# Microbial community and metagenome dynamics during biodegradation of dispersed oil reveals potential key-players in cold Norwegian seawater

Deni Ribicic[1*], Roman Netzer[2], Terry C. Hazen[3], Stephen M. Techtmann[4], Finn Drabløs[1], Odd Gunnar Brakstad[2]

[1]NTNU Norwegian University of Science and Technology, Department of Clinical and Molecular Medicine, Trondheim, Norway.

[2]SINTEF OCEAN, Trondheim, Norway.

[3]University of Tennessee Knoxville, Department of Civil and Environmental Engineering, Knoxville TN, USA.

[4]Michigan Technological University, Department of Biological Sciences, Houghton MI, USA.

*Address correspondence to Deni Ribicic, deni.ribicic@ntnu.no

Running Title: Microbial dynamics in oil amended microcosm

**Abstract**

Oil biodegradation as a weathering process has been extensively investigated over the years, especially after the Deepwater Horizon blowout. In this study, we performed microcosm experiments at 5°C with small droplet chemical dispersion in non-amended seawater. We link biodegradation processes with microbial community and metagenome dynamics and explain the succession based on substrate specialization. Reconstructed genomes and 16S rRNA gene analysis revealed that *Bermanella* and *Zhongshania* were the main contributors to initial n-alkane breakdown, while subsequent abundances of *Colwellia* and microorganisms closely related to *Porticoccaceae* were involved in secondary n-alkane breakdown and beta-oxidation. *Cycloclasticus, Porticoccaceae and Spongiiabcteraceae* were associated with degradation of mono- and poly- cyclic aromatics. Successional pattern of genes coding for hydrocarbon degrading enzymes at metagenome level, and reconstructed genomic content, revealed a high differentiation of biodegraders for a hydrocarbon substrate. A cooperation among oil degrading microorganisms is thus needed for the complete substrate transformation.

## 1 Introduction

The fate of oil released to the marine environment is largely dependent on environmental conditions and oil properties, being influenced by various processes like physical, chemical and biological weathering of the oil, spreading and dispersion in the water column, and even sedimentation (1, 2). However, biodegradation is the only process that completely mineralizes oil compounds. Oil biodegradation has been extensively studied over the years, and more frequently in a combination with oil dispersants after the Deep Water Horizon (DWH) oil spill incident. Significant microbial activity will start shortly after an oil spill, depending on the indigenous microbial community structure, oil characteristics and environmental conditions present at the time in the affected environment. Microbial concentrations close to the spill site may increase in numbers (3), and the community compositions will temporarily shift towards bacteria able to utilize the oil compounds (3-5). While microbes generally involved in hydrocarbon biodegradation are represented within many phylogenetic groups (6), biodegradation of

aromatic hydrocarbons in marine environments has been associated with genera like *Cycloclasticus*, *Pseudoalteromonas* and *Colwellia* (4, 7-10). Typical alkane-degrading bacteria include members of the genera *Alcanivorax*, *Oleiphilus*, *Oleispira* and *Thalassolituus* (9, 11, 12). Following the DWH spill novel molecular biology techniques were used, both to characterize the changes in microbial communities related to the deep water plume (mostly based on 16S rRNA gene analysis), and to identify essential microbial processes involved in oil biodegradation (3, 13-15). These data coupled to detailed chemical analysis of targeted oil compounds has a potential for elucidating the driving mechanisms involved in oil biodegradation. Many of the oil biodegradation studies, however, focus solely on microbial community (4, 16, 17) or purely on oil compound decay dynamics (18-20). Studies that combine both of the approaches, however, have not exploited the full potential of metagenome analysis (18, 21). Nevertheless, to study oil biodegradation in detail, experiments were designed using enrichment cultures as microbial source (13, 17). This approach is rather different from real oil spill conditions and can potentially result in incomplete conclusions, since pure cultures do not have the capacity to mimic interactions between numerous groups of microorganisms found in environment. In this study our aim is to bridge the gap between processes involved in microbiological degradation of chemically dispersed oil with chemistry by performing detailed analysis of both and mimicking close-to-real oil spill conditions. For elucidating the microbial community response, we employed 16S rRNA gene and metagenome shotgun analysis in combination with binning approach. GC-MS analysis we used for identifying degradation dynamics of targeted oil compounds. The studies were conducted in a system developed for studying biodegradation of dispersed oil (18, 22) and the source of microbial community is local seawater, rather than enrichment cultures.

## 2 Materials and methods

### 2.1 Oil, seawater and dispersant

Seawater (salinity 34 PSU) was collected from a depth of 80 m (below thermocline) in a Norwegian fjord (Trondheimsfjord; 63°26'N, 10°23'E), supplied by a pipeline system to our laboratories. The seawater was incubated at 5°C overnight before start of the experiments.

Dispersions with nominal median diameter of 10 µm droplets were prepared from premixed fresh paraffinic oil (Statfjord crude, batch 1998-0170), pre-mixed with the dispersant Slickgone NS (Dasic International Ltd., Romsey, Hampshire, UK) at dispersant to oil ratio (DOR) 1:100, as previously described (18, 22). Stock oil dispersions (200 mg/L) were diluted with seawater to a final concentration of 3 mg/L in 2-L pre-sterilized (autoclaved 120°C, 15 min) flasks (SCHOTT), based on Coulter Counter measurements (see below). Natural seawater with oil dispersions (NSOD) were generated in unfiltered non-amended seawater, while sterilized seawater with oil dispersions (hereinafter referred to as “chemical control”) were prepared in seawater filtered through 1 µm Nalgene™ Rapid-Flow™ filters (ThermoFisher Scientific, MA USA), autoclaved (120°C, 15 min) and poisoned with 100 mg/L (final concentration) $HgCl_2$. In addition, flasks of natural seawater without oil were included as biological controls (hereinafter referred to as "biological control"). The flasks were mounted on a carousel system with continuous slow rotation (0.75 r.p.m.) and incubated at 5°C for up to 64 days. Flasks with dispersions (NSOD and chemical controls) and biological controls were sacrificed for analyses after 0, 3, 6, 9, 13, 16, 32 and 64 days. At each sampling date flasks with NSOD (triplicate), chemical control (duplicate) and biological control (one replicate) were sampled. Each sample was analyzed for oil droplet size and semi-volatile and volatile oil compounds, while microbiological analyses (microbial enumerations, community characterization, and metagenome analyses) were performed on NSOD or control treatment from all samples. The experimental and analytical approach is described in FIG S1.

### 2.2 Microbiological analyses

Microbiological analyses included fluorescence microscopy for determination of total cell counts, and most probable number (MPN) analysis of oil-degrading microorganism (ODM). 16S rRNA gene amplicon sequencing was used for bacterial community analyses and shotgun sequencing (Whole Genome Sequencing) was employed for mapping the metabolic potential of microbial communities. Detailed information on microbiological analysis can be found in supplemental material.

### 2.2.1 Total and viable microbial cell counts

Cells were stained with 4',6-diamidino-2-phenylindole (DAPI) and counted using an epifluorescence microscope (1250 x magnification) (23). Most probable number (MPN) analysis of hydrocarbonoclastic prokaryotes was conducted in accordance with Rand *et al* and Brakstad and Lødeng (24, 25).

### 2.2.2 16S rRNA library and analysis

Detailed description of 16S rRNA gene workflow analysis can be found in supplemental material (S1). Briefly, raw pair-end reads were assembled with fastq-join in QIIME 1.9.1 (26). Assembled sequences were demultiplexed and quality filtered to remove low quality reads (Phred score <20; -q 19). UCHIME was employed for chimera detection on assembled quality filtered reads (27). Operational Taxonomic Units (OTUs) were determined by clustering assembled sequences on 97% nucleotide identity using UCLUST (28) with open reference clustering option. Representative sequences were aligned with PyNAST (29) and taxonomy assignment was performed with RDP classifier (30) based on SILVA-123 database (31). In order to visualize differences in taxonomical composition, relative abundances for OTUs on each sampling point were calculated. For the purpose of statistical analysis of OTUs, DESeq2 (32), an R package, was used to standardize the counts between samples rather than rarifying to the number of reads present in the sample with smallest number of reads.

