Potential for oil hydrocarbon biodegradation by microbial communities in marine environments

A Dissertation Presented for the

Doctor of Philosophy

Degree

The University of Tennessee, Knoxville

John Izaak Miller

December 2019

Abstract

In 2010 the Deepwater Horizon spill released 4.9 million barrels into the Gulf of Mexico before being contained in September of that year. The native microbial community degraded petroleum in the water column, and a variety of hydrocarbon-degrading microorganisms were present in the deep-sea plume. In the years following the DWH spill (2011-present), there have been 1,917 reported spill or contamination events resulting in more than 2,989,675 barrels of petroleum released into the environment. Traditional methods for cleaning up petroleum spills are expensive in efficient; bioremediation (i.e., biodegradation by microorganisms) of petroleum can be effective while having minimal additional adverse affects on the environment However, environmental variables such as temperature, oxygen concentration, and available nutrients impact the rate and extent of microbial biodegradation of petroleum. This dissertation examines microbial communities in six marine basins around the world with particular attention to the petroleum degrading sub-population to determine the potential for oil biodegradation in each basin. Oil degrading bacteria were ubiquitous in all six marine environments. The Eastern Atlantic receives nutrient enrichment from the Angola-Benguela Front, which is likely to facilitate oil biodegradation in that basin. Microcosm experiments were conducted to investigate the effect of oxygen amendment to communities from hypoxic environments, which one might expect to enhance oil biodegradation. Most of the individual oil hydrocarbons were degraded more efficiently in anoxic microcosms. Anoxic microcosms degraded a larger proportion of the shorter aliphatics ($<$ 22 carbons; including branched aliphatics, e.g., pristane and phytane) than oxic microcosms. Conversely, oxic microcosms more completely degraded most of the longer aliphatics ($\geq$ 22 carbons). Anoxic microcosms were similar to or exceeded oxic microcosms in the degradation of all aromatic hydrocarbons. When the communities are exposed to

oxygen, aerobic organisms may grow rapidly on labile carbon and dominate the community, limiting overall oil biodegradation. On the other hand, these effects would be limited under anoxic or hypoxic conditions. These results suggest that deep water communities in the Caspian Sea are adapted for better oil biodegradation under anoxic or hypoxic conditions when compared to oxic conditions.

Table of Contents

1 Introduction **1**

1.1 Petroleum spill events . . . 2

1.2 Formation, composition, and structure of petroleum . . . 3

1.3 Microbial biodegradation of petroleum hydrocarbons . . . 4

1.4 Oxygen minimum zones . . . 15

1.5 Central Hypothesis . . . 16

1.6 Specific Aims . . . 16

2 Composition of microbial communities in oxygen minimum zones **18**

2.1 Introduction . . . 19

2.1.1 Marine basins . . . 19

2.2 Methods . . . 28

2.2.1 Sample collection and environmental factors . . . 28

2.2.2 Acridine orange direct cell counts . . . 30

2.2.3 DNA extraction and 16S rRNA gene amplicon sequencing . . . 30

2.2.4 Sequence read processing . . . 30

2.2.5 Statistical analysis . . . 31

2.2.6 Alpha diversity analysis . . . 31

2.2.7 Beta diversity analysis . . . 31

2.2.8 16S rRNA Gene enrichment analysis using Fisher's exact test . . . 32

2.3 Results and Discussion . . . 32

2.3.1 Site descriptions . . . 32

2.3.2 Within-community diversity (alpha-diversity) . . . 34

2.3.3 Between-community diversity (beta-diversity) . . . 45
2.3.4 Characterization of microbial communities in oxygen minimum zones 46
2.3.5 Possible effects of oil spill on microbial communities . . . 50

3 Differences in functional potential of microbial communities from eutrophic and oligotrophic environments 51

3.1 Introduction . . . 52
3.2 Methods . . . 54
3.2.1 Site Descriptions and Sample Collection . . . 54
3.2.2 Physical and Chemical Measurements . . . 54
3.2.3 DNA Extraction and Sequencing . . . 55
3.2.4 Automated pipeline for extraction of 16S reads and taxonomic assignment . . . 55
3.2.5 Identification of potential functional genes . . . 57
3.2.6 Statistical analysis . . . 58
3.2.7 Functional annotation of sequence reads with mi-faser . . . 62
3.3 Results and Discussion . . . 64
3.3.1 Site descriptions: Physical and geochemical features . . . 64
3.3.2 Sequencing results . . . 66
3.3.3 Diversity, richness, and enriched taxa . . . 66
3.3.4 Functional potential of microbial communities . . . 70

4 Oil Hydrocarbon Degradation by Caspian Sea Microbial Communities 90

4.1 Introduction . . . 91
4.2 Materials and methods . . . 94
4.2.1 Sample collection and environmental parameters . . . 94
4.2.2 Acridine orange direct cell counts . . . 97
4.2.3 Microcosms experiments . . . 97
4.2.4 Hydrocarbon analysis . . . 98
4.2.5 DNA extraction, 16S rRNA gene amplicon sequencing, and data pre-preprocessing . . . 98

4.2.6 Bioinformatics processing of sequence data 99
4.2.7 Statistical analysis 99
4.3 Results and discussion 100
4.3.1 Distinct environmental niches exist in the Caspian Sea 100
4.3.2 Microbial community composition changes with environmental niche . 101
4.3.3 A microbial community from a deep water OMZ rapidly degrades recalcitrant hydrocarbons 105

5 Conclusions and future directions **117**
5.1 Research summary and final conclusions 118
5.2 Future Directions 119

Bibliography **121**

Vita **165**

List of Tables

1.1 Microbes known to degrade polycyclic aromatic hydrocarbons. Table from [48]. 7

2.1 Table: Comparison of physical and chemical factors in basins around the globe that have a history of oil biodegradation. Table from Hazen and Techtmann [116]. 29

2.2 Range of OTUs observed in each basin. The distribution of OTUs observed per sample in basins marked with an asterisk * was not normal (Shapiro-Wilk test P-value <0.05). 37

2.3 *Chao*1 richness estimates for each basin. The number of observed OTUs per sample in basins marked with an asterisk * was not drawn from a normal distribution (Shapiro-Wilk test p-value <0.05). 42

3.1 Sample locations. 65

3.2 Overview of shotgun metagenomic sequencing. 67

4.1 The important seafloor features near each sampling site are listed. Sample sites were chosen in part based on important sea floor features. Some features are man-made (M), while others result unintentionally from anthropogenic activities (A). All other features are assumed to be natural (N). 96

4.2 Summary of environmental parameters by depth category. Environmental parameters for shallow ($\leq$ 50 m) and deep ($>$ 50 m) waters were tested for statistically different means using a two-tailed t-test. Statistically significant environmental parameters (p-value < 0.05) are indicated by *. 108

4.3 List of taxa that were enriched in an anoxic, oil amended microcosm as determined by TEXMEX analysis. Changes in the microbial community composition of the anoxic, oil-amended microcosm were compared with that of the anoxic, non-amended microcosm. Microbial taxa that were enriched in the oil-amended microcosm relative to the non-amended microcosm are listed here. 111
4.4 List of taxa that were enriched in an oxic, oil amended microcosm as determined by TEXMEX analysis. Changes in the microbial community composition of the oxic, oil-amended microcosm were compared with that of the oxic, non-amended microcosm. Microbial taxa that were enriched in the oil-amended microcosm relative to the non-amended microcosm are listed here. 112

List of Figures

1.1 Structural formulas, solubility, and carcinogenicities of model PAHs. Recalcitrance (or resistance to enzymatic degradation) and carcinogenicity increase with additional aromatic rings, while, simultaneously, solubility decreases. Figure from [48]. 5

1.2 Petroleum degrading microbial taxa. Petroleum degrading microbial phyla, highlighted in red, have been identified from all three domains of life. Figure from [116]. 8

1.3 Overview of aerobic degradation of hydrocarbons by microorganisms. Figure from [62]. 9

1.4 Potential pathways for anaerobic biodegradation of n-alkanes. (A) Proposed pathway for biodegradation by sulfure-reducing bacteria. (B) Proposed pathway for biodegradation by denitrifying bacteria. Figure from [111]. . . 11

1.5 Microbial biodegradation of aromatic hydrocarbons. Initial steps in the microbial pathways for oxidation of polycyclic aromatic hydrocarbons. Figure from [48]. 13

1.6 Methods for the removal of polycyclic aromatic hydrocarbons (PAHs) from contaminated environments. Non-biological processes convert PAHs directly to non-toxic compounds, while biological processes (bioremediation) converts PAHs to intermediate metabolites. Figure from [48]. 14

2.1 Idealized model of the flow of phosphate through the food web of microbial communities. Phytoplankton require phosphate for growth; i.e., low phosphate concentrations limit phytoplankton growth and thus limit primary production. Solid arrows represent the flow of phosphate, and broken arrows represent input of limiting nutrients. Figure from Thingstad et al. [284]. . . 23
2.2 Physical variables for each basin. 35
2.3 Chemical variables and nutrients for each basin. 36
2.4 Alpha-diversity metrics of microbial communities from different basins and water masses (by depth category). 39
2.5 Alpha-diversity metrics of microbial communities from OMZs compared to all other communities. 40
2.6 PCoA with confidence ellipses illustrates the dissimilarity between microbial communities. (A) Confidence ellipses drawn around microbial communities from each basin illustrate communities from most basins are generally similar to each other, with the exception of communities from the Caspian Sea, which are distinct from other communities as indicated be the lack of overlap between confidence ellipses. (B) Confidence ellipses drawn around microbial communities for each basin and depth location illustrate that some communities are distinct from each other by depth. 47
2.7 Heatmap and heirarchical clustering of Bray-Curtis dissimilarity matrix reveals clusters of microbial communities. Higher color intensity indicates similarity between samples. Dendrograms were constructed using agglomerative heirarchical clustering with the average linkage method. Color bars indicate either basin or depth (higher blue intensity indicates deep water sample). Caspian Sea microbial communities cluster together and are distinct from other communities. Shallow water and deep water communities tend to cluster together. 48

3.1 Overview of workflows in popular automated methods for analysis of shotgun sequences. (A) Metagenomic data uploaded to MG- RAST is analyzed and annotated with this pipeline. (B) Metagenomic data uploaded to the EBI Metagenomics (EMG) portal is analyzed and annotated with this pipeline. Figure adapted from Meyer [190], Meyer et al. [189], Mitchell et al. [195], and pip [3]. 56
3.2 Pipeline used by the mi-faser web-server. 63
3.3 Comparison of sequencing depth (number of reads) obtained for the 16S rRNA gene from the amplicon and from the shotgun results. Sequencing depth was similar for most of the sampled communities, with the exception of the shallow water communities of the Eastern Mediterranean Sea, in which sequencing depth was two-fold higher. (A) Sequencing depth based on 16S sequences extracted from shotgun metagenome sequence. (B) Sequencing depth based on 16S rRNA amplicon sequencing (as reported in Chapter 2). PCR amplification may be biased toward particular sequences (i.e., those derived from specific taxa), and thus affect the diversity analysis of these data. 69
3.4 Richness and alpha-diversity of water masses in the Caspian Sea and Eastern Mediterranean Sea. Results are based on 16S rRNA gene sequences extracted from shotgun metagenomic data. Both observed richness (number of OTUs) and estimated richness (chao1) showed similar trends across water masses. The shallow water communities of the Caspian Sea tend to be slightly less rich than deep water communities; while the shallow water communities of the Eastern Mediterranean Sea show the opposite trend. Two alpha-diversity metrics (Simpson, Shannon) also showed similar trends across water masses. The shallow water communities of the Caspian Sea were much lower in diversity compared to the deep water communities; this trend was similar, but not as extreme, in the Eastern Mediterranean communities. 71

3.5 For comparison of diversity based on 16S amplicon sequencing (this figure) to diversity based on 16S annotations from shotgun sequences (above). Both observed richness (number of OTUs) and estimated richness (chao1) showed similar trends across water masses. The shallow water communities of the Caspian Sea tend to be slightly less rich than deep water communities; while the shallow water communities of the Eastern Mediterranean Sea show the opposite trend. Two alpha-diversity metrics (Simpson, Shannon) also showed similar trends across water masses. The shallow water communities of the Caspian Sea were much lower in diversity compared to the deep water communities; this trend was similar, but not as extreme, in the Eastern Mediterranean communities. 72
3.6 Benzoate degradation (KEGG map00362). 74
3.7 Aminobenzoate degradation (KEGG map00627). 75
3.8 Xylene degradation (KEGG map00622). 77
3.9 Styrene degradation (KEGG map00643). 78
3.10 Phenylalanine, tyrosine and tryptophan biosynthesis (KEGG map00400). . 79
3.11 Phenylalanine metabolism (KEGG map00360). 80
3.12 Tryptophan metabolism (KEGG map00380). 82
3.13 Environmental sources of aromatic compounds. All organisms catabolize aromatic amino acids, although animals are not able to synthesize them. Plants produce aromatic compounds in various forms including lignin, which is the most abundant polymer on earth. Petroleum and other chemicals contain aromatic compounds in various forms including BTEX (benzene, toluene, ethylbenzene and xylene), many of which are toxic and/or carcinogenic. Aromatic hydrocarbons are degraded through a wide variety of pathways, but ultimately are funnelled into the central metabolic pathways. Figure from Fuchs et al. [86]. 83
3.14 Polycyclic aromatic hydrocarbon degradation (KEGG map00624). 84
3.15 Nitrogen metabolism (KEGG map00910). 86
3.16 Sulfur metabolism (KEGG map00920). 87

3.17 Methane metabolism (KEGG map00680). 89

4.1 Environmental factors varied with depth. Environmental variables were measured throughout the water column from the surface water to near sea floor. A thermocline was observed at 50 m below sea level. Below the thermocline, dissolved oxygen and pH declined with depth. 95

4.2 Acridine orange direct cell counts (AODCs). 102

4.3 Relative abundance of microbial taxa in natural communities at Order level. 104

4.4 Non-metric multidimensional scaling (NMDS) correlation biplot of shallow and deep-water communities and influential environmental variables. NMDS of weighted UniFrac distances (stress 0.075, P-value <0.05) are shown. Shallow water communities are marked with squares, and deep-water communities are marked with circles. The blue intensity of the marker indicates the depth, with the shallowest samples in gray and the deepest samples in dark blue. Environmental variables were fit to the two-dimensional representation of the microbial communities; arrow directions indicate correlation with each axis and the size of the length of the arrow indicates the correlation coefficient. . 106

4.5 TEXMEX plots. 110

4.6 Hydrocarbons from microcosm experiments were quantified and are reported according to type (aromatic or aliphatic). Significant differences in the amount of each hydrocarbon type are indicated (alpha < 0.05, alpha < 0.001). Anoxic microcosm communities (purple, blue) degraded significant amounts of both aromatic and aliphatic hydrocarbons, whereas oxic microcosm communities (green, yellow) did not. At day 17, the amount of both aromatic and aliphatic hydrocarbons was significantly less in anoxic microcosms (blue) compared to oxic microcosms (yellow). 115

4.7 The most abundant microbial taxa that were enriched in either the natural or microcosm communities. Relative abundance of each taxon is represented by intensity of the heat map. Enriched taxa were determined using Fisher's exact test. 116

Chapter 1

Introduction

1.1 Petroleum spill events

In April of 2010 an explosion on the Deepwater Horizon (DWH) petroleum drilling rig initiated the largest recorded petroleum spill, releasing 4.9 million barrels into the Gulf of Mexico before being contained in September of that year [292]. It was later reported that the native microbial community degraded petroleum in the water column, and a variety of hydrocarbon-degrading microorganisms were present in the deep-sea plume [117]. Furthermore, the microbial communities appeared to change rapidly as the composition of the petroleum changed during the degradation process [117]. After this event, BP (formerly British Petroleum) was under increased scrutiny by countries in which they had leases for petroleum production and exploration to ensure that a similar event would not happen again (personal correspondence). Therefore, BP set out to do a comprehensive risk assessment of several marine basins around the world in which the company held petroleum leases. A component of this risk assessment was the investigation and characterization of the in situ microbial community at these sites, which was motivated by the finding of rapid biodegradation of petroleum by the native microbial community in the Gulf of Mexico [117].

In the years following the DWH spill (2011-present), there have been 1,917 reported spill or contamination events resulting in more than 2,989,675 barrels of petroleum released into the environment [5]. These incidents may occur during petroleum exploration or production-related activities, or due to improper storage or transport [282, 311, 113, 101]. Petroleum spills and petroleum hydrocarbon contamination have severe, detrimental effects in the immediate exposure site as well as nearby ecosystems into which even small amounts of hydrocarbons are transmitted [273, 251]. Traditional methods for cleaning up petroleum spills include skimming (to recover petroleum from the surface of water), shoreline cleanup, administration of dispersants, burning petroleum from the surface of the water, and bioremediation (Figure 1.6) [5]. Of these available methods, bioremediation (i.e., biodegradation by microorganisms) of petroleum can be effective while having minimal additional adverse affects on the environment [265, 126, 220]. However, environmental variables such as temperature, oxygen concentration, and available nutrients impact the rate and extent of microbial biodegradation of petroleum [176, 19, 118].

1.2 Formation, composition, and structure of petroleum

Petroleum is a naturally occurring component of marine biomes [174]. The term "petroleum" can refer both to "crude oil" and to refined petroleum products such as gasoline; in this work the term "petroleum" is used exclusively to refer to "crude oil", and I attempt to avoid the term "oil", which may refer to many hydrophobic liquids. Petroleum is a mixture of hydrocarbons that is formed naturally from algal biomass over millions of years [119, 116]. An estimated 600,000 tonnes of petroleum enters marine environments around the world each year [59]. Natural petroleum seeps are hypothesized to contribute the largest fraction of this estimate, with anthropogenic runoff accounting for the next largest fraction. However, petroleum spill events (not including major events such as the DWH) routinely contribute approximately 45,000 tonnes of petroleum per year, on average.

Petroleum is an extremely complex mixture containing hundreds or thousands of different compounds [181]. Petroleum is ~90% comprised of hydrocarbons, with linear/branched alkanes, cyclic alkanes, and aromatics making up equal proportions, on average [118, 286]. The remaining ~10% is comprised of various sulfur- and nitrogen-containing compounds. During biodegradation of petroleum, the lighter, more labile components are degraded first, and overtime the remaining petroleum becomes more dense and recalcitrant [299]. Microbial biodegradation at natural seeps is thought to produce "unresolved complex mixtures" containing thousands of component chemicals [85, 120].

Linear, branched, and cyclic alkanes (naphthenes) together comprise ~60% of petroleum [273, 137]. Linear alkanes are often the most abundant component of petroleum. Short-chain alkanes (C1-C9) are more toxic than long-chain (>C9) because they can pass through the cell membrane [262]. Alkanes with fewer than ~40 carbon atoms are typically degraded in 1-2 months [118].

Cyclic and polycyclic aromatic hydrocarbons (PAHs) comprise as much as 30% of petroleum [137]. Benzene is the simplest aromatic hydrocarbon (a single aromatic ring), and naphthalene is the simplest PAH (two fused aromatic rings). These two chemicals are often the most abundant aromatic compounds in petroleum [212, 187]. Other model PAHs include acenaphthene, anthracene, phenanthrene, fluoranthene, pyrene, benz[a]anthracene,

and benzo[a]pyrene (Figure 1.1) [48]. PAHs in the environment are sometimes classified as either petrogenic or pyrogenic [68, 176]. Petrogenic PAHs are components of petroleum that can enter an environment either through natural seeps or through petroleum spills, while pyrogenic PAHs are combustion products that are spread through the atmosphere. PAHs are highly hydrophobic and therefore they can persist in the environment for years [142]. Biodegradation of PAHs is possible, but these molecules are often considered to be recalcitrant [60, 144, 219]. The preferential biodegradation of more labile hydrocarbons can lead to persistence of PAHs [266]. Because of their low solubility in water, PAHs are often more concentrated in soil and sediments [21, 48].

1.3 Microbial biodegradation of petroleum hydrocarbons

The half-life of dispersed petroleum in marine environments vary widely from days to months [118]. Methods for cleanup of spilled petroleum include burning, UV oxidation, fixation, and solvent extraction [91, 205, 88, 5]. These methods, however, are often high in cost, difficult to implement in situ, and may not completely break down the petroleum components. On the other hand, microbial biodegradation can completely mineralize petroleum hydrocarbons to carbon dioxide and small organic molecules that can be incorporated into microbial biomass [153, 178]. Many marine microbes are able to degrade petroleum hydrocarbons, and these organisms are now considered to be ubiquitous in marine environments [119, 107]. The advantages of microbial biodegradation make this an attractive option for bioremediation of contaminated environments [19].

During an petroleum spill event, the petroleum presents a carbon surplus that may contribute to microbial growth [118]. However, nutrients including nitrogen, phosphorous, potassium, and iron may become limiting, preventing microbial biodegradation of the available petroleum hydrocarbons [91]. Furthermore, environmental conditions such as temperature, available dissolved oxygen, mixing of the water column, and diffusion of petroleum impact the rate of microbial biodegradation. One way to increase the rate

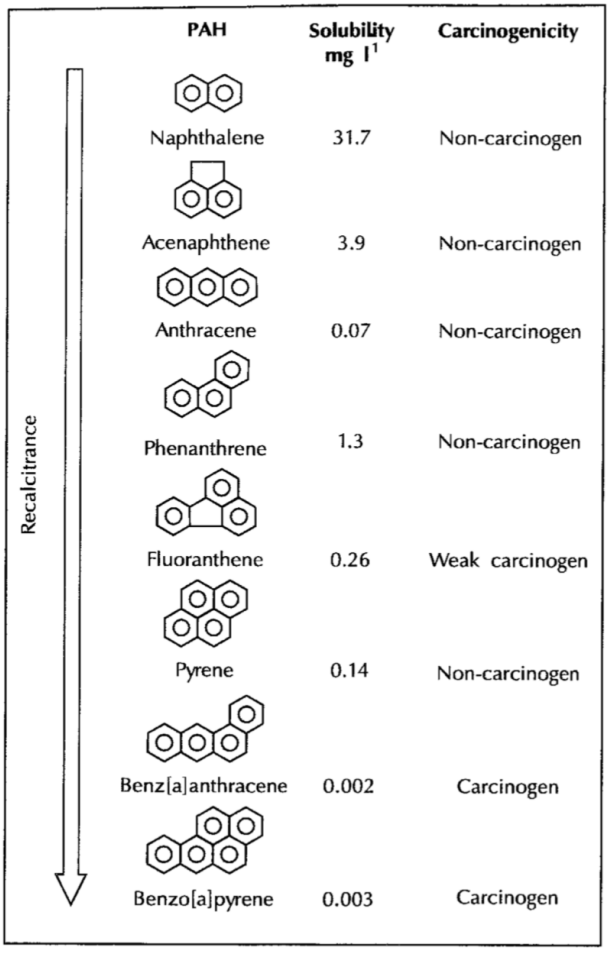


Figure 1.1: Structural formulas, solubility, and carcinogenicities of model PAHs. Recalcitrance (or resistance to enzymatic degradation) and carcinogenicity increase with additional aromatic rings, while, simultaneously, solubility decreases. Figure from [48].

of microbial biodegradation is to add fertilizer to increase the available nitrogen and phosphorous [223]. This method was used with success to assist in the cleanup of the Exxon Valdez spill in Prince William Sound, Alaska over a period of three years [224]. Both laboratory-experiments and field-testing suggested that this strategy stimulated the growth of hydrocarbon-degraders and increased the rate of petroleum biodegradation [18]. Nevertheless, natural, labile carbon sources are likely to be consumed by microorganisms prior to contaminant carbon sources [266]. Petroleum biodegradation may be increased by increasing the accessibility of recalcitrant petroleum components (e.g., polycyclic aromatic hydrocarbons; Table 1.1) [91]. Oxygen limitation may be an additional factor affecting the rate of petroleum biodegradation, but the extent to which petroleum contamination impacts oxygen concentration in the water column is disputed [118, 176].

Microbial communities are a vital component of carbon cycling in the ocean [118]. Petroleum that enters marine environments through natural seeps on the ocean floor can be a carbon source for microbial communities (Figure 1.2) [59, 116]. Most petroleum that enters aerobic marine waters is degraded within weeks, but if it becomes trapped in anaerobic sediments, in can remain there for much longer (Figure 1.3). Petroleum is a complex, heterogeneous mixture, and its biodegradation also requires a complex mixture of microorganisms [117, 182, 151, 238]. Following the DWH oil spill, a few petroleum-degrading microbes were consistently enriched, which was accompanied with an overall decrease in diversity. This process has been describes as "bloom and succession" [70]. Knowledge about microbial community composition, diversity, and functional potential will enable the development of strategies to predict petroleum biodegradation by microbial communities in situ.

Microbial biodegradation of petroleum is complex and involves interactions between microbial taxa and the environment [176]. Nutrient availability, temperature, and solubility of hydrocarbons are a few of the important environmental factors that affect petroleum biodegradation in impacted sites. This makes it difficult to link specific microbial taxa to specific petroleum component chemicals. Further complicating the issue is the fact that sequencing bias leads to systemic distortions in diversity analysis in microbial communities, making accurate assignment specific taxa to specific functions difficult [10, 285].

Table 1.1: Microbes known to degrade polycyclic aromatic hydrocarbons. Table from [48].

Bacteria	Cyanobacteria and algae
Acinetobacter calcoaceticus	Agmenellum quadruplicatum
Acinobacter sp.	Anabaena sp (strain CA)
Aeromonas sp.	Anabaena sp (strain 1F)
Alcaligenes denitrificans	Amphora sp.
Acaligenes faecaolis	Aphanocapsa sp.
Arthrobacter polychromogenes	Chlorella autotrophica
Arthrobacter sp.	Chlorella sorokiniana
Bacillus cereus	Chlamydomonas angulosa
Beijerinckia sp.	Coccochloris elabens
Corynebacterium renale	Cylindrotheca sp.
Flavobacterium sp.	Dunaliella tertiolecta
Micrococcus sp.	Microcoleus chthonoplastes
Moraxella sp.	Navicula sp.
Mycobacterium sp.	Nitzschia sp.
Nocardia sp.	Nostoc sp.
Pseudomonas cepacia	Oscillatoria sp. (strain JCM)
Pseudomonas fluorescencs	Oscillatoria sp. (strain MEV)
Pseudomonas paucimobilis	Porphyridium cruentum
Pseudomonas putida	Selenastrum capricornutum
Pseudomonas testeroni	Synedra sp.
Pseudomonas vesicularis	Ulva fasciata
Pseudomonas sp.	
Rhodococcus sp.	
Staphylococcus auriculans	
Streptomyces griseus	
Streptomyces sp.	
Vibrio sp.	

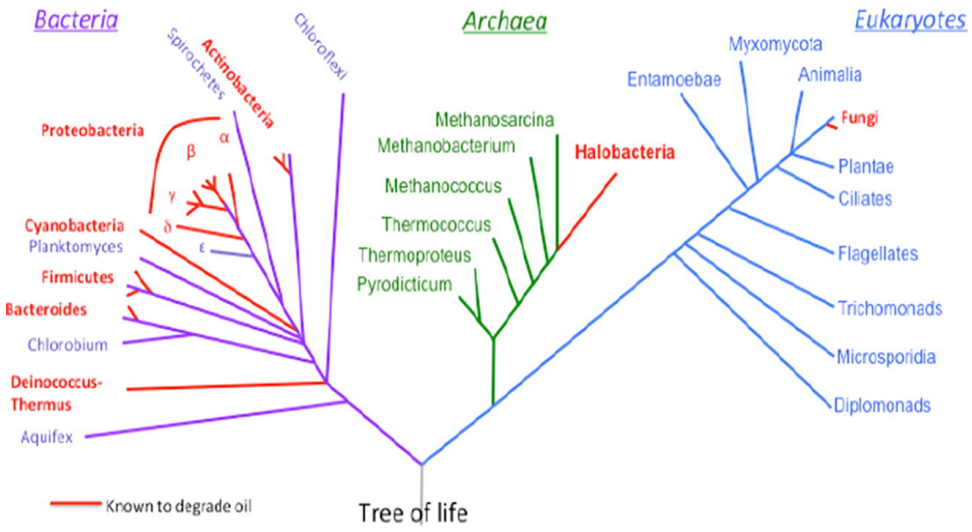


Figure 1.2: Petroleum degrading microbial taxa. Petroleum degrading microbial phyla, highlighted in red, have been identified from all three domains of life. Figure from [116].

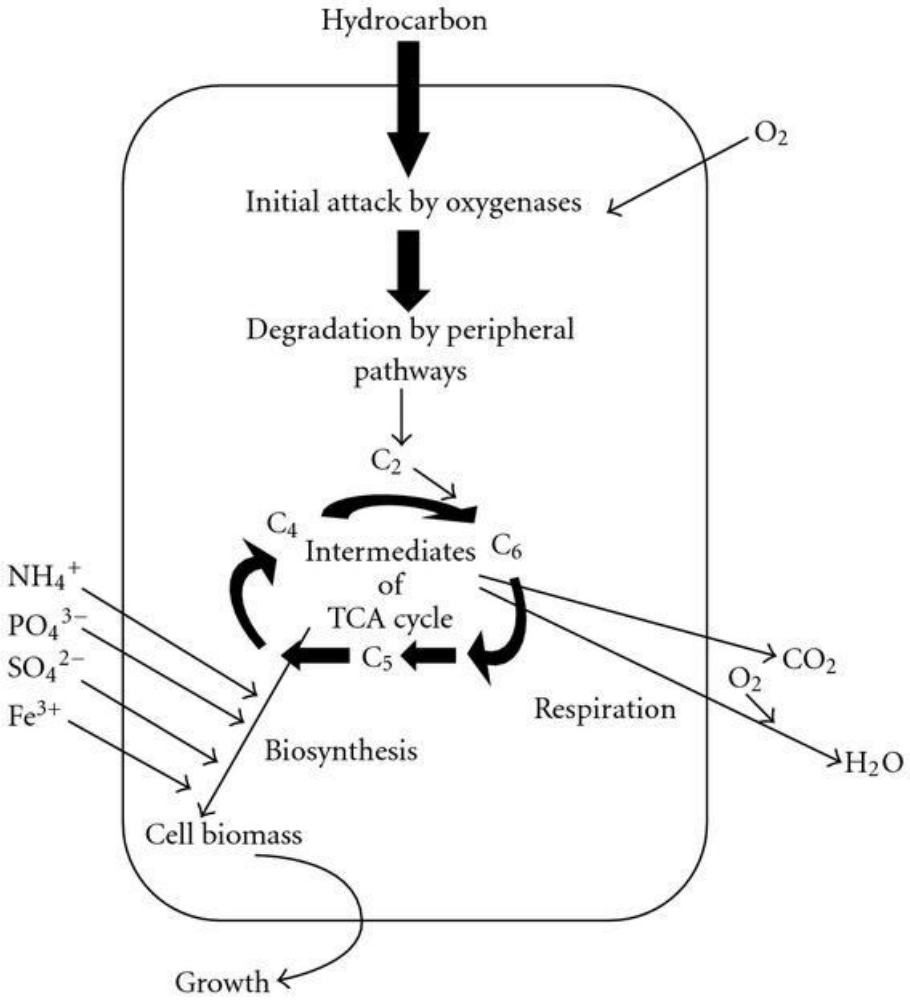


Figure 1.3: Overview of aerobic degradation of hydrocarbons by microorganisms. Figure from [62].

It has been estimated that more than 67,000 tonnes of petroleum enter the Gulf of Mexico through natural seeps each year, and that the microbial community is therefore well-adapted for its biodegradation [59]. Following the DWH spill, *Oceanospirillales* quickly dominated the community, and it was hypothesized that these bacteria degrade alkanes [117, 182, 70, 152]. Approximately 6 weeks later, the microbial community composition had changed and was dominated by *Colwellia* and *Cycloclasticus*. When the well was finally closed approximately 12 weeks later, the community composition had again shifted and was dominated by *Flavobacteria, Alteromonadaceae,* and *Rhodobacteraceae,* which are hypothesized to degrade high molecular weight organic compounds. Microcosm experiments have confirmed these results, indicating that *Oceanospirillales, Colwellia,* and *Cycloclasticus* are enriched following amendment of sea-water samples with petroleum hydrocarbons [173].

To date, 175 Bacteria and Archaea have been identified that are capable of growing on petroleum as the sole carbon source [118]. Nitrogen, phosphorous, and other nutrients can be limiting to microbial growth, but the concentrations of these nutrients in marine environments is usually adequate for microbial growth. However, some marine basins, such as the Eastern Mediterranean Sea, are known to have quite low phosphorous content; it is unknown what affect this would have on petroleum biodegradation in the event of a spill [157, 284, 158]. The local nutrient conditions could substantially affect overall petroleum biodegradation rates, even when the initial microbial community composition is similar.

Bacteria are capable of petroleum hydrocarbon biodegradation under both aerobic and anaerobic conditions (Figure 1.4) [119, 225]. Aerobic heterotrophic Bacteria and Archaea are found throughout the water column, even in areas of hypoxia, while anaerobic heterotrophs are more localized to anoxic sediments and hydrocarbon seeps [226, 298]. Under anoxic conditions, respiration of hydrocarbons requires the presence of an alternative electron acceptor such as iron(III), sulfate, or nitrate [82, 122, 58]. It is assumed that anaerobic hydrocarbon biodegradation is much slower that aerobic biodegradation, due to the lower reduction potential of the typical alternative electron acceptors. Dissolved oxygen content in some marine basins fluctuates due to vertical mixing (caused by temperature changes in the water column) or currents, but other basins, such as the Caspian Sea, have semi-permanent anoxic or hypoxic zones [218, 204, 139].

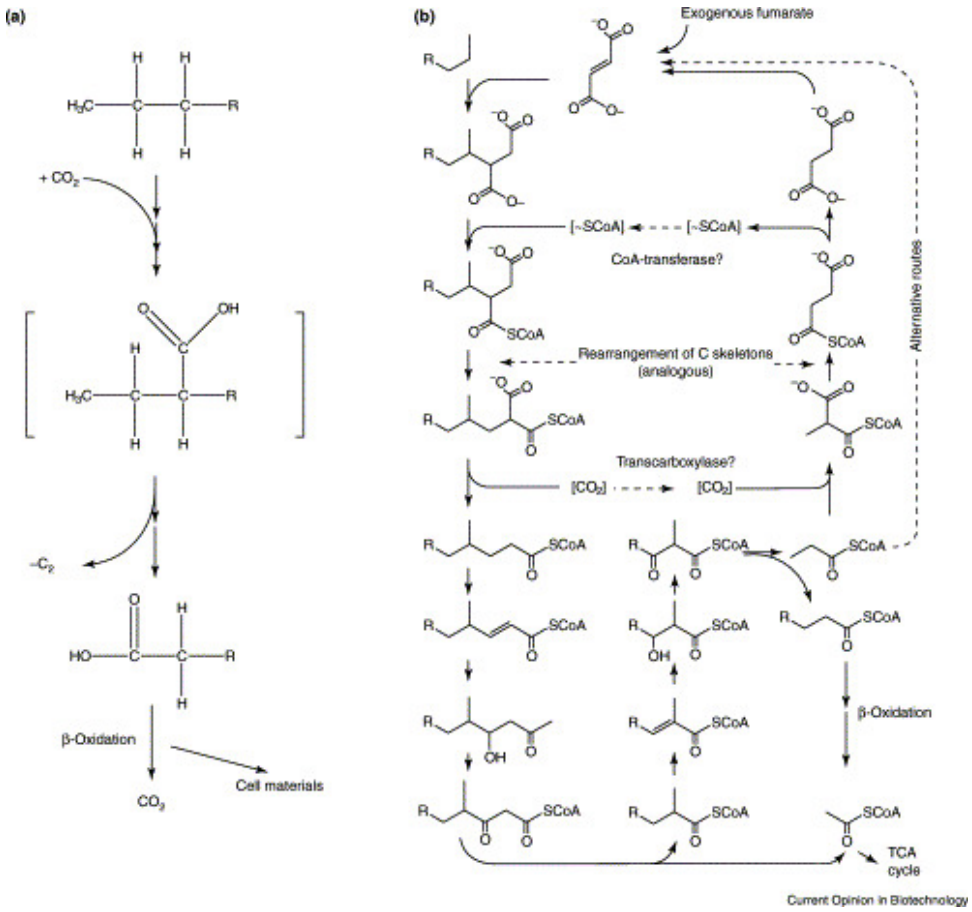


Figure 1.4: Potential pathways for anaerobic biodegradation of n-alkanes. (A) Proposed pathway for biodegradation by sulfure-reducing bacteria. (B) Proposed pathway for biodegradation by denitrifying bacteria. Figure from [111].

Degradation of many recalcitrant hydrocarbons including 4 and 5 ring PAHs by both bacteria and fungi has been documented [219]. Fungi in particular are capable of degrading a wide assortment of hydrocarbons that are similar in structure to constituent compounds found in petroleum [219]. While some organisms may degrade only alkanes or only aromatics, others are capable of degrading both (Figure 1.5) [109, 62, 207]. For example, *Colwellia* degrades propane, ethane, and butane, while *Cycloclasticus* degrades BTEX (butane, toluene, ethylbenzene, and xylene) and some PAHs [118, 234]. The ability to degrade petroleum hydrocarbons is believed to be not only the result of adaptation (in and around hydrocarbon seeps, for example), but also as a result of broad substrate specificity by some enzymes [109, 212].

Petroleum hydrocarbons and other complex organic contaminants are sometimes degraded in a step-wise manner, in which multiple organisms are required to completely degrade the larger hydrocarbons [103]. This may be the norm in hypoxic or anaerobic environments in which microorganisms rely on alternative electron acceptors [303]. It is well-known that fungi are able to degrade many recalcitrant PAHs [31, 21, 47]. Most fungi, however, are not capable of growth on high molecular weight PAHs as the sole carbon source, and it has been hypothesized that bacteria in PAH-contaminated soil metabolize the PAH-degradation products produced by fungi [47, 140]. Furthermore, synergistic metabolism involving bacteria and fungi can improve overall hydrocarbon biodegradation [31, 52, 150].

The topic of microbial biodegradation of petroleum hydrocarbons has been visited numerous times over the years [17, 48, 188, 180, 288]. Heterotrophic bacterial degradation of hydrocarbons is an energetically favorable, 'assimilatory' process [273]. The most easily degraded hydrocarbons are alkanes, followed by mono-aromatic hydrocarbons, then cyclic alkanes, and then polycyclic aromatic hydrocarbons [107]. The most common pathway for microbial hydrocarbon biodegradation begins with incorporation of one atom of molecular oxygen into the substrate catalyzed by an oxygenase or peroxidase enzyme [118]. Degradation of aromatic compounds is typically initiated by a monooxygenase or dioxygenase enzyme that incorporates one or both atoms of molecular oxygen into the substrate [21]. The products of these reactions are converted to intermediates of the tricarboxylic acid (TCA) cycle. Alkanes are similarly converted into fatty acids which are then metabolized via beta-oxidation [239].

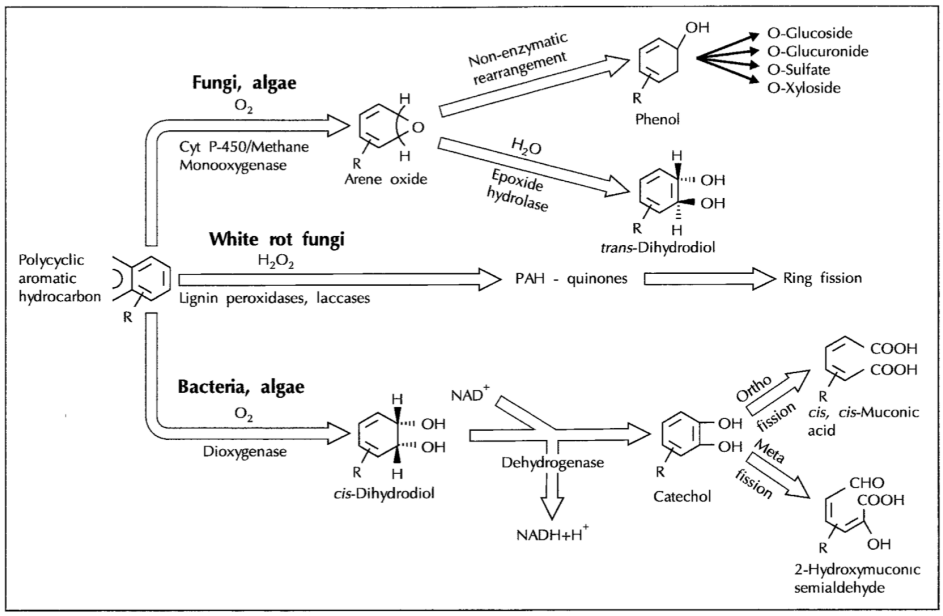


Figure 1.5: Microbial biodegradation of aromatic hydrocarbons. Initial steps in the microbial pathways for oxidation of polycyclic aromatic hydrocarbons. Figure from [48].

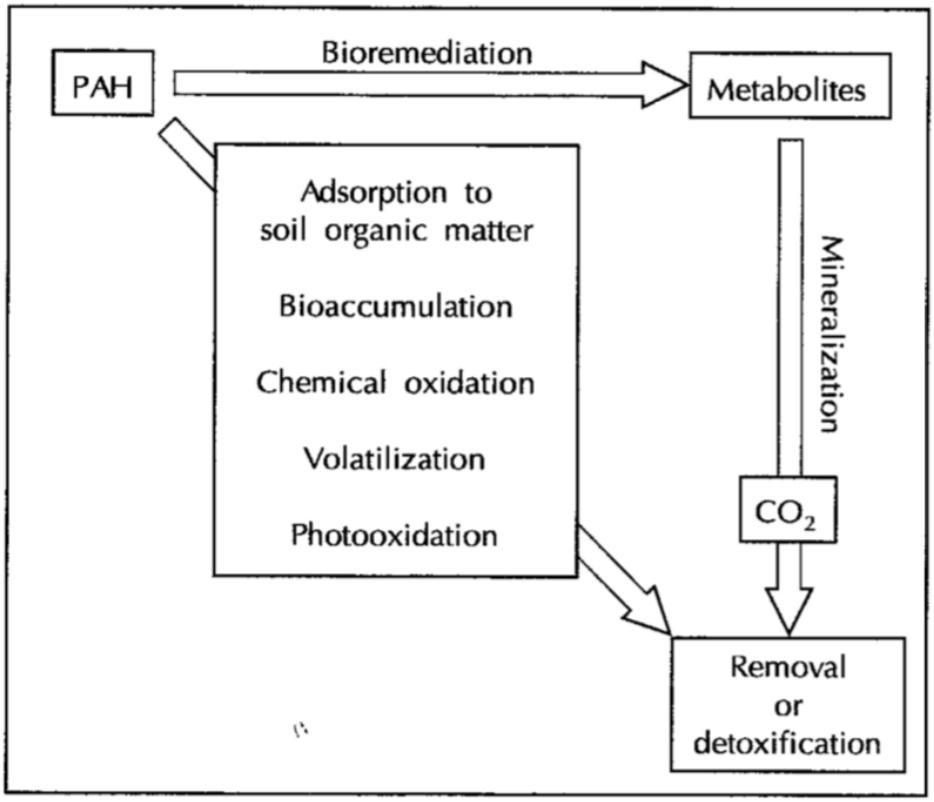


Figure 1.6: Methods for the removal of polycyclic aromatic hydrocarbons (PAHs) from contaminated environments. Non-biological processes convert PAHs directly to non-toxic compounds, while biological processes (bioremediation) converts PAHs to intermediate metabolites. Figure from [48].

1.4 Oxygen minimum zones

Microbial communities are a critical component of the global carbon and nitrogen cycles [61, 310, 172]. The composition and functional capacity of marine microbial communities can be altered by pollution in the form of nutrient enrichment ("eutrophication") [271, 191, 255]. Nutrient-rich industrial and agricultural runoff into marine environments has increased substantially in the last century [32]. Eutrophication in the form of an influx of nitrates in marine waters causes increased microbial growth in shallow water, chiefly algae [129]. When these microbes die, the biomass is consumed by aerobic heterotrophic bacteria, which depletes the oxygen in the water column, forming OMZs. Stratification between the shallow and deep waters causes OMZs to persist in deep waters in the open ocean [169, 218, 139]. The shallow, warm water is less dense than the deep, cold water, and this prevents mixing of the two water masses.

In addition to pollution, human activity also affects marine environments through climate change in the form of acidification and increasing water temperatures, which may exacerbate OMZs [300]. Predicting the direct and indirect effects of climate change on microbial production in the oceans is complex [279]. Some preliminary models indicate the possibility of a positive feedback mechanism between climate change and OMZ formation, primarily by increased global nitrous oxide production. OMZs often produce nitrous oxide, which would increase along with increases in both the number of OMZs and their persistence. Other research, on the other hand, suggests that OMZs will increase biological carbon storage, acting as a negative feedback with climate change [46]. Microaerophilic bacteria in the OMZ may contribute substantially to the net release or capture of carbon dioxide [175]. Microaerophiles are organisms that require low levels of oxygen and often require high levels of carbon dioxide, but they are inhibited by either anoxia or normal atmospheric concentrations. These organisms are hypothesized to be tightly coupled to photosynthetic cyanobacteria in OMZs to consume oxygen as soon as it is produced. They respond rapidly to fluctuating oxygen concentrations; however, their actual response to seasonal and daily fluctuations in produced oxygen, and thus their contribution to carbon cycling, is not yet fully understood. Predicting the direct and indirect effects of climate change on

microbial production in the oceans is complex, and no definitive predictions have yet been made. Current models are improving but are still not sufficiently accurate to make reliable predictions based on microbial activity [247]. Further study is needed to understand the microbial processes within OMZs and their impacts on global biogeochemical cycles.

