Soil Biology & Biochemistry 86 (2015) 126–134

Contents lists available at ScienceDirect

Soil Biology & Biochemistry

journal homepage: www.elsevier.com/locate/soilbio

Multi-scale variability analysis reveals the importance of spatial distance in shaping Arctic soil microbial functional communities

Yu Shi [a, e], Paul Grogan [b], Huaibo Sun [a], Jinbo Xiong [c], Yunfeng Yang [d], Jizhong Zhou [d, f, g], Haiyan Chu [a, *]

[a] State Key Laboratory of Soil and Sustainable Agriculture, Institute of Soil Science, Chinese Academy of Sciences, East Beijing Road 71, Nanjing 210008, China
[b] Department of Biology, Queen's University, Kingston, ON K7L 3N6, Canada
[c] Faculty of Marine Sciences, Ningbo University, Ningbo 315211, China
[d] State Key Laboratory of Environment Simulation and Pollution Control, School of Environment, Tsinghua University, Beijing 100084, China
[e] University of Chinese Academy of Sciences, Beijing 100049, China
[f] Institute for Environmental Genomics and Department of Microbiology and Plant Biology, University of Oklahoma, Norman, OK 73019, USA
[g] Earth Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA

ARTICLE INFO

Article history:
Received 22 December 2014
Received in revised form
30 March 2015
Accepted 31 March 2015
Available online 11 April 2015

Keywords:
GeoChip
Arctic
PCNM
Biogeography
Functional gene

ABSTRACT

Understanding biological diversity and distribution patterns at multiple spatial scales is a central issue in ecology. Here, we investigated the biogeographical patterns of microbial functional genes in 24 heath soils from across the Arctic using GeoChip-based metagenomics and principal coordinates of neighbour matrices (PCNM)-based analysis. Functional gene richness varied considerably among sites, while the proportions of each major functional gene category were evenly distributed. Functional gene composition varied significantly at most medium to large spatial scales, and the PCNM analyses indicated that 14–20% of the variation in total and major functional gene categories could be attributed primarily to relatively large-scale spatial effects that were consistent with broad-scale variation in soil pH and total nitrogen. The combination of variance partitioning and multi-scales analysis indicated that spatial distance effects accounted for 12% of the total variation in functional gene composition, whereas environmental factors accounted for only 3%. This small but significant influence of spatial variation in determining functional gene distributions contrasts sharply with typical microbial phylotype/species-based biogeographical patterns (including these same Arctic soil samples), which are primarily determined by contemporary environmental heterogeneities. Therefore, our results suggest that historical contingencies such as disturbance events, physical heterogeneities, community interactions or dispersal barriers that occurred in the past, have some significant influence on soil functional gene distribution patterns.

1. Introduction

Understanding the diversity and distribution patterns of populations and communities at multiple spatial scales is a central issue in ecology (Levin, 1992; Borcard and Legendre, 2002; Tuomisto et al., 2003; Zhou et al., 2008). For soil microbes, it is well documented that population and community distributions in natural environments are spatially patterned (Martiny et al., 2006; Ramette and Tiedje, 2007; Hanson et al., 2012). The classic microbiological tenet "Everything is everywhere, but the environment selects" (Baas Becking, 1934) proposes that dispersal is ubiquitous and contemporary environmental factors are the primary mechanism determining distributions of microbial communities. Many recent studies support this hypothesis by demonstrating significant correlations across multiple sites between microbial community structure and environmental variables over large spatial scales (Fierer and Jackson, 2006; Lauber et al., 2009; Chu et al., 2010; Griffiths et al., 2011). Meanwhile, historical contingencies (factors that were important in the past but that are not currently influential such as disturbance events, physical heterogeneities, community interactions, dispersal barriers etc.) have

* Corresponding author. Tel.: +86 2586881356; fax: +86 02586881000.
E-mail address: hychu@issas.ac.cn (H. Chu).

http://dx.doi.org/10.1016/j.soilbio.2015.03.028
0038-0717/

also been suggested as important determinants of microbial distributions (Martiny et al., 2006). Several field studies support this latter hypothesis (Cho and Tiedje, 2000; Whitaker et al., 2003), and since the impacts of historical contingencies on microbial communities are likely correlated with spatial distance (Ramette and Tiedje, 2007), it is now believed that microbial species/phylotype distributions are shaped not just by local environmental heterogeneities but also to at least some minor extent by geographic distance (Ge et al., 2008; Griffiths et al., 2011) or dispersal limitation (Martiny et al., 2011).

A wide range of studies have investigated spatial patterning in microbial communities at scales from centimeters to meters (Franklin and Mills, 2003; Philippot et al., 2009), and at the landscape scale (Yergeau et al., 2009; Enwall et al., 2010; Bru et al., 2011; Shi et al., 2015). Spatial autocorrelation has been commonly observed, and can occur at sampling distances up to 739 km (Bru et al., 2011). Techniques to analyze spatial patterns across multiple scales have recently been used in microbial community studies (Martiny et al., 2011; Franklin and Mills, 2003). For example, Ramette and Tiedje (2007) showed that variation in species abundances and community composition within the Burkholderia bacterial group in an agricultural ecosystem was greatest at small scales (between individual plant roots) rather than larger spatial scales (across a field). However, because functional redundancy among microbial phylotypes seems to be very frequent (Lozupone et al., 2012), these taxonomy-based biogeographical studies may be very limited in terms of providing insights as to how spatial heterogeneity in microbial community structure influences biogeochemical processes within and among ecosystems. Nevertheless, some recent studies have demonstrated that for some specific biogeochemical processes, there can be strong spatial linkages between abundances of the functional groups responsible for those processes and activity rates. For example, denitrifier functional gene abundances were highly spatially correlated with a N_2O production in soils at sampling distances up to 5 m at three different Arctic sites (Banerjee and Siciliano, 2012a), and across a pasture at distances from 6 to 16 m (Philippot et al., 2009). Furthermore, the abundances of ammonia-oxidizing genes within archaeal and bacterial communities were spatially correlated with aerobic ammonia oxidation potential rates at distances up to 4 m in the same Arctic soils referred to above (Banerjee and Siciliano, 2012b). However, the question of whether the relationship between phylogenetic structure and functioning in terrestrial soil microbial communities applies across multiple biogeochemical functions and across landscape and larger spatial scales has not yet been investigated.

Understanding spatial distributions of soil functional genes at large scales is a high priority in terms of predicting terrestrial ecosystem responses to global land use and climate changes (He et al., 2010a; Zhou et al., 2012; Chan et al., 2013; Feng et al., 2014). In a previous study, we documented the pattern and influences of environmental heterogeneity and geographic distance on bacterial community structure in heath tundra soils that were sampled from across a large part of the Arctic (Chu et al., 2010). Our overall goal in this current study was to investigate the patterns and controls on functional gene distributions in those same soils to directly compare taxonomically-based and trait-based biogeographical patterns and the relative influences of environmental and spatial factors. Characterizing the 'pure' influence of spatial scale on the biogeography of microbial communities is complex because other categories of potential explanatory variables such as those associated with environmental heterogeneity can also vary across space. Most previous studies of this issue have investigated the influences of local environment and spatial distance separately, without accounting for potential covariation (Fierer and Jackson, 2006; Lauber et al., 2009; Chu et al., 2010; Feng et al., 2014). The Principal Coordinates of Neighbor Matrices (PCNM) analytical approach was specifically developed to model community structures across a wide range of scales and to characterize the relative influences of the explanatory factors both separately, and in combination, at multiple different spatial scales (Borcard and Legendre, 2002).