### 2.2.3 Metagenome sequencing and analysis

Detailed analysis description can be found in supplemental material (S1). Briefly, five samples were used for metagenome exploration; two controls (day 0 and day 64) and three NSOD (day 9, 16 and 31). Illumina MiSeq paired raw reads were subject to quality filtering using Sickle (33) and assembled into contigs with MEGAHIT assembler (34). PROKKA pipeline was used to find and annotate genes using the default settings (35). Reads were mapped to contigs with Bowtie2 (36) and counting was performed with HTSeq (37). Counts were standardized based on “transcripts per million” (TPM) calculation (38). In our case, transcripts correspond to reads. Annotations containing enzyme commission number (EC number) were matched against list of gene ontology (GO) terms to produce file containing GO of hydrocarbon

degrading genes and their abundances. In order to determine the contribution of microorganisms to specific biodegradation processes, binning of metagenomic reads was performed. Reads from all five samples were co-assembled with MEGAHIT assembler with default parameters. Annotation and mapping was done as described previously. Resulting co-assembled file with contigs and BAM files from five samples were used as input for Anvi'o v2.2.2, binning and analysis tool (39). Bins were further manually curated to achieve desired completeness and redundancy. Additional quality check and taxonomical assignment was done with CheckM (40). We reconstructed phylogeny additionally using Phylosift (41) and FastTree (42) on bins that were taxonomically poorly resolved or not in consensus after Anvi'o and CheckM analysis.

#### 2.2.4 Nucleotide sequence data

Raw metagenome, 16S rRNA sequences and genome assemblies were deposited in the European Nucleotide archive (ENA) under the study accession number PRJEB14899 entitled as "Oil spill dispersant strategies and bioremediation efficiency". Raw metagenome sequences can be found from sample accession ID ERS1289858 to ERS1289862, while 16S rRNA sequences from samples accession ID ERS1265011 to ERS1265037. Under the sample accession ID ERS1867669 to ERS1867687, 19 genome assemblies were deposited.

### 2.3 Chemical analyses

#### 2.3.1 GC-MS analyses

Chemical analyses included GC-MS analyses of targeted oil compounds. NSOD and chemical control samples were solvent-solvent extracted with dichloromethane (DCM) for measurements of semi-volatile organic compounds (SVOC) by the gas chromatographic methods. The glass wall surfaces of the flasks were also rinsed with DCM after removal of dispersions to extract oil compounds attached to the glass walls. Samples for analyses of volatile compounds (VOCs) were transferred to glass vials (40 ml), acidified (pH < 2) with HCl before analyses by Purge&Trap GC-MS. Approximately 115 individual SVOC and VOC compounds were analyses by the GC-MS methods, including C5-C36 n-alkanes,

decalins, BTEX, phenols, naphthalenes, PAH and 17α(H),21β(H)-Hopane (30ab Hopane). Solvent extracts of SVOC compounds and acidified (pH < 2) dispersions for VOC compound quantification were analyzed as previously described (18).

The SVOC target compound concentrations were normalized against 17α(H),21β(H)-Hopane (30ab Hopane) (43) and depletion calculated as % compounds of concentrations in corresponding sterilized controls as follows:

$$\%\ depletion = 100\ \frac{(t_{c/Hop_c)nSW}}{(t_{c/Hop_c)sterSW}}$$, where

$t_c$ – target compound concentration; $Hop_c$ – Hopane concentration; nSW – normal seawater sample; sterSW – sterilized seawater sample.

The VOC compounds were not normalized against any internal standard, and % depletion of target compound concentrations in the regular samples was calculated as % of concentrations in sterilized samples of dispersions.

Multivariate statistics were performed in R studio v.3.2.1 using Vegan package v.2.3-0 (44) on generated Bray-Curtis distances.

## 3 Results and discussion

The focus of this study was to examine microbial community structure and metagenome dynamics during biodegradation of chemically dispersed oil in cold Norwegian seawater, and at low oil concentrations relevant for oil spills.

We have recently performed a study of microbial successions in relation to biodegradation of the Macondo oil originating from the DWH incident (45), but in the current study a Norwegian paraffinic oil was used in the biodegradation experiments, and more in-depth studies were performed on community structures and genes associated with degradation of targeted oil compounds.

### 3.1 Microbial community and hydrocarbon degradation dynamics

We examined microbial community structure in oil-contaminated microcosms by inspecting 16S rRNA gene profile in order to understand the community dynamics when exposed to chemically dispersed oil at low seawater temperature. 16S rRNA gene sequence analysis revealed dramatic changes in community structure in seawater- oil dispersion (NSOD) samples over the incubation period. *Gammaproteobacteria* (*Oceanospirillaceae*, *Porticoccaceae*, *Spongiibacteraceae* and *Piscirickettsiaceae*) was the predominant class over the entire experimental period, accounting for more than 90% relative abundance at certain sampling points (FIG 1C).

*Bacteroidetes* (*Flavobacteriaceae*) showed increased abundance as well (29% in relative sequence abundance at incubation day 64). Samples were found to be increasingly less diverse over time, in terms of richness and evenness, than communities in control samples at corresponding time points. From the start of the biodegradation experiment (day 0), where more than 2000 OTUs were determined, the diversity decreased to approximately 250 OTUs in some of the replicates on incubation day 13, 16 and 31 (FIG S2A and S2B). Similar patterns were observed for *Shannon* diversity index where the lowest values were reported for day 9 and 13. *Simpson* diversity index exhibited the lowest values for day 9, while the second lowest was represented by one replicate from incubation day 6 (FIG S2A). In oil-free control samples, *Gammaproteobacteria* was also the most abundant class, however with less than 40% in relative sequence abundance (Table S2). A similar pattern of community change during oil incubation with dispersed oil was recently reported, except that the starting community was predominant with *Alphaproteobacteria* instead of *Gammaproteobactera* (45). This may be due to seasonal variation in community composition of local seawater when the experiments have been conducted (winter 2011 and summer 2013 season). The strong community shift towards certain oil degraders has been observed elsewhere (3, 4, 17, 21, 23, 24, 46).

Semi-volatile n-alkanes were primarily degraded between days 6 and 31, and single compound analysis revealed that degradation was correlated to HC chain length (FIG 2A). The n-alkane degradation period

correlated with high relative sequence abundances of *Oceanospirillaceae* (up to 55% at day 9) and *Colwelliaceae* (up to 52% at day 13) (FIG 1A). During the same period (days 6-16), increased concentrations of total microbes (DAPI) and viable oil-degrading microbes (ODM) were determined, peaking at day 16 (FIG 1A), and therefore matching the high abundances of *Oceanospirillaceae* and *Colwelliaceae*. Oil-degrading microbes increased in abundance by three orders of magnitude (from $3x10^1$ on day 0, to $3x10^3$ on day 6). While nC10 – nC20 alkanes were depleted by ≥ 50% within 2 weeks, nC30 – nC36 alkanes were not significantly depleted before 31 days (FIG 2A). Similar dynamics were observed during the DWH oil spill, where *Oceanospirillaceae* was the main contributor of aliphatics degradation (15). On the other hand, *Spongiibacteraceae*, as a novel established family (47), was not associated with the DWH oil spill. However, a *Spongiibacteraceae* genus, *Zhongshania*, have been reported to degrade aliphatic hydrocarbons (48). Semi-volatile aromatic hydrocarbons (PAH) generally showed slower degradation than the n-alkanes, mainly between days 13 and 64 (FIG 1A). This corresponded to increased abundances of *Porticoccaceae* (up to 28% on day 31), *Piscirickettsiaceae* (up to 12% on day 31) and *Flavobacteriaceae* (up to 29% on day 64) (FIG 1A). ODM concentrations were also high during the period with high abundances of bacterial families associated with PAH degradation. Biodegradation rates of PAHs were related to alkyl substitution level, with increasing alkyl substitution resulting in slower biotransformation (FIG 2B). Even after 64 days, with more than 95% of the PAH depleted, this trend was still observed. In the DWH oil spill, biodegradation of PAH was associated with high abundances of *Piscirickettsiaceae* (mainly the genus *Cycloclasticus*), *Alteromonadaceae*, *Flavobacteriaceae* and *Rhodobacteraceae* (4, 17). Again, *Porticoccaceae* as a novel established family within a novel order of *Cellvibrionales* (separated from *Alteromonadaceae*)(47) was not associated with degradation of hydrocarbon compounds during DWH. Nevertheless, a recent genome report presents *Porticoccus hydrocarbonoclasticus* as an obligate hydrocarbonoclastic marine bacteria (49).

