UNIVERSITY OF CALIFORNIA

Santa Barbara

Genomic Insights into the Marine Microbial Response to Oil Spills: Biogeographic Priming, Cryptic Hydrocarbon Cycling, and Substrate Specialization

A dissertation submitted in partial satisfaction of the

requirements for the degree Doctor of Philosophy

in Marine Science

by

Eleanor C. Arrington

Committee in charge:

Professor David L. Valentine, Chair

Professor Alyson E. Santoro

Professor Craig A. Carlson

Professor Elizabeth G. Wilbanks

December 2021

The dissertation of Eleanor C. Arrington is approved.

_____

Alyson E. Santoro

_____

Craig A. Carlson

_____

Elizabeth G. Wilbanks

_____

David L. Valentine, Committee Chair

December 2021

ACKNOWLEDGEMENTS

This work would not have been possible without the abundant support, encouragement, and guidance provided by my friends, family, committee, lab group members, and many others. To David Valentine, Lizzy Wilbanks, Craig Carlson, and Alyson Santoro: your advice, assistance, and mentorship has been invaluable. Your wisdom and perspective have helped build a solid foundation for my career and life. Thank you for helping me to see further. To my friends, family, partner Miguel, and IGPMS peers: Thank you for always being there with anything and everything I needed. To my dog Coby: I know you can't read this, but thanks for making sure I went to the beach for some exercise while working on this.

VITA OF ELEANOR C ARRINGTON
December 2021

EDUCATION

Bachelor of Science in Ecology and Evolutionary Biology, University of Rochester, May 2014
Bachelor of Science in Environmental Science, University of Rochester, May 2014
Doctor of Philosophy in Marine Science, University of California, Santa Barbara, December 2021 (expected)

PROFESSIONAL EMPLOYMENT

2014-2021: Graduate Student Researcher, Marine Science Institute, University of California, Santa Barbara
2015-2018: Teaching Assistant, Department of Earth Science, University of California, Santa Barbara
2013-2014: Teaching Assistant, Department of Biology, University of Rochester
2013-2014: Teaching Assistant, Department of Environmental Science, University of Rochester
2011-2015: Technical Assistant, University of Rochester Medical Center

PUBLICATIONS

Love, C. R., Arrington, E. C.*, Gosselin, K. M., Reddy, C. M., Van Mooy, B. A., Nelson, R. K., & Valentine, D. L. (2021). Microbial production and consumption of hydrocarbons in the global ocean. Nature Microbiology,
1-10. (*:co-lead author)

Vallota-Eastman, A., Arrington, E. C., Meeken, S., Roux, S., Dasari, K., Rosen, S., ... & Paul, B. G. (2020). Role of diversity-generating retroelements for regulatory pathway tuning in cyanobacteria. BMC genomics, 21(1), 1-13.

Leonte, M., Kessler, J. D., Kellermann, M. Y., Arrington, E. C., Valentine, D. L., & Sylva, S. P. (2017). Rapid rates of aerobic methane oxidation at the feather edge of gas hydrate stability in the waters of Hudson Canyon, US Atlantic Margin. Geochimica et Cosmochimica Acta, 204, 375-387.

Chan, E. W., Shiller, A. M., Joung, D. J., Arrington, E. C., Valentine, D. L., Redmond, M. C., ... & Kessler, J. D. (2019). Investigations of aerobic methane oxidation in two marine seep environments: Part 1—Chemical kinetics. Journal of Geophysical Research: Oceans, 124(12), 8852-8868.

Chan, E. W., Shiller, A. M., Joung, D. J., Arrington, E. C., Valentine, D. L., Redmond, M. C., ... & Kessler, J. D. (2019). Investigations of aerobic methane oxidation in two marine seep environments: Part 2—Isotope kinetics. Journal of Geophysical Research: Oceans, 124(11), 8392-8399.

AWARDS

Pacific Section- American Association of Petroleum Geologists Scholarship (May 2018)

Preston Cloud Memorial Fellowship (December 2017)
The Wendell Philips Woordring Memorial Graduate Fellowship (June 2017)
Richard and Eleanor Migues Field Research Prize (June 2016)
Annalise Kjolhede Memorial Prize for Academic Excellence in Environmental Science (May 2014)

FIELDS OF STUDY

Major Field: Marine Microbial Ecology

Studies in Petroleum Microbiology with Professor David L. Valentine

Studies in Metagenomics with Professor David L. Valentine

Studies in Methane Oxidation with Professor John Kessler

ABSTRACT

Genomic Insights into the Marine Microbial Response to Oil Spills: Biogeographic Priming, Cryptic Hydrocarbon Cycling, and Substrate Specialization

by

Eleanor C. Arrington

Our seas, oceans, and coastal zones are under great stress and pollution, particularly by crude oil, which fuels the global economy. Subsurface petroleum reservoirs originate from geo-thermo-chemical reactions on biological debris over millions of years, resulting in a complex heterogeneous mixture of hydrocarbons, with major components consisting of alkanes with different chain lengths and branch points, cycloalkanes, branched cycloalkanes, mono-aromatic, and polycyclic aromatic hydrocarbons. Populations of hydrocarbon-degrading bacteria, including many species that cannot utilize other carbon sources, are present in all marine systems and play an important role in turnover and fate of these compounds. In this dissertation, the microbial response to petroleum components is probed in multiple environments to understand the role different chemical fractions play in eliciting different niches of oil consumers, and to identify factors controlling basal seed populations of hydrocarbon degraders poised to bloom to petroleum disasters.

Through study in the subtropical North Atlantic Ocean, a cryptic long-chain alkane cycle has been confirmed, originating from cyanobacteria, dwarfing the quantity of other petroleum inputs to the ocean. In Chapter 1, I demonstrate waters in the mesopelagic underlying the photic zone hosted *n*-alkane degrading bacteria that bloomed rapidly when fed pentadecane, exhibiting exponential oxygen loss due to respiration within a week. Parallel experiments performed with sinking particles collected in situ from beneath the deep chlorophyll maximum—representing an export flux of particulate-phase pentadecane and its microbial consumers from the euphotic zone—exhibited similarly rapid bloom timing

with pentadecane, but with greater oxygen decline. Notably, bloom onset timing for other petroleum compounds with no biogenic origin in the mesopelagic is an order of magnitude slower compared to biogenic alkanes. Metagenomic analyses of pentadecane blooms exposes the metabolic pathways used for pentadecane consumption. Analysis of gene abundance in unaltered seawater from oligotrophic settings reveals long-chain alkane genes are prolific in this setting and highlights the much lower prevalence of genes related to aromatic and short-chain alkane consumption. This work emphasizes the impact of phytoplankton-derived alkanes on the widespread abundance of long-chain alkane degraders in the ocean.

The lack of biological hydrocarbon accumulation in the ocean points to their efficient consumption by networks of microorganisms. From analysis of sinking particles out of the photic zone, we know biogenic alkane consumption largely occurs in the sunlit ocean. To gain an understanding of which microbes consume biogenic alkanes we analyze the Tara Oceans global dataset for the presence of the alkane-1-monooxygenase gene (*alkB)*. Stations within the North Atlantic subtropical gyre reveal *alkB*-related genes are abundant in the surface ocean and deep chlorophyll maximum and these genes are phylogenetically distinct from the ancestrally related delta-9 fatty acid desaturase and xylene monooxygenase. Notably, a dominant clade of *alkB*-like monooxygenases belongs to the globally abundant Marine Group II (MGII) archaea and is consistently present in all surface and DCM stations. This highly successful group of surface-ocean-dwelling archaea is known for a chemoorganoheterotrophic lifestyle targeting lipids, proteins and amino acids and can utilize photoheterotrophy, but a key role in biohydrocarbon cycling was unexpected as MGII archaea are not among the ~300 genera of bacteria and archaea previously identified as hydrocarbon degraders. Our further analysis of MGII genomes show alkane-1-monooxygenase genes are present in every genus of the MGII taxonomic order, and that

other genes required to shunt alkane-derived carboxylic acids into central carbon metabolism are common among MGII.

In the Gulf of Mexico (GOM), the biodegradation of *n*-pentane in the deep ocean was investigated along a gradient of natural seepage influences, illuminating the regional influence natural seeps have on priming petroleum hydrocarbon consumption. This seawater-soluble, volatile, compound is known to partition to the ocean's interior following release from the seafloor and is abundant in petroleum reservoirs and refined products. The predominant member of the microbial community in *n*-pentane blooms is *Cycloclasticus,* and interestingly one *Cycloclasticus* ecotype favors the seep-ridden region of the GOM, whereas the other favors the open ocean environment far from natural seepage. Metagenomic analysis of the contrasting *Cycloclasticus* variants indicate the open ocean adapted variant encodes more general pathways for alkane consumption including short-chain alkanes, aromatics, and long-chain alkanes and possess pathways for dissimilatory nitrate reduction and thiosulfate oxidation, whereas the near-seep variant specializes solely on short-chain alkanes and aerobic metabolism. Metagenomic reconstruction and phylogenetic analysis of *Cycloclasticus* from publicly available sequence data reveals distinct strategies in hydrocarbon generalization and specialization for each major clade within the *Cycloclasticus* genera.

Cycloalkanes are a major component of petroleum and many refined petroleum products. Methylated cycloalkane incubations using methylcyclohexane and methylcyclopentane were conducted in stations spanning a transect across the Gulf of Mexico. In three out of four stations, a single strain belonging to a novel genus within the *Poricccaceae* family and class *Gammaproteobacteria* bloomed to consume the methyl-cycloalkane provided within 18-21 days. Metagenomic reconstruction and analysis of the central carbon metabolism reveal a distinct strategy to oxidize cycloalkanes. A larger phylogenomic analysis reveals this methyl-cycloalkane consumer belongs to a

monophyletic clade of genomes belonging to the same genera which originate from other environments with petroleum influence at subseafloor aquifers, hydrothermal vents, and petroleum mesocosms from a variety of marine sources. A defining feature of this clade is the presence of a divergent particulate hydrocarbon monooxygenase, which in the phylogenetic tree of the CuMMO enzyme superfamily (of ammonia and hydrocarbon monooxygenases) forms a distinct novel monophyletic clade from all other ammonia, methane, ethane, and ethylene monooxygenases.

TABLE OF CONTENTS

Introduction

Petroleum-based products are the major source of energy for modern industry and daily life. Leaks and accidental spills occur regularly during the exploration, production, refining, transport, storage, and use of petroleum and petroleum products. Hydrocarbons naturally enter the ocean through petroleum seepage from geologic strata and enzymatic reactions in the biosphere. Together, natural and anthropogenic inputs add more than 1,300,000 metric tons of petroleum to the marine environment annually (National Resource Council, 2003). Release of petroleum hydrocarbons into the environment is a major cause of water and soil pollution. Contamination with petroleum causes extensive damage to local ecosystems as accumulation of pollutants within animal and plant tissue causes mutations and death.

Petroleum originates from geo-thermo-chemical reactions on biological debris over millions of years. As such, it is an extremely diverse mixture of organic compounds consisting of several thousand individual structures, each with unique physical-chemical properties, reactivity, and environmental fates. One of the primary mechanisms by which petroleum and other hydrocarbon pollutants are eliminated from the environment is microbial attack and degradation. General biodegradation rates have been shown to be highest for the saturates, followed by light aromatics, with high-molecular-weight aromatics and polar compounds exhibiting extremely low rates of degradation (Leahy and Colwell, 1990).

The work in this dissertation began four years after the *Deepwater Horizon* blowout caused the release of several million barrels of oil and gas to the deep Gulf of Mexico, prompting the collection and analysis of the most comprehensive dataset in existence for hydrocarbons in the marine setting (Diercks et al., 2010; Hazen et al., 2010a; Valentine et al., 2010, 2012, 2014; Kessler et al., 2011; Mason et al., 2012; Reddy et al., 2012; Rivers et al., 2013; Dombrowski et al., 2016). The context of the hydrocarbon release at the Macondo

wellhead (1,544 m below the sea surface) posed a challenge to scientists and responders alike and highlighted a major knowledge gap regarding the fate of petroleum compounds in both the oligotrophic surface waters and the deep ocean, which together comprise the majority of marine ecosystems. This bias became clear in the microbiology realm when the bacterial response was dominated mainly by unknown or unexpected genera, or genera performing unexpected functions (Hazen et al., 2010b; Valentine et al., 2010; Eren et al., 2015; Hu et al., 2017). The deep hydrocarbon intrusion prevented the typical loss of volatile compounds to the atmosphere, which also highlighted another major knowledge gap regarding this class of compounds in the ocean (Valentine et al., 2010; Redmond and Valentine, 2012a). Together, these unknowns regarding the fate of volatile compounds in the oligotrophic and deep ocean prompted the questions that led to Chapter 3 and 4.

In the first year of this dissertation work a finding by (Lea-Smith et al., 2015) implicating cyanobacteria as the source of a major hydrocarbon cycle in the ocean changed the course of this dissertation. As our lab began working on observing and constraining this cycle in nature it became clear that microbial consumption was a key factor in this cycle which warranted study and resulted in Chapter 1 and 2. Each of the four chapters are unified in elucidating the processes (whether it is natural seepage or biogenic alkane production) that maintain populations of hydrocarbon degraders in uncontaminated seawater and how those populations respond to colonize petroleum inputs in various settings.

The complex lifestyle of hydrocarbon-consuming marine microbes governs their distribution in marine ecosystems, which in turn controls the rate and degree consumption of petroleum occurs. Until recently, our studies on petroleum microbiology have been focused on a select number of obligate petroleum-consuming cultivated isolates. These isolates have provided our foundational knowledge of the enzymes used to convert various petroleum compounds to the central carbon currency used to generate energy in cells.

Using these metabolic blueprints, we can now utilize current sequencing technology to scan for similar genomic pathways in environmental DNA and infer function from phylogenetic analysis. This type of analysis is utilized in every chapter of this dissertation and leads to many findings regarding uncultivated microbes performing key functions in the marine response to oil spills and the potential utility of facultative petroleum degraders making a living off other natural organic compounds in the ocean.

Chapter 1

1. Biogenic Pentadecane Consumption in the North Atlantic Mesopelagic

In Chapter 1, I explore the fate of hydrocarbons produced biologically by cyanobacteria and other photosynthetic organisms in oligotrophic regions. This work derives from a recent hypothesis presented by (Lea-Smith et al., 2015) of a widespread cyanobacterial hydrocarbon cycle and the work conducted in the dissertation of my lab mate, Connor Love, demonstrating the existence of this cycle in nature and its magnitude in the ocean. The work presented in this chapter has been published in (Love and Arrington et al., 2021).

1.1 Introduction

Petroleum hydrocarbons, totaling ~1.3 Tg per year, enter the ocean via natural means—such as seepage from geologic strata (Kvenvolden and Cooper, 2003)—as well as from anthropogenic sources, such as industrial oil spills associated with extraction, transportation, and consumption of petroleum (National Resource Council, 2003). Oceanic microorganisms also contribute to natural hydrocarbon sources in the form of methane in both anoxic (Saunois et al., 2020) and oxic (Repeta et al., 2016) settings, as well as microalgal production of isoprene (McGenity et al., 2018) and ethylene (Seifert et al., 1999). Among these sources, petroleum inputs to the ocean have received significant attention because of petroleum's inherent toxicity to marine life, the negative impacts high concentration of pollutants have in ecologically sensitive areas, and because discharge sometimes occurs through catastrophic and preventable events such as tanker spills (e.g., Exxon Valdez), pipeline breaches (e.g., Kalamazoo River), and blowouts (e.g., Deepwater Horizon).

1.2 Background

1.2.1 Cyanobacterial Hydrocarbons

(Lea-Smith et al., 2015) estimated the global scale production of pentadecane ($C_{15}H_{32}$; nC_{15}), heptadecane ($C_{17}H_{36}$; nC_{17}), and heptadecene ($C_{17}H_{34}$; ΔC_{17}) from the two most abundant marine cyanobacteria *Prochlorococcus* and *Synechococcus.* This work involved scaling cultivation-observed alkane production to the global production of these genera, concluding a hypothetical global source of ~308-771 million tons of hydrocarbons per year. Authors further noted all *Prochlorococcus* and *Synechococcus* strains have the predicted capacity for alkane production via the acyl-ACP reductase (FAR) and aldehyde deformylating oxygenase (FAD), enzymes which catalyze the decarboxylation of the acyl chains typically used for membrane lipid tails (Schirmer et al., 2010; Lea-Smith et al., 2015).

nC_{15}-nC_{19} hydrocarbons play a key physiological role for cyanobacteria, whereby they accumulate in thylakoid and cytoplasmic membranes promoting membrane flexibility and curvature (Lea-Smith et al., 2016). Hydrocarbons may be particularly useful to alter membrane structure for phytoplankton under nutrient limitation because unlike polar lipids, hydrocarbons lack headgroups with phosphate, nitrogen, and sulfur. Indirect evidence for this global alkane and alkene cycle is also present in past literature where long-chain alkanes/alkenes were first noted in blue green algae by (Han and Calvin, 1969), as well as in active cyanobacterial mats and sediments/fossils associated with cyanobacterial inputs (Grimalt et al., 1992; Arp et al., 1999). Alkanes and alkenes have also been detected in the form of spot measurements in water-column samples from cyanobacterial blooms, albeit with little dynamical context (Winters et al., 1969; Gschwend et al., 1980; Carpenter et al., 1997; Wang et al., 2011; White et al., 2019). Collectively, studies have provided pieces to a past and current complex biogeochemical cycle of cyanobacterial derived alkanes and alkenes. However, the biogeochemical cycle of oceanic hydrocarbons has not been directly observed in nature or closed, and the ecological ramifications of this input are scarcely considered beyond an untested hypothesis that biohydrocarbons prime the oceans for consumption of petroleum.

1.2.2 Pentadecane Production in the North Atlantic

In our study (Love and Arrington *et al.*, 2021) we directly observe and constrain the biogenic alkane/alkene cycle in nature. The dissertation of Connor Love covers his work on constraining the pentadecane stock, production rates, and global budget, and is briefly summarized here as background information. Our collective efforts focus on the North Atlantic subtropical oligotrophic gyre for which productivity is dominated by hydrocarbon-producing cyanobacteria *Prochlorococcus* and *Synechococcus*, genera estimated to account for ~25% of the global ocean's net primary production (Chisholm et al., 1988; Field et al., 1998; Flombaum et al., 2013). Subtropical oligotrophic gyres comprise ~40% of the planet's surface (Polovina et al., 2008; Karl and Church, 2014), tend to host predominantly

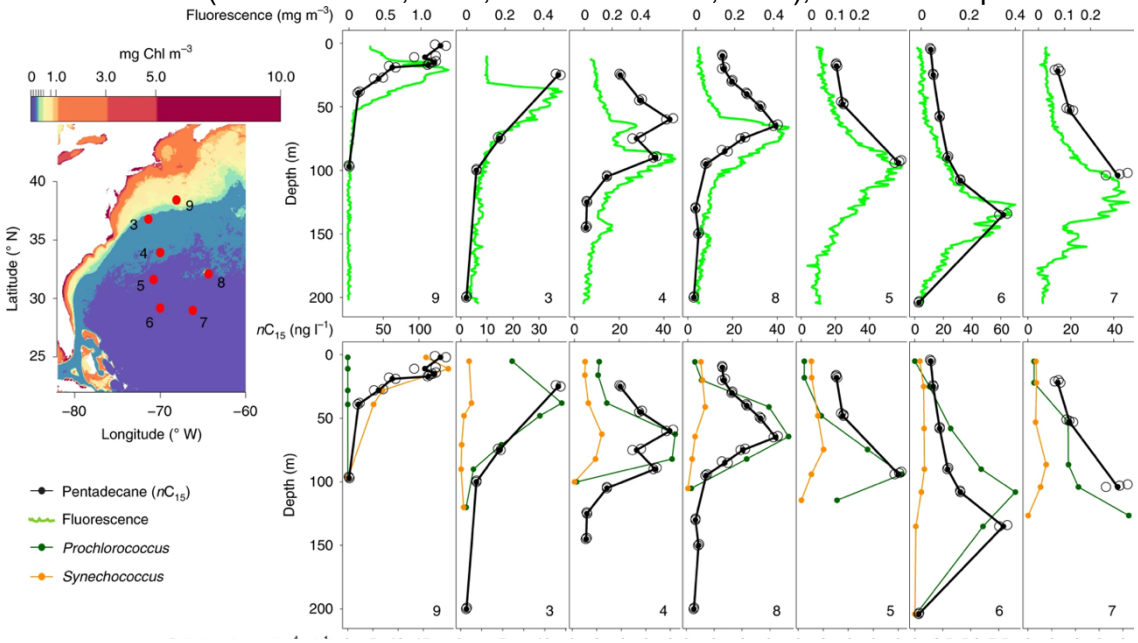


Figure 1. This work was completed by Connor Love and will be included in his dissertation. Study area (at left) shows station coordinates mapped onto 4-km resolution MODIS-Aqua satellite chlorophyll (Chl) concentration for 2017. Station 3 was located in the Gulf Stream and station 9 targeted a *Synechococcus* bloom; all other stations captured more 'typical' *Prochlorococcus*-dominated oligotrophic water. Pentadecane depth distributions for each station are displayed with fluorescence (top row) and cyanobacterial abundance (bottom row). Depth distributions are organized by descending latitude with pentadecane distribution and station number duplicated for ease of comparison. Open black circles show biologically independent pentadecane measurements; each data point represents the contents of one distinct sample bottle. Replicates are sequentially moved 1 m below the other for visualization (water was taken from same depth, depth of top replicate); solid black circles indicate mean of $n = 2$ at stations 9, 4, 8 and 6 and represent mean of $n = 3$ for stations 3, 5 and 7.

cyanobacterial productivity (Polovina et al., 2008) and are far from the continents and associated petroleum sources that could mask the signal of cyanobacterial hydrocarbons.

We report that for over 441 particulate samples (≥ 0.2μm), spanning seven locations in the western North Atlantic, cyanobacteria mainly produce *n*-pentadecane (Figure 1). Concentrations of *n*-pentadecane in oligotrophic gyres range from 2-65 ng L^{-1}. We also observed pentadecane concentrations within the Gulf Stream at a maximum value of ~80 ng L^{-1} and within a *Synechococcus* bloom at a maximum of ~130 ng L^{-1}. We observe that pentadecane in oligotrophic waters presents as a distinct subsurface concentration maximum that coincides with both fluorescence and cyanobacteria cell counts, aligning with the deep chlorophyll maximum (DCM). Compared to the DCM, concentrations of pentadecane above the DCM are lower (10-15 ng L^{-1}), however they become undetectable at the base of the euphotic zone (150-200 m). Using ^{13}C-enriched bicarbonate as an isotopic tracer we observed production of pentadecane in shipboard incubations conducted at ambient temperature and light levels (Love and Arrington et al., 2021). Using this method, we observed the greatest pentadecane production near the DCM (~30 ng nC_{15} L^{-1} d^{-1}) where approximately 1% of photosynthetically active radiation penetrates. Diel variability in pentadecane concentration is also greatest at the DCM and 1% PAR, further consistent with the concept that pentadecane production and consumption is concentrated to the lower euphotic zone.

Importantly, cyanobacteria lack alkane degradation pathways to consume hydrocarbons once they are formed, indicating a widespread external degradation process (Lea-Smith et al., 2015). Pentadecane and other long chain *n*-alkanes can also be major components of spilled oil (Head et al., 2003, 2006; Reddy et al., 2012), and thus a priming effect has been proposed by which populations of petroleum degraders are sustained in a latent hydrocarbon cycle enabling a rapid response to oil spills (Lea-Smith et al., 2015). However, petroleum contains thousands of compounds in addition to *n*-alkanes (Frysinger

et al., 2003; Wardlaw et al., 2008; McKenna et al., 2014), leading us to question the extent to which steady state biohydrocarbon consumption primes the ocean with a microbial community capable of rapidly consuming this myriad of other compounds. This chapter includes my work to constrain the consumption component of the biogenic pentadecane cycle in the North Atlantic Ocean and to understand how the cycle may impact the microbial response to oil spills. Biodegradation was thus investigated to differentiate factors driving consumption of biological versus petroleum hydrocarbons.

1.3 Methods

1.3.1 Respiration Experiment

Pentadecane respiration incubations were conducted at station 3 (36° 50.93' N, 71° 23.94' W) and station 6 (29° 4.79' N, 69° 44.38' W) with water collected from 500 m. Pentane respiration incubations were conducted at stations 2 (40° 9.14' N, 68° 19.889' W), 4 (33° 58.21' N, 69° 43.38' W), 10 (27° 30.41' N, 87° 12.41' W), 11 (27° 15.00' N, 89° 05.05' W), 12 (27° 11.60' N, 90° 41.75' W) and 13 (27° 38.40' N, 90° 54.98' W) with water collected from 1,000 m. Water samples from the CTD Niskin bottles were transferred to 250-ml glass serum vials using a small length of Tygon tubing. Vials were filled with at least three volumes of water to overflow. Care was taken to ensure no bubbles were present before sealing with a Teflon-coated rubber stopper and crimp cap. Abiotic controls were amended with 14 g of zinc chloride before sealing. All bottles except for unamended blank controls immediately received 10 µl of pentadecane or pentane using a gas-tight syringe and were maintained in the dark at in situ temperature (15°C for pentadecane, 4°C for pentane). Sediment traps at stations 3 and 6 were deployed for 24 h at 150 m. For each particle addition, 10 ml of particle-laden seawater was vortexed lightly for 1 min, then 2 ml of the vortexed seawater was added to the bottom of each serum bottle with a pipet via displacement. Each serum bottle was fixed with a contactless optical oxygen sensor (OXSP5, Pyroscience) on the inner side with silicone glue and oxygen content was

monitored approximately every 12 h with a fibre optic oxygen meter (FireStingO2, Pyroscience). Observed changes in oxygen content were normalized to abiotic and unamended seawater controls to correct for variability due to temperature and background respiration not caused by alkane addition. In the case of the pentadecane particle incubations, oxygen losses from particles and seawater were subtracted from particle plus pentadecane treatments to isolate pentadecane respiration. Bloom onset is operationally defined as three consecutive time points with oxygen loss >0.21 μM h^{-1}. At the end of each respiration experiment incubations were sacrificially harvested and filtered on a 0.22-μm polyethersulfone filter.

1.3.2 DNA Extraction

DNA extraction was performed from ¼ of each filter. The PowerSoil DNA Extraction (Qiagen) was used according to manufacturer recommendations with the following modifications: 200 μl of bead beating solution was removed in the initial step and replaced with phenol chloroform isoamyl alcohol, the C4 bead binding solution was supplemented with 600 μl of 100% ethanol and we added an additional column-washing step with 650 μl of 100% ethanol. Extracts were purified and concentrated with ethanol precipitation.

1.3.3 16S rRNA Gene Amplification and Analysis

We amplified and barcoded the V4 region of the 16S rRNA gene using the method described (Kozich et al., 2013) previously with small modifications to the 16Sf and 16Sr primers (Caporaso et al., 2012). Amplicon PCR reactions contained 1 μl of template DNA, 2 μl of forward primer, 2 μl of reverse primer and 17 μl of AccuPrime Pfx SuperMix. Thermocycling conditions consisted of 95°C for 2 min; 30 cycles of 95 °C for 20 s, 55 °C for 15 s and 72 °C for 5 min; and a final elongation at 72 °C for 10 min. Sample DNA concentrations were normalized using the SequelPrep Normalization Kit, cleaned using the DNA Clean and Concentrator kit, visualized on an Agilent Tapestation and quantified using a Qubit fluorometer. Samples were sequenced and demultiplexed at the University of

California Davis Genome Center on the Illumina MiSeq platform with 250-nucleotide paired-end reads. A PCR-grade water sample was included in extraction, amplification, and sequencing as a negative control to assess for DNA contamination.

Trimmed fastq files were quality-filtered using the fastqPairedFilter command within the dada2 R package, v.1.9.3 (Callahan et al., 2016), with the following parameters: truncLen = c(190,190), maxN = 0, maxEE = c(2,2), truncQ = 2, rm.phix = TRUE, compress = TRUE, multithread = TRUE. Quality-filtered reads were dereplicated using derepFastq command. Paired dereplicated fastq files were joined using the mergePairs function with the default parameters. A single-nucleotide variant table was constructed with the makeSequenceTable command and potential chimeras were removed de novo using removeBimeraDenovo. Taxonomic assignment of the sequences was done with the assignTaxonomy command using the Silva taxonomic training dataset formatted for DADA2 v.132. If sequences were not assigned, they were left as not applicable (NA).

1.3.4 Metagenome Assembly, Binning, and Relative Abundance

Metagenomic library preparation and shotgun sequencing were conducted at the University of California Davis DNA Technologies Core. DNA was sequenced on the Illumina HiSeq4000 platform, producing 150-base pair (bp) paired-end reads with a targeted insert size of 400 bp. Quality control and adaptor removal were performed with Trimmomatic (Bolger et al., 2014) (v.0.36; parameters: leading 10, trailing 10, sliding window of 4, quality score of 25, minimum length 151 bp) and Sickle (Joshi and Fass, 2011) (v.1.33 with paired-end and Sanger parameters). Concatenation of high-quality reads for replicate samples (for coassembly) was conducted before assembly (see Table 1 for more details on coassembly). The trimmed high-quality reads were assembled using metaSPAdes (Nurk et al., 2017) (v.3.8.1; parameters k = 21, 33, 55, 77, 88, 127). The quality of assemblies was determined using QUAST (Gurevich et al., 2013) (v.5.0.2 with default parameters). Sequencing coverage (and differential coverage for coassemblies) was determined for each

assembled scaffold by mapping high-quality reads to the assembly using Bowtie2 (Langmead and Salzberg, 2012) (v.2.3.4.1; default parameters) with Samtools (Li et al., 2009) (v.1.7). Contigs greater than 2,500 bp were manually binned using Anvi'o with Centrifuge (Eren et al., 2015; Kim et al., 2016) (v.1.0.1) based on coverage uniformity (v.5). Quality metrics for metagenome-assembled genomes (MAGs) were determined using CheckM (Parks et al., 2015) (v.1.0.7; default parameters). The taxonomic identity of each MAG was determined using GTDB-Tk (Chaumeil et al., 2020) (v.1.0.2) against The Genome Taxonomy Database (https://data.ace.uq.edu.au/public/gtdb/data/releases/release89/89.0/, v.r89). The length-normalized relative abundance of MAGs was determined for each sample as by (Tully et al., 2017).

1.3.5 Metagenomic Annotation of Hydrocarbon Degradation Genes

Open reading frames were predicted for MAGs using Prodigal (Hyatt et al., 2010) (v.2.6.3; default parameters). Functional annotation was determined using HMMER3 (Eddy, 2011)(v.3.1b2) against the Pfam database (v.31.0) with an expected value (e-value) cutoff of 1×10^{-7} and KofamScan (v.1.1.0) (Aramaki et al., 2020) against the Hidden Markov model (HMM) profiles for Kyoto Encyclopedia of Genes and Genomes and Kegg Orthology (KEGG/KO) with a score cutoff of 1×10^{-7}. To find the number of hits for *almA* we used Pfam (PF00743), for *rhdA* we used Pfam (PF00848) and for *pHMO* we summed Pfam hits (subunit a: PF02461; subunit b: PF04744; subunit c: PF04896).

For alkane-1-monooxygenase (*alkB*) detection we used both HMMER3 against the Pfam database (PF00487) and KofamScan against the KEGG/KO HMM profiles (K00496). Each hit was manually curated using Geneious Prime v.2019.2.3 (https://www.geneious.com) to search for the eight-histidine residues considered catalytically essential for function (Shanklin et al., 1994a). The base seed alignments for both PF00487 and K00496 include the ancestrally related proteins, fatty acid desaturase

and xylene monooxygenase; therefore, we found it necessary to phylogenetically analyze each hit to determine relation to *alkB*. Through this method we learned that the HMMER3 method with Pfam ID PF00487 identifies more hits within each MAG for *alkB* than KofamScan with K00496; however, those additional hits were generally more closely related to fatty acid desaturases than *alkB*. We excluded any hits that formed a well-supported monophyletic clade with xylene monooxygenase or fatty acid desaturase from our final number of copies of *alkB*. In total, we used KofamScan with K00496 to search for *alkB*, manually curated the results to ensure presence of eight-histidine residues essential for function and phylogenetically analyzed each hit for relation to *alkB* compared with fatty acid desaturase and xylene monooxygenase.

1.3.6 Phylogenetic analyses

Each putative *alkB* hit was aligned using MUSCLE (Edgar, 2004) (v.3.8.425). All columns with >95% gaps were removed using TrimAL (Capella-Gutiérrez et al., 2009) (v.1.2). Phylogenetic analysis of concatenated *alkB* was inferred by RAxML (Stamatakis, 2014) (v.8.2.9; parameters: raxmlHPC -T 4 -s input -N autoMRE -n result -f a -p 12345 -x 12345 -m PROTCATLG). Resulting trees were visualized using FigTree (FigTree) (v.1.4.3).

1.3.7 Hydrocarbon Gene Abundance in the Tara Oceans Dataset

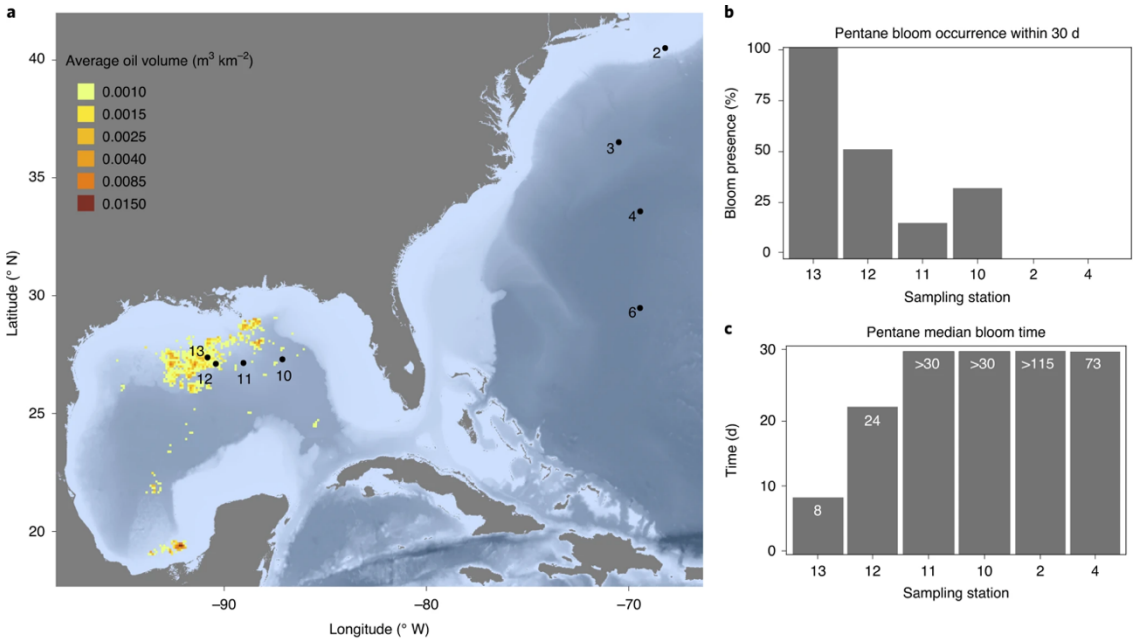
To quantify the abundance of genes involved with hydrocarbon degradation we queried the Ocean Microbial Reference Gene Catalogue (OM-RGC) dataset (see http://ocean-microbiome.embl.de/companion.html) from the Tara Oceans expedition (Sunagawa et al., 2015) for KEGG identifiers of interest. These included genes for the activation of alkanes such as alkane-1-monooxygenase (K00496), flavin-binding monooxygenase (K10215) and particulate hydrocarbon monooxygenase (K10944, K10945, K10946), as well as aromatic hydrocarbons such as toluene dioxygenase (K03268), naphthalene 1,2-dioxygenase (K14579, K14580, K14578, K14581), toluene methyl-monooxygenase (K15757 and K15758), p-cymene methyl-monooxygenase

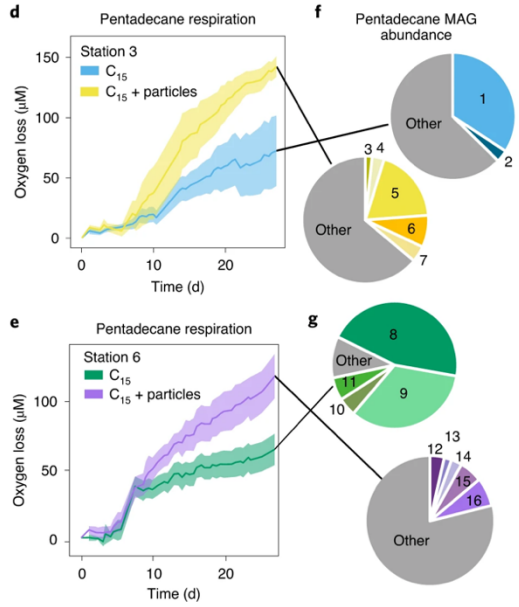
(K10616), benzene/toluene/chlorobenzene dioxygenase (K18089) and biphenyl 2,3-dioxygenase (K08689, K15750). We extracted the abundance of each gene from the Tara Oceans OM-RGC profiles dataset which was calculated from read counts mapped to each reference gene normalized by the gene length (Sunagawa et al., 2015). The total abundance of OM-RGC sequences matching the reference gene identifier was normalized to the total abundance of the single-copy gene *recA* (KEGG identifier: K03553), as performed in previous studies, to calculate abundance on a per-genome level (Martinez et al., 2010; Sosa et al., 2019). The resulting data are included in Supplementary Data 1.

