

ORIGINAL ARTICLE

Functional gene diversity of soil microbial communities from five oil-contaminated fields in China

Yuting Liang^{1,2,3}, Joy D Van Nostrand^{3,4}, Ye Deng^{3,4}, Zhili He^{3,4}, Liyou Wu^{3,4}, Xu Zhang², Guanghe Li² and Jizhong Zhou^{2,3,4,5}

¹Changzhou University, Jiangsu, China; ²Department of Environmental Science and Engineering, Tsinghua University, Beijing, China; ³Institute for Environmental Genomics and Department of Botany and Microbiology, University of Oklahoma, Norman, OK, USA; ⁴Virtual Institute for Microbial Stress and Survival, University of Missouri-Columbia, Columbia, SC, USA and ⁵Earth Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA

To compare microbial functional diversity in different oil-contaminated fields and to know the effects of oil contaminant and environmental factors, soil samples were taken from typical oil-contaminated fields located in five geographic regions of China. GeoChip, a high-throughput functional gene array, was used to evaluate the microbial functional genes involved in contaminant degradation and in other major biogeochemical/metabolic processes. Our results indicated that the overall microbial community structures were distinct in each oil-contaminated field, and samples were clustered by geographic locations. The organic contaminant degradation genes were most abundant in all samples and presented a similar pattern under oil contaminant stress among the five fields. In addition, alkane and aromatic hydrocarbon degradation genes such as monooxygenase and dioxygenase were detected in high abundance in the oil-contaminated fields. Canonical correspondence analysis indicated that the microbial functional patterns were highly correlated to the local environmental variables, such as oil contaminant concentration, nitrogen and phosphorus contents, salt and pH. Finally, a total of 59% of microbial community variation from GeoChip data can be explained by oil contamination, geographic location and soil geochemical parameters. This study provided insights into the *in situ* microbial functional structures in oil-contaminated fields and discerned the linkages between microbial communities and environmental variables, which is important to the application of bioremediation in oil-contaminated sites.

The ISME Journal (2011) 5, 403–413; doi:10.1038/ismej.2010.142; published online 23 September 2010

Subject Category: microbial population and community ecology

Keywords: GeoChip; functional gene arrays; metagenomics; oil contamination; functional gene diversity

Introduction

With a large demand for crude oil as an energy source, oil contamination occurs quite often as a result of exploration, production, maintenance, transportation, storage and accidental release, leading to significant ecological impact (Allen *et al.*, 2007; Paise *et al.*, 2008). Oil contamination is particularly severe and has become a global issue, such as in Kuwait (Al-Sarawi and Massoud, 1998; Al-Hashem *et al.*, 2007; Din *et al.*, 2008), India (Gogoi *et al.*, 2003), Libya (Hamid *et al.*, 2008),

China (Xiong *et al.*, 1997; Liang *et al.*, 2009a) and the United States (Lundegard and Johnson, 2006).

Indigenous microbial communities have an important role in oil contaminant degradation. Once the site is contaminated, the microbial community composition will be greatly changed. Both laboratory and field studies have shown that hydrocarbon contamination shifts the overall microbial community structure. Certain hydrocarbon-degrading taxa become dominant in oil-impacted environments because of natural selection resulting from the pressure of oil contaminants (Margesin *et al.*, 2003; Head *et al.*, 2006). Other environmental variables also influence microbial distribution, such as regional climate (Bhattacharya *et al.*, 2003; Maila *et al.*, 2006), soil type and characteristics (Hamamura *et al.*, 2006) and vegetation (Joner *et al.*, 2001). To date, most of the efforts to describe microbial communities in oil-contaminated fields have been focused on phylogenetic composition. Information

Correspondence: J Zhou, Institute for Environmental Genomics, University of Oklahoma, Norman, OK 73019, USA.

E-mail: jzhou@ou.edu

or G Li, Department of Environmental Science and Engineering, Tsinghua University, Beijing, 100084, China.

E-mail: ligh@tsinghua.edu.cn

Received 10 June 2010; accepted 9 July 2010; published online 23 September 2010

is still lacking with regard to microbial functional structures and its possible relationship with environmental variables in oil-contaminated fields.

Understanding the microbial functional diversity and the factors influencing microbial functions is important for bioremediation of oil-contaminated soils. However, analyzing microbial functional structure in a rapid and comprehensive manner is difficult using conventional molecular ecology approaches. Thus, high-throughput large-scale sequencing and/or microarray-based metagenomics tools are needed. GeoChip, a functional gene array, is a powerful metagenomics technology for analyzing microbial community functional structure (He *et al.*, 2007). GeoChip 2.0 contains 24 243 oligonucleotide (50-mer) probes, which cover >10 000 genes in >150 functional groups involved in nitrogen, carbon, sulfur and phosphorus cycling, metal reduction and resistance, as well as in organic contaminant degradation (He *et al.*, 2007). Several studies have demonstrated that GeoChip is an ideal tool for dissecting the microbial community functional structure in both natural and contaminated environments (Leigh *et al.*, 2007; Yergeau *et al.*, 2007; Wu *et al.*, 2008; Zhou *et al.*, 2008; Van Nostrand *et al.*, 2009; Wang *et al.*, 2009; He *et al.*, 2007, 2010; Xu *et al.*, 2010).

In this study, we determined microbial functional gene diversity from oil-contaminated soils with GeoChip. Samples were obtained from surface soils of five well-known oil fields from different geographic regions of China. This study aimed at addressing the following questions: (i) What is the microbial functional structure in oil-contaminated fields? (ii) How does oil contamination affect microbial functional gene diversity? (iii) How do environmental factors affect microbial community functional diversity in oil-contaminated fields? Our results indicated that microbial functional gene structure is distinct in each oil-contaminated field. The organic contaminant degradation genes were most abundant in all samples and presented similar pattern under oil contaminant stress. Moreover, such microbial functional patterns were highly correlated to oil contamination and environmental factors.

Materials and methods

Site description and sampling

Soil samples were obtained from oil-contaminated sites in five oil fields (fields in areas where oil is extracted) located at different geographic regions of China (660–2030 km apart): Daqing (DQ) in northeast China, Yumen (YM) and Changqing (CQ) in northwest China, Shengli in the Yellow River area in north China and Jiangnan in the Yangtze River area in central China (Figure 1a). The climates in the five regions vary greatly. DQ (46°35'N, 125°18'E) has a temperate continental monsoon climate with a mean annual rainfall of 427.5 mm. YM (39°49'N, 97°35'E)

