UC Merced
UC Merced Previously Published Works

Title

Transcriptional response of bathypelagic marine bacterioplankton to the Deepwater Horizon oil spill.

Permalink

https://escholarship.org/uc/item/1359k3ts

Journal

The ISME Journal: Multidisciplinary Journal of Microbial Ecology, 7(12)

Authors

Rivers, Adam

Sharma, Shalabh

Tringe, Susannah

et al.

Publication Date

2013-12-01

DOI

10.1038/ismej.2013.129

Peer reviewed

npg

ORIGINAL ARTICLE

Transcriptional response of bathypelagic marine bacterioplankton to the Deepwater Horizon oil spill

Adam R Rivers[1], Shalabh Sharma[1], Susannah G Tringe[2], Jeffrey Martin[2], Samantha B Joye[1] and Mary Ann Moran[1]

[1]Department of Marine Sciences, University of Georgia, Athens, GA, USA and [2]DOE Joint Genome Institute, Walnut Creek, CA, USA

The Deepwater Horizon blowout released a massive amount of oil and gas into the deep ocean between April and July 2010, stimulating microbial blooms of petroleum-degrading bacteria. To understand the metabolic response of marine microorganisms, we sequenced ~66 million community transcripts that revealed the identity of metabolically active microbes and their roles in petroleum consumption. Reads were assigned to reference genes from ~2700 bacterial and archaeal taxa, but most assignments (39%) were to just six genomes representing predominantly methane- and petroleum-degrading Gammaproteobacteria. Specific pathways for the degradation of alkanes, aromatic compounds and methane emerged from the metatranscriptomes, with some transcripts assigned to methane monooxygenases representing highly divergent homologs that may degrade either methane or short alkanes. The microbial community in the plume was less taxonomically and functionally diverse than the unexposed community below the plume; this was due primarily to decreased species evenness resulting from Gammaproteobacteria blooms. Surprisingly, a number of taxa (related to SAR11, *Nitrosopumilus* and *Bacteroides*, among others) contributed equal numbers of transcripts per liter in both the unexposed and plume samples, suggesting that some groups were unaffected by the petroleum inputs and blooms of degrader taxa, and may be important for re-establishing the pre-spill microbial community structure.

Introduction

The Deepwater Horizon blowout released 6.9×10^{11} g of oil and gas into Gulf of Mexico (GOM) waters (Reddy *et al.*, 2012). Approximately one-third of the hydrocarbons initially concentrated in a lens near 1100 m depth (Camilli *et al.*, 2010; Ryerson *et al.*, 2012), inducing a shift in the composition of the microbial community and the proliferation of Gammaproteobacteria related to *Oceanospirillales*, *Colwellia* and *Cycloclasticus* (Hazen *et al.,* 2010). Previous work has examined community composition and the abundance of functional genes and transcripts (Hazen *et al.,* 2010; Lu *et al,.* 2011; Mason *et al.,* 2012), as well as microbial consumption of hydrocarbons (Redmond and Valentine, 2012). This paper extends that work by analyzing ~66 million microbial messenger RNAs (mRNAs) and >350 000 16S rRNA genes for a comprehensive metabolic and ecological reconstruction of microbial processes in response to the oil and gas inputs.

The petroleum released from the Macondo Well contained short- to long chain alkanes, aromatic compounds, polycyclic aromatic hydrocarbons, asphaltenes and polar compounds. Methane accounted for 15% of the total petroleum and gas leaving the riser pipe, short alkanes excluding methane (C_2–C_5) comprised 9% and medium chain alkanes (n-C_6 to n-C_{20}) comprised 13% (Reddy *et al.,* 2012). High rate constants were reported for the oxidation of ethane and propane (Valentine *et al.,* 2010), while bacteria enriched from the subsurface plume had the potential to degrade long chain n-alkanes and branched alkanes (Hazen *et al.,* 2010).

The role of bacteria in methane metabolism is less clear, however, as methane may have been removed by microbial degradation or by venting to the atmosphere. Methane oxidation rates measured in May (3 weeks after the uncontrolled blowout began) were high (Joye, unpublished data), but rate constants for methane oxidation measured a few weeks later and again in August and September (after the well had been capped) were low (Valentine *et al.,* 2010; Kessler *et al.,* 2011) and functional gene microarrays showed no enrichment of methane monooxygenase genes in the plume communities (Lu *et al.,* 2011). Counts of 16S rRNA genes

diagnostic for methanotrophs varied depending on the timing of sampling and methodology (Hazen et al., 2010; Valentine et al., 2010; Kessler et al., 2011; Redmond and Valentine, 2012; Mason et al., 2012). Estimates of atmospheric venting, the alternative route for methane loss, did not resolve the gas fate question as flux estimates were high when based on surface water measurements (Joye, unpublished data), but low when based on remote-sensing data (Yvon-Lewis et al., 2011).

This study builds on the previous meta-omic analyses conducted following the Deepwater Horizon accident (Hazen et al., 2010; Lu et al., 2011; Mason et al., 2012) by generating a sizeable transcript data set with low levels of rRNA contamination and robust coverage of more than 561 000 functional genes from exposed and unexposed communities. The high coverage of expressed oxygenase sequences allowed application of evolutionary placement methods to address the relative importance of methane, ethane and propane metabolism, whereas the use of internal standards generated information on changes in transcript numbers per liter. These approaches created a detailed picture of the metabolic and ecological responses of the bathypelagic microbial community.

Materials and methods

Microbial mRNA was sequenced from four samples collected at two different stations located ∼6–8 km from the wellhead. Two samples were collected within the oil/gas plume (samples P16 at 1116 m and P52 at 1198 m) from areas with elevated colored dissolved organic matter (CDOM, a proxy for oil) fluorescence and oxygen depletion. Two samples were collected below the plume, and although intended to represent two non-impacted communities, the sample IP16 from 1240 m was adjacent to an oxygen anomaly and subsequent data analysis suggested that it represented an intermediate condition. The sample NP52 from 1286 m was the least affected by the oil/gas release, and although it did not have detectable oil or oxygen anomalies by instruments on the conductivity, temperature and depth (CTD) rosette instrumentation, gas chromatographic analysis revealed that methane was elevated above background levels (Table 1). This paired-in/below-plume sampling scheme minimized the influence of geographic location. In addition to the transcript data, seventeen 16S rRNA gene libraries were created from seven stations at depths above, in and below the plume. An oceanographic station map and depth profiles (Supplementary Figure S1) were plotted using the R package OCE (Kelley, 2013).

Sample collection and sequencing
Samples were collected between 26 May 2010 and 3 June 2010 aboard the RV Walton Smith using a rinsed CTD rosette. For RNA analysis, 9 l of seawater was immediately filtered through a 0.45-µm-pore

Table 1 Characteristics of metatranscriptome samples

	P16	P52	IP16	NP52
Physical				
Depth (m)	1116	1198	1240	1286
Hydrocarbon plume	Yes	Yes	No	No
Oxygen anomaly	Yes	Yes	Yes	No
Chemical[†]				
CDOM (relative fluorescence units)	2.14	47.44	0.44	0.44
Methane (nmol l^{-1})	210 171	235 980	4517	106
Nitrate (µmol l^{-1})	21.26	15.14	26.21	27.08
Nitrite (µmol l^{-1})	0.28	0.89	0.13	0.05
Phosphate (µmol l^{-1})	1.46	0.96	1.78	1.72
Biological				
Cell density (cells ml^{-1})	4.95E + 05	1.39E + 05	3.30E + 04	1.74E + 04
Total RNA (µg per l seawater)	1.90E − 04	4.20E − 05	7.12E − 07	3.37E − 07
Sequencing				
Median read length (bases)	217	208	207	233
Total joined reads	28 599 525	23 871 295	22 614 967	29 171 472
rRNA reads (%)	1 780 537 (6%)	9 664 134 (43%)	8 538 576 (36%)	6 104 284 (21%)
Internal standard reads (%)	NA	3 619 342 (16%)	3 918 756 (16%)	5 007 057 (17%)
Possible proteins (%)	26 792 869 (94%)	9 331 491 (39%)	11 413 963 (50%)	18 060 131 (62%)
Archaeal and bacterial hits	10 054 034 (35%)	6 075 577 (25%)	5 611 064 (35%)	1 945 004 (35%)
RefSeq genes	341 983	421 562	428 614	242 768
RefSeq taxa	3627	3628	3742	3485
KEGG assignments	6 141 041	3 781 411	3 680 128	1 097 381

Abbreviation: CDOM, colored dissolved organic matter.
[†]Nutrient data are from the closest depth available, corresponding to 1125 m for P16, 1180 m for P52, 1280 m for IP16 and 1250 M for NP52.

size, 142-mm-diameter polyethersulfone membrane filter (Pall Corporation, Port Washington, NY, USA). The larger pore size was chosen to minimize clogging by oil droplets and to keep average filtration times to <10 min to better preserve mRNA integrity, although it may have biased cell collection toward larger cells or those associated with oil droplets. All tubing, filters and containers were soaked in 1 M HCl and rinsed in distilled water before sampling. Filters were immediately preserved in RNAlater (Life Technologies, Carlsbad, CA, USA) and kept at 4 °C. A smaller volume of water (0.5–1 l) was filtered through a 0.45-μm pore size, 47-mm diameter polyethersulfone filter and stored on dry ice for subsequent DNA extraction.

