Microbial gene functions enriched in the Deepwater Horizon deep-sea oil plume

Zhenmei Lu[1,2], Ye Deng[2], Joy D. Van Nostrand[2], Zhili He[2], James Voordeckers[2], Aifen Zhou[2], Yong-Jin Lee[2], Olivia Mason[3], Eric Dubinsky[3], Krystle Chavarria[3], Lauren Tom[3], Julian Fortney[3], Regina Lamendella[3], Janet K. Jansson[3], Patrik D'haeseleer[3], Terry C. Hazen[3], Jizhong Zhou[2, 3,4,*]

[1]College of Life Sciences, Zhejiang University, Hangzhou 310058, China; [2]Institute for Environmental Genomics, Department of Botany and Microbiology, University of Oklahoma, Norman, OK 73019, USA; [3]Earth Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720; [4]Department of Environmental Science and Engineering, Tsinghua University, Beijing 100084, China

[*]Corresponding author: Dr. Jizhong Zhou
Institute for Environmental Genomics
University of Oklahoma
Norman, OK 73019
Phone: (405) 325-6073
Fax: (405) 325-7552
Email: jzhou@ou.edu

ABSTRACT

The Deepwater Horizon oil spill in the Gulf of Mexico is the deepest and largest offshore spill in U.S. history and its impacts on marine ecosystems are largely unknown. Here, we showed that the microbial community functional composition and structure were dramatically altered in a deep-sea oil plume resulting from the spill. A variety of metabolic genes involved in both aerobic and anaerobic hydrocarbon degradation were highly enriched in the plume compared to outside the plume, indicating a great potential for *intrinsic* bioremediation or natural attenuation in the deep-sea. Various other microbial functional genes relevant to carbon, nitrogen, phosphorus, sulfur and iron cycling, metal resistance, and bacteriophage replication were also enriched in the plume. Together, these results suggest that the indigenous marine microbial communities could play a significant role in biodegradation of oil spills in deep-sea environments.

Key words: Oil spill; deep-sea plume; microbial community; metagenomics, functional gene arrays, GeoChip

INTRODUCTION

On April 20, 2010, a massive oil leak occurred in the Gulf of Mexico's Mississippi Canyon area at a depth of 1,544 m releasing approximately 4.9 million barrels of crude oil into the deep ocean before the wellhead was finally capped on July 15, 2010 (The Federal Interagency Solutions Group, Oil Budget Calculator Science and Engineering Team, November 2010). Chemical dispersants, including COREXIT EC9500A and COREXIT EC9527A, were used on site at one of the highest rates in history to accelerate oil dispersal. A deep-water oil plume was initially detected at a depth of 1,000-1,200 m below the surface (Camilli *et al*., 2010; Hazen *et al*., 2010; Mascarelli, 2010a), but at last account (Mascarelli, 2010b) could no longer be detected, presumably as a result of dispersion and microbial degradation (OSAT, 2010). Significant environmental differences in the deep-sea of Gulf of Mexico from other historic offshore oil spills present an urgent need to better understand the fate and impacts of the oil in this specific habitat (Kerr *et al*., 2010a; Kerr *et al*., 2010b).

In marine ecosystems, microorganisms are known to play predominant roles in degradation of oil contaminants (Head *et al*., 2003; Larter *et al*., 2003). Therefore, it was expected that the indigenous microbial communities would play a significant role in degradation of the deep oil plume. This hypothesis was supported by two recent studies that explored the microbial and chemical properties of samples collected from the deep oil plume (Camilli *et al*., 2010; Hazen *et al*., 2010). Hazen *et al*. (2010) used a combination of molecular, chemical and physiological approaches to investigate the microbial and chemical composition in the deep-sea plume compared to uncontaminated water from the same depth outside the plume. They demonstrated that the oil depletion was due to a combination of mixing, dispersion and biodegradation by microbes residing in the deep-sea (Hazen *et al*., 2010).

In this study, samples from the deep-sea plume, oil-contaminated seawater (hereafter referred to as "oil plume" in the following text) and non-plume controls (seawater samples at

same depth that were not contaminated with oil) were analyzed with a functional gene microarray, the GeoChip 4.0 (Hazen *et al*., 2010), to address the following questions: (i) How did the oil contamination affect the marine microbial community functional composition and structure? (ii) How did different microbial functional genes involved in key microbial processes shift in response to the oil spill? (iii) Were functional genes specific to hydrocarbon degradation processes enriched in the oil plume? Our results indicated that the oil spill dramatically altered microbial community functional structure, the marine microbial communities present were metabolically diverse, and that these communities were able to respond to the oil spill.

Materials and Methods

The following is the summary of methods used in this study. More detailed information is provided in the Supplemental Data A.

Sample description

Between May 27 and June 2, 2010, seawater samples were collected from the Gulf of Mexico during two monitoring cruises on the R/V Ocean Veritas and R/V Brooks McCall (Table S1) as previously described (Hazen *et al*., 2010). Briefly, two colored dissolved organic matter (CDOM) WETstar fluorometers (WET Labs, Philomath, OR) were attached to a CTD sampling rosette (Sea-Bird Electronics Inc., Bellevue, WA) and used to detect the presence of oil. The fluorometer results were subsequently confirmed by laboratory hydrocarbon analysis. Niskin bottles attached to the CTD rosette were used to capture water samples at various depths with detected hydrocarbons. Eight samples (BM053, BM054, BM057, BM058, BM064, OV201, OV401 and OV501) from the MC252 dispersed oil plume, and five samples (OV003, OV004, OV009, OV013, OV014) from non-plume at depth of 1099-1219m were analyzed in this study.

To better define the geochemical properties of the plume and non-plume samples, two sets of variables were measured: (i) seawater variables (dissolved oxygen (DO), temperature, small particle counts, total ammonia nitrogen (TAN), nitrite (NO_2-N), total Iron (Tot Fe), Ortho-phosphate (PO_4-P), and acridine orange direct count (AODC)), and (ii) oil composition variables (Fluorometer detection of oil, benzene, toluene, ethylbenzene, isopropylbenzene, n-propylbenzene, 1,3,5-trimethylbenzene, tert-butylbenzene, 1,2,4-trimethylbenzene, sec-butylbenzene, p-isopropyltoluene, n-butylbenzene, naphthalene, o-xylene, m, p-xylenes) (Hazen *et al.*, 2010).

DNA amplification and labeling

Approximately 100 ng of DNA that was previously extracted from the samples (Hazen *et al*., 2010) was amplified using a modification of the Templiphi kit (GE Healthcare, Piscataway, NJ). The amplified DNA (2 µg) was then labeled with Cy3 using random primers and the Klenow fragment of DNA polymerase I (Wu *et al*., 2006), and then purified and dried in a SpeedVac (45°C, 45 min; ThermoSavant) before hybridization.

GeoChip 4.0 hybridization and data pre-processing

The GeoChip 4.0, containing 83,992 50-mer oligonucleotide probes targeting 152,414 genes in 410 gene categories for different microbial functional and biogeochemical processes, was synthesized by NimbleGen (Madison, WI). All hybridizations were carried out at 42°C with 40% formamide for 16 h on a MAUI hybridization station (BioMicro, Salt Lake City, UT). After hybridization, the arrays were scanned (NimbleGen MS200) at a laser power of 100%. Signal intensities were measured based on scanned images, and spots with signal-to-noise ratios (SNR) lower than 2 were removed prior to statistical analysis as described previously (He *et al*., 2010).

