nature microbiology

LETTERS

PUBLISHED: 9 MAY 2016 | ARTICLE NUMBER: 16057 | DOI: 10.1038/NMICROBIOL.2016.57

Reconstructing metabolic pathways of hydrocarbon-degrading bacteria from the Deepwater Horizon oil spill

Nina Dombrowski[1], John A. Donaho[1], Tony Gutierrez[2], Kiley W. Seitz[1], Andreas P. Teske[3] and Brett J. Baker[1]*

The Deepwater Horizon blowout in the Gulf of Mexico in 2010, one of the largest marine oil spills[1], changed bacterial communities in the water column and sediment as they responded to complex hydrocarbon mixtures[2–4]. Shifts in community composition have been correlated to the microbial degradation and use of hydrocarbons[2,5,6], but the full genetic potential and taxon-specific metabolisms of bacterial hydrocarbon degraders remain unresolved. Here, we have reconstructed draft genomes of marine bacteria enriched from sea surface and deep plume waters of the spill that assimilate alkane and polycyclic aromatic hydrocarbons during stable-isotope probing experiments, and we identify genes of hydrocarbon degradation pathways. Alkane degradation genes were ubiquitous in the assembled genomes. *Marinobacter* was enriched with *n*-hexadecane, and uncultured Alpha- and Gammaproteobacteria populations were enriched in the polycyclic-aromatic-hydrocarbon-degrading communities and contained a broad gene set for degrading phenanthrene and naphthalene. The repertoire of polycyclic aromatic hydrocarbon use varied among different bacterial taxa and the combined capabilities of the microbial community exceeded those of its individual components, indicating that the degradation of complex hydrocarbon mixtures requires the non-redundant capabilities of a complex oil-degrading community.

Marine microorganisms derive energy and carbon from the degradation of petroleum hydrocarbons and drive the bioremediation process during anthropogenic oil spills, such as the Deepwater Horizon (DWH) spill[2,4,7–9]. Various uncultured bacteria, primarily belonging to *Oceanospirillales* and other Gammaproteobacteria, were enriched during the DWH spill and are believed to have played a major role in oil degradation[2,3,10–12]. However, this assumption relies on the dominance of these organisms in environmental sequencing surveys, whereas the metabolic potential and pathways for degrading hydrocarbons in bacterial populations, which demonstrably assimilate these compounds, remain to be determined. To address this question, we obtained ~113 Gb of shotgun metagenomic data from hydrocarbon-degrading enrichments using stable-isotope probing (SIP) experiments[13]. Enrichments were obtained from weathered hydrocarbons floating on the sea surface and from the subsurface hydrocarbon plume, collected during the DWH oil spill in 2010. The plume sample was incubated with ^{13}C-labelled *n*-hexadecane (HEX), and the sea surface sample was separately incubated with ^{13}C-labelled naphthalene (NAP) or phenanthrene (PHE) in SIP experiments[13]. *De novo* assembly and (tetranucleotide signature) binning of metagenomic sequences from the ^{13}C-labelled fractions allowed the reconstruction of 7 high-quality (completeness of 52–95%) and 17 fragmentary draft genomes (completeness <50%), as well as ~13% of unbinned scaffolds (see Methods, Supplementary Table 1 and Supplementary Fig. 1).

We first annotated genes within all assembled genomes. Taxonomic assignment of annotated genes in the HEX, NAP and PHE assemblies revealed distinct bacterial communities (Supplementary Fig. 2a,b). *Marinobacter* and *Alcanivorax* dominated the HEX-degrading community, and *Alteromonas* and *Cycloclasticus* were abundant in NAP and PHE samples, respectively. This community spectrum coincided with results from the previous SIP-based analysis of the same samples using 16S rRNA clone sequencing[13]. To identify taxa representing the seven high-quality draft genomes, we constructed a phylogenetic tree using up to 15 ribosomal proteins (Fig. 1 and Supplementary Table 1). We recovered a 95% complete *Marinobacter* genome from the HEX enrichment, designated H-Mar (99% inferred protein similarity to *Marinobacter salarius* R9SW1, Fig. 1 and Supplementary Table 1). From the NAP-degrading community we reconstructed two gammaproteobacterial genomes belonging to *Oceanospirillales* (N-Alc, 73% similar to *Alcanivorax* sp. 43B_GOM-46m, 84% complete) and *Alteromonadales* (N-Alt, 86% similar to *Alteromonas macleodii* English Channel 673, 80% complete). An additional 93% complete alphaproteobacterial genome belonged to the order *Rhodospirillales* (N-Tha, 99% similar to *Thalassospira profundimaris* WP0211). Three genomes from the PHE-degrading communities included a bacterium 83% similar to *Cycloclasticus pugetii* PS-134H (P-Cyc, 52% complete), an *Oceanospirillales* member with 73% similarity to *Neptuniibacter caesariensis* MED92 (P-Nep, 86% complete) and a member of the *Alteromonadales* with 64% similarity to *Colwellia psychrerythraea* 34H (P-Col, 70% complete).

