Metagenomics, metatranscriptomics and single cell genomics reveal functional response of active Oceanospirillales to Gulf oil spill

Olivia U. Mason[a], Terry C. Hazen[a], Sharon Borglin[a], Patrick S. G. Chain[b,c], Eric A. Dubinsky[a], Julian L. Fortney[a], James Han[b,c], Hoi-Ying N. Holman[a], Jenni Hultman[a], Regina Lamendella[a], Rachel Mackelprang[c], Stephanie Malfatti[c,d], Lauren M. Tom[a], Susannah G. Tringe[c], Tanja Woyke[c], Jizhong Zhou[e,f], Edward M. Rubin[c], and Janet K. Jansson[a,c,*]

[a] *Ecology Department, Lawrence Berkeley National Laboratory, Berkeley National Lab, CA 94720*

[b] *Metagenomics Applications Team, Genome Science Group, Los Alamos National Laboratory, Los Alamos, NM 87545*

[c] *Department of Energy Joint Genome Institute, Walnut Creek, CA 94598*

[d] *Biosciences and Biotechnology Division, Lawrence Livermore National Laboratory, Livermore, CA 94550*

[e] *Institute for Environmental Genomics and Department of Botany and Microbiology, University of Oklahoma, OK 73072*

[f] *School of Environment, Tsinghua University, Beijing 100084*

**To whom correspondence may be addressed: Dr. Janet K. Jansson, Lawrence Berkeley National Laboratory, M/S 70A-3317, One Cyclotron Road, Berkeley, CA 94720*. Email: jrjansson@lbl.gov

13 May 2012

ACKNOWLEDGMENTS:

This work was supported by a subcontract from the University of California at Berkeley, Energy Biosciences Institute (EBI) to Lawrence Berkeley National Laboratory under its U.S. Department of Energy contract DE-AC02-05CH11231. The SR-FTIR work was conducted at the infrared beamline at the Advanced Light Source, which is supported by the Director, Office of Science, Office of Basic Energy Sciences, of the U.S. Department of Energy. We are extremely grateful to Michael D. Curtin for his assistance with Paracel BLAST. We are thankful for the help of Theresa Pollard with handling shipping, ordering, and transportation of supplies and people to and from the field. We also thank the captain, crew, and science teams aboard the R/V Ocean Veritas and the R/V Brooks McCall. Lastly, we thank Damon Tighe and Janey Lee at JGI for

single cell sorting and amplification and Chien-Chi and Matthew Scholz for metagenome assemblies

DE-AC02-05CH11231 from the Office of Science of the U.S. Department of Energy to the DOE Joint Genome Institute.

DISCLAIMER:

Metagenomics, metatranscriptomics and single cell genomics reveal functional response of active *Oceanospirillales* to Gulf oil spill

Authors: Olivia U. Mason[a], Terry C. Hazen[a], Sharon Borglin[a], Patrick S. G. Chain[b,c], Eric A. Dubinsky[a], Julian L. Fortney[a], James Han[b,c], Hoi-Ying N. Holman[a], Jenni Hultman[a], Regina Lamendella[a], Rachel Mackelprang[c], Stephanie Malfatti[c,d], Lauren M. Tom[a], Susannah G. Tringe[c], Tanja Woyke[c], Jizhong Zhou[e,f], Edward M. Rubin[c], and Janet K. Jansson[a,c,*]

Affiliations:
[a]Ecology Department, Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, CA 94720, USA
[b]Metagenomics Applications Team, Genome Science Group Los Alamos National Laboratory, Los Alamos, NM 87545, USA
[c]Department of Energy Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
[d]Biosciences and Biotechnology Division, Lawrence Livermore National Laboratory, Livermore, CA 94550, USA
[e]Institute for Environmental Genomics and Department of Botany and Microbiology, University of Oklahoma, OK, 73072, USA
[f]School of Environment, Tsinghua University, Beijing 100084

*Corresponding Author:
Dr. Janet K. Jansson
MS 70A-3317
One Cyclotron Rd.
Lawrence Berkeley National Laboratory
Berkeley, CA
94720
Email: jrjansson@lbl.gov
Phone: 510-486-7487

Abstract

The Deepwater Horizon oil spill in the Gulf of Mexico resulted in a deep-sea hydrocarbon plume that caused a shift in the indigenous microbial community composition with unknown ecological consequences. Early in the spill history a bloom of uncultured, thus uncharacterized, members of the *Oceanospirillales* was previously detected, but their role in oil disposition was unknown. Here our aim was to determine the functional role of the *Oceanospirillales* and other active members of the indigenous microbial community using deep sequencing of community DNA and RNA, as well as single-cell genomics. Shotgun metagenomic and metatranscriptomic sequencing revealed that genes for motility, chemotaxis, and aliphatic hydrocarbon degradation were significantly enriched and expressed in the hydrocarbon plume samples compared to uncontaminated seawater collected from plume depth. By contrast, although genes coding for degradation of more recalcitrant compounds such as benzene, toluene, ethylbenzene, total xylenes, and polycyclic aromatic hydrocarbons were identified in the metagenomes, they were expressed at low levels, or not at all based on analysis of the metatranscriptome. Isolation and sequencing of two *Oceanospirillales* single cells revealed that both cells possessed genes coding for n-alkane and cycloalkane degradation. Specifically, the near complete pathway for cyclohexane oxidation in the *Oceanospirillales* single cells was elucidated and supported by both metagenome and metatranscriptome data. The draft genome also included genes for chemotaxis, motility, and nutrient acquisition strategies that were also identified in the metagenomes and metatranscriptomes. These data point towards a mechanism for rapid response of members of the *Oceanospirillales* to aliphatic hydrocarbons in the deep-sea.

Introduction

On April 20, 2010 the Deepwater Horizon oil rig exploded and sank resulting in an unremitting flow of oil from April to July 2010 into the Gulf of Mexico, for a total of approximately 4.9 million barrels (779 million liters) ±10% (Command, 2010). The MC252 oil fraction was comprised of a complex mixture of hydrocarbons including saturated hydrocarbons (74%), aromatic hydrocarbons (16%), including polycyclic aromatic hydrocarbons (PAH), which reached maximal concentrations of 1200 μg/L at the surface (3) and polar hydrocarbons (10%) (Reddy et al, 2011). During the spill an oil plume was detected at depths of approximately 1000-1300 m (Camilli et al, 2010, Hazen et al, 2010). The deep-sea oil plume was reported to contain gaseous components (Kessler et al, 2011, Valentine et al, 2010), as well as non-gaseous, more recalcitrant compounds such as benzene, toluene, ethylbenzene, and total xylenes (BTEX) at concentrations ranging from 50-150 μg/L (Camilli et al, 2010, Hazen et al, 2010). This influx of hydrocarbons significantly impacted the indigenous microbial community structure (Hazen et al, 2010, Kessler et al, 2011, Redmond and Valentine, 2011, Valentine et al, 2010), including enrichment of uncultivated members of the *Oceanospirillales* early in the spill history (Hazen et al, 2010, Redmond and Valentine, 2011). The lack of a cultivated isolate of the *Oceanospirillales* from the plume precluded a clear understanding of the direct physiological and ecological consequences of the hydrocarbons on this group of microorganisms.

The documented shifts in the microbial community structure over time in response to the deep-sea plume of hydrocarbons have been shown by DNA based

methods such as cloning and sequencing of 16S rRNA genes (Hazen et al, 2010, Kessler et al, 2011, Redmond and Valentine, 2011, Valentine et al, 2010) and microarray analysis of functional genes (Lu et al, 2011). Cloning and sequencing revealed a clear temporal succession of Bacteria in the deep-sea hydrocarbon plume from a community dominated by *Oceanospirillales* (Hazen et al, 2010, Redmond and Valentine, 2011) to *Colwellia* and *Cycloclasticus* (Redmond and Valentine, 2011, Valentine et al, 2010), and finally to methylotrophic bacteria (Kessler et al, 2011). To date, however, no deep-sequencing approach has been used to analyze the microbial community structure, including rare members of the community, and their function. In addition, there is no information about what microorganisms were active or which functional genes were actually expressed in response to the oil spill.

Here we aimed to determine the specific roles of the *Oceanospirillales* that were enriched in the plume early in the spill history. In addition, we aimed to determine which functional genes and pathways were expressed in the deep-sea plume. To address these aims we not only analyzed the functional gene repertoire in total DNA extracted from metagenomic samples, we also extracted and sequenced total RNA metatranscriptomes to determine which genes were highly expressed and representative of active members of the community. In addition, to specifically characterize the functional roles of the dominant *Oceanospirillales* we isolated and sequenced single representative cells. For all of these analyses we used the Illumina sequencing platform, which resulted in over 60 GB of data. To analyze these large datasets, including, raw, unassembled reads and to integrate the different 'omics' we used several novel bioinformatics approaches, which are outlined in Figure 1. For this study, we focused on samples that were collected

during the oil spill between May 27-31, 2010 (Hazen et al, 2010) for in-depth phylogenetic and functional analyses: two plume samples, one proximal (1.5 km from the wellhead) and one distal (11 km from the wellhead), and one uncontaminated sample collected at plume depth (40 km from the wellhead) (Figure S1).

Methods

Sample Collection

From each station 1-5 L of seawater were filtered through a 0.2 μm diameter filters from the Gulf of Mexico during two monitoring cruises from May 27-June 2 2010 on the R/V Ocean Veritas and R/V Brooks McCall. Detailed information regarding sample collection can be found in Hazen, et al. (Hazen et al, 2010).

DNA Extraction

DNA was extracted from microbial cells collected onto filters using a modified Miller method (Miller et al, 1999), with the addition of a pressure lysis step to increase cell lysis efficiency. One half of each filter was placed into a Pressure Biosciences FT500 Pulse Tube (Pressure Biosciences, Easton, MA). 300 μL of Miller phosphate buffer and 300μL of Miller SDS lysis buffer were added and mixed. 600 μL phenol:chloroform:isoamyl alcohol (25:24:1) was then added. The samples were subjected to pressure cycling at 35,000 psi for 20 s and 0 psi for 10 s for a total of 20 cycles using the Barocycler NEP3229 (Pressure Biosciences). After pressure cycling the sample material was transferred to a Lysing Matrix E tube (MP Biomedicals, Solon, OH) and the samples were subjected to bead beating at 5.5m/s for 45sec in a FastPrep

131 instrument (MP Biomedicals). The tubes were centrifuged at 16,000 x g for 5 min at 4°C,
132 540 µL of supernatant was transferred to a 2 ml tube and an equal volume of chloroform
133 was added. The individual samples were mixed by inversion and then centrifuged at
134 10,000 x g for 5 min. 400 µL of the aqueous phase was transferred to another tube and 2
135 volumes of Solution S3 (MoBio, Carlsbad, CA) were added and mixed by inversion. The
136 rest of the clean-up procedures followed the instructions in the MoBio Soil DNA
137 extraction kit. Samples were recovered in 60µL Solution S5 and stored at -20°C.
138
139 *16S rRNA gene sequencing and analysis*
140 16S rRNA gene sequences were amplified from the DNA extracts using the
141 primer pair 926wF (5'-AAACTYAAAKGAATTGRCGG-3') and 1392R as previously
142 described (Kunin et al, 2010). The reverse primer included a 5 bp barcode for
143 multiplexing of samples during sequencing. Emulsion PCR and sequencing of the PCR
144 amplicons was performed at DOE's Joint Genome Institute (JGI) following
145 manufacturer's instructions for the Roche 454 GS Titanium technology (Allgaier et al,
146 2010). A total of 87,000 pyrotag sequences were obtained and analyzed using QIIME
147 (Caporaso et al, 2010b). Briefly, 16S rRNA gene sequences were clustered with uclust
148 (Edgar, 2010) and assigned to operational taxonomic units (OTUs) with 97% similarity.
149 Representative sequences from each OTU were aligned with Pynast (Caporaso et al,
150 2010a) using the Greengenes (DeSantis et al, 2006) core set. Taxonomy was assigned
151 using the Greengenes 16S rRNA gene database. As the number of sequence reads in each
152 sample varied, the dataset was rarified prior to alpha diversity calculations.
153

RNA extraction and amplification

Immediately following sampling and filtration at the proximal sampling station, samples intended for RNA extractions were placed in RNAlater (Ambion, Foster City, CA) to prevent RNA degradation. Total RNA was extracted as previously described (DeAngelis et al, 2010) and amplified using the Message Amp II-Bacteria Kit (Ambion) following the manufacturers' instructions. First strand synthesis of cDNA from the resulting antisense RNA was carried out with the SuperScript III First Strand Synthesis System (Invitrogen, Carlsbad, CA). The SuperScript Double-Stranded cDNA Synthesis Kit (Invitrogen) was used to synthesize double stranded cDNA. cDNA was purified using a QIAquick PCR purification kit (Qiagen, Valencia, CA). Poly(A) tails were removed by digesting purified DNA with *Bpm*I for 3 h at 37°C. Digested cDNA was purified with QIAquick PCR purification kit (Qiagen).

Emulsion PCR

To increase yields required for sequencing, DNA and cDNA were amplified by emulsion PCR. A detailed description of this method can be found in Blow, et al. (Blow et al, 2008). Briefly, DNA for metagenomic samples was sheared (cDNA was not sheared) using the Covaris S-Series instrument (Covaris, Woburn, MA). DNA and cDNA were end-repaired using the End-It DNA End-Repair Kit (EPICENTRE Biotechnologies, Madison, WI). End-repaired DNA and cDNA was then ligated with Illumina Paired End Adapters 1 and 2. For each sample 10 ng was used for emulsion PCR. Emulsion PCR reagents and thermal cycler protocols were as previously described (Blow et al, 2008). Amplified products were cleaned with a PCR mini-elute column (Qiagen), visualized,

177 and ~300 bp fragments were excised from a 2% agarose gel.

178

179 *Sequencing*

180 Metagenomic shotgun sequencing libraries of the samples were sequenced using
181 the Illumina GAIIx 2x 114 bp pair-end technology. The Illumina sequencing platform
182 was used to generate 14-17 Gb of sequence data per sample.

183 cDNA was sequenced using the Illumina GAIIx sequencing platform. cDNA was
184 quantified and clustered accordingly onto one lane of a flow cell on Illumina's cBot
185 Cluster Generation System. After cluster generation, the flow cell was transferred to a
186 GAIIx and was sequenced for 100 cycles for read 1. Then, turnaround chemistry was
187 performed by the Paired End Module, which prepared the flow cell for read 2 sequencing.
188 Another 100 cycles of sequencing followed to result for 100 bp paired-end reads.

189

190 *Sequence assembly and analysis*

191 Raw Illumina metagenomic reads (~113 bp in length) were trimmed using a
192 minimum quality cutoff of 3. Both trimmed and untrimmed reads were kept for further
193 assembly. Paired-end Illumina reads were assembled using SOAP*denovo*
194 (http://soap.genomics.org.cn/soapdenovo.html) at a range of Kmers (21,23,25,27,29,31)
195 for both trimmed and untrimmed reads. Default settings for all SOAPdenovo assemblies
196 were used (flags: –d 1 and –R). Contigs generated by each assembly (12 total contig
197 sets), were merged using a combination of in-house Perl script. Contigs were then sorted
198 into two pools based on length. Contigs smaller than 1800 bp were assembled using
199 Newbler (Life Technologies, Carlsbad, CA) in an attempt to generate larger contigs

(flags: -tr, -rip, -mi 98, -ml 60). All assembled contigs larger than 1800 bp, as well as the contigs generated from the final Newbler run were combined using minimus 2 (AMOS:, http://sourceforge.net/projects/amos) and the default parameters for joining. Minimus2 is an overlap based assembly tool that is useful for combining low numbers of longer sequences, as are found in assembled contigs. Assembly of the total of 368 million paired-end quality filtered metagenome sequence reads that averaged 113 bp in length (45 Gb) resulted in 1.1 million contigs. These contigs had an average N50 length of 382 bp (N50 is the length of the smallest contig in the set of largest contigs that have a combined length that represents at least 50% of the assembly (Miller et al, 2010)). Assembled data was annotated in IMG (Markowitz et al, 2008). COG annotations for both plume samples and the uncontaminated sample, including average fold, were exported. A pairwise statistical comparison of COGs in each of the three samples was carried out using STAMP (Parks and Beiko, 2010). Raw Illumina metatranscriptomic reads (~100 bp in length) were assembled using the CLC Genomics Workbench (version 4.0.3, CLC Bio). Paired end reads were assembled using the following parameters: mismatch cost 2, insertions cost 3, deletion cost 3, length fraction 0.5, similarity 0.8. The minimum contig length was set to 200 bp. Assembled metatranscriptomic data was annotated using CAMERA (v2.0.6.2) (Seshadri et al, 2007).

blastn

Single reads from each metagenomic and metatranscriptomic sample was searched against the Greengenes (DeSantis et al, 2006) database of 16S rRNA genes

using blastn with an a bit score cutoff of > 100. For each sequence the blast result with the highest bit score was selected.

tblastn

Raw metagenomic, metatranscriptomic, and single cell reads were searched against a subset of proteins (~12,000) involved in hydrocarbon degradation from the GeoChip (He et al, 2010) database. This database was selected because, to our knowledge, this is the only curated database of nearly complete pathways for hydrocarbon degradation. Paracel blast was used with the tblastn algorithm allowing all possible hits and using a bit score cutoff of > 40. For each sequence the blast result with the highest bit score was selected. While putative and potential proteins were part of the overall database searched, only characterized proteins were included in the final data analysis and presentation. A pairwise statistical comparison of the results of the metagenomic and metatranscriptomic blast analyses was carried out using STAMP (Parks and Beiko, 2010) using the a two-sided Chi-square test (with Yates) statistic with the DP: Asymptotic-CC confidence interval method and the Bonferroni multiple test correction. A p-value of > 0.05 was used with a double effect size filter (difference between proportions effect size < 1.00 and a ratio of proportions effect size < 2.00.