RNA was extracted according to Poretsky *et al.* (2009). Six internal RNA standards (artificial mRNAs synthesized by *in vitro* transcription of DNA templates) were added as internal controls (1.4×10^{10} transcripts per sample) (Satinsky *et al.*, 2013). The two smallest standards (~200 nt) were not appreciably recovered on the solid phase extraction matrix used for RNA purification and were not considered further. RNA extracts were depleted of rRNA by creating sample-specific removal probes targeting the large and small subunits of archaeal, bacterial and eukaryotic rRNA genes using the DNA of each sample (Stewart *et al.*, 2010). The protocol was modified by pooling probes from multiple DNA samples (see below) and including an additional round of probe removal. rRNA-depleted mRNA was amplified using the MessageAmp II bacterial kit (Life Technologies) and complementary DNA was synthesized using the universal Riboclone complementary DNA synthesis system (Promega, Madison, WI, USA). Complementary DNA was sheared ultrasonically to ~225 bp, and each sample was run on one lane of an Illumina Genome Analyzer IIx (Illumina Inc., San Diego, CA, USA) using 150-bp paired-end chemistry. Sample processing and bioinformatic analysis is summarized in Supplementary Figure S2.

DNA was extracted from samples using the DNA Power Water Kit (Mo Bio Laboratories Inc., Carlsbad, CA, USA) and further purified by ethanol precipitation. The V6–V8 region of the 16S rRNA gene was amplified using the primers 926F (5′-cctatccctg tgtgccttggcagtctcag AAA CTY AAA KGA ATT GRC GG-3′) and 1392R (5′-ccatctcatccctgcgtgtctccgactcag-[barcode] -ACGGGCGGTGTGTRC-3′), with 454A or B adapter sequences indicated in lower case. PCR amplicons were purified using Ampure SPRI Beads (Beckman Coulter, Brea, CA, USA). Emulsion PCR and sequencing of the PCR amplicons were performed following the Roche/454 (Branford, CT, USA) GS FLX Titanium technology instructions.

Bioinformatics
16S rRNA gene sequence data were analyzed through the QIIME pipeline, version 1.3 (Caporaso *et al.*, 2010). Sequencing artifacts were minimized using the Denoiser package of QIIME (Reeder and

Knight, 2010). Default settings were used except that the Greengenes database (DeSantis *et al.*, 2006) was used to train the Ribosomal Database Project classifier (Wang *et al.*, 2007). Technical replicates showed good agreement and the technical replicate with the highest number of reads was used for analysis. The Unifrac distance matrix (Lozupone *et al.*, 2011) generated by QIIME was used for single-linkage clustering of the sequences and for nonmetric multidimensional scaling and environmental variable fitting in the R package Vegan (Jari *et al.*, 2013).

Transcriptomic data were checked for quality using FastX-Toolkit (Gordon and Hannon, 2011, unpublished data). Paired-end reads were joined using the program SHERA with default parameters (Rodrigue *et al.*, 2010) and a quality metric of 0.5. Joined reads were trimmed using Seqtrim (Falgueras and Lara, 2010) with default values. To remove remaining rRNA and the internal standard sequences, a Blastn search was performed against a small database containing ~1000 sequences that included the internal standards and representative ribosomal and tRNA sequences; matching reads with a bit score >50 were removed (Gifford *et al.*, 2011). The remaining paired and trimmed reads were queried against RefSeq protein database version 45 using Blastx for the best match with a bit score >40. Reads were mapped to KEGG ortholog groups (Kanehisa and Goto, 2000) and COG categories (Tatusov *et al.*, 2003) by querying the matching RefSeq gene against the KEGG gene database or the COG gene database (bit scores >40). Blast results were stored for analysis in a MySQL database. Internal standards were tallied and used to measure sequencing accuracy, as well as to estimate absolute numbers of transcripts (transcripts l^{-1}; Gifford *et al.*, 2011; Satinsky *et al.*, 2013), except for sample P16 that had an error in the amount of standard added. The plot of log-fold change and COG category assignment was generated with Circos (Krzywinski *et al.*, 2009).

The analysis of differential abundance was carried out using the R package DESeq (Anders and Huber, 2010) by comparing P16 and P52 (plume samples) with NP52 (non-plume sample); IP16 was not included because of its intermediate physical and biological characteristics. Significance was calculated using dispersion values estimated from the two plume samples, as no replicate was available for NP52, and was defined as a *P*-value of <0.05 in a negative binomial test following correction for false discovery rate (Benjamini and Hochberg, 1995). For pathway analysis, statistical significance was calculated based on the abundance of reads in KEGG orthology groups (K numbers). For organism-specific analyses, significance was calculated for each RefSeq gene present for the reference genome. The pathways inferred to be involved in hydrocarbon degradation were generated by modifying existing pathway maps (Anthony, 1982; Caspi *et al.*, 2012) with manual curation.

The *de novo* assembly of operons was carried out using the Rnnotater pipeline (Martin *et al.,* 2010); the mean assembly length was 886 bp (Supplementary Figure S3). The candidate particulate methane monooxygenase (*pmoCAB)* operons in the assembled data were identified by querying known *pmoA* and *pxmA* genes from *Methylobacter tundripaludum* against the database of transcript assemblies using Tblastn with an expect value cutoff of 10^{-5}. Open reading frames were identified using MetaGeneMark (Zhu *et al.,* 2010). Full-length monooxygenase subunit A ORFs were extracted from the operons with Blastx and aligned to known *pmoA, pxmA* and *hmoA* genes using MAFFT (Katoh *et al.,* 2002) (L-INS-I algorithm, Blosum62 matrix) and a high-quality alignment region consisting of 232 amino-acid sites was selected using TrimAl (Capella-Gutierrez *et al.,* 2009). PhyML mixture was used to generate a maximum likelihood mixture model tree using the UL3 substitution matrices with a gamma rate model, assuming four rate categories and estimating the invariant sites (Le *et al.,* 2008). Hmmalign was run with the *pmoA/amoA* hidden Markov model (PF02461) to align 33 257 short reads, 5 short reference sequences and 28 full-length reference sequences (Eddy, 2011). After alignment, the reads were placed on the tree with the maximum likelihood placement algorithm in RAxML (Stamatakis, 2006). The results were converted to phyloXML format (Han and Zmasek, 2009) by pplacer/guppy (Matsen *et al.,* 2010) and visualized in Archaeopteryx (Han and Zmasek, 2009). The tree of pyrroloquinoline quinone (PQQ) alcohol and methanol dehydrogenases was constructed using the same method. The site variability of assembled *pmoA* genes was assessed by realigning reads from evolutionary tree placement to the reference sequence using Geneious version 5.3.6 (Biomatters, Auckland, NZ, USA). The number of positions with variation (defined as >25% of reads different from the assembly) ranged from 3 (DH4550) to 120 (DH311), and variation co-occurred on reads with other variations, suggesting that reads within pmoA assemblies originated from just a few closely related bacteria.

To assess the fate of the pre-spill microbial community, transcripts per liter were estimated across samples for 50 taxa that were most abundant in sample NP52. The ratios of transcripts per liter for NP52:P52 and NP52:IP16 were used to create a hierarchical clustering of the taxa (complete linkage, Manhattan distance). Two major clusters of the tree were tested to see if there were significant differences in the number of reads per sample; testing was done with a Welch corrected one-way analysis of variance on the log-transformed read counts.

Sequences from metatranscriptomic and 16S rRNA sequencing are available from the Community Cyberinfrastructure for Advanced Microbial Ecology Research and Analysis database under project number CAM_P_0001088. The *pmoA* homolog assemblies are available at the National Center for Biotechnology Information database under accession numbers KF296332-KF296332.

Results

The depth of the Deepwater Horizon hydrocarbon plume was identified at seven stations in the vicinity of the well by CDOM fluorescence and low-oxygen anomalies (Supplementary Figure S1 B-H, Table 1) (Joye *et al.,* 2011). Seventeen seawater samples were collected at depths from directly in, above or below the plume for analysis of bacterial community composition based on amplified 16S rRNA genes. Four of these samples were selected for further metatranscriptome analysis as follows: cast 16 at 1116 and 1198 m, and cast 52 at 1240 and 1286 m. Chemical data indicated that two of the metatranscriptome samples represented plume conditions (characterized by both a hydrocarbon signal and an oxygen anomaly; these were designated as P16 and P52) and one represented largely non-impacted seawater (below both the hydrocarbon and oxygen signals; NP52). The fourth sample, initially selected as a second below-plume metatranscriptome replicate because of a low hydrocarbon signal, was later found to have a small oxygen anomaly and was designated as an intermediate plume sample (IP16); this sample might represent a region receiving low hydrocarbon inputs or an aged portion of the main plume. The two in-plume samples had greater biological activity than the non-plume or intermediate plume samples, with cell concentrations that were almost 10-fold higher (geometric mean in plume: 2.6×10^5, intermediate/non-plume: 2.4×10^4 cells ml^{-1}) and total RNA yields that were $\sim$200-fold higher (Table 1, Supplementary Figure S4).

