RESEARCH ARTICLE

Microarray-based analysis of microbial functional diversity along an oil contamination gradient in oil field

Yuting Liang[1,2], Guanghe Li[2], Joy D. Van Nostrand[3,4], Zhili He[3,4], Liyou Wu[3,4], Ye Deng[3,4], Xu Zhang[2] & Jizhong Zhou[3,4]

[1]Jiangsu Polytechnic University, Jiangsu, China; [2]Department of Environmental Science and Engineering, Tsinghua University, Beijing, China; [3]Department of Botany and Microbiology, Institute for Environmental Genomics, University of Oklahoma, Norman, OK, USA; and [4]Virtual Institute for Microbial Stress and Survival, Berkeley, CA, USA

Correspondence: Guanghe Li, Department of Environmental Science and Engineering, Tsinghua University, Beijing 100084, China. Tel.: +8610 62773232; fax: +8610 62785687; e-mail: ligh@tsinghua.edu.cn

Received 29 April 2009; revised 2 July 2009; accepted 4 August 2009.
Final version published online 23 September 2009.

DOI:10.1111/j.1574-6941.2009.00774.x

Editor: James I. Prosser

Keywords
oil contamination; GeoChip; functional genes; oil field; microbial functional diversity.

Abstract

To understand better the *in situ* microbial functional diversity under oil contamination stress, soils were sampled along a contamination gradient at an oil field in north-east China. Microbial community functional structure was examined with a functional gene array, termed GeoChip. Multivariate statistical analysis and meta-analysis were conducted to study the functional gene responses to oil concentrations. The total functional gene abundance and diversity decreased along the gradient of increasing contamination. The overall abundance of soil bacteria, archaea and fungi decreased to 10%, 40% and 80% of those in the pristine soil. Several functional genes in the families *pgl*, *rbcL*, *nifH* and *nor* and those encoding cellulase, laccase, chitinase, urease and key enzymes in metabolizing organic compounds were significantly decreased with oil contamination, especially under high contamination stress. However, a few genes encoding key enzymes for catechol, protocatechuate, and biphenyl degradation and in the gene families of *nir*, *rbcL* and *pgl* showed a significant increase at a medium level of oil contamination. Oil content and soil available nitrogen were found to be important factors influencing the microbial community structure. The results provide an insight into microbial functional diversity in oil-contaminated soils, providing potential information for on-site management and remediation measures.

Introduction

With an increasing demand for oil, occurrences of oil contamination have become widespread as a result of exploration, production, maintenance, transportation, storage and accidental release (Atlas, 1981; Leahy & Colwell, 1990; Van Hamme *et al.*, 2003; Head *et al.*, 2006). Oil contamination in oil fields is particularly serious across the world (Xiong *et al.*, 1997; Al-Hashem *et al.*, 2007; Hamid *et al.*, 2008). Oil contaminants accumulate in soils and migrate to other habitats such as groundwater, posing a huge threat to the ecosystem and to human health. This is especially true for polycyclic aromatic hydrocarbons (PAHs), which are one of the main components in oils and have high carcinogenic and toxicological properties (Totsche *et al.*, 2006; Wehrer & Totsche, 2008).

The degradation of contaminants in natural environments is critically dependent on the metabolic capabilities of the indigenous microbial communities. Both laboratory and field experiments have shown that the overall microbial community is responsive to oil contamination. Certain microorganisms become prevalent under long-term stress. However, due to the extreme diversity and complexity of microbial communities, and their as yet uncultured characteristics, microbial detection, characterization and quantification in natural environments remain challenging. Difficulties in characterizing indigenous microbial communities impacted by oil contamination are exacerbated by the myriad of substrates and metabolic products (Van Hamme *et al.*, 2003). The rapid development of genomic tools has greatly advanced the knowledge of microbial communities in complex environments. For example, 16S rRNA gene-based PCR amplification has been widely used to examine microbial community structure and responses under stressful conditions. Most studies focus on the taxonomic-based characterization of microbial community structure and

FEMS MICROBIOLOGY ECOLOGY

dynamics and the dominant species. However, it might be difficult to directly link microbial phylogenetic diversity to physiological functions because only a limited number of organisms are well known for their narrowly defined metabolic capabilities. Thus, from a practical perspective, tracking specific functional genes in oil-impacted environments may be more useful for understanding microbial metabolic potential and ecological activities. Sotsky *et al.* (1994) detected bacterial populations containing both *xylE* and *alkB* genes in sediments affected by an oil spill. Several other functional genes, such as *ndoB*, *nahAc* and *phnAc*, have been used to assess the composition and degradation capacity of microbial communities in oil-contaminated sites or during bioremediation (Laurie & Lloyd-Jones, 2000; Siciliano *et al.*, 2003; Tuomi *et al.*, 2004; Park & Crowley, 2006). However, comprehensive information regarding the indigenous microbial functional diversity at oil-contaminated sites is still lacking.

Here, the *in situ* microbial community structure was examined along an oil contamination gradient in a long-term oil field in north-east China. In order to better understand the dynamics of microbial functional diversity, a robust functional gene array, GeoChip (He *et al.*, 2007), was used, containing 24 243 oligonucleotide probes covering $>$ 10 000 genes in $>$ 150 functional groups involved in nitrogen, carbon, sulfur and phosphorus cycling, metal reduction and resistance, and organic contaminant degradation. Multivariate statistical analysis and meta-analysis were conducted to study the functional gene responses to a gradient of oil concentration. Our results indicated a significant decrease in the microbial functional gene number and diversity along the contaminant gradient. Key functional genes encoding carbon and nitrogen cycling and organic contaminant degradation showed variable responses to oil contamination. Oil associated with soil available nitrogen was the most significant environmental factor controlling microbial functional gene diversity.