Statistical analysis

Pre-processed GeoChip data were further analyzed with different statistical methods: (i)

Microbial diversity index, the two-tailed t-test, and response ratio (RR) (Luo *et al.*, 2006); (ii) Hierarchical clustering for microbial community structure and composition (de Hoon *et al.*, 2004~~)~~; (iii) Analysis of similarity (ANOSIM), permutational multivariate analysis of variance using distance matrices (Adonis), and multi response permutation procedure (MRPP) analysis of differences of microbial communities (Anderson, 2001); (iv) Canonical correspondence analysis (CCA) for linking microbial communities to environmental variables (Ramette and Tiedje, 2007; Zhou *et al.*, 2008); (v) partial CCA for co-variation analysis of wellhead distance and environmental variables (variation partitioning analysis, VPA). Details for all methods are provided in the Supplementary Information.

Results

Functional gene changes in response to oil spill

To assess the dynamic changes of microbial communities in response to oil spill, microbial community functional composition and structure was analyzed using functional gene arrays (GeoChip 4.0). Significantly more functional genes ($p < 0.01$) were detected in the oil plume samples than non-plume control (Table S2). The overall microbial functional diversity was also significantly ($p<0.01$) higher in the plume samples based on Shannon-Weiner (H') and Simpson's (1/D) indices. Consistent with geochemical ordination patterns, hierarchical clustering analysis showed that all plume samples were clustered together and well separated from non-plume samples (Fig. 1 and S1), as also shown for the microbial communities at a phylogenetic level (Hazen *et al.*, 2010). However, considerable variability in functional gene distribution was observed among different samples and some functional genes were common to all samples while others were unique to oil plume samples (Fig. 1). For example, Group 6, with 1439 or 20.14% of all genes detected, largely involved in organic remediation, carbon degradation, denitrification, sulfate reduction, metal resistance, and

stress response were generally detected in all samples. Group 1, 2, 10 and 17, with 2.2%, 3.9%, 20.5%, and 10.8% of all genes detected, were mainly detected in the plume samples (Fig. 1). In addition, the microbial community functional structure was significantly ($p<0.05$) different between the plume and non-plume samples as revealed by three complimentary non-parametric multivariate statistical tests (ANOSIM, adonis, and MRPP) (Table 1).

Oil as a predominant factor shaping microbial community functional structure

Canonical correspondence analysis (CCA) was performed to determine the most significant environmental variables shaping microbial community structure. Based on variance inflation factors (VIFs), seven variables were selected: DO, temperature, total volatile hydrocarbon [HC], total extractable petroleum HC, fluorometer detection of oil, phosphate, and iron. The specified CCA model was significant (p=0.026). Of these, the total volatile HC, extractable petroleum HC, fluorometer detection of oil, and DO were most significantly correlated with plume samples (Fig. 2). To separate the effects of seawater geochemical variables, geographic distance and oil composition on microbial community structure, a CCA-based variation partitioning analysis (VPA) (Ramette and Tiedje, 2007; Zhou *et al*., 2008) was performed. Seawater geochemical variables, oil composition and wellhead distance showed a significant correlation (p=0.041) with the functional gene structure of the community. Oil composition explained substantially more variations (48.34%, p=0.03) than seawater variables (21.76%, p=0.017), whereas distance independently explained 9.1% (p=0.43) of the observed variation (Fig. 3). About 28% of the community functional variation based on GeoChip data remained unexplained by the above selected variables, which is significantly lower than those observed in other systems such as soils (Ramette and Tiedje, 2007; Zhou *et al*., 2008). These results indicate that oil contaminants could be a dominant factor shaping microbial community functional structure, and potentially regulating

associated microbial functional processes.

Oil spill stimulated increase in functional genes for hydrocarbon degradation

A substantial number of genes involved in hydrocarbon degradation were detected in the oil plume samples (Hazen *et al*., 2010), especially those involved in degrading alkanes, alkynes and cycloalkanes, BTEX and related aromatics, chlorinated aromatics, heterocyclic aromatics, nitroaromatics, polycyclic aromatics and aromatic carboxylic acids. For example, gene *alkB* encoding alkane 1-monooxygenase, a key enzyme responsible for the initial oxidation of inactivated alkanes, showed a significantly ($p<0.05$) higher abundance with 19 to 26 genes detected in the oil contaminated samples and 11 to 15 detected in the non-oil contaminated samples. The *alkB* genes derived from *Rhodospirillum centenum* SW, *Bdellovibrio bacteriovorus* HD100, *Prauserella rugosa*, *Roseobacter* sp. CCS2, *Mycobacterium bovis* AF2122/97, *Bacillus* sp. BTRH40, *Gordonia* sp. Cg and *Rhodococcus* sp. RHA1 appeared to be dominant in all oil plume samples (Fig. S2).

GeoChip analysis also detected many aerobic PAH degradation genes from a variety of microorganisms (Fig. 4 and S3). PAH degradation genes were more abundant in the plume samples while some were unique to the plume samples. Although oxygen was still present in the plume samples (Camilli *et al*., 2010; Hazen *et al*., 2010), the gene *bbs* (beta-oxidation of benzylsuccinate) for anaerobic toluene degradation was also enriched in plume samples. These *bbs* genes were derived from putative E-phenylitaconyl-CoA hydratase of *Azoarcus* sp. EbN1 and *Thauera aromatic*, and benzylsuccinyl-CoA dehydrogenase of *Azoarcus* sp. EbN1 (Fig. 5).

Shifts of the genes involved in key biogeochemical cycling processes

Carbon. Among the carbon cycling genes detected, 798 genes involved in the degradation of complex carbon compounds such as starch, hemicellulose, cellulose, chitin, lignin and aromatics showed positive hybridization signals. Most of these genes (e.g., *pula*, *xylA*, *xynA*,

lip, *limEH*, and *vanA*) showed significantly ($p < 0.05$) higher abundances in plume than in non-plume samples (Fig. S4). These types of genes could also be important in degradation of various oil components and their intermediates.

In this study, 9 to 14 *mcrA* genes encoding the alpha subunit of methyl coenzyme M reductase, and 5 to 8 *pmoA* genes for methane monooxygenase were detected in the plume samples. Specifically, *mcrA* genes from *Methanococcus aeolicus* Nankai-3, *Methanoculleus marisnigri* JR1 and *Methanocorpusculum labreanum* Z were detected in all of the oil plume samples, but most of them were from uncultured microorganisms. Significantly ($p<0.05$) higher signal intensities were observed for *mcrA* in the plume than non-plume samples (Fig. S5). However, no significant differences were found for *pmoA* and *mmoX* (particulate methane monooxygenase) between plume and non-plume samples.

Nitrogen. Petroleum generally contains about 0.1-2% nitrogen, and given the large quantities of oil involved it may act as a N pool in this ecosystem. Interestingly, *nasA* (nitrate reductase) and *nir* (nitrite reductase) for assimilatory N reduction and *gdh* (glutamate dehydrogenase) for ammonia assimilation exhibited significantly ($p < 0.05$ or 0.01) higher signal intensities in plume samples (Fig. 6). The observed stimulation of N assimilation processes could be due to an increase of microbial biomass (Hazen *et al*., 2010). However, no significant differences were observed for other N-cycling genes, e.g. nitrification, denitrification and N fixation (Fig. 6).

Sulfur. Sulfite reduction genes were highly abundant in the deep-sea plume: 81 to 102 *dsrA/B* genes for dissimilatory sulfite reductase, and 8 to 12 *AprA* genes for dissimilatory adenosine-5'-phosphosulfate (APS) reductase were detected with significantly ($p<0.05$) higher abundance in the plume than in non-plume samples (Fig. S6). Microbial populations similar to *Alkalilimnicola ehrlichei* MLHE-1, *Chlorobium ferrooxidans* DSM 13031, *Clostridium leptum* DSM 753, *Desulfomicrobium thermophilum*, *Pyrobaculum calidifontis*

JCM 11548, *Thermodesulforhabdus norvegica*, *Magnetococcus* sp. MC-1, *Pyrobaculum aerophilum* str. IM2, *Alkalilimnicola ehrlichei* MLHE-1, *Desulfohalobium retbaense* DSM 5692, sulfate-reducing bacterium QLNR1 and *Syntrophobacter fumaroxidans* MPOB were frequently detected in each sample, while most of the genes detected were from uncultured microorganisms (e.g, sulfate-reducing bacteria, SRB) from various environments. The results suggest that sulfate reduction could be enhanced when coupled with hydrocarbon degradation.