16S rRNA composition
The 17 microbial communities were characterized based on >350 000 16S rRNA amplicon sequences (Figure 1b), and composition was compared using nonmetric multidimensional scaling of a Unifrac taxonomic distance matrix (Lozupone *et al.,* 2011). The communities clustered in patterns that were generally consistent with CDOM and oxygen anomaly features (Figure 1d). Among the plume samples, four bacterial families accounted for 95% of the amplicon sequences. These had closest 16S rRNA sequence matches to Colwelliaceae (genus *Colwellia)*, Oceanospirillaceae (genera *Neptuniibacter* and *Bermanella*), Piscirickettsiaceae (genus *Cycloclasticus*) and Methylococcaceae (genera *Methylobacter* and *Methylococcus*) families (Figure 1b). The importance of these taxonomic groups in the samples was significantly positively correlated with increased concentrations of methane
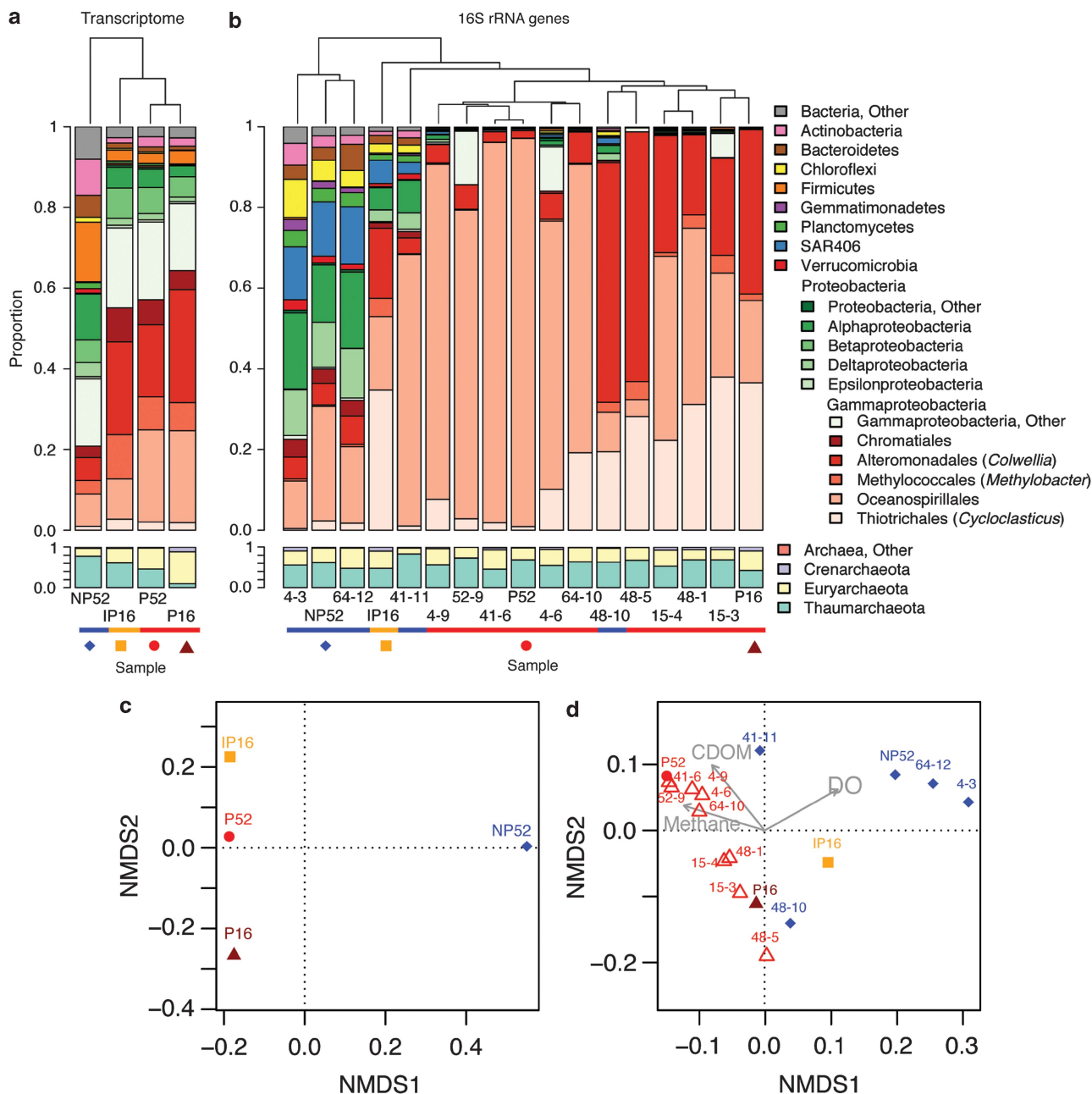
Figure 1 The taxonomic composition of the transcript pool based on ~23 million identified bacterial and archaeal transcripts (**a**) and the rRNA pool based on ~350 000 16S rRNA gene sequences (**b**). Transcript libraries are ordered based on single-linkage clustering of individual. genes, while 16S rRNA libraries are ordered based on single-linkage clustering of the Unifrac metric (Lozupone *et al.*, 2011). Taxa are vertically arranged according to the order shown in the key, and indentations in the key indicate bacterial groups at the class and order level. Sequence libraries were ordinated by nonmetric multidimensional scaling (**c** and **d**) using the same distance matrices and environmental variables fit to the 16S rRNA data.

and oil, and significantly negatively correlated with concentrations of dissolved oxygen ($P < 0.04$ for each factor, Figure 1d). In the non-plume sample NP52, the top four taxonomic groups were members of the family Oceanospirillaceae (28%), the class Alphaproteobacteria (11%), the SAR406 group (14%) and the class Deltaproteobacteria (11%).

The composition of the samples selected for metatranscriptome sequencing was representative of the broader microbial community. 16S rRNA

amplicons from the two plume samples, P16 and P52, and clustered with other plume samples (Figure 1). Sample NP52 was taxonomically and chemically distinct from plume samples, and clustered with non-plume samples, and sample IP16 had an intermediate taxonomic profile. The four dominant families indicative of the plume (see above) constituted 99% of 16S rRNA amplicons in plume samples P16 and P52, 75% in intermediate sample IP16 and 36% in non-plume sample NP52.

2320

Transcript composition

The 104 million joined (that is, with overlapping paired ends), quality-filtered reads had an average length of 217 bp (2.2×10^{10} bp total). Internal RNA standards added at the initiation of cell lysis (467, 576, 917 and 970 bp) were used to measure the error rate due to sequencing or mis-joining of the paired-end reads, and the combined error rate was estimated at 1.40% per bp, or three errors in an average 217-bp sequence. After removal of rRNAs, tRNAs and internal standards, 66 million possible protein-encoding sequences remained. Of those, ~ 23 million sequences were classified as one of 591 000 bacterial or archaeal genes using Blastx searches against the NCBI RefSeq database (Table 1). The rest did not have a hit in the database that met our cutoff criteria or were assigned to sequences from eukaryotes or viruses. RefSeq hits were matched to KEGG ortholog groups by Blastp, with 15 million sequences assigned to 5511 KEGG groups.

In the plume transcriptomes (P16 and P52), most reads binned to six reference genomes as follows: *Bermanella marisrubri* RED65 (19%), *Colwellia psychrerythraea* 34H (16%), *M. tundripaludum* SV96 (5.6%), *Methylococcus capsulatus* str. Bath (1.8%), *Neptuniibacter caesariensis* MED92 (1.2%) and *Methylophaga thiooxidans* DMS010 (1.1%) (Supplementary Table S1). Because sequences binning to a genome were quite divergent in some cases, these estimates are only approximate. For example, the *Cycloclasticus*-like bacterium that was abundant in 16S rRNA amplicon libraries from the plume had no reference genome available at the time of our analysis or that of Mason *et al.* (2012); transcripts from this taxon likely binned to the genome of a Gammaproteobacterial relative, with *M. tundripaludum* SV96 being the closest based on 16S rRNA gene similarity. Of the dominant plume taxa, the *Colwellia*-like reads most closely resembled their reference genome based on protein identity (86%) and consistency of genome coverage (Supplementary Figure S5). The transcripts that recruited to other reference genomes were more distantly related (75% identity for *B. marisrubri*, 72% for *M. tundripaludum*, 69% for *M. capsulatus*, 67% for *N. caesariensis* and 68% for *M. thiooxidans*). The non-plume transcriptome was more diverse. Reference genomes accumulating the most transcripts were *B. marisrubri* (5.5%), the archaeal ammonia oxidizer *Nitrosopumilus maritimus* (2.9%), *M. tundripaludum* (2.6%) and *C. psychrerythraea* (1.4%), with the remaining reads broadly distributed across multiple taxa.

Community metabolic reconstruction

Relative abundance of transcripts matching the KEGG database pathways for the oxidation of hydrocarbons was determined (Figure 2), and statistical comparisons were made between plume and non-plume samples (R package DESeq)

(Anders and Huber, 2010). With only a single non-plume sample (NP52), the dispersion calculations for the statistic were based only on the plume samples (P16 and P52; see Materials and methods) and significance values should be interpreted as approximate.

Alkanes were among the most abundant hydrocarbons measured in the Deepwater Horizon plume (Reddy *et al.*, 2012), and correspondingly transcripts for the enzyme alkane-1-monooxygenase, which degrades C_5–C_{13} n-alkanes, were abundant in the transcriptomes ($n = 7304$ total reads) (Figure 2). Monooxygenases and dioxygenases involved in aromatic compound degradation were significantly enriched in the plume transcriptomes ($P < 0.05$), including phenol-2 monooxygenase (involved in both phenol and toluene metabolism, *phn* genes), cyclohexanone monooxygenase, aryl alcohol dehydrogenase, benzoate 1,2-dioxygenase and biphenyl-2,3-diol 1,2, monooxygenase (Figure 2). Consistent with aromatic compounds comprising $< 2\%$ of the hydrocarbons from the Deepwater Horizon reservoir fluid (Reddy *et al.* 2012), the abundance of transcripts related to aromatic degradation was considerably lower than for alkane degradation. Mason *et al.* (2012) also reported high levels of alkane-1-monooxygenase and cyclohexanone monooxygenase in metatranscriptomes, and obtained a single-cell genome of an *Oceanospirillales* member with the capacity to degrade cyclohexanone.