Materials and methods

Site description and sampling

Daqing oil field in north-east China was explored in the 1960s and has the largest production of crude oil among all oil fields in China. Soil contamination is particularly serious during long-term exploration, production and maintenance. Soil samples were obtained from a contaminated site (46°38′56″N, 125°04′18″E) in an area about 500 m × 500 m at Daqing oil field. Five contaminated samples were collected along an oil contaminant gradient near oil pumping wells where the contamination accumulated over a number of years through spillages. Uncontaminated samples were obtained from pristine soils. All soil samples were collected from the surface soil (10 cm deep). Several soil cores with similar oil concentrations were mixed as replicates for homogenization (200 g in total). Soils were sealed in sterile sampling bags and transported to the laboratory on ice. Soil physical and chemical parameters [pH, water content, total nitrogen (nitrogen in all organic and inorganic forms), available nitrogen (N, nitrate (NO_3^--N), nitrite (NO_2^--N) and ammonium (NH_4^+-N)), total phosphorus (phosphorus in all organic and inorganic forms), available phosphorus (P, phosphate (PO_4^{3-}-P)) and organic matter] were determined according to recommended soil testing procedures (Lu, 1999). Crude oil concentrations in the contaminated soil were determined using an Ultrasonic-Soxhlet extraction gravimetric method (Huesemann, 1995). Aliquots of soil samples were stored at -80 °C for molecular analysis.

Soil microbial community DNA extraction

The microbial community genomic DNA was extracted from 5 g of well-mixed soil samples by combining freeze-grinding and sodium dodecyl sulfate (SDS) for cell lysis as detailed in Zhou *et al.* (1996). The 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.) and a Quant-It™ PicoGreen® kit (Invitrogen, Carlsbad, CA).

GeoChip analysis

An aliquot (100 ng) of DNA from each sample was amplified in triplicate using the TempliPhi kit (Amersham Biosciences, Piscataway, NJ) in a modified buffer containing single-strand-binding protein (200 ng μL^{-1}) and spermidine (0.04 mM) to increase the sensitivity of amplification and was incubated at 30 °C for 3 h (Wu *et al.*, 2006). All of the amplified DNA was denatured and then fluorescently labeled and purified as described previously (Wu *et al.*, 2008). The labeled products were dried and resuspended in 130 μL hybridization solution containing 50% formamide, 3 × SSC, 0.3% SDS, 0.7 μg μL^{-1} herring sperm DNA, 0.02 mM dithiothreitol and water. Hybridization to GeoChip 2.0 was performed on an HS4800 Hybridization Station (TECAN US, Durham, NC) in triplicate at 42 °C for 10 h. Detailed probe information of GeoChip 2.0 was described by He *et al.* (2007). Microarrays were scanned using a ScanArray 5000® Microarray Analysis System (PerkinElmer, Wellesley, MA) at 95% laser power and 68% PMT (photomultiplier tube gain).

Data analysis

Signal intensities of each spot were measured with IMAGENE 6.0 (Biodiscovery Inc., El Segundo, CA). Only the spots

automatically scored as positive in the raw output data were used for further analysis. The signal intensities used for final analysis were subtracted from the background. Intensities of three replicates were normalized with mean signal intensities as described previously (Wu *et al.*, 2008). Spots with signal-to-noise ratio (SNR) < 2.0 [SNR = (signal intensity − background intensity)/background SD] and outliers of replicates (> 2 SDs) were removed. Gene detection was considered positive when a positive hybridization signal was obtained from $\geq 51\%$ of spots targeting the gene in all replicates. A matrix was generated from the normalized pixel intensities of all protein-encoding genes. The functional-gene-normalized signal intensities were defined as 'species' abundance. The normalized hybridization data for individual functional gene sequences were then reorganized based on phylogenetic groups (e.g. fungi, archaea, bacteria, Gram-positive bacteria, Gram-negative bacteria, *Alpha-*, *Beta-*, *Gamma-* and *Deltaproteobacteria*) (Zhou *et al.*, 2008). Hierarchical cluster analysis of the total functional genes was performed using the unweighted pair-wise average-linkage clustering algorithm (Eisen *et al.*, 1998) with R and VEGAN statistics packages.

A meta-analysis (Luo *et al.*, 2006) of the response of functional genes to increasing oil content was conducted. Noncontaminated samples (DQ1–3), medium-contaminated samples (DQ4–6) and high-contaminated samples (DQ7, 8) were combined separately and filtered with those genes appearing in all samples of each combination. The response ratio of gene i was calculated between medium- and noncontaminated samples and high- and noncontaminated samples as follows:

$$rr_i = \ln(\bar{x}_i/\bar{y}_i) \quad (i = 1, \ldots, n)$$

where $\bar{x}$ is the mean of normalized signal intensity in low- or high-contaminated samples; $\bar{y}$ is the mean of the normalized signal intensity in noncontaminated samples.

The variance (v) is

$$v_i = \frac{s_{x_i}^2}{m_{x_i}\bar{x}_i^2} + \frac{s_{y_i}^2}{m_{y_i}\bar{y}_i^2} \quad (i = 1, \ldots, n)$$

where s is the SD of gene i in noncontaminated samples (DQ1–3), medium-contaminated samples (DQ4–6) and high-contaminated samples (DQ7, 8); m is the number of gene i in noncontaminated samples (DQ1–3), medium-contaminated samples (DQ4–6) and high-contaminated samples (DQ7, 8).

The 90%, 95% and 99% confidence interval (CI) for the response ratio is

$$\mathrm{CI}_i = rr_i \pm \lambda\sqrt{v_i} \quad (i = 1, \ldots, n)$$

$\lambda = 1.64$, 1.96 and 2.58 for 90% CI, 95% CI and 99% CI, respectively.

The difference is significant only if 90% CI, 95% CI or 99% CI of a response variable does not overlap with zero.

Canonical correspondence analysis (CCA) was performed to identify the relationship between geochemical parameters and microbial functional genes using CANOCO for Windows version 4.5 (ter Braak & Smilauer, 1998). The functional-gene-normalized signal intensities were log transformed [$Y = \log(A \times X + B)$, $A = 1.0$, $B = 10.0$] and downweighted for a rare species for CCA. Monte Carlo tests were used to assess the significance of the environmental variables with 999 permutations.

Results

Microbial functional gene diversity

The oil content in contaminated soils ranged from 4.4 to 157 mg g^{-1} (Table 1). Such high contamination levels have been shown to influence the soil ecosystem, as Saterbak *et al.* (1999) reported a decrease in earthworm survival and seed germination at more than 10 mg hydrocarbon g^{-1} soil. Other geochemistry values varied considerably among the samples (Table 1).