The NAP-enriched taxa *Rhodospirillales*, *Oceanospirillales and Alteromonadales* (N-Tha, N-Alc and N-Alt) were previously associated with hydrocarbon degradation. For example, isolates of the order *Alteromonadales* showed alkane and polycyclic aromatic hydrocarbon (PAH)-degrading activity in previous SIP experiments[13]. The order *Rhodospirillales* accumulated during the DWH spill in sea surface samples, and isolates of this order, such as *Thalassospira tepidiphila*, degrade PAHs[3,14]. We also assembled one genome belonging to *Alcanivorax* from the NAP-degrading community (N-Alc), a genus that was barely detected in the spill itself but was present in the previous SIP experiments[13]. Additionally, *Alcanivorax* was abundant in HEX enrichments (Supplementary Fig. 2b). Although *Alcanivorax* isolates degrade alkane hydrocarbons and none have been described to metabolize PAHs[13,15], *Alcanivorax* was previously detected in bacterial consortia growing

[1]Department of Marine Science, University of Texas Austin, Marine Science Institute, Port Aransas, Texas 78373, USA. [2]School of Life Sciences, Heriot-Watt University, Edinburgh EH14 4AS, UK. [3]Department of Marine Sciences, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599 USA. *e-mail: acidophile@gmail.com

Figure 1 | Phylogenetic characterization of genomic bins reconstructed from three SIP enrichments. Maximum-likelihood-based phylogenetic tree of up to 15 concatenated ribosomal proteins (rpL2, 3, 4, 5, 6, 14, 15, 18, 22, 24 and rpS3, 8, 10, 17, 19) from seven high-quality genomes assembled from the *n*-hexadecane (H), naphthalene (N) and phenanthrene (P) SIP enrichments (bold). Two genomes from previously published DWH plume-derived metagenomes are indicated in bold italics[10,46]. Bootstrap values were generated using MrBayes. Two Planctomycetes strains were used as outgroup. Black, grey and white circles: nodes with bootstrap values of 100, 75–100 and <75% (1,000 replicates).

on pyrene as the sole carbon source[16,17]. Thus, the NAP enrichment of *Alcanivorax* supports a broader versatility in *Alcanivorax* for degrading not only alkanes but also PAHs, potentially by participating in bacterial consortia that fully metabolize PAHs.

Focusing on the PHE-degrading community, the genera *Cycloclasticus*, *Colwellia* and *Neptuniibacter* (corresponding to P-Cyc, P-Col and P-Nep) represented 66, 13 and 3% of clones in previous PHE-SIP experiments, respectively[13]. Notably, the *Cycloclasticus* 16S rRNA phylotype recovered from the previous SIP study was >99% similar to the most dominant *Cycloclasticus* phylotype that was enriched at the sea surface during the spill[13]. Concordantly, *Colwellia* was enriched in laboratory experiments in dispersant-oil microcosms[18]. Thus, the assembled genomes represent abundant, as well as rare, hydrocarbon-degrading DWH spill community members.

16S rRNA gene sequences recovered from the SIP assemblies co-clustered with sequences derived from the previous SIP experiments and partially with sequences enriched in surface and plume samples during the spill (Supplementary Fig. 3)[2,3,13]. These clustering patterns therefore indicate that the bacterial lineages present in our metagenomic assemblies were present during the spill and do not represent artefacts of the SIP experiment. To confirm this inference, we mapped our SIP assemblies to published plume-derived metagenomic and metatranscriptomic data sets[10]. The HEX-degrading assembly recruited 2.7 and 0.8% of metagenomic reads from plume and unpolluted samples, respectively (Supplementary Table 2). Annotated genes from the HEX assembly matched 6.3 and 1.3% of metatranscriptomic reads from samples collected proximal and distal, respectively, from the plume (Supplementary Table 2 and Supplementary Figs 4 and 5). This suggests that genes from our assembled genomes were enriched and more highly expressed with increasing hydrocarbon exposure. However, despite the percentage of mapped reads, we were unable to locate specific alkane-degradation genes, possibly due to insufficient sequencing depth that would allow covering non-abundant taxa (Supplementary Fig. 5). No metagenomic data set from the sea surface of the spill is available; but based on previous 16S rRNA gene studies, Alphaproteobacteria (that is, order *Rhodospirillales*) as well as Gammaproteobacteria (orders *Alteromonadales* and *Oceanospirillales*) were enriched in oil-contaminated sea surface samples[3]. Thus, the lineages reconstructed in our assemblies reflect (or resemble) those enriched in hydrocarbon-contaminated sea surface communities.

To identify genes that enabled bacterial community members to be enriched during the DWH spill, we compared the gene content between the different hydrocarbon-degrading communities obtained in the SIP assemblies. Of the detected 4,756 unique gene functions, ~40% were shared and 4, 13 and 22% were unique among the HEX, NAP and PHE assemblies, respectively (Supplementary Fig. 6 and Supplementary Data 1). The PHE enrichment showed selection for serine/threonine kinases, while the HEX samples encoded for abundant short-chain dehydrogenases and the NAP assembly was enriched for tripartite ATP-independent periplasmic (TRAP) transporters. Serine/threonine kinases phosphorylate serine, threonine and tyrosine are proposed mediators of bacteria–bacteria interactions, while TRAP transporters function in osmoregulated solute transport[19,20]. Additionally, several functional categories were present in individual genomic bins (Fig. 2 and Supplementary Data 2). Five of seven genomes carried genes for bacterial chemotaxis and six of seven for flagella biosynthesis, hinting at active motility and therefore the