Multivariate statistics of the microbial community and chemical composition of measured oil compounds reveals similarity in pattern evolution during the experimental period (FIG 1B and FIG 1D). Successions

of microbial communities were manifested in a counterclockwise directed PCoA plot based on weighted-UniFrac distance metric (explaining 61% and 18.8% differences on Axis.1 and Axis.2, respectively), showing structural differences between control samples and NSOD over the 64 days experimental period. Replicates from the same time-point were clustering together and apart from the replicates from different time-points, showing that those samples shared unique qualitative and quantitative phylogenetic features, suggesting metagenome changes. Unique phylogenetic and functional characteristics of different sample types were observed during the DWH oil spill as well. Clustering based on community composition was distinct between pre- and spill- samples (4). Also samples characterized as pristine or contaminated were well defined based on functional potential and community composition (3, 17, 50, 51). Similar to the microbial community development in a two-dimensional space, composition of the targeted oil compounds exhibited distinguishable patterns, with a clear development from day 3 to day 64 in the clockwise direction. This is the product of unique susceptibility of targeted oil compounds to biodegradation at different rates due substrate complexity (FIG 2). Development of the microbial community occurred on a finer scale than the development of the targeted chemical composition, comparing incubation days 3 – 16 in regard to the most explanatory axis (Axis.1 61%), revealing high sensitivity of the community to small changes in measured oil composition (day 3- day 16).

### 3.2 Metagenome succession

Two control samples (day 0 and 64) and three NSOD samples (day 9, 16 and 31) were subjected to full metagenome analysis in order to elucidate dynamics of genes coding for hydrocarbonoclastic enzymes. The count data were standardized using “transcripts per million” (TPM) estimates (38) and gene composition was followed at metagenome level, along experiment timeline, rather than at different taxonomic levels. Our findings suggest a cascade-wise change in abundance of gene ontology (GO) terms related to aliphatics and aromatics degradation. Namely, aliphatics degradation was observed to shift from NSOD-9, where alkane 1-monooxygenase genes peaked in abundance (486 TPM counts) to alkanal monooxygenase exhibiting highest values in NSOD-31 (1413 TPM counts) (FIG 3B).

Across previously mentioned GOs, we observed the rest of alkane hydroxylase system which includes rubredoxin/ferredoxin reductases (peaked at NSOD-16), alcohol and aldehyde dehydrogenases. The latter two exhibited higher abundances in all three NSOD samples. The mechanism for alkane degradation is well known and starts by terminal (in some cases sub-terminal) activation of hydrocarbons with addition of molecular oxygen to the chain and creating corresponding alcohol, subsequently aldehyde and carboxylic acid. This followed by conjugation to CoA and finally transformation to acetyl-CoA to allow beta-oxidation (52, 53). C5-C10 alkane were already degraded >70% by day 9 (NSOD-9) (Fig. 2C), whereas C10-C20 were degraded by >20% in the same period. This corresponded to high abundance of alkane 1-monooxygenase, which exhibited a decreasing pattern afterwards. Since, we were not able to observe GO patterns before sample NSOD-9 and based on already highly degraded substrate, we can only speculate that the abundance of alkane 1-monooxygenase showed even higher values in period prior to NSOD-9. Aliphatics degradation genes were followed by alkane beta-oxidation genes, distributed evenly, but slightly peaked in the last metagenome sample (NSOD-31). The most prominent one was the medium/very long-chain-acyl-CoA dehydrogenase, which suggested increased transformation potential of degradation byproducts of medium to long chain alkanes. Aromatics degradation genes started to increase in abundance in sample NSOD-16 and peaked in the last metagenome sample NSOD-31. Degradation pattern of PAH compounds coincided to change in profile of genes coding for aromatics degradation (Fig. 1A and 2B). Most abundant genes related to aromatics degradation proved to be genes coding for enzymes involved in phenol (phenol 2-mononxygenase, 329 TPM counts), keton (phenylacetone monooxygenase, 433 TPM counts), benzene (biphenyl 2, 3-dioxygenase, 687 TPM counts), benzoate (4-hydroxybenzoate octaprenyltransferase, evenly represented in all three NSOD samples) and naphthalene degradation (naphthalene 1,2-dioxygenase, 206 TPM counts).

Observed cascade-wise change in abundance of genes along experimental timeline, or metagenome succession, has analogy to microbial community succession in an oil contaminated environment. The concept itself is rather basic; genes encoding enzymes relevant for degradation of simple hydrocarbons are

successively substituted with genes encoding enzymes potentially involved in degradation of more recalcitrant hydrocarbon substrate. Accordingly, in our experiment, genes coding for enzymes involved in initial aliphatics degradation were first to increase in abundance (NSOD-9), followed by genes responsible for beta oxidation (present evenly with slight peak in NSOD 31) and ending with increased abundance of genes coding for enzymes responsible for degradation of aromatic compounds (NSOD-16 and NSOD-31). As proposed by the metagenome succession concept, the shift in gene abundances followed degradation pattern of targeted hydrocarbons (FIG 1A and 2B). In addition, gene succession was observed for the whole metagenome which was recaptured by multivariate PCA plot (FIG S4). The phenomenon of metagenome succession could be observed elsewhere (50, 54, 55), and it is likely to have a common pattern on a global scale. This can be supported by the fact that similar biodegradation and real case studies containing seawater from different parts of the world (4, 15, 17, 21, 46), or seawater from the same location, but different seasons and therefore different ambient community (45), converge to a community having similar key players and functional potential when exposed to hydrocarbons.

### 3.3 Key-players contribution to hydrocarbon degradation gene abundances

For the purposes of resolving biodegradation potential of different taxa, genome bins were reconstructed and annotated from co-assembled metagenomes originating from five samples that were subjected to full metagenome sequencing (see previous section). Bins that originally were of satisfying quality (>45% completeness, <10% redundancy) are denoted in text as “Bin” followed by index number. Otherwise, bins that needed refinement are denoted as “Refined” followed by index number. Details about analysis can be found in supplementary 1. Reconstruction of genomes revealed that genus *Bermanella* was contributing the most to initial n-alkane degradation pathway followed by other *Oceanospirillaceae* related genera *Oleispira/Oceanobacter*. The annotations were differing between genera of *Oleispira* and *Oceanobacter* depending on whether CheckM (metagenomics tool for bin quality assessment and annotation) was used for taxonomy assignment or the Anvi’o internal taxonomy tool based on Centrifuge metagenome classifier. Furthermore, *Zhongshania* a *Spongiibacteraceae* genus, contributed to initial n-alkane

degradation as well. *Oceanospirillaceae* (Bin_8, Bin_23, Refined_1 and Refined_2) and *Spongiibacteraceae* (Bin_32) exhibited higher abundances of genes responsible for initial oxidation of n-alkanes (alkane 1-monooxygenase, rubredoxin NAD+ reductase and ferredoxin NADP+ reductase) for sampling days 9 and 16 (FIG 3B), where the respective bins displayed to be most abundant (FIG 4). *Cycloclasticus* assigned bins showed to contain initial alkane oxidizing genes as well. However, they were rather low in abundance. *Porticoccaceae* (Refined_4 and Refined_5) contained higher number of secondary alkane degrading genes (alkanesulfonate monooxygenase, alcohol and aldehyde dehydrogenase and alkanal monooxygenase), as well as *Colwellia* and *Zhongshania* (Bin_18, Bin_21 and Bin_32). While *Porticoccaceae* exhibited highest abundance at day 31, *Colwellia* and *Zhongshania* dominated day 9 and 16, in addition to day 31(FIG 4).