Following initial oxidation, the hydrocarbons appeared to be incorporated into biomass primarily by the glyoxylate shunt of the tricarboxylic acid (TCA) cycle (Figure 2). In this pathway, acetyl-CoA (likely originating from the oxidation of alkanes via fatty acid beta oxidation or from aromatic compounds) are processed by isocitrate lyase and malate synthase, and both of these enzymes were significantly overrepresented in the plume transcriptomes. There was also evidence that bacterioplankton were using the methylcitrate pathway to metabolize propionyl-CoA generated from the oxidation of odd chain fatty acids, shorter fatty acids or possibly aromatic compounds (Figure 2).

Methane oxidation is initiated by the methane monooxygenase-mediated conversion of methane to methanol. Transcripts with homology to genes encoding methane monooxygenases were highly abundant in the transcriptomes from all samples (all subunits, $n = 74405$) and significantly enriched in the plume samples ($P < 0.05$) (Figure 2). Most of these were homologous to copper-containing *pmo* rather than soluble methane monooxygenases (Figure 2), as was also reported by Mason *et al.* (2012). The hits to *pmo* genes typically had low amino-acid identity with experimentally verified sequences; however, (Supplementary Figure S6), prompting construction of a phylogenetic tree with full-length monooxygenase assemblies from the GOM transcriptomes and *pmoA* homologs demonstrated experimentally to oxidize methane,
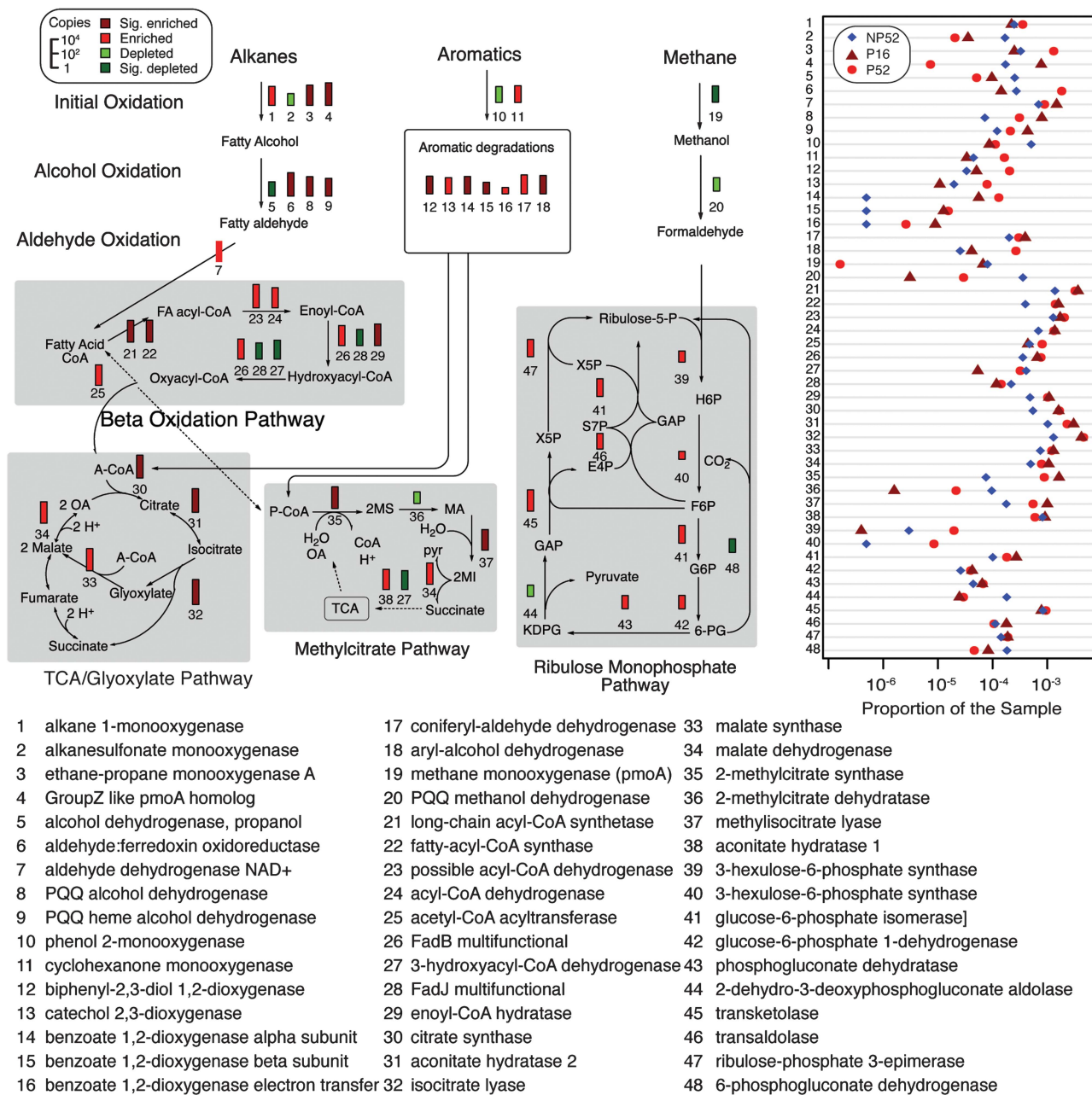
Figure 2 KEGG assignments to pathways related to hydrocarbon metabolism (left) and relative abundance of transcripts mapping to each step (right). The height of the bars in the pathway diagrams represents the number of transcripts (log scale at upper left) summed over samples P16, P52 and NP52, whereas the colors represent enrichment (red) or depletion (green) in the plume samples and darker colors indicate statistically significant differences.

1 alkane 1-monooxygenase	17 coniferyl-aldehyde dehydrogenase	33 malate synthase
2 alkanesulfonate monooxygenase	18 aryl-alcohol dehydrogenase	34 malate dehydrogenase
3 ethane-propane monooxygenase A	19 methane monooxygenase (pmoA)	35 2-methylcitrate synthase
4 GroupZ like pmoA homolog	20 PQQ methanol dehydrogenase	36 2-methylcitrate dehydratase
5 alcohol dehydrogenase, propanol	21 long-chain acyl-CoA synthetase	37 methylisocitrate lyase
6 aldehyde:ferredoxin oxidoreductase	22 fatty-acyl-CoA synthase	38 aconitate hydratase 1
7 aldehyde dehydrogenase NAD+	23 possible acyl-CoA dehydrogenase	39 3-hexulose-6-phosphate synthase
8 PQQ alcohol dehydrogenase	24 acyl-CoA dehydrogenase	40 3-hexulose-6-phosphate synthase
9 PQQ heme alcohol dehydrogenase	25 acetyl-CoA acyltransferase	41 glucose-6-phosphate isomerase]
10 phenol 2-monooxygenase	26 FadB multifunctional	42 glucose-6-phosphate 1-dehydrogenase
11 cyclohexanone monooxygenase	27 3-hydroxyacyl-CoA dehydrogenase	43 phosphogluconate dehydratase
12 biphenyl-2,3-diol 1,2-dioxygenase	28 FadJ multifunctional	44 2-dehydro-3-deoxyphosphogluconate aldolase
13 catechol 2,3-dioxygenase	29 enoyl-CoA hydratase	45 transketolase
14 benzoate 1,2-dioxygenase alpha subunit	30 citrate synthase	46 transaldolase
15 benzoate 1,2-dioxygenase beta subunit	31 aconitate hydratase 2	47 ribulose-phosphate 3-epimerase
16 benzoate 1,2-dioxygenase electron transfer	32 isocitrate lyase	48 6-phosphogluconate dehydrogenase

ammonia, ethane and propane (Figure 3a). The GOM metatranscriptomic assemblies fell into four clades as follows: Group 1 contained the canonical Gammaproteobacterial methanotroph clade, Group 2 contained the Group Z sequence with unknown function (Tavormina et al., 2010), Group 3 contained the ethane-oxidizing cluster as defined by strains ET-HIRO and ET-SHO (Suzuki et al., 2012), along with environmental ethane-assimilating clone pmo-7 (Redmond et al., 2010), and Group X clones (Tavormina et al., 2010), and Group 4 contained a cluster of highly divergent assemblies, most