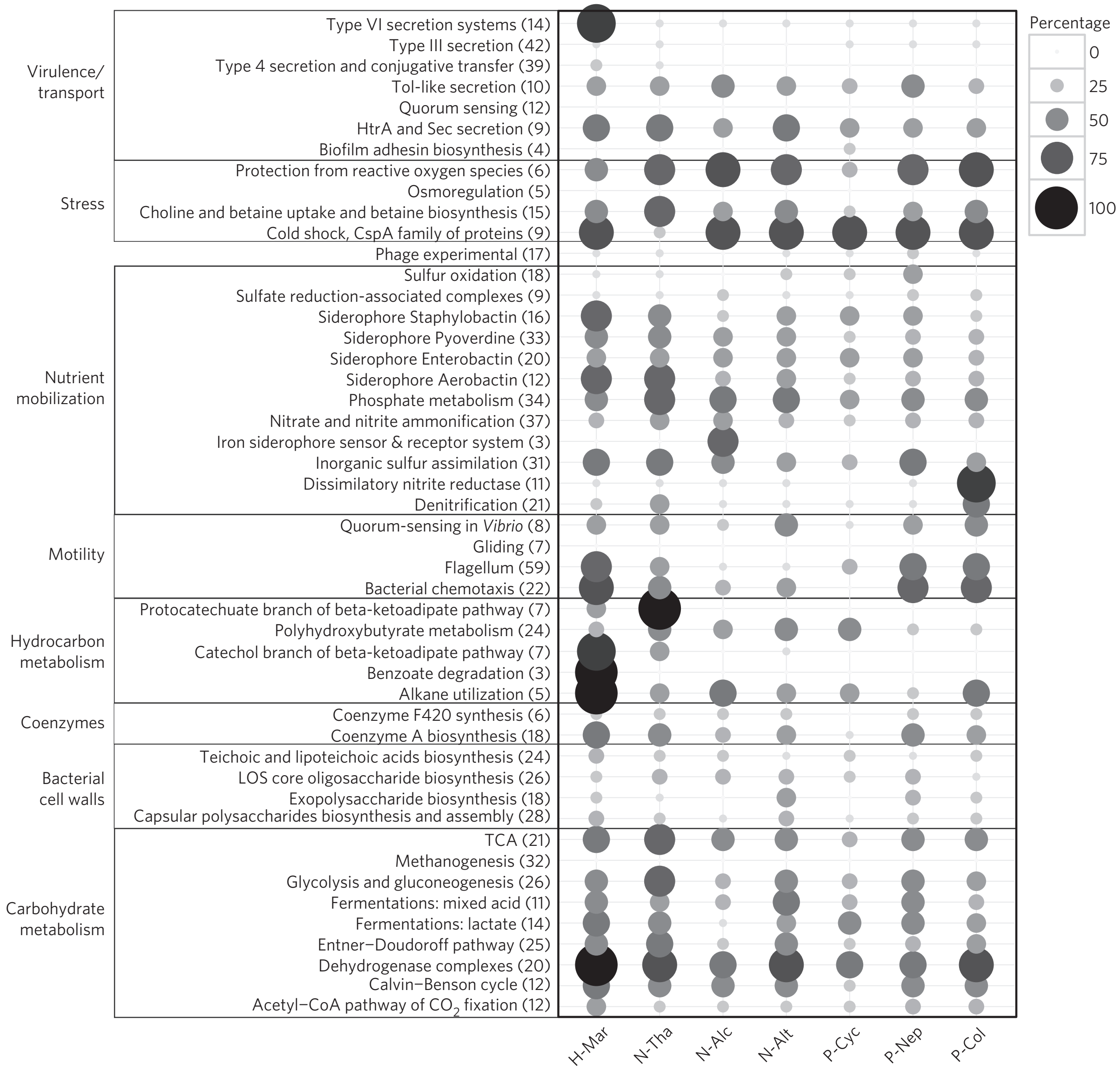

Figure 2 | Abundance of functional categories detected within the reconstructed genomes. Functional categories detected in the seven draft genomes assembled from the *n*-hexadecane (H), naphthalene (N) and phenanthrene (P) SIP enrichments. Functional categories are based on the SEED subsystems[36] and were compared against genes annotated within each metagenomic bin using blastp (e-value cutoff of 1×10^{-20}). Numbers in parentheses: total number of genes within each functional category. Percentage: per cent of total number of genes per subsystem found within each bin.

ability to move towards oil substrates. Additionally, P-Col encoded for several denitrification genes (13 of 21 tested genes), including *norB/C*, *nosR* and *nosD*, while P-Nep appears able to oxidize sulfur (that is, including *soxA-D* and sulfite oxidase), implying relevance for degrading sulfur-containing components of oil and the applied dispersant Corexit. Suggesting physiological adaptations to phosphorous and iron limitation, most genomic bins contained genes relevant for phosphate starvation or siderophore biosynthesis, such as staphylobactin (that is, *phoH* and *sbnA*; Supplementary Data 2). The presence of major genes for motility and for use of scarce nutrients suggests that the enriched organisms are well adapted for chemotactic motility towards their hydrocarbon substrates and for physiological responses to nutrient-limiting conditions that characterize oil-induced bacterial blooms[21]. Alternatively, these genes might be of general importance for survival and growth in the Gulf of Mexico.

To resolve the hydrocarbon degradation pathways in the assembled bacterial genomes, we searched for homologues to known degradation genes. An entire pathway for alkane degradation was found in H-Mar (*Marinobacter*), the dominant genus in the HEX-degrading community (Fig. 3a and Supplementary Figs 1 and 2). This result is consistent with isolates of *Marinobacter* being able to degrade HEX, increasing *Marinobacter* abundance in DWH plume samples as validated by fluorescence *in situ* hybridization (FISH), and enrichment of genes for alkane degradation in metatranscriptomes from the DWH spill and the Gulf coast[2,10,22,23]. Additionally, we reconstructed the alkane-degradation pathway in most of the genomes from the NAP- and PHE-degrading bins (Fig. 3a). However, based on the SIP approach, only a subset of bacteria appears to actively employ these pathways—*Marinobacter* (H-Mar)—and potentially outcompete other alkane-degrading bacteria, such as *Alcanivorax* (N-Alc). To validate this finding, metatranscriptomic sequencing could be used to map active gene expression to our assembled draft genomes. One unique feature of the active alkane-degrader H-Mar is the presence of genes encoding for the type VI secretion system[24,25], which were mostly absent from all other genomic bins (Fig. 2). The type VI secretion system is involved in inter-bacterial competition and has been discussed to be relevant for alkane assimilation[24,25], and thus could provide H-

