DIVERSITY OF BIODEGRADATIVE GENE POPULATIONS IN AQUATIC SEDIMENTS EXAMINED BY GENE-TARGETED METAGENOMICS

by

ELYSE ANNE RODGERS-VIEIRA

A Dissertation submitted to the

Graduate School-New Brunswick

Rutgers, The State University of New Jersey

and

The Graduate School of Biomedical Sciences

University of Medicine and Dentistry of New Jersey

In partial fulfillment of the requirements

For the degree of

Doctor of Philosophy

Graduate Program in Microbiology and Molecular Genetics

Written under the direction of

Professor Gerben J. Zylstra

And approved by

_____

_____

_____

_____

_____

New Brunswick, New Jersey

January, 2012

ABSTRACT OF THE DISSERTATION

DIVERSITY OF BIODEGRADATIVE GENE POPULATIONS IN AQUATIC SEDIMENTS EXAMINED BY GENE-TARGETED METAGENOMICS

by

ELYSE ANNE RODGERS-VIEIRA

Dissertation Director: Professor Gerben J. Zylstra

Alkanes are common environmental pollutants in soil and water. The degradation of medium length *n*-alkanes is initiated under aerobic conditions by alkane monooxygenases which add one atom of molecular oxygen to the terminal carbon resulting in an alkanol product. Alkane monooxygenases fall into two distinct classes: the integral membrane bound AlkB family and the cytoplasmic cytochrome P450 family. Gene-targeted metagenomics was used to examine the microbial diversity and distribution of these two types of alkane monooxygenases in sediments in the United States and Central Asia.

The Passaic River in Newark, New Jersey has a long history of industrial pollution making it an ideal site to study monooxygenase diversity. 16S rRNA and alkane monooxgyenase gene populations were analyzed by pyrosequencing to determine if sampling location on the river influenced the microbial community and if triplicate enrichments yield comparable results. Samples were collected at an arbitrary start point

(0 meters) and at 10 and 1000 meters down the river. The replicates were similar to each other at two of the three sampling locations and differed slightly at 1000 meters.

Sediments from rivers and streams in Central Asia were compared to determine if novel alkane monooxygenase families could be found in a largely unstudied geographic region. The 16S rRNA and monooxygenase gene communities recovered from sediment and enrichments originating from disparate environments with varied anthropogenic influence were compared by pyrosequencing. Novel alkane monooxygenase populations were recovered from sites in Central Asia and comparisons between sites showed that each population was distinct due to their distant geographic origins.

The effect of salinity on alkane monooxygenase populations was examined in sediments obtained from Puerto Rico. Samples were collected from the Port of San Juan, an estuary, mangroves, and shore locations. Salinity was not the major determinant of alkane monooxygenase community composition in hexadecane enrichment cultures. The type of environment (mangrove compared to shore or port locations) had the greatest affect on the gene populations recovered.

ACKNOWLEDGMENTS

First, I would to express my gratitude to my thesis advisor Dr. Gerben Zylstra. I would like to thank him for supporting my research and allowing me many opportunities for scientific growth. This work would have not been possible without his guidance, advice, and patience throughout my graduate studies.

I am indebted to Drs. Jerry Kukor, Loren Launen, Costantino Vetriani, and Lily Young for taking the time to serve as members of my doctoral dissertation committee and for offering their expertise and suggestions throughout the process.

I would like to thank fellow lab members, graduate students, and colleagues including Hung-Kuang Chang, Michael Murillo, Ruyang Han, Hisako Masuda, Lora McGuiness, Sinead Ni Chadhain, Jong-Chan Chae, Boris Wawrik, Maria Rivera, Adam Mumford, and Abigail Porter. Thank you for all of the helpful discussions and technical assistance. I also had the pleasure of working with two wonderful undergraduate researchers, Virian Serei and Arwa Abbas.

Several colleagues at Michigan State University were instrumental in assisting me with the pyrosequencing and data analysis including, Jim Cole, Qiong Wang, Benli Chai, Jordon Fish, and John Quensen. I greatly appreciate their willingness to share their bioinformatics knowledge with me. I would also like to extend a thank you to Jose Perez-Jimenez at University of Turabo at Gurabo for assistance with the Puerto Rico project.

I am grateful to many scientific mentors and teachers for their encouragement and introducing me to science and research. Their influence was instrumental in my decision to pursue a career in science.

Also, thank you to my family and friends for their support. This would have not been possible without you.

TABLE OF CONTENTS

ABSTRACT OF THE DISSERTATION……ii

ACKNOWLEDGEMENTS……iv

TABLE OF CONTENTS……vi

LIST OF FIGURES

Chapter One

Figure 1: Aerobic hydrocarbon degrading bacteria……..25

Figure 2: Alkane degradation by monooxygenases……..26

Figure 3: Alkane degradation pathway……27

Figure 4: AlkB monooxygenase system found in *P. putida*……..29

Figure 5: Alkane hydroxylase system P450 Class I……...31

Figure 6: Ribosomal database project pyrosequencing pipeline……...32

Chapter Two

Figure 1: 16S rRNA phyla before and after enrichment……...47

Figure 2: 16S rRNA proteobacterial classes……..48

Figure 3: The principal component analysis of each sampling location in relationship to each other……..49

Figure 4: Cluster dendrograms and heatmaps……..51

Figure 5: Neighbor joining phylogenetic trees…….53

Chapter Three

Figure 1: Central Asia map……...75

Figure 2: TRFLP cluster analysis…….……76

Figure 3: AlkB phylogenetic tree……..…….80

Figure 4: *alkB* gene family distribution……..……..81

Figure 5: 16S rRNA phyla before and after enrichment……..…….83

Figure 6: 16S rRNA proteobacteria classes……..…….84

Figure 7: AlkB cluster dendrograms and heatmap……..….85

Figure 8: P450 cluster dendrograms and heatmap……....86

Figure 9: Percent similarity of experimental sequences to known sequences…..88

Chapter Four

Figure 1: Map of sampling locations……104

Figure 2: 16S rRNA phyla……..…107

Figure 3: 16S rRNA Proteobacterial classes…….108

Figure 4: AlkB diversity……109

Figure 5: P450 diversity……111

Figure 6: AlkB cluster dendrogram and heatmap…….113

Figure 7: P450 cluster dendrogram and heatmap…….114

LIST OF TABLES

Chapter One

Table 1: AlkB and P450 enzyme classes involved in alkane oxidation……..…28

Table 2: Genes involved in alkane degradation in *P. putida* GP01……..…30

Chapter Two

Table 1: Protein family composition…….….55

Chapter Three

Table 1: Central Asia sampling locations…….74

Table 2: 16S rRNA sequencing statistics……...77

Table 3: *alkB* sequencing statistics……...78

Table 4: P450 monooxygenase gene sequencing statistics……79

Chapter Four

Table 1: Number of sequences obtained from pyrosequencing……..105

Table 2: Water sample measurements……106

CHAPTER ONE

INTRODUCTION

Petroleum pollution and oxygenases…….…..……1

DNA sequencing technologies……...…….…1

Alkane degradation……..……2

Alkane monooxygenases……..……...3

AlkB monooxygenases……..….…4

P450 monooxygenases……..…..….4

Gene organization……7

Gene regulation……7

Hydrocarbon uptake…….8

Enzyme specificity……...8

Differential expression...……10

Survival in a range of environments……..11

Commercial application of oxygenases…….11

Culturing environmental samples……..13

Biogeography…….14

Bioinformatics and next generation sequencing……15

EXPERIMENTAL PLAN OVERVIEW……...16

CHAPTER TWO

SHIFTS IN ALKANE MONOOXGENASE GENE POPULATIONS IN PASSAIC RIVER SEDIMENT FOLLOWING SELECTIVE ENRICHMENT……33

Abstract……..34

Introduction……35

Materials and Methods……...…37

Results……39

Discussion……..41

CHAPTER THREE

ALKANE MONOOXYGENASE DIVERSITY RECOVERED FROM CENTRAL ASIA SEDIMENTS……..…56

Abstract……..……57

Introduction……..….….58

Materials and Methods……..…….60

Results……..……..64

Discussion……..66

CHAPTER FOUR

EFFECT OF BIOGEOGRAPHY AND SALINITY ON ALKANE MONOOXGENASE POPULATIONS RECOVERED FROM PUERTO RICAN SEDIMENT……89

Abstract……..90

Introduction……91

Materials and Methods……...93

Results……96

Discussion……..98

CHAPTER FIVE

CONCLUSIONS AND SIGNIFICANCE……...…115

REFERENCES

Chapter One……18

Chapter Two……44

Chapter Three……..71

Chapter Four……101

Curriculum Vitae……118

CHAPTER ONE

INTRODUCTION

Petroleum pollution and oxygenases

Fuel-based and industrial pollution contribute major environmental contaminants in both soil and water. Petroleum is also one of the most common pollutants in the marine environment. Oxygenases are metalloenzymes that are expressed in many naturally occurring bacteria in these environments. These enzymes catalyze the first step in the biodegradation of various pollutants. Microbes involved in biodegradation are metabolically able to transform organic contaminants into less hazardous forms. Microbes with the ability to degrade alkanes are isolated from both contaminated and non-contaminated sites in the environment. Bioremediation of alkanes in released in oil spills, and in polluted soils by natural attenuation and nutrient addition is a widely accepted method for site cleanup (1, 37, 41, 43, 52, 65).

DNA sequencing technologies

Until the advent of molecular techniques, knowledge of microbial diversity was limited to laboratory culturing techniques. The availability of high-throughput DNA sequencing technologies have made the study of microbial diversity more attainable and allowed for better estimates of that diversity (49). Metagenomics applies this technology to understand how genes affect microbial function within the community as a whole (15). Gene-targeted metagenomics focuses sequencing efforts on a certain functional gene of interest. It is imperative that primers are designed to be as inclusive of an entire gene

family as possible, and that conserved regions are present in the gene within the limits for amplification by emulsion PCR (29). Operational taxonomic unit (OTU)-based analysis does not sufficiently detect variations within the environment unless there is sufficient sequencing coverage. If environments are similar, a greater sequencing depth is required to obtain an accurate definition of microbial diversity (36). Gene-targeted metagenomics allows for better determination of community structure at a greater sequencing depth, while focusing on genes that are associated with functions that are key to understanding community dynamics. In this work, the functional genes studied were alkane monooxygenases, which are involved in alkane degradation.

Alkane degradation

It is important to understand the alkane degradation process because these compounds are such prevalent environmental contaminants. Alkanes are hydrocarbons that comprise 20-50% of crude oil (71). Straight chain *n*-alkanes are produced by geological processes through the reduction of organic matter, and they are also naturally occurring products of plants and microbes (8). Alkanes are composed completely of carbon and hydrogen atoms, and are chemically very stable due to the strength of carbon-carbon bonds. Using enzyme systems and metabolic pathways, many bacteria have garnered the ability to utilize *n*-alkanes as a carbon and energy source. Alkane degraders are rather cosmopolitan, as most soil, sand, and sediment possess microbes with hydrocarbon degradation capabilities (24, 71, 83). However, in most cases they are more abundant in oil-polluted environments (24, 25). The ability to degrade saturated hydrocarbons is characteristic of many bacteria. Hydrocarbon degraders are distributed

throughout many different types of bacteria, and more are being discovered with the use of metagenomics (Figure 1).

Alkane monooxygenases

The initial step in alkane metabolism under aerobic conditions is performed by a hydroxylase that acts on the terminal carbon. Oxygen is used as a reactant in the activation of the otherwise stable *n*-alkane. The hydroxylase introduces an oxygen atom gained from molecular oxygen, creating 1-alkanols. This reaction results in the creation of a reactive oxygen species (51). Two families of alkane monooxygenases catalyze this regio- and stereo-specific reaction: AlkB and P450s (Figure 2). These enzymes serve an important role in oil degradation and bioremediation in the environment. They each have certain specificities, acting upon different groups of alkanes (Table 1). Following the formation of the 1-alkanol, the alcohol is further oxidized to an aldehyde by alcohol dehydrogenase. The aldehyde is they acted upon by an aldehyde dehydrogenase resulting in a fatty acid. The fatty acids are then paired with CoA and metabolized further by β-oxidation to carbon dioxide (Figure 3) (5, 77, 79, 83). Alkane monooxygenases can have negative effects on microbes because they produce active oxygen species or hydrogen peroxide that can affect the stability of the oxygenase and the host (35, 66). Alkane hydroxylases can loosely be divided into three groups: those that act on short chain alkanes (C_1-C_4, which are acted upon by methane monooxygenase type enzymes), those that act on medium length alkanes (C_5-C_{16}, oxidized by AlkB or P450 enzymes), and those that act on long-length (C_{17}, or greater that are oxidized by P450s) (8).

AlkB monooxygenases

AlkB enzymes are non-heme diiron integral membrane-bound proteins. The first was discovered in a hexane degrading isolate named *Pseudomonas oleovorans* that was later reclassified as *Pseudomonas putida* (7, 73, 76). Homologs are present in an array of alkane degrading Alphaproteobacteria, Betaproteobacteria, Gammaptroteobacteria, in Gram-positive high G+C content bacteria (68). In *Pseudomonas putida* GPo1 the alkane hydroxylase system is made up of three components: an alkane hydroxylase (AlkB), rubredoxin (AlkG), and rubredoxin reductase (AlkT) (Figure 4). AlkB is a cytoplasmic bound protein that works in conjunction with two electron transfer proteins, AlkG and AlkT. The rubredoxin reductase (AlkT) is a cytoplasmic bound enzyme uses cofactor FAD to transfer electrons from NADH to the rubredoxin (AlkG). There is currently no crystal structure for AlkB, but evidence shows that it probably has six transmembrane segments with a catalytic site that is on the cytoplasmic end of the enzyme that contains four conserved His sequence motifs (60, 74).

Alkane hydroxylase systems in two *Rhodococcus* strains (Q15 and B-16531) have at least four alkane monooxygenase gene homologs (alkB1, 2, 3 and 4). Other systems that contain multiple degradative enzymes have also been described in *Rhodococcus* (85).

P450 monooxygenases

A second type of alkane oxygenase is Cytochrome P450. P450 monooxygenases are soluble enzymes that comprise one of the largest known heme-containing superfamilies. P450 enzymes have been identified in every kingdom of life (44). P450s

mediate monooxygenation reactions in a regio-specific manner. They are known to perform many functions including expoxidation, heteroatom oxygenation, dealkylation, desaturation, and also dehalogenation (28). Known alkane hydroxylating P450s fall into two groups, class I and class II depending on the electron transfer system type they employ. The class I P450s consist of three parts: the oxygenase, an iron-sulfur ferredoxin, and an FAD-containing ferredoxin reductase (Figure 5) (46, 68). This class includes bacteria and mitochondrial P450s (46). Class II P450s have two parts: a membrane-bound oxygenase and a reductase (68). They are microsomal P450s and receive electrons from NADPH cytochrome P450 reductase. The reductase contains cofactors, FAD and FMN (46).

In addition to alkane degradation, bacterial cytochrome P450s are involved in the biosynthesis of antibiotics and other secondary metabolites and monooxgenation of various organic compounds (33). The first soluble cytochrome P450 enzyme that was shown to hydroxylate alkanes is CYP153A6. Short length aliphatic and cyclic hydrocarbons can effectively bind to the active site, but medium and long chain length alkanes are preferred enzyme substrates. This P450 is regio-selective, producing greater than 95% 1-alkanol product (18). Many other bacterial strains have since been identified that metabolize linear alkanes (71). However, at this time the overall number of known bacterial P450s is low. It is likely that many unknown genes exist in diverse bacterial species that are difficult to culture or isolate in various environments.

The cytochrome P450 enzymes contain a cysteine thiolate which is coordinated to heme iron atom. The heme center is the site where molecular oxygen is activated. This activation results in the insertion of molecular oxygen into the substrate (46). The

thiolate acts as a ligand that is necessary for O-O bond cleavage by the enzyme and also facilitates proper protein folding (6). Mutation studies show that alterations in the cysteine thiolate render P450s catalytically inactive (86). Once a substrate binds to a P450 monooxygenase it becomes catalytically active. In most cases, substrate binding is required for activation which involves the transfer of one electron to the iron center. However, there are some P450s that can become activated prior to binding (45). The heme is then reduced from the ferric to the ferrous state with electrons transferred from NADPH through the ferredoxin reductase and ferredoxin to the P450 (45, 46). The ferrous center then binds the oxygen followed by the transfer of a second electron to the complex yielding a ferro-hydroperoxo complex. Breaking the O-O bond of the oxygen results in the formation of water. The mechanisms for hydroxylation of hydrocarbons is still up for debate. Two major theories exist, while the most widely accepted is that the ferryl oxygen removes the hydrogen from the hydrocarbon, and the oxygen and hydrogen bound to the iron center of the P450 is then recombined with hydrocarbon (45). In a P450 system, oxygen is incorporated into the alkane substrate resulting in water and a 1-alkanol. Reducing equivalents are provided by NADPH (19).

A cytochrome P450 enzyme from *Novosphingobium aromaticivorans* DSM12444 that oxidizes alkanes and alkyl side chains, was recently purified and crystalized. However, further resolution of the crystal structure is needed until the full structure is known (88). At this time the number of known bacterial P450s is low. However, it is likely that many unknown genes exist in diverse bacterial species that are difficult to culture or isolate in various environments (33).

Gene organization

The genes involved in alkane oxidation are organized in various ways among alkane-degrading bacteria, and even the clustering of these is not necessarily the same in all bacteria. In some strains the genes are not linked to each other, and many are distributed throughout the genome (63). Sometimes the hydroxylase is not near the rubredoxin and the rubredoxin reductase or the alcohol and aldehyde dehydrogenases. Even though there can be more than one hydroxylase in one bacterial strain, they are not necessarily found near each other, and the genes that regulate their expression may not be near the hydroxylases either (51). The alkane degradation genes in *Pseudomonas putida* GP01 (previously known as *Pseudomonas oleovorans*) are well characterized. The genes are clustered into two groups, one with *alkBGFHJKL*, and the other with *alkST* (Table 2). These genes are mapped to the well-characterized OCT plasmid which codes for a number of proteins involved in the oxidation of (C_6 –C_{12}) *n*-alkanes in *Pseudomonas putida* GP01 (13, 67).

Gene regulation

Regulation of the alkane degradation pathway is generally highly controlled. There are regulators that control gene expression so that they are only expressed when the bacteria is exposed to alkanes that it can metabolize. Regulators include LuxR/MalT, AraC/XylS, GntR, and other regulator families (51). The regulators belong to non-related families. Evidence suggests that increasing amounts of alkanes accumulate in the cytoplasmic membrane where they interact with the regulators. The nature of this interaction is still under investigation, however, the transcription of the alkane

hydroxylase gene *alkB* in *Alkanivorax borkumensis* and *Pseudomonas putida* is regulated by AlkS (32, 57, 72). This regulator appears to be associated with the cytoplasmic side of the membrane where it may interact with the alkanes that act as effectors which in turn result in hydroxylase expression (51, 54). Alkanes are not usually the choice carbon source for bacteria. They usually have the ability to utilize different carbon sources, but prefer certain growth substrates over others in a process called catabolite repression. In *P. putida, P. aeruginosa, B. cepacia, P. butanovora,* and *Acinetobacter* sp. alkane degradation can actually be downregulated when the preferred carbon sources are present (16, 42, 48, 64).

Hydrocarbon uptake

Microbial cells must come in contact with aliphatic alkanes to facilitate uptake of the hydrocarbons and to upregulate genes necessary for their degradation. In alkane degraders that metabolize alkanes of 10 carbons in length or greater, there are two ways that bacteria come in contact with the hydrocarbons. The first is by direct contact and the second is with the help of a biosurfactant to put the cell in contact with the hydrocarbon droplets in a process of micellular transfer (9, 10, 20, 22, 26, 61, 83). Many microbes that are alkane degraders produce an array of surfactants that vary in composition, to assist in hydrocarbon uptake (26, 53).

Enzyme specificity

AlkB hydroxylases can only accommodate certain alkane chain-lengths, due to restrictions in the enzyme active site. Certain amino acids in alkane hydroxylases

determine the length of alkane that the enzyme can act upon. For instance, AlkB hydroxylase mutants that gained the ability to oxidize long chain alkanes had a tryptophan residue mutated in the middle of one of the six transmembrane helices. This tryptophan limits the substrate specificity of AlkB (75). AlkBs show protein similarities to a larger family of enzymes that include desaturases, epoxsidases, decarbonylases, and methyl oxidases, due to a conserved eight histidine motif (59).

Some alkane degrading bacteria grow on alkanes that vary in carbon lengths. To increase their substrate range they often possess more than one alkane hydroxylase gene, with mixtures of *alkB* and P450 genes (2, 69, 70, 71). There are several other reasons why bacteria would find it beneficial to have more than one alkane monooxygenase. Different enzymes may be expressed in different stages of bacterial growth such as in *Pseudomonas aeruginosa* PAO1 which contains two AlkB hydroxylases with similar substrate ranges (42, 62). The enzymes may have different affinity constants for different alkanes, so utilizing one monooxygenase may be most favorable with one alkane substrate and not with another. Also, some of the *alkB* genes detected in bacteria may be pseudogenes (68).

Recombinant hosts for alkane hydroxylases can provide functional rubredoxin and rubredoxin reductase to evaluate the activity of novel *alkB* genes. This approach was employed to determine the length of alkane that certain alkane hydroxylases are able to oxidize. *E. coli* and *P. putida* were selected as hosts to test growth on medium length alkanes, while *P. fluorescens* was selected to test activity with long alkanes. Only 10-20% of the normal hydroxylase activity was necessary for the recombinants to survive on alkanes. Testing *alkB* genes in these hosts allows for a mechanism to select for a certain

alkane substrate range in cloned *alkB* genes and provides a better understanding of what residues are involved in substrate binding (63).

Differential expression

When more than one *alkB* or P450 is present in a bacterial strain, the cell separately regulates expression of each of the hydroxylase genes. The expression is dependent upon the chain length of an alkane, or may be only expressed when certain physiological parameters are met. This may be regulated by global regulators that receive physiological signals, or by regulators with greater specificity. For example, *Pseudomonas aeruginosa* and *Alkanivorax borkumensis* have more than one alkane hydroxylase systems which are differentially expressed (23, 42).

Gram-positive *Dietzia* sp. H0B was one of the dominant bacterial strains isolated from the microbial community following the Prestige Oil Spill. The oil spill occurred just off the coast of Galicia, in northwestern Spain when an oil tanker was damaged in a storm. The ship then traveled northwest and then south to find harbor, but ended up sinking and spilling 20 million gallons of oil in the process. Fuel oils like that released during the spill, are mixtures of aliphatic compounds including *n*-alkanes, aromatics, resins, and asphaltenes. Characterization of the Cornebacterinaeae *Dietzia* sp. H0B revealed three alkane hydroxylase genes; one *alkB* homolog, and two P450 homologs. These genes differed from known hydroxylases in their sequence, and also showed novel activity with 8-hexadecene as an intermediate in the degradation pathway. 8-hexadecene was produced by the strain when grown on hexadecane suggesting that 8-hexadecene may be an intermediate in the degradation pathway. This was the first reported instance

where the oxidation occurred at the middle of the alkane instead of at the terminal or subterminal positions (2). Gram-negative *Alkcanivorax* sp. SK2 responded to the oil spill in even higher numbers than *Dietzia* sp. H0B due to a quicker doubling time. Similar opposing growth between Gram-positive and negative bacteria was also observed previously by Margesin *et al* at an oil contaminated site (39).

Survival in a range of environments

Alkane degradation thrives in a range of habitats and temperatures. Biologically mediated petroleum hydrocarbon degradation was observed in psychrophilic microbes in Arctic, Alpine and Antarctic soils (1, 37, 39). Contaminated, as well as pristine environments contained a high frequency of *alkB* genes, and no correlation was found between the extent of contamination at a site and *alkB* expression (40). At the other end of the temperature spectrum, *alkB* genes were amplified from thermophilic *Geobacillus.* In *Geobacillus* the expression of *alkB* was dependent on the temperature of culture and soil sample incubation. Positive RT-PCR signals were only obtained at 55°C, and not at mesophilic temperatures (38). Evidence of alkane hydroxylase expression at both temperature extremes demonstrates how ubiquitous they are in the environment.

Commercial application of oxygenases

Oxygenases are important in many environmental processes, and many have been applied to commercial processes. Often, the biggest setback applying an enzyme industrially is developing the processes for practical large-scale application in industry (11, 30). The alkane hydroxylase and xylene monooxygenase in *Pseudomonas putida*

transforms alcohols to aldehydes, and the latter can also produce carboxylic acid. This process was scaled-up for production (11, 66). Since the oxidation process is regio- and stereo-specific, oxygenases can provide specific products that are useful in the chemical industry. AlkB can produce an (*R)*-epoxide with high enantiomeric specificity from compounds with a terminal double bond. An optically active epoxide produced by AlkB could be used to produce several chemicals that act as precursors to valuable products. Many roadblocks can hamper large-scale production in this case including substrate toxicity and uptake, product toxicity, low compound turnover, product recovery, and uncoupling of the process (50). Greater understanding of these enzymes is crucial for their potential application.

Alkane hydroxylases have the potential for commercial applications if they are single-component, soluble, and stable. The ideal enzyme would not require a cofactor for functionality, and would be very precise in its action by being very stereo-, regio- and enantiomerically selective. There is a cytochrome P450 heat-stable enzyme that is close to meeting these requirements since it is a single-component, soluble enzyme (55). Additionally, some enzymes are going through a process of directed evolution in the laboratory in an attempt to change substrate specificity and increase their potential commercial value (50). Following several iterations of directed evolution, a fast-acting Class II P450 cloned from *B. megaterium* which originally favored fatty acid substrates, acted upon octane and other alkanes (19). Through directed evolution, favorable characteristics of one alkane hydroxylase system may be utilized for industrial applications through optimization of other aspects of the enzyme.

Alkane monooxygenases are also of interest due to their bioremediation capabilities in soil and aquatic environments. For example, in remote areas such as in the polar regions, bioremediation of petroleum spills is the only viable option for cleanup (1). Also, since oil exploration is increasingly moving towards remote areas, the potential for an increase in the need of bioremediation technology in that area is growing. Contaminated soils may also benefit from bioremediation efforts. Temperature is crucial in the efficiency of hydrocarbon degradation by biotic processes and is an important factor contributing to the success of *in situ* bioremediation projects (34). This strategy has been investigated in crude-oil contaminated soils in Brazil, Antarctic soils, as well as many other environments (3, 31, 47).

