1 Bacterial communities in oil contaminated soils: biogeography and co-occurrence
2 patterns

3
4

5 Shuo Jiao, Zhenshan Liu, Yanbing Lin, Yang Jun, Weimin Chen*, Gehong Wei*

6 State Key Laboratory of Crop Stress Biology in Arid Areas, College of Life Sciences,
7 Northwest A&F University, Yangling Shaanxi 712100, P. R. China^a

8
9

10 *Correspondence: Gehong Wei and Weimin Chen

11 E-mail: weigehong@nwsuaf.edu.cn, chenwm029@nwsuaf.edu.cn

12 Tel: +86 29 87091175

13 Fax: +86 29 87091175

14

Abstract

To improve the knowledge about the biogeographic patterns of bacteria in soils contaminated with crude oil, we studied the effects of local geochemical properties and geographic distance on bacterial community structure in oil contaminated soil in five oil refineries (46 to 360 km apart). The microbial structure was significantly affected by soil environmental factors such as pH, total petroleum hydrocarbons (TPH), total nitrogen, and cadmium level. Microbial alpha-diversity was positively correlated with pH but negatively correlated with TPH. Among sampling sites, the community dissimilarities increased with spatial distance. Variation of bacterial community was mostly attributed to simultaneous effects of spatial distance and environmental factors, and purely spatial distance contributed more to the variation. Microbial generalist OTUs were broadly distributed and dominant in contaminated soils. Their populations were in low proportion (15.75%), but they had high relative abundance (65.05%), and some were associated with TPH-degradation. Network analysis indicated that microbial communities had non-random co-occurrence patterns. Keystone taxa were *Rubrivivax*, *Nitrospira*, *Methylothermobacter*, *Methyloversatilis* and *Acidaminobacter*. Microbial taxa from the same module had strong ecological linkages and were involved in biological electron-transfer, C and N-cycles, and organic contaminant degradation. Our results indicate that the same microbial groups with TPH-degradation ability can be assembled from indigenous microorganisms in separate regions through long term exposure to contamination.

Keywords: biogeography; oil contamination; co-occurrence patterns; biodiversity

1. Introduction

Microbial biogeography, the study of microorganism distribution patterns across space and time, has received increasing attention since the development of molecular tools (Fenchel, 2003; Martiny et al., 2006; Hanson et al., 2012). One viewpoint of microbial biogeography is that “everything is everywhere; the environment selects” (Baas-Becking, 1934). Environmental factors were initially thought to be the dominant factors affecting variation among microbial communities. Many studies have demonstrated that environmental factors such as soil pH, nutrients, soil texture, and climatic conditions can significantly affect microbial community distribution (Fierer and Jackson, 2006; Bissett et al., 2010). However, some studies indicate that the non-random distributions of microorganisms are not only affected by environmental factors but can also be significantly correlated to geographic distance (Xiong et al., 2014; Liang et al., 2015; Wang et al., 2015). Spatial distance can limit the dispersal of microorganisms and contribute to accelerated speciation (Diniz-Filho and Telles, 2000).

Crude oil contamination is a global problem that causes great damage to natural environments and animal health. Contamination frequently occurs during oil exploration, transportation, storage, usage, or accidental spills (Liang et al., 2015). Removal of crude oil from soil is a difficult and long-term process (Peterson et al., 2003). Microbial remediation is an effective option for environment cleanup due to its low cost and limited impact on other aspects of the environment (Liebeg and Cutright, 1999). Soil is a huge reservoir for diverse microbial populations. Many

microorganisms adapted to harmful environments are naturally abundant and able to degrade organic pollutants. Large numbers of functional genes contributing to degradation of alkane fractions have been detected in oil polluted soil (Fierer and Jackson, 2006). Microorganisms therefore can play important roles in bioremediation of oil damaged ecosystems. Determining the structure of microbial communities in oil contaminated soils could provide valuable clues for ecosystem restoration and environmental management.

Microorganisms within a specific ecological niche form complex interaction webs (Faust and Raes, 2012). Co-occurrence network analysis may provide insight into the structure of complex microbial communities and the interactions between microorganisms (Barberán et al., 2012). Using the high-throughput sequencing technology, broad and deep surveys of microbial communities have been made in diverse environments including marine water (Beman et al., 2011), soil (Barberán et al., 2012) and activated sludge (Ju et al., 2014). Distinct environments are affected by abiotic factors and predict specific habitat limits which can support coexistence of species within these communities (Stegen et al., 2012). However, the co-occurrence patterns of species in oil contaminated soils are still unclear.

In this study, we conducted a metagenomic analysis of bacterial communities from oil contaminated soils via high-throughput sequencing of the 16S rRNA gene. The results were analyzed for correlations with environmental factors and spatial distances. Samples were obtained from petroleum contaminated surface soils around five oil refineries in the Shaanxi Province of China. Our main goal was to investigate

the biogeographic pattern and biodiversity of soil bacterial communities in fields with a long history of oil contamination, especially to study (i) the bacterial diversity in oil-contaminated fields; (ii) the relationships among the bacterial community and soil environmental factors; (iii) the effects of environmental factors and geographic distance on bacterial community; (iv) the co-occurrence patterns among bacterial taxa responding to oil contamination.

2. Materials and Methods

2.1. Sample collection

A total of 38 soil samples were collected around five oil refineries in Shaanxi Province of China in October 2013, including Yulin (JB, 9 samples), Yanchang (YC, 4 samples), Yanchuan (YP, 9 samples), Luochuan (JKH, 7 samples) and Xianyang (XY, 9 samples). Each soil sample was a mixture of five soil cores (0-15 cm) from an area of 500×500 m² around an oil refinery. Soil samples were transported to the laboratory in sterile plastic bags on dry ice. A subset of the soil was air-dried for analysis of edaphic properties. Aliquots of soil samples were stored at -80°C for microbial analyses. Total petroleum hydrocarbon (TPH) concentrations in the soil samples were measured using the Ultrasonic-Soxhlet extraction gravimetric method (Huesemann, 1995). Physical and chemical parameters of the soils, including pH, total nitrogen (TN), available nitrogen (AN), total phosphorus (TP), available phosphorus (AP), available potassium (AK) and the heavy metals cadmium (Cd), chromium (Cr) and lead (Pb), were measured using standard soil testing procedures (Bao, 2000).