The total number of functional genes decreased under oil contamination (Table 2). An average of 964 genes was detected in uncontaminated soils and 399 genes in contaminated soils. Also, diversity indices (both Simpson's reciprocal and Shannon–Weaver) indicated lower levels of functional gene diversity in the contaminated soils. Evenness was similar across all samples. Although samples

Table 1. Geochemical data of soil samples from Daqing oil field in north-east China

No.	Oil (mg g^{-1})	pH	Water (%)	TN (mg kg^{-1})	N (mg kg^{-1})	TP (mg kg^{-1})	P (mg kg^{-1})	OM (mg g^{-1})
DQ1	0	7.8	12.4	599.5	44.5	523.7	3.7	3.9
DQ2	0	7.7	10.3	311.6	66.8	985.8	5.3	9.3
DQ3	0	7.2	4.7	249.3	14.8	581.7	3.2	15.6
DQ4	4.4	7.9	21.3	–	–	–	–	30.4
DQ5	22.2	8.6	16.4	507.5	4.9	511.2	37.7	95.0
DQ6	49.1	7.7	7.3	750.0	50.0	580.0	11.8	106.8
DQ7	87.0	7.3	23.4	516.4	64.3	561.0	39.1	178.7
DQ8	156.8	7.5	12.3	800.1	69.3	378.8	30.5	217.7

TN, total nitrogen; N, nitrogen; TP, total phosphorus; P, phosphorous; OM, organic matter.

Table 2. Total number of functional genes detected, unique and overlapping genes and diversity indices

	DQ1	DQ2	DQ3	DQ4	DQ5	DQ6	DQ7	DQ8
% of genes detected								
DQ1	**26.3***	*36.3*†	*32.9*	*23.9*	*23.9*	*21.72*	*23.2*	*18.2*
DQ2		**26.6**	*32.7*	*25.4*	*22.6*	*21.2*	*26.4*	*19.6*
DQ3			**17.7**	*28.2*	*27.8*	*24.3*	*26.1*	*21.5*
DQ4				**14.8**	*25.8*	*27.5*	*26.6*	*23.7*
DQ5					**12.7**	*24.6*	*24.2*	*23.6*
DQ6						**11.0**	*28.2*	*27.9*
DQ7							**13.3**	*29.3*
DQ8								**10.6**
Total genes detected	1047	1115	730	479	419	355	459	283
Diversity indices								
Simpon's (1/*D*)	486.6	467.7	333.5	181.5	189.9	190.3	211.1	127.8
Shannon–Weaver *H*′	6.56	6.59	6.21	5.69	5.65	5.57	5.74	5.26
Evenness	0.94	0.94	0.94	0.92	0.94	0.95	0.94	0.93

*Values in bold represent unique genes in each sample.

†Values in italics represent genes overlapping between two samples.

DQ1–3 were all from uncontaminated sites, microbial heterogeneity was observed, as only 33–36% of detected genes were shared among all of the uncontaminated soils. An even lower overlap was observed among contaminated soils (24–29%), and between uncontaminated and contaminated soils (18–28%). Hierarchical cluster analysis of all functional genes was performed to generate a dendrogram to examine the relationship between the bacterial community structures. As shown in Fig. 1, samples fell into three groups: noncontaminated (DQ1–3), low- to medium-contaminated (DQ4–6) and high-contaminated (DQ7, 8), which is similar to the results observed with hydrocarbon-contaminated coastal sediments (Paisse *et al.*, 2008).

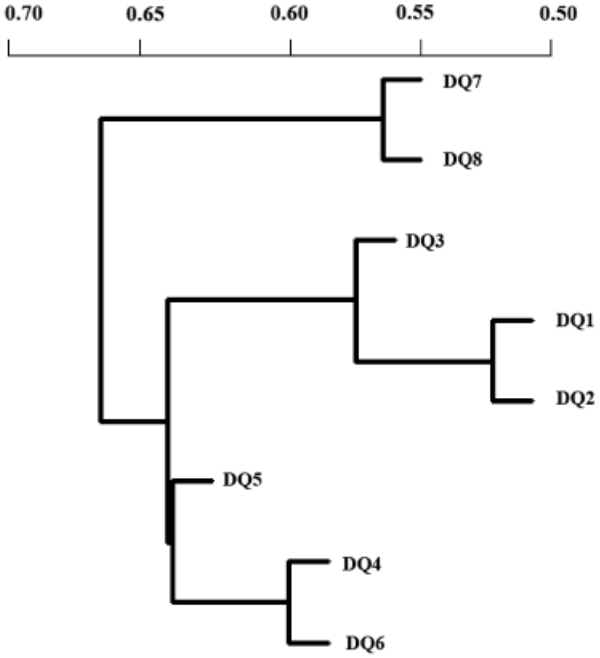

Fig. 1. Hierarchical cluster analysis of all functional genes. DQ1–3 were pristine soil samples; DQ4–8 were contaminated samples with increasing oil concentration. Samples grouped with oil contamination level.

Microbial community patterns

The richness of bacteria, archaea and fungi in the eight samples is shown in Fig. 2. Bacteria decreased to around 60% of the average uncontaminated level in low- to medium-contaminated soils, and around 40% in high-contaminated soils (Fig. 2a). Archaea decreased dramatically in most contaminated samples to 10% of the average uncontaminated level (Fig. 2b). Extensive changes in archaeal communities following exposure to oil were also reported in beach microcosms (Roling *et al.*, 2004a). Fungi in most contaminated samples decreased to about 80% of the average uncontaminated level (Fig. 2c).

The composition and abundance of *Actinobacteria*, *Firmicute* and *Proteobacteria* (*Alpha-*, *Beta-*, *Gamma-* and *Deltaproteobacteria*) were compared (Fig. 2a). *Alpha-* and *Gammaproteobacteria* were the dominant bacterial groups across all samples, representing 22.2–32.5% and 28.4–43.5% of the total bacterial abundance, respectively. A significant negative relationship was observed between oil content with *Gammaproteobacteria* ($r = -0.82$, $P = 0.007$), *Alphaproteobacteria* ($r = -0.72$, $P = 0.022$), *Deltaproteobacteria* ($r = -0.69$, $P = 0.030$), *Actinobacteria* ($r = -0.67$, $P = 0.033$) and *Betaproteobacteria* ($r = -0.63$, $P = 0.048$), but oil content did not significantly influence *Firmicutes*.