Culturing environmental samples

It is estimated that only 1% of microorganisms are culturable using standard laboratory methodology, although a greater portion has the potential for cultivation with improvements in protocols (4, 27, 87). Results from culture-dependent studies suggest that some bacterial divisions are ubiquitous in the environment, while others are limited (27, 56). Some bacteria that are difficult to culture, yet abundant in the environment, have been cultured by use of extinction dilution and low-nutrient media methods (80). The majority of microbes in the environment have not yet been cultured in the laboratory through conventional methods. Many efforts have focused on recreating the natural conditions present in the environment and providing lower nutrient and substrate concentrations to encourage the growth of “unculturable” organisms (12, 14, 17, 58, 87).

The advent of molecular tools has allowed an efficient way to study microorganisms from diverse and hard to culture environments.

Biogeography

Alkane monooxygenases can be found in many habitats across the globe and their distribution can be influenced by several factors. Environmental microbial diversity and biogeography patterns can be affected by limited dispersal, environmental heterogeneity, and also by the size of a site studied (84). Comparing microbial gene families found in environments ranging from pristine to polluted, as well as in different areas of the world will help to further understand microbial genetic diversity and microbial biogeography patterns.

Many studies seek to determine if all microbes are in all environments, or if different microbial populations are present in different geographical locations. Using TRFLP, actinomycete polyketide synthase genes from soil in Central Asia and New Jersey were compared to find out if similar communities were present in both locations. It was concluded that different populations exist in New Jersey in comparison to Central Asia (82). It is expected that the same trend should apply to other microbial groups and other genes as well. Such results suggest that a wealth of microbial and genetic diversity is yet to be discovered. Despite the small portion of culturable bacteria, DNA extracted from environmental soil samples is very diverse. Studies indicate as many as 4,000 bacterial genomes are present in just 180g of soil (81). In contrast to the results of the Central Asia and New Jersey comparison study by Wawrik et al., other studies of spatial

diversification show evidence that suggest a lower amount of microbial diversity across regions.

In Australia a study of microbial turnover at the local and regional levels suggests that despite a high level of local diversity, only a moderate level of microbial diversity exists across regions (21). The investigation of hydrocarbon degradation in Alpine soils revealed no correlation between the numbers of hydrocarbon degraders, and the level of contamination, but revealed known petroleum degraders that were adapted to cold environments (39). In the Timor Sea, novel AlkB sequences and high gene copy numbers suggest an augmented capacity for alkane degradation due to constant alkane exposure at the hydrocarbon seeps (78). Much still remains to be understood regarding the spatial relationship between microbial populations as well as the diversity of microorganisms in the environment.

Bioinformatics and next generation sequencing

Bioinformatics is an integral part of next generation sequencing. Therefore, it is also necessary for analysis of gene-targeted metagenomics data. The Ribosomal Database Project (RDP) Pyrosequencing Pipeline is a tool available for public use online. Sequences can be uploaded through the RDP website for processing using their programs. In order to process DNA sequences the raw reads must go through an initial processing step (Figure 6) where the sequences are sorted to retain only those of sufficient quality and trimmed. These sequences then are infernally aligned and clustered using the RDP pipeline. Following clustering, different diversity measurements can be obtained. For example, β diversity can be measured using either the Jaccard or Sorensen

Index output to create heatmaps and cluster dendrograms comparing populations. Sequencing technologies will likely become increasingly useful as bioinformatics tools continue to improve in their efficiency. Currently, the limiting factor in analyzing large sequencing datasets is computing capacity, however this issue is quickly being resolved as technology advances.

Experimental plan overview

Alkane hydroxylases encompass a diverse group of enzymes that serve an important function in the bioremediation of petroleum components in the environment. To better understand the biogeographic distribution, diversity, and role of this group of monooxygenases in the environment sediment samples from various locations in the United State and Central Asia were analyzed using molecular techniques. Gene-targeted metagenomics was utilized to focus on 16S rRNA and alkane monooxygenases, *alkB* and P450, to observe the gene populations present in sediment and enrichment cultures, and perhaps uncover novel monooxygenase families.

Sediment from rivers and streams in Kyrgyzstan and Tajikistan were compared to see if novel alkane monooxygenase families could be found in a largely unstudied area. Locations were chosen from varied environments with disparate amounts of anthropogenic influence. The monooxygenase communities at each of these sites, and in alkane enrichment cultures were expected to exhibit greater similarity within countries, or in sites with similar characteristics (ie. near urban centers).

The Passaic River is a historically polluted waterway that runs through NJ and out to the Atlantic Ocean. Sediment samples were collected down the center of the river over a 1000 meter area for enrichment with hexadecane. Analysis of the microbial communities in the sediment and following enrichment for 16S rRNA and alkane monooxygenases was performed to determine if: sampling location down the river influenced the microbial community, and also to see if the results from triplicate enrichments were reliable when the populations were analyzed by pyrosequencing. It was expected that the triplicate samples would be similar in composition. Samples collected at closer proximity to each other should have closely related microbial communities, while the community retrieved from sediment collected 1000 meters down river from another site could have a distinct community profile.

Alkane monooxygenase populations in aquatic sediments from the island of Puerto Rico were also investigated to determine the affect of salinity on monooxygenase diversity, if monooxygenases were different on the western and eastern shores of the Port of San Juan, and to see the effect of mangroves on the microbial population. Salinity changes were expected to have a strong effect on monooxygenase diversity as different salinities should exert a selective pressure on the community. Additionally, geographical differences were expected to result in disparate microbial communities between distant locations. Mangroves create a unique environment in sediments, and were therefore anticipated to have an affect on the microbial communities, however, salinity was hypothesized to be the largest factor influencing bacterial survival and growth in the enrichment cultures.

References

1. **Aislabie, J., D. J. Saul, and J. M. Foght**. 2006. Bioremediation of hydrocarbon-contaminated polar soils. Extremophiles: life under extreme conditions **10**:171-179.
2. **Alonso-Gutierrez, J., M. Teramoto, A. Yamazoe, S. Harayama, A. Figueras, and B. Novoa**. 2011. Alkane-degrading properties of Dietzia sp. H0B, a key player in the Prestige oil spill biodegradation (NW Spain). Journal of applied microbiology **111**:800-810.
3. **Alvarez, V. M., S. C. Santos, C. Casella Rda, R. L. Vital, G. V. Sebastin, and L. Seldin**. 2008. Bioremediation potential of a tropical soil contaminated with a mixture of crude oil and production water. Journal of microbiology and biotechnology **18**:1966-1974.
4. **Amann, R. I., W. Ludwig, and K. H. Schleifer**. 1995. Phylogenetic identification and in situ detection of individual microbial cells without cultivation. Microbiological reviews **59**:143-169.
5. **Ashraf, W., A. Mihdhir, and J. C. Murrell**. 1994. Bacterial oxidation of propane. FEMS microbiology letters **122**:1-6.
6. **Auclair, K., P. Moenne-Loccoz, and P. R. Ortiz de Montellano**. 2001. Roles of the proximal heme thiolate ligand in cytochrome P450(cam). Journal of the american chemical society **123**:4877-4885.
7. **Baptist, J. N., R. K. Gholson, and M. J. Coon**. 1963. Hydrocarbon oxidation by a bacterial enzyme system. I. Products of octane oxidation. Biochimica et biophysica acta **69**:40-47.
8. **Beilen, J. B. V., and E. G. Funhoff**. 2007. Alkane hydroxylases involved in microbial alkane degradation. Applied microbiol biotechnol. **74**:13-21.
9. **Bouchez-Naitali, M., H. Rakatozafy, R. Marchal, J. Y. Leveau, and J. P. Vandecasteele**. 1999. Diversity of bacterial strains degrading hexadecane in relation to the mode of substrate uptake. Journal of applied microbiology **86**:421-428.
10. **Boulton, C. A., and C. Ratledge** (1984). The physiology of hydrocarbon-utilizing microorganisms. Enzyme and Fermentation Biotechnology. A. Wieseman. New York, Halstead Press**.** 11-77.
11. **Buhler, B., B. Witholt, B. Hauer, and A. Schmid**. 2002. Characterization and application of xylene monooxygenase for multistep biocatalysis. Applied and environmental microbiology **68**:560-568.
12. **Bussmann, I., B. Philipp, and B. Schink**. 2001. Factors influencing the cultivability of lake water bacteria. Journal of microbiological methods **47**:41-50.
13. **Chakrabarty, A. M., G. Chou, and I. C. Gunsalus**. 1973. Genetic regulation of octane dissimilation plasmid in Pseudomonas. Proceedings of the national academy of sciences of the United States of America **70**:1137-1140.
14. **Connon, S. A., and S. J. Giovannoni**. 2002. High-throughput methods for culturing microorganisms in very-low-nutrient media yield diverse new marine isolates. Applied and environmental microbiology **68**:3878-3885.

15. **National Research Council.** (2007). Why Metagenomics? The new science of metagenomics. Washington, D.C., The national academies press.
16. **Doughty, D. M., L. A. Sayavedra-Soto, D. J. Arp, and P. J. Bottomley**. 2006. Product repression of alkane monooxygenase expression in Pseudomonas butanovora. Journal of bacteriology **188**:2586-2592.
17. **Ferrari, B. C., S. J. Binnerup, and M. Gillings**. 2005. Microcolony cultivation on a soil substrate membrane system selects for previously uncultured soil bacteria. Applied and environmental microbiology **71**:8714-8720.
18. **Funhoff, E. G., U. Bauer, I. García-rubio, B. Witholt, and J. B. V. Beilen**. 2006. CYP153A6, a soluble P450 oxygenase catalyzing terminal-alkane hydroxylation. J. Bacteriol **188**:5220-5227.
19. **Glieder, A., E. T. Farinas, and F. H. Arnold**. 2002. Laboratory evolution of a soluble, self-sufficient, highly active alkane hydroxylase. Nature biotechnology **20**:1135-1139.
20. **Goswami, P., and H. D. Singh**. 1991. Different modes of hydrocarbon uptake by two Pseudomonas species. Biotechnology and bioengineering **37**:1-11.
21. **Green, J. L., A. J. Holmes, M. Westoby, I. Oliver, D. Briscoe, M. Dangerfield, M. Gillings, and A. J. Beattie**. 2004. Spatial scaling of microbial eukaryote diversity. Nature **432**:747-750.
22. **Haferburg, D., R. Hommel, R. Claus, and H. P. Kleber**. 1986. Extracellular micorbial lipids as biosurfactants. Advances in biochemical engineering biotechnology **33**:53-93.
23. **Hara, A., K. Syutsubo, and S. Harayama**. 2003. Alcanivorax which prevails in oil-contaminated seawater exhibits broad substrate specificity for alkane degradation. Environmental microbiology **5**:746-753.
24. **Harayama, S., Y. Kasai, and A. Hara**. 2004. Microbial communities in oil-contaminated seawater. Current opinion in biotechnology **15**:205-214.
25. **Head, I. M., D. M. Jones, and W. F. Roling**. 2006. Marine microorganisms make a meal of oil. Nature reviews. Microbiology **4**:173-182.
26. **Hommel, R. K.** (1994). Formation and function of biosurfactants for degradation of water-insoluble substrates. Biochemistry of microbial biodegradation. C. Ratledge. Dordrecht, Kluwer academic publishers**.** 63-87.
27. **Hugenholtz, P., B. M. Goebel, and N. R. Pace**. 1998. Impact of culture-independent studies on the emerging phylogenetic view of bacterial diversity. J Bacteriol **180**:4765-4774.
28. **Isin, E. M., and F. P. Guengerich**. 2007. Complex reactions catalyzed by cytochrome P450 enzymes. Biochimica et biophysica acta **1770**:314-329.
29. **Iwai, S., B. Chai, W. J. Sul, J. R. Cole, S. A. Hashsham, and J. M. Tiedje**. 2010. Gene-targeted-metagenomics reveals extensive diversity of aromatic dioxygenase genes in the environment. The ISME journal **4**:279-285.
30. **Katopodis, A. G., K. Wimalasena, J. Lee, and S. W. May**. 1984. Mechanistic studies on non-heme iron monooxygenase catalysis: expodiation, aldehyde formation, and demethylation by the w-hydroxylation system of Pseudomonas oleovorans. Journal of the american chemical society **106**:7928-7935.

31. **Katsivela, E., E. R. Moore, D. Maroukli, C. Strompl, D. Pieper, and N. Kalogerakis**. 2005. Bacterial community dynamics during in-situ bioremediation of petroleum waste sludge in landfarming sites. Biodegradation **16**:169-180.
32. **Kok, M., R. Oldenhuis, M. P. van der Linden, P. Raatjes, J. Kingma, P. H. van Lelyveld, and B. Witholt**. 1989. The Pseudomonas oleovorans alkane hydroxylase gene. Sequence and expression. The Journal of biological chemistry **264**:5435-5441.
33. **Kubota, M., M. Nodate, M. Yasumoto-Hirose, T. Uchiyama, O. Kagami, Y. Shizuri, and N. Misawa**. 2005. Isolation and functional analysis of cytochrome P450 CYP153A genes from various environments. Bioscience, biotechnology, and biochemistry **69**:2421-2430.
34. **Leahy, J. G., and R. R. Colwell**. 1990. Microbial degradation of hydrocarbons in the environment. Microbiological reviews **54**:305-315.
35. **Lee, K.** 1999. Benzene-induced uncoupling of naphthalene dioxygenase activity and enzyme inactivation by production of hydrogen peroxide. Journal of bacteriology **181**:2719-2725.
36. **Lemos, L. N., R. R. Fulthorpe, E. W. Triplett, and L. F. Roesch**. 2011. Rethinking microbial diversity analysis in the high throughput sequencing era. Journal of microbiological methods **86**:42-51.
37. **Luz, A. P., R. C. Gamba, M. S. Lauretto, M. C. Bicego, S. Taniguchi, R. C. Montone, and V. H. Pellizari**. 2006. Potential for bioremediation of hydrocarbon polluted soils in the Maritime Antarctic. Antarctic Science **18**:335-343.
38. **Marchant, R., F. H. Sharkey, I. M. Banat, T. J. Rahman, and A. Perfumo**. 2006. The degradation of n-hexadecane in soil by thermophilic geobacilli. FEMS microbiology ecology **56**:44-54.
39. **Margesin, R., D. Labbe, F. Schinner, C. W. Greer, and L. G. Whyte**. 2003. Characterization of hydrocarbon-degrading microbial populations in contaminated and pristine Alpine soils. Applied and environmental microbiology **69**:3085-3092.
40. **Margesin, R., D. Labbe, F. Schinner, C. W. Greer, and L. G. Whyte**. 2003. Characterization of hydrocarbon-degrading microbial populations in contaminated and pristine Alpine soils. Applied and environmental microbiology **69**:3085-3092.
41. **Margesin, R., and F. Schinner**. 2001. Bioremediation (natural attenuation and biostimulation) of diesel-oil-contaminated soil in an alpine glacier skiing area. Applied and environmental microbiology **67**:3127-3133.
42. **Marin, M. M., L. Yuste, and F. Rojo**. 2003. Differential expression of the components of the two alkane hydroxylases from Pseudomonas aeruginosa. Journal of bacteriology **185**:3232-3237.
43. **Mishra, S., J. Jyot, R. C. Kuhad, and B. Lal**. 2001. Evaluation of inoculum addition to stimulate in situ bioremediation of oily-sludge-contaminated soil. Applied and environmental microbiology **67**:1675-1681.
44. **Nelson, D. R.** (2009). "The Cytochrome P450 Homepage." Human Genomics **4**:59-65.
45. **Ortiz de Montellano, P. R., and J. J. De Voss** (2005). Substrate oxidation by cytochrome P450 enzymes. Cytochtome P450: structure, mechanism, and biochemistry. P. R. Ortiz de Montellano. New York, Plenum publishers**.** 183-245.

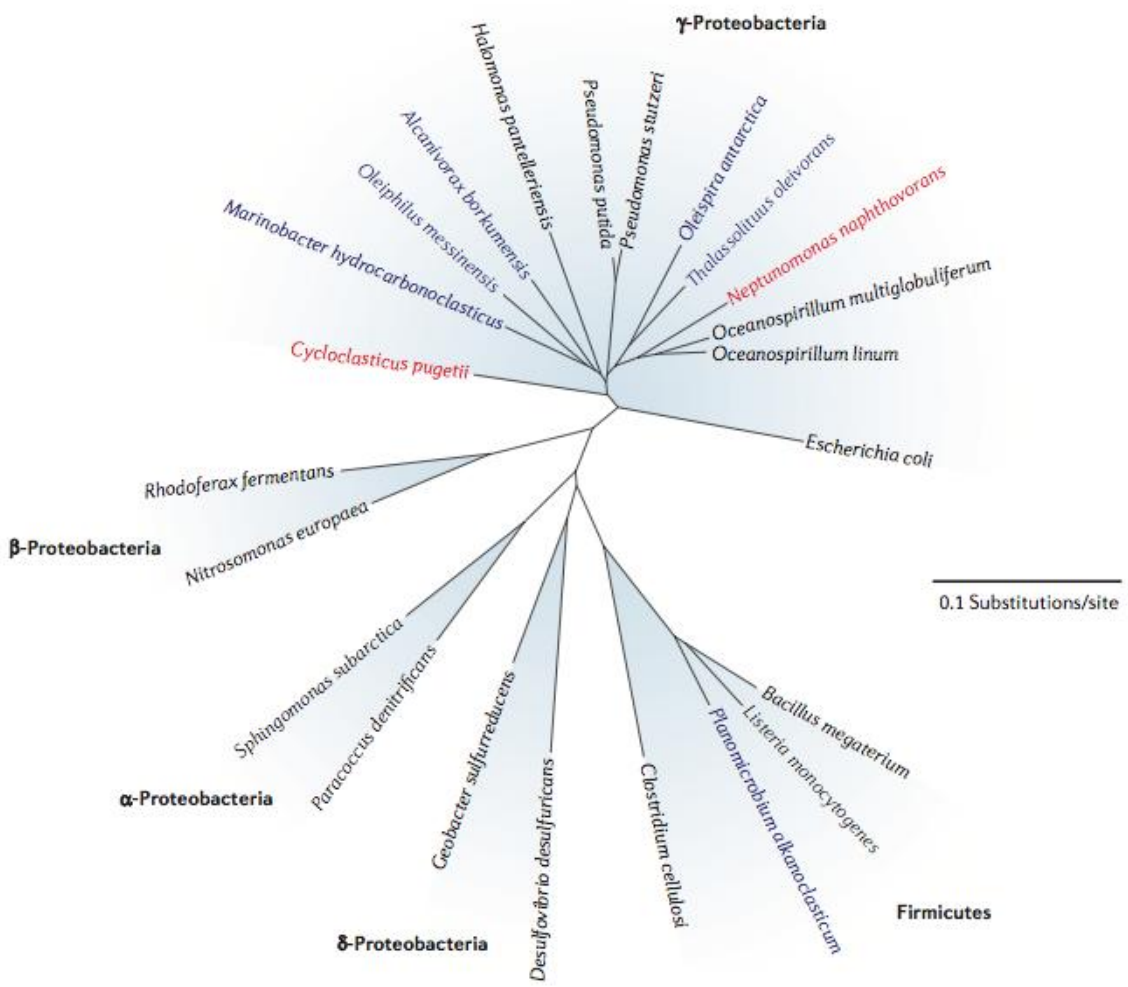
46. **Paine, M. J. I., N. S. Scrutton, A. W. Munro, A. Gutierrez, G. C. K. Roberts, and C. R. Wolf** (2005). Electron transfer partners of cytochrome P450. Cytochrome P450: structure, mechanism, and biochemistry. P. R. Ortiz de Montellano. New York, Plenum publishers. 115-148.
47. **Powell, S. M., S. H. Ferguson, J. P. Bowman, and I. Snape**. 2006. Using real-time PCR to assess changes in the hydrocarbon-degrading microbial community in Antarctic soil during bioremediation. Microbial ecology **52**:523-532.
48. **Ratajczak, A., W. Geissdorfer, and W. Hillen**. 1998. Expression of alkane hydroxylase from Acinetobacter sp. Strain ADP1 is induced by a broad range of n-alkanes and requires the transcriptional activator AlkR. Journal of bacteriology **180**:5822-5827.
49. **Roesch, L. F., R. R. Fulthorpe, A. Riva, G. Casella, A. K. Hadwin, A. D. Kent, S. H. Daroub, F. A. Camargo, W. G. Farmerie, and E. W. Triplett**. 2007. Pyrosequencing enumerates and contrasts soil microbial diversity. The ISME journal **1**:283-290.
50. **Rojo, F.** 2005. Specificity at the end of the tunnel: understanding substrate length discrimination by the AlkB alkane hydroxylase. J Bacteriol **187**:19-22.
51. **Rojo, F.** 2009. Degradation of alkanes by bacteria. Environmental microbiology **11**:2477-2490.
52. **Röling, W. F. M., M. G. Milner, D. M. Jones, K. Lee, F. Daniel, R. J. P. Swannell, and I. Head**. 2002. Robust hydrocarbon degradation and dynamics of bacterial communities during nutrient-enhanced oil spill bioremediation. Applied and environmental microbiology **68**:5537-5548.
53. **Ron, E. Z., and E. Rosenberg**. 2002. Biosurfactants and oil bioremediation. Current opinion in biotechnology **13**:249-252.
54. **Sabirova, J. S., M. Ferrer, D. Regenhardt, K. N. Timmis, and P. N. Golyshin**. 2006. Proteomic insights into metabolic adaptations in Alcanivorax borkumensis induced by alkane utilization. Journal of bacteriology **188**:3763-3773.
55. **Salazar, O., P. C. Cirino, and F. H. Arnold**. 2003. Thermostabilization of a cytochrome p450 peroxygenase. Chembiochem: a European journal of chemical biology **4**:891-893.
56. **Schlegel, H. G., and H. W. Jannasch** (1992). Prokaryotes and their habitats. The prokaryotes. A. Balows, H. G. Truper, M. Dworkin, W. Harder and S. K.-H. New York, NY, Springer-Verlag. **1:**75-125.
57. **Schneiker, S., V. A. Martins dos Santos, D. Bartels, T. Bekel, M. Brecht, J. Buhrmester, T. N. Chernikova, R. Denaro, M. Ferrer, C. Gertler, A. Goesmann, O. V. Golyshina, F. Kaminski, A. N. Khachane, S. Lang, B. Linke, A. C. McHardy, F. Meyer, T. Nechitaylo, A. Puhler, D. Regenhardt, O. Rupp, J. S. Sabirova, W. Selbitschka, M. M. Yakimov, K. N. Timmis, F. J. Vorholter, S. Weidner, O. Kaiser, and P. N. Golyshin**. 2006. Genome sequence of the ubiquitous hydrocarbon-degrading marine bacterium Alcanivorax borkumensis. Nature biotechnology **24**:997-1004.
58. **Schut, F., E. J. de Vries, J. C. Gottschal, B. R. Robertson, W. Harder, R. A. Prins, and D. K. Button**. 1993. Isolation of Typical Marine Bacteria by Dilution Culture: Growth, Maintenance, and Characteristics of Isolates under Laboratory Conditions. Applied and environmental microbiology **59**:2150-2160.

59. **Shanklin, J., and E. Whittle**. 2003. Evidence linking the Pseudomonas oleovorans alkane omega-hydroxylase, an integral membrane diiron enzyme, and the fatty acid desaturase family. FEBS Lett **545**:188-192.
60. **Shanklin, J., E. Whittle, and B. G. Fox**. 1994. Eight histidine residues are catalytically essential in a membrane-associated iron enzyme, stearoyl-CoA desaturase, and are conserved in alkane hydroxylase and xylene monooxygenase. Biochemistry **33**:12787-12794.
61. **Singer, M. E., and W. R. Finnerty** (1984). Microbial metabolism of straight- chain and branched alkanes. Microbial metabolism of straight-chain and branched alkanes. R. M. Atlas. New York, Macmillan Publishing Company**.** 1-60.
62. **Smits, T. H., B. Witholt, and J. B. van Beilen**. 2003. Functional characterization of genes involved in alkane oxidation by Pseudomonas aeruginosa. Antonie van Leeuwenhoek **84**:193-200.
63. **Smits, T. H. M., B. Witholt, and J. B. V. Beilen**. 2002. Functional analysis of alkane hydroxylases from gram-negative and gram-positive bacteria. Applied and Environmental microbiology. **184**:1733-1742.
64. **Staijen, I. E., R. Marcionelli, and B. Witholt**. 1999. The PalkBFGHJKL promoter is under carbon catabolite repression control in Pseudomonas oleovorans but not in Escherichia coli alk+ recombinants. Journal of bacteriology **181**:1610-1616.
65. **Swannell, R. P., K. Lee, and M. McDonagh**. 1996. Field evaluations of marine oil spill bioremediation. Microbiological reviews **60**:342-365.
66. **van Beilen, J. B., W. A. Duetz, A. Schmid, and B. Witholt**. 2003. Practical issues in the application of oxygenases. Trends in biotechnology **21**:170-177.
67. **van Beilen, J. B., G. Eggink, H. Enequist, R. Bos, and B. Witholt**. 1992. DNA sequence determination and functional characterization of the OCT-plasmid-encoded alkJKL genes of Pseudomonas oleovorans. Molecular microbiology **6**:3121-3136.
68. **van Beilen, J. B., and E. G. Funhoff**. 2005. Expanding the alkane oxygenase toolbox: new enzymes and applications. Current opinion in biotechnology **16**:308-314.
69. **van Beilen, J. B., E. G. Funhoff, A. V. Loon, A. Just, L. Kaysser, M. Bouza, M. Ro, Z. Li, and B. Witholt**. 2006. Cytochrome P450 alkane hydroxylases of the CYP153 family are common in alkane-degrading eubacteria lacking integral membrane alkane hydroxylases. Applied and environmental microbiology. **72**:59-65.
70. **van Beilen, J. B., E. G. Funhoff, A. van Loon, A. Just, L. Kaysser, M. Bouza, R. Holtackers, M. Rothlisberger, Z. Li, and B. Witholt**. 2006. Cytochrome P450 alkane hydroxylases of the CYP153 family are common in alkane-degrading eubacteria lacking integral membrane alkane hydroxylases. Applied and environmental microbiology **72**:59-65.
71. **van Beilen, J. B., Z. Li, W. A. Duetz, T. H. M. Smits, and B. Witholt**. 2003. Diversity of alkane hydroxylase systems in the environment. Oil and gas science and technology **58** 427-440.
72. **van Beilen, J. B., M. M. Marin, T. H. Smits, M. Rothlisberger, A. G. Franchini, B. Witholt, and F. Rojo**. 2004. Characterization of two alkane hydroxylase genes

from the marine hydrocarbonoclastic bacterium Alcanivorax borkumensis. Environmental microbiology **6**:264-273.