To characterize soil microbial functional genes in these soils, we utilized the GeoChip 4.0 array which contains probes for ~152,000 biogeochemically important functional genes (Hazen et al., 2010; He et al., 2010b; Yang et al., 2013; Tu et al., 2014). Using the GEOCHIP and PCNM analytical approaches on the same triplicate samples from the 24 heath tundra sites across the Arctic that we had used in our study of microbial community phylogenetic composition (Chu et al., 2010), we specifically address the following three questions:

I) Are distributions of microbial functional genes in Arctic soils more spatially structured at large or small scales?
II) Can the spatial structure of soil microbial functional genes be categorized into discrete spatial scales that are associated with heterogeneities in environmental variables?
III) What is the relative importance of spatial distance as compared to local environment in determining the distribution patterns of microbial functional genes in heath soils across the Arctic?

2. Methods and materials

2.1. Soil sampling

Surface soil organic samples were collected from 24 heath tundra sites (at least 190 km apart from each other) across the Canadian, Alaskan and European Arctic in the summer of 2007 and 2008 as described by Chu et al. (2010). At each site, soil samples were collected close to the top of exposed ridges at three similar locations (20–100 m apart) from below dry heath vegetation in which at least one of the following plant species was common: *Empetrum* spp., *Cassiope* spp. or *Dryas* spp. The soil type immediately underlying this vegetation was typically an Orthic Dystric Static Cryosol (Paré, 2011. Canadian Soil Classification System - http://sis.agr.gc.ca/cansis/taxa/cssc3/CY/SC/index.html). Samples of the top surface dark brown/black organic soil were cut out with a serrated knife (that was wiped off with a clean tissue before sampling from subsequent locations) from a ~12 cm × 12 cm area to 2–5 cm depth and placed in a separate plastic bag. All the samples were immediately shipped to Kingston, Canada where they were stored at −20 °C until processing. Details of the soil sampling and of each site's geographical, ecological, and biogeochemical characteristics (including methods for the latter) have been described previously (Chu et al., 2010). Unfortunately, climatic data were not available for many of the sites because they did not have local weather stations.

2.2. GeoChip analysis

DNA was extracted from 5 g fresh weight of soil from each sampling location (n = 72 in total) using a freeze-grinding mechanical lysis method as described previously (Zhou et al., 1996). We used GeoChip 4.0 to analyze DNA samples as described previously (Lu et al., 2012; Yang et al., 2013; Tu et al., 2014). Briefly, DNA was labeled with the fluorescent dye Cy-5 using a random priming method and then purified with the QIA quick purification kit (Qiagen, Valencia, CA, USA) according to the manufacturer's

instructions. After measuring dye incorporation on a NanoDrop ND-1000 spectrophotometer (Nano-Drop Technologies Inc., Wilmington, DE, USA), DNA was dried in a Speed Vac (ThermoSavant, Milford, MA, USA) at 45 °C for 45 min. Subsequently, labeled DNA was resuspended in 50 μl hybridization solution containing 40% formamide, 3 × SSC, 10 μg of unlabeled herring sperm DNA (Promega, Madison, WI, USA), and 0.1% SDS, and hybridizations were performed with a MAUI hybridization station (BioMicro, Salt Lake City, UT, USA) according to the manufacturer's recommended method. After washing and drying, microarrays were scanned with a NimbleGen MS200 scanner (Roche, Madison, WI, USA) at 633 nm using a laser power of 100% and a photomultiplier tube (PMT) gain of 75%. Signal intensities were subsequently quantified.

GeoChip 4.0 contained 83,992 50-mer oligo probes which covered 152,414 gene variants from 401 distinct functional gene categories associated with bacteria, archaea and fungi (Lu et al., 2012; Tu et al., 2014). There are 12 main gene categories: biogeochemical cycling of carbon, nitrogen, phosphorus, and sulfur; resistance to metal and antibiotics; energy process; organic compound remediation; stress response; bacterial phage-related; virulence-related; and others (Tu et al., 2014).

2.3. Data analysis

Raw data were analyzed using a data analysis pipeline as described previously (He et al., 2010b). In brief, the following steps were performed: (i) spots flagged or with a signal to noise ratio (SNR) less than 2.0 were removed as poor-quality spots; (ii) after removing poor spots, normalized intensity of each spot was calculated by dividing the signal intensity of each spot by the total intensity of the microarray and then multiplying it with a constant value of 58,000,740, which is the average signal intensity of all of GeoChip data; (iii) intensities were transformed to natural logarithm values; and (iv) a minimum of two valid values of three biological replicates were remained from the same sites. In order to quantify geographical linkages among all the sites, we first determined the threshold spatial index above which all the sites were connected (Borcard and Legendre, 2002) using the location data (latitude and longitude). The threshold (shortest distance index which could keep all the sites link together) was chosen to be as short as possible to ensure that all points remain connected by links smaller than or equal to the truncation distance (Fig. S1). Before principal coordinates of neighbour matrices (PCNM) (Borcard and Legendre, 2002) analysis, Moran's I (Moran, 1950) was used to measure spatial correlation of univariate quantitative variables (Fig. S2). Values below E(I) indicate negative spatial correlation, and values above E(I) indicate positive correlation. E(I) is close to 0 when n (the total number of observations) is large (Borcard et al., 2011). After that, we performed a correlogram of these spatial correlation values against the distance classes (Fig. S3), degrees (longitude) were used to construct the distance class index. The number of classes was computed using Sturge's rule [number of classes $= 1 + (3.3219 \times \log 10n)$, where n is the number of elements (number of samples in this case)]. The Mantel correlogram was computed, tested and plotted by using vegan's function mantel. correlog (Fig. S3). Then we used PCNM method to decompose the total spatial variation into a discrete set of explanatory spatial variables, each of which corresponds to a specific scale. The spatial trend was removed from the GeoChip data by linear regression of the x, y coordinates (latitude, longitude) of all samples from the study sites before PCNM analysis was performed. The detrended biological variables were regressed on the PCNM variables, and the significance coefficients were tested on 999 permutations of the residuals (Legendre and Legendre, 1998; Borcard et al., 2011). A forward selection as in CANOCO (TerBraak, 1988) based on the 999 Monte Carlo permutation procedure using residuals from the reduced model by the "packfor" package (Blanchet et al., 2008; Dray et al., 2011) was used to identify those environmental parameters that explained the spatial patterns at different scales. Multivariate linear regressions were then applied to calculate the explained variance for each variable. In addition, in order to compare the correlation between the functional gene richness and bacterial OTU richness, we only calculated the bacterial functional gene richness (There are very few archaeal and fungal functional gene richness were detected by Geo Chip).