Beta-oxidation genes were found to be abundant on day 9, 16 and peaked on day 31 (FIG 3A). Most contributing bins to genes associated with beta-oxidation were found to be *Bermanella* (Bin 8) and *Porticoccaceae* related bins (Refined_4 and_5) on day 31 (FIG 3B). *Colwellia* (Bin_21, Bin_18 and Bin_36) showed to be most abundant on incubation days 9 and 16 containing respective genes (FIG 4). *Spongiibacteraceae* (Bin_32 and Bin_15) displayed abundances of beta-oxidation genes for incubation days 9 and 16, where the genus was the most abundant (FIG 3B and FIG 4). This corresponds well to 16S rRNA gene profile, where *Spongiibacteraceae* peaked at day 16 (2% in relative sequence abundance) (FIG 1C). *Cycloclasticus* (Refined_3, Refined_6 and Bin_27) and *Flavobacteriaceae* (Bin_22 and Bin_25) show as well potential for beta-oxidation.

*Cycloclasticus* (Bin_27, Refined_3 and Refined_6) displayed plethora of genes coding for aromatics degradation enzymes for day 16 and 31 (FIG 3B). Most abundant genes were associated with degradation of phenol, cresol, catechol, benzene, benzoate and biphenyl. In addition, *Porticoccaceae* (Refined_4 and_5) contributed to abundance of phenylacetone, hydroxyacetonephenone, biphenyl, naphthalene and benzoate degrading genes on incubation day 31 (FIG 3B). Ubiquitous *Zhongshania* genome (Bin_32) displayed as well potential in aromatics degradation (catechol, biphenyl, benzoate, cyclopentanol). One

313 *Colwellia* genome (Bin_18) showed potential to be involved in degradation of phenol, cresol, anthranilate,
314 benzoate and cyclopentanol. Reconstructed bins assigned to *Flavobacteriaceae* (Bin 22 and 25) and
315 *Rhodobacteraceae* (Bin 10) displayed abundance on incubation days 16 and 31, but also for controls
316 (*Flavobacteriaceae*) (FIG 4). However, respective bins contributed to a lesser extent to the abundance of
317 genes coding for hydrocarbon degrading enzymes (FIG 3B).

318 Two recent studies (21, 45) have conducted similar microcosm experiments with low oil dispersion
319 concentration at defined droplet size and have analyzed chemical (GC-MS of targeted compounds) and
320 microbiological composition (whole metagenome sequencing- WMS) in an attempt to correlate
321 microbiological succession to targeted compound degradation. Although WMS was conducted, only
322 taxonomical analyses were carried out, hindering the full potential of metagenomic dataset. Conclusions
323 based solely on taxonomy correlation to chemistry are rather incomplete. For instance, *Colwellia* (from
324 the very same location as in current study) was attributed as one of the main n-alkane degraders (45).
325 However, in current study *Colwellia* genome assembly, by a closer inspection of gene content, did not
326 contain initial alkane degradation gene at all (i.e. alkane 1-monooxygenase), rather had a potential to
327 consume downstream compounds following n-alkane breakdown and beta-oxidation. Furthermore, unlike
328 *Colwellia* from DWH study (10), reconstructed *Colwellia* genomes in current experiment did not contain
329 high abundance of genes coding for enzymes involved in aromatics degradation (except one species-
330 Bin_18). One way to argue this finding is that *Colwellia* found in the Gulf of Mexico (GoM) could have
331 an evolutionary advantage of being adapted to oil which is occurring via natural oil seeps (56).
332 *Bermanella* genus reported in the study by Brakstad et al., and associated with alkane degradation (45),
333 was also found to be the main n-alkane consumer in the current study. Similar observations were made by
334 Hu et al. (57) in a recent study which adopted a system for studying biodegradation, that was developed at
335 SINTEF (18, 22) and used also in the present study. Similarities between respective study and current one
336 in microbial community, metagenome and targeted oil compounds succession is apparent although

different oil, seawater and dispersant was used (Macondo light crude, GoM seawater and Corexit EC9500A, respectively).

Hu et al. (57) have observed substrate specialization based on genomic content and species abundance compared to chemical analysis of targeted compounds. Correspondingly, in present study, each bin contained a specific set of genes coding for enzymes responsible for hydrocarbon degradation that other bins lack, filling a gap in degradation pattern. E.g., *Porticoccaceae* (Refined_4 and Refined_5) contains phenylacetone monooxygenase, 4-hydroxyacetophenon monooxygenase, cyclopentanol dehydrogenase, coniferyl-aldehyde dehydrogenase and 4-hydroxyphenylacetate 3-monooxygenase, perfectly filling the gap in *Cycloclasticus* (Refined_3, Refined_6 and Bin_36) array of aromatics degradation genes (Fig 3B). Same can be observed for the genes encoding initial degradation of n-alkanes, where all of *Oceanospirillaceae* bins and *Zhongshania* bin contain alkane 1-monooxygenase, rubredoxin-$NAD^+$ reducates and ferredoxin-$NADP^+$ reductase, while the other bins are mostly lacking respective genes (Fig 3B). On the other hand, *Oceanospirillaceae* and *Zhonghania* seem to be lacking most of the secondary alkane degradation enzymes (alkanesulfonate monooxygenase, aldehyde dehydrogenase ($NADP^+$) and alkanal monooxygenase). However, *Porticoccaceae* and *Colwellia* have the capacity to deputy (Fig 3B). Similar pattern is observed for beta oxidation genes, where *Colwellia*, *Porticoccaceae*, *Cycloclasticus* and *Flavobacteriaceae* substitute for absence in *Oceanospirillaceae* and *Spongiibacteraceae* genome (Fig 3B). This implies that there is a cooperation necessity among microorganisms for absolute hydrocarbon transformation. We are not referring here, however, only to cooperation in a sense of consuming different types of crude oil hydrocarbons; it is instead a cross-feeding of partially oxidized or dead-end products. de Lorenzo (58, 59) proposed an idea known as the ecotopic concept, which is based on epi-metabolome formed by a pool of compounds that diffuse, or are being actively secreted, out of the cells between two steps of a metabolic pathway. This means that microorganisms have the ability to share intermediate products for the sake of efficient hydrocarbon mineralization. Disentangling of the ecotopic metabolism, however, requires a more pragmatic approach using metatranscriptomics and metabolomics, which can

provide a detailed insight into activity status of specific community members and the metabolites produced. Information based solely on (meta)genome content should be carefully interpreted, as it is not revealing actual gene activity nor metabolite production status, but rather a potential for it. Metagenomics, nevertheless, may deliver a glimpse into the catabolic capability of the biodegrader community.

## 4 Conclusions

This study provides an attempt to reconstruct metabolic pathways for hydrocarbon degradation from metagenomes obtained solely from the incubation studies under close-to-realistic oil spill conditions (oil dispersion concentration, oil dispersion droplet size and the source of microbial community) and to explain community succession based on hydrocarbon specialization.

Metagenome and ribosomal gene (16S rRNA) screening revealed succession pattern in microbial community as well as in metagenome composition. Relatively complete genomes (>80%) for main biodegraders were reconstructed from metagenomic datasets. The obtained information revealed a narrow niches specificity for hydrocarbon substrate. *Oceanospirillaceae* (mainly *Bermanella*) and *Spongiibacteraceae* (*Zhongshania*) contained primarily genes coding for short/medium alkane degradation (i.e. alkane 1-monooxygenase). *Colwelliaceae* (*Colwellia*) and *Porticoccaceae* seemed to specialize in consumption of initial alkane degradation byproducts, while *Pisciriketsiaceae* (*Cycloclasticus*), *Rhodobacteraceae, Porticoccaceae* and ubiquitous *Spongiibaceteraceae* contained genes encoding for aromatics transformation. At the metagenome level, we observed a successional pattern of genes coding for hydrocarbon degrading enzymes, which together with reconstructed genomic content revealed a high specificity and differentiation for hydrocarbon substrate, accentuating a need for cooperation among oil biodegraders for a successful substrate transformation. Among the main biodegraders we found novel-degrading microorganisms. *Spongiibacteraceae* and *Porticoccaceae* were not detected in any of the DWH studies to our knowledge. Based on our conclusions, this study provides a novel insight into microbial community potential for oil biodegradation in Norwegian seawater.

## 5 Funding information

This study is a part of a project which is funded by The Research Council of Norway (project # 22827/E30: Oil Spill Dispersant Strategies and Biodegradation Efficiency) and the oil companies Statoil ASA, AkerBP ASA, ExxonMobil Production Norway Inc., Total E&P and ConocoPhillips Skandinavia AS.