Phosphorus and iron reduction. Since phosphorus is often a limiting factor for oil bioremediation, it is essential to understand phosphorus cycling in marine ecosystems. Genes encoding exopolyphosphatase (*ppx*) for inorganic polyphosphate degradation and phytase for phytate degradation were detected with significantly ($p<0.01$ and $p<0.05$, respectively) increased abundances in plume samples (Fig. S7). These results suggested that organic phosphorus release could be stimulated by oil contamination. In addition, higher ($p<0.1$) signal intensities for 61 detected cytochrome C genes were observed in plume samples (Fig S8), suggesting that hydrocarbon degradation coupled with metal reduction could occur in the deep water.

Metal resistance. A substantial number (917) of the genes involved in resistance to various metals were detected, many of which showed significantly ($p<0.05$) increased abundance in plume samples (Fig. S9). Genes encoding reductases for As (*arsC*) and Hg (*mer*), efflux transporters for Cd (*cadA*), Cu, Co, and Zn (*czcA* and *czcD*), Cr (*ChrA*), Cu (*copA*), Hg (*merT*), Ag (*silC*), and Zn (*zntA*), and the proteins involved in Te resistance (*terC*, *terD* and *terZ*) were more ($p<0.05$ or 0.01) abundant in the plume samples.

Bacteriophages were also significantly stimulated

In total, 52 bacterial phage genes associated with host recognition, lysis, replication, and structure were observed in all samples. The signal intensities for many of the genes

involved in replication were significantly ($p<0.05$) higher in the plume than non-plume samples (Fig. 7), supporting the suggestion by Head *et al*. (2006) that bacteriophages could be an important factor for intrinsic bioremediation of hydrocarbons.

Discussion

The Deepwater Horizon oil spill in the Gulf of Mexico was one of the worst environmental disasters in U.S. history. The impact of an oil spill of such an unprecedented magnitude and depth on marine ecosystems is largely unknown. Using the GeoChip-based high throughput microarray technology, we showed that diverse microbial functional groups (a group of genes involved in certain functional processes), including those important to hydrocarbon degradation, carbon metabolism, methanogenesis, nitrogen assimilation, sulfate reduction, phosphorus release, metal resistance, and bacteriophage replication, were more highly represented in the oil plume samples than in non-plume samples from the same depth. Also, the changes in community functional structure were highly correlated to the changes in geochemistry, with oil being the predominant factor shaping the functional composition and structure of the microbial communities. Our results support the phylogeny-based study by Hazen *et al*. (2010) that the deep-sea marine microbial communities underwent a dynamic change in response to the oil spill and associated geochemical changes. Our results are also consistent with previous studies of oil spill and petroleum contamination (Bordenave *et al*., 2007; Harayama *et al*., 2004; Head *et al*., 2006), which showed that microorganisms able to utilize hydrocarbons became dominant in oil-contaminated sites. Such functional gene information is useful for assessing the impacts of oil spills and should facilitate design of appropriate strategies and approaches to deal with petroleum contamination.

The clean-up of the deep sea oil plume will primarily depend on the indigenous microbes present in this environment since current technology does not allow removing the dispersed oil and gas at such great depths. One of the critical environmental questions is whether

microorganisms for degrading various hydrocarbons exist in the community and whether they respond to oil spill. Our GeoChip results indicated that many functional genes/populations involved in both aerobic and anaerobic degradation of various oil components are detected and/or enriched in the oil plume, indicating that the indigenous hydrocarbon-degrading populations are capable of responding to the oil spill. For example, *alkB* for alkanes, *Xamo* for alkene, genes *bco*, *ohbAB*, *GCoADH* and *pimF* for benzoate, genes *mdlA*, *mdlB* and *mdlC* for mandelate, genes *Apc*, *catB* for BTEX metabolic pathway exhibited a significantly ($p<0.05$) higher abundance in oil plume than in the non-oil plume. The changes in relative abundance of these genes/populations were significantly correlated with the concentration of various oil contaminants in the samples (Hazen *et al*., 2010). Especially, several genes for PAH degradation were enriched in the oil plume samples, which could be important in determining the long-term effects of the oil spill on the marine ecosystems. Also, consistent with phylogenetic gene distribution obtained using a phylogenetic microarray "PhyloChip" (Hazen *et al*., 2010), functional genes representative of the order *Oceanospirillales* appeared to have significantly higher ($p < 0.01$) abundance in the plume samples than in non-plume samples. While the dominance of the *Oceanospirillum* population consuming the oil in the plume was based on clone library and sequence analysis of 16S rRNA genes (Hazen *et al*., 2010). GeoChip was not originally designed to link the detection of functional genes to the existence of related microbial population and it contains 567 functional genes derived from the order *Oceanospirillales* with 25 genes detected in this study. In addition, large number of metal resistance genes were enriched in plume samples, which are usually linked to organic degradation genes, for example on plasmids (Kunapuli *et al*., 2007; Parales and Haddock, 2004). Our GeoChip results demonstrated that there is a great potential for *intrinsic* bioremediation of oil contamination in the deep-sea environment.

Anaerobic hydrocarbon degradation associated with sulfate reduction, denitrification and

methanogenesis has long been considered the prevailing mechanism for petroleum biodegradation in the deep subsurface (Aitken *et al*., 2004; Head *et al*., 2003; Jones *et al*., 2008; Kniemeyer *et al*., 2007). Recent investigations have demonstrated that several classes of petroleum hydrocarbons including alkanes (So *et al*., 2003), mono- and polycyclic aromatic compounds (Meckenstock *et al*., 2000; Widdel and Rabus, 2001), and short-chain hydrocarbons (Kniemeyer *et al*., 2007), can be degraded anaerobically under nitrate-, iron-, or sulfate-reducing conditions, or under methanogenic conditions (Harayama *et al*., 2004; Jones *et al*., 2008). Indeed, a substantial number of *dsrA/B* genes for sulfate reduction, *mcrA* genes for methanogenesis, *narG*, *nirS*, *nirK*, and *nosZ* responsible for denitrification, and populations for metal reduction were detected in this study. Also, *dsrA/B* and *mcrA* genes showed significantly ($p<0.05$ or 0.01) higher abundances in the plume than non-plume samples. In addition, *bbs* for the strict anaerobic toluene degradation were detected and enriched in the plume samples. It is possible that anaerobic hydrocarbon degradations could have most likely occured through microaggregate formation as reported in Hazen *et al*., 2010.

Hydrocarbon degradation is generally limited by nutrient availability, which can be improved by nutrient recycling through phage-mediated biomass turnover (Head *et al*., 2006; Jiang *et al*., 1998; Paul, 2008). Because significant biomass increase was observed (Hazen *et al*., 2010) in the plume samples, bacteriophages could play critical roles in hydrocarbon degradation. Approximately 43% of marine bacterial isolates have been found to contain prophages (Jiang *et al*., 1998; Paul, 2008), which are induced by various environmental contaminants such as fuel oil (Cochran *et al*., 1998). The oil spill may stimulate the growth of pathogenic bacteria in marine environments and many pathogens are capable of efficiently degrading hydrocarbons (Rojo and Martínez, 2010). The research on phages has been heavily slanted to those that affect human related activities, health/medical and industry. Since no target genes for *Oceanospirillum* phages were designed on GeoChip

4.0, the *Oceanospirillum* phages were not detected. Genes for both iron uptake (*iro*) and adherence (*pap* and *pilin*) were significantly ($p < 0.01$ or 0.05) enriched in the plume samples. The increase in the abundance of microorganisms capable of producing siderophores, highly specific iron chelating compounds, may facilitate microbial acquisition of iron, a limiting nutrient in marine systems (Barbeau *et al*., 2001a; Barbeau *et al*., 2001b), thereby potentially increasing hydrocarbon degradation.