2.2. DNA extraction and purification

Community DNA was extracted from 0.5 g of soil samples using the MP FastDNA®SPIN Kit for soil (MP Biochemicals, Solon, OH, USA) according to the

manufacturer protocol. The V4-V5 hypervariable regions of the 16S rRNA gene was amplified using primers 515F (5'-GTG CCA GCM GCC GCG GTA A-3') and 907R (5'-CCG TCA ATT CCT TTG AGT TT-3') with the forward primer modified to contain a unique 6 nt barcode at the 5' end. All PCR reactions were performed with 30 μ L system with 15 μ L of Phusion® High-Fidelity PCR Master Mix (New England Biolabs), 0.2 μ M of forward and reverse primers and about 10 ng template DNA under the following thermal cycling conditions: initial denaturation at 98 °C for 1 min, followed by 30 cycles of denaturation at 98 °C for 10 s, annealing at 50 °C for 30 s, and extension at 72 °C for 60 s, and an extension step at 72 °C for 5 min after cycling was complete. All samples were amplified in triplicate, and no-template controls were included in all steps of the process. Triplicate PCR amplicons were pooled together and then mixed with the same volume of 1× loading buffer (containing SYB green). They were detected by electrophoresis in a 2% (w/v) agarose gel. PCR products with bright band between 400-450bp were mixed in equal density ratios and purified with GeneJET Gel Extraction Kit (Thermo Scientific). The purified PCR amplicons were sequenced using the Illumina Miseq (300-bp paired-end reads) platform at Novogene Bioinformatics Technology Co., Ltd. (Beijing, China).

2.3. Sequence analysis of the 16S rRNA amplicons

Paired-end reads were merged using FLASH (V1.2.7, <http://ccb.jhu.edu/software/FLASH/>), and quality filtering of reads was performed according to the literature (Caporaso et al., 2011). The acquired sequences were chimera detected and removed using USEARCH software based on the UCHIME algorithm (Edgar et al., 2011). The sequences were then assigned to each sample with unique barcodes. Sequence analysis was performed by the UPARSE software package using the UPARSE-OTU and UPARSE-OTUref algorithms. Sequences with $\geq 97\%$

similarity were assigned to the same OTUs. The low abundance OTUs were eliminated from the OTU table if they did not have a total of at least 2 counts across all the samples in the experiment. The representative sequences for each OTU were assigned to taxonomic groups using the RDP classifier (Caporaso et al., 2011).

2.4. Data analyses

Alpha diversity and beta diversity based on Bray-Curtis distance were calculated using 36,920 reads per sample (minimum number of sequences required to normalize the differences in sequencing depth) using QIIME (<http://qiime.org/index.html>), with multiple indices (observed species, Pielou's evenness, Shannon-Wiener index, and phylogenetic diversity) and the Bray-Curtis distance between samples.

Maps of sampling sites were made using GenGIS II (Parks et al., 2013). The relationships between geographical coordinates and sample ordination on taxonomic (Bray-Curtis) or environmental dissimilarity (Euclid distance based on environmental variables) were determined. The Mantel test was used to test the correlation between taxonomic and environmental dissimilarity.

Principal component analysis (PCA) was calculated using soil environment parameters as variables and alpha diversity indices with significance < 0.05 were fitted as factors onto the ordination. Constrained analysis of principal coordinates (CAP) based on Bray-Curtis distance was performed to investigate the relationship between bacterial community composition and environmental factors. A BEST was used to rank the importance of environmental features in influencing beta-diversity community comparisons, with feature significance confirmed by permutational MANOVA. Partial least squares regression (PLSR) analysis was used to model associations between normalized mean values for significant environmental conditions (X variables) and taxonomic features (Y variables).

The rate of distance-decay of the bacterial communities was calculated as the slope of a linear least-squares regression on the relationship between the geographic distance and the bacterial similarity based on 1- dissimilarity of the Bray-Curtis metric. Spatial variables were derived from geographic coordinates using both linear trend and the principle coordinates of neighbor matrices (PCNM) procedure (Griffith and Peres-Neto, 2006), that could minimize the autocorrelation between spatial distance and environmental variables, to capture all the detectable spatial scales in the dataset. Based on CAP model, forward selection procedures were used to select subsets of environmental, trend, and PCNM variables, respectively. The forward selection was stopped if the significance level alpha ($P > 0.05$) was reached or if no improvement of selection criterion (R^2) was seen when adding additional variables. Variation partitioning was performed with adjusted R^2 to determine the proportion of variation in bacterial communities explained purely by environmental or spatial factors and shared variation of environmental and spatial factors.

Network was used to explore co-occurrence patterns of bacterial taxa. The genera with relative abundances above 0.05% were selected. A Spearman's correlation between two genera was considered statistically robust if the Spearman's correlation coefficient (ρ) was > 0.6 and the P-value was < 0.01 (Barberán et al., 2012). All the robust correlations identified from pairwise comparison of the genera abundance form a correlation network where each node represents one genus, and each edge stands for a strong and significant correlation between the nodes. To describe the topology of the resulting networks, a set of measures (number of nodes and edges, average path length, network diameter, average degree, graph density, clustering coefficient and modularity) was calculated using igraph (Csardi and Nepusz, 2006) packages in R environment and networks were visualized using the interactive platform Gephi

(Newman, 2003, 2006; Bastian et al., 2009). Meanwhile, 10,000 Erdős–Rényi random networks were generated to compare with the topology of real network, with each edge having the same probability of being assigned to any node (Erdős and Rényi, 1960). Genera with highest betweenness centrality value, which indicates the relevance of a node as capable of holding together communicating nodes, were considered as keystone species (González et al., 2010; Vick-Majors et al., 2014).

All statistics analyses were performed in R environment (<http://www.r-project.org>) unless otherwise indicated.

3. Results

3.1. Geochemical characteristics of oil-contaminated soils

The major geographical and physicochemical characteristics of the oil-contaminated soils are summarized in Table 1. The soil types for all the samples were sandy loam. TPH in the soils ranged from 0.582 to 60.376 g kg⁻¹ of dry soil, and it was significantly higher for JB samples than that for samples in the other sites. All soils were alkalescent and pHs ranged from 7.49 to 9.20. Heavy metals Cd, Cr and Pb were also detected. The Pb level was 138.34 mg kg⁻¹ in XY which was 10× higher than that in the other sites. The nutrient factors TP, AP and TN from JB and XY were higher than those in JKH, YP and YC. The concentration of AN was not significantly different among the sites.

3.2. Distribution of taxa and phylotypes

A data set of 1,613,390 quality sequences was produced from the soil samples. The mean number of sequences per sample (n=38) was 42,458 (max = 48,177, min=36,926, SD=2720). Most sequences (99.73%) could be classified to the phyla of bacteria. The total OTU number was 9604 defined by 97% sequence similarity. Six phyla including Proteobacteria, Bacteroidetes, Actinobacteria, Chloroflexi, Firmicutes

and Acidobacteria were predominant (relative abundance > 5%), and accounted for 84.37% of the total sequences. The phylum Proteobacteria, with a relative abundance of 47.37%, was represented by the following classes: Betaproteobacteria (18.82%), Gammaproteobacteria (13.24%), Deltaproteobacteria (7.50%), and Alphaproteobacteria (7.27%). In addition, Gemmatimonadetes, Spirochaetes, Planctomycetes, Verrucomicrobia, Chlorobi and Nitrospirae were detected in most contaminated soils at low relative abundances.