Key functional gene responses under oil contamination stress

The changes observed in each functional gene group were examined in detail (Fig. 3, Supporting Information,

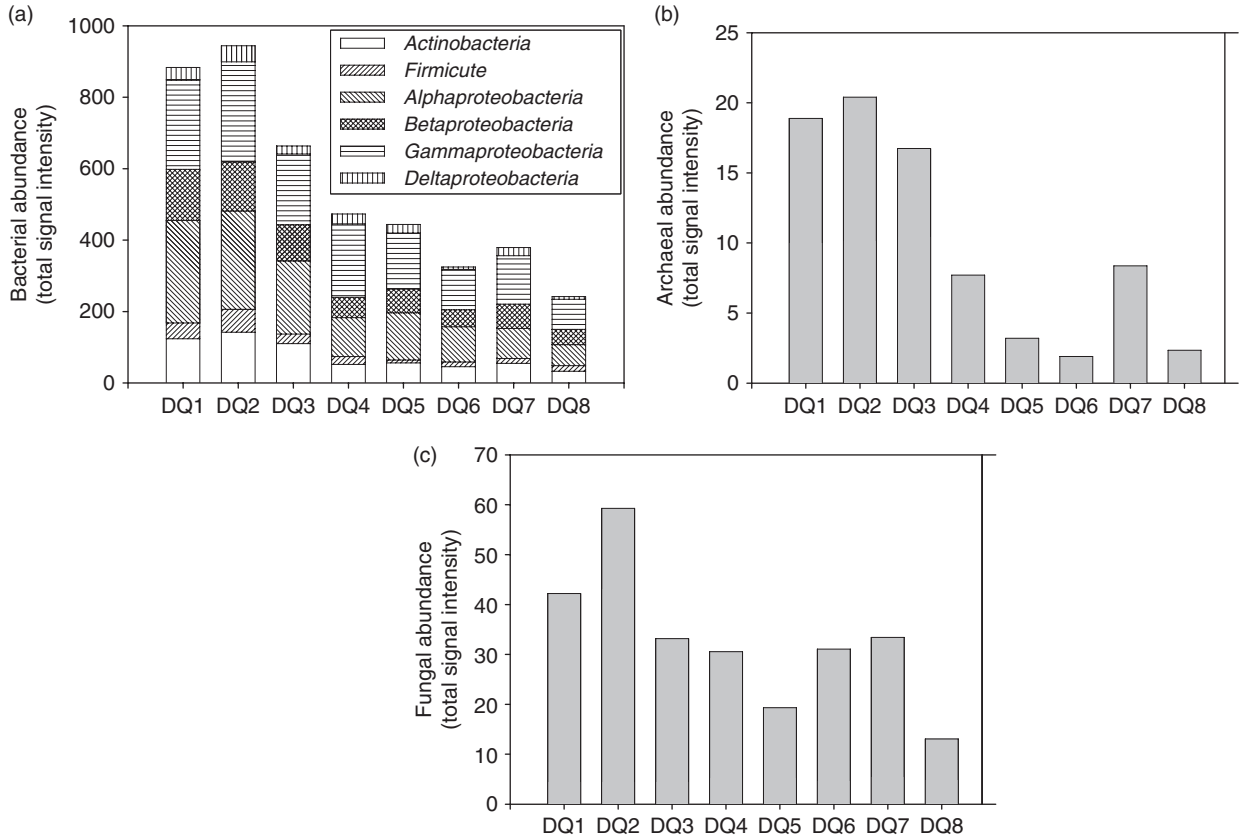

Fig. 2. Abundance of (a) bacteria (*Actinobacteria*, *Firmicute*, *Alpha-*, *Beta-*, *Gamma-* and *Deltaproteobacteria*), (b) archaea and (c) fungi based on measurement from the GeoChip hybridization. The *y*-axis represents the total signal intensity of genes derived from bacteria, archaea and fungi.

Figs S1–S4). Response ratios were calculated to reveal the gene changes in response to increasing oil contamination levels (Fig. 3). A positive value indicated an increase in gene abundance relative to the uncontaminated soil, whereas a negative value indicated the opposite. Three CIs, 90%, 95% and 99%, were used to examine the significance of changes. In total, 83 genes were present in all medium- and non-contaminated soils with response ratios in the range -1.87 to 1.48 (-0.28 average), while only 75 genes were common to all high- and noncontaminated soils with response ratios in the range of -2.12 to 0.86 (-0.61 average). For each functional gene category, a significant decrease in medium-contaminated soils was observed: 38% in carbon cycling, 33% in nitrogen cycling and 19% in organic contaminant degradation. In high-contaminated soils, a greater decrease was observed: 67% in carbon cycling, 71% in nitrogen cycling and 44% in organic contaminant degradation. The response ratios among high- to medium-contaminated soils ranged from -1.72 to 0.65 (-0.35 average) (Fig. S1). Many response ratios in the medium- or high-contaminated soils were below zero, which indicated the negative impact of contamination on soil microbial communities.

Several functional genes involved in carbon degradation and fixation dramatically decreased with increasing oil concentration, including genes encoding cellulase, laccase and Rubisco. However, none of the functional genes encoding chitinase showed a significant response to oil contamination. For nitrogen fixation-related functional genes, most *nifH* genes were not significantly influenced by medium contamination. However, oil contamination had a negative impact on most genes encoding urease. For nitrogen reduction genes, *nasA*, *norB* and *nosZ* decreased in medium-contaminated soils. The genes *narG*, *nirS*, *norB* and *nosZ* decreased in high-contaminated soils.