has a temperate continental arid climate with a mean annual rainfall of 55.1 mm. CQ (34°20'N, 107°10'E) has a temperate continental monsoon climate with a mean annual rainfall of 470 mm. Shengli (37°28'N, 118°29'E) has a warm temperate continental semihumid monsoon climate with a mean annual rainfall of 550 mm. Jiangnan (30°21'N, 114°20'E) has a subtropical humid monsoon climate with a mean annual rainfall of 1159.8 mm. Contaminated soils were collected adjacent to crude oil pumping wells in which contamination occurred several years ago. Uncontaminated soils were collected from undisturbed pristine soils. Sampling sites in the same regions were located in an area of 1 km². All were collected from surface soil (0–10 cm). Several soil cores were composited for microbial and chemical analysis (200 g in total). Soils were sealed in sterile sampling bags and transported to the lab on ice. The following physical and chemical parameters of soil were measured according to the recommended soil testing procedures (Lu, 1999): pH, water content, total nitrogen (nitrogen in all organic and inorganic forms), available nitrogen (N, NO₃⁻-N, NO₂⁻-N and NH₄⁺-N), total phosphorus (phosphorus in all organic and inorganic forms), available phosphorus (P, PO₄³⁻-P), available micronutrients (Cu, Zn, Fe, Mn), organic matter, soluble salts and soil texture. The crude oil concentration in the contaminated soil was determined using an Ultrasonic-Soxhlet extraction gravimetric method (Huesemann, 1995). Uncontaminated soils were also extracted with the Ultrasonic-Soxhlet to confirm that they were 'non-contaminated'. The four components (aliphatic hydrocarbons, aromatic hydrocarbons, polar fraction with heteroatoms of nitrogen, sulfur and oxygen (NSO fraction) and asphaltenes) of crude oil were separated using silica gel (0.15–0.18 mm) and alumina (0.07–0.15 mm) column chromatography (Liang *et al.*, 2009b). Aliquots of soil samples were stored at -80 °C for molecular analysis.

Soil microbial DNA extraction

Microbial genomic DNA was extracted from 5 g of well-mixed soil for each sample by combining freeze-grinding and sodium dodecyl sulfate for cell lysis as previously described (Zhou *et al.*, 1996). Crude DNA was purified by agarose gel electrophoresis, followed by phenol–chloroform–butanol extraction. The purified DNA was quantified with agarose gel electrophoresis, an ND-1000 spectrophotometer (Nanodrop Inc., Wilmington, DE, USA) and Quant-It PicoGreen kit (Invitrogen, Carlsbad, CA, USA).

GeoChip hybridization

A detailed microbial functional diversity study was conducted with GeoChip. A pair of contaminated and uncontaminated soils from each field was selected for GeoChip analysis, with three replicates

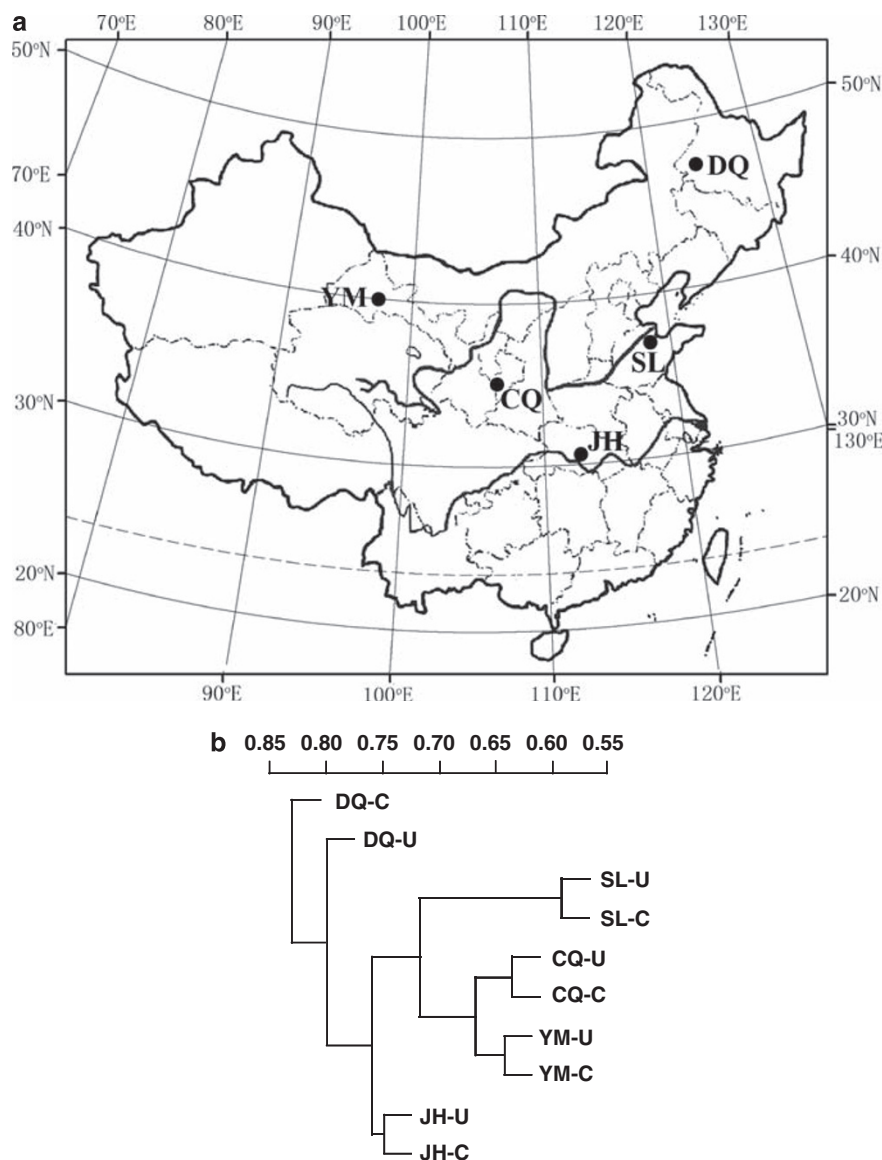


Figure 1 (a) Map of the five sampling oil-contaminated fields located in different geoclimate regions across China: Daqing (DQ) in northeast, Yumen (YM) and Changqing (CQ) in northwest, Shengli (SL) in north and Jiangnan (JH) in central China; and (b) hierarchical clustering analysis of microbial communities in the five oil-contaminated fields based on all functional genes. U: uncontaminated; C: contaminated.

for each sample. An aliquot of 100 ng DNA was amplified using the TempliPhi kit (Amersham Biosciences, Piscataway, NJ, USA) in a modified buffer containing single-strand binding protein ($200 \text{ ng } \mu\text{l}^{-1}$) and spermidine (0.04 mM) to increase the sensitivity and representation of amplification, and incubated at $30 \text{ }^\circ\text{C}$ for 3 h (Wu *et al.*, 2006). All of the amplified DNA was labeled, purified and resuspended in $130 \mu\text{l}$ hybridization solution as described previously (Wu *et al.*, 2006). The fluorescently labeled DNA was hybridized with GeoChip 2.0 on a HS4800 Hybridization Station (Tecan US, Durham, NC, USA) in triplicate at $42 \text{ }^\circ\text{C}$ for 10 h. Microarrays were scanned by a ScanArray 5000 Microarray Analysis System (PerkinElmer,

Wellesley, MA, USA) at 95% laser power and 68% photomultiplier tube gain.

Data analysis

For GeoChip analysis, signal intensities of each spot were measured with ImaGene 6.0 (Biodiscovery Inc., El Segundo, CA, USA). Only the spots automatically scored as positive in the output of raw data were used for further data analysis. The signal intensities used for final analysis were background subtracted. Intensities of three replicates were normalized with mean signal intensities as described previously (Wu *et al.*, 2008). Spots with signal intensity-background intensity/background standard deviation < 2.0) and

outliers of replicates (>2 standard deviation) were removed. A gene was considered present when a positive hybridization signal was obtained from $\geq 51\%$ of spots across all replicates. A matrix was generated from the normalized pixel intensities of all protein coding genes.