closely related to the ethane-oxidizing Group 3. Mapping of the 33 285 pmoA-like short reads from the GOM metatranscriptome indicated that 50% of reads clustered with Group 2 (39% locus 180 + 11% locus 311), 42% with Group 3, 4% with Group 1 and 2% with Group 4. Thus reads assigned to M. tundripaludum and M. capsulatus by Blast analysis against existing genome sequences were distributed widely across the phylogenetic tree, with most resembling the Group Z-like pmoA genes of unknown function or an ethane dehydrogenase function (Group 3) (Figure 3b).
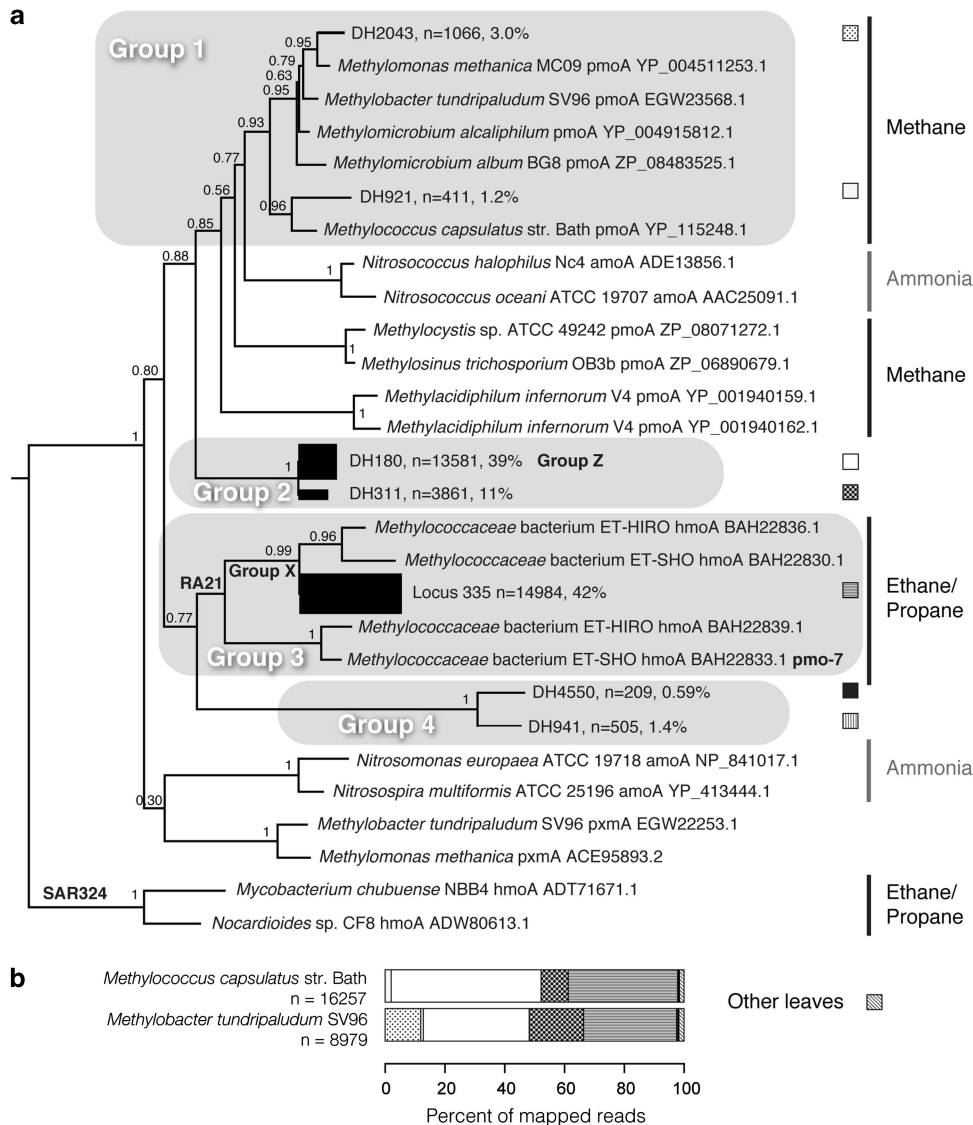
a

b

Figure 3 A maximum likelihood tree of *pmoA* homologs (**a**) and the proportion of reads originally assigned to the *M. capsulatus* or *M. tundripaludum* genome bins that cluster at each node on the tree (**b**). Five short *pmoA* reference sequences (in bold text) and 33 257 GOM reads identified as *pmoA* homologs by Blastx were placed on a reference tree made with full-length sequences. The thickness of the tree branches indicates the number of reads assigned.

The subsequent oxidation of methanol to formaldehyde is mediated by a PQQ-containing methanol dehydrogenase (Anthony, 1982), and homologs of this gene were significantly enriched in the plume transcriptomes ($P < 0.05$) (Figure 2). The taxonomic distribution again suggests that some genes in this category were used to oxidize methanol, but most to oxidize ethanol or larger alcohols (Supplementary Figure S7). In the case of the former, known pathways for the resultant formaldehyde include oxidation for energy or assimilation via the serine pathway (typical of alphaproteobacterial 'type II' and Gammaproteobacterial 'type X' methanotrophs) or the ribulose monophosphate pathway (RuMP; typical of Gammaproteobacteria 'type I' and 'type X' methanotrophs). In the GOM transcriptome, RuMP sequences were found in all samples, whereas three

key enzymes in the serine pathway (glycerate-2-kinase, hydroxypyruvate reductase and malate thiokinase) were either not found or were present in low abundance. The best-supported dissimilatory pathways for the generation of energy from methane oxidation were the dissimilatory branch of the RuMP pathway (used by most obligate type I methanotrophs) and the tetrahydrofolate pathway (used by obligate and facultative methylotrophs).

Taxon-specific metabolic reconstruction
The six bacterial genomes recruiting the most reads from the plume transcriptomes included three taxa specializing on non-C1 hydrocarbons (*C. psychrerythraea, B. marisrubri* and *N. caesariensis*) and three specializing on C1 compounds (*M. tundripaludum,*

M. capsulatus and *M. thiooxidans*). The genome of *B. marisrubri*, a bacterium related to known oil degraders, recruited the largest number of transcripts (3 409 771; Supplementary Table S1). These populations expressed several genes for alkane degradation, including alkane hydroxylase (*alkB*), rubredoxin (*alkG)* and rubredoxin reductase (*alkT*); of those mentioned genes, *alkG* was significantly enriched in the plume (Supplementary Tables S2, S4). An iron-containing alcohol dehydrogenase was also overrepresented in the plume and may be responsible for subsequent oxidation steps. Twenty flagellar proteins from this genome bin were enriched in plume samples and eight were depleted, suggesting that motility was regulated in response to the hydrocarbon inputs; several chemotaxis-related proteins were also enriched (Supplementary Table S4), and this was seen by Mason *et al.* (2012), as well.

The genome of Gammaproteobacterium *C. psychrerythraea*, a psychrophilic hydrocarbon degrader (Methe *et al.,* 2005), recruited 3 310 791 transcripts (Supplementary Table S1). The GOM population represented by this genome bin appeared to be consuming alkanes, as putative monooxygenases and dioxygenases identified in the *Colwellia* genome were enriched in samples collected within the plume (Supplementary Table S4). Homologs of known hydrocarbon-degrading genes included the ring-degrading genes *phnAB* and their related transcriptional regulators *phnS* and *phnR* (Supplementary Table S2). The PQQ alcohol dehydrogenase of *C. psychrerythraea* was relatively more abundant in the plume and a number of genes encoding its cofactor were significantly overrepresented (Supplementary Table S4). The genes isocitrate lyase and malate synthase were abundant in plume samples, whereas isocitrate dehydrogenase was underrepresented, indicating that the glyoxylate shunt of the TCA cycle was highly active in the *Colwellia*-like populations. Genes for motility, including three flagellar and five type IV pilus genes, were significantly overrepresented in the plume, as was the ferrous iron transporter *feoB*. Multiple subunits of cytochrome cbb(3) were enriched in the plume, a cytochrome with high affinity for oxygen that is typically expressed in suboxic conditions (Pitcher and Watmough, 2004).

In the *N. caesariensis* transcriptome bin (2 98 158 reads), homologs of the ring-hydroxylating dioxygenase *phnAb*, phenol-1-monooxygenase and the permease component of an ABC transporter for toluene suggested populations involved in the consumption of aromatic hydrocarbons (Supplementary Table S4). Quinoprotein-containing alcohol dehydrogenase transcripts were well expressed, as were glyoxylate shunt genes isocitrate lyase and malate synthase. One of the most abundant transcripts was for poly(R)-hydroxyalkanoic acid synthase, implying that *Neptuniibacter*-like populations were storing carbon derived from the hydrocarbons.

Among the C1-utilizing taxa, the largest number of transcripts was recruited to *M. tundripaludum* (1 408 081 reads; Supplementary Table S4). At least one of the populations recruiting to this genome was living at the expense of methane, based on the presence of transcripts clustering with the verified *pmo* genes (*pmoCAB*), although the dominance of transcripts for ethane monooxygenases and divergent particulate monooxygenases (see above) suggests that this reference genome captured a number of populations with varying substrate specialization. Methanol dehydrogenase sequences were evident (*mxaFI*), although read counts were two orders of magnitude lower than for *pmo*, consistent with the possibility that most *pmo*-like genes were degrading other short chain alkanes. All three pathways for oxidizing formaldehyde predicted in this genome (Svenning *et al.,* 2011) were represented in the transcriptome (the pterin-dependent pathways involving H4folate and H4MPT, as well as formaldehyde dehydrogenase), along with a formaldehyde-activating enzyme that accelerates the condensation of formaldehyde with H4MPT (Trotsenko and Murrell, 2008). Transcripts for carbon assimilation via the RuMP pathway were evident, as well as for all the steps in the TCA cycle. Although the *M. tundripaludum* reference genome contains genes for nitrogen fixation and denitrification, these were not represented in the transcripts of the GOM populations. Genes for nitrite transport and assimilation were represented, including nitrite reductase (*nirBD*). Overall, the plume populations contained relatively more transcripts for flagellum synthesis, chemotaxis, pilus assembly, and cobalamin and pterin synthesis than the non-plume sample (Supplementary Table S3).