To better assess the effects of space and environmental soil parameters on functional gene distributions, we conducted a variance partitioning analysis by combining the PCNM output with a modified variation partitioning diagram (Fig. S4) derived from Borcard et al. (2011). In this analysis, the location coordinates (i.e. latitude, but not longitude because it did not significantly correlate with site gene composition), environmental soil variables, and PCNM scales were independently forward selected before variance partitioning. The variance partitioning of the spatial and environmental variations were further examined by redundancy analysis ordination (Borcard et al., 1992; Legendre and Legendre, 1998), using the "varpart" function in the "vegan" package (Oksanen et al., 2007), which computes the adjusted canonical R^2 in multiple regression (Peres-Neto et al., 2006). During the variance partitioning analysis, the GeoChip data were undetrended, because a linear trend can be considered as a source of variation like any other and is likely to act on the response as well as the explanatory variables. All of the analyses were performed by functions in R v. 3.0.1 (R Development Core Team, 2006).

2.4. Data accessibility

The GeoChip dataset was deposited in the National Center for Biotechnology Information (NCBI) Gene Expression Omnibus (http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE63927) under study number GSE63927.

3. Results

3.1. Functional gene distribution and diversity

From the 24 Arctic heath sites that were sampled in triplicate (i.e. 72 soil samples in total), we detected a total of 44,714 different genes involved in carbon, nitrogen, sulfur and phosphorus cycling, organic contaminant degradation, stress tolerance, metal tolerance, fungi function and antibiotic resistance. Although the proportions of these major functional gene categories were very similar across all sites (Fig. 1), the number of hybridized gene probes per category differed among sites (Fig. S5; Table S2), indicating that the richness of the genes associated with each of these major biological functions varied among Arctic sites. For example, the number of the detected genes that were associated with carbon degradation varied from 2000 to 2800 among sites, while the number of genes associated with nitrogen fixation varied from 300 to 400, and likewise for other functional gene categories (Fig. S5). The hierarchical clustering analysis of functional gene composition showed that the three sampling locations within each site clustered together except Bl site (Fig. S6), indicating great similarities in functional gene composition at this spatial scale (20–100 m apart). Furthermore, at larger scales (i.e. among the sites, and therefore at least 190 km apart), the similarities in gene composition among pairs of sites was high (65–88%) (Table S3). Conversely, very few detected genes (0.1–0.8%) were unique to only one site in the study (Table S3).

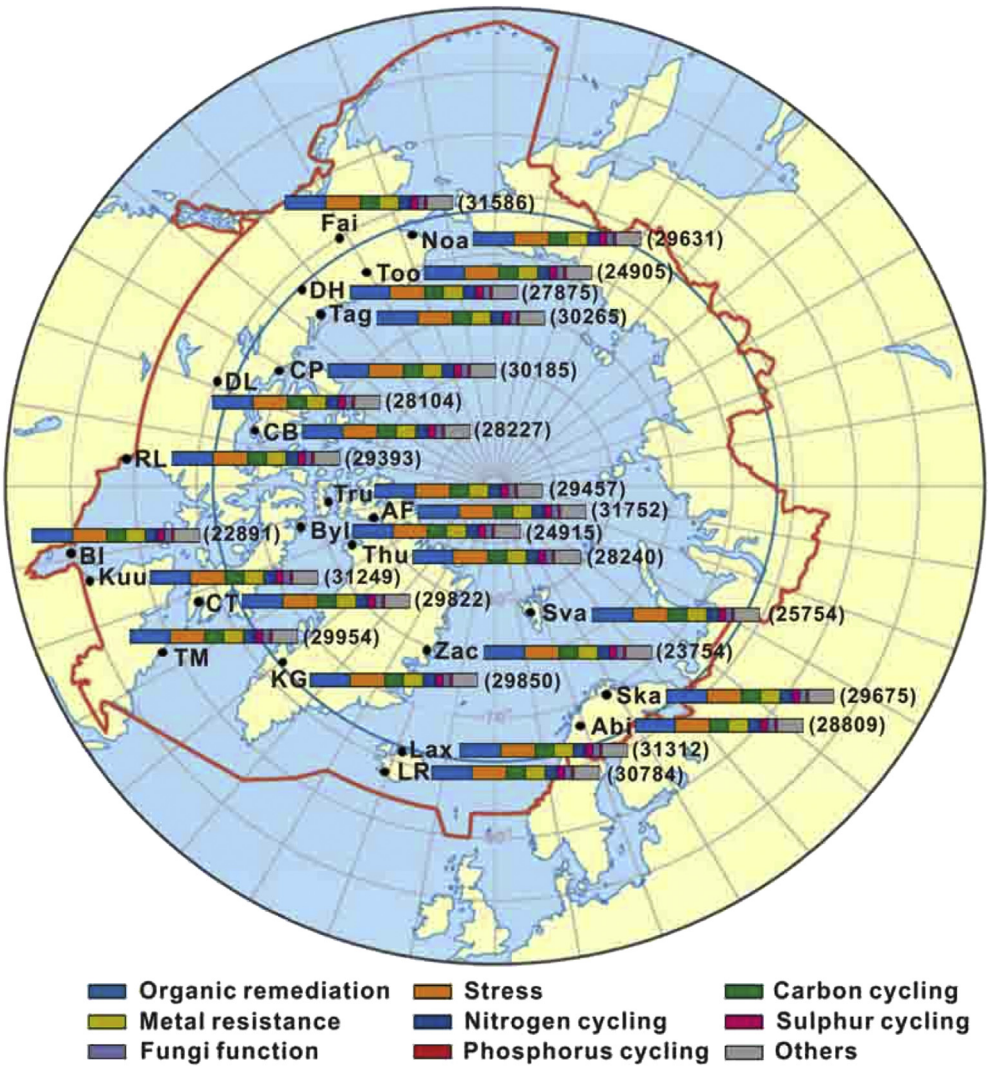

Fig. 1. Functional genes detected by GeoChip 4.0 in 24 similar dry heath tundra ecosystem sites across the Arctic. The length of the bar at each site is proportional to the total number of functional genes sampled there (mean number of functional genes across the 3 sampling locations per site—numerical value indicated at the end of each bar). The section lengths within each bar indicate the relative proportions of total genes associated with each of the different functional gene categories. 'Others' refer to the remaining, less abundant functional genes.

In this study, we found richness (i.e. patterns of richness of hybridized gene probes among sites) was not significantly correlated with soil physiochemical properties such as total carbon (TC), total nitrogen (TN), available phosphorus, pH, moisture, or with latitude (Fig. S7). Finally, site-level functional gene richness was not significantly correlated with bacterial OTU richness as determined in our previous study (Chu et al., 2010) (Fig. S8).