3.3. Bacterial communities associated with environment factors

Based on soil factors, the PCA analysis showed that soil samples from JKH, YC and YP clustered together except for one sample from JKH. Soil samples from JB and XY distributed in the first and fourth quadrant, indicating that these soils samples hold large environmental heterogeneity (Fig. 1A). When alpha diversities including observed species (Richness), Pielou's evenness (Evenness), Shannon–Wiener index (Shannon) and phylogenetic diversity (PD) were fitted as factors onto the Ordination, alpha diversities were positively correlated with pH, but negatively correlated with TPH, TN and AP. The significant correlations were confirmed by the linear regression relationships between pH/TPH and Richness, and spearman's correlation test as shown in Fig. 1B/C and Supplementary Table S1.

The associations between environment and bacterial community structure from different sites are shown in Fig. 2. The geographical location and average phylum-level taxonomic compositions of sampling sites were exhibited visually. The clusters based on soil environment factors (left diagram) and bacterial community structure (right diagram) were similar, which showed significant corresponding relationships (the mantel test, $r=0.4195$, $P<0.001$). Constrained analysis of principal coordinates (CAP) based on Bray-Curtis distance (Fig. 3A) demonstrated that

environmental variables pH, TPH, TN and Cd significantly affected the bacterial community (permutation testing). The first axis was positively correlated with TN, Cd and TPH but negatively correlated with pH. The second axis was positively correlated with Cd, but negatively correlated with TPH. Among the variables, pH had the strongest influence on the difference among the bacterial communities of all samples (Table 3). The best variables rank correlation test (BEST) confirmed pH, TN, Cd and TPH as the most important environmental variables to explain the greatest degree of β -diversity dissimilarity in bacterial communities for multiple sample subsets.

The partial least squares regression (PLSR) uncovered the specific microbial populations that may be influenced by the environmental factors, that was used to model covariance between environmental factors and bacterial taxa at multiple taxonomic levels (Fig. 3B). PLSR projections were initially made between all phylum- to genus-level taxa with a relative abundance above 0.5% and used to select taxa explaining >40% of the variance in the first two components. Cd was strongly associated with Desulfovibrionales, *Thauera*, *Desulfomicrobium* and Rhodocyclaceae; TN was positively associated with Xanthomonadaceae, Caulobacteraceae and *Arenimonas*; TPH was weakly associated with Gammaproteobacteria, Caulobacteraceae and Xanthomonadaceae; and pH was positively associated with Acidobacteria-6, Solirubrobacterales and Ellin6529. Spearman correlation coefficients generated between these variables and factors revealed significant ($P < 0.05$) correlations for many of these observations (Supplementary Table S2).

3.4. Spatial distribution of bacterial community

The importance of spatial distance in bacterial community variability estimated using beta-diversity based on Bray-Curtis distance revealed the distance-decay relationships (DDR; Fig. 4) of community similarity vs. geographic distance for each

pairwise set of samples (shown in Supplementary Table S3). A significant negative distance–decay of similarity relationship (slope= -0.034, $P<0.001$) was found in linear regression between the two sites and the corresponding spatial distance. The Mantel test also showed that spatial distance significantly affects community structure ($r=0.342$ and $P<0.001$).

To further investigate the similarity of microbial distribution in oil contaminated soils, the broadly distributed microbial taxa were described. A total of 573 OTUs presenting in all the 38 samples were identified and defined as generalists, that accounted for a mean of 15.75% (range: 13.20%-19.91%) of the total OTUs in each sample, but contributed a substantial fraction of the reads (mean = 65.05%, range between 45.34% and 82.00%). Generalist OTUs were composed of a diverse range of phylogenetic groups such as Proteobacteria, Actinobacteria, Bacteroidetes and Acidobacteria. The Heatmap plotted the relative abundances of top 30 most abundant generalist OTUs (Supplementary Fig. S1) and showed that generalist OTUs contained some common genera such as *Thiobacillus*, *Thauera*, *Hydrogenophaga*, *Arenimonas*, *Lysobacter* and *Geobacter*.

3.5. Variation partitioning of bacterial community

To quantify the contribution of spatial distance (as a proxy for dispersal barriers) and soil properties to bacterial community variation, variance partitioning analysis was performed based on CAP model and illustrated using a modified variation partitioning diagram (Fig. 5). Through a forward selection procedure, four environmental variables (pH, TPH, Cd and TN), two PCNM variables, and two linear trend variables were selected as explanatory variables. All the fractions of variation partitioning were significant in an ANOVA permutation test ($P<0.01$). The complete set of spatial (linear trend and PCNM) and soil environmental variables explained

43.09% of the variation in the bacterial communities. Simultaneous effects of soil environmental variables and spatial factors (PCNM plus linear trend) dominantly accounted for 29.04% of the variation, which suggested that the effects of spatial factors and environmental variables were dependent on each other to a great extent. The pure effect of soil environmental variables explained 1.81% of the variation. The PCNM and linear trend variables purely explained 3.74% and 8.09% of the variation, respectively. Moreover, the PCNM and linear trend variables simultaneously explained 0.41% of variation, indicating the purely spatial factors contributed 12.24% of the variation in shaping bacterial communities. Therefore, both soil environmental factors and spatial distance were principal factors affecting the structure of bacterial communities.

3.6. Co-occurrence network analysis

The soil microbial network consisted of 95 nodes (genera) and 446 edges (with a mean of 9.389 edges per node) (Fig. 6, Table 2), for which significant topological properties were calculated to describe the complex pattern of inter-relationships among nodes, and to compare this real network with an identically sized Erdős-Rényi random network (Table 2). The average path length (APL) was 4.230 edges with a diameter of 7 edges. The clustering coefficient (CC) was 0.657 and the modularity index (MD) was 0.551 (values >0.4 suggest that the network has a modular structure). The structural properties of the real-world network were greater than the Erdős-Rényi random networks of APL_r (2.257), CC_r (0.0995) and MD_r (0.241). Overall, the microbial network was comprised of highly connected genera and formed a “small world” topology (that is, real-world networks are more significantly clustered than

random graphs).

The nodes in the network were assigned to eleven bacteria phyla (Supplementary Fig. S2). Among these, five phyla (Proteobacteria, Actinobacteria, Bacteroidetes, Firmicutes, and Chloroflexi) were widely distributed, accounting for 90% of all nodes. When the distribution of nodes was modularized, all nodes were grouped into four major modules. The nodes from module I mostly belonged to Proteobacteria and Chloroflexi; nodes from module II were mostly Proteobacteria and Actinobacteria; nodes from module III were mostly Bacteroidetes; nodes for module IV all belonged to Proteobacteria. Based on betweenness centrality scores, the top five genera identified as keystone taxa were *Rubrivivax*, *Nitrospira*, *Methylothermobacter*, *Methylothermobacter* and *Acidaminobacter*.

4. Discussion

Determining the biogeographic distribution of microorganisms and its response to oil contamination is an important issue in microbial ecology. However, characterization and quantification of bacterial community biogeography is challenging due to variability in environmental factors which results in complex bacterial communities. In this study, we analyzed the effect of environmental factors and spatial distance on the structure of bacterial communities in five separate oil contaminated fields spaced 46 to 360 km apart.