## 6 Acknowledgments

We would like to thank the staff at the SINTEF OCEAN SeaLab for performing the chemical analysis (Marianne Rønsberg, Kjersti Amås and Inger Steinsvik).

# 7 References


1. **Brakstad OG, Bonaunet K, Nordtug T, Johansen Ø.** 2004. Biotransformation and Dissolution of Petroleum Hydrocarbons in Natural Flowing Seawater at Low Temperature. Biodegradation **15:**337-346.
2. **Haritash AK, Kaushik CP.** 2009. Biodegradation aspects of Polycyclic Aromatic Hydrocarbons (PAHs): A review. Journal of Hazardous Materials **169:**1-15.
3. **Hazen TC, Dubinsky EA, DeSantis TZ, Andersen GL, Piceno YM, Singh N, Jansson JK, Probst A, Borglin SE, Fortney JL, Stringfellow WT, Bill M, Conrad ME, Tom LM, Chavarria KL, Alusi TR, Lamendella R, Joyner DC, Spier C, Baelum J, Auer M, Zemla ML, Chakraborty R, Sonnenthal EL, D'haeseleer P, Holman H-YN, Osman S, Lu Z, Van Nostrand JD, Deng Y, Zhou J, Mason OU.** 2010. Deep-Sea Oil Plume Enriches Indigenous Oil-Degrading Bacteria. Science **330:**204-208.
4. **Dubinsky EA, Conrad ME, Chakraborty R, Bill M, Borglin SE, Hollibaugh JT, Mason OU, M. Piceno Y, Reid FC, Stringfellow WT, Tom LM, Hazen TC, Andersen GL.** 2013. Succession of Hydrocarbon-Degrading Bacteria in the Aftermath of the Deepwater Horizon Oil Spill in the Gulf of Mexico. Environmental Science & Technology **47:**10860-10867.
5. **King GM, Kostka JE, Hazen TC, Sobecky PA.** 2015. Microbial Responses to the Deepwater Horizon Oil Spill: From Coastal Wetlands to the Deep Sea. Annual Review of Marine Science **7:**377-401.
6. **Prince RC.** 2005. The Microbiology of Marine Oil Spill Bioremediation, p 317-335, Petroleum Microbiology doi:doi:10.1128/9781555817589.ch16. American Society of Microbiology.
7. **Dyksterhouse SE, Gray JP, Herwig RP, Lara JC, Staley JT.** 1995. Cycloclasticus pugetii gen. nov., sp. nov., an Aromatic Hydrocarbon-Degrading Bacterium from Marine Sediments. International Journal of Systematic and Evolutionary Microbiology **45:**116-123.
8. **Geiselbrecht AD, Hedlund BP, Tichi MA, Staley JT.** 1998. Isolation of Marine Polycyclic Aromatic Hydrocarbon (PAH)-Degrading Cycloclasticus Strains from the Gulf of Mexico and Comparison of Their PAH Degradation Ability with That of Puget Sound Cycloclasticus Strains. Applied and Environmental Microbiology **64:**4703-4710.
9. **Harayama S, Kasai Y, Hara A.** 2004. Microbial communities in oil-contaminated seawater. Current Opinion in Biotechnology **15:**205-214.
10. **Mason OU, Han J, Woyke T, Jansson JK.** 2014. Single-cell genomics reveals features of a Colwellia species that was dominant during the Deepwater Horizon oil spill. Frontiers in Microbiology **5:**332.

11. **Hara A, Syutsubo K, Harayama S.** 2003. Alcanivorax which prevails in oil-contaminated seawater exhibits broad substrate specificity for alkane degradation. Environmental Microbiology **5:**746-753.

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

13. **Bælum J, Borglin S, Chakraborty R, Fortney JL, Lamendella R, Mason OU, Auer M, Zemla M, Bill M, Conrad ME, Malfatti SA, Tringe SG, Holman H-Y, Hazen TC, Jansson JK.** 2012. Deep-sea bacteria enriched by oil and dispersant from the Deepwater Horizon spill. Environmental Microbiology **14:**2405-2416.

14. **Lu Z, Deng Y, Van Nostrand JD, He Z, Voordeckers J, Zhou A, Lee Y-J, Mason OU, Dubinsky EA, Chavarria KL, Tom LM, Fortney JL, Lamendella R, Jansson JK, D'Haeseleer P, Hazen TC, Zhou J.** 2012. Microbial gene functions enriched in the Deepwater Horizon deep-sea oil plume. ISME J **6:**451-460.

15. **Mason OU, Hazen TC, Borglin S, Chain PSG, Dubinsky EA, Fortney JL, Han J, Holman H-YN, Hultman J, Lamendella R, Mackelprang R, Malfatti S, Tom LM, Tringe SG, Woyke T, Zhou J, Rubin EM, Jansson JK.** 2012. Metagenome, metatranscriptome and single-cell sequencing reveal microbial response to Deepwater Horizon oil spill. ISME J **6:**1715-1727.

16. **Valentine DL, Kessler JD, Redmond MC, Mendes SD, Heintz MB, Farwell C, Hu L, Kinnaman FS, Yvon-Lewis S, Du M, Chan EW, Tigreros FG, Villanueva CJ.** 2010. Propane Respiration Jump-Starts Microbial Response to a Deep Oil Spill. Science **330:**208-211.

17. **Kostka JE, Prakash O, Overholt WA, Green SJ, Freyer G, Canion A, Delgardio J, Norton N, Hazen TC, Huettel M.** 2011. Hydrocarbon-Degrading Bacteria and the Bacterial Community Response in Gulf of Mexico Beach Sands Impacted by the Deepwater Horizon Oil Spill. Applied and Environmental Microbiology **77:**7962-7974.

18. **Brakstad OG, Nordtug T, Throne-Holst M.** 2015. Biodegradation of dispersed Macondo oil in seawater at low temperature and different oil droplet sizes. Marine Pollution Bulletin **93:**144-152.

19. **McFarlin KM, Prince RC, Perkins R, Leigh MB.** 2014. Biodegradation of Dispersed Oil in Arctic Seawater at -1°C. PLoS ONE **9:**e84297.

20. **Valentine DL, Fisher GB, Bagby SC, Nelson RK, Reddy CM, Sylva SP, Woo MA.** 2014. Fallout plume of submerged oil from Deepwater Horizon. Proceedings of the National Academy of Sciences **111:**15906-15911.

21. **Wang J, Sandoval K, Ding Y, Stoeckel D, Minard-Smith A, Andersen G, Dubinsky EA, Atlas R, Gardinali P.** 2016. Biodegradation of dispersed Macondo crude oil by indigenous Gulf of Mexico microbial communities. Science of The Total Environment **557–558:**453-468.

22. **Nordtug T, Olsen AJ, Altin D, Meier S, Overrein I, Hansen BH, Johansen Ø.** 2011. Method for generating parameterized ecotoxicity data of dispersed oil for use in environmental modelling. Marine Pollution Bulletin **62:**2106-2113.

23. **Brakstad OG, Nonstad I, Faksness L-G, Brandvik PJ.** 2007. Responses of Microbial Communities in Arctic Sea Ice After Contamination by Crude Petroleum Oil. Microbial Ecology **55:**540-552.

24. **Brakstad OG, Lødeng AGG.** 2005. Microbial Diversity during Biodegradation of Crude Oil in Seawater from the North Sea. Microbial Ecology **49:**94-103.

25. **Rand MC, Greenberg AE, Taras MJ.** 1976. Standard methods for the examination of water and wastewater. 14th edition. Prepared and published jointly by American Public Health Association, American Water Works Association, and Water Pollution Control Federation.

26. **Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, Fierer N, Pena AG, Goodrich JK, Gordon JI, Huttley GA, Kelley ST, Knights D, Koenig JE, Ley RE, Lozupone CA, McDonald D, Muegge BD, Pirrung M, Reeder J, Sevinsky JR, Turnbaugh PJ, Walters WA, Widmann J, Yatsunenko T, Zaneveld J, Knight R.** 2010. QIIME allows analysis of high-throughput community sequencing data. Nat Meth **7:**335-336.

27. **Edgar RC, Haas BJ, Clemente JC, Quince C, Knight R.** 2011. UCHIME improves sensitivity and speed of chimera detection. Bioinformatics **27:**2194-2200.