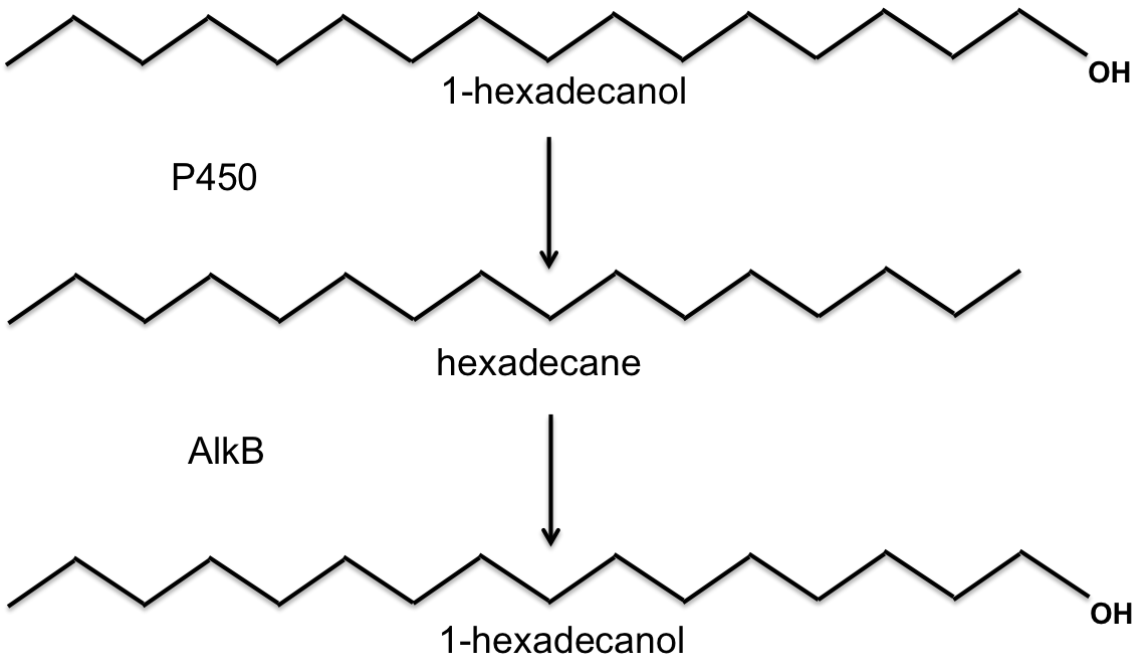
73. **van Beilen, J. B., S. Panke, S. Lucchini, A. G. Franchini, M. Rothlisberger, and B. Witholt**. 2001. Analysis of Pseudomonas putida alkane-degradation gene clusters and flanking insertion sequences: evolution and regulation of the alk genes. Microbiology **147**:1621-1630.
74. **van Beilen, J. B., D. Penninga, and B. Witholt**. 1992. Topology of the membrane-bound alkane hydroxylase of Pseudomonas oleovorans. The Journal of biological chemistry **267**:9194-9201.
75. **van Beilen, J. B., T. H. Smits, F. F. Roos, T. Brunner, S. B. Balada, M. Rothlisberger, and B. Witholt**. 2005. Identification of an amino acid position that determines the substrate range of integral membrane alkane hydroxylases. Journal of bacteriology **187**:85-91.
76. **van Beilen, J. B., M. G. Wubbolts, and B. Witholt**. 1994. Genetics of alkane oxidation by Pseudomonas oleovorans. Biodegradation **5**:161-174.
77. **van Hamme, J. D., A. Singh, and O. P. Ward**. 2003. Recent advances in petroleum microbiology. Microbiology and molecular biology reviews **67**:503-549.
78. **Wasmund, K., K. A. Burns, D. I. Kurtboke, and D. G. Bourne**. 2009. Novel alkane hydroxylase gene (alkB) diversity in sediments associated with hydrocarbon seeps in the Timor Sea, Australia. Applied and environmental microbiology **75**:7391-7398.
79. **Watkinson, R. J., and P. Morgan**. 1990. Physiology of aliphatic hydrocarbon-degrading microorganisms. Biodegradation **1**:79-92.
80. **Wawrik, B., L. Kerkhof, J. Kukor, and G. Zylstra**. 2005. Effect of different carbon sources on community composition of bacterial enrichments from soil. Applied and environmental microbiology **71**:6776-6783.
81. **Wawrik, B., L. Kerkhof, J. Kukor, and G. Zylstra**. 2005. Effect of different carbon sources on community composition of bacterial enrichments from soil. Applied and environmental microbiology **71**:6776-6783.
82. **Wawrik, B., D. Kutliev, U. A. Abdivasievna, J. J. Kukor, G. J. Zylstra, and L. Kerkhof**. 2007. Biogeography of actinomycete communities and type II polyketide synthase genes in soils collected in New Jersey and Central Asia. Applied environmental microbiology **73**:2982-2989.
83. **Wentzel, A., T. E. Ellingsen, H. K. Kotlar, S. B. Zotchev, and M. Throne-Holst**. 2007. Bacterial metabolism of long-chain n-alkanes. Applied microbiology and biotechnology **76**:1209-1221.
84. **Whitfield, J.** 2005. Biogeography. Is everything everywhere? Science **310**:960-961.
85. **Whyte, L. G., T. H. Smits, D. Labbe, B. Witholt, C. W. Greer, and J. B. van Beilen**. 2002. Gene cloning and characterization of multiple alkane hydroxylase systems in Rhodococcus strains Q15 and NRRL B-16531. Applied and environmental microbiology **68**:5933-5942.
86. **Yoshioka, S., S. Takahashi, H. Hori, K. Ishimori, and I. Morishima**. 2001. Proximal cysteine residue is essential for the enzymatic activities of cytochrome P450cam. European journal of biochemistry **268**:252-259.

87. **Zengler, K., G. Toledo, M. Rappe, J. Elkins, E. J. Mathur, J. M. Short, and M. Keller**. 2002. Cultivating the uncultured. Proceedings of the national academy of sciences of the United States of America **99**:15681-15686.
88. **Zhou, R., C. Huang, A. Zhang, S. G. Bell, W. Zhou, and L. L. Wong**. 2011. Crystallization and preliminary X-ray analysis of CYP153C1 from Novosphingobium aromaticivorans DSM12444. Acta crystallographica. Section F, Structural biology and crystallization communications **67**:964-967.



From Head *et al*., 2006.

Figure 1: Aerobic hydrocarbon degrading bacteria: The bacteria in blue are known to degrade saturated hydrocarbons, while those in red degrade polycyclic aromatic hydrocarbons. Those shown in black do not degrade any type of hydrocarbon.



Redrawn from van Beilen *et al*., 2005.

Figure 2. Alkane degradation by monooxygenases: The P450 and AlkB reactions catalyzed by alkane hydroxylases with hexadecane as the carbon source

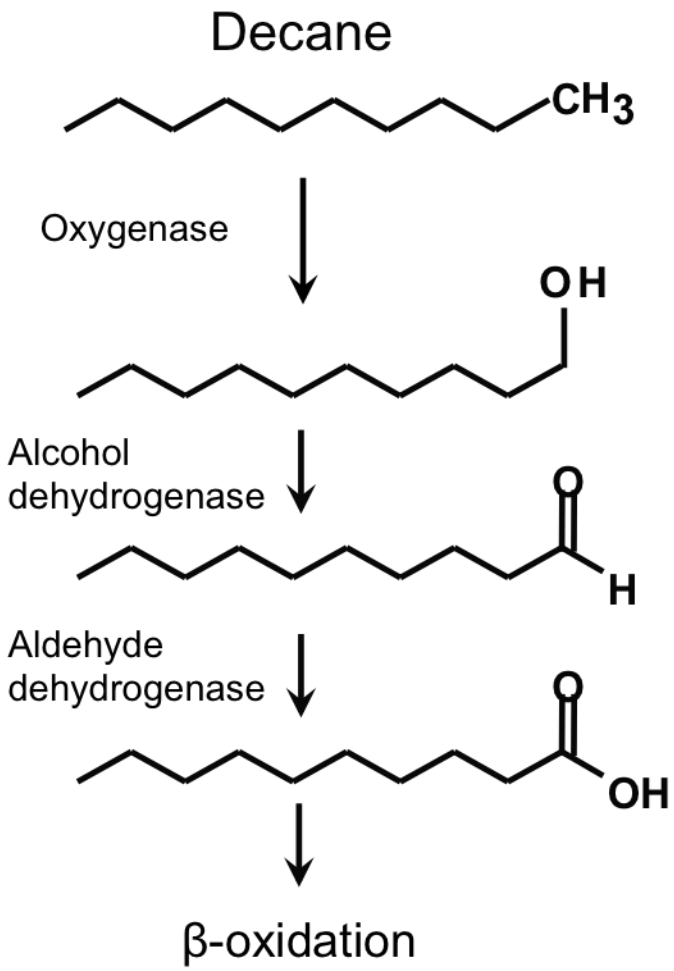
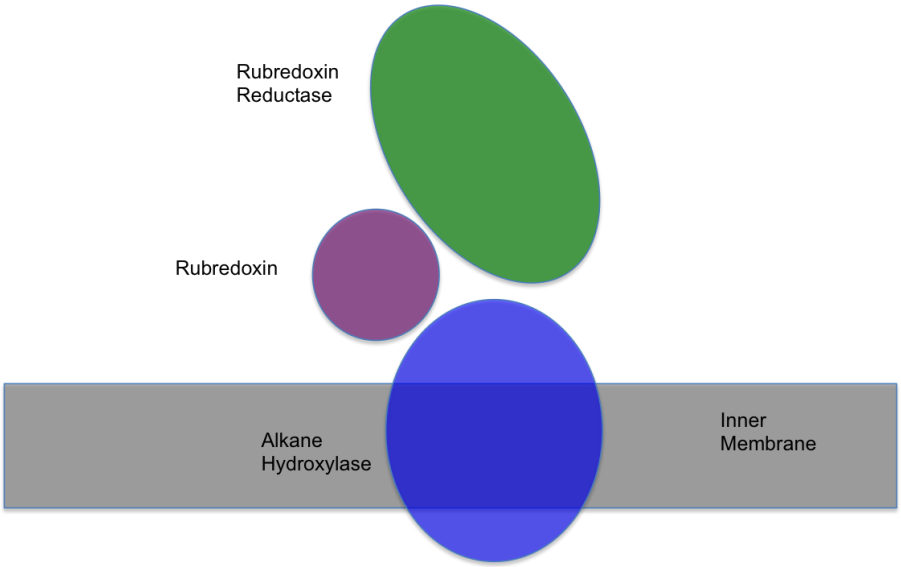


Figure 3. Alkane degradation pathway: Initial steps in the degradation of alkanes involves the action of an oxygenase and two dehydrogenases that result in formation of a fatty acid. Further reactions in the β-oxidation pathway provide carbon and energy for growth of the microbe.

Enzyme class	Composition and cofactors	Substrate range	Presence shown in
Eukaryotic P450 (CYP52, class II))	Microsomal oxygenase: P450 heme reductase: FAD, FMN, NADPH	C10-C16 *n*-alkanes C4-C16 *n*-alkanes	*Candida maltosa, Candida tropicalis, Yarrowia lipolytica*
Bacterial P450 oxygenase systems (CYP153, class I)	P450 oxygenase: P450 heme ferredoxin: iron-sulfur ferredoxin reductase: FAD, NADH	Probably C4-C10 *n*-alkanes	*R. rhodochrous* 7E1C *Acinetobacter* sp. EB104
***alkB*-related alkane hydroxylases**	Membrane hydroxylase: dinuclear iron rubredoxin: iron rubredoxin reductase: FAD, NADH	C10-C30 *n*-alkanes C4-C10 *n*-alkanes	*Acinetobacter , Alcanivorax, Burkholderia, Mycobacterium, Pseudomonas, Rhodococcus, etc.*

Adapted from van Beilen *et al.* 2003.

Table 1: AlkB and P450 enzyme classes involved in alkane oxidation. Class I P450 and AlkB monooxygenase systems are involved in microbial alkane degradation.



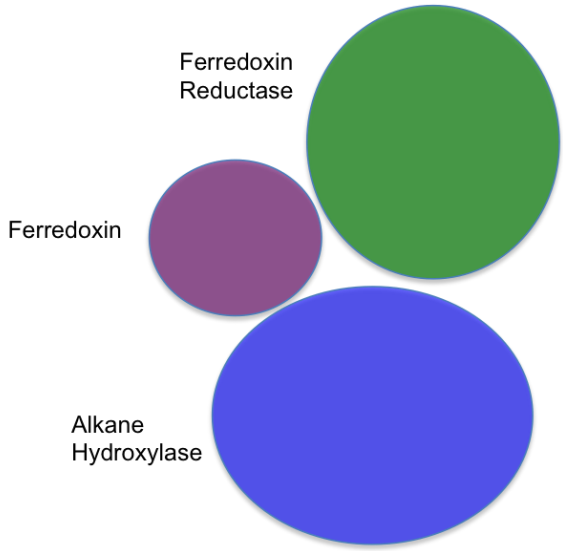
Redrawn from van Beilen *et al.*, 2003.

Figure 4: AlkB monooxygenase system found in *P. putida*. The rubredoxin reductase (AlkT) transfers electrons to the rubredoxin (AlkG) from NADH using the cofactor FAD. Rubredoxin then transfers the elctrons to the alkane hydroxylase.

Gene	Function of the encoded protein
alkB	Hydroxylase of monooxygenase
alkFG	Rubredoxin of monooxygenase
alkH	Aldehyde dehydrogenase
alkJ	Alcohol dehydrogenase
alkK	Acyl-CoA synthetase
alkL	Outer membrane protein
alkM	Methyl-accepting chemotaxis protein
alkS	Regulator
alkT	Rubredoxin reductase of monooxygenase

From van Beilen *et al. 1992.*

Table 2: Genes involved in alkane degradation in *P. putida* GP01



Redrawn from van Beilen and Funhoff, 2005.

Figure 5: Alkane hydroxylase system P450 Class I. This enzyme system consists of a P450 alkane hydroxylase, ferredoxin, and ferredoxin reductase. The ferredoxin and ferredoxin reductase transfer electrons to the alkane hydroxylase which incorporates an oxygen at the terminal carbon of *n*-alkanes.

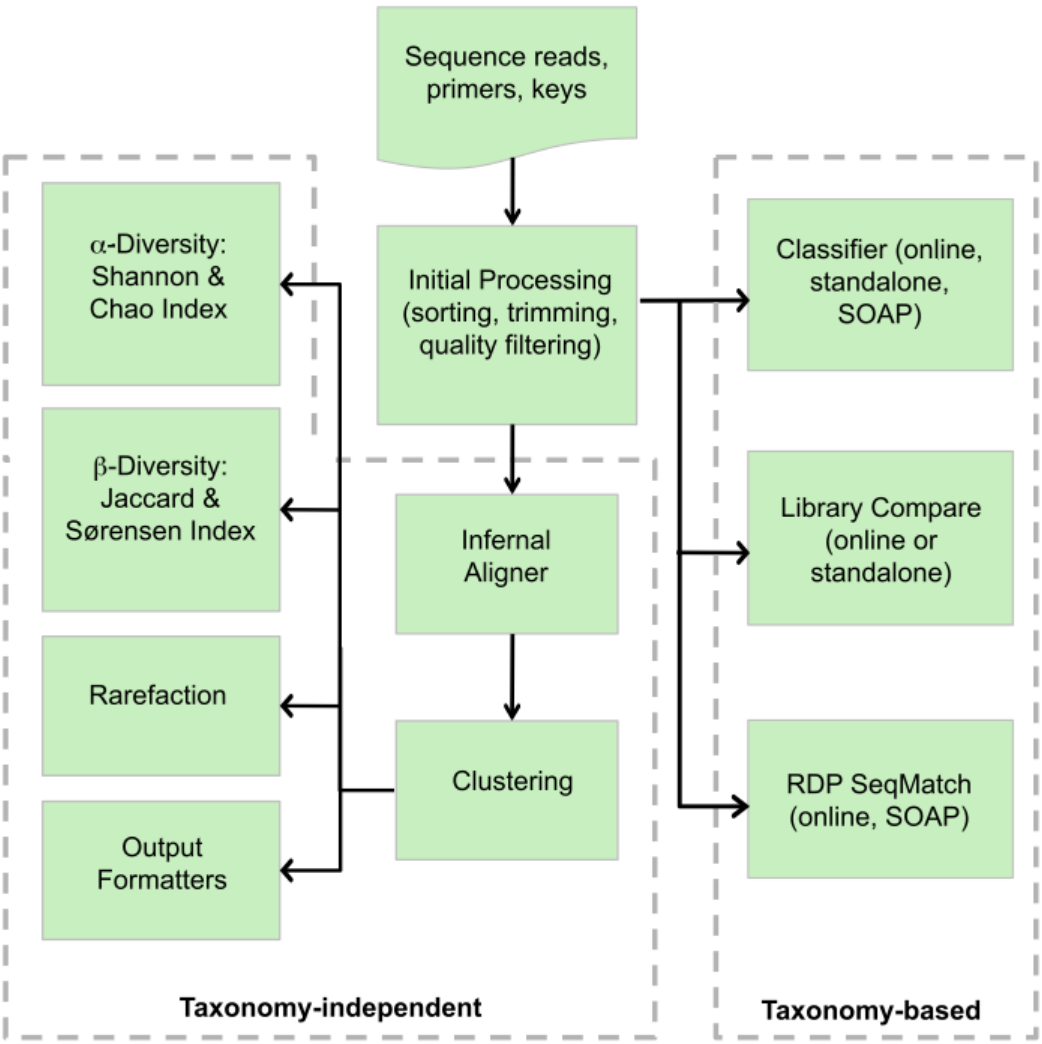


Figure from Benli Chai, RDP, Michigan State University.

Figure 6: Ribosomal Database Project pyrosequencing pipeline. This diagram shows the path of sequencing data during analysis.

CHAPTER TWO

Spatial effects of alkane monooxygenase gene populations in Passaic River sediment following selective enrichment

---

Elyse, A. Rodgers-Vieira [1], Qiong Wang[3], Jim R. Cole[3], Jerome J. Kukor[2],

Jim M. Tiedje[3], and Gerben J. Zylstra[1]

Department of Biochemistry and Microbiology[1], Rutgers University, School of Environmental and Biological Sciences[2], Rutgers, The State University of New Jersey, New Brunswick, New Jersey 08901, USA. Center for Microbial Ecology and Department of Microbiology and Molecular Genetics, Michigan State University, East Lansing, Michigan 48824, USA[3].

Abstract

The Passaic River in New Jersey has been heavily impacted by human activities and polluted by industrial processes. A prominent industrial presence along the shores of the Passaic in Newark dates back to the late 1800's, and many petroleum refineries and chemical manufacturing plants still operate along the river. The aim of this study was to examine the bacterial population diversity and degradative gene distribution at different points along the river and determine how this may affect the microbial response to a hydrocarbon challenge. Triplicate sediment samples were collected in the middle of the Passaic River from three sites in a heavily industrialized area (0 meters, 10 meters, and 1 km in relationship to the start point). Each of the nine sediment samples were enriched in minimal media with hexadecane as the sole carbon source. Total DNA was extracted at the beginning of the enrichment and after three to four days. Cell growth was monitored by DAPI staining. PCR was performed on the extracted DNA using primers for the 16S rRNA gene and primers for two different families of alkane monooxygenase encoding genes (P450 and AlkB type). In order to obtain a complete picture of the genetic diversity of each sample, the 54 PCR products were subjected to pyrosequencing. Parallels between each of the sample populations were observed, however, there were also gene sequences that were unique to each of the sample locations. As expected, a shift in the microbial community and in the population of the two different alkane monooxygenase genes was observed in the hexadecane enrichment cultures.

Introduction

An estimated 99% of microorganisms in the environment are not readily cultured in the laboratory given current methods and technologies (23, 24). As a result, culture independent strategies may be applied to recover functional gene populations within an environmental sample. Gene-targeted metagenomics focuses sequencing efforts on genes of particular interest (12). This approach was employed in order to probe and compare the community diversity of alkane monooxygenases in the Passaic River sediment, and to determine if the proximity of sampling locations along the Passaic River affects the microbial community composition recovered when the locations are compared.

Industrial impacts to the lower six miles of the tidal portion of the Passaic River date back to the nineteenth century. Industrialization left behind a host of organic and inorganic pollutants in the river sediment that vary in concentration, including many unidentified chemical compounds (1, 2, 10, 15). Direct and indirect discharges from manufactured gas plants, chemical manufacturing, paper manufacturing, recycling, and other facilities have been released into the Passaic from the mid 1800's to the present (4, 8, 11, 18). As a result, chemicals and pollutants including metals, dioxins, polychlorinated biphenyls, pesticides, volatile and semi-volatile compounds, and petroleum hydrocarbons from these facilities are just some of the pollutants present in the Passaic (2, 10, 18). Leaking oil tanks, and a burst oil pipeline in the late 1800's covered the Passaic in a sheet of oil (13). Lightweight fuel oil was routinely spread in the surrounding marshes to control insects in the early 1900's (9). Sediments in the Lower Passaic River region pose high risks to human health and that of the river ecosystem. In the lower eight miles of the river, where samples for this study were collected, up to 25

feet of sediment has accumulated since the last major dredging efforts of the 1940s and 1980s. This contaminated sediment has been moved and distributed by the current and tidal mixing throughout the Lower Passaic (18).

Alkanes are saturated hydrocarbons. Depending on the source, up to 50% of crude oil is composed of alkanes. Plants, algae, bacteria, and animals also naturally produce alkanes in the environment. Alkane monooxygenases activate the terminal carbon in straight chain alkanes by the regio-specific addition of an oxygen atom to a compound. This marks the first step in the degradation of alkanes, resulting in the formation of an alcohol product (30). Two types of alkane monooxygenases involved in the oxidation process are cytochrome P450 and AlkB. Often multiple, divergent copies of the alkane monooxygenase genes are present in bacteria such as *Alkanivorax*, *Rhodococcus*, *Parvibaculum*, and *Pseudomonas*, that each confer an increased substrate range (3, 17, 27, 29). The process of alkane degradation by microbes in the environment is especially of interest in historically polluted areas such as the Passaic.

Using gene-targeted metagenomics, alkane monooxygenases were recovered directly from Passaic River sediment and following enrichment with the alkane hexadecane to determine and compare: 1) the diversity of the alkane monooxygenase populations; and 2) if alkane monooxygenase populations differ at certain distances along the river. Novel P450 and AlkB families were observed in Passaic River sediment and enrichments when compared to known protein sequences in GenBank. Additionally, populations of alkane monooxygenases in unenriched sediment were distinct from the community recovered following enrichment on hexadecane. Differences were also seen between sediment sampling locations on the river.

Materials and Methods

Sediment collection. Sediment samples were collected at three points in the Passaic River in the Newark, New Jersey area. A grab sampler was lowered in the center of the river to collect the sediment at an arbitrary start point (0 meters) (N 40.75990, W 74.16311), and at 10 meters and 1000 meters (N 40.75890, W 74.16491) down river in relationship to the start point. The sediment collected was put into 3 plastic containers to serve as triplicate replicates. These sediment samples were stored at 4°C.

Growth conditions and cell enumeration. 50 ml of minimal salts base (MSB) media as described by Stanier *et al.*, was inoculated with .5 g sediment and amended with 50 µl of hexadecane as the sole carbon source (Sigma Aldrich, St. Louis, MO) (21). The cultures were incubated shaking at 30°C for three to four days until cell density reached approximately 2 x 10^{10} cells/ml. Growth was measured by DAPI counts. For the DAPI staining, 1 ml of the enrichment culture was fixed with 100 µl of 37% formaldehyde overnight. The sample was vortexed and diluted 1:200 with sterile MSB medium. It was then sonicated in an ice water bath for 5 minutes before staining. They were stained with 5 µg/ml of 4', 6'- diamidino-2-phenylindole (DAPI) in the dark for 1 hour (Sigma Aldrich, St. Louis, MO). The sample was then filtered using a 0.2 µm black 22-mm diameter filter from Millipore. A Zeiss Axiovert 200 M epifluorescent microscope was used. Cell counts from 10 grid fields on each slide were tallied (19, 20).

DNA extraction and amplification for pyrosequencing. DNA was extracted using the PowerSoil kit before and after enrichment and quantified using a NanoDrop Spectrophotometer (MoBio, Carlsbad, CA and Thermo Scientific, Wilmington, DE). DNA was amplified using the FastStart High Fidelity PCR system (Roche, Indianapolis,

IN). The composition of the PCR mixtures were 1 µM each of the forward and reverse primers, 0.2 mM dNTPs, and 150 ng/µl BSA in a total volume of 20 µl which supplied 1X buffer with 1 unit of enzyme per reaction. Approximately 25 ng of template DNA was added to each reaction. For *alkB* amplification, barcoded degenerate primers alkBF (AATACHGSVCAYGAGCTCRGYCAYAAR) and alkBR (GCRTGRTGATCAGARTGHCGYTG) were used (16). For *alkB* amplification the following protocol was used: 95°C for 5 minutes, followed by 35 cycles of 95°C for 45 seconds, 55°C for 45 seconds, and 72°C for 1 minute, with a final elongation of 72°C for 4 minutes. The tagged primers used to amplify P450 monooxygenase genes were P450F (GTSGGCGGCAACGACACSAC) and P450R (GCASCGGTGGATGCCGAAGCCRAA) (26). The PCR protocol used for P450 amplification was: 95°C for 5 minutes, followed by 35 cycles of 95°C for 30 seconds, 58°C for 30 seconds, and 72°C for 40 seconds, with a final elongation at 72°C for 4 minutes. The 16S rRNA primers used were 16S_577F (AYTGGGYDTAAAGNG) and 16S_926R (CCGTCAATTCMTTTRAGT) (14, 22). For 16S rRNA amplification the PCR protocol was: 95°C for 2 minutes, followed by 30 cycles of 95°C for 45 seconds, 55°C for 45 seconds, and 72°C for 1 minute, with a final elongation at 72°C for 4 minutes.