Hierarchical cluster analysis of whole functional genes was performed using the unweighted pairwise average-linkage clustering algorithm (Eisen *et al.*, 1998) with R, using vegan and stats packages. Detailed hierarchical clustering of functional genes was performed using CLUSTER (<http://rana.stanford.edu>), and visualized in TREEVIEW (<http://rana.stanford.edu/>).

The contributions of soil geochemical parameters (E), geographic location (L) and oil contamination (O) to microbial community variations were further evaluated with variance partitioning analysis using (partial) canonical correspondence analysis by CANOCO for Windows Version 4.5 (ter Braak and Smilauer, 1998, Plant Research International, Wageningen, The Netherlands). All soil geochemical data, except pH values, were $\log_2(x+1)$ transformed for standardization. The difference in soil texture was not considered, as clay, silt and sand compositions were only available for some samples. Spatial variables measured as latitude–longitude coordinates were converted into projected coordinates and represented by a cubic trend-surface polynomial to capture broad-scale spatial trends (Legendre and Legendre, 1998). A forward selection procedure was performed on all environmental and spatial variables to select the most parsimonious subset of variables for further variance partitioning analysis. The significance test was carried out by Monte Carlo permutation (999 times).

Results

Geochemical data variation

The oil content and other geochemical parameters of the soil samples are shown in Supplementary Table S1. Contamination level varied substantially among these sites investigated, ranging from 2 to 238 mg g⁻¹. The four components (aliphatic hydrocarbons, aromatic hydrocarbons, polar fractions with heteroatoms of nitrogen, sulfur and oxygen

(NSO fraction) and asphaltenes) in crude oil extracted from soils were analyzed (Supplementary Table S2). Generally, aliphatic and aromatic hydrocarbons were the main components of oil (>50%). The concentration of oil components varied substantially among the fields. For example, soils from DQ had a relatively high abundance of aliphatic hydrocarbons (>50%) and a relatively low abundance of NSO fraction and asphaltenes, whereas soils from YM and CQ had a relatively high abundance of asphaltenes (18–27%). Compared with the concentrations of oil components from oil-pumping wells (Supplementary Table S3), there was a dramatic decrease of aliphatic hydrocarbons in soils, which was due to both volatilization of low molecular weight compounds and biodegradation by indigenous microbes (Liang *et al.*, 2009b). Other soil parameters such as nutrient level (nitrogen, phosphorous), water content and soil texture also varied considerably among the five fields. Among the soil geochemical variables, significant correlation was observed between water content and oil ($r = -0.58$, $P < 0.01$), which might be due to increase in water evaporation caused by higher soil temperature with oil contamination. Detrended correspondence analysis was performed on all geochemical data, including oil content (Supplementary Figure S1). The samples from YM and CQ, both located in northwest China, grouped separately, whereas other samples were not obviously clustered together.

Overall microbial functional patterns

The microbial functional diversities of both contaminated and uncontaminated soils from each field were analyzed with GeoChip 2.0. A total of 1814 functional genes were detected. Overall, the gene numbers showed considerable differences among sites and between contaminated and uncontaminated soils (Table 1). For example, the gene numbers were higher in CQ and lower in Jianghan and DQ. Within each site, gene numbers decreased in most contaminated soils ($P < 0.05$). Furthermore, the overall microbial functional diversity was significantly decreased with contaminants based on both the Shannon–Weaver index (H) and Simpson's diversity index ($1/D$) ($P < 0.05$). However, the evenness of

Table 1 Diversity indices of all functional genes in soil samples

	DQ-U	DQ-C	YM-U	YM-C	CQ-U	CQ-C	SL-U	SL-C	JH-U	JH-C
Richness ^a	177	145	349	380	1076	541	558	325	196	151
H^b	4.70	4.43	5.55	5.54	6.56	5.88	5.95	5.45	4.75	4.59
$1/D^c$	69.2	49.9	184.6	167.1	484.9	227.0	270.6	167.0	71.1	58.5
Evenness ^d	0.91	0.89	0.95	0.93	0.94	0.93	0.94	0.94	0.90	0.91

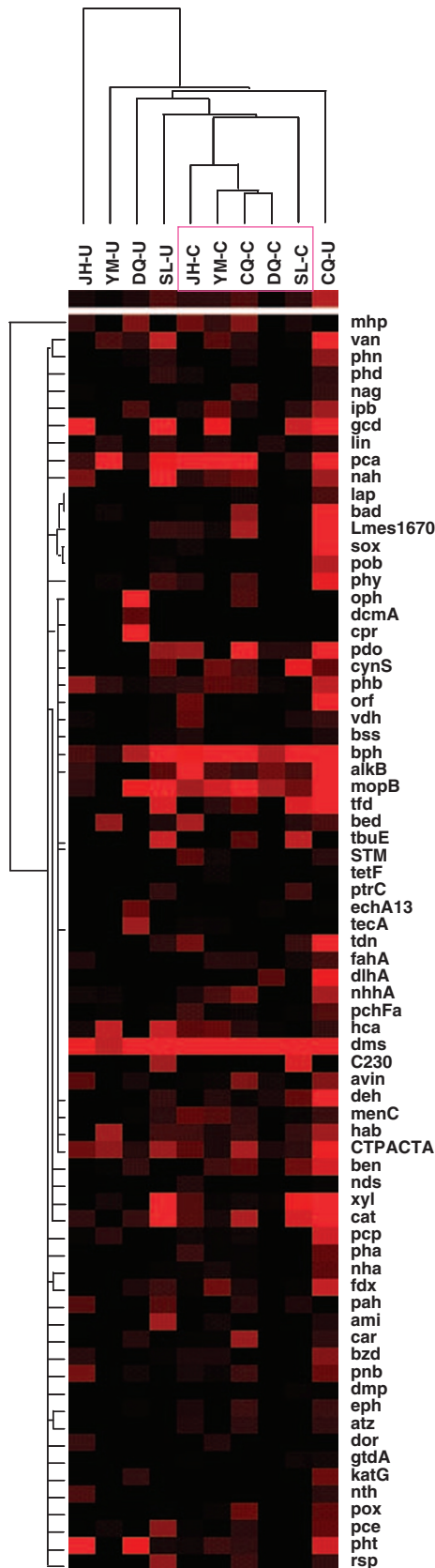
DQ, Daqing; YM, Yumen; CQ, Changqing; SL, Shengli; JH, Jianghan; U, uncontaminated; C, contaminated.

^aDetected gene number.

^bShannon–Weaver index, higher number represents higher diversity.

^cReciprocal of Simpson's index, higher number represents higher diversity.

^dEvenness index.



all samples did not show obvious differences and all were close to 1, indicating an even distribution of soil microbial communities.