Single-cell sorting, whole genome amplification and screening

Cells were collected following the clean sorting procedures detailed by Rodrigue et al. (Rodrigue et al, 2009). Briefly, single cells from the proximal plume water sample were sorted by the Cytopeia Influx Cell Sorter (BD Biosciences, Franklin Lakes, NJ) into

three 96 well plates containing 3 µl of UV treated TE. The cells were stained with SYBR Green I (Invitrogen) and illuminated by a 488 nm laser (Coherent Inc., Santa Clara, CA). The sorting window was based on size determined by side scatter and green fluorescence (531/40 bp filter). Single cells were lysed for 20 min at room temperature using alkaline solution from the Repli-G UltraFast Mini Kit (Qiagen) according to manufacturer's instructions. After neutralization, the samples were amplified using the RepliPHI Phi29 reagents (EPICENTRE Biotechnologies). Each 50 µl reaction contained Phi29 Reaction Buffer (1X final concentration), 50 mM random hexamers with the phosphorothioate bonds between the last two nucleotides at the 3' end d (IDT), 0.4 mM dNTP, 5% DMSO (Sigma, St. Louis, MO), 10 mM DTT (Sigma), 100 U Phi29 and 0.5 mM Syto 13 (Invitrogen). A mastermix of MDA reagents minus the Syto 13 sufficient for a 96-well plate was UV treated for 60 min for decontamination. Syto 13 was then added to the mastermix, which was added to the single cells for real time multiple displacement amplification (MDA) on the Roche LightCycler 480 for 17 hours at 30ºC. All steps of single cell handling and amplification were performed under most stringent conditions to reduce the introduction of contamination. Single cell MDA products were screened using Sanger sequencing of 16S rRNA gene amplicons derived from each MDA product. A total of 16 *Oceanospirillales* cells were obtained. Three single amplified genome (SAGs) were identified as being 95% similar to the dominant *Oceanospirillales* OTU, and of high sequence quality (16S rRNA gene) and pursued for whole genome sequencing.

Single cell Illumina sequencing, QC and assembly

Single cell amplified DNA of three *Oceanospirillales* cells was used to generate normalized, indexed Illumina libraries. Briefly, 3 μg of MDA product was sheared in 100 μl using the Covaris E210 (Covaris) with the setting of 10% duty cycle, intensity 5, and 200 cycle per burst for 6 min per sample and the fragmented DNA purified using QIAquick columns (Qiagen) according to the manufacturer's instructions. The sheared DNA was end-repaired, A-tailed, and ligated to the Illumina adaptors according to the Illumina standard PE protocol. The ligation product was purified using AMPure SPRI beads, then underwent normalization using the Duplex-Specific Nuclease Kit (Axxora, San Diego, CA). The normalized libraries were then amplified by PCR for 12 cycles using a set of two indexed primers and the library pool was sequenced using an Illumina GAIIx sequencer according to the manufacturer's protocols (run mode 2x150 bp). Approximately 2.5 Gbp (16,797,846 reads) of sequence data was collected from the *Oceanospirillales* single cell genomes. The Illumina SAG data was QC'd using GC content and blast analysis and no contamination was detectable in two of the SAGs, while the third SAG was excluded from the analysis due to the presence of contaminating sequences. Reads from these two single cells were assembled using Velvet (Zerbino and Birney, 2008). To estimate genome sequence completeness the annotated, assembled draft genome data was compared to core COGs for Proteobacteria and Gammaproteobacteria (number of identified core COGs/number of expected core COGs).

Mapping and analysis

Unassembled metatranscriptomic reads were mapped to the *Oceanospirillales* single cell draft genome using the CLC Genomics Workbench (CLC bio) using the

following parameters: mismatch cost 2, insertions cost 3, deletion cost 3, length fraction 0.5, similarity 0.8. Assembled single cell data was annotated using CAMERA (v2.0.6.2) (Seshadri et al, 2007). The Interactive Pathways Explorer (Letunic et al, 2008) v2 was used to map the assembled, annotated metatranscriptome with an assembled, annotated *Oceanospirillales* single cell draft genome. Clustered regularly interspaced short palindromic repeat (CRISPR) regions were identified in the draft genome using CRISPRFinder (Grissa et al, 2007).

Infrared spectromicroscopy and data processing

SR-FTIR measurements and analyses were conducted at the infrared beamline of the Advanced Light Source (http://infrared.als.lbl.gov/) on thin layers of fresh samples placed between a gold-coated Si wafer and a SiN_x window. Photons emitted over a mid-infrared wavenumber range of 4000 to 650 cm^{-1} were focused through the samples by the Nicolet Nic-Plan IR microscope (with a numerical aperture objective of 0.65), which was coupled to a Nicolet Magna 760 FTIR bench (Thermo Scientific Inc., MA, USA). The entire view-field was typically divided into equal-sized 2-μm×2-μm squares before raster scanning. The SR-FTIR transflectance spectra at each position were collected using a single-element MCT detector at a spectral resolution of 4 cm^{-1} with 32 co-added scans and a peak position accuracy of 1/100 cm^{-1}. In transflectance, the synchrotron infrared beam transmitted through the cells, reflected off the gold-coated surface, and then transmitted through the sample a second time before reaching the detector. Background spectra were acquired from neighboring locations without any cells, and used as reference spectra for both samples and standards to remove background H_2O and CO_2 absorptions. Background spectra were obtained and used as reference spectra for both

314 samples and standards to remove background H_2O and CO_2 absorptions. All SR-FTIR
315 transflectance spectra were subjected to an array of data preprocessing and processing
316 calculations using Thermo Electron's Omnic version 7.3. The processing includes the
317 computation conversion of transflectance to absorbance, spectrum baseline removal, and
318 univariate analysis. In the univariate analysis, the calculated infrared absorbance at each
319 wavenumber in the mid-infrared region can also be related to the relative concentration of
320 a particular chemical component through the Beer-Lambert Law. Because analysis of
321 each spectral absorption band provides a single absorption value (representing the
322 relative abundance of a chemical component), we also constructed two-dimensional
323 images to visualize the relative abundance of petroleum products and microbial
324 biomolecules.
325
326 *Hydrocarbon analysis*
327 The profile of Macondo crude oil (collected 5/22/10 directly from the Discovery
328 Enterprise drill ship located above the well-head) was determined by GC/MS using an
329 Agilent 6890N (Agilent, Santa Clara, CA). Triplicate samples of 0.2 μL of raw oil were
330 directly injected to the column with no sample cleanup. This method was used to enable
331 detection of low molecular weight compounds that would be lost during sample
332 processing or masked due to interference from solvent peaks. The Agilent 6890N was
333 equipped with a 5972 mass selective detector and operated in SIM/SCAN mode. The
334 injection temperature was 250°C, detector temperature was 300°C, and column used was
335 60 m Agilent HP-1 MS with a flow rate of 2 ml/min. The oven temperature program
336 included a 50°C hold for 3 min ramped to 300°C at 4°C/min with a final 10 min hold at

300°C. Compound identification was determined from selective ion monitoring coupled with comparison to known standards and compound spectra in the NIST 08 MS library. Compounds were reported as fractions of total oil in Figure S3 from averages of triplicate injections, the error bars indicating standard deviation.

Hydrocarbon concentrations in plume samples (Table S2) were determined from plume water that was collected in the field and directly filtered through Sterivex filters (0.22 μm, Millipore, Billerica, MA) as described previously (Hazen et al, 2010). Oil biomarkers from the plume samples matched those observed from the Macondo well.

Volatile aromatic hydrocarbons were measured using USEPA methods 5030/8260b on an Agilent 6890 GC with a 5973 mass spectrometer detector. Initial oven temperature 10°C, initial time 3.00 min, ramp 8°C/min to 188C, then 16°C/min to 220°C, hold for 9.00 min. Split ratio 25:1. Restek Rtx-VMS capillary column, 60 meter length by 250 micron diameter, 1.40 micron film. Scan 50 to 550 m/z.

Results and Discussion

Throughout our analyses we found differences in the microbial community structures of the samples collected from the two plume sites due to the differences in the amount of time the respective indigenous deep-sea microbes were exposed to hydrocarbons. Our samples were collected during the Deepwater Horizon spill within 24 h following the failed top kill effort (May 29, 2010; proximal station). This effort resulted in a large influx of hydrocarbons into the deep-sea on the dates that we sampled. Due to the movement of water in marine currents we took the current velocity into account (6.7 km/day (2,3)) when calculating the length of time that microbes in our samples had been

360 exposed to hydrocarbons from the oil spill. Based on these calculations the microbial
361 communities would have been exposed to hydrocarbons for approximately 6 h by the
362 time the plume reached the proximal station, whereas by the time the plume reached the
363 distal station the microbes would have been exposed to hydrocarbons for approximately
364 39 h.

365 Analysis of our combined DNA sequence data (16S rRNA gene sequences from
366 454 pyrotag sequences and total metagenome DNA) revealed that the plume samples had
367 a lower microbial diversity than samples outside the plume (Figure S2 and Table 1), with
368 an enrichment of *Oceanospirillales* (Figure 2 and Tables S1 and S2), as previously
369 reported (Hazen et al, 2010, Redmond and Valentine, 2011). In the pyrotag data one
370 *Oceanospirillales* OTU comprised up to 80% to 90% of the proximal and distal plume
371 communities, respectively, whereas it comprised only 3% of the total community in the
372 uncontaminated sample (Figure 2 and Table S1). Similarly, in the metagenome data the
373 *Oceanospirillales* comprised > 60% of both plume samples, compared to 5% in the
374 uncontaminated sample in the metagenome data (Figure 2 and Table S2). This observed
375 bloom of *Oceanospirillales* corresponded with an increase in bacterial cell densities in
376 the plume, from 8.3×10^3 cells/ml in the uncontaminated sample to 1.5×10^4 cells/ml in
377 the proximal plume and 2.2×10^4 cells/ml in the distal plume.

378 Recently we used a GeoChip (He et al, 2010) functional gene microarray to
379 determine which functional genes were prevalent in the plume and found several
380 hydrocarbon degradation genes having a higher relative abundance in the plume (Lu et al,
381 2011). However, those data were not sufficient for determination of biodegradation
382 pathways or whether such pathways were actually expressed or attributed to a particular

383 microorganism in the plume. Here we examined deep metagenome sequence data for
384 genes and pathways involved in hydrocarbon degradation. We found that the entire
385 pathway for degradation of n-alkanes was represented and abundant in the metagenome
386 data from the plume samples (Figure 3). Alkane oxidation is initiated by
387 monooxygenases yielding alcohols as intermediates, which are converted to aldehydes
388 and fatty acids by alcohol and aldehyde dehydrogenases (Sabirova et al, 2006). In our
389 study we observed genes corresponding to alkane monooxygenases, a group of enzymes
390 with broad substrate specificity. In addition, the nearly complete pathway for
391 cyclohexane degradation (alkane monooxygenase → cyclohexanol dehydrogenase →
392 cyclohexanone monooxygenase →→ beta oxidation) (Sabirova et al, 2006) was observed
393 and abundant in the metagenomes (Figure 3). We also found a specific alkane gene
394 (alkane-1 monooxygenase), as also reported by Lu et al (Lu et al, 2011), that was more
395 abundant in the plume than outside of the plume. However, in contrast to Lu et al (Lu et
396 al, 2011), we found that genes involved in degradation of aromatic compounds were less
397 abundant than those involved in alkane degradation (Figure 3; see Figure S3 for Macondo
398 crude oil constituents and Table S3 for n-alkane, cyclohexane, methylcyclohexane,
399 BTEX, and PAH concentrations in the plume samples). For example, genes coding for
400 ethylbenzene, toluene and PAH degradation were significantly ($p < 0.05$) less abundant in
401 both plume samples compared to the uncontaminated sample. The abundance of genes
402 involved in alkane degradation compared to those involved in degradation of aromatic
403 compounds in our dataset is consistent with the ease of degradation of the respective
404 hydrocarbons (Das and Chandran, 2011) and suggested that the plume was enriched with
405 populations having the capacity for degradation of alkanes. Additional evidence for

406 biodegradation of alkanes in the plume samples was presented in our previous study
407 (Hazen et al, 2010) that reported oil half-lives in the plume of 1.2 to 6.1 days for C_{13} to
408 C_{26} n-alkanes. It should be noted that biodegradation of hydrocarbons in the plume was
409 carried out without significant oxygen depletion (oxygen saturation averaged 59% - 67%
410 inside and outside the plume, respectively) (Hazen et al, 2010).
411 To determine the active microbial community composition and expressed
412 functions in the plume interval we extracted total RNA from the proximal and distal
413 plume stations and sequenced the samples using the Illumina platform, resulting in a total
414 of 140 million paired-end reads (15 Gb). To assign microbial identities, the unassembled
415 metatranscriptome data (70 million single reads) was compared to a Greengenes
416 (DeSantis et al, 2006) database using blastn. We found that *Oceanospirillales* was not
417 only the most abundant member of the community; it was also active with a relative
418 abundance of transcripts of 46% in the proximal plume station sample and 69% in the
419 distal plume station sample (Figure 2 and Table S4). Other members of the community
420 that were active included Alteromonadales (11% relative abundance proximal plume/9%
421 relative abundance distal plume), Deltaproteobacteria (10%/1%), Pseudomonadales
422 (6%/4%), and SAR86 (3%/1%) (Figure 2 and Table S4). These community members
423 were also relatively abundant in our metagenome data (Figure 2 and Table S2). Therefore,
424 the dominant members of the community that were enriched by the deep-sea plume were
425 also active in the plume.
426 Previous analysis of samples from the deep-sea plume using DNA–based analyses
427 reported other microbial clades that were more or less abundant at different sampling
428 times. For example, members of the *Colwelliaceae* were detected as dominant

community members in the deep-sea plume in samples collected in mid-June 2010 (Valentine et al, 2010). In addition, microcosm experiments with labeled ethane and propane were dominated by *Colwellia*, with some *Oceanospirillales* increasing in abundance (Redmond and Valentine, 2011). Thus these authors suggested that *Colwellia* was primarily responsible for *in situ* ethane and propane oxidation, with perhaps, *Oceanospirillales* also playing a role (Redmond and Valentine, 2011). However cross-feeding could not be excluded (Redmond and Valentine, 2011). Although the *Colwelliaceae* were not abundant at < 1% relative abundance in our samples collected in late May, we found that they were represented in the active microbial community in both of our plume samples (Figure 2 and Table S4). However, other members of the community that were previously reported to be abundant (Valentine et al, 2010), such as *Cycloclasticus* which has members that are able to degrade simple and PAH aromatics (Dyksterhouse et al, 1995), while present in the pyrotag data at low abundances (Table S1) were not represented in our metagenome or metatranscriptome data (Tables S2 and S4). In addition, the methylotrophs (*Methylococcales* and *Methylophaga*) while rare, at less than 1% relative abundance in the plume samples, were active (Table S4). The metatranscriptome data thus revealed for the first time that *Oceanospirillales* was the dominant active member of the microbial community in the deep-sea plume in late May, in addition to some other members of the community, including some rare members.