454 pyrosequencing. The amplified products were separated on a 1% agarose gel. All PCR reactions were done in triplicate or higher and pooled during gel purification using the Qiagen Gel Extraction Kit and quantified using a Nanodrop Spectrophotometer (Qiagen, Valencia, CA and Thermo Scientific, Wilmington, DE). The samples were run

at the Michigan State University core facility using a Roche 454 FLX Titanium sequencer.

Data analysis. The raw sequencing reads were sorted, trimmed, and then put through a quality filter in the Ribosomal Database Project Pipeline (RDP) (6, 7). This quality filter checks for the presence of the primer sequences and overall sequence quality.

16S rRNA reads that passed the quality filter were used for further analysis in the RDP Pyrosequencing Pipeline. The 16S rRNA sequences were clustered at 80% sequence similarity for analysis and clusters were identified using the RDP Classifier (28). Functional gene reads deemed to be of sufficient quality were translated and frame shift corrected with RDP's FrameBot tool using the Hidden Markov Model (HMM) built specifically for each functional gene using known sequences from GenBank (5). The HMM model uses the known sequences to determine the most conserved regions and which variations within the sequence are the most likely to occur. Further analysis of the *alkB* and P450 sequences was done using the RDP Fungene Pipeline (6, 7).

Results

The purpose of this enrichment and sampling scheme was to determine the effect of sample location in the river and selective enrichment on alkane monooxygenase gene diversity. Sampling locations were selected down the length of the Passaic River in Newark, New Jersey, within a relatively small area (1000 meters). Samples were collected at an arbitrary 0 meter point, 10 meters, and 1000 meters to determine if sampling in one location is sufficient to gain a good picture of the microbial community,

or if multiple locations are necessary within a short distance to determine genetic diversity of the alkane monooxygenase populations. Classification of the 16S rRNA sequences recovered from pyrosequencing showed a dramatic shift in the microbial population following enrichment (Figure 1). There was an increase in the Firmicutes and Bacteriodetes, with a decrease in Acidobacteria and Chloroflexi. There was a significant portion of unclassified bacteria before enrichment, but this sector decreased with hexadecane enrichment. Before enrichment there was a predominance of proteobacteria that only increased in proportion to the total population following enrichment. Within the proteobacteria, 16S rRNA sequences for Betaproteobacteria stayed relatively constant during enrichment while those for Gammaproteobacteria, went up and those for the Deltaproteobacteria and Alphaproteobacteria went down (Figure 2).

Analysis of the alkane monooxygenase functional gene populations was performed at the deduced protein level. Principal Coordinate Analysis was performed for both AlkB and P450 alkane monooxyenases (Figures 3A and 3B). A clear divide was seen between samples before and after enrichment as the sites fell into two separate clusters, the original sediment and the enrichment culture, on the PCA plot. The cluster dendrograms also showed a similar trend with very separate groups of clades falling within the dendrograms (Figures 4A and 4B). Heatmaps showed an increase in similarity between hydrocarbon enriched samples, and conversely when unenriched sediment was compared to other unenriched samples. These trends were more pronounced for the AlkB data than for P450.

To determine to what extent known versus novel AlkB and P450 alkane hydroxylase families were being recovered using the gene-targeted metagenomics

approach, neighbor joining phylogenetic trees were constructed (Figures 5A and 5B). The phylogenetic trees visually convey the proportions of known and novel alkane hydroxylases families revealed by pyrosequencing. All known protein sequences for both P450 and AlkB that are part of GenBank were placed in two separate libraries to act as reference sequences. The sequence data obtained from pyrosequencing was then added to the reference sequences. All of the protein sequences were clustered at 80% identity before being loaded into ARB for analysis. Clades within the resulting trees were collapsed into a manageable number of gene families. The size of the shaded area of the clade is in proportion to the number of sequences present. Those clades marked with circles, and highlighted (Figures 5A and 5B) do not contain reference sequences from GenBank, and are therefore completely comprised of experimental sequences obtained from enriched and unenriched sediment. These clades represent novel AlkB or P450 sequences. For AlkB, 23% of 34 clades were void of previously known protein sequences. In the P450 neighbor-joining tree, 53% of 36 clades did not contain known protein sequences from GenBank. In both cases these novel clades are distributed throughout the phylogenetic trees.

Discussion

Alkane monooxygenases play a key role in the activation and subsequent degradation of alkanes in the environment. The two types of alkane hydroxylases, AlkB and P450, perform a similar enzymatic reaction, converting linear alkanes into alcohols. However, they represent different enzyme families. To capture an entire snapshot of the

alkane monooxygenases present in Passaic River sediment from a highly polluated section of the river, primers designed for each of the two gene families were utilized to amplify a portion of the gene. From pyrosequencing data for the alkane monooxygenases as well as 16S rRNA, a greater understanding of the diversity of monooxygenases in sediment and following enrichment was gained. Additionally, this data also probes the question of whether or not multiple environmental sampling sites are needed to investigate gene diversity in sediment, or if it doesn't matter where you sample within an environment.

From the 16S rRNA pyrosequencing data small differences are seen between the different sites. Classification at the phyla level reveals a large portion of unclassified bacteria before enrichment according to RDP. This unclassified portion decreases in abundance at all sample sites following enrichment. Since more is known about microbes that are readily cultured in the laboratory, it is expected that more unclassified bacteria would be recovered from the sediment before enrichment. The increase in Proteobacteria is expected since many microbes that belong to that group, especially Gamma and Betaproteobacteria, are known to be involved in alkane oxidation (25). Overall, the triplicate samples are similar at all sites except for the site 1000 meters down the river. This difference may be attributed to microniches within the sediment at that site. One of the goals during the sediment handling was to mix the sample to achieve as much homogeneity as possible. However, due to the presence of small rocks and organic debris the situation was not ideal.

A significant proportion of the clades in the AlkB and P450 phylogenetic trees are comprised solely of novel sequences from this study. The absence of known protein

sequences from the GenBank database in many of the clades suggests that the primers used in this study were effective in capturing new alkane monooxygenases from the diverse Passaic River sediment and enrichment cultures. Pyrosequencing efforts revealed diversity in these gene families that was previously unknown when compared to known protein sequences. Gathering larger numbers of sequences allowed for greater sequencing depth and resolution of alkane monooxygenases in the Passaic River sediment. Additionally, when comparing the different sites down the river and replicates to each other, a clear divide is seen between samples enriched with the hydrocarbon and those before enrichment. This shift supports that there was a response by the bacterial populations from each site following exposure to hexadecane. The response is more pronounced in AlkB than P450 hydroxylases, which may be due to the differences in known diversity between the two groups. Since the cytochrome P450 enzyme family is very diverse with many different types of substrates, the primers used may select for a larger group of enzymes than the *alkB* primers.

Sampling location along the Passaic did have some affect on the 16S populations recovered, showing that one location is not sufficient to obtain an accurate idea of community diversity. Despite continual river mixing in the tidal portion of the Passaic, sediment collected from sites within closer proximity are more likely to be similar than a site further downstream.

References

1. **Adams, D. A., O'Conner, J. S., Weisberg S. B.** (1998). Sediment quality of the NY/NJ harbor system. Environmental Protection Agency.
2. **National Oceanographic and Atmospheric Association** (1995). Magnitude and extent of sediment toxicity in the Hudson-Raritan Estuary. Silver Spring, MD, Coastal monitoring and bioeffects assessment division.
3. **Beilen, J. B. V., and E. G. Funhoff**. 2007. Alkane hydroxylases involved in microbial alkane degradation. FEMS Microbiology Letters **74**:13-21.
4. **Brydon, N. F.** (1974). The Passaic River: Past, present, future. New Brunswick, NJ, Rutgers University.
5. **Chai, B., Q. Wang, J. Fish, D. M. McGarrel, J. M. Tiedje, and C. J. R** (2010). The ribosomal database project: improved tools and services for rRNA analysis 100th General meeting of the american society for microbiology. San Diego, California.
6. **Cole, J. R., B. Chai, R. J. Farris, Q. Wang, A. S. Kulam-Syed-Mohideen, D. M. McGarrell, A. M. Bandela, E. Cardenas, G. M. Garrity, and J. M. Tiedje**. 2007. The ribosomal database project (RDP-II): introducing myRDP space and quality controlled public data. Nucleic acids research **35**:D169-172.
7. **Cole, J. R., Q. Wang, E. Cardenas, J. Fish, B. Chai, R. J. Farris, A. S. Kulam-Syed-Mohideen, D. M. McGarrell, T. Marsh, G. M. Garrity, and J. M. Tiedje**. 2009. The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. Nucleic acids research **37**:D141-145.
8. **Galishoff, S.** (1988). Newark, The Nation's unhealthiest city. New Brunswick, New Jersey, Rutgers University press.
9. **Headlee, T. J.** (1945). The mosquitoes of New Jersey and their control. New Brunswick, New Jersey, Rutgers University press.
10. **Iannuzzi, T. J., T. N. Armstrong, J. B. Thelen, D. F. Ludwig, and C. E. Firstenberg**. 2005. Characteriziation of chemical contamination in shallow-water estuarine habitats of an industrialized river. Part 1: Organic compounds. Soil and sediment contamination **14**:13-33.
11. **Iannuzzi, T. J., D. F. Ludwig, J. M. Wallin, J. C. Kinnell, R. W. Dunford, and D. W. H.** (2002). A common tradegy: History of an urban river Amherst, MA, Amherst scientific publishing.
12. **Iwai, S., B. Chai, W. J. Sul, J. R. Cole, S. A. Hashsham, and J. M. Tiedje**. 2010. Gene-targeted-metagenomics reveals extensive diversity of aromatic dioxygenase genes in the environment. The ISME journal **4**:279-285.
13. **Jacobson, D.** 1958. The pollution problem of the Passaic River. Proceedings of the New Jersey historical society **76**:186-198.
14. **Jesus, E. d. C., E. Susilawati, S. L. Smith, Q. Wang, B. Chai, R. Farris, J. L. M. Rodrigues, K. D. Thelen, and J. M. Tiedje**. 2010. Bacterial Communities in the rhizosphere of biofuel crops grown on marginal lands as evaluated by 16S rRNA gene pyrosequences. Bioenergy research **3**:20-27.

15. **Kay, D. P., J. L. Newsted, M. T. Benkinney, T. J. Iannuzzi, and J. P. Giesy**. 2008. Passaic river sediment interstitial water phase I toxicity identification evaluation. Chemosphere **70**:1737-1747.
16. **Kloos, K., J. C. Munch, and M. Schloter**. 2006. A new method for the detection of alkane-monooxygenase homologous genes (alkB) in soils based on PCR-hybridization. J. microbiolo. methods **66**:486 - 496.
17. **Liu, C., W. Wang, Y. Wu, Z. Zhou, Q. Lai, and Z. Shao**. 2011. Multiple alkane hydroxylase systems in a marine alkane degrader, Alcanivorax dieselolei B-5. Environmental microbiology **13**:1168-1178.
18. **Malcolm Pirnie, I.** (2007). Lower Passaic River restoration project: Draft source control early action focused feasibility study.
19. **Ni Chadhain, S. M., R. S. Norman, K. V. Pesce, J. J. Kukor, and G. J. Zylstra**. 2006. Microbial dioxygenase gene population shifts during polycyclic aromatic hydrocarbon biodegradation. Applied and environmental microbiology **72**:4078-4087.
20. **Porter, K. G., and Y. S. Feig**. 1980. The use of DAPI for identifying and counting aquatic microflora. Limnology and oceanography **25**:943-948.
21. **Stanier, R. Y., N. J. Palleroni, and M. Doudoroff**. 1966. The aerobic pseudomonads: a taxonomic study. Journal of general microbiology **43**:159-271.
22. **Sul, W. J., J. R. Cole, C. Jesus Eda, Q. Wang, R. J. Farris, J. A. Fish, and J. M. Tiedje**. 2011. Bacterial community comparisons by taxonomy-supervised analysis independent of sequence alignment and clustering. Proceedings of the national academy of sciences of the United States of America **108**:14637-14642.
23. **Torsvik, V., J. Goksoyr, and F. L. Daae**. 1990. High diversity in DNA of soil bacteria. Applied and environmental microbiology **56**:782-787.
24. **Torsvik, V., and L. Ovreas**. 2002. Microbial diversity and function in soil: from genes to ecosystems. Current opinion in microbiology **5**:240-245.
25. **van Beilen, J. B., and E. G. Funhoff**. 2005. Expanding the alkane oxygenase toolbox: new enzymes and applications. Current opinion in biotechnology **16**:308-314.
26. **van Beilen, J. B., E. G. Funhoff, A. V. Loon, A. Just, L. Kaysser, M. Bouza, M. Ro, Z. Li, and B. Witholt**. 2006. Cytochrome P450 alkane hydroxylases of the CYP153 family are common in alkane-degrading eubacteria lacking integral membrane alkane hydroxylases. Applied and environmental microbiology **72**:59-65.
27. **Wang, L., W. Wang, Q. Lai, and Z. Shao**. 2010. Gene diversity of CYP153A and AlkB alkane hydroxylases in oil-degrading bacteria isolated from the Atlantic Ocean. Environmental microbiology **12**:1230-1242.
28. **Wang, Q., G. M. Garrity, J. M. Tiedje, and J. R. Cole**. 2007. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. Applied and environmental microbiology **73**:5261-5267.
29. **Whyte, L. G., T. H. Smits, D. Labbe, B. Witholt, C. W. Greer, and J. B. van Beilen**. 2002. Gene cloning and characterization of multiple alkane hydroxylase systems in Rhodococcus strains Q15 and NRRL B-16531. Applied and environmental microbiology **68**:5933-5942.

30. **Whyte, L. G., T. H. M. Smits, B. Witholt, C. W. Greer, and J. B. V. Beilen**. 2002. Gene cloning and characterization of multiple alkane hydroxylase systems in *Rhodococcus* strains Q15 and NRRL B-16531. Applied and environmental microbiology **68**:5933-5942.

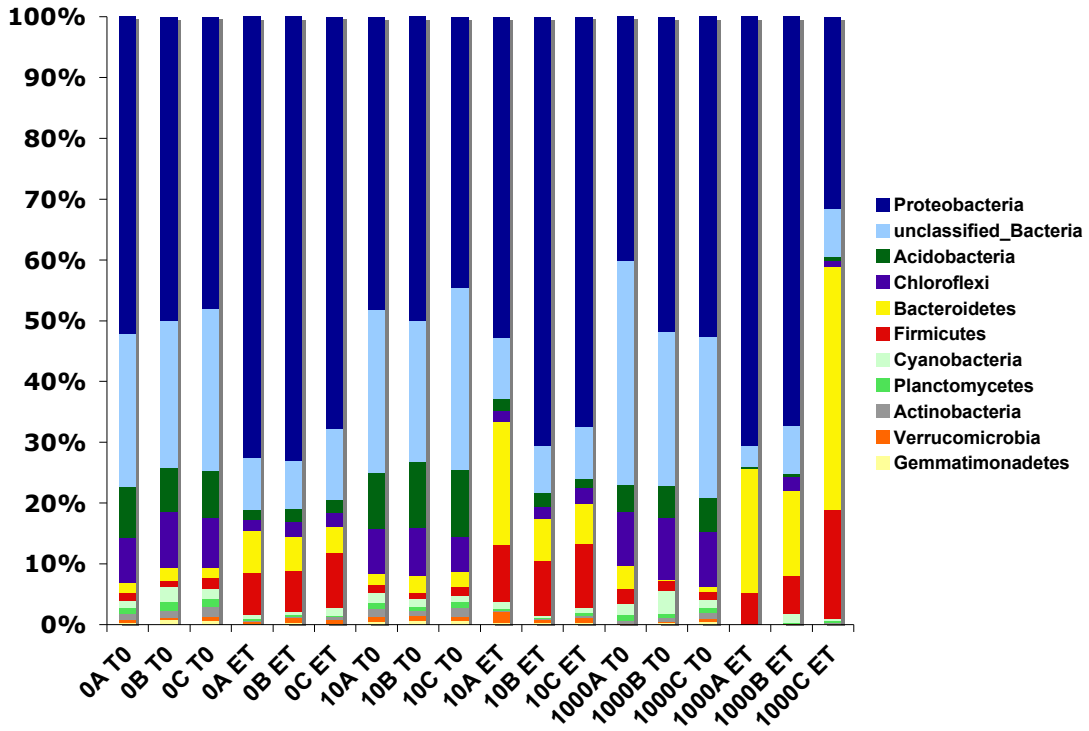


Figure 1. 16S rRNA phyla before and after enrichment: The x-axis indicates the site in meters (0, 10, or 1000) followed by the replicate letter (A, B, C). T0 indicates the sediment sample before enrichment and ET indicates the sediment sample after enrichment. The y-axis indicates the percent of the total sequences at the phyla level.

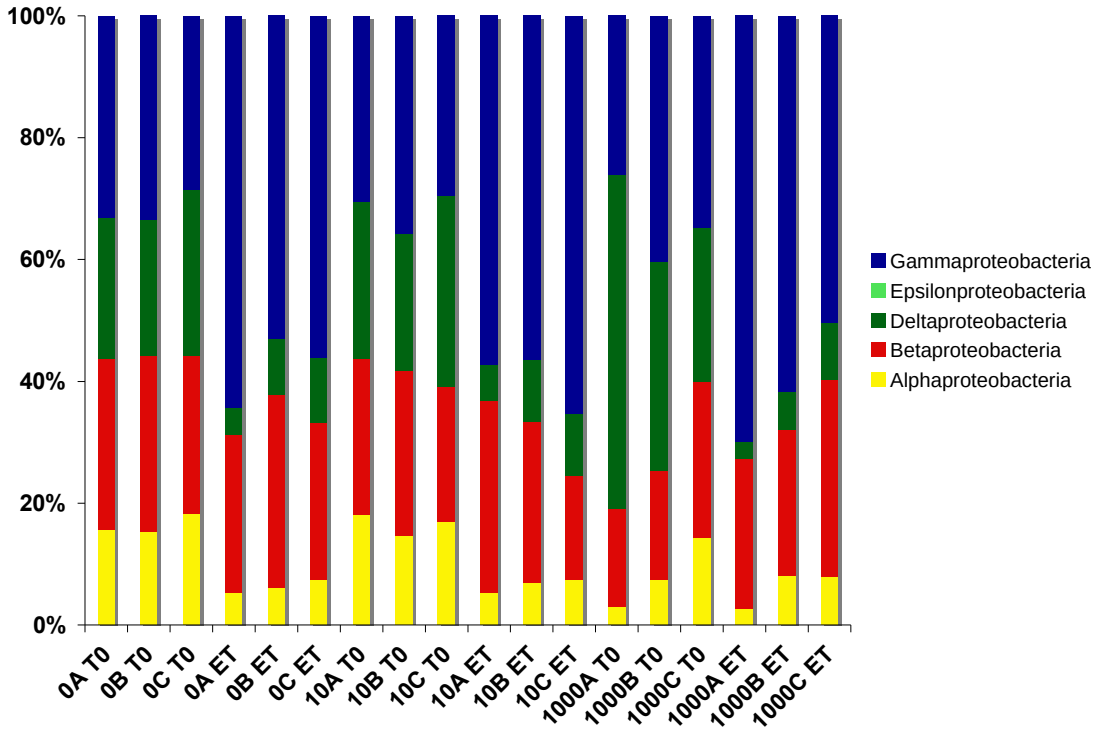
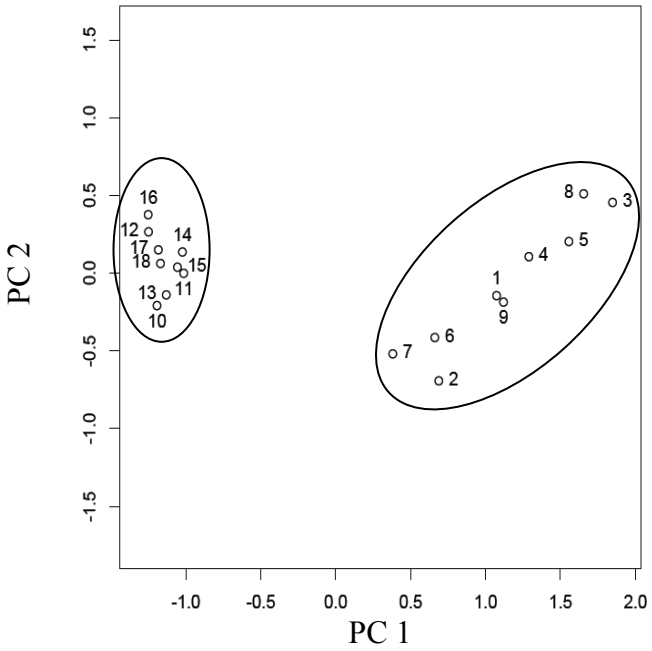
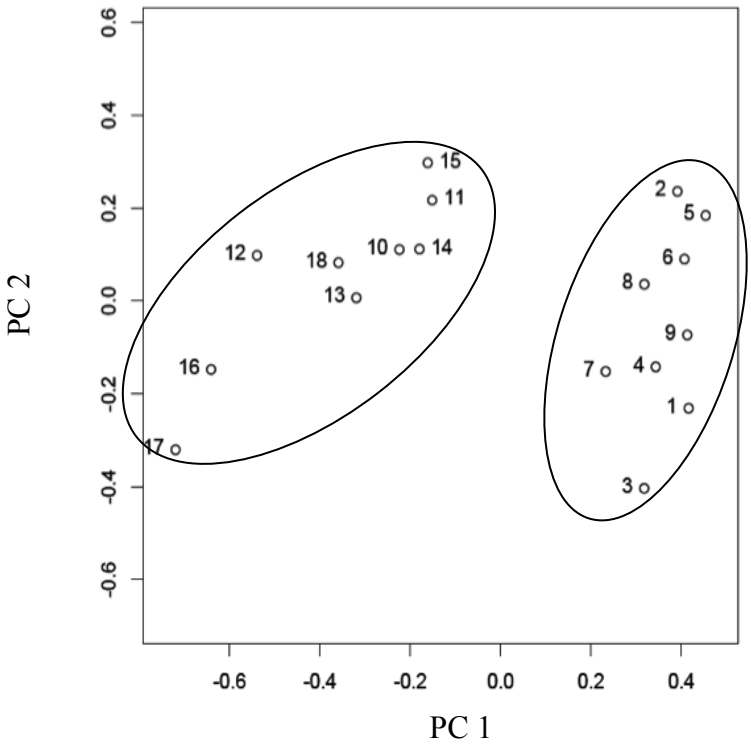


Figure 2. 16S rRNA proteobacterial classes: The x-axis indicates the site in meters (0, 10, or 1000) followed by the replicate letter (A, B, C). T0 indicates the sediment sample before enrichment and ET indicates the sediment sample after enrichment. On the y-axis is the percent of total proteobacteria sequences.

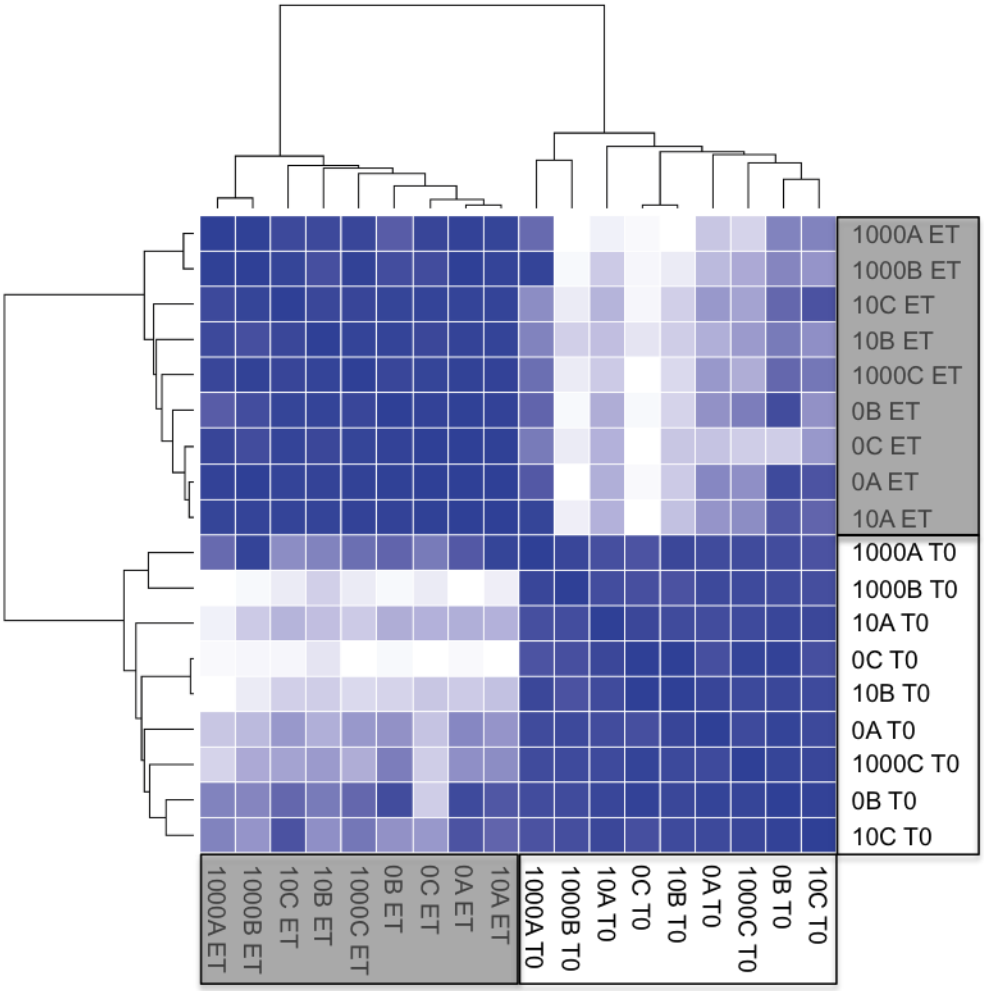

PC 2
PC 1
1.5
1.0
0.5
0.0
-0.5
-1.0
-1.5
-1.0
-0.5
0.0
0.5
1.0
1.5
2.0
16
12
17
18
14
15
13
11
10
8
3
5
4
1
9
6
7
2

3A.