Several functional genes involved in organic contaminant degradation were at similar levels in contaminated soils, especially in medium-contaminated soils. Genes involved in degradation of catecholic PAH derivatives were stable in both medium- and high-contaminated soils. Some genes, such as *pcaG* and *menC*, even showed a significant increase in medium-contaminated soils. However, functional genes involved in benzoate, aniline and protocatechuate degradation were significantly influenced by high oil contamination.

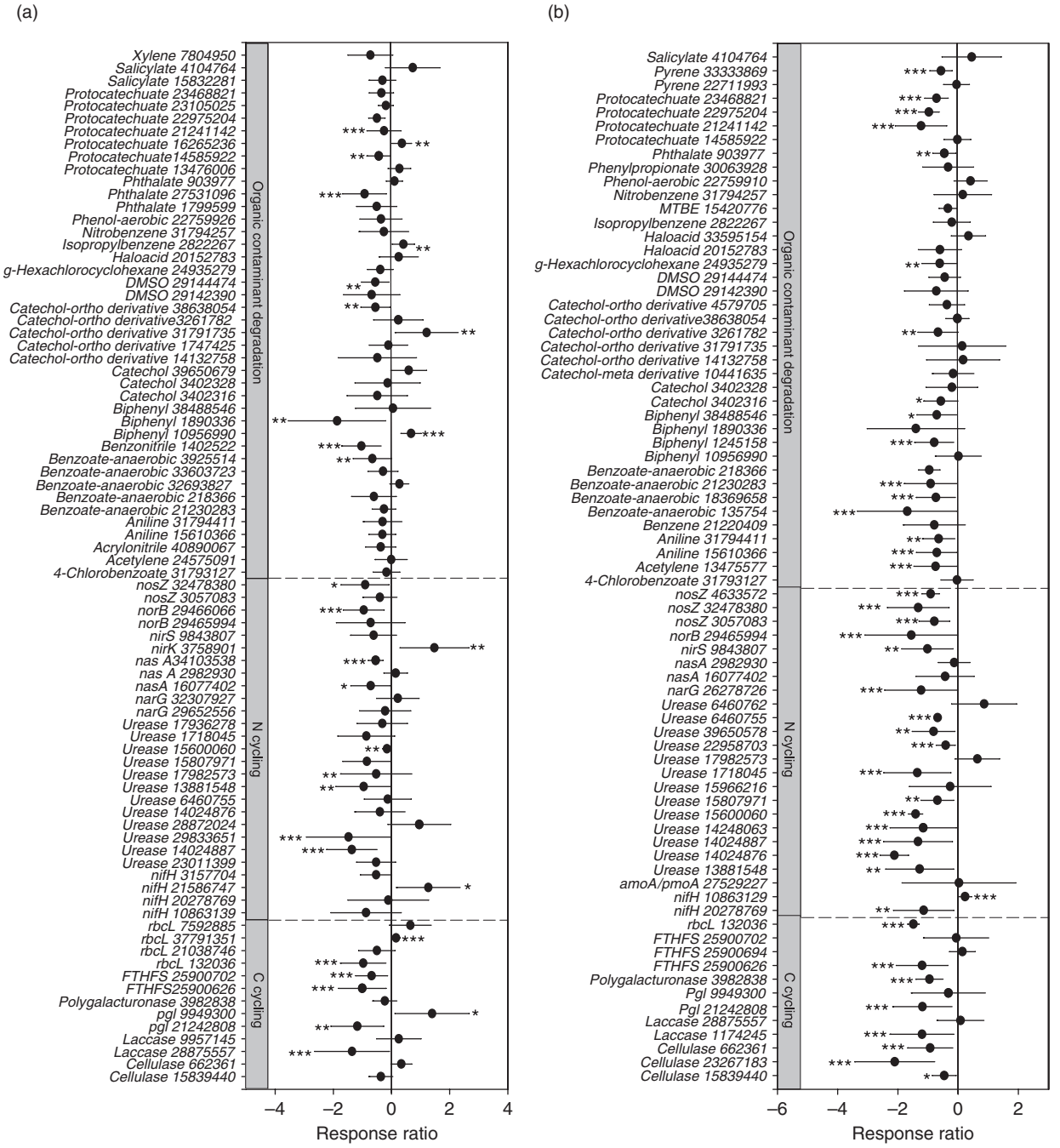

Fig. 3. Response ratios (mean ± SE for each gene) of (a) medium contamination (DQ4–6) to noncontaminated (DQ1–3) and (b) high contamination (DQ7–8) to noncontaminated (DQ1–3). *Significant at 90% CI; **significant at 95% CI; ***significant at 99% CI. Gene number is the protein ID number for each gene as listed in the GenBank database.

Relationships between community structure and environmental variables

CCA was performed to discern possible linkages between geochemical parameters and microbial community functional structure (Fig. 4). Only geochemical parameters that were significant were included in the CCA biplot. For all functional genes, 33.4% of the total variance could be explained by the first two constrained axes, with the first axis explaining 17.4%. Both the first canonical axis ($P = 0.018$) and the sum of all canonical axes ($P = 0.001$) were significant. Samples showed grouping patterns similar

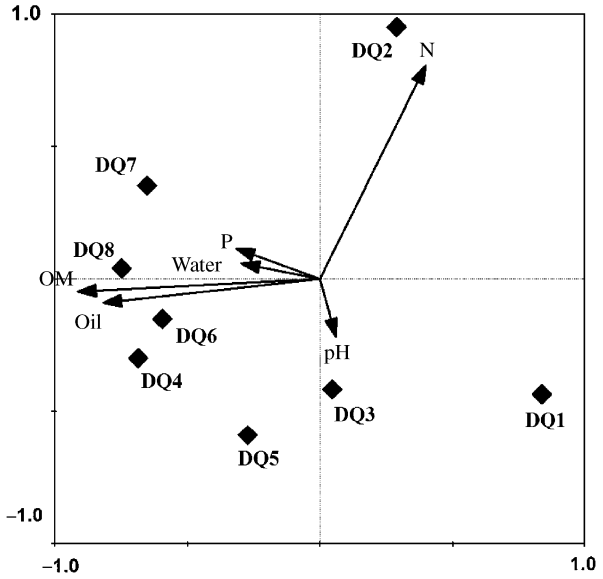

Fig. 4. CCA of GeoChip hybridization signal intensities and soil geochemical variables: oil concentration, pH, water content, organic matter (OM), nitrogen (N) and phosphorus (P). DQ1–3 are pristine soil samples; DQ4–8 are contaminated samples with increasing oil concentration.

to those observed with hierarchical clustering, with non-, low- to medium- and high-contaminated soils grouped together, respectively. The first axis was negatively correlated with oil concentration, soil organic matter, water content and soil available phosphorus and the second axis was positively correlated with soil available nitrogen and negatively with pH. Because oil accounted for the majority of total soil organic matter, it was considered that of all geochemical variables, soil available nitrogen and oil concentration had a stronger influence on microbial functional gene structure. Pristine soils were highly correlated with the nitrogen content. Contaminated soils, especially those with high contamination, were controlled by oil concentration.