We next determined what functions were expressed in the active microbial community enriched in the plume, with a focus on hydrocarbon degradation genes. A total of 70 million single, unassembled reads resulting from the metatranscriptome sequences were compared to a hydrocarbon degradation gene database. Differences in

452 relative abundances of active degradation genes (RNA transcripts) in the plume samples
453 were more pronounced compared to the DNA analyses (Figure 4). The metatranscriptome
454 data largely supported our metagenome data; for example finding that alkane
455 monooxygenases were highly expressed, with the same pathways for alkane, and
456 specifically for cyclohexane degradation present and abundant (Figure 5). This finding
457 suggests that alkane degradation was the dominant hydrocarbon degradation pathway
458 expressed in the plume at the time interval we sampled. Genes coding for degradation of
459 simple and PAH aromatics were either expressed at low levels, or not at all (Figure 5).
460 Reddy, et al (Reddy et al, 2011) determined the composition of oil and gas that was
461 emitted from the Macondo well and reported that BTEX compounds were the most
462 abundant hydrocarbons larger than C_1 to C_5 in the plume. However our findings indicate
463 that of the BTEX compounds, only those genes coding for ethylbenzene degradation
464 were expressed, and only in the proximal plume sample. This finding suggests that the
465 more recalcitrant compounds were not being actively degraded at the time that we
466 sampled. Although the samples analyzed by Reddy, et al (Reddy et al, 2011) were
467 collected at later time points than ours (mid to late June) their findings of negligible
468 biodegradation of BTEX compounds over 4 d in the deep-sea plume is consistent with
469 our findings.
470 Our study also revealed that particulate methane monooxygenase (*Pmo*) was
471 expressed in the plume at higher levels with distance from the wellhead and over time (i.e.
472 1.5-3 days to reach the distal station). Although *pmo* genes were expressed in the oil
473 plume, their relative levels were still less than those for genes coding for alkane
474 degradation (Figure 5). These results were surprising given that methane was the most

475 abundant hydrocarbon released during the spill (Kessler et al, 2011) with concentrations
476 ranging from 20-50 fold higher than background levels (Valentine et al, 2010) (and
477 references therein). Our data, as well as those of Valentine et al. (Valentine et al, 2010)
478 and Kessler et al. (Kessler et al, 2011) suggested a lag time in the response of
479 methanotrophs to the plume, relative to the initial bloom of *Oceanospirillales* capable of
480 oxidation of alkanes. However, our findings suggest that methane oxidation was actively
481 occurring earlier in the spill history than has previously been suggested (Kessler et al,
482 2011, Valentine et al, 2010).

483 Due to the dominance of members of the *Oceanospirillales* in the plume samples
484 and the recalcitrant nature of members of this order to cultivation, we specifically
485 targeted this group for single cell genome sequencing. We sorted water collected from the
486 proximal plume station by fluorescence-activated cell sorting. The single cells were lysed
487 and genomic DNA was amplified using multiple displacement amplification (MDA).
488 Subsequently, the single cells were screened on the basis of their 16S rRNA gene
489 sequences for those with high sequence quality and that were >95% similar to the
490 dominant *Oceanospirillales* OTU. After sequencing on the Illumina platform, two of
491 these cells yielded high quality sequences, which were concatenated and assembled
492 resulting in a single draft genome. The single cells were most closely related (16S rRNA
493 gene) to an uncultured *Oceanospirillales* (99% similar) from the oil spill (Redmond and
494 Valentine, 2011). Closest cultured representatives were *Oleispira Antarctica* (97%
495 similar) and *Thalassolituus oleivorans* (97% similar), both of which degrade aliphatic
496 hydrocarbons (C_{10}-C_{18} and C_7-C_{20}, respectively). However, genome sequences are not
497 available for either of these isolates. There are ten *Oceanospirillales* genome sequences

498 available in IMG (Markowitz et al, 2008) the most well characterized being *Alcanivorax*
499 *borkumensis* (Schneiker et al, 2006). As a rough estimate, the assembled single cell
500 *Oceanospirillales* draft genome (1.9 Mb genome with 876 contigs, N50 of 5,030 bp,
501 longest contig 25,481 bp) represented more than half a complete genome based on
502 comparisons to the 3.1 Mb genome of *Alcanivorax borkumensis*. *A. borkumensis* is
503 typically found at low abundance in unpolluted marine environments (Schneiker et al,
504 2006), but can represent as much as 90% of petroleum degrading microbial communities
505 (Harayama et al, 1999). The 16S rRNA gene sequences for our single cells were less than
506 88% similar to *A. borkumensis*, and thus represent a different genus within the
507 *Oceanospirillales*. Additionally, by comparison of the annotated COGs from the draft
508 genome assembly to those within the Gammaproteobacteria the draft genome was 53%
509 complete at the phylum level and 52% complete at the sub-phylum level. We also
510 examined all of the raw, unassembled reads for each single cell genome to ensure that all
511 of the sequence data was analyzed.
512 Within the draft genome we used CAMERA (Seshadri et al, 2007) to obtain gene
513 annotations in the assembled contigs. The annotations included putative genes encoding
514 methyl-accepting chemotaxis proteins, flagella, pili, and signal transduction mechanisms,
515 all of which were present in the metagenomes and expressed in the plume interval (Figure
516 6, S4, and S5). Physical evidence of microbial cell attraction to oil in the proximal plume
517 sample was also provided by synchrotron radiation-based Fourier-transform infrared (SR-
518 FTIR) spectromicroscopy that revealed sharp absorptions at ~1640 and ~1548 cm^{-1} in the
519 fingerprint region (between 1800 and 900 cm^{-1}) that are interpreted as Macondo oil
520 droplets surrounded by microorganisms (Figure S6). Together the physical and molecular

521 evidence suggest that bacterial cells were actively attracted to and interacted with oil in
522 the hydrocarbon plume.
523 Several key functions were recently identified as important for several low
524 abundance marine surface bacteria to rapidly respond and bloom when conditions
525 become more energy-rich (Yooseph et al, 2010). These included the capacity for
526 chemotaxis and motility, which we found in the draft genome, the metagenomes, and
527 metatranscriptome. Clustered regularly interspaced short palindromic repeat (CRISPR)
528 regions to protect from phage predation (Yooseph et al, 2010) were also identified in the
529 *Oceanospirillales* draft genome, suggesting a mechanism for avoiding phage predation.
530 Closer investigation of the draft genome revealed genes for uptake of a suite of
531 nutrients (Figure 6), all of which were also found in the metagenomes and expressed in
532 the plume metatranscriptome. For example, COGs involved in uptake of nitrogen
533 (ammonia permease), phosphate (ABC-type phosphate/phosphonate transport system,
534 permease component), iron (ABC-type Fe^{3+} siderophore transport system, permease
535 component, siderophore interacting protein, and Fe^{2+} transport system proteins), sulfur
536 (sulfate permease and related transporters), and Cobalt, Cadmium, and Zinc (transporters)
537 were detected in all three data sets (see Table S5).
538 We also analyzed the unassembled *Oceanospirillales* single cell reads for genes
539 involved in hydrocarbon degradation and searched for genes with closest similarities to
540 previously characterized genes based on bit scores ≥ 40. Consistent with what we
541 observed in the metagenomic and metatranscriptomic data, the *Oceanospirillales* draft
542 genome had genes with closest similarities to those coding for the cyclohexane
543 degradation pathway (Figure 6). This aliphatic degradation pathway is similar to what

544 was proposed for *A. borkumensis* (Schneiker et al, 2006). We did not find evidence in the
545 draft genome for ethane or propane oxidation, which Redmond and Valentine (2011)
546 suggested as a potential metabolic role for the *Oceanospirillales* observed in their SIP
547 experiments.
548
549 **Conclusion**
550 In this study we determined that the dominant and active, yet uncultured,
551 *Oceanospirillales* single cells possessed genes that encode the near complete pathway for
552 cyclohexane degradation. This pathway was present in the single cells, the metagenomes
553 and expressed in the plume metatranscriptome. The capacity of the *Oceanospirillales*
554 representatives for chemotaxis, motility, and for degradation of alkanes, may have
555 enabled these cells to actively aggregate and increase in numbers in the plume, and to
556 scavenge nutrients using a suite of transporters and siderophores. In addition, by using a
557 shotgun metatranscriptome approach, for the first time, we were able to determine which
558 hydrocarbon degradation pathways and other functions were actively expressed in the
559 deep-sea, to ascribe these pathways to particular groups of microorganisms, and to
560 elucidate how these active processes shifted in response to the hydrocarbon plume.
561 Given that the Gulf of Mexico experiences frequent, natural oil spills, elucidating the role
562 of *Oceanospirillales* in oil disposition provides critical data in understanding how
563 members of the deep-sea microbial community can rapidly respond and become enriched
564 in the presence of hydrocarbons.
565
566 **Acknowledgements**

This work was supported by a subcontract from the University of California at Berkeley, Energy Biosciences Institute (EBI) to Lawrence Berkeley National Laboratory under its U.S. Department of Energy contract DE-AC02-05CH11231. The SR-FTIR work was conducted at the infrared beamline at the Advanced Light Source, which is supported by the Director, Office of Science, Office of Basic Energy Sciences, of the U.S. Department of Energy. We are extremely grateful to Michael D. Curtin for his assistance with Paracel BLAST. We are thankful for the help of Theresa Pollard with handling shipping, ordering, and transportation of supplies and people to and from the field. We also thank the captain, crew, and science teams aboard the R/V Ocean Veritas and the R/V Brooks McCall. Lastly, we thank Damon Tighe and Janey Lee at JGI for single cell sorting and amplification and Chien-Chi and Matthew Scholz for metagenome assemblies.

References

Allgaier M, Reddy A, Park JI, Ivanova N, D'Haeseleer P, Lowry S *et al* (2010). Targeted Discovery of Glycoside Hydrolases from a Switchgrass-Adapted Compost Community. *PLoS One* **5:** e8812.

Blow MJ, Zhang T, Woyke T, Speller CF, Krivoshapkin A, Yang DY *et al* (2008). Identification of ancient remains through genomic sequencing. *Genome Res* **18:** 1347-1353.

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

Caporaso JG, Bittinger K, Bushman FD, DeSantis TZ, Andersen GL, Knight R (2010a). PyNAST: a flexible tool for aligning sequences to a template alignment. *Bioinformatics* **26:** 266-267.

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

Chao A (1984). Non-parametric estimation of the number of classes in a population. *Scandinavian Journal of Statistics* **11:** 265-270.

Chao A, Lee SM, Jeng SL (1992). Estimating population size for capture-recapture data when capture probabilities vary by time and individual animal. *Biometrics* **48:** 201-216.

Chao A, Hwang WH, Chen YC, Kuo CY (2000). Estimating the number of shared species in two communities. *Stat Sinica* **10:** 227-246.

Colwell RK, Coddington JA (1994). Estimating Terrestrial Biodiversity through Extrapolation. *Philos T Roy Soc B* **345:** 101-118.

Command OSATOUA (2010). Summary Report for Sub-Sea and Sub-Surface Oil and Dispersant Detection: Sampling and Monitoring.

Das N, Chandran P (2011). Microbial Degradation of Petroleum Hydrocarbon Contaminants: An Overview. *Biotechnology Research International* **2011**.

DeAngelis KM, Silver WL, Thompson AW, Firestone MK (2010). Microbial communities acclimate to recurring changes in soil redox potential status. *Environmental Microbiology* **12:** 3137-3149.

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.

Dyksterhouse S, Gray J, Herwig R, Lara J, Staley J (1995). Cycloclasticus pugetii gen. nov., sp. nov., an Aromatic Hydrocarbon-Degrading Bacterium from Marine Sediments. *International Journal of Systematic Bacteriology* **45:** 116-123.

Edgar RC (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26:** 2460-2461.

Grissa I, Vergnaud G, Pourcel C (2007). CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats. *Nucleic Acids Research* **35:** W52-W57.

Harayama S, Kishira H, Kasai Y, Shutsubo K (1999). Petroleum biodegradation in marine environments. *J Mol Microbiol Biotechnol* **1:** 63-70.

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.

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* **DOI:10.1126/science.1199697**.

Kunin V, Engelbrektson A, Ochman H, Hugenholtz P (2010). Wrinkles in the rare biosphere: pyrosequencing errors can lead to artificial inflation of diversity estimates. *Environmental Microbiology* **12:** 118-123.

Letunic I, Yamada T, Kanehisa M, Bork P (2008). iPath: interactive exploration of biochemical pathways and networks. *Trends in biochemical sciences* **33:** 101-103.

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*.

Markowitz VM, Ivanova NN, Szeto E, Palaniappan K, Chu K, Dalevi D *et al* (2008). IMG/M: a data management and analysis system for metagenomes. *Nucleic Acids Research* **36:** D534-D538.

669 Miller DN, Bryant JE, Madsen EL, Ghiorse WC (1999). Evaluation and Optimization of
670 DNA Extraction and Purification Procedures for Soil and Sediment Samples. *Appl*
671 *Environ Microbiol* **65:** 4715-4724.
672
673 Miller JR, Koren S, Sutton G (2010). Assembly algorithms for next-generation
674 sequencing data. *Genomics* **95:** 315-327.
675
676 Parks DH, Beiko RG (2010). Identifying biologically relevant differences between
677 metagenomic communities. *Bioinformatics* **26:** 715-721.
678
679 Reddy CM, Arey JS, Seewald JS, Sylva SP, Lemkau KL, Nelson RK *et al* (2011).
680 Composition and fate of gas and oil released to the water column during the
681 Deepwater Horizon oil spill. *Proceedings of the National Academy of Sciences*
682 **doi:10.1073/pnas.1101242108**.
683
684 Redmond MC, Valentine DL (2011). Natural gas and temperature structured a
685 microbial community response to the Deepwater Horizon oil spill. *Proceedings of the*
686 *National Academy of Sciences* **doi:10.1073/pnas.1108756108**.
687
688 Rodrigue S, Malmstrom RR, Berlin AM, Birren BW, Henn MR, Chisholm SW (2009).
689 Whole genome amplification and de novo assembly of single bacterial cells. *PLoS*
690 *One* **4:** e6864.
691
692 Sabirova JS, Ferrer M, Regenhardt D, Timmis KN, Golyshin PN (2006). Proteomic
693 insights into metabolic adaptations in Alcanivorax borkumensis induced by alkane
694 utilization. *J Bacteriol* **188:** 3763-3773.
695
696 Schneiker S, dos Santos VAPM, Bartels D, Bekel T, Brecht M, Buhrmester J *et al*
697 (2006). Genome sequence of the ubiquitous hydrocarbon-degrading marine
698 bacterium Alcanivorax borkumensis. *Nat Biotech* **24:** 997-1004.
699
700 Seshadri R, Kravitz SA, Smarr L, Gilna P, Frazier M (2007). CAMERA: A Community
701 Resource for Metagenomics. *PLoS Biol* **5:** e75.
702
703 Simpson EH (1949). Measurement of Diversity. *Nature* **163:** 688-688.
704
705 Valentine DL, Kessler JD, Redmond MC, Mendes SD, Heintz MB, Farwell C *et al*
706 (2010). Propane respiration jump-starts microbial response to a deep oil spill.
707 *Science* **330:** 208-211.
708
709 Yooseph S, Nealson KH, Rusch DB, McCrow JP, Dupont CL, Kim M *et al* (2010).
710 Genomic and functional adaptation in surface ocean planktonic prokaryotes. *Nature*
711 **468:** 60-66.
712
713 Zerbino D, Birney E (2008). Velvet: Algorithms for De Novo Short Read Assembly
714 Using De Bruijn Graphs. *Genome Research* **18:** 821-829.

715
716
717

718 **Figure legends**

719 Figure 1. Methods schematic. Each type of molecular method -metagenomics,
720 metatranscriptomics, and single cell genomics are shown, as are the subsequent, novel
721 bioinformatics approaches that were used to analyze the various datasets.

722

723 Figure 2. Relative abundance of Bacteria and Archaea in the proximal and distal plume
724 samples and in the uncontaminated sample collected from plume depth. A) Relative OTU
725 abundance of rarified 16S rRNA gene 454-pyrotag data. Universal primers for Archaea
726 and Bacteria were used. Taxonomy was assigned using the Greengenes (DeSantis et al,
727 2006) 16S rRNA gene database. B) Raw, unassembled metagenomic and
728 metatranscriptomic reads were compared to the Greengenes (DeSantis et al, 2006)
729 database. The complete list of Bacteria and Archaea observed in these analyses are
730 presented in Table S1, S2, and S4.

731

732 Figure 3. Analysis of genes involved in hydrocarbon degradation in the metagenome data.
733 Blue bars denote the distal station metagenome; black bars denote the uncontaminated
734 sample metagenome; red bars denote the proximal station metagenome. Raw,
735 unassembled metagenomic reads were compared to proteins involved in hydrocarbon
736 degradation, using a custom database using the tblastn algorithm. A bit score cutoff of ≥
737 40 was used. Genes were grouped according to function. [A] indicates that a corrected p-
738 value was not significant. Gene categories denoted with an ‡ indicates a similar substrate
739 degradation pathway. A complete list of all gene categories is provided in Table S6.

740

741 Figure 4. Analysis of genes involved in hydrocarbon degradation in the metagenome and
742 metatranscriptome data. Raw, unassembled metagenomic and metatranscriptomic reads
743 were compared to proteins involved in hydrocarbon degradation, using a custom database
744 using the tblastn algorithm. A bit score cutoff of ≥ 40 was used. Genes were grouped
745 according to function. Gene categories denoted with an ‡ indicates a similar substrate
746 degradation pathway. A complete list of all gene categories is provided in Table S5.
747
748 Figure 5. Analysis of genes involved in hydrocarbon degradation in the
749 metatranscriptome data. Blue bars denote the distal station metagenome and red bars
750 denote the proximal station metagenome. Raw, unassembled metatranscriptome reads
751 were compared to proteins involved in hydrocarbon degradation, using a custom database
752 using the tblastn algorithm. A bit score cutoff of ≥ 40 was used. Genes were grouped
753 according to function. An asterisk indicates that the difference in relative abundance of a
754 particular gene group in the proximal station metatranscriptome compared to the distal
755 station metatranscriptome was statistically significant. Gene categories denoted with an ‡
756 indicates a similar substrate degradation pathway. A complete list of all gene categories is
757 provided in Table S6.
758
759 Figure 6. *Oceanospirillales* single cell metabolic reconstruction using COG annotations
760 of assembled sequence data and the blast comparison of unassembled single cell reads to
761 genes involved in hydrocarbon degradation. All genes in the single cell metabolic
762 reconstruction were present in the metagenomes and most were expressed in the
763 metatranscriptome, except for those with an asterisk following the gene name.