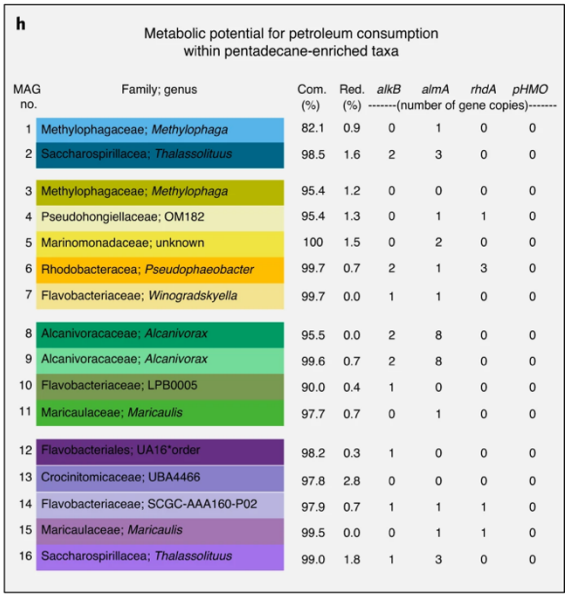
1.4. Results and Discussion

1.4.1 Biohydrocarbon consumption decoupled from petroleum

Our findings indicate a biological pentadecane cycle at steady state based on rapid production, consistent concentrations, and the tight coupling to cyanobacterial physiology – spanning ≥40% of the Earth. Ocean-going experimentation revealed that waters in the mesopelagic underlying the North Atlantic subtropical gyre photic zone hosted *n*-alkane degrading bacteria that bloomed rapidly when fed pentadecane, exhibiting exponential oxygen decline within ~1 week (Fig. 2d-e). Parallel experiments performed with sinking particles collected in-situ from beneath the DCM – representing an export flux of particulate phase pentadecane and its bacterial consumers from the euphotic zone – exhibited similarly rapid bloom timing with pentadecane, but with greater oxygen decline. Despite similar timing of respiratory blooms on pentadecane with and without particles, each station and treatment displayed a distinctive bacterial response by a limited number of taxa (Fig.







h

Metabolic potential for petroleum consumption within pentadecane-enriched taxa

MAG no.	Family; genus	Com. (%)	Red. (%)	*alkB*	*almA*	*rhdA*	*pHMO*
				(number of gene copies)			
1	Methylophagaceae; *Methylophaga*	82.1	0.9	0	1	0	0
2	Saccharospirillacea; *Thalassolituus*	98.5	1.6	2	3	0	0
3	Methylophagaceae; *Methylophaga*	95.4	1.2	0	0	0	0
4	Pseudohongiellaceae; OM182	95.4	1.3	0	1	1	0
5	Marinomonadaceae; unknown	100	1.5	0	2	0	0
6	Rhodobacteracea; *Pseudophaeobacter*	99.7	0.7	2	1	3	0
7	Flavobacteriaceae; *Winogradskyella*	99.7	0.0	1	1	0	0
8	Alcanivoracaceae; *Alcanivorax*	95.5	0.0	2	8	0	0
9	Alcanivoracaceae; *Alcanivorax*	99.6	0.7	2	8	0	0
10	Flavobacteriaceae; LPB0005	90.0	0.4	1	0	0	0
11	Maricaulaceae; *Maricaulis*	97.7	0.7	0	1	0	0
12	Flavobacteriales; UA16*order	98.2	0.3	1	0	0	0
13	Crocinitomicaceae; UBA4466	97.8	2.8	0	0	0	0
14	Flavobacteriaceae; SCGC-AAA160-P02	97.9	0.7	1	1	1	0
15	Maricaulaceae; *Maricaulis*	99.5	0.0	0	1	1	0
16	Saccharospirillacea; *Thalassolituus*	99.0	1.8	1	3	0	0

Figure 2. Microbial respiratory blooms on pentane and pentadecane quantified via contactless optical oxygen sensors, followed by metagenomic analysis. **a**, Oceanographic sampling stations relative to natural petroleum seepage with increasing distance from intense seepage as follows: 13, 12, 11, 10, 6, 4, 3 and 2. **b**, Occurrence of respiratory blooms on pentane (petroleum proxy compound) at 1,000 m with increasing distance from seepage ($n = 6$ at stations 10–13, $n = 8$ at stations 2 and 4). **c**, Timing of respiratory blooms on pentane with experiment duration 30 d at stations 10–13 and 115 d for stations 2 and 4. Numbers listed in white are median bloom times over the entire duration of the experiment, which varied from 30 d to 115 d, assuming all bottles bloom given enough time. **d**, Pentadecane respiration at station 3 (500 m) ± particles (solid line indicates mean; shading indicates ± standard deviation; $n = 6$ at station 3). **e**, Pentadecane respiration at station 6 (500 m) ± particles (solid line indicates mean; shading indicates ± standard deviation; $n = 4$ at station 6). **f**, Relative abundance of metagenomes (MAGs) at final time point of pentadecane incubations (~28 d) from station 3. **g**, Relative abundance of metagenomes at final time point of pentadecane incubations (~28 d) from station 6. **h**, Genome quality and metabolic potential for MAGs. Abbreviations include completion (Com); redundancy (Red); *alkB* (alkane-1-monooxygenase); *almA* (flavin-binding monooxygenase); *rhdA* (ring-hydroxylating dioxygenase subunit a); and *pHMO* (particulate hydrocarbon monooxygenase subunits A, B and C). *MAG 12 is unclassified at the family and genus level, and therefore we have listed the class and order. For panels **b**–**e**, '*n*' describes the number of biologically independent incubations.

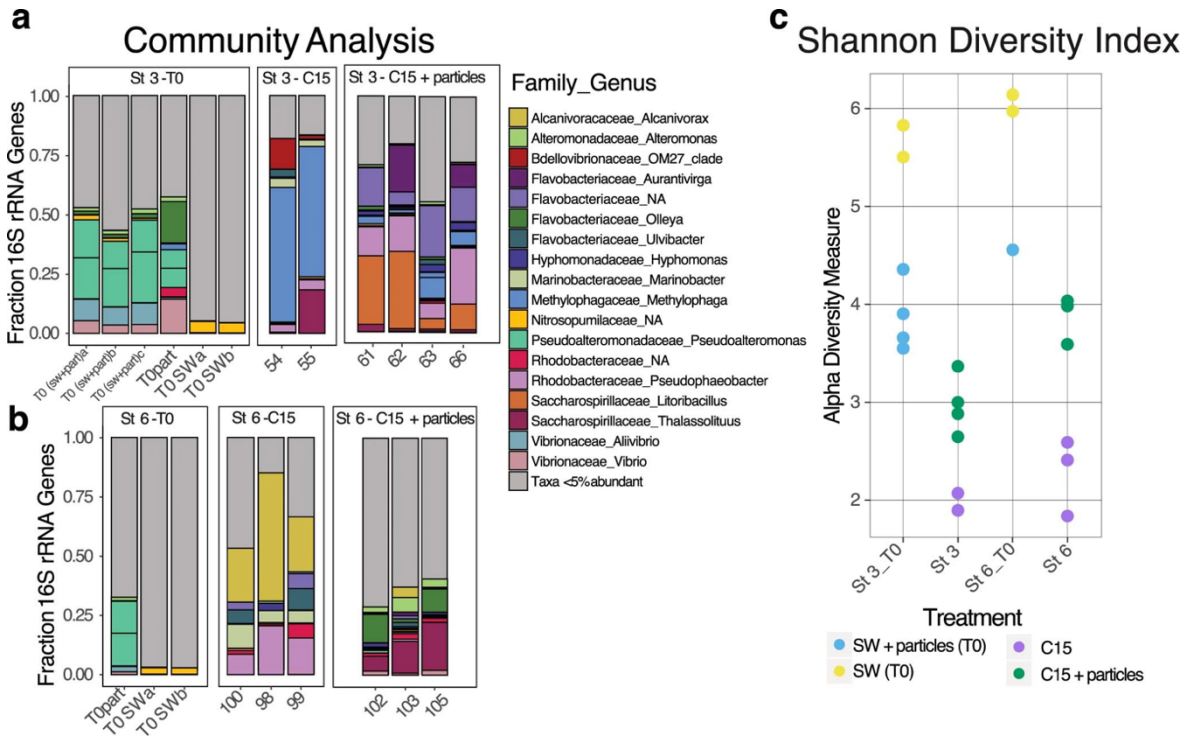


Figure 3. a-b, Microbial community composition within pentadecane incubations informed via the V4 region of the 16 S rRNA gene for initial samples and those harvested at 27 days (station 3) and 29 days (station 6). Labels on x axis are sample IDs of biologically independent DNA samples with the following abbreviations (T0: time point 0, T0part: initial sediment trap particle community, T0SW (a and b): initial seawater community, T0sw + part (a, b, and c): initial seawater community immediately after particles added, #: pentadecane enrichment). Nucleotide variants are grouped by genus and are listed under associated family and genus; if genus is unclassified than it is listed as NA. All taxa less than 5% are aggregated and shaded gray. **c**, Shannon diversity index for each biologically independent DNA sample. Shannon indices for pentadecane (n = 2 at station 3, n = 3 at station 6) and pentadecane + particles (n = 4 at station 3, n = 3 at station 6).

2f-h, Fig. 3a). Blooms on pentadecane were dominated by *Alcanivorax* at station 6 and *Methylophaga* at station 3, whereas the addition of particles favored *Thalassolituus* and an uncharacterized genus belonging to the family *Marinomonadaceae*, at these respective stations (Fig. 2f-h, Fig. 3). Using metagenomics, we compiled genomes for the dominant organisms that bloomed on pentadecane, finding the alkB and almA genes (Fig. 2h; Table 1) – genes known to encode proteins that act on medium (C_5 to C_{11}) and long chain *n*-alkanes (C_{12}-C_{30}) – to be common among these taxa, with up to 10 copies (almA + alkB) per genome (Smits et al., 2002; van Beilen et al., 2003; Wang and Shao, 2012). It was necessary to use a phylogenetic analysis to differentiate *alkB* genes from ancestrally

related xylene monooxygenase and fatty acid desaturase genes (Fig. 4). Each genome also encodes beta oxidation functionality, essential for shunting alkane-derived carboxylic acids into central carbon metabolism.

MAG name	Assembly Method (sample ID)	Comp. (%)	Red. (%)	# Contigs	N50 (bp)	GC Content	Length (bp)	Taxonomy
1_P54_P55_methylophagaceae	Coassembly (54 + 55)	82.1	0.9	211	25666	45.2	2556684	d Bacteria;p Proteobacteria;c Gammaproteobacteria;o Nitrosococcales;f Methylophagaceae;g Methylophaga;s Methylophaga sp002696735
2_P54_P55_saccharospirillaceae	Coassembly (54 + 55)	98.5	1.6	45	126959	46.7	3730513	d Bacteria;p Proteobacteria;c Gammaproteobacteria;o Pseudomonadales;f Saccharospirillaceae;g Thalassolituus;s Thalassolituus oleivorans
3_P61_P62_methylophagaceae	Coassembly (61 + 62)	95.4	1.2	94	44824	45.8	2556822	d Bacteria;p Proteobacteria;c Gammaproteobacteria;o Nitrosococcales;f Methylophagaceae;g Methylophaga;s Methylophaga thiooxydans
4_P61_P62_pseudohongiellaceae	Coassembly (61 + 62)	95.4	1.3	74	79032	53.3	2910293	d Bacteria;p Proteobacteria;c Gammaproteobacteria;o Pseudomonadales;f Pseudohongiellaceae;g OM182;s OM182 sp001438145
5_P61_P62_marinomonadaceae	Coassembly (61 + 62)	100	1.5	17	479598	42.5	4846406	d Bacteria;p Proteobacteria;c Gammaproteobacteria;o Pseudomonadales;f Marinomonadaceae;g ;s unknown
6_P61_P62_rhodobacteraceae	Coassembly (61 + 62)	99.7	0.7	40	243772	59.1	4474503	d Bacteria;p Proteobacteria;c Alphaproteobacteria;o Rhodobacterales;f Rhodobacteraceae;g Pseudophaeobacter;s unknown
7_P61_P62_flavobacteriaceae	Coassembly (61 + 62)	99.7	0.0	23	267233	33.0	2816890	d Bacteria;p Bacteroidota;c Bacteroidia;o Flavobacteriales;f Flavobacteriaceae;g Winogradskyella;s Winogradskyella sp002163855
8_P98_alcanivoracaceae	Single assembly (98)	95.5	0.0	25	422209	58.3	3726153	d Bacteria;p Proteobacteria;c Gammaproteobacteria;o Pseudomonadales;f Alcanivoracaceae;g Alcanivorax;s Alcanivorax sp000155615
9_P100_alcanivoracaceae	Single assembly (100)	99.6	0.7	27	1001868	58.0	4606115	d Bacteria;p Proteobacteria;c Gammaproteobacteria;o Pseudomonadales;f Alcanivoracaceae;g Alcanivorax;s Alcanivorax sp000155615
10_P100_P98_flavobacteriaceae	Coassembly (98 + 100)	90.0	0.4	64	56874	40.5	2497634	d Bacteria;p Bacteroidota;c Bacteroidia;o Flavobacteriales;f Flavobacteriaceae;g LPB0005;s unknown
11_P100_P98_maricaulaceae	Coassembly (98 + 100)	97.7	0.7	44	108979	64.1	3165928	d Bacteria;p Proteobacteria;c Alphaproteobacteria;o Caulobacterales;f Maricaulaceae;g Maricaulis;s unknown
12_P102_P105_flavobacteriales	Coassembly (102 + 105)	98.2	0.3	219	35785	44.3	3641863	d Bacteria;p Bacteroidota;c Bacteroidia;o Flavobacteriales;f UA16;g ;s unknown
13_P102_P105_crocinitomicaceae	Coassembly (102 + 105)	97.8	2.8	115	59205	41.1	3641863	d Bacteria;p Bacteroidota;c Bacteroidia;o Flavobacteriales;f Crocinitomicaceae;g UBA4466;s unknown
14_P102_P105_flavobacteriaceae	Coassembly (102 + 105)	97.9	1.7	132	29412	31.1	2365969	d Bacteria;p Bacteroidota;c Bacteroidia;o Flavobacteriales;f Flavobacteriaceae;g SCGC-AAA160-P02;s unknown
15_P102_P105_maricaulaceae	Coassembly (102 + 105)	99.5	0.0	8	627500	62.7	3372584	d Bacteria;p Proteobacteria;c Alphaproteobacteria;o Caulobacterales;f Maricaulaceae;g Maricaulis;s Maricaulis maris
16_P105_saccharospirillaceae	Single assembly (105)	99.0	1.8	45	122390	46.6	3878442	d Bacteria;p Proteobacteria;c Gammaproteobacteria;o Pseudomonadales;f Saccharospirillaceae;g Thalassolituus;s Thalassolituus oleivorans

Table 1. Metagenome assembled genome (MAG) information and statistics. Quality metrics for MAGs were determined using CheckM2 (Version 1.0.7; default parameters). The taxonomic identity of each MAG was determined using GTDB-Tk3 (Version 1.0.2) against The Genome Taxonomy Database.

The genomes containing the greatest number of *almA* + *alkB* copies belong to the genera *Alcanivorax* and *Thalassolituus*, neither of which contain key genes for catabolism of aromatic (*rhdA* - ring hydroxylating dioxygenases) or short chain (*pHMO* - particulate hydrocarbon monooxygenase subunit A, B, C) alkanes, and both of which bloomed at the North Atlantic subtropical gyre stations. We interpret these genomes to indicate a specialization in long-chain *n*-alkanes (i.e., biohydrocarbons) with an undefined upper limit on the carbon chain length. Testing for potential crossover catabolism to aromatic hydrocarbons using the approach of (González-Gaya et al., 2019), we also analyzed the

genomes for rhdA, finding copies in *Pseudophaobacter*, *Flavobacteria, Maricaulis* and *Pseudohongiellaceae* genomes. However, further analysis of protein hits for rhdA reveals that hits within our pentadecane-enriched taxa could originate from amino acid metabolism rather than aromatic hydrocarbon catabolism and point to a need to analyze rhdA hits in close detail before assuming hydrocarbon oxidation functionality (Table 2). Despite ambiguity with the function of *rhdA*, observed blooms of *n*-alkane specialists underlying the oligotrophic ocean point to a decoupling between biohydrocarbons and dissimilar petroleum hydrocarbons such as aromatics.

Metagenome	Contig#_ORF#	Assigned KO ID	KO Threshold	Hmmsearch Score	E-value	KO definition
14_P102_P105_flavobacteriaceae	c_000000000444_9	K00499	313.87	319.3	2.00E-99	choline monooxygenase
15_P102_P105_maricaulaceae	c_000000000002_229	K00499	313.87	367	6.30E-114	choline monooxygenase
4_P61_P62_pseudohongiellaceae	c_000000000337_16	K00499	313.87	173.9	2.60E-55	choline monooxygenase
6_P61_P62_rhodobacteraceae	c_000000000007_301	K00479	305.97	552.6	2.30E-170	glycine betaine catabolism
6_P61_P62_rhodobacteraceae	c_000000000010_117	K00479	305.97	269.4	2.40E-84	glycine betaine catabolism
6_P61_P62_rhodobacteraceae	c_000000000010_118	K00479	305.97	269.8	1.80E-84	glycine betaine catabolism

Table 2. Further analysis of hits for rhdA (ring-hydroxylating dioxygenase subunit A) within metagenomes. KofamScan against the Kegg database hmms was performed for each hit to PF00848 (ring-hydroxylating dioxygenase subunit A). Resulting Kegg Orthology (KO) IDs and definitions indicate rhdA hits are for genes related to amino acid metabolism and catabolism. This is explained by the seed alignment for PF00848 (rhdA) including dioxygenases for aromatic hydrocarbon catabolism as well as genes for cytochrome P450 oxidoreductase, glutathione S-transferase, choline monooxygenase, proline dehydrogenase, as well as ambiguous genes containing Rieske-type iron-sulfur clusters.

1.4.2 Genes for Alkane Activation Abundant in the Photic Zone

To further probe alkane specialization among oligotrophic microbial populations, we analyzed gene abundance from the Tara Oceans dataset. The relative abundance of *alkB*-like genes in the upper oligotrophic ocean (relative to the single copy gene, *recA*) was substantially greater than for genes involved in the activation of other hydrocarbons including C1-C5 alkanes, phenanthrene, benzene, toluene, naphthalene, xylene, cymene, and biphenyl (Supplementary Data 1), supporting a greater capacity for long-chain *n*-alkane degradation relative to other hydrocarbons.

1.4.3 Effect of Particle Additions on the Community

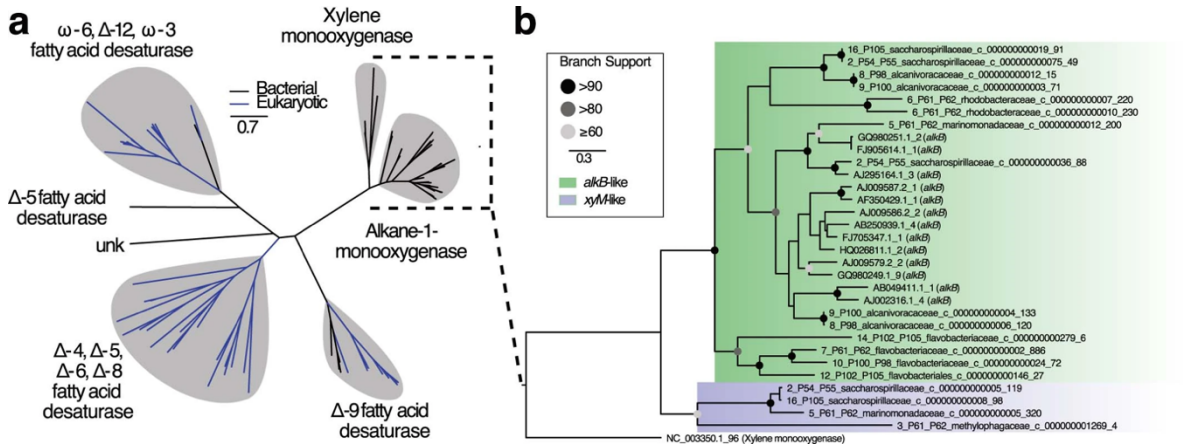


Figure 4. **a**, Maximum likelihood tree of *alkB* hits within metagenomes compared to fatty acid desaturase, xylene monooxygenase, and *alkB* functionally expressed/characterized representatives. Δ-X indicates activity X carbons from the carboxylic end of the fatty acid and ω-X indicates activity X carbons from the methyl end of the fatty acid. **b**, Expanded view of the alkane-1-monooxygenase, xylene monooxygenase, and related hits from metagenomes. Coloration in panel **b** is according to position in panel **a**. Gene copies for *alkB* in MAGS (in green) used in Fig. 2h.

The pentadecane incubations with particle treatments at both sites elicited response by a greater number of taxa compared to pentadecane only treatments. This response was quantified using Shannon diversity indices whereby (particle + pentadecane) enrichments had a higher Shannon diversity index compared to the pentadecane only treatments (two tailed t-test: Station 3, $p < 0.023$; Station 6, $p < 0.017$) (Fig. 3c). Presumably, the particles provided a diverse range of substrates beyond the added pentadecane and/or more taxa that could consume pentadecane. Nonetheless, several taxa that contain opportunistic hydrocarbon degraders were observed to bloom in these treatments including Litoribacillus (Saccharospirillaceae), Pseudophaeobacter (Rhodobacteraceae), and Aurantivirga (Flavobacteriaceae) at station 3 and Thalassolituus (Saccharospirillaceae) and Olleya (Flavobacteriaceae) at station 6. From this data, we postulate a possible association of alkane specialists with the hydrophobic phase and alkane generalists with particulate matter.

1.4.4 Natural Seeps Prime Petroleum Hydrocarbon Consumption

Results from our investigation indicate that the biohydrocarbon cycle primes the ocean for consumption of long-chain *n*-alkanes, but that this effect is at-least partially decoupled from the consumption of petroleum hydrocarbons by microbial *n*-alkane specialization. We therefore explore an alternative hypothesis, that priming for petroleum hydrocarbon degradation occurs by proximity to petroleum sources. The test of this biogeographic hypothesis is a bloom response experiment conducted at a single depth across seven stations representing a gradient of natural oil seep intensity, spanning from the seep-ridden northern Gulf of Mexico to the North Atlantic subtropical gyre (Fig. 2a). *n*-Pentane was used for these experiments as it is a model non-biogenic compound that is unique to and abundant in petroleum, a structural analog to pentadecane (linear chain with an odd number of carbons), readily bioavailable based on its higher vapor pressure and aqueous solubility, relatively-low in toxicity, and is known to partition to the ocean's interior following release from the seafloor (Ryerson et al., 2011, 2012b). We find the microbial response to *n*-pentane to be structured by proximity to seepage, with ~10X more rapid bloom onset in the Gulf of Mexico versus the North Atlantic subtropical gyre (Fig. 2c). Notably, the bloom onset for pentane underlying the North Atlantic subtropical gyre is ~9X slower than for pentadecane in the same region, albeit with experiments conducted at different depths and stations. Results demonstrate a clear biogeographic dependence on natural seepage for biodegradation of a petroleum hydrocarbon, providing another example of decoupling between petroleum versus biological hydrocarbon consumption, and pointing to source-specific priming by which the capacity for rapid consumption of a petroleum-derived hydrocarbon is defined by proximity to petroleum inputs.

1.5 Conclusion

Through our studies in the subtropical North Atlantic Ocean, we have confirmed the existence and magnitude of a cryptic hydrocarbon cycle as proposed by Lea-Smith et al., (2015). We further demonstrate a decoupling between biological alkanes and petroleum-

derived hydrocarbons that points to a complex interplay of chemical composition and biogeography that structure the Ocean's response to oil spills. Importantly, our findings are most applicable to the oligotrophic regions of the Ocean, encompassing ~40% of Earth's surface. Other oceanic regions may harbor abundant Eukaryotic phytoplankton, many of which are capable of producing hydrocarbons (Sorigué et al., 2016, 2017; Aleksenko et al., 2020). Based on our qualitative observations from productive waters on the continental shelf of the Northwest Atlantic Ocean, we expect such environments to harbor a dynamic and complex hydrocarbon cycle including biological alkanes and alkenes, structured in-part by proximity to continental sources and interaction with the sea floor. Cryptic hydrocarbon cycling and its relationship to biogeographic structuring of microbial populations represents an important factor in understanding the metabolic response capacity of the oceanic microbiome to oil inputs and should be incorporated as a predictive tool in oil spill response planning.

Chapter 2

2. Genomic Evidence for Marine Group II Euryarchaeota Consumption of Biogenic Alkanes

While analyzing data for the (Love and Arrington et al., 2021) manuscript we realized most biogenic alkane consumption likely occurs in the photic zone and not in the mesopelagic where our experiments were conducted. At sea, we were limited to mesopelagic incubations due to the nature of our experimental design, which relied on ambient nutrient availability to enable a robust growth of microbes to observe via a large respiratory signal. This chapter encompasses my work to use publicly available genomic datasets to begin to fill in our knowledge gap regarding alkane degradation in the oligotrophic euphotic zone. By analyzing the phylogenetic placement of putative hits to alkane-1-monooxygenase from multiple Tara Oceans stations I noted the presence of an "unknown" clade at each surface (5 meter) and deep chlorophyll maximum station. The homology search for close representatives to sequences within this unknown clade indicated all sequences belonged to the phylum Euryarchaeota. As I began to look for more evidence that these archaeal *alkB*-like sequences could be used to consume biogenic alkanes I generated the remaining data for this chapter. Aspects of this chapter (Figure 7 and 8, and some discussion) were included in the publication (Love and Arrington et al., 2021).

2.1 Introduction

Biogenic hydrocarbons (or cryptic microbial hydrocarbons) originate from cyanobacteria and other microalgae in the form of long-chain alkanes and alkenes, and are released into the ocean in a massive and rapid cycle that greatly exceeds all oceanic petroleum inputs (Lea-Smith et al., 2015; Love and Arrington et al., 2021). Our recent study constrained this global hydrocarbon cycle in oligotrophic waters with stock measurements throughout the water-column, production experiments in the sunlit ocean, and consumption

studies in the mesopelagic (Love and Arrington et al., 2021). However, our analysis of the export flux of biogenic alkanes on sinking particles collected *in situ* beneath the euphotic zone indicates only ~$1x10^{-4}$% of pentadecane production was exported below 150 m, highlighting a major gap in our understanding of the 131-649 Tg of biogenic hydrocarbons consumed in the oligotrophic photic zone annually. In this chapter, I use bioinformatic approaches to elucidate which microorganisms encode genes for alkane degradation in the photic zone of the North Atlantic Ocean.

2.1 Background

2.1.1 Hydrocarbon Degradation Across the Tree of Life

The ability to use hydrocarbons as the sole or major source of carbon is present in all three domains of life, including approximately 320 bacterial and 115 fungal genera (Head et al., 2006; Hazen et al., 2016; Prince et al., 2019a). Approximately 18 genera of archaeal isolates currently use hydrocarbons as major or sole sources of carbon, with smaller taxonomic diversity likely related to obstacles with cultivation techniques (Prince et al., 2019b, 2019a). Major environmental controls on hydrocarbon degradation include temperature, salinity, pressure, radiation and desiccation, with various reviews summarizing these findings (Atlas, 1981; Leahy and Colwell, 1990; Van Hamme et al., 2003; Head et al., 2006; Hazen et al., 2016; Scoma et al., 2019). Despite the considerable amount of literature on microbial hydrocarbon degradation, until recently we have had limited knowledge about which microorganisms are environmentally relevant during pollution events. With the advent of novel cultivation-independent approaches in stable isotopes and next generation sequencing we have significantly increased our breadth and depth of understanding of hydrocarbon degrading taxa over the last decade.

2.1.2 Kinetic Theory of Dilute Alkanes

Many planktonic chemoheterotrophic organisms regulate the concentrations of dissolved biogenic organics down to nanomolar concentrations in the ocean (Button, 1985;

Button and Jüttner, 1989) until concentrations decrease to levels that are sufficient for growth of only organisms with good transport systems, or organisms that can grow very slowly. We sought to understand the extent dilute pentadecane concentrations could sustain obligate hydrocarbon degraders in the photic zone. Assuming a pentadecane to biomass conversion of 5-50% and a carbon mass of hydrocarbon consuming microbes of 120 fg C $cell^{-1}$ (Valentine et al., 2012) pentadecane in the lower photic zone (1% PAR) would support microbial production of the order of ~10-100 cells ml^{-1} d^{-1} and the upper photic zone (30% PAR) could support ~2-20 cells ml^{-1} d^{-1} (Love and Arrington et al., 2021). The size of the supported community further depends on cellular turnover time, and an assumed turnover rate of 0.1 day^{-1} equates to a steady state population of ~10^2 -10^3 cells ml^{-1} in the lower photic zone (1% PAR) and ~20-200 cells ml^{-1} in the upper photic zone (30 % PAR) of the oligotrophic ocean.

We further sought to understand whether the collision frequency of pentadecane molecules to microbial cells could sustain growth of free-living obligate degraders at ecologically relevant cell sizes and rates. Using a theoretical model for molecular diffusion and microbial growth developed after work by (Berg and Purcell, 1977; Button et al., 1998) on the organism *Cycloclasticus* and the substrate toluene. This work assumes the cells are perfect spheres and they sequester every hydrocarbon molecule that reaches its surface.

$$\text{Doubling Time} = \text{m} \cdot \text{Z} \cdot \frac{1}{\text{J}} \qquad \text{(Equation 1)}$$

Where m is the dry mass of cells of radius *r*, Z is the number of molecules of hydrocarbon required to double per cell dry mass, and J is the flux of molecules into the cell per unit time.

$$\text{Cell Dry Mass (m)} = 717 \ \frac{\text{fg}}{\mu\text{m}^3} \ \text{r}^3, \qquad \text{(Equation 2)}$$

where r *Cycloclasticus* cell radius in μm

$$\text{Z}_{\text{toluene}} = \frac{3.65 \text{ x } 10^{\ 9} \text{ molecules toluene}}{40.1 \text{ fg}},$$

values for *Cycloclasticus* and toluene (Button et al., 1998)

$$Z_{\text{pentadecane}} = \frac{1.47\text{x } 10^{\,9} \text{ molecules}}{40.1 \text{ fg}}$$

The free-energy yield from toluene is -3,653 kJ mol^{-1}, whereas for pentadecane it is greater at -9,102 kJ mol^{-1}, therefore we expect 2.49x less pentadecane molecules required for doubling.

$$\text{Flux (J)} = 4\pi rDC\left(\frac{2.451\text{x}10^{13}\text{molecules pentadecane}\cdot\text{L}}{\mu\text{m}\cdot\mu\text{g}\cdot\text{cm}^2}\right), \qquad \text{(Equation 3)}$$

Where r is the cell radius in μm, D is the diffusion constant (described below), and C is the concentration far from the cell in μg L^{-1}.

$$\text{Diffusion Coefficient (D)} = 8.6\text{x}10^{-6}\,\frac{\text{cm}^2}{\text{sec}},$$

We assume here the diffusion coefficient is the same for toluene and pentadecane (Oelkers, 1991; Button et al., 1998).

$$\text{Doubling Time} = 717\,\frac{\text{fg}}{\mu\text{m}^3}\,\text{r}^3\,\text{x}\,\frac{1.47\text{x } 10^{\,9} \text{ molecules}}{40.1 \text{ fg}}\,\text{x}\,\frac{1}{\text{rC}\cdot 2.65\text{x}10^{9}\,\frac{\text{molecules}\cdot\text{L}}{\text{sec}\cdot\mu\text{m}\cdot\mu\text{g}}}$$

$\text{Doubling Time} = \frac{9.87\,\text{r}^2}{\text{C}}$, r is radius in μm and C is concentration in μg L^{-1}.

Pentadecane concentrations reach a maximum of 0.08 μg L^{-1} in the Gulf stream and reach 0.13 μg L^{-1} for a *Synechococcus* bloom (Love and Arrington et al., 2021). From the results of our theoretical diffusion model, we expect free-living degraders of pentadecane with doubling times between 0.5-2 day^{-1} likely use other sources of carbon and energy coupled to pentadecane consumption. If an organism were to use biogenic pentadecane as a sole source of carbon and energy, we expect the cell radius to be small (i.e., 0.35μm) and have a slow doubling time of ~13 days. If an organism were to be an obligate degrader of pentadecane we would expect the low but sustained production of pentadecane to act as a selective pressure for a small cell size and low carbon content. Another ecological strategy

for an obligate degrader could be to spatially exist on hydrophobic phases such as necromass or particles where pentadecane may accumulate to higher concentrations per volume of seawater.

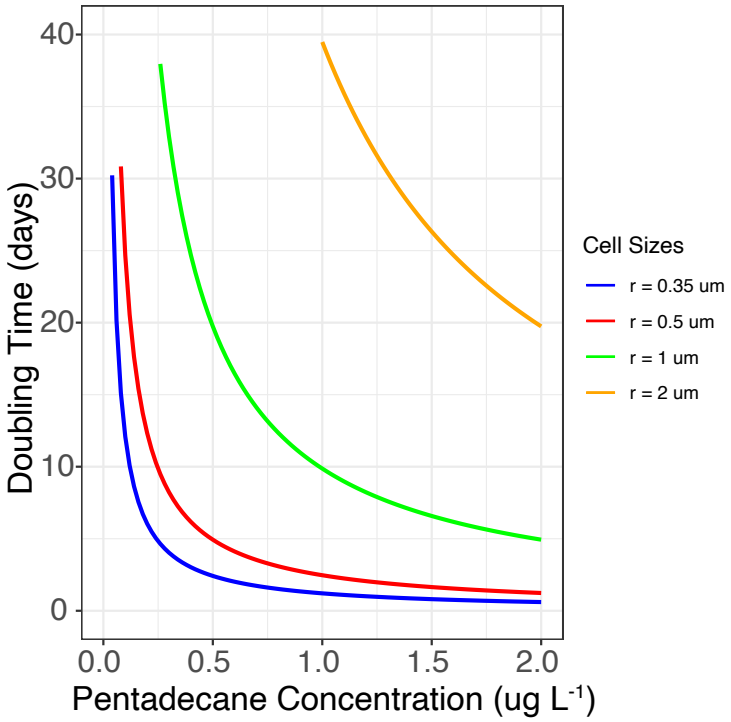


Figure 5. Theoretical relationship between doubling time and substrate concentration for cell radius of a perfect sphere.