Discussion

In this study, we examined changes in microbial functional genes along an oil contamination gradient of an oil field using a functional gene array, GeoChip, to gain an insight into *in situ* microbial community responses to oil contamination. The GeoChip has been shown to provide robust results in tracking hundreds to thousands of functional genes and to be a powerful approach to dissection of microbial functional community structure in complex environments (Leigh *et al.*, 2007; Yergeau *et al.*, 2007; Wu *et al.*, 2008; Zhou *et al.*, 2008; Liang *et al.*, 2009; Mason *et al.*, 2009; Waldron *et al.*, 2009; Wang *et al.*, 2009). This type of array is useful for monitoring environmental microbial community structure and linking genomic information with the key biogeochemical processes. Crude oil is a complex mixture of a diverse group of organic compounds. A large number of probes for genes involved in organic contaminant degradation are targeted by probes on the array, as well as functional genes in nitrogen, carbon and sulfur cycling. Meta-analysis and other statistical analyses indicated that changes in microbial functional diversity were associated with oil contamination in the studied area.

Organic chemical pollution has been reported to significantly affect microbial communities from a variety of environments. Oil-induced diminished microbial population diversity was observed using both culture-dependent and culture-independent approaches (Van Hamme *et al.*, 2003). In this study, both detected functional gene number and diversity decreased with increasing oil contamination level. The decrease in diversity may be due to the toxicity of some oil components or metabolic products (Cheung & Kinkle, 2001), or to selection for metabolic generalists (Roling *et al.*, 2004b). Until now, studies on archaeal community dynamics under oil contamination have mainly focused on some extreme or strictly anaerobic environments. Roling *et al.* (2004a) reported a negative effect on archaea by oil spills in beach sediments in laboratory experiments. Sandaa *et al.* (1999) also observed a decrease in archaeal numbers with increasing heavy-metal contamination in soils. In this study, we found a dramatic decrease in *in situ* archaea in oil-contaminated soils as indicated by the limited archaeal genes on the array. Although some archaea have the capacity to utilize petroleum hydrocarbon (Le Borgne *et al.*, 2008), their true functions in the field, especially in oil-contaminated aerobic environments, are not understood. Thus, further work is needed.

Several indigenous fungi such as *Trichoderma* sp., *Penicillium* sp., *Aspergillus* sp. and *Phanerochaete* sp. have been isolated from oil-contaminated soils and reported to be capable of degrading both aliphatic and aromatic hydrocarbons (Bokhary & Parvez, 1993; Bishnoi *et al.*, 2008; Husaini *et al.*, 2008). In this study, fungi appeared to be more resistant to oil contamination than bacteria and archaea, which may be due to their higher ability to colonize, penetrate and spread in various soils (Novotny *et al.*, 1999; Husaini *et al.*, 2008). Most fungal functional genes detected are involved in carbon degradation, encoding laccase, cellulase, chitinase and polygalacturonase. The genes mainly derived from *Piloderma byssinum*, *Trametes* sp., *Polyporus ciliatus*, *Pleurotus* sp., *Penicillium griseoroseum* and *Aspergillus aculeatus*, some of which are believed to catalyze PAH degradation by an extracellular ligninolytic enzyme system (Bezalel *et al.*, 1996; Pickard *et al.*, 1999).

Detecting microbial catabolic genes in the environment is a more direct way of linking microbial metabolic potential to changing environmental conditions (Junca & Pieper, 2004). For example, catabolic genes such as *ndoB*, *alkB* and *xylE* in the bulk soil were affected by different treatments,

while no detectable shift in the 16S rRNA gene composition occurred during a PAH phytoremediation field trial (Siciliano *et al.*, 2003). The abundance of functional genes, such as *nahAc*, *alkB* and *xylE*, has been used to estimate the toluene and naphthalene microbial degradation potential in petroleum hydrocarbon-contaminated sites (Tuomi *et al.*, 2004; Salminen *et al.*, 2008). In the current study, although the overall gene number and diversity decreased along the oil contamination gradient, some genes remained at the same level as in the original pristine soils regardless of the contamination level. This was especially true for functional genes involved in organic contaminant degradation, such as functional genes involved in the degradation of naphthalene, pyrene, biphenyl and catecholic PAH derivatives. An increase in the abundance of genes encoding catechol 2,3-dioxygenases in isolates was also reported with BTEX pollution in soil samples (Junca & Pieper, 2004). In crude oil-contaminated soils, there was a drastic decrease in alkanes and aromatic hydrocarbons with low volatility, which was largely due to *in situ* microbial degradation (Liang *et al.*, 2009). The high abundance of organic contaminant degradation genes detected in oil-contaminated soils indicated their potential importance in oil degradation.

Oil contamination influences soil microbial activities. Several functional genes involved in carbon and nitrogen cycling decreased in both medium- and high-contaminated soils, such as *rbcL* (132036), *pgl* (21242808), urease (14024887, 13881548, 1580797 and 15600060) and *nosZ* (32478380), indicating the negative effect of oil on soil microorganisms and their potential functions in carbon and nitrogen cycling. Also, in other studies, soil microbial urease and dehydrogenase activity showed a significant negative correlation with oil concentration (Megharaj *et al.*, 2000; Li *et al.*, 2005). Both functional gene and enzyme activities could be used as the most sensitive indicator of oil contamination. Also, some functional genes involved in nitrogen and carbon cycling and organic contaminant degradation showed an initial increase and then a decrease with increasing oil level, indicating a possible selection for those organisms at a low to medium contamination level, but toxicity at high contamination level.