Tables

Table 1. Diversity metrics of rarified 16S rRNA 454-pyrotag sequences.

Sample	Chao1[a]	Chao1 (lower bound)[b]	Chao1 (upper bound)[b]	ACE[c]	Simpson[d]	Singletons[e]	Doubletons[e]
Distal plume	394.53	273.32	628.27	443.80	0.58	91.78	15.96
Proximal plume	806.71	626.34	1093.50	911.07	0.57	198.93	38.47
Uncontaminated	1722.58	1507.22	2007.66	1849.08	0.96	481.80	126.04

[a]Species richness (Chao, 1984).
[b]Confidence intervals (Chao et al, 1992).
[c]Species richness (Chao et al, 2000).
[d]Species diversity (Simpson, 1949).
[e]Singeltons are species with only one individual. Doubletons are species with only two individuals (Colwell and Coddington, 1994).

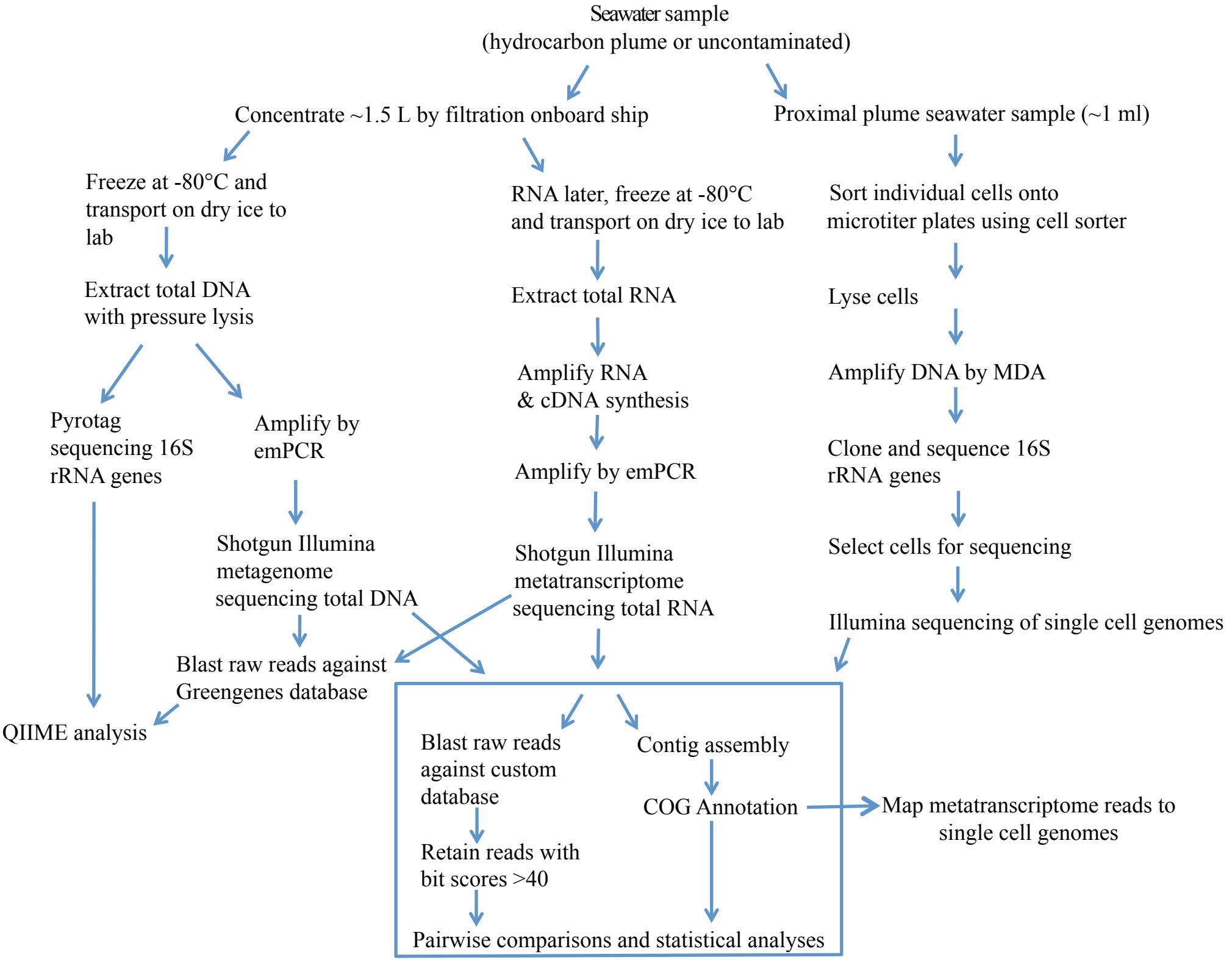

Seawater sample
(hydrocarbon plume or uncontaminated)
Concentrate ~1.5 L by filtration onboard ship
Proximal plume seawater sample (~1 ml)
Freeze at -80°C and transport on dry ice to lab
RNA later, freeze at -80°C and transport on dry ice to lab
Sort individual cells onto microtiter plates using cell sorter
Extract total DNA with pressure lysis
Extract total RNA
Lyse cells
Pyrotag sequencing 16S rRNA genes
Amplify by emPCR
Amplify RNA & cDNA synthesis
Amplify DNA by MDA
Amplify by emPCR
Clone and sequence 16S rRNA genes
Shotgun Illumina metagenome sequencing total DNA
Shotgun Illumina metatranscriptome sequencing total RNA
Select cells for sequencing
Illumina sequencing of single cell genomes
Blast raw reads against Greengenes database
QIIME analysis
Blast raw reads against custom database
Contig assembly
COG Annotation
Map metatranscriptome reads to single cell genomes
Retain reads with bit scores >40
Pairwise comparisons and statistical analyses

Figure 1.

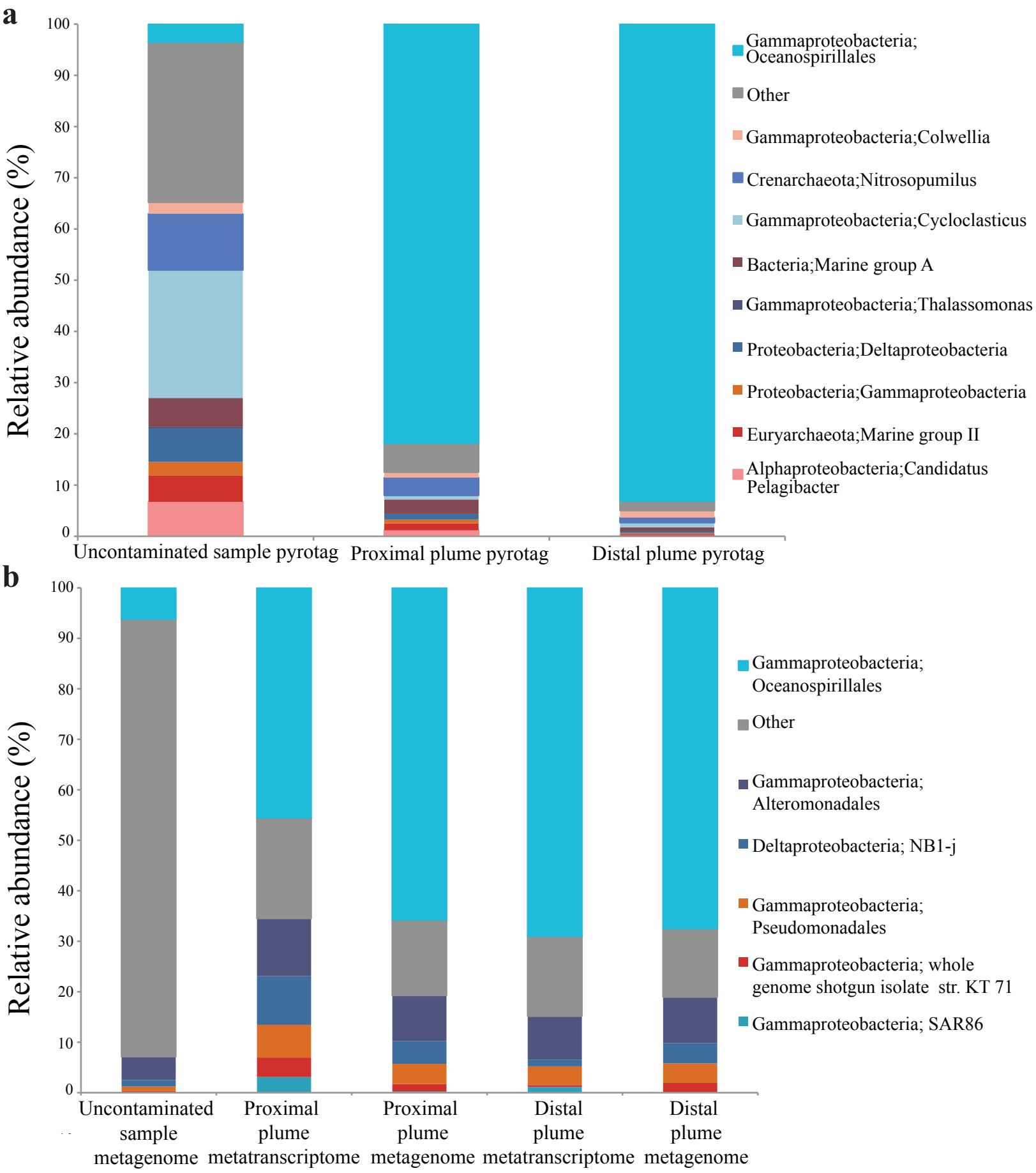

a
Relative abundance (%)
100
90
80
70
60
50
40
30
20
10
0
Uncontaminated sample pyrotag
Proximal plume pyrotag
Distal plume pyrotag
Gammaproteobacteria; Oceanospirillales
Other
Gammaproteobacteria;Colwellia
Crenarchaeota;Nitrosopumilus
Gammaproteobacteria;Cycloclasticus
Bacteria;Marine group A
Gammaproteobacteria;Thalassomonas
Proteobacteria;Deltaproteobacteria
Proteobacteria;Gammaproteobacteria
Euryarchaeota;Marine group II
Alphaproteobacteria;Candidatus Pelagibacter
b
Relative abundance (%)
100
90
80
70
60
50
40
30
20
10
0
Uncontaminated sample metagenome
Proximal plume metatranscriptome
Proximal plume metagenome
Distal plume metatranscriptome
Distal plume metagenome
Gammaproteobacteria; Oceanospirillales
Other
Gammaproteobacteria; Alteromonadales
Deltaproteobacteria; NB1-j
Gammaproteobacteria; Pseudomonadales
Gammaproteobacteria; whole genome shotgun isolate str. KT 71
Gammaproteobacteria; SAR86

Figure 2.

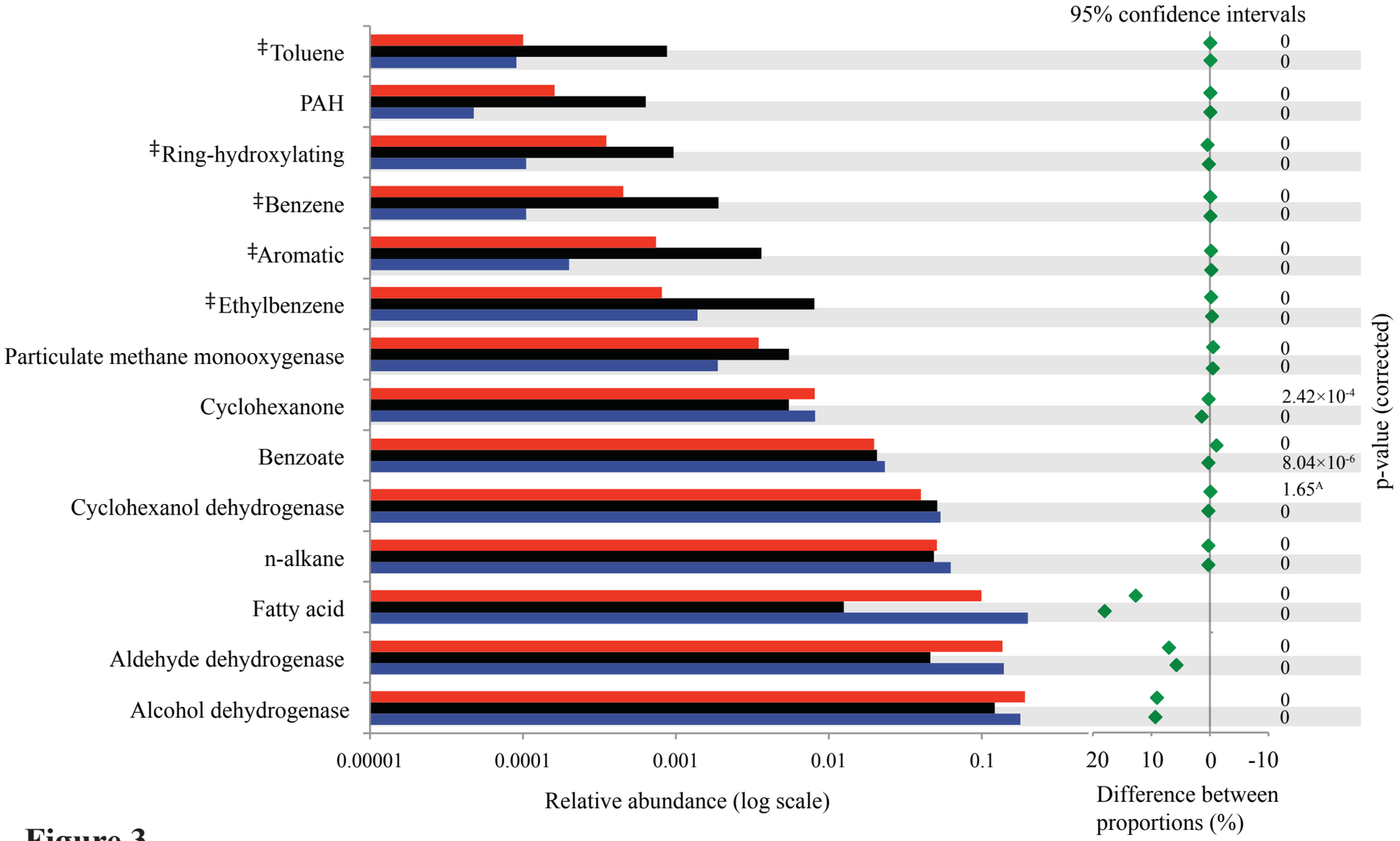

95% confidence intervals
‡Toluene
PAH
‡Ring-hydroxylating
‡Benzene
‡Aromatic
‡Ethylbenzene
Particulate methane monooxygenase
Cyclohexanone
Benzoate
Cyclohexanol dehydrogenase
n-alkane
Fatty acid
Aldehyde dehydrogenase
Alcohol dehydrogenase
0
0
0
0
0
0
0
0
0
0
0
0
0
0
0
0
0
0
0
0
0
0
0
0
2.42×10-4
0
0
8.04×10-6
1.65A
0
0
0
0
0
0
0
0
0
p-value (corrected)
0.00001
0.0001
0.001
0.01
0.1
20
10
0
-10
Relative abundance (log scale)
Difference between proportions (%)

Figure 3.