2.1.3 Metabolism of Dilute Substrates

Much of our knowledge of alkane biodegradation comes from studies on cultivated isolates, which introduces a bias toward organisms that can form colonies on agar and that excel at conditions of high nutrients and high substrate loading – a bias against the conditions common in the oligotrophic ocean. This bias towards pure culture hydrocarbon metabolism causes us to underestimate the diversity of organisms that govern the fate of dilute biogenic alkanes in the environment in multiple ways. First, it is becoming clear that auxotrophic organisms may grow and metabolize biogenic alkanes while living dependent on interspecies cooperation or transfer of essential precursors and metabolites (Kanaly et al., 2002; Allen and Banfield, 2005; Brenner et al., 2008; McGenity et al., 2012; Hays et al.,

2015). Second, organisms may co-metabolize compounds (a strategy often utilized in the bioremediation field) whereby a microorganism uses a non-specific enzyme to oxidize an alkane without using it as a growth-substrate, while sustaining its own growth by assimilating a different substrate. Co-metabolism provides no apparent benefit to the microorganism yet enables biodegradation of hydrocarbons at concentrations too low to solely sustain a consumer, and is particularly relevant in the case of hydrocarbon monooxygenases which are strong oxidizers that can have activity on over 300 substrates (Hazen, 2010; Nzila, 2013; Hazen et al., 2016). Finally, active hydrocarbon degraders in the oligotrophic environment may simultaneously metabolize other substrates such as organic acids, lipids, or amino acids as major sources of carbon and use dilute alkanes as a supplemental source of carbon and energy.

2.1.4 Fatty Acid and Alkane Metabolism

As an ecological strategy, fatty acid metabolism coupled to biogenic alkane metabolism is a rational tactic for multiple reasons. First, the metabolic pathways for fatty acid and medium-to-long chain alkane incorporation into central carbon metabolism overlap in their use of hydrophobic transporters belonging to the FadL family (in bacteria) and ABC (ATP-binding cassette) transporters in archaea, and in the utilization of beta-oxidation (Van Den Berg, 2005). These compounds also spatially co-exist because biogenic alkanes tend to reside within the biological membranes of cyanobacteria or to adsorb to biomass/necromass (Love and Arrington et al., 2021). This is further supported by work done by (Lea-Smith et al., 2015, 2016), and the low solubility of straight-chain hydrocarbons 15–17 carbons in length.

2.1.5 Enzyme Classes Involved in the Oxidation of Alkanes

The initial terminal hydroxylation of *n*-alkanes in aerobic environments can be carried out by enzymes belonging to multiple different families. Each enzyme uses oxygen as a reactant, and overcomes the low chemical reactivity of hydrocarbons by generating

reactive oxygen species (van Beilen et al., 2003; Rojo, 2010). Bacterial strains consuming medium-chain alkanes (~C_5-C_{16}) often encode multiple copies of enzymes for alkane-1-monooxygenase (*alkB*), cytochrome P450 belonging to the CPY153 family (CYP), or flavin-binding monooxygenases (*almA*). One of the most well studied model isolates for alkane oxidation, *Alcanivorax dieselolei* strain B5, has been the subject of numerous studies (Liu and Shao, 2005; Qiao and Shao, 2010; Liu et al., 2011). Strain B5 actively degrades numerous alkanes (C_6–C_{36}), including branched alkanes and encodes a complex alkane monooxygenase system that includes two membrane-bound non-heme iron alkane hydroxylases (*alkB*), one cytochrome P450 (CYP) and a putative flavin-binding monooxygenase (*almA*) (Wang and Shao, 2014). Each enzyme underwent heterologous expression and showed differential expression with *alkB1* and alkB2 responding to C_{12}–C_{26}, the P450 to C_8–C_{16}, and *almA* was upregulated in the presence of C_{22}–C_{36} alkanes (Liu et al., 2011).

A large body of work exists on the functional analysis of the *alkB* enzyme family across multiple model organisms (i.e., *Pseudomonas putida*, *Pseudomonas aeruginosa* PAO1, *Pseudomonas fluorescens* CHA0, *Alcanivorax borkumensis* AP1, *Mycobacterium tuberculosis* H37Rv, *Acinetobacter* sp. ADP1, and *Prauserella rugosa* NRRL B-2295) as well as the ancestrally related proteins xylene monooxygenase and fatty acid desaturase (in eukaryotes and bacteria). Together, this body of work makes the *alkB* enzyme class more tractable to comparative genomics and is the focus of this chapter. Importantly, there is a major gap in our knowledge of aerobic hydrocarbon degrading archaea. None of the archaeal strains reported to use hydrocarbons as sole or major sources of carbon currently exist in publicly accessible culture collections and no genomic studies were reported on them as well (Oren, 2019). Previous work on genomic surveys of *alkB* exist. In one study, 3,979 microbial genomes and 137 metagenomes from terrestrial, freshwater, and marine environments were surveyed for the

presence of *alkB* in archaea, finding *alkB* to be absent from the domain (Nie et al., 2014). This study was likely flawed due to the narrow nature of the custom Hidden Markov Model created to scan for *alkB*, which was created based off an alignment of only 28 experimentally validated bacterial *alkB* sequences.

2.2 Methods

2.2.1 Tara Oceans *AlkB* Analysis

For select Tara Oceans stations underlying the North Atlantic subtropical gyre (Fig. 6) we conducted a survey of the diversity of alkane-1-monooxygenase (*alkB*) genes. First, we took the assembled Tara Oceans data (see http://ocean-microbiome.embl.de/companion.html) and used Prodigal (v.2.6.3; default parameters) to identify open reading frames and translate each gene to amino acid sequences (Hyatt et al., 2010; Sunagawa et al., 2015). We then used multiple methods to scan for alkane-1-monooxygenase (*alkB*): we used both HMMER3 against the Pfam database (PF00487) and KofamScan against the KEGG/KO HMM (Hidden Markov Models) profiles (K00496) (Eddy, 2011; El-Gebali et al., 2019; Aramaki et al., 2020). Importantly, both methods use Hidden Markov Models which encompass fatty acid desaturase, *alkB*, and xylene monooxygenase due to their ancestral nature and shared catalytic motifs. By using this method to scan for potential hits we are broadly scanning for this protein family, not just *alkB.* Each putative hit

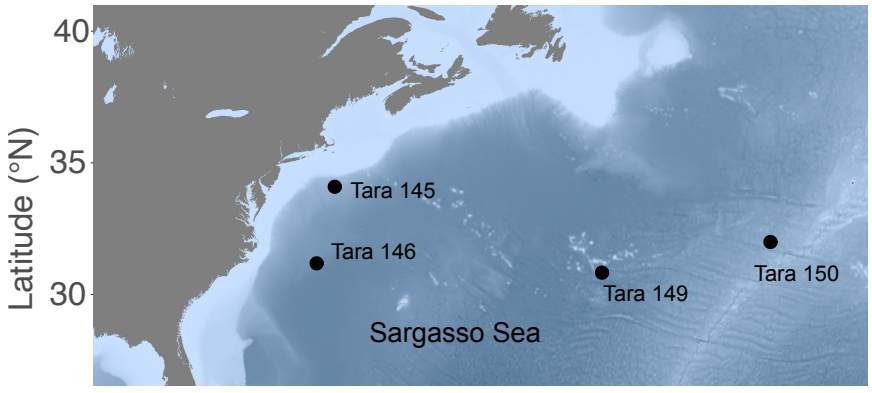


Figure 6. Tara Oceans stations analyzed in this chapter. Depths included the surface (5m), deep chlorophyll maximum, and the mesopelagic (590-740m). The size fractions analyzed were 0.22-3µm.

was then manually curated using Geneious Prime v.2019.2.3 (https://www.geneious.com) to search for the eight-histidine residues considered catalytically essential for function (Shanklin et al., 1994b). This manual refinement excluded many sequences that contained 5-7 histidine residues, and future work is needed to determine whether these sequences can also perform the function of interest. To differentiate between *alkB,* xylene monooxygenase, and fatty acid desaturase, we found it necessary to phylogenetically analyze each hit to determine relation to *alkB* or fatty acid desaturase. Through this method we learned that the HMMER3 method with Pfam ID PF00487 identifies more putative hits for *alkB* in a given environmental sequence than KofamScan with K00496; however, those additional hits were generally more closely related to fatty acid desaturases than *alkB*. Regardless of the initial scanning method, the phylogenies produced are necessary to determine similarity to the related xylene monooxygenase protein which acts on the methyl groups of xylene. In total, we used KofamScan with K00496 to search for *alkB*, manually curated the results to ensure presence of eight-histidine residues essential for function and phylogenetically analyzed each hit for relation to *alkB* compared with fatty acid desaturase and xylene monooxygenase.

Following phylogenetic analysis, each curated hit was assigned a taxonomic classification through homology search using BLAST (v.2.7.1) against the nr database (v.38 accessed December 2019) (Camacho et al., 2009). Read mapping of high-quality reads from each respective station using Bowtie2 (v.2.3.4.1,--very sensitive preset) was used to determine the abundance of each unique *alkB*-like protein at each station (Langmead and Salzberg, 2012).

2.2.2 Analysis of Marine Group II Archaea Genomes

We analyzed 268 metagenomes of Marine Group II Archaea (MGII), each genome was originally gathered by (Rinke et al., 2019). To scan for the presence of *alkB* we used the same general method as the Tara Oceans analysis: briefly, we used the program

KofamScan against the KEGG/KO HMM profile (K00496) for the monooxygenase/fatty acid desaturase enzyme family, manually refined the sequences for histidine residues, and phylogenetically analyzed each putative hit alongside reference sequences for bacterial *alkB,* fatty acid desaturase, and xylene monooxygenase. To further assess the presence of *alkB* across the phylogenetic diversity of MGII we constructed a ribosomal protein tree based on 122 concatenated archaeal marker genes as in (Rinke et al., 2019).

We also searched for other genes in the metabolic pathway of alkane consumption with both HMMER3 and Pfam, as well as Kofamscan and KEGG (Eddy, 2011; El-Gebali et al., 2019; Aramaki et al., 2020). The following HMM models were used to scan for the function of interest: PQQ-dependent alcohol dehydrogenase (PF13360), aldehyde dehydrogenase (PF00171), acyl-CoA synthase (FadD; K01897), acyl-CoA dehydrogenase (FadE; K00249), enoyl-CoA hydratase/3-hydroxyacyl-CoA dehydrogenase (FadB; PF00378), and acetyl-CoA C-acyltransferase (FadA; K00632 and K00626). In our broader analysis of alkane metabolism, we deviate from the alkane degradation pathway in the KEGG database in two ways: first, the alcohol dehydrogenase encoded by MGII is likely a $NAD(P)^+$ independent pyrroloquinoline (PQQ) alcohol dehydrogenase, a common metabolic strategy among alkane degraders (Rojo, 2010) and, second to search for the FadB/YfcX protein we used an annotation method with Pfam (PF00378) for the crotonase family. This broader search for the crotonase family was performed because previous methods to scan for FadB/YfcX using the KEGG identifier (K01692) and (K07516) have shown these genes are missing from many MGII genera (Tully, 2019). Authors of this previous work note the absence of FadB/YfcX could be the result of uncharacterized family-specific analogs of this function and due to the abundance of genes belonging to the crotonase family in MGII this is likely the case.

2.2.3 Phylogenetic Trees

Two separate phylogenetic trees were created for *alkB* and the archaeal concatenated marker genes. In each case, protein sequences were aligned using MUSCLE (Edgar, 2004) (v.3.8.425). All columns with >95% gaps were removed using TrimAL (Capella-Gutiérrez et al., 2009) (v.1.2). Phylogenetic analysis of concatenated alignments were performed with RAxML (Stamatakis, 2014) (v.8.2.9; parameters: raxmlHPC -T 4 -s input -N autoMRE -n result -f a -p 12345 -x 12345 -m PROTCATLG). Resulting trees were visualized using (FigTree) (v.1.4.3).

2.3 Results and Discussion

2.3.1 *AlkB* Diversity and Abundance in the North Atlantic

A detailed analysis of the Tara Oceans dataset in the North Atlantic reveals *alkB*-related genes are abundant in the surface ocean and DCM and are phylogenetically distinct from the related delta-9 fatty acid desaturase (Fig. 7). We observe diverse bacterial classes belonging to Alphaproteobacteria, Gammaproteobacteria, Flavobacteria, Bacteroidetes, and Chlorobi encode *alkB*-related genes (Supplementary Data 2). The most abundant bacterial *alkB* hits belong to a newly proposed phylum *Candidatus Marinimicrobia* (Hawley et al., 2017). Many well-studied obligate hydrocarbon degraders are infrequently represented among these data (i.e., *Cycloclasticus, Oleiphilus, Oleispira, Thalassolituus* genera are absent and only two *Alcanivorax* sequences detected at one station) (Supplementary Data 2).

Notably, a dominant clade of *alkB*-like monooxygenases belongs to the globally abundant Marine Group II (MGII) archaea (with possible occurrence also in Marine Group III archaea) and is consistently present in all surface and DCM stations (Fig. 7, Fig. 8, Supplementary Data 2). This highly successful group of surface-ocean-dwelling archaea is known for a chemoorganoheterotrophic lifestylee targeting lipids, proteins and amino acids (Rinke et al., 2019; Tully, 2019) and can utilize photoheterotrophy, but a key role in biohydrocarbon cycling was unexpected as MGII archaea are not among the ~300 genera

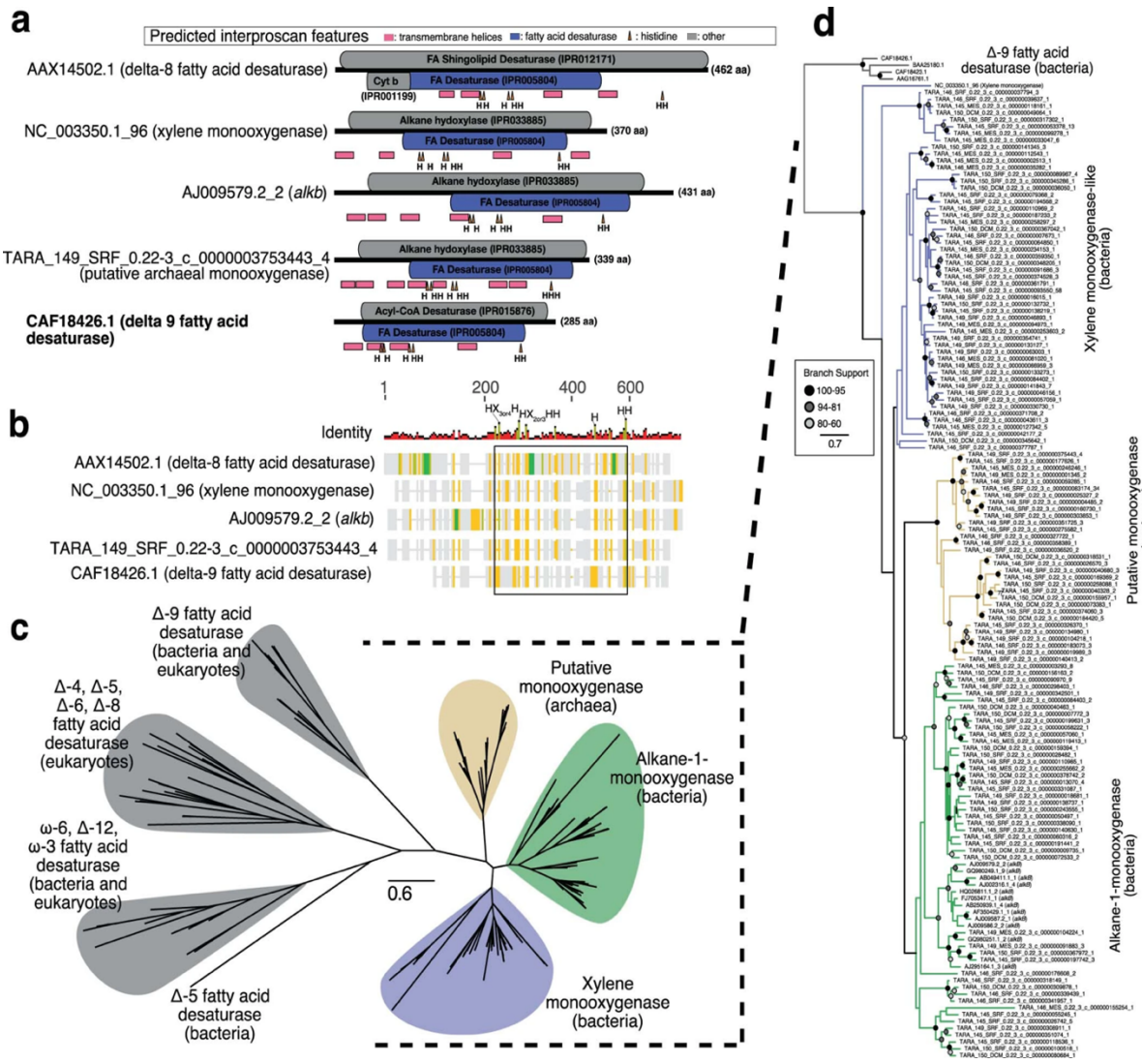
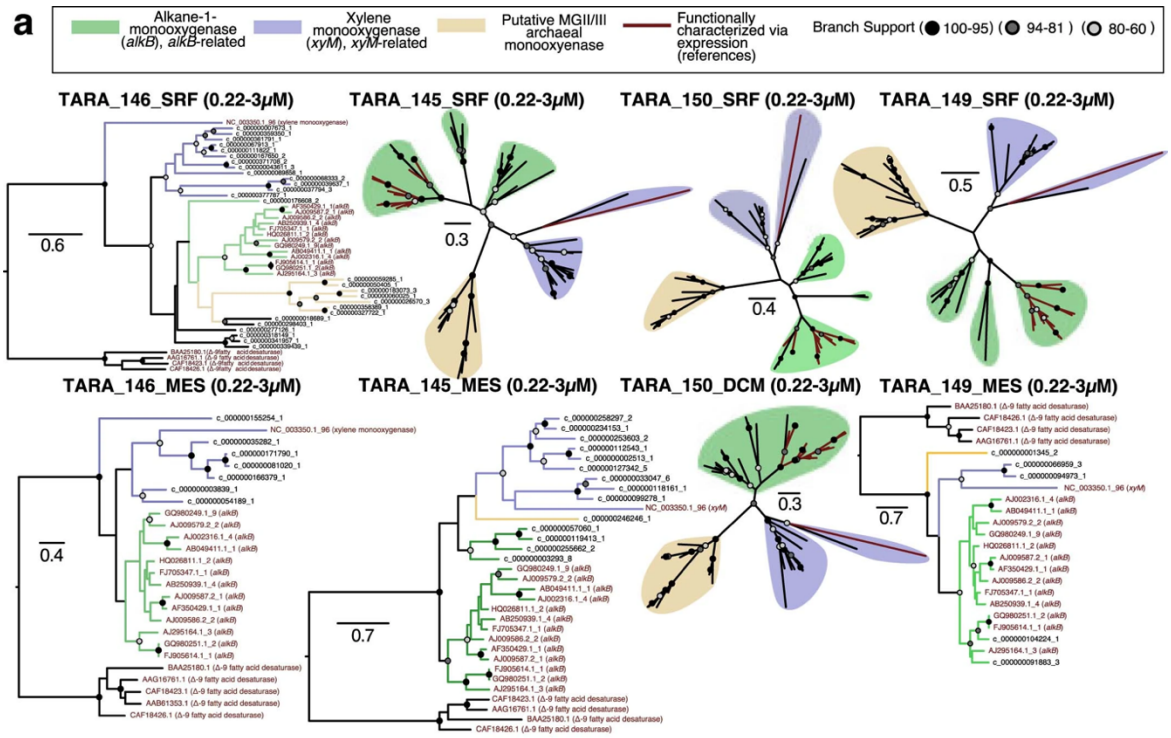


Figure 7. Phylogenetic analysis of genes closely related to *alkB* from Tara Ocean dataset reveal bacterial and archaeal clades distinct from xylene monooxygenase and fatty acid desaturases. **a**, Protein domain architecture across select representatives of xylene monooxygenase, alkane-1 monooxygenase, fatty acid desaturase, and related proteins from the Tara Oceans dataset which share a core fatty acid desaturase-like region (blue) expanded on in panel **b**. **b**, Abbreviated protein alignment for phylogenetic analyses (for details see Methods). Each column of alignment figure represents a sliding window of 5 bp with the following identity to consensus sequence coloration: green (100%), mustard (80-99% similar), yellow (60-79% similar), gray (<60% similar). The black box represents the region containing the eight histidine residues considered catalytically essential which were used for phylogenetic analyses in panels **c**-**d**. **c**, Maximum-likelihood phylogenetic tree with scale bar of substitutions per site. For clarity, bootstrap values are not shown for the full tree. Δ-X indicates activity X carbons from the carboxylic end of the fatty acid and ω-X indicates activity X carbons from the methyl end of the fatty acid. **d**, Expanded subtree of membrane monooxygenases and delta-9 fatty acid desaturases (outgroup). Clade coloration in panel **d** is according to position in panel **c**. NCBI accession codes are given for functional representatives in the subtree (accession_ORF#).



b

Sample ID	Latitude (degrees North)	Longitude (degrees East)	Sampling depth (m)	Abundance (% coverage)	Diversity (# unique)
TARA_146_MES_0.22-3	34.6663	-71.2907	640	0.0	0
				100.0	7
				0.0	0
TARA_146_SRF_0.22-3	34.6712	-71.3093	5	27.6	7
				53.7	13
				37.4	7
TARA_149_MES_0.22-3	34.0771	-49.8233	740	34.2	1
				36.4	2
				29.5	3
TARA_149_SRF_0.22-3	34.1132	-49.9181	5	39.0	15
				44.8	12
				16.2	8
TARA_145_MES_0.22-3	39.2392	-70.0343	590	8.0	1
				65.8	9
				26.2	4
TARA_145_SRF_0.22-3	39.2305	-70.0377	5	34.5	9
				24.8	14
				40.7	15
TARA_150_DCM_0.22-3	35.8427	-37.1526	40	40.7	14
				28.7	10
				30.6	13
TARA_150_SRF_0.22-3	35.9346	-37.3032	5	35.5	9
				30.0	9
				38.5	11

Clade Type
Archaea
Xylene
AlkB

Figure 8. a, Maximum-likelihood phylogenetic analysis for each station with scale bar of substitutions per site. Clade designations as follows: green (alkane-1-monoxygenase representatives and related Tara hits), blue (xylene monooxygenase representative and related Tara hits), yellow (putative Marine Group II/III archaeal monooxygenase). See Supplementary Data 2 for homology search results for putative MG II/III monooxygenase hits. Trees <27 unknown Tara sequences are out-grouped with delta-9 fatty acid desaturases, whereas trees with >27 unknown Tara sequences are left unrooted. **b**, Meta-data for each Tara station and abundance of unique hits derived from read mapping. % Coverage indicates the fraction of reads that map to genes within each clade (xylene monooxygenase, *alkB*, or archaeal monooxygenase) over the total reads mapped to all *alkB*-like, xylene-like, and archaeal monooxygenases found at each station.

of bacteria and archaea previously identified as hydrocarbon degraders (Prince et al., 2019b, 2019a).

Archaea are ubiquitous and abundant members of marine water-column communities, yet their roles in biogeochemical cycles of carbon, nitrogen, phosphate, and sulfur are only beginning to be understood. The planktonic archaea comprise three major phylogenetic groups, each with distinct physiology and ecology: Marine Group I Thaumarchaeota (MGI), Group II Euryarchaeota (MGII), and Group III Euryarchaeota (MGIII) (DeLong, 1992; Fuhrman and McCallum, 1992). Marine Group II and Marine Group III have recently been proposed as *Candidatus* Poseidoniales and Pontarchaea, however for this chapter we maintain the use of MGII and MGIII (Adam et al., 2017; Rinke et al., 2019). Only one group, the marine Thaumarchaeota (MGI), have cultivated representatives, limiting our understanding of the functional role for other archaea in the ocean (Könneke et al., 2005; Qin et al., 2014).

Advances in genomic sequencing, in-situ measurements, and computational approaches are revolutionizing our understanding of planktonic archaea. MGII archaea have been found spanning from polar to tropical ocean regimes (Bano et al., 2004; Galand et al., 2010). MGII are often detected in surface sunlit waters (comprising up to 30% of the prokaryotic community), are consistent members of the microbial community present within the deep chlorophyll maximum (Orsi et al., 2015), and generally decrease in abundance beneath the photic zone (Massana et al., 1997; DeLong et al., 2006), yet have been observed in deeper mesopelagic regions (Parada and Fuhrman, 2017). Another feature of MGII is that they often bloom coincident with or just after phytoplankton blooms (Galand et al., 2010; Hugoni et al., 2013; Needham and Fuhrman, 2016).

2.3.2 Evidence for Alkane Utilization in MGII

Our analysis of 268 MGII metagenomes (MAGs) originally gathered by Rinke et al., 2019 suggests that 83% encode homologs to alkane-1-monooxygenase (*alkB)* spanning

every genus within the taxonomic order (Fig. 9). Expanding beyond *alkB*, MGII also encode the beta-oxidation genes necessary to shunt alkane-derived carboxylic acids into the tricarboxylic acid cycle (Fig. 10). Recent metatranscriptomic work confirmed the expression of MGII *alkB*-like genes in coastal waters offshore Georgia (Damashek et al., 2021). I note within the MGII lineage, some subclades exhibit distinct biogeographic ranges, depth,

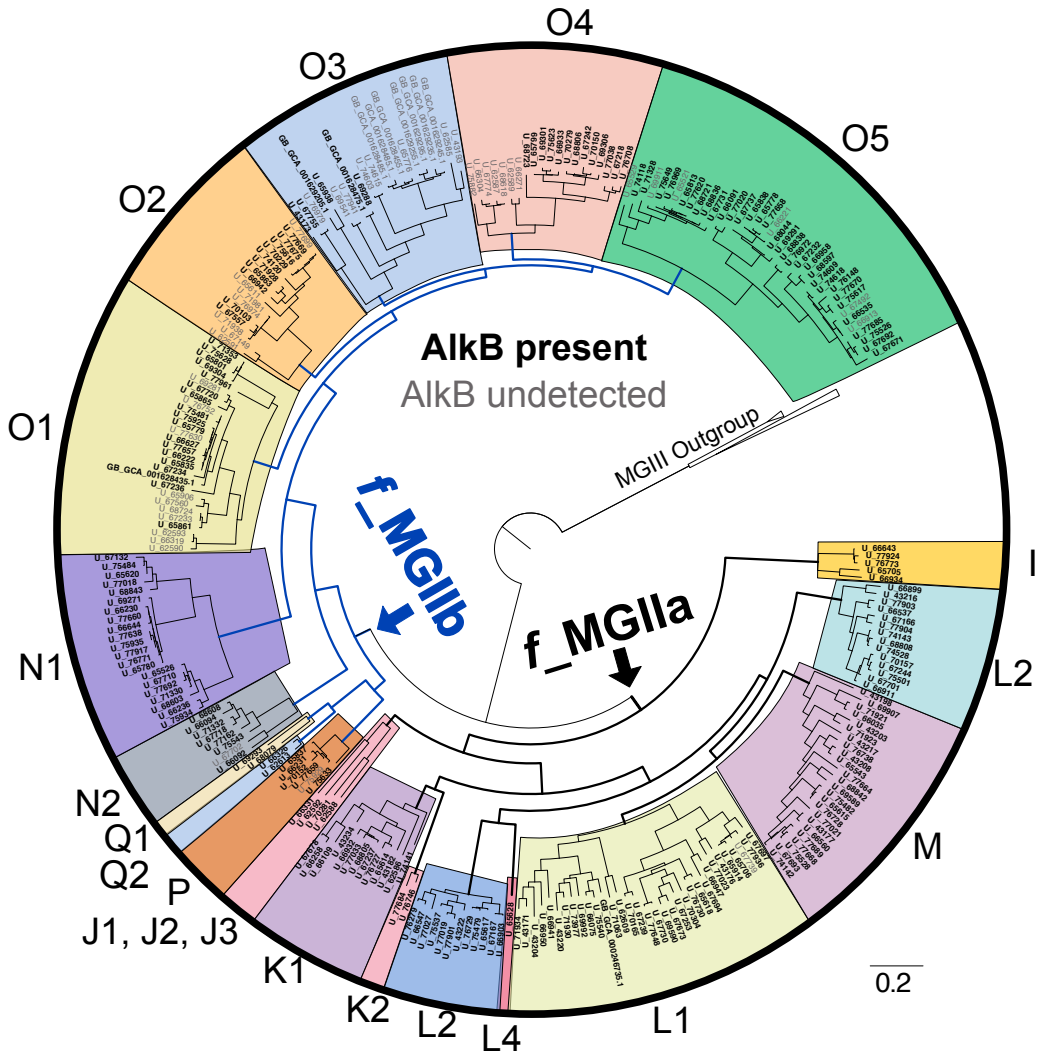


Figure 9. Prevalence of *alkB* homologs among the MGII taxonomic order. This phylogenomic tree is based on 122 concatenated archaeal marker gene alignments and modeled after Rinke et al., 2019. Each tip label represents a single MAG, bold black typeface indicates at least one *alkB*-like gene was detected, gray indicates none were detected. In the larger tree topology, the entire tree represents the MGII order, with the families MGIIb and MGIIa highlighted in blue and black line colors, each highlighted clade represents a single genus. The outgroup is MGIII. The scale bar (lower right corner) represents 0.2 substitutions per site.

specialization, and nutrient preferences, however with respect to alkane degradation all MGIIs are of interest because of the widespread conservation of these genomic features.

2.4 Conclusion

The coincidence of efficient pentadecane recycling, euphotic zone niche specialization for MGII archaea and the high prevalence of MGII archaeal *alkB*-like monooxygenases in oligotrophic surface waters collectively point to a potentially important role for MGII archaea in biohydrocarbon cycling. A capacity for MGII to consume *n*-alkanes, coupled with an inability to outcompete bacteria in a spill scenario, is consistent with existing theories of archaea's ecological specialization (Valentine, 2007) and, if correct, provides a further example of decoupling from petroleum, given the lack of reported MGII in oil spill bloom response.

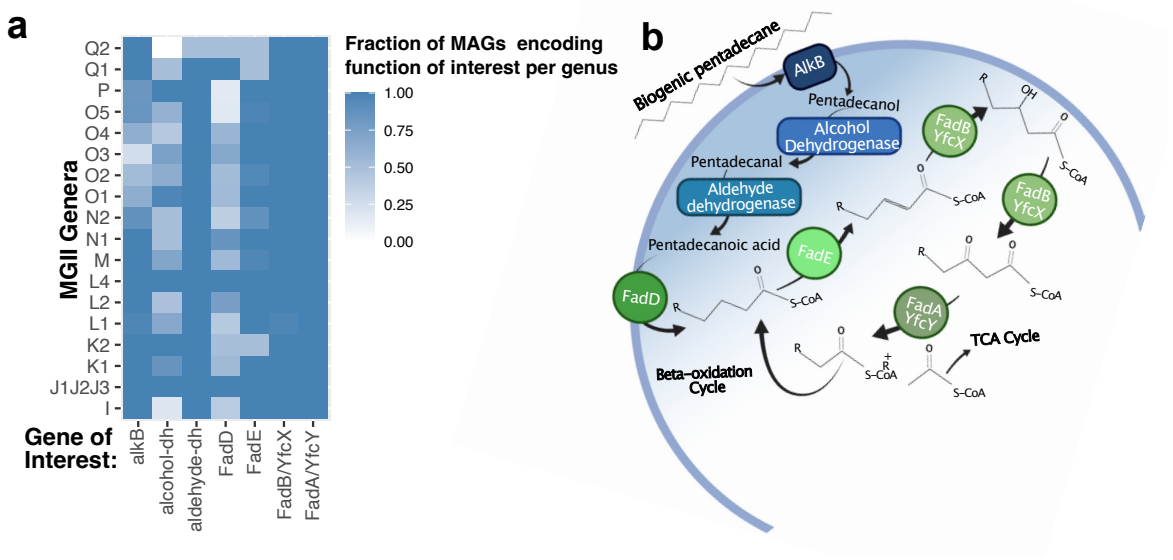


Figure 10. a Alkane metabolism encoded among MGII archaea. Heatmap is scaled from 0 to 1, where 1 indicates all MAGs within the designated genus possess the gene of interest. **b** Model of cell cytoplasmic membrane whereby blue enzymes indicate alkane specific pathway: *alkB* (alkane-1-monooxygenase; PF00487), alcohol-dh (PQQ-dependent alcohol dehydrogenase; PF13360), aldehyde-dh (aldehyde dehydrogenase; PF00171). Beta-oxidation (lipid) specific genes in green: FadD (acyl-CoA synthase; K01897), FadE (acyl-CoA dehydrogenase; K00249), FadB (enoyl-CoA hydratase/3-hydroxyacyl-CoA dehydrogenase; PF00378), FadA (acetyl-CoA C-acyltransferase; K00632 and K00626).

Chapter 3

3. Hydrocarbon Metabolism and Petroleum Seepage as Ecological and Evolutionary Drivers for *Cycloclasticus*

This chapter is in the final stages of editing (with co-author input) before journal submission. The *n*-pentane respiration data from the GOM was included in Chapter 1 and within the (Love and Arrigton et al., 2021) manuscript and therefore is only mentioned in the background as it applies to our observations of the *Cycloclasticus* ecotypes. This chapter focuses on the microbiology and metabolism we observe within the *Cycloclasticus* organisms that bloom on *n*-pentane.

3.1 Abstract

Petroleum from natural seeps and spills typically includes abundant aqueous-soluble hydrocarbons that can dissolve to the ocean's interior and structure the metabolic response of deep-sea microbial populations at a regional scale. This study identifies two distinct free-living *Cycloclasticus* strains that bloom in response to *n*-pentane, with blooms seemingly differentiated by proximity to natural seepage within the deep Gulf of Mexico. *n*-Pentane is a seawater-soluble, volatile compound that is abundant in petroleum products and reservoirs and can partition to the deep-water column following release from the seafloor. Using both 16S rRNA amplicon sequencing and comparative genomics from experimental treatments we explore the ecology and niche partitioning of these two *Cycloclasticus* ecotypes and distinguish them as an open ocean variant and a seep-proximal variant, each with distinct capabilities for hydrocarbon catabolism. Metagenomic analysis indicates the open ocean adapted variant encodes more general pathways for hydrocarbon consumption including short-chain alkanes, aromatics, and longer-chain alkanes and possess pathways for respiratory nitrate reduction, thiosulfate oxidation, and nitrite oxidation; in contrast, the seep variant specializes on short-chain alkanes and relies on aerobic metabolism. Metagenomic reconstruction of 14 other *Cycloclasticus* genomes

along with publicly available sequences reveals distinct strategies in hydrocarbon metabolism for each major clade and lays an improved foundation to predict microbial response to deep sea petroleum spills.

3.2 Introduction

The abundant supply of hydrocarbons from subseafloor reservoirs fuels the global economy and their recovery as fuel products plays a pervasive role in modern society. Marine petroleum inputs, on the order of 1.3 Tg per year, range from catastrophic spills to chronic discharges, posing an array of environmental risks for ecologically sensitive areas and playing an important role in biogeochemical cycles (Oil in the Sea III, 2003). The introduction of oil and natural gas to the water-column stimulates development of hydrocarbon-degrading microbial communities, which in turn undergo complex ecological transformations and sequentially consume components of oil and gas substrates. The development of the hydrocarbon degrading community with time is a function of the composition of the source crude oil, interactions with viruses and predators, as well as environmental parameters such as nutrient availability, temperature, and salinity (Atlas, 1981; Leahy and Colwell, 1990; Van Hamme et al., 2003; Head et al., 2006).