3.2. Multi-scale analysis of functional gene variation

We used the same approach as Borcard and Legendre (2002) and classified the different spatial scales into large, medium and small scales according to the PCNM orders. Analysis of the functional gene distributions (Moran's I—Methods in Supplementary material) indicated significant positive correlations with seven of these spatial scale categories. The remaining spatial scale categories showed non-significant or negative spatial correlations, and were not included in subsequent PCNM analyses (Fig. S9). We observed significant positive spatial correlation below 20 (distance class index), indicating that closer sites tended to have more similar communities. The significant negative correlation after 60 suggests that pairs of sites that were substantially distant from each other may have had contrasting soil conditions that resulted in different functional gene composition (Fig. S3). The PCNM analysis was conducted using the geographic coordinates of the locations within each of the 72 plots, and the derived spatial scales decreased with increasing PCNM orders. This spatial scale decomposition method yielded 14 spatial scale categories across all of our Arctic samples (Table S4), ranging from large scale (long distance – S1) to small scale (short distance – S14). Across these 14 spatial scale categories, the 7 largest scale categories (excepting S2) were significantly positively correlated with each other (Table S5), indicating of spatial autocorrelation. Using the forward selection method (Blanchet et al., 2008), we found that variation in composition of functional genes within at least some of the functional gene categories was significantly related to scales 1, 3 and 4–8 (Table 1). For the carbon cycling, nitrogen cycling, sulphur cycling, phosphorus cycling, organic remediation, stress tolerance, fungi function and other functional gene categories, about 16–21% of the variation in composition across sites could be explained by these 7 spatial

Table 1
Percentages of variation in functional genes for major biological processes that are explained at spatial scale categories ranging from large ('1') to medium ('8') – See text for details. Values for all functional genes combined (i.e. total genes detected by GeoChip 4.0) are included at the bottom of the table. Significance of the models and variables were tested by Monte Carlo permutations (**$P \leq 0.01$; *$P \leq 0.05$).

	Scale 1	Scale 3	Scale 4	Scale 5	Scale 6	Scale 7	Scale 8	Residuals
Carbon cycling	2.32**	2.05*	2.12**	2.14*	2.11**	5.42**	3.29**	80.56
Nitrogen cycling	2.25*	1.94*	2.16**	2.22*	1.94.	5.56**	3.12**	80.82
Sulphur cycling	2.21**	1.94*	1.99**	2.13*	2.02*	5.71**	3.31**	80.7
Fungi function	2.55**	2.14**	2.17*	2.02*	2.22*	5.58**	3.64**	79.68
Phosphorus cycling	2.25**	2.05*	2.13**	1.91*	2.03**	5.48**	3.07**	81.08
Organic Remediation	2.25*	1.96*	1.95*	2.10**	1.99*	5.57**	3.24**	80.94
Stress	2.25**	1.92*	2.07**	2.12*	2.07**	5.49**	3.38**	80.71
cold shock (stress)	2.15*		2.21*			7.46**	4.07**	84.11
heat shock (stress)	2.39**	1.88*	1.94*	2.21*	1.96*	5.4**	3.16**	81.05
Soil borne pathogen	2.21*	1.96*	1.99.	1.92*	2.16*	5.77**	3.31**	80.69
Soil_benefit	2.03.	1.94*	2.22*	2.14*	2.07**	5.16**	3.08**	81.36
Bacterial phage	2.49**	2.05*	2.73**	2.45**	2.49**	5.67**	3.27**	78.84
Virulence	2.32**	2.01*	2.23*	2.12*	1.91*	5.36**	3.32**	80.74
Energy process	2.14*	2.19*	2.17**	2.07**	2.06*	4.74**	3.27**	81.35
Antibiotic resistance	2.24**	2.03*	2.03*	2.05*	2.01**	5.24**	3.19**	81.21
Metal resistance	2.21**	1.91*	2.08*	2.03*	1.93*	5.55**	3.18**	81.12
All functional genes			2.21*			7.46**	4.07**	86.26

Note: Soil benefit contains the functional genes which are helpful to the soil.

scales, while the cold shock gene variation (16%) was explained by just 4 of the significant spatial scales (Table 1). Moreover, functional gene subcategories such as *nifH* (nitrogen fixation), *nirS* (denitrification), *pmoA* (methane oxidation) and *mcrA* (methane production) also varied significantly at different spatial scales (Table S6). Although each of the larger scale categories (i.e., 1, 3, 4, 5 and 6) was not as correlated with the variation within the functional gene categories as were the two finer scales (7 and 8), overall, the combination of these large scales together had more explanatory power (Table 1). These analyses strongly suggest that large-scale processes are more important than medium or small-scale processes in determining spatial variation in functional gene groups within soil communities across the Arctic.

In order to understand how these spatial patterns could be explained by soil characteristics, partial standard regression coefficients were calculated for the soil physicochemical variables (Table 2). In this analysis, the specific contribution of each environmental variable to each scale was determined, and only those that were statistically significant were included in the models. The soil physicochemical variables in all explained 1.4%–30.3% of the spatial variation with the relationships being strongest for the scale 5 category and weakest for scale 12 (Table 2). The greatest explanatory power (highest regression coefficients) was for variation in soil total nitrogen, DOC, DON and pH at the broad scales (1, 2, 5 and 6). As the variations in functional gene groups were also generated mainly by broad scale processes (Table 1), our results suggest that soil pH and total N were the main contributors to the spatial patterning of functional genes in Arctic heath soils (Table S7 and S8).

3.3. Functional gene variation partitioning according to the soil variables and spatial scale categories

The significant PCNM scales were split into broad and fine-scale groups, and the latitudinal coordinate of sampling site was represented by a linear trend. Together, the whole set (environmental component, linear trend and broad scale) of spatial and environmental soil variables explained 14.5% of variation in the undetrended (see method in Borcard et al., 2011) GeoChip data (i.e. for all functional genes combined). The spatial variables (broad scale plus linear trend) alone explained 12.2% of variation in functional genes, and the environmental soil variables explained 3.1% (among them 0.6% was spatially structured) (Fig. 2). The covariations (i.e. overlaps) between environmental variables and spatial distance/latitude were very small (Fig. 2; Fig. S9), indicating that the effects of spatial scales and environmental variables were largely independent of each other. However, a large part of the variation (85.5%) across all functional genes together was not explained by any of the model parameters, or their interactions. The combination of PCNM and variance partitioning analysis was also performed on each of the categories of major functional genes associated with carbon cycling, nitrogen cycling and phosphorus cycling, and about 18% of the variations in these gene distributions were explained by the spatial and environmental soil variables together (Figs. 2 and 3,

Table 2
Partial regression coefficients relating soil physicochemical variables to a range of spatial scale categories from large ('1') to small ('13') – See text for details. 'Total % explained' refers to the percentages of the total variance explained by significant environmental variables in each of the scale categories. PCNM scales 3, 4 and 7 are not shown because they were not significantly correlated with any of the environmental variables. DOC: dissolved organic carbon; DON: dissolved organic nitrogen; P: Phosphorus.