A substantial quantity of methane gas was released together with the oil (Oil budget calculator science and engineering team, 2010; Kessler *et al*., 2011), which may result in more methane in the oil plume ecosystem and have the potential to greatly impact methane metabolism. GeoChip targets three key genes/enzymes involved in methane metabolism with *mcrA* encoding methyl coenzyme M reductase for methanogenesis and two enzymes/genes (methane monooxygenase/*mmoX* and particulate methane monooxygenase/*pmoA*) for methanotrophy (He *et al*., 2010). In this study, *pmoA* and *mmoX* genes for aerobic methane oxidation did not show a statistically significant change though their abundance was higher in plume samples than in non-plume samples. There are two possible explanations for this: one is that the aerobic methane oxidation was inhibited due to the presence of easier to degrade alkanes in the deep-sea, and the other is that the methane gas was moved up to the surface more directly and did not accumulate in the deep oil plume. Also, unlike propane, methane may form gas hydrates at the deep plume temperature and pressure, making it unavailable to microorganisms (Valentine *et al*., 2010). However, significantly ($p<0.05$) higher signal intensities were observed for *mcrA* in the plume than non-plume samples, indicating that those enriched *mcrA* genes derived from methanogens likely link to hydrocarbon degradation rather than plume methane release (Harayama *et al*., 2004; Jones *et al*., 2008). Enzymes or genes involved in anaerobic methane oxidation, however, remain unclear, thus we could not detect this functional process.

In this study, many functional genes were detected in the uncontaminated samples that were not detected in the contaminated samples (Table S3 and Fig S10). These results suggest that oil spills can select against those populations containing these genes, or that specific members of the community have a selective advantage if they are capable of hydrocarbon degradation and these grow to represent a greater proportion of the functional gene repertoire.

In conclusion, our results indicate that a variety of hydrocarbon-degrading functional genes were enriched in response to oil contamination and associated environmental changes. Our results also imply that there is a great potential for *in situ* bioremediation of oil contaminants in the deep-sea water ecosystem, and such oil-degrading populations and associated microbial communities may play a significant role in determining the ultimate fates and consequences of the spilled oil. However, to further understand and evaluate the potential impacts of this unprecedented oil spill on the marine ecosystem structure and function, it is essential to launch an integrated and comprehensive monitoring program to track the dynamics and adaptive responses of microbial communities together with other physical and chemical analysis of tracing oil contaminants and their products.

Acknowledgements

We thank the Captain, crew, and science teams aboard the R/V Ocean Veritas and R/V Brooks McCall. This work was part of ENIGMA, a Scientific Focus Area Program supported by the US Department of Energy, Office of Science, Office of Biological and Environmental Research, Genomics: GTL Foundational Science through Contract DE-AC02-05CH11231 between Lawrence Berkeley National Laboratory and the U.S. Department of Energy. This study was also supported by University of Oklahoma Research Foundation and the National Key Science and Technology Project of China: Water Pollution Control and Treatment (NO.2008ZX07101-006).

Table 1. Significance of the effects of the oil spill on the overall microbial community structure and geochemical pattern using three statistical analyses.

	Geochemical parameters[d]		Microbial community	
Method	Statistic	P value	Statistic	P value
MRPP[a]	233.112	0.037	53.617	0.003
ANOSIM[b]	0.057	0.046	0.501	0.002
Adonis[c]	0.258	0.043	0.192	<0.001

[a]Multiple Response Permutation Procedure, a nonparametric procedure that does not depend on assumptions such as normally distributed data or homogeneous variances, but rather depends on the internal variability of the data.

[b]Analysis of Similarities.

[c]Non-parametric multivariate analysis of variance (MANOVA) with the adonis function.

[d]Geochemical parameters included: temperature, DO concentration, fluorometer detection of oil, small particle concentrations, Fe, nitrate, phosphate, benzene, toluene, naphthalene, ethylbenzene, isopropylbenzene, n-propylbenzene, 1,3,5-trimethylbenzene, tert-butylbenzene, 1,2,4-trimethylbenzene, sec-butylbenzene, p-isopropyltoluene, n-butylbenzene, total xylenes, total volatile HC, and total petroleum hydrocarbons - extractable (DRO).

All three tests are non-parametric multivariate analyses based on dissimilarities among samples.

Figure legend

Fig 1. Hierarchical cluster analysis of all genes present in at least two out of five samples. Results were generated in CLUSTER and visualized using TREEVIEW. Red indicates signal intensities above background while black indicates signal intensities below background. Brighter red coloring indicates higher signal intensities. All oil plume samples clustered together and were well separated from non-plume samples.

Fig. 2 Canonical correspondence analysis (CCA) compares the GeoChip hybridization signal intensities (symbols) and environmental variables (arrows). Environmental variables were chosen based on significance calculated from individual CCA results and variance inflation factors (VIFs) calculated during CCA. The percentage of variation explained by each axis is shown, and the relationship is significant (p=0.026).

Fig 3. Variation partitioning based on canonical correspondence analysis (CCA) for all functional gene signal intensities. A CCA-based variance inflation factor (VIF) was performed to identify common sets of oil composition and sea water variables important to the microbial community structure. Oil composition variables included fluorometer detection of oil, the concentration of total volatile hydrocarbons, xylenes, and petroleum hydrocarbons - extractable (DRO). Sea water geochemical variables included temperature, DO, Fe, and phosphate.

Fig 4. The normalized signal intensity of the *nahA* genes (naphthalene 1,2-diooxygenase) for the initial oxidation of naphthalene. The signal intensity for each sequence was the average of the total signal intensity from all the replicates. Gene number is the protein ID number for each gene as listed in the GenBank database. All data are presented as mean ± SE. ***p<0.01,

**$p<0.05$, *$p<0.1$.

Fig 5. The normalized signal intensity of *bbs* (beta-oxidation of benzylsuccinate) genes for anaerobic toluene degradation. The signal intensity for each sequence was the average of the total signal intensity from all the replicates. Gene number is the protein ID number for each gene as listed in the GenBank database. All data are presented as mean ± SE. ***$p<0.01$, **$p<0.05$, *$p<0.1$. In a total, 7 probes were designed for *bbs* genes in GeoChip 4.0 and 3 probes were detected in the samples.

Fig. 6 The relative changes of the detected genes involved in the N cycle in oil plume. The signal intensity for each gene detected was normalized by all detected gene sequences using the mean. The percentage of a functional gene in a bracket was the sum of signal intensity of all detected sequences of this gene divided by the grand sum of signal intensity of the detected N cycle genes, and weighted by the fold change of the signal intensity of this gene in plume to that in non-plume. For each functional gene, red indicates that this gene had a higher signal intensity in plume than in non-plume and their significance was indicated with two stars (**) at $p<0.01$, while blue indicates that this gene had a lower signal intensity in oil-plume than in non-plume. Grey-colored genes were not targeted by this GeoChip, or not detected in those samples. It remains unknown if *nosZ* homologues exist in nitrifiers. Description of the genes: (A) *gdh*, encoding glutamate dehydrogenase, *ureC*, encoding urease responsible for ammonification. (B) *nasA*, encoding nitrate reductase, *NiR*, encoding nitrite reductase, responsible for assimilatory N reduction; (C) *nifH*, encoding nitrogenase responsible for N_2 fixation; (D) *narG* encoding nitrate reductase, *nirS* and *nirK-D* (with denitrification activity), encoding nitrite reductase; *nosZ*, encoding nitrous oxide reducatse, *norB*, encoding nitric oxide reducatse, responsible for denitrification (E) *napA*, encoding

periplasmic nitrate reductase, *nrfA*, encoding c-type cytochrome nitrite reducatse, responsible for dissimilatory N reduction to ammonium; (F) *hao*, encoding hydroxylamine oxidoreductase, and *nirK-N* encoding nitrite reductase for nitrifiers (an indication of nitrification activity), responsible for nitrification.