Although we generally assume that aerobic processes are faster than anaerobic processes, preliminary evidence shows that the opposite may be true under certain circumstances [241, 263]. Very few studies have directly compared aerobic and anaerobic degradation of hydrocarbons, and most of our current knowledge is focused on aerobic degradation [122]. Preliminary evidence indicates that marine microbial communities rapidly degrade recalcitrant hydrocarbons such as oil under anoxic conditions [90]. Increased biodegradation under anoxic conditions is likely caused by one of two methods: One possibility is that hypoxia causes selective pressure on the microbial community, resulting in high relative abundance of microaerophiles that degrade recalcitrant hydrocarbons rapidly. Another possibility is that one or more organisms have adapted to hypoxia by evolving unique proteins to rapidly degrade recalcitrant hydrocarbons; and the expression of those proteins is induced in hypoxia.

1.5 Central Hypothesis

Microbial communities in oxygen minimum zones are adapted to rapidly degrade recalcitrant hydrocarbons.

1.6 Specific Aims

1. Describe the diversity and composition of communities from OMZs in multiple marine basins using taxonomic information. Microbial communities from the following marine basins will be characterize: the Western and Eastern Atlantic Ocean; the Central and Eastern Mediterranean; the Caspian Sea; and the Great Australian Bight. Microbial taxa will be identified using 16S rRNA gene sequencing. The abundance of each

microbial taxon will be compared to environmental characteristics using statistical methods for enrichment and correlation analysis.

2. Compare microbial communities from a chronically eutrophic OMZ to that of a low-nutrient, oxic environment using community-wide genomic information. Microbial communities from a eutrophic OMZ (the Caspian Sea) and a low-nutrient, oxic environment (the Eastern Mediterranean Sea) will be compared using whole genome shotgun sequencing (metagenomics). From the metagenomic data, protein coding genes will be predicted and grouped into protein families to compare functional potential of the two environments. Statistical methods for "enrichment analysis" can then be used to determine which protein families are over-represented within each environmental niche.

3. Determine the microbial community composition from a eutrophic environment and test its ability to degrade oil hydrocarbons under oxic or anoxic conditions. Water samples will be taken from a eutrophic OMZ (the Caspian Sea) for microcosm experiments. These experiments will be prepared under either oxic or anoxic conditions, and then crude oil will be added to the sample. Microbial taxa that comprise the communities in these experiments will be determined using 16S rRNA gene sequencing, and hydrocarbon composition of the microcosms will be determined by GC-MS.

Chapter 2

Composition of microbial communities in oxygen minimum zones

This chapter is derived from a manuscript in progress. Data and statistical analysis, and preparation and writing of the manuscript were carried out by Miller. Sampling was carried out by Techtmann, Fortney, Mahmoudi, Joyner, Fernandez, Gardinali, GaraJayeva, and Askerov. Guidance on experimental design and data analysis was provided by Fordyce and Jacobson. Guidance on manuscript preparation was provided by Hazen.

2.1 Introduction

2.1.1 Marine basins

Atlantic Ocean

Water samples were obtained from the Eastern Atlantic (off-shore Angola). There are several different current systems that interact to form distinct water masses by depth in this region. Two current systems (the Angola and Benguela currents) intersect in this region to for the Angola Current [28]. The Angola current runs south along the coast of Angola, is generally warm (~24°C), with salinity typical of ocean waters (36.4 PSU). The current is strong in the upper layers forms part of a cyclonic gyre in the Angola Basin. The Poleward Under Current is deep to the Angola Current and runs southward, while the Benguela Current flows northward along the west coast of Africa.

Four water masses are formed as a result of currents in the region: the surface water (<100 m), South Atlantic Central Water (100 – 500 m), Antarctic Intermediate Water (500 – 1200 m), and the North Atlantic Deep Water (>1200 m) [276, 76]. All of these water masses move southward and then become part of the Angola Gyre.

The Benguela current brings cold, nutrient-rich water from the antarctic up the coast, resulting in an upwelling effect. This current meets the Angola current to form the Angola-Benguela front.

The shallow waters have high primary productivity and are saturated or supersaturated with oxygen. This leads to high rates of aerobic, heterotrophic degradation of the resulting biomass. Heterotrophic metabolism depletes oxygen in the water column, which decreases to about 10% saturation at 400 m depth, thus forming the oxygen minimum zone (OMZ). The

deep Benguela current is oxygenated, thus the waters below the OMZ have higher oxygen concentrations [196]. Oxygen concentrations in some areas are sufficiently low that anoxic processes occur at sufficient rates for sulfide to accumulate [37, 36]. Produced sulfide can be then co-metabolized during nitrate-reduction (e.g., by some *Gammaproteobacteria*) [167].

Previous analysis of microbial communities in this region identified nutrient transport as an important feature [198, 199]. Transporter proteins for nitrogenous compounds were identified, but no transporter proteins for phosphate were identified; this is consistent with the fact that nitrate is low while phosphate is high in this region. Furthermore, the Archaeal ammonia monooxygenase (Amo) protein, a marker for ammonium oxidation, was identified. The shallow waters were commonly dominated by *Prochlorococcus* spp. and SAR11 , depending on the specific sample location, while *Actinobacteria*, *Roseobacteria*, and OCS116 were also commonly enriched. However, the relative abundance of these taxa was highly variable. This variability could be due to selection (e.g., based on nutrient availability) or the result of different sampling coverage.

Nitrogen cycling in the region varies by water mass. In the Benguela upwelling region, where nutrient content is high, rates of nitrogen fixation are low, which might be due to low iron concentrations [270]. Denitrification may be the predominant mechanism for nitrate removal. Although oxygen concentrations are low, there is still sufficient oxygen to prevent anaerobic ammonium oxidation. Within the OMZ, bacteria adhere to particulate matter to carry out anaerobic ammonium oxidation [162, 309].

The primary source of hydrocarbons for hydrocarbon degrading bacteria is algal and bacterial lipids [45]. Degradation of these lipids in aerobic heterotrophic processes contributes to the formation of OMZs, as described above. Large quantities of these lipids may be present and sink through the water column. As they do so, labile hydrocarbons are degraded first, leaving more recalcitrant hydrocarbons in the deeper waters. Bacteria in the OMZs may also be capable of oil biodegradation if they have adapted to degrade these recalcitrant hydrocarbons. Additionally, hydrocarbons from seeps on the ocean floor may be degraded by heterotrophic processes. Microbes found near hydrocarbons seeps are capable of degrading oil [26]. Furthermore, members of the *Oceanospirillaceae*, which were enriched during the Deepwater Horizon spill in the Gulf of Mexico, have been reported in the Angola basin

[117, 249]. This evidence supports the hypothesis that oil degrading bacteria may be present in the Eastern Atlantic.

Oil exploration and production has been ongoing since the 1960s [302]. To date there have been only 'minor' oil spills in this region, releasing, e.g., approximately 2,000,000 barrels of oil into the Eastern Atlantic Ocean. However, it is unknown to what extent the native microbial communities may or may not be capable of degrading oil in the case of a large spill. The extent of microbial oil biodegradation has important implications for clean up of oil spills both large and small. Based on the current evidence, I expect to find putative oil degrading bacteria in this basin. Furthermore, very few reports describe the microbial communities of the deep water and OMZs. In the case of an event such as the Deepwater Horizon, the deep water communities will be the first to be effective, and potentially the most important in mitigating the harmful effects of the released oil. Both ecological and experimental studies are required to investigate the oil biodegradation potential of these microbial communities.

Several different water masses intersect in the Western Atlantic Ocean, north of South America [116]. Interesting features in this region include the output of the Amazon river into the Atlantic ocean. Methanogenic archaea are present in these rivers, and they might be an important influence on these coastal ocean waters. As with other basins in this study, mud volcanoes are present on the sea floor in this region. The deep waters in this region show evidence of multiple petroleum-degrading microbes including *Colwelliaceae*, and *Alcanivoracaceae* [40].

Mediterranean Sea

The International Hydrographic Organization has established that are two primary basins in the Mediterranean Sea: the western and eastern basins, which are separated by Italy and Sicily [1]. Water samples for this report were obtained from the eastern basin, which we divided into two regions: Central Mediterranean (off-shore Libya) and Eastern Mediterranean (off-shore Egypt). Major inputs into the Mediterranean Sea include precipitation, run off from the surrounding land, and water from the Atlantic Ocean via the Strait of Gibraltar.

Water from the Atlantic Ocean flows into the Mediterranean Sea from the Strait of Gibraltar in the west to east forming currents below the surface waters [106]. This water is generally lower in salinity than other water masses in the Mediterranean Sea. The surface waters flow near the coast, also in an easterly direction, although there are many eddies and gyres [237]. Stratification of the water masses is relatively strong during the summer months, when the surface waters are warm, and subsides during the winter months, when the surface waters are cooler. The Levantine Intermediate Water lies below the Atlantic Water at 200-500m below sea level. This water mass is colder (~15°C) and higher in salinity (~39 PSU) [159]. This water flows in to the west, in the opposite direction of the Atlantic water, and exits the Eastern basin through the strait of Sicily. The Levantine Deep Water lies below the Levantine Intermediate Water (>500 m), and is similar in temperature (~13.5°C) and salinity (~38 PSU) [293]. The high temperatures of the deep water masses and high salinity in general make the Eastern Mediterranean a unique body of water, and these environmental stresses may select for unique microbial functions.

The nutrient content of the waters of the Eastern Mediterranean is also unique. The shallow waters are very low in phosphorous, with a Redfield ratio (N:P) 29:1 (recall that expected N:P ratio is 16:1), while nitrogen and iron concentrations are relatively high [159, 284] (Figure 2.1). The low phosphate concentrations and resulting low primary productivity make the Eastern Mediterranean an ultra-oligotrophic body of water [233, 275]. The low primary productivity in the surface waters yields little biomass for aerobic heterotrophic activity by bacteria, thus oxygen concentrations are high throughout the water column.

The Mediterranean Sea is home to a diversity of microbial communities, and cell counts are typical of marine environments (1×10^5 cells/mL in shallow waters decreasing to 1×10^4 cells/mL in deep waters) [89, 29, 164, 23, 232, 313]. Bacterial classes including *Gammaproteobacteria*, *Actinobacteria*, and *Firmicutes* are commonly found in the Mediterranean Sea [313]. However, the observed composition of the microbial communities varies by both sampling location and depth of sample [29]. This observation could be the result of selective pressures that differ by water mass, or could be the result of normal temporal fluctuations in the community composition.

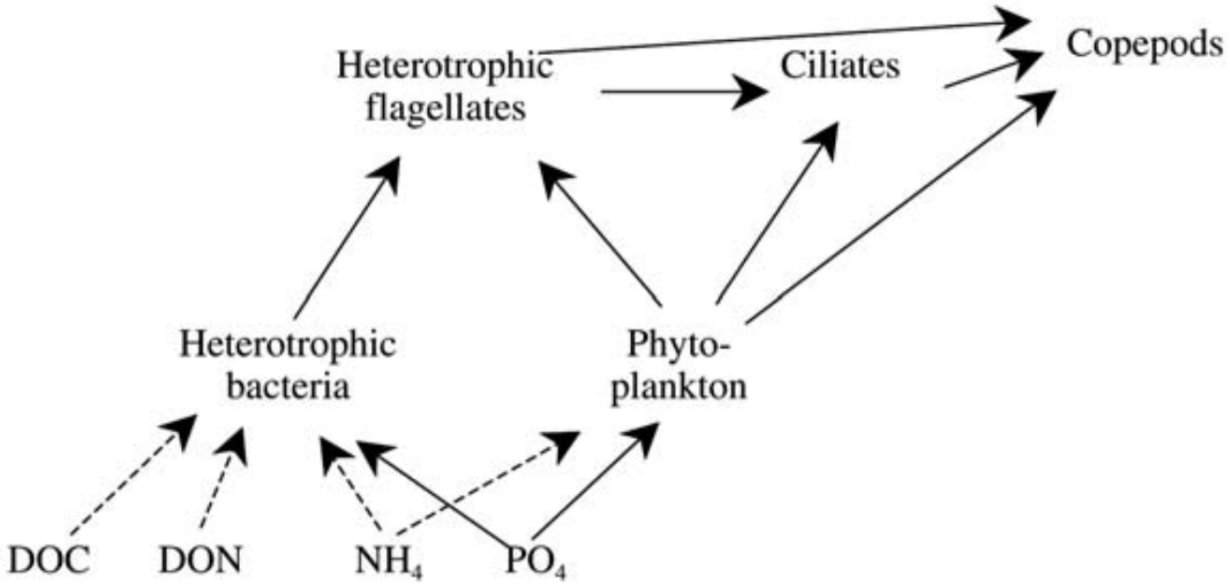


Figure 2.1: Idealized model of the flow of phosphate through the food web of microbial communities. Phytoplankton require phosphate for growth; i.e., low phosphate concentrations limit phytoplankton growth and thus limit primary production. Solid arrows represent the flow of phosphate, and broken arrows represent input of limiting nutrients. Figure from Thingstad et al. [284].

Hydrocarbon seeps have been located along the sea floor of the Mediterranean Sea, however, the amount and type of hydrocarbon input into the system from this source has not been fully investigated [260, 72, 133]. The microbial communities around these hydrocarbon seeps have been investigated [123, 124, 210, 211, 214, 94, 100]. Reports indicate that these observed microbial communities are diverse but are often dominated by *Gammaproteobacteria* [210, 94, 100, 221]. These microbial communities have been studied using a combination of experimental techniques and metagenomics. Oil enrichment experiments indicate that oil degrading bacteria are common in these waters, and hydrocarbon degrading bacteria are estimated to be as abundant as 1×10^2 cells/mL in the Mediterranean Sea [77, 53, 314, 316, 138, 200, 77, 248]. Interestingly, Cyanobacteria may be associated with this hydrocarbon degrading community [316]. It remains unknown how these Cyanobacteria are involved in hydrocarbon biodegradation; it is possible that the observed Cyanobacteria are either hydrocarbon degraders or that they form a synergistic relationship with hydrocarbon degrading bacteria by providing oxygen for aerobic degradation. Another study reported that the dominant microbes in hydrocarbon degrading communities were *Proteobacteria*, *Firmicutes*, *Actinobacteria*, and *Bacteriodetes* [248]. Experimental studies have shown that these communities are capable of degrading benzene and toluene [221, 183]. Enrichment microcosm experiments with oil combined with metagenomics revealed several novel esterase enzymes with potential applications in biotechnology [79]. However, this study did not characterize the overall community structure, so it is unknown what affect oil had on the microbial community. This study demonstrated the potential for identification of novel enzymes for industrial use, but future work should also characterize the microbial community composition. Petroleum hydrocarbon pollution in the region has been relatively minor to-date, however, hydrocarbon degrading microbes are reported to be enriched in local regions were hydrocarbon pollution is high [203, 77].

Other microbial studies on the communities of the Mediterranean Sea have focused on how these communities are adapted to the ultra-oligotrophic conditions (i.e., low phosphate and high nitrogen content). These conditions may select for small microbes in favor of larger ones [157]. Metagenomic analysis indicates a high number of proteins involved in phosphate transport and processing [78]. Additionally, microbiological experiments have shown that

some microbes from this region are capable of swimming toward inorganic phosphate [78, 135]. Nitrogen fixing microbes are highly active in the Eastern Mediterranean, and it is hypothesized that the high proportion of nitrogen in these waters is due to microbial nitrogen fixation [237, 160, 157, 242, 215].

Current information on the microbial community of the Mediterranean Sea is disparate and inconsistent, which makes it difficult to draw conclusions regarding the composition and functional potential of these microbial communities. Important questions regarding the hydrocarbon biodegradation potential of these microbial communities as well as basic ecological questions regarding the consistency and biodiversity of the community composition remain unanswered. The existing literature in this area has focused on microbial communities from sediments and coastal waters, leaving open questions about the open-water and deep-water communities [53, 316, 138, 200, 248]. These questions have important implications for assessing the potential microbial biodegradation of oil, e.g., in the event of an oil spill in this basin. Oil exploration and production is increasing in the Mediterranean Sea, as in many basins around the world. With increased production comes increased risk of an accidental spill event. The existing evidence indicates that there is a strong likelihood of finding novel hydrocarbon degrading organisms in the Mediterranean Sea that are adapted to the unique environmental conditions found there. This following chapter describes the analysis of microbial communities from multiple sites and depths within the Mediterranean Sea to answer these questions.

Caspian Sea

The Caspian Sea is the largest landlocked body of water in the world. It receives freshwater inputs from river runoff and precipitation, making the water brackish with salinity ~1/3 of ocean salinity (~12 PSU) [290]. The Caspian Sea is divided into three geographic regions: Northern, Middle, and Southern [156]. The Northern basin is substantially more shallow than the Middle and Southern basins (maximum depths are ~20 m, 788m, and 1025 m, respectively). The temperatures in the shallow waters of the Middle and Southern vary seasonally, but there is thermocline near 50 m and below 100 m the temperatures are

consistent year-round (~10°C) [290]. The currents in the Caspian Sea are cyclonic, running south along the western coast and north along the eastern coast [54].

More than 130 rivers flow into the Caspian Sea, and these inputs substantially impact the sea level and nutrient content [290]. Input of riverine freshwater has increased steadily since the mid-20th century bringing increased industrial and agricultural pollution into the Caspian Sea [291, 54]. During the same time, the oxygen concentration in the deep waters of the Middle and Southern basins have decreased substantially (from ~45% to 7% in the Middle basin and from ~26% saturation to 5% in the Southern basin). Nutrient concentrations in the shallow waters are kept low (relative to the deep waters), likely due to primary productivity in these waters [245]. Nutrient concentrations increase with depth until ~300 m, and then decrease again in the deepest waters [54]. The combination of sustained nutrient enrichment and persistent hypoxia make the deep waters of the Caspian Sea unique among marine basins.

In addition to pollution from rivers, the Caspian Sea receives an estimated 70,000-90,000 tons of hydrocarbon input from oil exploration and production, and from hydrocarbon seeps on the sea floor [54]. Hydrocarbon seeps (aka "mud volcanoes") are an important source of hydrocarbon input into the Caspian Sea, and they're presence indicates that these microbial communities have a long history of hydrocarbon exposure [93]. Furthermore, although hydrocarbon seeps are found on the sea floor in many marine basins (e.g., the Mediterranean Sea), the deep waters of most marine basins have moderate to high oxygen concentration. I.E., the hypoxic conditions of the deep waters of the Caspian Sea make these hydrocarbon seeps particularly unique.

Microbial cell counts vary seasonally, ranging from 1×10^5 to 1×10^6 cells/mL in shallow waters and decline to 1×10^4 cells/mL in the deep waters [245]. As expected, aerobic microbes tend to dominate the shallow water communities while microaerophiles tend to dominate the deep water communities. However, the impact on the unique combination of hypoxia and hydrocarbon exposure from hydrocarbon seeps in the deep water communities has not been investigated. Based on the history of exposure to hydrocarbons from natural seeps, the microbial community of the deep waters of the Caspian Sea is expected to have the potential for high hydrocarbon biodegradation.

Oil degrading bacteria and fungi have been isolated from the Caspian Sea [115, 114, 259, 296, 245, 297]. Furthermore the number of heterotrophic bacteria in general and hydrocarbon degraders specifically are both high (1×10^5 and 1×10^4 cells/mL, respectively; Hassanshahian et al. [115]). The identities of some of the oil degrading isolates have been reported (e.g., *Pseudomonas*, *Gordonia*), and the potential for bioremediation in the event of an oil spill has been reported, however the community composition in these heavily hydrocarbon impacted waters has not yet been investigated [115, 296].

Great Australian Bight

The Great Australian Bight (GAB) lies off of Australia's southern coast and receives input from the Indian Ocean and the Antarctic Ocean (aka Southern Ocean). Several different ocean currents and water masses intersect along Australia's coastlines [112]. The primary currents that affect the GAB are the Leeuwin current, the Southern Australian current, and the Flinders current [24, 268, 192]. The Leeuwin current is warm with low salinity and carries water in a southerly direction. Below the Leeuwin current is the Leeuwin undercurrent, which carries water in a northerly direction. The Flinders current is driven by wind and carries water in a westerly direction. The Southern Australian current is also driven by wind but carries high salinity water in an easterly direction.

There is also an upwelling region that brings nutrient-rich water to the surface periodically [146, 145, 193]. Primary productivity in the GAB is generally low, but elevated productivity has been observed during the summer [294, 295]. The Cyanobacteria *Prochlorococcus* and *Synechococcus* were abundant in these microbial communities. The deep waters (>185m) are dominated by *Thaumarchaeota* [281]. These microbes, along with SAR11, *Rhodobacteriales*, *Oceanospirillales*, *Alteromonadales*, and *Bacteroidetes* are commonly observed in the nearby waters of the Antarctic Ocean [256, 38].

As in other basins included in this study, the Bight is of commercial interest for oil exploration, and is one of the deepest marine basins in consideration [174, 116]. However, unlike some of the other basins, no natural hydrocarbon seeps have been identified in the GAB [277]. There is a particular dearth of information available on the microbial community of the GAB. As such, the potential for microbial hydrocarbon biodegradation in the event

of an oil spill is purely speculative. Here, we report on the microbial community of the GAB with particular attention to potential oil degrading microbes.

0cm

2.2 Methods

2.2.1 Sample collection and environmental factors

Water samples were collected as part of BP's oceanographic survey from six marine basins in 2013: Western Atlantic, Eastern Atlantic (off-shore Angola), Central Mediterranean (off-shore Libya), Eastern Mediterranean (off-shore Egypt), the Great Australian Bight (GAB), and the Caspian Sea. In total, 142 samples were obtained.

Niskin bottles were deployed for water collection. Environmental variables (temperature, dissolved oxygen, salinity, pH, turbidity) were measured with a MIDAS CTD + Profiler (Valeport Ltd, St. Peter's Quay, UK) on the sampling rosette.

At each sample location, approximately 100 liters of ambient seawater was filtered using a large volume pump (McLane Research Laboratories, East Falmouth, MA). The water was filtered through a 142 mm nylon membrane with a pore size of 0.2 µm (Sterlitech, Kent, WA) and then stored at -20°C. One third of the filter was used for DNA analysis.

Recovered seawater was dispensed into 4-liter amber bottles using clean TYGON Tubing (Saint-Gobain, La Défense, Courbevoie, France) to limit aeration. Bottles were stored on-ship at 4°C and then shipped on wet ice. Forty mL of recovered seawater was fixed in 4% formaldehyde and stored at 4°C for acridine orange direct counts (AODCs).

One hundred mL of water was frozen at -20°C for analysis of dissolved organic carbon and nutrients. Total organic carbon and total nitrogen were analyzed with TOC-L analyzer (Shimadzu Scientific Instruments, Columbia, MD), and inorganic nutrients were analyzed with a SEAL AutoAnalyzer 3 HR (SEAL Analytical Inc., Mequon, Wisconsin). Nutrients (nitrate, nitrite, ammonia, total nitrogen, inorganic phosphate, silicate) were measured at each of the sampling location.

Table 2.1: Table: Comparison of physical and chemical factors in basins around the globe that have a history of oil biodegradation. Table from Hazen and Techtmann [116].

	Nit. (g/kg)	Pho. (g/kg)	Amm. (g/kg)	Fe (g/kg)	Sul. (g/kg)	Sal. (PSU)	Tem. (°C)	DO (mg/L)
Eastern Mediterranean (Egypt)	178.4	39.7	28.0	6.5	4.7	38.9	13.8	6.4
Central Mediterranean (Libya)	160.0	29.3	30.0	ND	ND	38.6	13.7	5.5
Great Australian Bight	144.7	153.5	125.0	5.8	1.6	34.8	2.5	5.0
Caspian Sea	127.8	56.9	739.6	4.7	1.9	11.3	6.8	0.5
North Sea	569.6	33.5	12.5	ND	ND	33.7	3.0	ND
Eastern Atlantic (Angola)	20.0	1.6	10.2	ND	ND	36.0	4.0	3.4
Western Atlantic (Brazil)	25.0	1.1	ND	ND	ND	37.0	5.0	ND
Gulf of Mexico	208.0	205.0	78.0	56.0	ND	35.0	4.8	ND

2.2.2 Acridine orange direct cell counts

Acridine orange direct cell counts (AODCs) were performed as describe by Francisco et al. [83]. Cell counts were performed with Zeiss Axioskop epifluorescence microscope (Carl Zeiss, Inc., Germany).

2.2.3 DNA extraction and 16S rRNA gene amplicon sequencing

Genomic DNA was extracted the method of Miller et al. [194] with modifications as described in Hazen et al. [117]. DNA was cleaned using the Genomic DNA Clean & Concentrator kit (Zymo Research, Irvine, California). Quality of extracted DNA was determined by measuring the 260/280 and 260/230 ratios on a NanoDrop spectrophotometer (Thermo Fisher Scientific, Waltham, MA). DNA concentration was determined by PicoGreen (Thermo Fisher Scientific, Waltham, MA).

The 16S rRNA gene libraries were prepared as described by Caporaso et al. [42]. The V4 region of the 16S rRNA gene was amplified by PCR using universal primers 515f and barcoded 806r, which anneals to both *Bacterial* and *Archaeal* sequences, with Phusion DNA polymerase (Thermo Fisher Scientific, Waltham, Massachusetts). A 12 base pair barcode index was included in the reverse primer to multiplex samples for sequencing analysis. The 16S rRNA gene amplicons were then pooled together, and the quality and size of the amplicons was analyzed using a Bioanalyzer (Agilent Technologies, Santa Clara, CA). The 16S rRNA gene libraries were sequenced using a MiSeq with a V2 kit (Illumina, San Diego, CA).

2.2.4 Sequence read processing

The resulting DNA sequences were analyzed using the following QIIME (v1.9) pipeline [42]. The paired-end sequences were joined using fastq-join [16]. The joined sequences were then demultiplexed and sequences with a phred score below 20 were removed. Chimeric sequences were detected using UCHIME [74, 75] and removed. Sequences were clustered into operation taxonomic units (OTUs) at 97% sequence similarity using UCLUST [74] with the "open-reference OTU picking" protocol. Taxonomy was assigned to a representative sequence from

each OTU using UCLUST against the SILVA 132 QIIME compatible database [229, 312]. Samples with fewer than 20,000 sequences were removed from the dataset.

2.2.5 Statistical analysis

Samples were divided into two water masses (shallow and deep) for analysis of environmental parameters and microbial communities; shallow waters were defined as waters $\leq$50 m in depth, and deep waters were defined as waters >50 m in depth, which is consistent with the observed thermoclines. The hypothesis that environmental parameters were the same across basins and water masses was tested using a t-test for independent (unpaired) samples (alpha = 0.05), and *P*-values were corrected with the Bonferroni method.

2.2.6 Alpha diversity analysis

Each sample was sub-sampled (i.e., "rarefied") to 20,490 sequences prior to computing alpha-diversity metrics in order to control for differences in sequencing coverage across samples. Alpha-diversity metrics were calculated using the scikit-bio (v0.4.2) package in Python (v3.6.0). The following metrics were calculated:

- chao1 richness estimator [50]
- Shannon's index (Shannon entropy) [257, 274, 306]
- Simpson's index (Gini-Simpson form; the probability of specific encounter) [264, 134, 143]

Observed OTUs in each sample community were also recorded. Calculations were performed on both the raw (filtered) read counts and the rarefied read counts.

2.2.7 Beta diversity analysis

Bray-Curtis dissimilarity between microbial communities was calculated using SciPy (v0.19.1) in Python (v3.6.0) [34]. The resulting pair-wise dissimilarity values were used as input for principal coordinates analysis to visualize the similarity of all samples to one-another. The

dissimilarity values were also used as input for hierarchical clustering using the average linkage method to identify clusters of communities across basins and water masses.

2.2.8 16S rRNA Gene enrichment analysis using Fisher's exact test

Fisher's exact test is a classical statistical test that is commonly used for gene enrichment analysis [81]. Gene enrichment analysis is generally used to test if a set of genes is enriched in one condition vs another, where the set of genes is often a group of related genes based on some ontological assignment. In this case, there is only one gene of interest, the 16S rRNA gene, which is clustered into sets (OTUs) based on sequence similarity. The null hypothesis for each OTU is that the proportion of reads assigned to that OTU is not different among the environments of interest. Fisher's exact test was applied in a one-vs-all manner; i.e., for each sampling location (basin and depth category), if the proportion of that OTU in that location vs that OTU in all other basins compared to the proportion of all other OTUs in that basin vs all other OTUs in all other basins. Before the test was applied, a pseudo count (+1 read) was added for consistency with DESeq2 (below). Based on empirical evidence, the pseudo count did not change the determination that an OTU was enriched or not, but, in the case of low read counts, the odds ratio decreased by orders of magnitude.

2.3 Results and Discussion

The primary goal of this study was to identify important features of marine environments that shape microbial communities, with particular attention to oxygen minimum zones.

2.3.1 Site descriptions

In total, 142 samples were obtained from six marine basins: Western Atlantic, Eastern Atlantic (off-shore Angola), Central Mediterranean (off-shore Libya), Eastern Mediterranean (off-shore Egypt), the Great Australian Bight (GAB), and the Caspian Sea. There were important distinctions between the sampled environments. A thermocline was observed

near 50 m depth in all basins, however temperatures in the GAB fluctuated more than other basins until stabilizing near 1000 m. The deep waters of the Caspian Sea, Eastern Atlantic, and GAB were the coldest, rapidly dropping below 10°C and approaching 4°C in the deepest waters. The waters of the Mediterranean (off-shore Libya and off-shore Egypt) were the warmest, with waters never dropping much below ~15°C.

High cell counts are typically observed in shallow waters and then decrease with depth [313]. AODCs were generally 10-fold greater in shallow water samples (1×10^6 cells/mL) compared to deep water samples (1×10^4 to 1×10^5 cells/mL) from the same basin, which is typical for open ocean waters. In the Eastern Atlantic, AODCs were as much as 1000-fold greater in shallow waters compared to deep waters.

Salinity in each of the basins was generally constant throughout the water column. The waters of the Mediterranean were the most saline (nearly 40 PSU), while waters of the Eastern Atlantic and GAB were slightly less saline (~37 PSU). Salinity in the Caspian Sea, which is known to be brackish, was much lower; salinity was ~1/3 of that observed in other basins (~12 PSU).

Dissolved oxygen was highest in the shallow waters, as expected, and generally declined with depth. Oxygen minimum zones were observed in both the Caspian Sea and Eastern Atlantic, and oxygen concentrations declined steadily until ~500 m. Oxygen concentrations in both of these OMZs fell below 3 mg/L (the upper limit of hypoxia). In the Eastern Atlantic, deep water currents flow northward from the Antarctic, bringing both nutrients and oxygen to the deep waters here. The Caspian Sea, on the other hand, being landlocked, receives no input of oxygen to the deep waters. Oxygen concentrations in the GAB and in Egypt declined steadily with depth, but never dropped below 5 mg/L. In the deep waters of the Central Mediterranean Sea, oxygen fell below 5 mg/L in the deepest waters, but no mid-depth OMZ was observed. The primary cause of OMZ formation is eutrophication, resulting in increased aerobic activity in shallow waters. Therefore, the waters of the Mediterranean Sea and the GAB are unlikely to be substantially impacted by eutrophication.

Total Nitrogen was different among the basins. The highest values were observed in the Caspian Sea (~0.4 mg/L) and in the deep waters (>800 m) of the GAB (>0.4 mg/L). In the GAB, these high values were attributable to very high nitrate concentrations (30

micromolar). Nitrate levels in the Caspian Sea were also relatively high (~10 micromolar), but were much closer to values observed in other basins (~8 micromolar). In the Caspian Sea, high total N values were attributable to high nitrite (0.1 mg/L). Nitrite concentrations in all other samples were below detectable limits. In this shallow waters of the GAB, however, Total N was much lower (<0.2 mg/L), consistent with values observed in other basins. Total N increased slightly with depth in the Eastern Mediterranean, but never increased above 0.2 mg/L. Finally, ammonia levels were typically low, with highest values (~10 micromolar) observed in two of the deep water samples from the Eastern Mediterranean.

Inorganic phosphate concentrations were closely correlated with total nitrogen. Namely, inorganic phosphate was highest in the deep waters of the GAB and in the Caspian Sea. However, inorganic phosphate was more more highly correlated with depth than total N. Low phosphate concentrations were observed in the Eastern Mediterranean, which, although unusual for a large marine basin, is consistent with previous reports [280, 157, 284]. Additionally, silicate concentrations were correlated with total N and Pi. Silicate is indicative of diatoms. Lastly, total organic carbon (TOC) was substantially higher in the Caspian Sea (mean ~7 micromolar) than in any of the other basins (mean ~1 micromolar). This typically indicates that biological productivity is high. However, oil production in the Caspian Sea, and oil content in the water has been documented to be high as well. Therefore, these TOC measurements might be detecting oil hydrocarbons rather than active and/or recently dead biomass.

2.3.2 Within-community diversity (alpha-diversity)

The most fundamental measure of richness in a community is simply the number of species or OTUs observed The median number of observed OTUs was highest in Egypt (6595) and GAB (6122) indicating that microbial biodiversity is highest in these environments. Median observed OTUs in Caspian (3843) and Eastern Atlantic (i.e., off-shore Angola; 4810) indicate intermediate biodiversity, while Libya (1804) and West Atlantic (1865) had the lowest biodiversity. The number of observed OTUs per sample in each basin is summarized in Table 2.2.

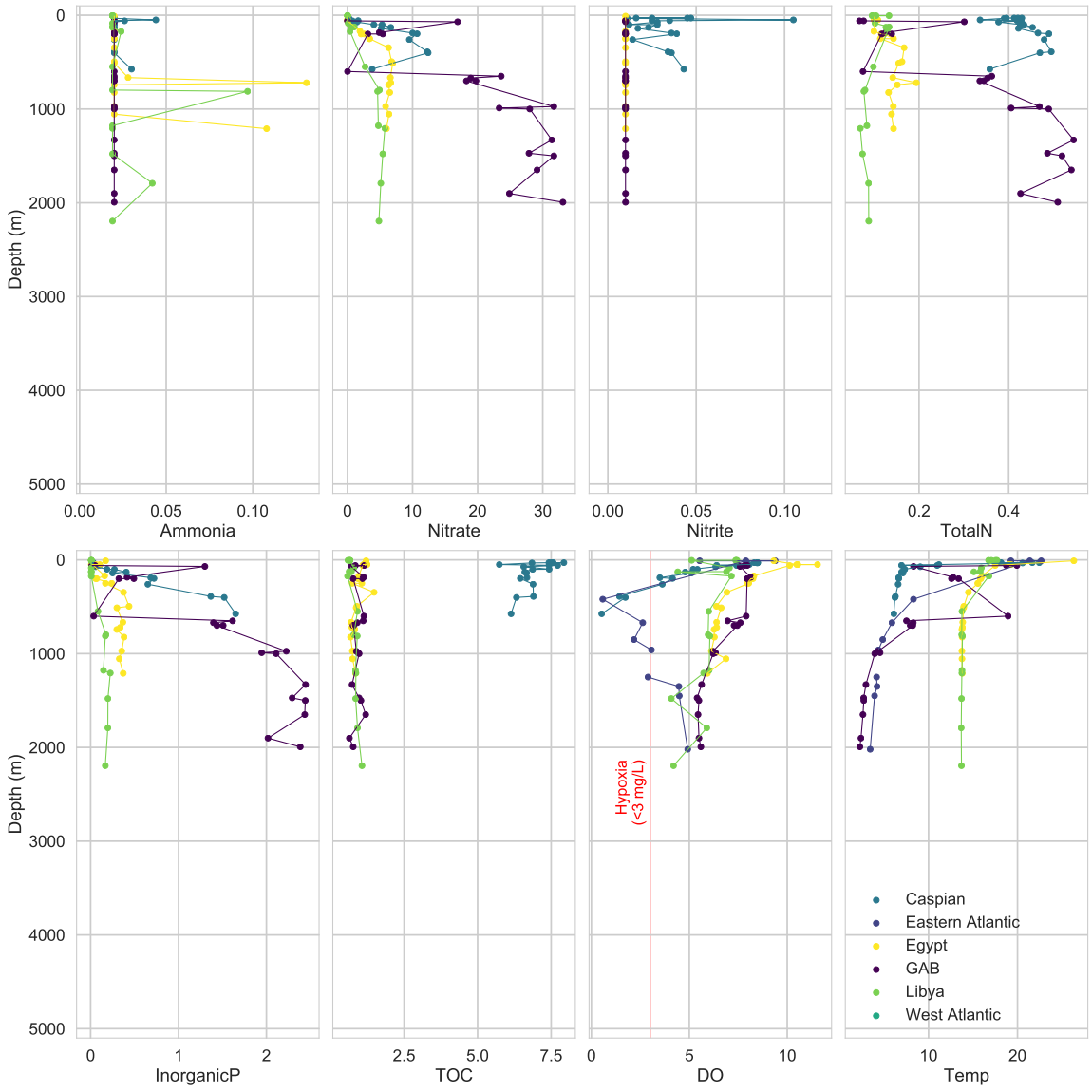


Figure 2.2: Physical variables for each basin.

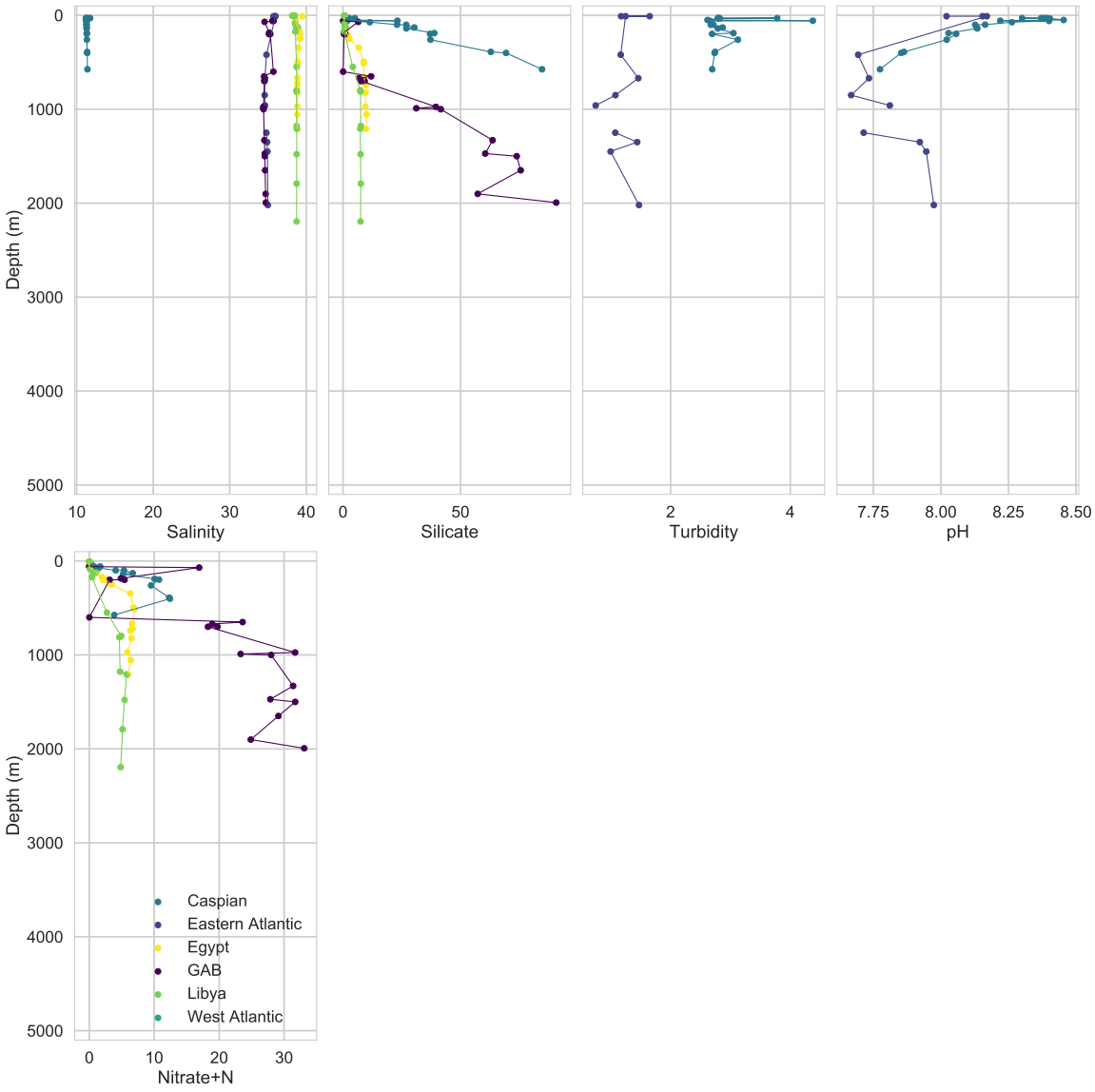


Figure 2.3: Chemical variables and nutrients for each basin.

Table 2.2: Range of OTUs observed in each basin. The distribution of OTUs observed per sample in basins marked with an asterisk * was not normal (Shapiro-Wilk test P-value <0.05).

Basin	min	median	max
Eastern Atlantic*	2339	4810.0	9486
West Atlantic*	586	1865.0	6853
Eastern Mediterranean (Egypt)	2779	6595.0	11440
Central Mediterranean (Libya)	1269	1804.0	2562
Caspian*	1662	3843.0	11460
GAB*	4071	6122.5	22771

There were significant differences between microbial communities from deep and shallow waters (Figure 2.4, Figure 2.5). Richness (observed and estimated) and alpha diversity metrics were similar across basins and depths. These metrics attempt to quantify the overall biodiversity of a community, and higher values correspond to higher biodiversity. Therefore, richness and alpha-diversity can be seen as high-level indicators of the 'health' of a community. Beta-diversity metrics indicate that microbial communities cluster according to basin and/or depth. Furthermore, 16S rRNA gene enrichment analysis revealed genera that are over-represented in each sampling location (basin, depth).

The number of OTUs observed in OMZ communities was also investigated to test the hypothesis that hypoxia significantly influences richness, e.g., by selecting against aerophilic bacteria. The number of OTUs observed in microbial communities within OMZs was not significantly different from microbial communities from higher oxygen environments.

The *chao*1 metric attempts to measure the actual richness, and this metric is discussed below [50]. Diversity attempts to consider not only the number of species, but the distribution of those species. There are many possible metrics for diversity, two of which are discussed below and used in this work. Sample sizes (number of reads) differed substantially between samples; therefore, samples were rarefied to 20,490 reads (the number of reads in the smallest sample) to in an attempt to make comparisons more comparable. The following results for alpha-diversity metrics were conducted using the rarefied samples.

The *chao*1 metric is an attempt to estimate the actual species richness S from the observed richness S_{obs}. The metric is based on the assumption that rare taxa are likely to be sampled infrequently, whereas highly abundant taxa are likely to be sampled frequently. Therefore, it is further assumed that the observed rare taxa contain information about undetected taxa, while the highly abundant taxa contain no useful information about undetected taxa. In particular, the observed rare taxa that are of interest are singletons (species observed once) and doubletons (species observed twice). If every observed taxa has at least two counts, then it can be assumed that the community has been completely sampled. The equation takes the following form:

$$\hat{S} = \begin{cases} S_{obs} + \frac{f_1^2}{2f_2} & if f_2 > 0 \\ S_{obs} + \frac{f_1(f_1-1)}{2} & if f_2 = 0 \end{cases}$$

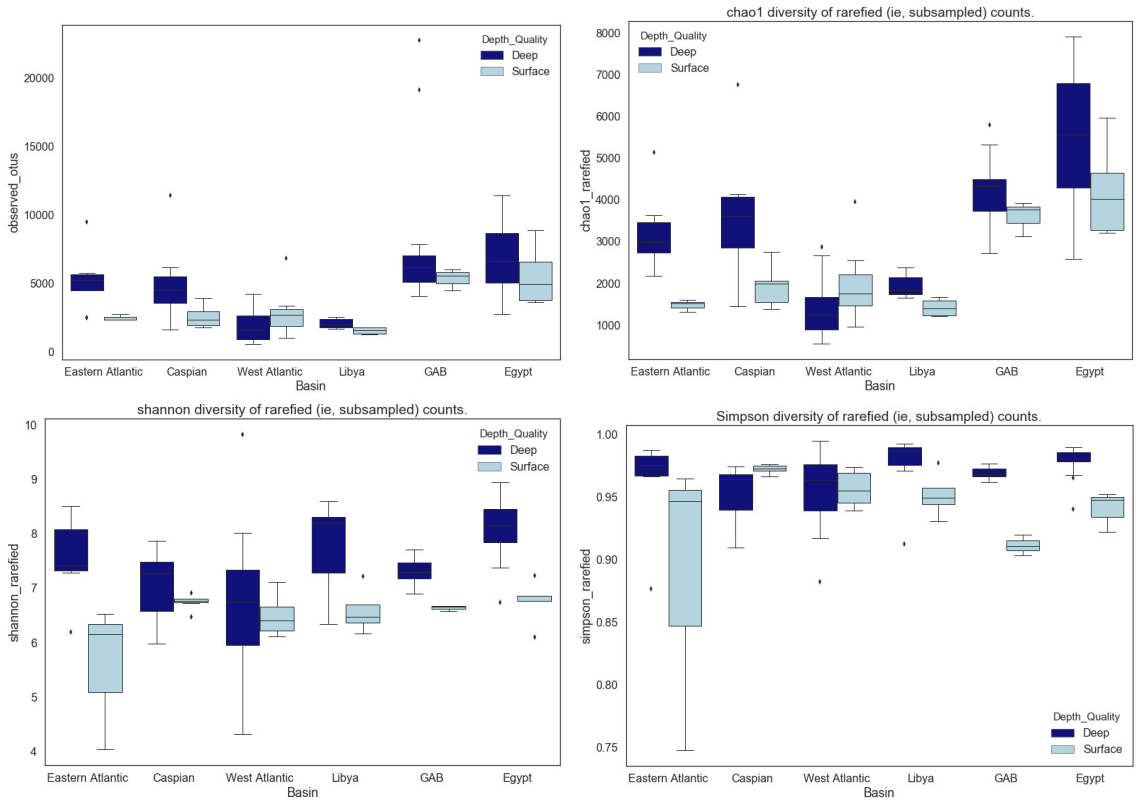


Figure 2.4: Alpha-diversity metrics of microbial communities from different basins and water masses (by depth category).