Soil variables	PCNM scale								
	Scale 1	Scale 2	Scale 5	Scale 6	Scale 9	Scale 10	Scale 11	Scale 12	Scale 13
C						0.15			−0.01
N	−70.19	−29.11	−12.44	4.07					
CN ratio					0.10				
DOC	0.02					0.00			0.00
DON	1.13						−0.01	0.00	
NH_4^+						−0.05			
P							0.04	−0.01	
pH		−17.17	14.86	6.05		1.53			
Moisture					0.02				
Total % explained	17.30	16.29	30.32	5.11	23.29	27.95	15.03	1.43	17.20

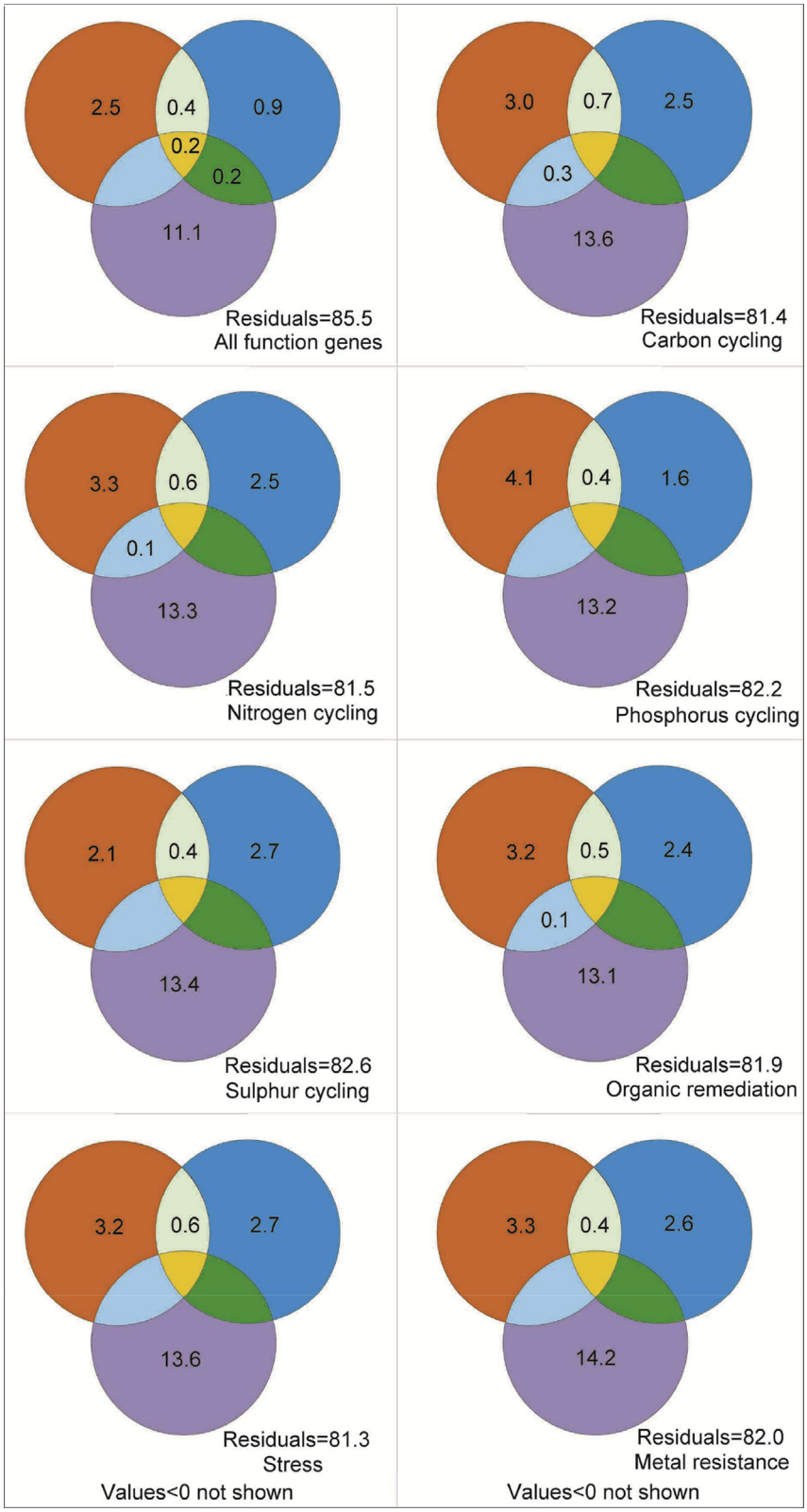

Fig. 2. Variance partitioning (as % of total variance) of the undetrended functional gene distributions across the 24 Arctic sites into a pure environmental component (upper left-hand orange circle), a pure trend (latitude) scale (upper right-hand blue circle) and a pure broad spatial scale derived from the PCNM spatial components and their covariation (lower purple circle). The empty sectors in the plots indicate relationships with small negative R^2_{adj} values that were not presented. Small spatial-scale variables are absent from the variance partitioning diagrams because they were not included in the analysis since they had negative or no correlation to functional genes.

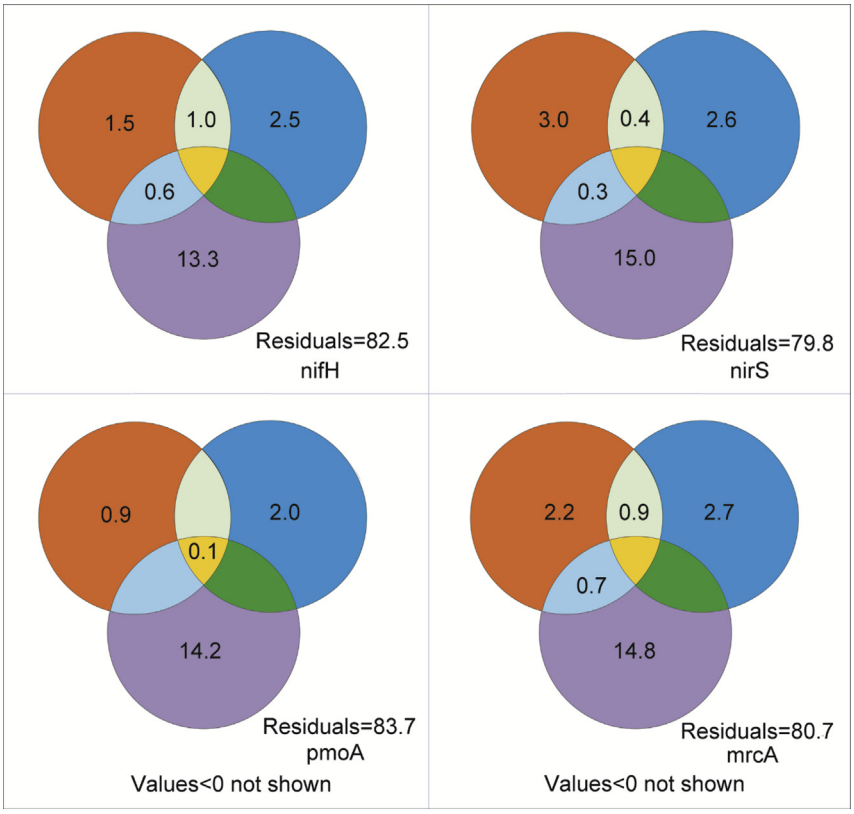

Fig. 3. Variance partitioning (as % of total variance) of the undetrended distributions of functional genes associated with sub-categories of nitrogen cycling (nifH-nitrogen fixation; nirS-nitrate denitrification; pmoA-methane oxidation; mrcA-methane production) across the 24 Arctic sites into a pure environmental component (upper left-hand orange circle), a pure trend scale (latitude) and a pure broad spatial scale derived from the PCNM spatial components and their covariation (lower purple circle). Empty sectors in the plots occur where the relationships had small negative R^2_{adj} values.