Hierarchical clustering analysis was performed on all functional genes to examine overall patterns of variation among the soil samples (Figure 1b). Interestingly, cluster results showed that samples grouped by geographic locations, indicating distinct microbial communities in the five fields

Changes in microbial functional genes

Identifying microbial functional genes in the environment is important to establish the linkage of community structure to functions. Microbial functional genes detected for major biogeochemical/metabolic processes among the five fields are listed in Supplementary Table S4. Functional genes involved in organic contaminant degradation were most abundant in all samples. However, there was little difference in the proportion of contaminant degradation genes detected in contaminated versus uncontaminated soils, although the overall diversity of functional genes in contaminated soils was lower than that in uncontaminated soils. Owing to their potential importance in bioremediation of oil-contaminated fields, we primarily focused on the organic contaminant degradation genes below. Genes involved in other processes were described only briefly.

Organic contaminant degradation genes. Biodegradation of hydrocarbons is the most important process for the removal of oil contaminants from the environment (Kanaly and Harayama, 2000; Van Hamme *et al.*, 2003). Microbial functional genes may represent good indicators for the biodegradation potential of microbial communities. In total, 618 genes from 73 gene families involved in organic contaminant degradation showed positive hybridization signals. Hierarchical clustering analysis was performed on all degradation genes (Figure 2). All the samples with contamination grouped together, indicating that the microbial communities might present similar functional pattern in organic contaminant degradation under oil contaminant stress.

Aliphatic hydrocarbons degradation genes. Aliphatic hydrocarbons were major components of oil contaminants (~50%) (Supplementary Table S2). The most common aerobic degradation pathway of alkanes is terminal oxidation, which is initiated by alkane monooxygenase-catalyzed oxidation of the

Figure 2 Clustering analysis of functional genes involved in organic contaminant degradation. The protein id number and its derived organism for each gene are presented. Heat maps were generated in CLUSTER and visualized using TREEVIEW. Red indicates signal intensities above background, whereas black indicates signal intensities below background. Brighter red coloring indicates higher signal intensities. DQ-Daqing, YM-Yumen, CQ-Changqing, SL-Shengli, JH-Jiangnan. U: uncontaminated; C: contaminated.

terminal methyl group of alkane to 1-alkanol. Genes encoding alkane monooxygenase in the soil samples from the five fields are listed in Supplementary Table S5. Alkane-1-monooxygenase genes derived from *Rhodococcus erythropolis* and *Rhodococcus* sp. Q15 (14331004, 15420776) were detected across all samples. *Rhodococcus* strains have been reported to degrade a broad range of aliphatics, C₆–C₃₆ n-alkanes and branched alkanes (Whyte *et al.*, 1998; van Beilen *et al.*, 2002). The alkane hydroxylase gene (13750769) from unidentified nocardia, rhodococci and mycobacteria (CNM-group) bacterium HXN1000 and alkane 1-monooxygenase (10185046) from *Burkholderia cepacia* were also detected in most samples. *B. cepacia* has been reported to degrade C₁₂–C₃₀ n-alkanes (Yuste *et al.*, 2000). Further work is needed to study the expression of alkane degradation genes of *in situ* microbial communities.

Aromatics degradation genes. Aromatic hydrocarbons are also major components of crude oil contaminants, and most polycyclic aromatic hydrocarbons (PAHs) have high carcinogenic and toxicological properties. In total, 161 genes involved in aromatic hydrocarbon degradation showed positive hybridization signal intensity. Seven genes encoding benzene dioxygenase were detected. The dominant genes were benzene 1,2-dioxygenase, ferredoxin protein gene (21232202) derived from *Xanthomonas campestris* pv. *campestris* str. ATCC 33913 and benzene 1,2-dioxygenase system ferredoxin-NAD(+) reductase component gene (115088) derived from *Pseudomonas putida*. Genes involved in toluene aerobic and anaerobic degradation and xylene degradation were detected, especially in contaminated soils, with dominant genes derived from *Pseudomonas* sp. and *Thauera aromatica*. A total of 24 genes involved in biphenyl degradation were detected, with dominant genes derived from *Rhodococcus* sp., *Comamonas testosteroni* and *Pseudomonas* sp.

A previous study indicated that PAHs with 2–5 rings and their alkylated derivatives were detected in oil-contaminated fields (Liang *et al.*, 2009b). Genes involved in the degradation of PAHs such as naphthalene, phenanthrene and pyrene were also studied (Supplementary Tables S6–S8). For naphthalene catabolism, genes encoding the naphthalene dioxygenase Fe–S protein small subunit derived from *Pseudomonas* sp. ND6 (38638603) and the naphthalene degradation regulator protein derived from *Rhodococcus* sp. (37683584) were dominant across all samples. Other genes encoding 1,2-dihydroxybenzylpyruvate aldolase (10956960), naphthalene catabolism enzyme (150231) and 2-hydroxychromene-2-carboxylate isomerase protein (17544992) were also detected in most samples. For phenanthrene catabolism, the genes encoding 1-hydroxy-2-naphthoate transporter, derived from *Nocardioides* sp. KP7 (6093708), and LysR-type transcriptional regulator PhnS, derived from *Burkholderia* sp. RP007 (3820514), were detected

in high abundance. For pyrene catabolism, genes encoding PAH ring-hydroxylating dioxygenase, large subunit two derived from *Mycobacterium* sp. 6PY1 (27657409), putative ring-hydroxylating dioxygenase large subunit derived from *Mycobacterium* sp. S65 (33333869) and catalase-peroxidase protein KatG derived from *Burkholderia pseudomallei* (22711993) were dominant.

Catechol is a key dihydroxylated intermediate in PAH catabolic pathways. The *ortho*- and *meta*-cleavage pathways generate the tricarboxylic acid cycle intermediated from catechol. The dioxygenase genes involved in catechol degradation were detected in all samples, with considerable variations in the number of genes from 4 to 55 (Supplementary Figure S2). Generally, 10–55 genes involved in catechol degradation were observed in uncontaminated soils and 4–26 genes in contaminated soils. The impact of oil contaminants on genes involved in catechol degradation varied substantially. Of all catechol degradation-related genes, four genes (*tdnJ* from *Pseudomonas* sp., *orf8* from *Rhodococcus* sp., catechol-*meta* derivative gene from *Mesorhizobium* sp. and catechol-*ortho* derivative gene from *Nostoc* sp.) were positively correlated with oil concentration ($r=0.55$ – 0.78 , $P\leq 0.05$). However, two genes (*catB2* from *Frateuria* sp. and *ORF291* from *Rubrivivax* sp.) were negatively correlated with oil concentration ($r=-0.67$ to -0.53 , $P\leq 0.05$). The most dominant genes (3402316, 3402328) both encoded catechol 2,3-dioxygenase from unidentified bacteria that were isolated from a mixed culture of crude oil-degrading bacteria and a mixed culture of phenol-degrading bacteria.

Carbon-cycling genes. Cellulose, lignin and chitin are the most abundant carbon sources derived from plant tissues or organisms in soil ecosystems. Microbial functional genes related to cellulase, laccase and chitinase were present in high abundance as expected (Supplementary Figures S3A–C). Of these, two genes (cellulase from *Pectobacterium* sp. and laccase from *Piloderma* sp.) were detected across all samples. Several carbon degradation genes were significantly affected by oil contamination. Three genes (two laccase genes from *Polyporus* sp. and *Trametes* sp. and a chitinase gene from an environmental isolate) were negatively correlated with oil concentration ($r=-0.73$ – 0.62 , $P\leq 0.05$). In contrast, three genes (cellulase from *Rhodospirillum* sp. and *Fusarium* sp. and polygalacturonase from *Phytophthora* sp.) were positively correlated with nitrogen available in soil ($r=0.55$ to -0.64 , $P\leq 0.05$). Most of the *rbcL* genes involved in carbon fixation were from uncultured bacteria (Supplementary Figure S3D) and were negatively correlated with oil concentrations ($r=-0.68$ to -0.78 , $P\leq 0.05$).