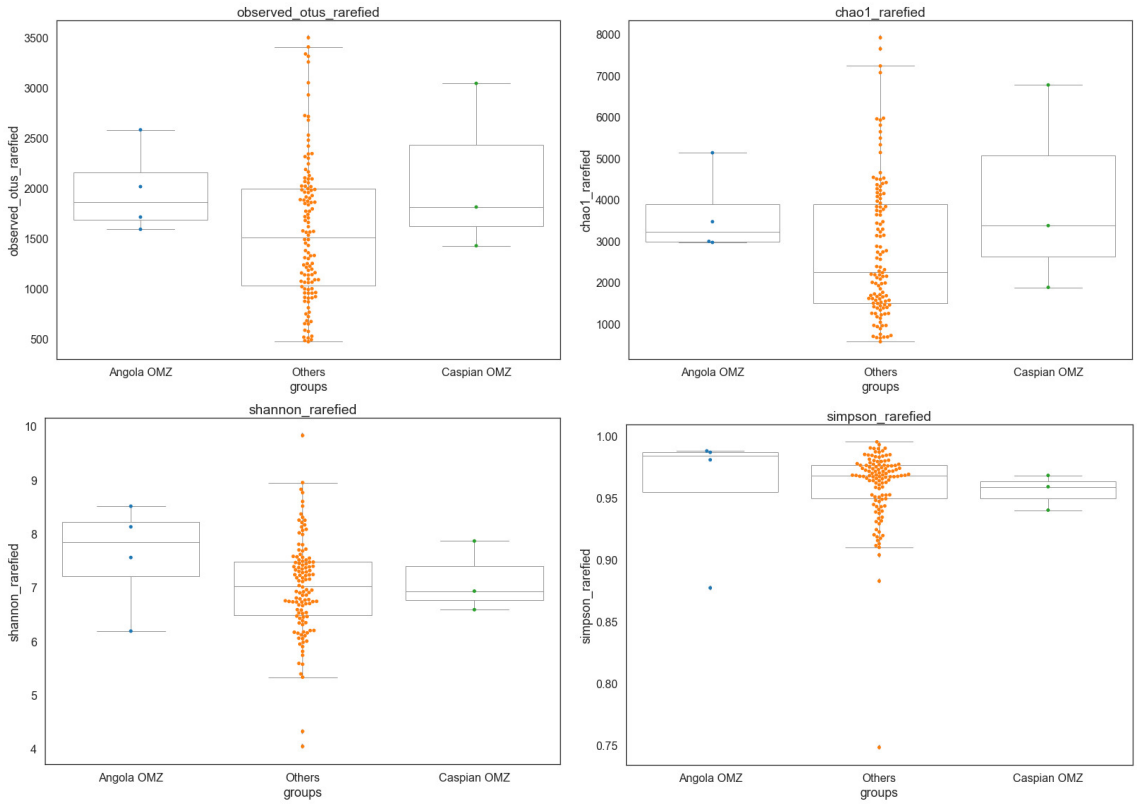


Figure 2.5: Alpha-diversity metrics of microbial communities from OMZs compared to all other communities.

Where S_{obs} is the observed number of species, f_1 is the number of singletons, and f_2 is the number of doubletons. The trends for richness estimation across basins were similar to the trends in observed OTUs, which is not surprising, considering that the equation for richness estimation takes as a parameter the number of observed OTUs. Richness estimation can be interpreted as a lower limit for the number of OTUs/species in an environment, and therefore informs estimates for future sampling (e.g., with respect to the amount of water or biomass required in order to reach some sequencing coverage goal). Based on the richness estimates, the observed OTUs represent only ~50% of the total number of actual number of OTUs in each basin. The observed OTUs for the Libya samples are nearest to the estimated richness (mean 76.94% +/- 0.06%). However, when sequencing coverage is very low, and very few singletons have been observed, chao1 is likely to give an estimate that is too low. The Libya samples had the lowest sequencing coverage of all of the samples, and therefore it is likely that sampling was too low in these basins to get good estimates of richness. Sequencing coverage for many of the Western Atlantic samples was low, while coverage for a few of these samples was quite high, and this variation carries over to the chao1 richness estimates. Overall, the chao1 richness estimates indicate that perhaps half of the microbial community members were not observed in our sequencing efforts. This fact, combined with the likelihood that sequencing coverage for some samples was perhaps too low to even produce good richness estimates means that future sampling efforts in all of these basins should be expanded substantially.

Actual richness using the *chao*1 estimator in OMZ communities was also investigated to test the hypothesis that hypoxia significantly influences potential richness. The estimated richness of microbial communities within OMZs was not significantly different from microbial communities from higher oxygen environments.

The traditional Simpson index is defined as:

$$\lambda = \sum_{i=1}^{R} p_i^2,$$

Where R is the total number of OTUs observed (i.e., richness) and p_i is the proportional abundance of the i^{th} OTU.

Table 2.3: *Chao*1 richness estimates for each basin. The number of observed OTUs per sample in basins marked with an asterisk * was not drawn from a normal distribution (Shapiro-Wilk test p-value <0.05).

Basin	min	median	max
West Atlantic*	558.1	1440.8	3956.4
Eastern Atlantic	1323.2	2843.7	5126.8
Libya	1207.6	1712.3	2376.3
Egypt	2581.5	5132.7	7909.1
Caspian*	1387.2	2849.9	6766.3
GAB	2723.5	4189.1	5794.0

The value of the Simpson index equals the probability that two microbes sampled from the same community will be the same species, assuming random sampling with replacement. In its original form, the Simpson index is *small* when diversity is *high* and vice versa; i.e., when fewer species are present in the community, the relative abundance of each species is higher. The Gini-Simpson index is the complement of the Simpson index.

$$1 - \lambda$$

The interpretation of the Gini-Simpson index is often considered to be more intuitive, and therefore it is commonly used in ecology. The value of the Gini-Simpson index is *large* when diversity is *high*; if there is only one species in the data, then its value is 0. The value of the Gini-Simpson index equals the probability that two microbes sampled from the same community will be of different species, and is bounded between 0 and 1. Gini-Simpson diversity was generally similar across basins in deep water communities. Diversity varied more across basins in the shallow water communities, and tended to be lower than diversity in deep water communities of the same basin. The microbial communities in the shallow waters of the Caspian Sea were the most diverse shallow water communities. One community in the Eastern Atlantic had a very low diversity (despite similar sequencing coverage to other shallow water communities). This indicates that this sample is likely to be dominated by one (or a few) OTUs.

The calculated Gini-Simpson diversity of OMZ communities was also investigated to test the hypothesis that hypoxia significantly influences biodiversity, e.g., by selecting for microaerophilic and/or facultative-anaerobic bacteria. The diversity of microbial communities within OMZs was not significantly different from microbial communities from higher oxygen environments as determined by ANOVA.

The value of the Shannon entropy represents the uncertainty of predicting the species of a microbe randomly sampled from a community. It is bounded between 0 and `ln(R)`. If all species are equal in relative abundance, then the value is `ln(R)`. The value approaches 0 when one (or a few) species dominates the community and all other species are present in low relative abundance. Trends across basins and depths were similar when comparing

Shannon entropy (diversity) and Gini-Simpson diversity (above). I.E., Shannon diversity is generally higher in the deep water communities and lower in the shallow water communities. However, the microbial communities of the shallow waters of the Caspian Sea were less diverse compared to the communities from the deep waters. The sample from the Eastern Atlantic that had a very low Gini-Simpson diversity still had a low Shannon diversity, but it was more similar to the other communities overall.

The Shannon diversity index (aka, Shannon entropy) is defined as:

$$H' = -\sum_{i=1}^{R} p_i \ln p_i$$

where R is the total number of OTUs observed (i.e., richness) and p_i is the proportional abundance of the i^{th} OTU. The Shannon index quantifies the uncertainty in predicting the next OTU that will be observed when drawing OTUs from the sample randomly. When one or a few OTUs dominate the community, it is more likely that an OTU draw randomly will be accurately predicted; this is represented by a lower Shannon entropy. In an evenly distributed community, the different OTUs are roughly equal in relative abundance; this is represented by a higher Shannon entropy. The Shannon diversity of OMZ communities was also investigated to test the hypothesis that hypoxia significantly influences biodiversity, e.g., when microbes with a selective advantage become enriched in a community. The diversity of microbial communities within OMZs was not significantly different from microbial communities from higher oxygen environments.

Previous studies with similar data have observed statistically significant differences in alpha-diversity for microbial communities from different water masses within the same basin, however, none were observed in this study [11, 9, 280]. This could be indicate that depth categories used in this study were too broad, and therefore do not accurately represent different microbial communities. Alternatively, significant differences observed in previous studies might be due to small sample sizes.

2.3.3 Between-community diversity (beta-diversity)

Read counts for OTUs were converted to relative abundance and the Bray-Curtis dissimilarity was calculated between all pairs of samples. The resulting dissimilarity matrix was then used as input for principal coordinate analysis (PCoA) and hierarchical clustering (HC). Both of these methods take the same distance (or dissimilarity) matrix as input, and therefore yield similar results, but have slightly different purposes. PCoA is a form of dimensionality reduction that allows the user to visualize at a high level which communities are similar to each based on how close together they appear, but does not cluster communities together. Hierarchical clustering, as the name indicates, clusters the communities together by merging communities that are nearest (least dissimilar) to each other; these clusters of communities can then be further analyzed. The dissimilarity between clusters is represented as a dendrogram, which is sometimes difficult to interpret accurately because it represents dissimilarity between *clusters* of samples and not dissimilarity between *samples*. Furthermore, dividing the dendrogram into clusters is often based on a used defined threshold, making the resulting clusters somewhat subjective.

PCoA indicates that Caspian Sea microbial communities are distinct from communities in other marine basins. The other marine basins were generally similar to each other, with no clear distinctions by basin. However, is a general trend across basins that communities are separated by depth. There are some important correlations with microbial community depth that seem to be generalizable across marine basins. Shallow communities were generally correlated with higher salinity, higher temperature, higher dissolved oxygen, and higher total organic carbon. The Benguela current carries cold, nutrient-rich water up the coast, which contributes to these influences. This current meets the warmer, nutrient-poor waters of the Angola current, creating the Angola-Benguela Front, which continues north. Deep water communities were correlated with nitrate, total nitrogen, inorganic phosphate, and silicate (which is indicative of diatoms).

HC using agglomerative clustering with the average linkage method indicates similar trends as those observed with PCoA, but more clearly indicates the presence of three potential clusters of microbial communities. Microbial communities from the Caspian Sea

clustered together, distinct from other basins, with sub clusters for shallow and deep water communities. In addition, HC indicates that shallow water communities (not including the Caspian Sea) tend to cluster, with no clear sub-clusters by marine basin. Deep water communities (not including the Caspian Sea) formed a large cluster with sub-clusters for Western Atlantic, Eastern Atlantic, and Great Australian Bight communities. Microbial communities from the Mediterranean Sea (off shore Libya and off shore Egypt) also tended to cluster together, but a few communities from the Great Australian Bight also clustered here.

2.3.4 Characterization of microbial communities in oxygen minimum zones

Important differences exist between microbial communities from the photic and aphotic zones, but unexpected differences have also been found between deep water communities within the same basin [11, 9, 280]. A goal of this study was to identify key features of oxygen minimum zones, which typically form below the photic zone in eutrophic water masses .

Overall, microbial communities from oxygen minimum zones were not significantly different from other marine microbial communities. Richness and alpha-diversity metrics for microbial communities from OMZs were similar to those observed for other microbial communities. However, 16S rRNA gene enrichment analysis revealed genera that are significantly over-represented in OMZ microbial communities.

Beta-diversity showed that microbial communities from the Caspian Sea clustered together, and the OMZ communities did not show particular similarity to each other apart from other Caspian Sea communities. The Caspian Sea communities have likely been conditioned over a period of ~50 years to steadily increasing eutrophication and decreasing oxygen concentrations. It may be the case that a substantial fraction of the community is able to survive in a wide range of oxygen concentrations, and therefore oxygen concentration is a relatively minor influence on this community composition. However, the community from CMS01NB seems to be an outlier. This is likely because this community is (1) the deepest sample from the Caspian Sea, (2) in close proximity to a mud volcano, and (3) the

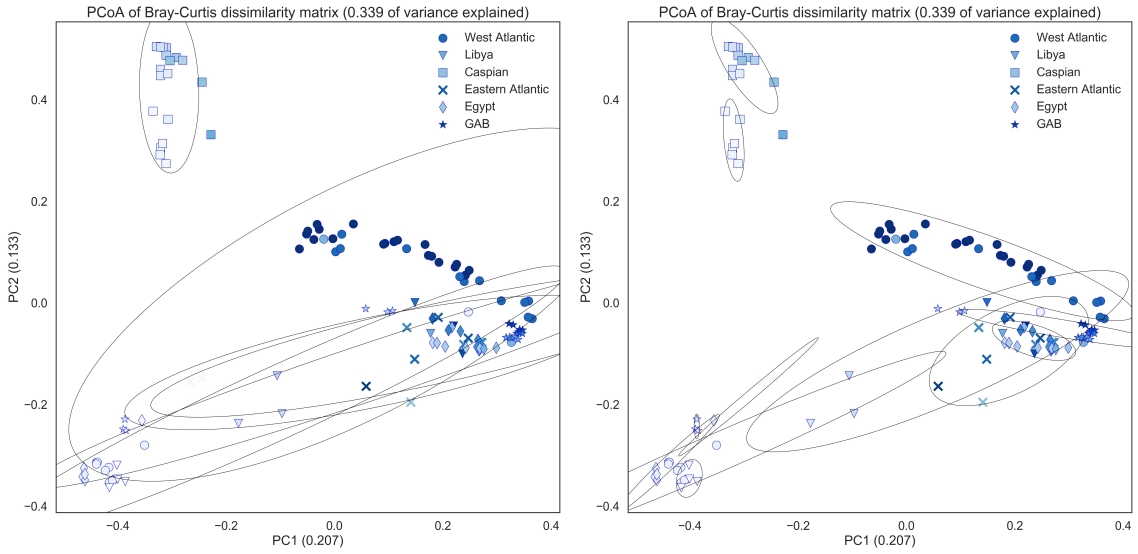


Figure 2.6: PCoA with confidence ellipses illustrates the dissimilarity between microbial communities. (A) Confidence ellipses drawn around microbial communities from each basin illustrate communities from most basins are generally similar to each other, with the exception of communities from the Caspian Sea, which are distinct from other communities as indicated be the lack of overlap between confidence ellipses. (B) Confidence ellipses drawn around microbial communities for each basin and depth location illustrate that some communities are distinct from each other by depth.

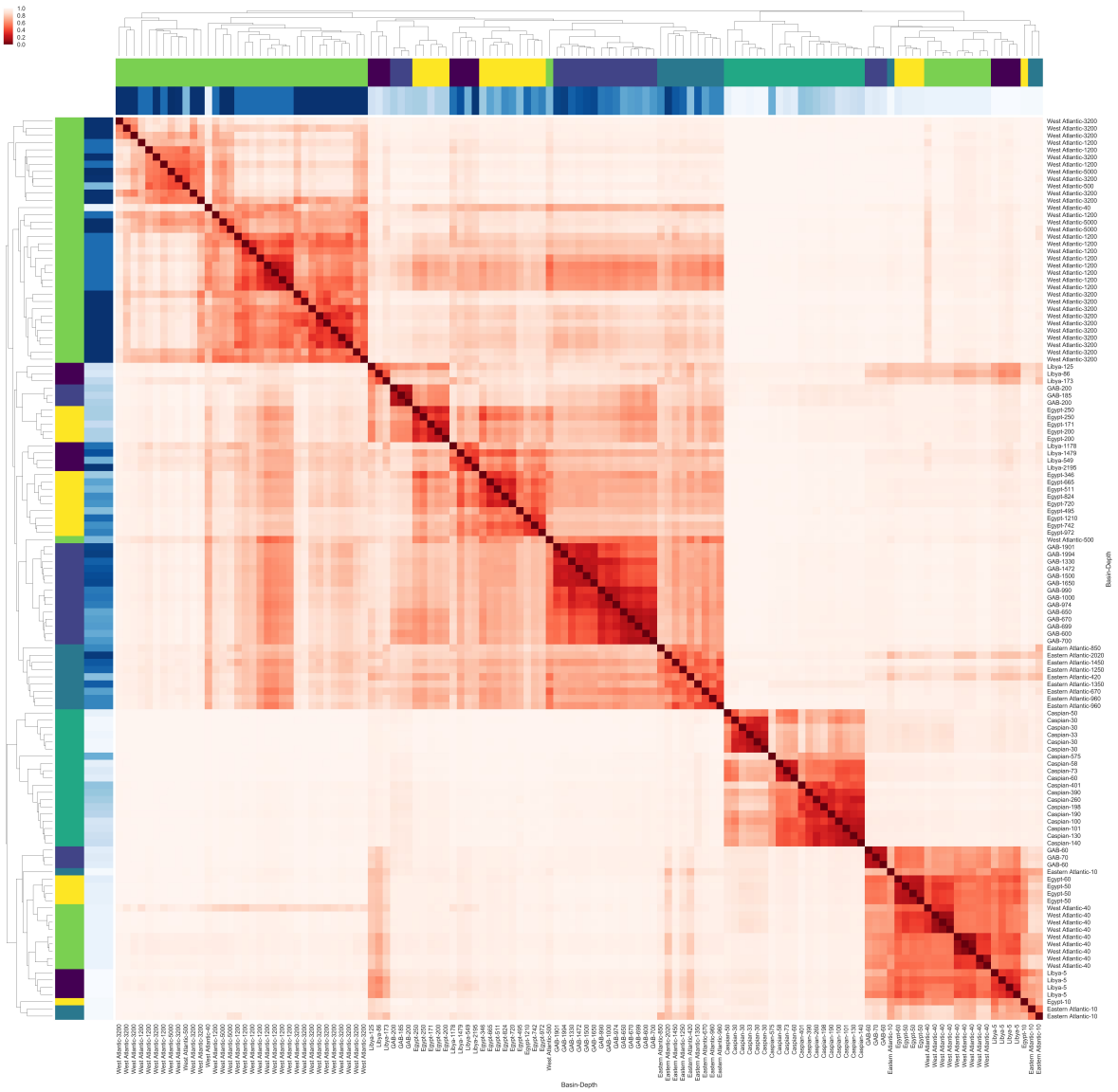


Figure 2.7: Heatmap and heirarchical clustering of Bray-Curtis dissimilarity matrix reveals clusters of microbial communities. Higher color intensity indicates similarity between samples. Dendrograms were constructed using agglomerative heirarchical clustering with the average linkage method. Color bars indicate either basin or depth (higher blue intensity indicates deep water sample). Caspian Sea microbial communities cluster together and are distinct from other communities. Shallow water and deep water communities tend to cluster together.

farthest location from known oil production activities. Future work might investigate which of these mechanisms is the most influential.

Genera that were enriched in the Caspian Sea OMZ were identified with Fisher's exact test. These enriched genera are found in specific clusters of all of the observed genera (across basins). These genera in these clusters tend to collectively exhibit one of the following trends: (1) enriched in the Caspian Sea in general, or, at least, in the deep waters of the Caspian Sea; (2) observed frequently in both the Caspian Sea and Eastern Atlantic in lower abundance, and otherwise rarely observed. One of the clusters was comprised of genera that were frequently observed across all samples, and there were two sub-clusters within this group. The genera of one of the sub-clusters were observed in similar abundance in both shallow and deep water samples, while genera in the other sub-cluster were observed in lower abundance in shallow waters and higher abundance in deep water samples.

Another cluster was formed from genera that were enriched in the Caspian Sea. This cluster was also divided into two sub-clusters. The genera of one of these sub-clusters showed a trend toward decreased abundance in deep waters in other marine basins, while the genera of the other sub-cluster were more consistently observed across samples. Two additional clusters were comprised of genera that were infrequently observed across samples. However, there are several genera within these clusters that seem to show a trend toward increased abundance in deep water samples. Future sampling may be required to determine if this phenomenon is real.

There were important differences between the microbial communities observed in the two OMZs. First of all, Archaea are enriched in the Caspian Sea, while few Archaea were enriched in the Eastern Atlantic. Furthermore, Archaea of the family *Nitrosopumilaceae* were among the five most abundant genera enriched in the Caspian Sea. On the other hand, Archaea that were enriched in the Eastern Atlantic comprised a small fraction of reads (<=123 reads). Sequence reads that were completely unidentified were also significantly enriched in the Caspian Sea OMZ communities. These could represent as yet uncharacterised taxa, or they could represent chimeric sequences that were not removed during quality filtering steps. Several *Gammaproteobacteria* including *Pseudoalteromonas*, *Limnobacter*, and *Psychrobacter* were enriched in the Eastern Atlantic OMZ communities.

There were some important similarities between the enriched genera in the two OMZs. The genera *Blastopirellula*, (*Thermoanaerobaculaceae*) Subgroup 10, (*Cellvibrionaceae*) and (*Cyclobacteriaceae*) were enriched across both OMZs. *Methyloprofundus* was also significantly enriched in both OMZs, but was observed in high abundance in only one of the Eastern Atlantic communities. The communities from the deepest water samples in the two OMZs were the most similar to each other out of all of the OMZ communities, supporting the hypothesis that depth is strongly influential, perhaps more influential than oxygen concentration.

2.3.5 Possible effects of oil spill on microbial communities

Oil degrading bacteria are ubiquitous in marine environments [118]. In the Caspian Sea and Eastern Atlantic Ocean, we observed typical oil degrading bacteria. In the event of a spill event in the Eastern Atlantic, we expect that these microbes will degrade oil as observed in other spill events. Importantly, this environment receives nutrient enrichment from the Angola-Benguela Front. This nutrient input will likely enhance degradation potential in this basin relative to other basins. However, high rates of in situ biodegradation similar to those observed in the Gulf of Mexico [118] may not be reached for days or weeks following a spill. During the course of biodegradation, a succession of bacteria similar to that observed during other spills is expected based on the existing community composition. We further expect that oil components will be degraded sequentially in an order similar to previous observations, i.e.: (1) volatile organics, (2) alkanes, (3) SVOC , and finally (4) polycyclic aromatic hydrocarbons [130].

Chapter 3

Differences in functional potential of microbial communities from eutrophic and oligotrophic environments

This chapter is derived from a manuscript in progress. Data and statistical analysis, and preparation and writing of the manuscript were carried out by Miller. Sampling was carried out by Techtmann, Fortney, Mahmoudi, Joyner, Fernandez, Gardinali, GaraJayeva, and Askerov. Guidance on experimental design and data analysis was provided by Jacobson. Guidance on manuscript preparation was provided by Hazen.

3.1 Introduction

Pollution in marine environments is increasing around the globe [105]. One important form of pollution is the addition of nutrients to the environment (called eutrophication) [191, 57, 236]. A major source of nutrient enrichment is via industrial and agricultural runoff, which affects virtually every marine ecosystem [105]. Eutrophication causes an increase in the abundance of microorganisms, chiefly algae, which is sometimes called a "bloom" [12]. When the algae die, their biomass is respired by bacteria in aerobic processes, decreasing oxygen saturation in the water, resulting in hypoxia or anoxia. These hypoxic or anoxic water masses may persist in deep waters due to differences in temperature (and thus density) between the shallow and deep waters, preventing mixing; these water masses are called oxygen minimum zones (OMZs) [216]. Furthermore, as dead microbial biomass sinks, the more easily degraded (labile) components are preferentially consumed [108, 116]. The result is that deeper waters are enriched in organic carbon that is more difficult to degrade (recalcitrant). Recalcitrant hydrocarbons from algal biomass include polyphenols, cellulosic fibers, and lignin-like components [304]. The combined effect is that the deep-water microbial communities are chronically exposed to hypoxia and high proportions of recalcitrant hydrocarbons.

Marine microbial communities are an important component of biogeochemical cycles such as the carbon and nitrogen cycles [119]. The persistent combination of hypoxia and exposure to recalcitrant hydrocarbons are important influences on microbial community composition and function [128]. Novel pathways for degradation of recalcitrant hydrocarbons have been discovered from isolates from these water masses, however, progress in this area is slow due to difficulties in isolating microbes from environmental samples [187, 187]. So far,

genomic studies indicate that these organisms are only distantly related [263, 44]. Deep ocean communities are still poorly characterized, despite early work in this area [132, 197]. Whole genome shotgun sequencing might help us to study these communities without the need to isolate microbes exhibiting these specific, novel functions.

Degradation of aromatic hydrocarbons is carried out predominantly by aerobic and anaerobic bacteria, aerobic fungi, and a few archaea [86]. However, it has been observed that environments polluted with aromatic hydrocarbons often become anoxic [44]. Because of the variance in reduction potential of alternative electron acceptors, different enzymatic pathways have developed in different organisms to degrade aromatic hydrocarbons under the various environmental conditions. The result is that there is a greater variety of anaerobic methods for aromatic hydrocarbon biodegradation compared to aerobic pathways.

The impact of the combination of hypoxia in OMZs and chronic exposure to recalcitrant hydrocarbons has not been studied. Understanding how different microbial communities respond to pollutants might help us to develop recovery strategies for the affected environments. Although some attempts have been made to develop models if marine petroleum biodegradation, the degradation capacity of marine microbial communities, especially in OMZs, is still uncertain [7, 235]. The aim of this work is to use shotgun metagenomics to investigate the potential of two microbial communities for oil hydrocarbon biodegradation: one eutrophic, hypoxic environment (the Caspian Sea) and one oligotrophic, oxic environment (the Eastern Mediterranean Sea).

The Caspian Sea has the highest concentration of mud volcanoes anywhere in the world [54]. Its southern basin has the highest concentration of putative petroleum-derived hydrocarbons [155]. Below the photic zone (>200 m), the Caspian Sea microbial communities are often dominated by *Proteobacteria* and *Thaumarchaeota* [116].

Natural hydrocarbon seeps can be found on the floor of the Eastern Mediterranean Sea [123, 183, 211]. Deep water microbial communities (>400 m) are typically dominated by *Proteobacteria*, *Actinobacteria*, *Chloroflexi*, and *Thaumarchaeota* [280].

3.2 Methods

3.2.1 Site Descriptions and Sample Collection

Water samples were obtained from the Southern basin of the Caspian Sea for shotgun metagenomic sequencing. Temperature, salinity, pH, oxygen concentrations, and turbidity were measured with a MIDAS CTD+ sensor array (Valeport Ltd, St. Peter's Quay, UK). Each of four sites was sampled at two depths (the first sample from above the thermocline and the second sample near the sea floor): site 1 at 50 m and 575 m; site 2 at 30 m and 190 m; site 3 at 30 m and 130 m; site 5 at 33 m and 390 m. The bottom waters of the Caspian Sea are persistently hypoxic (<2 mg/L of dissolved oxygen) [66]. Water samples were collected using a Stand Alone Particle Sampler (SAPS, Challenger Oceanic, UK) with controller, battery and pump upgrades (Oceanlab, University of Aberdeen, Scotland). Approximately 100 L of seawater was filtered at depth through a 292 mm diameter nylon filter with a pore size of 0.2 m. The filter was sectioned into thirds and one-third was used for DNA analysis.

Water samples were obtained from the Eastern basin of the Mediterranean Sea between Oct 11 and Oct 15, 2012 for shotgun metagenomic sequencing. Temperature, salinity, oxygen saturation, pH, and turbidity were measured as described above Each of three sites were sample at three depths (one sample above the thermocline, one sample near the sea floor, and one 'intermediate' depth sample): site 1 at 50 m, 250 m, and 1210 m; site 2 at 60 m, 200 m, and 1055 m; and site 3 at 10 m, 171 m, and 495 m. Water samples were collected as described above. Approximately 100 L of seawater was filtered at depth through a 292 mm diameter nylon filter with a pore size of 0.2 m The filter was sectioned into thirds and one-third was used for DNA analysis.

3.2.2 Physical and Chemical Measurements

The similarity of the sample sites to each other based on physical and chemical variables (temperature, salinity, oxygen saturation, pH, and turbidity) was investigated using Principal Components Analysis (PCA). The data was first standardized (centered to have mean of 0 and scaled to a standard deviation of 1) with the `sklearn.preprocessing.StandardScaler`

method. PCA was performed on the standardized data with the `sklearn.decomposition.PCA` method from the scikit-learn library (v0.19.0) in Python (v3.6.0) [217, 4]. The two components that explained the largest proportion of variance were then plotted to visualize the results. Confidence ellipses (95% confidence) were drawn around each sample location for visualization purposes.

3.2.3 DNA Extraction and Sequencing

Genomic DNA was extracted from filters (above) for sequencing using the method of Miller et al. [194] with modifications as described in Hazen et al. [117]. The quality of the DNA was determined by measuring 260/280 and 260/230 ratios on a NanoDrop spectrophotometer (Thermo Scientific, Waltham, MA). The DNA concentration was determined using PicoGreen (Life Technologies, Carlsbad CA). Sequencing was performed with a MiSeq (Illumina, San Diego, CA) as described by Caporaso et al. [43]. All forward and reverse reads were trimmed for quality and to remove Nextera adapter sequences with Trimmomatic [30]. Read quality was visualized with FastQC to verify trimming was effective [14]. The resulting `fastq` files were further processed as described below.

3.2.4 Automated pipeline for extraction of 16S reads and taxonomic assignment

The filtered sequencing reads were uploaded to MG-RAST [189] for analysis and are publicly available (accession numbers mgm4619055.3–62.3, mgm4619143.3–47.3, and mgm4620509.3–12.3). An overview of the MG-RAST pipeline can be seen in Figure 3.1. Following completion of analysis, annotations were downloaded via the MG-RAST application program interface (API). Taxonomy was assigned using the Greengenes database with an alignment length of 50 bp and an e-value of `1e-25`. Functional annotations were assigned using the SEED Subsystems ontology with an alignment length of 50 bp and e-value of `1e-25`.

0cm

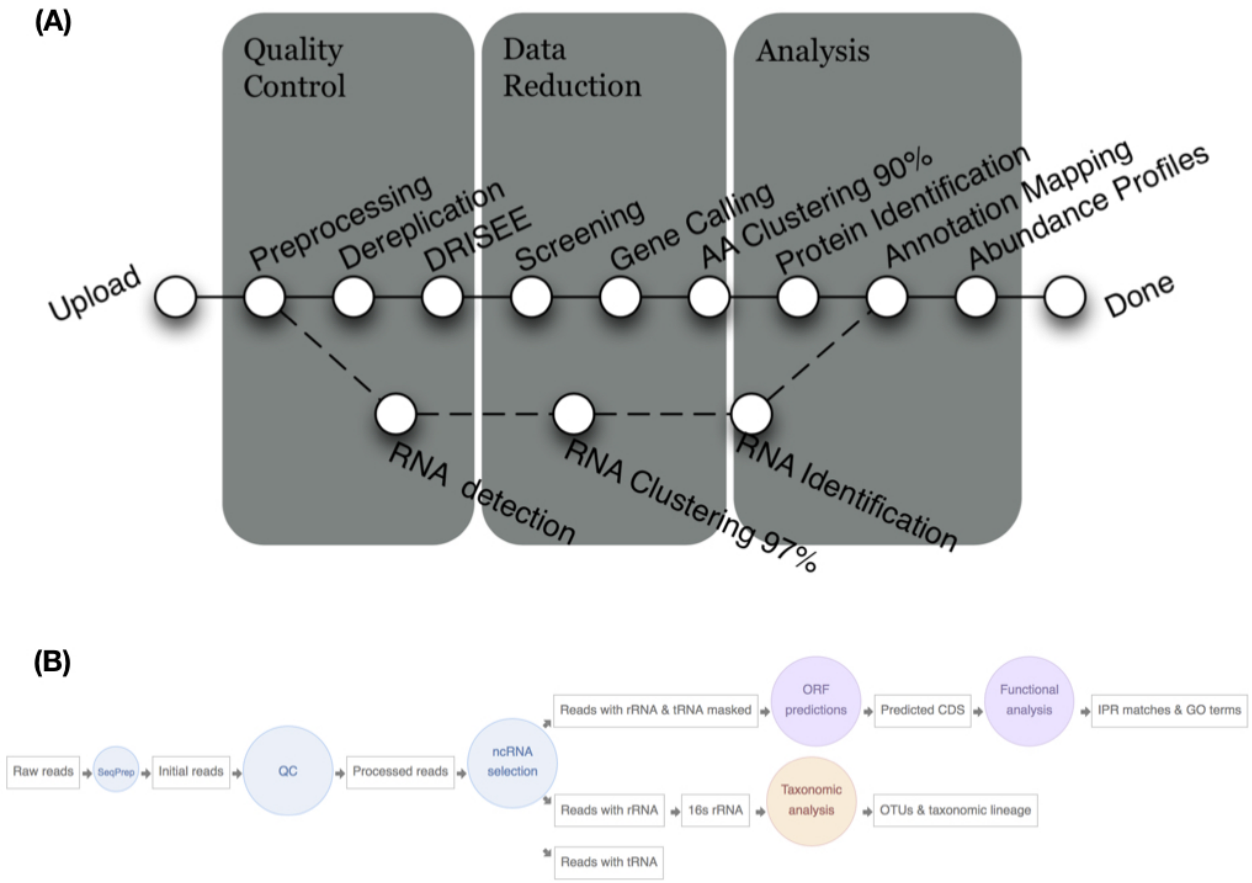


Figure 3.1: Overview of workflows in popular automated methods for analysis of shotgun sequences. (A) Metagenomic data uploaded to MG- RAST is analyzed and annotated with this pipeline. (B) Metagenomic data uploaded to the EBI Metagenomics (EMG) portal is analyzed and annotated with this pipeline. Figure adapted from Meyer [190], Meyer et al. [189], Mitchell et al. [195], and pip [3].

3.2.5 Identification of potential functional genes

Assemble sequences into contigs: SPAdes

Quality filtered and trimmed sequencing reads were assembled using SPAdes [22, 206]. The SPAdes software attempts to address some of the problems with other assembly algorithms including lack of uniform read coverage, variable insert sizes in paired-end reads, and chimeric reads that cover discontinuous regions of a genome. For Illumina reads, the SPAdes pipeline begins by correcting sequencing read errors with the BayesHammer software. Next, SPAdes assembles reads into contigs using multiple k-mer lengths; the values of k are selected automatically based on the sequencing read length. Although SPAdes includes the 'mismatch corrector' software, it is not recommended to run this on metagenomic sequencing reads, as reported here.

Predict protein coding genes and annotate: Prokka + Swiss-Prot, AromaDeg, CAZy

The Prokka software is useful for generating robust annotations of bacterial genes [253]. Prokka uses Prodigal to identify open reading frames and putative proteins [136]. Prodigal accurately recognizes translation start sites in microbial genomes and therefore achieves more accurate protein structure prediction and reduces false positives compared to similar software. Prodigal is able to do this by learning various 'properties' (e.g., ribosome binding site motif usage, start codon usage) from the sequence information. Additionally, Prodigal is able to identify genes across gaps and to predict partial genes. Following this step, Prokka attempts to annotate those proteins as follows:

1. Using BLAST+ `blastp` against a custom, user-supplied database [39].

    - The goal is to identify proteins that might be involved in biodegradation of aromatic hydrocarbons.
    - The custom database is compiled from AromaDeg and CAZy auxiliary enzymes [69, 41, 131].

2. Using BLAST+ `blastp` against a subset of proteins in UniProt (~16,000 proteins with "real protein or transcript evidence" that are not fragments) [15].

    - Swiss-Prot contains more than 500,000 non-redundant, manually curated protein sequences [283]. Functional annotations in Swiss-Prot are assigned to proteins based on experimental evidence; therefore, this database is well suited to the task of making inferences about the putative function of predicted proteins [222].

3. Using BLAST+ `blastp` against proteins from "finished" genomes in RefSeq [227].

4. Using HMMER against HMM profile databases (Pfam, TIGRFAMs) [73, 228, 104].

Finally, if no confident annotations are assigned (at an e-value threshold of `1e-6`) in the previous steps, the protein is labeled as a 'hypothetical protein'.

Map sequence reads to predicted proteins: Bowtie 2

The Bowtie 2 software was used to map sequence reads back to assembled and annotated contigs to estimate the abundance of each contig in the metagenome [165]. Bowtie 2 was chosen for alignment because it is typically more accurate than most other alignment methods.

3.2.6 Statistical analysis

Unless otherwise noted, all statistical analyses were performed R (v3.3.1) or in Python (v3.6.0) using either the SciPy (v0.19.1), scikit-learn (v0.19.0), or scikit-bio (v0.4.2) library [230, 4, 141, 217, 6].

Environmental variables

Physical and geochemical variables were standardized (centered on the mean and scaled by standard deviation) with the `sklearn.preprocessing.StandardScaler` method in scikit-learn.

Analysis of taxonomic diversity

Taxonomic assignments were made with MG-RAST against the RefSeq database [189, 227]. Taxonomic assignments were then downloaded from the MG-RAST server and used for the following analyses.

Rarefaction curves Rarefaction curves were constructed for each sample using the taxonomic assignments. Rarefaction curves are useful for comparing diversity across samples with different sample sizes [246, 134, 121]. Rather than attempting to encapsulate diversity into a single statistic, rarefaction curves represent the diversity of each sample as a line or curve at different hypothetical sampling depths. The curves produced from this analysis are also called species abundance curves because they illustrate the number of species that would be expected in a sample if a given number of individuals from the community were observed. The shape of the curve reflects the fact that at the outset of sampling, new species are observed at a nearly constant rate with increasing number of individuals [246, 95, 51]. As more individuals are observed, however, new species are observed at a continuously decreasing rate.

What is interesting about this form of diversity analysis is that different environments have characteristic curves [246]. The shape of the curve can be attributed to the interaction of two ecological principles: physical control vs biological accommodation of communities. A community that is primarily under physical control is one in which there is substantial selective pressure, e.g., in extremes of salinity or temperature. These communities are often dominated by a few species that comprise a large fraction of the community. A community that is 'biologically accommodated' is one in which environment variables are relatively mild or temperate. Examples include deep-sea marine waters; tropical, shallow marine waters; and tropical rain forests. Over time, the organisms in these communities become adapted to the local conditions. These environments tend to support a more diverse community.

These principles inform the interpretation of the rarefaction curves. A community with a shallow rarefaction curve and a few number of species is expected to be primarily under physical control. A community with a steep rarefaction curve and a high number of species is expected to be primarily biologically accommodated. Both of these principles always affect

microbial communities, and the rarefaction curves helps in the determination of which of these principles as more strongly affecting a particular community.

Richness and alpha-diversity Richness and alpha diversity metrics were calculated using the RefSeq-based taxonomy. Richness for each sample was reported as the number of observed OTUs and the estimated richness using the `chao1` metric. The read counts for each sample were rarefied to a number of reads equal to that of the smallest sample before performing further alpha diversity calculations. Alpha diversity for samples was calculated using both the Shannon and Gini-Simpson indices. All alpha diversity metrics were calculated using the `skbio.diversity.alpha_diversity` method from the scikit-bio library.

To test the hypothesis that richness and alpha diversity metrics were significantly different between water masses, ANOVA was used followed by Tukey's honest significant difference (HSD) test. ANOVA was calculated using the SciPy (v0.19.1) package, and Tukey's HSD was calculated using the Statsmodels (v0.8.0) package in Python. The resulting p-values were corrected for multiple comparisons using the Statsmodels package.

Analysis of abundance matrices

Before performing further analysis, the read counts for each gene (or OTU) were converted to a simple proportion by dividing by the total number of reads in that sample. Although simple proportions may not be appropriate for all statistical analyses, they yield accurate results when, e.g., comparing samples to each other using an appropriate similarity (or dissimilarity) metric [186].

Dissimilarity and distance metrics All samples were compared in a pairwise manner based on physical and geochemical variables, taxonomic annotations, and functional annotations. All distances were calculated using the `pdist` method from SciPy. Distance between samples based on physical and chemical parameters was calculated using Euclidean distance. Dissimilarity between samples based on taxonomic annotations and functional annotations was calculated using the Bray-Curtis dissimilarity metric [34].

To test the hypothesis that the environments and microbial communities were similar based on depth category, analysis of similarities (ANOSIM) and permutational analysis of variance (PERMANOVA) were performed on the matrices of Euclidean distances (for environmental data) or Bray-Curtis dissimilarities (for taxonomic and functional data) [55, 13]. ANOSIM was calculated with the `anosim` function, and PERMANOVA was calculated with the `permanova` function, both from the scikit-bio library. To test whether the dissimilarity (or distance) matrices for physical and chemical parameters, taxonomic annotations, and functional annotations, were similar to each other, the Mantel test was used [179]. The Mantel test is a permutation test, and therefore does not require that the values within the matrices follow any distribution. The Mantel test calculation was performed with the `mantel` function in the scikit-bio library using Spearman's rank correlations for each pair of distance matrices (e.g., environmental matrix vs taxonomic matrix, etc.). The Mantel test statistics were calculated for each pair of distance matrices with the `mantel` function in the scikit-bio library using Spearman's rank correlations.

Hierarchical clustering The distance matrices were used as input for hierarchical clustering of samples. Hierarchical clustering was performed using the `linkage` method in SciPy. To test the hypothesis that the identified clusters were not due to chance, ANOSIM and PERMANOVA were performed on the matrices in a manner similar to that described above. Resulting clusters and associated dendrograms were visualized with the `clustermap` method from the seaborn library (v0.8.0) in Python.

Gene enrichment and feature importance

Gene enrichment analysis with Fisher's exact test Identification of enriched genes was performed using Fisher's exact test, and the resulting p-values were corrected for multiple tests. Fisher's exact test was implemented using the `fisher_exact` method from SciPy, and p-values were corrected with the `multipletests` method from Statsmodels. The null hypothesis for each gene is that the proportion of reads assigned to that gene is not different among the sample locations. Fisher's exact test was applied in a one-vs-all manner; i.e., for each sampling location (basin and depth category), the proportion of each gene

across sample locations was compared to the proportion of all genes across sample locations. Before the test was applied, a pseudo count (+1 read) was added.

3.2.7 Functional annotation of sequence reads with mi-faser

Whenever possible, shotgun metagenomic reads should be assembled into larger contigs for gene prediction and annotation. However, when the sequencing depth is low and the microbial community is sufficiently complex, assembly becomes difficult [252]. In these cases, it may be practical to attempt to annotate reads directly; however, directly annotation of short reads by sequence similarity is likely to be inaccurate [315]. Furthermore, a substantial fraction of annotations that have been assigned using automated approaches are inaccurate [250]. The mi-faser web-server attempts to overcome some of these challenges (Figure 3.2; Zhu et al. [315]). This online tool uses the faser (functional annotation of sequencing reads) software to annotate reads against a curated database of verified, experimentally-characterized proteins, and achieves higher accuracy than BLAST or PSI-BLAST. The faser software uses protein (rather than nucleotide) alignments and penalizes low bp alignments. The "gold standard" (GS) database used by the online server (mi-faser) is compiled from Swiss-Prot and the Catalytic Site Atlas databases. Proteins from the Swiss-Prot database were only included if the Swiss-Prot evidence attribute is "1" (the best support) and the protein has been assigned a complete enzyme commission (EC) number (i.e., an EC number of "1.1.1.-" would be excluded). Proteins from the Catalytic Site Atlas were included if they had a "literature-based annotation". Only bacterial proteins that were found in both of these two sets of proteins were included in the final GS database. All shotgun sequencing reads were submitted to mi-faser for annotation by faser against the GS database [315]. The identical set of sequences that was uploaded to MG-RAST (above) was uploaded to mi-faser. These annotations are likely highly accurate, but represent a small fraction of the enzymatic functions that exist in these microbial communities.

0cm

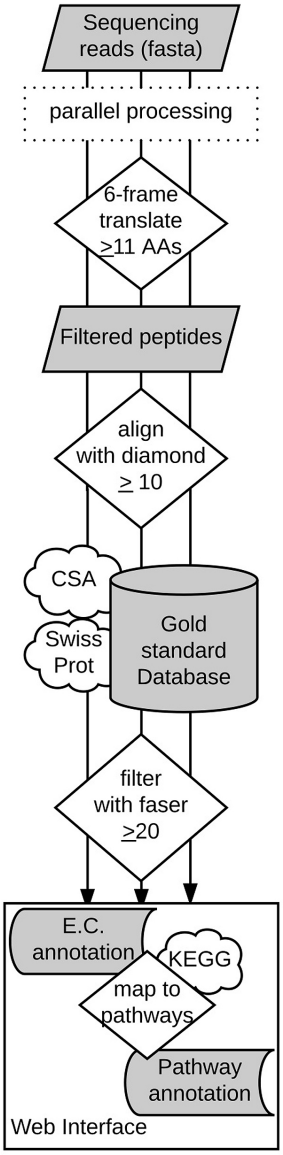


Figure 3.2: Pipeline used by the mi-faser web-server.

3.3 Results and Discussion

3.3.1 Site descriptions: Physical and geochemical features

Water samples were obtained from four sites in the Caspian Sea in the summer of 2013. At each site, water samples were taken at two depth categories: near surface (NS; 30-50 m) and near bottom (NB; 130-575 m) (Table 3.1). The mean temperature was 18.9°C in the shallow waters and 6.5°C of the deep waters. The mean salinity salinity was 11.3 PSU, which is approximately one-third of ocean salinity and consistent with the characterization of the Caspian Sea as brackish. Mean dissolved oxygen varied from 7.9 mg/L in the shallow waters to 2.3 mg/L in deep waters, which is below the threshold of 3.0 mg/L to be considered hypoxic. Turbidity was consistent through the water column at 2.8 FTU, while mean pH varied from 8.4 in the shallow waters to 8.0 in the deep waters.