NATURE MICROBIOLOGY DOI: 10.1038/NMICROBIOL.2016.57

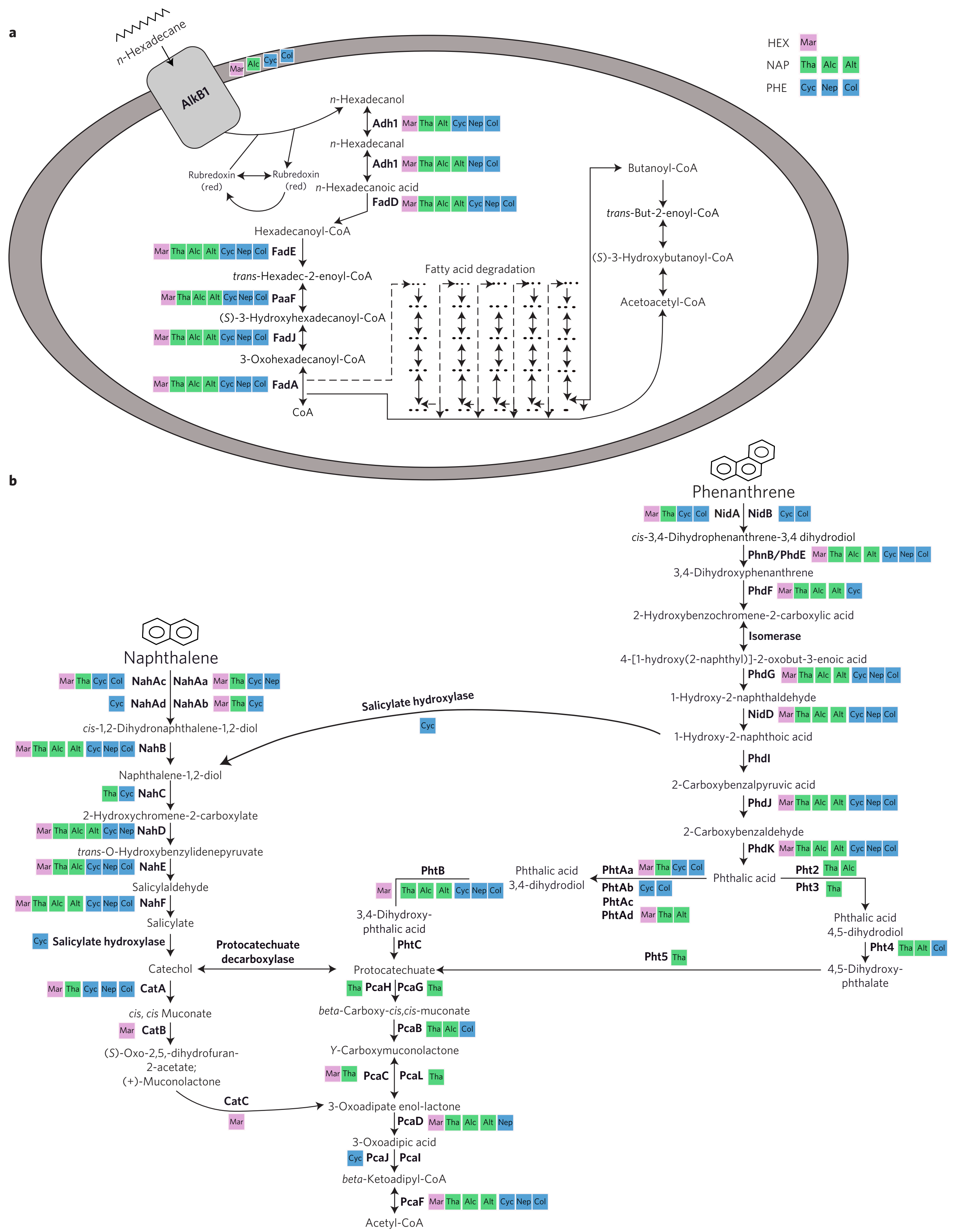

Figure 3 | Distribution of key alkane and polycyclic hydrocarbon degradation pathways in reconstructed genomes. a, Metabolic reconstruction of the *n*-hexadecane (HEX) degradation pathway within seven high-quality metagenomic bins. **b**, Metabolic reconstruction of naphthalene (NAP) and phenanthrene (PHE) degradation pathways within seven high-quality genomes. Purple, green, blue: HEX, NAP and PHE-degrading bins, respectively.

Mar with a competitive advantage over other non-active alkane degraders.

To characterize PAH degradation pathways in SIP-enriched bacteria, we identified genes encoding for enzymes in both NAP- and PHE-degradation pathways in our assembled genomes (Fig. 3b). Of 41 described NAP- and PHE-degradation genes, 19 were present in >50% of genomes, while a subset of genes, such as *nahC*, salicylate hydroxylase or *phdF*, were only detectable in a few genomes. For example, within N-Tha, N-Alc and N-Alt, we were unable to annotate the full known NAP-degradation pathway. N-Tha (*Thalassospira profundimaris*) is equipped to degrade NAP to salicylate, but lacks the salicylate hydroxylase, which converts salicylate to catechol. Only the fragmentary genome NAP-25 (*Oceanospirillales*) encoded the gene for this step among the assembled genomes from the NAP enrichments. Similarly, the extradiol dioxygenase (PhdF), which converts 3,4-dihydroxyphenanthrene to 2-hydroxybenzochromene-2-carboxylic acid early on in the PHE-degradation pathway, was only detected in P-Cyc (*Cycloclasticus*). Unless all pathway gaps are without exception ascribed to sequencing gaps within the genomic bins, a more parsimonious interpretation of these results suggests that alternative genes and enzymes for these individual steps exist, which are not represented in gene databases. Alternatively, a coordinated community activity might be required to completely degrade polycyclic hydrocarbons during the DWH spill.