3.3 Background

3.3.1 Water-Soluble Hydrocarbons in the Deep Ocean

As a scientific community our understanding of hydrocarbon biodegradation has been focused on liquid phase oil in the form of slicks existing at the ocean-atmosphere interface. Most studies have also focused on coastal environments with high nutrient loading but have generally excluded the bulk of the ocean including both oligotrophic surface waters and the deep ocean. However, much of the petroleum entering the ocean is released from the seafloor e.g., during pipeline ruptures, release from sinking tankers, or natural seeps and there are several compounds which may dissolve to the surrounding water-column due to their aqueous solubility (Ryerson et al., 2012a; Gros et al., 2017).

These aqueous-soluble, volatile compounds have largely been overlooked in biodegradation studies due to their tendency to evaporate to the atmosphere in surface oil slicks, yet these compounds are abundant in petroleum and refined petroleum products, and are semi-soluble (and as such bioavailable); therefore they may be major drivers of oil biogeochemistry following oil release from the seafloor (Reddy et al., 2012). Here, we focus on biodegradation in the deep ocean and the compound *n*-pentane, which is known to partition to the deep ocean following release from the seafloor (Ryerson et al., 2012a; Gros et al., 2017), and is representative of similar abundant compounds including butane and propane.

3.3.2 The Origin of Seed Populations

Exposure of seawater to a petroleum source drives a remarkable decrease in prokaryotic diversity due to a strong selection for hydrocarbon-degrading microorganisms coupled to toxic effects on other taxa (Röling et al., 2002; Teira et al., 2007; Hazen et al., 2010; Païssé et al., 2010; Redmond and Valentine, 2012a; Love and Arrington et al., 2021). Models of in-situ hydrocarbon biodegradation indicate that as a water parcel encounters a hydrocarbon source, a seed population of hydrocarbon degraders begins to grows in abundance (Hara et al., 2003; Yakimov et al., 2007; Valentine et al., 2012; Harris et al., 2014; Nie et al., 2014; Mahmoudi et al., 2015; Techtmann et al., 2015; Liu et al., 2017). However, the origin and ecology of these seed populations is largely hypothetical. Many studies have suggested seed populations are prolonged and sustained by hydrocarbon substrates originating from various sources including hydrothermal vents (Zhou et al., 2020), cyanobacteria and eukaryotic phytoplankton (Lea-Smith et al., 2015; Love et al., 2021), as well as natural gas seepage (Rubin-Blum et al., 2017; Saunois et al., 2020). As an example, the ubiquitous alkane degrader, *Alcanivorax*, exhibits basal cell populations that range from 10 to 5,000 cells per mL in uncontaminated seawater (Hara et al., 2003; Coulon et al., 2007), with recent work suggesting high native abundance is subsidized by

widespread biosynthesis of long chain *n*-alkanes by marine phytoplankton (Chapter 1, Lea-Smith et al., 2015; Love et al., 2021). Recent evidence has also shown that methanotrophs can be physically transported on bubbles from a gas seep (Jordan et al., 2020, 2021b), pointing to seeps as a potential mechanism to seed the water column with other hydrocarbon degraders. Alternatively, facultative hydrocarbon degraders could be present that rely on other metabolic inputs such as amino acids, carbohydrates, lipids, or other organic acids and switch to hydrocarbons under appropriate conditions (Button et al., 1998; Chung and King, 2001; Fernández-Martínez et al., 2003). Currently, very few studies have focused on the how these factors control the development of a petroleum degrading community during spills in previously uncontaminated waters.

3.3.3 Pentane Bloom Kinetics

In Chapter 1, our investigation of the ocean's biological hydrocarbon cycle (Love and Arrington et al., 2021) revealed the microbial response to *n*-pentane is structured by proximity to seepage (Fig. 2). Chapter 1 and this current chapter focus on sea-going incubations from deep ocean seawater samples (1,000 m), taken along a transect spanning

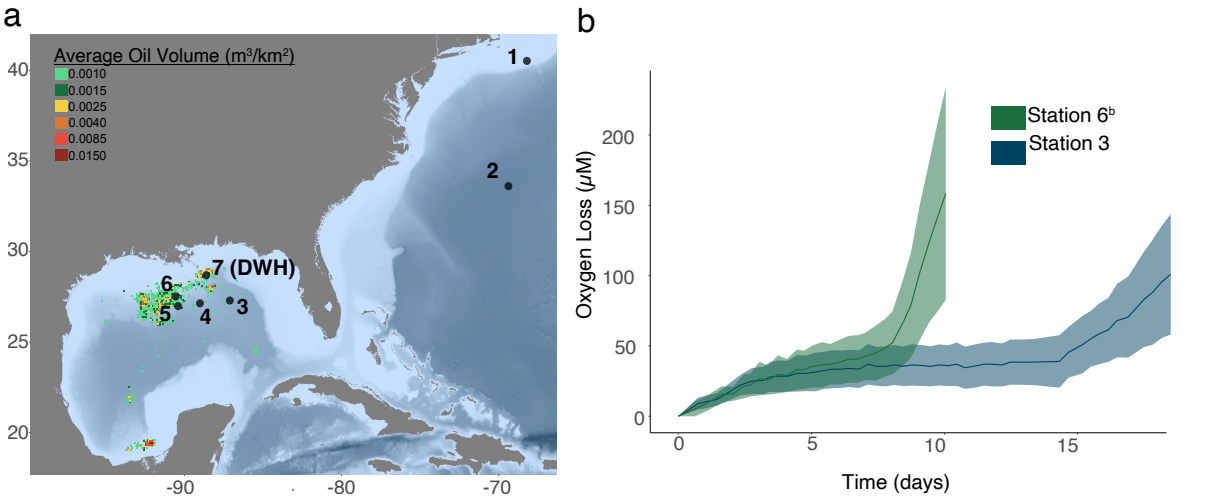


Figure 11. Sampling stations relevant to this chapter. **a** Sampling stations relative to natural petroleum seepage (MacDonald et al., 2015). DWH indicates the location of the sample collected during the DWH disaster. Pentane biodegradation observed via respiratory blooms in experiments conducted at station 1-6. **b** Kinetics of pentane blooms, data published in Love et al. 2021. **c** Oxygen loss with time for blooms at station 6 vs station 3. [b] indicates that one bloom (sample 180) was omitted as it was an outlier with delayed bloom response (23 days) compared to the other 5 replicates which bloomed in 8 days.

both the Gulf of Mexico (GOM) and the North Atlantic. *n*-Pentane metabolism was observed through a closed-system optical oxygen technique in which population blooms were defined by the exponential rise in respiration. We observed distinct bloom response times to *n*-pentane in relation to natural seepage whereby bloom onset on pentane is ~9X faster in the seep-ridden Northwest Gulf of Mexico compared to the water underlying the North Atlantic subtropical gyre (MacDonald et al., 2015; Love et al., 2021). Median bloom times varied from 72.9 days furthest from natural seepage to 8.3 days closest to natural seepage. The fraction of samples that exhibited a bloom response also coincided with proximity to natural seepage (Fig. 2b)

In this chapter, we investigate the influence of biogeography on microbial hydrocarbon metabolism by analyzing the genomic content of organisms with contrasting bloom response times and source water origins. We extract and analyze the 16S rRNA amplicon content of each bloom, then assemble high-quality draft genomes from bloom experiments and perform a complementary proteomic analysis to evaluate metabolism. We find the predominant microbial member of the *n*-pentane enriched community across the Gulf of Mexico belongs to the *Cycloclasticus* genus of ubiquitous hydrocarbon degraders, originally named for their metabolic capability to consume polycyclic aromatic hydrocarbons (PAHs). Interestingly, we find that among near-identical *Cycloclasticus* genomes, one variant favors the seep-influenced region of the northwestern GOM (hereafter referred to as the seep variant, SV), whereas the other favors the open ocean region far from natural seepage (hereafter referred to as the open ocean variant, OOV).

3.4 Methods

3.4.1 Incubation Design and Sample Collection

Seawater samples were collected on two research cruises aboard RV Atlantis in June 2015 and the RV Neil Armstrong in May 2017. *n*-Pentane incubations were conducted at stations 1 (40° 9.14' N, 68° 19.889' W), 2 (33° 58.21' N, 69° 43.38' W), 3 (27° 30.41' N,

87° 12.41' W), 4 (27° 15.00' N, 89° 05.05' W), 5 (27° 11.60' N, 90° 41.75' W) and 6 (27° 38.40' N, 90° 54.98' W) with seawater collected from 1,000 m. Respiration data and methods are available from (Love et al., 2021), with sample numbers re-named for this chapter to exclude irrelevant pentadecane data. Seawater collected from the CTD Niskin bottles were transferred to 250 mL glass serum vials using a small length of Tygon tubing. Vials were filled for at least 3 volumes of water to overflow. Care was taken to ensure no bubbles were present before sealing with a polytetrafluoroethylene (PTFE) coated chlorobutyl rubber stopper and crimp cap seal. All bottles, except for unamended blank controls, immediately received 10 μL of pentane using a gas-tight syringe (Hamilton) and were maintained in the dark at in-situ temperature (4ºC). Prior to filling, each serum bottle was fixed with a contactless optical oxygen sensor (Pyroscience, Aachen, Germany) on the inner side with silicone glue, and afterward were cleaned from organic contaminants with rinses of ethanol, 3% hydrogen peroxide, 10% hydrochloric acid, and MilliQ water, and were sterilized via autoclave. Oxygen concentration was monitored approximately every 8 hours with a fiber optic oxygen meter (Pyroscience, Aachen, Germany). Observed changes in oxygen content were normalized to unamended controls to correct for oxygen loss from background respiration processes and variability due to temperature changes.

3.4.2 DNA extraction, 16S rRNA community analysis

After 27-30 days each bottle was harvested and filtered on a 0.22 μm polyethersulfone filter. DNA extraction was performed from ¼ of each filter using the PowerSoil DNA extraction kit with the following modifications: 200 μl of bead beating solution was removed at the initial step and replaced with phenol chloroform, the C4 bead binding solution was supplemented with 600 μL of 100% ethanol, and we added an additional column washing step with 650 μL of 100% ethanol. Extracts were purified and concentrated by ethanol precipitation, then stored at -80ºC. The V4 region of the 16S rRNA gene was amplified using the method described by (Kozich et al., 2013) with small

modifications to the 16Sf and 16Sr primers according to (Apprill et al., 2015; Parada et al., 2016). Amplicon PCR reactions contained 1 µL of template DNA, 2 µL of forward primer, 2 µL of reverse primer, and 17 µL of AccuPrime Pfx SuperMix. Thermocycling conditions consisted of 95° 2 min, 30 cycles of 95°C for 20 secs, 55°C for 15 secs, 72°C for 5 min, and a final elongation at 72°C for 10 min. Sample DNA concentrations were normalized using the SequelPrep Normalization Kit, cleaned using the DNA Clean and Concentrator kit, visualized on an Agilent Tapestation, and quantified using a Qubit Flurometer. Samples were sequenced at the UC Davis Genome Center on the Illumina MiSeq platform with 250nt, paired end reads. A PCR-grade water sample was included in extraction, amplification, and sequencing as negative control to assess for DNA contamination.

Trimmed fastq files were quality filtered using the fastqPairedFilter command within the dada2 R package, version 1.9.3 (Callahan et al., 2016) with following parameters: truncLen=c(190,190), maxN=0, maxEE=c(2,2), truncQ=2, rm.phix=TRUE, compress=TRUE, multithread=TRUE. Quality filtered reads were dereplicated using derepFastq command. Paired dereplicated fastq files were joined using the mergePairs function with the default parameters. A Single Nucleotide Variant (SNV) table was constructed with the makeSequenceTable command and potential chimeras were removed denovo using removeBimeraDenovo. Taxonomic assignment of the sequences was done with the assignTaxonomy command using the Silva taxonomic training dataset formatted for DADA2 v132 (Pruesse et al., 2007; Glöckner et al., 2017). If sequences were not assigned, they were left as NA.

3.4.3 Metagenomic Reconstruction

Metagenomic library preparation and shotgun sequencing were conducted at the University of California Davis DNA Technologies Core. DNA was sequenced on the Illumina HiSeq4000 platform, producing 150-base pair (bp) paired-end reads with a targeted insert size of 400 bp. Quality control and adaptor removal were performed with Trimmomatic

(Bolger et al., 2014) (v.0.36; parameters: leading 10, trailing 10, sliding window of 4, quality score of 25, minimum length 151 bp) and Sickle (Joshi and Fass, 2011) (v.1.33 with paired-end and Sanger parameters). The trimmed high-quality reads were assembled using metaSPAdes (Nurk et al., 2017) (v.3.8.1; parameters $k = 21, 33, 55, 77, 88, 127$). The quality of assemblies was determined using QUAST (Gurevich et al., 2013) (v.5.0.2 with default parameters). Sequencing coverage was determined for each assembled scaffold by mapping high-quality reads to the assembly using Bowtie2 (Langmead and Salzberg, 2012)(v.2.3.4.1; default parameters) with Samtools (Li et al., 2009) (v.1.7). Contigs greater than 2,500 bp were manually binned using Anvi'o with Centrifuge (Eren et al., 2015; Kim et al., 2016) (v.1.0.1) based on coverage uniformity (v.5). Quality metrics for metagenome-assembled genomes (MAGs) were determined using CheckM (Parks et al., 2015) (v.1.0.7; default parameters). The taxonomy of each MAG was classified using GTDB-Tk (v.1.0.2) against The Genome Taxonomy Database (Chaumeil et al., 2020) (https://data.ace.uq.edu.au/public/gtdb/data/releases/release89/89.0/, v.r89). The average nucleotide identity of each genome was determined with the ANI Matrix via the Enveomics tool collection (Rodriguez-R and Konstantinidis, 2016).

Using a 16S search tool through the Joint Genome Institute- Integrated Microbial Genomes and Microbiomes (JGI-IMG) portal, we identified public unprocessed environmental metagenomic datasets with *Cycloclasticus* representation. These datasets were downloaded and metagenomic reconstruction was performed according to the above protocol with the following modifications: binning was performed with the automated binning software, MetaBAT 2 (Kang et al., 2019). This tool at JGI-IMG has since been removed, however we have received explicit permission from authors of each dataset to include their data in our analyses.

3.4.4 Phylogenetics

To define genome phylogenomic relationships of MAGs, 16 universal ribosomal proteins (RPs) were used (L2-L6, L14-L16), L18, L22, L24, S3, S8, S10, S17, and S19. For phylogenies of metabolic genes as well ribosomal proteins, all representative sequences and concatenated alignments that contained <25% informative sites were excluded in tree construction. Each protein was aligned using MUSCLE (v.3.8.425) (Edgar, 2004). All columns with >95% gaps were removed using TrimAL (Capella-Gutiérrez et al., 2009). Maximum-likelihood phylogenetic analysis of concatenated alignment was inferred by RAxML (Stamatakis, 2014) (v.8..9; parameters: raxmlHPC -T 4 -s input -N autoMRE -n result -f a -p 12345 -x 12345 -m PROTCATLG). Resulting trees were visualized using FigTree (FigTree) (v.1.4.3).

3.3.5 Metaproteomics

We analyzed metaproteomes from two of the SV (seep variant) samples with corresponding metagenomes ("3_C5_1" and "3_C5_2") as the reference databases. Proteins from each sample were extracted and prepared from ¼ filter (equivalent to ~60 mL of filtered water and ~$1.3x10^8$ bacteria) for liquid chromatography and tandem mass spectrometry (LC-MS/MS) using a protocol adapted from (Timmins-Schiffman et al., 2016). Briefly, filters were submerged in 100 µL of 6M urea and 600 µL of 50 mM NH_4HCO_3 and sonicated (5 x 20s) to lyse cells. Proteins within the lysate were reduced and alkylated using dithiothreitol and iodoacetamide, respectively, digested with Trypsin (12 h; 1:20 enzyme to protein) and desalted with C18 centrifugal spin columns. Peptides were dried down and resuspended in 2% ACN, 0.1% formic acid prior to analysis with a nanoAcquity UPLC (Waters Corp., Milford, MA) in line with a Q-Exactive-HF (Thermo Fisher Scientific, Waltham, MA). All database searches were performed with Comet (Eng et al., 2013). All relevant peptide hits with an e-value less than 0.01 were used to define protein presence. All peptides tandem mass spectra discussed here were manually investigated to verify b and y ions.

3.5 Results and Discussion

3.5.1 Variant Biogeography

Incubations conducted with seawater from 1,000 m depth containing ambient nutrients successfully exhibited robust blooms of bacteria, when supplied with *n*-pentane as a source of carbon and energy. Blooms are characterized by bacterioplankton abundance, increasing by ~10X (Supplementary Data 3), significant drawdowns in inorganic phosphate and nitrate concentration (Fig. 12c, Fig. 12d, Fig. 12e; Supplementary Data 3), and the emergence of dominant taxa comprising >70% of the microbial community at the termination of the experiment (Fig 12a). Community analysis of incubations that failed to bloom within experimental timeframe (27-30 days) revealed instances where the community was dominated by a limited number of taxa, indicating a microbial community shift precedes major respiratory signals (Supplementary Data 4). Over 60% of all pentane blooms in the deep GOM were dominated by *Cycloclasticus* (Fig 12a). Variation among replicate incubations was observed as a shift in the dominant taxa, potentially related to greater microbial diversity and abundance of alkane degraders near seepage. At station 6 *Colwellia* and an unclassified genus belonging to *Cellvibrionaceae* bloomed, and the same unclassified *Cellvibrionaceae* bloomed at station 5 (Fig 12a). Among blooms, we observed two dominant *Cycloclasticus* variants with one (SV) being the primary blooming population at station 6 within the northwestern GOM seep field and the other (OOV) being the primary blooming organism at station 3 within the more seep deplete region (Fig. 11a, Fig. 12a, Supplementary Data 3). The occurrence of SV and OOV at stations 4 and 5 is patchier with both variants occurring at station 4 and only OOV blooming at station 5 (Fig. 12a). The success of OOV over SV at station 5 may be explained by the western flow of eastern sourced seep-deplete waters observed at the time seawater was collected (Fig. 13). Cell-specific respiration is higher for SV compared to the OOV (Fig. 12b), and the respiration

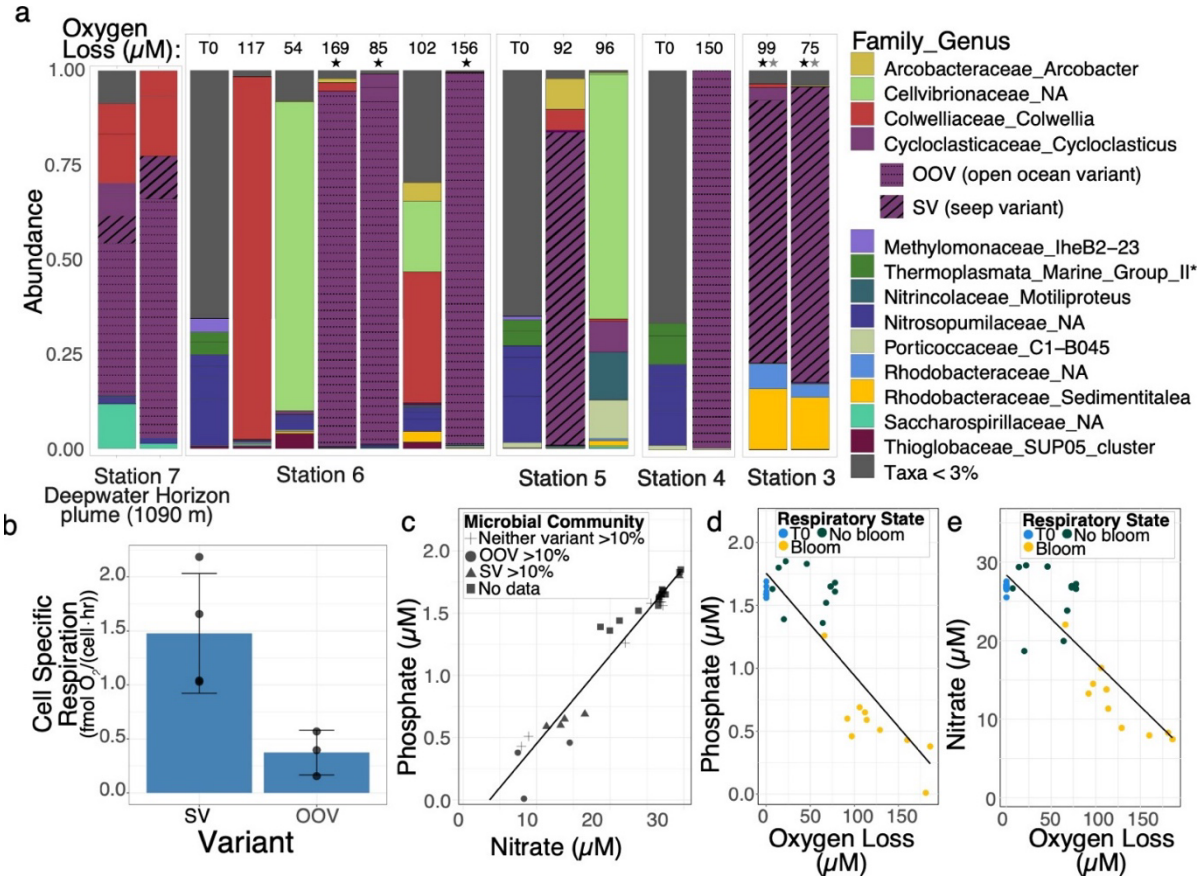


Figure 12. Microbial and geochemical characteristics of blooms. **a** Microbial community composition of pentane blooms informed via the V4 region of the 16S rRNA gene for initial environmental samples (denoted "T0") and pentane incubations harvested after significant oxygen loss was observed. DWH sample collected on 5/30/10 during the DWH event. Variant 1 of *Cycloclasticus* blooms in 8-20 days in locations close to natural seepage and was abundant on 5/30/10 during the Deepwater Horizon disaster. Variant 2 of *Cycloclasticus* blooms in 16-20 days and dominates seawater originating further from natural seepage inputs. *No taxonomic representative at the family or genera level. **b** Cell specific respiration in incubations dominated (>80%) by *Cycloclasticus* variant 1 (close to seepage) and 2 (far from seepage/open ocean). **c** Dissolved phosphate vs dissolved nitrate concentration in initial samples (T0) and at sample harvest. **d** Dissolved phosphate concentration in final samples is depleted compared to unamended controls and initial samples (T0). **e** Dissolved nitrate concentration in final samples is depleted compared to unamended controls and initial samples (T0).

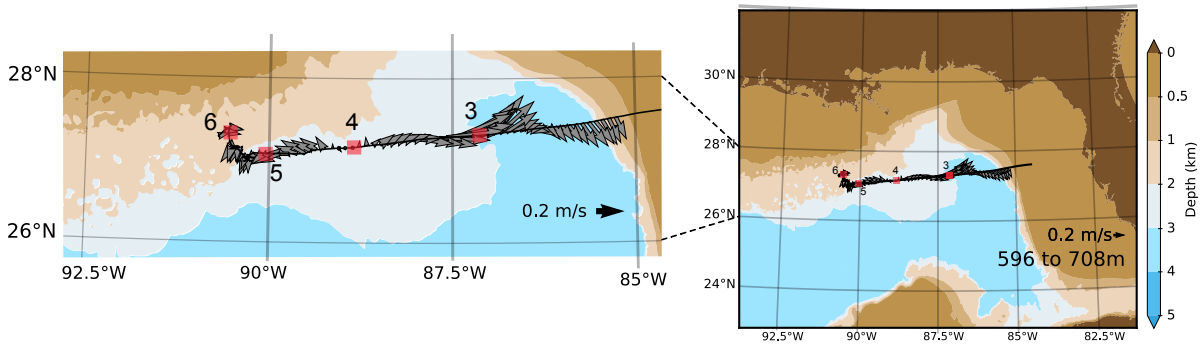


Figure 13. ADCP data showing deep ocean currents over the course of the GOM sampling expedition (Atlantis, 2015).

profile (Fig. 11c) of these variants also showed distinct patterns, where OOV presents more gradual oxygen loss with time. The OOV also was observed in small relative abundances (<1%) at station 6 (Supplementary Data 4). In each pentane incubation where the two variants co-occurred, the SV numerically dominated by ~3 orders of magnitude except for the *Colwellia*/*Cellvibrionaceae* bloom at station 6 where both variants were detected at <1% abundance. This suggests that SV is more adapted to conditions associated with natural seepage, as it outcompeted OOV, which only boomed at stations that are distant from seeps.

3.5.2 Pentane Metabolism

We reconstructed high quality metagenomes from five pentane bloom samples, with completeness >97% and redundancy <2% (black stars in Fig 12). Three MAGs, named “6_C5_1”, “6_C5_2”, and “6_C5_3”, originated from station 6 (natural seep region) and two MAGs, named “3_C5_1” and “3_C5_2”, originated from station 3 (open ocean region). Our analyses were further supported by proteomic analysis of the two MAGs from station 3 (gray stars in Fig. 12a; Fig 14b). Each MAG was recovered from biologically independent incubations, yet every major component of metabolism and every taxonomic marker was nearly identical across each station, therefore we will refer to “SV-MAG” as the three MAGs from station 6 and “OOV-MAG” as the two MAGs from station 3. Within both SV- and OOV-MAGs we found genomic potential for pentane utilization for catabolism and anabolism (Fig. 14a).

The first step in the consumption of pentane is the oxidation to pentanol, and we hypothesize that this step is catalyzed by the copper-containing membrane associated monooxygenase, called particulate hydrocarbon monooxygenases (pHMOs). The most well-characterized pHMO is the particulate methane monooxygenase, which oxidizes methane to methanol (Jordan et al., 2021a). We found multiple copies of genes encoding pHMOs in both *Cycloclasticus* MAG variants (Fig. 14a). Each copy of pHMO varies

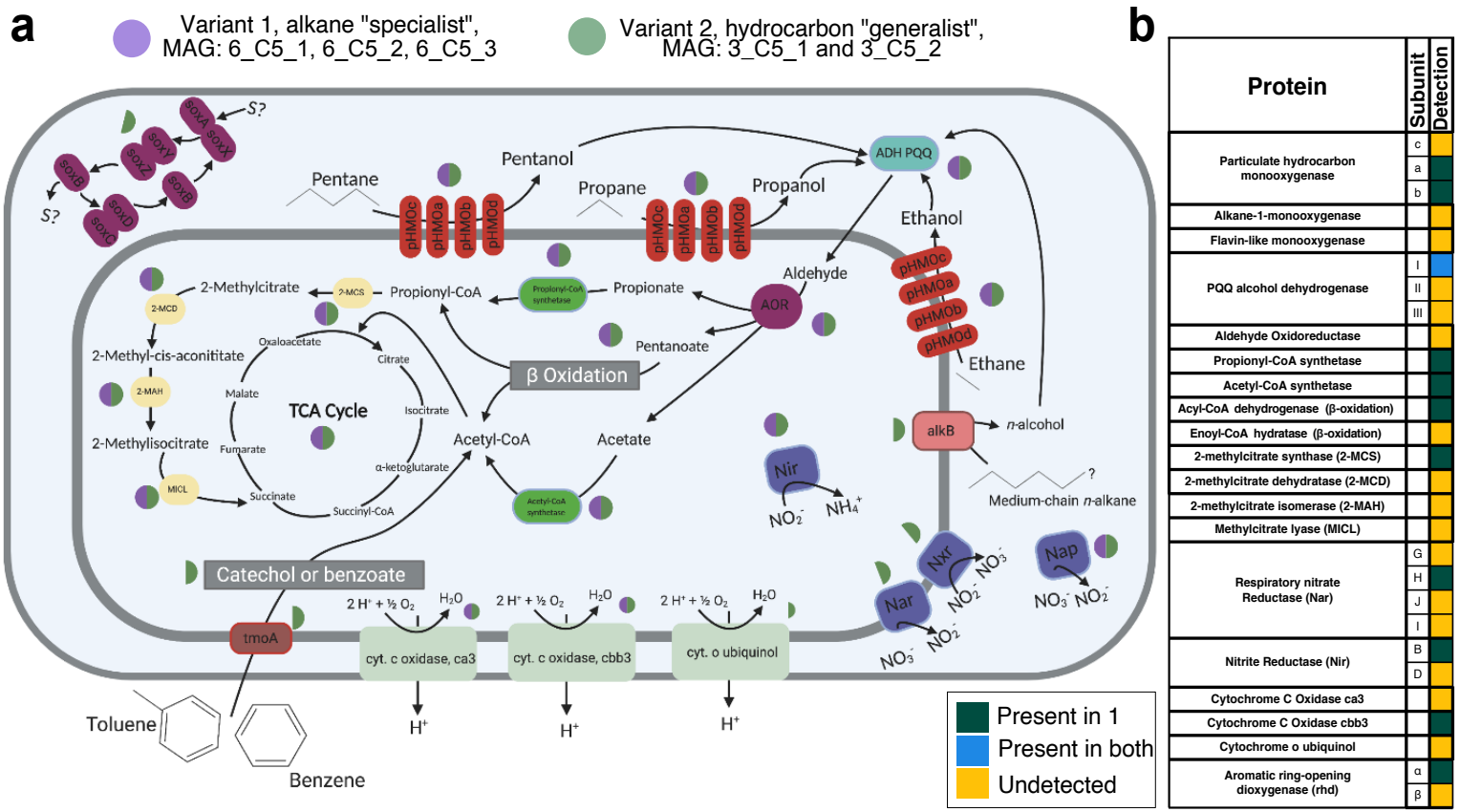


Figure 14. Carbon, nitrogen, and sulfur metabolism present in *Cycloclasticus* variants blooming on pentane. **a** pathways used for carbon assimilation and respiration via oxygen, nitrate, and sulfate. Color of half circles indicate the gene or pathway is present in MAG variant 1 (SV, purple) or MAG variant 2 (OOV, green). Enzyme abbreviations: particulate hydrocarbon monooxygenase pHMO(a,b,c,d); PQQ-dependent alcohol dehydrogenase (ADH PQQ); aldehyde oxidoreductase (AOR); malate dehydrogenase (MD); isocitrate lyase (ICL); malate synthase (MS); 2-methylcitrate synthase (2-MCS); 2-methylcitrate dehydratase (2-MCD); 2-methylaconitate isomerase (2-MAH); methylisocitrate lyase (MICL); respiratory nitrate reductase (Nir); thiosulfate oxidation complex (Sox); alkane-1-monooxygenase (alkB). **b** presence and absence matrix of proteomic data for MAG variant 2 only ("3_C5_1" and "3_C5_2"), where dark green indicates peptides observed in one of the two samples, light blue indicates peptides observed in both samples, and yellow indicates peptides were undetected.

phylogenetically from the other copies within the same genome, suggesting each operon may have different substrate specificities (Fig. 15). Both MAG-SV and MAG-OOV have pHMO sequences that form monophyletic clades with reference sequences with demonstrated affinity for ethane and butane. Both variants also contain a sequence that forms a monophyletic clade that is only distantly related to pMMOs; however, this clade contains no currently validated reference sequences, and we refer to its function as "unknown". Proteomics confirmed the expression of pHMOs, specifically subunit a and b (Fig.14b) in the presence of pentane. The pHMO for which peptides were detected belongs to a sequence from OOV-MAG in the "unknown" clade of pHMOs. *AlkB*, a gene known to function on medium to long-chain alkanes, was found encoded in the MAG-OOV, however no peptides were observed in the proteomics analyses (Fig. 14b). Still, given the minimal sample size analyzed for proteomics and false negatives are a possibility due to ionization and extraction efficiencies, we do not exclude the possibility that *alkB* could also be active in these samples and used to consume pentane by the OOV.

The second step in the consumption of pentane is the conversion of pentanol to an aldehyde. In many bacteria that oxidize alcohols, this reaction is catalyzed by pyrroloquinoline quinone-dependent alcohol dehydrogenases (PQQ-ADH). We found genes encoding PQQ-ADHs in both *Cycloclasticus* MAG variants, as well as proteomic expression of PQQ-ADH in both OOV-MAG samples, "3_C5_1" and "3_C5_2" (Fig. 14b, Fig.16). None of the PQQ-ADH's formed a monophyletic clade with reference sequences known to act on methane, providing evidence against methane metabolism in SV and OOV. In the third step of pentane consumption the aldehydes are oxidized to carboxylic acids, which could be achieved via a tungsten-containing aldehyde ferredoxin reductase (AOR), known to use short-chain alkane derived aldehydes as their substrate (White et al., 1993). This conversion can also be performed by PQQ-ADH, as activity on aldehydes has been confirmed with reference sequences related to those encoded by SV and OOV (Fig. 16).

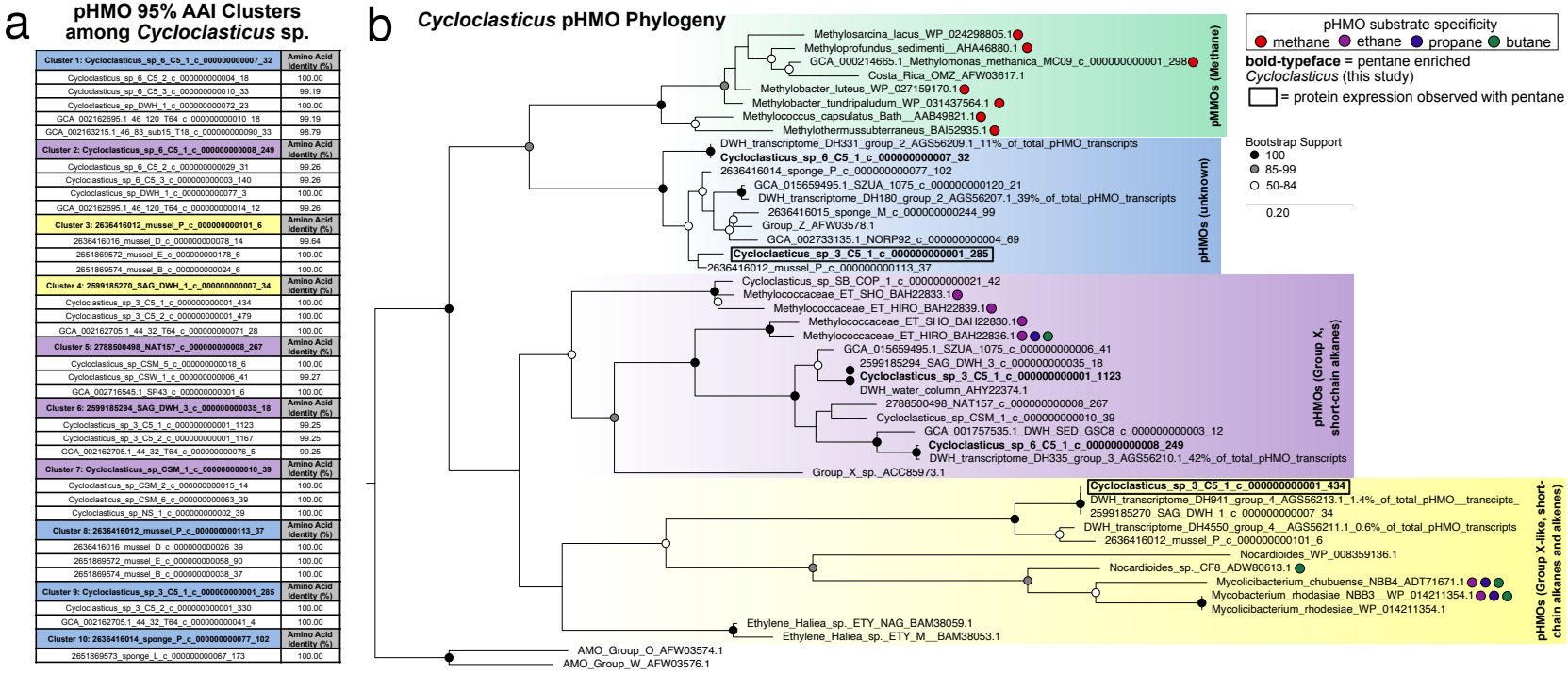


Figure 15. Phylogeny of pHMO subunit A protein sequences. **a** Description of pHMO clusters containing 95% amino acid identity. The representative sequence for each cluster is colored according to the phylogentic placement. **b** Maximum likelihood tree is drawn to scale, with branch lengths representing the number of substitutions per site. Bootstrap values below 50% are not shown. Each major clade is color coded for readability with green representing pMMOs with activity on methane, the blue clade is "unknown" and lacks any known substrate specificity, the purple clade represents group X pHMOs (ethane/ethylene, propane, and butane activity), and the yellow clade are group-X like (ethane, propane, and butane activity). Sequences from the pentane enriched *Cycloclasticus* MAGs are in bold, boxed values indicate pHMOs detected in proteomic data.