The C1-utilizing genome of *M. capsulatus* recruited 4 53 969 reads from the GOM transcriptomes. The reads assigned to the two *pmo* operons (*pmoCAB*; 53 947 reads) also had greater similarity to non-methane oxidizing enzymes or those with unknown function (see above and Supplementary Table S4). There was lower but measurable expression of the soluble methane monooxygenase operon (378 reads; *mmoX,Y,Z,B,C,D*). Similar to *M. tundripaludum*, the *M. capsulatus* genome bin contained transcripts for all three types of formaldehyde oxidation in approximately equal numbers. *M. capsulatus* is categorized as a type X methanotroph based on evidence for low levels of enzymes from the serine pathway and Calvin cycle that may supplement carbon assimilation by the RuMP pathway (Hanson and Hanson, 1996), and transcript patterns in the GOM populations are consistent with this assignment (Supplementary Table S3). Whether or not *M. capsulatus* has a functional TCA cycle has been a point of controversy (Ward *et al.,* 2004; Kelly *et al.,* 2005), but the GOM populations were indeed operating the TCA cycle based on significant expression of the 2-oxoglutarate dehydrogenase E1 and E2 subunits. Nitrite reductase transcripts

suggest that nitrite was a major source of N; the nitrate transporter was not expressed and the ammonium transporter had low relative expression. Plume populations had greater relative expression of a fatty acid degradation gene (long chain fatty acid CoA ligase) (Supplementary Table S3).

The C1-utilizing *M. thiooxidans* (recruiting 266 672 reads) is capable of oxidizing methanol but not methane (Boden *et al.*, 2011). Transcripts for methanol dehydrogenase were present in this genome bin and the two formaldehyde oxidation pathways involving pterins (H4folate and H4MPT) were being transcribed. A surprisingly large number of genes encoding extracellular processes were among the most highly expressed genes in the *M. thiooxidans* genome bin, including type I, II and III secretion systems, a type IV pilus system, signal transduction and second messenger molecules, chemotaxis genes and LPS and lipid A synthesis. Transcripts for transport of both nitrite and ammonia were abundant, as were those for nitrite reductase.

Ecology of the plume community

The transcription shifts induced in the GOM bathypelagic bacterial community by the Deepwater Horizon hydrocarbons could represent either of the two ecological scenarios: (1) blooming hydrocarbon-utilizing taxa that competed directly (for example, for nutrients or substrates) with the pre-existing bacterioplankton community, and whose rapid growth displaced typical GOM bathypelagic bacteria; or (2) hydrocarbon-utilizing taxa that responded to the new substrates without displacing the pre-existing species, with the latter representing a smaller proportion of the total transcriptome yet not adversely affected in numbers or activity. To address this issue, we calculated the numbers of transcripts l^{-1} for each genome bin in each sample, based on recovery of the internal standards added during sample processing (see Materials and Methods). The 50 best-represented taxa in the NP transcriptome fell into four qualitative ecological categories based on abundance patterns (Figure 4).
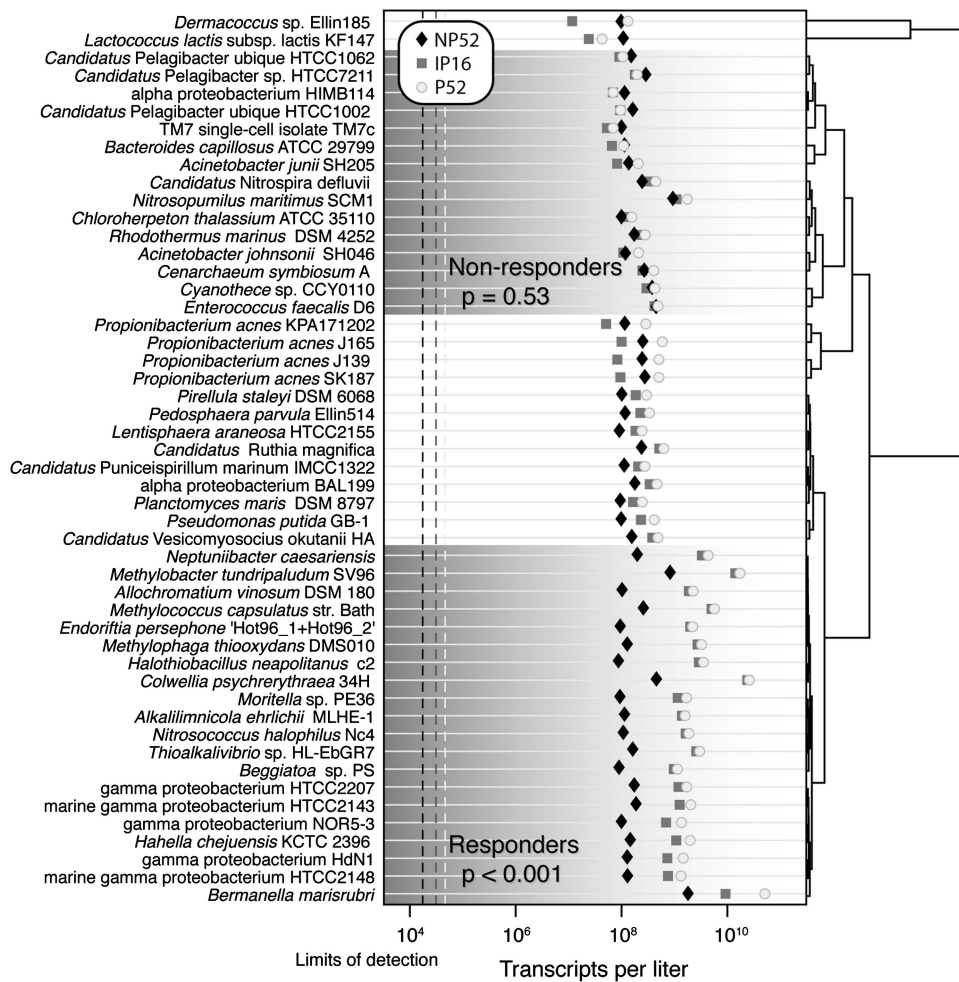

Figure 4 The transcripts per liter of seawater for the 50 most abundant bacterial and archaeal taxa in non-plume sample NP52. Taxa were clustered by the ratios of plume to non-plume transcripts, and major groupings are highlighted with shading. Sample P16 is not included because of an internal standard addition error. Transcripts from marine taxa without a close reference genome may be binned to an uninformative species name; for example, marine Actinobacteria sequences bin to *Propionibacterium* genomes and SAR406 transcripts to the *Rhodothermus marinus* genome. The dashed lines indicate the limit of detection for each transcriptome.

Approximately 1/3 of the taxa had significantly higher numbers of transcripts in the plume by up to two orders of magnitude (one-way analysis of variance, $P < 0.001$); this category consisted entirely of Gammaproteobacteria and included all six dominant bloom taxa (*Bermanella*, *Colwellia*, *Neptuniibacter*, *Methylobacter*, *Methylococcus* and *Methylophaga*) along with several members of the Oligotrophic Marine Gammaproteobacteria group (HTCC2207, HTCC2143, HTCC2148 and NOR5-3) and taxa affiliated with the Chromatiales (Figure 4). An additional ~1/4 of the taxa similarly had increased contribution to the plume transcriptome, but only by two- to fivefold; these included Actinobacteria, marine Rhodospirallales relatives, Planctomyces and groups related to sulfur-oxidizing Gammaproteobacterial symbionts. Transcript numbers for another 1/3 of the taxa, including SAR11 and marine Thaumarchaeota, were not detectibly affected by the presence of the fast growing Gammaproteobacteria in the plume (analysis of variance, $P = 0.53$). Finally, a surprisingly low 5% of the taxa, including two binning to Firmicutes genomes, had lower numbers of transcripts in the plume (Figure 4), apparently adversely affected by either competition or hydrocarbon toxicity.

Discussion

Microbial community composition and gene expression were profoundly altered by petroleum contamination, even in samples where plume signals were not detectable with *in situ* instrumentation. In NP52, CDOM fluorescence was below the 1 p.p.m. limit of detection for crude oil, yet 36% of 16S rRNA sequences and 17% of transcripts were from the four families that dominated plume-associated communities. These same organisms made up < 4% of a nearby bathypelagic community before the spill (King *et al.*, 2012). Methane was indeed elevated in the control sample compared with historical background concentrations (106 nM in NP52 versus ~2 nM background; Lamontagne *et al.*, 1973). Yet with concentrations still ~2000-fold less than the in-plume samples, the microbial community response was quite strong. Because sample NP52 contained low-level hydrocarbon contamination, differences in relative expression detected between impacted and non-impacted samples are likely to underestimate the effects of hydrocarbon inputs on bacterial gene expression.

mRNAs from the four metatranscriptomes mapped to over 2700 bacterial and archaeal taxa, but the similarity of reads to their reference sequences varied considerably. Of the six highest-recruiting genome bins, only the *Colwellia*-like sequences closely resembled the reference genome (87% average protein identity; Supplementary Figure S5), whereas fragment recruitment plots for the other five taxa show genome-wide protein identities of ~70%. The low sequence similarity to reference genomes affects interpretation of the metatranscriptomic data, as physiologies of the captured populations may vary considerably among each other, as well as compared with the sequenced representative.