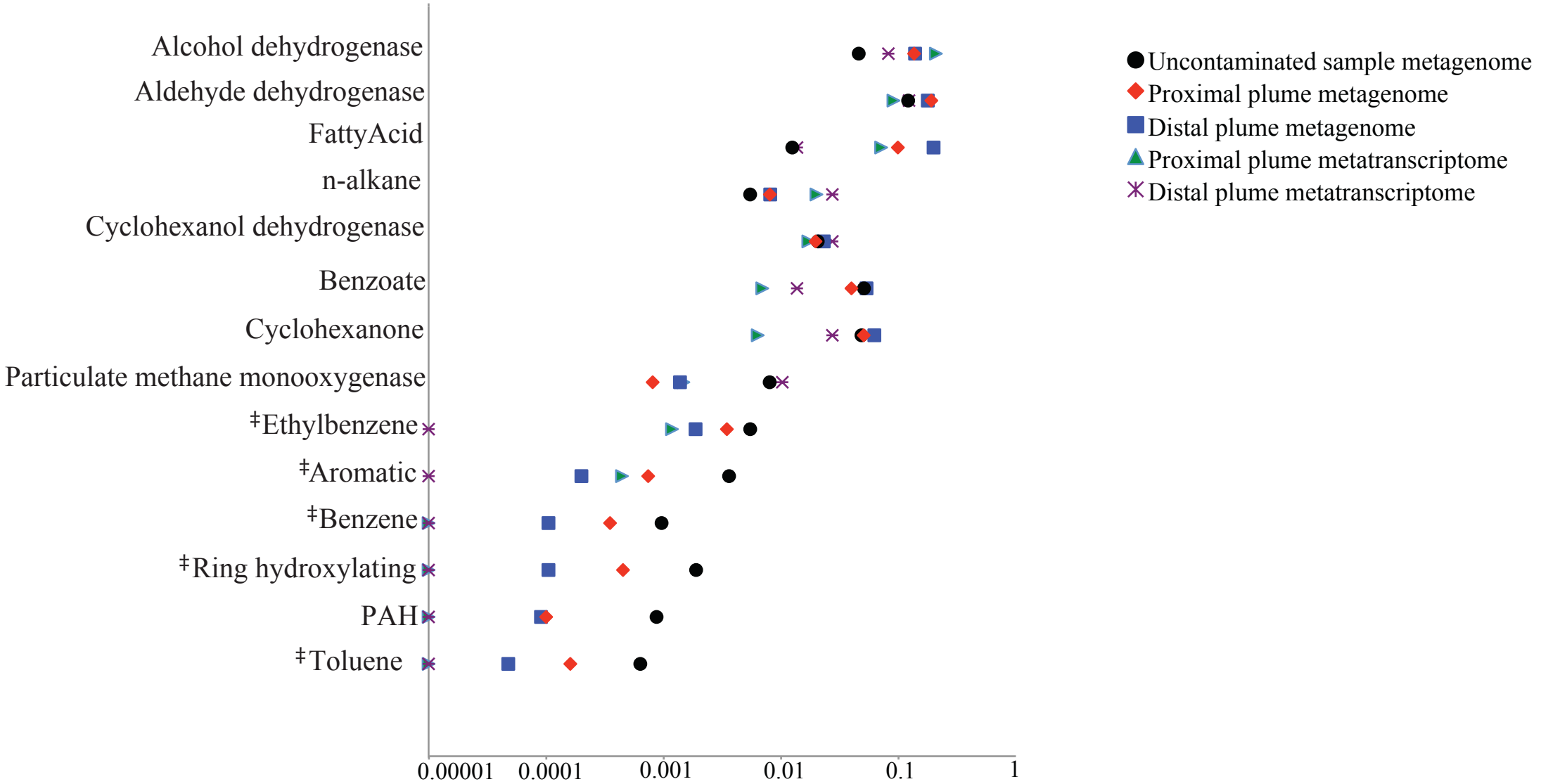

Alcohol dehydrogenase
Aldehyde dehydrogenase
FattyAcid
n-alkane
Cyclohexanol dehydrogenase
Benzoate
Cyclohexanone
Particulate methane monooxygenase
‡Ethylbenzene
‡Aromatic
‡Benzene
‡Ring hydroxylating
PAH
‡Toluene
Uncontaminated sample metagenome
Proximal plume metagenome
Distal plume metagenome
Proximal plume metatranscriptome
Distal plume metatranscriptome
0.00001
0.0001
0.001
0.01
0.1
1
Gene relative abundance (log scale)

Figure 4.

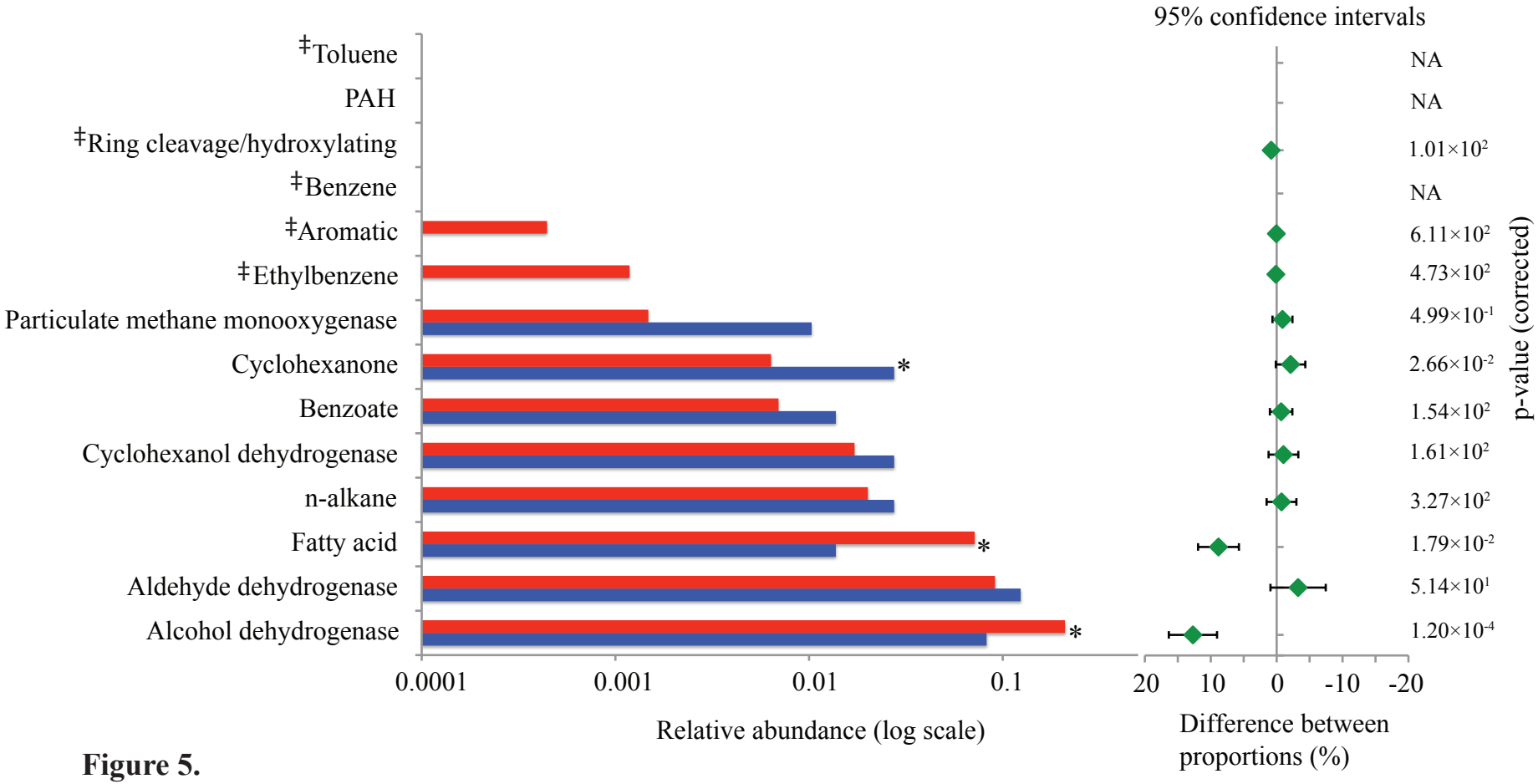

95% confidence intervals
‡Toluene
PAH
‡Ring cleavage/hydroxylating
‡Benzene
‡Aromatic
‡Ethylbenzene
Particulate methane monooxygenase
Cyclohexanone
Benzoate
Cyclohexanol dehydrogenase
n-alkane
Fatty acid
Aldehyde dehydrogenase
Alcohol dehydrogenase
*
*
*
NA
NA
1.01×10²
NA
6.11×10²
4.73×10²
4.99×10⁻¹
2.66×10⁻²
1.54×10²
1.61×10²
3.27×10²
1.79×10⁻²
5.14×10¹
1.20×10⁻⁴
p-value (corrected)
0.0001
0.001
0.01
0.1
Relative abundance (log scale)
20
10
0
-10
-20
Difference between proportions (%)

Figure 5.

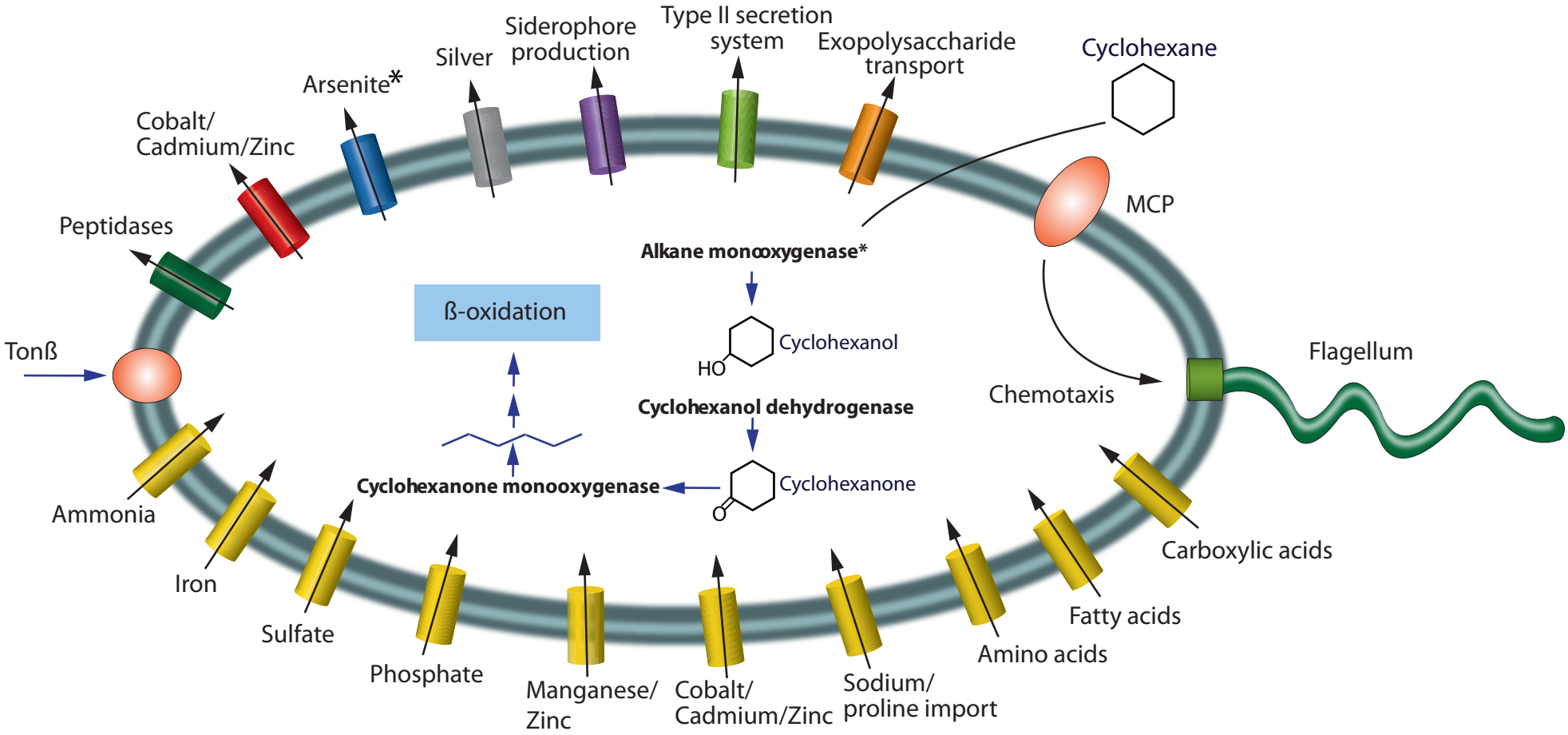

Type II secretion system
Siderophore production
Silver
Exopolysaccharide transport
Cyclohexane
Arsenite*
Cobalt/ Cadmium/Zinc
MCP
Peptidases
Alkane monooxygenase*
ß-oxidation
Cyclohexanol
HO
Tonß
Flagellum
Chemotaxis
Cyclohexanol dehydrogenase
Cyclohexanone monooxygenase
Cyclohexanone
Ammonia
Carboxylic acids
Iron
Fatty acids
Sulfate
Amino acids
Phosphate
Sodium/ proline import
Manganese/ Zinc
Cobalt/ Cadmium/Zinc

Figure 6.

Supplemental Material

Supplementary Tables

Table S1.16S rRNA gene pyrotag sequence data of percent relative abundance for all observed Bacteria and Archaea.

Taxonomy	Distal plume	Proximal plume	Uncontaminated
Proteobacteria;Oceanospirillales	92.97	80.62	3.98
Proteobacteria;Colwellia	1.39	1.07	1.99
Crenarchaeota;Nitrosopumilus	0.96	3.39	9.62
Proteobacteria;Cycloclasticus	0.92	0.80	21.96
Bacteria;Marine group A	0.65	2.68	4.92
Proteobacteria;Thalassomonas	0.30	0.22	0.41
Proteobacteria;Deltaproteobacteria	0.22	0.96	5.66
Proteobacteria;Gammaproteobacteria	0.22	0.72	2.34
Euryarchaeota;Marine group II	0.21	1.32	4.49
Proteobacteria;Candidatus Pelagibacter	0.16	1.16	5.90
Proteobacteria;Kangiella	0.15	0.30	1.76
Crenarchaeota;Cenarchaeaceae	0.14	0.91	3.24
Proteobacteria;Microbulbifer	0.14	0.12	0.00
Euryarchaeota;Marine group III	0.13	0.71	2.89
Chloroflexi;Chloroflexi-4	0.12	0.38	1.39
Actinobacteria;ZA3409c	0.10	0.26	1.04
Gemmatimonadetes;Gemmatimonadetes (class)	0.08	0.29	1.50
Planctomycetes;Pirellulales	0.06	0.15	1.29
Acidobacteria;Acidobacteriales	0.06	0.19	0.97
Cyanobacteria;Synechococcus	0.06	0.05	0.63
Bacteroidetes;Flavobacteriaceae	0.05	0.11	0.78
Proteobacteria;Candidatus Portiera	0.05	0.10	0.42
Acidobacteria;Acidobacteria	0.05	0.06	0.28
Proteobacteria;Chromatiales	0.05	0.31	2.36
Proteobacteria;Rhodospirillaceae	0.05	0.27	1.25
Bacteroidetes;Flavobacteria	0.04	0.24	0.92
Planctomycetes;Phycisphaerales	0.04	0.11	0.51
Proteobacteria;Sinobacteraceae	0.04	0.08	0.62
Proteobacteria;Nitrospina	0.03	0.32	1.41
Proteobacteria;Methylococcales	0.03	0.02	1.00
Acidobacteria;BPC102	0.03	0.02	0.02
Planctomycetes;Planctomyces	0.03	0.04	0.37
Acidobacteria;BPC102	0.03	0.04	0.39
Bacteroidetes;Sphingobacteriales	0.02	0.07	0.25
Cyanobacteria;Prochlorococcus	0.02	0.07	0.18
Proteobacteria;Coxiellaceae	0.02	0.02	0.23
Proteobacteria;Comamonadaceae	0.02	0.00	0.09
Verrucomicrobia;Puniceicoccaceae	0.02	0.05	0.28
Planctomycetes;OM190	0.01	0.02	0.35
Bacteroidetes;Sediminicola	0.01	0.00	0.00
Planctomycetes;CL500-15	0.01	0.02	0.14
Proteobacteria;Ectothiorhodospiraceae	0.01	0.01	0.00

Table S2.16S rRNA gene metagenome sequence data of percent relative abundance for all observed Bacteria and Archaea.