4.1. Important abiotic factors shaping bacterial community structure

Microbial communities can respond rapidly to the environmental changes caused by crude oil contamination. Contaminants may result in either an increase (Kaplan

and Kitts, 2004) or a decrease (Vinas et al., 2005; Bordenave et al., 2007) in microbial diversity. In the present study, the significantly and negatively correlations between TPH and alpha diversity (including richness, evenness, phylotype and phylogenetic diversity) (Figs. 1 and 3) was likely caused by the selective pressure of TPH contaminants. Selection pressure might occur from (i) the toxic effect of TPH and biodegradation products (Bordenave et al., 2007; Stroud et al., 2007), and (ii) electron acceptor limitations for microbial community due to TPH mineralization (Zhou and Crawford, 1995). Abundant nutrients would be expected to facilitate the growth of some microorganisms, which might partially explain our results that nutrient factors TN, AN, and AP were negatively correlated with bacterial diversity. The effects of nutrients on bacterial diversity may have resulted from imbalanced C: N: P ratios, stemming from an influx of carbon-rich TPH, leading to competition within the microbial community for nutrients and ultimately reducing nutrient availability (Zytner et al., 2006; Margesin et al., 2007). The pH was positively associated with microbial diversity within a limited range of 1.7 pH units (pH 7.49 to 9.20). However, some studies in uncontaminated environments showed different results. For example, in lake sediments, pH values from 6.88 to 10.37 were negatively correlated with phylotype richness and phylogenetic diversity (Xiong et al., 2012). In acid soils, pH was positively correlated with bacterial alpha-diversity (Griffiths et al., 2011). The phylogenetic diversity of bacterial communities was highest with near-neutral pH (Lauber et al., 2009). This discrepancy may be related to the unique and complex environments present in oil-contaminated fields.

This was confirmed by a Similar to previous study (Liang et al., 2011), our study also showed that oil contaminants exerted significant impact on soil pH, TPH, TN and Cd which in turn changed the microbial functional communities. Caulobacteraceae and Xanthomonadaceae were significantly correlated with an increase of TPH (Fig. 3B). These groups are associated with the bioremediation of oil contaminated soils (Lladó et al., 2012). The pH was the main environment variable controlling the bacterial community structure and this was consistent with results from a previous study (Xiong et al., 2012). This implies that pH is a robust and useful predictor of bacterial community structure in oil contaminated soils. However, the influence of other environmental factors such as TN and Cd can also impact the distribution of selected bacterial phyla. For example, the relative abundance of *Desulfomicrobium*, a group of bacteria tolerant to heavy metals, was positively correlated with Cd (Azabou et al., 2007). These results indicate a deterministic process (environment filtering) acting to shape the bacterial community in oil-contaminated soils.

4.2. Microbial spatial structure in oil-contaminated area

An important question in biogeographic research is the influence of spatial distance on microbial assemblages (Martiny et al., 2011). We found a significant negative correlation between bacterial community variation and spatial distance, indicating a spatial structure of bacterial communities. This result may be influenced by two main processes: dispersal limitation (Lear et al., 2014) and heterogeneous habitat (Ranjard et al., 2013). Variation partitioning analysis quantified the contribution of spatial distance and environmental variables to bacterial community

variation. A high proportion (30.85%) of the variation of bacterial communities can be explained by a combination of environmental variables. This indicated that environmental selection was a major process in microbial assemble. This could be comparable with a few previous studies (Van der Gucht et al., 2007; Xiong et al., 2014). The unexplained variation may be due to unmeasured additional factors and ecologically neutral processes of diversification (Allison and Martiny, 2008).

Most variation partitioning of the environmental variables were dependent on spatial distance. Bacterial beta-diversity is dominantly explained by the shared variation of soil environmental variables and spatial factors (29.04%). In addition, purely spatial variables (the linear trend and PCNM variables) contributed more to the variation than purely environmental variables. This suggests that dispersal limitation (i.e. isolation) plays an important role in structuring bacterial communities in the oil-contaminated fields. Other researchers have shown similar microbial dispersal limitations between distant environments (Dumbrell et al., 2010; Caruso et al., 2011; Wang et al., 2015). Environmental and spatial variability can significantly affect the structure of bacterial communities (Xiong et al., 2014; Liang et al., 2015; Wang et al., 2015). Wang et al (2015) noted that bacterial community structure is mainly affected by environmental variables and dispersal-related processes should not override species distribution in the highly connected marine areas. Pure spatial variability should not be overestimated because of the limitation in understanding of all related environmental traits. In oil contaminated soils, unmeasured environmental variables like soil texture and elements of oil contaminants could contribute to the unexplained

variation, which may also be geographically correlated. It is practically impossible to completely quantify all selective variables, including biotic variables, such as the abundance or composition of other organisms present. The spatial pattern of microbial communities is probably jointly driven by deterministic and stochastic processes (Liang et al., 2015). Stochasticity was dominant in non-contaminated soils and long-term oil contamination could increase the influence of deterministic processes (Liang et al., 2015). Oil contaminants can significantly affect soil microbes due to toxicity and selective pressures (Labud et al., 2007) and increasing competition for nutrients resulting from imbalance of C:N:P (Margesin et al., 2007). Overall, our results suggest that the spatial bacterial communities in oil-contaminated areas are jointly determined by species sorting and dispersal limitations. Thus, biogeography studies of bacterial communities should consider the influence of dispersal-related processes.

4.3. Distribution and co-occurrence patterns of microorganisms in oil-contaminated fields

Generalist microorganisms were broadly distributed in oil-contaminated soils. Their OTUs were a low proportion (15.75%) of the total but they had a high relative abundance in the total reads (65.05%). This suggests that generalists are a dominant part of the bacterial community. The soil samples in long-term oil contaminated fields analyzed in our study had significant environmental variability: different habitats and a broad spatial scale. This suggests that the generalists are adapted to a variety of conditions and are candidate indicator microbial groups for oil contamination. Some

microbes within this group are associated with toxic pollutant degradation. For example, the abundant genera *Hydrogenophaga* and *Thauera* were dominant among microorganisms degrading PAHs (Jiao et al., 2016). In addition, the co-occurrence of these abundant generalists in soils separated by great distances (46 to 360 km) and environmental heterogeneity also suggest the action of stochastic processes (dispersal) in the bacterial community assembly.