Nitrogen-cycling genes. Detailed changes of microbial functional genes involved in nitrogen cycling such as *nifH*, *amo*, urease, *nar* and *nir* are shown in

Supplementary Figure S4A–E. Most of the *nifH*, *nar* and *nir* genes observed were from environmental clones rather than from pure cultures. Of these nitrogen-cycling genes, two urease genes from *Deinococcus* sp. and *Mesorhizobium* sp. were abundant and present in all samples. Correlation analysis indicated that several genes were significantly negatively correlated with oil concentration: urease from *Wautersia*, *nosZ* from *Paracoccus* sp. and *nifH*, *narG*, *nirS* and *norB* from uncultured bacteria ($r = -0.71$ to -0.55 , $P \leq 0.05$).

As nitrogen could be a major limiting factor in bioremediation of oil-contaminated soils (Braddock *et al.*, 1997; Gogoi *et al.*, 2003; Liang *et al.*, 2009b), the relationship between nitrogen-cycling genes and nitrogen concentrations in soils was further examined. Several genes such as urease from *Mycobacterium* sp., *Streptomyces* sp., *Mesorhizobium* sp., *Bacillus* sp., *Brucella* sp. and *Pseudomonas* sp., and *nifH*, *nar* and *nir* from uncultured bacteria were positively correlated with soil nitrogen concentrations ($r = 0.65$ – 0.77 , $P \leq 0.05$). These results suggested that nitrogen-cycling activities could be negatively affected by oil contamination but positively affected by soil nitrogen concentrations.

Relationship between microbial functional diversity and environmental variables

Canonical correspondence analysis was performed to discern possible linkages between microbial functional structure and soil geochemical parameters (Figure 3). Only significant geochemical

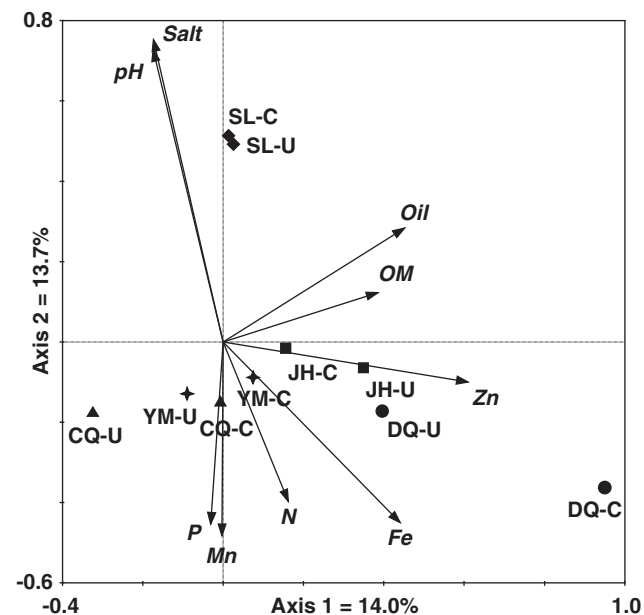


Figure 3 Canonical correspondence analysis (CCA) of GeoChip hybridization signal intensities and soil geochemical data that were significantly related to microbial variations: oil concentration, pH, salt content, organic matter, nitrogen (N), phosphorus (P), Fe, Zn and Mn. DQ-Daqing, YM-Yumen, CQ-Changqing, SL-Shengli, JH-Jianghan. U: uncontaminated; C: contaminated.

parameters were included in the canonical correspondence analysis biplot (oil, pH, N, P, Mn, Fe, Zn, salt and organic matter, based on a forward selection procedure and variance inflation factors with 999 Monte Carlo permutations). Samples were grouped together on the basis of their geographic locations, which was consistent with the clustering result of all functional genes. The first axis was positively correlated with organic matter, oil content and Zn. The second axis was positively correlated with salt content, pH and oil content, but negatively correlated with N, P and Mn. On the basis of the relationship between environmental variables and microbial functional structure, oil content seemed to be a major factor influencing the microbial functional structure in DQ and Jianghan, which have a long-term exploration history with the highest oil contamination level. In addition, N and P seemed to be major factors influencing the microbial functional structure in YM and CQ in northwest China, whereas oil content, salt and pH could have key roles in shaping microbial functional structure in Shengli in north China, which has high oil contamination levels and severe soil saline alkalization.

Variance partitioning analysis was further performed to quantify the contributions of geographic location (L), soil environmental parameters (E) and oil contamination (O) to the microbial community variation. We considered oil contamination as an independent factor, as it was due to human activity. The total variation was partitioned into the pure effects of L, E and O (when the effects of all other factors were removed), interactions between any two components ($L \times E$, $E \times O$ and $O \times L$), common interactions of all three components ($L \times E \times O$) and the unexplained portion (Figure 4a). On the basis of GeoChip data, a total of 58.7% of the variation was significantly explained ($P = 0.006$) by these three

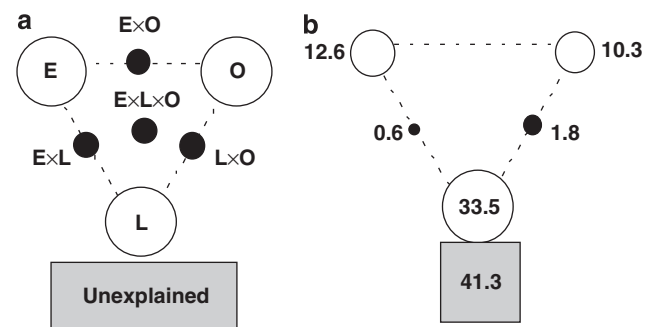


Figure 4 Variation partitioning analysis of microbial diversity explained by soil geochemical parameters (E), geographic location (L) and oil contamination (O). (a) General outline, (b) all functional genes. Each diagram represents the biological variation partitioned into the relative effects of each factor or a combination of factors, in which geometric areas were proportional to the respective percentages of explained variation. The edges of the triangle presented the variation explained by each factor alone. The sides of the triangles presented interactions of any two factors and the middle of the triangles represented interactions of all three factors.

components (Figure 4b). Oil contamination, geographic location and soil environmental parameters were able to independently explain 10.3, 33.5 and 12.6% of the total variations observed, respectively. Interactions among the three major components seemed to have less influence than did individual components, and were only observed between soil environmental parameters and geographic location (0.6%) and between oil contamination and geographic location (1.8%). More than 40% of the community variation based on GeoChip data could not be explained by these three components.