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

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

30. **Wang Q, Garrity GM, Tiedje JM, Cole JR.** 2007. Naïve Bayesian Classifier for Rapid Assignment of rRNA Sequences into the New Bacterial Taxonomy. Applied and Environmental Microbiology **73:**5261-5267.

31. **Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glöckner FO.** 2013. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Research **41:**D590-D596.

32. **Love MI, Huber W, Anders S.** 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biology **15:**1-21.

33. **Joshi N, Fass J.** 2011. Sickle: a sliding-window, adaptive, quality-based trimming tool for FastQ files. Available from: github com/najoshi/sickle.

34. **Li D, Liu C-M, Luo R, Sadakane K, Lam T-W.** 2015. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. Bioinformatics **31:**1674-1676.

35. **Seemann T.** 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics **30:**2068-2069.

36. **Langmead B, Salzberg SL.** 2012. Fast gapped-read alignment with Bowtie 2. Nat Meth **9:**357-359.

37. **Anders S, Pyl PT, Huber W.** 2015. HTSeq—a Python framework to work with high-throughput sequencing data. Bioinformatics **31:**166-169.

38. **Wagner GP, Kin K, Lynch VJ.** 2012. Measurement of mRNA abundance using RNA-seq data: RPKM measure is inconsistent among samples. Theory in Biosciences **131:**281-285.

39. **Eren AM, Esen ÖC, Quince C, Vineis JH, Morrison HG, Sogin ML, Delmont TO.** 2015. Anvi'o: an advanced analysis and visualization platform for 'omics data. PeerJ **3:**e1319.

40. **Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW.** 2015. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Research **25:**1043-1055.

41. **Darling AE, Jospin G, Lowe E, Matsen FAIV, Bik HM, Eisen JA.** 2014. PhyloSift: phylogenetic analysis of genomes and metagenomes. PeerJ **2:**e243.

42. **Price MN, Dehal PS, Arkin AP.** 2009. FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. Mol Biol Evol **26:**1641-1650.

43. **Prince RC, Elmendorf DL, Lute JR, Hsu CS, Haith CE, Senius JD, Dechert GJ, Douglas GS, Butler EL.** 1994. 17.alpha.(H)-21.beta.(H)-hopane as a conserved internal marker for estimating the biodegradation of crude oil. Environmental Science & Technology **28:**142-145.

44. **Dixon P, Palmer MW.** 2003. VEGAN, a package of R functions for community ecology. Journal of Vegetation Science **14:**927-930.

45. **Brakstad OG, Throne-Holst M, Netzer R, Stoeckel DM, Atlas RM.** 2015. Microbial communities related to biodegradation of dispersed Macondo oil at low seawater temperature with Norwegian coastal seawater. Microbial Biotechnology **8:**989-998.

46. **Kleindienst S, Seidel M, Ziervogel K, Grim S, Loftis K, Harrison S, Malkin SY, Perkins MJ, Field J, Sogin ML, Dittmar T, Passow U, Medeiros PM, Joye SB.** 2015. Chemical dispersants can suppress the activity of natural oil-degrading microorganisms. Proceedings of the National Academy of Sciences **112:**14900-14905.

47. **Spring S, Scheuner C, Göker M, Klenk H-P.** 2015. A taxonomic framework for emerging groups of ecologically important marine gammaproteobacteria based on the reconstruction of evolutionary relationships using genome-scale data. Frontiers in Microbiology **6**.

48. **Lo N, Kang HJ, Jeon CO.** 2014. Zhongshania aliphaticivorans sp. nov., an aliphatic hydrocarbon-degrading bacterium isolated from marine sediment, and transfer of Spongiibacter borealis Jang et al. 2011 to the genus Zhongshania as Zhongshania borealis comb. nov. Int J Syst Evol Microbiol **64:**3768-3774.

49. **Gutierrez T, Whitman WB, Huntemann M, Copeland A, Chen A, Kyrpides N, Markowitz V, Pillay M, Ivanova N, Mikhailova N, Ovchinnikova G, Andersen E, Pati A, Stamatis D, Reddy TBK, Ngan CY, Chovatia M, Daum C, Shapiro N, Cantor MN, Woyke T.** 2015. Genome Sequence of Porticoccus hydrocarbonoclasticus Strain MCTG13d, an Obligate Polycyclic Aromatic Hydrocarbon-Degrading Bacterium Associated with Marine Eukaryotic Phytoplankton. Genome Announcements **3:**e00672-00615.

50. **Rodriguez-R LM, Overholt WA, Hagan C, Huettel M, Kostka JE, Konstantinidis KT.** 2015. Microbial community successional patterns in beach sands impacted by the Deepwater Horizon oil spill. ISME J doi:10.1038/ismej.2015.5.

51. **Rivers AR, Sharma S, Tringe SG, Martin J, Joye SB, Moran MA.** 2013. Transcriptional response of bathypelagic marine bacterioplankton to the Deepwater Horizon oil spill. ISME J **7:**2315-2329.

52. **Rojo F.** 2009. Degradation of alkanes by bacteria. Environmental Microbiology **11:**2477-2490.

53. **Grund A, Shapiro J, Fennewald M, Bacha P, Leahy J, Markbreiter K, Nieder M, Toepfer M.** 1975. Regulation of alkane oxidation in Pseudomonas putida. Journal of Bacteriology **123:**546-556.

54. **Dombrowski N, Donaho JA, Gutierrez T, Seitz KW, Teske AP, Baker BJ.** 2016. Reconstructing metabolic pathways of hydrocarbon-degrading bacteria from the Deepwater Horizon oil spill. Nat Microbiol **1:**16057.

55. **Mason OU, Scott NM, Gonzalez A, Robbins-Pianka A, Balum J, Kimbrel J, Bouskill NJ, Prestat E, Borglin S, Joyner DC, Fortney JL, Jurelevicius D, Stringfellow WT, Alvarez-Cohen L, Hazen TC, Knight R, Gilbert JA, Jansson JK.** 2014. Metagenomics reveals sediment microbial community response to Deepwater Horizon oil spill. ISME J **8:**1464-1475.

56. **Macdonald IR, Guinasso NL, Ackleson SG, Amos JF, Duckworth R, Sassen R, Brooks JM.** 1993. Natural oil slicks in the Gulf of Mexico visible from space. Journal of Geophysical Research: Oceans **98:**16351-16364.

57. **Hu P, Dubinsky EA, Probst AJ, Wang J, Sieber CMK, Tom LM, Gardinali PR, Banfield JF, Atlas RM, Andersen GL.** 2017. Simulation of Deepwater Horizon oil plume reveals substrate specialization within a complex community of hydrocarbon degraders. Proceedings of the National Academy of Sciences **114:**7432-7437.

58. **de Lorenzo V.** 2008. Systems biology approaches to bioremediation. Curr Opin Biotechnol **19:**579-589.

59. **de Lorenzo V, Fraile S, Jiménez JI.** 2010. Emerging Systems and Synthetic Biology Approaches to Hydrocarbon Biotechnology, p 1411-1435. *In* Timmis KN (ed), Handbook of Hydrocarbon and Lipid Microbiology doi:10.1007/978-3-540-77587-4_97. Springer Berlin Heidelberg, Berlin, Heidelberg.

