CR, Ticknor LO, Miller ME, et al. Comparison of soil bacterial communities in rhizospheres of three plant species and the interspaces in an arid grassland. Appl Environ Microbiol 2002;68:1854–63.

Lauber CL, Hamady M, Knight R, et al. Pyrosequencing-based assessment of soil pH as a predictor of soil bacterial community structure at the continental scale. Appl Environ Microbiol 2009;75:5111–20.

Lauber CL, Strickland MS, Bradford MA, et al. The influence of soil properties on the structure of bacterial and fungal communities across land-use types. Soil Biol Biochem 2008;40:2407–15.

Liu Z, Fu B, Zheng X, et al. Plant biomass, soil water content and soil N:P ratio regulating soil microbial functional diversity in a temperate steppe: A regional scale study. Soil Biol Biochem 2010;42:445–50.

Lozupone CA, Knight R. Global patterns in bacterial diversity. Proc Natl Acad Sci U S A 2007;104:11436–40.

McArthur JV, Kovacic DA, Smith MH. Genetic diversity in naturalpopulations of a soil bacterium across a landscape gradient. Proc Natl Acad Sci U S A 1988;85:9621–4.

McCaig AE, Glover LA, Prosser JI. Molecular analysis of bacterial community structure and diversity in unimproved and improved upland grass pastures. Appl Environ Microbiol 1999;65:1721–30.

Magoc T, Salzberg SL. FLASH: fast length adjustment of short reads to improve genome assemblies. Bioinformatics 2011;27:2957–63.

Mannisto MK, Tiirola M, Haggblom MM. Bacterial communities in Arctic fjelds of Finnish Lapland are stable but highly pH-dependent. FEMS Microbiol Ecol 2007;59:452–65.

Marschner P, Yang CH, Lieberei R, et al. Soil and plant specific effects on bacterial community composition in the rhizosphere. Soil Biol Biochem 2001;33:1437–45.

Martiny JBH, Bohannan BJM, Brown JH, et al. Microbial biogeography: putting microorganisms on the map. Nat Rev Microbiol 2006;4:102–12.

Nielsen UN, Ayres E, Wall DH, et al. Global-scale patterns of assemblage structure of soil nematodes in relation to climate and ecosystem properties. Global Ecol Biogeogr 2014;23: 968–78.

Oksanen JF, Blanchet FG, Kindt R, et al. vegan: Community Ecology Package. R Package version 2.0-10. 2013.

Padmanabhan P, Padmanabhan S, DeRito C, et al. Respiration of C-13-labeled substrates added to soil in the field and subsequent 16S rRNA gene analysis of C-13-labeled soil DNA. Appl Environ Microbiol 2003;69:1614–22.

Papke RT, Ramsing NB, Bateson MM, et al. Geographical isolation in hot spring cyanobacteria. Environ Microbiol 2003;5:650–9.

Peay KG, Garbelotto M, Bruns TD. Evidence of dispersal limitation in soil microorganisms: Isolation reduces species richness on mycorrhizal tree islands. Ecology 2010;91:3631–40.

Pointing SB, Chan YK, Lacap DC, et al. Highly specialized microbial diversity in hyper-arid polar desert. Proc Natl Acad Sci U S A 2009;106:19964–9.

Ramette A, Tiedje JM. Multiscale responses of microbial life to spatial distance and environmental heterogeneity in a patchy ecosystem. Proc Natl Acad Sci U S A 2007;104: 2761–6.

Ricklefs RE. A comprehensive framework for global patterns in biodiversity. Ecol Lett 2004;7:1–15.

Rillig MC, Mummey DL. Mycorrhizas and soil structure. New Phytol 2006;171:41–53.

Rousk J, Baath E, Brookes PC, et al. Soil bacterial and fungal communities across a pH gradient in an arable soil. ISME J 2010;4:1340–51.

Stephan A, Meyer AH, Schmid B. Plant diversity affects culturable soil bacteria in experimental grassland communities. J Ecol 2000;88:988–98.

Talbot JM, Bruns TD, Taylor JW, et al. Endemism and functional convergence across the North American soil mycobiome. Proc Natl Acad Sci U S A 2014;111:6341–6.

Tedersoo L, Bahram M, Toots M, et al. Towards global patterns in the diversity and community structure of ectomycorrhizal fungi. Mol Ecol 2012;21:4160–70.

Wang C, Wang X, Liu D, et al. Aridity threshold in controlling ecosystem nitrogen cycling in arid and semi-arid grasslands. Nat Commun 2014;5:4799.

Wardle DA. Communities and Ecosystems: Linking the Aboveground and Belowground Components. Princeton: Princeton University Press, 2002.

Wardle DA, Bardgett RD, Klironomos JN, et al. Ecological linkages between aboveground and belowground biota. Science 2004;304:1629–33.

Ward DM, Ferris MJ, Nold SC, et al. A natural view of microbial biodiversity within hot spring cyanobacterial mat communities. Microbiol Mol Biol Rev 1998;62:1353–70.

Whitaker RJ, Grogan DW, Taylor JW. Geographic barriers isolate endemic populations of hyperthermophilic archaea. Science 2003;301:976–8.

Yang Y, Fang J, Ma W, et al. Large-scale pattern of biomass partitioning across China’s grasslands. Global Ecol Biogeogr 2010;19:268–77.

Yao M, Rui J, Li J, et al. Rate-specific responses of prokaryotic diversity and structure to nitrogen deposition in the Leymus chinensis steppe. Soil Biol Biochem 2014;79:81–90.

Zhou J, Bruns MA, Tiedje JM. DNA recovery from soils of diverse composition. Appl Environ Microbiol 1996;62:316–22.

Zhou J, Xia B, Treves D, et al. Spatial and resource factors influencing high microbial diversity in soil. Appl Environ Microbiol 2002;68:326–34.