Nitrogen limitation is always considered as one of the key factors that influence bioremediation of organic compounds (Atlas, 1981; Oh *et al.*, 2001). The optimal carbon/nitrogen ratio has been known to be 10 : 1 (Atlas, 1981). The ratio of oil content to total nitrogen is much higher in the site studied, about 40 to 200 : 1, than the optimal value. Organic nitrogen accounted for 87–99% of total nitrogen, and the available nitrogen [N, nitrate (NO_3^--N), nitrite (NO_2^--N) and ammonium (NH_4^+-N)] that microbial directly used for oil degradation only accounted for a really small portion. If the microbial activities in nitrogen cycling were reduced, as preliminarily evidenced by the decrease of functional genes, nitrogen limitation must be a critical issue for *in situ* bioremediation of oil contaminants in oil fields.

The *in situ* response of oil-stressed microbial functional diversity indicated that the potential metabolic capacity of microbial communities should be considered when designing site-specific approaches for bioremediation. Most studies on functional genes in natural environments, including this study, however, focus mainly on the characterization of the presence and/or the abundance of a family of catabolic genes rather than gene expression. More practical information can be obtained if workable methods for mRNA analysis can be developed (Van Hamme *et al.*, 2003). Recently, a new approach, termed whole-community RNA amplification, has been developed to provide sufficient amounts of mRNA from environmental samples for microarray analysis (Gao *et al.*, 2007). Further work is needed to examine *in situ* microbial functional gene expression under contamination stress via RNA analysis. In addition, more sampling efforts should be made to study *in situ* microbial functional diversity and ecological functions under oil contamination stress in oil fields.

In conclusion, the present study profiles *in situ* microbial functional diversity changes along an oil contamination gradient across an oil field. The microbial functional gene numbers and diversity decreased with increasing oil concentration and exhibited similar patterns under similar contamination levels. Soil microbial functional community structure varied in response to contamination stress. Oil content and soil available nitrogen were found to be important factors influencing the microbial community structure. The results not only extend our understanding of the microbial ecology of contaminated environments but also provide potential information for on-site management and remediation measure.

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 the Genomics: GTL program through the Virtual Institute of Microbial Stress and Survival (VIMSS; http://vimss.lbl.gov) and by the Oklahoma Center for the Advancement of Science and Technology under Oklahoma Applied Research Support Program.

References

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.

Atlas R (1981) Microbial degradation of petroleum hydrocarbons: and environmental perspective. *Microbiol Rev* **45**: 180–209.

Bezalel L, Hadar Y, Fu P, Freeman J & Cerniglia C (1996) Metabolism of phenanthrene by the white rot fungus *Pleurotus ostreatus*. *Appl Environ Microb* **62**: 2547–2553.

Bishnoi K, Kumar R & Bishnoi N (2008) Biodegradation of polycyclic aromatic hydrocarbons by white rot fungi *Phanerochaete chrysosporium* in sterile and unsterile soil. *J Sci Ind Res* **67**: 538–542.

Bokhary H & Parvez S (1993) Fungi from petroleum-contaminated Saudi Arabian soils. *Arid Soil Res Rehab* **7**: 191–195.

Cheung P & Kinkle B (2001) *Mycobacterium* diversity and pyrene mineralization in petroleum-contaminated soils. *Appl Environ Microb* **67**: 2222–2229.

Eisen M, Spellman P, Brown P & Botstein D (1998) Cluster analysis and display of genome-wide expression patterns. *P Natl Acad Sci USA* **95**: 14863–14868.

Gao H, Yang Z, Gentry T, Wu L, Schadt C & Zhou J (2007) Microarray-based analysis of microbial community RNAs by whole-community RNA amplification. *Appl Environ Microb* **73**: 563–571.

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 Z, Gentry T, Schadt C *et al.* (2007) GeoChip: a comprehensive microarray for investigating biogeochemical, ecological and environmental processes. *ISME J* **1**: 67–77.

Head I, Jones D & Roling W (2006) Marine microorganisms make a meal of oil. *Nat Rev Microbiol* **4**: 173–182.

Huesemann M (1995) Predictive model for estimating the extent of petroleum hydrocarbon biodegradation in contaminated soils. *Environ Sci Technol* **29**: 7–18.

Husaini A, Roslan H, Hii K & Ang C (2008) Biodegradation of aliphatic hydrocarbon by indigenous fungi isolated from used motor oil contaminated sites. *World J Microb Biot* **24**: 2789–2797.

Junca H & Pieper D (2004) Functional gene diversity analysis in BTEX contaminated soils by means of PCR-SSCP DNA fingerprinting: comparative diversity assessment against bacterial isolates and PCR-DNA clone libraries. *Environ Microbiol* **6**: 95–110.

Laurie A & Lloyd-Jones G (2000) Quantification of *phnAc* and *nahAc* in contaminated New Zealand soils by competitive PCR. *Appl Environ Microb* **66**: 1814–1817.

Leahy J & Colwell R (1990) Microbial degradation of hydrocarbons in the environment. *Microbiol Rev* **54**: 305–315.

Le Borgne S, Paniagua D & Vazquez-Duhalt R (2008) Biodegradation of organic pollutants by halophilic bacteria and archaea. *J Mol Microb Biotech* **15**: 74–92.

Leigh M, Pellizari V, Uhlik O, Sutka R, Rodrigues J, Ostrom N, Zhou J & Tiedje J (2007) Biphenyl-utilizing bacteria and their functional genes in a pine root zone contaminated with polychlorinated biphenyls (PCBs). *ISME J* **1**: 134–148.

Li H, Chen G, Yang T & Zhang C (2005) Impact of petroleum-containing waste water irrigation on microbial population and enzyme activities in paddy soil of Shenfu irrigation area. *Chin J Appl Ecol* **16**: 1355–1359.

Liang Y, Nostrand J, Wang J, Zhang X, Zhou J & Li G (2009) Microarray-based functional gene analysis of soil microbial communities during ozonation and biodegradation of crude oil. *Chemosphere* **75**: 193–199.