Fig. S9). Furthermore, in our separate analyses of distinct categories of functional genes, spatial variables explained more variation (14.8%–16.8%) than the environmental soil variables (2.5%–4.0%). Similar to the analysis of all functional genes combined, the analysis of individual functional gene categories also indicated that a large part of the variation (>80%) was not explained by either the environmental, latitudinal or spatial components of the PCNM, or their interactions.

4. Discussion

Our investigation of the spatial structure of functional genes in heath tundra soils from across the Arctic indicated that the proportions of many important categories of functional genes were similarly distributed (Fig. 1). By contrast, we found substantial variation in bacterial phylotype community structure among these same soil samples in our previous study (Chu et al., 2010). Burke et al. (2011)'s study of bacterial community structure on a green macro algal host species also found high microbial phylogenetic variability but similar microbial functional gene distributions. In addition, the functional gene distributions in our samples were structured according to multiple different spatial scales that correspond with heterogeneity in specific environmental variables, suggesting that scale-dependent ecological processes are an important determinant of soil functional gene biogeography. Spatial factors were consistently more important than latitude or environmental factors in explaining the compositional variation in our analysis of all functional genes combined, and in our analyses of each separate category of functional genes (Figs. 2 and 3). Banerjee and Siciliano (2012b) also found that spatial scale was an important control on ammonia oxidation communities in Arctic soils, but over a much smaller distance range (up to 4 m) than our study. Despite the statistical significance of spatial distance in determining microbial gene (and phylotype) distributions, it generally accounts for a fairly small proportion of the total variation within such data. For example, at least 80% of the variation in functional gene composition across our Arctic sites was not explained by either the spatial, latitudinal or environmental factors (or their interactions). Likewise, Ramette and Tiedje (2007)'s reported that 73% of the genetic variation in *Burkholderia ambifaria* bacterial populations could not be explained by environmental and spatial variables across a patchy agricultural field. Similarly, only about one fifth of the total variation in microbial community composition in a forest soil could be explained by environmental heterogeneity (20.7%) or geographic distance (18.3%), or their interaction (5.8%) (Zhou et al., 2008). High levels of unexplained variation have been attributed to unmeasured environmental factors and inadequate sampling methodologies (Zhou et al., 2008), and former probably also include factors that were spatially structured and possibly auto-correlated with latitude. The recently developed neutral theory in community ecology (Hubbell, 2001) suggests that some of the unexplained variation could be due to stochastically structured gene distributions that occur when phylotype immigration and extinction rates have a stronger influence than natural selection in determining community composition and therefore functional gene biogeographies. By contrast, classic niche theory predicts that natural selection would tend to favour functional genes that were appropriate to the environment, and to eliminate the remainder. These two processes are not mutually exclusive, and both seem to be operating simultaneously in a wide range of ecological

communities (Gravel et al., 2011). To fully characterize the controls on trait- and taxonomically-based spatial patterning, future studies are needed that specifically address the unmeasured variability, and that evaluate and contrast the potential influences of niche and neutral-based processes in determining the distributions of soil functional genes.

A central goal in ecology is to develop a comprehensive multi-scale understanding of spatial patterns in community structure and functional gene distribution (Zhou et al., 2008), and the development of the PCNM method (Borcard and Legendre, 2002) is an important step forward in this process. In this study, both the analysis of all functional genes together, and the analyses of individual gene categories such as carbon cycling, nitrogen cycling and phosphorus cycling indicated significant variation at spatial scales ranging from S9 through to S1, with the larger scales explaining more of the total variation. By contrast, some studies found that the microbial community were shaped primarily by small-scale spatial factors. For example, using geostatistical variogram analysis, Franklin and Mills (2003) found that significant spatial autocorrelation in taxonomy-based microbial community structure at scales ranging from 30 cm to more than 6 m. Through nested sampling scheme at scales ranging from 2 to 2000 m, King et al. (2010) found that bacterial community composition has significant spatial autocorrelation up to a distance of 240 m. Several studies have specifically focused on spatial patterning of individual functional genes related to denitrification. For example, Philippot et al. (2009) observed a non-random distribution pattern of the size of the denitrifier community estimated by quantification of the denitrification genes (*narG*, *napA*, *nirS*, *nirK* and *nosZ*) with a spatial dependence (6–16 m). Banerjee and Siciliano (2012a, b) found that denitrifier functional gene (*nirK*, *nirS* and *nosZ*) distributions were highly spatially correlated within a scale of 5 m. In our study, we found that *nirS* denitrifier functional genes were shaped primarily by very large spatial scale factors. Together, these studies complement each other because they were done at non-overlapping sample distances. They suggest that microbial gene distributions (at least for the denitrifier function) may be determined by very small scale spatial factors (acting at distances up to ~16 m) and also by very large scale spatial factors.

Spatial structuring in microbial communities may result from several environmental variables acting at different spatial scales (Legendre and Legendre, 1998). For example, in Arctic soils, Banerjee and Siciliano (2012b) found that gene abundances were spatially structured within 4 m, but the biochemical processes were structured within 40 m. In our study, soil pH was significantly correlated with large- (scale 2, 5 and 6) and small-scale (scale 10) spatial patterning in functional gene distributions, which was consistent with the results from other ecosystems showing that soil pH was a key determinant of taxonomic distributions in bacteria (Lauber et al., 2009). Similarly, Feng et al. (2014) found that soil pH was a key factor determining the community structure of the Anoxygenic purple phototrophic bacteria functional group in the same soil samples as studied here. On the basis of all our studies with these heath soils, we conclude that pH in particular plays a very important role in determining not just the distributions and community structure of Arctic soil microorganisms (Chu et al., 2010; Feng et al., 2014), but also the distributions of functional genes.

Many studies have investigated the role of environmental heterogeneity and spatial distance on the microbial phylotype distributions (Fierer and Jackson, 2006; Martiny et al., 2006; Lauber et al., 2009; Chu et al., 2010; Griffiths et al., 2011). Furthermore, the relative contribution of contemporary factors and historical contingency in shaping functional gene distributions has also been investigated, but only in experimental manipulation contexts or over small geographical areas (Zhou et al., 2008, 2012; He et al., 2010a; Chan et al., 2013). For example, Yang et al. (2013) found that elevation, vegetation, and soil variables contributed 7.7%, 25.5%, and 22.7%, respectively, to the variation in functional gene distributions. In our study here, we found that spatial distance played an important role in shaping soil functional gene distributions across the Arctic region. By contrast, bacterial community structure in those same Arctic soils was unaffected by spatial distance (Chu et al., 2010). Here, we found no correlation between soil bacterial phylotype richness and total microbial functional gene richness across the 24 heath sites, suggesting that the structure of microbial functional gene distributions is fundamentally different to that of bacterial taxonomic distributions perhaps because of distinct archaeal and fungal gene contributions. In any event, these results support the conclusion that the functional significance of microbial community assembly and structure in natural environments relates more to spatial patterns in functional genes than to spatial patterns in species/phylotypes (Burke et al., 2011).