Discussion

Analyzing microbial functional genes encoding key enzymes involved in major biogeochemical processes and contaminant degradation is important to link microbial community structure to their potential ecological functions (Torsvik and Ovreas, 2002). However, characterization and quantification of microbial community functional structure are great challenges, especially for those in oil-contaminated fields. On one hand, oil contaminants could be toxic to many microbial populations, and thus reduce microbial diversity, on the other hand, the vast range of carbon substrates (alkanes, aromatics, PAHs, branched hydrocarbons and heterocyclics) and subsequent metabolites (carboxylic acids, *n*-aldehydes and so on) present in oil-contaminated soils could facilitate the development of rather complex microbial communities (Van Hamme *et al.*, 2003). It is extremely difficult to detect microbial metabolic genes involved in contaminant degradation, as well as other important geochemical cycles, in a rapid and comprehensive way using conventional molecular ecology approaches. Microarray-based genomic technology provides a powerful tool for monitoring various microbial functional genes (Leigh *et al.*, 2007; Yergeau *et al.*, 2009; Liang *et al.*, 2009a).

In this study, microbial functional gene diversity in five oil-contaminated fields located in different geographic regions of China was analyzed using GeoChip. The abundance of several functional genes involved in organic contaminant degradation, such as catechol dioxygenase genes and *bph*, was found to increase with oil concentration. The biodegradation of PAHs by microorganisms through a complex enzymatic process was well documented (Kanaly and Harayama, 2000). Langworthy *et al.* (1998) found that high PAH concentrations increased the frequencies of *nahA* and *alkB* genes. Gomes *et al.* (2005) also reported a positive correlation between naphthalene contamination and abundance of *phnAc* genes. The increase of these functional genes might be due to the natural selection of organisms capable of utilizing hydrocarbon compounds. The dominant detected genes involved in high molecular weight PAH degradation, such as pyrene dioxygenase, were mainly derived from *Mycobacterium*

sp., which has been shown to use PAHs as carbon and energy sources. For example, a series of enzymes from some *Mycobacterium* cultures were shown to be responsible for pyrene catabolism (Heitkamp *et al.*, 1988). The existence of metabolic genes indicates a potential metabolic capacity of indigenous microbial communities in oil-contaminated fields. Further work would be needed to connect microbial functional gene expression with contaminants degradation *in situ*.

Many genes involved in carbon and nitrogen cycling had negative correlations with oil contaminant concentrations. This is consistent with the report that oil contamination had a negative influence on soil urease and dehydrogenase activity (Megharaj *et al.*, 2000). Lindstrom *et al.* (1999) also found that N mineralization was lower in the oiled soils. The general consistency of GeoChip results with other studies further supports the robustness of GeoChip analysis as demonstrated by many of our previous studies (Rhee *et al.*, 2004; Tiquia *et al.*, 2004; Wu *et al.*, 2001, 2006; He *et al.*, 2007; Wang *et al.*, 2009).

Understanding the factors that influence microbial community structure is an important goal in microbial ecology. Our results indicate that oil contamination had a significant impact on soil microbial functional communities in oil-contaminated fields. Although the samples were from different geographic locations, the microbial organic contaminant degradation genes presented similar pattern under oil contaminant stress. This is expected because of natural selection. Aislabie *et al.* (2004) reported an increase in the numbers of hydrocarbon-degrading bacteria, typically *Rhodococcus*, *Sphingomonas* and *Pseudomonas*, after a hydrocarbon spill on Antarctic soils, although overall microbial diversity declined. Saul *et al.* (2005) also found that microbial communities in oil-contaminated soils were dominated by Proteobacteria, specifically, by members of the genera *Pseudomonas*, *Sphingomonas* and *Variovorax*, some of which degrade hydrocarbons.

Variation partitioning analysis in this study showed that geographic location contributed the most (33.5%) to microbial functional gene variation, which is in accordance with the hierarchical clustering of overall microbial functional genes, indicating a significant impact of local environmental conditions on the composition and structures of microbial communities. Although this is an interesting finding, the sampling strategy and the number of samples analyzed in this study limit our ability to fully address this question. Studies using a more extensive sampling effort are currently underway to examine the microbial distribution patterns in oil-contaminated sites across different geographic locations, with many replicated samples from each location.

A large part (41.3%) of the variations of microbial community structure could not be explained by any

of the three components (geographic location, soil geochemical variables and oil contamination). This is also consistent with several other recent studies. Zhou *et al.* (2008) showed that more than 50% of variations in a forest soil community could not be explained by both environmental variables and geographic distance. Ramette and Tiedje (2007b) showed that 34–80% of microbial variations could not be explained by measured variables in agricultural soils. All the above studies indicate that considerable amounts of variations could not be explained by environmental variables measured. This could be because of unmeasured environmental variables. For instance, it has been reported that the types of oil contaminants and age of site contamination could influence the microbial genetic diversity (Bhattacharya *et al.*, 2003). The composition and structure of hydrocarbons in oil contaminants are highly variable in different oil-contaminated fields, which could affect the microbial functional gene composition. Other unmeasured factors such as vegetation and plant species and soil texture may also influence microbial diversity (He *et al.*, 2006; Rooney-Varga *et al.*, 2007; Ramette and Tiedje, 2007b). Moreover, some other environmental factors, such as soil and water content, only represent the conditions when sampling occurred, which may have been significantly different during other periods. Moreover, sampling effects and ecologically neutral processes of diversification may contribute to the unexplained portion of microbial community variation (Ramette and Tiedje, 2007a; Zhou *et al.*, 2008).

One of the central goals of investigating microorganisms in oil-contaminated fields is for environmental cleanup. One practical application of studying microbial functional diversity and the related environmental variables in oil-contaminated fields is for assisting *in situ* bioremediation. High abundance of genes involved in organic contaminant degradation existed in all of the oil-contaminated sites (Supplementary Table S4), indicating the existence of indigenous microorganisms for oil contaminant degradation. However, the degradation process might be influenced by low nutrients (Atlas, 1981; Oh *et al.*, 2001). Specifically, nitrogen could be limited because of the decrease in nitrogen-cycling genes with oil concentration, which may indicate a decrease in nitrogen-cycling activity. Adjustment of the carbon/nitrogen ratio by adding nitrogen may be important for *in situ* bioremediation of oil-contaminated fields. The influence of several other environmental factors on hydrocarbon biodegradation has been studied. Ward and Brock (1978) showed that the rates of metabolism of hydrocarbon compounds decreased as salinity increased. Shiaris (1989) reported a generally negative correlation between salinity and rates of mineralization of phenanthrene and naphthalene. In this study, we also found a significant influence of salt in microbial functional diversity, especially in fields with severe

soil saline alkalization. These findings suggest the significant challenges that we may face when *in situ* bioremediation is performed on oil-contaminated sites.

The existence of key genes involved in hydrocarbon degradation across all oil-contaminated fields implies that stimulating indigenous microorganisms could be a valid option for remediating oil-contaminated sites. However, the microbial community structures were substantially different at different sites and within sites; hence, biostimulation of indigenous microorganisms could be a complicated process, and a variety of optimized strategies to stimulate and maintain the desired populations may need to be considered to achieve bioremediation goals. Alternatively, some key populations involved in degrading the desired hydrocarbon compounds may not be present and/or may be difficult to stimulate to the desired level. In this case, bioaugmentation through introducing desired microorganisms could be an option. However, on the basis of spatial isolation hypothesis (Zhou *et al.*, 2002; Treves *et al.*, 2003), although any introduced strain may be able to successfully colonize a highly diverse community, it would be difficult to achieve dominance (Zhou *et al.*, 2002). Understanding microbial functional potential, the mechanisms shaping microbial community functional structure and spatial distribution patterns will help design appropriate strategies for successful *in situ* bioremediation through biostimulation and/or bioaugmentation.