Among the communities of bacteria studied here, the phyla Proteobacteria, Bacteroidetes and Actinobacteria were dominant, and this relationship has been found in other contaminated soils (Kasai et al., 2005; Milton et al., 2010). Network analysis also showed that most of the nodes belonged to these three phyla. The identification of *Rubrivivax*, *Nitrospira*, *Methylothera*, *Methyloversatilis* and *Acidaminobacter* as top five of the betweenness centrality scores in this study implies that they may play critical roles in maintaining the structure and function of ecological communities. Discerning the modules maintaining the connectivity in network, betweenness centrality has highlighted its usefulness in defining the keystone species in the system (Vick-Majors et al., 2014; Banerjee et al., 2015). *Rubrivivax* is a photosynthetic bacterium that can participate in the process of high redox potential electron transfer and aromatic hydrocarbon degradation (Ramana et al., 2006). *Nitrospira* contained nitrite-oxidizing bacteria which is part of the biogeochemical nitrogen cycle (Altmann et al., 2003). *Methylothera* and *Methyloversatilis* are methylotrophic groups that are obligate methyl utilizers (Kalyuzhnaya et al., 2006a; Kalyuzhnaya et al., 2006b). *Methyloversatilis* has also been reported to contribute to Benazolin-Ethyl

biodegradation in activated sludge (Cai et al., 2011). *Acidaminobacter*, an amino acid fermentation taxon, has been detected in PAH-degrading methanogenic communities (Chang et al., 2005). Therefore, all the five keystone taxa may play important roles in ecological function processes.

Co-occurrence patterns were explored to offer insight into microbial interactions. Comparison of our network with a random network demonstrated non-random co-occurrence and connectedness of bacterial communities across the oil-contaminated fields, illustrating the role of deterministic processes in the structure of these communities. Moreover, modular structure was evident - the nodes were grouped into four major modules. The nodes in different modules perform different functions (Newman, 2006). In this case, strong ecological linkages manifesting as a cluster of robust co-occurrence correlations were observed. In module I of the network, some bacteria are involved with electron-transfer. In previous studies, electron transfer was seen in the sulfate-reducing bacteria of *Desulfomicrobium*, *Desulfovibrio* and *Sulfurimonas*, sulfur-oxidizing bacteria of *Sulfuritalea* and *Sulfuricurvum*, iron reduction bacteria of *Thiobacillus* (Fortin et al., 1996) and iron-oxidizing bacteria of *Gallionella* (Hallbeck et al., 1993). Biological electron-transfer plays an important role in microbial degradation of organic matter (Stams et al., 2006). Certain bacteria in module I are also involved with organic matter degradation, including *Syntrophus* (Jackson et al., 1999), *Syntrophobacter* (Boone and Bryant, 1980), *Comamonas* (Felföldi et al., 2010) and *Alcanivorax* (Kasai et al., 2002). The main taxa in module II may be bacteria involved in the

biogeochemical C- and N-cycles. For example, *Nitrospira* is a well-known nitrite-oxidizing bacterium. *Mycobacterium* is common NO_3^- reducers and *Planctomyces* is ammonia oxidation bacteria (Schmid et al., 2005). Other bacteria in module II included the Bradyrhizobiaceae, Hyphomicrobiaceae, Microbacteriaceae, Mycobacteriaceae, Rhodospirillaceae, Streptomycetaceae, Nocardioidaceae, and Sphingomonadaceae, that are also involved in C and N cycling (Anderson et al., 2011). Bacteria in module IV are mainly associated with organic contaminant degradation, including *Hydrogenophaga*, *Rubrivivax*, *Methylothermobacter*, and *Rhodobacter*. Thus, the complexity of community structure and functional processes and the co-occurrence patterns of bacterial communities in contaminated soil systems appear to be non-random and function-driven.

5. Conclusion

In the present study, we analyzed and quantified the effects of environmental factors and spatial distance on the structure of bacterial communities in five oil contaminated fields. Spatial turnover of microbes in oil-contaminated areas suggests that biogeography studies of oil-contaminated fields should consider dispersal-related processes. Non-random co-occurrence and ecological function-driven modular patterns occurred in the bacterial communities, providing a new perspective of microbial assembly in oil-contaminated fields. In future works, time-series or seasonal scale sampling is needed to be conducted to predict the dynamic pattern in bacterial assembly. More complete soil physiochemical information and climate conditions

should be integrated with microbial taxonomic and functional data to better
characterize the biogeographic distribution of community and functional structures.

Acknowledgments

This work was funded by the National Science Foundation of China (31270529)
and the Cheung Kong Scholars Programme.

All authors contributed intellectual input and assistance to this study and the
manuscript preparation. S.J., G.W. and W.C. developed the original framework. S.J.
and Z.L. collected the samples. S.J. did data analysis and wrote the paper with help
from W.C.

Reference

- Allison, S.D., Martiny, J.B., 2008. Resistance, resilience, and redundancy in microbial communities. *Proceedings of the National Academy of Sciences* 105, 11512-11519.
- Altmann, D., Stief, P., Amann, R., De Beer, D., Schramm, A., 2003. In situ distribution and activity of nitrifying bacteria in freshwater sediment. *Environmental Microbiology* 5, 798-803.
- Anderson, C.R., Condon, L.M., Clough, T.J., Fiers, M., Stewart, A., Hill, R.A., Sherlock, R.R., 2011. Biochar induced soil microbial community change: implications for biogeochemical cycling of carbon, nitrogen and phosphorus. *Pedobiologia* 54, 309-320.
- Azabou, S., Mechichi, T., Patel, B.K., Sayadi, S., 2007. Isolation and characterization of a mesophilic heavy-metals-tolerant sulfate-reducing bacterium *Desulfomicrobium* sp. from an enrichment culture using phosphogypsum as a sulfate source. *Journal of Hazardous Materials* 140, 264-270.
- Baas-Becking, L.G.M., 1934. *Geobiologie; of inleiding tot de milieukunde*. WP Van Stockum & Zoon NV.
- Banerjee, S., Baah - Acheamfour, M., Carlyle, C.N., Bissett, A., Richardson, A.E., Siddique, T., Bork, E.W., Chang, S.X., 2015. Determinants of bacterial communities in Canadian agroforestry systems. *Environmental Microbiology*.
- Bao, S.D., 2000. *Soil and agricultural chemistry analysis*. Agriculture Publication, Beijing, 355-356.
- Barberán, A., Bates, S.T., Casamayor, E.O., Fierer, N., 2012. Using network analysis to explore co-occurrence patterns in soil microbial communities. *The ISME journal* 6, 343-351.
- Bastian, M., Heymann, S., Jacomy, M., 2009. Gephi: an open source software for exploring and manipulating networks. *ICWSM* 8, 361-362.
- Beman, J.M., Steele, J.A., Fuhrman, J.A., 2011. Co-occurrence patterns for abundant marine archaeal and bacterial lineages in the deep chlorophyll maximum of coastal California. *The ISME journal* 5, 1077-1085.
- Bissett, A., Richardson, A., Baker, G., Wakelin, S., Thrall, P., 2010. Life history determines biogeographical patterns of soil bacterial communities over multiple spatial scales. *Molecular Ecology* 19, 4315-4327.
- Boone, D.R., Bryant, M.P., 1980. Propionate-degrading bacterium, *Syntrophobacter wolinii* sp. nov. gen. nov., from methanogenic ecosystems. *Applied and Environmental Microbiology* 40, 626-632.
- Bordenave, S., Goñi-Urriza, M.S., Caumette, P., Duran, R., 2007. Effects of heavy fuel oil on the bacterial community structure of a pristine microbial mat. *Applied and Environmental Microbiology* 73, 6089-6097.
- Cai, T., Qian, L., Cai, S., Chen, L., 2011. Biodegradation of benazolin-ethyl by strain *Methyloversatilis* sp. cd-1 isolated from activated sludge. *Current Microbiology* 62, 570-577.
- Caporaso, J.G., Lauber, C.L., Walters, W.A., Berg-Lyons, D., Lozupone, C.A., Turnbaugh, P.J., Fierer, N., Knight, R., 2011. Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proceedings of the National Academy of Sciences* 108, 4516-4522.
- Caruso, T., Chan, Y., Lacap, D.C., Lau, M.C., McKay, C.P., Pointing, S.B., 2011. Stochastic and deterministic processes interact in the assembly of desert microbial communities on a global scale. *The ISME journal* 5, 1406-1413.
- Chang, W., Um, Y., Hoffman, B., Holoman, T.R.P., 2005. Molecular characterization of polycyclic aromatic hydrocarbon (PAH) - degrading methanogenic communities. *Biotechnology Progress* 21, 682-688.
- Csardi, G., Nepusz, T., 2006. The igraph software package for complex network research. *InterJournal*,