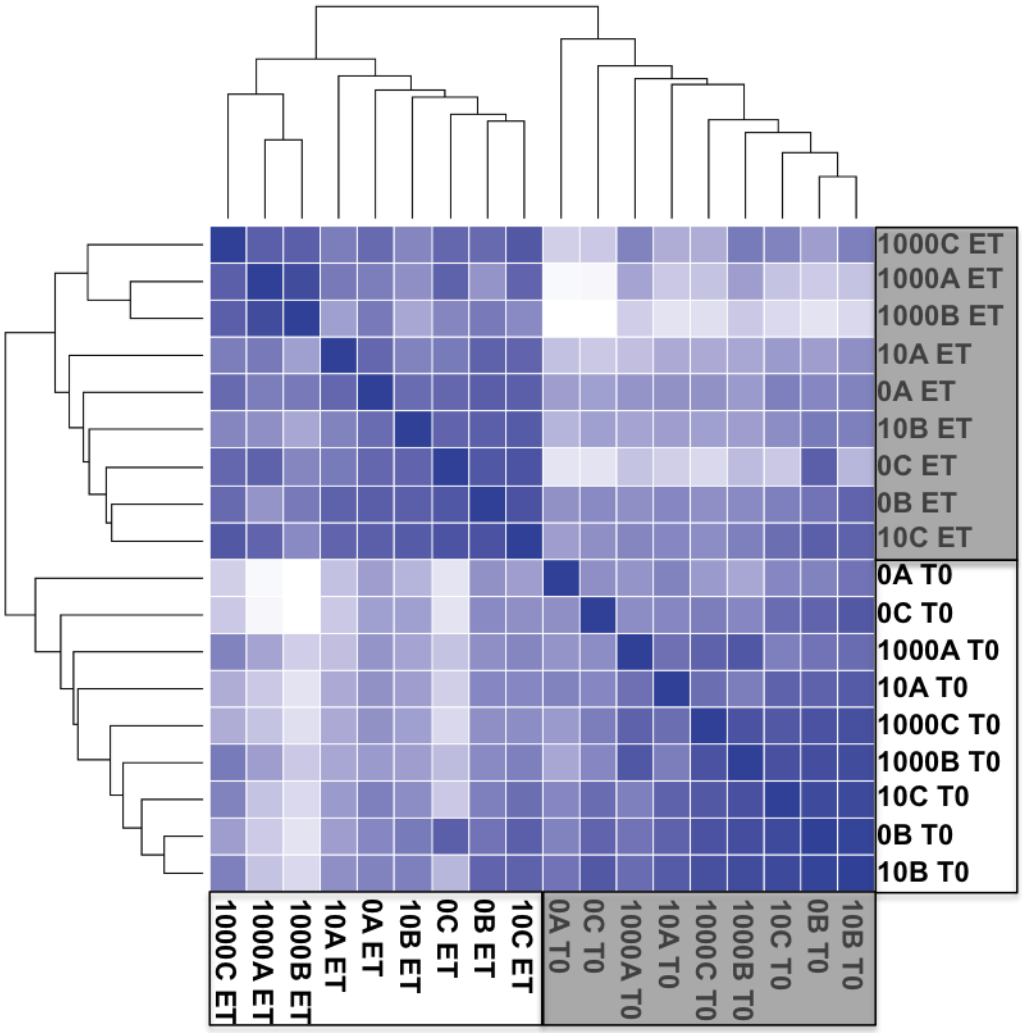

PC 2
PC 1
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.6
-0.4
-0.2
0.0
0.2
0.4
15
11
12
18
10
14
13
16
17
2
5
6
8
9
7
4
1
3

3B.

Figure 3. The principal component analysis of each sampling location in relationship to each other: A. Principal Component Analysis of AlkB sequences. B. Principal Component Analysis of P450 sequences. The site in meters is followed by the replicate (A, B, C) and before (T0) and after enrichment (ET). Points are 1, 0A ET; 2, 0B ET; 3, 0C ET; 4, 10A ET; 5, 10B ET; 6, 10C ET; 7, 1000A ET; 8, 1000B ET; 9, 1000C ET; 10, 0A ET; 11, 0B ET; 12, 0C ET; 13, 10A ET; 14, 10B ET; 15, 10C ET; 16, 1000A ET; 17, 1000B ET; and 18, 1000C ET.

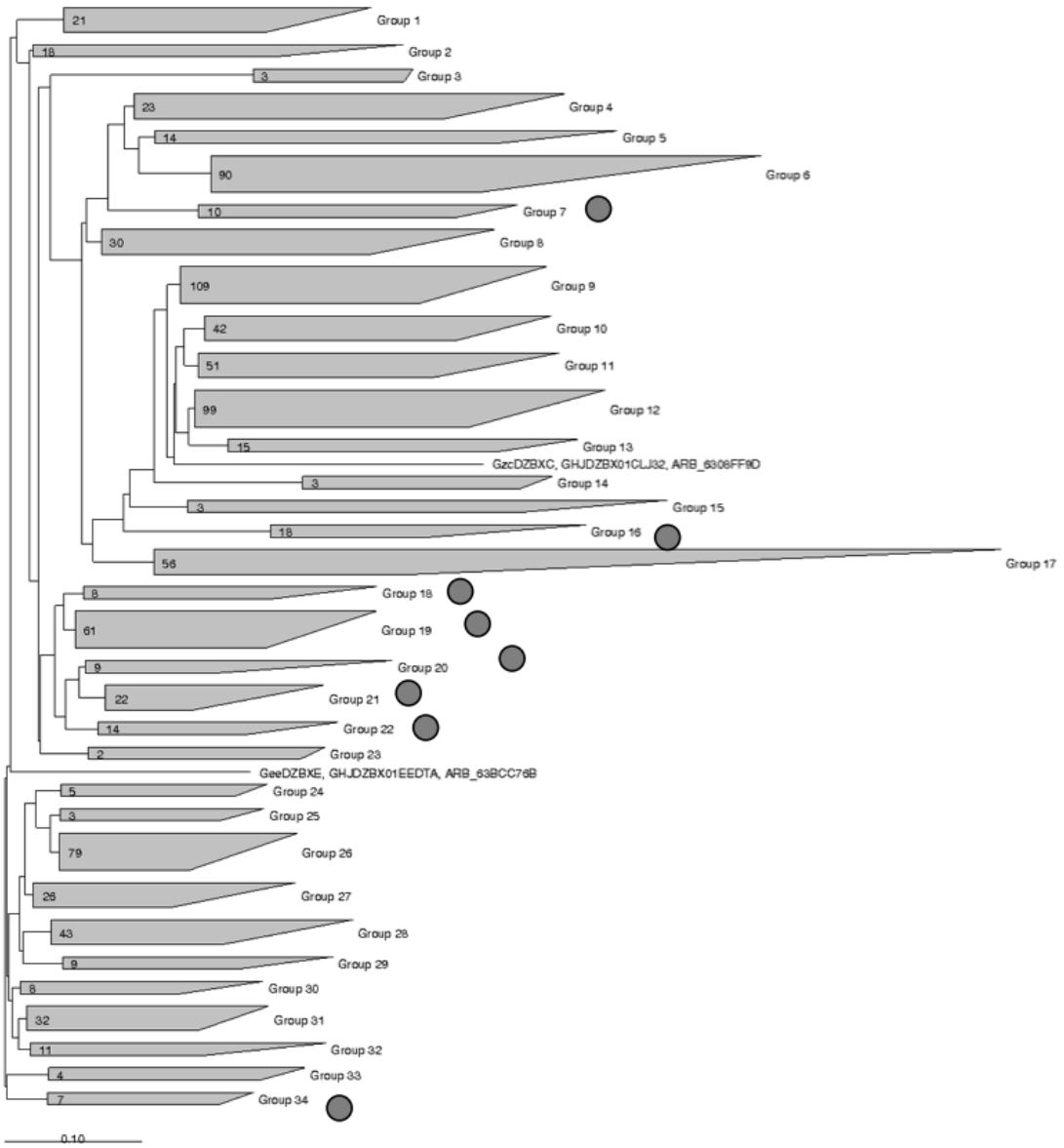

1000A ET
1000B ET
10C ET
10B ET
1000C ET
0B ET
0C ET
0A ET
10A ET
1000A T0
1000B T0
10A T0
0C T0
10B T0
0A T0
1000C T0
0B T0
10C T0
1000A T0
1000B T0
10A T0
0C T0
10B T0
0A T0
1000C T0
0B T0
10C T0
1000A ET
1000B ET
10C ET
10B ET
1000C ET
0B ET
0C ET
0A ET
10A ET

4A.

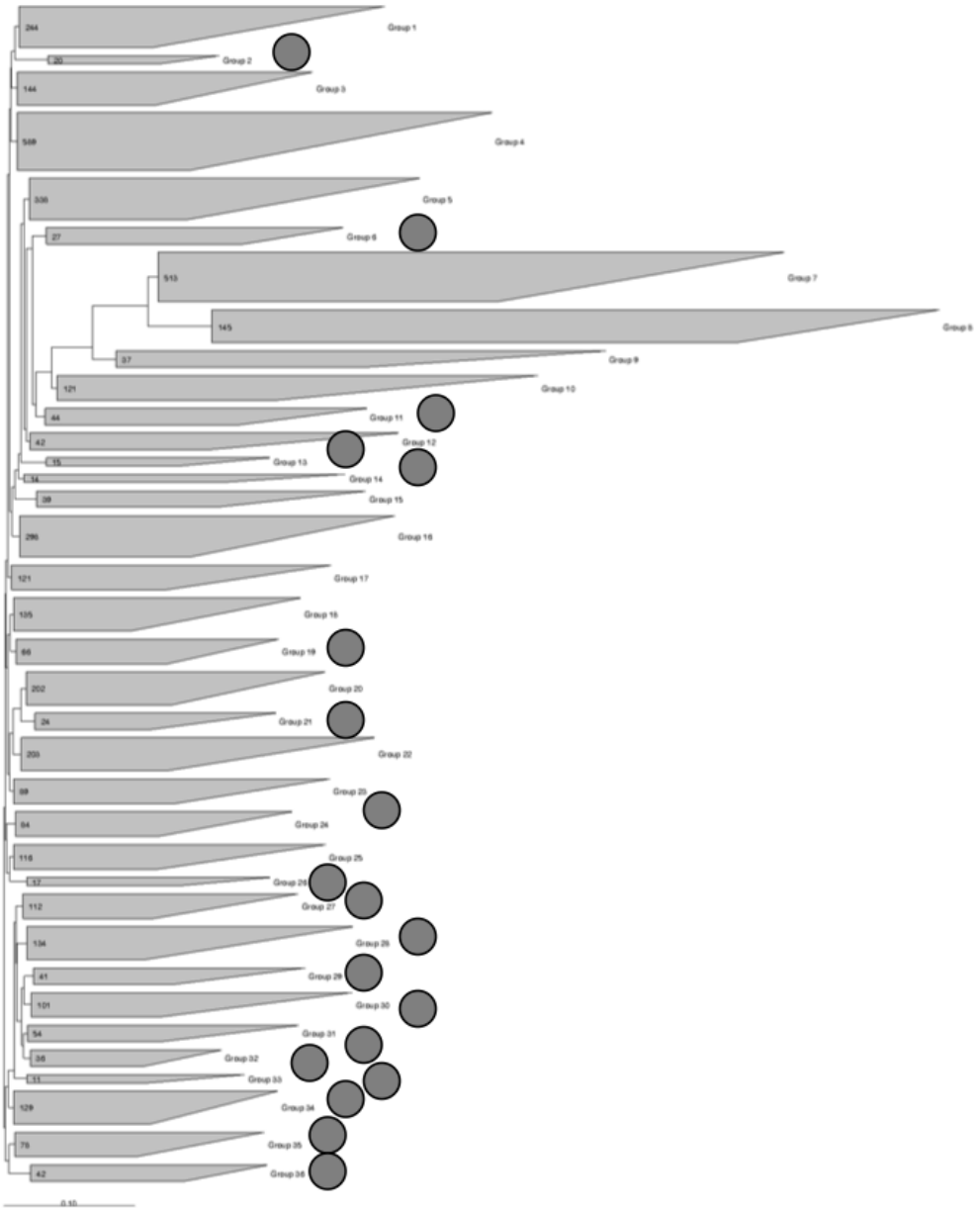


4B.

Figure 4. Cluster dendrograms and heatmaps: Dendrograms for A. AlkB, and B. P450 monooxgenases comparing sites of each sampling location in relationship to each other are on the top and left of the diagrams. The site is in meters is followed by replicate (A, B, C), and before enrichment (T0) or following enrichment (ET). The heatmap also compares the sites to each other showing the degree of similarity between the sampling sites. The darker blue the greater the similarity between sites.


21
Group 1
18
Group 2
3
Group 3
23
Group 4
14
Group 5
90
Group 6
10
Group 7
30
Group 8
109
Group 9
42
Group 10
51
Group 11
99
Group 12
15
Group 13
GzcDZBXC, GHJDZBX01CLJ32, ARB_6306FF9D
3
Group 14
3
Group 15
18
Group 16
56
Group 17
8
Group 18
61
Group 19
9
Group 20
22
Group 21
14
Group 22
2
Group 23
GeeDZBXE, GHJDZBX01EEDTA, ARB_63BCC76B
5
Group 24
3
Group 25
79
Group 26
26
Group 27
43
Group 28
9
Group 29
8
Group 30
32
Group 31
11
Group 32
4
Group 33
7
Group 34
0.10

5A.



5B.

Figure 5. Neighbor joining phylogenetic trees: A. AlkB, and B. P450 phylogenetic trees were constructed at 80% similarity. Shaded circles indicate novel clades composed only of Passaic River sequences.

Group	Reference Sequences	Passaic River Sequences
1	3	18
2	11	7
3	2	1
4	6	17
5	3	11
6	37	53
7	0	10
8	5	25
9	91	18
10	10	32
11	22	29
12	6	93
13	9	9
14	1	2
15	2	1
16	0	18
17	1	55
18	0	8
19	0	61
20	0	9
21	0	22
22	0	14
23	1	1
24	1	4
25	1	2
26	51	28
27	6	20
28	8	35
29	1	8
30	4	4
31	14	18
32	2	9
33	3	1
34	0	7

1A.

Group	Reference Sequences	Passaic River Sequences
1	4	240
2	0	20
3	6	138
4	7	582
5	81	257
6	0	27
7	188	325
8	101	44
9	3	34
10	8	113
11	0	44
12	3	39
13	0	15
14	0	14
15	1	38
16	19	279
17	28	93
18	2	133
19	0	66
20	14	188
21	0	24
22	5	198
23	18	71
24	0	84
25	1	115
26	0	17
27	0	112
28	0	134
29	0	41
30	0	101
31	0	54
32	0	36
33	0	11
34	0	129
35	0	78
36	0	42

1B.

Table 1. Protein family composition. The number of known GenBank reference sequences and Passaic River sequences are indicated for A. AlkB, and B. P450 type monooxygenases.

CHAPTER THREE

Alkane monooxygenase diversity recovered from Central Asia sediments

---

Elyse A. Rodgers-Vieira[1], Tinatin Doolotkeldieva[4], Jerome J. Kukor[2], Lee J. Kerkhof[3], and Gerben J. Zylstra[1]

Department of Biochemistry and Microbiology[1], Department of Environmental Sciences[2], and Institute of Marine and Coastal Sciences[3], New Brunswick, New Jersey 08901, Kyrgyz-Turkish International University, Kyrgyzstan[4].

ABSTRACT

Previous studies in our lab and elsewhere showed that different microbial gene groups are obtained from geographically distant sites. Alkane monooxygenases activate terminal carbon-carbon bonds in the first step of alkane degradation with the addition of an oxygen to the compound, resulting in an alcohol product. The aim of this study was to test the hypothesis that unique microbial populations and alkane monooxygenase (*alkB* and Cytochrome P450) genes can be recovered from geographically distinct sites, with varying degrees of anthropogenic impact. Sediment was collected from six river and stream sites with various levels of human impact located in Bishkek, Kyrgyzstan as well as Dushanbe and Khorog, Tajikistan, in Central Asia. The sites included rural and mountain streams, as well as rivers near or in cities. All sediments were enriched in media containing hexadecane as a carbon source and subcultured into fresh media with the alkane two times. DNA was extracted from each enrichment culture at various points throughout the enrichment for 16S rRNA and *alkB* clone library construction. The enrichment cultures were also analyzed by terminal restriction fragment length polymorphism (TRFLP) analysis of both 16S rRNA and *alkB* genes. The *alkB* and P450 genes, as well as 16S rRNA, were targeted with 454 pyrosequencing to obtain an estimate of microbial and alkane hydroxylase diversity at the sample locations and enrichments. The data show that distinct populations of *alkB* and P450 genes can be enriched from different locations and countries in Central Asia confirming that there is a wide diversity of alkane degrading genes in the environment.

INTRODUCTION

Our understanding of microbial diversity has increased immensely in recent decades. This is in part due to the advent of molecular techniques to study microbes directly in the environment as well as phylogenetic tools to organize the information gained from such studies (6). Evidence from several studies suggests that biogeographical distribution of microbes in the environment varies across the globe (12, 20, 28). Soil is very heterogeneous, and can contain thousands of different bacterial species per gram. The same holds true in aquatic sediments. Only a small component of this diversity (perhaps <1%) is captured using standard isolation techniques (6). A much larger proportion of these microbes have the potential to be cultured with increasingly efficient culturing techniques. However, molecular methods allow us to explore this untapped diversity directly in environmental samples.

To investigate bacterial distribution we focused on a largely unstudied geographical area. Other studies have found alkane monooxygenases in diverse environments (8, 11, 14, 27). The purpose of our current study was to determine the diversity of alkane monooxygenase genes (*alkB* and P450 type) present in a variety of spatially distinct environments in Central Asia, and investigate if novel alkane hydroxylase gene families are present at the sites.

Alkanes are a type of saturated hydrocarbon composed exclusively of carbon and hydrogen atoms and exhibit low chemical reactivity at ambient temperatures. They are the major component of crude oil, are formed by geochemical reactions, and are ubiquitous in the environment (30). Medium length alkanes (C_{10}–C_{20}) are readily activated by alkane monooxygenases for biodegradation (31). These enzymes act by

addition of an oxygen to the terminal carbon of the alkane resulting in an alcohol product. The alcohol product is then ultimately metabolized into fatty acids by dehydrogenases. Most alkane degrading bacteria act upon alkanes that contain more than 10 carbon atoms. Alkane degraders also occur in unpolluted environments where alkanes are produced in low concentrations by plants, algae, and other organisms (15). Two types of alkane monooxygenases are active in hydroxylating alkanes. AlkB is an integral membrane-bound diiron enzyme while cytochrome P450s are soluble enzymes (5). Multiple AlkB or P450 enzymes can be present in one bacterial strain, increasing the potential for hydrocarbon utilization (21, 24, 29).

The bacterial communities were examined both prior to and following enrichment on the 16 carbon *n*-alkane hexadecane. Functionally active populations and genes were identified through molecular based methods. Environmental samples were collected from a variety of sites in Central Asia, including both rural and isolated habitats and urban environments. For comparison samples were also collected from rural areas which would be minimally, if at all, exposed to hydrocarbons. Community composition for each location was compared by terminal restriction length polymorphism (TRFLP) for both 16S rRNA and *alkB*, as well as by 454 pyrosequencing for 16S rRNA and the alkane hydroxylases *alkB*, and P450.

MATERIALS AND METHODS

Sampling Locations. Sediment was obtained from six sites in Central Asia (Table 1, Figure 1). Samples were collected from the Panj River near the Kozideh Village along the Dushanbe-Khorog road on the Afghanistan border (RMI001), from two sites in the Dushanbinka River in Dushanbe (RMI005, RMI006), from a small high mountain stream along the Bishkek-Osh road (RMI009), and from the Chua River in Bishkek (RMI010A). All samples were stored in sterile plastic jars at room temperature for transport.

Enrichment Conditions. 50 ml of minimal salts base (MSB) media was inoculated with .5 g sediment and amended with 50 μl of hexadecane (Sigma Aldrich, St. Louis, MO) (22). The cultures were incubated shaking at 30°C for 6-7 days. Each was subcultured 2 times, at the beginning of the enrichment (first time point), the middle of the enrichment, and the end time point.

DNA Extraction, PCR and Clone Libraries. DNA was extracted using the PowerSoil kit (MoBio, Carlsbad, CA). PCR was performed in a 50 μl volume using ReadyMix Taq PCR Reaction Mix with $MgCl_2$ with ~25 ng of template (Sigma, Saint Louis, MO). The final concentrations for the 16S PCR were: 1.5 units Taq DNA polymerase, 10 mM Tris-HCl, 50mM KCl, 1.5 mM $MgCl_2$, 0.001% gelatin and 0.2 mM dNTPs with 3.2 pmol of each primer. 27F (AGAGTTTGATCMTGGCTCAG) and 1525R (AAGGAGGTGATCCAGCC) primers were used for 16S rRNA amplification (9). Cycling conditions were: 95°C for 5 minutes, 35 cycles at 95°C for 30 seconds, 55°C for 30 seconds, 72°C for 1 minute 45 seconds, followed by a final extension step at 72°C for 15 minutes. For *alkB* amplification degenerate primers alkBF (AATACHGSVCAYGAGCTCRGYCAYAAR) and alkBR

(GCRTGRTGATCAGARTGHCGYTG) (7) were used with the same reagent concentrations as for the 16S rRNA amplification except for an increase in $MgCl_2$ concentration with the addition of 0.8 µl 50 mM $MgCl_2$. Cycling conditions were: 95°C for 5 minutes, 35 cycles at 95°C for 30 seconds, 55°C for 30 seconds, 72°C for 1 minute 45 seconds, followed by a final extension step at 72°C for 15 minutes. Each PCR product was gel purified using the GeneCleanII kit (MP Biomedicals, Solon, OH) and ligated into the pGem-T Easy Vector (Promega, Madison, WI). The plasmids were transformed into Top10 cells and the libraries were screened by colony PCR using Sp6 and T7P primers to amplify the insert. Exonuclease I was used to remove excess nucleotides and primers from the PCR product for direct sequencing (USB, Cleveland, OH). Sequencing was performed for the 16S rRNA products using the 27F primer while SP6 and T7P were used to sequence the *alkB* product.

Phylogenetic analysis of clone library sequences. *alkB* sequences were assembled in groups by date and sample site using SeqMan Lasergene software with a sequence identity of 97% (DNAStar, Madison, WI). The contigs were exported and aligned in ClustalX (10). The alignment was imported into Phylip to produce a lower distance matrix for use in Mothur (4). Mothur software was used to group the sequences into OTUs and to create phylogenetic trees (16, 17, 18, 19).

TRFLP. Amplified PCR products were tagged in a reaction using primers that were fluorescently labeled on the 5' end with 6-FAM ([6]-carboxy-fluorescein)(25). 16S rRNA genes were tagged using labeled 27F (AGAGTTTGATCMTGGCTCAG) while *alkB* PCR products were tagged using labeled alkBR (GCRTGRTGATCAGARTGHCG) with the same reagents and concentrations as stated previously (7). The same PCR

conditions were used for the labeling reactions except only 5 cycles of amplification were performed. The amplified products were quantified and sized on a 1% agarose gel and compared to 5 µl of Lambda HindIII size standard. Fluorescently labeled PCR products were digested with either MnlI or AluI restriction enzymes separately for TRFLP analysis in 20 µl reactions at 37°C for 4 to 6 hours with 15 ng DNA using the manufacturer supplied buffers and 1 U of enzyme (New England Bio-labs, Beverly, MA). DNA was precipitated in 95% ethanol for 20 minutes at -20°C and centrifuged at 14,000 rpm at 4°C for 15 minutes. The pelleted DNA was washed with 100 µl of 70% ethanol and centrifuged at 14,000 rpm for 15 minutes at 4°C and dried for 30 minutes in a vacuum centrifuge. Samples were run in 20 µl volumes with ROX size standard on an ABI Prism 310 autoanalyzer (Applied Biosystems, Foster City, CA).

Amplification for Pyrosequencing. DNA was quantified using a NanoDrop Spectrophotometer (Thermo Scientific, Wilmington, DE). The DNA was amplified with fusion tagged primers using the FastStart High Fidelity PCR system (Roche). The composition of the PCR mixtures using the FastStart High Fidelity PCR System were 1 U enzyme, 1X FastStart High Fidelity Reaction Buffer, 1 µM each of the forward and reverse primers, 0.2 mM dNTPs, and 150 ng/µl BSA in a final volume of 20 µl per reaction. Approximately 25 ng of template DNA was added to each reaction. For *alkB* amplification, barcoded degenerate primers alkBF (AATACHGSVCAYGAGCTCRGYCAYAAR) and alkBR (GCRTGRTGATCAGARTGHCGYTG) were used (7). For *alkB* amplification the following protocol was used: 95°C for 5 minutes, followed by 35 cycles of 95°C for 45 seconds, 55°C for 45 seconds, and 72°C for 1 minute, with a final elongation of 72°C for

4 minutes. The tagged primers used to amplify P450 genes were P450F (GTSGGCGGCAACGACACSAC) and P450R (GCASCGGTGGATGCCGAAGCCRAA) (23). The PCR protocol used for P450 amplification was: 95°C for 5 minutes, followed by 35 cycles of 95°C for 30 seconds, 58°C for 30 seconds, and 72°C for 40 seconds, with a final elongation at 72°C for 4 minutes. The 16S rRNA primers used were 16S_577F (AYTGGGYDTAAAGNG) and 16S_926R (CCGTCAATTCMTTTRAGT). For 16S rRNA amplification the PCR protocol was: 95°C for 2 minutes, followed by 30 cycles of 95°C for 45 seconds, 55°C for 45 seconds, and 72°C for 1 minute, with a final elongation at 72°C for 4 minutes.