Our study combines metagenomics and SIP, allowing the assembly of genomes from active hydrocarbon-degrading bacteria present during the DWH spill, and enriched from deep plume and surface samples based on alkane and PAH assimilation. All assembled genomes possessed pathways for alkane and PAH degradation, suggesting that several taxa actively degraded hydrocarbons during the spill (Fig. 3). For example, H-Mar (*Oceanospirillales*) corresponds to *Marinobacter salarius*, which degrades alkanes in the plume, and the bacterium that constitutes N-Tha (*Rhodospirillales*) was responsible for degradation of PAHs at the sea surface. These taxa were enriched during the spill and their genes for hydrocarbon degradation were detected in the plume and sea surface[2,3,10,22,26]. However, the plume-derived genome (H-Mar) appears to be enriched only at a low relative abundance compared to previous data sets, where an uncharacterized *Oceanospirillales* represented >60% of sequence reads[10]. This disparity suggests that low-abundant community members were active during the spill, as observed in deep-sea hydrothermal plumes[27]. Additionally, differences between bacterial taxa detected during the DWH spill compared to SIP enrichments could be a reflection of the experimental set-up. For example, no dispersant was added during the experiment, but Corexit was added during the spill[13,28] and found to select against *Marinobacter* in microcosm experiments[18]. Additionally, the SIP procedure required incubation steps at room temperature instead of 4 °C to obtain sufficient incorporation of the ^{13}C label into DNA for SIP[13]. Such temperature shifts have been demonstrated to induce changes in plume bacterial community composition[3]. Whole-community gene expression studies will be required to evaluate the activity of our assembled genomes in their native community context.

Within the constraints inherent to the SIP experimental approach, combining SIP with metagenomics enabled us to identify and analyse hydrocarbon-degrading pathways within several uncultured bacteria. Bacteria that are usually not associated with PAH degradation (that is, members of *Alcanivorax* in the NAP enrichments) persisted at sufficiently high levels for metagenomic reconstruction, indicating that complete pathways for the degradation of PAHs may not be obligatory for individual community members that thrive in PAH-degrading enrichments. Although we cannot exclude that genes for hydrocarbon degradation were missed in our assemblies, the apparent partitioning of key pathway steps to individual community members suggests an alternate hypothesis, in that complete degradation of a complex mixture of PAHs requires the coordinated response of a complex bacterial community and the coordination of its metabolic pathways.

Methods

Sample collection and preparation. DNA samples for metagenomic reconstruction were obtained from a previously published stable-isotope probing (SIP) experiment[13]. Briefly, an oil-contaminated sea surface sample (original ID PE5) was collected near a site of the DWH spill (28°44.175′ N, 88°22.335′ W) in the Gulf of Mexico on 5 May 2010. Two deep hydrocarbon plume samples (original ID B3 and B6, biological replicates) were collected near the Macondo wellhead (28°41.686′ N, 88°26.081′ W) at depths of 1,170 and 1,210 m on 31 May 2010 and combined for the SIP experiment. Sea surface and plume samples were collected at ~0.86 and 3.5 m from the wellhead. SIP experiments for the plume sample (B3+B6) were performed with ^{13}C-labelled HEX and for the sea surface sample (PE5) with ^{13}C-labelled NAP or PHE (yielding a total of three samples). Incubations containing solely ^{12}C-unlabelled substrate were run in parallel to act as unlabelled controls. Total DNA from labelled and unlabelled incubations from each of the three SIP experiments was extracted as described previously[29]. Extracted labelled and unlabelled DNA (0.7 to 1.0 μg for each sample) was separated using CsCl gradient ultracentrifugation[13]. To prepare these previously processed SIP samples for sequencing, DNA concentrations were measured using a Qubit 3.0 Fluorometer and a final concentration of 10 ng μl^{-1} for each sample (using a total amount of 100 ng) from the heavy-labelled fraction was used to prepare libraries for paired-end Illumina (HiSeq 2500) sequencing.

Metagenomic assembly and binning. Illumina library preparation and sequencing was performed by the Genome Sequencing and Analysis Facility (GSAF) at the University of Texas at Austin. For each of the three samples, two Illumina libraries were prepared and sequenced (technical replicates). Sequencing was performed on an Illumina HiSeq 2500 with the following specifications: high-throughput run mode, run type paired end 2 × 125 bp, 6 × 4.0E8 target reads (millions), insert size of ~360–420 bp and ~5% PhiX control spike-in. This sequencing approach provided ~113 Gb of sequencing data (373,279,006, 289,782,708 and 246,243,202 reads from the HEX, NAP and PHE samples, respectively). Raw Illumina shotgun genomic reads were separated from Illumina artefacts by removing the adaptors and DNA spike-ins from the forward and reverse reads. Therefore, a sliding window approach using a kmer size of 28 and a step size of 1 was used. Reads with ≥3 Ns (unambigous nucleotide), an average quality score <20 and a read length of <50 bps were removed using cutadapt (yielding 886,823,725 total reads)[30]. Afterwards, reads were interleaved using interleave_fasta.py and the interleaved sequences were trimmed using Sickle and a minimum quality cutoff of 5 (yielding 877,414,119 total reads)[31]. The script for interleave_fasta.py can be found at https://github.com/jorvis/biocode/blob/master/fasta/interleave_fasta.py. Metagenomic reads from all SIP samples were individually assembled using IDBA-UD and the following parameters: –pre_correction, -mink 75, -maxk 105, –step 10, –seed_kmer 55 (ref. 32). This yielded a total of 2,739,076 scaffolds from all three SIP enrichments. The minimum and maximum scaffold length ranged from 200 to 130,027 bp.