Taxonomy	Distal plume	Proximal plume	Uncontaminated
Proteobacteria;Oceanospirillales	62.54	60.31	4.67
Proteobacteria;Alteromonadales	9.25	9.09	3.75
Proteobacteria;marine boreal sponge Isops phlegraei Norway:Sula Ridge	5.84	5.47	0.29
Proteobacteria;NB1-j	4.10	4.50	1.04
Proteobacteria;Pseudomonadales	3.92	4.00	0.74
Proteobacteria;Reinekea	2.12	1.66	0.08
Proteobacteria;1097664134132 whole genome shotgun isolate str. KT 71	1.81	1.47	0.10
Proteobacteria;Env Seq	1.51	1.35	1.66
Proteobacteria;Sphingomonadales	1.30	1.26	0.76
Proteobacteria;seawater isolate str. sw-11	1.23	1.15	0.07
Proteobacteria;Aeromonadales	1.21	1.11	0.19
Proteobacteria;Env Seq	1.02	0.95	0.00
Proteobacteria;Boston Harbor surface water isolate str. UMB6E	0.62	0.63	0.08
Proteobacteria;Thiotrichales	0.61	0.55	20.25
Proteobacteria;Env Seq	0.24	0.24	0.00
Proteobacteria;Enterobacteriales	0.20	0.27	0.18
Proteobacteria;Consistiales	0.17	0.85	11.93
Proteobacteria;Methylophaga	0.16	0.14	6.89
Proteobacteria;SAR86	0.15	0.26	0.97
Proteobacteria;Boston Harbor surface water isolate str. UMB3E	0.14	0.21	0.02
Proteobacteria;Vibrionales	0.13	0.19	0.32
Thaumarchaeota;Cenarchaeum	0.12	0.51	4.20
Proteobacteria;HOC28	0.09	0.14	0.00
Proteobacteria;Rhizobiales	0.09	0.09	0.28
Proteobacteria;Pseudomonadaceae	0.08	0.08	0.45
Actinobacteria;Actinomycetales	0.08	0.13	0.50
Proteobacteria;agg47	0.08	0.13	1.64
Proteobacteria;Boston Harbor surface water isolate str. UMB6D	0.06	0.06	0.14
Proteobacteria;Sva0853	0.05	0.29	3.59
Proteobacteria;OM60	0.04	0.05	0.24
Proteobacteria;ZD0417	0.04	0.15	1.29
Proteobacteria;Thiomicrospira	0.04	0.03	0.01
Proteobacteria;Arctic96B-1	0.03	0.03	0.17
Proteobacteria;isolate str. EHK-1	0.03	0.01	1.14
Proteobacteria;ZA2333c	0.03	0.02	0.04
Proteobacteria;SUP05	0.03	0.12	1.76
Proteobacteria;Rhodobacterales	0.03	0.07	0.32
Proteobacteria;BD7-8	0.03	0.01	0.00
Proteobacteria;sulfur-oxidizing symbionts	0.02	0.09	0.26
Proteobacteria;Pasteurellales	0.02	0.01	0.02
Marine_group_A;Env Seq	0.02	0.09	0.73
Proteobacteria;Methylococcales	0.02	0.03	0.99
Proteobacteria;Legionellales	0.02	0.05	0.24
Thermoplasmata_Eury;SB95-72	0.02	0.15	1.45
Firmicutes;Clostridiales	0.02	0.02	0.06
Proteobacteria;Vibrionaceae	0.02	0.01	0.00
Proteobacteria;Chromatiales	0.02	0.01	0.37
Marine_group_A;Env Seq	0.02	0.13	0.85
Chloroflexi;SAR307	0.02	0.06	0.95
Marine_group_A;Env Seq	0.01	0.04	0.35
Proteobacteria;marine str. HTCC2080	0.01	0.02	0.03
Marine_group_A;Env Seq	0.01	0.11	1.62
Actinobacteria;Microthrixineae	0.01	0.10	1.38
Verrucomicrobia;Verruco-3	0.01	0.09	1.12
Proteobacteria;Xanthomonadales	0.01	0.01	0.08
Proteobacteria;Halomonadaceae	0.01	0.01	0.00
Proteobacteria;Env Seq	0.01	0.01	0.44
Proteobacteria;Betaproteobacteria	0.01	0.01	0.31
Bacteroidetes;Flavobacteriales	0.01	0.01	0.35
Proteobacteria;str. NEP2	0.01	0.01	0.04

Table S3. BTEX, cycloalkanes, and PAH concentrations in the distal and proximal plume samples, including, when determined, the range of concentrations over the plume intervals.

	Depth (mbsl)	Total alkanes (ug/L)	Cyclohexane (ug/L)	Methylcyclohexane (ug/L)	Benzene (ug/L)	Ethylbenzene (ug/L)	Toluene (ug/L)	Xylenes, Total (ug/L)	PAH (ug/L)
Proximal plume	1207	292.6	49.35	65.75	32.15	10.19	63.00	71.25	9.50
(plume interval range)	1181-1207	n.d.	49.35 - 99.90	65.75 - 121.00	32.15 - 70.50	10.19 - 21.70	63.00 - 135.00	71.25 - 150.452	n.d.
Proximal plume (metatranscriptome sample)	1194	238	148.00	154.00	73.60	28.10	158.00	199.00	10.23
Distal plume	1179	323.8	79.20	86.40	48.80	13.90	106.00	107.00	8.10
(plume interval range)	1136-1179	n.d.	19.10 - 79.20	20.90 - 86.40	15.10 - 48.80	4.67 - 13.90	33.80 - 106.00	30.30 - 107.00	n.d.

n.d. indicates not determined.

Table S4. 16S rRNA sequence data from the plume metatranscriptomes of percent relative abundance for all observed Bacteria and Archaea.

Taxonomy	Proximal station metatranscriptome	Distal station metatranscriptome
Bacteria; Proteobacteria; Gammaproteobacteria; Oceanospirillales	45.52	68.99
Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales	11.45	8.59
Bacteria; Proteobacteria; Deltaproteobacteria; NB1-j	9.65	1.35
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales	6.48	3.72
Bacteria; Proteobacteria; Gammaproteobacteria; 1097664134132 whole genome shotgun isolate 71KT str. KT 71	3.87	0.38
Bacteria; Proteobacteria; Gammaproteobacteria; SAR86	3.11	1.13
Bacteria; Oxidative ultraviolet C treatment culturable harvested and identified wells groundwater isolate C1groundwater biofilm str. C1	2.65	1.48
Bacteria; symbiont	2.45	2.01
Bacteria; Oceaniserpentilla; Oceaniserpentilla haliotidis	2.41	2.45
Bacteria; Proteobacteria; isolate str. M3-6	1.78	1.05
Bacteria; Proteobacteria; Gammaproteobacteria; deep-sea sediment clone BD1-7	1.46	0.97
Bacteria; Boston Harbor surface water isolate str. UMB8C	1.01	0.62
Bacteria; Proteobacteria; Gammaproteobacteria; Thiotrichales	0.93	0.48
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales	0.70	0.55
Bacteria; Proteobacteria; Gammaproteobacteria; Boston Harbor surface water isolate str. UMB3E UMB3E	0.50	0.37
Bacteria; Proteobacteria; Gammaproteobacteria; Aeromonadales	0.39	0.12
Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales	0.37	0.21
Bacteria; Proteobacteria; Gammaproteobacteria; Methylophaga	0.31	0.11
Bacteria; Sponge isolate str. Ex6	0.30	0.22
Bacteria; Proteobacteria; Gammaproteobacteria; libraries sediment Kings Bay Svalbard Arctic clone SS1_B_04_41	0.30	0.33
Bacteria; Proteobacteria; Gammaproteobacteria; OM60	0.28	0.43
Bacteria; Proteobacteria; Alphaproteobacteria; Sphingomonadales	0.27	0.14
Bacteria; Proteobacteria; Gammaproteobacteria; deep-sea octacoral clone ctg_NISA145	0.22	0.19
Bacteria; Proteobacteria; Gammaproteobacteria; HOC28	0.20	0.06
Bacteria; Proteobacteria; Alphaproteobacteria; Rhodobacterales	0.19	0.13
Bacteria; Proteobacteria; Gammaproteobacteria; Thiomicrospira	0.17	0.08
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales	0.16	0.11
Bacteria; Proteobacteria; Gammaproteobacteria; sea water isolate sw-11sw-11 str. sw-11	0.16	0.25
Bacteria; Proteobacteria; Gammaproteobacteria; libraries sediment Kings Bay Svalbard Arctic clone SS1_B_06_26	0.14	0.06
Bacteria; Endosymbiont Pogonophora sp. JT-1 clone JTPE-1 Pogonophora	0.14	0.09
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadaceae	0.11	0.14
Bacteria; Proteobacteria; Gammaproteobacteria; Reinekea	0.10	0.43
Bacteria; Proteobacteria; Gammaproteobacteria; agg47	0.10	0.06
Bacteria; Proteobacteria; Gammaproteobacteria; Halomonadaceae	0.10	0.02
Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales	0.09	0.22
Bacteria; Proteobacteria; Gammaproteobacteria; sulfur-oxidizing symbionts	0.08	0.12
Bacteria; Proteobacteria; Alphaproteobacteria; Consistiales	0.07	0.03
Bacteria; Proteobacteria; Gammaproteobacteria; BD1-1	0.06	0.02
Bacteria; Proteobacteria; Gammaproteobacteria; marine str. HTCC2080	0.06	0.00
Bacteria; Proteobacteria; Alphaproteobacteria; Bradyrhizobiales	0.05	0.11
Bacteria; Proteobacteria; Gammaproteobacteria; cold seep sediment clone JT58-12	0.04	0.04
Bacteria; Proteobacteria; Gammaproteobacteria; landfill leachate clone GZKB44	0.04	0.01
Bacteria; Proteobacteria; Gammaproteobacteria; BPC036	0.04	0.02
Bacteria; tetracycline and tylosin resistance genes impacted swine effluent holding pit	0.04	0.03
Bacteria; Proteobacteria; Gammaproteobacteria; sulfur_oxidizing_symbionts	0.04	0.03
Bacteria; Proteobacteria; Genetic and functional microbial symbionts new species polychaete worms symbiont Osedax MB4 clone T933_2_H5	0.04	0.03

Table S4. 16S rRNA sequence data from the plume metatranscriptomes of percent relative abundance for all observed Bacteria and Archaea (cont).

Bacteria; Proteobacteria; Alphaproteobacteria; Rhodospirillales	0.04	0.03
Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionaceae	0.04	0.06
Bacteria; Proteobacteria; Gammaproteobacteria; Legionellales	0.03	0.06
Bacteria; Proteobacteria; Alphaproteobacteria; Pacific arctic surface sediment clone S26-7	0.03	0.03
Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales	0.03	0.02
Bacteria; Firmicutes; Clostridia; Clostridiales	0.03	0.18
Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales	0.03	0.01
Bacteria; Actinobacteria; Acidimicrobidae; identification marine actinobacteria sediment isolate YM22-133Acidimicrobidae str. YM22-133	0.02	0.01
Bacteria; Proteobacteria; Gammaproteobacteria; BD7-8	0.02	0.14
Bacteria; Proteobacteria; Gammaproteobacteria; Arctic96B-1	0.02	0.01
Bacteria; Proteobacteria; Gammaproteobacteria; determined library mangrove clone DS095	0.02	0.01
Bacteria; Firmicutes; Bacillales; Bacillaceae	0.02	0.02
Bacteria; Proteobacteria; Gammaproteobacteria; Chromatiales	0.02	0.02
Bacteria; Influence starvation ozone and selected isolates on survival southern Jasus edwardsii healthy larval rock lobster	0.02	0.03
Bacteria; Actinobacteria; Acidimicrobidae; EB1017_group	0.02	0.01
Bacteria; Proteobacteria; Gammaproteobacteria; ZA2333c	0.02	0.02
Bacteria; Proteobacteria; Gammaproteobacteria	0.02	0.02
Bacteria; Unexpected symbionts sp. cold seeps eastern Mediterranean: new evolution symbiosis mytilids gill tissue Idas sp clone M2.41	0.02	0.02
Bacteria; Proteobacteria; Gammaproteobacteria; cold seep sediment clone JT75-103	0.02	0.02
Bacteria; Proteobacteria; Gammaproteobacteria; polyphasic description Pocillopora meandrina Palmyra Atoll Calcinus obscurus abdominal flora clone Cobs2TisB5	0.02	0.02
Bacteria; Proteobacteria; Gammaproteobacteria; Oceanimonaceae	0.02	0.00
Bacteria; Proteobacteria; Gammaproteobacteria; SUP05	0.02	0.02
Bacteria; Firmicutes; Mollicutes; Mycoplasmatales	0.01	0.01
Bacteria; Proteobacteria; Gammaproteobacteria; bacterioplankton clone ZA3235c	0.01	0.01
Bacteria; Proteobacteria; Alphaproteobacteria; Anderseniella	0.01	0.01
Bacteria; Proteobacteria; Gammaproteobacteria; Methylococcales	0.01	0.01
Bacteria; Proteobacteria; Gammaproteobacteria; Endozoicimonas	0.01	0.11
Bacteria; Proteobacteria; Gammaproteobacteria; deep-sea octacoral clone ctg_CGOF019	0.01	0.00
Bacteria; Proteobacteria; Gammaproteobacteria; marine str. HTCC2178	0.01	0.03
Bacteria; Proteobacteria; Gammaproteobacteria; Alkalimonas	0.01	0.00
Bacteria; Proteobacteria; Genetic and functional microbial symbionts new species polychaete worms symbiont Osedax MB3 clone T931_1_B1	0.01	0.00
Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales	0.01	0.01
Bacteria; Bacteroidetes; Flavobacteria; Flavobacteriales	0.01	0.00
Bacteria; Proteobacteria; Gammaproteobacteria; marine sponge clone HOC34	0.01	0.00
Bacteria; Proteobacteria; Gammaproteobacteria; Nitrincola	0.01	0.00
Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales	0.01	0.03
Bacteria; Proteobacteria; Gammaproteobacteria; Abundance and microbial life ocean crust seafloor lavas Loi'hi Seamount South Rift X3 clone P7X3b4C08	0.01	0.00
Bacteria; Proteobacteria; Gammaproteobacteria; ZD0417	0.01	0.01
Bacteria; Proteobacteria; Gammaproteobacteria; Microbial Adherent Sediment Particles Heavy Metal Contaminated North Sea Surface Sediments marine sediments clone	0.01	0.01
Bacteria; Proteobacteria; Gammaproteobacteria; marine boreal sponge Geodia baretti Norway:Korsfjord	0.01	0.00
Bacteria; Proteobacteria; Gammaproteobacteria; isolate str. EHK-1	0.01	0.01
Bacteria; Actinobacteria; Acidimicrobidae; identification marine actinobacteria sand isolate YM16-303Acidimicrobidae str. YM16-303	0.01	0.00
Bacteria; Chloroflexi; Chloroflexi-4; SAR307	0.01	0.00
Bacteria; Proteobacteria; Deltaproteobacteria; Myxococcales	0.01	0.02
Bacteria; Proteobacteria; Alphaproteobacteria; Acetobacterales	0.01	0.00
Bacteria; Proteobacteria; Alphaproteobacteria; Brucella_spHJ114	0.01	0.00
Bacteria; Proteobacteria; Gammaproteobacteria; Deselenobacterium	0.01	0.00
Bacteria; Proteobacteria; Gammaproteobacteria; Moraxellaceae	0.01	0.01
Bacteria; Proteobacteria; Gammaproteobacteria; Oceanrickettsia	0.01	0.01
Bacteria; Proteobacteria; Deltaproteobacteria; CTD005-82B-02	0.01	0.00
Bacteria; Actinobacteria; Acidimicrobidae; CL500-29	0.01	0.00
Bacteria; Isolation and identification hyper-ammonia producing swine storage pits manure	0.01	0.01
Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales	0.01	0.00
Bacteria; Proteobacteria; Deltaproteobacteria; Sva0853	0.01	0.00
Other	0.56	0.95

Table S5. Cluster of orthologous groups involved in chemotaxis and nutrient acquisition.

COG	Description	Class	Class description
COG0840	Methyl-accepting chemotaxis protein	NT	Multiple classes
COG0004	Ammonia permease	P	Inorganic ion transport and metabolism
COG3639	ABC-type phosphate/phosphonate transport system, permease component	P	Inorganic ion transport and metabolism
COG0609	ABC-type Fe^{3+}-siderophore transport system, permease component	P	Inorganic ion transport and metabolism
COG2375	Siderophore-interacting protein	P	Inorganic ion transport and metabolism
COG1918	Fe^{2+} transport system protein A	P	Inorganic ion transport and metabolism
COG0370	Fe^{2+} transport system protein B	P	Inorganic ion transport and metabolism
COG0659	Sulfate permease and related transporters (MFS superfamily)	P	Inorganic ion transport and metabolism
COG0053	Predicted Co/Zn/Cd cation transporters	P	Inorganic ion transport and metabolism

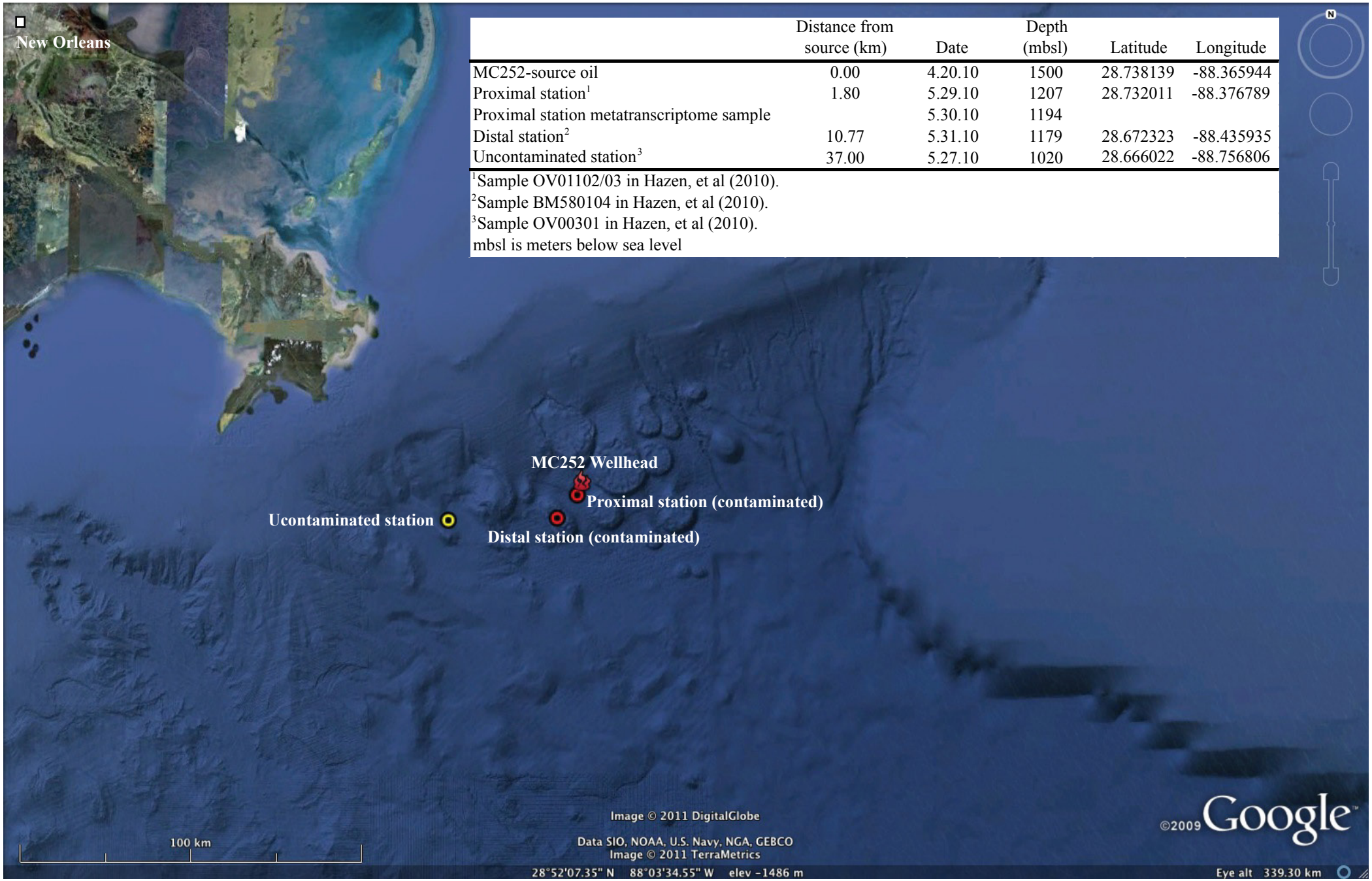
Table S6. Complete list of gene categories involved in hydrocarbon degradation from Figure 2. The genes in each gene category represent several genes, thus do not have accession numbers. Gene names in the table are directly from GenBank.