454 Sequencing. DNA was amplified with fusion tagged primers using the FastStart high Fidelity PCR system (Roche). The amplified products were sized on a 1% agarose gel. All PCR reactions were done in triplicate or higher and pooled during gel purification using the Qiagen Gel Extraction Kit and quantified using a Nanodrop Spectrophotometer (Qiagen, Valencia, CA and Thermo Scientific, Wilmington, DE).

The raw sequencing reads were run through a quality filter in the Ribosomal Database Project Pipeline (RDP) (2, 3). 16S rRNA reads that passed the quality filter were used for further analysis in the RDP Pyrosequencing Pipeline. The 16S rRNA sequences were clustered at 80% for analysis and clusters were identified using the RDP Classifier (26).

Functional gene reads deemed to be of sufficient quality were frame shift corrected with RDPs FrameBot tool using the Hidden Markov Model (HMM) built specifically for each functional gene using known sequences from GenBank (1). The HMM model uses known sequences to determine the most conserved regions and which variations within

the sequence are the most likely to occur. Further analysis of the *alkB* and P450 sequences was performed using the RDP Fungene Pipeline (2, 3). Sequences were clustered at 80% identity for analysis.

RESULTS

DNA extracted from each sediment was amplified using 16S rRNA specific primers and degenerate AlkB primers for clone libraries. The same primers, with added fluorescent labels, were used to amplify products for TRFLP analysis. TRFLP profiles of the *alkB* and 16S rRNA PCR products were generated with the restriction enzymes AluI and MnlI. Clustering analysis of the resulting data revealed no distinct clustering based on the sample location (Figure 2). The dendrograms are very deeply branched suggesting that the sampling sites are very dissimilar to each other possibly due to distance between sampling sites and a plethora of ecological differences from site to site. Clone libraries for 16S rRNA and *alkB* were prepared and sequenced. Since approximately 50 clones were sequenced per library, it is expected that only a portion of the total diversity present in these environments would be recovered. These sequences were translated and compared to known protein sequences in GenBank. A dendrogram was constructed of all the Central Asia AlkB sequences and representative GenBank sequences (Figure 3). Phylogenetic analysis revealed known and unknown groups of AlkB sequences, which are indicated with brackets in the phylogenetic tree. To determine if a difference in alkane hydroxylase populations was present at the different sites, the clone library data for *alkB* was broken up by site (Figure 4). Clustering of all the *alkB* sequences at 80% identity resulted in 30 OTUs. Certain OTUs were present in some sites, and not in

others. No OTU was found in all of the sites. However, some OTUs represented the majority of clones recovered from certain sites.

The clone libraries demonstrated that there was alkane hydroxylase diversity at these sites that warranted further study. As a result, 454 pyrosequencing for 16S rRNA, and the alkane hydroxylase genes P450 and *alkB* were performed using the same DNA as amplified for the clone libraries. 16S rRNA sequences quality filtered (Table 2) and clustered using the RDP Pyrosequencing Pipeline Classifier. At the phyla level, a high proportion of sequences before and after enrichment belonged to the Proteobacteria (Figure 5). The fraction of sequences that represented Proteobacteria increased to a majority following enrichment at all sites regardless of country of origin. More diversity was seen before enrichment in all sediments. Before enrichment a significant portion of the population belonged to the Acidobacteria phyla, which was not detectable following enrichment. The Proteobacteria were not dominated by any one class before enrichment (Figure 6). In all sediment samples Gammaproteobacteria, Alphaproteobacteria, Betaproteobacteria, and Deltaproteobacteria comprised significant portions of the microbial population. However, following enrichment the Deltaproteobacteria disappeared, making way for an increase in Betaproteobacteria in all but one of the samples (the Tajik City enrichment).

The *alkB* (Table 3) and P450 (Table 4) sequences obtained from pyrosequencing were translated and grouped by their site of origin and time point in the enrichment. From this data, cluster dendrograms and heatmaps were constructed to determine relatedness between the samples (Figures 7 and 8). Each branch on the dendrogram represents a sample site and timepoint. The dendrograms are placed perpendicular to

each other to compare the samples in a matrix showing the degree to which each is related. Sites that are most similar to each other are represented by dark blue in the matrix of the heatmap, the least related sites are conversely indicated by varying shades of blue to white (no relatedness) depending on the degree of relatedness.

To determine if new AlkB or P450s were present in the Central Asia sites, the sequences were compared to known protein sequences in the GenBank database (Figure 8). Each pie chart shows the percent similarity of the experimental sequences to known GenBank sequences. An increase in sequences with a higher similarity to known alkane hydroxylases was observed following enrichment with hexadecane. The highest proportion (56%) of the AlkB sequences before enrichment was only 50-59% similar to known AlkB amino acid sequences. Following enrichment this number increased to 60-69% and represented 47% of the total. A similar trend was seen with P450 where the highest proportion (51%) before enrichment with the alkane was 70-79%, which increased to 90-99% after enrichment comprising 32%.

DISCUSSION

The TRFLP and clone library survey of *alkB* populations in Central Asia sediments revealed that diverse families of alkane hydroxylases were present that warranted further study. Alkane hydroxylase sequences from the clone libraries belonged to 30 OTUs when clustered at 80% nucleotide similarity (Figure 3). None of these OTUs are present in all of the libraries suggesting the presence of divergent alkane hydroxylase populations at the geographically distant sites of sample collection. This trend is

supported by the TRFLP clustering diagrams which are deeply branching for both AlkB and 16S rRNA, suggesting that the sample sites are quite divergent in their microbial populations (Figure 1). Clustering is not seen based on country, or degree of anthropogenic impact at the sample site. Sediments originating from rural, suburban, and urban waterways appear to have few similarities in their microbial populations suggesting that both country of origin and level of human and industrial impact affects alkane hydroxylase populations in the environment.

AlkB clone library sequences are similar to a variety of known AlkB sequences from organisms such as *Rhodococcus*, *Mycobacterium*, *Pseudomonas*, *Limnobacter*, and *Acinetobacter* (Figure 2). While some experimental sequences fell within clades with known sequences, others formed novel clades apart from known protein sequences throughout all stages of the enrichment indicating that new alkane hydroxylase sequences were found in Central Asia sediments.

Pyrosequencing of the 16S rRNA gene revealed shifts within the microbial community with enrichment on the medium length alkane hexadecane (Figure 4). Prior to enrichment, Acidobacteria and many other phyla including Crenarchaeota and Gemmatimonadetes comprised a significant percentage of the total 16S rRNA sequences. Proteobacteria represented the largest portion of 16S rRNA diversity prior to enrichment and increased dramatically after enrichment. Regardless of the country the sediment originated from, Proteobacteria accounted for 85% or more of the 16S rRNA sequences following enrichment. Differences in the Proteobacteria composition of the communities at different sites was observed However, there were similar patterns in the phyla present before enrichment and a shift in the population resulted from hexadecane exposure.

Since the Proteobacteria were the dominant phyla, it is pertinent to look at the proportion of each class in regard to each sampling site (Figure 5). Before enrichment with hexadecane Gammaproteobacteria, Alphaproteobacteria, Betaproteobacteria, and Deltaproteobacteria comprised significant portions of the microbial community at all sample sites in Kyrgyzstan and Tajikistan. Following enrichment the Deltaproteobacteria sequences disappeared in all the samples while Betaproteobacteria dominated in the enrichments. Betaproteobacteria include many aerobic and facultative bacteria that are known hydrocarbon degraders such as *Burkholderia, Comamonas, Poloromonas, Ralstonia,* and many others. Not much is known about Betaproteobacteria that degrade straight chain alkanes (13). So, it is significant that such a large increase was seen in that class of Proteobacteria in this study. Interestingly, the Tajikistan City enrichment followed a different trend than other sediment samples in regards to Proteobacteria. Approximately 90% of the Proteobacterial sequences recovered in that enriched sample were Gammaproteobacteria. The best studied group of bacteria for alkane degradation is *Pseudomonas* which are Gammaproteobacteria. While it is not surprising that the Gammaproteobacteria sequences went up in numbers, it is surprising that in four of the five enrichments, the Betaproteobacteria dominated.

One of the purposes of this study was to see if different alkane hydroxylase populations could be found in different countries in unique environments in Central Asia. Cluster dendrograms and heatmaps compiled for both AlkB and P450 alkane hydroxylase sequences obtained from pyrosequencing show diverse and unrelated populations of monooxygenases (Figure 6 and 7). The degree of relatedness between populations is indicated by varying shades of blue within a matrix associated with branches on cluster

dendrograms. The darker squares within the heatmap matrix indicate a higher level of relatedness between the corresponding sample sites. Overall, sediments before enrichment are more similar to each other than following enrichment for the AlkB sequences. Even so, these relationships are very deeply branched suggesting significant differences between the populations. No clear trends were observed based on country of origin or by site type (City, Suburb, or Rural). However, in some cases such as the AlkB Tajik suburb samples, the enriched samples clustered together. The Cytochrome P450 data suggested a slightly different picture of alkane monooxygenase diversity. Some clustering was seen between Tajik City and Suburb sites, as well as Kyrgyz samples, but this trend did not hold in all cases. The P450 dendrogram shows mostly very deep branching associations between sample sites, even between samples in the same country.

To determine if new alkane monooxygenase families were found in Central Asia sediment and enrichments, the experimental sequences obtained were compared to known sequences (Figure 8). Before enrichment, almost all of the AlkB sequences (84%) were less than 80% similar to known alkane monooxygenase sequences in GenBank and 5% of the sequences were less than 50% similar. After enrichment the Central Asia sequences were more similar to GenBank sequences. However, a majority of the AlkB sequences (61%) were still less than 80% similar to known GenBank sequences. Only 21% were similar to known sequences above 90%. A similar trend was seen with cytochrome P450. Prior to enrichment the largest portion of the total (51%) was 70-79% similar to the known P450s. After enrichment with hexadecane, together the portions that were 90-99% and 80-89% similar comprised the majority (51%) of the total. This trend is logical, since most of the sequences in the GenBank database are associated with

culturable organisms. One would expect to see more unknown, or novel gene families before culturing since much of the microbial diversity seen in sediments and soils is lost using current laboratory culturing methods. Enrichment with a carbon source such as hexadecane, as used in this experiment, applies a selective pressure towards the success of medium length alkane degraders that are also amenable to the culturing conditions used. As a result, it is not surprising that there is a trend towards known monooxygenase sequences from known cultured organisms following enrichment. The *alkB* and P450 primers did successfully capture alkane monooxygenases from Central Asia sediment that are novel and more distantly related to many sequences that are currently published.

References

1. **Chai, B., Q. Wang, J. Fish, D. M. McGarrel, J. M. Tiedje, and C. J. R** (2010). The ribosomal database project: improved tools and services for rRNA analysis 100th General meeting of the american society for microbiology. San Diego, California.
2. **Cole, J. R., B. Chai, R. J. Farris, Q. Wang, A. S. Kulam-Syed-Mohideen, D. M. McGarrell, A. M. Bandela, E. Cardenas, G. M. Garrity, and J. M. Tiedje**. 2007. The ribosomal database project (RDP-II): introducing myRDP space and quality controlled public data. Nucleic acids research **35**:D169-172.
3. **Cole, J. R., Q. Wang, E. Cardenas, J. Fish, B. Chai, R. J. Farris, A. S. Kulam-Syed-Mohideen, D. M. McGarrell, T. Marsh, G. M. Garrity, and J. M. Tiedje**. 2009. The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. Nucleic acids research **37**:D141-145.
4. **Felsenstein, J.** (2005). PHYLIP (Phylogeny Inference Package) version 3.6. Distributed by the author. Department of Genome Sciences, University of Washington, Seattle.
5. **Funhoff, E. G., U. Bauer, I. García-rubio, B. Witholt, and J. B. V. Beilen**. 2006. CYP153A6, a soluble P450 oxygenase catalyzing terminal-alkane hydroxylation. J. Bacteriol **188**:5220-5227.
6. **Hugenholtz, P., B. M. Goebel, and N. R. Pace**. 1998. Impact of culture-independent studies on the emerging phylogenetic view of bacterial diversity. J Bacteriol **180**:4765-4774.
7. **Kloos, K., J. C. Munch, and M. Schloter**. 2006. A new method for the detection of alkane-monooxygenase homologous genes (alkB) in soils based on PCR-hybridization. J microbiological methods **66**:486 - 496.
8. **Kuhn, E., G. S. Bellicanta, and V. H. Pellizari**. 2009. New alk genes detected in Antarctic marine sediments. Environmental Microbiology **11**:669-673.
9. **Lane, D. J.** (1991). rRNA sequencing. Nucleic acid techniques in bacteria systematics. E. Stackebrandt and M. Goodfellow. New York, John Wiley & sons**:**115-175.
10. **Larkin, M. A., G. Blackshields, N. P. Brown, R. Chenna, P. A. McGettigan, H. McWilliam, F. Valentin, I. M. Wallace, A. Wilm, R. Lopez, J. D. Thompson, T. J. Gibson, and D. G. Higgins**. 2007. Clustal W and Clustal X version 2.0. Bioinformatics **23**:2947-2948.
11. **Margesin, R., D. Labbe, F. Schinner, C. W. Greer, and L. G. Whyte**. 2003. Characterization of hydrocarbon-degrading microbial populations in contaminated and pristine Alpine soils. Applied and environmental microbiology **69**:3085-3092.
12. **Martiny, J. B., B. J. Bohannan, J. H. Brown, R. K. Colwell, J. A. Fuhrman, J. L. Green, M. C. Horner-Devine, M. Kane, J. A. Krumins, C. R. Kuske, P. J. Morin, S. Naeem, L. Ovreas, A. L. Reysenbach, V. H. Smith, and J. T. Staley**. 2006. Microbial biogeography: putting microorganisms on the map. Nature reviews. Microbiology **4**:102-112.

13. **Parales, R. E.** (2010). Handbook of hydrocarbon and lipid microbiology, Part 4: Hydrocarbon degradation by betaproteobacteria. Berlin, Heidelberg, Springer-Verlag. 1715-1724.
14. **Powell, S. M., S. H. Ferguson, J. P. Bowman, and I. Snape**. 2006. Using real-time PCR to assess changes in the hydrocarbon-degrading microbial community in Antarctic soil during bioremediation. Microbial ecology **52**:523-532.
15. **Rojo, F.** 2005. Specificity at the end of the tunnel: understanding substrate length discrimination by the AlkB alkane hydroxylase. J Bacteriol **187**:19-22.
16. **Schloss, P. D., and J. Handelsman**. 2005. Introducing DOTUR, a computer program for defining operational taxonomic units and estimating species richness. Appl environ microbiol **71**:1501-1506.
17. **Schloss, P. D., and J. Handelsman**. 2006. Introducing SONS, a tool for operational taxonomic unit-based comparisons of microbial community memberships and structures. Appl environ microbiol **72**:6773-6779.
18. **Schloss, P. D., and J. Handelsman**. 2006. Introducing TreeClimber, a test to compare microbial community structures. Appl environ microbiol **72**:2379-2384.
19. **Schloss, P. D., B. R. Larget, and J. Handelsman**. 2004. Integration of microbial ecology and statistics: a test to compare gene libraries. Appl environ microbiol **70**:5485-5492.
20. **Sliwinski, M. K., and R. M. Goodman**. 2004. Spatial heterogeneity of crenarchaeal assemblages within mesophilic soil ecosystems as revealed by PCR-single-stranded conformation polymorphism profiling. Applied and environmental microbiology **70**:1811-1820.
21. **Sotsky, J. B., C. W. Greer, and R. M. Atlas**. 1994. Frequency of genes in aromatic and aliphatic hydrocarbon biodegradation pathways within bacterial populations from Alaskan sediments. Canadian journal of microbiology **40**:981-985.
22. **Stanier, R. Y., N. J. Palleroni, and M. Doudoroff**. 1966. The aerobic pseudomonads: a taxonomic study. Journal of general microbiology **43**:159-271.
23. **van Beilen, J. B., E. G. Funhoff, A. V. Loon, A. Just, L. Kaysser, M. Bouza, M. Ro, Z. Li, and B. Witholt**. 2006. Cytochrome P450 alkane hydroxylases of the CYP153 family are common in alkane-degrading eubacteria lacking integral membrane alkane hydroxylases. Applied and environmental microbiology **72**:59-65.
24. **van Beilen, J. B., E. G. Funhoff, A. van Loon, A. Just, L. Kaysser, M. Bouza, R. Holtackers, M. Rothlisberger, Z. Li, and B. Witholt**. 2006. Cytochrome P450 alkane hydroxylases of the CYP153 family are common in alkane-degrading eubacteria lacking integral membrane alkane hydroxylases. Applied and environmental microbiology **72**:59-65.
25. **Vetriani, C., H. V. Tran, and L. J. Kerkhof**. 2003. Fingerprinting microbial assemblages from the oxic/anoxic chemocline of the Black Sea. Applied and environmental microbiology **69**:6481-6488.
26. **Wang, Q., G. M. Garrity, J. M. Tiedje, and J. R. Cole**. 2007. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. Applied and environmental microbiology **73**:5261-5267.
27. **Wasmund, K., K. A. Burns, D. I. Kurtboke, and D. G. Bourne**. 2009. Novel alkane hydroxylase gene (alkB) diversity in sediments associated with

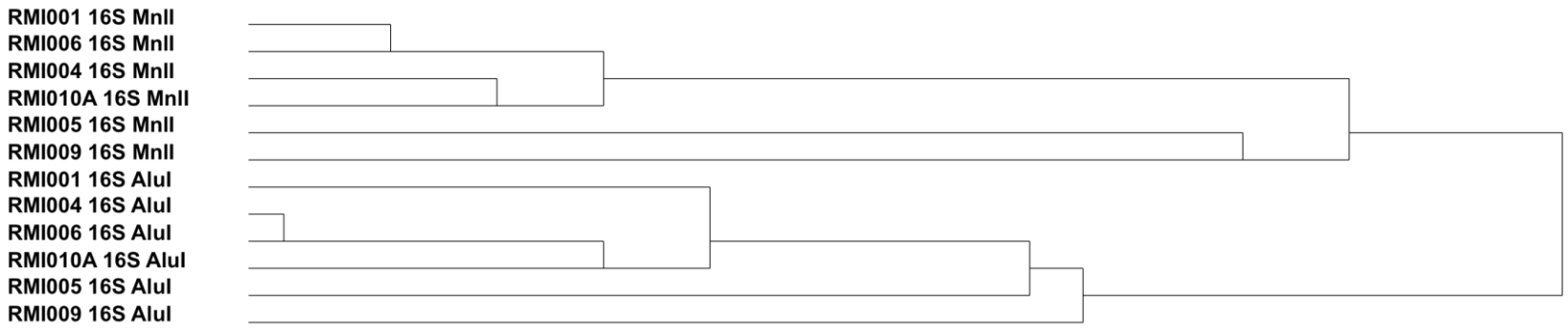
hydrocarbon seeps in the Timor Sea, Australia. Applied and environmental microbiology **75**:7391-7398.

28. **Wawrik, B., D. Kutliev, U. A. Abdivasievna, J. J. Kukor, G. J. Zylstra, and L. Kerkhof**. 2007. Biogeography of actinomycete communities and type II polyketide synthase genes in soils collected in New Jersey and Central Asia. Applied and environmental microbiology **73**:2982-2989.
29. **Whyte, L. G., T. H. Smits, D. Labbe, B. Witholt, C. W. Greer, and J. B. van Beilen**. 2002. Gene cloning and characterization of multiple alkane hydroxylase systems in Rhodococcus strains Q15 and NRRL B-16531. Applied and environmental microbiology **68**:5933-5942.
30. **Widdel, F., and R. Rabus**. 2001. Anaerobic biodegradation of saturated and aromatic hydrocarbons. Current opinion in biotechnology **12**:259-276.
31. **Yuste, L., M. E. Corbella, M. J. Turiegano, U. Karlson, A. Puyet, and F. Rojo**. 2000. Characterization of bacterial strains able to grow on high molecular mass residues from crude oil processing. FEMS microbiology ecology **32**:69-75.

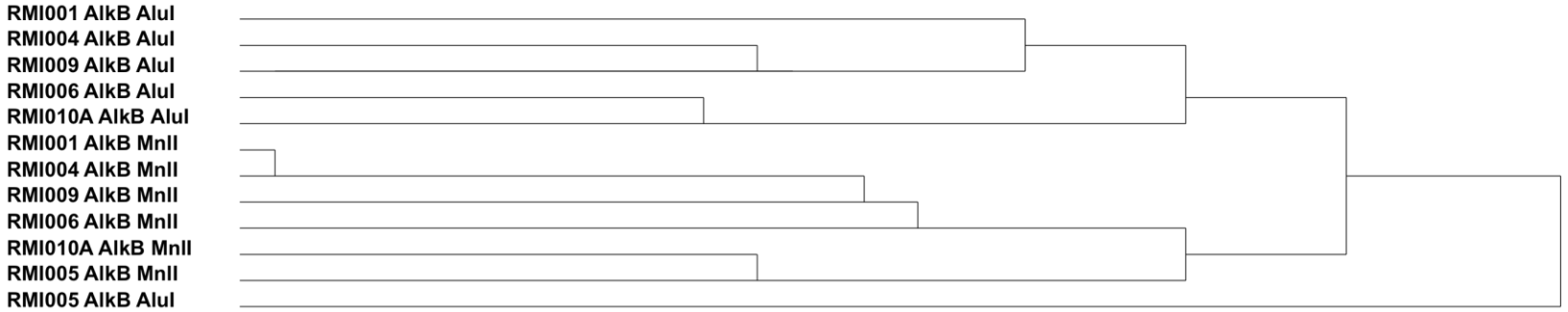
RMI001	Tajik Rural	Panj River near Khorog, sandy
RMI004	Tajik Suburb	Dushanbinka River in Dushanbe
RMI005	Tajik City	Dushanbinka river in Dushanbe, sandy
RMI006	Kyrgyz Suburb	Rural stream near Bishkek
RMI009	Kyrgyz Rural	High mountain stream south of Bishkek
RMI010A	Kyrgyz City	Chua River Canal through Bishkek

Table 1. Central Asia sample site descriptions: The first column includes the sample identifier, the second column is the site name, and the third includes a description of the site location.

Figure 1. Central Asia map: Map of the locations where samples were collected. RMI001 was collected near Khrog. RMI004 and RMI005 were collected in Dushanbe. RMI006, RMI009, and RMI0010 were collected near Bishkek.



A.



B.

Figure 2. TRFLP cluster analysis: A). 16S rRNA and **B).** *alkB* PCR products clustered by the presence or absence of peaks in TRFLP. Labeled by sample site, gene, and restriction enzyme used.

Sample Name	Match Barcode	Match QC	Avg. Length	% Passed QC
Tajik Rural FT	10718	8305	331	77.49%
Tajik Rural MT	10638	7908	331	74.34%
Tajik Rural ET	7661	5520	331	72.05%
Tajik Suburb FT	2109	1678	331	79.56%
Tajik Suburb MT	4082	3235	330	79.25%
Tajik Suburb ET	22622	19507	330	86.23%
Tajik City FT	4539	3570	331	78.65%
Tajik City MT	7776	6493	331	83.50%
Tajik City ET	9451	7837	330	82.92%
Kyrgyz Suburb FT	4919	3862	331	78.51%
Kyrgyz Suburb MT	11126	9391	330	84.41%
Kyrgyz Suburb ET	7442	6384	330	85.78%
Kyrgyz Rural FT	25839	19839	331	76.78%
Kyrgyz Rural MT	3464	2548	331	73.56%
Kyrgyz Rural ET	4648	3584	330	77.11%
Krygyz City FT	10812	8629	331	79.81%
Krygyz City MT	4638	3847	330	82.95%
Krygyz City ET	7224	5878	330	81.37%

Table 2. 16S rRNA sequencing statistics. FT is the first time point (sediment), MT is the middle time point, and ET is the end time point. The number of sequences that matched the barcode is in column 2, the percent that passed the quality control is in column 5.

Sample Name	Match Barcode	Match QC	Avg. Length	% Passed QC
Tajik Rural FT	2050	1140	436	55.61%
Tajik Rural MT	4603	3172	448	68.91%
Tajik Rural ET	6297	4627	451	73.48%
Tajik Suburb FT	4414	2545	429	57.66%
Tajik Suburb MT	7041	4672	449	66.35%
Tajik Suburb ET	2690	1824	452	67.81%
Tajik City FT	3692	1892	423	51.25%
Tajik City MT	2537	1092	462	43.04%
Tajik City ET	718	394	459	54.87%
Kyrgyz Suburb FT	2383	1511	436	63.41%
Kyrgyz Suburb MT	2263	1498	452	66.20%
Kyrgyz Suburb ET	1063	898	449	84.48%
Kyrgyz Rural FT	86	35	387	40.70%
Kyrgyz Rural MT	4260	2726	465	63.99%
Kyrgyz Rural ET	1466	1162	465	79.26%
Krygyz City FT	4586	2673	427	58.29%
Krygyz City MT	670	318	447	47.46%
Krygyz City ET	759	482	444	63.50%

Table 3. *alkB* sequencing statistics. FT is the first time point (sediment), MT is the middle time point, and ET is the end time point of enrichment. The number of sequences that matched the barcode is in column two, the percent that passed quality control is in column 5.

Sample Name	Match Barcode	Match QC	Avg. Length	% Passed QC
Tajik Rural FT	15938	14163	294	89%
Tajik Rural MT	11617	9969	295	86%
Tajik Rural ET	5690	3195	296	56%
Tajik Suburb FT	2218	946	296	43%
Tajik Suburb MT	11580	10176	295	88%
Tajik Suburb ET	3178	2620	294	82%
Tajik City FT	15490	13129	294	85%
Tajik City MT	8923	7911	295	89%
Tajik City ET	28745	17345	295	60%
Kyrgyz Suburb FT	6196	4700	295	76%
Kyrgyz Suburb MT	17233	14963	293	87%
Kyrgyz Suburb ET	11731	8850	289	75%
Kyrgyz Rural FT	12829	10150	295	79%
Kyrgyz Rural MT	6545	2757	294	42%
Kyrgyz Rural ET	7566	4411	295	58%
Krygyz City FT	4737	2318	294	49%
Krygyz City MT	14	13	295	93%
Krygyz City ET	2798	1761	292	63%

Table 4. P450 monooxygenase gene sequencing statistics. FT is the first time point (sediment), MT is the middle time point, and ET is the end time point of enrichment. The number of sequences that matched the barcode is in column two, the percent that passed quality control is in column 5.