Here, pentanoate is likely beta-oxidized using acyl-CoA dehydrogenase and enoyl-CoA hydratase and shunted into central carbon metabolism via the citric acid cycle (Fig. 14a).

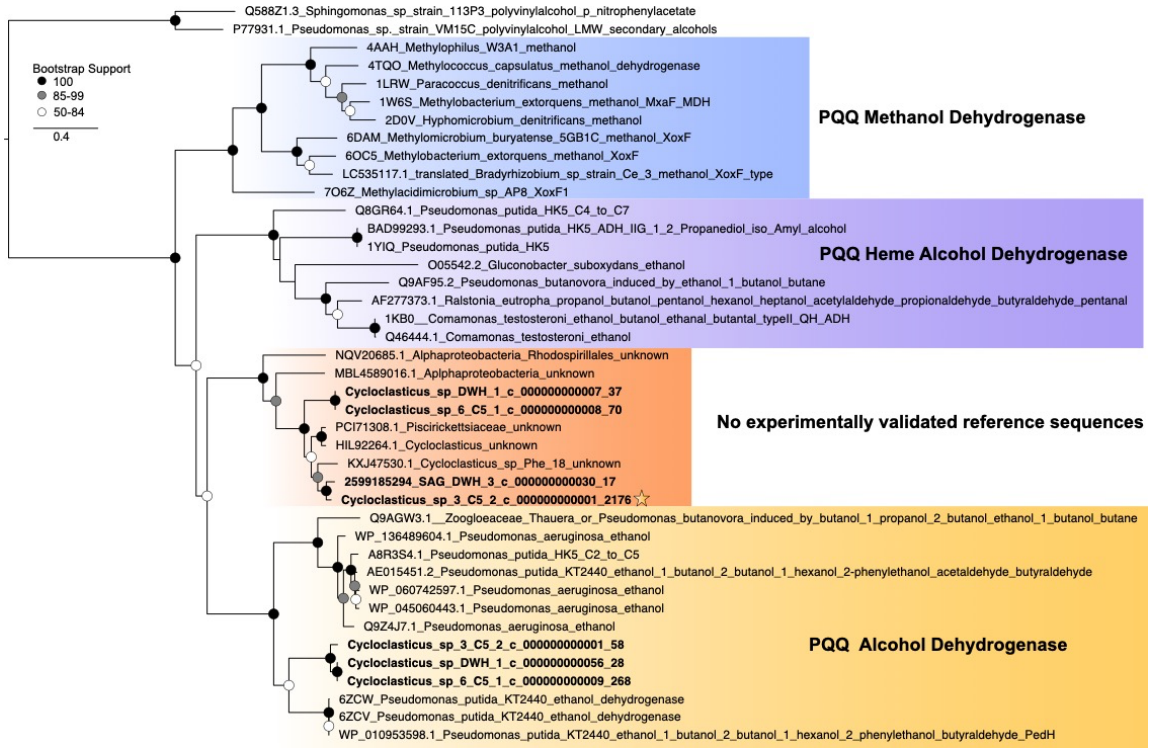


Figure 16. Maximum-likelihood phylogenetic tree of PQQ-dependent alcohol dehydrogenase sequences rooted to polyvinylalcohol reference sequences. Experimentally validated reference sequences are annotated with known substrate activity. Sequence denoted with a star was observed in proteomic samples from pentane enrichment. No experimentally validated reference sequences present in the orange clade.

3.5.3 Differences in Variant Metabolic Capability

Interestingly, the metabolic capabilities of the SV-MAG and OOV-MAG differ substantially (Fig. 17b). The OOV-MAG encodes for general hydrocarbon metabolism that includes the nearly complete pathway for toluene consumption via the toluene monooxygenase conversion of toluene to benzoate (7 of 8 genes), benzoate conversion to catechol (3 of 4 genes), and the catechol metacleavage to acetyl-CoA which enters the tricarboxylic acid cycle (13 of 13 genes). The OOV-MAG also encodes toluene 2-monooxyenase which converts benzene to catechol (6 of 6 genes) and utilizes the catechol metacleavage pathway to acetyl-CoA (13 of 13 genes) and the tricarboxylic acid cycle. The

OOV could also use the toluene-2 monooxygenase system to convert toluene to 3-methylcatechol (6 of 6 genes) and then convert 3-methylcatechol to acetyl-CoA and shunt to the tricarboxylic acid cycle (3 of 5 genes). Furthermore, the OOV-MAG encodes for *alkB* (1 of 1 gene), which is commonly used by other organisms for consumption of long chain alkanes via beta oxidation (OOV encodes 7 of 7 genes) resulting in propionyl-CoA and acetyl-CoA which are also incorporated into the tricarboxylic acid cycle. Interestingly, neither OOV or SV (or any other *Cycloclasticus* genome analyzed in this study) encode a complete canonical naphthalene degradation pathway (naphthalene 1,2, dioxygenase is missing from all genomes) although *Cycloclasticus* SP-1 isolate has been experimentally validated to use naphthalene as a sole carbon source (Fig. 17b) (Wang et al., 2018). Both *Cycloclasticus* SP-1 and MAG-OOV encode 3 of 10 genes for naphthalene degradation whereas SV encodes 0 of 10 genes. In instances of fragmented pathways with highly conserved genes, it is likely that divergent genes are present which are undetected by our protein scanning methods.

Overall, the SV-MAG lacks many metabolic pathways for longer chain alkanes and aromatic compounds compared to the OOV-MAG, seemingly limiting its hydrocarbon metabolism potential (Fig. 14, Fig. 17b). These observed differences are consistent with SV specialization on short chain, aqueous-soluble alkanes and biogeography that includes seeding from the petroleum-rich source region in the Northern Gulf of Mexico. The genomic capacity for catabolism of multiple hydrocarbon classes in the OOV-MAG is consistent with an ability to capitalize on hydrocarbon sources that expand beyond natural seeps to include atmospheric deposition, biogenic inputs, and oil spills. This enhanced capacity in OOV is consistent with an expanded biogeographic range relative to SV, which appears to be more highly reliant on substrate sourced from natural seepage.

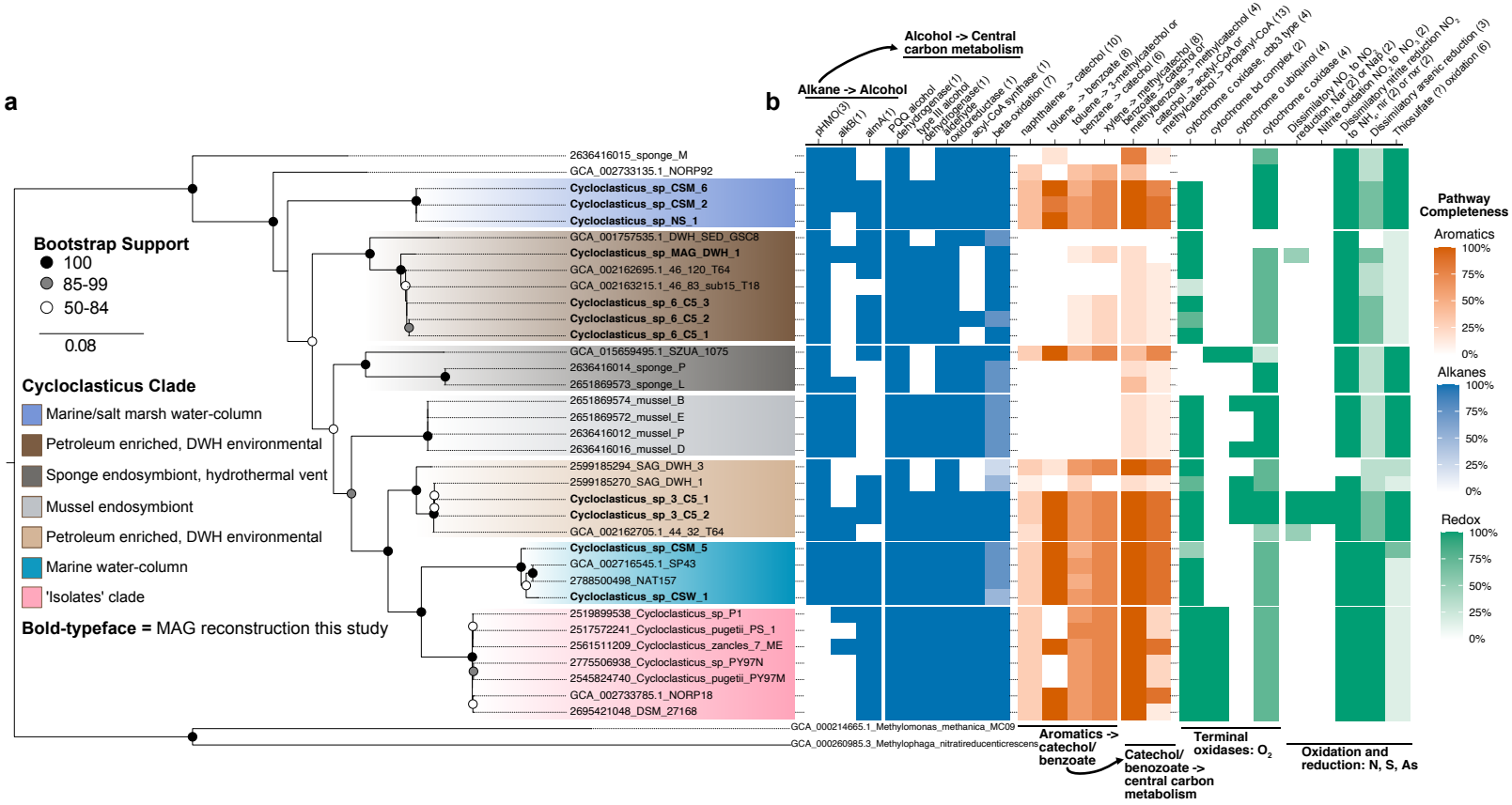


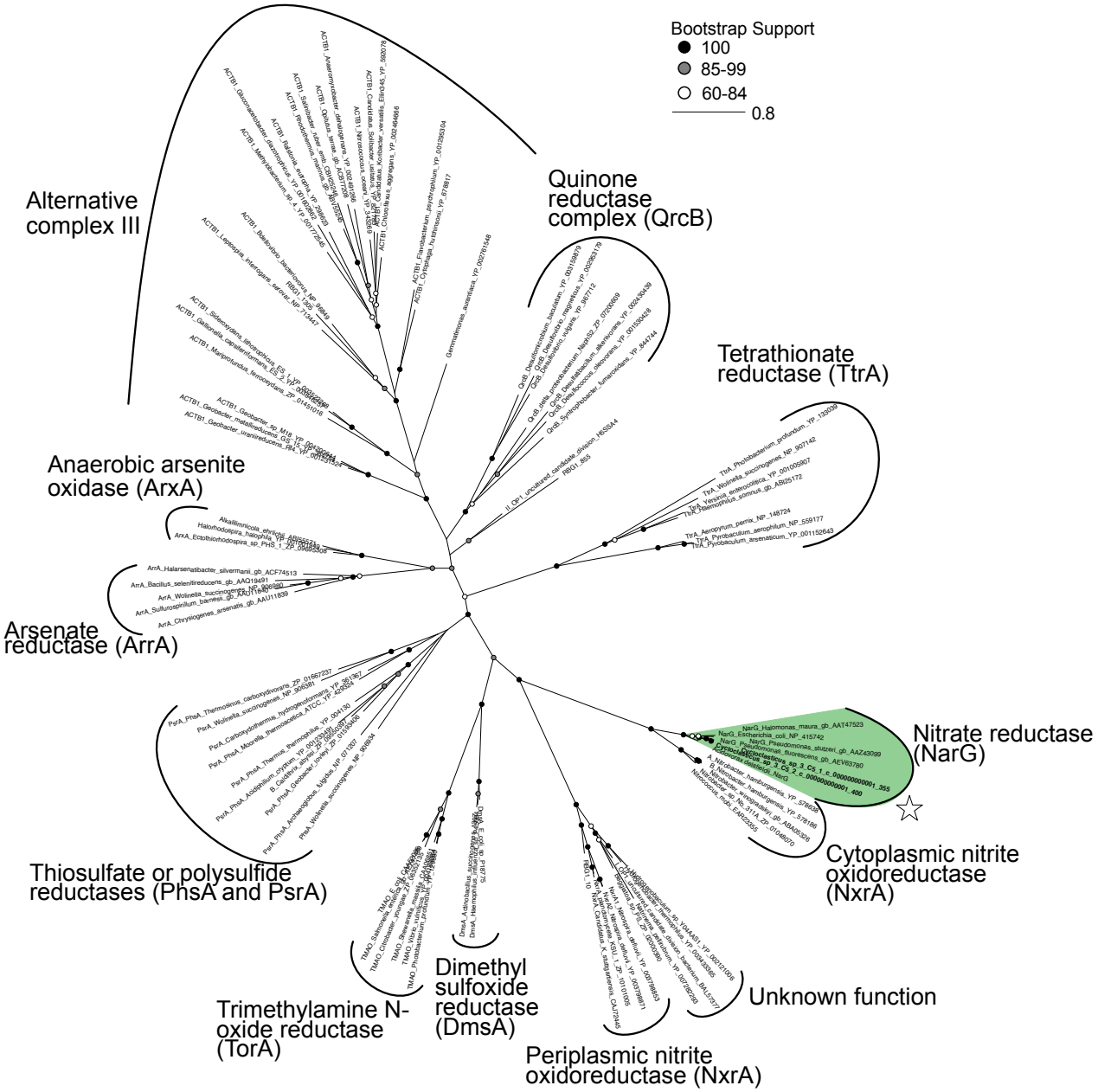
Figure 17. Phylogeny and capabilities for hydrocarbon consumption of *Cycloclasticus*. **a** Phylogeny constructed using 16 ribosomal proteins, subclades are designated based on relative distance from the root and supported by %ANI (supplementary material). **b** Completeness of select metabolic pathways relating to alkane (blue), aromatic (red) metabolism, and redox (green). The number of genes considered in calculating pathway completeness are shown in parentheses.

3.5.4 Anaerobic Metabolism in *Cycloclasticus*

Anaerobic metabolism has not yet been observed in *Cycloclasticus,* and it remains unknown how these bacteria could contribute to hydrocarbon cycling in oxygen minimum zones or anoxic sediments. Here, we show OOV *Cycloclasticus* exhibits adaptations for life so is without oxygen, including the occurrence of genes for respiratory nitrate reductase (*Nar*), as well as a potential linkage to thiosulfate metabolism (Fig. 14). In OOV we identify a complete canonical *nar* operon (*narGHJI*) encoding: i) the α subunit that catalyze NO_3^- reduction to NO_2^- (*narG*); ii) the iron–sulfur-containing β subunit (*narH*) that transfers electrons to the molybdenum cofactor of *narG*; iii) the transmembrane cytochrome *b*-like γ subunit (*narI*) involved in electron transfer from membrane quinols to *narH*; and iv) the *narJ* chaperone involved in enzyme formation. Among these genes, *narH* peptides were observed in the proteomic analysis (Fig. 14b). Phylogenetic placement of *Cycloclasticus narG s*equences also confirms relation to *narG* reference sequences (Fig. 18). Furthermore, OOV also encodes the nitrite oxidation genes *nxrAB* which is also reversible and can catalyze the reduction of nitrate to nitrite (Fig. 17b) (Starkenburg et al., 2006). OOV also contains the *sox* operon (*soxCDXYZAB*), which encodes periplasmic sulfur oxidizing proteins (Fig. 14). This operon can be used as a means of detoxification in some Gammaproteobacteria (Spring, 2014), however we do not exclude the possibility that *Cycloclasticus* could employ a lithoheterotrophic strategy. Use of thiosulfate to supplement heterotrophy is a strategy that has been demonstrated in other proteobacteria and could be useful in seeps and other benthic environments (Moran et al., 2004).

Samples for DNA and proteomics were C5_1" and "3_C5_2", which is greater than the so-called hypoxic barrier (~20 μM) in marine environments (Giovannoni et al., 2021). Interestingly, within the same sample ("3_C5_1") we see evidence for aerobic metabolism in the form of terminal oxidases (cytochrome c. cbb3 type) and particulate hydrocarbon

monooxygenase, as well as the anaerobic process of nitrate reduction via *Nar*. We conclude that the *Cycloclasticus* community was likely transitioning from aerobic to anaerobic metabolism in which some portion of the community was performing aerobic processes while another portion was functioning anaerobically. However, we cannot



exclude the possibility that individual cells within that population use oxygen concurrently

Figure 18. Maximum-likelihood phylogenetic tree with scale bar of substitutions per site of DMSO reductase superfamily modeled after Castelle et al. 2013. NarG sequences from Cycloclasticus OOV variant in bold and noted with star symbol.

with nitrate as electron acceptors. It is currently unknown how *Cycloclasticus* might initiate the first step of pentane consumption while functioning anaerobically, given pHMOs are dependent on oxygen, but the potential to do so is important to the biogeochemical processing of hydrocarbons in the ocean.

3.5.5 Deepwater Horizon *Cycloclasticus*

The microbial response to the 2010 Deepwater Horizon blowout in the Gulf of Mexico induced blooms of *Cycloclasticus* in the deep ocean from large scale intrusions of dissolved hydrocarbons (Valentine et al., 2010). These DWH blooms included multiple *Cycloclasticus* 16S rRNA sequence variants (Redmond and Valentine, 2012b), which led us to ask whether SV and OOV were among those DWH variants. Therefore, we extracted and analyzed two replicate archive DNA samples collected while the wellhead was still leaking into the Gulf of Mexico on 5/30/10 from a depth of 1,090 m. At this depth, there was an oxygen anomaly characteristic of the respiratory response associated with the DWH subsurface intrusions (Kessler et al., 2011). Upon initial analysis of the microbial community via the V4 region (252 bp) of the 16S rRNA gene we found the SV-MAG to be identical to the dominant member of the DWH sample, and OOV-MAG to be identical to the second most abundant *Cycloclasticus* 16S rRNA single nucleotide variant (Fig. 12a).

We carried out additional shotgun sequencing on DNA from one of the DWH samples and reconstructed a high-quality metagenome, here named “MAG_DWH_1”, which is 94% complete and 3.3% redundant. Unfortunately, the second variant related to OOV was unable to be recovered with metagenomics due to issues with assembly fragmentation and binning of the two closely related strains. To obtain a high-quality draft genome of “MAG_DWH_1”, we subsampled our reads by 50% until the OOV-related sequences were a small fraction of the assembled data. Upon expanding our analysis to the full length 16S rRNA gene we find that the SV-MAG is 99.5% identical to MAG_DWH_1, and through a phylogenomic analysis of 16 ribosomal proteins, we find MAG_DWH_1 forms

a monophyletic clade with SV-MAG (Fig. 17a). For comparison, we also drew from our previously published single amplified genomes (SAGs) from DWH, which are 71%, 49%, and 46% complete and herein referred to as “SAG_DWH_3”, “SAG_DWH_1”, and “SAG_DWH_2” (Rubin-Blum et al., 2017). We find “SAG_DWH_1” and “SAG_DWH_3” are closely related to OOV whereas “SAG_DWH_2” appears related to SV. For the relation of SV and OOV to the SAGs and MAG_DWH_1 we also find supporting evidence in analysis of Average Nucleotide Identity and the 16S rRNA phylogeny (Fig. 19 and Fig. 20). These phylogenomic results indicate a previously unrecognized distinction in the microbial response to the DWH event – that SV-like *Cycloclasticus* may have responded specifically

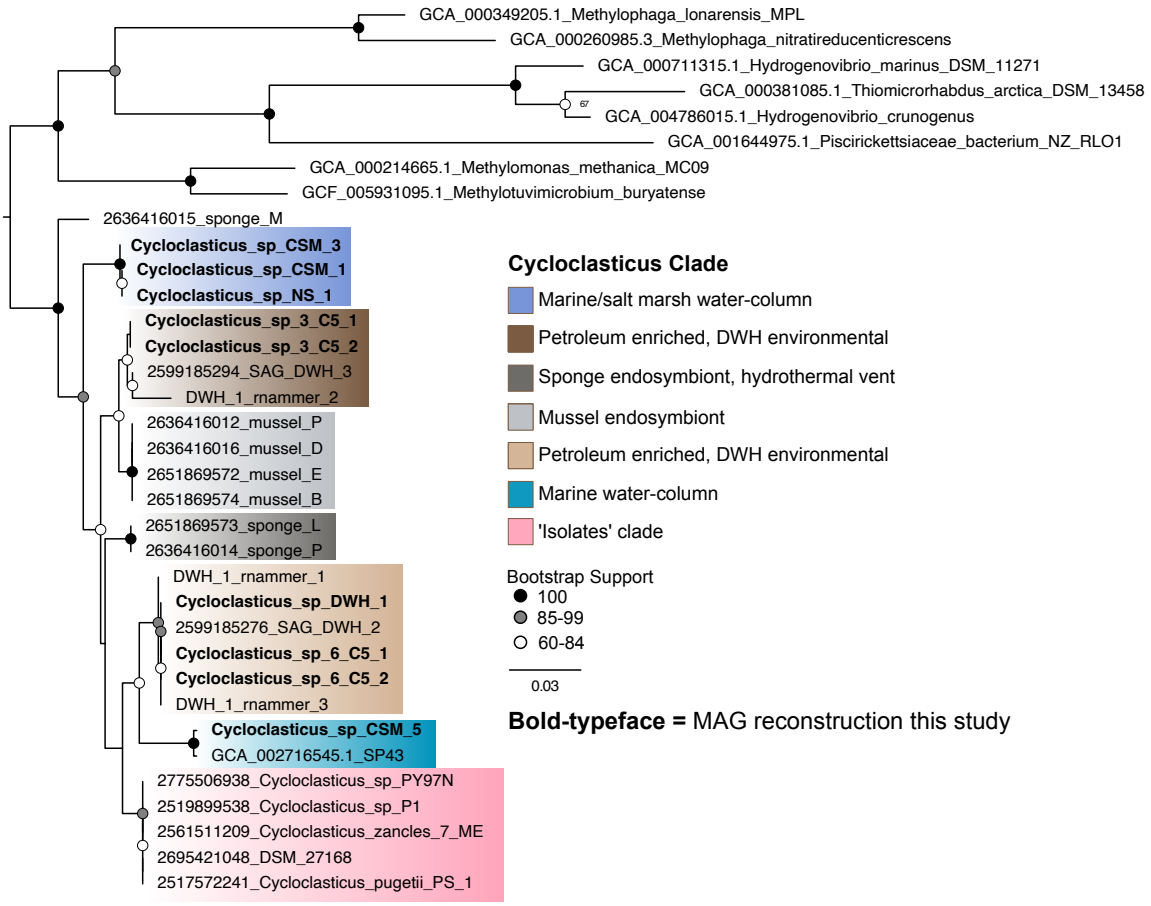


Figure 19. Maximum-likelihood phylogenetic tree of16S rRNA rooted to representatives from *Piscirickettsiaceae* and *Methylococcacae* family. “_rnammer” denotes 16S sequences recovered from assembled contigs prior to binning for the DWH sample. Alignment includes 1,547 amino acid residues.

to the highly abundant soluble *n*-alkanes whereas OOV-like *Cycloclasticus* may have responded to soluble *n*-alkanes as well as other components including benzene and toluene.

To further asses the ecological relevance of SV and OOV *Cycloclasticus* to DWH we compared the similarities in the pHMO phylogenetic placement between SV- and OOV-MAGs and previously published transcripts from DWH subsurface plumes (Fig. 15) (Rivers et al., 2013). These results indicate that pHMOs most closely related to SV and OOV *Cycloclasticus* were expressed at high relative abundance during the DWH event, consistent with data showing the rapid microbial response by *Cycloclasticus* to short chain n-alkanes (but not methane) concurrent with active discharge (Valentine et al., 2010). The pulse of bacterial growth in the deep ocean from the DWH event has been estimated at > 10^{23} cells, with a substantial fraction being SV *Cycloclasticus* (Valentine et al., 2012). We therefore questioned if this level of ecological disturbance might have structured the hydrocarbon-degrading community in the GOM through 2015, when samples for this work were collected. While there is insufficient data to rigorously assess this hypothesis, previous work suggests that methanotrophic biomass remained elevated in the years following the DWH event, perpetuating elevated methanotrophic activity above the background levels existing prior to the disaster (Rogener et al., 2018). Therefore, it remains possible SV *Cycloclasticus* observed in our pentane incubations were poised to bloom 5 years following the spill due to some form of a memory effect from the large influx of biomass caused by the disaster.

3.5.6 Hydrocarbon Metabolisms Across *Cycloclasticus*

To gain an understanding for how hydrocarbon metabolic capability within *Cycloclasticus* relates to ecological and evolutionary patterns, we reconstructed *Cycloclasticus* metagenomes from a variety of environments using publicly available

datasets. This effort resulted in the curation of eight high-quality MAGs with completion >80% and <2% redundancy. These eight MAGs are in addition to the five pentane MAGs and the one DWH MAG, and include one from the uncontaminated North Sea “NS_1”, six from a coastal salt marsh in Skidaway Island, Georgia, “CSM_1”, “CSM_2”, “CSM_3”, “CSM_4”, “CSM_5”, and “CSM_6”, and one metagenome from coastal seawater near Pivers Island, North Carolina “CSW_1”. The 14 genomes reconstructed for this chapter along with other publicly available genomes were used to form a phylogenomic tree of all *Cycloclasticus* (Fig. 17a). Each genome was then scanned for hydrocarbon-related

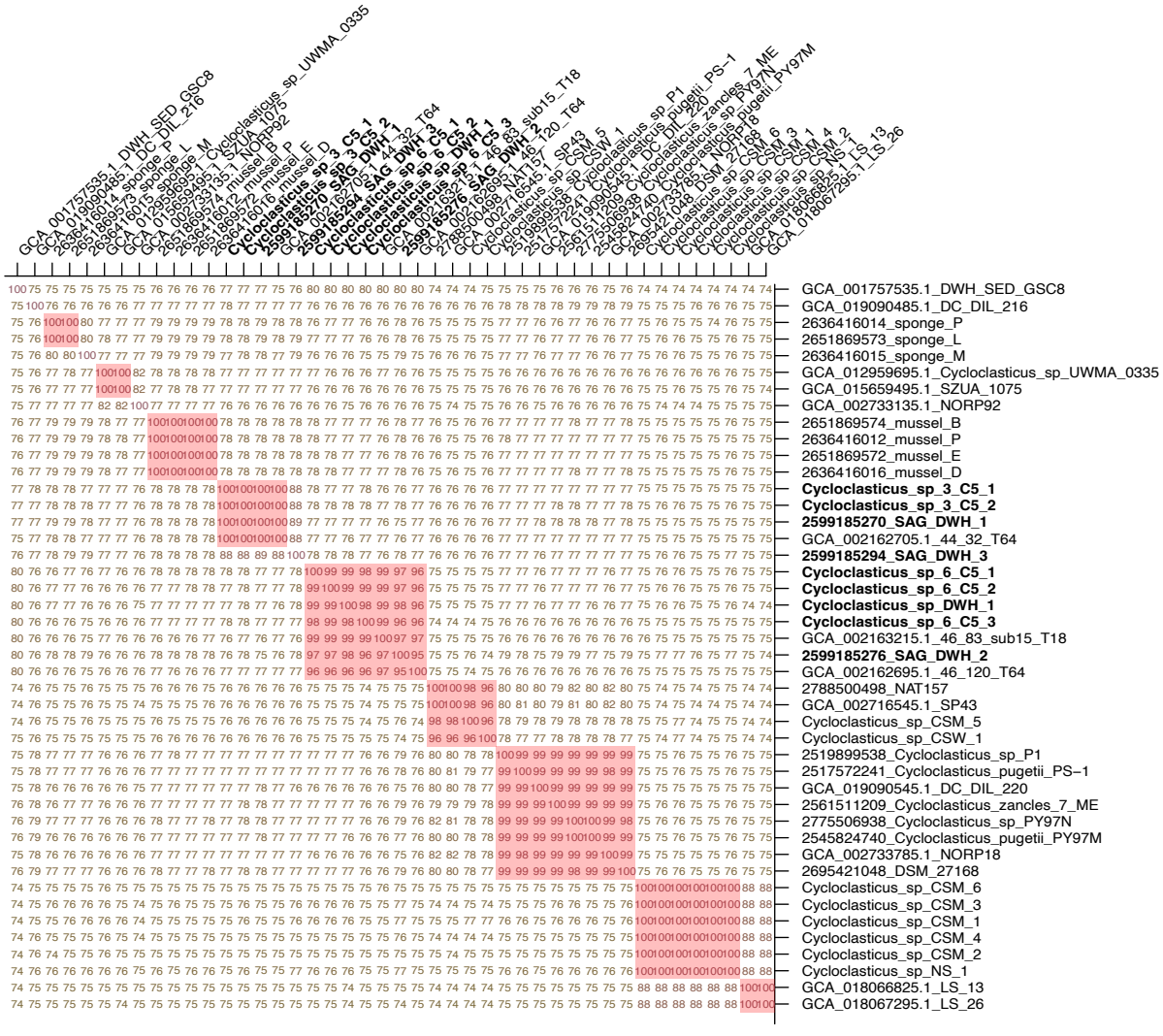


Figure 20. Average nucleotide identity of *Cycloclasticus* genomes. SV and OOV are highlighted in bold.

pathways of interest and for other metabolic functions related to energy generation (Fig. 17b).

From this analysis, we observe distinct metabolic strategies by each major clade within the *Cycloclasticus* genera (Fig. 17b). Notably, all cultivated *Cycloclasticus* are very closely related to each other (Fig. 5a, bottom clade, denoted 'isolates'), and within this clade we found no evidence for genes involved in the use of short-chain alkanes. This is a major bias in our present understanding of *Cycloclasticus*, especially notable because all other *Cycloclasticus* genomes analyzed contained pHMO genes. These pHMO genes were genetically distinct from pMMOs across all currently analyzed *Cycloclasticus* genomes (Fig. 15). We also observe two water-column clades from uncontaminated seawater that harbor diverse pathways for short-chain and long chain alkanes, as well as near-complete pathways for naphthalene and xylene, and complete pathways for toluene and benzene consumption. In all, we find a minimum of seven clades within the *Cycloclasticus*, seemingly unified as marine organisms that grow from aqueous soluble hydrocarbons. One key factor distinguishing the clades is the evolved preference to access certain classes of aqueous soluble hydrocarbons and not others.

3.6 Conclusion

Our study provides genomic and proteomic evidence for pentane metabolism by free-living *Cycloclasticus* in contrasting oceanic regimes, one with prolific natural seep influence and another far removed from prolific seepage. Through the comparison of *Cycloclasticus* genomes and metagenomes we show that the hydrocarbon metabolism within this genus is not limited to PAH degradation, with genomic variability enabling different ecotypes to access different ecological niches and structural classes of hydrocarbons. The apparent commonality among *Cycloclasticus* is not the ability to consume aromatic hydrocarbons, as the genus name suggests, but rather metabolic

specialization among the subset of hydrocarbons that exhibit aqueous solubility in marine settings.

Our results further expand on previous findings that illuminated the contrasting strategies between cultivated *Cycloclasticus* relying solely on aromatic hydrocarbons and mussel and sponge symbionts that primarily consume short-chain alkanes. We identify distinct clades of free-living *Cycloclasticus* that further expand on these contrasting specializations: a seep variant clade that selectively targets short-chain alkanes using pHMO, and an open ocean variant clade that exhibits broader hydrocarbon versatility and the ability for anaerobic metabolism. A versatile metabolism could be advantageous for *Cycloclasticus* in the ephemeral natural seep environment or in the context of accessing non-seep hydrocarbon sources.

The specialization of SV *Cycloclasticus* on short-chain alkanes is perhaps more perplexing because of the apparent ecological risk. The apparent strategy of these SV *Cycloclasticus* requires a consistent supply of short-chain alkanes for growth, which are supplied almost exclusively from petroleum-rich environments such as seeps or spills. Given the geographic constraints on petroleum seepage and the ephemeral nature of discharge, it is perhaps surprising that specialist bacterioplankton have evolved into this niche. Nonetheless, results of our experiments and the clear success of SV *Cycloclasticus* during DWH indicate that such specialization results in a successful ecological strategy. The success of SV *Cycloclasticus* is potentially related to rapid cellular respiration that presumably enables rapid response and growth upon exposure to substrate.

The insight gained from this work provides a new vantage to consider the deep ocean microbial response to hydrocarbon discharge during DWH. SV *Cycloclasticus* were well adapted to bloom in response to the massive intrusions of aqueous-soluble *n*-alkanes that accompanied this event. The OOV *Cycloclasticus* may have engaged in direct competition by consuming these same compounds, but they may have also accessed other

compounds in parallel or in sequence. This work goes beyond the DWH and provides a predictive capacity for understanding the ocean's response to future industrial incidents on a variety of scales such as a rupture of a subsea pipeline or the sinking of a tanker vessel carrying gas condensate, light crude oil, or diluted bitumen; or another well blowout.

Chapter 4

4. Methylated Cycloalkanes Fuel a Novel Genera in the *Porticoccaceae* Family

4.1 Background

4.1.1 Cycloalkanes in Petroleum

In this chapter, we focus on the microbial metabolism of methylcyclohexane (MCH) and methylcyclopentane (MCP) in the deep (1,000m) Gulf of Mexico. Cycloalkanes were chosen for study for several reasons. First, they are abundant in subsurface reservoirs and are thus quantitatively important to both natural seeps and to spills. For example, over 50 different cycloalkane structures were identified in one bitumen sample from Green River Shale and authors noted many structures were composed of cyclopentane and cyclohexane rings originating from ancient isoprenoid lipids and carotenoid pigments (Anders and Robinson, 1971). The specific cycloalkanes studied in this chapter, MCH and MCP, are often the most abundant cycloalkanes observed in petroleum. Cycloalkanes composed 16% of the total oil and gas released in the Macondo oil from the Deepwater Horizon event, and MCP and MCH together represented 2.4% of the total (Reddy et al., 2012). Second, cycloalkanes belong to the class of water-soluble volatile alkanes (Discussed in Chapter 3.3.1) that display distinct behaviors compared with traditional oil. Cycloalkanes will partition into seawater or the atmosphere depending on the context by which they enter the ocean, a set of characteristics which make them unsuitable for traditional fate and transport models that govern our understanding of liquid oil. Third, these compounds form their own product streams – known as natural gas condensate or natural gas liquids – with potential for ocean discharge (Mokhatab and Poe, 2012). These product streams are also used as diluents to enable pipeline transport of heavy oils and bitumen, with multiple high-profile ruptures of diluent-containing pipelines in the recent past (Fitzpatrick et al., 2016; National Academies Press, 2016; Perez et al., 2016; Walker et al., 2016). Finally, these volatile liquid compounds are known to have toxicological impacts on

aquatic organisms and pose human health risks via air emissions, a major concern during oil spill response operations and for communities near spill areas (Peterson, 1994; McKenzie et al., 2012).