Water samples from the Eastern Mediterranean Sea were obtained from three sites in the summer of 2013 (Table 3.1). At each site, water samples were taken at three depth categories: near surface (NS; 10-60 m), mid (MD; 171-250 m), and near bottom (NB; 495-1210 m). These depth categories are consistent with previous reports suggesting three distinct water masses within the Eastern Mediterranean [280]. The mean temperature was 21.0°C in the shallow waters and 13.8°C in the deep waters. The mean salinity was 39.0 PSU , which is typical of the Eastern Mediterranean, although slightly higher than the open ocean. Mean dissolved oxygen was 9.0 mg/L in the shallow waters and decreased to 7.4 mg/L and 5.5 mg/L in the mid and deep waters, respectively. Turbidity and pH were consistent throughout the water column at 1.3 FTU and 8.2 , respectively.

Profiles of environmental factors indicate important differences between the Eastern Mediterranean Sea and the Caspian Sea. The Eastern Mediterranean is warmer than the Caspian, especially the deep waters, and has higher salinity and higher oxygen concentration. The turbidity in the Caspian is higher than the Eastern Mediterranean, which is consistent with reports that the Caspian Sea is eutrophic (having high microbial productivity and thus biomass) and that the Eastern Mediterranean Sea is oligotrophic (having low microbial productivity). Euclidean distances between samples were calculated across environmental factors to quantify the dissimilarity between the water masses (i.e., sample site and depth).

Table 3.1: Sample locations.

Basin	ID	Date	Depth (m)	Depth Cat.	Label	Lat., Lon.
Caspian	1	2013-07-27	50	N. Surface	CS.1.50	39.75,50.48
Caspian	1	2013-07-27	575	N. Bottom	CS.1.575	39.75,50.48
Caspian	2	2013-07-31	30	N. Surface	CS.2.30	39.99,51.50
Caspian	2	2013-07-31	190	N. Bottom	CS.2.190	39.99,51.50
Caspian	3	2013-07-31	30	N. Surface	CS.3.30	40.04,51.35
Caspian	3	2013-07-31	130	N. Bottom	CS.3.130	40.04,51.35
Caspian	5	2013-08-01	33	N. Surface	CS.5.33	40.02,51.26
Caspian	5	2013-08-01	390	N. Bottom	CS.5.390	40.02,51.26
East Mediterranean	1	2012-10-13	50	N. Surface	EM.1.50	31.81,29.57
East Mediterranean	1	2012-10-13	250	Mid	EM.1.250	31.81,29.57
East Mediterranean	1	2012-10-12	1210	N. Bottom	EM.1.1210	31.81,29.57
East Mediterranean	2	2012-10-14	60	N. Surface	EM.2.60	31.96,29.89
East Mediterranean	2	2012-10-14	200	Mid	EM.2.200	31.96,29.89
East Mediterranean	2	2012-10-14	1055	N. Bottom	EM.2.1055	31.96,29.89
East Mediterranean	3	2012-10-11	10	N. Surface	EM.3.10	31.97,30.14
East Mediterranean	3	2012-10-11	171	Mid	EM.3.171	31.97,30.14
East Mediterranean	3	2012-10-11	495	N. Bottom	EM.3.495	31.97,30.14

Principal Components Analysis (PCA) and Hierarchical Clustering (HC) of Euclidean distances demonstrate that samples from the same water mass cluster together, which is consistent with previous reports [280]. Permutational analysis of variance (PERMANOVA) and analysis of similarities (ANOSIM) of clusters demonstrated that samples from the same water mass were significantly different from samples from the other water masses ($F = 481.7$, p-value $= 0.0001$, permutations $= 9999$; and $R = 0.94$, p-value $= 0.0001$, permutations $= 9999$; respectively).

3.3.2 Sequencing results

All sequencing reads were submitted to MG-RAST for analysis in their automated pipeline [189]. The number of sequences per sample from the Caspian Sea samples ranged from 4.8-13.0 million sequences and that of the Eastern Mediterranean Sea samples ranged from 3.7-13.7 million sequences (Table 3.2). Of those, the number of sequences that were identified as 16S rRNA from the Caspian Sea samples range from 3.3-9.9 thousand sequences and that of the Eastern Mediterranean Sea samples ranged from 3.1-21.6 thousand sequences. Taxonomy was assigned to these sequences using the Silva SSU database [229]. Additionally, taxonomy was assigned to all possible reads using the RefSeq database for comparison [227]. The remaining (non-16S rRNA) sequences were annotated against the SEED Subsystems (SS) and KEGG Orthology (KO) databases for functional analysis [213, 148].

3.3.3 Diversity, richness, and enriched taxa

Various methods can be used to determine abundance of taxa in shotgun metagenomic samples [35]. In this study, taxonomic classification was performed using MG-RAST web-server. Briefly, 16S reads were extracted from the shotgun data and taxonomy was assigned against the Silva database for comparison with 16S amplicon sequencing (Chapter 2).

Coverage of the 16S rRNA gene was highest in the shallow water communities from the Eastern Mediterranean Sea, but consistent across the other basins (Figure 3.3). This was different from the amplicon sequencing results (previous chapter) in which the highest number of 16S reads were observed in the deep water communities of the Western Atlantic

Table 3.2: Overview of shotgun metagenomic sequencing.

Label	Total BPs	Total Sequences	Mean Seq. Length (BPs)
CS.1.50	1,581,683,635	8,336,132	189.738
CS.1.575	2,652,775,718	13,039,064	203.448
CS.2.30	1,767,137,086	9,050,050	195.263
CS.2.190	1,028,419,886	5,899,500	174.323
CS.3.30	1,790,850,166	9,463,964	189.228
CS.3.130	1,603,815,100	8,581,962	186.882
CS.5.33	1,464,420,865	8,014,417	182.723
CS.5.390	1,105,999,940	4,817,229	229.593
EM.1.50	3,063,673,380	13,608,202	225.134
EM.1.250	1,627,487,981	7,805,517	208.505
EM.1.1210	1,341,997,638	6,421,584	208.982
EM.2.60	780,525,454	3,715,055	210.098
EM.2.200	2,095,631,983	9,874,033	212.237
EM.2.1055	1,019,230,954	5,774,224	176.514
EM.3.10	2,808,262,366	12,484,572	224.939
EM.3.171	3,018,743,363	13,696,233	220.407
EM.3.495	2,418,302,540	11,727,667	206.205

and in the deep water communities of the Great Australian Bight. It's possible that some of the microbial taxa from the shallow waters of the Eastern Mediterranean Sea were more amenable to DNA extraction, or simply that some of these samples were sequenced more deeply, perhaps as a result of loading more DNA at the time of sequencing. When compared to the rarefied 16S amplicon sequencing data, ~10x fewer OTUs were observed in each basin. This could be the result of PCR amplification picking up rare OTUs, or it could be the result of PCR amplification introducing spurious OTUs, despite removing chimeric reads in the pre-processing stages.

Diversity metrics were calculated for each of the water masses in the Caspian Sea and in the Eastern Mediterranean Sea in a manner consistent with Chapter 2 but using the 16S rRNA gene data from the shotgun data (Figure 3.4). The number of OTUs observed in microbial communities within the Caspian Sea deep water OMZ was not significantly different from microbial communities from other water masses. The actual richness (number of OTUs) in each sample was estimated using the chao1 estimator [50]. The trends for richness estimation across water masses were similar to the trends in observed OTUs, which is not surprising, considering that the equation for richness estimation takes as a parameter the number of observed OTUs. The most notable deviation from the estimated richness was higher than the observed richness for the microbial communities from the shallow waters of the Eastern Mediterranean Sea. In the Caspian Sea, both observed and estimated richness was lower in the shallow water communities and higher in the deep water communities, but the opposite was true of the Eastern Mediterranean Sea. The Gini-Simpson and Shannon alpha-diversity metrics yielded similar results. Alpha-diversity was lower in the shallow water communities of the two basins, which implies that a few OTUs are dominant in these communities. Alpha-diversity was highest in the "near bottom" communities of the Caspian Sea and in the "mid" depth communities of the Eastern Mediterranean Sea. This might indicate that these communities are relatively similar in structure. Importantly, even though these two communities were categorized differently ("near bottom" vs "mid" depth), these communities are actually at similar depths, because these sites in the Caspian Sea were relatively shallow. This might indicate that absolute depth from sea level is more important than proximity to sea floor through most of the water column, which is in contrast to

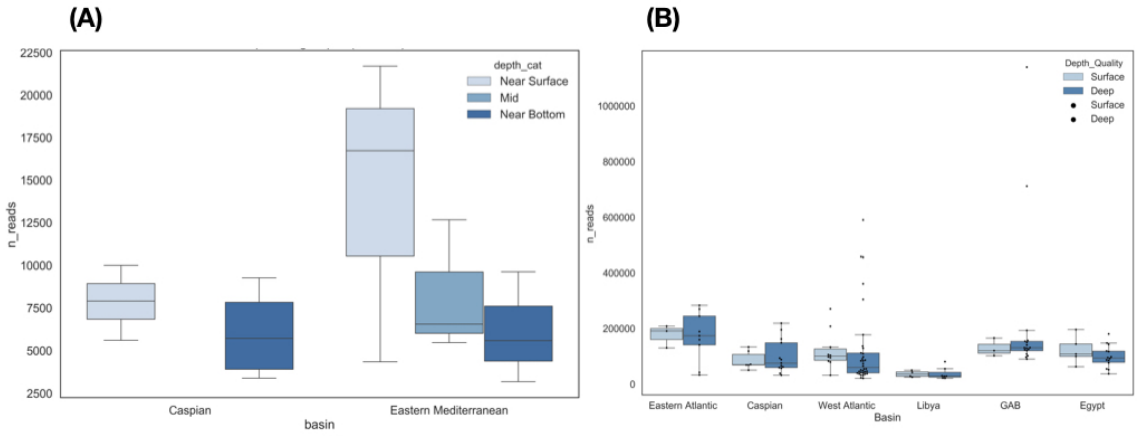


Figure 3.3: Comparison of sequencing depth (number of reads) obtained for the 16S rRNA gene from the amplicon and from the shotgun results. Sequencing depth was similar for most of the sampled communities, with the exception of the shallow water communities of the Eastern Mediterranean Sea, in which sequencing depth was two-fold higher. (A) Sequencing depth based on 16S sequences extracted from shotgun metagenome sequence. (B) Sequencing depth based on 16S rRNA amplicon sequencing (as reported in Chapter 2). PCR amplification may be biased toward particular sequences (i.e., those derived from specific taxa), and thus affect the diversity analysis of these data.

previous observations [280]. Alpha-diversity metrics in the deepest water ("near bottom") communities of the Eastern Mediterranean Sea were intermediate to the other water masses. In general, these results were consistent with the diversity metrics calculated from the 16S amplicon data, with the only exception being that the Gini-Simpson diversity for the shallow water communities of the Caspian Sea were much higher when calculated with the amplicon data (Figure 3.5)

3.3.4 Functional potential of microbial communities

All shotgun sequencing reads were submitted to mi-faser for annotation by faser against the GS database [315]. The identical set of sequences that was uploaded to MG-RAST (above) was uploaded to mi-faser. The mean number of total reads per sample was 7.5 (+/-2.1) million reads. The mean number of annotated reads per sample was 0.3 (+/-0.1) million reads, which comprised 4.2% (+/-0.7%) of total reads. The low number of annotations is likely the result of a combination of the limitation of the GS database, which is small, and that these samples are from ocean communities that might contain novel proteins. The results were comprised of 1185 (983+/-44 per sample) enzymatic functions (unique EC numbers). These annotations are likely highly accurate, but represent a small fraction of the enzymatic functions that exist in these microbial communities.

Of the resulting EC number annotations, thirty-six were chosen for further study due to their potential association with hypoxic or anoxic conditions. The number of enzymatic functions for each of the pathways that was included in the GS database varied substantially. This illustrates the disparity between the work that has been done to identify the intermediates of metabolic pathways and the work that has been done to experimentally characterize the enzymes associated with those pathways. Additional enzymatic functions were chosen for further study due to their potential association with petroleum hydrocarbon biodegradation [182, 315].

Fisher's exact test was used to test for enriched enzymatic functions in each water mass. There were no significantly enriched functions identified in any of the water masses. Spearman's rank correlations were then calculated between all enzymatic functions and environmental factors across water masses, and p-values were corrected for multiple tests.

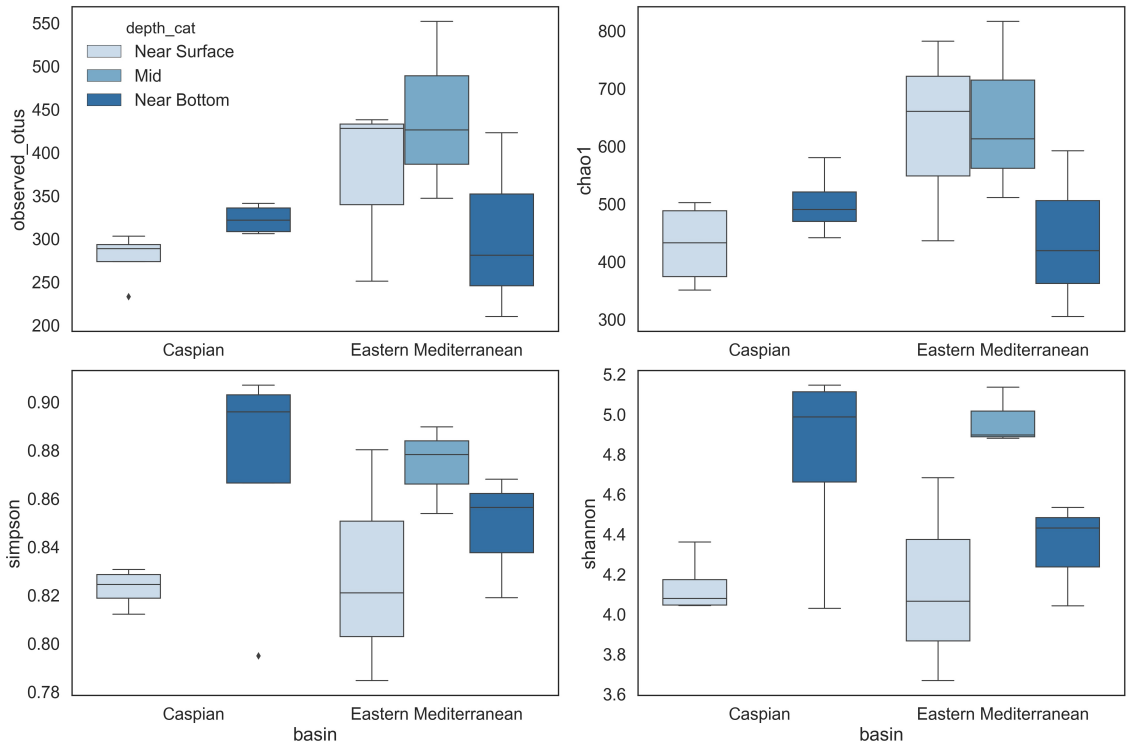


Figure 3.4: Richness and alpha-diversity of water masses in the Caspian Sea and Eastern Mediterranean Sea. Results are based on 16S rRNA gene sequences extracted from shotgun metagenomic data. Both observed richness (number of OTUs) and estimated richness (chao1) showed similar trends across water masses. The shallow water communities of the Caspian Sea tend to be slightly less rich than deep water communities; while the shallow water communities of the Eastern Mediterranean Sea show the opposite trend. Two alpha-diversity metrics (Simpson, Shannon) also showed similar trends across water masses. The shallow water communities of the Caspian Sea were much lower in diversity compared to the deep water communities; this trend was similar, but not as extreme, in the Eastern Mediterranean communities.

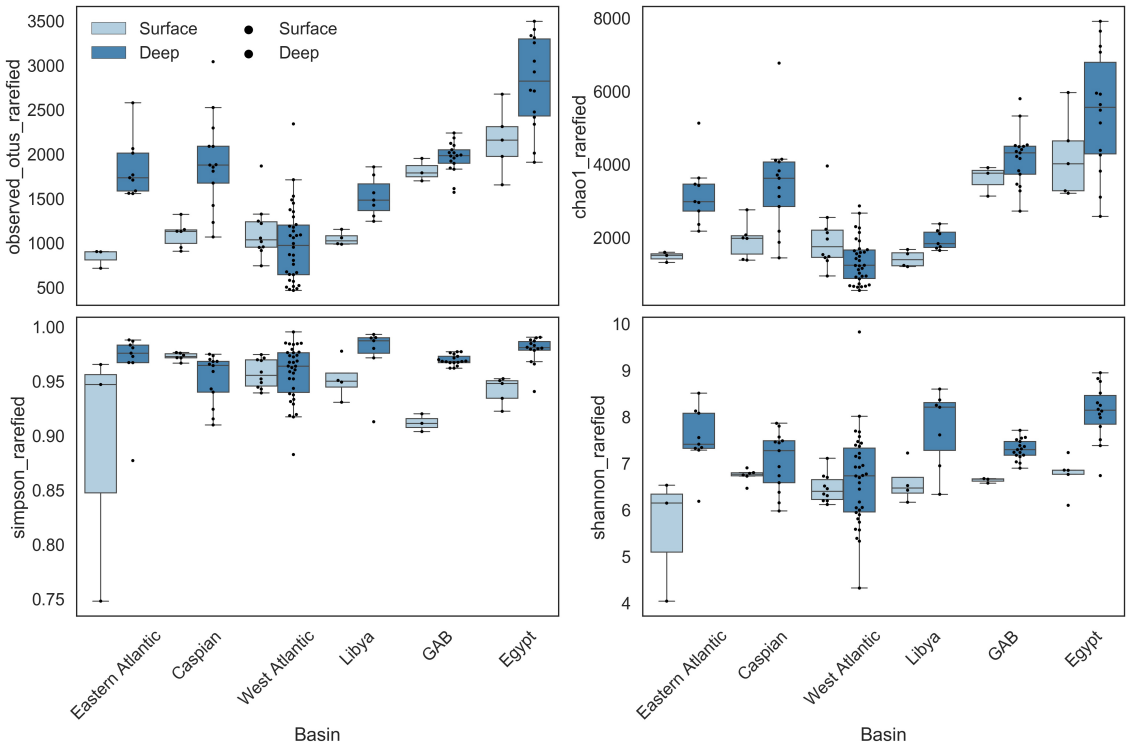


Figure 3.5: For comparison of diversity based on 16S amplicon sequencing (this figure) to diversity based on 16S annotations from shotgun sequences (above). Both observed richness (number of OTUs) and estimated richness (chao1) showed similar trends across water masses. The shallow water communities of the Caspian Sea tend to be slightly less rich than deep water communities; while the shallow water communities of the Eastern Mediterranean Sea show the opposite trend. Two alpha-diversity metrics (Simpson, Shannon) also showed similar trends across water masses. The shallow water communities of the Caspian Sea were much lower in diversity compared to the deep water communities; this trend was similar, but not as extreme, in the Eastern Mediterranean communities.

Enzymatic functional potential for petroleum hydrocarbon biodegradation

Benzoate degradation Benzoate (or benzoic acid) is composed of a benzene ring with a single carboxylic acid substituent (Figure 3.6). Benzoic acid biodegradation is categorized with xenobiotics biodegradation and metabolism in the BRITE hierarchy [147, 2]. Benzoate degradation in turn is composed of several sub-modules involving the degradation of benzoate, benzoyl-CoA, benzene, toluene, and catechol. Aerobic benzoate degradation involves three main steps: (1) activation of the aromatic ring, (2) de-aromatization (ring cleavage), and (3) degradation to central metabolites [65]. Four enzymes involved in benzoate degradation from the GS database were reportedly enriched in microbial communities following the DWH spill: 4-hydroxybenzoate 3-monooxygenase (EC:1.14.13.2), 3-hydroxybenzoate 6-monooxygenase (EC:1.14.13.24), benzoate 4-monooxygenase (EC:1.14.14.92), and acetaldehyde dehydrogenase (EC:1.2.1.10) [315, 182]. Of these, 3-hydroxybenzoate 6-monooxygenase (EC:1.14.13.24) is notable because it can utilize hydroxybenzoate substituted at multiple positions and is able to use either NADH or NADPH as a cofactor [97]. Acetaldehyde dehydrogenase (EC:1.2.1.10) is also notable because it is a component of the central metabolic pathway converting acetaldehyde to acetyl CoA which is can then be utilized in the tricarboxylic acid (TCA) cycle. Although this enzyme was reportedly enriched in metagenomic and metatranscriptomic studies, it may simply be the result of generally increased heterotrophic metabolic activity, and is likely not unique to petroleum exposure.

0cm

Aminobenzoate degradation Aminobenzoate is a benzoic acid molecule with an amino group substitution on the aromatic ring (Figure 3.7). Degradation of these compounds generally feed into the benzoate degradation pathway. One enzyme in this pathway has reportedly been enriched following petroleum hydrocarbon exposure: a subunit of (vanillate) 4-hydroxybenzoate decarboxylase. This enzyme converts 4-hydroxybenzoate to phenol and CO2 [289]. Phenol might then be converted to catechol and passed into the benzoate degradation pathway.

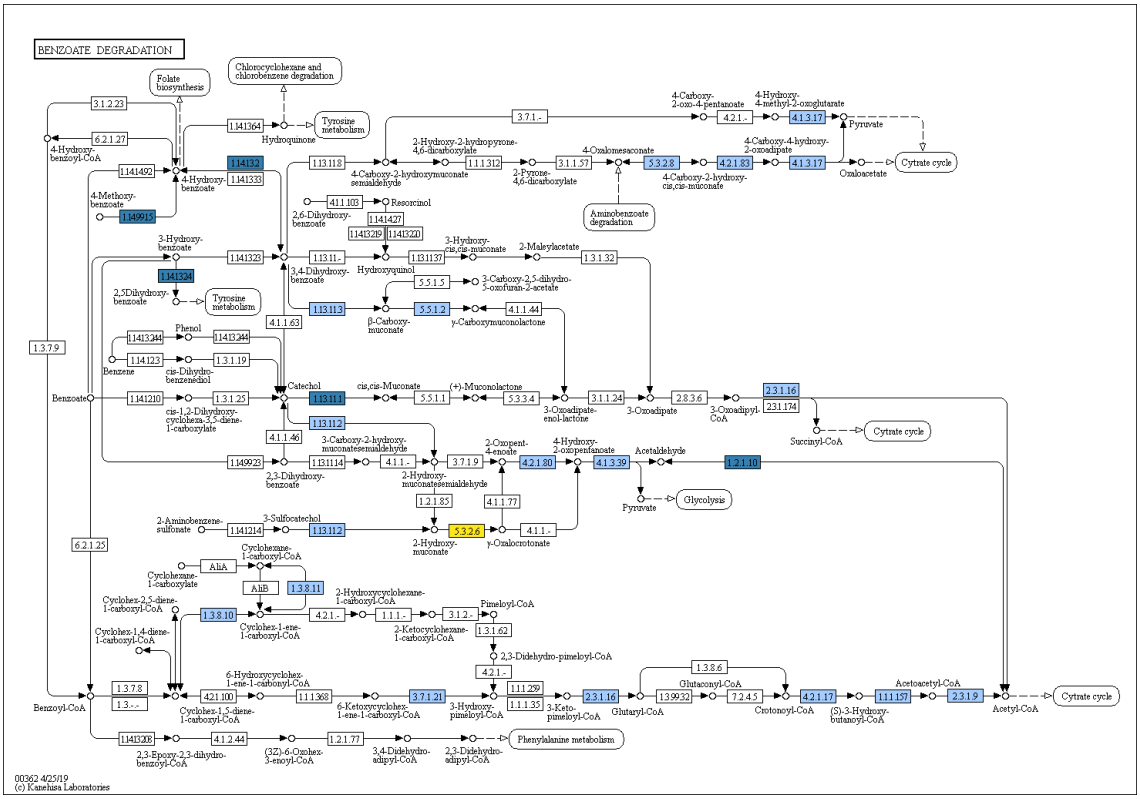


Figure 3.6: Benzoate degradation (KEGG map00362).

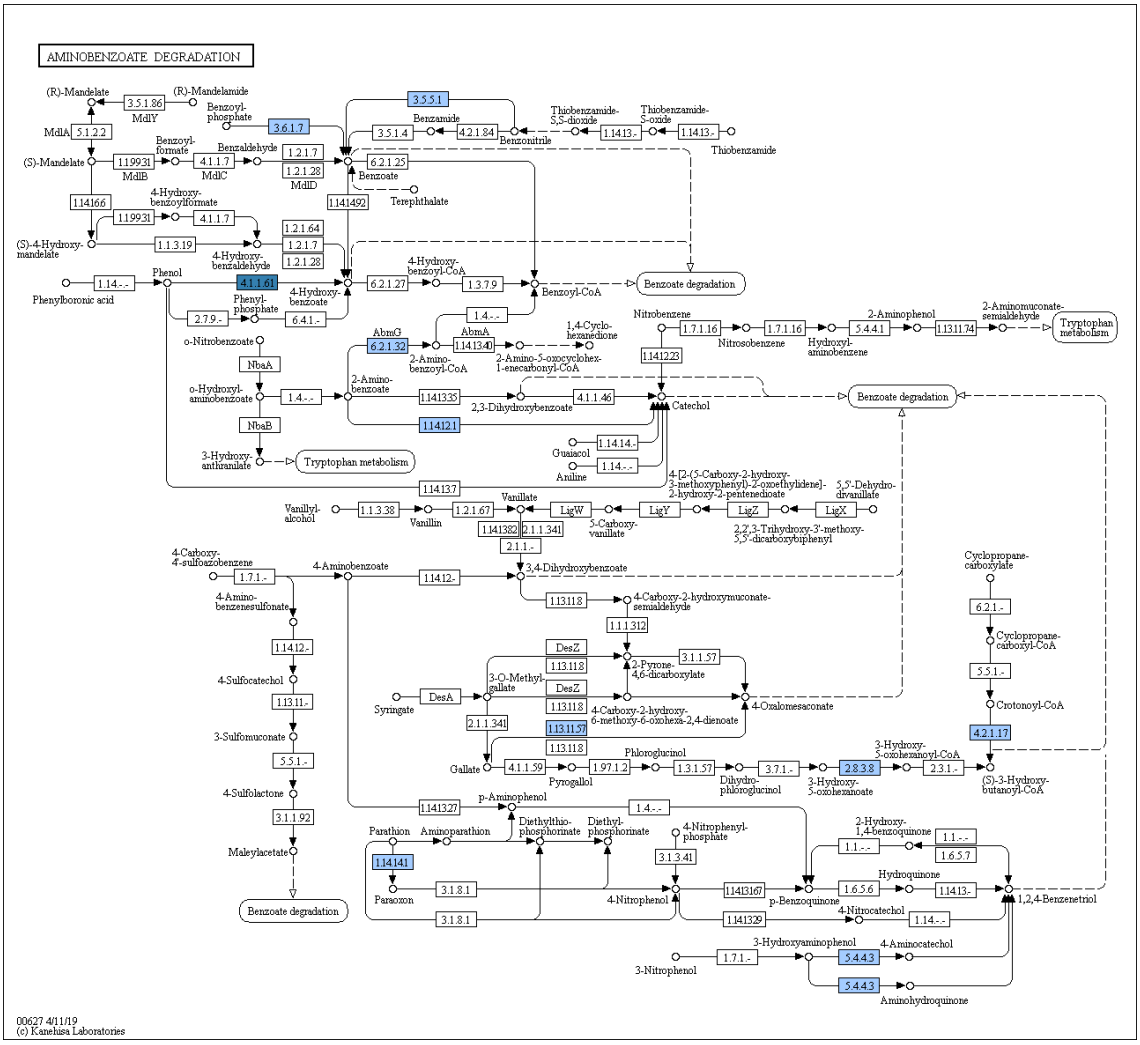


Figure 3.7: Aminobenzoate degradation (KEGG map00627).

Xylene degradation Xylene is an important component of petroleum and petroleum-derived products [67]. The chemical consists of a benzene ring with two methyl substituents (in any positions). The KEGG pathway for xylene degradation consists of several sub-modules including cymene degradation and cumate degradation as well as some overlap with benzoate and catechol degradation (Figure 3.8). However, only five of the enzymatic functions in the pathway are included in the GS database. Of these, four were found in the shotgun sequences. Degradation products feed into toluene degradation and the central metabolic pathways.

Styrene degradation Styrene is another aromatic hydrocarbon, consisting of a single aromatic ring and a single ethene substituent. Although the pathway is small, only four of the enzymes were in the GS database, and all of these were found in the shotgun sequences. Ethylbenzene degradation feeds into the styrene degradation pathway, and styrene degradation in turn feeds into the benzoate degradation pathway as well as the central metabolic pathways (Figure 3.9).

Phenylalanine, tyrosine, and tryptophan biosynthesis The aromatic amino acids (phenylalanine, tyrosine, and tryptophan) are synthesized in part through the shikimate pathway [125]. Aromatic amino acids are metabolized by all organisms (but not synthesized in animals), and therefore linked with metabolism of aromatic compounds derived from other sources. Some enzymes associated with metabolism of specific aromatics have associated with petroleum hydrocarbon exposure. This pathway has been well characterized, and approximately half of the steps in the pathway were found in the GS database (Figure 3.10).

Phenylalanine metabolism The phenylalanine module from KEGG includes one additional sub-module for trans-cinnamate degradation (Figure 3.11). Overall, the pathway was well-represented in the GS database, and all of these enzymes were present in the shotgun sequences. Of these, one enzyme, aldehyde dehydrogenase (NAD(P)+; EC:1.2.1.5) was enriched following the DWH spill. Similar to acetaldehyde dehydrogenase (EC:1.2.1.10), this enzyme is found in many pathways, and it may have simply appeared to be enriched following the DWH spill due to increased heterotrophic activity.

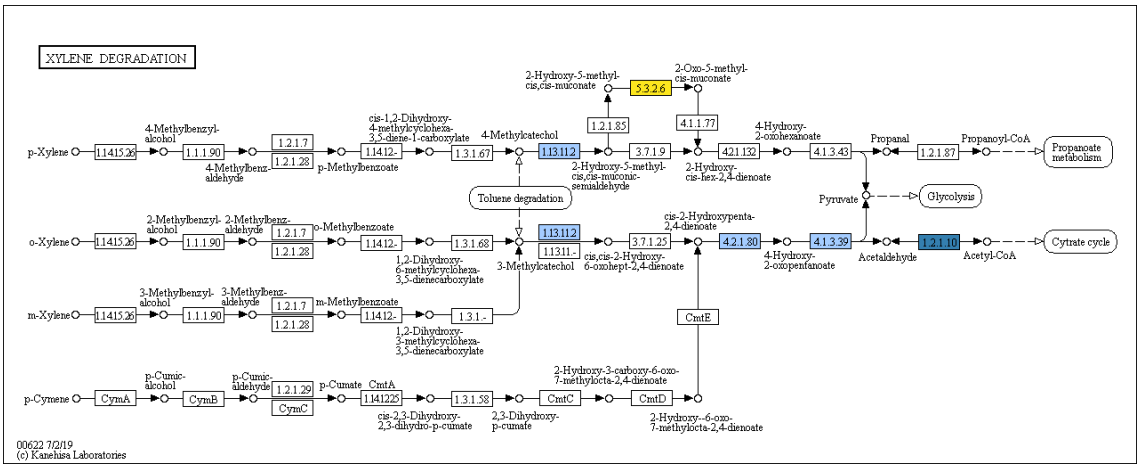


Figure 3.8: Xylene degradation (KEGG map00622).

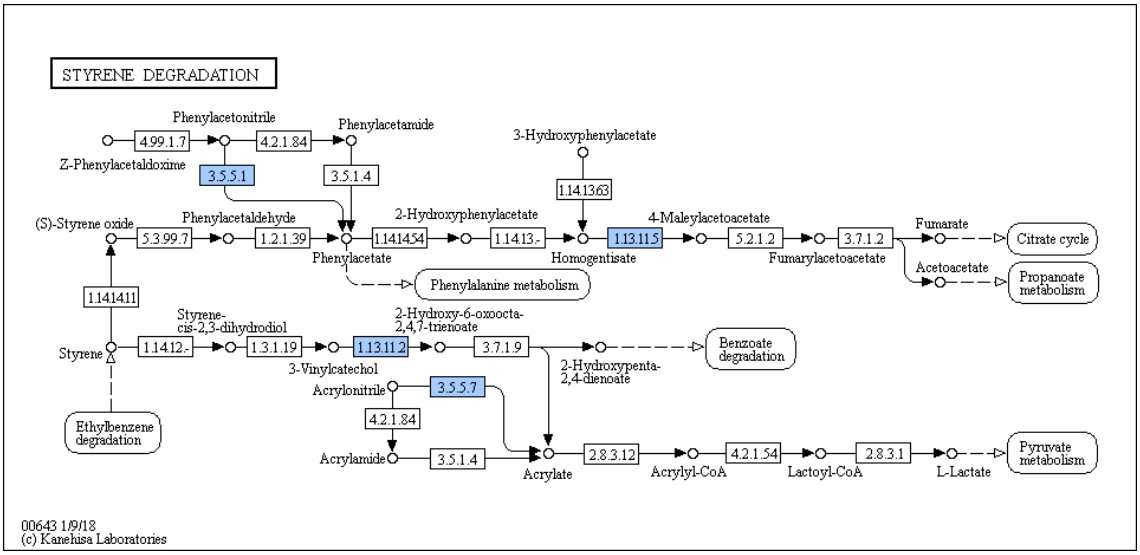


Figure 3.9: Styrene degradation (KEGG map00643).

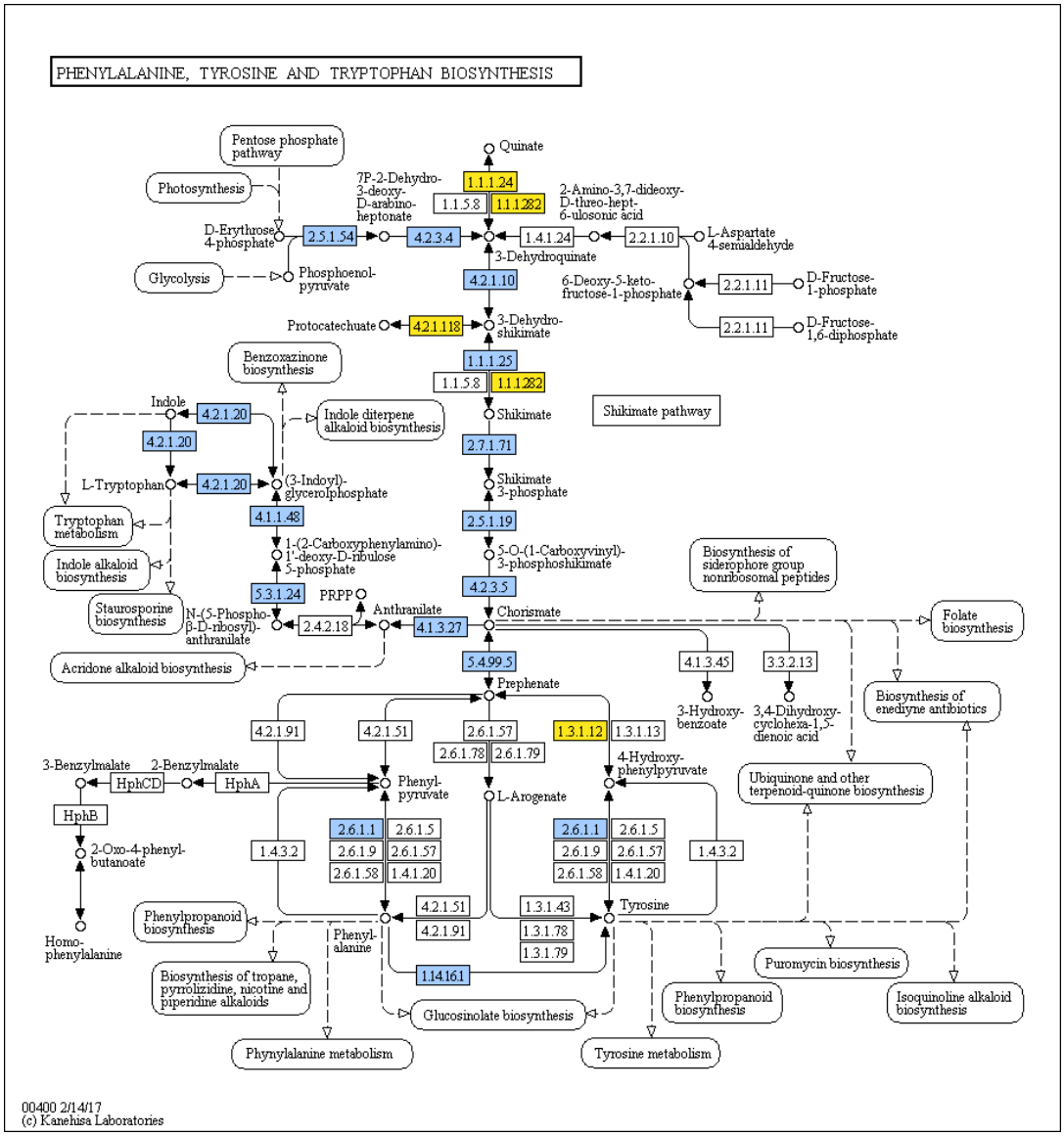


Figure 3.10: Phenylalanine, tyrosine and tryptophan biosynthesis (KEGG map00400).

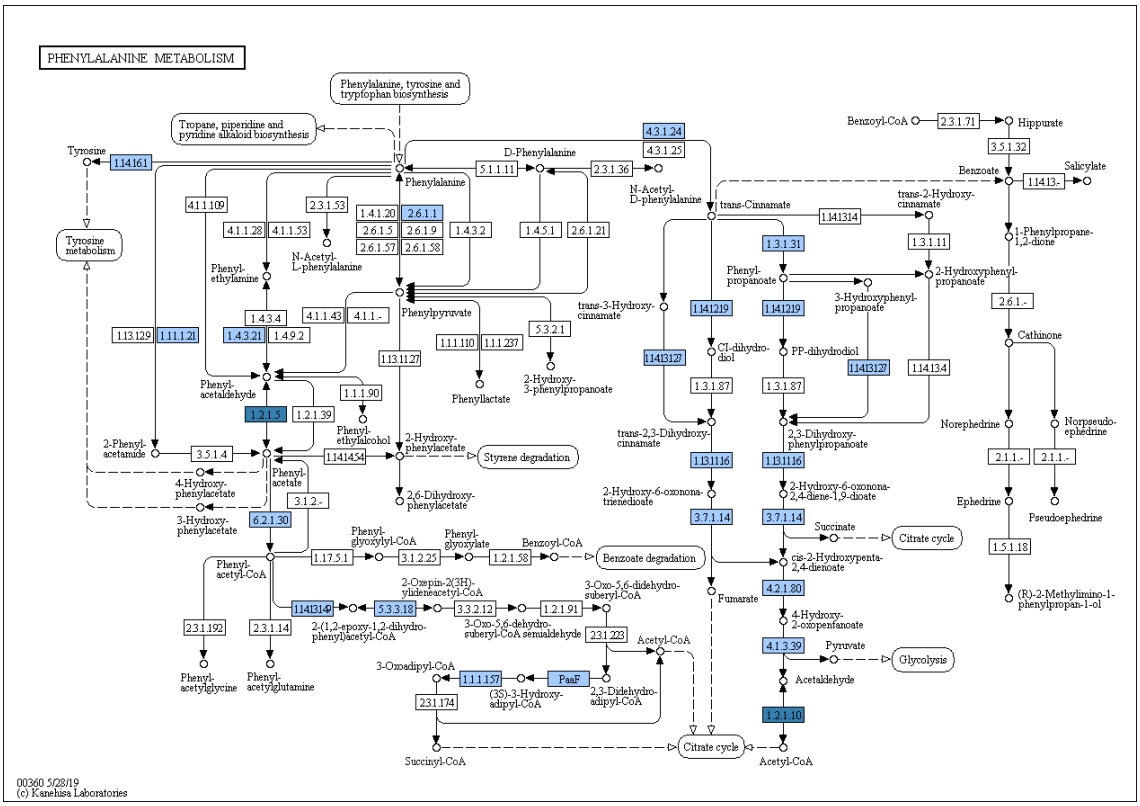


Figure 3.11: Phenylalanine metabolism (KEGG map00360).

Tryptophan metabolism The tryptophan module from KEGG includes two additional sub-modules: melatonin biosynthesis, glucosinolate biosynthesis (Figure 3.12). Many enzymes involved in tryptophan metabolism were missing from the GS database, but all of the enzymes in the database were present in the shotgun sequences. Similar to phenylalanine metabolism, one enzyme, aldehyde dehydrogenase (NAD+; EC:1.2.1.3) was enriched following the DWH spill. Again, this enzyme is found in many pathways, and it may have simply appeared to be enriched following the DWH spill due to increased heterotrophic activity.

Polycyclic aromatic hydrocarbon degradation I was particularly curious about the presence of PAH degrading enzymes in the data. PAHs are an important, recalcitrant component of petroleum, but these compounds can enter an environment from many other sources as well (Figure 3.13). Unfortunately, only one enzyme in the PAH degradation pathway was present in the GS database (Figure 3.14). This enzyme, protocatechuate 3,4-dioxygenase (EC:1.13.11.3), is also associated with benzoate degradation. Future work should focus on characterizing specific proteins involved with PAH biodegradation.

Functional evidence for hypoxic heterotrophic metabolism

Correlations were calculated between oxygen concentration and enzyme abundance across microbial communities. To identify enzymatic functions that might be associated with hypoxia, I looked for significant inverse correlations between enzyme abundance and oxygen concentration. Pathways for nitrogen, sulfur, and methane metabolism are associated with hypoxic and anoxic microbial metabolic activity [172]. Therefore, the enzymes in these pathways were analyzed for correlations.

Nitrogen metabolism The nitrogen cycle is a complex biogeochemical cycle involving interactions of many organisms with multiple environmental factors [300]. Enzymes associated with dissimilatory nitrate reduction, denitrification, nitrification, and anammox were found in the shotgun sequences, however, several enzymatic functions in the pathway have not been experimentally verified. Although these enzymatic activities were not

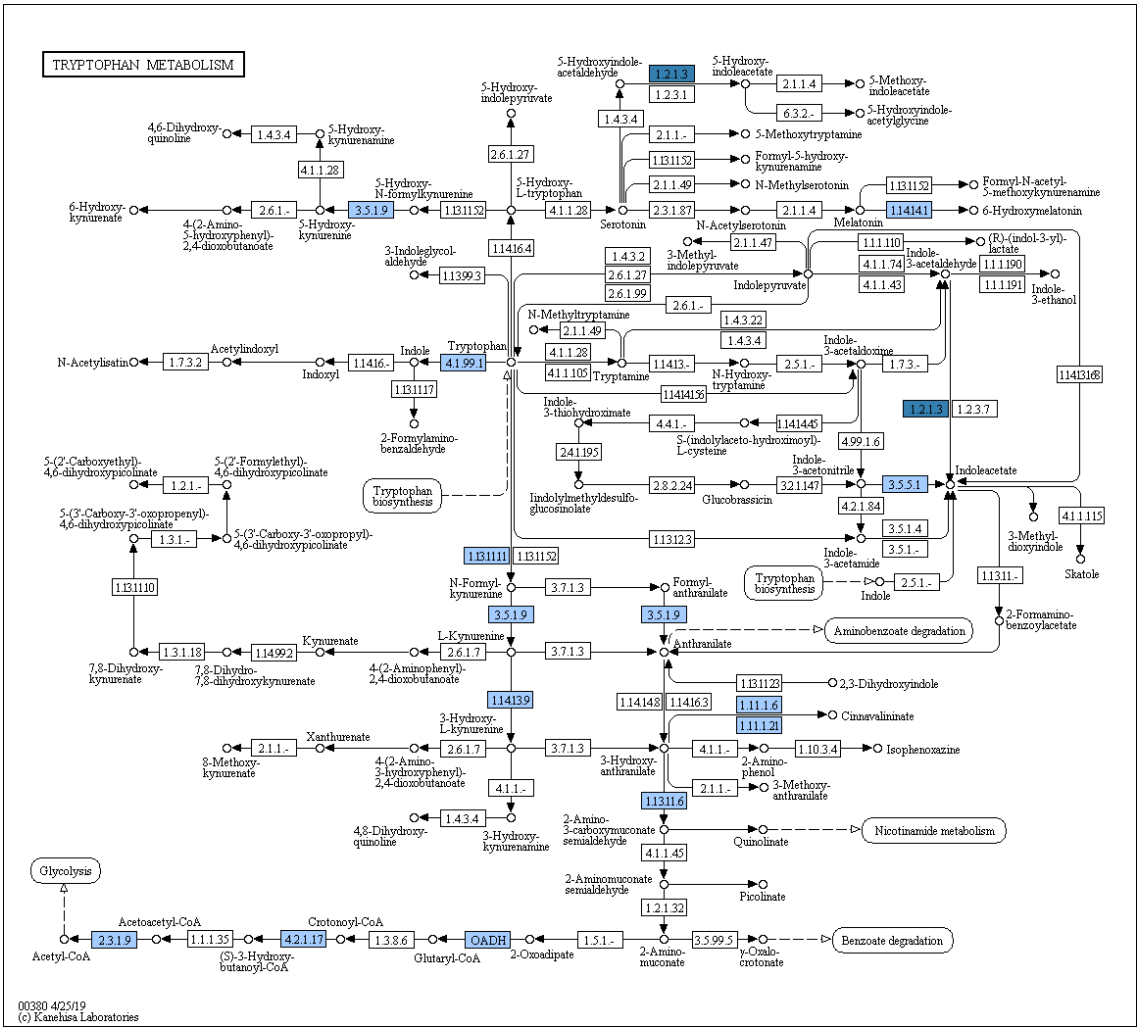


Figure 3.12: Tryptophan metabolism (KEGG map00380).