Metagenomic binning was performed on assembled SIP enrichments using tetranucleotide frequencies on scaffolds with a length ≥4,000 bp (including a total of 4,874 scaffolds)[33]. The resulting emerging self-organizing maps (ESOMs) were manually sorted and curated (Supplementary Fig. 1)[33]. Metagenomic binning was enhanced by incorporating reference genomes as genetic signatures for the assembled contigs into ESOMs[33,34]. In this way we assembled 7 high-quality and 17 fragmentary metagenomic bins (completeness threshold of 50%). The seven high-quality genomes showed a completeness ranging from 52–95%, included 1,492/4,874 scaffolds with a scaffold length between 4,006 and 257,386 bp. The 17 fragmentary genomes displayed a completeness below 50% and comprised 2,737/4,874 scaffolds that ranged from 4,000 to 44,889 bp. Additionally, 645/4,874 scaffolds remained unbinned (4,002–77,080 bp length). After binning, draft genomes were linked to the SIP enrichments based on their unique scaffold ID. CheckM was used to evaluate the accuracy of the binning approach by determining the percentage of completeness and contamination (Supplementary Table 1)[35]. Contaminants that were identified based on their phylogenetic placement (wrong taxonomic assignment compared to the average taxonomic assignment of the genes assigned to each bin; see section on taxonomic assignment), GC content (>25% difference compared to the mean of all scaffolds assigned to each bin) or confidence level (below 0.5) were manually removed from the bins.

Gene calling, taxonomic assignment and functional characterization. Gene calling and taxonomic assignment of the three SIP enrichments and seven high-quality draft genomes were performed using the Joint Genome Institute–Production Genomics Facility (JGI-PGF) Integrated Microbial Genomes with Microbiome (IMG/M) system. To achieve this, the complete assemblies from the three SIP enrichments were uploaded onto the JGI server. The retrieved JGI output was linked to the seven draft genomes using their unique scaffold IDs. The JGI output summarizing the taxonomic assignment of all gene annotations was used to depict the taxonomic composition as shown in Supplementary Fig. 2. The JGI-based amino-acid sequences were used as a database for blastp searches (with the used quality parameters as indicated in the text).

For the functional characterization of our draft genomes, a reference gene database was assembled by downloading SEED subsystems using the network-based SEED API (function svr_all_subsystems)[36]. The SEED-based reference database was manually curated to include only subsystems of interest by downloading the gene functions and gene IDs of individual subsystems (functions svr_subsystem_roles and svr_subsystem_spreadsheet), which were then used to retrieve their respective fasta sequences (function svr_fasta). The SEED subsystems included are shown in Fig. 2, and the numbers in parentheses include the total number of genes screened per subsystem (Fig. 2 and Supplementary Data 2). Additionally, we manually curated a reference database including only hydrocarbon degradation genes by searching the KEGG and NCBI databases for hydrocarbon degradation pathways and corresponding genes (Fig. 3). Both of these reference databases were screened against annotated genes from the seven draft genomes using their JGI-derived amino-acid sequences as input for blastp (e-value threshold of 1×10^{-20})[37].

Phylogenetic analyses. Phylosift was used to extract marker genes for the phylogenetic placement of the assembled metagenomic bins[38]. These marker genes consist of up to 15 syntenic ribosomal protein genes that have been demonstrated to undergo limited lateral gene transfer (rpL2, 3, 4, 5, 6, 14, 15, 18, 22, 24 and rpS3, 8, 10, 17, 19)[39]. This gene set was derived from a reference database as detailed in ref. 40. To search for ribosomal protein sequences, all metagenomic bins (fasta files) were used as an input in Phylosift, which was used with default parameters. Moreover, we searched NCBI to include amino-acid sequences from bacterial reference strains for phylogenetic analyses. Amino-acid alignments of the individual ribosomal protein genes were generated using MUSCLE and curated manually[41]. Afterwards, the curated alignments of the ribosomal proteins were concatenated for further phylogenetic analyses. In addition to analysing ribosomal proteins, we employed EMIRGE to retrieve 16S rRNA genes from the three SIP enrichments[42]. We therefore ran an EMIRGE analysis comparing the SILVA database against the raw, short reads from the three SIP enrichments using default parameters. Retrieved 16S rRNA genes were aligned to the SILVA database using the ARB alignment tool and were curated manually[43]. For both the ribosomal protein and 16S rRNA gene alignments, phylogenetic trees were generated using a maximum likelihood-based analysis (RAxML; rate distribution models: PROTGAMMA for ribosomal proteins and GTRGAMMA for 16S rRNA gene sequences)[44]. Bootstrap values were calculated using MrBayes with 100,000 generations of Markov chain Monte Carlo (MCMC) analyses with 100 sample and print frequencies[45]. As a point of comparison, we included published 16S rRNA gene sequences from the plume and sea surface as well as from the previous SIP experiment (16S rRNA genes from clone libraries and isolated bacterial strains) for the 16S rRNA phylogenetic tree[2,3,10,13,46].