Acknowledgements

This work was supported by the National Natural Scientific Foundation of China (No. 40730738), the United States Department of Energy under the Environmental Remediation Science Program, and was part of the Virtual Institute for Microbial Stress and Survival (<http://VIMSS.lbl.gov>), which was supported by the U. S. Department of Energy, Office of Science, Office of Biological and Environmental Research, Genomics Program: GTL through contract DE-AC02-05CH11231 between Lawrence Berkeley National Laboratory and the U. S. Department of Energy, and was also supported by the Oklahoma Center for the Advancement of Science and Technology under the Oklahoma Applied Research Support Program.

References

- Aislabie J, Balks M, Foght J, Waterhouse E. (2004). Hydrocarbon spills on Antarctic soils: Effects and management. *Environ Sci Technol* **38**: 1265–1274.
- Al-Hashem M, Brain P, Omar S. (2007). Effects of oil pollution at Kuwait's greater Al-Burgan oil field on polycyclic aromatic hydrocarbon concentrations in the tissues of the desert lizard *Acanthodactylus*

- scutellatus* and their ant prey. *Ecotoxicology* **16**: 551–555.
- Al-Sarawi M, Massoud M. (1998). Preliminary assessment of oil contamination levels in soils contaminated with oil lakes in the greater Burgan oil fields, Kuwait. *Water Air Soil Pollut* **106**: 493–504.
- Allen J, Atekwana E, Atekwana E, Duris J, Werkema D, Rossbach S. (2007). The microbial community structure in petroleum-contaminated sediments corresponds to geophysical signatures. *Appl Environ Microbiol* **73**: 2860–2870.
- Atlas R. (1981). Microbial-degradation of petroleum-hydrocarbons: an environmental perspective. *Microbiol Rev* **45**: 180–209.
- Bhattacharya D, Sarma P, Krishnan S, Mishra S, Lal B. (2003). Evaluation of genetic diversity among *Pseudomonas citronellolis* strains isolated from oily sludge-contaminated sites. *Appl Environ Microbiol* **69**: 1435–1441.
- Braddock J, Ruth M, Catterall P, Walworth J, Mccarthy K. (1997). Enhancement and inhibition of microbial activity in hydrocarbon-contaminated arctic soils: Implications for nutrient-amended bioremediation. *Environ Sci Technol* **31**: 2078–2084.
- Din S, Dousari A, Literathy P. (2008). Evidence of hydrocarbon contamination from the Burgan oil field, Kuwait—Interpretations from thermal remote sensing data. *J Environ Manage* **86**: 605–615.
- Eisen M, Spellman P, Brown P, Botstein D. (1998). Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci USA* **95**: 14863–14868.
- Gogoi B, Dutta N, Goswami P, Mohan T. (2003). A case study of bioremediation of petroleum-hydrocarbon contaminated soil at a crude oil spill site. *Adv Environ Res* **7**: 767–782.
- Gomes N, Kosheleva I, Abraham W, Smalla K. (2005). Effects of the inoculant strain *Pseudomonas putida* KT2442 (pNF142) and of naphthalene contamination on the soil bacterial community. *FEMS Microbiol Ecol* **54**: 21–33.
- Hamamura N, Olson S, Ward D, Inskeep W. (2006). Microbial population dynamics associated with crude-oil biodegradation in diverse soils. *Appl Environ Microbiol* **72**: 6316–6324.
- Hamid H, Kassim W, El Hishir A, El-Jawashi S. (2008). Risk assessment and remediation suggestion of impacted soil by produced water associated with oil production. *Environ Monit Assess* **145**: 95–102.
- He J, Xu Z, Hughes J. (2006). Molecular bacterial diversity of a forest soil under residue management regimes in subtropical Australia. *FEMS Microbiol Ecol* **55**: 38–47.
- He Z, Gentry T, Schadt C, Wu L, Liebich J, Chong S *et al*. (2007). GeoChip: a comprehensive microarray for investigating biogeochemical, ecological and environmental processes. *ISME J* **1**: 67–77.
- He Z, Xu M, Deng Y, Kang S, Kellogg L, Wu L *et al*. (2010). Metagenomic analysis reveals a marked divergence in the structure of belowground microbial communities at elevated CO₂. *Ecol Lett* **13**: 564–575.
- Head I, Jones D, Roling W. (2006). Marine microorganisms make a meal of oil. *Nat Rev Microbiol* **4**: 173–182.
- Heitkamp M, Franklin W, Cerniglia C. (1988). Microbial-metabolism of polycyclic aromatic-hydrocarbons-isolation and characterization of a pyrene-degrading bacterium. *Appl Environ Microbiol* **54**: 2549–2555.
- Huesemann M. (1995). Predictive model for estimating the extent of petroleum hydrocarbon biodegradation in contaminated soils. *Environ Sci Technol* **29**: 7–18.
- Joner E, Johansen A, Loibner A, Dela Cruz M, Szolar O, Portal J *et al*. (2001). Rhizosphere effects on microbial community structure and dissipation and toxicity of polycyclic aromatic hydrocarbons (PAHs) in spiked soil. *Environ Sci Technol* **35**: 2773–2777.
- Kanaly R, Harayama S. (2000). Biodegradation of high-molecular-weight polycyclic aromatic hydrocarbons by bacteria. *J Bacteriol* **182**: 2059–2067.
- Langworthy D, Stapleton R, Saylor G, Findlay R. (1998). Genotypic and phenotypic responses of a riverine microbial community to polycyclic aromatic hydrocarbon contamination. *Appl Environ Microbiol* **64**: 3422–3428.
- Legendre L, Legendre P. (1998). *Numerical ecology*. Elsevier Press: Amsterdam, The Netherlands.
- Leigh M, Pellizari V, Uhlik O, Sutka R, Rodrigues J, Ostrom N *et al*. (2007). Biphenyl-utilizing bacteria and their functional genes in a pine root zone contaminated with polychlorinated biphenyls (PCBs). *ISME J* **1**: 134–148.
- Liang Y, Li G, Van Nostrand J, He Z, Wu L, Den Y *et al*. (2009a). Microarray-based analysis of microbial functional diversity along an oil contamination gradient in oil field. *FEMS Microbiol* **70**: 324–333.
- Liang Y, Van Nostrand J, Wang J, Zhang X, Zhou J, Li G. (2009b). Microarray-based functional gene analysis of soil microbial communities during ozonation and biodegradation of crude oil. *Chemosphere* **75**: 193–199.
- Lindstrom J, Barry R, Braddock J. (1999). Long-term effects on microbial communities after a subarctic oil spill. *Soil Biol Biochem* **31**: 1677–1689.
- Lu RK. (1999). *Soil agricultural chemical analysis*. China Agricultural Science and Technology Press: Nanjing, China.
- Lundegard P, Johnson P. (2006). Source zone natural attenuation at petroleum hydrocarbon spill sites-II: Application to a former oil field. *Ground Water Monit Remed* **26**: 93–106.
- Maila M, Randima P, Dronen K, Cloete T. (2006). Soil microbial communities: Influence of geographic location and hydrocarbon pollutants. *Soil Biol Biochem* **38**: 303–310.
- Margesin R, Labbe D, Schinner F, Greer C, Whyte L. (2003). Characterization of hydrocarbon-degrading microbial populations in contaminated and pristine alpine soils. *Appl Environ Microbiol* **69**: 3085–3092.
- Megharaj M, Singleton I, McClure N, Naidu R. (2000). Influence of petroleum hydrocarbon contamination on microalgae and microbial activities in a long-term contaminated soil. *Arch Environ Contam Toxicol* **38**: 439–445.
- Oh Y, Sim D, Kim S. (2001). Effects of nutrients on crude oil biodegradation in the upper intertidal zone. *Mar Pollut Bull* **42**: 1367–1372.
- Paisse S, Coulon F, Goni-Urriza M, Peperzak L, McGenity T, Duran R. (2008). Structure of bacterial communities along a hydrocarbon contamination gradient in a coastal sediment. *FEMS Microbiol Ecol* **66**: 295–305.
- Ramette A, Tiedje J. (2007a). Biogeography: an emerging cornerstone for understanding prokaryotic diversity, ecology, and evolution. *Microb Ecol* **53**: 197–207.
- Ramette A, Tiedje J. (2007b). Multiscale responses of microbial life to spatial distance and environmental heterogeneity in a patchy ecosystem. *Proc Natl Acad Sci USA* **104**: 2761–2766.
- Rhee S, Liu X, Wu L, Chong S, Wan X, Zhou J. (2004). Detection of genes involved in biodegradation and