Complex Systems 1695, 1-9.

Diniz-Filho, J.A.F., Telles, M.P.d.C., 2000. Spatial pattern and genetic diversity estimates are linked in stochastic models of population differentiation. *Genetics and Molecular Biology* 23, 541-544.

Dumbrell, A.J., Nelson, M., Helgason, T., Dytham, C., Fitter, A.H., 2010. Relative roles of niche and neutral processes in structuring a soil microbial community. *The ISME journal* 4, 337-345.

Edgar, R.C., Haas, B.J., Clemente, J.C., Quince, C., Knight, R., 2011. UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* 27, 2194-2200.

Erdős, P., Rényi, A., 1960. On the evolution of random graphs. *Publ. Math. Inst. Hung. Acad. Sci* 5, 17-61.

Faust, K., Raes, J., 2012. Microbial interactions: from networks to models. *Nature Reviews Microbiology* 10, 538-550.

Felföldi, T., Székely, A.J., Gorál, R., Barkács, K., Scheirich, G., András, J., Rácz, A., Márialigeti, K., 2010. Polyphasic bacterial community analysis of an aerobic activated sludge removing phenols and thiocyanate from coke plant effluent. *Bioresource Technology* 101, 3406-3414.

Fenchel, T., 2003. Biogeography for bacteria. *Science* 301, 925-926.

Fierer, N., Jackson, R.B., 2006. The diversity and biogeography of soil bacterial communities. *Proceedings of the National Academy of Sciences* 103, 626-631.

Fortin, D., Davis, B., Beveridge, T., 1996. Role of *Thiobacillus* and sulfate-reducing bacteria in iron biocycling in oxic and acidic mine tailings. *FEMS Microbiology Ecology* 21, 11-24.

González, A.M.M., Dalsgaard, B., Olesen, J.M., 2010. Centrality measures and the importance of generalist species in pollination networks. *Ecological Complexity* 7, 36-43.

Griffith, D.A., Peres-Neto, P.R., 2006. Spatial modeling in ecology: the flexibility of eigenfunction spatial analyses. *Ecology* 87, 2603-2613.

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.

Hallbeck, L., Ståhl, F., Pedersen, K., 1993. phytoeny and phenotypic characterization of the stalk-forming and iron-oxidizing bacterium *Gallionella ferruginea*. *Journal of General Microbiology* 139, 1531-1535.

Hanson, C.A., Fuhrman, J.A., Horner-Devine, M.C., Martiny, J.B., 2012. Beyond biogeographic patterns: processes shaping the microbial landscape. *Nature Reviews Microbiology* 10, 497-506.

Huesemann, M.H., 1995. Predictive model for estimating the extent of petroleum hydrocarbon biodegradation in contaminated soils. *Environmental Science and Technology* 29, 7-18.

Jackson, B.E., Bhupathiraju, V.K., Tanner, R.S., Woese, C.R., McInerney, M.J., 1999. *Syntrophus aciditrophicus* sp. nov., a new anaerobic bacterium that degrades fatty acids and benzoate in syntrophic association with hydrogen-using microorganisms. *Archives of Microbiology* 171, 107-114.

Jiao, S., Chen, W., Wang, E., Wang, J., Liu, Z., Li, Y., Wei, G., 2016. Microbial succession in response to pollutants in batch-enrichment culture. *Scientific reports* 6.

Ju, F., Xia, Y., Guo, F., Wang, Z., Zhang, T., 2014. Taxonomic relatedness shapes bacterial assembly in activated sludge of globally distributed wastewater treatment plants. *Environmental Microbiology* 16, 2421-2432.

Kalyuzhnaya, M.G., Bowerman, S., Lara, J.C., Lidstrom, M.E., Chistoserdova, L., 2006a. *Methylothermobacter mobilis* gen. nov., sp. nov., an obligately methylamine-utilizing bacterium within the family Methylophilaceae. *International Journal of Systematic and Evolutionary Microbiology* 56, 2819-2823.

Kalyuzhnaya, M.G., De Marco, P., Bowerman, S., Pacheco, C.C., Lara, J.C., Lidstrom, M.E., Chistoserdova,

580 L., 2006b. *Methyloversatilis universalis* gen. nov., sp. nov., a novel taxon within the Betaproteobacteria
581 represented by three methylotrophic isolates. International Journal of Systematic and Evolutionary
582 Microbiology 56, 2517-2522.

583 Kaplan, C.W., Kitts, C.L., 2004. Bacterial succession in a petroleum land treatment unit. Applied and
584 Environmental Microbiology 70, 1777-1786.

585 Kasai, Y., Kishira, H., Sasaki, T., Syutsubo, K., Watanabe, K., Harayama, S., 2002. Predominant growth of
586 *Alcanivorax* strains in oil - contaminated and nutrient - supplemented sea water. Environmental
587 Microbiology 4, 141-147.

588 Kasai, Y., Takahata, Y., Hoaki, T., Watanabe, K., 2005. Physiological and molecular characterization of a
589 microbial community established in unsaturated, petroleum - contaminated soil. Environmental
590 Microbiology 7, 806-818.

591 Labud, V., Garcia, C., Hernandez, T., 2007. Effect of hydrocarbon pollution on the microbial properties
592 of a sandy and a clay soil. Chemosphere 66, 1863-1871.

593 Lauber, C.L., Hamady, M., Knight, R., Fierer, N., 2009. Pyrosequencing-based assessment of soil pH as a
594 predictor of soil bacterial community structure at the continental scale. Applied and Environmental
595 Microbiology 75, 5111-5120.

596 Lear, G., Bellamy, J., Case, B.S., Lee, J.E., Buckley, H.L., 2014. Fine-scale spatial patterns in bacterial
597 community composition and function within freshwater ponds. The ISME journal 8, 1715-1726.