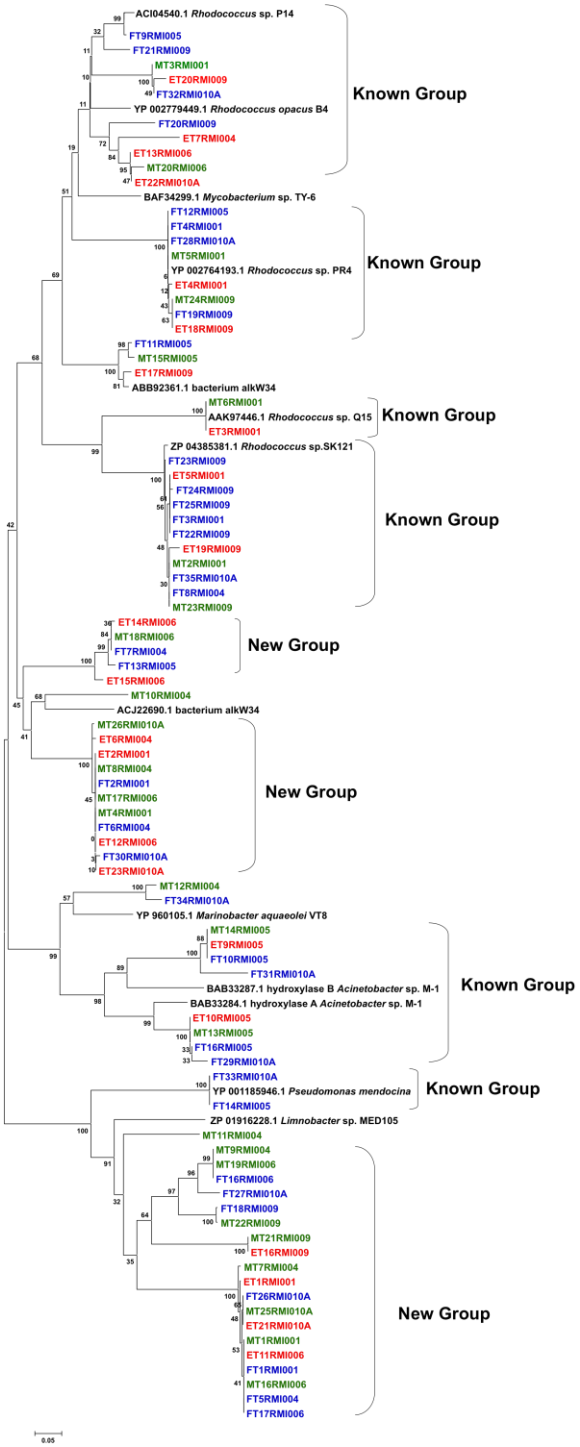


Figure 3. AlkB phylogenetic tree: *alkB* clone library sequences were translated into amino acid and were used to construct a phylogenetic tree with known AlkB sequences obtained from GenBank. Each sample sequence is labeled by enrichment time point: first time point (FT), mid time point (MT), and end time point (ET). The site name follows the timepoint. Known and new groups are indicated with brackets.

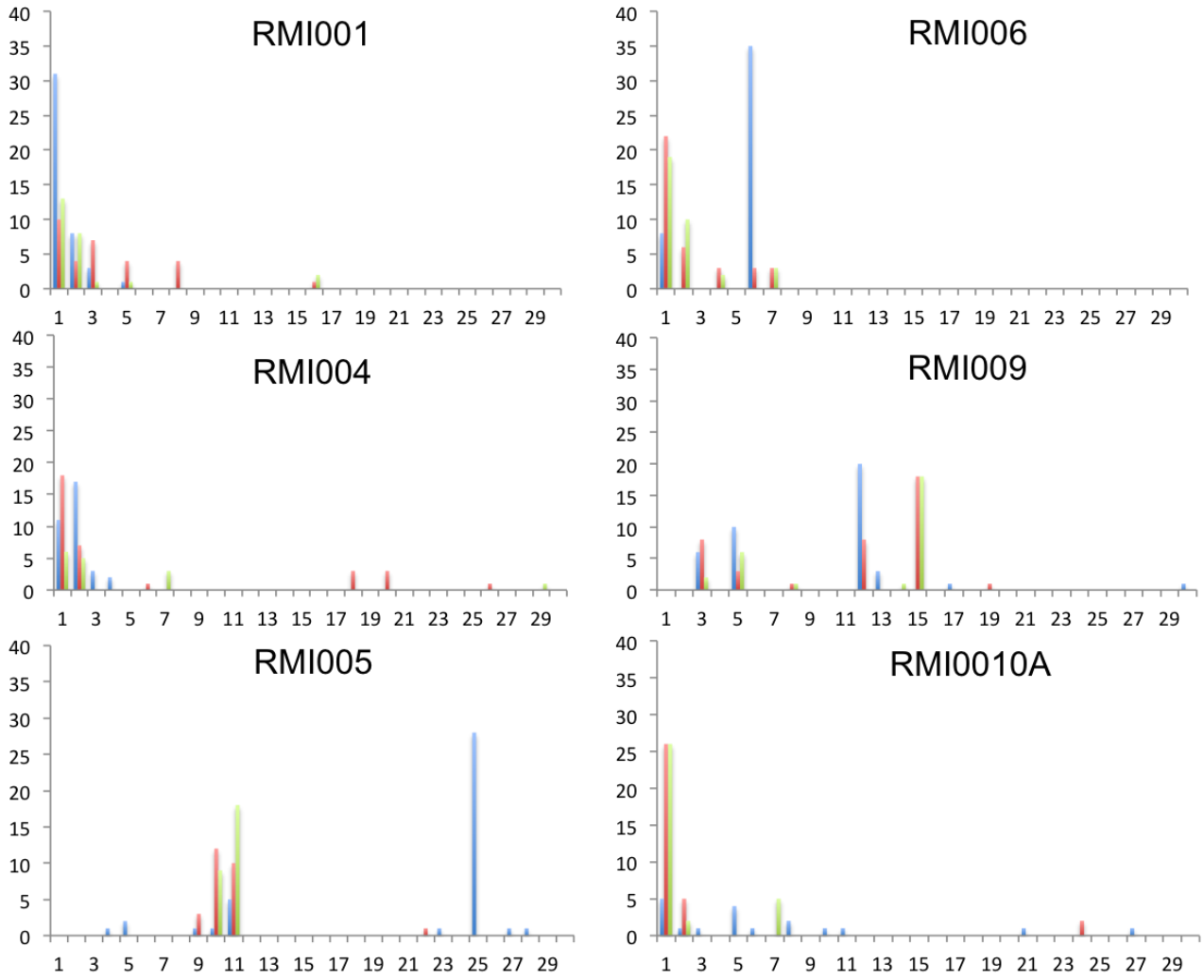

RMI001
RMI006
RMI004
RMI009
RMI005
RMI0010A
0 5 10 15 20 25 30 35 40
1 3 5 7 9 11 13 15 17 19 21 23 25 27 29

Figure 4. *alkB* gene family distribution: *alkB* sequences were translated into amino acid sequence and grouped at 80% identity for analysis. The bar graphs show the number of total families on the x axis (30 total) when the sequences were grouped at 80% identity. The y axis shows the total number of clones within each family at each site. Each sample location is shown on a separate graph. The time points are indicated by the following: first time point (blue), mid time point (red), and end time point (green).

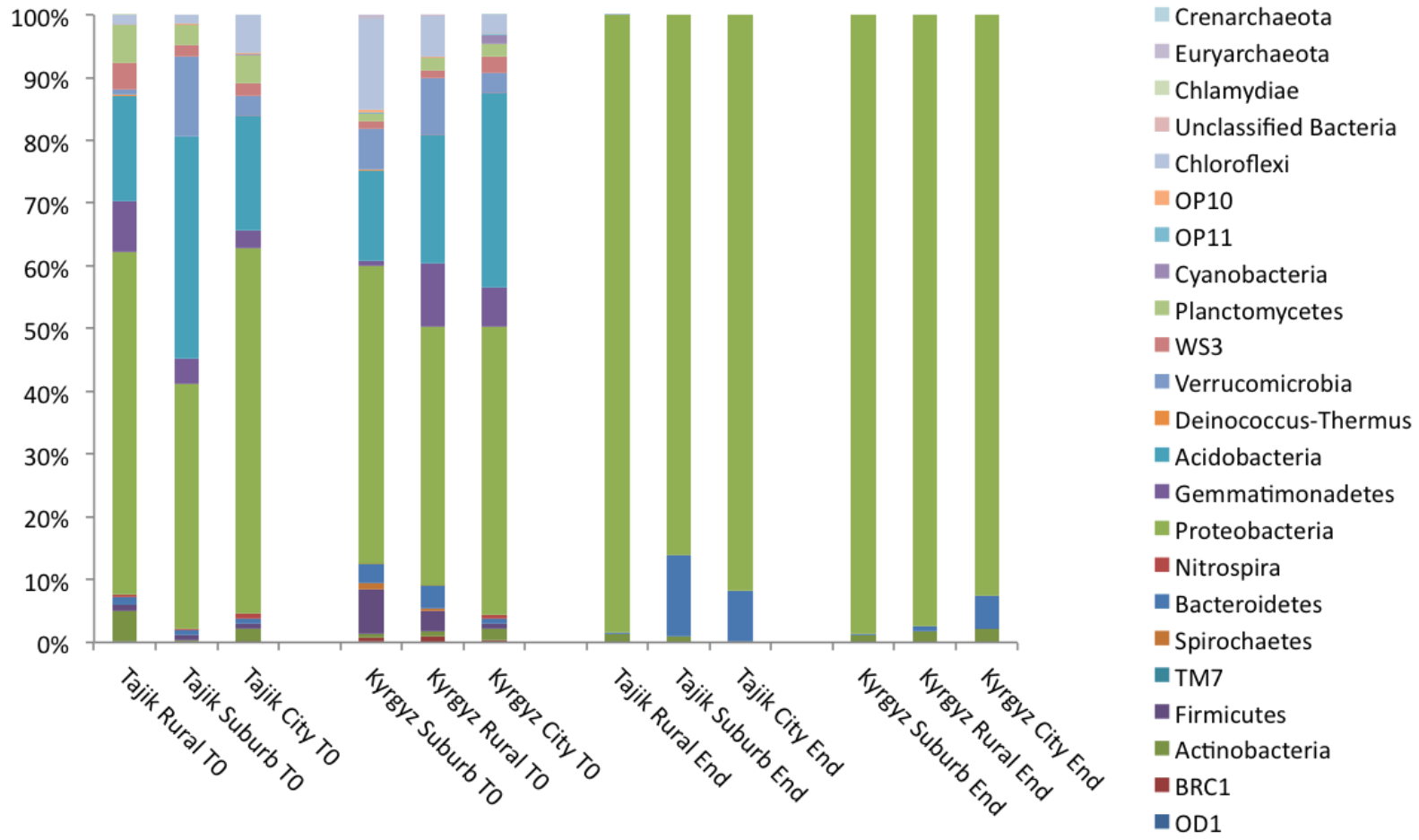


Figure 5. 16S rRNA phyla before and after enrichment: Sample sites labeled by country and type of environment, and before (T0) or after (ET) enrichment with hexadecane. Classified using the RDP Classifier.

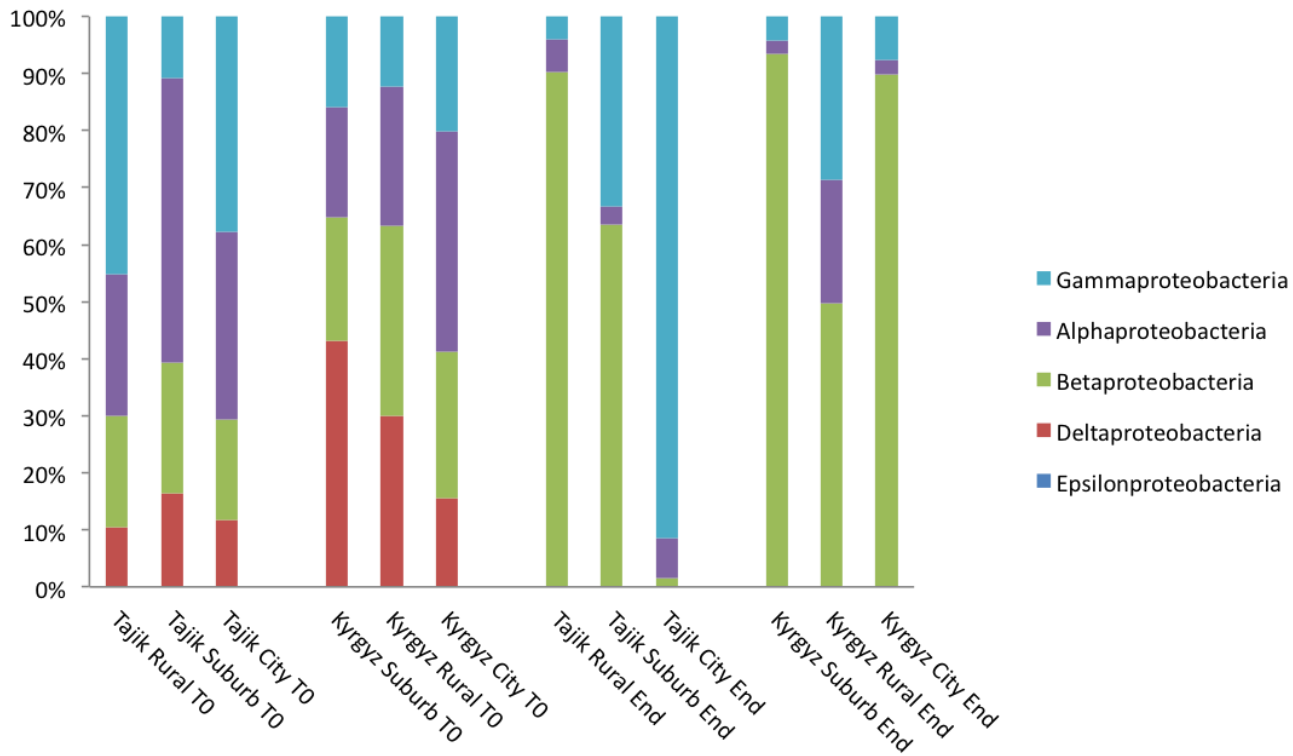


Figure 6. 16S rRNA proteobacteria classes: Sample sites labeled by country and type of environment, and before (T0) or after (ET) enrichment with hexadecane. Classified using the RDP Classifier.

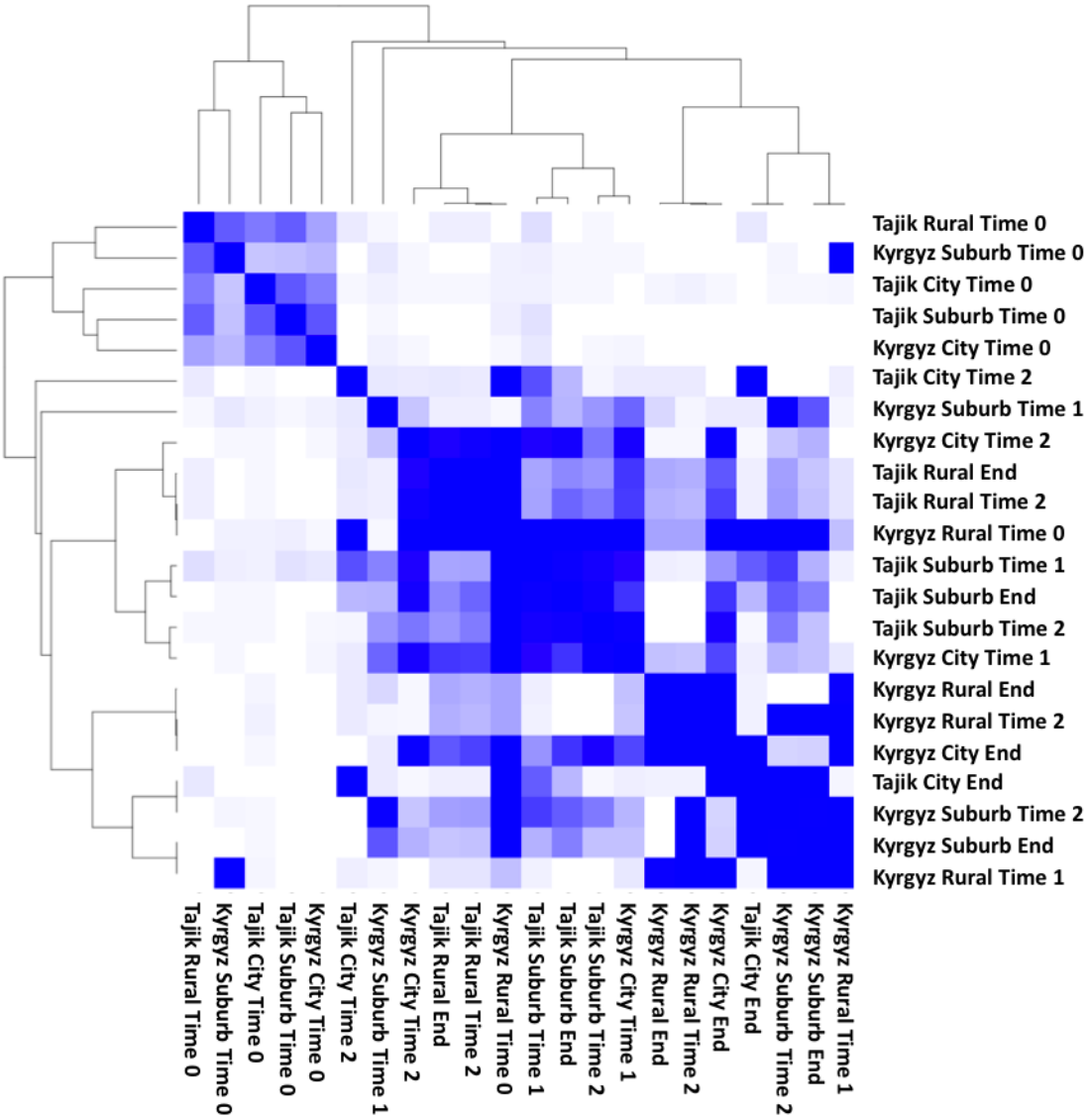


Figure 7. AlkB cluster dendrograms and heatmap: The sites and enrichment conditions were compared to each other by cluster dendrograms on the top and left of the heatmap. The heatmap shows relatedness between sites and enrichment conditions, showing the degree of similarity. The more similar two samples are, the darker the blue, and the least similar sites are associated with a white square in the corresponding portion

of the matrix. Sample sites labeled by country and type of environment, and before (T0) or after (ET) enrichment with hexadecane.

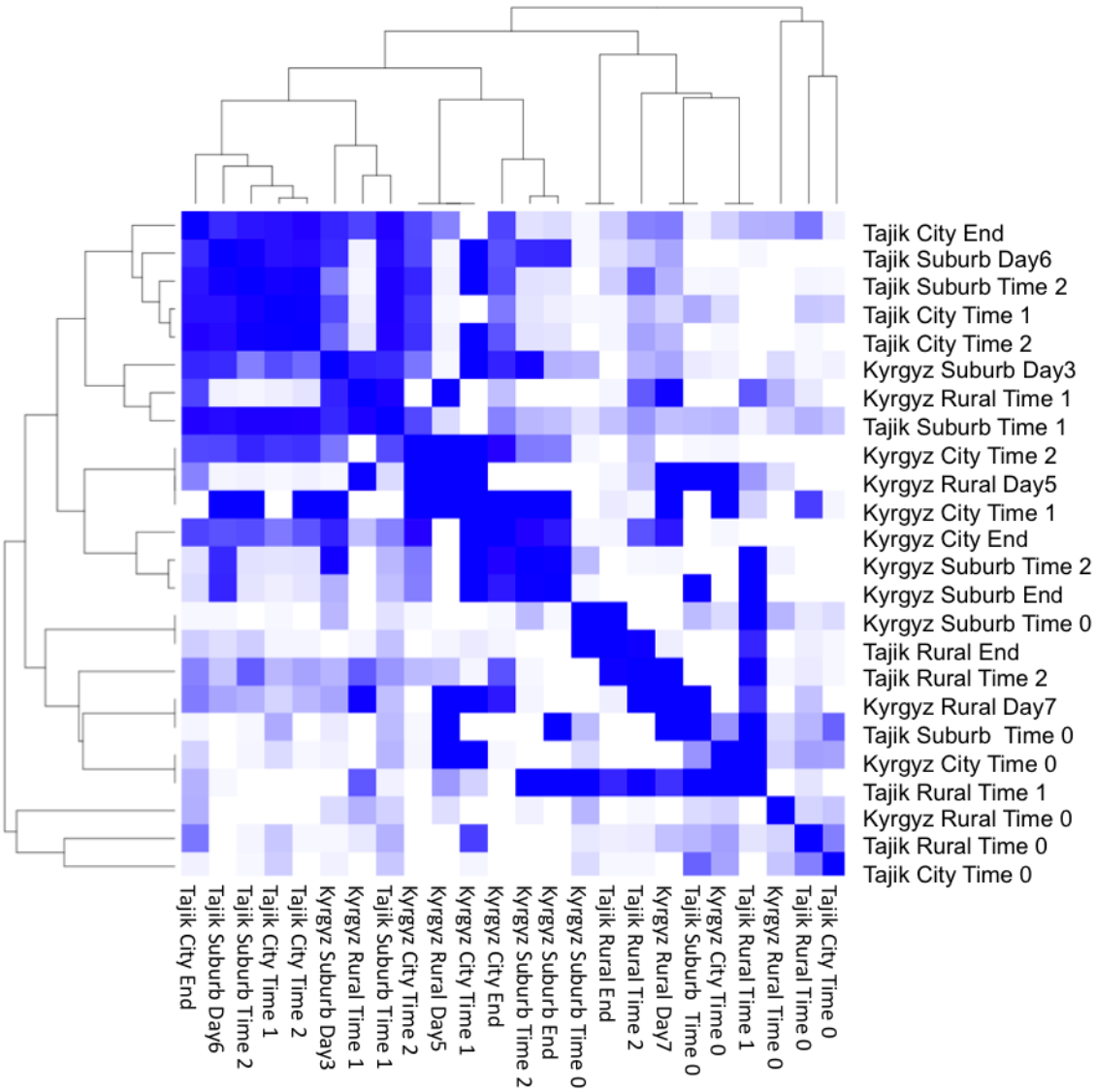


Figure 8. P450 cluster dendrograms and heatmap: The sites and enrichment conditions were compared to each other by cluster dendrograms on the top and left of the heatmap. The heatmap shows relatedness between sites and enrichment conditions,

showing the degree of similarity. The more similar two samples are, the darker the blue, and the least similar sites are associated with a white square in the corresponding portion of the matrix. Sample sites labeled by country and type of environment, and before (T0) or after (ET) enrichment with hexadecane.

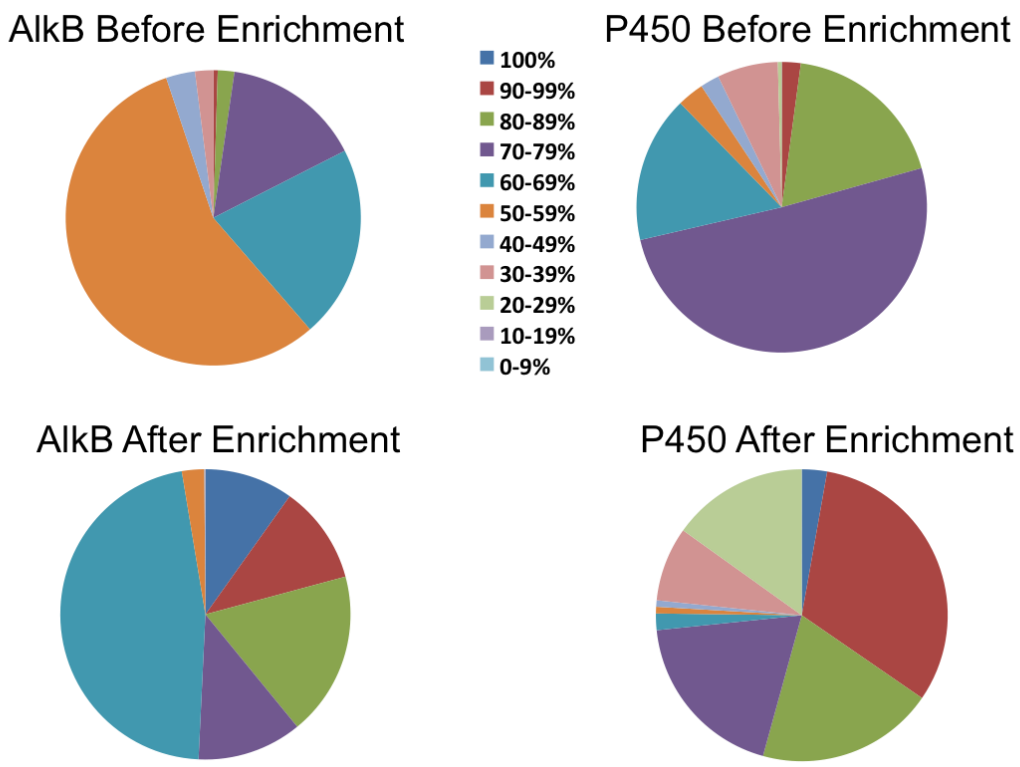


Figure 9. Percent similarity of experimental sequences to known sequences: Each chart shows the percent similarity of translated nucleotide sequences in relationship to known protein sequences from GenBank.

CHAPTER FOUR

Effect of biogeography and salinity on alkane monooxygenase populations recovered from Puerto Rican sediment

---

Elyse A. Rodgers-Vieira[1], Qiong Wang[2], Jim R. Cole[2], José R. Pérez-Jiménez[3], and Gerben J. Zylstra[1]

Department of Biochemistry and Microbiology, Rutgers, The State University of New Jersey[1], Center for Microbial Ecology and Department of Microbiology and Molecular Genetics, Michigan State University[2], Interdisciplinary Research Institute, Universidad del Turabo[3].