4.1.2 Cycloalkane Metabolism

Despite their abundance in petroleum, only a few seminal studies have focused on cycloalkane metabolism. This work began with Imelik in 1948 with the isolation of the first strain of bacteria (*Pseudomonas aeruginosa*) capable of degrading cyclohexane (Imelik, 1948). Insights from early cultivation studies suggest that in comparison to *n*-alkanes such as pentane, cycloalkanes are more resistant to microbial attack due to additional metabolic steps required for ring cleavage (Ooyama and Foster, 1965; Perry and Gibson, 1977a). In cultivated isolates, degradation of substituted cycloalkanes appears to occur more readily than the degradation of unsubstituted forms, particularly if there is an *n*-alkane substitution of adequate chain length (Stirling et al., 1977). This preference towards the alkylated saturated rings occurs because microbes beta-oxidize the *n*-alkyl chains on cyclic compounds prior to attacking the saturated ring. Longer chain lengths allow for more rounds of beta-oxidation before the *n*-alkyl chains are 1-2 carbons in length which blocks beta-oxidation and initiates ring cleavage (Willetts and Cain, 1972; Beam and Perry, 1974). MCH and MCP specifically are of interest because their carboxylic acid counterparts (cyclohexane carboxylic acid and cyclopentane carboxylic acid) are often intermediate compounds in beta-oxidized alkylated cycloalkane and alkylated cyclo-carboxylic acid (also called naphthenic acid) degradation (Perry and Gibson, 1977b; Herman et al., 1993).

Stirling et al. 1977 first characterized cyclohexane metabolism in Nocardia finding growth, respiration, and enzymes from culture extracts are consistent with cyclohexane degradation via cyclohexanol, cyclohexanone, caprolactone, and ε-hydroxycaproate (Stirling et al., 1977). This metabolic route for cyclohexane consumption was further corroborated in *Pseudomonas* and *Xanthobacter* cultures. Enzymes used in this pathway

include cyclohexanol dehydrogenase, cyclohexanone monooxygenase, ε caprolactone hydrolase (Anderson et al., 1980; Trower et al., 1985). Tonge and Higgins found the isolate *Norcardia petroleophilia* grows on methylcyclohexane and found 3-methylcyclohexanol and 3-methylhexanone in the growth medium during exponential growth (Tonge and Higgins, 1974). The cyclopentane degradation pathway, first elucidated by (Griffin and Trudgill, 1972), occurs via cyclopentanol, cyclopentanone, 5-valerolactone, 5-hydroxyvalerate, glutarate, and acetyl-CoA (Griffin and Trudgill, 1972). This pathway has been also observed in *Comamonas* sp. and *Acinetobacter* with significant efforts to define key enzymes including: cyclopentane monooxygenase, cyclopentanol dehydrogenase, cyclopentanone 1,2-monooxygenase, a ring-opening 5-valerolactone hydrolase, 5-hydroxyvalerate dehydrogenase, and 5-oxovalerate dehydrogenase (Iwaki et al., 2002). These seminal works on cycloalkane enzymology are vital points of comparison for our study.

4.1.3 Particulate Ammonia/Hydrocarbon Monooxygenases

The first step in hydrocarbon consumption is oxidation. For methylated cycloalkanes initial oxidative attack can be performed by a number enzymes including the soluble non-heme iron monooxygenase (s*MMO*), soluble cytochrome P450, non-heme iron monooxygenase (*alkB*), flavin-binding monooxyenases (*alma*), or the particulate copper-containing monooxygenase (*phmoCAB)* (Moreno and Rojo, 2017). In this chapter, I focus on the p*hmoCAB* enzyme due to its conserved nature within the metagenomes of methylcyclohexane and methylcyclopentane consumers. The particulate hydrocarbon monooxygenase (*phmoCAB*) enzyme class was originally named *pMMO* (particulate methane monooxygenase) due to its original classified activity on methane, however in recent years its substrate specificity has expanded to ethane, ethylene, propane, butane, pentane and therefore we use the term particulate hydrocarbon monooxygenase (*phmo*) (Tavormina et al., 2011). Genes encoding *phmo* and ammonia monooxygenase share high

sequence identity and structural analysis has confirmed they are evolutionarily related (Holmes et al., 2006).

4.2 Methods

4.2.1 Incubation Design and Sample Collection

Seawater samples were collected aboard RV Atlantis in June 2015. MCP and MCH incubations were conducted at stations 1 (27° 30.41' N, 87° 12.41' W), 2 (27° 15.00' N, 89° 05.05' W), 3 (27° 11.60' N, 90° 41.75' W) and 4 (27° 38.40' N, 90° 54.98' W) with seawater collected from 1,000 m. Seawater collected from the CTD Niskin bottles was transferred to 250 mL glass serum vials using a small length of Tygon tubing. Vials were filled for at least 3 volumes of water to overflow. Care was taken to ensure no bubbles were present before sealing with a polytetrafluoroethylene (PTFE) coated chlorobutyl rubber stopper and crimp cap seal. All bottles, except for unamended blank controls, immediately received 10 μL of MCH or MCP using a gas-tight syringe (Hamilton) and were maintained in the dark at in-situ temperature (4ºC). Prior to filling, each serum bottle was fixed with a contactless optical oxygen sensor (Pyroscience, Aachen, Germany) on the inner side with silicone glue, and afterward were cleaned from organic contaminants with rinses of ethanol, 3% hydrogen peroxide, 10% hydrochloric acid, and MilliQ water, and were sterilized via autoclave. Oxygen concentration was monitored approximately every 8 hours with a fiber optic oxygen meter (Pyroscience, Aachen, Germany). Observed changes in oxygen content were normalized to unamended controls to correct for oxygen loss from background respiration processes and variability due to temperature changes. Bloom onset is operationally defined as three consecutive time points with oxygen loss >0.21 μM h^{-1}. At the end of each respiration experiment incubations were harvested and filtered on a 0.22-μm polyethersulfone filter.

4.2.2 DNA extraction, 16S rRNA community analysis

DNA extraction was performed from ¼ of each filter using the PowerSoil DNA extraction kit with the following modifications: 200 µl of bead beating solution was removed at the initial step and replaced with phenol chloroform, the C4 bead binding solution was supplemented with 600 µL of 100% ethanol, and we added an additional column washing step with 650 µL of 100% ethanol. Extracts were purified and concentrated by ethanol precipitation, then stored at -80ºC. The V4 region of the 16S rRNA gene was amplified using the method described by (Kozich et al., 2013) with small modifications to the 16Sf and 16Sr primers according to (Apprill et al., 2015; Parada et al., 2016). Amplicon PCR reactions contained 1 µL of template DNA, 2 µL of forward primer, 2 µL of reverse primer, and 17 µL of AccuPrime Pfx SuperMix. Thermocycling conditions consisted of 95° 2 min, 30 cycles of 95°C for 20 secs, 55°C for 15 secs, 72°C for 5 min, and a final elongation at 72°C for 10 min. Sample DNA concentrations were normalized using the SequelPrep Normalization Kit, cleaned using the DNA Clean and Concentrator kit, visualized on an Agilent Tapestation, and quantified using a Qubit Flurometer. Samples were sequenced at the UC Davis Genome Center on the Illumina MiSeq platform with 250nt, paired end reads. A PCR-grade water sample was included in extraction, amplification, and sequencing as negative control to assess for DNA contamination.

Trimmed fastq files were quality filtered using the fastqPairedFilter command within the dada2 R package, version 1.9.3 (Callahan et al., 2016) with following parameters: truncLen=c(190,190), maxN=0, maxEE=c(2,2), truncQ=2, rm.phix=TRUE, compress=TRUE, multithread=TRUE. Quality filtered reads were dereplicated using derepFastq command. Paired dereplicated fastq files were joined using the mergePairs function with the default parameters. A Single Nucleotide Variant (SNV) table was constructed with the makeSequenceTable command and potential chimeras were removed denovo using removeBimeraDenovo. Taxonomic assignment of the sequences was done with the assignTaxonomy command using the Silva taxonomic training dataset formatted

for DADA2 v132 (Pruesse et al., 2007; Glöckner et al., 2017). If sequences were not assigned, they were left as NA.

4.2.3 Metagenomic Reconstruction

Metagenomic library preparation and shotgun sequencing were conducted at the University of California Davis DNA Technologies Core. DNA was sequenced on the Illumina HiSeq4000 platform, producing 150-base pair (bp) paired-end reads with a targeted insert size of 400 bp. Quality control and adaptor removal were performed with Trimmomatic (Bolger et al., 2014) (v.0.36; parameters: leading 10, trailing 10, sliding window of 4, quality score of 25, minimum length 151 bp) and Sickle (Joshi and Fass, 2011) (v.1.33 with paired-end and Sanger parameters). The trimmed high-quality reads were assembled using metaSPAdes (Nurk et al., 2017) (v.3.8.1; parameters $k = 21, 33, 55, 77, 88, 127$). The quality of assemblies was determined using QUAST (Gurevich et al., 2013) (v.5.0.2 with default parameters). Sequencing coverage was determined for each assembled scaffold by mapping high-quality reads to the assembly using Bowtie2 (Langmead and Salzberg, 2012)(v.2.3.4.1; default parameters) with Samtools (Li et al., 2009) (v.1.7). Contigs greater than 2,500 bp were manually binned using Anvi'o with Centrifuge (Eren et al., 2015; Kim et al., 2016) (v.1.0.1) based on coverage uniformity (v.5). Quality metrics for metagenome-assembled genomes (MAGs) were determined using CheckM (Parks et al., 2015) (v.1.0.7; default parameters). The taxonomy of each MAG was classified using GTDB-Tk (v.1.0.2) against The Genome Taxonomy Database (Chaumeil et al., 2020) (https://data.ace.uq.edu.au/public/gtdb/data/releases/release89/89.0/, v.r89). The average nucleotide identity of each genome was determined with the ANI Matrix via the Enveomics tool collection (Rodriguez-R and Konstantinidis, 2016).

4.2.4 Metagenome Annotation

Open reading frames were predicted for MAGs using Prodigal (Hyatt et al., 2010) (v.2.6.3; default parameters). Functional annotation was determined using HMMER3 (Eddy,

2011)(v.3.1b2) against the Pfam database (v.31.0) with an expected value (e-value) cutoff of 1×10^{-7} and KofamScan (v.1.1.0) (Aramaki et al., 2020) against the Hidden Markov model (HMM) profiles for Kyoto Encyclopedia of Genes and Genomes and Kegg Orthology (KEGG/KO) with a score cutoff of 1×10^{-7}. To find hits for *almA* we used Pfam (PF00743), for *rhdA* we used Pfam (PF00848) and for *pHMO* we summed KO hits (subunit a: K10944; subunit b: K10945; subunit c: K10946). For alkane-1-monooxygenase (*alkB*) detection we used KofamScan with K00496 to search for *alkB*.

4.2.5 Phylogenetics

To define genome phylogenomic relationships of MAGs, 16 universal ribosomal proteins (RPs) were used (L2-L6, L14-L16), L18, L22, L24, S3, S8, S10, S17, and S19. For phylogenies of metabolic genes as well ribosomal proteins, all representative sequences and concatenated alignments that contained <25% informative sites were excluded in tree

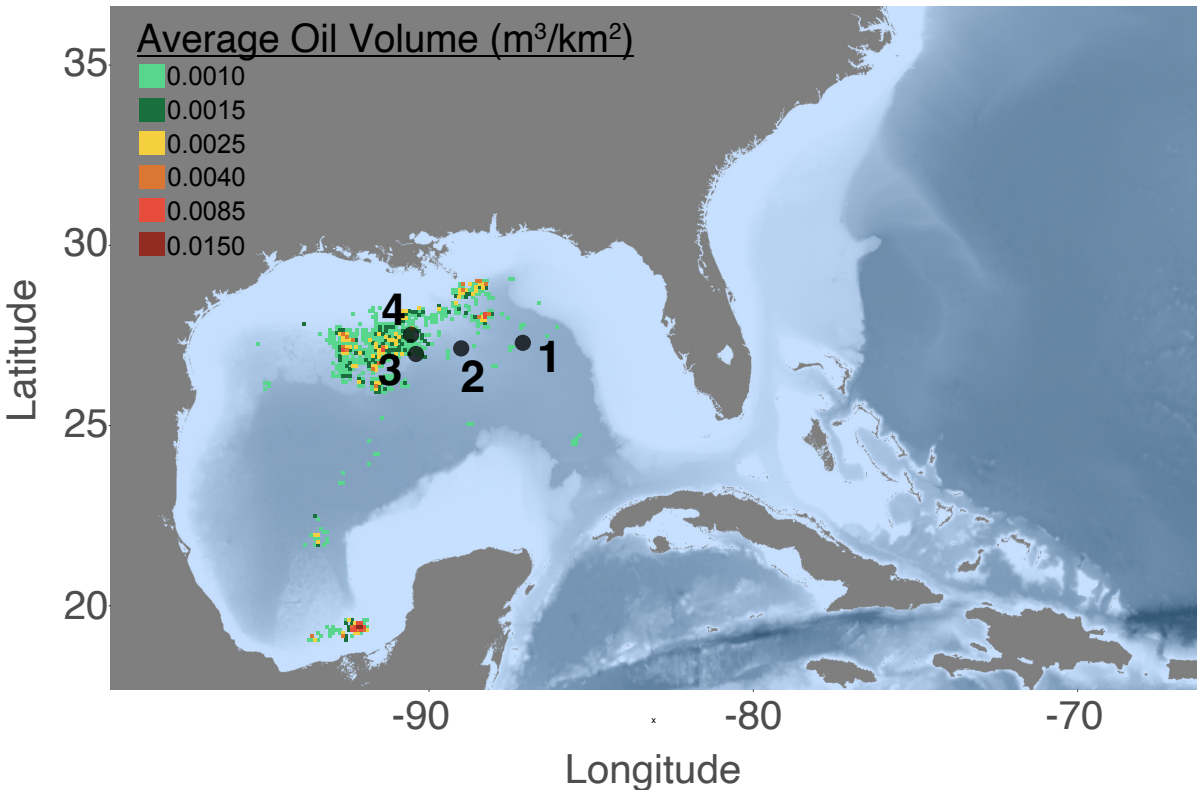


Figure 21. Sampling stations relative to natural oil seepage in the Gulf of Mexico.

construction. Each protein was aligned using MUSCLE (v.3.8.425) (Edgar, 2004). All columns with >95% gaps were removed using TrimAL (Capella-Gutiérrez et al., 2009). Maximum-likelihood phylogenetic analysis of concatenated alignment was inferred by RAxML (Stamatakis, 2014) (v.8..9; parameters: raxmlHPC -T 4 -s input -N autoMRE -n result -f a -p 12345 -x 12345 -m PROTCATLG). Resulting trees were visualized using FigTree (FigTree) (v.1.4.3).

4.3 Results and Discussion

4.3.1 Methylated Cycloalkane Bloom Occurrence

Cycloalkane incubations were conducted in parallel with pentane incubations in the Gulf of Mexico (Fig. 21, Chapter 1, and Chapter 3). Incubations included enrichments with MCH and MCP. Three out of eleven incubations at the station furthest from natural seepage (Station 1) bloomed with MCP whereas zero out of eleven incubations bloomed with MCH at the same station. In contrast, at locations closet to seepage(stations 2, 3, and 4) over 50% of the incubations bloomed within 18-21 days for both MCP and MCH treatments. The average respiration profile for blooms on MCP and MCH showed similar patterns in oxygen

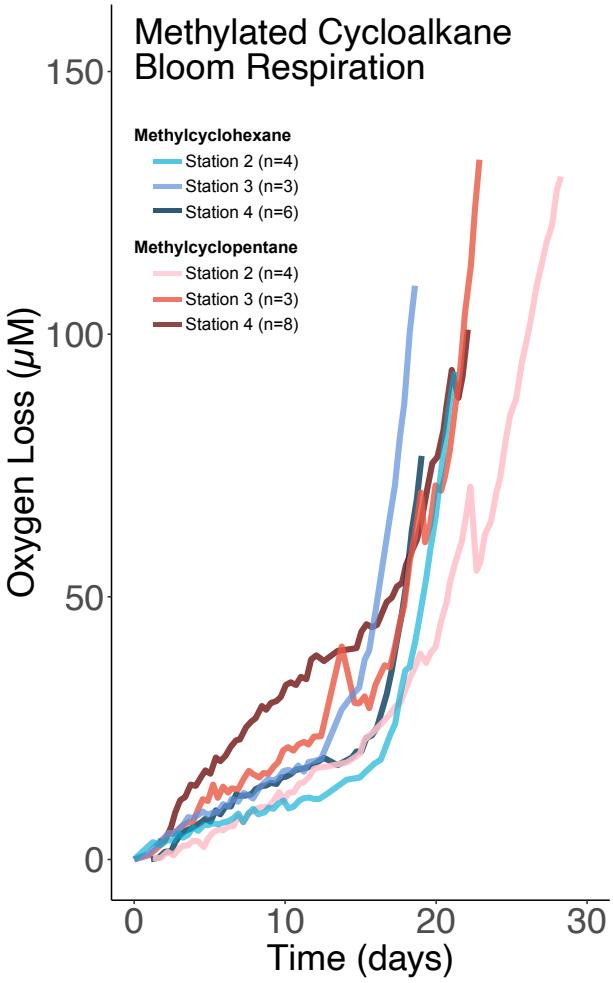


Figure 22. Average oxygen loss with time for blooms (defined by oxygen loss >0.21 μM h^{-1}) on methylcyclopentane and methylcyclohexane at station 2, 3, and 4.

loss with time across each station (Fig. 22).

4.3.2 Cycloalkane Community Analysis

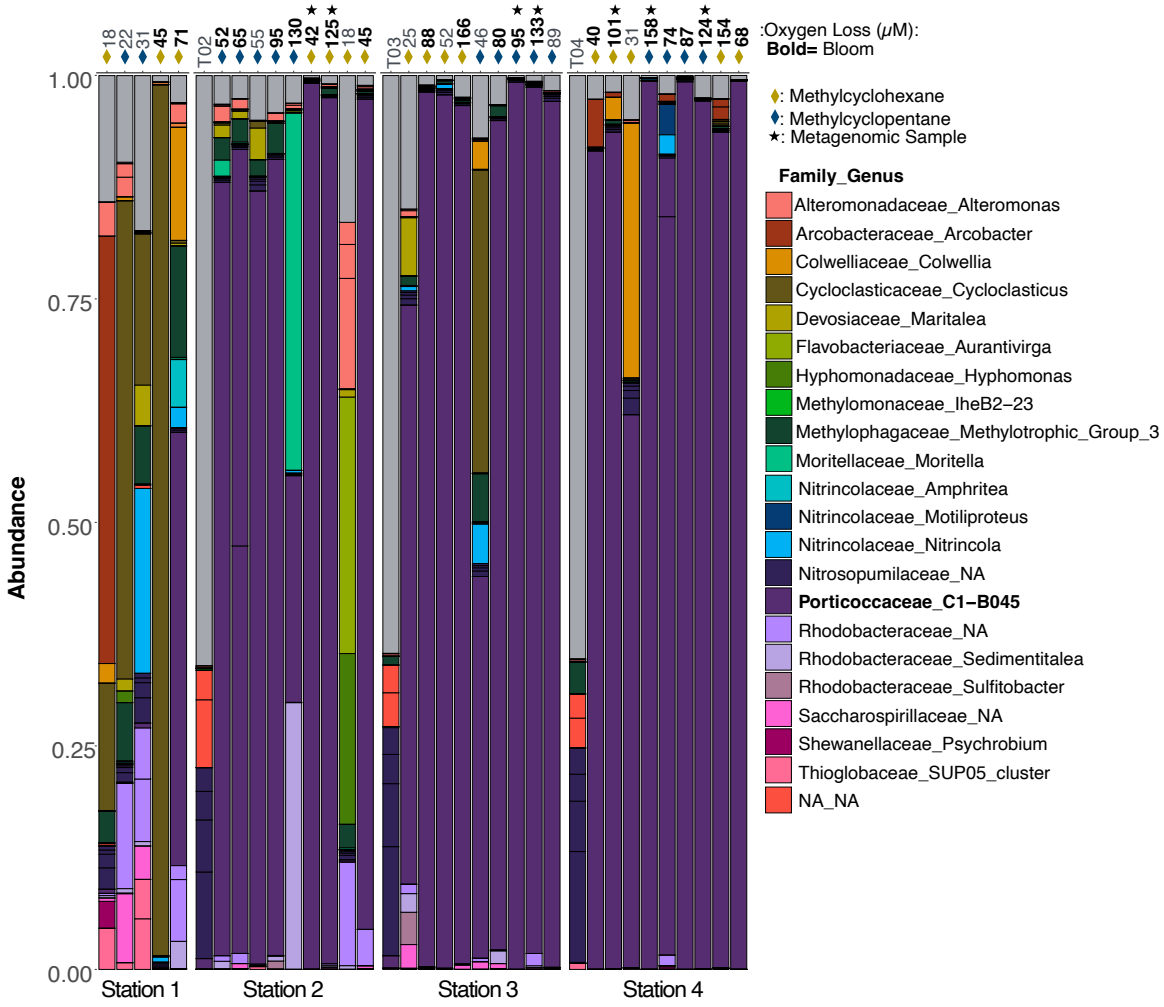


Figure 23. Microbial characteristics of blooms. Microbial community composition of methylcyclohexane (MCH) and methylcyclopentane (MCP) blooms informed via the V4 region of the 16S rRNA gene for initial environmental samples (denoted "T0") and incubations harvested after significant oxygen loss was observed. "NA_NA" taxonomic representative at the family or genera level. Diamond color indicates which compound was added (MCP or MCH). Oxygen loss at DNA sample harvest is labeled for each sample with bold samples indicating bloom occurrence (i.e., sample had 3 time points with >0.21 μM O_2 loss per hour). Black stars note which samples were sequenced further for metagenomic analyses.

Methylcyclohexane and methylcyclopentane enrichments typically elicited the emergence of a single dominant taxa comprising >50% of the 16S rRNA amplicon sequences across all stations (Fig. 23). The full taxonomy of the closest match to the dominant SNV (observed in all but two samples regardless of bloom status) is as follows in

the Silva and NCBI database: Bacteria; Proteobacteria; Gammaproteobacteria; *Cellvibrionales*; *Porticoccaceae*; *C1-B045* (Glöckner et al., 2017). Blooms containing abundant taxa other than *C1-B045* occurred four times out of the twenty-two methylated cycloalkane blooms sequenced. These other blooming taxa belong to the genera *Cycloclasticus*, *Colwellia,* and *Moritellaceae* (each >30% abundant in a single bloom). Only one bloom at station 1 with *Cycloclasticus* occurred without the co-occurrence of *C1-B045*. The *C1-B045* SNV is a novel (uncultivated) bacteria at the genera level that was first detected in deep-sea hydrothermal vents (Teske et al., 2002), has been detected in other oil contaminated regions (Liao et al., 2015), and is distantly related to SAR92 (Stingl et al., 2007). To the best of our knowledge, a complete phylogenomic analysis of the *Porticoccacea* family has never been conducted.

4.3.3 Genome Reconstruction and Taxonomic Inferences

We reconstructed high quality metagenomes from seven MCP and MCH treatments in a state of active respiratory bloom, with completeness >97% and redundancy <2% (black stars in Fig. 23, Table 3). Two MAGs, named “MCH_2_108” and “MCH_2_109”, originated from station 2 (further from natural seepage) and MCH enrichment, and two MAGs, named “MCP_3_146” and “MCP_3_148”, originated from MCP enrichments at station 3; and three MAGs, named “MCP_4_184”, “MCP_4_160”, and “MCH_4_158” originated from MCP and MCH blooms in the natural seep region at station 4. Each MAG was recovered from biologically independent incubations, yet every major component of metabolism and every taxonomic marker was nearly identical across each; therefore, we will refer to them as “B045-MAG”. Using our high-quality genomes as queries to the Genome Taxonomy Database we found the family *Porticoccacea* has recently been reclassified to resolve polyphyletic phylogenetic relationships among the *Cellvibrionales* family (Parks et al., 2020). The taxonomy of the B045-MAGs according to GTDB-Tk is Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales; *Porticoccacea; 500-400-T64.* The genera

designation "500-400-T64" refers to non-standard placeholder name assigned by GTDB-Tk. There have been no genus names proposed formally for this group of genomes, so for the purpose of this study we refer to the genera as B045.

MAG Name	Number of Contigs	Genome Size	Main genome contig N/L50:	Completion	Redundancy
MCH_2_108	4	3.156 MB	2/832.349 KB	100.0%	0.5%
MCH_2_109	3	3.118 MB	2/1.058 MB	99.8%	0.5%
MCP_3_148	4	3.141 MB	2/1.167 MB	99.8%	0.5%
MCP_3_146	4	3.159 MB	1/1.823 MB	99.8%	0.6%
MCP_4_184	5	3.113 MB	2/732.488 KB	99.8%	0.5%
MCP_4_160	11	3.115 MB	2/579.437 KB	99.8%	0.5%
MCH_4_158	156	2.762 MB	24/36.15 KB	96.0%	0.5%

Table 3. Bin statistics for metagenomic constructions from this study. Completion and redundancy were calculated with CheckM.

4.3.4 Methylcyclohexane Metabolism

Within the B045-MAGs we found genomic potential for cycloalkane utilization for catabolism and anabolism. We found evidence for both MCP and MCH metabolism, however given there is a large amount of redundancy between the two pathways and the larger number of studies conducted on cyclohexane metabolism in cultured isolates, we have focused on the compound MCH in this chapter (Fig. 24 and Table 4). The first step in the consumption of MCH in our hypothesized pathway is the oxidation to 3-methylcyclohexanol. Previous work on *Norcardia petroleophilia* suggests oxidation acts on the 3- carbon position during methylcyclohexane metabolism, therefore we have reconstructed the methylcyclohexane pathway from 3-methycyclohexanol; however, it is possible B045-MAG initially oxidizes another carbon position or the methyl group as we also observe genes for cyclohexane carboxylic acid metabolism (Tonge and Higgins, 1974).

We hypothesize that the first step MCH metabolism is catalyzed by the copper-containing membrane associated monooxygenase, called particulate hydrocarbon monooxygenases (*phmo*).The most well-characterized *phmo* is the particulate methane monooxygenase, which oxidizes methane to methanol (Jordan et al., 2021a). We searched for other possible enzymes to catalyze the initial oxidation of methylcyclohexane including flavin binding monooxygenases (*almA;* PF00743*)* and alkane-1-monooxgyenase (*alkB;* K00496) finding all *almA* hits corresponded to aldehyde dehydrogenases further down in the cycloalkane metabolic pathway and *alkB* hits were absent from B045-MAGs (hits are closely related to the ancestrally related fatty acid desaturase instead). We did find a single copy of a gene belonging to the cytochrome P450 enzyme family within each B045-MAG, and we do not exclude the possibility that this gene could function to oxidize cycloalkanes. The second step of methylcyclohexane consumption utilizes an alcohol dehydrogenase to convert 3-methylcyclohexanol to 3-methylcyclohexanone.

The third step in the consumption of methylcyclohexane is the conversion of 3-methylcyclohexanol to delta-methyl-epsilon caprolactone using a cyclohexanone monooxygenase that is evolutionarily related to *almA* and catalyzes Baeyer-Villiger oxidations on a number of ketones including cyclopentanones and cyclobutanones (Kim et al., 2008). We found 7 copies of cyclohexanone monooxygenase genes present in each genome of the B045-MAG, possibly illustrating the diversity of cyclic compounds targeted by the organism. In the fourth step of methylcyclohexane consumption the delta-methyl-epsilon-caprolactone is hydrolyzed to from 6-hydroxy-5-methyl-hexanoic acid, which is a function likely performed by the epsilon-lactone hydrolase. Another alcohol dehydrogenase is used in the fifth step of consumption to convert the 6-hydroxy-5-methyl-hexanoic acid to 5-methy adipate semialdehyde. The sixth step of consumption requires an aldehyde dehydrogenase to convert 5-methyl adipate semialdehyde to 5-methyl adipate. In the

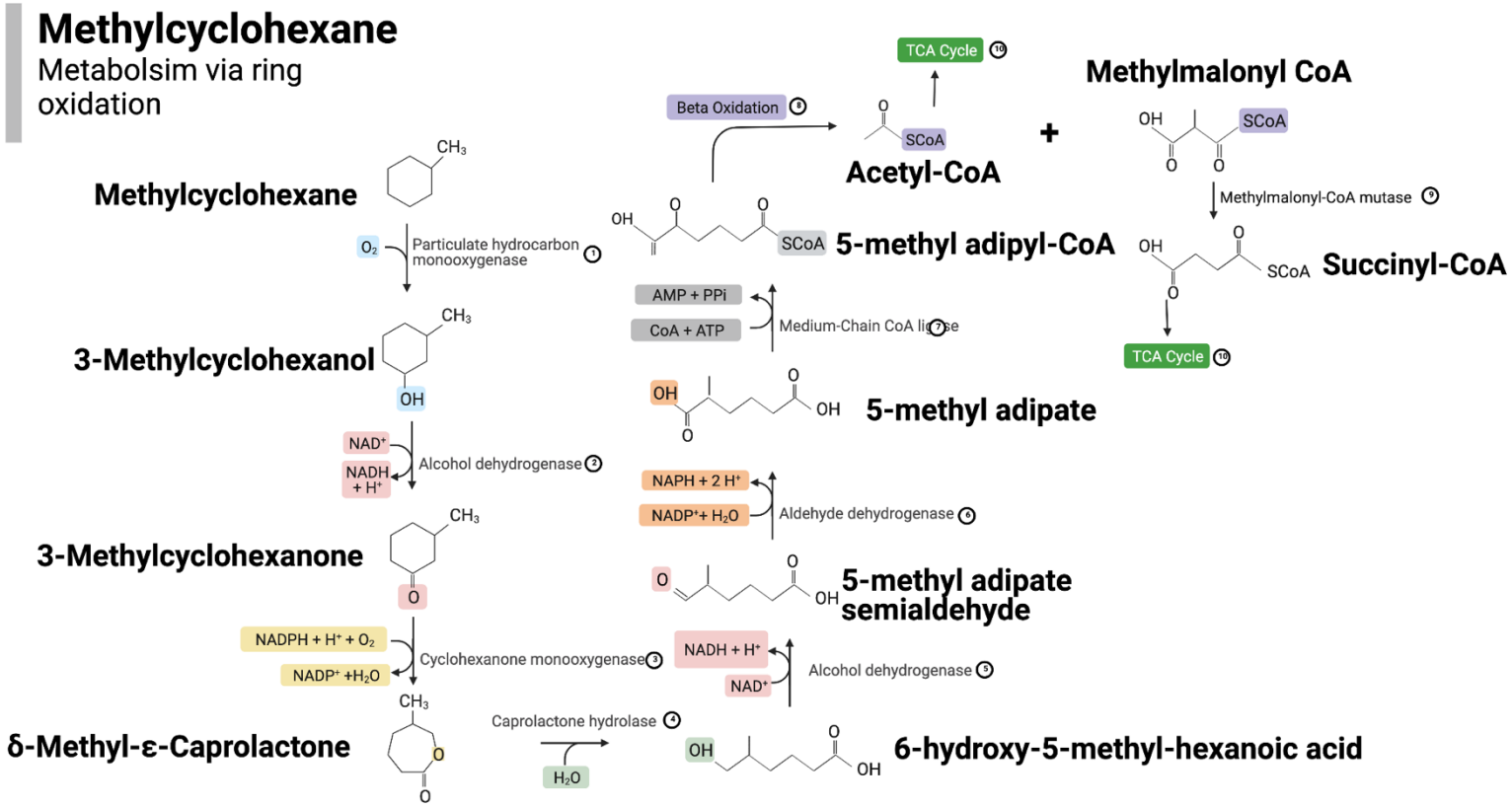


Figure 24. Hypothetical pathway for methylcyclohexane consumption via ring oxidation. Enzymes are denoted by each number with a circle around it. Numbers correspond to the information in Table 4.

seventh step, medium-chain CoA ligase prepares the compound for beta oxidation by adding CoA resulting in 5-methyl adipyl CoA. Beta-oxidation subsequently occurs with acetyl-CoA and methyl malonyl CoA as the products. Interestingly, the B045-MAG contains many copies of the genes required for beta-oxidation, including 17 individual copies of acyl-CoA dehydrogenase per genome, which may suggest B045-MAG specializes on beta-oxidation of a variety of chemical substrates. The remaining methyl malonyl CoA contains a methyl group in the beta-carbon position preventing traditional beta-oxidation. The methyl malonyl CoA mutase is used in the final step to overcome this obstacle and convert methyl malonyl CoA to succinyl-CoA. Both the acetyl-CoA and succinyl-CoA resulting from the metabolic pathway are shunted into central carbon metabolism via the tricarboxylic acid cycle (TCA).

Metabolic Step	Enzyme Name	KO/KEGG ID
1	Particulate Hydrocarbon Monooxygenase	K10944
2	Cyclohexanol Dehydrogenase	K19960
3	Cyclohexanone Monooxygenase	K00499
4	Gluconolactonase	K03379
	Epsilon-Lactone Hydrolase	K14731
5	Alcohol Dehydrogenase	K13954
6	Aldehyde Dehydrogenase	K00138
7	Medium-Chain CoA Ligase	K23617
8	Acyl-CoA Dehydrogenase	K00249
	Enoyl-CoA hydratase	K01692
	3-Hydroxyacyl-CoA Dehydrogenase	K00022
9	Methylmalonyl-CoA mutase	K01847

Table 4. KEGG Gene Identifiers used for Fig. 24.

4.3.5 Phylogenomic Analysis of B045-MAG

To gain an understanding for how hydrocarbon metabolic capability within the *Porticoccacea* family relates to ecological and evolutionary patterns, we compiled a total of 176 high quality (>70% complete, <5% redundant) *Porticoccacea* genomes from a variety of environments using publicly available data. These environments included marine oligotrophic surface waters, marine mesopelagic waters, coastal sand, hydrothermal vents, marine phytoplankton blooms at a variety of latitudes, contaminated groundwater, and oil mesocosms from a variety of sources including Labrador Sediment, Gulf of Mexico, and the Douglas Channel. Each genome was analyzed using the Genome Taxonomy Database (GTDB-Tk) (Chaumeil et al., 2020) (v.1.0.2) which uses 120 bacterial marker genes to classify the taxonomic identity of each genome and then uses s a threshold of average nucleotide identity >95% to designate microbial species. The 176 genomes were used to construct a phylogenomic tree of the *Porticoccacea* family and each genus was labeled according to designations provided by GTDB-Tk (Fig. 25). For all genera without a cultivated isolate, GTDB-Tk provides a “non-standard placeholder name” which is typically a MAG “isolate name” from the National Center Biotechnology Information (NCBI) database of a genome within that clade (Fig. 25).