One relevant example is that reads mapping to the *pmoCAB* genes of *M. tundripaludum* and *M. capsulatus* represent multiple clusters with varying percent identities and likely originating from several populations (Supplementary Figure S6). The evolution of bacterial monooxygenases is complex: *pmoA* genes are paraphyletic; the superfamily encodes enzymes targeting methane, ammonia, ethane and propane; and several newly discovered homologs have unknown function, including Group Z *pmoA* and *pxmA* (Tavormina *et al.*, 2011). Of the 33 000 reads annotated as *pmoA* based on RefSeq Blast analysis, only 4% clustered with experimentally verified methane oxidation genes, 50% clustered with a Group 2 (Group Z-like) *pmoA* of unknown function and 42% clustered with ET-SHO/ET-HIRO-like short chain hydrocarbon monooxygenases (Figure 3). Approximately 2% of the *pmoA* homologs were from a new, highly divergent group that has not been described previously (Group 4; Figure 3). Thus, the metatranscriptomic data cannot definitively address the question of the importance of methane oxidation in the Deepwater Horizon plume until the diverse *pmo*-like sequences are better characterized functionally.

Likewise, only 3% of the 36 000 PQQ and PQQ-heme dehydrogenase sequences that catalyze the next step in methane oxidation grouped with verified methanol dehydrogenases. Transcripts representing the alkane-1-monooxygenase (*alkB*), rubredoxin and rubredoxin reductase system responsible for degradation of C_6 to C_{13} n-alkanes were also abundant and enriched in the plume, whereas transcripts for oxidation of medium chain (C_{10}–C_{30}) alkanes (*alkM*) were particularly abundant in the *C. psychrerythraea* and *B. marisrubri* genome bins. With the exception of C_1 metabolism, transcripts counts for alkane degradation were negatively related to hydrocarbon length and positively related to hydrocarbon solubility and concentration.

Nitrate and nitrite transporters were abundant in the plume, and nitrite concentrations were elevated (Table 1). Concentrations of nitrate were generally lower in plume samples P16 and P52 and other hydrocarbon plume samples collected from the spill site (data not shown) compared with non-impacted samples, suggesting that dissimilatory nitrate reduction was occurring. Partial nitrification by ammonia oxidizers (such as *N. maritimus*) could also have been producing nitrite. The most abundant organism bins had enriched nitrite transporter transcripts and may have been taking advantage of the nitrite as a nutrient source. Nitrogen and phosphate did not appear to be limiting growth at
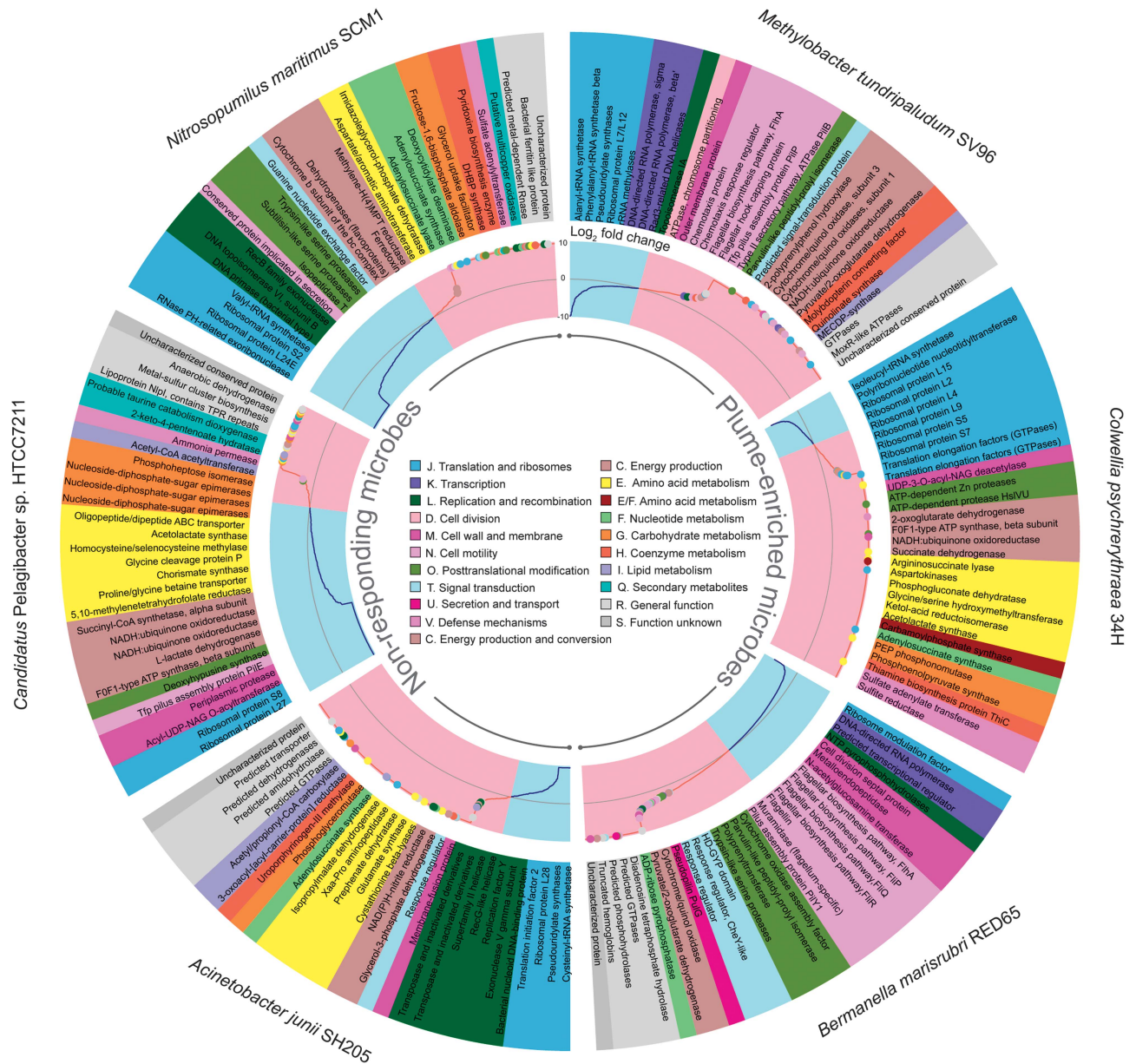
Figure 5 The expression of genes from three plume-enriched microbes (right half) and three non-responding microbes (left half). The inside track indicates log 2 fold change in abundance, with positive values indicating more transcripts in P16 and P52 after being normalized for the number of reads per sample and organism, and negatives values indicating more transcripts in NP52. Dots on the inside track indicate the genes present on the outside track. The outside track represents the 30 most significant genes that were enriched in the plume samples. Both are color-coded by COG category.

the time of sampling based on measured concentrations (Table 1).

The natural seepage of hydrocarbons from the GOM seafloor ($4–10 \times 10^{10}$ g per year; National Research Council, 2003) may set the context for the observed ecological response of the bathypelagic microbial community by sustaining a native community of hydrocarbon-degrading bacteria that quickly responds to periodic fluxes of oil and gas. In support of this idea, most of the *pmoA* homologs represented in the plume and non-plume transcriptomes were similar to the monooxygenases observed

at natural hydrocarbon seep sites and submarine canyons (Redmond *et al.*, 2010; Tavormina *et al.,* 2010; Lesniewski *et al.*, 2012; Suzuki *et al.*, 2012; Li *et al.*, 2013). Further, populations related to several of the bloom-forming taxa (*Cycloclasticus* and *Methylococcus*) were present in bathypelagic GOM seawater 1 month before the Deepwater Horizon event, although at low abundance (King *et al.,* 2012). On the other hand, the presence of taxa whose transcript abundance did not vary between plume and unexposed samples suggests the presence of a community of non-responding

bacterioplankton with a potential role in community structure recovery. Expression patterns of representatives of these non-responding microbes showed fewer differences in gene transcription patterns between control and plume samples than bloom taxa (Figure 5), and may have continued to occupy pre-spill niches in the petroleum-impacted seawater.

This metatranscriptomic analysis has revealed the primary pathways used for hydrocarbon degradation following the Deepwater Horizon accident, based on highly abundant transcripts for degradation of C2–C4 gaseous alkanes and C5–C12 alkanes, and less abundant but still considerable transcripts for degradation of C12–C18 alkanes, aromatic compounds and methane. Many of the monooxygenase sequences binned to methane monooxygenases from sequenced bacteria, but are more likely to be involved in short chain alkane oxidation based on recent studies from deep-sea seeps. The hydrocarbons were metabolized by a few key pathways, predominantly beta oxidation, glyoxylate, 2-methylcitrate and RuMP. A few groups of Gammaproteobacteria accounted for most of the bacterial response. Despite the enormous bloom of hydrocarbon-degrading Gammaproteobacteria that increased bacterial cell counts by two orders of magnitude, members of the natural microbial community persisted at their pre-bloom activity levels and may be important in the re-establishment of the original microbial community.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgements

References

Anders S, Huber W. (2010). Differential expression analysis for sequence count data. *Genome Biol* **11**: R106.

Anthony C. (1982). *The biochemistry of methylotrophs*. Academic Press: London, UK.

Benjamini Y, Hochberg Y. (1995). Controlling the false discovery rate - a practical and powerful approach to multiple testing. *J R Stat Soc Series B Stat Methodol* **57**: 289–300.

Boden R, Ferriera S, Johnson J, Kelly DP, Murrell JC, Schäfer H. (2011). Draft genome sequence of the chemolithoheterotrophic, halophilic methylotroph *Methylophaga thiooxydans* DMS010. *J Bacteriol* **193**: 3154–3155.

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.