598 Liang, Y., Van Nostrand, J.D., Deng, Y., He, Z., Wu, L., Zhang, X., Li, G., Zhou, J., 2011. Functional gene
599 diversity of soil microbial communities from five oil-contaminated fields in China. The ISME journal 5,
600 403-413.

601 Liang, Y., Zhang, X., Zhou, J., Li, G., 2015. Long-term oil contamination increases deterministic
602 assembly processes in soil microbes. Ecological Applications 25, 1235-1243.

603 Liebeg, E.W., Cutright, T.J., 1999. The investigation of enhanced bioremediation through the addition
604 of macro and micro nutrients in a PAH contaminated soil. International biodeterioration &
605 biodegradation 44, 55-64.

606 Lladó, S., Solanas, A., de Lapuente, J., Borrás, M., Viñas, M., 2012. A diversified approach to evaluate
607 biostimulation and bioaugmentation strategies for heavy-oil-contaminated soil. Science of the Total
608 Environment 435, 262-269.

609 Margesin, R., Hämmerle, M., Tschérko, D., 2007. Microbial activity and community composition during
610 bioremediation of diesel-oil-contaminated soil: effects of hydrocarbon concentration, fertilizers, and
611 incubation time. Microbial Ecology 53, 259-269.

612 Martiny, J.B., Eisen, J.A., Penn, K., Allison, S.D., Horner-Devine, M.C., 2011. Drivers of bacterial
613 β -diversity depend on spatial scale. Proceedings of the National Academy of Sciences 108, 7850-7854.

614 Martiny, J.B.H., Bohannan, B.J., Brown, J.H., Colwell, R.K., Fuhrman, J.A., Green, J.L., Horner-Devine,
615 M.C., Kane, M., Krumins, J.A., Kuske, C.R., 2006. Microbial biogeography: putting microorganisms on
616 the map. Nature Reviews Microbiology 4, 102-112.

617 Militon, C., Boucher, D., Vachelard, C., Perchet, G., Barra, V., Troquet, J., Peyretailade, E., Peyret, P.,
618 2010. Bacterial community changes during bioremediation of aliphatic hydrocarbon-contaminated soil.
619 FEMS Microbiology Ecology 74, 669-681.

620 Newman, M.E., 2003. The structure and function of complex networks. Siam Review 45, 167-256.

621 Newman, M.E., 2006. Modularity and community structure in networks. Proceedings of the National
622 Academy of Sciences 103, 8577-8582.

623 Parks, D.H., Mankowski, T., Zangooei, S., Porter, M.S., Armanini, D.G., Baird, D.J., Langille, M.G., Beiko,

R.G., 2013. GenGIS 2: Geospatial analysis of traditional and genetic biodiversity, with new gradient algorithms and an extensible plugin framework. *PloS one* 8, e69885.

Peterson, C.H., Rice, S.D., Short, J.W., Esler, D., Bodkin, J.L., Ballachey, B.E., Irons, D.B., 2003. Long-term ecosystem response to the Exxon Valdez oil spill. *Science* 302, 2082-2086.

Ramana, C.V., Sasikala, C., Arunasri, K., Kumar, P.A., Srinivas, T., Shivaji, S., Gupta, P., Süling, J., Imhoff, J.F., 2006. *Rubrivivax benzoatilyticus* sp. nov., an aromatic, hydrocarbon-degrading purple betaproteobacterium. *International Journal of Systematic and Evolutionary Microbiology* 56, 2157-2164.

Ranjard, L., Dequiedt, S., Prévost-Bouré, N.C., Thioulouse, J., Saby, N., Lelievre, M., Maron, P., Morin, F., Bispo, A., Jolivet, C., 2013. Turnover of soil bacterial diversity driven by wide-scale environmental heterogeneity. *Nature communications* 4, 1434.

Schmid, M.C., Maas, B., Dapena, A., van de Pas-Schoonen, K., van de Vossenberg, J., Kartal, B., Van Niftrik, L., Schmidt, I., Cirpus, I., Kuenen, J.G., 2005. Biomarkers for in situ detection of anaerobic ammonium-oxidizing (anammox) bacteria. *Applied and Environmental Microbiology* 71, 1677-1684.

Stams, A.J., De Bok, F.A., Plugge, C.M., Eekert, V., Miriam, H., Dolfing, J., Schraa, G., 2006. Exocellular electron transfer in anaerobic microbial communities. *Environmental Microbiology* 8, 371-382.

Stegen, J.C., Lin, X., Konopka, A.E., Fredrickson, J.K., 2012. Stochastic and deterministic assembly processes in subsurface microbial communities. *The ISME journal* 6, 1653-1664.

Stroud, J., Paton, G., Semple, K.T., 2007. Microbe - aliphatic hydrocarbon interactions in soil: implications for biodegradation and bioremediation. *Journal of Applied Microbiology* 102, 1239-1253.

Van der Gucht, K., Cottenie, K., Muylaert, K., Vloemans, N., Cousin, S., Declerck, S., Jeppesen, E., Conde-Porcuna, J.-M., Schwenk, K., Zwart, G., 2007. The power of species sorting: local factors drive bacterial community composition over a wide range of spatial scales. *Proceedings of the National Academy of Sciences* 104, 20404-20409.

Vick-Majors, T.J., Priscu, J.C., Amaral-Zettler, L.A., 2014. Modular community structure suggests metabolic plasticity during the transition to polar night in ice-covered Antarctic lakes. *The ISME journal* 8, 778-789.

Vinas, M., Sabaté, J., Espuny, M.J., Solanas, A.M., 2005. Bacterial community dynamics and polycyclic aromatic hydrocarbon degradation during bioremediation of heavily creosote-contaminated soil. *Applied and Environmental Microbiology* 71, 7008-7018.

Wang, K., Ye, X., Chen, H., Zhao, Q., Hu, C., He, J., Qian, Y., Xiong, J., Zhu, J., Zhang, D., 2015. Bacterial biogeography in the coastal waters of northern Zhejiang, East China Sea is highly controlled by spatially structured environmental gradients. *Environmental Microbiology* 17, 3898-3913.

Xiong, J., Liu, Y., Lin, X., Zhang, H., Zeng, J., Hou, J., Yang, Y., Yao, T., Knight, R., Chu, H., 2012. Geographic distance and pH drive bacterial distribution in alkaline lake sediments across Tibetan Plateau. *Environmental Microbiology* 14, 2457-2466.

Xiong, J., Ye, X., Wang, K., Chen, H., Hu, C., Zhu, J., Zhang, D., 2014. Biogeography of the sediment bacterial community responds to a nitrogen pollution gradient in the East China Sea. *Applied and Environmental Microbiology* 80, 1919-1925.

Zhou, E., Crawford, R.L., 1995. Effects of oxygen, nitrogen, and temperature on gasoline biodegradation in soil. *Biodegradation* 6, 127-140.