Meta analysis. To determine whether genes from the assembled metagenomic bins in this study were enriched and active in the plume during the spill, we compared our three SIP enrichments with previously published plume-derived metagenomic and metatranscriptomic data sets (http://mason.eoas.fsu.edu/DWH_plume/)[10]. Two reference databases were generated, which were compared against the published metagenomic and metatranscriptomic data sets. Reference database 1 consisted of DNA sequences comprising scaffolds of >2,500 bp that were derived from the three SIP assemblies. Reference database 2 included amino-acid sequences from the seven high-quality draft genomes. Reference database 1 (DNA sequences) was mapped against the published metagenomic and metatranscriptomic data set using BWA. Database 2 (amino-acid sequences) was used to search against the published metatranscriptomic data using Rapsearch2 to retrieve the exact genes that mapped against this data set[47,48]. For BWA the default parameters were used and an e-value cutoff of 0.001 was used for RapSearch2.

Accession codes. The genomes are available in NCBI Genbank under BioProjectID PRJNA301966. The whole genome shotgun projects have been deposited under accession nos. LSMM00000000 (H-Mar), LSMN00000000 (N-Tha), LSMO00000000 (N-Alc), LSMP00000000 (N-Alt), LSMQ00000000 (P-Cyc), LSMR00000000 (P-Nep) and LSMS00000000 (P-Col). The reference numbers for the original raw sequencing data are SRX1562986 (*n*-hexadecane), SRX1585241 (naphthalene) and SRX1586894 (phenanthrene).

Received 11 December 2015; accepted 29 March 2016; published 9 May 2016

References

1. Crone, T. J. & Tolstoy, M. Magnitude of the 2010 Gulf of Mexico oil leak. *Science* **330,** 634–634 (2010).
2. Hazen, T. C. *et al.* Deep-sea oil plume enriches indigenous oil-degrading bacteria. *Science* **330,** 204–208 (2010).
3. Redmond, M. C. & Valentine, D. L. Natural gas and temperature structured a microbial community response to the Deepwater Horizon oil spill. *Proc. Natl Acad. Sci. USA* **109,** 20292–20297 (2012).
4. Yang, T. *et al.* Pulsed blooms and persistent oil-degrading bacterial populations in the water column during and after the Deepwater Horizon blowout. *Deep Sea Res. II* http://dx.doi.org/10.1016/j.dsr2.2014.01.014 (2014).
5. Crespo-Medina, M. *et al.* The rise and fall of methanotrophy following a deepwater oil-well blowout. *Nature Geosci.* **7,** 423–427 (2014).
6. Valentine, D. L. *et al.* Propane respiration jump-starts microbial response to a deep oil spill. *Science* **330,** 208–211 (2010).
7. Chauhan, A., Oakeshott, J. G. & Jain, R. K. Bacterial metabolism of polycyclic aromatic hydrocarbons: strategies for bioremediation. *Indian J. Microbiol.* **48,** 95–113 (2008).
8. Kimes, N. E., Callaghan, A. V., Suflita, J. M. & Morris, P. J. Microbial transformation of the Deepwater Horizon oil spill—past, present, and future perspectives. *Front. Microbiol.* **5,** 603 (2014).
9. Lea-Smith, D. J. *et al.* Contribution of cyanobacterial alkane production to the ocean hydrocarbon cycle. *Proc. Natl Acad. Sci. USA* **112,** 13591–13596 (2015).
10. Mason, O. U. *et al.* Metagenome, metatranscriptome and single-cell sequencing reveal microbial response to Deepwater Horizon oil spill. *ISME J.* **6,** 1715–1727 (2012).
11. Bælum, J. *et al.* Deep-sea bacteria enriched by oil and dispersant from the Deepwater Horizon spill. *Environ. Microbiol.* **14,** 2405–2416 (2012).
12. Kleindienst, S. *et al.* Diverse, rare microbial taxa responded to the Deepwater Horizon deep-sea hydrocarbon plume. *ISME J.* **10,** 400–415 (2015).
13. Gutierrez, T. *et al.* Hydrocarbon-degrading bacteria enriched by the Deepwater Horizon oil spill identified by cultivation and DNA-SIP. *ISME J.* **7,** 2091–2104 (2013).
14. Kodama, Y., Stiknowati, L. I., Ueki, A., Ueki, K. & Watanabe, K. *Thalassospira tepidiphila* sp. nov., a polycyclic aromatic hydrocarbon-degrading bacterium isolated from seawater. *Int. J. Syst. Evol. Microbiol.* **58,** 711–715 (2008).
15. Schneiker, S. *et al.* Genome sequence of the ubiquitous hydrocarbon-degrading marine bacterium *Alcanivorax borkumensis*. *Nature Biotechnol.* **24,** 997–1004 (2006).
16. Lai, Q. *et al.* *Alcanivorax pacificus* sp. nov., isolated from a deep-sea pyrene-degrading consortium. *Int. J. Syst. Evol. Microbiol.* **61,** 1370–1374 (2011).
17. Wang, B., Lai, Q., Cui, Z., Tan, T. & Shao, Z. A pyrene-degrading consortium from deep-sea sediment of the West Pacific and its key member *Cycloclasticus* sp. P1. *Environ. Microbiol.* **10,** 1948–1963 (2008).
18. Kleindienst, S. *et al.* Chemical dispersants can suppress the activity of natural oil-degrading microorganisms. *Proc. Natl Acad. Sci. USA* **112,** 14900–14905 (2015).
19. Cozzone, A. J. Role of protein phosphorylation on serine/threonine and tyrosine in the virulence of bacterial pathogens. *J. Mol. Microbiol. Biotechnol.* **9,** 198–213 (2005).
20. Grammann, K., Volke, A. & Kunte, H. J. New type of osmoregulated solute transporter identified in halophilic members of the bacteria domain: TRAP transporter TeaABC mediates uptake of ectoine and hydroxyectoine in *Halomonas elongata* DSM 2581T. *J. Bacteriol.* **184,** 3078–3085 (2002).
21. Edwards, B. R. *et al.* Rapid microbial respiration of oil from the Deepwater Horizon spill in offshore surface waters of the Gulf of Mexico. *Environ. Res. Lett.* **6,** 035301 (2011).
22. McKay, L. J., Gutierrez, T. & Teske, A. P. Development of a group-specific 16S rRNA-targeted probe set for the identification of *Marinobacter* by fluorescence in situ hybridization. *Deep Sea Res. II* http://dx.doi.org/10.1016/j.dsr2.2013.10.009 (2014).
23. Lamendella, R. *et al.* Assessment of the Deepwater Horizon oil spill impact on Gulf coast microbial communities. *Aquat. Microbiol.* **5,** 130 (2014).
24. Salomon, D. *et al.* Type VI secretion system toxins horizontally shared between marine bacteria. *PLoS Pathogens* **11,** e1005128 (2015).
25. Vaysse, P.-J. *et al.* Proteomic analysis of *Marinobacter hydrocarbonoclasticus* SP17 biofilm formation at the alkane–water interface reveals novel proteins and cellular processes involved in hexadecane assimilation. *Res. Microbiol.* **160,** 829–837 (2009).
26. Arnosti, C., Ziervogel, K., Yang, T. & Teske, A. Oil-derived marine aggregates—hot spots of polysaccharide degradation by specialized bacterial communities. *Deep Sea Res. II* http://dx.doi.org/10.1016/j.dsr2.2014.12.008 (2015).
27. Baker, B. J., Lesniewski, R. A. & Dick, G. J. Genome-enabled transcriptomics reveals archaeal populations that drive nitrification in a deep-sea hydrothermal plume. *ISME J.* **6,** 2269–2279 (2012).
28. Kujawinski, E. B. *et al.* Fate of dispersants associated with the Deepwater Horizon oil spill. *Environ. Sci. Technol.* **45,** 1298–1306 (2011).
29. Tillett, D. & Neilan, B. A. Xanthogenate nucleic acid isolation from cultured and environmental cyanobacteria. *J. Phycol.* **36,** 251–258 (2000).
30. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17,** 10–12 (2011).
31. Sickle v. 1.33 (Joshi, N. & Fass, J., 2011); https://github.com/najoshi/sickle
32. Peng, Y., Leung, H. C. M., Yiu, S. M. & Chin, F. Y. L. IDBA-UD: a *de novo* assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics* **28,** 1420–1428 (2012).
33. Dick, G. J. *et al.* Community-wide analysis of microbial genome sequence signatures. *Genome Biol.* **10,** R85 (2009).