The B045-MAGs forms a well-supported monophyletic clade with other genomes within the same genera. Genomes within the *B045* genera all appear to respond to petroleum compounds including two genomes from Gulf of Mexico enrichments with Macondo oil, two genomes from Labrador Shelf sediment inoculated with diesel, a single genome from the Guaymas Basin and a single genome from North Pond subseafloor aquifer. Both Guaymas Basin and North Pond sites are locations where petroleum compounds have been noted in previous studies (Didyk and Simoneit, 1989; Zhou et al., 2020). The most closely related genera to *B045* are the *HB2-32-21* and *FW300_53*. The *HB2-32-21* include genomes from the Douglas Channel incubated with diluted bitumen,

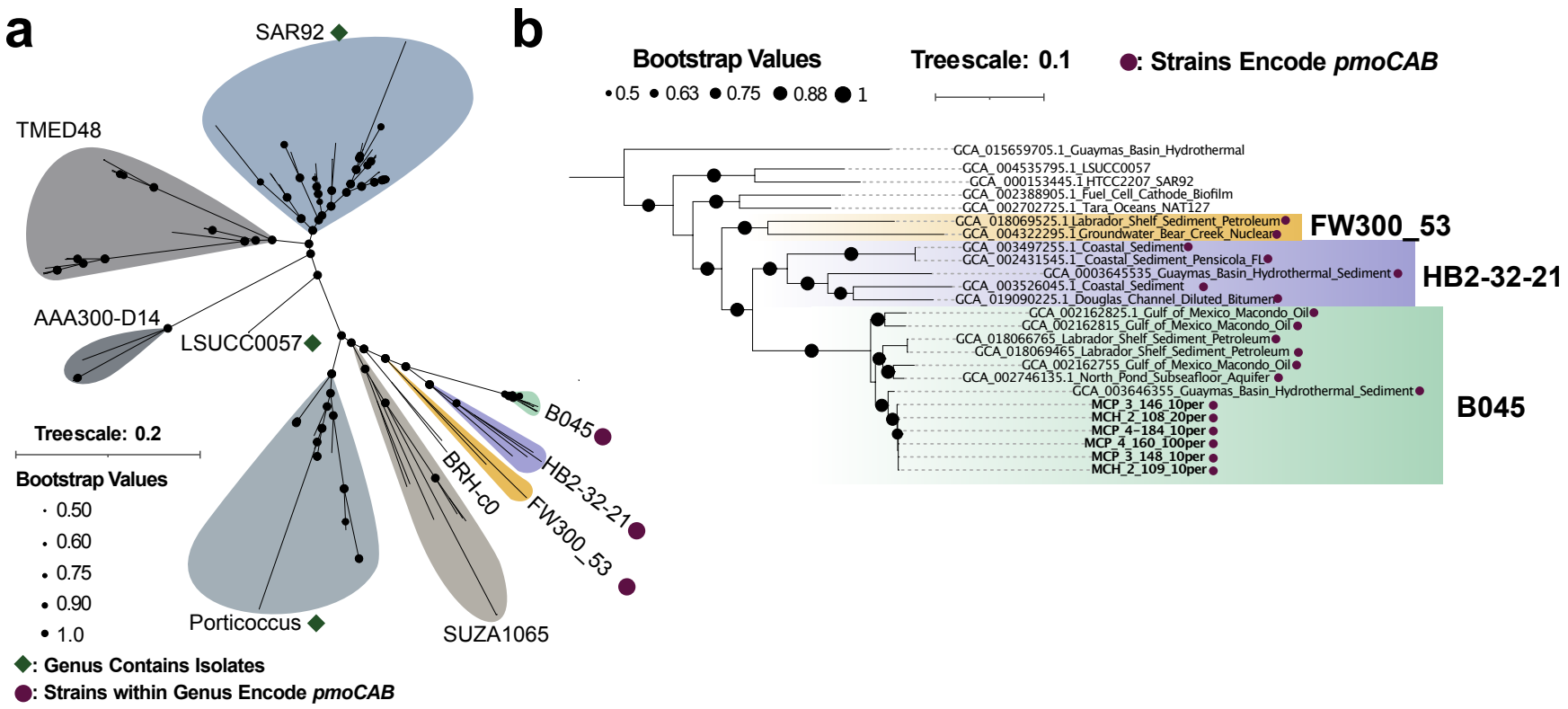


Figure 25. Phylogenomic analysis of *Porticoccacea* ribosomal proteins. **a** Phylogeny constructed using 16 ribosomal proteins, colored subclades are designated and labeled based on GTDB-Tk classification of genera. Cultivated isolates are noted with green diamond. Genera that encode *phmo*CAB are highlighted with maroon circle. **b** Phylogenomic tree of B045-MAG and close relative ribosomal proteins rooted on SAR92. Genomes that encode *phmo*CAB are noted with maroon circle. The tree scale bar represents substitutions per site.

coastal sediment from Pensacola, Florida, and another genome from a Guaymas Basin hydrothermal vent (Fig. 25). The *FW300-53* genera include two genomes: one from Labrador Shelf sediment incubated with diesel and another from groundwater in Tennessee listed as an uncontaminated well nearby sites exposed to early nuclear research under the Manhattan Project (Tian et al., 2020).

4.3.6 CuMMO Phylogeny

Interestingly, out of the 176 genomes analyzed in the *Porticoccacea* family, the only genomes to encode particulate hydrocarbon monooxygenase (*phmo*CAB) belong to the *HB2-32-21*, *B045*, *FW300_53* genera (Fig. 25). This suggests that the acquisition of *phmo* and the resulting metabolic function could be intertwined with the evolution of these three closely related genera. Within the *B045* genera the amino acid identity of the *phmoA* subunit varies from 82-100%. Representatives from the *phmo* and *amo* enzyme class, called the CuMMO enzyme superfamily, can be found across the tree of life of bacteria and archaea and we sought to understand where the *phmo* from *B045* and its close relatives exists within the diversity of the CuMMO superfamily. Here, we analyze the phylogenetic relationship of all archaeal and bacterial *amo* and *phmo* sequences by using the alpha subunit as a phylogenetic marker similar to the method previously done in other studies (Tavormina et al., 2011, 2013). By using B045-MAG *phmoA* sequences to query the National Center Biotechnology Information (NCBI) and Department of Energy Joint Genome Institute (DOE-JGI) databases we also found more distantly related sequences to the novel *phmo* sequences, which originate from diverse environments including contaminated groundwater from mine drilling fluid, agricultural soil, brackish Black Sea water, and hydrothermal vents.

From our phylogenetic analysis, we observe all proteobacterial and Verrucomicrobia *amoA* and *pmo*A (including sequences that function on methane, ethane, butane, and ethylene) form a monophyletic clade which excludes B045-related *phmoA* sequences. Each

sequence related to B045-*phmo*, HB2-32-21-*phmo, FW300_53-phmo,* and the environmental relatives form single monophyletic clade (Fig 26). Our phylogeny is largely consistent with a recently published tree on the CuMMO enzyme superfamily except for the placement of the Verrucomicrobia and NC10 clades, which in our tree branch with the proteobacterial sequences with high bootstrap support. In the previous study these clades branched closer to the gram-positive actinobacteria with poor bootstrap support (Alves et al., 2018). Future research examining the expression of these genes at the RNA or protein level in cultures or environmental incubations, in tandem with extended geochemical analyses, may aid in extending our understanding of the function *B045 phmo* and the role it plays in the metabolism of the *B045* genera.

4.4 Conclusion

Our study provides genomic evidence for methyl-cycloalkane metabolism by the free-living novel *B045* genera within the *Porticoccaceae* family. We have elucidated the pathway for MCH consumption which is the first environmental genomic study of methyl-cycloalkane consumption. Through the comparison of *Porticoccaceae* genomes and metagenomes we show that the acquisition of the particulate hydrocarbon monooxygenase (*phmoCAB*) may have been a key event in the evolution of hydrocarbon metabolism within the *B045* genera and the closely related *HB2-32-21* and *FW300_53* genera. The phylogenetic analysis of the CuMMO enzyme superfamily revealed the placement of the *phmo* from *B045* (and its close relatives) is separated from all other proteobacterial *amo*/*mmo*/*phmo* and it forms its own novel monophyletic clade.

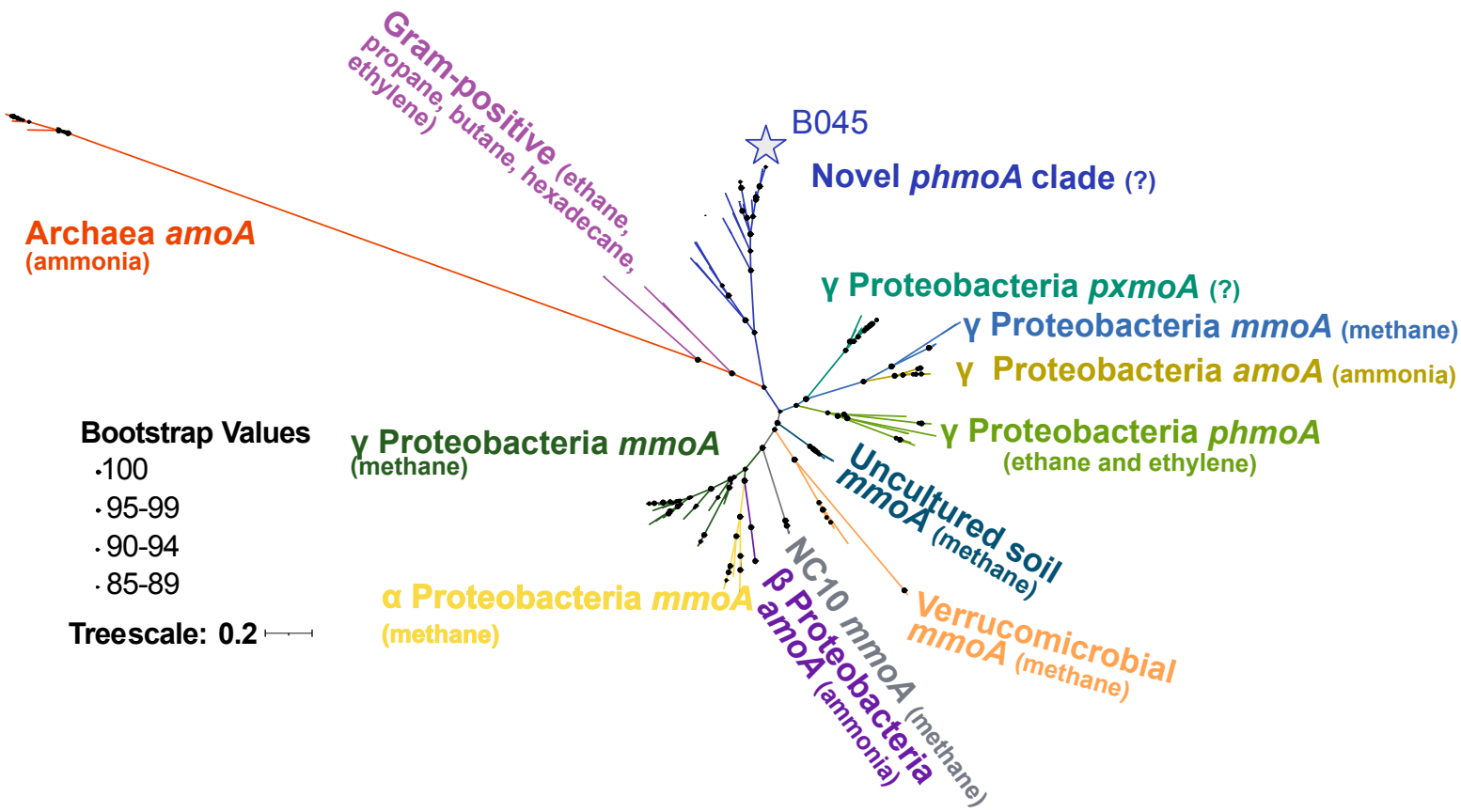


Figure 26. Phylogeny of *phmo* and *amo* subunit A amino acid sequences. Maximum likelihood tree is drawn to scale, with branch lengths representing the number of substitutions per site. Bootstrap values below 50% are not shown. Each major clade is color coded for readability. Tree scale represents the number of substitutions per sits. The B045 sequences are denoted by a star symbol.

References

Adam, P. S., Borrel, G., Brochier-Armanet, C., and Gribaldo, S. (2017). The growing tree of Archaea: New perspectives on their diversity, evolution and ecology. *ISME J.* 11, 2407–2425. doi:10.1038/ismej.2017.122.

Aleksenko, V. A., Anand, D., Remeeva, A., Nazarenko, V. V., Gordeliy, V., Jaeger, K. E., Krauss, U., and Gushchin, I. (2020). Phylogeny and structure of fatty acid photodecarboxylases and glucose-methanol-choline oxidoreductases. *Catalysts* 10, 1–21. doi:10.3390/catal10091072.

Allen, E. E., and Banfield, J. F. (2005). Community genomics in microbial ecology and evolution. *Nat. Rev. Microbiol.* 3, 489–498. doi:10.1038/nrmicro1157.

Alves, R. J. E., Minh, B. Q., Urich, T., Von Haeseler, A., and Schleper, C. (2018). Unifying the global phylogeny and environmental distribution of ammonia-oxidising archaea based on amoA genes. *Nat. Commun.* 9, 1–17. doi:10.1038/s41467-018-03861-1.

Anders, D. E., and Robinson, W. E. (1971). Cycloalkane constituents of the bitumen from Green River Shale. *Geochim. Cosmochim. Acta* 35, 661–678. doi:10.1016/0016-7037(71)90065-2.

Anderson, M. S., Hall, R. A., and Griffin, M. (1980). Microbial metabolism of alicyclic hydrocarbons: cyclohexane catabolism by a pure strain of Pseudomonas sp. *J. Gen. Microbiol.* 120, 89–94. doi:10.1099/00221287-120-1-89.

Apprill, A., Mcnally, S., Parsons, R., and Weber, L. (2015). Minor revision to V4 region SSU rRNA 806R gene primer greatly increases detection of SAR11 bacterioplankton. *Aquat. Microb. Ecol.* 75, 129–137. doi:10.3354/ame01753.

Aramaki, T., Blanc-Mathieu, R., Endo, H., Ohkubo, K., Kanehisa, M., Goto, S., and Ogata, H. (2020). KofamKOALA: KEGG Ortholog assignment based on profile HMM and adaptive score threshold. *Bioinformatics* 36, 2251–2252. doi:10.1093/bioinformatics/btz859.

Arp, G., Reimer, A., and Reitner, J. (1999). Calcification in cyanobacterial biofilms of alkaline salt lakes. *Eur. J. Phycol.* 34, 393–403. doi:10.1080/09670269910001736452.

Atlas, R. M. (1981). Microbial degradation of petroleum hydrocarbons: an environmental perspective. *Microbiol. Rev.* 45, 180–209. Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC281502/.

Bano, N., Ruffin, S., Ransom, B., and Hollibaugh, J. T. (2004). Phylogenetic Composition of Arctic Ocean Archaeal Assemblages and Comparison with Antarctic Assemblages. *Appl. Environ. Microbiol.* 70, 781–789. doi:10.1128/AEM.70.2.781-789.2004.

Beam, H. W., and Perry, J. J. (1974). Microbial degradation and assimilation of n alkyl substituted cycloparaffins. *J. Bacteriol.* 118, 394–399. doi:10.1128/jb.118.2.394-399.1974.

Berg, H. C., and Purcell, E. M. (1977). Physics of chemoreception. *Biophys. J.* 20, 193–219. doi:10.1016/S0006-3495(77)85544-6.

Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. doi:10.1093/BIOINFORMATICS/BTU170.

Brenner, K., You, L., and Arnold, F. H. (2008). Engineering microbial consortia: a new frontier in synthetic biology. *Trends Biotechnol.* 26, 483–489. doi:10.1016/j.tibtech.2008.05.004.

Button, D. K. (1985). Kinetics of Nutrient-Limited Transport and Microbial Growth. *Microbiol. Rev.* 49, 270–297. Available at: https://journals.asm.org/journal/mr [Accessed November 17, 2021].

Button, D. K., and Jüttner, F. (1989). Terpenes in Alaskan waters: Concentrations, sources, and the microbial kinetics used in their prediction. *Mar. Chem.* 26, 57–66.

doi:10.1016/0304-4203(89)90064-9.
Button, D. K., Robertson, B. R., Lepp, P. W., and Schmidt, T. M. (1998). A small, dilute-cytoplasm, high-affinity, novel bacterium isolated by extinction culture and having kinetic constants compatible with growth at ambient concentrations of dissolved nutrients in seawater. *Appl. Environ. Microbiol.* 64, 4467–4476. doi:10.1128/AEM.64.11.4467-4476.1998.
Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J. A., and Holmes, S. P. (2016). DADA2: High-resolution sample inference from Illumina amplicon data. *Nat. Methods 2016 137* 13, 581–583. doi:10.1038/nmeth.3869.
Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., and Madden, T. L. (2009). BLAST+: Architecture and applications. *BMC Bioinformatics* 10, 1–9. doi:10.1186/1471-2105-10-421/FIGURES/4.
Capella-Gutiérrez, S., Silla-Martínez, J. M., and Gabaldón, T. (2009). trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25, 1972–1973. doi:10.1093/BIOINFORMATICS/BTP348.
Caporaso, J. G., Lauber, C. L., Walters, W. A., Berg-Lyons, D., Huntley, J., Fierer, N., Owens, S. M., Betley, J., Fraser, L., Bauer, M., Gormley, N., Gilbert, J. A., Smith, G., and Knight, R. (2012). Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *ISME J.* 6, 1621–1624. doi:10.1038/ismej.2012.8.
Carpenter, E. J., Harvey, H. R., Brian, F., and Capone, D. G. (1997). Biogeochemical tracers of the marine cyanobacterium Trichodesmium. *Deep. Res. Part I Oceanogr. Res. Pap.* 44, 27–38. doi:10.1016/S0967-0637(96)00091-X.
Chaumeil, P.-A., Mussig, A. J., Hugenholtz, P., and Parks, D. H. (2020). GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. *Bioinformatics* 36, 1925–1927. doi:10.1093/BIOINFORMATICS/BTZ848.
Chisholm, S. W., Olson, R. J., Zettler, E. R., Goericke, R., Waterbury, J. B., and Welschmeyer, N. A. (1988). A novel free-living prochlorophyte abundant in the oceanic euphotic zone. *Nat. 1988 3346180* 334, 340–343. doi:10.1038/334340a0.
Chung, W. K., and King, G. M. (2001). Isolation, Characterization, and Polyaromatic Hydrocarbon Degradation Potential of Aerobic Bacteria from Marine Macrofaunal Burrow Sediments and Description of Lutibacterium anuloederans gen. nov., sp. nov., and Cycloclasticus spirillensus sp. nov. *Appl. Environ. Microbiol.* 67, 5585–5592. doi:10.1128/AEM.67.12.5585-5592.2001.
Committee On Oil In The Sea, Ocean Studies Board and Marine Board, N. R. C. (2013). *Oil in the Sea III: Inputs, Fates, and Effects*. National Academies Press.
Coulon, F., McKew, B. A., Osborn, A. M., McGenity, T. J., and Timmis, K. N. (2007). Effects of temperature and biostimulation on oil-degrading microbial communities in temperate estuarine waters. *Environ. Microbiol.* 9, 177–186. doi:10.1111/j.1462-2920.2006.01126.x.
Damashek, J., Okotie-Oyekan, A. O., Gifford, S. M., Vorobev, A., Moran, M. A., and Hollibaugh, J. T. (2021). Transcriptional activity differentiates families of Marine Group II Euryarchaeota in the coastal ocean. *ISME Commun.* 1, 1–11. doi:10.1038/s43705-021-00002-6.
DeLong, E. F. (1992). Archaea in coastal marine environments. *Proc. Natl. Acad. Sci. U. S. A.* 89, 5685–5689. doi:10.1073/pnas.89.12.5685.
DeLong, E. F., Preston, C. M., Mincer, T., Rich, V., Hallam, S. J., Frigaard, N. U., Martinez, A., Sullivan, M. B., Edwards, R., Brito, B. R., Chisholm, S. W., and Karl, D. M. (2006). Community genomics among stratified microbial assemblages in the ocean's interior. *Science (80-.).* 311, 496–503. doi:10.1126/science.1120250.
Didyk, B. M., and Simoneit, B. R. T. (1989). Hydrothermal oil of Guaymas Basin and implications for petroleum formation mechanisms. *Nature* 342, 65–69.

doi:10.1038/342065a0.
Diercks, A.-R., Highsmith, R. C., Asper, V. L., Joung, D., Zhou, Z., Guo, L., Shiller, A. M., Joye, S. B., Teske, A. P., Guinasso, N., Wade, T. L., and Lohrenz, S. E. (2010). Characterization of subsurface polycyclic aromatic hydrocarbons at the Deepwater Horizon site. *Geophys. Res. Lett.* 37, n/a-n/a. doi:10.1029/2010GL045046.
Dombrowski, N., Donaho, J. A., Gutierrez, T., Seitz, K. W., Teske, A. P., and Baker, B. J. (2016). Reconstructing metabolic pathways of hydrocarbon-degrading bacteria from the Deepwater Horizon oil spill. *Nat. Microbiol.* 1, 1–7. doi:10.1038/nmicrobiol.2016.57.
Eddy, S. R. (2011). Accelerated profile HMM searches. *PLoS Comput. Biol.* 7, e1002195. doi:10.1371/journal.pcbi.1002195.
Edgar, R. C. (2004). MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinforma. 2004 51* 5, 1–19. doi:10.1186/1471-2105-5-113.
El-Gebali, S., Mistry, J., Bateman, A., Eddy, S. R., Luciani, A., Potter, S. C., Qureshi, M., Richardson, L. J., Salazar, G. A., Smart, A., Sonnhammer, E. L. L., Hirsh, L., Paladin, L., Piovesan, D., Tosatto, S. C. E., and Finn, R. D. (2019). The Pfam protein families database in 2019. *Nucleic Acids Res.* 47, D427–D432. doi:10.1093/nar/gky995.
Eng, J. K., Jahan, T. A., and Hoopmann, M. R. (2013). Comet: An open-source MS/MS sequence database search tool. *Proteomics* 13, 22–24. doi:10.1002/pmic.201200439.
Eren, A. M., Esen, Ö. C., Quince, C., Vineis, J. H., Morrison, H. G., Sogin, M. L., and Delmont, T. O. (2015). Anvi'o: an advanced analysis and visualization platform for 'omics data. *PeerJ* 3, e1319. doi:10.7717/PEERJ.1319.
Fernández-Martínez, J., Pujalte, M. J., García-Martínez, J., Mata, M., Garay, E., and Rodríguez-Valera, F. (2003). Description of Alcanivorax venustensis sp. nov. and reclassification of Fundibacter jadensis DSM 12178T (Bruns and Berthe-Corti 1999) as Alcanivorax jadensis comb. nov., members of the emended genus Alcanivorax. *Int. J. Syst. Evol. Microbiol.* 53, 331–338. doi:10.1099/ijs.0.01923-0.
Field, C. B., Behrenfeld, M. J., Randerson, J. T., and Falkowski, P. (1998). Primary production of the biosphere: Integrating terrestrial and oceanic components. *Science (80-.).* 281, 237–240. doi:10.1126/SCIENCE.281.5374.237/SUPPL_FILE/982246E_THUMB.GIF.
FigTree Available at: http://tree.bio.ed.ac.uk/software/figtree/ [Accessed October 15, 2021].
Fitzpatrick, F. A., Johnson, R., Zhu, Z., Waterman, D., McCulloch, R. D., Hayter, E. J., Garcia, M. H., Boufadel, M., Dekker, T., Soong, D. T., Hoard, C. J., and Lee, K. (2016). Integrated Modeling Approach for Fate and Transport of Submerged Oil and Oil-Particle Aggregates in a Freshwater Riverine Environment. *Proceedings, Fed. Interag. Sediment. Progr. Conf.*, 1783–1794.
Flombaum, P., Gallegos, J. L., Gordillo, R. A., Rincón, J., Zabala, L. L., Jiao, N., Karl, D. M., Li, W. K. W., Lomas, M. W., Veneziano, D., Vera, C. S., Vrugt, J. A., and Martiny, A. C. (2013). Present and future global distributions of the marine Cyanobacteria Prochlorococcus and Synechococcus. *Proc. Natl. Acad. Sci. U. S. A.* 110, 9824–9829. doi:10.1073/PNAS.1307701110/-/DCSUPPLEMENTAL.
Frysinger, G. S., Gaines, R. B., Xu, L., and Reddy, C. M. (2003). Resolving the unresolved complex mixture in petroleum-contaminated sediments. *Environ. Sci. Technol.* 37, 1653–1662. doi:10.1021/es020742n.
Fuhrman, J. A., and McCallum, K. (1992). Novel major archaebacterial group from marine plankton. *Nature* 356, 148–149. doi:10.1038/356148a0.
Galand, P. E., Gutiérrez-Provecho, C., Massana, R., Gasol, J. M., and Casamayor, E. O. (2010). Inter-annual recurrence of archaeal assemblages in the coastal NW Mediterranean Sea (Blanes Bay Microbial Observatory). *Limnol. Oceanogr.* 55, 2117–2125. doi:10.4319/lo.2010.55.5.2117.

Giovannoni, S., Chan, F., Davis, E., Deutsch, C., and Wolf, S. (2021). Biochemical Barriers on the Path to Ocean Anoxia? *MBio* 12, e0133221. doi:10.1128/mBio.01332-21.
Glöckner, F. O., Yilmaz, P., Quast, C., Gerken, J., Beccati, A., Ciuprina, A., Bruns, G., Yarza, P., Peplies, J., Westram, R., and Ludwig, W. (2017). 25 years of serving the community with ribosomal RNA gene reference databases and tools. *J. Biotechnol.* 261, 169–176. doi:10.1016/J.JBIOTEC.2017.06.1198.
González-Gaya, B., Martínez-Varela, A., Vila-Costa, M., Casal, P., Cerro-Gálvez, E., Berrojalbiz, N., Lundin, D., Vidal, M., Mompeán, C., Bode, A., Jiménez, B., and Dachs, J. (2019). Biodegradation as an important sink of aromatic hydrocarbons in the oceans. *Nat. Geosci.* 12, 119–125. doi:10.1038/s41561-018-0285-3.
Griffin, M., and Trudgill, P. W. (1972). The metabolism of cyclopentanol by Pseudomonas N.C.I.B. 9872. *Biochem. J.* 129, 595–603. doi:10.1042/bj1290595.
Grimalt, J. O., de Wit, R., Teixidor, P., and Albaigés, J. (1992). Lipid biogeochemistry of Phormidium and Microcoleus mats. *Org. Geochem.* 19, 509–530. doi:10.1016/0146-6380(92)90015-P.
Gros, J., Socolofsky, S. A., Dissanayake, A. L., Jun, I., Zhao, L., Boufadel, M. C., Reddy, C. M., and Arey, J. S. (2017). Petroleum dynamics in the sea and influence of subsea dispersant injection during Deepwater Horizon. *Proc. Natl. Acad. Sci. U. S. A.* 114, 10065–10070. doi:10.1073/pnas.1612518114.
Gschwend, P., Zafiriou, O. C., and Gagosian, R. B. (1980). Volatile organic compounds in seawater from the Peru upwelling region1,2. *Limnol. Oceanogr.* 25, 1044–1053. doi:10.4319/lo.1980.25.6.1044.
Gurevich, A., Saveliev, V., Vyahhi, N., and Tesler, G. (2013). QUAST: quality assessment tool for genome assemblies. *Bioinformatics* 29, 1072–1075. doi:10.1093/BIOINFORMATICS/BTT086.
Han, J., and Calvin, M. (1969). Hydrocarbon distribution of algae and bacteria, and microbiological activity in sediments. *Proc. Natl. Acad. Sci. U. S. A.* 64, 436–443. doi:10.1073/pnas.64.2.436.
Hara, A., Syutsubo, K., and Harayama, S. (2003). Alcanivorax which prevails in oil-contaminated seawater exhibits broad substrate specificity for alkane degradation. *Environ. Microbiol.* 5, 746–753. doi:10.1046/j.1468-2920.2003.00468.x.
Harris, A. P., Techtmann, S. M., Stelling, S. C., Utturkar, S. M., Alshibli, N. K., Brown, S. D., and Hazen, T. C. (2014). Draft Genome Sequence of Pseudoalteromonas sp. Strain ND6B, an Oil-Degrading Isolate from Eastern Mediterranean Sea Water Collected at a Depth of 1,210 Meters. *Genome Announc.* 2, e01212-14. doi:10.1128/genomeA.01212-14.
Hawley, A. K., Nobu, M. K., Wright, J. J., Durno, W. E., Morgan-Lang, C., Sage, B., Schwientek, P., Swan, B. K., Rinke, C., Torres-Beltrán, M., Mewis, K., Liu, W. T., Stepanauskas, R., Woyke, T., and Hallam, S. J. (2017). Diverse Marinimicrobia bacteria may mediate coupled biogeochemical cycles along eco-thermodynamic gradients. *Nat. Commun.* 8, 1–10. doi:10.1038/s41467-017-01376-9.
Hays, S. G., Patrick, W. G., Ziesack, M., Oxman, N., and Silver, P. A. (2015). Better together: Engineering and application of microbial symbioses. *Curr. Opin. Biotechnol.* 36, 40–49. doi:10.1016/j.copbio.2015.08.008.
Hazen, T. C. (2010). "Cometabolic Bioremediation," in *Handbook of Hydrocarbon and Lipid Microbiology*, 2505–2514. doi:10.1007/978-3-540-77587-4_185.
Hazen, T. C., Dubinsky, E. A., DeSantis, T. Z., Andersen, G. L., Piceno, Y. M., Singh, N., Jansson, J. K., Probst, A., Borglin, S. E., Fortney, J. L., Stringfellow, W. T., Bill, M., Conrad, M. E., Tom, L. M., Chavarria, K. L., Alusi, T. R., Lamendella, R., Joyner, D. C., Spier, C., Baelum, J., Auer, M., Zemla, M. L., Chakraborty, R., Sonnenthal, E. L., D'haeseleer, P., Holman, H.-Y. N., Osman, S., Lu, Z., Van Nostrand, J. D., Deng, Y.,

Zhou, J., and Mason, O. U. (2010a). Deep-Sea Oil Plume Enriches Indigenous Oil-Degrading Bacteria. *Science (80-.).* 330, 204–208. doi:10.1126/science.1195979.
Hazen, T. C., Dubinsky, E. A., DeSantis, T. Z., Andersen, G. L., Piceno, Y. M., Singh, N., Jansson, J. K., Probst, A., Borglin, S. E., Fortney, J. L., Stringfellow, W. T., Bill, M., Conrad, M. E., Tom, L. M., Chavarria, K. L., Alusi, T. R., Lamendella, R., Joyner, D. C., Spier, C., Baelum, J., Auer, M., Zemla, M. L., Chakraborty, R., Sonnenthal, E. L., D'haeseleer, P., Holman, H.-Y. N., Osman, S., Lu, Z., Van Nostrand, J. D., Deng, Y., Zhou, J., and Mason, O. U. (2010b). Deep-Sea Oil Plume Enriches Indigenous Oil-Degrading Bacteria. *Science (80-.).* 330, 204–208. doi:10.1126/science.1195979.
Hazen, T. C., Prince, R. C., and Mahmoudi, N. (2016). Marine Oil Biodegradation. *Environ. Sci. Technol.* 50, 2121–2129. doi:10.1021/acs.est.5b03333.
Head, I. M., Jones, D. M., and Larter, S. R. (2003). Biological activity in the deep subsurface and the origin of heavy oil. *Nature* 426, 344–352. Available at: http://dx.doi.org/10.1038/nature02134.
Head, I. M., Jones, D. M., and Roling, W. F. M. (2006). Marine microorganisms make a meal of oil. *Nat Rev Micro* 4, 173–182. Available at: http://dx.doi.org/10.1038/nrmicro1348.
Herman, D. C., Fedorak, P. M., and Costerton, J. W. (1993). Biodegradation of cycloalkane carboxylic acids in oil sand tailings. *Can. J. Microbiol.* 39, 576–580. doi:10.1139/m93-083.
Holmes, A. J., Costello, A., Lidstrom, M. E., and Murrell, J. C. (2006). Evidence that participate methane monooxygenase and ammonia monooxygenase may be evolutionarily related. *FEMS Microbiol. Lett.* 132, 203–208. doi:10.1111/j.1574-6968.1995.tb07834.x.
Hu, P., Dubinsky, E. A., Probst, A. J., Wang, J., Sieber, C. M. K., Tom, L. M., Gardinali, P. R., Banfield, J. F., Atlas, R. M., and Andersen, G. L. (2017). Simulation of Deepwater Horizon oil plume reveals substrate specialization within a complex community of hydrocarbon degraders. *Proc. Natl. Acad. Sci. U. S. A.* 114, 201703424. doi:10.1073/pnas.1703424114.
Hugoni, M., Taib, N., Debroas, D., Domaizon, I., Dufournel, I. J., Bronner, G., Salter, I., Agogué, H., Mary, I., and Galand, P. E. (2013). Structure of the rare archaeal biosphere and seasonal dynamics of active ecotypes in surface coastal waters. *Proc. Natl. Acad. Sci. U. S. A.* 110, 6004–6009. doi:10.1073/pnas.1216863110.
Hyatt, D., Chen, G. L., LoCascio, P. F., Land, M. L., Larimer, F. W., and Hauser, L. J. (2010). Prodigal: Prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 11, 1–11. doi:10.1186/1471-2105-11-119.
Imelik, B. (1948). Oxidation of cyclohexane by Pseudomonas aeruginosa. *C. R. Hebd. Seances Acad. Sci.* 226, 2082. Available at: https://pubmed.ncbi.nlm.nih.gov/18882697/ [Accessed November 29, 2021].
Iwaki, H., Hasegawa, Y., Wang, S., Kayser, M. M., and Lau, P. C. K. (2002). Cloning and characterization of a gene cluster involved in cyclopentanol metabolism in Comamonas sp. strain NCIMB 9872 and biotransformations effected by Escherichia coli-expressed cyclopentanone 1,2-monooxygenase. *Appl. Environ. Microbiol.* 68, 5671–5684. doi:10.1128/AEM.68.11.5671-5684.2002.
Jordan, S. F. A., Gräwe, U., Treude, T., Lee, E. M., Schneider von Deimling, J., Rehder, G., and Schmale, O. (2021a). Pelagic methane sink enhanced by benthic methanotrophs ejected from a gas seep. *Geophys. Res. Lett.* 63, 183–229. doi:10.1029/2021gl094819.
Jordan, S. F. A., Gräwe, U., Treude, T., Lee, E. M. van der, Deimling, J. S. von, Rehder, G., and Schmale, O. (2021b). Pelagic Methane Sink Enhanced by Benthic Methanotrophs Ejected From a Gas Seep. *Geophys. Res. Lett.* 48, e2021GL094819.

doi:10.1029/2021GL094819.
Jordan, S. F. A., Treude, T., Leifer, I., Janßen, R., Werner, J., Schulz-Vogt, H., and Schmale, O. (2020). Bubble-mediated transport of benthic microorganisms into the water column: Identification of methanotrophs and implication of seepage intensity on transport efficiency. *Sci. Rep.* 10, 1–15. doi:10.1038/s41598-020-61446-9.
Joshi, N., and Fass, J. (2011). Sickle: A sliding-window, adaptive, quality-based trimming tool for FastQ files (Version 1.33) [Software]. *Available at https://github.com/najoshi/sickle.*, 2011. Available at: https://github.com/najoshi/sickle [Accessed October 14, 2021].
Kanaly, R. A., Harayama, S., and Watanabe, K. (2002). Rhodanobacter sp. strain BPC1 in a benzo[a]pyrene-mineralizing bacterial consortium. *Appl. Environ. Microbiol.* 68, 5826–5833. doi:10.1128/AEM.68.12.5826-5833.2002/ASSET/642763CA-3405-4155-BF90-FA7FE64C23C3/ASSETS/GRAPHIC/AM1220786005.JPEG.
Kang, D. D., Li, F., Kirton, E., Thomas, A., Egan, R., An, H., and Wang, Z. (2019). MetaBAT 2: An adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. *PeerJ* 2019, e7359. doi:10.7717/peerj.7359.
Karl, D. M., and Church, M. J. (2014). Microbial oceanography and the Hawaii Ocean Time-series programme. *Nat. Rev. Microbiol.* 12, 699–713. doi:10.1038/nrmicro3333.
Kessler, J. D., Valentine, D. L., Redmond, M. C., Du, M., Chan, E. W., Mendes, S. D., Quiroz, E. W., Villanueva, C. J., Shusta, S. S., Werra, L. M., Yvon-Lewis, S. A., and Weber, T. C. (2011). A Persistent Oxygen Anomaly Reveals the Fate of Spilled Methane in the Deep Gulf of Mexico. *Science (80-.).* 331, 312–315. doi:10.1126/science.1199697.
Kim, D., Song, L., Breitwieser, F. P., and Salzberg, S. L. (2016). Centrifuge: rapid and sensitive classification of metagenomic sequences. *Genome Res.* 26, 1721–1729. doi:10.1101/GR.210641.116.
Kim, Y. M., Jung, S. H., Chung, Y. H., Yu, C. B., and Rhee, I. K. (2008). Cloning and characterization of a cyclohexanone monooxygenase gene from Arthrobacter sp. L661. *Biotechnol. Bioprocess Eng.* 13, 40–47. doi:10.1007/s12257-007-0162-1.
Könneke, M., Bernhard, A. E., De La Torre, J. R., Walker, C. B., Waterbury, J. B., and Stahl, D. A. (2005). Isolation of an autotrophic ammonia-oxidizing marine archaeon. *Nature* 437, 543–546. doi:10.1038/nature03911.
Kozich, J. J., Westcott, S. L., Baxter, N. T., Highlander, S. K., and Schloss, P. D. (2013). Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the miseq illumina sequencing platform. *Appl. Environ. Microbiol.* 79, 5112–5120. doi:10.1128/AEM.01043-13.
Kvenvolden, K. A., and Cooper, C. K. (2003). Natural seepage of crude oil into the marine environment. *Geo-Marine Lett.* 23, 140–146. doi:10.1007/s00367-003-0135-0.
Langmead, B., and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods 2012 94* 9, 357–359. doi:10.1038/nmeth.1923.
Lea-Smith, D. J., Biller, S. J., Davey, M. P., Cotton, C. A. R., Sepulveda, B. M. P., Turchyn, A. V., Scanlan, D. J., Smith, A. G., Chisholm, S. W., and Howe, C. J. (2015). Contribution of cyanobacterial alkane production to the ocean hydrocarbon cycle. *Proc. Natl. Acad. Sci. U. S. A.* 112, 13591–13596. doi:10.1073/pnas.1507274112.
Lea-Smith, D. J., Ortiz-Suarez, M. L., Lenn, T., Nürnberg, D. J., Baers, L. L., Davey, M. P., Parolini, L., Huber, R. G., Cotton, C. A. R., Mastroianni, G., Bombelli, P., Ungerer, P., Stevens, T. J., Smith, A. G., Bond, P. J., Mullineaux, C. W., and Howe, C. J. (2016). Hydrocarbons are essential for optimal cell size, division, and growth of Cyanobacteria. *Plant Physiol.* 172, 1928–1940. doi:10.1104/pp.16.01205.
Leahy, J. G., and Colwell, R. R. (1990). Microbial degradation of hydrocarbons in the environment. *Microbiol. Rev.* 54, 305–315. doi: .