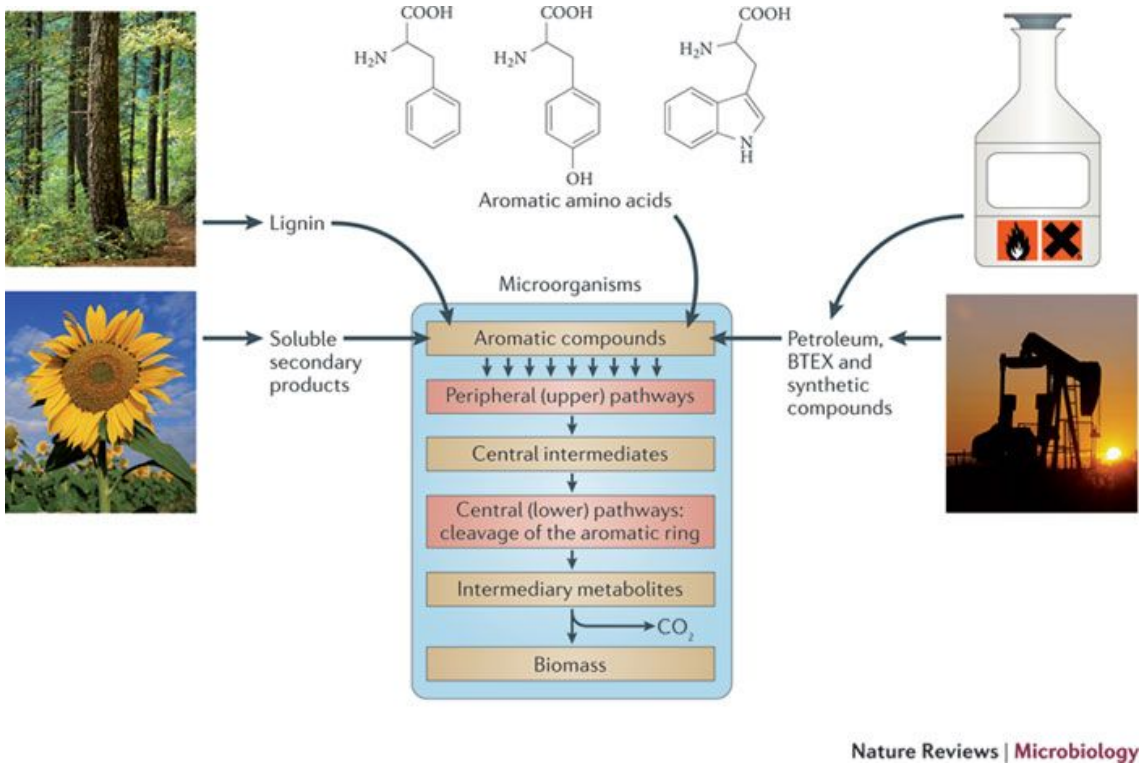


Figure 3.13: Environmental sources of aromatic compounds. All organisms catabolize aromatic amino acids, although animals are not able to synthesize them. Plants produce aromatic compounds in various forms including lignin, which is the most abundant polymer on earth. Petroleum and other chemicals contain aromatic compounds in various forms including BTEX (benzene, toluene, ethylbenzene and xylene), many of which are toxic and/or carcinogenic. Aromatic hydrocarbons are degraded through a wide variety of pathways, but ultimately are funnelled into the central metabolic pathways. Figure from Fuchs et al. [86].

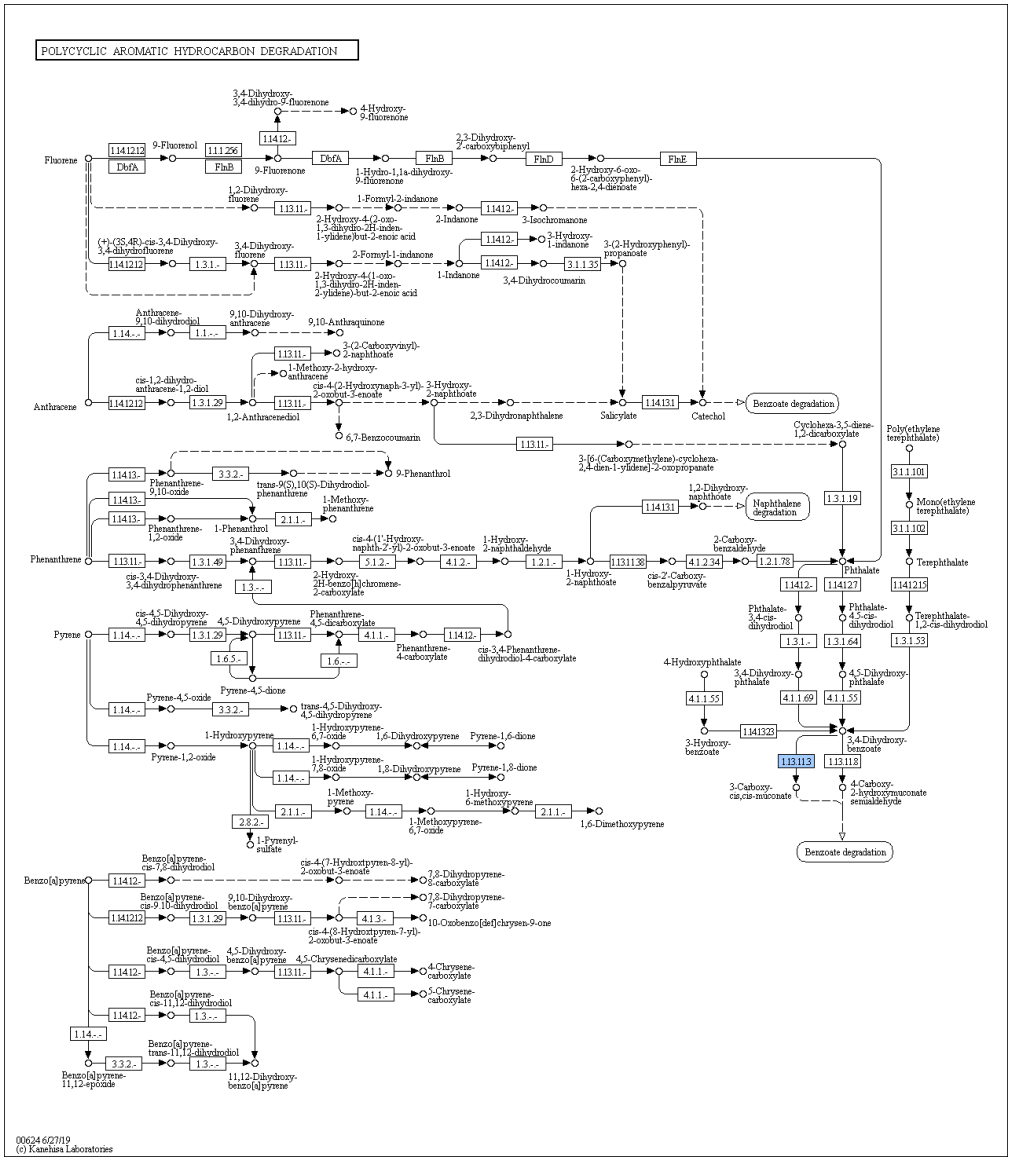


Figure 3.14: Polycyclic aromatic hydrocarbon degradation (KEGG map00624).

significantly enriched in any water masses, there were potentially important correlations between the abundance of two enzymatic functions and decreasing oxygen concentration (Figure 3.15). The two correlated enzymes were nitrite reductase (EC:1.7.2.1; NirK, NirS) and nitrous-oxide reductase (EC:1.7.2.4; NosZ). Nitrite reductase catalyzes the conversion of nitrite to nitric oxide, and nitrous-oxide reductase catalyzes the conversion of nitrous oxide to diatomic nitrogen. Both of these enzymes are associated with denitrification, and nitrite reductase is also associated with anammox. Both of these functions are carried out by anaerobic or facultative anaerobic heterotrophic bacteria. The enzymes that carry out the conversion of nitrate to nitrate, the first step in denitrification, have not been experimentally validated and were not present in the GS database. Future work should investigate the possibility that this function is also present in these microbial communities.

Sulfur metabolism Sulfate can be reduced by assimilatory ("energy consuming") or dissimilatory ("energy producing") pathways. The dissimilatory pathway is only found in anaerobic Bacteria and Archaea, and uses sulfate as a terminal electron acceptor [96]. The three principal enzymes of dissimilatory sulfate reduction were present in both the GS database and in the shotgun sequence data, but none were correlated with hypoxia (Figure 3.16). It has been hypothesized that these enzymes might function "in reverse" to oxidize sulfur compounds to sulfate in chemolithoautotrophic organisms, so it's possible that the presence of these enzymes is not associated with hypoxia in these communities [84].

Methane metabolism The methane metabolism pathway on KEGG includes pathways used by methanogens, which produced methane, and methanotrophs, which consume methane [122]. Methanogens produce methane under anoxic conditions by three different pathways: (1) conversion of CO2 to methane, (2) conversion of C1 compounds (e.g., methanol, methyl-amines) to methane, or (3) conversion of acetate to methane (Figure 3.17). Less than half of the enzymes associated with methane metabolism were in the GS database, and none of the ones that were present in the shotgun sequences were correlated with hypoxia. Of the enzymatic functions associated specifically with methanogenesis in the KEGG pathway, only two were present in the GS database. Other enzymes associated

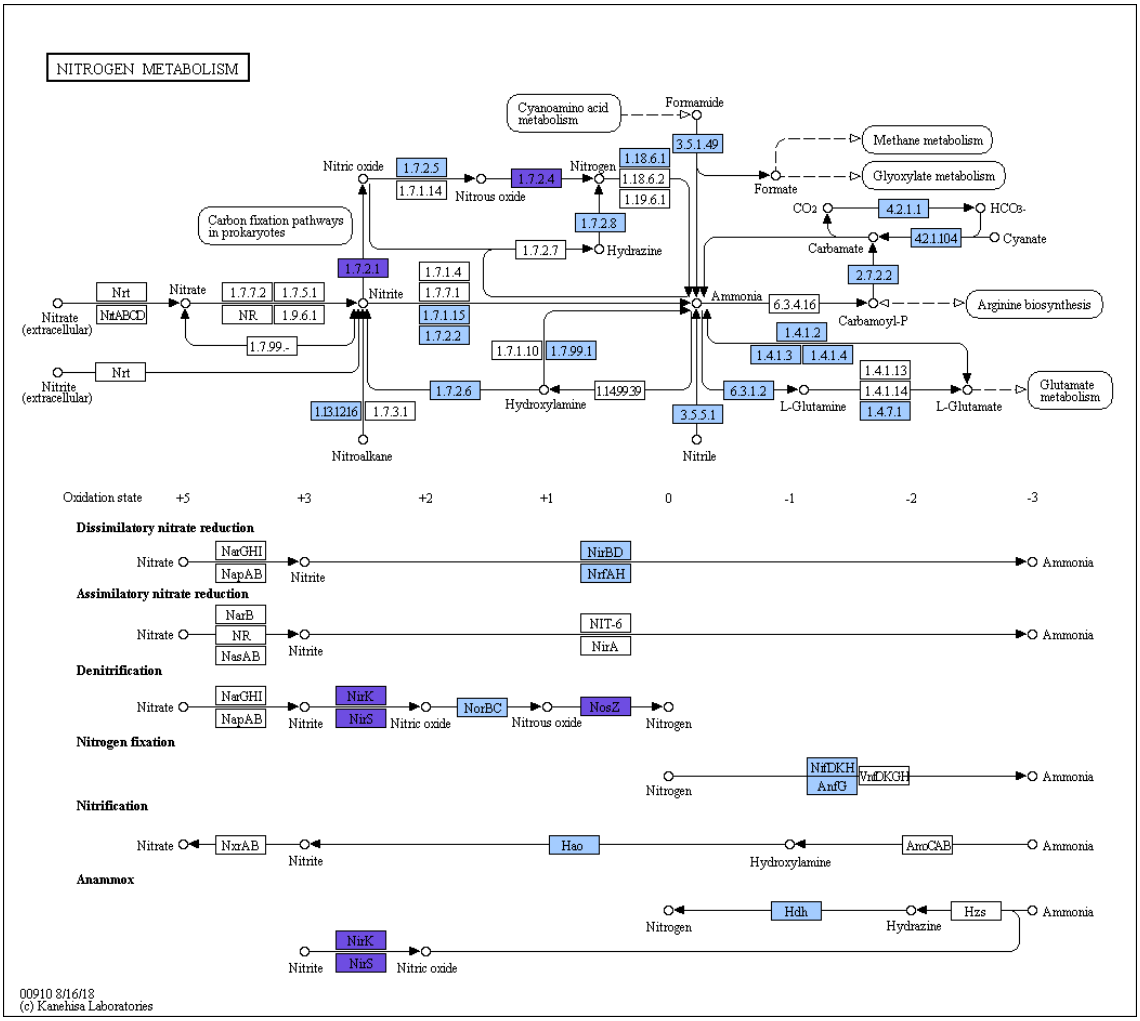


Figure 3.15: Nitrogen metabolism (KEGG map00910).

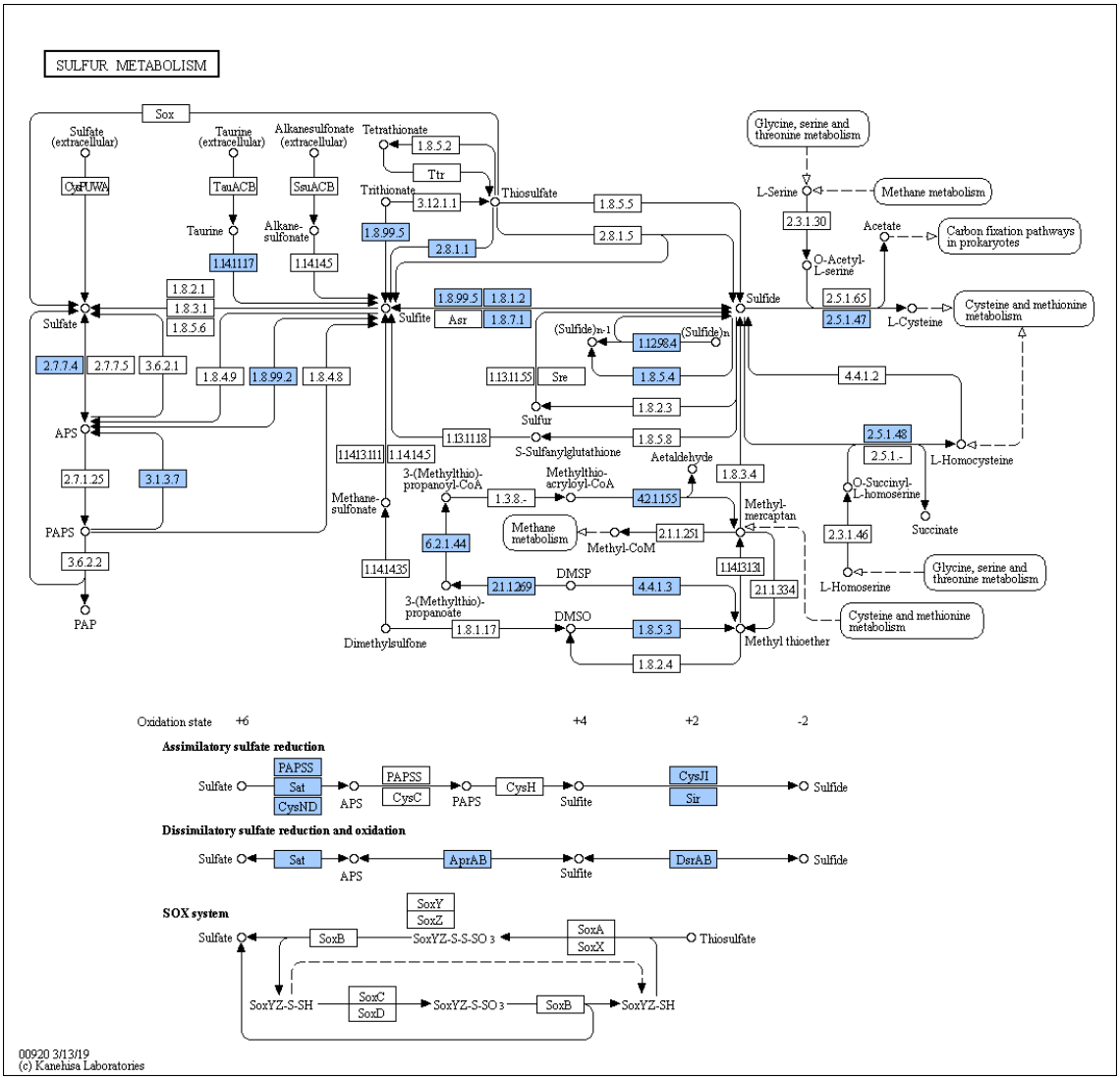


Figure 3.16: Sulfur metabolism (KEGG map00920).

with methane metabolism in the KEGG pathway are also involved in other carbon cycling pathways. The enzymes in the GS database have poor coverage over anaerobic methane metabolism.

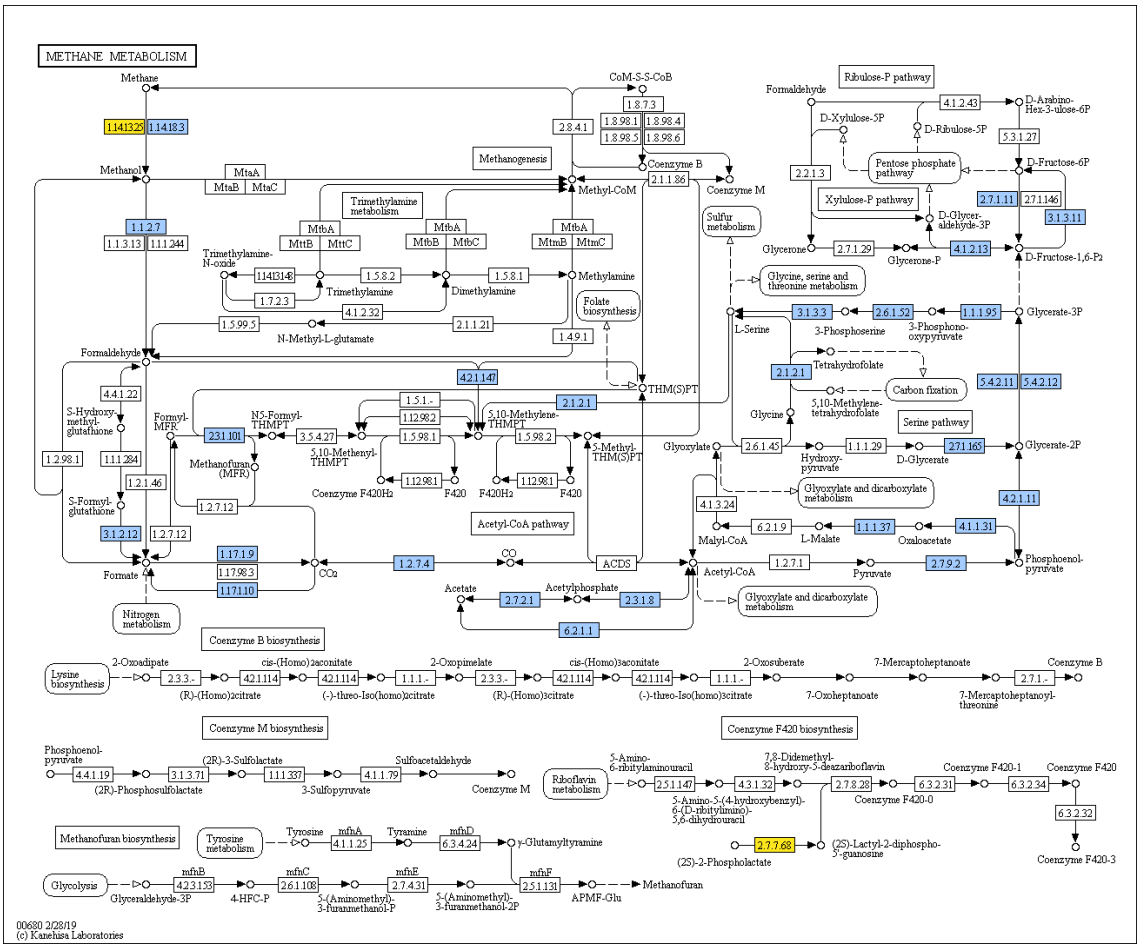


Figure 3.17: Methane metabolism (KEGG map00680).

Chapter 4

Oil Hydrocarbon Degradation by Caspian Sea Microbial Communities

This chapter was originally published as:

Miller, J. I., Techtmann, S., Fortney, J., Mahmoudi, N., Joyner, D., Liu, J., et al. (2019). Oil Hydrocarbon Degradation by Caspian Sea Microbial Communities. Frontiers in Microbiology 10, 995. doi:10.3389/fmicb.2019.00995.

Data and statistical analysis, and preparation and writing of the manuscript were carried out by Miller. Sampling was carried out by Techtmann, Fortney, Mahmoudi, Joyner, Fernandez, Gardinali, GaraJayeva, and Askerov. Microcosm experiments were performed by Liu. The TEXMEX method was developed by Olesen and Alm. AODCs, guidance, and feedback on experimental design, analysis, and manuscript preparation were provided by Hazen.

4.1 Introduction

The Caspian Sea is the largest enclosed body of water on earth. It has a volume of 78,000 km^3 and a surface area of $3.8 \times 10^5\,km^2$ [71, 291]. The sea was once connected to the oceans but has been landlocked for five million years. Primary fresh water inputs are from river runoff, with more than 130 freshwater river inputs, and precipitation; thus, the waters of the Caspian Sea are brackish, with salinity values approximately one third of ocean seawater [171]. The Caspian Sea is divided into three 'basins', and the southern basin is the deepest (1025 m). The surface waters are warm (~25°C) but fluctuate seasonally, and temperatures in the deepest waters (> 100 m) are constant year-round at 6.8°C [291].

The Caspian Sea is polluted with nutrients enrichment from agricultural and industrial runoff. Freshwater input from rivers is nutrient-rich, deriving from industrial and agricultural runoff, and therefore contributes substantial pollution to the Caspian Sea, and this has increased in recent years (Zonn 2005). This causes the waters to undergo eutrophication, resulting in bottom waters that are oxygen deficient with zones of hypoxia (Diaz 2001). Since the middle of the 20^{th} century, the bottom waters have declined from ~26% oxygen saturation to ~5% [54, 291]. The microbial communities in the Caspian Sea have been influenced by eutrophication for at least 50 years. Eutrophication in the Caspian Sea has caused the deep waters to become severely hypoxic [244, 201]. Hypoxic zones may have negative

environmental consequences such as increased mortality of benthic organisms, decreased biodiversity, and altered biochemical cycles (Diaz and Rosenberg 1995).

There is a large influx of oil hydrocarbons into the Caspian Sea from both natural and anthropogenic sources. Although this environment has a long history of natural exposure to oil hydrocarbons, the concentration of oil hydrocarbons has increased due to increased oil exploration and production in the recent past. Furthermore, oil and gas exploration and production from the Caspian Sea have also increased in recent years. Between 70,000 - 90,000 tons of petroleum hydrocarbons are released into the Caspian Sea annually, and total petroleum hydrocarbon concentrations range from 0.12 mg/L in the shallow water to <0.02 mg/L in the deep water [54, 155]. The highest petroleum hydrocarbon concentrations have been reported in the southern basin and in surface waters and sediments, but are 'almost undetectable' in the water column below 500 m (Korshenko and Gul 2005).

Hydrocarbon degrading microorganisms have been isolated from the coastal waters and sediments. Many contaminants, especially organic compounds (e.g., oil hydrocarbons) are derived from natural compounds or have natural analogs in the environment [33]. Biodegradation of contaminants by microbial communities is an important process. Hydrocarbon degrading *Bacteria* and fungi have been isolated from the coastal waters and sediments of the Caspian Sea [114, 115, 161, 177, 243, 245, 259, 296, 297], but the taxonomic identity of these isolates was often uninvestigated. Several of the identified isolates (e.g., *Pseudomonas* sp. and *Gordonia* sp.) demonstrated growth on crude oil [114]. Most probable number experiments indicate that heterotrophic *Bacteria* in coastal sediments are high (2.5×10^5 cells/mL), and that hydrocarbon degrading *Bacteria* are also high [115]. Many oil degrading microbes exist in the shallow waters of the Caspian Sea, but the structure of the community has not been investigated.

Microbial communities contribute substantially to oil biodegradation under both aerobic and anaerobic conditions [58]. Anaerobic biodegradation of hydrocarbons is a significant process in many environments [92, 99, 166, 305] and may be an important process in the deep-water communities of the Caspian Sea. Anaerobic hydrocarbon degradation is most frequently reported to be slower than aerobic hydrocarbon degradation, with biodegradation coefficients of 0.445/d and 0.522/d, respectively [307, 278]. Anaerobic hydrocarbon

degradation has been studied in a variety of environments and conditions including the following: fossil hydrocarbon reserves, thermophilic communities, and in microcosms inoculated with sediment under sulfate-reducing conditions [27, 166, 258]. Previous investigations of the Caspian Sea microbial community most frequently characterized isolates from shallow, coastal waters, with fewer reports on the deep, oxygen deficient waters (Diarov and Serikov 2006). Opportunistic organisms capable of metabolizing oil hydrocarbons may be ubiquitous globally in deep sea basins, but they vary in both biodegradation capabilities and proportional contribution to the community structure [118]. Population blooms of these hydrocarbon degrading *Bacteria* are expected with an influx of oil hydrocarbons into a system that is likely to occur from increased oil exploration and recovery. Indigenous microbial taxa and communities may adapt to contaminants if the exposure continues for a period of time [33, 267]. Furthermore, opportunistic organisms are known to become enriched in the event of an influx of hydrocarbons into a system, but vary in their relative abundance and contaminant degrading abilities [20, 118]. The unique combination of physical and chemical parameters may have caused adaptations within the microbial community of the deep waters of the Caspian Sea.

The deep waters of the Caspian Sea represent a unique environment that is cold, hypoxic, and has a large influx of oil hydrocarbons. The deep waters of the Caspian Sea are known to be severely hypoxic, and the microbial community in these waters has not previously been investigated. Due to the unique combination of exposure history to oil hydrocarbons and persistent hypoxia, the deep waters of the Caspian may harbor communities with novel adaptations for anoxic hydrocarbon degradation. However, little knowledge exists about anaerobic hydrocarbon biodegradation in psychrophilic communities like those from the deep waters of the Caspian Sea.

The purpose of this study was to characterize the natural, *in-situ* community structure, and then to investigate the role of that community in oil hydrocarbon degradation. Baseline sampling provides a snapshot of the native, ambient community structure on the date of sampling, representing non-perturbed environmental conditions, which is an important reference for experimental studies. We used microcosm experiments to investigate changes in the microbial community composition and the potential of this community for oil

hydrocarbon biodegradation under oxic and anoxic conditions. We report here the changes in community composition and evaluate those changes with respect to the baseline community, and we also report the extent of degradation of various oil hydrocarbons and the differences in degradation.

4.2 Materials and methods

4.2.1 Sample collection and environmental parameters

Water samples were collected from six sites in the southern basin of the Caspian Sea between July 27, 2013 and August 1, 2013. Sampling was conducted as part of BP's oceanographic survey by Fugro, aboard the MV Svetlomor II. Niskin bottles were deployed for water collection. A MIDAS CTD + Profiler (Valeport Ltd, St. Peter's Quay, UK) was attached to the sampling rosette for continuous monitoring of physical and chemical water parameters. Environmental variables (temperature, dissolved oxygen, salinity, pH, turbidity) were measured continually through the water column (Figure 4.1).

Sample sites were chosen to sample diverse sea floor features, both natural and man-made (sites 2-6), and one control site with no known man-made features (site 1) (Table 4.1) Man-made features include drill cuttings splay, debris, and oil and gas wells (active and abandoned). Water samples were collected from 2-4 depths at each sample site, depending on the depth to seafloor. Sample depths for each site were selected to approximate the following categories (where appropriate): near surface, one-third depth from surface, two-thirds depth from surface, and near bottom. In total, nineteen samples were collected.

In situ sampling of ambient seawater was conducted as follows. Between 62 and 123 liters of ambient seawater were filtered using a large volume pump (McLane Research Laboratories, East Falmouth, MA). Different amounts of water were sampled due to the differences in the amount of particular matter at different sample locations, which affected filtration. The water was filtered through a 142 mm nylon membrane with a pore size of 0.2 m (Sterlitech, Kent, WA) and then stored at -20°C. One third of the filter was used for DNA analysis reported here.

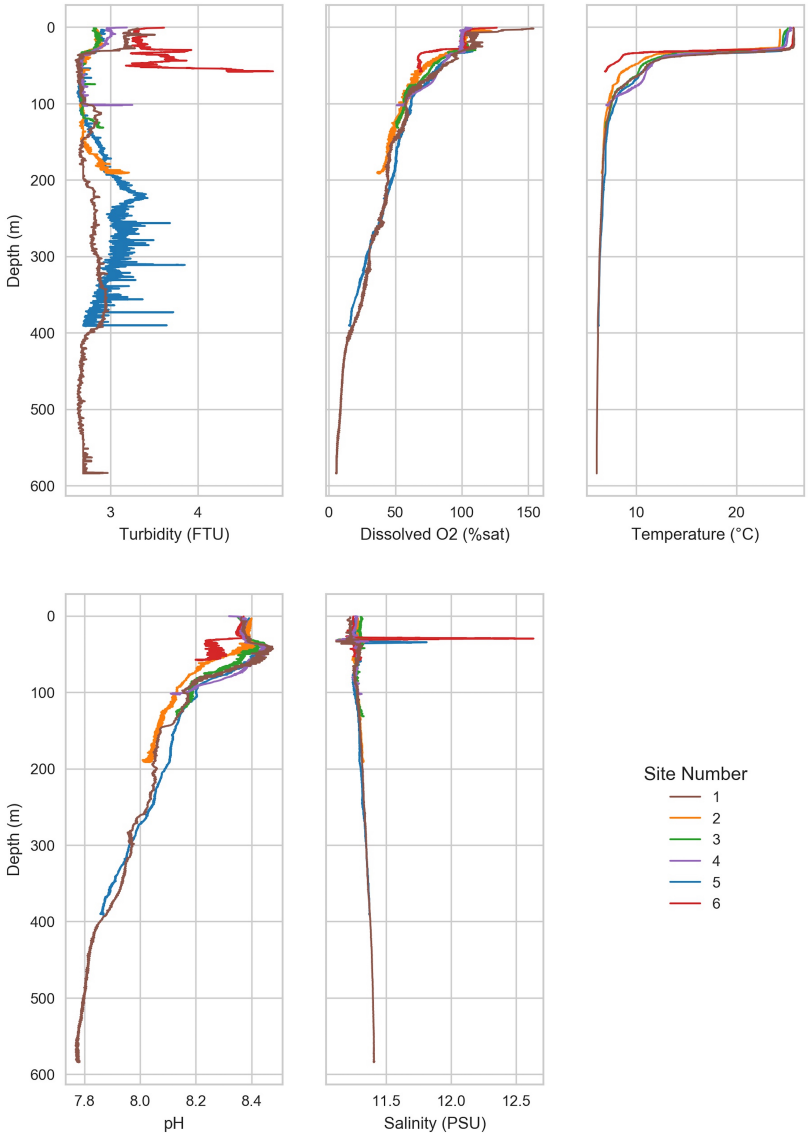


Figure 4.1: Environmental factors varied with depth. Environmental variables were measured throughout the water column from the surface water to near sea floor. A thermocline was observed at 50 m below sea level. Below the thermocline, dissolved oxygen and pH declined with depth.

Table 4.1: The important seafloor features near each sampling site are listed. Sample sites were chosen in part based on important sea floor features. Some features are man-made (M), while others result unintentionally from anthropogenic activities (A). All other features are assumed to be natural (N).

Site:	1	2	3	4	5	6
Drill cuttings splay (M)		Y	Y	Y	Y	
Embayment bounding ridge (N)				Y		
Eroded seabed (A)			Y			
Exposed landslide deposit (N)				Y		
Landslide visible at seabed (N)			Y			
Main landslide headwall scarp (N)			Y	Y		
Main landslide shear (N)			Y		Y	
Major crustal fault at seabed (N)				Y		
Man-made objects, debris (M)			Y	Y	Y	
Mud mound or mud volcano flow track (N)			Y		Y	
Mud volcano (N)	Y		Y	Y		
Plugged and abandoned oil and gas well (M)			Y	Y	Y	
Plugged and abandoned oil well (M)			Y	Y	Y	
Prospecting (M)			Y	Y	Y	
Tension crack (A)			Y			

Immediately following string recovery, seawater was dispensed into 4-liter amber bottles using clean TYGON Tubing (Saint-Gobain, La Défense, Courbevoie, France) to limit aeration. Bottles were stored on-ship at 4°C and then shipped on wet ice. Forty mL of water was fixed in 4% formaldehyde and stored at 4°C for acridine orange direct counts (AODC)s.

One hundred mL of water was frozen at -20°C for analysis of dissolved organic carbon and nutrients. Total organic carbon and total nitrogen were analyzed with TOC-L analyzer (Shimadzu Scientific Instruments, Columbia, MD), and inorganic nutrients were analyzed with a SEAL AutoAnalyzer 3 HR (SEAL Analytical Inc., Mequon, Wisconsin). Nutrients (nitrate, nitrite, ammonia, total nitrogen, inorganic phosphate, silicate) were measured at each of the sampling sites and depths.

4.2.2 Acridine orange direct cell counts

Acridine orange direct cell counts (AODC) were performed as describe by Francisco et al. [83]. Cell counts were performed with Zeiss Axioskop epifluorescence microscope (Carl Zeiss, Inc., Germany).

4.2.3 Microcosms experiments

Laboratory microcosm experiments were prepared in order to investigate (1) the change in microbial community composition and structure and (2) it's potential for oil biodegradation under oxic and anoxic conditions. Microcosms were prepared in triplicate at 6°C using the deepest waters from stations (575 m) with either nitrogen (anoxic) or atmospheric (oxic) headspace. They were amended with either 100 ppm native oil hydrocarbons (E10 slot crude oil) or 100 ppm oil and 1 ppm Corexit dispersant (Nalco, Sugar Land, TX) for comparison with Bælum et al. [20]. Control microcosms were killed and then amended as above. Carbon dioxide evolution was measured continually with a Micro-Oxymax Respirometer (Columbus Instruments International, Columbus, OH) for 17 days. Microcosms were removed from the experiment in triplicate and killed at days 0, 3, and 17. One microcosm was used for 16S rRNA gene sequencing and and two were used for hydrocarbon quantification.

4.2.4 Hydrocarbon analysis

Native crude oil "E10 slot" was provided by BP. At 15°C, the API gravity of E10 slot oil is approximately 33, and the density is 0.86 g/L. Samples from each microcosm were sent to Florida International University and processed by liquid-liquid extraction with methylene chloride for hydrocarbon quantification. Saturated hydrocarbons and an estimation of total petroleum hydrocarbons (TPHs) were performed by gas chromatography-flame ionization detector (GC-FID) using a modification of SW-846 Method 8015. A subset of relevant polycyclic aromatic hydrocarbons (PAHs) including the 16 EPA priority PAHs and their alkylated homologues were analyzed using gas chromatography-mass spectrometry (GC/MS) in selected ion monitoring (SIM). The recalcitrant biomarker C30-hopane was also measured in the GC/MS method. This protocol is described in EARL-SOP-2000-O-109 and is based on a modification of previously reported procedures expanded to accommodate additional analytes [63]. 1998).

4.2.5 DNA extraction, 16S rRNA gene amplicon sequencing, and data pre-preprocessing

Genomic DNA was extracted using methodology described by Miller, et al. (1999) with modifications as described by Hazen et al. [117]]. DNA was cleaned using the Genomic DNA Clean & Concentrator (DCC) kit (Zymo Research, Irvine, California). Quality of extracted DNA was determined by measuring the 260/280 and 260/230 ratios on a NanoDrop spectrophotometer (Thermo Fisher Scientific, Waltham, MA). DNA concentration was determined by PicoGreen (Thermo Fisher Scientific, Waltham, MA).

The 16S rRNA gene libraries were prepared as described by Caporaso et al. (2012). The V4 region of the 16S rRNA gene was amplified by PCR using universal primers 515f and barcoded 806r, which anneals to both *Bacterial* and *Archaeal* sequences, with Phusion DNA polymerase (Thermo Fisher Scientific, Waltham, Massachusetts). A 12 base pair barcode index was included in the reverse primer to multiplex samples for sequencing analysis. The 16S rRNA gene amplicons were then pooled together, and the quality and size of the amplicons was analyzed using a Bioanalyzer (Agilent Technologies, Santa Clara, CA). The

16S rRNA gene libraries were sequenced using a MiSeq with a V2 kit (Illumina, San Diego, CA).

4.2.6 Bioinformatics processing of sequence data

The resulting sequence information was pre-processed with the QIIME wrapper software (Caporaso, et al. 2010) on BioLinux [80]. Paired-end reads were joined using fastq-join (Aronesty 2013), and then paired sequences were demultiplexed and quality filtered to remove reads with phred scores below 20 (using the QIIME script `split_libraries_fastq.py -q 19`). Chimera detection was then performed on demultiplexed, quality-filtered reads using UCHIME including de novo and reference-based chimera detection [74, 75]. Chimera-checked sequences were clustered into OTUs at 97% sequence similarity using UCLUST using the open reference clustering protocol with the GreenGenes 16S rRNA gene sequence database as reference [74, 64]. Counts of the reads comprising each OTU were exported to biom format for statistical analysis [184]. Read counts for each sample were rarefied to a number of reads equal to that of the smallest sample. Then, OTUs that comprised $< 0.005\%$ of total reads across all samples were removed from the data before downstream analysis.

Illumina MiSeq analysis of 16S rRNA gene amplicon sequences from the microcosms resulted in mean 389,492 (477,528) reads per sample (after filtering). A total of 14,913 OTUs were identified, with mean 6,892 (5,197) OTUs per sample. All microcosm 16S rRNA gene OTU data are available in NCBI's SRA.

4.2.7 Statistical analysis

Samples were divided into two water masses (shallow and deep) for analysis of environmental parameters and microbial communities; shallow waters were defined as waters 50 m in depth, and deep waters were defined as waters $>$ 50 m in depth, which is consistent with the observed thermocline. To test the hypothesis that environmental parameters were different in the two water masses, the means of each parameter were compared by t-test for independent (unpaired) samples (alpha = 0.05). Non-metric multidimensional scaling (NMDS) was performed to test the hypothesis that environmental parameters influence

the community structure. A weighted Unifrac distance matrix was constructed and used as input for NMDS. NMDS was performed and environmental parameters were fit to the ordination result using the vegan package in R [208]. To test the hypothesis that the two water masses were enriched in different taxa, the OTU read counts were grouped at the order and genus level and tested with Fisher's exact test (alpha = 0.05); P-values were corrected with the Šidák method [261]. To determine which organisms responded to oil amendment during microcosm experiments, the OTU read counts were summarized at genus level and then transformed using the TEXMEX method [209]. Following transformation, the change in community structure of the oil amended microcosms was compared to that of the un-amended microcosms in each atmospheric condition.

4.3 Results and discussion

4.3.1 Distinct environmental niches exist in the Caspian Sea

There is a thermocline at ~50-100 m in all sample sites. The shallow waters (50 m) were warm (20.2°C 5.1°C), but temperature declined rapidly and finally stabilized below 100 m (7.2°C 1.4°C). The most rapid decline in temperature was observed at site 6, which dropped to 7.0°C at 58 m. The lowest observed temperature was 6.1°C in the deepest waters (Site 1, 575 m). These measurements were consistent with previous reports [291].

Dissolved oxygen (DO) declined with depth at all sample sites. Dissolved oxygen in shallow waters was near saturation (8.38 0.17 mg/L) but declined rapidly until ~50 m (Figure 4.1). Below 50 m, DO declined slowly with depth reaching 5 mg/L by 100 m, and continued to decline to <3 mg/L (the upper limit for hypoxia) below 250 m [127]. The lowest oxygen concentration (0.5 mg/L) was observed in the deep waters of site 1, which is considered severe hypoxia.

Salinity was similar throughout the water column, but pH generally changed with depth. Mean salinity was 11.3 0.06 PSU, which is approximately one-third of mean ocean salinity (Wilson 1975). The mean pH was 8.2 0.2 and declined with depth from 8.4 (site 3) in the

shallow waters to 7.8 (site 1) in the deepest waters. Except for site 6, all sites increased in pH slightly around 50 m before declining steadily with depth.

Ratios of carbon, nitrogen, and phosphorous were different from the expected Redfield ratio and were higher than reported in other marine basins [177]. The N:P ratio (63:1) for the sample collected from site 4 at 60 m depth was identified as an outlier and therefore removed from the calculations [185]. The mean N:P ratio was near the Redfield ratio (12.1 6.9) but varied substantially (range 2.30 - 25.0; Figure 4.1).

4.3.2 Microbial community composition changes with environmental niche

The number of cells in the water column was highest in near surface waters and declined with depth. Acridine orange direct counts (AODCs) ranged from 4.6×10^3 to 2.3×10^5 cells/mL across all samples (Figure 4.2). The shallowest sites (3, 4, and 6) had the greatest range of AODCs throughout the water column. Although AODCs generally declined with depth across all sites, they varied little at site 1 throughout the water column (site 1 is the deepest site and also the farthest from known man-made structures).

The shallow water communities were dominated by *Bacteria*. *Bacteria* dominated the shallow waters, comprising 98.8% (1.6%) of all reads in shallow water communities. *Archaea* collectively comprised only about 1% of reads in the shallow waters. *Proteobacteria* were the dominant phylum (38.4 7.0% of reads) followed by *Bacteroidetes* (20.4 2.0%), and *Cyanobacteria* (12.3 3.9%). The most abundant Bacterial orders were *Flavobacteriales* (15.6 1.4%), uncharacterized *Proteobacteria* (10.3 2.2%), and MWH-UniP1 (2.2 0.7%). The most abundant phyla were *Proteobacteria* (30% of all reads), *Crenarchaeota* (18%), and *Bacteroidetes* (13%) (Figure 4.3). *Proteobacteria* dominated the shallow water communities (38%) but comprised a substantial fraction at all sites and depths. *Proteobacteria* may be enriched during oil spills in other regions of the world [20, 98, 117]. *Bacteroidetes* were observed in highest abundance in the shallow water communities (20%) and declined with depth. *Bacteroidetes* have been observed as co-dominant in oil-impacted seawater [149] Our results indicate that *Proteobacteria* and *Bacteroidetes* are dominant phyla in the Caspian

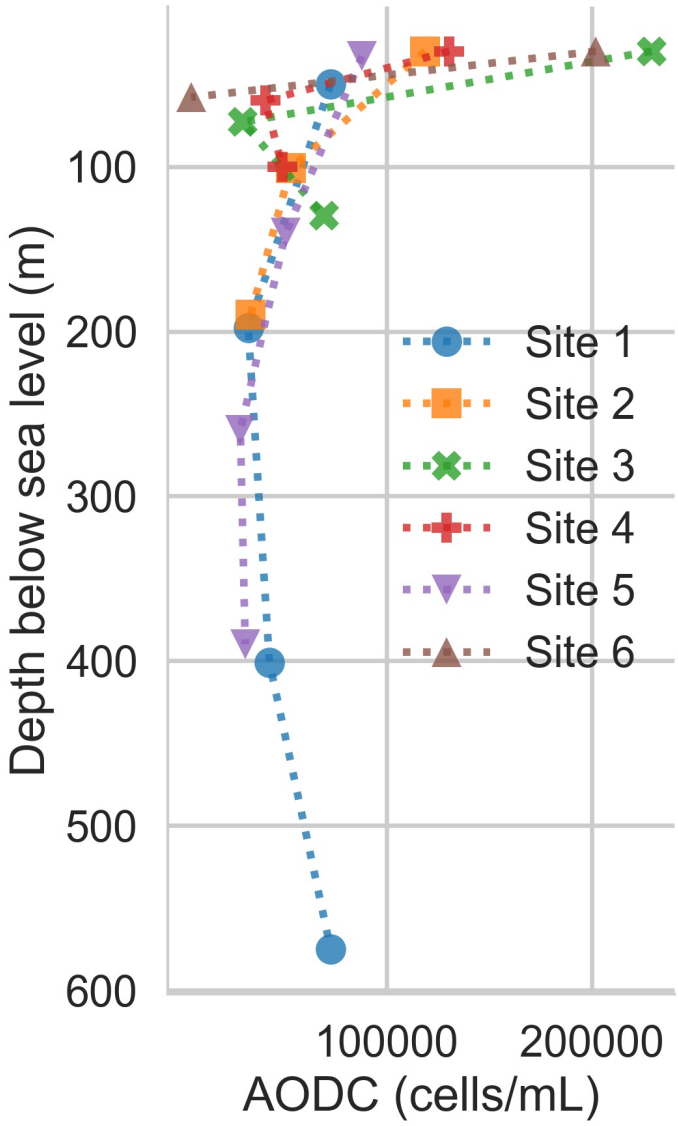


Figure 4.2: Acridine orange direct cell counts (AODCs).

Sea, which is consistent with reports of these phyla in other oil impacted basins and with reports that the Caspian Sea receives large inputs of petroleum hydrocarbons [54] The fact that *Proteobacteria* and *Bacteroidetes* comprise large fractions of the shallow and deep-water communities, respectively. At the order level, *Flavobacteriales*, *Actinomycetales*, and *Oceanospirillales* co-dominated the ambient communities. *Flavobacteriales* have previously been reported as dominant in shallow water communities, and members of this order are commonly reported as chemoheterotrophs and hydrocarbon degraders [173, 8]. *Gammaproteobacteria* are commonly found in oil-impacted marine environments and dominate oil amended microcosm experiments for $\geq$ 20 days; their (co)dominance therefore suggests that these communities are capable of oil biodegradation [118, 240]. Interestingly, *Proteobacteria* that were unclassified at the class level were a substantial fraction of the shallow water communities. It is unknown what role these unclassified microbes may play, if any, in hydrocarbon degradation.