- biotransformation in microbial communities by using 50-mer oligonucleotide microarrays. *Appl Environ Microbiol* **70**: 4303–4317.
- Rooney-Varga J, Giewat M, Duddlestone K, Chanton J, Hines M. (2007). Links between archaeal community structure, vegetation type and methanogenic pathway in Alaskan peatlands. *FEMS Microbiol Ecol* **60**: 240–251.
- Saul D, Aislabie J, Brown C, Harris L, Foght J. (2005). Hydrocarbon contamination changes the bacterial diversity of soil from around Scott Base, Antarctica. *FEMS Microbiol Ecol* **53**: 141–155.
- Shiaris M. (1989). Seasonal biotransformation of naphthalene, phenanthrene, and benzo[a]pyrene in surficial estuarine sediments. *Appl Environ Microbiol* **55**: 1391–1399.
- ter Braak CJF, Smilauer P. (1998). *CANOCO Reference Manual and User's Guide to CANOCO for Windows: Software for Canonical Community Ordination*, Version 4 edn. Microcomputer Power: New York, USA.
- Tiquia S, Wu L, Chong S, Passovets S, Xu D, Xu Y *et al*. (2004). Evaluation of 50-mer oligonucleotide arrays for detecting microbial populations in environmental samples. *Biotechniques* **36**: 664–675.
- Torsvik V, Ovreas L. (2002). Microbial diversity and function in soil: from genes to ecosystems. *Curr Opin Microbiol* **5**: 240–245.
- Treves D, Xia B, Zhou J, Tiedje J. (2003). A two-species test of the hypothesis that spatial isolation influences microbial diversity in soil. *Microb Ecol* **45**: 20–28.
- van Beilen J, Smits T, Whyte L, Schorcht S, Rothlisberger M, Plaggemeier T *et al*. (2002). Alkane hydroxylase homologues in Gram-positive strains. *Environ Microbiol* **4**: 676–682.
- Van Hamme J, Singh A, Ward O. (2003). Recent advances in petroleum microbiology. *Microbiol Mol Biol Rev* **67**: 503–549.
- Van Nostrand J, Wu W, Wu L, Deng Y, Carley J, Carroll S *et al*. (2009). GeoChip-based analysis of functional microbial communities during the reoxidation of a bioreduced uranium-contaminated aquifer. *Environ Microbiol* **11**: 2611–2626.
- Wang F, Zhou H, Meng J, Peng X, Jiang L, Sun P *et al*. (2009). GeoChip-based analysis of metabolic diversity of microbial communities at the Juan de Fuca Ridge hydrothermal vent. *Proc Natl Acad Sci USA* **106**: 4840–4845.
- Ward D, Brock T. (1978). Hydrocarbon biodegradation in hypersaline environments. *Appl Environ Microbiol* **35**: 353–359.
- Whyte L, Hawari J, Zhou E, Bourbonniere L, Inniss W, Greer C. (1998). Biodegradation of variable-chain-length alkanes at low temperatures by a psychrotrophic *Rhodococcus* sp. *Appl Environ Microbiol* **64**: 2578–2584.
- Wu L, Kellogg L, Devol A, Tiedje J, Zhou J. (2008). Microarray-based characterization of microbial community functional structure and heterogeneity in marine sediments from the gulf of Mexico. *Appl Environ Microbiol* **74**: 4516–4529.
- Wu L, Liu X, Schadt C, Zhou J. (2006). Microarray-based analysis of subnanogram quantities of microbial community DNAs by using whole-community genome amplification. *Appl Environ Microbiol* **72**: 4931–4941.
- Wu L, Thompson D, Li G, Hurt R, Tiedje J, Zhou J. (2001). Development and evaluation of functional gene arrays for detection of selected genes in the environment. *Appl Environ Microbiol* **67**: 5780–5790.
- Xiong Z, Hu H, Wang Y, Fu G, Tan Z, Yan G. (1997). Comparative analyses of soil contaminant levels and plant species diversity at developing and disused oil well sites in Qianjiang Oilfield, China. *Bull Environ Contam Toxicol* **58**: 667–672.
- Xu M, Wu W, Wu L, He Z, Nostrand DVJ, Deng Y *et al*. (2010). Responses of microbial community functional structures to pilot-scale uranium *in situ* bioremediation. *ISME J* **1**: 1–11.
- Yergeau E, Arbour M, Brousseau R, Juck D, Lawrence J, Masson L *et al*. (2009). Microarray and real-time PCR analyses of the responses of high Arctic soil bacteria to hydrocarbon pollution and bioremediation treatments. *Appl Environ Microbiol* **75**: 6258–6267.
- Yergeau E, Kang S, He Z, Zhou J, Kowalchuk G. (2007). Functional microarray analysis of nitrogen and carbon cycling genes across an Antarctic latitudinal transect. *ISME J* **1**: 163–179.
- Yuste L, Corbella M, Turiegano M, Karlson U, Puyet A, Rojo F. (2000). Characterization of bacterial strains able to grow on high molecular mass residues from crude oil processing. *FEMS Microbiol Ecol* **32**: 69–75.
- Zhou J, Bruns M, Tiedje J. (1996). DNA recovery from soils of diverse composition. *Appl Environ Microbiol* **62**: 316–322.
- Zhou J, Kang S, Schadt C, Garten C. (2008). Spatial scaling of functional gene diversity across various microbial taxa. *Proc Natl Acad Sci USA* **105**: 7768–7773.
- Zhou J, Xia B, Treves D, Wu L, Marsh T, O'Neill R *et al*. (2002). Spatial and resource factors influencing high microbial diversity in soil. *Appl Environ Microbiol* **68**: 326–334.

Supplementary Information accompanies the paper on The ISME Journal website (<http://www.nature.com/ismej>)