Fig. 7 The normalized signal intensity of the replication genes for bacteriophage. The signal intensity for each sequence was the average of the total signal intensity from all the replicates. Gene number is the protein ID number for each gene as listed in the GenBank database. All data are presented as mean ± SE. ***$p<0.01$, **$p<0.05$, *$p<0.1$.

References

Aitken CM, Jones DM, Larter SR. (2004). Anaerobic hydrocarbon biodegradation in deep subsurface oil reservoirs. *Nature* **431**: 291-294.

Anderson M. (2001). A new method for non-parametric multivariate analysis of variance. *Austral Ecol* **26**: 32-46.

Barbeau K, Rue EL, Bruland KW, Butler A. (2001a). Photochemical cycling of iron in the surface ocean mediated by microbial iron(III)-binding ligands. *Nature* **413**: 409-413.

Barbeau K, Zhang G, Live DH, Butler A. (2001b). Petrobactin, a photoreactive siderophore produced by the oil-degrading marine bacterium *Marinobacter hydrocarbonoclasticus*. *J Am Chem Soc* **124**: 378-379.

Bordenave S, Goni-Urriza MS, Caumette P, Duran R. (2007). Effects of heavy fuel oil on the bacterial community structure of a pristine microbial mat. *Appl Environ Microbiol* **73**: 6089-6097.

Camilli R, Reddy CM, Yoerger DR, Van Mooy BAS, Jakuba MV, Kinsey JC *et al.* (2010). Tracking hydrocarbon plume transport and biodegradation at Deepwater Horizon. *Science* **330**: 201-204.

Cochran PK, Kellogg CA, Paul JH. (1998). Prophage induction of indigenous marine lysogenic bacteria by environmental pollutants. *Mar Ecol Prog Ser* **164**: 125-133.

de Hoon MJ, Imoto S, Nolan J, Miyano S. (2004). Open source clustering software. *Bioinformatics* 20: 1453-1454.

Harayama S, Kasai Y, Hara A. (2004). Microbial communities in oil-contaminated seawater. *Curr Opin Biotechnol* **15**: 205-214.

Hazen TC, Dubinsky EA, DeSantis TZ, Andersen GL, Piceno YM, Singh N *et al.* (2010). Deep-sea oil plume enriches indigenous oil-degrading bacteria. *Science* **330**: 204-208.

He Z, Deng Y, Van Nostrand JD, Tu Q, Xu M, Hemme CL *et al.* (2010). GeoChip 3.0 as a high-throughput tool for analyzing microbial community composition, structure and functional activity. *ISME J* **4**: 1167-1179.

Head IM, Jones DM, Larter SR. (2003). Biological activity in the deep subsurface and the origin of heavy oil. *Nature* **426:** 344-352.

Head IM, Jones DM, Roling WFM. (2006). Marine microorganisms make a meal of oil. *Nat Rev Micro* **4:** 173-182.

Jiang SC, Kellogg CA, Paul JH. (1998). Characterization of marine temperate phage-host systems isolated from Mamala Bay, Oahu, Hawaii. *Appl Environ Microbiol* **64**: 535-542.

Jones DM, Head IM, Gray ND, Adams JJ, Rowan AK, Aitken CM *et al.* (2008). Crude-oil biodegradation via methanogenesis in subsurface petroleum reservoirs. *Nature* **451**: 176-180.

Kerr R, Kintisch E, Stokstad E. (2010a). Will deepwater horizon set a new standard for catastrophe? *Science* **328**: 674-675.

Kerr RA, Kintisch E, Schenkman L, Stokstad E. (2010b). Five questions on the spill. *Science* **328**: 962-963.

Kessler JD, Valentine DL, Redmond MC, Du MR, Chan EW, Mendes SD *et al*. (2011). A persistent oxygen anomaly reveals the fate of spilled methane in the deep Gulf of Mexico.

Science **331**: 312-315.

Kniemeyer O, Musat F, Sievert SM, Knittel K, Wilkes H, Blumenberg M *et al.* (2007). Anaerobic oxidation of short-chain hydrocarbons by marine sulphate-reducing bacteria. *Nature* **449**: 898-901.

Kunapuli U, Lueders T, Meckenstock RU. (2007) The use of stable isotope probing to identify key iron-reducing microorganisms involved in anaerobic benzene degradation. *ISME J* **1**: 643–653.

Larter S, Wilhelms A, Head I, Koopmans M, Aplin A, Di Primio R *et al.* (2003). The controls on the composition of biodegraded oils in the deep subsurface-part 1: biodegradation rates in petroleum reservoirs. *Org Geochem* **34**: 601-613.

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.

Mascarelli A. (2010a). Extent of lingering Gulf oil plume revealed. *Nature*. (doi:10.1038/news.2010.1420).

Mascarelli A. (2010b). Deepwater Horizon: After the oil. *Nature* 467: 22-24.

Meckenstock RU, Annweiler E, Michaelis W, Richnow HH, Schink B. (2000). Anaerobic naphthalene degradation by a sulfate-reducing enrichment culture. *Appl Environ Microbiol* **66**: 2743-2747.

Operational Science Advisory Team (OSAT). Summary report for sub-sea and sub-surface oil and dispersant detection: sampling monitoring. December 17, 2010. (http://www.restorethegulf.gov/release/2010/12/16/data-analysis-and-findings).

Parales RE, Haddock JD. (2004). Biocatalytic degradation of pollutants. *Curr Opin Biotechnol* **15**: 374-379.

Paul JH. (2008). Prophages in marine bacteria: dangerous molecular time bombs or the key to survival in the seas? *ISME J* **2**: 579-589.

Ramette A, Tiedje J. (2007). Biogeography: an emerging cornerstone for understanding prokaryotic diversity, ecology, and evolution. *Microb Ecol* **53**: 197-207.

Rojo F, Martínez JL. (2010). Oil Degraders as Pathogens. In: Timmis KN (ed). *Handbook of Hydrocarbon and Lipid Microbiology*. Springer Berlin Heidelberg. pp 3293-3303.

So CM, Phelps CD, Young LY. (2003). Anaerobic transformation of alkanes to fatty acids by a sulfate-reducing bacterium, strain Hxd3. *Appl Environ Microbiol* **69**: 3892-3900.

The Federal Interagency Solutions Group, Oil Budget Calculator Science and Engineering Team. Oil budget calculator deepwater horizon technical documentation. November 2010 (http://www.restorethegulf.gov/sites/default/files/documents/pdf/OilBudgetCalc_Full_HQ-Print_111110.pdf).

Valentine DL, Kessler JD, Redmond MC, Mendes SD, Heintz MB, Farwell C *et al.* (2010). Propane respiration jump-starts microbial response to a deep oil spill. *Science* **330**: 208-211.

Widdel F, Rabus R. (2001). Anaerobic biodegradation of saturated and aromatic hydrocarbons. *Curr Opin Biotechnol* **12**: 259-276.

Wu L, Liu X, Schadt CW, 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.

Zhou J, Kang S, Schadt CW, Garten CT. (2008). Spatial scaling of functional gene diversity across various microbial taxa. *Proc Natl Acad Sci USA* **105**: 7768-7773.