The deep water communities were dominated by Archaeal taxa. When taken together, *Bacteria* dominated the deep waters but declined from >98% to 70.8% (7.6%) of all reads in deep water communities. Although *Bacteria* dominated all samples (at all depths), they decreased in abundance with depth, while *Archaea* increased in abundance. *Planctomycetes*, however, were enriched in the deep waters in contrast with other *Bacterial* phyla. When taken together, *Archaea* were present in large numbers comprising 28.2% (7.3%) of all reads in deep water communities. The most abundant phylum was *Crenarchaeota* (26.4 6.7%), followed closely by *Proteobacteria* (25.0 9.9%). *Phycisphaerales* were commonly observed in the deep-water communities (2%), which was unexpected because they are known to be enriched in aerobic seawater [301]. Curiously, although *Gammaproteobacteria* tended to decline with depth (i.e., in a manner consistent with *Bacteria* in general), the highest fraction was observed in the deepest sample (36%, site 1). Two-thirds of these *Gammaproteobacteria* were unclassified at the order level, and these unclassified reads dominated the community (23%). Other *Bacteria* that increased in abundance uniquely in this sample included SAR406 AB16 Arctic96B-7 (3% compared to $<$ 1% in general) and WS3 PRR-12 GN03 ($\sim$ 1% of reads but rarely observed otherwise). *Crenarchaeota* dominated most of the deep-water communities (26%), suggesting that these communities

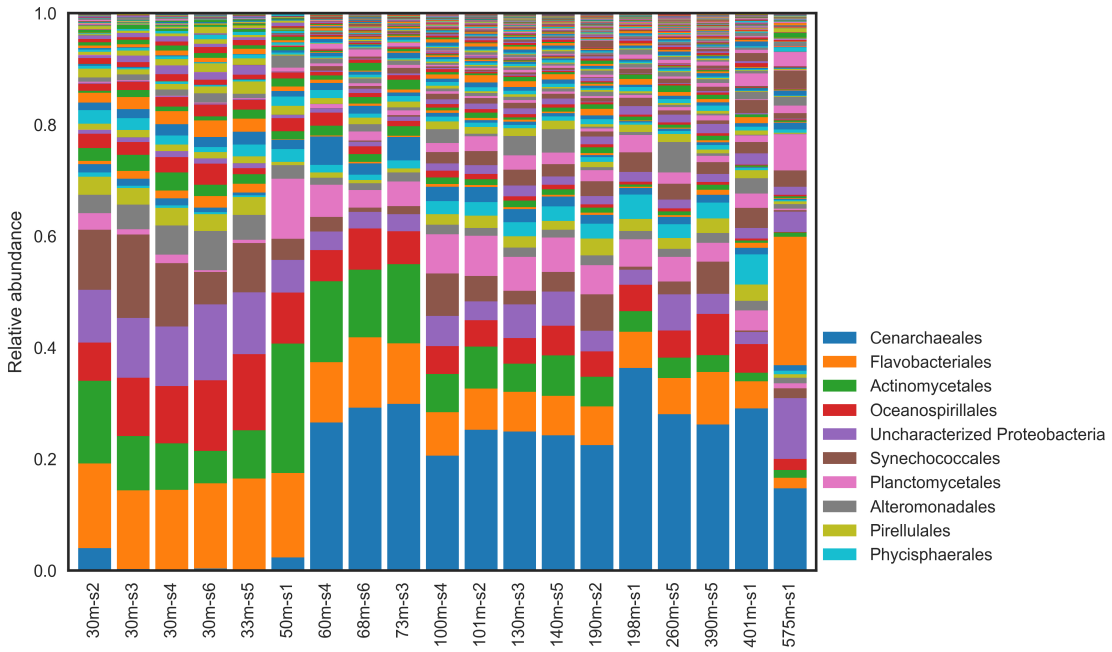


Figure 4.3: Relative abundance of microbial taxa in natural communities at Order level.

may be nutrient limited since this group survives predominantly by ammonia-oxidation [118]. Other highly abundant Bacterial phyla included *Bacteroidetes* (9.7 3.4%), *Chloroflexi* (4.7 2.5%), and *Cyanobacteria* (3.4 2.6%), although *Bacteroidetes* and *Cyanobacteria* declined to less than half of their relative abundance compared to shallow waters. The most abundant taxonomic order was *Cenarchaeales* (26.4 6.7%), which comprised nearly all of the *Archaea*. *Thaumarchaeota* are generally considered to be ammonia-oxidizers and play an important role in nitrogen cycling. Prior studies showed that they are important in the sediments and deep waters of the Caspian Sea [177, 281]. They have also been reported to dominate the communities in the Baltic Sea, which, like the Caspian Sea, is characterized by low salinity and low oxygen concentration [163]. It has been hypothesized that these *Thaumarchaeota* are adapted to eutrophic conditions and are not obligate autotrophic ammonia oxidizers [281]. *Cenarchaeales* have been observed in polychlorinated biphenyl and PAH contaminated sand samples and contribute to the degradation of biphenyl in activated biosludge [110, 272]. Following *Cenarchaeales*, the remaining highly abundant orders were all *Bacteria* and were more evenly distributed. These included *Flavobacteriales* (7.8 2.9%), uncharacterized *Proteobacteria* (4.2 2.3%), and *Phycisphaerales* (2.6 1.3%). Many of the reads observed in the microcosm communities were not able to be classified at high taxonomic levels.

There were important correlations between environmental factors and microbial communities. Non-metric multidimensional scaling (NMDS) of weighted Unifrac distances between samples implies clusters of the communities according to depth (Figure 4.4). Temperature and DO decreased with depth and are influential factors on the shallow water communities (Table 4.2). Inorganic phosphate, silicate, nitrate, salinity and total nitrogen increased in with depth and are likely to be influential on the deep-water communities.

4.3.3 A microbial community from a deep water OMZ rapidly degrades recalcitrant hydrocarbons

Microbial communities changed in the presence of oil differently if they were prepared under oxic or anoxic conditions. TEXMEX highlights organisms that increased in abundance following oil amendment compared to organisms that increased in both control and

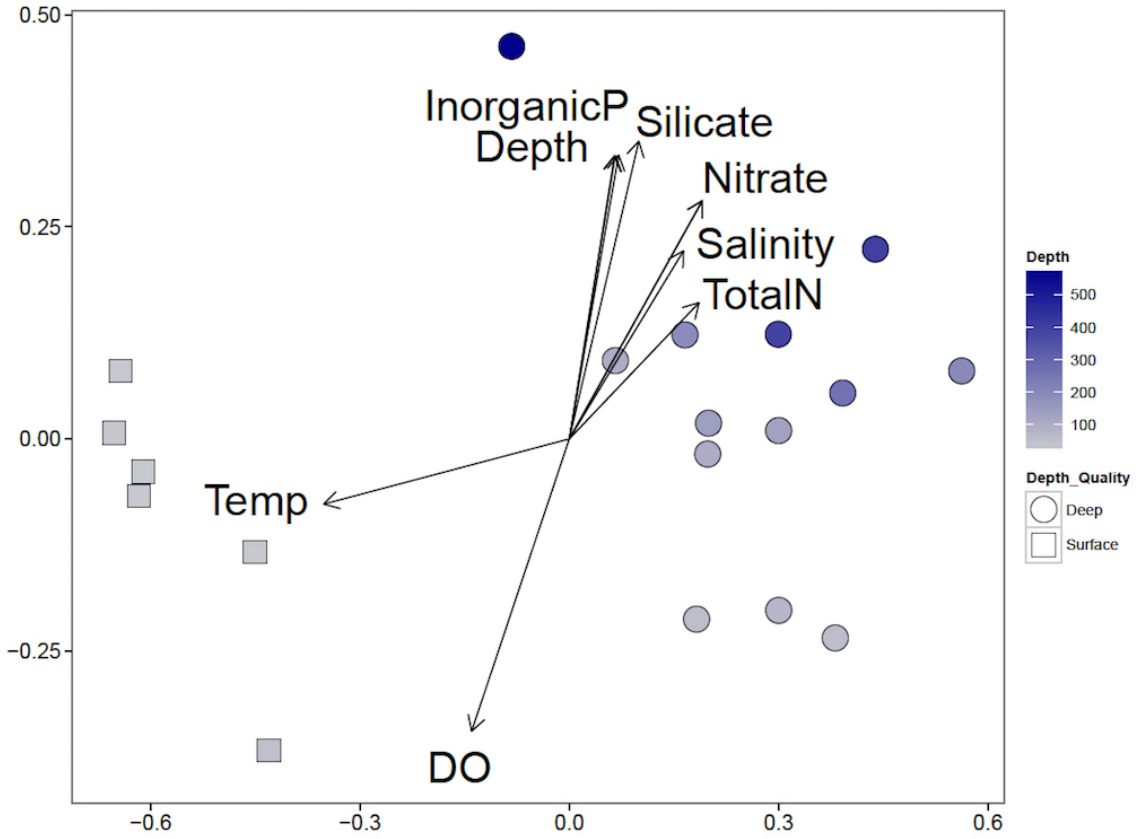


Figure 4.4: Non-metric multidimensional scaling (NMDS) correlation biplot of shallow and deep-water communities and influential environmental variables. NMDS of weighted UniFrac distances (stress 0.075, P-value <0.05) are shown. Shallow water communities are marked with squares, and deep-water communities are marked with circles. The blue intensity of the marker indicates the depth, with the shallowest samples in gray and the deepest samples in dark blue. Environmental variables were fit to the two-dimensional representation of the microbial communities; arrow directions indicate correlation with each axis and the size of the length of the arrow indicates the correlation coefficient.

experimental conditions, thus mitigating the impact of so-called "bottle effects" [209]. The change in abundance during the control experiment (x-axis) and the treatment experiment (y-axis) is positive if the OTU increased in abundance and negative if the OTU decreased in abundance. The taxa that appear furthest from the x=y diagonal line (shifted toward the upper left of the plot) responded strongly to the experimental condition and are thus the most likely of the observed taxa to be involved in oil degradation (Figure 4.4, Figure 4.3, Figure 4.5). TEXMEX revealed that *Oceanospirillales*, which includes known oil degraders, responded to oil amendment in both oxic and anoxic microcosm experiments, although a greater diversity within this order responded in oxic microcosms [118]. The *Oceanospirillales* that responded in both oxic and anoxic microcosms may be facultative anaerobes or aerotolerant, while those responding only in oxic conditions may be facultative anaerobes or obligate aerobes. The order *Oceanospirillales* responded strongly to oil amendment in both oxic and anoxic microcosms and were classified as *Oceanospirillaceae Oleispira*. Notably, a greater diversity of *Oceanospirillales* was observed responding to oil amendment in oxic microcosms, including some that were not able to be classified below the family level. In anoxic microcosms, the *Alphaproteobacterial* order *Sphingomonadales* responded strongly to oil amendment, which is consistent with reports that members of this order are hydrocarbon degraders [25]. Members of *Sphingomonadales* are known to occupy aerobic waters; however, those identified in this study were not able to be classified below the order level and may represent novel taxa capable of hydrocarbon degradation in anoxic environments [8]. *Alphaproteobacteria* were also enriched in oxic microcosms. Taxa of the class *Alphaproteobacteria* responded to oil amendment in both oxic and anoxic conditions as well, but different orders responded depending on the atmospheric condition of the microcosm. In anoxic microcosms, the order *Sphingomonadales* responded strongly, but were not able to be classified below the order level. Other than the *Oceanospirillales*, *Sphingomonadales* were the only microbes to respond in the anoxic condition. In oxic microcosms, the orders *Kiloniellales*, *Rhizobiales*, and *Rhodobacterales* responded strongly to oil amendment. *Kiloniellas* were not classified below the order level. Although many members of *Kiloniellas* are host associated, , one member of this order was isolated and reported as a chemoheterotrophic aerobe that may be involved in denitrification

Table 4.2: Summary of environmental parameters by depth category. Environmental parameters for shallow ($\leq$ 50 m) and deep ($>$ 50 m) waters were tested for statistically different means using a two-tailed t-test. Statistically significant environmental parameters (p-value < 0.05) are indicated by *.

	Shallow ($\leq$ 50 m, $\mu \pm \sigma^2$)	Deep ($>$ 50 m, $\mu \pm \sigma^2$)
Temperature *	20.239 (5.097)	7.240 (1.362)
Dissolved oxygen (mg/L) *	8.381 (0.170)	4.278 (2.069)
Salinity (PSU) *	11.245 (0.055)	11.327 (0.042)
Nitrate (mol/L) *	0.152 (0.053)	6.441 (4.138)
Nitrite (mol/L)	0.042 (0.034)	0.028 (0.010)
Ammonia (mol/L)	0.024 (0.010)	0.021 (0.003)
Total nitrogen (mg/L) *	0.397 (0.033)	0.440 (0.041)
Inorganic phosphate (mol/L) *	0.026 (0.008)	0.606 (0.569)
Silicate (mol/L) *	2.081 (1.626)	36.450 (23.317)
TOC (mg/L)	7.169 (0.784)	6.820 (0.463)

[56, 168, 269, 308]. The *Rhizobiales* were classified as *Methylobacteriaceae Methylobacterium*, and the *Rhodobacterales* were classified as *Rhodobacteraceae Phaeobacter*. Members of the genus *Methylobacterium* also responded to oil amendment. This genus is widespread and is the model for aerobic methylotrophy [154, 202]. The genus *Phaeobacter* of the order *Rhodobacterales* also responded to oil amendment, which is consistent with previous reports [49]. Commonly reported members of *Phaeobacter* are marine heterotrophs capable of growing in a wide range of nitrogen and phosphorous concentrations and a wide range of temperatures [170, 287, 308]. Many of the reads observed in the microcosm communities were not able to be classified, even at high taxonomic levels. Many of these uncharacterized taxa are oil degraders, based on the enrichments observed in this study. These reads represent potentially novel organisms that may be amenable to growth in laboratory conditions, either in isolation or as a consortium, and should be investigated further.

Hydrocarbon degradation was different between oxic and anoxic microcosms. Total aliphatic and aromatic hydrocarbons at days 3 and 17 were quantified relative to a standard (1,3-dichlorobenzene; Figure 4.6). At day 3, both aliphatic and aromatic hydrocarbon degradation were similar in oxic and anoxic conditions. However, at day 17, anoxic microcosms degraded a larger fraction of total hydrocarbons compared to oxic, with a striking increase in aromatic hydrocarbon degradation. In particular, benz(a)anthracene and chrysene were completely depleted in anoxic microcosm experiments. Naphthalene was also depleted similarly in oxic microcosms. Degradation of aliphatics decreased inversely with size up to ~22 carbons in both oxic and anoxic conditions. Anoxic microcosms degraded more of these aliphatic hydrocarbons. Aliphatic hydrocarbons of length 19 and 21 carbons were poorly degraded in oxic conditions relative to anoxic conditions. On the other hand, the heaviest aliphatics (> 28) were degraded to a greater extent in oxic microcosms compared to anoxic microcosms. Pristane, norpristane, and phytane were also depleted to a greater extent in anoxic microcosms, with only a minor fraction degraded in oxic microcosms. Trimethyldodecane and trimethyltridecane were similarly lost from both oxic and anoxic conditions. Aromatic hydrocarbons were degraded to a lesser extent than aliphatic hydrocarbons. This was particularly noticeable under oxic conditions. Both oxic and anoxic microcosms

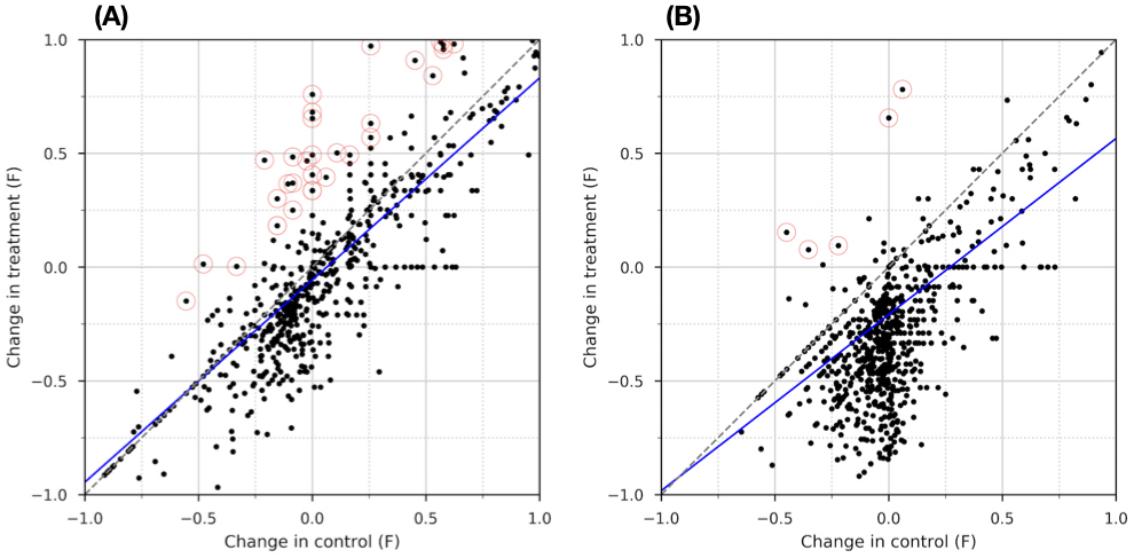


Figure 4.5: TEXMEX plots.

Table 4.3: List of taxa that were enriched in an anoxic, oil amended microcosm as determined by TEXMEX analysis. Changes in the microbial community composition of the anoxic, oil-amended microcosm were compared with that of the anoxic, non-amended microcosm. Microbial taxa that were enriched in the oil-amended microcosm relative to the non-amended microcosm are listed here.

Class	Order	Family	Genus
Bacteroidia	*Bacteroidales*	*Porphyromonadaceae*	
Gammaproteobacteria	*Oceanospirillales*		
Gammaproteobacteria	*Oceanospirillales*	*Oceanospirillaceae*	
Gammaproteobacteria	*Oceanospirillales*	*Oceanospirillaceae*	*Oleispira*

Table 4.4: List of taxa that were enriched in an oxic, oil amended microcosm as determined by TEXMEX analysis. Changes in the microbial community composition of the oxic, oil-amended microcosm were compared with that of the oxic, non-amended microcosm. Microbial taxa that were enriched in the oil-amended microcosm relative to the non-amended microcosm are listed here.

Class	Order	Family	Genus
Actinobacteria	*Actinomycetales*	*Frankiaceae*	
Actinobacteria	*Actinomycetales*	*Intrasporangiaceae*	*Phycicoccus*
Actinobacteria	*Actinomycetales*	*Intrasporangiaceae*	*Terracoccus*
Bacteroidia	*Bacteroidales*		
Cytophagia	*Cytophagales*	*Cytophagaceae*	*Flectobacillus*
Flavobacteriia	*Flavobacteriales*	*Flavobacteriaceae*	*Ulvibacter*
Nostocophycideae	*Nostocales*		
Clostridia	*Clostridiales*	*Eubacteriaceae*	*Acetobacterium*
Clostridia	*Clostridiales*	*Veillonellaceae*	
Alphaproteobacteria	*Kiloniellales*		
Alphaproteobacteria	*Kiloniellales*	*Kiloniellaceae*	
Alphaproteobacteria	*Kordiimonadales*	*Kordiimonadaceae*	
Alphaproteobacteria	*Rhizobiales*	*Bradyrhizobiaceae*	*Bosea*
Alphaproteobacteria	*Rhizobiales*	*Hyphomicrobiaceae*	*Parvibaculum*
Alphaproteobacteria	*Rhodobacterales*	*Hyphomonadaceae*	*Hyphomonas*
Alphaproteobacteria	*Rhodobacterales*	*Rhodobacteraceae*	*Phaeobacter*
Alphaproteobacteria	*Sphingomonadales*	*Erythrobacteraceae*	*Erythrobacter*
Epsilonproteobacteria			
Epsilonproteobacteria	*Campylobacterales*		
Epsilonproteobacteria	*Campylobacterales*	*Campylobacteraceae*	
Epsilonproteobacteria	*Campylobacterales*	*Campylobacteraceae*	*Sulfurospirillum*
Epsilonproteobacteria	*Campylobacterales*	*Helicobacteraceae*	*Sulfuricurvum*
Gammaproteobacteria	*Alteromonadales*	*Alteromonadaceae*	*Glaciecola*
Gammaproteobacteria	*Alteromonadales*	*Alteromonadaceae*	*Simiduia*
Gammaproteobacteria	*Alteromonadales*	*Colwelliaceae*	*Thalassomonas*
Gammaproteobacteria	*Oceanospirillales*	*Oceanospirillaceae*	
Gammaproteobacteria	*Oceanospirillales*	*Oceanospirillaceae*	*Amphritea*
Gammaproteobacteria	*Oceanospirillales*	*Oceanospirillaceae*	*Neptunomonas*
Gammaproteobacteria	*Oceanospirillales*	*Oceanospirillaceae*	*Oceaniserpentilla*
Gammaproteobacteria	*Oceanospirillales*	*Oceanospirillaceae*	*Oleispira*
Gammaproteobacteria	*Pseudomonadales*	*Pseudomonadaceae*	

degraded similar proportions of 1,3-dimethylnapthalene, 1,6,7-trimethylapthalene, 1,6-dimethylnapthalene, 1-methylnaphthalene, 2,6-dimethylnaphthalene, 2-methylnaphthalene, biphenyl, and benzothiophenes. Oxic microcosms degraded phenanthrene more efficiently than anoxic microcosms. In addition to the EPA16 compounds already mentioned, higher proportions of dibenzothiophene were degraded in anoxic microcosms.

Although the oxic and anoxic microcosms degraded similar amounts of hydrocarbons by day 3, there were important differences in hydrocarbon degradation by day 17, which may be attributed to the changes in the microcosm communities observed during the same period (Figure 4.7). Interestingly, anoxic microcosms degraded more total oil hydrocarbons by day 17 compared to oxic microcosms. The half-life for total hydrocarbon degradation was estimated to be 11 days in anoxic microcosms and 15 days in oxic microcosms. Most of the individual oil hydrocarbons were degraded more efficiently in anoxic microcosms. Anoxic microcosms degraded a larger proportion of the shorter aliphatics ($<$ 22 carbons; including branched aliphatics, e.g., pristane and phytane) than oxic microcosms. Conversely, oxic microcosms more completely degraded most of the longer aliphatics ($\geq$ 22 carbons). Anoxic microcosms were similar to or exceeded oxic microcosms in the degradation of all aromatic hydrocarbons. In particular, degradation of benz(a)anthracene, chrysene, dibenzothiophene, fluorene, and phenanthrene in anoxic microcosms far exceeded that of oxic microcosms. Phytane, pristane, and hopane are used as analytic standards when attempting to determine the origin of crude oil and its rate of biodegradation. Pristane and phytane are common components of crude oil, and the pristane/phytane ratio may be used as an indicator of the oil source [231, 254]. However, pristane and phytane are known to be susceptible to biodegradation, which was also observed in this study [102]. Hopane is considered less susceptible to biodegradation than phytane and pristane [87, 102]. However, some biodegradation of hopane was also observed in this study. This finding has important implications for the analysis of crude oil biodegradation in the Caspian Sea, and elsewhere. Hydrocarbons may be degraded more efficiently at low temperatures compared to warmer temperatures, which could explain high hydrocarbon degradation in the deep waters of the Caspian. However, improved hydrocarbon degradation under anoxic (compared to oxic) atmospheric headspace, as observed and reported here, is not

well understood. Microogranisms from this environment may have adapted by developing efficient methods for degrading recalcitrant hydrocarbons. When the communities are exposed to oxygen, those biodegradation pathways may be inhibited, or aerobic organisms may grow rapidly and dominate the community, limiting overall oil biodegradation. On the other hand, these effects would be limited under anoxic or hypoxic conditions. These results suggest that deep water communities in the Caspian Sea are adapted for better oil biodegradation under anoxic and (*in situ* hypoxic) conditions when compared to oxic conditions. Potential oil degradation pathways have been identified in *Gammaproteobacteria*, including pathways for the degradation of low molecular weight alkanes, aromatics, BTEX, proline, catechol, cyclohexanone, and nitroalkanes [130]. It is therefore reasonable to expect that the *Gammaproteobacteria* observed in these results are important in the biodegradation results observed here, which is consistent with other reports [20].

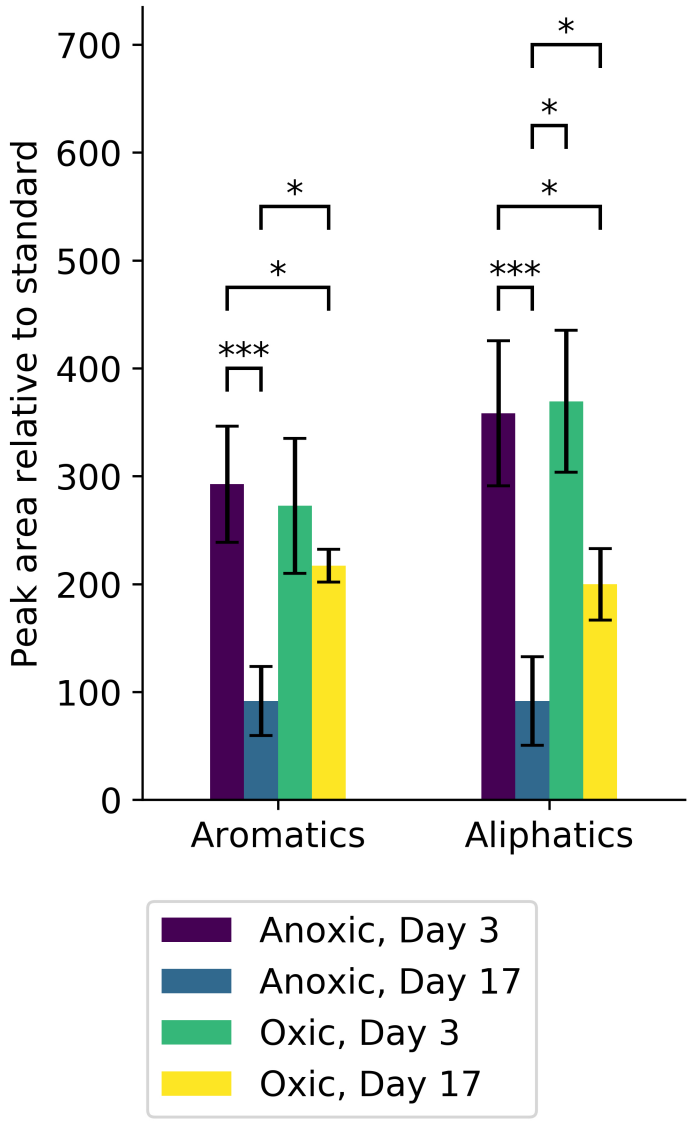


Figure 4.6: Hydrocarbons from microcosm experiments were quantified and are reported according to type (aromatic or aliphatic). Significant differences in the amount of each hydrocarbon type are indicated (alpha $<$ 0.05, alpha $<$ 0.001). Anoxic microcosm communities (purple, blue) degraded significant amounts of both aromatic and aliphatic hydrocarbons, whereas oxic microcosm communities (green, yellow) did not. At day 17, the amount of both aromatic and aliphatic hydrocarbons was significantly less in anoxic microcosms (blue) compared to oxic microcosms (yellow).

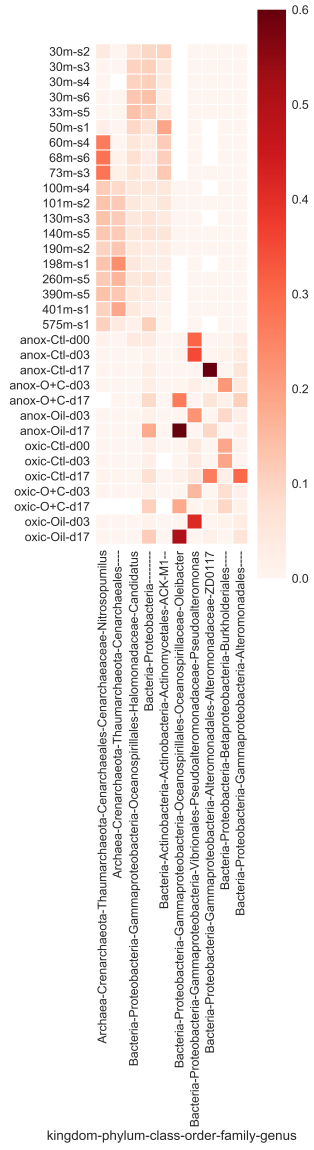


Figure 4.7: The most abundant microbial taxa that were enriched in either the natural or microcosm communities. Relative abundance of each taxon is represented by intensity of the heat map. Enriched taxa were determined using Fisher's exact test.

Chapter 5

Conclusions and future directions

5.1 Research summary and final conclusions

The goal of the work presented in this dissertation was to advance our understanding of the influence of hypoxia on petroleum hydrocarbon biodegradation. In this work, microbial communities from six marine basins around the world were analyzed using multiple techniques providing new information about the potential for petroleum hydrocarbon biodegradation in these discrete environments. These communities were analyzed with both 16S rRNA gene amplicon sequencing to understand their composition and diversity and with shotgun sequencing to understand the influence of hypoxia on the potential for petroleum hydrocarbon biodegradation. Data from multiple basins was analyzed comprehensively in order to make general inferences about the effect of hypoxia on microbial communities.

Aim 1 demonstrated that although microbial communities from discrete basins have different composition, potential petroleum-degrading microbes are ubiquitous. In general, microbial communities tended to cluster more strongly by depth than by sample site, however, there were important differences between the basins. In particular, the microbial community from the Caspian Sea was unique among the basins, which is not surprising given that it has been landlocked for more than 5 million years. Regardless, it shares with the other basins petroleum-degrading bacteria such as *Oceanospirillales*. However, the communities from oxygen minimum zones (OMZs) in the Caspian Sea and in the Eastern Atlantic Ocean (Angola) were different in important ways. For example, Archaea were enriched in the Caspian Sea OMZ, while they appear to be less important in the Angola OMZ. Taken together, these results indicate that petroleum-degrading microorganism exist in microbial communities around the globe.

Aim 2 demonstrated that hypoxia is likely to be a stronger influence on microbial community functional potential than eutrophication in general. Microbial communities from the Eastern Mediterranean Sea and the Caspian Sea were analyzed using shotgun metagenomics to understand the functional potential. The abundance of enzymatic functions was similar across the two basins, but there were important correlations with hypoxia. In particular, heterotrophic denitrification is likely to be an important function in hypoxic environments. However, hypoxia (O_2 <3 mg/L) was only present in the Caspian Sea.

Previous reports indicate the diverse pathways for aromatic hydrocarbon biodegradation are present in sub-oxic environments, due to the varying reduction potential of alternative electron acceptors. Therefore, it seems likely that the mechanisms of petroleum hydrocarbon biodegradation might be influenced by the available oxygen concentration.

Aim 3 demonstrated that petroleum hydrocarbon biodegradation by Caspian Sea microbial communities is faster under anoxic conditions than under oxic conditions in microcosm experiments, which was unexpected. The Caspian Sea has been landlocked for >5 million years, and the sea floor may have the highest concentration of mud volcanoes leaking hydrocarbons. Furthermore, this basin has become increasingly eutrophic and hypoxic since the 1950s due to substantial industrial and agricultural runoff, as well as being an important site of petroleum production. These factors might have influenced the microbial communities in one of two ways. One possibility is that these communities have become enriched in petroleum-degrading microbes, which were capable of rapid petroleum biodegradation. Another possibility is that selective pressure has influenced petroleum-degrading microbes in these communities for more efficient petroleum biodegradation under hypoxic (or anoxic) conditions, without substantial change to the overall community structure. The work in this dissertation demonstrated that the microbial communities of the Caspian Sea are significantly different from communities in other basins; therefore current evidence supports the former hypothesis.

5.2 Future Directions

The DWH oil spill was a great disaster, but petroleum spills continue to occur around the world. Understanding how the in situ microbial community will respond will assist cleanup efforts in the event of a spill. Furthermore, bioremediation of petroleum hydrocarbons in contaminated environments is an attractive strategy. To date, applying fertilizers to contaminated environments is the most effective method to increase rates of biodegradation. However, there is still interest in developing strategies to influence the in situ microbial communities for biodegradation of petroleum hydrocarbons and other contaminants.

This work demonstrated the ability of microbial communities from the Caspian Sea to demonstrate petroleum hydrocarbons in microcosm experiments. However, there are important questions that remain open. Future work should investigate the specific genes and proteins involved in petroleum hydrocarbon biodegradation in microcosm experiments using metatranscriptomics. Although it is historically difficult, future work should continue to isolate microbes from environment samples because this is the only way to verify our hypotheses about the role of specific organisms in petroleum hydrocarbon biodegradation, such as *Oceanospirillales*. Extrapolating the results of microcosm experiments to in situ environments can be ineffective. More field studies are needed to understand petroleum hydrocarbon biodegradation in situ in contaminated environments. These studies might use radiocarbon (14°C) signatures from microbial biomass to determine which microbes are involved in biodegradation of petroleum hydrocarbons. Shotgun metagenomics and appropriate binning methods enable the assembly of whole genomes from metagenomes, and these data can be used to build a better reference database for analysis of these microbial communities.

Future microcosm experiments are needed to better understand how the Caspian Sea microbial communities degrade petroleum hydrocarbons in anoxic conditins. High resolution quantification of hydrocarbons is necessary if we are to understand the mechanisms of petroleum hydrocarbon biodegradation. This work should use shotgun metagenomics and binning, as mentioned above, to assemble whole genomes and build a reference database of the organisms that are contained within the microcosms. This work should also use metatranscriptomics to identify which genes are enriched following amendment of the microcosms with petroleum hydrocarbons. A time course of 16S rRNA amplicon sequences will also help to understand the succession of microorganisms in the microcosms following petroleum amendment. I have obtained additional samples from the Caspian Sea, and this experiment is currently underway.

Bibliography

[1] Limits of Oceans and Seas. URL https://www.iho.int/iho_pubs/standard/S-23/S-23_Ed3_1953_EN.pdf. 21

[2] KEGG BRITE: KEGG Orthology (KO). URL https://www.kegg.jp/kegg-bin/get_htext?query=ko00362&htext=ko00001.keg. 73

[3] EBI metagenomics (pipeline). URL https://www.ebi.ac.uk/metagenomics/pipelines/3.0;jsessionid=0F3D7BA1E245EACE814778D14DFABA55. 00000. xi, 56

[4] Python Software Foundation. URL https://www.python.org/. 00000. 55, 58

[5] Raw Incident Data | NOAA. URL https://incidentnews.noaa.gov/raw/index. 2, 4

[6] Scikit-bio. URL https://github.com/biocore/scikit-bio. 58

[7] Ole Morten Aamo, Mark Reed, and Keith Downing. Oil spill contingency and response (oscar) model system: Sensitivity studies. 1997(1):429–438. ISSN 2169-3366. doi: 10.7901/2169-3358-1997-1-429. 53

[8] Firouz Abbasian, Robin Lockington, Megharaj Mallavarapu, and Ravi Naidu. A pyrosequencing-based analysis of microbial diversity governed by ecological conditions in the Winogradsky column. 31(7):1115–1126. ISSN 1573-0972. doi: 10.1007/s11274-015-1861-y. 103, 107

[9] Hélène Agogué, Dominique Lamy, Phillip R. Neal, Mitchell L. Sogin, and Gerhard J. Herndl. Water mass-specificity of bacterial communities in the North Atlantic revealed by massively parallel sequencing. 20(2):258–274. ISSN 1365-294X. doi: 10.1111/j.1365-294X.2010.04932.x. 44, 46

[10] Daniel Aird, Michael G. Ross, Wei-Sheng Chen, Maxwell Danielsson, Timothy Fennell, Carsten Russ, David B. Jaffe, Chad Nusbaum, and Andreas Gnirke. Analyzing and minimizing PCR amplification bias in Illumina sequencing libraries. 12(2):R18. ISSN 1474-760X. doi: 10.1186/gb-2011-12-2-r18. 6

[11] Nelson Alves Junior, Pedro Milet Meirelles, Eidy de Oliveira Santos, Bas Dutilh, Genivaldo G. Z. Silva, Rodolfo Paranhos, Anderson S. Cabral, Carlos Rezende, Tetsuya Iida, Rodrigo L. de Moura, Ricardo Henrique Kruger, Renato C. Pereira, Rogério Valle, Tomoo Sawabe, Cristiane Thompson, and Fabiano Thompson. Microbial community diversity and physical–chemical features of the Southwestern Atlantic Ocean. 197(2): 165–179. ISSN 0302-8933, 1432-072X. doi: 10.1007/s00203-014-1035-6. 44, 46

[12] Donald M. Anderson, Patricia M. Glibert, and Joann M. Burkholder. Harmful algal blooms and eutrophication: Nutrient sources, composition, and consequences. 25(4): 704–726. ISSN 0160-8347. doi: 10.1007/BF02804901. 52

[13] Marti J. Anderson. A new method for non-parametric multivariate analysis of variance. 26(1):32–46. ISSN 1442-9993. doi: 10.1111/j.1442-9993.2001.01070.pp.x. 05719. 61

[14] Simon Andrews. FastQC: A Quality Control tool for High Throughput Sequence Data. URL http://www.bioinformatics.babraham.ac.uk/projects/fastqc/. 00003. 55

[15] Rolf Apweiler, Amos Bairoch, Cathy H. Wu, Winona C. Barker, Brigitte Boeckmann, Serenella Ferro, Elisabeth Gasteiger, Hongzhan Huang, Rodrigo Lopez, Michele Magrane, Maria J. Martin, Darren A. Natale, Claire O'Donovan, Nicole Redaschi, and Lai-Su L. Yeh. UniProt: The Universal Protein knowledgebase. 32:D115–D119. ISSN 0305-1048. doi: 10.1093/nar/gkh131. 58

[16] Erik Aronesty. Ea-utils: Command-line tools for processing biological sequencing data. URL https://expressionanalysis.github.io/ea-utils/. 30

[17] R M Atlas. Microbial degradation of petroleum hydrocarbons: An environmental perspective. 45(1):180–209. ISSN 0146-0749. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC281502/. 12

[18] Ronald Atlas and James Bragg. Bioremediation of marine oil spills: When and when not – the Exxon Valdez experience. 2(2):213–221. ISSN 1751-7915. doi: 10.1111/j.1751-7915.2008.00079.x. 6

[19] Ronald M. Atlas and Terry C. Hazen. Oil Biodegradation and Bioremediation: A Tale of the Two Worst Spills in U.S. History. 45(16):6709–6715. ISSN 0013-936X. doi: 10.1021/es2013227. 00280. 2, 4

[20] Jacob Bælum, Sharon Borglin, Romy Chakraborty, Julian L. Fortney, Regina Lamendella, Olivia U. Mason, Manfred Auer, Marcin Zemla, Markus Bill, Mark E. Conrad, Stephanie A. Malfatti, Susannah G. Tringe, Hoi-Ying Holman, Terry C. Hazen, and Janet K. Jansson. Deep-sea bacteria enriched by oil and dispersant from the Deepwater Horizon spill. 14(9):2405–2416. ISSN 1462-2920. doi: 10.1111/j.1462-2920.2012.02780.x. 93, 97, 101, 114

[21] Selina M Bamforth and Ian Singleton. Bioremediation of polycyclic aromatic hydrocarbons: Current knowledge and future directions. 80(7):723–736. ISSN 0268-2575, 1097-4660. doi: 10.1002/jctb.1276. 4, 12

[22] Anton Bankevich, Sergey Nurk, Dmitry Antipov, Alexey A. Gurevich, Mikhail Dvorkin, Alexander S. Kulikov, Valery M. Lesin, Sergey I. Nikolenko, Son Pham, Andrey D. Prjibelski, Alexey V. Pyshkin, Alexander V. Sirotkin, Nikolay Vyahhi, Glenn Tesler, Max A. Alekseyev, and Pavel A. Pevzner. SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. 19(5):455–477. ISSN 1557-8666. doi: 10.1089/cmb.2012.0021. 57

[23] E. Bar-Zeev, T. Berman, E. Rahav, G. Dishon, B. Herut, N. Kress, and I. Berman-Frank. Transparent exopolymer particle (TEP) dynamics in the eastern Mediterranean Sea. 431:107–118. ISSN 0171-8630, 1616-1599. doi: 10.3354/meps09110. 22

[24] Mary L. Batteen and Henry A. Miller. Process-oriented modeling studies of the 5500-km-long boundary flow off western and southern Australia. 29(4):702–718. ISSN 0278-4343. doi: 10.1016/j.csr.2008.11.011. 27

[25] Melanie J. Beazley, Robert J. Martinez, Suja Rajan, Jessica Powell, Yvette M. Piceno, Lauren M. Tom, Gary L. Andersen, Terry C. Hazen, Joy D. Van Nostrand, Jizhong Zhou, Behzad Mortazavi, and Patricia A. Sobecky. Microbial Community Analysis of

a Coastal Salt Marsh Affected by the Deepwater Horizon Oil Spill. 7(7):e41305. ISSN 1932-6203. doi: 10.1371/journal.pone.0041305. 107

[26] Zvi Ben-Avraham, George Smith, Moshe Reshef, and Eric Jungslager. Gas hydrate and mud volcanoes on the southwest African continental margin off South Africa. 30 (10):927–930. ISSN 0091-7613, 1943-2682. doi: 10.1130/0091-7613(2002)030<0927:GHAMVO>2.0.CO;2. 00038. 20

[27] Carolina Berdugo-Clavijo and Lisa M. Gieg. Conversion of crude oil to methane by a microbial consortium enriched from oil reservoir production waters. 5. ISSN 1664-302X. doi: 10.3389/fmicb.2014.00197. 93

[28] W. H. Berger, G. Wefer, C. Richter, C. B. Lange, J. Giraudeau, O. Hermelin, and Shipboard Scientific Party. The Angola-Benguela Upwelling System: Paleoceanographic Synthesis Of Shipboard Results From Leg 175. In *Proceedings Ocean Drilling Program, Initial Reports*, volume 175, pages 505–531. URL http://www.iodp.tamu.edu/publications/175_IR/VOLUME/CHAPTERS/CHAP_17.PDF. 00039. 19

[29] Martina Blümel, Jörg Süling, and Johannes F. Imhoff. Depth-specific distribution of Bacteroidetes in the oligotrophic Eastern Mediterranean Sea. 46(3):209–224. ISSN 0948-3055, 1616-1564. doi: 10.3354/ame046209. 22

[30] Anthony M. Bolger, Marc Lohse, and Bjoern Usadel. Trimmomatic: A flexible trimmer for Illumina sequence data. 30(15):2114–2120. ISSN 1367-4803, 1460-2059. doi: 10.1093/bioinformatics/btu170. 00096. 55

[31] S. Boonchan, M. L. Britz, and G. A. Stanley. Degradation and Mineralization of High-Molecular-Weight Polycyclic Aromatic Hydrocarbons by Defined Fungal-Bacterial Cocultures. 66(3):1007–1019. ISSN 0099-2240, 1098-5336. doi: 10.1128/AEM.66.3.1007-1019.2000. 12

[32] A. B. A. Boxall and L. Maltby. The Effects of Motorway Runoff on Freshwater Ecosystems: 3. Toxicant Confirmation. 33(1):9–16. ISSN 0090-4341, 1432-0703. doi: 10.1007/s002449900216. 00101. 15

[33] Odd G. Brakstad, Per S. Daling, Liv-G. Faksness, Inger K. Almås, Siv-H. Vang, Line Syslak, and Frode Leirvik. Depletion and biodegradation of hydrocarbons in dispersions and emulsions of the Macondo 252 oil generated in an oil-on-seawater mesocosm flume basin. 84(1):125–134. ISSN 0025-326X. doi: 10.1016/j.marpolbul.2014.05.027. 92, 93

[34] J. Roger Bray and J. T. Curtis. An Ordination of the Upland Forest Communities of Southern Wisconsin. 27(4):325–349. ISSN 1557-7015. doi: 10.2307/1942268. 07022. 31, 60

[35] Florian P. Breitwieser, Jennifer Lu, and Steven L. Salzberg. A review of methods and databases for metagenomic classification and assembly. ISSN 1467-5463, 1477-4054. doi: 10.1093/bib/bbx120. 66

[36] Volker Brüchert, Bronwen Currie, and Kathleen R. Peard. Hydrogen sulphide and methane emissions on the central Namibian shelf. 83(1):169–179, . ISSN 0079-6611. doi: 10.1016/j.pocean.2009.07.017. 00035. 20

[37] Volker Brüchert, Bo Barker Jørgensen, Kirsten Neumann, Daniela Riechmann, Manfred Schlösser, and Heide Schulz. Regulation of bacterial sulfate reduction and hydrogen sulfide fluxes in the central namibian coastal upwelling zone. 67(23):4505–4518, . ISSN 0016-7037. doi: 10.1016/S0016-7037(03)00275-8. 00139. 20

[38] Kathryn A Burns, Paul F Greenwood, Roger E Summons, and Gregg J Brunskill. Vertical fluxes of hydrocarbons on the Northwest Shelf of Australia as estimated by a sediment trap study. 32(10):1241–1255. ISSN 0146-6380. doi: 10.1016/S0146-6380(01)00088-2. 27

[39] Christiam Camacho, George Coulouris, Vahram Avagyan, Ning Ma, Jason Papadopoulos, Kevin Bealer, and Thomas L. Madden. BLAST+: Architecture and applications. 10(1):421. ISSN 1471-2105. doi: 10.1186/1471-2105-10-421. 57

[40] Mariana E. Campeão, Luciana Reis, Luciana Leomil, Louisi de Oliveira, Koko Otsuki, Piero Gardinali, Oliver Pelz, Rogerio Valle, Fabiano L. Thompson, and Cristiane C.