Capella-Gutierrez S, Silla-Martinez JM, Gabaldon T. (2009). trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses *Bioinformatics* **25**: 1972–1973.

Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK *et al.* (2010). QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* **7**: 335–336.

Caspi R, Altman T, Dreher K, Fulcher CA, Subhraveti P, Keseler IM *et al.* (2012). The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. *Nucleic Acids Res* **40**: D742–D753.

DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL, Keller K *et al.* (2006). Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl Environ Microbiol* **72**: 5069–5072.

Eddy SR. (2011). Accelerated Profile HMM Searches. *PLoS Comput Biol* **7**: e1002195.

Falgueras J, Lara A. (2010). SeqTrim: a high-throughput pipeline for pre-processing any type of sequence read. *BMC Bioinformatics* **11**: 38.

Gifford SM, Sharma S, Rinta-Kanto JM, Moran MA. (2011). Quantitative analysis of a deeply sequenced marine microbial metatranscriptome. *ISME J* **5**: 461–472.

Han M, Zmasek C. (2009). phyloXML: XML for evolutionary biology and comparative genomics. *BMC Bioinformatics* **10**: 356.

Hanson RS, Hanson TE. (1996). Methanotrophic bacteria. *Microbiol Rev* **60**: 439–471.

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.

Jari O, GB F, Roeland K, Pierre L, Peter RM, BO R *et al.* (2013). *Vegan: Community Ecology Package.*

Joye SB, MacDonald IR, Leifer I, Asper V. (2011). Magnitude and oxidation potential of hydrocarbon gases released from the BP oil well blowout. *Nat Geosci* **4**: 160–164.

Kanehisa M, Goto S. (2000). KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* **28**: 27–30.

Katoh K, Misawa K, Kuma K, Miyata T. (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res* **30**: 3059–3066.

Kelley D. (2013). *Oce: Analysis of Oceanographic data.*

Kelly DP, Anthony C, Murrell JC. (2005). Insights into the obligate methanotroph *Methylococcus capsulatus*. *Trends Microbiol* **13**: 195–198.

Kessler JD, Valentine DL, Redmond MC, Du M, 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.

King GM, Smith CB, Tolar B, Hollibaugh JT (2012). Analysis of composition and structure of coastal to mesopelagic bacterioplankton communities in the northern Gulf of Mexico. *Front Microbio* **3**: 438.

Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D *et al.* (2009). Circos: an information aesthetic for comparative genomics. *Genome Res* **19**: 1639–1645.

Lamontagne RA, Swinnerton JW, Linnenbom VJ, Smith WD. (1973). Methane concentrations in various marine environments. *J Geophys Res* **78**: 5317–5324.

Le SQ, Lartillot N, Gascuel O. (2008). Phylogenetic mixture models for proteins. *Philos Trans R Soc Lond B Biol Sci* **363**: 3965–3976.

Lesniewski RA, Jain S, Anantharaman K, Schloss PD, Dick GJ. (2012). The metatranscriptome of a deep-sea hydrothermal plume is dominated by water column methanotrophs and lithotrophs. *ISME J* **6**: 2257–2268.

Li M, Jain S, Baker BJ, Taylor C, Dick GJ. (2013). Novel hydrocarbon monooxygenase genes in the metatranscriptome of a natural deep-sea hydrocarbon plume. *Environ Microbiol* e-pub ahead of print Doi:10.1111/1462-2920.12182.

Lozupone C, Lladser ME, Knights D, Stombaugh J, Knight R. (2011). UniFrac: an effective distance metric for microbial community comparison. *ISME J* **5**: 169–172.

Lu Z, Deng Y, Van Nostrand JD, He Z, Voordeckers J, Zhou A *et al.* (2011). Microbial gene functions enriched in the Deepwater Horizon deep-sea oil plume. *ISME J* **6**: 451–460.

Martin J, Bruno VM, Fang Z, Meng X, Blow M, Zhang T *et al.* (2010). Rnnotator: an automated de novo transcriptome assembly pipeline from stranded RNA-Seq reads. *BMC Genomics* **11**: 663.

Mason OU, Hazen TC, Borglin S, Chain PSG, Dubinsky EA, Fortney JL *et al.* (2012). Metagenome, metatranscriptome and single-cell sequencing reveal microbial response to Deepwater Horizon oil spill. *ISME J* **6**: 1715–1727.

Matsen FA, Kodner RB, Armbrust EV. (2010). Pplacer: linear time maximum-likelihood and Bayesian phylogenetic placement of sequences onto a fixed reference tree. *BMC Bioinformatics* **11**: 538.

Methe BA, Nelson KE, Deming JW, Momen B, Melamud E, Zhang XJ *et al.* (2005). The psychrophilic lifestyle as revealed by the genome sequence of *Colwellia psychrerythraea* 34H through genomic and proteomic analyses. *Proc Natl Acad Sci USA* **102**: 10913–10918.

National Research Council. (2003). *Oil in the Sea III: Inputs, Fates, and Effects*. National Academies Press: Washington, DC.

Pitcher RS, Watmough NJ. (2004). The bacterial cytochrome cbb3 oxidases. *Biochim Biophys Acta* **1655**: 388–399.

Poretsky RS, Gifford S, Rinta-Kanto J, Vila-Costa M, Moran MA. (2009). Analyzing gene expression from marine microbial communities using environmental transcriptomics. *J Vis Exp* **3**: 1–6.

Reddy CM, Arey JS, Seewald JS, Sylva SP, Lemkau KL, Nelson RK *et al.* (2012). Composition and fate of gas and oil released to the water column during the Deepwater Horizon oil spill. *Proc Natl Acad Sci USA* **109**: 20229–20234.

Redmond MC, Valentine DL. (2012). Natural gas and temperature structured a microbial community response to the Deepwater Horizon oil spill. *Proc Natl Acad Sci USA* **109**: 20292–20297.

Redmond MC, Valentine DL, Sessions AL. (2010). Identification of novel methane-, ethane-, and propane-oxidizing bacteria at marine hydrocarbon seeps by stable isotope probing. *Appl Environ Microbiol* **76**: 6412–6422.

Reeder J, Knight R. (2010). Rapidly denoising pyrosequencing amplicon reads by exploiting rank-abundance distributions. *Nat Methods* **7**: 668–669.

Rodrigue S, Materna AC, Timberlake SC, Blackburn MC, Malmstrom RR, Alm EJ *et al.* (2010). Unlocking short read sequencing for metagenomics. *PLoS ONE* **5**: e11840.

Ryerson TB, Camilli R, Kessler JD, Kujawinski EB, Reddy CM, Valentine DL *et al.* (2012). Chemical data quantify Deepwater Horizon hydrocarbon flow rate and environmental distribution. *Proc Natl Acad Sci USA* **109**: 20246–20253.

Satinsky BM, Gifford SM, Crump BC, Moran MA. (2013). Use of internal standards for quantitative metatranscriptome and metagenome analysis. *Meth Enzym* (in press).

Stamatakis A. (2006). RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22**: 2688–2690.

Stewart FJ, Ottesen EA, DeLong EF. (2010). Development and quantitative analyses of a universal rRNA-subtraction protocol for microbial metatranscriptomics. *ISME J* **4**: 896–907.

Suzuki T, Nakamura T, Fuse H. (2012). Isolation of two novel marine ethylene-assimilating bacteria, Haliea species ETY-M and ETY-NAG, containing particulate methane monooxygenase-like genes. *Microbes Environ* **27**: 54–60.

Svenning MM, Hestnes AG, Wartiainen I, Stein LY, Klotz MG, Kalyuzhnaya MG *et al.* (2011). Genome sequence of the Arctic methanotroph *Methylobacter tundripaludum* SV96. *J Bacteriol* **193**: 6418–6419.

Tatusov RL, Fedorova ND, Jackson JD, Jacobs AR, Kiryutin B, Koonin EV *et al.* (2003). The COG database: an updated version includes eukaryotes. *BMC Bioinformatics* **4**: 41.

Tavormina PL, Orphan VJ, Kalyuzhnaya MG, Jetten MSM, Klotz MG. (2011). A novel family of functional operons encoding methane/ammonia monooxygenase-related proteins in gammaproteobacterial methanotrophs. *Environ Microbiol Rep* **3**: 91–100.

Tavormina PL, Ussler W, Joye SB, Harrison BK, Orphan VJ. (2010). Distributions of putative aerobic methanotrophs in diverse pelagic marine environments *ISME J* **4**: 700–710.

Trotsenko YA, Murrell JC. (2008). Metabolic aspects of aerobic obligate methanotrophy. *Adv Appl Microbiol* **63**: 183–229.

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.

Wang Q, Garrity GM, Tiedje JM, Cole JR. (2007). Naive Bayesian classifier for rapid assignment of rRNA

sequences into the new bacterial taxonomy. *Appl Environ Microbiol* **73**: 5261–5267.

Ward N, Larsen O, Sakwa J, Bruseth L, Khouri H, Durkin AS *et al.* (2004). Genomic insights into methanotrophy: the complete genome sequence of *Methylococcus capsulatus* (Bath). *PLoS Biol* **2**: e303.

Yvon-Lewis SA, Hu L, Kessler J. (2011). Methane flux to the atmosphere from the Deepwater Horizon oil disaster. *Geophys Res Lett* **38**: L01602.

Zhu W, Lomsadze A, Borodovsky M. (2010). Ab initio gene identification in metagenomic sequences. *Nucleic Acids Res* **38**: e132.

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