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and Subgroup, 1000 Genome Project Data Processing (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25, 2078–2079. doi:10.1093/BIOINFORMATICS/BTP352.

Liao, J., Wang, J., and Huang, Y. (2015). Bacterial Community Features Are Shaped by Geographic Location, Physicochemical Properties, and Oil Contamination of Soil in Main Oil Fields of China. *Microb. Ecol.* 70, 380–389. doi:10.1007/s00248-015-0572-0.

Liu, C., and Shao, Z. (2005). Alcanivorax dieselolei sp. nov., a novel alkane-degrading bacterium isolated from sea water and deep-sea sediment. *Int. J. Syst. Evol. Microbiol.* 55, 1181–1186. doi:10.1099/IJS.0.63443-0/CITE/REFWORKS.

Liu, C., Wang, W., Wu, Y., Zhou, Z., Lai, Q., and Shao, Z. (2011). Multiple alkane hydroxylase systems in a marine alkane degrader, Alcanivorax dieselolei B-5. *Environ. Microbiol.* 13, 1168–1178. doi:10.1111/j.1462-2920.2010.02416.x.

Liu, J., Techtmann, S. M., Woo, H. L., Ning, D., Fortney, J. L., and Hazen, T. C. (2017). Rapid Response of Eastern Mediterranean Deep Sea Microbial Communities to Oil. *Sci. Rep.* 7, 5762. doi:10.1038/s41598-017-05958-x.

Love, C. R., Arrington, E. C., Gosselin, K. M., Reddy, C. M., Van Mooy, B. A. S., Nelson, R. K., and Valentine, D. L. (2021). Microbial production and consumption of hydrocarbons in the global ocean. *Nat. Microbiol.* 6, 489–498. doi:10.1038/s41564-020-00859-8.

MacDonald, I. R., Garcia-Pineda, O., Beet, A., Daneshgar Asl, S., Feng, L., Graettinger, G., French-Mccay, D., Holmes, J., Hu, C., Huffer, F., Leifer, I., Muller-Karger, F., Solow, A., Silva, M., and Swayze, G. (2015). Natural and unnatural oil slicks in the Gulf of Mexico. *J. Geophys. Res. Ocean.* 120, 8364–8380. doi:10.1002/2015JC011062.

Mahmoudi, N., Robeson, M. S., Castro, H. F., Fortney, J. L., Techtmann, S. M., Joyner, D. C., Paradis, C. J., Pfiffner, S. M., Hazen, T. C., and Hazen, T. C. (2015). Microbial community composition and diversity in Caspian Sea sediments. *FEMS Microbiol. Ecol.* 91, 1–11. doi:10.1093/femsec/fiu013.

Martinez, A., Tyson, G. W., and Delong, E. F. (2010). Widespread known and novel phosphonate utilization pathways in marine bacteria revealed by functional screening and metagenomic analyses. *Environ. Microbiol.* 12, 222–238. doi:10.1111/J.1462-2920.2009.02062.X.

Mason, O. U., Hazen, T. C., Borglin, S., Chain, P. S. G., Dubinsky, E. A., Fortney, J. L., Han, J., Holman, H.-Y. N., Hultman, J., Lamendella, R., Mackelprang, R., Malfatti, S., Tom, L. M., Tringe, S. G., Woyke, T., Zhou, J., Rubin, E. M., and Jansson, J. K. (2012). Metagenome, metatranscriptome and single-cell sequencing reveal microbial response to Deepwater Horizon oil spill. *ISME J* 6, 1715–1727. doi:http://www.nature.com/ismej/journal/v6/n9/suppinfo/ismej201259s1.html.

Massana, R., Murray, A. E., Preston, C. M., and DeLong, E. F. (1997). Vertical distribution and phylogenetic characterization of marine planktonic Archaea in the Santa Barbara Channel. *Appl. Environ. Microbiol.* 63.

McGenity, T. J., Crombie, A. T., and Murrell, J. C. (2018). Microbial cycling of isoprene, the most abundantly produced biological volatile organic compound on Earth. *ISME J.* 12, 931–941. doi:10.1038/s41396-018-0072-6.

McGenity, T. J., Folwell, B. D., McKew, B. A., and Sanni, G. O. (2012). Marine crude-oil biodegradation: a central role for interspecies interactions. *Aquat. Biosyst.* 8, 1–19. doi:10.1186/2046-9063-8-10.

McKenna, A. M., Williams, J. T., Putman, J. C., Aeppli, C., Reddy, C. M., Valentine, D. L., Lemkau, K. L., Kellermann, M. Y., Savory, J. J., Kaiser, N. K., Marshall, A. G., and Rodgers, R. P. (2014). Unprecedented ultrahigh resolution FT-ICR mass spectrometry and parts-per-billion mass accuracy enable direct characterization of nickel and vanadyl porphyrins in petroleum from natural seeps. *Energy and Fuels* 28, 2454–2464.

doi:10.1021/ef5002452.
McKenzie, L. M., Witter, R. Z., Newman, L. S., and Adgate, J. L. (2012). Human health risk assessment of air emissions from development of unconventional natural gas resources. *Sci. Total Environ.* 424, 79–87. doi:10.1016/j.scitotenv.2012.02.018.
Mokhatab, S., and Poe, W. A. (2012). *Handbook of natural gas transmission and processing*. Gulf Professional Pub.
Moran, M. A., Buchan, A., González, J. M., Heidelberg, J. F., Whitman, W. B., Kiene, R. P., Henriksen, J. R., King, G. M., Belas, R., Fuqua, C., Brinkac, L., Lewis, M., Johri, S., Weaver, B., Pai, G., Eisen, J. A., Rahe, E., Sheldon, W. M., Ye, W., Miller, T. R., Carlton, J., Rasko, D. A., Paulsen, I. T., Ren, Q., Daugherty, S. C., Deboy, R. T., Dodson, R. J., Durkin, A. S., Madupu, R., Nelson, W. C., Sullivan, S. A., Rosovitz, M. J., Haft, D. H., Selengut, J., and Ward, N. (2004). Genome sequence of Silicibacter pomeroyi reveals adaptations to the marine environment. *Nat. 2005 4327019* 432, 910–913. doi:10.1038/nature03170.
Moreno, R., and Rojo, F. (2017). "Enzymes for Aerobic Degradation of Alkanes in Bacteria," in *Aerobic Utilization of Hydrocarbons, Oils and Lipids* (Springer International Publishing), 1–25. doi:10.1007/978-3-319-39782-5_6-1.
National Academies Press (2016). *Spills of diluted bitumen from pipelines: A comparative study of environmental fate, effects, and response*. doi:10.17226/21834.
Needham, D. M., and Fuhrman, J. A. (2016). Pronounced daily succession of phytoplankton, archaea and bacteria following a spring bloom. *Nat. Microbiol.* 1, 16005. doi:10.1038/nmicrobiol.2016.5.
Nie, Y., Chi, C. Q., Fang, H., Liang, J. L., Lu, S. L., Lai, G. L., Tang, Y. Q., and Wu, X. L. (2014). Diverse alkane hydroxylase genes in microorganisms and environments. *Sci. Reports 2014 41* 4, 1–11. doi:10.1038/srep04968.
Nurk, S., Meleshko, D., Korobeynikov, A., and Pevzner, P. A. (2017). metaSPAdes: a new versatile metagenomic assembler. *Genome Res.* 27, 824–834. doi:10.1101/GR.213959.116.
Nzila, A. (2013). Update on the cometabolism of organic pollutants by bacteria. *Environ. Pollut.* 178, 474–482. doi:10.1016/j.envpol.2013.03.042.
Oelkers, E. H. (1991). Calculation of diffusion coefficients for aqueous organic species at temperatures from 0 to 350 °C. *Geochim. Cosmochim. Acta* 55, 3515–3529. doi:10.1016/0016-7037(91)90052-7.
Oil in the Sea III (2003). National Academies Press doi:10.17226/10388.
Ooyama, J., and Foster, J. W. (1965). Bacterial oxidation of cycloparaffinic hydrocarbons. *Antonie Van Leeuwenhoek* 31, 45–65. doi:10.1007/BF02045875.
Oren, A. (2019). "Aerobic Hydrocarbon-Degrading Archaea," in *Taxonomy, Genomics and Ecophysiology of Hydrocarbon-Degrading Microbes*, 41–51. doi:10.1007/978-3-030-14796-9_5.
Orsi, W. D., Smith, J. M., Wilcox, H. M., Swalwell, J. E., Carini, P., Worden, A. Z., and Santoro, A. E. (2015). Ecophysiology of uncultivated marine euryarchaea is linked to particulate organic matter. *ISME J.* 9, 1747–1763. doi:10.1038/ismej.2014.260.
Païssé, S., Goñi-Urriza, M., Coulon, F., and Duran, R. (2010). How a Bacterial Community Originating from a Contaminated Coastal Sediment Responds to an Oil Input. *Microb. Ecol.* 60, 394–405. doi:10.1007/s00248-010-9721-7.
Parada, A. E., and Fuhrman, J. A. (2017). Marine archaeal dynamics and interactions with the microbial community over 5 years from surface to seafloor. *ISME J.* 11, 2510–2525. doi:10.1038/ismej.2017.104.
Parada, A. E., Needham, D. M., and Fuhrman, J. A. (2016). Every base matters: Assessing small subunit rRNA primers for marine microbiomes with mock communities, time series and global field samples. *Environ. Microbiol.* 18, 1403–1414. doi:10.1111/1462-

2920.13023.
Parks, D. H., Chuvochina, M., Chaumeil, P.-A., Rinke, C., Mussig, A. J., and Hugenholtz, P. (2020). A complete domain-to-species taxonomy for Bacteria and Archaea. *Nat. Biotechnol. 2020 389* 38, 1079–1086. doi:10.1038/s41587-020-0501-8.
Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P., and Tyson, G. W. (2015). CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res.* 25, 1043–1055. doi:10.1101/GR.186072.114.
Perez, S., Furlan, P., Ellenberger, S., and Banker, P. (2016). Estimating diluted bitumen entrained by suspended sediments in river rapids using O2 absorption rate. *Int. J. Environ. Sci. Technol.* 13, 403–412. doi:10.1007/s13762-015-0874-2.
Perry, J. J., and Gibson, D. T. (1977a). Microbial metabolism of cyclic hydrocarbons and related compounds. *Crit. Rev. Microbiol.* 5, 387–412. doi:10.3109/10408417709102811.
Perry, J. J., and Gibson, D. T. (1977b). Microbial metabolism of cyclic hydrocarbons and related compounds. *Crit. Rev. Microbiol.* 5, 387–412. doi:10.3109/10408417709102811.
Peterson, D. R. (1994). Calculating the aquatic toxicity of hydrocarbon mixtures. *Chemosphere* 29, 2493–2506. doi:10.1016/0045-6535(94)90052-3.
Polovina, J. J., Howell, E. A., and Abecassis, M. (2008). Ocean's least productive waters are expanding. *Geophys. Res. Lett.* 35. doi:10.1029/2007GL031745.
Prince, R. C., Amande, T. J., and McGenity, T. J. (2019a). "Prokaryotic Hydrocarbon Degraders," in *Taxonomy, Genomics and Ecophysiology of Hydrocarbon-Degrading Microbes*, ed. T. J. McGenity (Cham: Springer International Publishing), 1–39. doi:10.1007/978-3-030-14796-9_15.
Prince, R. C., Amande, T. J., and McGenity, T. J. (2019b). "Taxonomy, Genomics and Ecophysiology of Hydrocarbon-Degrading Microbes," in *Taxonomy, Genomics and Ecophysiology of Hydrocarbon-Degrading Microbes* doi:10.1007/978-3-319-60053-6.
Pruesse, E., Quast, C., Knittel, K., Fuchs, B. M., Ludwig, W., Peplies, J., and Glöckner, F. O. (2007). SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Res.* 35, 7188–7196. doi:10.1093/NAR/GKM864.
Qiao, N., and Shao, Z. (2010). Isolation and characterization of a novel biosurfactant produced by hydrocarbon-degrading bacterium Alcanivorax dieselolei B-5. *J. Appl. Microbiol.* 108, 1207–1216. doi:10.1111/j.1365-2672.2009.04513.x.
Qin, W., Amin, S. A., Martens-Habbena, W., Walker, C. B., Urakawa, H., Devol, A. H., Ingalls, A. E., Moffett, J. W., Armbrust, E. V., and Stahl, D. A. (2014). Marine ammonia-oxidizing archaeal isolates display obligate mixotrophy and wide ecotypic variation. *Proc. Natl. Acad. Sci. U. S. A.* 111, 12504–12509. doi:10.1073/pnas.1324115111.
Reddy, C. M., Arey, J. S., Seewald, J. S., Sylva, S. P., Lemkau, K. L., Nelson, R. K., Carmichael, C. A., McIntyre, C. P., Fenwick, J., and Ventura, G. T. (2012). Composition and fate of gas and oil released to the water column during the Deepwater Horizon oil spill. *Proc. Natl. Acad. Sci.* 109, 20229–20234.
Redmond, M. C., and Valentine, D. L. (2012a). Natural gas and temperature structured a microbial community response to the Deepwater Horizon oil spill. *Proc. Natl. Acad. Sci. U. S. A.* 109, 20292–20297. doi:10.1073/pnas.1108756108.
Redmond, M. C., and Valentine, D. L. (2012b). Natural gas and temperature structured a microbial community response to the Deepwater Horizon oil spill. *Proc. Natl. Acad. Sci.* 109, 20292–20297. doi:10.1073/pnas.1108756108.
Repeta, D. J., Ferrón, S., Sosa, O. A., Johnson, C. G., Repeta, L. D., Acker, M., Delong, E. F., and Karl, D. M. (2016). Marine methane paradox explained by bacterial

degradation of dissolved organic matter. *Nat. Geosci.* 9, 884–887. doi:10.1038/ngeo2837.
Rinke, C., Rubino, F., Messer, L. F., Youssef, N., Parks, D. H., Chuvochina, M., Brown, M., Jeffries, T., Tyson, G. W., Seymour, J. R., and Hugenholtz, P. (2019). A phylogenomic and ecological analysis of the globally abundant Marine Group II archaea (Ca. Poseidoniales ord. nov.). *ISME J.* 13, 663–675. doi:10.1038/s41396-018-0282-y.
Rivers, A. R., Sharma, S., Tringe, S. G., Martin, J., Joye, S. B., and Moran, M. A. (2013). Transcriptional response of bathypelagic marine bacterioplankton to the Deepwater Horizon oil spill. *ISME J. 2013 712* 7, 2315–2329. doi:10.1038/ismej.2013.129.
Rodriguez-R, L., and Konstantinidis, K. (2016). The enveomics collection: a toolbox for specialized analyses of microbial genomes and metagenomes. doi:10.7287/peerj.preprints.1900.
Rogener, M. K., Bracco, A., Hunter, K. S., Saxton, M. A., and Joye, S. B. (2018). Long-term impact of the Deepwater Horizon oil well blowout on methane oxidation dynamics in the northern Gulf of Mexico. *Elem Sci Anth* 6, 73. doi:10.1525/elementa.332.
Rojo, F. (2010). "Enzymes for Aerobic Degradation of Alkanes," in *Handbook of Hydrocarbon and Lipid Microbiology*, 781–797. doi:10.1007/978-3-540-77587-4_59.
Röling, W. F. M., Milner, M. G., Jones, D. M., Lee, K., Daniel, F., Swannell, R. J. P., and Head, I. M. (2002). Robust hydrocarbon degradation and dynamics of bacterial communities during nutrient-enhanced oil spill bioremediation. *Appl. Environ. Microbiol.* 68, 5537–5548. doi:10.1128/AEM.68.11.5537-5548.2002.
Rubin-Blum, M., Antony, C. P., Borowski, C., Sayavedra, L., Pape, T., Sahling, H., Bohrmann, G., Kleiner, M., Redmond, M. C., Valentine, D. L., and Dubilier, N. (2017). Short-chain alkanes fuel mussel and sponge Cycloclasticus symbionts from deep-sea gas and oil seeps. *Nat. Microbiol.* 2, 17093. doi:10.1038/nmicrobiol.2017.93.
Ryerson, T. B., Aikin, K. C., Angevine, W. M., Atlas, E. L., Blake, D. R., Brock, C. A., Fehsenfeld, F. C., Gao, R. S., De Gouw, J. A., Fahey, D. W., Holloway, J. S., Lack, D. A., Lueb, R. A., Meinardi, S., Middlebrook, A. M., Murphy, D. M., Neuman, J. A., Nowak, J. B., Parrish, D. D., Peischl, J., Perring, A. E., Pollack, I. B., Ravishankara, A. R., Roberts, J. M., Schwarz, J. P., Spackman, J. R., Stark, H., Warneke, C., and Watts, L. A. (2011). Atmospheric emissions from the deepwater Horizon spill constrain air-water partitioning, hydrocarbon fate, and leak rate. *Geophys. Res. Lett.* 38. doi:10.1029/2011GL046726.
Ryerson, T. B., Camilli, R., Kessler, J. D., Kujawinski, E. B., Reddy, C. M., Valentine, D. L., Atlas, E., Blake, D. R., de Gouw, J., Meinardi, S., Parrish, D. D., Peischl, J., Seewald, J. S., and Warneke, C. (2012a). Chemical data quantify Deepwater Horizon hydrocarbon flow rate and environmental distribution. *Proc. Natl. Acad. Sci.* 109, 20246–20253. doi:10.1073/pnas.1110564109.
Ryerson, T. B., Camilli, R., Kessler, J. D., Kujawinski, E. B., Reddy, C. M., Valentine, D. L., Atlas, E., Blake, D. R., De Gouw, J., Meinardi, S., Parrish, D. D., Peischl, J., Seewald, J. S., and Warneke, C. (2012b). Chemical data quantify Deepwater Horizon hydrocarbon flow rate and environmental distribution. *Proc. Natl. Acad. Sci. U. S. A.* 109, 20246–20253. doi:10.1073/pnas.1110564109.
Saunois, M., R. Stavert, A., Poulter, B., Bousquet, P., G. Canadell, J., B. Jackson, R., A. Raymond, P., J. Dlugokencky, E., Houweling, S., K. Patra, P., Ciais, P., K. Arora, V., Bastviken, D., Zhuang, Q., et al. (2020). The global methane budget 2000-2017. *Earth Syst. Sci. Data* 12, 1561–1623. doi:10.5194/essd-12-1561-2020.
Schirmer, A., Rude, M. A., Li, X., Popova, E., and Del Cardayre, S. B. (2010). Microbial biosynthesis of alkanes. *Science (80-.).* 329, 559–562. doi:10.1126/science.1187936.
Scoma, A., Heyer, R., Rifai, R., Dandyk, C., Marshall, I., Kerckhof, F. M., Marietou, A., Boshker, H. T. S., Meysman, F. J. R., Malmos, K. G., Vosegaard, T., Vermeir, P.,

Banat, I. M., Benndorf, D., and Boon, N. (2019). Reduced TCA cycle rates at high hydrostatic pressure hinder hydrocarbon degradation and obligate oil degraders in natural, deep-sea microbial communities. *ISME J.* 13, 1004–1018. doi:10.1038/s41396-018-0324-5.

Seifert, R., Delling, N., Hermann Richnow, H., Kempe, S., Hefter, J., and Michaelis, W. (1999). Ethylene and methane in the upper water column of the subtropical Atlantic. *Biogeochemistry* 44, 73–91. doi:10.1007/bf00992999.

Shanklin, J., Whittle, E., and Fox, B. G. (1994a). Eight Histidine Residues Are Catalytically Essential in a Membrane-Associated Iron Enzyme, Stearoyl-CoA Desaturase, and Are Conserved in Alkane Hydroxylase and Xylene Monooxygenase. *Biochemistry* 33, 12787–12794. doi:10.1021/bi00209a009.

Shanklin, J., Whittle, E., and Fox, B. G. (1994b). Eight Histidine Residues Are Catalytically Essential in a Membrane-Associated Iron Enzyme, Stearoyl-CoA Desaturase, and Are Conserved in Alkane Hydroxylase and Xylene Monooxygenase. *Biochemistry* 33, 12787–12794. doi:10.1021/bi00209a009.

Smits, T. H. M., Balada, S. B., Witholt, B., and Van Beilen, J. B. (2002). Functional analysis of alkane hydroxylases from gram-negative and gram-positive bacteria. *J. Bacteriol.* 184, 1733–1742. doi:10.1128/JB.184.6.1733-1742.2002/ASSET/4B2960D0-059C-468A-837D-A760B0476C7A/ASSETS/GRAPHIC/JB0621401002.JPEG.

Sorigué, D., Légeret, B., Cuiné, S., Blangy, S., Moulin, S., Billon, E., Richaud, P., Brugière, S., Couté, Y., Nurizzo, D., Müller, P., Brettel, K., Pignol, D., Arnoux, P., Li-Beisson, Y., Peltier, G., and Beisson, F. (2017). An algal photoenzyme converts fatty acids to hydrocarbons. *Science (80-.).* 357, 903–907. doi:10.1126/science.aan6349.

Sorigué, D., Légeret, B., Cuiné, S., Morales, P., Mirabella, B., Guédeney, G., Li-Beisson, Y., Jetter, R., Peltier, G., and Beisson, F. (2016). Microalgae synthesize hydrocarbons from long-chain fatty acids via a light-dependent pathway. *Plant Physiol.* 171, 2393–2405. doi:10.1104/pp.16.00462.

Sosa, O. A., Repeta, D. J., DeLong, E. F., Ashkezari, M. D., and Karl, D. M. (2019). Phosphate-limited ocean regions select for bacterial populations enriched in the carbon–phosphorus lyase pathway for phosphonate degradation. *Environ. Microbiol.* 21, 2402–2414. doi:10.1111/1462-2920.14628.

Spring, S. (2014). Function and Evolution of the Sox Multienzyme Complex in the Marine Gammaproteobacterium Congregibacter litoralis. doi:10.1155/2014/597418.

Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30, 1312–1313. doi:10.1093/BIOINFORMATICS/BTU033.

Starkenburg, S. R., Chain, P. S. G., Sayavedra-Soto, L. A., Hauser, L., Land, M. L., Larimer, F. W., Malfatti, S. A., Klotz, M. G., Bottomley, P. J., Arp, D. J., and Hickey, W. J. (2006). Genome Sequence of the Chemolithoautotrophic Nitrite-Oxidizing Bacterium Nitrobacter winogradskyi Nb-255. *Appl. Environ. Microbiol.* 72, 2050–2063. doi:10.1128/AEM.72.3.2050-2063.2006.

Stingl, U., Desiderio, R. A., Cho, J. C., Vergin, K. L., and Giovannoni, S. J. (2007). The SAR92 clade: An abundant coastal clade of culturable marine bacteria possessing proteorhodopsin. *Appl. Environ. Microbiol.* 73, 2290–2296. doi:10.1128/AEM.02559-06.

Stirling, L. A., Watkinson, R. J., and Higgins, I. J. (1977). Microbial metabolism of alicyclic hydrocarbons: isolation and properties of a cyclohexane degrading bacterium. *J. Gen. Microbiol.* 99, 119–125. doi:10.1099/00221287-99-1-119/CITE/REFWORKS.

Sunagawa, S., Coelho, L. P., Chaffron, S., Kultima, J. R., Labadie, K., Salazar, G., Djahanschiri, B., Zeller, G., Mende, D. R., Alberti, A., Cornejo-Castillo, F. M., Costea, P. I., Cruaud, C., Bork, P., et al. (2015). Structure and function of the global ocean

microbiome. *Science (80-.).* 348. doi:10.1126/science.1261359.
Tavormina, P. L., Orphan, V. J., Kalyuzhnaya, M. G., Jetten, M. S. M., and Klotz, M. G. (2011). A novel family of functional operons encoding methane/ammonia monooxygenase-related proteins in gammaproteobacterial methanotrophs. *Environ. Microbiol. Rep.* 3, 91–100. doi:10.1111/j.1758-2229.2010.00192.x.
Tavormina, P. L., Ussler, W., Steele, J. A., Connon, S. A., Klotz, M. G., and Orphan, V. J. (2013). Abundance and distribution of diverse membrane-bound monooxygenase (Cu-MMO) genes within the Costa Rica oxygen minimum zone. *Environ. Microbiol. Rep.* 5, 414–423. doi:10.1111/1758-2229.12025.
Techtmann, S. M., Fortney, J. L., Ayers, K. A., Joyner, D. C., Linley, T. D., Pfiffner, S. M., and Hazen, T. C. (2015). The unique chemistry of Eastern Mediterranean water masses selects for distinct microbial communities by depth. *PLoS One* 10, 1–22. doi:10.1371/journal.pone.0120605.
Teira, E., Lekunberri, I., Gasol, J. M., Nieto-Cid, M., Álvarez-Salgado, X. A., and Figueiras, F. G. (2007). Dynamics of the hydrocarbon-degrading Cycloclasticus bacteria during mesocosm-simulated oil spills. *Environ. Microbiol.* 9, 2551–2562. doi:10.1111/j.1462-2920.2007.01373.x.
Teske, A., Hinrichs, K.-U., Edgcomb, V., Gomez, A. D. V., Kysela, D., Sylva, S. P., Sogin, M. L., and Jannasch, H. W. (2002). Microbial Diversity of Hydrothermal Sediments in the Guaymas Basin : Evidence for Anaerobic Methanotrophic Communities Microbial Diversity of Hydrothermal Sediments in the Guaymas Basin : Evidence for Anaerobic Methanotrophic Communities †. *Appl. Environ. Microbiol.* 68, 1994–2007. doi:10.1128/AEM.68.4.1994.
Tian, R., Ning, D., He, Z., Zhang, P., Spencer, S. J., Gao, S., Shi, W., Wu, L., Zhang, Y., Yang, Y., Adams, B. G., Rocha, A. M., Detienne, B. L., Lowe, K. A., Joyner, D. C., Klingeman, D. M., Arkin, A. P., Fields, M. W., Hazen, T. C., Stahl, D. A., Alm, E. J., and Zhou, J. (2020). Small and mighty: Adaptation of superphylum Patescibacteria to groundwater environment drives their genome simplicity. *Microbiome* 8, 1–15. doi:10.1186/S40168-020-00825-W/FIGURES/5.
Timmins-Schiffman, E., May, D. H., Mikan, M., Riffle, M., Frazar, C., Harvey, H. R., Noble, W. S., and Nunn, B. L. (2016). Critical decisions in metaproteomics: achieving high confidence protein annotations in a sea of unknowns. *ISME J. 2017 112* 11, 309–314. doi:10.1038/ismej.2016.132.
Tonge, G. M., and Higgins, I. J. (1974). Microbial metabolism of alicyclic hydrocarbons. Growth of Nocardia petroleophila (NCIB9438) on methylcyclohexane. *J. Gen. Microbiol.* 81, 521–524. doi:10.1099/00221287-81-2-521.
Trower, M. K., Buckland, R. M., Higgins, R., and Griffin, M. (1985). Isolation and characterization of a cyclohexane-metabolizing Xanthobacter sp. *Appl. Environ. Microbiol.* 49, 1282–1289. doi:10.1128/AEM.49.5.1282-1289.1985/FORMAT/PDF.
Tully, B. J. (2019). Metabolic diversity within the globally abundant Marine Group II Euryarchaea offers insight into ecological patterns. *Nat. Commun.* 10, 1–12. doi:10.1038/s41467-018-07840-4.
Tully, B. J., Wheat, C. G., Glazer, B. T., and Huber, J. A. (2017). A dynamic microbial community with high functional redundancy inhabits the cold, oxic subseafloor aquifer. *ISME J. 2018 121* 12, 1–16. doi:10.1038/ismej.2017.187.
Valentine, D. L. (2007). Adaptations to energy stress dictate the ecology and evolution of the Archaea. *Nat. Rev. Microbiol.* 5, 316–323. doi:10.1038/nrmicro1619.
Valentine, D. L., Fisher, G. B., Bagby, S. C., Nelson, R. K., Reddy, C. M., Sylva, S. P., and Wood, M. A. (2014). Fallout plume of submerged oil from Deepwater Horizon. *Proc. Natl. Acad. Sci. U. S. A.* 111, 15906–15911. doi:10.1073/pnas.1414873111.
Valentine, D. L., Kessler, J. D., Redmond, M. C., Mendes, S. D., Heintz, M. B., Farwell, C.,

Hu, L., Kinnaman, F. S., Yvon-Lewis, S., Du, M., Chan, E. W., Garcia Tigreros, F., and Villanueva, C. J. (2010). Propane respiration jump-starts microbial response to a deep oil spill. *Science* 330, 208–11. doi:10.1126/science.1196830.

Valentine, D. L., Mezić, I., Maćešić, S., Črnjarić-Žic, N., Ivić, S., Hogan, P. J., Fonoberov, V. A., and Loire, S. (2012). Dynamic autoinoculation and the microbial ecology of a deep water hydrocarbon irruption. *Proc. Natl. Acad. Sci. U. S. A.* 109, 20286–91. doi:10.1073/pnas.1108820109.

van Beilen, J. B., Li, Z., Duetz, W. A., Smits, T. H. M., and Witholt, B. (2003). Diversity of alkane hydroxylase systems in the environment. *Oil Gas Sci. Technol.* 58, 427–440. doi:10.2516/ogst:2003026.

Van Den Berg, B. (2005). The FadL family: Unusual transporters for unusual substrates. *Curr. Opin. Struct. Biol.* 15, 401–407. doi:10.1016/j.sbi.2005.06.003.

Van Hamme, J. D., Singh, A., and Ward, O. P. (2003). Recent Advances in Petroleum Microbiology Recent Advances in Petroleum Microbiology. *Microbiol. Mol. Biol. Rev.* 67, 503–549. doi:10.1128/MMBR.67.4.503.

Walker, A., Stern, C., Scholz, D., Nielsen, E., Csulak, F., and Gaudiosi, R. (2016). Consensus Ecological Risk Assessment of Potential Transportation-related Bakken and Dilbit Crude Oil Spills in the Delaware Bay Watershed, USA. *J. Mar. Sci. Eng.* 4, 23. doi:10.3390/jmse4010023.

Wang, K., Wommack, K. E., and Chen, F. (2011). Abundance and distribution of Synechococcus spp. and cyanophages in the Chesapeake Bay. *Appl. Environ. Microbiol.* 77, 7459–7468. doi:10.1128/AEM.00267-11.

Wang, W., and Shao, Z. (2012). Diversity of flavin-binding monooxygenase genes (almA) in marine bacteria capable of degradation long-chain alkanes. *FEMS Microbiol. Ecol.* 80, 523–533. doi:10.1111/j.1574-6941.2012.01322.x.

Wang, W., and Shao, Z. (2014). The long-chain alkane metabolism network of Alcanivorax dieselolei. *Nat. Commun. 2014 51* 5, 1–11. doi:10.1038/ncomms6755.

Wang, W., Wang, L., and Shao, Z. (2018). Polycyclic aromatic hydrocarbon (PAH) degradation pathways of the obligate marine PAH degrader Cycloclasticus sp. strain P1. *Appl. Environ. Microbiol.* 84. doi:10.1128/AEM.01261-18.

Wardlaw, G. D., Arey, J. S., Reddy, C. M., Nelson, R. K., Ventura, G. T., and Valentine, D. L. (2008). Disentangling Oil Weathering at a Marine Seep Using GC×GC: Broad Metabolic Specificity Accompanies Subsurface Petroleum Biodegradation. *Environ. Sci. Technol.* 42, 7166–7173. doi:10.1021/es8013908.

White, H., Huber, C., Feicht, R., and Simon, H. (1993). On a reversible molybdenum-containing aldehyde oxidoreductase from Clostridium formicoaceticum. *Arch. Microbiol.* 159, 244–249. doi:10.1007/BF00248479.

White, H. K., Marx, C. T., Valentine, D. L., Sharpless, C., Aeppli, C., Gosselin, K. M., Kivenson, V., Liu, R. M., Nelson, R. K., Sylva, S. P., and Reddy, C. M. (2019). Examining inputs of biogenic and oil-derived hydrocarbons in surface waters following the deepwater horizon oil spill. *ACS Earth Sp. Chem.* 3, 1329–1337. doi:10.1021/acsearthspacechem.9b00090.

Willetts, A. J., and Cain, R. B. (1972). Microbial metabolism of alkylbenzene sulphonates Enzyme system of a Bacillus species responsible for B-oxidation of the alkyl side chain of alkylbenzene sulphonates. *Antonie Van Leeuwenhoek* 38, 543–555. doi:10.1007/BF02328121.

Winters, K., Parker, P. L., and Van Baalen, C. (1969). Hydrocarbons of blue-green algae: Geochemical significance. *Science (80-.).* 163, 467–468. doi:10.1126/science.163.3866.467.

Yakimov, M. M., Timmis, K. N., and Golyshin, P. N. (2007). Obligate oil-degrading marine bacteria. *Curr. Opin. Biotechnol.* 18, 257–66. doi:10.1016/j.copbio.2007.04.006.

Zhou, Z., Liu, Y., Pan, J., Cron, B. R., Toner, B. M., Anantharaman, K., Breier, J. A., Dick, G. J., and Li, M. (2020). Gammaproteobacteria mediating utilization of methyl-, sulfur- and petroleum organic compounds in deep ocean hydrothermal plumes. *ISME J.* 14, 3136–3148. doi:10.1038/s41396-020-00745-5.