Thompson. The Deep-Sea Microbial Community from the Amazonian Basin Associated with Oil Degradation. 8:1019. ISSN 1664-302X. doi: 10.3389/fmicb.2017.01019. 21

[41] Brandi L. Cantarel, Pedro M. Coutinho, Corinne Rancurel, Thomas Bernard, Vincent Lombard, and Bernard Henrissat. The Carbohydrate-Active EnZymes database (CAZy): An expert resource for Glycogenomics. 37:D233–D238. ISSN 0305-1048, 1362-4962. doi: 10.1093/nar/gkn663. 57

[42] J. Gregory Caporaso, Justin Kuczynski, Jesse Stombaugh, Kyle Bittinger, Frederic D. Bushman, Elizabeth K. Costello, Noah Fierer, Antonio Gonzalez Peña, Julia K. Goodrich, Jeffrey I. Gordon, Gavin A. Huttley, Scott T. Kelley, Dan Knights, Jeremy E. Koenig, Ruth E. Ley, Catherine A. Lozupone, Daniel McDonald, Brian D. Muegge, Meg Pirrung, Jens Reeder, Joel R. Sevinsky, Peter J. Turnbaugh, William A. Walters, Jeremy Widmann, Tanya Yatsunenko, Jesse Zaneveld, and Rob Knight. QIIME allows analysis of high-throughput community sequencing data. 7(5):335–336, . ISSN 1548-7105. doi: 10.1038/nmeth.f.303. 30

[43] J Gregory Caporaso, Christian L Lauber, William A Walters, Donna Berg-Lyons, James Huntley, Noah Fierer, Sarah M Owens, Jason Betley, Louise Fraser, Markus Bauer, Niall Gormley, Jack A Gilbert, Geoff Smith, and Rob Knight. Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. 6(8):1621–1624, . ISSN 1751-7362. doi: 10.1038/ismej.2012.8. 55

[44] Manuel Carmona, María Teresa Zamarro, Blas Blázquez, Gonzalo Durante-Rodríguez, Javier F. Juárez, J. Andrés Valderrama, María J. L. Barragán, José Luis García, and Eduardo Díaz. Anaerobic Catabolism of Aromatic Compounds: A Genetic and Genomic View. 73(1):71–133. ISSN 1092-2172, 1098-5557. doi: 10.1128/MMBR.00021-08. 00184. 53

[45] Mary-Elena Carr. Estimation of potential productivity in Eastern Boundary Currents using remote sensing. 49(1):59–80. ISSN 0967-0645. doi: 10.1016/S0967-0645(01)00094-7. 00369. 20

[46] E. L. Cavan, M. Trimmer, F. Shelley, and R. Sanders. Remineralization of particulate organic carbon in an ocean oxygen minimum zone. 8:14847. ISSN 2041-1723. doi: 10.1038/ncomms14847. 15

[47] C E Cerniglia. Fungal metabolism of polycyclic aromatic hydrocarbons: Past, present and future applications in bioremediation. 19(5):324–333, . ISSN 1476-5535. doi: 10.1038/sj.jim.2900459. 12

[48] Carl E. Cerniglia. Biodegradation of polycyclic aromatic hydrocarbons. 4(3):331–338, . ISSN 0958-1669. doi: 10.1016/0958-1669(93)90104-5. vii, ix, 4, 5, 7, 12, 13, 14

[49] Romy Chakraborty, Sharon E. Borglin, Eric A. Dubinsky, Gary L. Andersen, and Terry C. Hazen. Microbial Response to the MC-252 Oil and Corexit 9500 in the Gulf of Mexico. 3:357. ISSN 1664-302X. doi: 10.3389/fmicb.2012.00357. 109

[50] Anne Chao. Nonparametric estimation of the number of classes in a population. pages 265–270. 02046. 31, 38, 68

[51] Anne Chao, Robert K. Colwell, Chih-Wei Lin, and Nicholas J. Gotelli. Sufficient sampling for asymptotic minimum species richness estimators. 90(4):1125–1133. ISSN 1939-9170. doi: 10.1890/07-2147.1. 00200. 59

[52] B Chávez-Gómez, R Quintero, F Esparza-Garcıá, A. M Mesta-Howard, F. J Zavala Dıáz de la Serna, C. H Hernández-Rodrıǵuez, T Gillén, H. M Poggi-Varaldo, J Barrera-Cortés, and R Rodrıǵuez-Vázquez. Removal of phenanthrene from soil by co-cultures of bacteria and fungi pregrown on sugarcane bagasse pith. 89(2):177–183. ISSN 0960-8524. doi: 10.1016/S0960-8524(03)00037-3. 12

[53] Z Chekroud, MK Gouda, and M Houhamdi. Biodegradation of crude oil in marine medium. 4:231–237. 24, 25

[54] O. V. Chicherina, A. V. Leonov, and D. Ya Fashchuk. Geographical and ecological characteristics of the Caspian Sea and modern tendencies in the evolution of its ecosystem. 31(3):271–289. URL http://www.springerlink.com/index/rt2514428961w4j0.pdf. 00003. 26, 53, 91, 92, 103

[55] K. R. Clarke. Non-parametric multivariate analyses of changes in community structure. 18(1):117–143. ISSN 0307-692X. doi: 10.1111/j.1442-9993.1993.tb00438.x. 61

[56] Daniel F. R. Cleary, Leontine E. Becking, Ana R. M. Polónia, Rossana M. Freitas, and Newton C. M. Gomes. Jellyfish-associated bacterial communities and bacterioplankton in Indonesian Marine lakes. 92(5):fiw064. ISSN 1574-6941. doi: 10.1093/femsec/fiw064. 109

[57] James E. Cloern. Our evolving conceptual model of the coastal eutrophication problem. 210:223–253. ISSN 0171-8630, 1616-1599. doi: 10.3354/meps210223. 52

[58] John D. Coates, Robert T. Anderson, Joan C. Woodward, Elizabeth J. P. Phillips, and Derek R. Lovley. Anaerobic Hydrocarbon Degradation in Petroleum-Contaminated Harbor Sediments under Sulfate-Reducing and Artificially Imposed Iron-Reducing Conditions. 30(9):2784–2789. ISSN 0013-936X. doi: 10.1021/es9600441. 10, 92

[59] Committee on Oil in the Sea: Inputs, Fates, and Effects, Ocean Studies Board, Marine Board, Transportation Research Board, Division on Earth and Life Studies, and National Research Council. *Oil in the Sea III: Inputs, Fates, and Effects*. National Academies Press. ISBN 978-0-309-08438-3. doi: 10.17226/10388. URL https://www.nap.edu/catalog/10388. 3, 6, 10

[60] Manuela da Silva, Carl E. Cerniglia, Jairaj V. Pothuluri, Vanderlei P. Canhos, and Elisa Esposito. Screening filamentous fungi isolated from estuarine sediments for the ability to oxidize polycyclic aromatic hydrocarbons. 19(4):399–405. ISSN 1573-0972. doi: 10.1023/A:1023994618879. 4

[61] Hongyue Dang, Haixia Zhou, Jinying Yang, Huangmin Ge, Nianzhi Jiao, Xiwu Luan, Chuanlun Zhang, and Martin G. Klotz. Thaumarchaeotal signature gene distribution in sediments of the northern South China Sea: An indicator of the metabolic intersection of the marine carbon, nitrogen, and phosphorus cycles? 79(7):2137–2147. ISSN 1098-5336. doi: 10.1128/AEM.03204-12. 15

[62] Nilanjana Das and Preethy Chandran. Microbial Degradation of Petroleum Hydrocarbon Contaminants: An Overview. 2011:1–13. ISSN 2090-3138, 2090-3146. doi: 10.4061/2011/941810. ix, 9, 12

[63] G. J. Denoux, P. Gardinali, and T. L. Wade. Quantitative determination of polynuclear aromatic hydrocarbons by gas chromatography/mass spectrometry (GC/MS)-selected ion monitoring (SIM) mode. page 129. 98

[64] T. Z. DeSantis, P. Hugenholtz, N. Larsen, M. Rojas, E. L. Brodie, K. Keller, T. Huber, D. Dalevi, P. Hu, and G. L. Andersen. Greengenes, a Chimera-Checked 16S rRNA Gene Database and Workbench Compatible with ARB. 72(7):5069–5072. ISSN 0099-2240. doi: 10.1128/AEM.03006-05. 02902. 99

[65] Eduardo Díaz, José Ignacio Jiménez, and Juan Nogales. Aerobic degradation of aromatic compounds. 24(3):431–442. ISSN 0958-1669. doi: 10.1016/j.copbio.2012.10.010. 73

[66] R. J. Diaz and R. Rosenberg. Marine benthic hypoxia: A review of its ecological effects and the behavioural responses of benthic macrofauna. 12(43):1250. ISSN 0967-0653. URL https://www.infona.pl//resource/bwmeta1.element.elsevier-35045107-6c73-3db5-b70c-beebf1cc124c. 54

[67] US DOE-OIT. Energy and Environmental Profile of the US Chemical Industry. 76

[68] Peter E. T. Douben. *PAHs: An Ecotoxicological Perspective*. John Wiley & Sons. ISBN 978-0-470-86429-6. 4

[69] Márcia Duarte, Ruy Jauregui, Ramiro Vilchez-Vargas, Howard Junca, and Dietmar H. Pieper. AromaDeg, a novel database for phylogenomics of aerobic bacterial degradation of aromatics. 2014:bau118. ISSN , 1758-0463. doi: 10.1093/database/bau118. 00015. 57

[70] Eric A. Dubinsky, Mark E. Conrad, Romy Chakraborty, Markus Bill, Sharon E. Borglin, James T. Hollibaugh, Olivia U. Mason, Yvette M. Piceno, Francine C. Reid,

William T. Stringfellow, Lauren M. Tom, Terry C. Hazen, and Gary L. Andersen. Succession of Hydrocarbon-Degrading Bacteria in the Aftermath of the Deepwater Horizon Oil Spill in the Gulf of Mexico. 47(19):10860–10867. ISSN 0013-936X. doi: 10.1021/es401676y. 6, 10

[71] H. J. Dumont. The Caspian Lake: History, biota, structure, and function. 43(1):44–52. ISSN 00243590. doi: 10.4319/lo.1998.43.1.0044. 91

[72] Stéphanie Dupré, John Woodside, Ingo Klaucke, Jean Mascle, and Jean-Paul Foucher. Widespread active seepage activity on the Nile Deep Sea Fan (offshore Egypt) revealed by high-definition geophysical imagery. 275(1):1–19. ISSN 0025-3227. doi: 10.1016/j.margeo.2010.04.003. 24

[73] S. R. Eddy. Accelerated Profile HMM Searches. 7(10):e1002195. ISSN 1553-7358. doi: 10.1371/journal.pcbi.1002195. 58

[74] Robert C. Edgar. Search and clustering orders of magnitude faster than BLAST. 26(19):2460–2461. ISSN 1367-4803, 1460-2059. doi: 10.1093/bioinformatics/btq461. 00597. 30, 99

[75] Robert C. Edgar, Brian J. Haas, Jose C. Clemente, Christopher Quince, and Rob Knight. UCHIME improves sensitivity and speed of chimera detection. 27(16):2194–2200. ISSN 1367-4803. doi: 10.1093/bioinformatics/btr381. 30, 99

[76] W.J. Emery. Water Types And Water Masses. In *Encyclopedia of Ocean Sciences*, pages 3179–3187. Elsevier. ISBN 978-0-12-227430-5. doi: 10.1006/rwos.2001.0108. URL http://linkinghub.elsevier.com/retrieve/pii/B012227430X001082. 00058. 19

[77] Soha Farag and Nadia A. Soliman. Biodegradation of crude petroleum oil and environmental pollutants by Candida tropicalis strain. 54(4):821–830. ISSN 1516-8913. doi: 10.1590/S1516-89132011000400023. 24

[78] Roi Feingersch, Marcelino T. Suzuki, Michael Shmoish, Itai Sharon, Gazalah Sabehi, Frédéric Partensky, and Oded Béjà. Microbial community genomics in eastern

Mediterranean Sea surface waters. 4(1):78–87. ISSN 1751-7370. doi: 10.1038/ismej.2009.92. 24, 25

[79] Manuel Ferrer, Olga V. Golyshina, Tatyana N. Chernikova, Amit N. Khachane, Vitor A. P. Martins dos Santos, Michail M. Yakimov, Kenneth N. Timmis, and Peter N. Golyshin. Microbial Enzymes Mined from the Urania Deep-Sea Hypersaline Anoxic Basin. 12(8):895–904. ISSN 1074-5521. doi: 10.1016/j.chembiol.2005.05.020. 24

[80] Dawn Field, Bela Tiwari, Tim Booth, Stewart Houten, Dan Swan, Nicolas Bertrand, and Milo Thurston. Open software for biologists: From famine to feast. 24(7):801–803. ISSN 1546-1696. doi: 10.1038/nbt0706-801. 99

[81] R. A. Fisher. On the Interpretation of $X2$ from Contingency Tables, and the Calculation of P. 85(1):87–94. ISSN 0952-8385. doi: 10.2307/2340521. 32

[82] J. Foght. Anaerobic Biodegradation of Aromatic Hydrocarbons: Pathways and Prospects. 15(2-3):93–120. ISSN 1660-2412. doi: 10.1159/000121324. 00177. 10

[83] Donald E. Francisco, Robert A. Mah, and Albert C. Rabin. Acridine Orange-Epifluorescence Technique for Counting Bacteria in Natural Waters. 92(3):416–421. ISSN 0003-0023. doi: 10.2307/3225245. 30, 97

[84] Niels-Ulrik Frigaard and Christiane Dahl. Sulfur Metabolism in Phototrophic Sulfur Bacteria. In *Advances in Microbial Physiology*, volume 54, pages 103–200. Elsevier. ISBN 978-0-12-374323-7. doi: 10.1016/S0065-2911(08)00002-7. URL https://linkinghub.elsevier.com/retrieve/pii/S0065291108000027. 85

[85] Glenn S. Frysinger, Richard B. Gaines, Li Xu, and Christopher M. Reddy. Resolving the Unresolved Complex Mixture in Petroleum-Contaminated Sediments. 37(8):1653–1662. ISSN 0013-936X. doi: 10.1021/es020742n. 3

[86] Georg Fuchs, Matthias Boll, and Johann Heider. Microbial degradation of aromatic compounds — from one strategy to four. 9(11):803–816. ISSN 1740-1534. doi: 10.1038/nrmicro2652. xii, 53, 83

[87] Simona Gagni and Darinn Cam. Stigmastane and hopanes as conserved biomarkers for estimating oil biodegradation in a former refinery plant-contaminated soil. 67(10): 1975–1981. ISSN 0045-6535. doi: 10.1016/j.chemosphere.2006.11.062. 113

[88] S. Gan, E.V. Lau, and H.K. Ng. Remediation of soils contaminated with polycyclic aromatic hydrocarbons (PAHs). 172(2-3):532–549. ISSN 03043894. doi: 10.1016/j.jhazmat.2009.07.118. 4

[89] Andrea Gärtner, Martina Blümel, Jutta Wiese, and Johannes F. Imhoff. Isolation and characterisation of bacteria from the Eastern Mediterranean deep sea. 100(3):421–435. ISSN 1572-9699. doi: 10.1007/s10482-011-9599-5. 22

[90] Maria Genovese, Francesca Crisafi, Renata Denaro, Simone Cappello, Daniela Russo, Rosario Calogero, Santina Santisi, Maurizio Catalfamo, Alfonso Modica, Francesco Smedile, Lucrezia Genovese, Peter N. Golyshin, Laura Giuliano, and Michail M. Yakimov. Effective bioremediation strategy for rapid in situ cleanup of anoxic marine sediments in mesocosm oil spill simulation. 5:162. ISSN 1664-302X. doi: 10.3389/fmicb.2014.00162. 16

[91] Debajyoti Ghosal, Shreya Ghosh, Tapan K. Dutta, and Youngho Ahn. Current State of Knowledge in Microbial Degradation of Polycyclic Aromatic Hydrocarbons (PAHs): A Review. page 1369. doi: 10.3389/fmicb.2016.01369. 00000. 4, 6

[92] Lisa M. Gieg and Courtney R. A. Toth. Anaerobic Biodegradation of Hydrocarbons: Metagenomics and Metabolomics. In *Consequences of Microbial Interactions with Hydrocarbons, Oils, and Lipids: Biodegradation and Bioremediation*, Handbook of Hydrocarbon and Lipid Microbiology, pages 1–42. Springer, Cham. ISBN 978-3-319-44535-9. doi: 10.1007/978-3-319-44535-9_16-1. URL https://link.springer.com/referenceworkentry/10.1007/978-3-319-44535-9_16-1. 92

[93] G. D. Ginsburg, R. A. Guseynov, A. A. Dadashev, G. A. Ivanova, S. A. Kazantsev, V. A. Solov'yev, E. V. Telepnev, R. Ye. Askeri-Nasirov, A. D. Yesikov, V. I. Mal'tseva, Yu. G. Mashirov, and Yu. Shabayeva. Gas Hydrates Of The Southern Caspian. 34(8): 765–782. ISSN 0020-6814, 1938-2839. doi: 10.1080/00206819209465635. 00071. 26

[94] A. C. Girnth, S. Grünke, A. Lichtschlag, J. Felden, K. Knittel, F. Wenzhöfer, D. de Beer, and A. Boetius. Novel, mat-forming Thiomargarita spp. thrive on a sulfidic fluid outflow at the Amon mud volcano (Eastern Mediterranean). 74 (12):A335–A335. ISSN 0016-7037. URL https://pure.mpg.de/pubman/faces/ViewItemOverviewPage.jsp?itemId=item_2485069. 24

[95] Nicholas J. Gotelli and Robert K. Colwell. Quantifying biodiversity: Procedures and pitfalls in the measurement and comparison of species richness. 4(4):379–391. ISSN 1461-0248. doi: 10.1046/j.1461-0248.2001.00230.x. 59

[96] Fabian Grein, Ana Raquel Ramos, Sofia S. Venceslau, and Inês A.C. Pereira. Unifying concepts in anaerobic respiration: Insights from dissimilatory sulfur metabolism. 1827 (2):145–160. ISSN 00052728. doi: 10.1016/j.bbabio.2012.09.001. 85

[97] Edye E. Groseclose and D. W. Ribbons. 3-Hydroxybenzoate 6-hydroxylase from Pseudomonas aeruginosa. 55(3):897–903. ISSN 0006-291X. doi: 10.1016/0006-291X(73)91228-X. 73

[98] M.J. Grossman, R.C. Prince, R.M. Garrett, K.K. Garrett, R.E. Bare, K.R. O'Neil, M.R. Sowlay, S.M. Hinton, K. Lee, G.A. Sergy, E.H. Owenns, and C.C. Guenette. *Microbial Diversity in Oiled and Un-Oiled Shoreline Sediments in the Norwegian Arctic*. Atlantic Canada Society for Microbial Ecology. 101

[99] Friederike Gründger, Núria Jiménez, Thomas Thielemann, Nontje Straaten, Tillmann Lüders, Hans-Hermann Richnow, and Martin Krüger. Microbial methane formation in deep aquifers of a coal-bearing sedimentary basin, Germany. 6. ISSN 1664-302X. doi: 10.3389/fmicb.2015.00200. 92

[100] S. Grünke, J. Felden, A. Lichtschlag, A.-C. Girnth, D. De Beer, F. Wenzhöfer, and A. Boetius. Niche differentiation among mat-forming, sulfide-oxidizing bacteria at cold seeps of the Nile Deep Sea Fan (Eastern Mediterranean Sea). 9(4):330–348. ISSN 1472-4669. doi: 10.1111/j.1472-4669.2011.00281.x. 24

[101] Michael R. Guerin, Ira B. Rubin, T. K. Rao, Bruce R. Clark, and James L. Epler. Distribution of mutagenic activity in petroleum and petroleum substitutes. 60(4): 282–288. ISSN 0016-2361. doi: 10.1016/0016-2361(81)90196-4. 2

[102] L. G. Guo, S. K. Liang, J. R. Lu, S. M. Yang, R. G. Su, and Y. Chen. [evaluation on biodegradability of hydrocarbon biomarkers in two crude oils under laboratory conditions]. 31(8):1897–1903. ISSN 0250-3301. URL http://europepmc.org/abstract/med/21090311. 113

[103] Sheridan K. Haack and Barbara A. Bekins. Microbial populations in contaminant plumes. 8(1):63–76. ISSN 1435-0157. doi: 10.1007/s100400050008. 12

[104] Daniel H. Haft, Jeremy D. Selengut, Roland A. Richter, Derek Harkins, Malay K. Basu, and Erin Beck. TIGRFAMs and Genome Properties in 2013. 41(D1):D387–D395. ISSN 0305-1048. doi: 10.1093/nar/gks1234. 58

[105] Benjamin S. Halpern, Shaun Walbridge, Kimberly A. Selkoe, Carrie V. Kappel, Fiorenza Micheli, Caterina D'Agrosa, John F. Bruno, Kenneth S. Casey, Colin Ebert, Helen E. Fox, Rod Fujita, Dennis Heinemann, Hunter S. Lenihan, Elizabeth M. P. Madin, Matthew T. Perry, Elizabeth R. Selig, Mark Spalding, Robert Steneck, and Reg Watson. A Global Map of Human Impact on Marine Ecosystems. 319(5865): 948–952. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.1149345. 52

[106] Najwa Hamad, Claude Millot, and Isabelle Taupier-Letage. A new hypothesis about the surface circulation in the eastern basin of the mediterranean sea. 66(2):287–298. ISSN 0079-6611. doi: 10.1016/j.pocean.2005.04.002. 22

[107] Jonathan D. Van Hamme, Ajay Singh, and Owen P. Ward. Recent Advances in Petroleum Microbiology. 67(4):503–549. ISSN 1092-2172, 1098-5557. doi: 10.1128/MMBR.67.4.503-549.2003. 4, 12

[108] Dennis A. Hansell. Recalcitrant Dissolved Organic Carbon Fractions. 5(1):421–445. doi: 10.1146/annurev-marine-120710-100757. 00084. 52

[109] Akihiro Hara, Kazuaki Syutsubo, and Shigeaki Harayama. Alcanivorax which prevails in oil-contaminated seawater exhibits broad substrate specificity for alkane degradation. 5(9):746–753. ISSN 1462-2920. doi: 10.1046/j.1468-2920.2003.00468.x. 12

[110] Renee M. Harada, Brandon A. Yoza, Stephen M. Masutani, and Qing X. Li. Diversity of Archaea Communities within Contaminated Sand Samples from Johnston Atoll. 17 (3):182–189. ISSN 1088-9868. doi: 10.1080/10889868.2013.808984. 105

[111] Shigeaki Harayama, Yuki Kasai, and Akihiro Hara. Microbial communities in oil-contaminated seawater. 15(3):205–214. ISSN 0958-1669. doi: 10.1016/j.copbio.2004.04.002. ix, 11

[112] Peter Harris, Andrew Heap, Vicki Passlow, Laura Sbaffi, Melissa Fellows, Rick Porter-Smith, Cameron Buchanan, and James Daniell. The Commonwealth Scientific and Industrial Research Organisation. 7(3):R126. ISSN 09609822. doi: 10.1016/S0960-9822(97)70976-X. 27

[113] Ronald G. Harvey. Mechanisms Of Carcinogenesis of Polycyclic Aromatic Hydrocarbons. 9(1-4):1–23. ISSN 1040-6638, 1563-5333. doi: 10.1080/10406639608031196. 2

[114] Mehdi Hassanshahian, Giti Emtiazi, and Simone Cappello. Isolation and characterization of crude-oil-degrading bacteria from the Persian Gulf and the Caspian Sea. 64(1):7–12, . ISSN 0025-326X. doi: 10.1016/j.marpolbul.2011.11.006. 00123. 27, 92

[115] Mehdi Hassanshahian, Giti Emtiazi, Rouha Kasra Kermanshahi, and Simone Cappello. Comparison of Oil Degrading Microbial Communities in Sediments from the Persian Gulf and Caspian Sea. 19(3):277–291, . ISSN 1532-0383. doi: 10.1080/15320381003695215. 00033. 27, 92

[116] Terry C. Hazen and Stephen M. Techtmann. Oil Biodegradation in Deep Marine Basins. In *Consequences of Microbial Interactions with Hydrocarbons, Oils, and*

Lipids: Biodegradation and Bioremediation, Handbook of Hydrocarbon and Lipid Microbiology, pages 1–18. Springer International Publishing. ISBN 978-3-319-44535-9. doi: 10.1007/978-3-319-44535-9_22-1. URL https://link.springer.com/referenceworkentry/10.1007/978-3-319-44535-9_22-1. vii, ix, 3, 6, 8, 21, 27, 29, 52, 53

[117] Terry C. Hazen, Eric A. Dubinsky, Todd Z. DeSantis, Gary L. Andersen, Yvette M. Piceno, Navjeet Singh, Janet K. Jansson, Alexander Probst, Sharon E. Borglin, Julian L. Fortney, William T. Stringfellow, Markus Bill, Mark E. Conrad, Lauren M. Tom, Krystle L. Chavarria, Thana R. Alusi, Regina Lamendella, Dominique C. Joyner, Chelsea Spier, Jacob Baelum, Manfred Auer, Marcin L. Zemla, Romy Chakraborty, Eric L. Sonnenthal, Patrik D'haeseleer, Hoi-Ying N. Holman, Shariff Osman, Zhenmei Lu, Joy D. Van Nostrand, Ye Deng, Jizhong Zhou, and Olivia U. Mason. Deep-Sea Oil Plume Enriches Indigenous Oil-Degrading Bacteria. 330(6001):204–208, . ISSN 0036-8075, 1095-9203. doi: 10.1126/science.1195979. 00479. 2, 6, 10, 21, 30, 55, 98, 101

[118] Terry C. Hazen, Roger C. Prince, and Nagissa Mahmoudi. Marine Oil Biodegradation. 50(5):2121–2129, . ISSN 0013-936X. doi: 10.1021/acs.est.5b03333. 00004. 2, 3, 4, 6, 10, 12, 50, 93, 103, 105, 107

[119] Ian M. Head, Neil D. Gray, and Stephen R. Larter. Life in the slow lane; biogeochemistry of biodegraded petroleum containing reservoirs and implications for energy recovery and carbon management. 5, . ISSN 1664-302X. doi: 10.3389/fmicb.2014.00566. 3, 4, 10, 52

[120] Ian M. Head, D. Martin Jones, and Steve R. Larter. Biological activity in the deep subsurface and the origin of heavy oil. 426(6964):344–352, . ISSN 1476-4687. doi: 10.1038/nature02134. 3

[121] Kenneth L. Heck, Gerald van Belle, and Daniel Simberloff. Explicit Calculation of the Rarefaction Diversity Measurement and the Determination of Sufficient Sample Size. 56(6):1459–1461. ISSN 1939-9170. doi: 10.2307/1934716. 59

[122] Johann Heider and Karola Schühle. Anaerobic Biodegradation of Hydrocarbons Including Methane (2013 ed.). In Eugene Rosenberg, Edward F. DeLong, Stephen Lory, Erko Stackebrandt, and Fabiano Thompson, editors, *The Prokaryotes*, pages 605–634. Springer Berlin Heidelberg. ISBN 978-3-642-30140-7 978-3-642-30141-4. doi: 10.1007/978-3-642-30141-4_80. URL http://link.springer.com/10.1007/978-3-642-30141-4_80. 10, 16, 85

[123] Sander K. Heijs, Anniet M. Laverman, Larry J. Forney, Pablo R. Hardoim, and Jan Dirk Van Elsas. Comparison of deep-sea sediment microbial communities in the Eastern Mediterranean. 64(3):362–377, . ISSN 0168-6496. doi: 10.1111/j.1574-6941.2008.00463.x. 24, 53

[124] Sander K. Heijs, Jaap S. Sinninghe Damsté, and Larry J. Forney. Characterization of a deep-sea microbial mat from an active cold seep at the Milano mud volcano in the Eastern Mediterranean Sea. 54(1):47–56, . ISSN 0168-6496. doi: 10.1016/j.femsec.2005.02.007. 24

[125] Klaus M Herrmann and Lisa M Weaver. The Shikimate Pathway. 50(1):473–503. ISSN 1040-2519. URL https://www.annualreviews.org/doi/10.1146/annurev.arplant.50.1.473. 76

[126] Rebecca Z Hoff. Responding to oil spills in coastal marshes: The fine line between help and hindrance. 2

[127] A. F. Hofmann, E. T. Peltzer, P. M. Walz, and P. G. Brewer. Hypoxia by degrees: Establishing definitions for a changing ocean. 58(12):1212–1226. ISSN 0967-0637. doi: 10.1016/j.dsr.2011.09.004. 100

[128] Robert Howarth, Francis Chan, Daniel J. Conley, Josette Garnier, Scott C. Doney, Roxanne Marino, and Gilles Billen. Coupled biogeochemical cycles: Eutrophication and hypoxia in temperate estuaries and coastal marine ecosystems. 9(1):18–26. ISSN 1540-9309. doi: 10.1890/100008. 52

[129] Robert W. Howarth and Roxanne Marino. Nitrogen as the limiting nutrient for eutrophication in coastal marine ecosystems: Evolving views over three decades. 51: 364–376. ISSN 1939-5590. doi: 10.4319/lo.2006.51.1_part_2.0364. 15

[130] Ping Hu, Eric A. Dubinsky, Alexander J. Probst, Jian Wang, Christian M. K. Sieber, Lauren M. Tom, Piero R. Gardinali, Jillian F. Banfield, Ronald M. Atlas, and Gary L. Andersen. Simulation of Deepwater Horizon oil plume reveals substrate specialization within a complex community of hydrocarbon degraders. page 201703424. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.1703424114. 00000. 50, 114

[131] Le Huang, Han Zhang, Peizhi Wu, Sarah Entwistle, Xueqiong Li, Tanner Yohe, Haidong Yi, Zhenglu Yang, and Yanbin Yin. dbCAN-seq: A database of carbohydrate-active enzyme (CAZyme) sequence and annotation. 46(D1):D516–D521. ISSN 0305-1048. doi: 10.1093/nar/gkx894. 57

[132] Julie A. Huber, David B. Mark Welch, Hilary G. Morrison, Susan M. Huse, Phillip R. Neal, David A. Butterfield, and Mitchell L. Sogin. Microbial Population Structures in the Deep Marine Biosphere. 318(5847):97–100. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.1146689. 53

[133] C. Huguen, J. P. Foucher, J. Mascle, H. Ondréas, M. Thouement, S. Gontharet, A. Stadnitskaia, C. Pierre, G. Bayon, L. Loncke, A. Boetius, I. Bouloubassi, G. de Lange, J. C. Caprais, Y. Fouquet, J. Woodside, and S. Dupré. Menes caldera, a highly active site of brine seepage in the Eastern Mediterranean sea: "In situ" observations from the NAUTINIL expedition (2003). 261(1):138–152. ISSN 0025-3227. doi: 10.1016/j.margeo.2009.02.005. 24

[134] Stuart H. Hurlbert. The Nonconcept of Species Diversity: A Critique and Alternative Parameters. 52(4):577–586. ISSN 1939-9170. doi: 10.2307/1934145. 31, 59

[135] Annemarie Hütz, Karin Schubert, and Jörg Overmann. Thalassospira sp. Isolated from the Oligotrophic Eastern Mediterranean Sea Exhibits Chemotaxis toward Inorganic Phosphate during Starvation. 77(13):4412–4421. ISSN 0099-2240, 1098-5336. doi: 10.1128/AEM.00490-11. 25

[136] Doug Hyatt, Gwo-Liang Chen, Philip F. LoCascio, Miriam L. Land, Frank W. Larimer, and Loren J. Hauser. Prodigal: Prokaryotic gene recognition and translation initiation site identification. 11(1):119. ISSN 1471-2105. doi: 10.1186/1471-2105-11-119. 57

[137] Norman J Hyne. *Nontechnical Guide to Petroleum Geology, Exploration, Drilling, and Production*. PennWell Corporation, 3rd ed edition. ISBN 978-0-87814-823-3. 3

[138] Ibraheem Borie Mohammad Ibraheem. BIODEGRADABILITY OF HYDROCARBONS BY CYANOBACTERIA1: BIODEGRADABILITY OF HCs BY CYANOBACTERIA. 46(4):818–824. ISSN 00223646. doi: 10.1111/j.1529-8817.2010.00865.x. 24, 25

[139] R A Ibrayev. Seasonal variability of the Caspian Sea three-dimensional circulation, sea level and air-sea interaction. page 19. 10, 15

[140] Anders R. Johnsen, Lukas Y. Wick, and Hauke Harms. Principles of microbial PAH-degradation in soil. 133(1):71–84. ISSN 0269-7491. doi: 10.1016/j.envpol.2004.04.015. 00809. 12

[141] Eric Jones, Travis Oliphant, and Pearu Peterson. SciPy: Open source scientific tools for Python. URL http://www.scipy.org/. 02493. 58

[142] K. C. Jones and P. de Voogt. Persistent organic pollutants (POPs): State of the science. 100(1):209–221. ISSN 0269-7491. doi: 10.1016/S0269-7491(99)00098-6. 4

[143] Lou Jost. Entropy and diversity. 113(2):363–375. ISSN 1600-0706. doi: 10.1111/j.2006.0030-1299.14714.x. 31

[144] Albert L Juhasz and Ravendra Naidu. Bioremediation of high molecular weight polycyclic aromatic hydrocarbons: A review of the microbial degradation of benzo[a]pyrene. 45(1-2):57–88. ISSN 09648305. doi: 10.1016/S0964-8305(00)00052-4. 4

[145] Jochen Kämpf. On preconditioning of coastal upwelling in the eastern Great Australian Bight. 115(C12). ISSN 2156-2202. doi: 10.1029/2010JC006294. 27

[146] Jochen Kämpf, Mark Doubell, David Griffin, Roger L. Matthews, and Tim M. Ward. Evidence of a large seasonal coastal upwelling system along the southern shelf of Australia. 31(9). ISSN 1944-8007. doi: 10.1029/2003GL019221. 27

[147] Minoru Kanehisa, Susumu Goto, Miho Furumichi, Mao Tanabe, and Mika Hirakawa. KEGG for representation and analysis of molecular networks involving diseases and drugs. 38:D355–D360, . ISSN 0305-1048. doi: 10.1093/nar/gkp896. 73

[148] Minoru Kanehisa, Yoko Sato, Masayuki Kawashima, Miho Furumichi, and Mao Tanabe. KEGG as a reference resource for gene and protein annotation. 44(D1): D457–462, . ISSN 1362-4962. doi: 10.1093/nar/gkv1070. 66

[149] Yuki Kasai, Hideo Kishira, Tetsuya Sasaki, Kazuaki Syutsubo, Kazuya Watanabe, and Shigeaki Harayama. Predominant growth of Alcanivorax strains in oil-contaminated and nutrient-supplemented sea water. 4(3):141–147. ISSN 1462-2920. doi: 10.1046/j.1462-2920.2002.00275.x. 101

[150] Jeong-Dong Kim and Choul-Gyun Lee. Microbial degradation of polycyclic aromatic hydrocarbons in soil by bacterium-fungus co-cultures. 12(4):410–416. ISSN 1226-8372. doi: 10.1007/BF02931064. 12

[151] Nikole E. Kimes, Amy V. Callaghan, Deniz F. Aktas, Whitney L. Smith, Jan Sunner, Bernardt Golding, Marta Drozdowska, Terry C. Hazen, Joseph M. Suflita, and Pamela J. Morris. Metagenomic analysis and metabolite profiling of deep-sea sediments from the Gulf of Mexico following the Deepwater Horizon oil spill. 4:50. ISSN 1664-302X. doi: 10.3389/fmicb.2013.00050. 6

[152] G. M. King, J. E. Kostka, T. C. Hazen, and P. A. Sobecky. Microbial responses to the Deepwater Horizon oil spill: From coastal wetlands to the deep sea. 7:377–401. ISSN 1941-0611. doi: 10.1146/annurev-marine-010814-015543. 10

[153] Y. V Kissin. Catagenesis and composition of petroleum: Origin of n-alkanes and isoalkanes in petroleum crudes. 51(9):2445–2457. ISSN 0016-7037. doi: 10.1016/0016-7037(87)90296-1. 4

[154] Claudia Knief, Vanina Dengler, Paul L. E. Bodelier, and Julia A. Vorholt. Characterization of Methylobacterium strains isolated from the phyllosphere and description of Methylobacterium longum sp. nov. 101(1):169–183. ISSN 1572-9699. doi: 10.1007/s10482-011-9650-6. 109

[155] Alexander Korshenko and Alvin Gasim Gul. Pollution of the Caspian Sea. In Andrey G. Kostianoy and Aleksey N. Kosarev, editors, *The Caspian Sea Environment*, volume 5P, pages 109–142. Springer Berlin Heidelberg. ISBN 978-3-540-28281-5. doi: 10.1007/698_5_006. URL http://link.springer.com/10.1007/698_5_006. 00000. 53, 92

[156] Alexey N. Kosarev. Physico-Geographical Conditions of the Caspian Sea. In Andrey G. Kostianoy and Aleksey N. Kosarev, editors, *The Caspian Sea Environment*, volume 5P, pages 5–31. Springer Berlin Heidelberg. ISBN 978-3-540-28281-5. doi: 10.1007/698_5_002. URL http://link.springer.com/10.1007/698_5_002. 00000. 25

[157] M. D. Krom, K-C Emeis, and P. Van Cappellen. Why is the Eastern Mediterranean phosphorus limited? 85(3–4):236–244, . ISSN 0079-6611. doi: 10.1016/j.pocean.2010.03.003. 00114. 10, 24, 25, 34

[158] M. D. Krom, N. Kress, S. Brenner, and L. I. Gordon. Phosphorus limitation of primary productivity in the eastern Mediterranean Sea. 36(3):424–432, . ISSN 00243590. doi: 10.4319/lo.1991.36.3.0424. 10

[159] Michael D Krom. The oceanography of the eastern Mediterranean Sea. 5:22–28. 22

[160] Michael D Krom, S Brenner, L Israilov, and B Krumgalz. Dissolved nutrients, preformed nutrients. 14(2):189–194, . 25

[161] I. Yu Kulikova, I. S. Dzerzhinskaya, and A. I. Netrusov. Study of an opportunity of application of the bioagent Phyloil for elimination of emergency oil spills in the sea. 65(3):112–115. ISSN 0096-3925, 1934-791X. doi: 10.3103/S0096392510030065. 00000. 92

[162] M. M. M. Kuypers, G. Lavik, D. Woebken, M. Schmid, B. M. Fuchs, R. Amann, B. B. Jorgensen, and M. S. M. Jetten. Massive nitrogen loss from the Benguela upwelling

system through anaerobic ammonium oxidation. 102(18):6478–6483. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.0502088102. 20

[163] Matthias Labrenz, Eva Sintes, Falko Toetzke, Anita Zumsteg, Gerhard J. Herndl, Marleen Seidler, and Klaus Jürgens. Relevance of a crenarchaeotal subcluster related to *Candidatus* Nitrosopumilus maritimus to ammonia oxidation in the suboxic zone of the central Baltic Sea. 4(12):1496–1508. ISSN 1751-7370. doi: 10.1038/ismej.2010.78. 105

[164] D. Lamy, C. Jeanthon, M. T. Cottrell, D. L. Kirchman, F. Van Wambeke, J. Ras, O. Dahan, M. Pujo-Pay, L. Oriol, L. Bariat, P. Catala, V. Cornet-Barthaux, and P. Lebaron. Ecology of aerobic anoxygenic phototrophic bacteria along an oligotrophic gradient in the Mediterranean Sea. 8(4):973–985. ISSN 1726-4189. doi: 10.5194/bg-8-973-2011. 22

[165] Ben Langmead and Steven L. Salzberg. Fast gapped-read alignment with Bowtie 2. 9 (4):357–359. ISSN 1548-7091. doi: 10.1038/nmeth.1923. 05557. 58

[166] Rafael Laso-Pérez, Gunter Wegener, Katrin Knittel, Friedrich Widdel, Katie J. Harding, Viola Krukenberg, Dimitri V. Meier, Michael Richter, Halina E. Tegetmeyer, Dietmar Riedel, Hans-Hermann Richnow, Lorenz Adrian, Thorsten Reemtsma, Oliver J. Lechtenfeld, and Florin Musat. Thermophilic archaea activate butane via alkyl-coenzyme M formation. 539(7629):396–401. ISSN 1476-4687. doi: 10.1038/nature20152. 92, 93

[167] Gaute Lavik, Torben Stührmann, Volker Brüchert, Anja Van der Plas, Volker Mohrholz, Phyllis Lam, Marc Mußmann, Bernhard M. Fuchs, Rudolf Amann, Ulrich Lass, and Marcel M. M. Kuypers. Detoxification of sulphidic African shelf waters by blooming chemolithotrophs. 457(7229):581–584. ISSN 0028-0836. doi: 10.1038/nature07588. 00194. 20

[168] Stephanie N. Lawler, Christina A. Kellogg, Scott C. France, Rachel W. Clostio, Sandra D. Brooke, and Steve W. Ross. Coral-Associated Bacterial Diversity Is

Conserved across Two Deep-Sea Anthothela Species. 7:458. ISSN 1664-302X. doi: 10.3389/fmicb.2016.00458. 109

[169] Taylor H. Leach, Beatrix E. Beisner, Cayelan C. Carey, Patricia Pernica, Kevin C. Rose, Yannick Huot, Jennifer A. Brentrup, Isabelle Domaizon, Hans-Peter Grossart, Bastiaan W. Ibelings, Stéphan Jacquet, Patrick T. Kelly, James A. Rusak, Jason D. Stockwell, Dietmar Straile, and Piet Verburg. Patterns and drivers of deep chlorophyll maxima structure in 100 lakes: The relative importance of light and thermal stratification. 63(2):628–646. ISSN 1939-5590. doi: 10.1002/lno.10656. 15

[170] Da-Eun Lee, Jinhwan Lee, Young-Mog Kim, Jeong-In Myeong, and Kyoung-Ho Kim. Uncultured bacterial diversity in a seawater recirculating aquaculture system revealed by 16S rRNA gene amplicon sequencing. 54(4):296–304. ISSN 1976-3794. doi: 10.1007/s12275-016-5571-4. 109

[171] S. A. G. Leroy, F. Marret, E. Gibert, F. Chalié, J. L. Reyss, and K. Arpe. River inflow and salinity changes in the Caspian Sea during the last 5500 years. 26(25):3359–3383. ISSN 0277-3791. doi: 10.1016/j.quascirev.2007.09.012. 91

[172] Yufang Li, Kai Tang, Lianbao Zhang, Zihao Zhao, Xiabing Xie, Chen-Tung Arthur Chen, Deli Wang, Nianzhi Jiao, and Yao Zhang. Coupled Carbon, Sulfur, and Nitrogen Cycles Mediated by Microorganisms in the Water Column of a Shallow-Water Hydrothermal Ecosystem. 9. ISSN 1664-302X. doi: 10.3389/fmicb.2018.02718. 15, 81

[173] Jiang Liu, Stephen M. Techtmann, Hannah L. Woo, Daliang Ning, Julian L. Fortney, and Terry C. Hazen. Rapid Response of Eastern Mediterranean Deep Sea Microbial Communities to Oil. 7(1). ISSN 2045-2322. doi: 10.1038/s41598-017-05958-x. 00000. 10, 103

[174] Graham A. Logan, Andrew T. Jones, John M. Kennard, Georgina J. Ryan, and Nadege Rollet. Australian offshore natural hydrocarbon seepage studies, a review and re-evaluation. 27(1):26–45. ISSN 0264-8172. doi: 10.1016/j.marpetgeo.2009.07.002. 3, 27

[175] Robert A. Ludwig. Microaerophilic bacteria transduce energy via oxidative metabolic gearing. 155(2):61–70. ISSN 0923-2508. doi: 10.1016/j.resmic.2003.11.001. 15

[176] Nagissa Mahmoudi. Assessing In Situ Degradation Of Petroleum Hydrocarbons By Indigenous Microbial Communities. URL https://macsphere.mcmaster.ca/handle/11375/13050. 2, 4, 6

[177] Nagissa Mahmoudi, Michael S. Robeson, Hector F. Castro, Julian L. Fortney, Stephen M. Techtmann, Dominique C. Joyner, Charles J. Paradis, Susan M. Pfiffner, and Terry C. Hazen. Microbial community composition and diversity in Caspian Sea sediments. 91(1):1–11. ISSN 1574-6941. doi: 10.1093/femsec/fiu013. 00006. 92, 101, 105

[178] Frank D. Mango. The light hydrocarbons in petroleum: A critical review. 26(7): 417–440. ISSN 0146-6380. doi: 10.1016/S0146-6380(97)00031-4. 4

[179] Nathan Mantel. The Detection of Disease Clustering and a Generalized Regression Approach. 27:209–220. ISSN 0008-5472, 1538-7445. URL http://cancerres.aacrjournals.org/content/27/2_Part_1/209. 09444. 61

[180] Mercedes M. Marıń, Theo H. M. Smits, Jan B. van Beilen, and Fernando Rojo. The Alkane Hydroxylase Gene of Burkholderia cepacia RR10 Is under Catabolite Repression Control. 183(14):4202–4209. ISSN 0021-9193, 1098-5530. doi: 10.1128/JB.183.14.4202-4209.2001. 12

[181] Alan G. Marshall and Ryan P. Rodgers. Petroleomics: The Next Grand Challenge for Chemical Analysis. 37(1):53–59. ISSN 0001-4842. doi: 10.1021/ar020177t. 3

[182] Olivia U. Mason, Terry C. Hazen, Sharon Borglin, Patrick S. G. Chain, Eric A. Dubinsky, Julian L. Fortney, James Han, Hoi-Ying N. Holman, Jenni Hultman, Regina Lamendella, Rachel Mackelprang, Stephanie Malfatti, Lauren M. Tom, Susannah G. Tringe, Tanja Woyke, Jizhong Zhou, Edward M. Rubin, and Janet K. Jansson. Metagenome, metatranscriptome and single-cell sequencing reveal microbial response

to Deepwater Horizon oil spill. 6(9):1715–1727. ISSN 1751-7362. doi: 10.1038/ismej.2012.59. 00122. 6, 10, 70, 73