5. Conclusion

Functional gene distributions in Arctic heath soils varied most at large spatial scales, and local heterogeneity in environmental variables had relatively little impact, together suggesting that historical contingencies are a major driver of trait-based biogeography. These findings differ markedly from previous reports showing that species/phylotypes distributions are strongly influenced by local environmental effects (Fierer and Jackson, 2006; Chu et al., 2010; Griffiths et al., 2011). We conclude that for Arctic soil microbes at least, patterns of functional gene distributions might be different from phylogenetic community distributions.

Acknowledgments

We sincerely thank the many research colleagues across the Arctic who collected soil samples on our behalf. We also thank Linda Cameron and several undergraduate students for help with soil processing and lab analyses. This work was supported by the Strategic Priority Research Program (XDB15010101) of the Chinese Academy of Sciences, the National Program on Key Basic Research Project (2014CB954002), and the National Natural Science Foundation of China (41371254) to H. Chu, by NSERC as part of the International Polar Year Project: Climate Change Impacts on Canadian Arctic Tundra (P. Grogan), and by an Ontario Early Researcher's Award (P. Grogan). J. Zhou's effort was partially supported by the US National Science Foundation (EF-1065844), by the Office of the Vice President for Research at the University of Oklahoma, and by the Collaborative Innovation Center for Regional Environmental Quality. The authors declare no conflicts of interest.

Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.soilbio.2015.03.028.

References

Baas Becking, L., 1934. Geobiologie of Inleiding Tot DeMilieukunde. Van Stockum and Zoon, The Hague, the Netherlands.
Banerjee, S., Siciliano, S.D., 2012a. Spatially tripartite interactions of denitrifiers in arctic ecosystems: activities, functional groups and soil resources. Environmental Microbiology 14, 2601–2613.
Banerjee, S., Siciliano, S.D., 2012b. Factors driving potential ammonia oxidation in Canadian Arctic ecosystems: does spatial scale matter? Applied and Environmental Microbiology 78, 346–353.
Blanchet, F.G., Legendre, P., Borcard, D., 2008. Forward selection of explanatory variables. Ecology 89, 2623–2632.