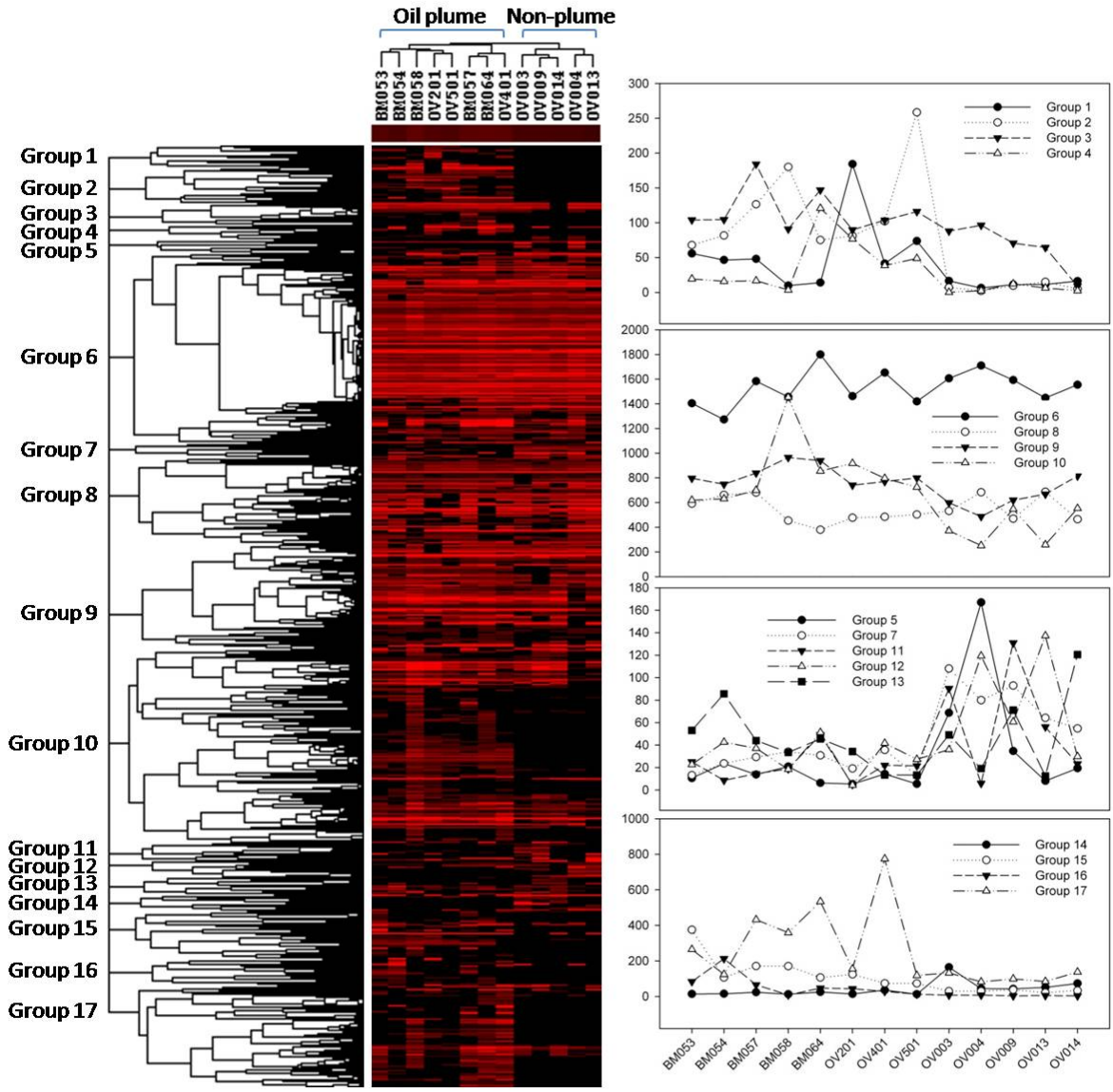

Fig. 1 Hierarchical cluster analysis of all genes present in at least two out of five samples. Results were generated in CLUSTER and visualized using TREEVIEW. Red indicates signal intensities above background while black indicates signal intensities below background. Brighter red coloring indicates higher signal intensities. All oil plume samples clustered together and were well separated from non-plume samples.

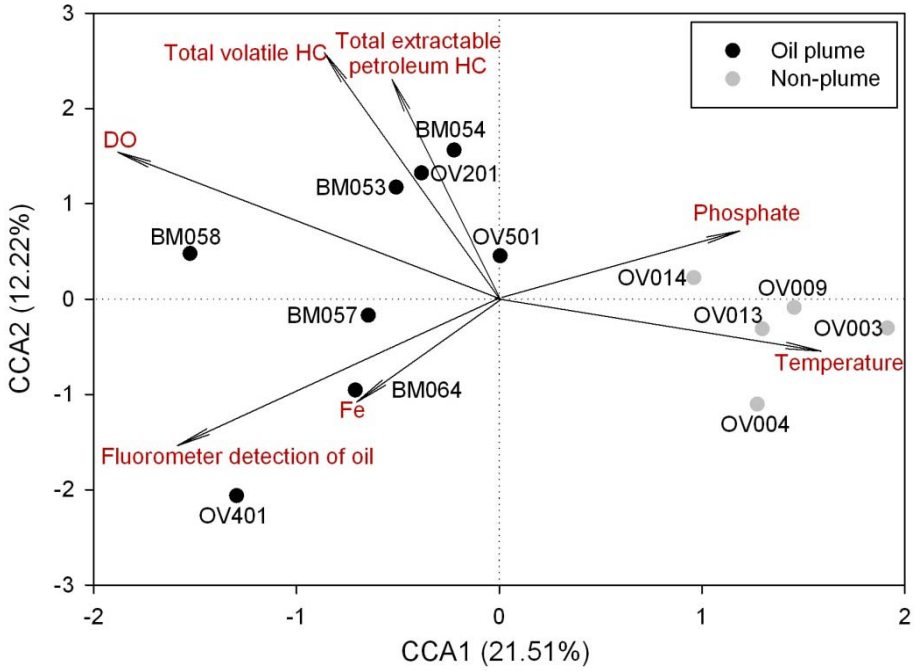

Fig. 2 Canonical correspondence analysis (CCA) compares the GeoChip hybridization signal intensities (symbols) and environmental variables (arrows). Environmental variables were chosen based on significance calculated from individual CCA results and variance inflation factors (VIFs) calculated during CCA. The percentage of variation explained by each axis is shown, and the relationship is significant (p=0.026).