## Tables and figures

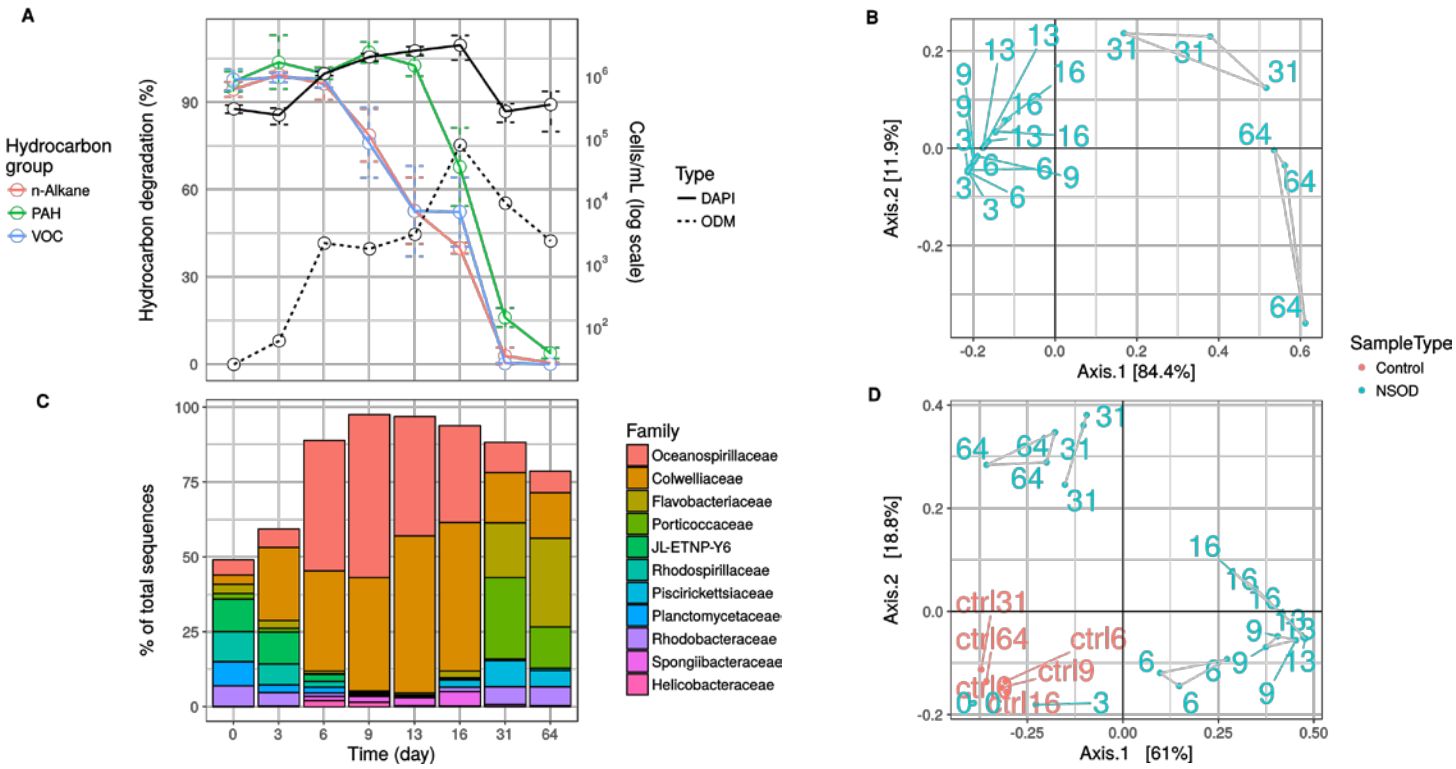


FIG 1- Comparison of microbial community structure on taxonomic level of family by 16S rRNA gene analysis (C) and total degradation rate in NSOD for n-alkanes, PAHs and VOCs after 64 days of incubation time (A, left hand side y axis) to changing abundance of total (DAPI) and oil

degrading (ODM) microbes (A, right hand side y axis). OTUs presented are contributing >1% in relative sequence abundance at least in two samples (C). Chemical and microbiological changes dynamics over the course of experimental period is displayed as Bray-Curtis and weighted-unifrack PCoA plot, respectively (B and D). Replicates from the same time-point are connected with a polygon. Oil free samples are annotated in red font with “ctrl” prefix. Oil dispersion samples are annotated in turquoise color. Numbers indicate incubation time in days.

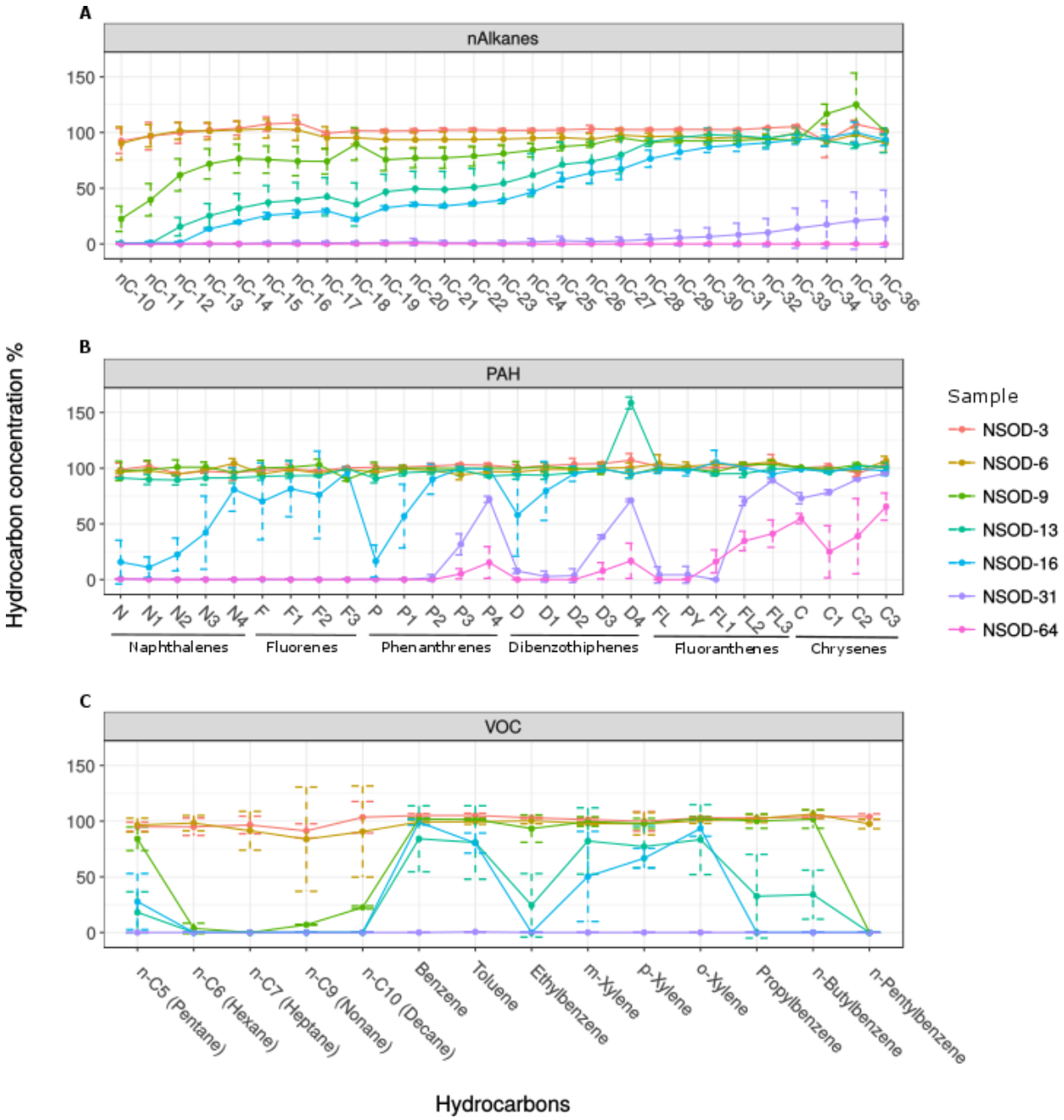


FIG 2- Degradation dynamics of targeted single compounds of n-alkanes (A), PAHs (B) and VOCs (C). Single compounds are listed on x-axis of each figure, while concentration of hydrocarbons standardized against hopane (except VOCs) is presented on y-axis. Different line colors represent different samples/sampling days. Standard deviations are presented as dashed error bars.