Borcard, D., Gillet, F.O., Legendre, P., 2011. Numerical Ecology with R. Springer, New York.
Borcard, D., Legendre, P., 2002. All-scale spatial analysis of ecological data by means of principal coordinates of neighbour matrices. Ecological Modelling 153, 51–68.
Borcard, D., Legendre, P., Drapeau, P., 1992. Partialling out the spatial component of ecological variation. Ecology 73, 1045–1055.
Bru, D., Ramette, A., Saby, N.P.A., Dequiedt, S., Ranjard, L., Jolivet, C., Arrouays, D., Philippot, L., 2011. Determinants of the distribution of nitrogen-cycling microbial communities at the landscape scale. The ISME Journal 5, 532–542.
Burke, C., Steinberg, P., Rusch, D., Kjelleberg, S., Thomas, T., 2011. Bacterial community assembly based on functional genes rather than species. Proceedings of the National Academy of Sciences, USA 108, 14288–14293.
Chan, Y., Van Nostrand, J.D., Zhou, J.Z., 2013. Pointing SB and Farrell RL. Functional ecology of an Antarctic dry Valley. Proceedings of the National Academy of Sciences, USA 110, 8990–8995.
Cho, J.C., Tiedje, J.M., 2000. Biogeography and degree of endemicity of fluorescent Pseudomonas strains in soil. Applied and Environmental Microbiology 66, 5448–5456.
Chu, H.Y., Fierer, N., Lauber, C.L., Caporaso, J.G., Knight, R., Grogan, P., 2010. Soil bacterial diversity in the Arctic is not fundamentally different from that found in other biomes. Environmental Microbiology 12, 2998–3006.
Dray, S., Legendre, P., Blanchet, G., 2011. Pack for: Forward Selection with Permutation. R package version 0.0–8/r100.
Enwall, K., Throback, I.N., Stenberg, M., Soderstrom, M., Hallin, S., 2010. Soil resources influence spatial patterns of denitrifying communities at scales compatible with land management. Applied and Environmental Microbiology 76, 2243–2250.
Fierer, N., Jackson, R.B., 2006. The diversity and biogeography of soil bacterial communities. Proceedings of the National Academy of Sciences, USA 103, 626–631.
Feng, Y.Z., Grogan, P., Caporaso, J.G., Zhang, H.Y., Lin, X.G., Knight, R., Chu, H.Y., 2014. pH is a good predictor of the distribution of anoxygenic purple phototrophic bacteria in Arctic soils. Soil Biology & Biochemistry 74, 193–200.
Franklin, R.B., Mills, A.L., 2003. Multi-scale variation in spatial heterogeneity for microbial community structure in an eastern Virginia agricultural field. Fems Microbiology Ecology 44, 335–346.
Ge, Y., He, J.Z., Zhu, Y.G., Zhang, J.B., Xu, Z.H., Zhang, L.M., Zheng, Y.M., 2008. Differences in soil bacterial diversity: driven by contemporary disturbances or historical contingencies? The ISME Journal 2, 254–264.
Gravel, D., Guichard, F., Hochberg, M.E., 2011. Species coexistence in a variable world. Ecology Letters 14, 828–839.
Griffiths, R.I., Thomson, B.C., James, P., Bell, T., Bailey, M., Whiteley, A.S., 2011. The bacterial biogeography of British soils. Environmental Microbiology 13, 1642–1654.
Hanson, C.A., Fuhrman, J.A., Horner-Devine, M.C., Martiny, J.B.H., 2012. Beyond biogeographic patterns: processes shaping the microbial landscape. Nature Reviews Microbiology 10, 497–506.
Hazen, T.C., Dubinsky, E.A., DeSantis, T.Z., Andersen, G.L., Piceno, Y.M., Singh, N., Jansson, J.K., Probst, A., Borglin, S.E., Fortney, J.L., Stringfellow, W.T., Bill, M., Conrad, M.E., Tom, L.M., Chavarria, K.L., Alusi, T.R., Lamendella, R., Joyner, D.C., Spier, C., Baelum, J., Auer, M., Zemla, M.L., Chakraborty, R., Sonnenthal, E.L., D'haeseleer, P., Holman, H.Y.N., Osman, S., Lu, Z.M., Van Nostrand, J.D., Deng, Y., Zhou, J.Z., Mason, O.U., 2010. Deep-sea oil plume enriches indigenous oil-degrading bacteria. Science 330, 204–208.
He, Z.L., Xu, M.Y., Deng, Y., Kang, S.H., Kellogg, L., Wu, L.Y., Van Nostrand, J.D., Hobbie, S.E., Reich, P.B., Zhou, J.Z., 2010a. Metagenomic analysis reveals a marked divergence in the structure of belowground microbial communities at elevated CO2. Ecology Letters 13, 564–575.
He, Z.L., Deng, Y., Van Nostrand, J.D., Tu, Q.C., Xu, M.Y., Hemme, C.L., Li, X.Y., Wu, L.Y., Gentry, T.J., Yin, Y.F., Liebich, J., Hazen, T.C., Zhou, J.Z., 2010b. GeoChip 3.0 as a high-throughput tool for analyzing microbial community composition, structure and functional activity. The ISME Journal 4, 1167–1179.
Hubbell, S.P., 2001. The Unified Neutral Theory of Biodiversity and Biogeography. Princeton University Press, Princeton.
King, A.J., Freeman, K.R., McCormick, K.F., Lynch, R.C., Lozupone, C., Knight, R., Schmidt, S.K., 2010. Biogeography and habitat modelling of high-alpine bacteria. Nature Communications 1, 1–6.
Lauber, C.L., Hamady, M., Knight, R., Fierer, N., 2009. Pyrosequencing-based assessment of soil pH as a predictor of soil bacterial community structure at the continental Scale. Applied and Environmental Microbiology 75, 5111–5120.
Legendre, P., Legendre, L., 1998. Numerical Ecology. Elsevier, Amsterdam; New York.
Levin, S.A., 1992. The problem of pattern and scale in ecology. Ecology 73, 1943–1967.
Lozupone, C.A., Stombaugh, J.I., Gordon, J.I., Jansson, J.K., Knight, R., 2012. Diversity, stability and resilience of the human gut microbiota. Nature 489, 220–230.
Lu, Z.M., Deng, Y., Van Nostrand, J.D., He, Z.L., Voordeckers, J., Zhou, A.F., Lee, Y.J., Mason, O.U., Dubinsky, E.A., Chavarria, K.L., Tom, L.M., Fortney, J.L., Lamendella, R., Jansson, J.K., D'haeseleer, P., Hazen, T.C., Zhou, J.Z., 2012. Microbial gene functions enriched in the deepwater horizon deep-sea oil plume. The ISME Journal 6, 451–460.
Martiny, J.B.H., Bohannan, B.J.M., Brown, J.N., Colwell, R.K., Fuhrman, J.A., Green, J.L., Horner-Devine, M.C., Kane, M., Krumins, J.A., Kuske, C.R., Morin, P.J., Naeem, S., Ovreas, L., Reysenbach, A.L., Smith, V.H., Staley, J.T., 2006. Microbial biogeography: putting microorganisms on the map. Nature Reviews Microbiology 4, 102–112.
Martiny, J.B.H., Eisen, J.A., Penn, K., Allison, S.D., Horner-Devine, M.C., 2011. Drivers of bacterial beta-diversity depend on spatial scale. Proceedings of the National Academy of Sciences, USA 108, 7850–7854.
Moran, P.A., 1950. Notes on continuous stochastic phenomena. Biometrika 37, 17–23.
Oksanen, J., Blanchet, F.G., Kindt, R., Legendre, P., Minchin, P.R., O'Hara, R.B., Simpson, G.L., Solymos, P., Stevens, M.H.H., Wagner, H., 2007. The Vegan Package. Community Ecology Package. Version: 1.8–5.
Paré, M.C. 2011. (PhD thesis). Organic matter quality in cryosols: effect on soil nitrogen dynamics and greenhouse gas emissions. In Partial Fulfillment of the Requirements For the Degree of Doctor of Philosophy. In the Department of Soil Science University of Saskatchewan Saskatoon By Maxime Charles Paré © Copyright Maxime Charles Paré, July 2011.
Peres-Neto, P.R., Legendre, P., Dray, S., Borcard, D., 2006. Variation partitioning of species data matrices: estimation and comparison of fractions. Ecology 87, 2614–2625.
Philippot, L., Cuhel, J., Saby, N.P.A., Cheneby, D., Chronakova, A., Bru, D., Arrouays, D., Martin-Laurent, F., Simek, M., 2009. Mapping field-scale spatial patterns of size and activity of the denitrifier community. Environmental Microbiology 11, 1518–1526.
Ramette, A., Tiedje, J.M., 2007. Multiscale responses of microbial life to spatial distance and environmental heterogeneity in a patchy ecosystem. Proceedings of the National Academy of Sciences, USA 104, 2761–2766.
R Development Core Team, 2006. R, a Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.
Shi, Y., Xiang, X.J., Shen, C.C., Chu, H.Y., Neufeld, J.D., Walker, V.K., Grogan, P., 2015. Vegetation-associated impacts on Arctic tundra bacterial and micro-eukaryotic communities. Applied and Environmental Microbiology 81, 482–501.
TerBraak, C., 1988. CANOCO-a FORTRAN Programme for Canonical Community Ordination by [partial][detrended][canonical] Correspondence Analysis, Principal Components Analysis and Redundancy Analysis (Version 2.1) (Agricultural Mathematics Group).
Tu, Q.C., Yu, H., He, Z.L., Deng, Y., Wu, L.Y., Van Nostrand, J.D., Zhou, A.F., Voordeckers, J., Lee, Y.J., Qin, Y.J., Hemme, C.L., Shi, Z., Xue, K., Yuan, T., Wang, A.J., Zhou, J.Z., 2014. GeoChip 4: a functional gene arrays-based high throughput environmental technology for microbial community analysis. Molecular Ecology Resource 14, 914–928.
Tuomisto, H., Ruokolainen, K., Yli-Halla, M., 2003. Dispersal, environment, and floristic variation of western Amazonian forests. Science 299, 241–244.
Whitaker, R.J, Grogan, D.W., Taylor, J.W., 2003. Geographic barriers isolate endemic populations of hyperthermophilic archaea. Science 301, 976–978.
Yang, Y.F., Wu, L.W., Lin, Q.Y., Yuan, M.T., Xu, D.P., Yu, H., Hu, Y.G., Duan, J.C., Li, X.Z., He, Z.L., Xue, K., Van Nostrand, J., Wang, S.P., Zhou, J.Z., 2013. Responses of the functional structure of soil microbial community to livestock grazing in the Tibetan alpine grassland. Global Change Biology 19, 637–648.
Yergeau, E., Schoondermark-Stolk, S.A., Brodie, E.L., Dejean, S., DeSantis, T.Z., Goncalves, O., Piceno, Y.M., Andersen, G.L., Kowalchuk, G.A., 2009. Environmental microarray analyses of Antarctic soil microbial communities. The ISME Journal 3, 340–351.
Zhou, J.Z., Bruns, M.A., Tiedje, J.M., 1996. DNA recovery from soils of diverse composition. Applied and Environmental Microbiology 62, 316–322.
Zhou, J.Z., 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.
Zhou, J.Z., Xue, K., Xie, J.P., Deng, Y., Wu, L.Y., Cheng, X.H., Fei, S.F., Deng, S.P., He, Z.L., Van Nostrand, J.D., Luo, Y.Q., 2012. Microbial mediation of carbon-cycle feedbacks to climate warming. Nature Climate Change 2, 106–110.