[183] Vincent Mastalerz, Gert J. de Lange, and Anke Dählmann. Differential aerobic and anaerobic oxidation of hydrocarbon gases discharged at mud volcanoes in the Nile deep-sea fan. 73(13):3849–3863. ISSN 0016-7037. doi: 10.1016/j.gca.2008.12.030. 24, 53

[184] Daniel McDonald, Jose C. Clemente, Justin Kuczynski, Jai Ram Rideout, Jesse Stombaugh, Doug Wendel, Andreas Wilke, Susan Huse, John Hufnagle, Folker Meyer, Rob Knight, and J. Gregory Caporaso. The Biological Observation Matrix (BIOM) format or: How I learned to stop worrying and love the ome-ome. 1(1):7. ISSN 2047-217X. doi: 10.1186/2047-217X-1-7. 00074. 99

[185] Robert McGill, John W. Tukey, and Wayne A. Larsen. Variations of Box Plots. 32(1): 12–16. ISSN 0003-1305. doi: 10.2307/2683468. 101

[186] Paul J. McMurdie and Susan Holmes. Waste Not, Want Not: Why Rarefying Microbiome Data Is Inadmissible. 10(4):e1003531. doi: 10.1371/journal.pcbi.1003531. 00050. 60

[187] Rainer U. Meckenstock, Matthias Boll, Housna Mouttaki, Janina S. Koelschbach, Paola Cunha Tarouco, Philip Weyrauch, Xiyang Dong, and Anne M. Himmelberg. Anaerobic Degradation of Benzene and Polycyclic Aromatic Hydrocarbons. 26(1-3): 92–118. ISSN 1464-1801, 1660-2412. doi: 10.1159/000441358. 3, 52

[188] F. M. Menn, B. M. Applegate, and G. S. Sayler. NAH plasmid-mediated catabolism of anthracene and phenanthrene to naphthoic acids. 59(6):1938–1942. ISSN 0099-2240, 1098-5336. URL https://aem.asm.org/content/59/6/1938. 12

[189] F Meyer, D Paarmann, M D'Souza, R Olson, Em Glass, M Kubal, T Paczian, A Rodriguez, R Stevens, A Wilke, J Wilkening, and Ra Edwards. The metagenomics RAST server – a public resource for the automatic phylogenetic and functional analysis

of metagenomes. 9(1):386. ISSN 1471-2105. doi: 10.1186/1471-2105-9-386. 01653. xi, 55, 56, 59, 66

[190] Folker Meyer. Metagenome Analysis With MG-RAST. xi, 56

[191] Lutz-Arend Meyer-Reil and Marion Köster. Eutrophication of Marine Waters: Effects on Benthic Microbial Communities. 41(1):255–263. ISSN 0025-326X. doi: 10.1016/S0025-326X(00)00114-4. 15, 52

[192] John F. Middleton and Mauro Cirano. A northern boundary current along Australia's southern shelves: The Flinders Current. 107(C9):12–1–12–11. ISSN 2156-2202. doi: 10.1029/2000JC000701. 27

[193] John F. Middleton and Guennadi Platov. The Mean Summertime Circulation along Australia's Southern Shelves: A Numerical Study. 33(11):2270–2287. ISSN 0022-3670. doi: 10.1175/1520-0485(2003)033<2270:TMSCAA>2.0.CO;2. 27

[194] D. N. Miller, J. E. Bryant, E. L. Madsen, and W. C. Ghiorse. Evaluation and Optimization of DNA Extraction and Purification Procedures for Soil and Sediment Samples. 65(11):4715–4724. ISSN 0099-2240, 1098-5336. URL https://aem.asm.org/content/65/11/4715. 30, 55

[195] Alex Mitchell, Francois Bucchini, Guy Cochrane, Hubert Denise, Petra ten Hoopen, Matthew Fraser, Sebastien Pesseat, Simon Potter, Maxim Scheremetjew, Peter Sterk, and Robert D. Finn. EBI metagenomics in 2016 - an expanding and evolving resource for the analysis and archiving of metagenomic data. 44(D1):D595–D603. ISSN 0305-1048. doi: 10.1093/nar/gkv1195. 00011. xi, 56

[196] V. Mohrholz, C. H. Bartholomae, A. K. van der Plas, and H. U. Lass. The seasonal variability of the northern Benguela undercurrent and its relation to the oxygen budget on the shelf. 28(3):424–441. ISSN 0278-4343. doi: 10.1016/j.csr.2007.10.001. 00092. 20

[197] Mary Ann Moran. The global ocean microbiome. 350(6266):aac8455. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.aac8455. 53

[198] Robert M. Morris, Christian D. Frazar, and Craig A. Carlson. Basin-scale patterns in the abundance of SAR11 subclades, marine Actinobacteria (OM1), members of the Roseobacter clade and OCS116 in the South Atlantic: Bacterioplankton communities in the South Atlantic. 14(5):1133–1144, . ISSN 14622912. doi: 10.1111/j.1462-2920.2011.02694.x. 00000. 20

[199] Robert M. Morris, Brook L. Nunn, Christian Frazar, David R. Goodlett, Ying S. Ting, and Gabrielle Rocap. Comparative metaproteomics reveals ocean-scale shifts in microbial nutrient utilization and energy transduction. 4(5):673–685, . ISSN 1751-7362. doi: 10.1038/ismej.2010.4. 00154. 20

[200] Ahmed S. Moursy and Mohamed M. El-Abagy. Microbial degradation of hydrocarbons in Ismailia canal water. 7(6):423–427. ISSN 0160-4120. doi: 10.1016/0160-4120(82)90158-1. 24, 25

[201] Hasan Saravi Nasrollahzadeh, Zubir Bin Din, Swee Yeok Foong, and Asieh Makhlough. Trophic status of the Iranian Caspian Sea based on water quality parameters and phytoplankton diversity. 28(9):1153–1165. ISSN 0278-4343. doi: 10.1016/j.csr.2008.02.015. 00043. 91

[202] Dipti D. Nayak and Christopher J. Marx. Genetic and phenotypic comparison of facultative methylotrophy between Methylobacterium extorquens strains PA1 and AM1. 9(9):e107887. ISSN 1932-6203. doi: 10.1371/journal.pone.0107887. 109

[203] Ahmed El Nemr, Tarek O. Said, Azza Khaled, Amany El-Sikaily, and Aly M. A. Abd-Allah. The Distribution and Sources of Polycyclic Aromatic Hydrocarbons in Surface Sediments Along the Egyptian Mediterranean Coast. 124(1):343–359. ISSN 1573-2959. doi: 10.1007/s10661-006-9231-8. 24

[204] J. Farley Nicholls, R. Toumi, and W. P. Budgell. Inertial currents in the Caspian Sea. 39(18). ISSN 1944-8007. doi: 10.1029/2012GL052989. 10

[205] Robert D. Norris. *Handbook of Bioremediation*. CRC Press. ISBN 978-1-56670-074-0. 4

[206] Sergey Nurk, Anton Bankevich, Dmitry Antipov, Alexey Gurevich, Anton Korobeynikov, Alla Lapidus, Andrey Prjibelsky, Alexey Pyshkin, Alexander Sirotkin, Yakov Sirotkin, Ramunas Stepanauskas, Jeffrey McLean, Roger Lasken, Scott R. Clingenpeel, Tanja Woyke, Glenn Tesler, Max A. Alekseyev, and Pavel A. Pevzner. Assembling Genomes and Mini-metagenomes from Highly Chimeric Reads. In Minghua Deng, Rui Jiang, Fengzhu Sun, and Xuegong Zhang, editors, *Research in Computational Molecular Biology*, Lecture Notes in Computer Science, pages 158–170. Springer Berlin Heidelberg. ISBN 978-3-642-37195-0. 57

[207] Oluwafemi S. Obayori, Matthew O. Ilori, Sunday A. Adebusoye, Ganiyu O. Oyetibo, Ayodele E. Omotayo, and Olukayode O. Amund. Degradation of hydrocarbons and biosurfactant production by Pseudomonas sp. strain LP1. 25(9):1615–1623. ISSN 1573-0972. doi: 10.1007/s11274-009-0053-z. 12

[208] Jari Oksanen, F. Guillaume Blanchet, Michael Friendly, Roeland Kindt, Pierre Legendre, Dan McGlinn, Peter R. Minchin, R. B. O'Hara, Gavin L. Simpson, Peter Solymos, M. Henry H. Stevens, Eduard Szoecs, and Helene Wagner. *Vegan: Community Ecology Package*. URL https://CRAN.R-project.org/package=vegan. 100

[209] Scott W. Olesen, Suhani Vora, Stephen M. Techtmann, Julian L. Fortney, Juan R. Bastidas-Oyanedel, Jorge Rodríguez, Terry C. Hazen, and Eric J. Alm. A Novel Analysis Method for Paired-Sample Microbial Ecology Experiments. 11(5):e0154804. ISSN 1932-6203. doi: 10.1371/journal.pone.0154804. 100, 107

[210] Enoma O. Omoregie, Vincent Mastalerz, Gert de Lange, Kristina L. Straub, Andreas Kappler, Hans Røy, Alina Stadnitskaia, Jean-Paul Foucher, and Antje Boetius. Biogeochemistry and Community Composition of Iron- and Sulfur-Precipitating Microbial Mats at the Chefren Mud Volcano (Nile Deep Sea Fan, Eastern Mediterranean). 74(10):3198–3215, . ISSN 0099-2240, 1098-5336. doi: 10.1128/AEM.01751-07. 24

[211] Enoma O. Omoregie, Helge Niemann, Vincent Mastalerz, Gert J. de Lange, Alina Stadnitskaia, Jean Mascle, Jean-Paul Foucher, and Antje Boetius. Microbial methane oxidation and sulfate reduction at cold seeps of the deep Eastern Mediterranean Sea. 261(1):114–127, . ISSN 0025-3227. doi: 10.1016/j.margeo.2009.02.001. 24, 53

[212] Beth N. Orcutt, Samantha B. Joye, Sara Kleindienst, Katrin Knittel, Alban Ramette, Anja Reitz, Vladimir Samarkin, Tina Treude, and Antje Boetius. Impact of natural oil and higher hydrocarbons on microbial diversity, distribution, and activity in Gulf of Mexico cold-seep sediments. 57(21):2008–2021. ISSN 0967-0645. doi: 10.1016/j.dsr2.2010.05.014. 3, 12

[213] Ross Overbeek, Tadhg Begley, Ralph M. Butler, Jomuna V. Choudhuri, Han-Yu Chuang, Matthew Cohoon, Valérie de Crécy-Lagard, Naryttza Diaz, Terry Disz, Robert Edwards, Michael Fonstein, Ed D. Frank, Svetlana Gerdes, Elizabeth M. Glass, Alexander Goesmann, Andrew Hanson, Dirk Iwata-Reuyl, Roy Jensen, Neema Jamshidi, Lutz Krause, Michael Kubal, Niels Larsen, Burkhard Linke, Alice C. McHardy, Folker Meyer, Heiko Neuweger, Gary Olsen, Robert Olson, Andrei Osterman, Vasiliy Portnoy, Gordon D. Pusch, Dmitry A. Rodionov, Christian Rückert, Jason Steiner, Rick Stevens, Ines Thiele, Olga Vassieva, Yuzhen Ye, Olga Zagnitko, and Veronika Vonstein. The Subsystems Approach to Genome Annotation and its Use in the Project to Annotate 1000 Genomes. 33(17):5691–5702. ISSN 0305-1048, 1362-4962. doi: 10.1093/nar/gki866. 00000. 66

[214] Maria G. Pachiadaki, Argyri Kallionaki, Anke Dählmann, Gert J. De Lange, and Konstantinos Ar. Kormas. Diversity and Spatial Distribution of Prokaryotic Communities Along A Sediment Vertical Profile of A Deep-Sea Mud Volcano. 62 (3):655–668. ISSN 1432-184X. doi: 10.1007/s00248-011-9855-2. 24

[215] Silvio Pantoja, Daniel J Repeta, Julian P Sachs, and Daniel M Sigman. Stable isotope constraints on the nitrogen cycle of the Mediterranean Sea water column. 49(9):1609–1621. ISSN 0967-0637. doi: 10.1016/S0967-0637(02)00066-3. 25

[216] A. Paulmier and D. Ruiz-Pino. Oxygen minimum zones (OMZs) in the modern ocean. 80(3):113–128. ISSN 0079-6611. doi: 10.1016/j.pocean.2008.08.001. 52

[217] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, and Vincent Dubourg. Scikit-learn: Machine learning in Python. 12:2825–2830. URL http://www.jmlr.org/papers/v12/pedregosa11a.html. 03975. 55, 58

[218] F. Peeters, R. Kipfer, D. Achermann, M. Hofer, W. Aeschbach-Hertig, U. Beyerle, D.M. Imboden, K. Rozanski, and K. Fröhlich. Analysis of deep-water exchange in the Caspian Sea based on environmental tracers. 47(4):621–654. ISSN 09670637. doi: 10.1016/S0967-0637(99)00066-7. 10, 15

[219] Ri-He Peng, Ai-Sheng Xiong, Yong Xue, Xiao-Yan Fu, Feng Gao, Wei Zhao, Yong-Sheng Tian, and Quan-Hong Yao. Microbial biodegradation of polyaromatic hydrocarbons. 32(6):927–955. ISSN 1574-6976. doi: 10.1111/j.1574-6976.2008.00127.x. 4, 12

[220] S. R. Pezeshki, M. W. Hester, Q. Lin, and J. A. Nyman. The effects of oil spill and clean-up on dominant US Gulf coast marsh macrophytes: A review. 108(2):129–139. ISSN 0269-7491. doi: 10.1016/S0269-7491(99)00244-4. 2

[221] Craig D. Phelps, Lee J. Kerkhof, and Lily Y. Young. Molecular characterization of a sulfate-reducing consortium which mineralizes benzene. 27(3):269–279. ISSN 0168-6496. doi: 10.1111/j.1574-6941.1998.tb00543.x. 24

[222] S. Poux, M. Magrane, C. N. Arighi, A. Bridge, C. O'Donovan, K. Laiho, and The UniProt Consortium. Expert curation in UniProtKB: A case study on dealing with conflicting and erroneous data. 2014(0):bau016–bau016. ISSN 1758-0463. doi: 10.1093/database/bau016. 58

[223] R. C. Prince, R. R. Lessard, and J. R. Clark. Bioremediation of Marine Oil Spills. 58 (4):463–468, . ISSN 1294-4475. doi: 10.2516/ogst:2003029. 6

[224] Roger C. Prince and James R. Bragg. Shoreline Bioremediation Following the Exxon Valdez Oil Spill in Alaska. 1(2):97–104. ISSN 1088-9868. doi: 10.1080/10889869709351324. 6

[225] Roger C. Prince and Clifford C. Walters. Chapter 11 - Biodegradation of oil hydrocarbons and its implications for source identification. In Zhendi Wang and Scott A. Stout, editors, *Oil Spill Environmental Forensics*, pages 349–379. Academic Press. ISBN 978-0-12-369523-9. doi: 10.1016/B978-012369523-9.50015-X. URL http://www.sciencedirect.com/science/article/pii/B978012369523950015X. 10

[226] Roger C. Prince, Kelly M. McFarlin, Josh D. Butler, Eric J. Febbo, Frank C. Y. Wang, and Tim J. Nedwed. The primary biodegradation of dispersed crude oil in the sea. 90 (2):521–526, . ISSN 0045-6535. doi: 10.1016/j.chemosphere.2012.08.020. 10

[227] Kim D. Pruitt, Tatiana Tatusova, and Donna R. Maglott. NCBI Reference Sequence (RefSeq): A curated non-redundant sequence database of genomes, transcripts and proteins. 33:D501–D504. ISSN 0305-1048. doi: 10.1093/nar/gki025. 58, 59, 66

[228] Marco Punta, Penny C. Coggill, Ruth Y. Eberhardt, Jaina Mistry, John Tate, Chris Boursnell, Ningze Pang, Kristoffer Forslund, Goran Ceric, Jody Clements, Andreas Heger, Liisa Holm, Erik L. L. Sonnhammer, Sean R. Eddy, Alex Bateman, and Robert D. Finn. The Pfam protein families database. 40(D1):D290–D301. ISSN 0305-1048. doi: 10.1093/nar/gkr1065. 58

[229] Christian Quast, Elmar Pruesse, Pelin Yilmaz, Jan Gerken, Timmy Schweer, Pablo Yarza, Jörg Peplies, and Frank Oliver Glöckner. The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools. 41(D1):D590–D596. ISSN 0305-1048. doi: 10.1093/nar/gks1219. 31, 66

[230] R Core Team. R: A Language and Environment for Statistical Computing. URL https://www.R-project.org/. 00540. 58

[231] M. A. Rashid. Pristane-phytane ratios in relation to source and diagenesis of ancient sediments from the Labrador Shelf. 25(1):109–122. ISSN 0009-2541. doi: 10.1016/0009-2541(79)90087-1. 113

[232] Robarts Rd, Zohary T, Waiser Mj, and Yacobi Yz. Bacterial abundance, biomass, and production in relation to phytoplankton biomass in the Levantine Basin of the southeastern Mediterranean Sea. 137:273–281. ISSN 0171-8630, 1616-1599. doi: 10.3354/meps137273. 22

[233] AC Redfield, BH Ketchum, and FA Richards. The Sea, ed Hill MN. 22

[234] Molly C. Redmond and David L. Valentine. Natural gas and temperature structured a microbial community response to the Deepwater Horizon oil spill. 109(50):20292–20297. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.1108756108. 12

[235] Deni Ribicic, Roman Netzer, Terry C. Hazen, Stephen M. Techtmann, Finn Drabløs, and Odd Gunnar Brakstad. Microbial community and metagenome dynamics during biodegradation of dispersed oil reveals potential key-players in cold Norwegian seawater. 129(1):370–378. ISSN 0025-326X. doi: 10.1016/j.marpolbul.2018.02.034. 53

[236] Katherine Richardson and Bo Barker Jørgensen. Eutrophication: Definition, History and Effects. In *Eutrophication in Coastal Marine Ecosystems*, pages 1–19. American Geophysical Union (AGU). ISBN 978-1-118-66548-0. doi: 10.1029/CE052p0001. URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/CE052p0001. 52

[237] A. R. Robinson, P. Malanotte-Rizzoli, A. Hecht, A. Michelato, W. Roether, A. Theocharis, Ü. Ünlüata, N. Pinardi, A. Artegiani, A. Bergamasco, J. Bishop, S. Brenner, S. Christianidis, M. Gacic, D. Georgopoulos, M. Golnaraghi, M. Hausmann, H. G. Junghaus, A. Lascaratos, M. A. Latif, W. G. Leslie, C. J. Lozano, T. Oguz, E. Özsoy, E. Papageorgiou, E. Paschini, Z. Rozentroub, E. Sansone, P. Scarazzato, R. Schlitzer, G. C. Spezie, E. Tziperman, G. Zodiatis, L. Athanassiadou, M. Gerges, and M. Osman. General circulation of the Eastern Mediterranean. 32(4):285–309. ISSN 0012-8252. doi: 10.1016/0012-8252(92)90002-B. 22, 25

[238] Luis M. Rodriguez-R, Will A. Overholt, Christopher Hagan, Markus Huettel, Joel E. Kostka, and Konstantinos T. Konstantinidis. Microbial community successional patterns in beach sands impacted by the Deepwater Horizon oil spill. 9(9):1928–1940. ISSN 1751-7370. doi: 10.1038/ismej.2015.5. 6

[239] Fernando Rojo. Degradation of alkanes by bacteria. 11(10):2477–2490. ISSN 1462-2920. doi: 10.1111/j.1462-2920.2009.01948.x. 12

[240] Wilfred F. M. Röling, Michael G. Milner, D. Martin Jones, Kenneth Lee, Fabien Daniel, Richard J. P. Swannell, and Ian M. Head. Robust Hydrocarbon Degradation and Dynamics of Bacterial Communities during Nutrient-Enhanced Oil Spill Bioremediation. 68(11):5537–5548. ISSN 0099-2240, 1098-5336. doi: 10.1128/AEM.68.11.5537-5548.2002. 103

[241] Hans Røy, Jens Kallmeyer, Rishi Ram Adhikari, Robert Pockalny, Bo Barker Jørgensen, and Steven D'Hondt. Aerobic Microbial Respiration in 86-Million-Year-Old Deep-Sea Red Clay. 336(6083):922–925. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.1219424. 16

[242] Julian P. Sachs and Daniel J. Repeta. Oligotrophy and Nitrogen Fixation During Eastern Mediterranean Sapropel Events. 286(5449):2485–2488. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.286.5449.2485. 25

[243] A. Safary, M. Roayayi Ardakani, A. Abolhasani Suraki, M. Akbarzade Khiavi, and H. Motamedi. Isolation and Characterization of Biosurfactant Producing Bacteria from Caspian Sea. 9(3):378–382. ISSN 1682296X. doi: 10.3923/biotech.2010.378.382.00012. 92

[244] Abolfazl Saleh, Ali Hamzehpour, Ali Mehdinia, Kazem Darvish Bastami, and Said Mazaheri. Hydrochemistry and nutrient distribution in southern deep-water basin of the Caspian Sea. 127:406–411. ISSN 0025-326X. doi: 10.1016/j.marpolbul.2017.12.013. 91

[245] M. A. Salmanov. Microbiological studies in the deepwater area of the Southern Caspian Sea. 75(2):206–212. ISSN 0026-2617, 1608-3237. doi: 10.1134/S0026261706020159. 00001. 26, 27, 92

[246] Howard L. Sanders. Marine benthic diversity: A comparative study. pages 243–282. 02383. 59

[247] S. Sankar, L. Polimene, L. Marin, N. N. Menon, A. Samuelsen, R. Pastres, and S. Ciavatta. Sensitivity of the simulated Oxygen Minimum Zone to biogeochemical processes at an oligotrophic site in the Arabian Sea. 372:12–23. ISSN 0304-3800. doi: 10.1016/j.ecolmodel.2018.01.016. 16

[248] R. Santas, A. Korda, A. Tenente, K. Buchholz, and Ph. Santas. Mesocosm assays of oil spill bioremediation with oleophilic fertilizers: Inipol, F1 or both? 38(1):44–48. ISSN 0025-326X. doi: 10.1016/S0025-326X(99)80011-3. 24, 25

[249] Regina Schauer, Christina Bienhold, Alban Ramette, and Jens Harder. Bacterial diversity and biogeography in deep-sea surface sediments of the South Atlantic Ocean. 4(2):159–170. ISSN 1751-7362. doi: 10.1038/ismej.2009.106. 00123. 21

[250] Alexandra M. Schnoes, Shoshana D. Brown, Igor Dodevski, and Patricia C. Babbitt. Annotation Error in Public Databases: Misannotation of Molecular Function in Enzyme Superfamilies. 5(12):e1000605. ISSN 1553-7358. doi: 10.1371/journal.pcbi.1000605. 62

[251] René P. Schwarzenbach, Philip M. Gschwend, and Dieter M. Imboden. *Environmental Organic Chemistry*. John Wiley & Sons. ISBN 978-1-118-76723-8. 2

[252] Alexander Sczyrba, Peter Hofmann, Peter Belmann, David Koslicki, Stefan Janssen, Johannes Dröge, Ivan Gregor, Stephan Majda, Jessika Fiedler, Eik Dahms, Andreas Bremges, Adrian Fritz, Ruben Garrido-Oter, Tue Sparholt Jørgensen, Nicole Shapiro, Philip D. Blood, Alexey Gurevich, Yang Bai, Dmitrij Turaev, Matthew Z. DeMaere, Rayan Chikhi, Niranjan Nagarajan, Christopher Quince, Fernando Meyer, Monika Balvoit, Lars Hestbjerg Hansen, Søren J. Sørensen, Burton K. H. Chia, Bertrand

Denis, Jeff L. Froula, Zhong Wang, Robert Egan, Dongwan Don Kang, Jeffrey J. Cook, Charles Deltel, Michael Beckstette, Claire Lemaitre, Pierre Peterlongo, Guillaume Rizk, Dominique Lavenier, Yu-Wei Wu, Steven W. Singer, Chirag Jain, Marc Strous, Heiner Klingenberg, Peter Meinicke, Michael Barton, Thomas Lingner, Hsin-Hung Lin, Yu-Chieh Liao, Genivaldo Gueiros Z. Silva, Daniel A. Cuevas, Robert A. Edwards, Surya Saha, Vitor C. Piro, Bernhard Y. Renard, Mihai Pop, Hans-Peter Klenk, Markus Göker, Nikos C. Kyrpides, Tanja Woyke, Julia A. Vorholt, Paul Schulze-Lefert, Edward M. Rubin, Aaron E. Darling, Thomas Rattei, and Alice C. McHardy. Critical Assessment of Metagenome Interpretation – a benchmark of computational metagenomics software. page 099127. doi: 10.1101/099127. 62

[253] Torsten Seemann. Prokka: Rapid prokaryotic genome annotation. 30(14):2068–2069. ISSN 1367-4811. doi: 10.1093/bioinformatics/btu153. 57

[254] E. Seklemova, A. Pavlova, and K. Kovacheva. Biostimulation-based bioremediation of diesel fuel: Field demonstration. 12(5):311–316. ISSN 0923-9820. URL https://doi.org/10.1023/A:1014356223118. 113

[255] Mindy Selman, Suzie Greenhalgh, Robert Diaz, and Zachary Sugg. Eutrophication And Hypoxia In Coastal Areas: A Global Assessment Of The State Of Knowledge. page 6. 15

[256] Justin R. Seymour, Martina A. Doblin, Thomas C. Jeffries, Mark V. Brown, Kelly Newton, Peter J. Ralph, Mark Baird, and James G. Mitchell. Contrasting microbial assemblages in adjacent water masses associated with the East Australian Current. 4 (5):548–555. ISSN 1758-2229. doi: 10.1111/j.1758-2229.2012.00362.x. 27

[257] C. E. Shannon. A Mathematical Theory of Communication. 27(3):379–423. ISSN 00058580. doi: 10.1002/j.1538-7305.1948.tb01338.x. 31

[258] A. Sherry, N. D. Gray, A. K. Ditchfield, C. M. Aitken, D. M. Jones, W. F. M. Röling, C. Hallmann, S. R. Larter, B. F. J. Bowler, and I. M. Head. Anaerobic biodegradation of crude oil under sulphate-reducing conditions leads to only modest enrichment of

recognized sulphate-reducing taxa. 81:105–113. ISSN 0964-8305. doi: 10.1016/j.ibiod.2012.04.009. 93

[259] A. N. Shkidchenko and M. U. Arinbasarov. Study of the Oil-Degrading Activity of Caspian Shore Microflora. 38(5):433–436. ISSN 0003-6838, 1608-3024. doi: 10.1023/A:1019912300840. 00001. 27, 92

[260] M. Sibuet and K. Olu-Le Roy. Cold Seep Communities on Continental Margins: Structure and Quantitative Distribution Relative to Geological and Fluid Venting Patterns. pages 235–251. doi: 10.1007/978-3-662-05127-6_15. 00106. 24

[261] Zbyněk Šidák. Rectangular Confidence Regions for the Means of Multivariate Normal Distributions. 62(318):626–633. ISSN 0162-1459. doi: 10.1080/01621459.1967.10482935. 100

[262] J. Sikkema, J. A. de Bont, and B. Poolman. Mechanisms of membrane toxicity of hydrocarbons. 59(2):201–222. ISSN 0146-0749. URL https://mmbr.asm.org/content/59/2/201. 3

[263] Anna Sikora, Anna Detman, Damian Mielecki, Aleksandra Chojnacka, and Mieczysław Błaszczyk. Searching for Metabolic Pathways of Anaerobic Digestion: A Useful List of the Key Enzymes. doi: 10.5772/intechopen.81256. 16, 53

[264] E. H. Simpson. Measurement of Diversity. 163(4148):688. ISSN 1476-4687. doi: 10.1038/163688a0. 31

[265] Ajay Singh and Owen P. Ward. *Biodegradation and Bioremediation*. Springer Science & Business Media. ISBN 978-3-540-21101-3. 2

[266] Gregory F. Slater, Helen K. White, Timothy I. Eglinton, and Christopher M. Reddy. Determination of Microbial Carbon Sources in Petroleum Contaminated Sediments Using Molecular 14C Analysis. 39(8):2552–2558. ISSN 0013-936X. doi: 10.1021/es048669j. 4, 6

[267] Mark B. Smith, Andrea M. Rocha, Chris S. Smillie, Scott W. Olesen, Charles Paradis, Liyou Wu, James H. Campbell, Julian L. Fortney, Tonia L. Mehlhorn, Kenneth A. Lowe, Jennifer E. Earles, Jana Phillips, Steve M. Techtmann, Dominique C. Joyner, Dwayne A. Elias, Kathryn L. Bailey, Richard A. Hurt, Sarah P. Preheim, Matthew C. Sanders, Joy Yang, Marcella A. Mueller, Scott Brooks, David B. Watson, Ping Zhang, Zhili He, Eric A. Dubinsky, Paul D. Adams, Adam P. Arkin, Matthew W. Fields, Jizhong Zhou, Eric J. Alm, and Terry C. Hazen. Natural Bacterial Communities Serve as Quantitative Geochemical Biosensors. 6(3):e00326–15, . ISSN 2150-7511. doi: 10.1128/mBio.00326-15. 00000. 93

[268] Robert L. Smith, Adriana Huyer, J. Stuart Godfrey, and John A. Church. The Leeuwin Current off Western Australia, 1986–1987. 21(2):323–345, . ISSN 0022-3670. doi: 10.1175/1520-0485(1991)021<0323:TLCOWA>2.0.CO;2. 27

[269] Nitzan Soffer, Jesse Zaneveld, and Rebecca Vega Thurber. Phage-bacteria network analysis and its implication for the understanding of coral disease. 17(4):1203–1218. ISSN 1462-2920. doi: 10.1111/1462-2920.12553. 109

[270] Jill A. Sohm, Jason A. Hilton, Abigail E. Noble, Jonathan P. Zehr, Mak A. Saito, and Eric A. Webb. Nitrogen fixation in the South Atlantic Gyre and the Benguela Upwelling System. 38(16):L16608. ISSN 1944-8007. doi: 10.1029/2011GL048315. 00046. 20

[271] Liyan Song. A multiomics approach to study the microbiome response to phytoplankton blooms. 101(12):4863–4870. ISSN 1432-0614. doi: 10.1007/s00253-017-8330-5. 15

[272] Mengke Song, Longfei Jiang, Dayi Zhang, Chunling Luo, Hua Yin, Yongtao Li, and Gan Zhang. Identification of biphenyl-metabolising microbes in activated biosludge using cultivation-independent and -dependent approaches. 353:534–541. ISSN 1873-3336. doi: 10.1016/j.jhazmat.2018.04.028. 105

[273] James G. Speight and Karuna K. Arjoon. *Bioremediation of Petroleum and Petroleum Products*. John Wiley & Sons. ISBN 978-1-118-52832-7. 2, 3, 12

[274] Ian F. Spellerberg and Peter J. Fedor. A Tribute to Claude Shannon (1916-2001) and a Plea for More Rigorous Use of Species Richness, Species Diversity and the 'Shannon-Wiener' Index. 12(3):177–179. ISSN 1466-8238. doi: 10.1046/j.1466-822X.2003.00015.x. 00344. 31

[275] P. J. Statham and V. Hart. Dissolved iron in the Cretan Sea (eastern Mediterranean). 50(4):1142–1148. ISSN 1939-5590. doi: 10.4319/lo.2005.50.4.1142. 22

[276] Lothar Stramma and Matthew England. On the water masses and mean circulation of the South Atlantic Ocean. 104:20863–20833. ISSN 0148-0227. doi: 10.1029/1999JC900139. 00448. 19

[277] H. I. M. Struckmeyer, A. K. Williams, R. Cowley, J. M. Totterdell, G. Lawrence, and G. W. O'Brien. Evaluation Of Hydrocarbon Seepage In The Great Australian Bight. 42(1):371–385. ISSN 2206-8996. doi: 10.1071/aj01020. 27

[278] Monica P. Suarez and Hanadi S. Rifai. Biodegradation Rates for Fuel Hydrocarbons and Chlorinated Solvents in Groundwater. 3(4):337–362. ISSN 1088-9868, 1547-6529. doi: 10.1080/10889869991219433. 92

[279] P. Suntharalingam and J. L. Sarmiento. Factors governing the oceanic nitrous oxide distribution: Simulations with an ocean general circulation model. 14(1):429–454. ISSN 1944-9224. doi: 10.1029/1999GB900032. 15

[280] Stephen M. Techtmann, Julian L. Fortney, Kati A. Ayers, Dominique C. Joyner, Thomas D. Linley, Susan M. Pfiffner, and Terry C. Hazen. The Unique Chemistry of Eastern Mediterranean Water Masses Selects for Distinct Microbial Communities by Depth. 10(3):e0120605, . ISSN 1932-6203. doi: 10.1371/journal.pone.0120605. 00003. 34, 44, 46, 53, 64, 66, 70

[281] Stephen M. Techtmann, Nagissa Mahmoudi, Kendall T. Whitt, Maria Fernanda Campa, Julian L. Fortney, Dominique C. Joyner, and Terry C. Hazen. Comparison of Thaumarchaeotal populations from four deep sea basins. 93(11), . ISSN 0168-6496. doi: 10.1093/femsec/fix128. 27, 105

[282] DR Thakker, H Yagi, W Levin, AW Wood, AH Conney, and DM Jerina. Polycyclic aromatic hydrocarbons: Metabolic activation to ultimate carcinogens. pages 177–242. 2

[283] The UniProt Consortium. UniProt: A hub for protein information. 43:D204–D212. ISSN 1362-4962, 0305-1048. doi: 10.1093/nar/gku989. 00000. 58

[284] T. F. Thingstad, M. D. Krom, R. F. C. Mantoura, G. a. F. Flaten, S. Groom, B. Herut, N. Kress, C. S. Law, A. Pasternak, P. Pitta, S. Psarra, F. Rassoulzadegan, T. Tanaka, A. Tselepides, P. Wassmann, E. M. S. Woodward, C. Wexels Riser, G. Zodiatis, and T. Zohary. Nature of Phosphorus Limitation in the Ultraoligotrophic Eastern Mediterranean. 309(5737):1068–1071. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.1112632. x, 10, 22, 23, 34

[285] Geng Tian, XuYang Yin, Hong Luo, XiaoHong Xu, Lars Bolund, and XiuQing Zhang. Sequencing bias: Comparison of different protocols of MicroRNA library construction. 10(1):64. ISSN 1472-6750. doi: 10.1186/1472-6750-10-64. 6

[286] B. P Tissot and D. H Welte. *Petroleum Formation and Occurrence*. Springer Berlin. ISBN 978-3-642-87815-2. OCLC: 864656145. 3

[287] Kathleen Trautwein, Christoph Feenders, Reiner Hulsch, Hanna S. Ruppersberg, Annemieke Strijkstra, Mirjam Kant, Jannes Vagts, Daniel Wünsch, Bernhard Michalke, Michael Maczka, Stefan Schulz, Helmut Hillebrand, Bernd Blasius, and Ralf Rabus. Non-Redfield, nutrient synergy and flexible internal elemental stoichiometry in a marine bacterium. 93(5). ISSN 1574-6941. doi: 10.1093/femsec/fix059. 109

[288] David Tropel and Jan Roelof van der Meer. Bacterial Transcriptional Regulators for Degradation Pathways of Aromatic Compounds. 68(3):474–500. ISSN 1092-2172, 1098-5557. doi: 10.1128/MMBR.68.3.474-500.2004. 12

[289] Andreas Tschech and Georg Fuchs. Anaerobic degradation of phenol via carboxylation to 4-hydroxybenzoate: In vitro study of isotope exchange between 14CO2 and 4-hydroxybenzoate. 152(6):594–599. ISSN 1432-072X. doi: 10.1007/BF00425493. 73

[290] Valentin S. Tuzhilkin and Alexey N. Kosarev. Thermohaline Structure and General Circulation of the Caspian Sea Waters. In Andrey G. Kostianoy and Aleksey N. Kosarev, editors, *The Caspian Sea Environment*, volume 5P, pages 33–57. Springer Berlin Heidelberg. ISBN 978-3-540-28281-5. doi: 10.1007/698_5_003. URL http://link.springer.com/10.1007/698_5_003. 00000. 25, 26

[291] Valentin S. Tuzhilkin, Damir N. Katunin, and Yury R. Nalbandov. Natural Chemistry of Caspian Sea Waters. In Andrey G. Kostianoy and Aleksey N. Kosarev, editors, *The Caspian Sea Environment*, The Handbook of Environmental Chemistry, pages 83–108. Springer Berlin Heidelberg. ISBN 978-3-540-31505-6. doi: 10.1007/698_5_005. URL https://doi.org/10.1007/698_5_005. 26, 91, 100

[292] United States, Coast Guard. On Scene Coordinator Report, Deepwater Horizon oil spill. URL https://repository.library.noaa.gov/view/noaa/283. 2

[293] Paul W. J. J. van der Wielen, Henk Bolhuis, Sara Borin, Daniele Daffonchio, Cesare Corselli, Laura Giuliano, Giuseppe D'Auria, Gert J. de Lange, Andreas Huebner, Sotirios P. Varnavas, John Thomson, Christian Tamburini, Danielle Marty, Terry J. McGenity, Kenneth N. Timmis, and BioDeep Scientific Party. The Enigma of Prokaryotic Life in Deep Hypersaline Anoxic Basins. 307(5706):121–123. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.1103569. 22

[294] Virginie van Dongen-Vogels, Justin R. Seymour, John F. Middleton, Jim G. Mitchell, and Laurent Seuront. Influence of local physical events on picophytoplankton spatial and temporal dynamics in South Australian continental shelf waters. 33(12):1825–1841. ISSN 0142-7873. doi: 10.1093/plankt/fbr077. 27

[295] Paul D. van Ruth, George G. Ganf, and Tim M. Ward. Hot-spots of primary productivity: An Alternative interpretation to Conventional upwelling models. 90 (3):142–158. ISSN 0272-7714. doi: 10.1016/j.ecss.2010.08.009. 27

[296] Ebrahim Vasheghani-Farahani and Mohammadreza Mehrnia. Bio-physicochemical treatment of oil contaminated sea water. 26(1–4):179–185. ISSN 0920-4105. doi: 10.1016/S0920-4105(00)00032-2. 00001. 27, 92

[297] M. G. Veliev, B. Danielsson, M. A. Salmanov, S. R. Alieva, and N. R. Bektashi. Biodegradation of Baku oil and hydrocarbons by micromycetes. 48(1):56–62. ISSN 0965-5441, 1555-6239. doi: 10.1134/S0965544108010118. 00004. 27, 92

[298] Frederick von Netzer, Giovanni Pilloni, Sara Kleindienst, Martin Krüger, Katrin Knittel, Friederike Gründger, and Tillmann Lueders. Enhanced Gene Detection Assays for Fumarate-Adding Enzymes Allow Uncovering of Anaerobic Hydrocarbon Degraders in Terrestrial and Marine Systems. 79(2):543–552. ISSN 0099-2240, 1098-5336. doi: 10.1128/AEM.02362-12. 10

[299] Sophie M. Vonk, David J. Hollander, and AlberTinka J. Murk. Was the extreme and wide-spread marine oil-snow sedimentation and flocculent accumulation (MOSSFA) event during the Deepwater Horizon blow-out unique? 100(1):5–12. ISSN 0025-326X. doi: 10.1016/j.marpolbul.2015.08.023. 3

[300] Maren Voss, Hermann W. Bange, Joachim W. Dippner, Jack J. Middelburg, Joseph P. Montoya, and Bess Ward. The marine nitrogen cycle: Recent discoveries, uncertainties and the potential relevance of climate change. 368(1621). ISSN 0962-8436. doi: 10.1098/rstb.2013.0121. 15, 81

[301] Hui Wang, Shuangfei Zhang, Amit Pratush, Xueying Ye, Jinli Xie, Huan Wei, Chongran Sun, and Zhong Hu. Acclimation of Culturable Bacterial Communities under the Stresses of Different Organic Compounds. 9:225, . ISSN 1664-302X. doi: 10.3389/fmicb.2018.00225. 103

[302] Qingren Wang, Shouan Zhang, Yuncong Li, and Waldemar Klassen. Potential Approaches to Improving Biodegradation of Hydrocarbons for Bioremediation of Crude Oil Pollution. 02(01):47–55, . ISSN 2152-2197, 2152-2219. doi: 10.4236/jep.2011.21005. 21

[303] Zhendi Wang and Scott Stout. *Oil Spill Environmental Forensics: Fingerprinting and Source Identification*. Elsevier. ISBN 978-0-08-046773-3. 12

[304] A.J. Ward, D.M. Lewis, and F.B. Green. Anaerobic digestion of algae biomass: A review. 5:204–214. ISSN 22119264. doi: 10.1016/j.algal.2014.02.001. 52

[305] Boris Wawrik, Margarita Mendivelso, Victoria A. Parisi, Joseph M. Suflita, Irene A. Davidova, Christopher R. Marks, Van Nostrand, Joy D, Yuting Liang, Jizhong Zhou, Brad J. Huizinga, Dariusz Strąpoć, and Amy V. Callaghan. Field and laboratory studies on the bioconversion of coal to methane in the San Juan Basin. 81(1):26–42. ISSN 0168-6496. doi: 10.1111/j.1574-6941.2011.01272.x. 92

[306] Robert J. Whittaker, Katherine J. Willis, and Richard Field. Scale and species richness: Towards a general, hierarchical theory of species diversity: Towards a general theory of diversity. 28(4):453–470. ISSN 03050270. doi: 10.1046/j.1365-2699.2001.00563.x. 31

[307] F. Widdel, K. Knittel, and A. Galushko. Anaerobic Hydrocarbon-Degrading Microorganisms: An Overview. In Kenneth N. Timmis, editor, *Handbook of Hydrocarbon and Lipid Microbiology*, pages 1997–2021. Springer Berlin Heidelberg. ISBN 978-3-540-77587-4. doi: 10.1007/978-3-540-77587-4_146. URL https://doi.org/10.1007/978-3-540-77587-4_146. 92

[308] Jutta Wiese, Vera Thiel, Andrea Gärtner, Rolf Schmaljohann, and Johannes F. Imhoff. Kiloniella laminariae gen. nov., sp. nov., an alphaproteobacterium from the marine macroalga Laminaria saccharina. 59:350–356. ISSN 1466-5026. doi: 10.1099/ijs.0.001651-0. 109

[309] Dagmar Woebken, Bernhard M. Fuchs, Marcel M. M. Kuypers, and Rudolf Amann. Potential Interactions of Particle-Associated Anammox Bacteria with Bacterial and Archaeal Partners in the Namibian Upwelling System. 73(14):4648–4657. ISSN 0099-2240, 1098-5336. doi: 10.1128/AEM.02774-06. 00163. 20

[310] Alexandra Z. Worden, Michael J. Follows, Stephen J. Giovannoni, Susanne Wilken, Amy E. Zimmerman, and Patrick J. Keeling. Rethinking the marine carbon cycle: Factoring in the multifarious lifestyles of microbes. 347(6223):1257594. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.1257594. 15

[311] Weiling Xue and David Warshawsky. Metabolic activation of polycyclic and heterocyclic aromatic hydrocarbons and DNA damage: A review. 206(1):73–93. ISSN 0041008X. doi: 10.1016/j.taap.2004.11.006. 2

[312] Pelin Yilmaz, Laura Wegener Parfrey, Pablo Yarza, Jan Gerken, Elmar Pruesse, Christian Quast, Timmy Schweer, Jörg Peplies, Wolfgang Ludwig, and Frank Oliver Glöckner. The SILVA and "All-species Living Tree Project (LTP)" taxonomic frameworks. 42(D1):D643–D648. ISSN 0305-1048. doi: 10.1093/nar/gkt1209. 31

[313] T Yokokawa, D De Corte, E Sintes, and Gj Herndl. Spatial patterns of bacterial abundance, activity and community composition in relation to water masses in the eastern Mediterranean Sea. 59(2):185–195. ISSN 0948-3055, 1616-1564. doi: 10.3354/ame01393. 22, 33

[314] M. Youssef, G. E. El-Taweel, A. Y. El-Naggar, S. E. El-Hawary, M. A. El-Meleigy, and S. A. Ahmed. Hydrocarbon degrading bacteria as indicator of petroleum pollution in Ismailia Canal, Egypt. 8(10):1226–1233. ISSN 1818-4952. URL https://www.cabdirect.org/cabdirect/abstract/20103195588. 24

[315] Chengsheng Zhu, Maximilian Miller, Srinayani Marpaka, Pavel Vaysberg, Malte C. Rühlemann, Guojun Wu, Femke-Anouska Heinsen, Marie Tempel, Liping Zhao, Wolfgang Lieb, Andre Franke, and Yana Bromberg. Functional sequencing read annotation for high precision microbiome analysis. 46(4):e23–e23. ISSN 0305-1048. doi: 10.1093/nar/gkx1209. 62, 70, 73

[316] Ines Zrafi-Nouira, Sonda Guermazi, Rakia Chouari, Nimer M. D. Safi, Eric Pelletier, Amina Bakhrouf, Dalila Saidane-Mosbahi, and Abdelghani Sghir. Molecular diversity analysis and bacterial population dynamics of an adapted seawater microbiota during the degradation of Tunisian zarzatine oil. 20(4):467–486. ISSN 1572-9729. doi: 10.1007/s10532-008-9235-x. 24, 25

Vita

John Miller received a B.S. in Biochemistry and a B.S. in Physiological Science from the University of California, Los Angeles in 2011. His main areas of interest are computational biology, bioinformatics, and data science.