Zytner, R., Salb, A., Stiver, W., 2006. Bioremediation of diesel fuel contaminated soil: comparison of individual compounds to complex mixtures. *Soil and Sediment Contamination: An International Journal* 15, 277-297.

Figure captions

Fig. 1. (A) Principal component analysis (PCA) based on soil physicochemical characteristics as variables. Alpha diversities indices were fitted as factors with significance < 0.05 onto the Ordination. (B) and (C) Linear regression relationships between pH/TPH and Richness. The TPH contents were log transformed.

Fig. 2. Geospatial biodiversity mapping analysis of the five sample sites. Dendrogram of Bray-Curtis distance based on bacterial communities and Euclid distance based on soil environmental factors were shown across geographical coordinates of sample sites. Branches are colored by the sample sites that they represent; white branches encompass two or more sites. Pie charts represent average phylum-level taxonomic compositions of all samples from each site.

Fig. 3. (A) Constrained analysis of principal coordinates (CAP) based on Bray-Curtis and environmental variables that were significantly related to microbial variation. (B) Partial least squares regression (PLSR) correlation loadings of four environmental variables (gray text) and select microbial taxa (black text).

Fig. 4. Correlation between community dissimilarities (Bray-Curtis distance) and geographic distances. Regression line, $y = -0.034x + 0.423$; $r = 0.494$; $P < 0.001$.

Fig. 5. Variation partitioning of bacterial communities by environmental variables

(top left) and the spatial factors including linear trend (Trend, top right) and PCNM variables (bottom).

Fig. 6. Network of co-occurring bacterial genera based on correlation analysis. A connection stands for a strong (Spearman's $\rho > 0.6$) and significant ($P < 0.01$) correlation. The size of each node is proportional to the relative abundance; the thickness of each connection between two nodes (edge) is proportional to the value of Spearman's correlation coefficients.

A. Co-occurring network colored by phylum.

B. Co-occurring network colored by modularity class.

702 **Table 1.** Environmental factors of the 5 sampling sites. The data represent the mean $\pm$
703 standard deviation. Different small letters indicate significant differences between
704 sites for that parameter. Means compared using one-way ANOVA, $P < 0.05$ level.

Environmental factor	Site				
	JB(9)	JKH(7)	XY(9)	YC(4)	YP(9)
Longitude (°E)	37°35.950'	35°39.180'	34°21.513'	36°37.542'	37°0.399'
Latitude (°N)	108°50.169'	109°20.866'	108°46.446'	111°0.695'	109°50.977'
pH	7.71 $\pm$ 0.25b	8.63 $\pm$ 0.19a	8.02 $\pm$ 0.3b	8.95 $\pm$ 0.22a	8.53 $\pm$ 0.23a
TPH (g/kg)	24.87 $\pm$ 19.35a	1.99 $\pm$ 0.70b	2.94 $\pm$ 1.21b	1.56 $\pm$ 0.86b	2.12 $\pm$ 2.73b
Cr (mg/kg)	36.39 $\pm$ 18.13a	22.74 $\pm$ 2.75ab	26.22 $\pm$ 5.29ab	22.51 $\pm$ 3.47ab	18.68 $\pm$ 4.16b
Pb (mg/kg)	11.01 $\pm$ 2.65b	16.51 $\pm$ 3.71b	138.34 $\pm$ 41.6a	12.06 $\pm$ 3.45b	15.82 $\pm$ 3.73b
Cd (mg/kg)	0.35 $\pm$ 0.18b	0.43 $\pm$ 0.22b	1.94 $\pm$ 0.61a	0.26 $\pm$ 0.15b	0.36 $\pm$ 0.22b
TP (mg/kg)	989.15 $\pm$ 479.83ab	607.28 $\pm$ 98.07b	1163.69 $\pm$ 505.71a	596.25 $\pm$ 67.86b	655.42 $\pm$ 86.98b
AP (mg/kg)	52.13 $\pm$ 33.47ab	20.87 $\pm$ 24.12bc	63.64 $\pm$ 21.91a	17.27 $\pm$ 6.68c	20.37 $\pm$ 8.26c
TN (mg/kg)	2776.25 $\pm$ 1598.3a	449.05 $\pm$ 264.96b	1907.08 $\pm$ 505.37a	316.76 $\pm$ 113.36b	496.83 $\pm$ 127.32b
AN (mg/kg)	37.98 $\pm$ 20.1a	26.15 $\pm$ 33.24a	28.31 $\pm$ 7.82a	8.39 $\pm$ 0.3a	14.29 $\pm$ 5.76a
AK (mg/kg)	87.5 $\pm$ 17.04b	158.71 $\pm$ 176.2ab	215.51 $\pm$ 44.33a	128.88 $\pm$ 100.43ab	138.42 $\pm$ 16.89ab

705

706 **Table 2.** Topological properties of co-occurring networks

	Modularity (MD)	Clustering coefficient (CC)	Average path length (APL)	Network diameter (ND)	Graph density (GD)	Average degree (AD)
Whole network	0.551	0.657	4.230	7.123	0.0998	9.389
random network	0.241 ($\pm$ 0.0155)	0.0995 ($\pm$ 0.0078)	2.257 ($\pm$ 0.0076)	4.0084 ($\pm$ 0.0956)	0.0998	9.389

707 Random network is an identically sized random network generated based on Erdős–Rényi model,

708 whose topological properties are calculated as the average value from 10 000 Erdős–Rényi

709 random networks.

710

Table 3. ANOVA of the environmental factors correlated to the bacteria

beta-diversity in CAP analysis.

Factor	<i>df</i>	<i>Variance</i>	<i>Pseudo-F</i>	<i>P-value</i>	<i>Signif.</i>
pH	1	0.9392	6.7211	0.001	***
TPH	1	0.3943	2.8219	0.003	**
TN	1	0.3841	2.7490	0.001	***
Cd	1	0.4803	3.4373	0.001	***
Residual	33	4.6114			

Axis	<i>df</i>	<i>Variance</i>	<i>Pseudo-F</i>	<i>P-value</i>	<i>Signif.</i>
CAP1	1	0.9817	7.0250	0.001	***
CAP2	1	0.7941	5.6824	0.001	***
CAP3	1	0.2856	2.0439	0.013	*
CAP4	1	0.1367	0.9780	0.441	
Residual	33	4.6114			

Significance: 0=***; 0.001=**; 0.01=*; 0.05='.'; 0.1="

df = degrees of freedom

Permutations = 999

Overall model significance: *Pseudo-F* = 3.9323, *P* = 0.001

Figure 1

A

B

C

Figure 2

Figure 3

A

B

Figure 4

Figure 5

Proteobacteria (49)	Actinobacteria (13)	Bacteroidetes (12)	Firmicutes (8)
Chloroflexi (6)	Spirochaetes (2)	Tenericutes (1)	Synergistetes (1)
Planctomycetes (1)	Acidobacteria (1)	Nitrospirae (1)	

Module I (28) Module II (27) Module III (20) Module IV (9)
Module V (7) Module VI (2) Module VII (2)