Lu R (1999) *Soil Agricultural Chemical Analysis*. China Agricultural Science and Technology Press, Nanjing.

Luo Y, Hui D & Zhang D (2006) Elevated CO_2 stimulates net accumulations of carbon and nitrogen in land ecosystems: a meta-analysis. *Ecology* **87**: 53–63.

Mason U, DiMeo-Savoie A, Van Nostrand J, Zhou J, Fisk M & Giovannoni S (2009) Prokaryotic diversity, distribution, and preliminary insights into their role in biogeochemical cycling in marine basalts. *ISME J* **3**: 231–242.

Megharaj M, Singleton I, McClure NC & Naidu R (2000) Influence of petroleum hydrocarbon contamination on microalgae and microbial activities in a long-term contaminated soil. *Arch Environ Con Tox* **38**: 439–445.

Novotny C, Erbanova P, Sasek V, Kubatova A, Cajthaml T, Lang E, Krahl J & Zadrazil F (1999) Extracellular oxidative enzyme production and PAH removal in soil by exploratory mycelium of white rot fungi. *Biodegradation* **10**: 159–168.

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.

Park J & Crowley D (2006) Dynamic changes in *nahAc* gene copy numbers during degradation of naphthalene in PAH-contaminated soils. *Appl Microbiol Biot* **72**: 1322–1329.

Pickard M, Roman R, Tinoco R & Vazquez-Duhalt R (1999) Polycyclic aromatic hydrocarbon metabolism by white rot fungi and oxidation by *Coriolopsis gallica* UAMH 8260 laccase. *Appl Environ Microb* **65**: 3805–3809.

Roling W, de Brito I, Swannell R & Head I (2004a) Response of archaeal communities in beach sediments to spilled oil and bioremediation. *Appl Environ Microb* **70**: 2614–2620.

Roling W, Milner M, Jones D, Fratepietro F, Swannell R, Daniel F & Head I (2004b) Bacterial community dynamics and hydrocarbon degradation during a field-scale evaluation of bioremediation on a mudflat beach contaminated with buried oil. *Appl Environ Microb* **70**: 2603–2613.

Salminen J, Tuomi P & Jorgensen K (2008) Functional gene abundances (*nahAc*, *alkB*, *xylE*) in the assessment of the efficacy of bioremediation. *Appl Biochem Biot* **151**: 638–652.

Sandaa R, Enger O & Torsvik V (1999) Abundance and diversity of Archaea in heavy-metal-contaminated soils. *Appl Environ Microb* **65**: 3293–3297.

Saterbak A, Toy R, Wong D, McMain B, Williams M, Dorn P, Brzuzy L, Chai E & Salanitro J (1999) Ecotoxicological and analytical assessment of hydrocarbon-contaminated soils and application to ecological risk assessment. *Environ Toxicol Chem* **18**: 1591–1607.

Siciliano S, Germida J, Banks K & Greer C (2003) Changes in microbial community composition and function during a polyaromatic hydrocarbon phytoremediation field trial. *Appl Environ Microb* **69**: 483–489.

Sotsky J, Greer C & Atlas R (1994) Frequency of genes in aromatic and aliphatic hydrocarbon biodegradation pathways within bacterial-populations from Alaskan sediments. *Can J Microbiol* **40**: 981–985.

ter Braak CJF & Smilauer P (1998) *CANOCO Reference Manual and User's Guide to CANOCO for Windows: Software for Canonical Community Ordination, Version 4*. Microcomputer Power, New York.

Totsche K, Jann S & Kogel-Knabner I (2006) Release of polycyclic aromatic hydrocarbons, dissolved organic carbon, and suspended matter from disturbed NAPL-contaminated gravelly soil material. *Vadose Zone J* **5**: 469–479.

Tuomi P, Salminen J & Jorgensen K (2004) The abundance of *nahAc* genes correlates with the C-14-naphthalene mineralization potential in petroleum hydrocarbon-contaminated oxic soil layers. *FEMS Microbiol Ecol* **51**: 99–107.

Van Hamme J, Singh A & Ward O (2003) Recent advances in petroleum microbiology. *Microbiol Mol Biol R* **67**: 503–549.

Waldron P, Van Nostrand J, Watson D, He Z, Wu L, Jardine P, Hazen T & Zhou J (2009) GeoChip analysis of subsurface microbial communities impacted by heavy metal and nitrate contamination. *Environ Sci Technol* **43**: 3529–3534.

Wang F, Zhou H, Meng X *et al.* (2009) GeoChip-based analysis of metabolic diversity of microbial communities at the Juan de Fuca Ridge hydrothermal vent. *P Natl Acad Sci USA* **106**: 4840–4845.

Wehrer M & Totsche K (2008) PAH release from tar-oil contaminated soil material in response to forced environmental gradients: implications for contaminant transport. *Eur J Soil Sci* **59**: 50–60.

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 Microb* **72**: 4931–4941.

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 Microb* **74**: 4516–4529.

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. *B Environ Contam Tox* **58**: 667–672.

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.

Zhou J, Bruns M & Tiedje J (1996) DNA recovery from soils of diverse composition. *Appl Environ Microb* **62**: 316–322.

Zhou J, Kang S, Schadt C & Garten C (2008) Spatial scaling of functional gene diversity across various microbial taxa. *P Natl Acad Sci USA* **105**: 7768–7773.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Fig. S1. Response ratios (RR, mean ± SE for each gene) of high contamination (DQ7–8) to medium contamination (DQ4–6).

Fig. S2. Linkage between the diversity of functional genes in carbon cycling: (a) cellulase, (b) laccase, (c) chitinase and (d) *rbcL*.

Fig. S3. Linkage between the diversity of the functional genes in nitrogen cycling: (a) *nifH*, (b) urease, (c) *nar*, (d) *nir*, (e) *nas*, *nor* and *nos*.

Fig. S4. Linkage between the diversity of the functional genes in organic contaminant degradation: (a) catechol, (b) protocatechuate, (c) naphthalene, (d) pyrene, (e) benzene and (f) biphenyl.

Please note: Wiley-Blackwell is not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.