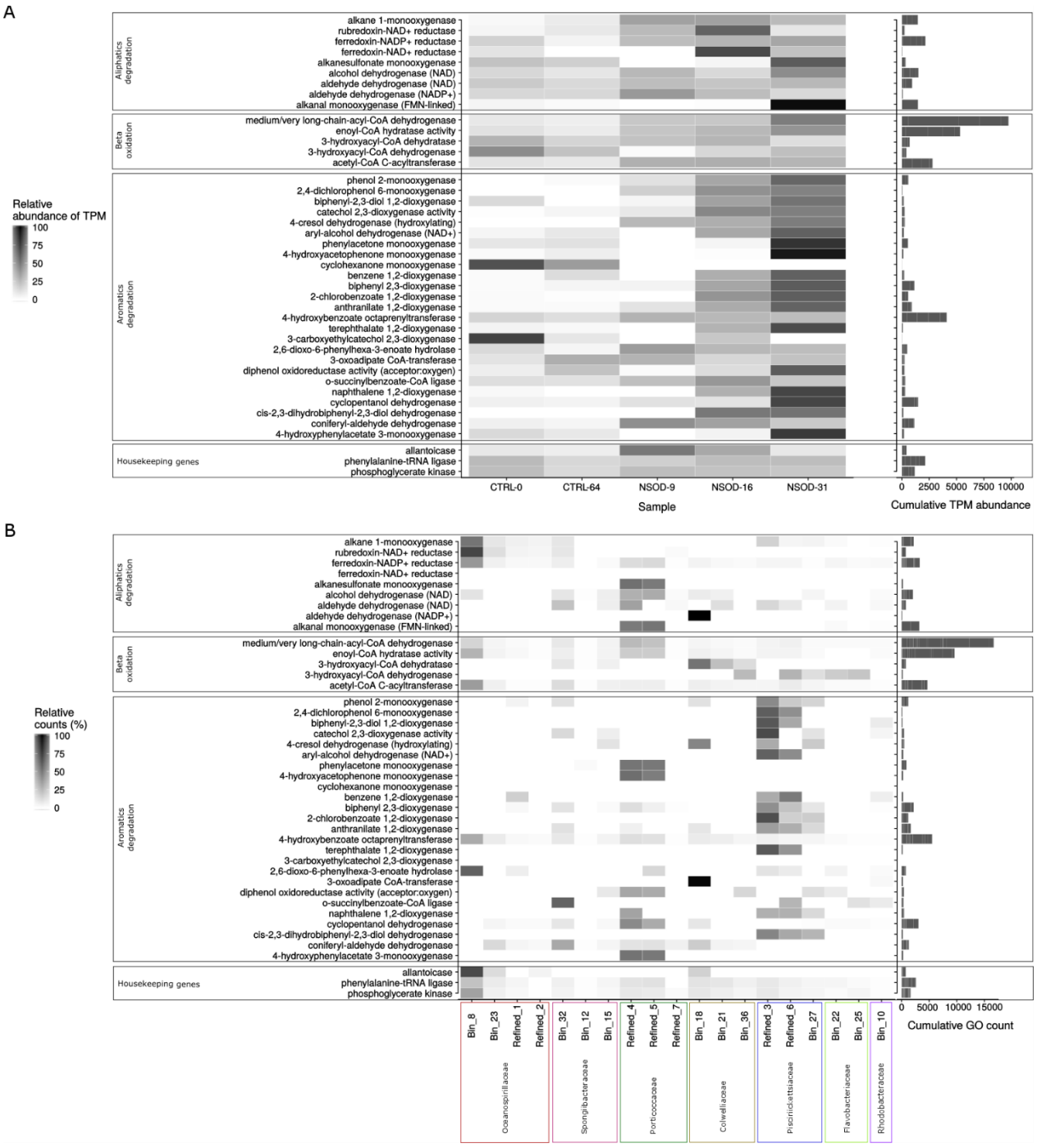


FIG 3- Heatmap representing relative abundance of selected GO terms across different samples with barchart indicating cumulative abundance of respective TPM values (A) Heatmap representing relative abundance of selected GO terms across identified bins with barchart indicating cumulative abundance of respective GO terms counted in each bin (B). Relative abundances were calculated across the samples or

identified bins for individual GO terms based on TPM (A) or total count values (B). Whereas, the cumulative TPM abundance (A) or cumulative GO count (B)was calculate by summing up all TMP or GO count values from each sample or bin for specific GO term.

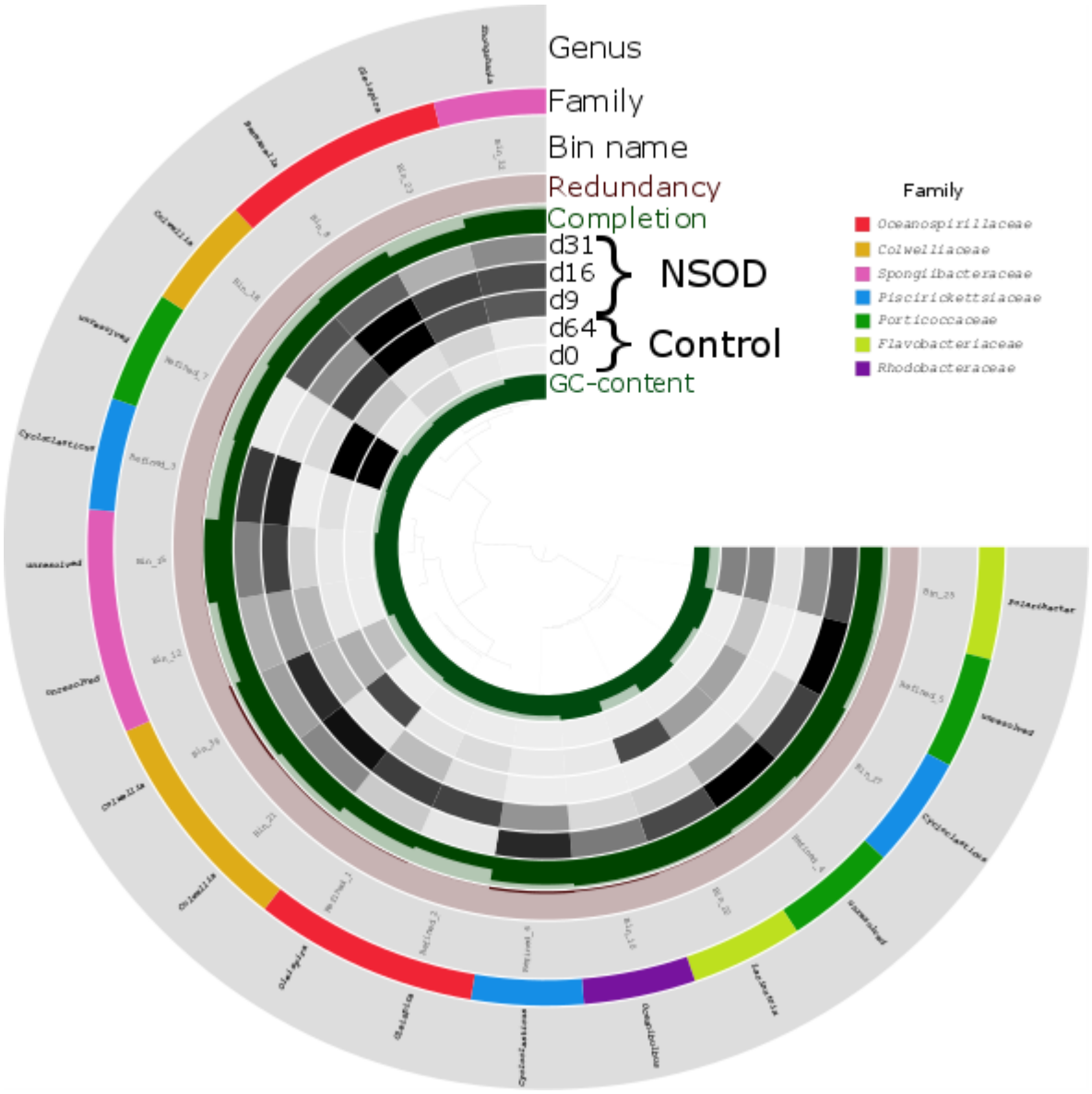


FIG 4- Anvi'o bin collection representation. Layers from inside out include following: (1) tree displays the coverage-based hierarchical clustering of 19 environmental draft genomes (bins) we determined from the co-assembly of metagenomic dataset. (2) GC-content layer. (3) The view layers for control and NSOD samples display the "mean coverage" of each bin in samples from the metagenomic dataset. Different shades of gray indicate "mean coverage" value, low (light gray) to high (dark gray). Next two layers depict the (4) completion and (5) redundancy of each bin ranging from 0 to 100, respectively. Following layer shows (6) names of each metagenomic bin. The two most outer layers show taxonomical annotation

on (7) family and (8) genus level obtained by CheckM, Anvi'o and manual phylogenetic tree curation consensus.