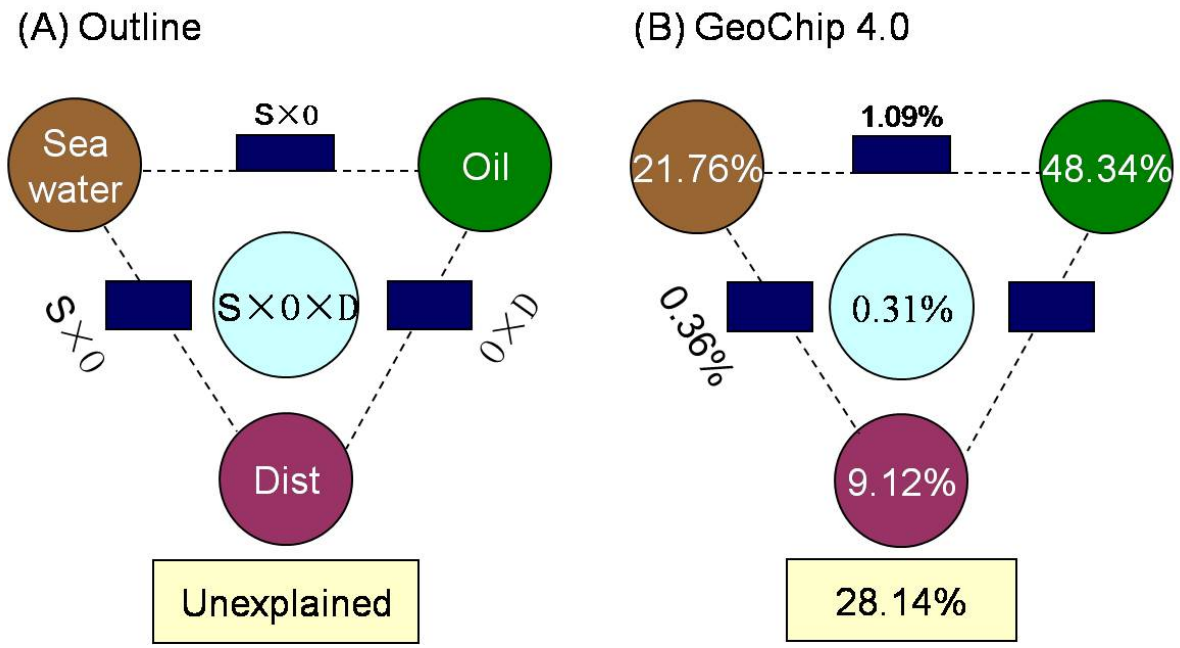

Fig. 3 Variation partitioning based on canonical correspondence analysis (CCA) for all functional gene signal intensities. A CCA-based variance inflation factor (VIF) was performed to identify common sets of oil composition and sea water variables important to the microbial community structure. Oil composition variables included fluorometer detection of oil, the concentration of total volatile hydrocarbons, xylenes, and petroleum hydrocarbons - extractable (DRO). Sea water geochemical variables included temperature, DO, Fe, and phosphate.

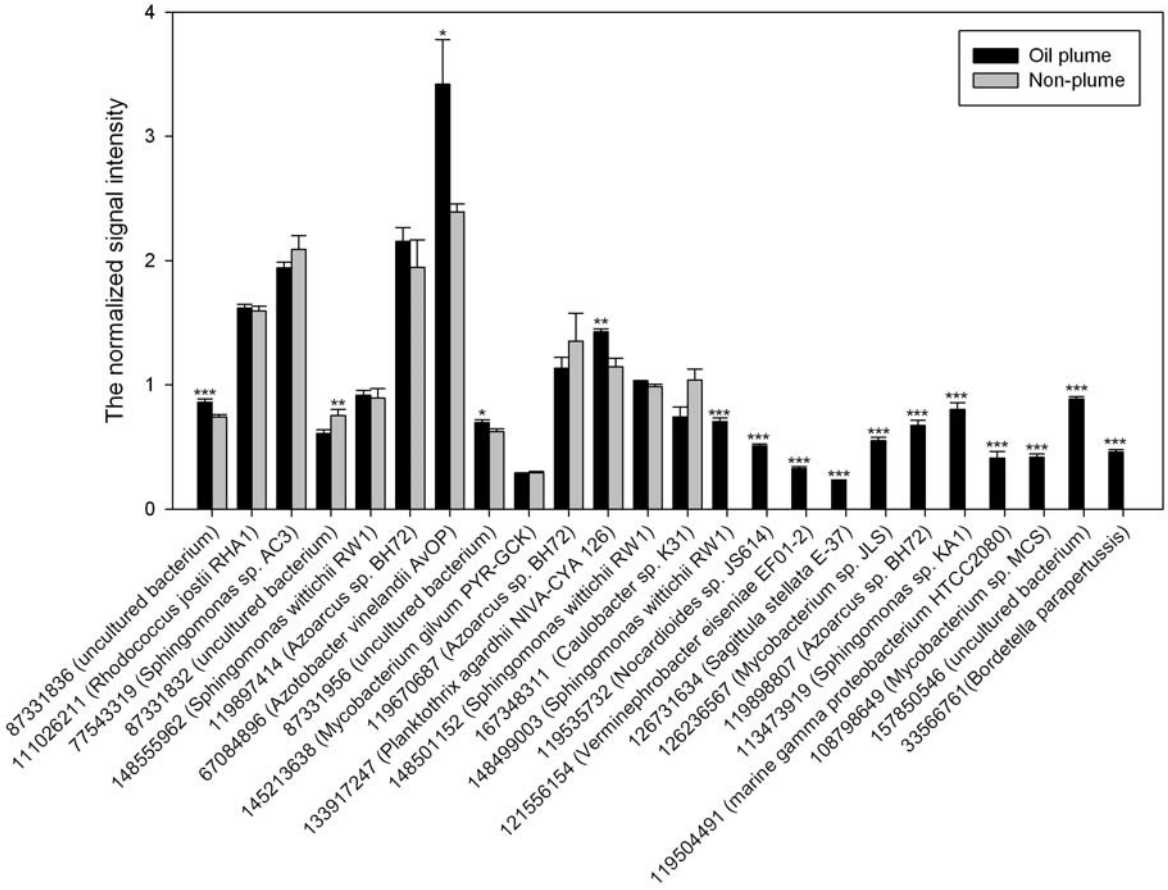

Fig. 4 The normalized signal intensity of the *nahA* genes (naphthalene 1,2-diooxygenase) for the initial oxidation of naphthalene. The signal intensity for each sequence was the average of the total signal intensity from all the replicates. Gene number is the protein ID number for each gene as listed in the GenBank database. All data are presented as mean ± SE. ***p<0.01, **p<0.05, *p<0.1.

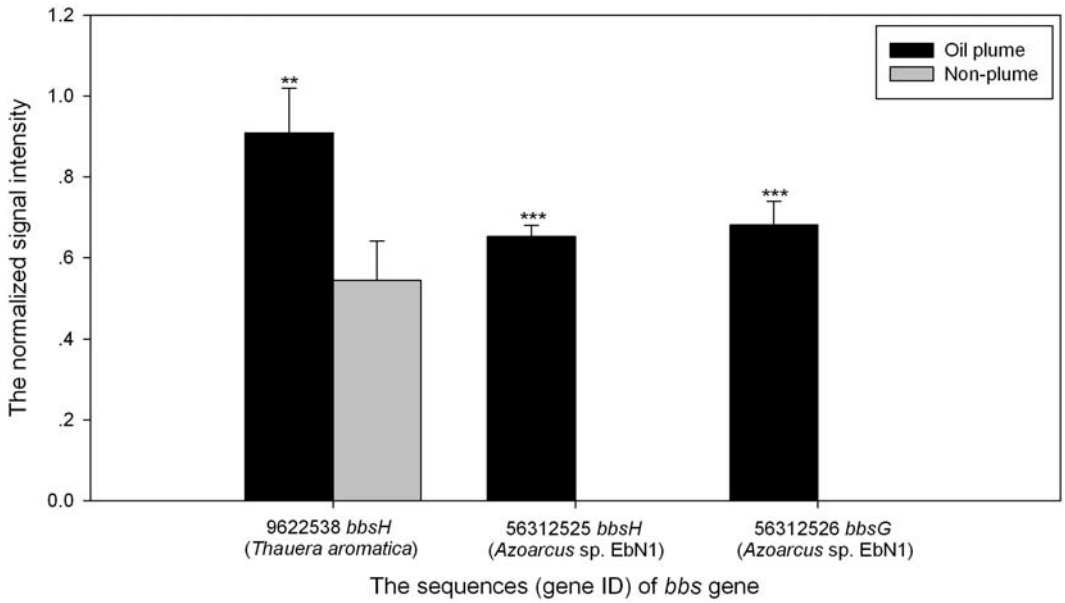

Fig. 5 The normalized signal intensity of *bbs* (beta-oxidation of benzylsuccinate) genes for anaerobic toluene degradation. The signal intensity for each sequence was the average of the total signal intensity from all the replicates. Gene number is the protein ID number for each gene as listed in the GenBank database. All data are presented as mean ± SE. ***p<0.01, **p<0.05, *p<0.1. In a total, 7 probes were designed for *bbs* genes in GeoChip 4.0 and 3 probes were detected in the samples.