34. Baker, B. J. Omic approaches in microbial ecology: charting the unknown. *Microbe* **8,** 353–360 (2013).
35. Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P. & Tyson, G. W. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res*. **25,** 1043–1055 (2015).
36. Overbeek, R. *et al.* The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). *Nucleic Acids Res.* **42,** D206–D214 (2014).
37. Altschul, S. F. *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res*. **25,** 3389–3402 (1997).
38. Darling, A. E. *et al.* PhyloSift: phylogenetic analysis of genomes and metagenomes. *PeerJ* **2,** e243 (2014).
39. Sorek, R. *et al.* Genome-wide experimental determination of barriers to horizontal gene transfer. *Science* **318,** 1449–1452 (2007).
40. Castelle, C. J. *et al.* Genomic expansion of domain archaea highlights roles for organisms from new phyla in anaerobic carbon cycling. *Curr. Biol.* **25,** 690–701 (2015).
41. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res*. **32,** 1792–1797 (2004).
42. Miller, C. S., Baker, B. J., Thomas, B. C., Singer, S. W. & Banfield, J. F. EMIRGE reconstruction of full-length ribosomal genes from microbial community short read sequencing data. *Genome Biol*. **12,** R44 (2011).
43. Ludwig, W. *et al.* ARB: a software environment for sequence data. *Nucleic Acids Res*. **32,** 1363–1371 (2004).
44. Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30,** 1312–1313 (2014).
45. Ronquist, F. & Huelsenbeck, J. P. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19,** 1572–1574 (2003).
46. Eren, A. M. *et al.* Anvi'o: an advanced analysis and visualization platform for 'omics data. *PeerJ* **3,** e1319 (2015).
47. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. Preprint at http://arXiv.org/1303.3997 (2013).
48. Zhao, Y., Tang, H. & Ye, Y. RAPSearch2: a fast and memory-efficient protein similarity search tool for next-generation sequencing data. *Bioinformatics* **28,** 125–126 (2012).

Acknowledgements
The metagenomic DNA originated from work that was supported by a Marie Curie International Outgoing Fellowship (PIOF-GA-2008-220129) to T.G. within the 7th European Community Framework Programme. Sampling in the Gulf of Mexico and SIP experiments underlying this study were made possible in part by a grant from The Gulf of Mexico Research Initiative and in part by a Marie Curie Fellowship to T.G. A.T. also acknowledges funding from the National Science Foundation (RAPID Response: the microbial response to the Deepwater Horizon Oil Spill; NSF-OCE 1045115). Data are publicly available through the Gulf of Mexico Research Initiative Information & Data Cooperative (GRIIDC) at https://data.gulfresearchinitiative.org (doi:10.7266/N7GH9FZ8). This is ECOGIG contribution 431.

Author contributions
N.D., T.G. and B.J.B. conceived this study. N.D. and B.J.B. supervised experiments and analyses. N.D., J.A.D., K.W.S. and B.J.B. performed analyses. N.D., T.G., A.P.T. and B.J.B. wrote the paper with contributions from all authors.

Additional information
Supplementary information is available online. Reprints and permissions information is available online at www.nature.com/reprints. Correspondence and requests for materials should be addressed to B.J.B.

Competing interests
The authors declare no competing financial interests.