Alcohol dehydrogenase	alcohol dehydrogenase (acceptor)
	alcohol dehydrogenase (zinc-containing)
	alcohol dehydrogenase [benzyl]
	alcohol dehydrogenase GroES domain protein
	alcohol dehydrogenase zinc-binding domain protein
	alcohol dehydrogenase zinc-binding domain-containing protein
	alcohol dehydrogenase zinc-binding type 2
	alcohol dehydrogenase, class III, bacterial-like protein
	alcohol dehydrogenase, zinc-binding
	alcohol dehydrogenase, NADP-dependent
Aldehyde dehydrogenase	aldehyde dehydrogenase (acceptor)
	aldehyde dehydrogenase (NAD)
	aldehyde dehydrogenase 1
	aldehyde dehydrogenase A
	aldehyde dehydrogenase family protein
	aldehyde dehydrogenase, NADP-dependent
Alkane	alkane 1-monooxygenase
	alkane hydroxylase
	alkane hydroxylase A
	alkane hydroxylase-rubredoxin
	alkane monooxygenase
	transmembrane alkane 1-monooxygenase AlkB
Aromatic	aromatic dioxygenase large subunit
	aromatic ring dioxygenase alpha subunit
	aromatic ring dioxygenase subunit A
	aromatic ring hydroxylation dioxygenase A
	aromatic ring monooxygenase
	aromatic ring-hydroxylating dioxygenase
	aromatic-ring hydroxylase
	aromatic-ring-hydroxylating dioxygenase, alpha subunit
	aromatic-ring-hydroxylating dioxygenase, alpha subunit-like protein
	aromatic 1,2-dioxygenase, alpha subunit
Benzene	benzene dioxygenase
	benzene dioxygenase large subunit
	benzene monooxygenase oxygenase subunit
Benzoate	benzoate 1,2 dioxygenase alpha subunit
	benzoate 1,2-dioxygenase alpha subunit / Toluate 1,2-dioxygenase alpha subunit
	benzoate 1,2-dioxygenase beta subunit
	benzoate 1,2-dioxygenase hydroxylase component,alpha subunit
	benzoate 1,2-dioxygenase subunit alpha BenA
	benzoate 1,2-dioxygenase, large subunit
	benzoate 1,2-dioxygenase, small subunit
	benzoate 4-monooxygenase cytochrome P450
	benzoate CoA ligase
	benzoate diol dehydrogenase
	benzoate dioxigenase alpha subunit
	benzoate dioxygenase large subunit
	benzoate dioxygenase, alpha subunit
	benzoate-CoA ligase
	benzoate-CoA ligase family
	benzoate-coenzyme A ligase
Cyclohexanol dehydrogenase	cyclohexanol dehydrogenase

Table S6. Complete list of gene categories involved in hydrocarbon degradation from Figure 2. The genes in each gene category represent several genes, thus do not have accession numbers. Gene names in the table are directly from GenBank.

Cyclohexanone	cyclohexanone 1,2-monooxygenase
	cyclohexanone monooxygenase
	cyclohexanone monooxygenase 1
	cyclohexanone monooxygenase 2
Ethylbenzene	ethylbenzene dehydrogenase, beta subunit
	ethylbenzene dioxygenase alpha subunit
	ethylbenzene dioxygenase large subunit
Fatty Acid	fatty acid oxidation complex alpha subunit
	fatty acid oxidation complex alpha subunit [includes: enoyl-co hydratase (ec 4.2.1.17); 3-hydroxyacyl-coa dehydrogenas (ec 1.1.1.35); 3-hydroxybutyryl-coa epimerase (ec 5.1.2.3)]
	fatty oxidation complex
	fatty oxidation complex alpha subunit
	fatty oxidation complex alpha subunit [includes: enoyl-CoA hydratase; 3-hydroxyacyl-CoA dehydrogenase and 3-hydroxybutyryl-CoA epimerase]
	fatty oxidation complex, alpha subunit FadB
	fatty oxidation complex, alpha subunit FadJ
	fatty acid oxidation complex subunit alpha, multifunctional
	fatty acid desaturase
PAH	PAH dioxygenase iron sulfer protein large subunit
	PAH dioxygenase large subunit
	PAH ring-hydroxylating dioxygenase alpha subunit
Particulate methane monooxygenase	particulate methane monooxygenase A subunit
	particulate methane monooxygenase alpha subunit
	particulate methane monooxygenase protein A
	particulate methane monooxygenase subunit
	particulate methane monooxygenase subunit A
	particulate methane monooxygenase-like
	particulate methane monoxygenase beta (PmoA) subunit
	particulate methane oxygenase
	particulate Methane Oxygenase A
	PmoA
Ring cleavage/hydroxylating	extradiol ring-cleavage dioxygenase, class III enzyme, subunit B
	intradiol ring-cleavage dioxygenase
	intradiol ring-cleavage dioxygenase:Catechol dioxygenase, N-terminal
	large subunit of ring-hydroxylating dioxygenase
	ring hydroxylating alpha subunit
	ring hydroxylating alpha subunit (catalytic domain) protein
	ring hydroxylating dioxygenase alpha subunit
	ring hydroxylating dioxygenase alpha subunit (catalytic domain)
	ring hydroxylating dioxygenase subunit
	ring hydroxylating dioxygenase, alpha subunit:Rieske
	ring hydroxylating dioxygenase, alpha subunit:Rieske (2Fe-2S) region
	ring hydroxylating dioxygenase, alpha subunit/Rieske (2Fe-2S) protein
	ring hydroxylating dioxygenase, Rieske (2Fe-2S) protein
	ring-hydroxylating dioxygenase
	ring-hydroxylating dioxygenase alpha subunit
	ring-hydroxylating dioxygenase large subunit
	ring-hydroxylating dioxygenase large terminal subunit
Toluene	toluene 4-monooxygenase protein A
	toluene 4-monooxygenase protein E
	toluene dioxygenase
	toluene dioxygenase small subunit
	toluene o-xylene monooxygenase component
	toluene-3-monooxygenase oxygenase subunit 1
	toluene-3-monooxygenase oxygenase subunit 2
	toluene-4-monooxygenase system protein A
	toluene/benzoate dioxygenase alpha subunit
	toluene/biphenyl Rieske non-heme iron oxygenase alpha subunit
	toluene 1.2-dioxygenase alpha subunit

Figures



	Distance from source (km)	Date	Depth (mbsl)	Latitude	Longitude
MC252-source oil	0.00	4.20.10	1500	28.738139	-88.365944
Proximal station[1]	1.80	5.29.10	1207	28.732011	-88.376789
Proximal station metatranscriptome sample		5.30.10	1194		
Distal station[2]	10.77	5.31.10	1179	28.672323	-88.435935
Uncontaminated station[3]	37.00	5.27.10	1020	28.666022	-88.756806

[1]Sample OV01102/03 in Hazen, et al (2010).
[2]Sample BM580104 in Hazen, et al (2010).
[3]Sample OV00301 in Hazen, et al (2010).
mbsl is meters below sea level

Figure S1. Map of sampling sites and MC252 Deepwater Horizon oil spill site.

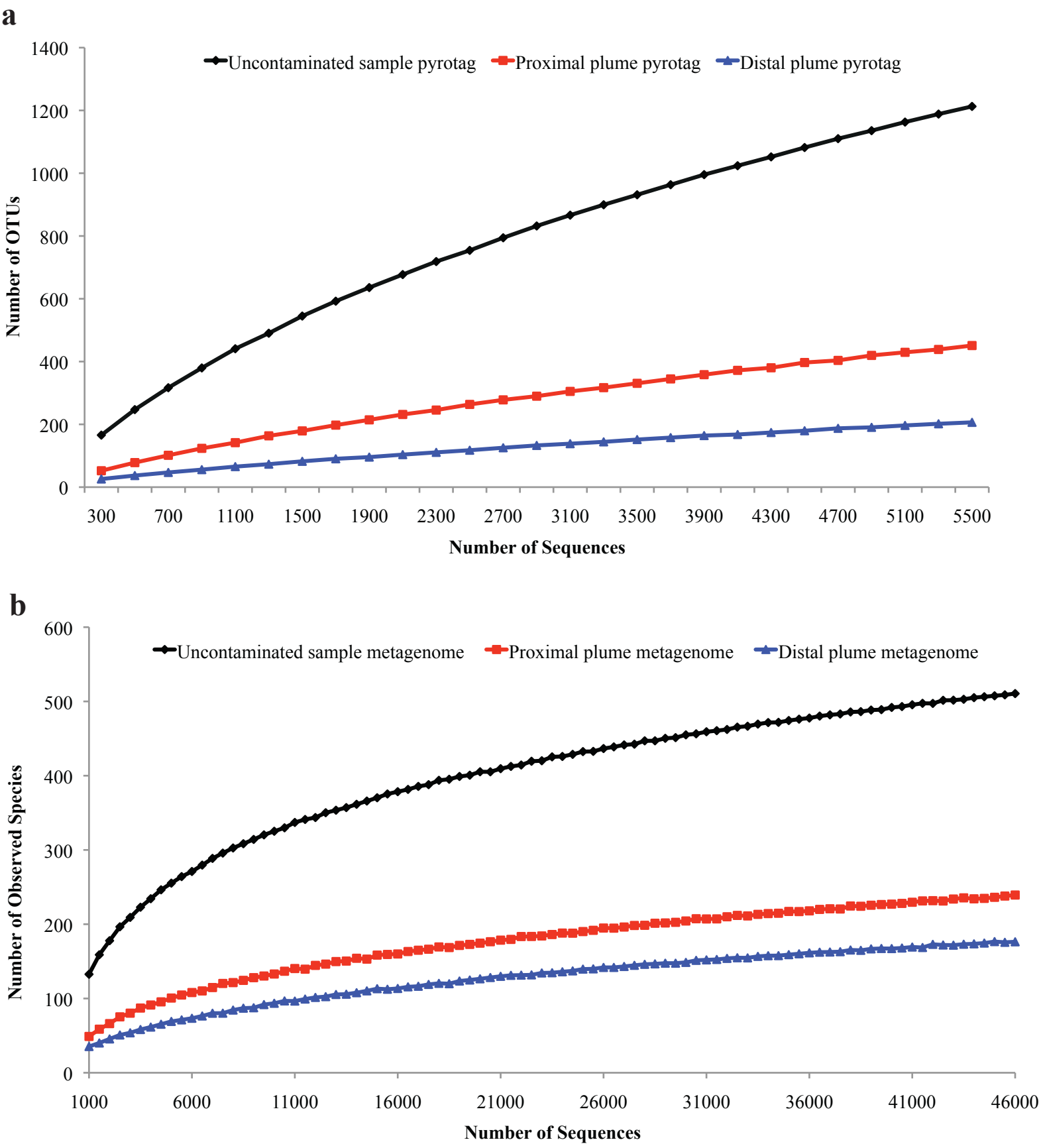
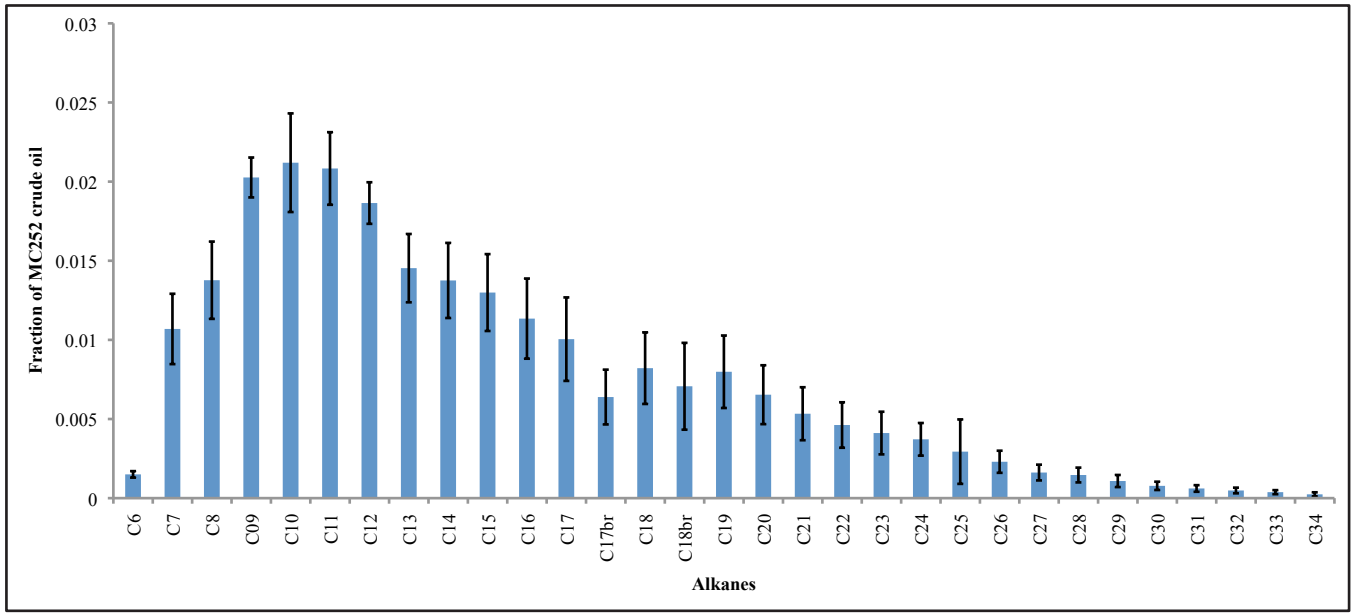
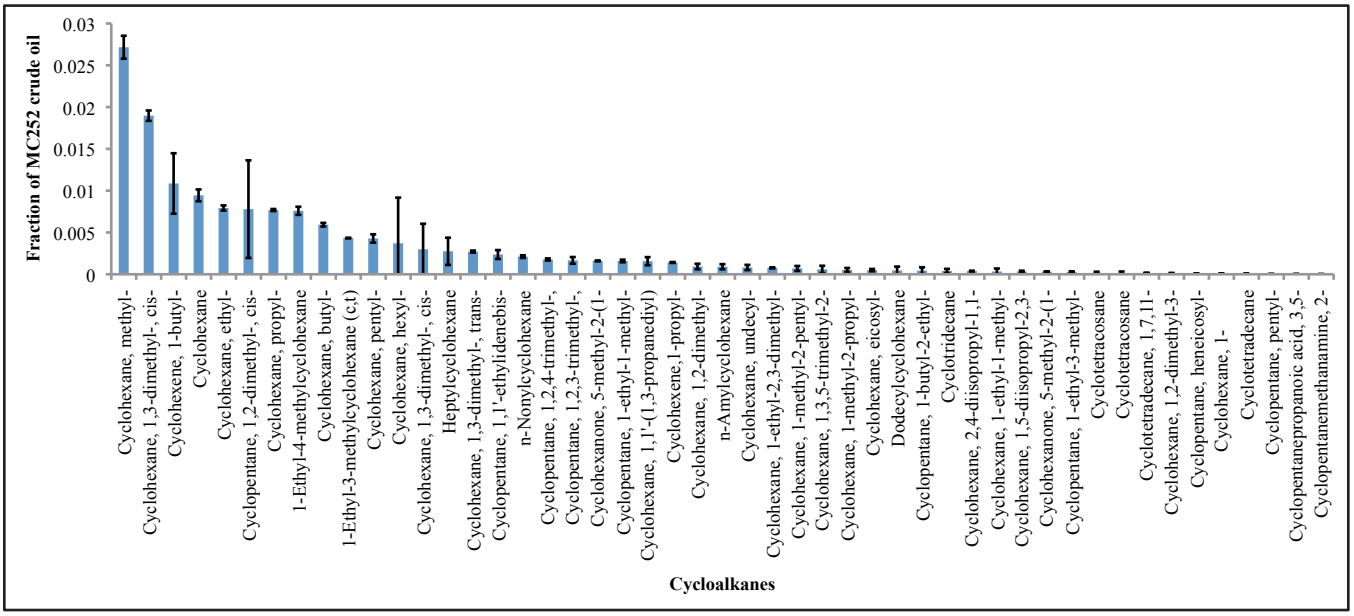


Figure S2. Rarefaction curves for 16S rRNA genes from the proximal and distal plume stations and from the uncontaminated sample collected from plume depth. Data was rarified prior to analysis. A) Rarefaction curve of pyrotag data. B) Rarefaction curve of metagenome data.