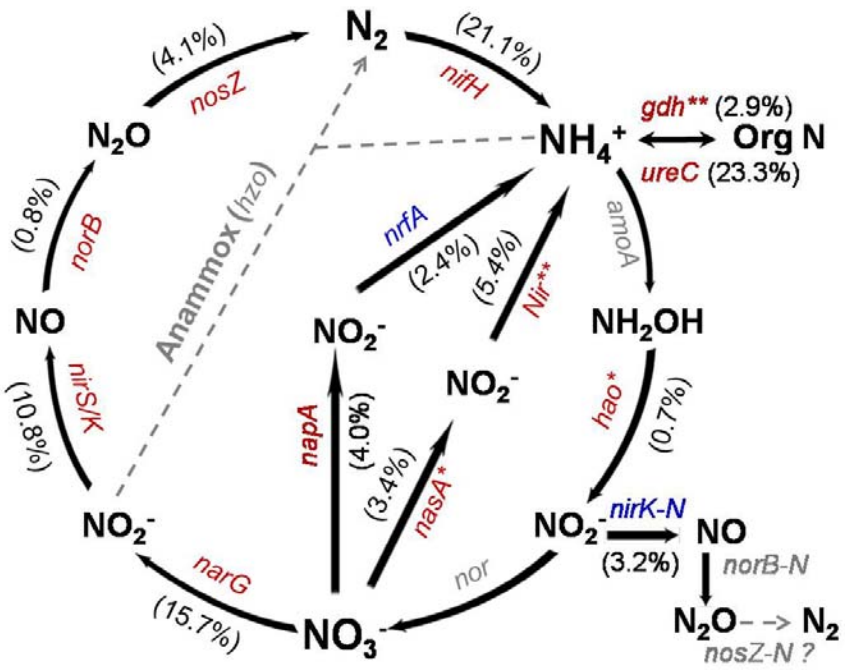

Fig. 6 The relative changes of the detected genes involved in the N cycle in oil plume. The signal intensity for each gene detected was normalized by all detected gene sequences using the mean. The percentage of a functional gene in a bracket was the sum of signal intensity of all detected sequences of this gene divided by the grand sum of signal intensity of the detected N cycle genes, and weighted by the fold change of the signal intensity of this gene in plume to that in non-plume. For each functional gene, red indicates that this gene had a higher signal intensity in plume than in non-plume and their significance was indicated with two stars (**) at $p<0.01$, while blue indicates that this gene had a lower signal intensity in oil-plume than in non-plume. Grey-colored genes were not targeted by this GeoChip, or not detected in those samples. It remains unknown if *nosZ* homologues exist in nitrifiers. Description of the genes: (A) *gdh*, encoding glutamate dehydrogenase, *ureC*, encoding urease responsible for ammonification. (B) *nasA*, encoding nitrate reductase, *NiR*, encoding nitrite reductase, responsible for assimilatory N reduction; (C) *nifH*, encoding nitrogenase responsible for N_2 fixation; (D) *narG* encoding nitrate reductase,

nirS and *nirK-D* (with denitrification activity), encoding nitrite reductase; *nosZ*, encoding nitrous oxide reducatse, *norB*, encoding nitric oxide reducatse, responsible for denitrification (E) *napA*, encoding periplasmic nitrate reductase, *nrfA*, encoding c-type cytochrome nitrite reducatse, responsible for dissimilatory N reduction to ammonium; (F) *hao*, encoding hydroxylamine oxidoreductase, and *nirK-N* encoding nitrite reductase for nitrifiers (an indication of nitrification activity), responsible for nitrification.

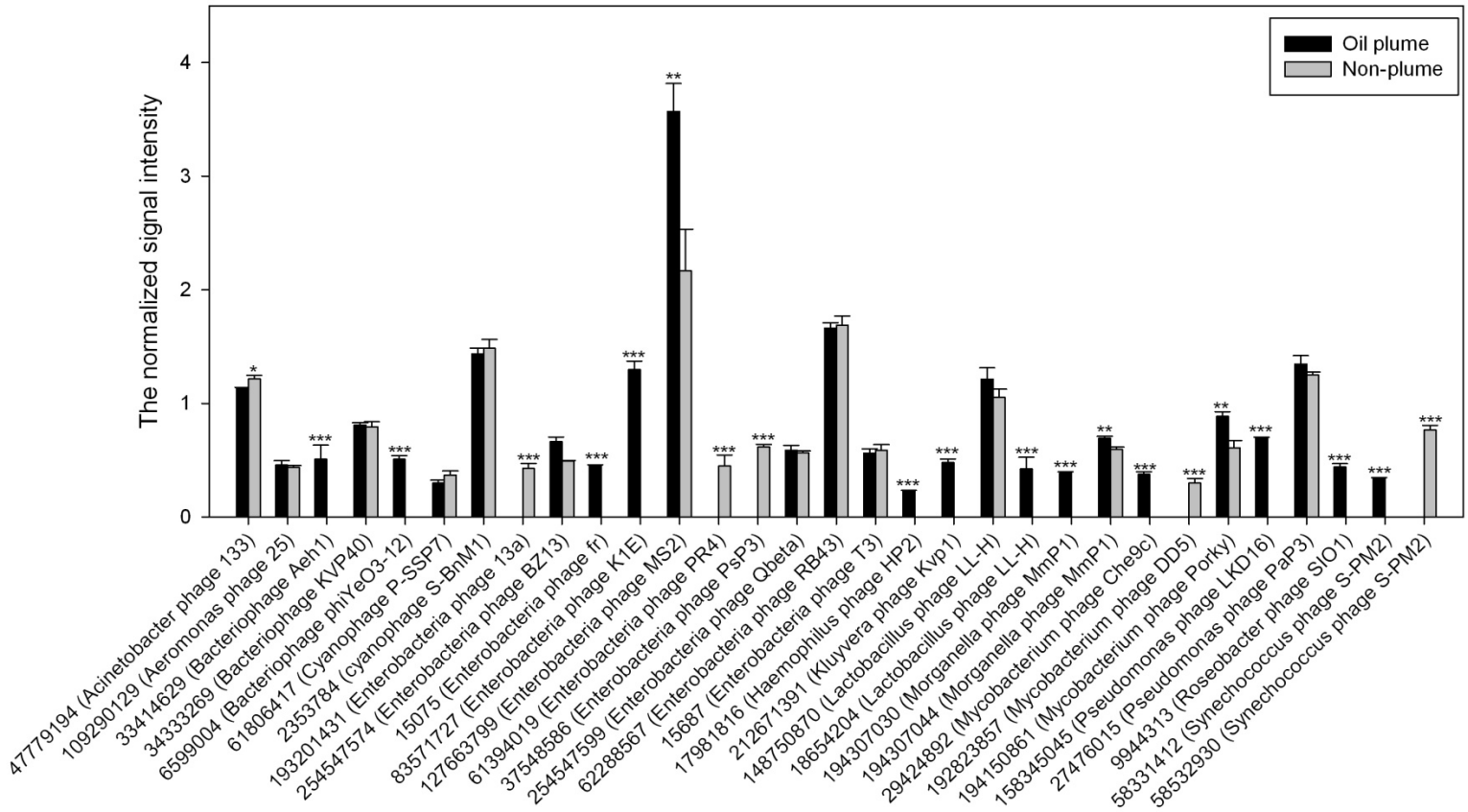

Fig. 7 The normalized signal intensity of the replication genes for bacteriophage. The signal intensity for each sequence was the average of the total signal intensity from all the replicates. Gene number is the protein ID number for each gene as listed in the GenBank database. All data are presented as mean ± SE. ***p<0.01, **p<0.05, *p<0.1.

DISCLAIMER

This document was prepared as an account of work sponsored by the United States Government. While this document is believed to contain correct information, neither the United States Government nor any agency thereof, nor The Regents of the University of California, nor any of their employees, makes any warranty, express or implied, or assumes any legal responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by its trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof, or The Regents of the University of California. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof or The Regents of the University of California.

Ernest Orlando Lawrence Berkeley National Laboratory is an equal opportunity employer.