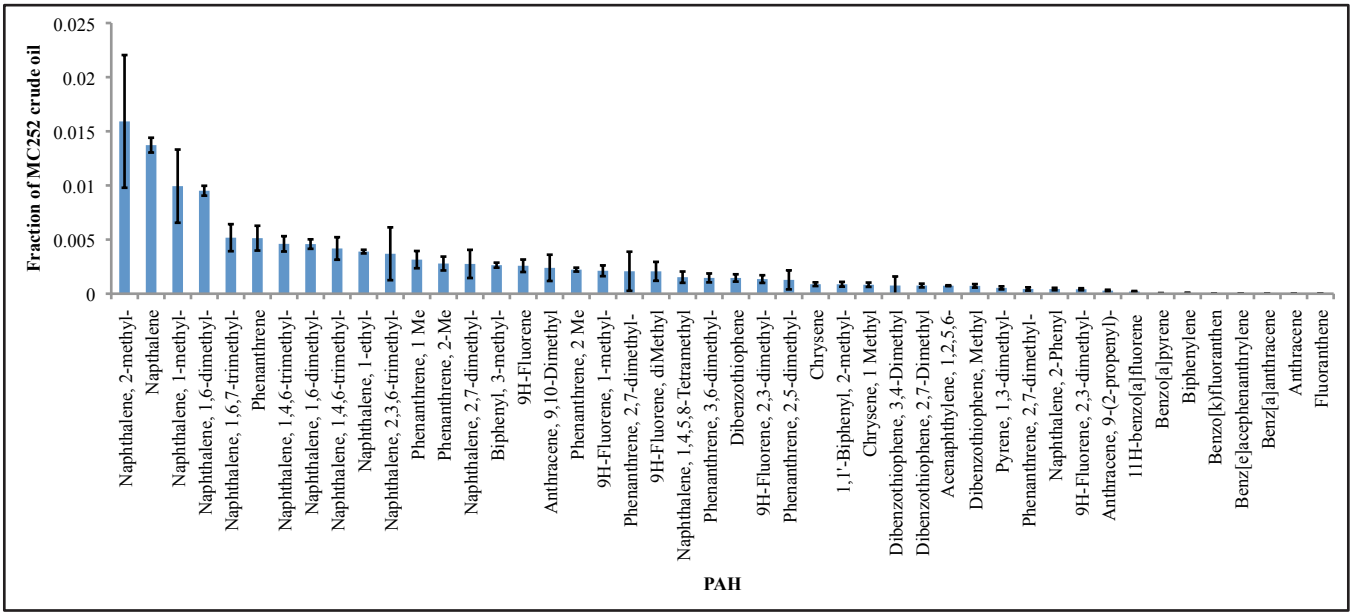


Figure S3. MC252 crude oil constituents >C5 determined by direct on-column GC/MS analysis. Bars indicate average (n = 3) fraction of each constituent in source oil.

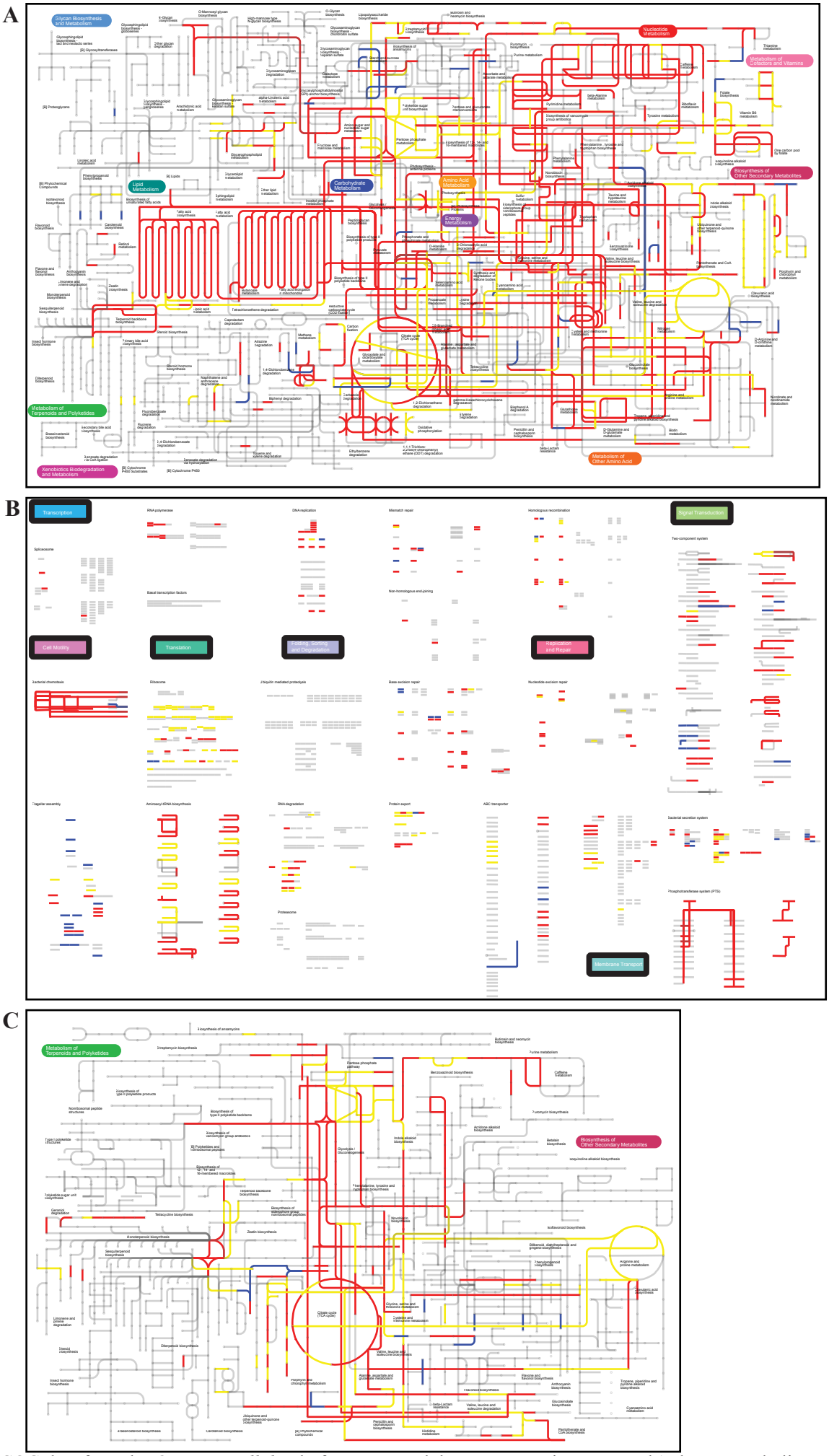


Figure S4. Mapped COG data from the *Oceanospirillales* draft genome and the metatranscriptome.Panel A shows metabolic pathways; panel B shows regulatory pathways; panel C shows the pathways for biosynthesis of secondary metabolites. Red indicates that an element was present in the draft genome and expressed; blue indicates that an element was present in the draft genome only; yellow indicates that an element was present in the metatranscriptome only.

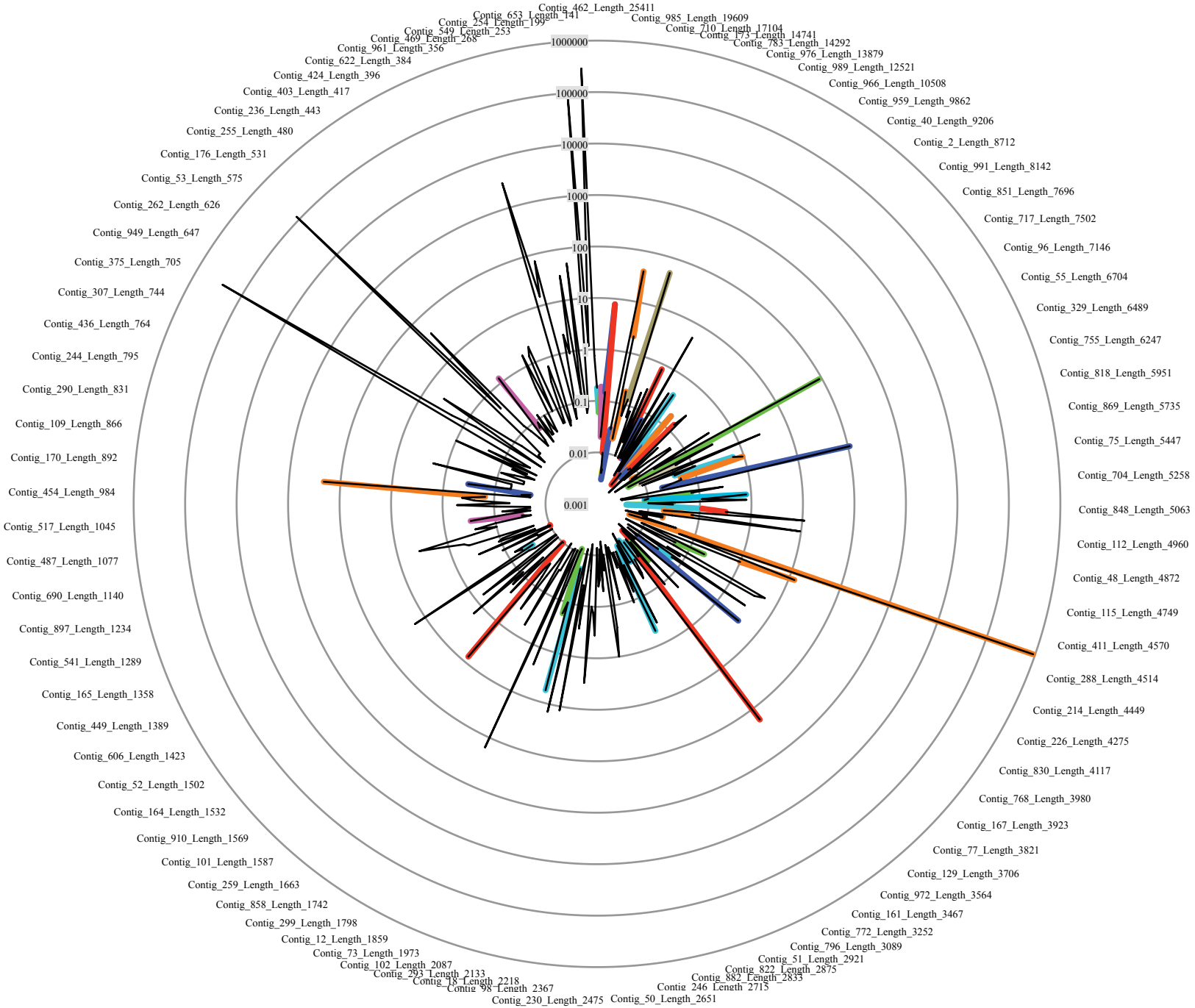


Figure S5. Unassembled paired-end reads (5.1 Gb) from the metatranscriptome were mapped to the *Oceanospirillales* draft genome. A total of 3.0 Gb from the metatranscriptome mapped to the assembled draft genome. The average contig coverage is shown above. Contigs containing COGs coding for methyl-accepting chemotaxis proteins are shown in red. Contigs containing COGs for flagellar processess are shown in green. COGs for pilin are shown in purple. COGs implicated in biofilm formation are shown in dark blue. COGs for nutrient acquisition are shown in light blue. COGs for signal transduction are shown in orange. COGs for plasmid maintenance and stabilization are shown in light brown. Contigs are shown in length order.

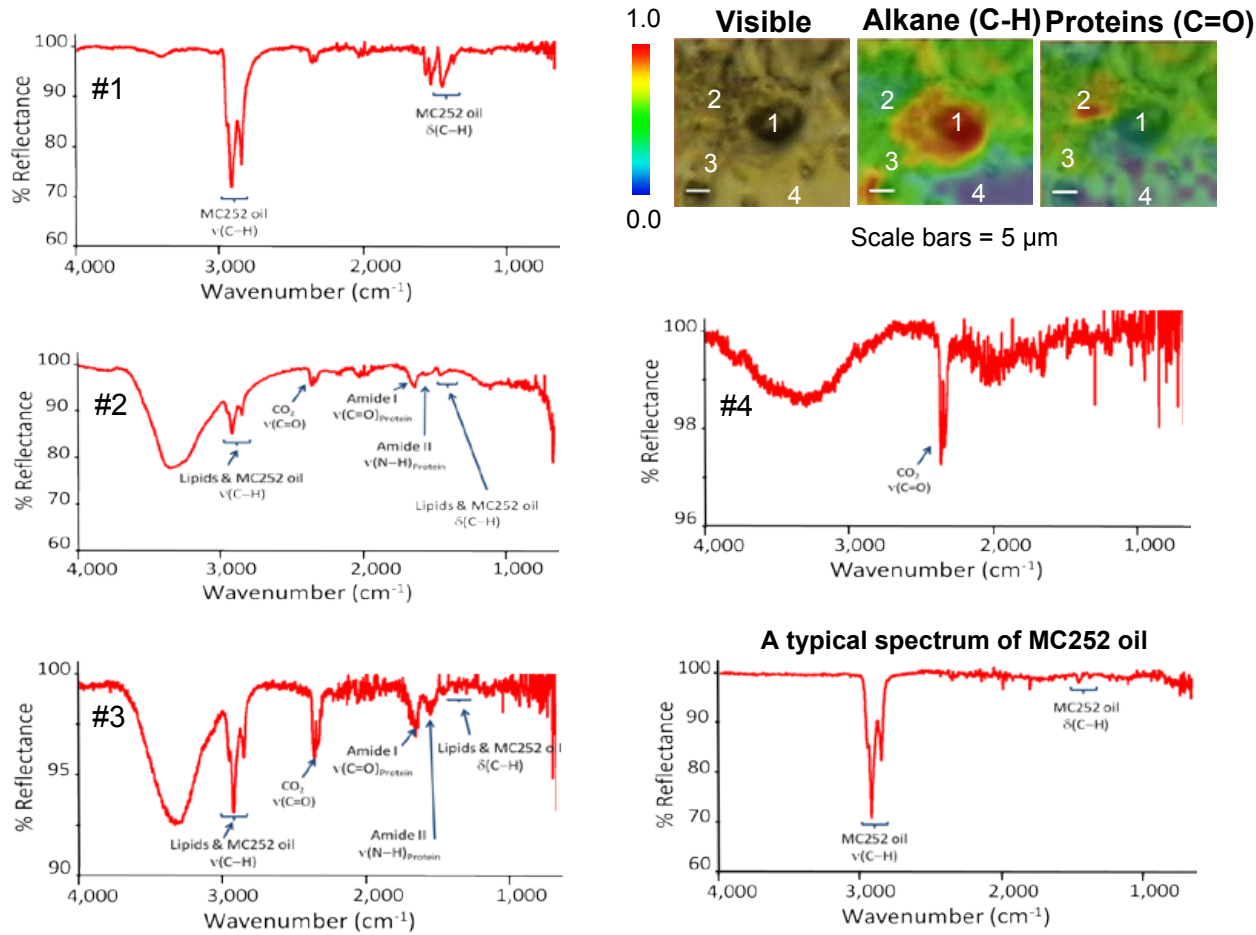


Figure S6. SR-FTIR analysis of proximal plume water. The spot labelled as 1 in the visible and infrared image corresponds to spectrum #1, which is interpreted as MC252 oil. Spot 2 corresponds with spectrum #2, which is interpreted as a a cluster of prokaryotic cells. Spot 3 corresponds with spectrum #3, which represents a mixture of prokaryotic cells and MC252 oil.