Abstract

Alkanes are common environmental pollutants as they represent a large component of crude oil. Alkane monooxygenases function in the first step of alkane degradation by activating the terminal carbon-carbon bonds resulting in an alcohol product. The aim of this study was to determine, through enrichment and 454 pyrosequencing the effect of hydrocarbon enrichment and salinity on alkane monooxygenase populations in a variety of aquatic sediments collected in Puerto Rico. Sediment was collected from several sites with varying historical pollution levels that ranged from the Bay and Port of San Juan in the north to a mangrove forest in an estuary preserve in the south. The sediment samples were enriched with hexadecane in minimal Artificial Saltwater Media at three different salt concentrations (33%, 66%, and 100% in relationship to the salt concentration in seawater). DNA was extracted from the enrichment cultures before and after enrichment. PCR was performed at both time points for 16S rRNA, as well as for genes encoding two different alkane monooxygenases (*alkB* and P450). Salinity was expected to have an effect on alkane monooxygenase populations observed following enrichment. Salinity did not have an affect on the alkane monooxygenase populations with hexadecane enrichment. Instead, the sample site type (mangrove or shore) was found to be the greatest determinant of 16S rRNA composition and alkane monooxygenase populations.

Introduction

Several studies have established a relationship between bacterial community composition and salinity along salinity gradients (4, 5, 11). Over short timescales the microbial communities may respond to changes in salinity with changes in gene expression, and not necessarily alteration of the microbial community (14). Bacteria are often sensitive to differences in salinity. For example, certain bacteria are only able to degrade hydrocarbons in marine environments and are halophilic while others prefer lower salinities (1).

The San Juan Harbor area receives heavy freight and industrial ship traffic and is also a popular cruise ship destination. The Jobos Bay Reserve is the second largest estuary in Puerto Rico spanning 48 km^2 within the municipalities of Guayama and Salinas in the southern arid region of the island (3, 31). Jobos Bay is a natural harbor sheltered by a series of mangrove islands and includes intertidal and subtidal zones (31). Mangrove forests are highly productive ecosystems found in tropical and subtropical regions. In mangroves, sulfate-reducing bacteria and methanogens thrive just below the oxic sediment layer (12, 19, 26).

A study conducted to investigate contaminants in Jobos Bay measured the mean surface salinity at 35.3 ± 0.1 ppt, with a temperature of 29.5 ± 0.1°C, and dissolved oxygen of 6.3 ± 0.2 ppm (22). This largely rural area contains two power generation plants, chemical and pharmaceutical facilities, and a petroleum refinery (31). The proportion of low and high molecular weight polycyclic aromatic hydrocarbons (PAHs) in the region was connected to pyrogenic origins (20, 22). Sediments from several locations were monitored within the Jobos Bay region for 130 organic chemical

pollutants. Among these, 58 PAHs, 31 organochlorine pesticides, 39 polychlorinated biphenyls, and 4 butyltins were identified. The Jobos watershed deposits surface water and groundwater into the bay from the Southern Coastal Plains of Puerto Rico, which may be largely responsible for the pollution observed. Impacts from chemical pollution in the area is likely, as all organic chemical contaminant levels measured were higher than sediment quality standards set by the National Oceanic and Atmospheric Association (22).

Gene-targeted metagenomics provides an approach for an in-depth investigation of functional genes in the environment. It requires that the gene family of interest have conserved regions for primer design that produce a product of appropriate length for pyrosequencing (16). *alkB* and P450 alkane monooxygenase genes, represent a diverse group of genes involved in alkane degradation in the environment. Alkanes are composed of carbon and hydrogen atoms and are chemically very stable. The alkane monooxygenase enzyme activates the terminal carbon in an alkane by insertion of an oxygen. This reaction results in the formation of an alcohol (15, 30).

In this study the relationship between salinity and alkane monooxygenase diversity was investigated by gene-targeted metagenomics. Gene-targeted metagenomics allows for in-depth analysis of the microbial community focusing on the gene of interest (16). Sediment was collected from several locations in Puerto Rico for enrichment. The locations studied included the Port of San Juan, an area of the shore near an oil-fueled power plant, and in the Jobos Bay National Estuarine Research Reserve. Alkane monooxygenase populations were not significantly affected by variations in salinity. Instead, the environment they originated from had a larger affect on the functional gene

populations. The presence of mangroves did result in different monooxygenase populations when compared to non-mangrove environments.

Materials and Methods

Sediment and water collection. Sediment was collected from six locations on the island of Puerto Rico. Water was also collected at each site for analysis. Two of the locations were in the Port of San Juan on the east and west side. In the southern portion of the island one location was along the shore near a power plant that is fueled by oil and another was taken from a mangrove-vegetated sediment in the estuary near the power plant. Another mangrove sample was taken farther into the estuary in protected area, while another was from a boat dock within the estuary. All locations are indicated on the map shown in Figure 1.

Growth conditions. 50 ml of Artificial Seawater (ASW) media at three salt concentrations (33%, 66%, and 100% in relationship to seawater) were inoculated with approximately 1.0 g sediment and amended with 50 μl of hexadecane (Sigma Aldrich, St. Louis, MO). Artificial Seawater Minimal Medium contained 23.6 g NaCl (at 100% salt concentration in relationship to seawater), 0.64 g KCl, 4.53 g $MgCl_2x6H_2O$, 5.94 g $MgSO_2x7H_2O$, 1.3 g $CaCl_2$ $x2H_2O$, 43.0 mg Na_2HPO_4 x $7H_2O$, 0.22 g $NaNO_3$, 0.65 g NH_4Cl, and 46.0 mg $NaHCO_3$ in a total volume of 1L. 1.0 ml each of Vitamin Solution for SRB, Trace Element Solution, and Vitamin-B12 Solution were added (6, 10). The cultures were incubated shaking for 5 days at 30°C.

Cell enumeration. The cultures were incubated until the cell density had reached approximately 2 x 10^{10} cells/ml. For the DAPI staining, 1 ml of the enrichment culture was fixed with 100 µl of 37% formaldehyde overnight. The sample was then vortexed and diluted 1:200 with sterile MSB medium. The dilution was then sonicated in an ice water bath for 5 minutes before staining. They were stained with 5 µg/ml of 4', 6'-diamidino-2-phenylindole (DAPI) in the dark for 1 hour (Sigma Aldrich, St. Louis, MO). The sample was filtered using a 0.2 µm black 22-mm diameter filter from Millipore. A Zeiss Axiovert 200 M epifluorescent microscope was used. Cell counts from 10 grid fields on each slide were tallied (21, 23).

Water testing. Water samples were collected at the same time as sediment at each of the sites. The water was stored at 4°C until testing. The testing was performed by the Soil Testing Laboratory at Rutgers University. The pH, salinity or electrical conductivity (EC), nitrogen, phosphorous, and iron were all determined.

DNA extraction and amplification for pyrosequencing. DNA was extracted using the PowerSoil kit at the beginning of enrichment and after 5 days (MoBio, Carlsbad, CA). DNA was quantified using a NanoDrop Spectrophotometer (Thermo Scientific, Wilmington, DE). The DNA was amplified using the FastStart High Fidelity PCR system (Roche, Indianapolis, IN). The composition of the PCR reactions were 1µM each of the forward and reverse primers, 0.2 mM dNTPs, and 150 ng/µl BSA in a total volume of 20 µl per reaction containing 1X reaction buffer and 1 unit of enzyme. Approximately 25ng of template DNA was added to each reaction. For *alkB* amplification, barcoded degenerate primers alkBF (AATACHGSVCAYGAGCTCRGYCAYAAR) and alkBR (GCRTGRTGATCAGARTGHCGYTG) were used (18). For *alkB* amplification the

following protocol was used: 95°C for 5 minutes, followed by 35 cycles of 95°C for 45 seconds, 55°C for 45 seconds, and 72°C for 1 minute, with a final elongation of 72°C for 4 minutes. The tagged primers used to amplify P450 alkane monooxgenase genes were P450F (GTSGGCGGCAACGACACSAC) and P450R (GCASCGGTGGATGCCGAAGCCRAA) (27). The PCR protocol used for P450 amplification was: 95°C for 5 minutes, followed by 35 cycles of 95°C for 30 seconds, 58°C for 30 seconds, and 72°C for 40 seconds, with a final elongation at 72°C for 4 minutes. The 16S rRNA primers used were 16S_577F (AYTGGGYDTAAAGNG) and 16S_926R (CCGTCAATTCMTTTRAGT) (17, 25). For 16S rRNA amplification the PCR protocol was: 95°C for 2 minutes, followed by 30 cycles of 95°C for 45 seconds, 55°C for 45 seconds, and 72°C for 1 minute, with a final elongation at 72°C for 4 minutes.

454 sequencing. Once DNA was amplified with fusion tagged primers, the amplified products were separated on a 1% agarose gel. All PCR reactions were performed in triplicate or higher and pooled during gel purification using the Qiagen Gel Extraction Kit and quantified using a Nanodrop Spectrophotometer (Qiagen, Valencia, CA and Thermo Scientific, Wilmington, DE). All samples were sequenced at the Michigan State University core facility using a Roche 454 FLX Titanium sequencer.

The raw sequencing reads were put through a quality filter in the Ribosomal Database Project Pipeline (RDP) (8, 9). 16S rRNA reads that passed the quality filter were used for further analysis in the RDP Pyrosequencing Pipeline. The 16S rRNA sequences were clustered at 80% identity for analysis and clusters were identified using the RDP Classifier (29).

Functional gene reads that were of sufficient quality were frame shift corrected with RDP's FrameBot tool using the Hidden Markov Model (HMM) trained specifically for each functional gene using known sequences from GenBank (7). The HMM model uses the known sequences to determine the most conserved regions and which variations within the sequence are the most likely to occur. Further analysis of the *alkB* and P450 sequences was done using the RDP Fungene Pipeline (8, 9). Sequences were clustered at 80% identity for analysis.

Results

Gene specific pyrosequencing was performed to determine the effect of salinity, enrichment, and location type on alkane monooxygenase diversity. The number of 16S rRNA sequences obtained at each sample site and that passed the initial processing criteria are shown in Table 1. The water collected at each site was analyzed for pH, salinity, nitrogen, phosphorous, and iron (Table 2). Concentrations of P, Fe, and inorganic N were extremely low and always below the analytical detection limit. The salinity for all samples varied between 59.30 and 64.50 dS/m. The pH varied between 7.22 and 8.22.

At the 16S rRNA phyla level, Proteobacteria comprised the majority of the bacterial population with and without enrichment (Figure 2). However, a decrease in diversity at the phyla level was seen with enrichment regardless of salinity. Many phyla, including Acidobacteria, decreased significantly with enrichment on hexadecane with varying salinity. Within the Proteobacteria phyla, Gammaproteobacteria dominated in 12

of the 18 samples (Figure 3). The Deltaproteobacteria were very prominent in both the unimpacted and impacted mangrove samples before enrichment. This population declined with enrichment at both 33% and 100% salt concentration in relationship to seawater with an increase in the Gammaproteobacteria. Deltaproteobacteria in the other four sediment samples (non-mangrove sites) decreased to less than 5% of the total upon enrichment. In all the sites there was a trend towards an increase in Gammaproteobacteria overall. Interestingly, the eastern side of San Juan Harbor had significantly more Alphaproteobacteria following enrichment while an increase in the Gammaproteobacteria occurred in the western San Juan Harbor enrichments.

AlkB pyrosequencing data was clustered at 80% identity and compared to known alkane hydroxylase protein sequences and the best match was found by sequence similarity (Figure 4). Several AlkB families only contained single sequences. The same analysis was done for the P450 data (Figure 5), where many families contained single sequences, 2-10 sequences, and 11-50 sequences. This indicates a larger number of families for P450 than AlkB.

The cluster dendrogram and heatmap for AlkB (Figure 6) compares four of the sample locations to each other and the populations resulting from enrichment on hexadecane with varying salinities. The two San Juan Harbor locations were compared to the impacted (near the power plant) and unimpacted mangrove locations in the Jobos Reserve. The harbor locations clustered together regardless of salinity. The same was seen with the mangrove locations. The same analysis was done for cytochrome P450 comparing the harbor sites to the mangrove locations, and boat dock location in the

southern portion of Puerto Rico (Figure 7). Clear clustering was seen by site and not by enrichment conditions.

Discussion

The objectives of this study were to utilize pyrosequencing technology to determine what effect, if any, salinity has on alkane monooxygenase populations obtained through sediment enrichment with the medium length alkane hexadecane, and to compare different geographical locations on the island of Puerto Rico to determine if different populations were seen in the eastern and western sides of the Port of San Juan.

Seawater has an average salinity of 55 dS/m. All of the sites measured were slightly more saline than average seawater (Table 2). The enrichment conditions in this study exposed the sediments to lower salt concentrations than observed at the time of sampling. The San Juan port locations would have less variation in salinity than the estuary sites in the southern portion of the island due to the natural influx of freshwater from waterways.

Decreases in the abundance of several 16S phyla were observed after enrichment. Non-enriched sediments contained a significant proportion of Chloroflexi, Planctomycetes, Cyanobacteria, Deinococcus, and Acidobacteria, all of which are common in aquatic environments. These groups decreased following enrichment with hexadecane and exposure to the three different salinities tested. Gammaproteobacteria increased in most cases with exposure to hexadecane in the enrichments. Selection for Gammaproteobacteria has been observed in oil-treated microcosms and has been linked

to alkane degradation (24). Large populations of Deltaproteobacteria were present in the sediment at the mangrove locations (power plant and unimpacted). The Deltaproteobacteria decreased in the enrichment cultures. This may be because mangroves are characterized by an active sulfur cycle in the sediments (19). Deltaprotebacteria contain a branch of sulfur and sulfate reducing bacteria that are strict anaerobes. When the selective pressure imposed by the mangrove habitat is removed and the sediment is put into an aerobic environment, this branch of bacteria would not be expected to thrive, which may explain the pattern seen in the mangrove sediments and enrichments. Although there are known sulfate reducers that also degrade n-alkanes, they are strictly anaerobic (2, 13).

Greater diversity was seen with P450 than AlkB (Figures 4 and 5). Cytochrome P450s are involved in many different types of reactions in addition to alkane utilization. While the P450 primers were designed from conserved regions of P450 alkane hydroxylase sequences, they are degenerate and may pick up some other related P450s due to the diverse functional nature of P450 enzymes (28). The primers may simply capture a larger group of P450's than the *alkB* primers do for that type of alkane hydroxylase.

The dendrograms and heatmap depicting the relationship between different sites and enrichment treatments for AlkB show a clear divide between the mangrove sites and the San Juan harbor locations (Figure 6). Sediment enrichments on the east and west sides of the harbor did not fall into separate clusters despite differences in proximity to the open ocean and industrial centers along the shore. The mangrove enrichments however, fell into distinct groups based on location.

The P450 dendrograms show a similar trend with the mangrove locations showing the most similarity to each other, and clustering separately from the harbor enrichments (Figure 7). The shore location adjacent to the oil-fueled power plant was more similar to the harbor samples than other locations. The boat dock located within the estuary preserve was also similar to the harbor samples.

The most distinct divisions remain between the mangrove and non-mangrove locations for P450 and AlkB monooxygnease populations. Since previous studies showed a variation on microbial populations with exposure to salinity gradients, the same was expected in these enrichments. Unexpectedly, salinity does not have much of an effect on the alkane monooxygenase populations obtained with enrichment. The divide was most pronounced between mangrove and non-mangrove locations, which suggests that *alkB* and P450 gene populations may be most affected by the presence of higher levels of sulfur and sulfate in sediments, or some other element unique to the mangrove habitat. Additionally, populations on either side of the port in San Juan varied slightly, but not enough to make a clear distinction between the populations.

References

1. **Abed, R. M., A. Al-Thukair, and D. de Beer**. 2006. Bacterial diversity of a cyanobacterial mat degrading petroleum compounds at elevated salinities and temperatures. FEMS microbiology ecology **57**:290-301.
2. **Aeckersberg, F., F. A. Rainey, and F. Widdel**. 1998. Growth, natural relationships, cellular fatty acids and metabolic adaptation of sulfate-reducing bacteria that utilize long-chain alkanes under anoxic conditions. Archives of microbiology **170**:361-369.
3. **Aldarondo-Torres, J. X., F. Samara, I. Mansilla-Rivera, D. S. Aga, and C. J. Rodriguez-Sierra**. 2010. Trace metals, PAHs, and PCBs in sediments from the Jobos Bay area in Puerto Rico. Marine pollution bulletin **60**:1350-1358.
4. **Bernhard, A. E., D. Colbert, J. McManus, and K. G. Field**. 2005. Microbial community dynamics based on 16S rRNA gene profiles in a Pacific Northwest estuary and its tributaries. FEMS microbiology ecology **52**:115-128.
5. **Bolhuis, H., and L. J. Stal**. 2011. Analysis of bacterial and archaeal diversity in coastal microbial mats using massive parallel 16S rRNA gene tag sequencing. The ISME journal **5**:1701-1712.
6. **Bruns, A., and L. Berthe-Corti**. 1999. Fundibacter jadensis gen. nov., sp. nov., a new slightly halophilic bacterium, isolated from intertidal sediment. International journal of systematic bacteriology **49 Pt 2**:441-448.
7. **Chai, B., Q. Wang, J. Fish, D. M. McGarrel, J. M. Tiedje, and C. J. R** (2010). The ribosomal database project: improved tools and services for rRNA analysis 100th General meeting of the American Society for Microbiology. San Diego, California.
8. **Cole, J. R., B. Chai, R. J. Farris, Q. Wang, A. S. Kulam-Syed-Mohideen, D. M. McGarrell, A. M. Bandela, E. Cardenas, G. M. Garrity, and J. M. Tiedje**. 2007. The ribosomal database project (RDP-II): introducing myRDP space and quality controlled public data. Nucleic acids research **35**:D169-172.
9. **Cole, J. R., Q. Wang, E. Cardenas, J. Fish, B. Chai, R. J. Farris, A. S. Kulam-Syed-Mohideen, D. M. McGarrell, T. Marsh, G. M. Garrity, and J. M. Tiedje**. 2009. The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. Nucleic acids research **37**:D141-145.
10. **Crespo-Medina, M., A. Chatziefthimiou, R. Cruz-Matos, I. Perez-Rodriguez, T. Barkay, R. A. Lutz, V. Starovoytov, and C. Vetriani**. 2009. Salinisphaera hydrothermalis sp. nov., a mesophilic, halotolerant, facultatively autotrophic, thiosulfate-oxidizing gammaproteobacterium from deep-sea hydrothermal vents, and emended description of the genus Salinisphaera. International journal of systematic and evolutionary microbiology **59**:1497-1503.
11. **Crump, B. C., C. S. Hopkinson, M. L. Sogin, and J. E. Hobbie**. 2004. Microbial biogeography along an estuarine salinity gradient: combined influences of bacterial growth and residence time. Applied and environmental microbiology **70**:1494-1505.
12. **Dar, S. A., R. Kleerebezem, A. J. Stams, J. G. Kuenen, and G. Muyzer**. 2008. Competition and coexistence of sulfate-reducing bacteria, acetogens and

methanogens in a lab-scale anaerobic bioreactor as affected by changing substrate to sulfate ratio. Applied microbiology and biotechnology **78**:1045-1055.

13. **Davidova, I. A., K. E. Duncan, O. K. Choi, and J. M. Suflita**. 2006. Desulfoglaeba alkanexedens gen. nov., sp. nov., an n-alkane-degrading, sulfate-reducing bacterium. International journal of systematic and evolutionary microbiology **56**:2737-2742.
14. **Edmonds, J. W., N. B. Weston, S. B. Joye, X. Mou, and M. A. Moran**. 2009. Microbial community response to seawater amendment in low-salinity tidal sediments. Microbial ecology **58**:558-568.
15. **Funhoff, E. G., U. Bauer, I. García-rubio, B. Witholt, and J. B. V. Beilen**. 2006. CYP153A6, a soluble P450 oxygenase catalyzing terminal-alkane hydroxylation. J. Bacteriol **188**:5220-5227.
16. **Iwai, S., B. Chai, W. J. Sul, J. R. Cole, S. A. Hashsham, and J. M. Tiedje**. 2010. Gene-targeted-metagenomics reveals extensive diversity of aromatic dioxygenase genes in the environment. The ISME journal **4**:279-285.
17. **Jesus, E. d. C., E. Susilawati, S. L. Smith, Q. Wang, B. Chai, R. Farris, J. L. M. Rodrigues, K. D. Thelen, and J. M. Tiedje**. 2010. Bacterial communities in the rhizosphere of biofuel crops grown on marginal lands as evaluated by 16S rRNA gene pyrosequences. Bioenergy research **3**:20-27.
18. **Kloos, K., J. C. Munch, and M. Schloter**. 2006. A new method for the detection of alkane-monooxygenase homologous genes (alkB) in soils based on PCR-hybridization. Growth (Lakeland) **66**:486 - 496.
19. **Lyimo, T. J., A. Pol, and H. J. Op den Camp**. 2002. Sulfate reduction and methanogenesis in sediments of Mtoni mangrove forest, Tanzania. Ambio **31**:614-616.
20. **Neff, J. M., S. A. Stout, and D. G. Gunster**. 2005. Ecological risk assessment of polycyclic aromatic hydrocarbons in sediments: Identifying sources and ecological hazard. Integrated environmental assessment and management **1**:23-33.
21. **Ni Chadhain, S. M., R. S. Norman, K. V. Pesce, J. J. Kukor, and G. J. Zylstra**. 2006. Microbial dioxygenase gene population shifts during polycyclic aromatic hydrocarbon biodegradation. Applied and environmental microbiology **72**:4078-4087.
22. **Pait, A. S., D. R. Whitall, A. Dieppa, S. E. Newton, L. Brune, C. Caldow, A. L. Mason, D. A. Apeti, and J. D. Christensen**. 2011. Characterization of organic chemical contaminants in sediments from Jobos Bay, Puerto Rico. Environmental monitoring and assessment.
23. **Porter, K. G., and Y. S. Feig**. 1980. The use of DAPI for identifying and counting aquatic microflora. Limnology and Oceanography **25**:943-948.
24. **Röling, W. F. M., M. G. Milner, D. M. Jones, K. Lee, F. Daniel, R. J. P. Swannell, and I. Head**. 2002. Robust hydrocarbon degradation and dynamics of bacterial communities during nutrient-enhanced oil spill bioremediation. Applied and environmental microbiology **68**:5537-5548.
25. **Sul, W. J., J. R. Cole, C. Jesus Eda, Q. Wang, R. J. Farris, J. A. Fish, and J. M. Tiedje**. 2011. Bacterial community comparisons by taxonomy-supervised analysis independent of sequence alignment and clustering. Proceedings of the National Academy of Sciences of the United States of America **108**:14637-14642.

26. **Taketani, R. G., C. A. Yoshiura, A. C. Dias, F. D. Andreote, and S. M. Tsai**. 2010. Diversity and identification of methanogenic archaea and sulphate-reducing bacteria in sediments from a pristine tropical mangrove. Antonie van Leeuwenhoek **97**:401-411.
27. **van Beilen, J. B., E. G. Funhoff, A. V. Loon, A. Just, L. Kaysser, M. Bouza, M. Ro, Z. Li, and B. Witholt**. 2006. Cytochrome P450 alkane hydroxylases of the CYP153 family are common in alkane-degrading eubacteria lacking integral membrane alkane hydroxylases. Applied and environmental microbiology **72**:59-65.
28. **van Beilen, J. B., E. G. Funhoff, A. van Loon, A. Just, L. Kaysser, M. Bouza, R. Holtackers, M. Rothlisberger, Z. Li, and B. Witholt**. 2006. Cytochrome P450 alkane hydroxylases of the CYP153 family are common in alkane-degrading eubacteria lacking integral membrane alkane hydroxylases. Applied and environmental microbiology **72**:59-65.
29. **Wang, Q., G. M. Garrity, J. M. Tiedje, and J. R. Cole**. 2007. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. Applied and environmental microbiology **73**:5261-5267.
30. **Whyte, L. G., T. H. Smits, D. Labbe, B. Witholt, C. W. Greer, and J. B. van Beilen**. 2002. Gene cloning and characterization of multiple alkane hydroxylase systems in Rhodococcus strains Q15 and NRRL B-16531. Applied and environmental microbiology **68**:5933-5942.
31. **Zitello, A. G., D. R. Whitall, A. Dieppa, J. D. Christensen, M. E. Monaco, and S. O. Rohmann** (2008). Characterizing Jobos Bay, Puerto Rico: a watershed Modeling analysis and monitoring plan. NOAA Technical Memorandum NOS NCCOS. **76:**81.

Figure 1. Map of sampling locations: Site 1, Shore at Power Plant; Site 2, Boat Dock; Site 3, Mangrove Unimpacted; Site 4, Mangrove Impacted; Site 5, San Juan Harbor East; Site 6, San Juan Harbor West.

Sample Site	Match Barcode	Passed Initial Process	% Passed Initial Process	Passed Primer Filter	% Passed Primer Filter
Shore at Power Plant	16,699	12,780	76.5	3,811	22.8
Boat Dock	41,529	33,518	80.7	7,801	18.8
Mangrove Unimpacted	61,412	54,796	89.2	11,929	19.4
Mangrove Impacted	64,000	48,437	75.7	14,837	23.2
San Juan Harbor East	61,595	50,750	82.4	10,568	17.2
San Juan Harbor West	68,417	54,157	79.2	13,871	20.3

Table 1. Number of sequences obtained from pyrosequencing: The number of 16S rRNA sequences that matched the pyrosequencing primer barcode, and the number that passed the initial quality filters are shown in columns two and three. The percentage of reads that passed the quality filters are shown in column six.

	pH	EC	NO_3-N	P	Fe
		dS/m	mg/L	mg/L	mg/L
Shore at Power Plant	7.7	59.3	-	< 0.1	< 0.1
Boat Dock	8.2	63.1	-	< 0.1	< 0.1
Mangrove Unimpacted	8.14	60.93	-	< 0.1	< 0.1
Mangrove Impacted	7.22	64.5	-	< 0.1	< 0.1
San Juan Harbor East	8.16	61.53	-	< 0.1	< 0.1
San Juan Harbor West	8.22	61.56	-	< 0.1	< 0.1

Table 2. Water sample measurements: Water samples from each site were analyzed for pH, electrical conductivity (EC), inorganic nitrogen, phosphorous, and iron.

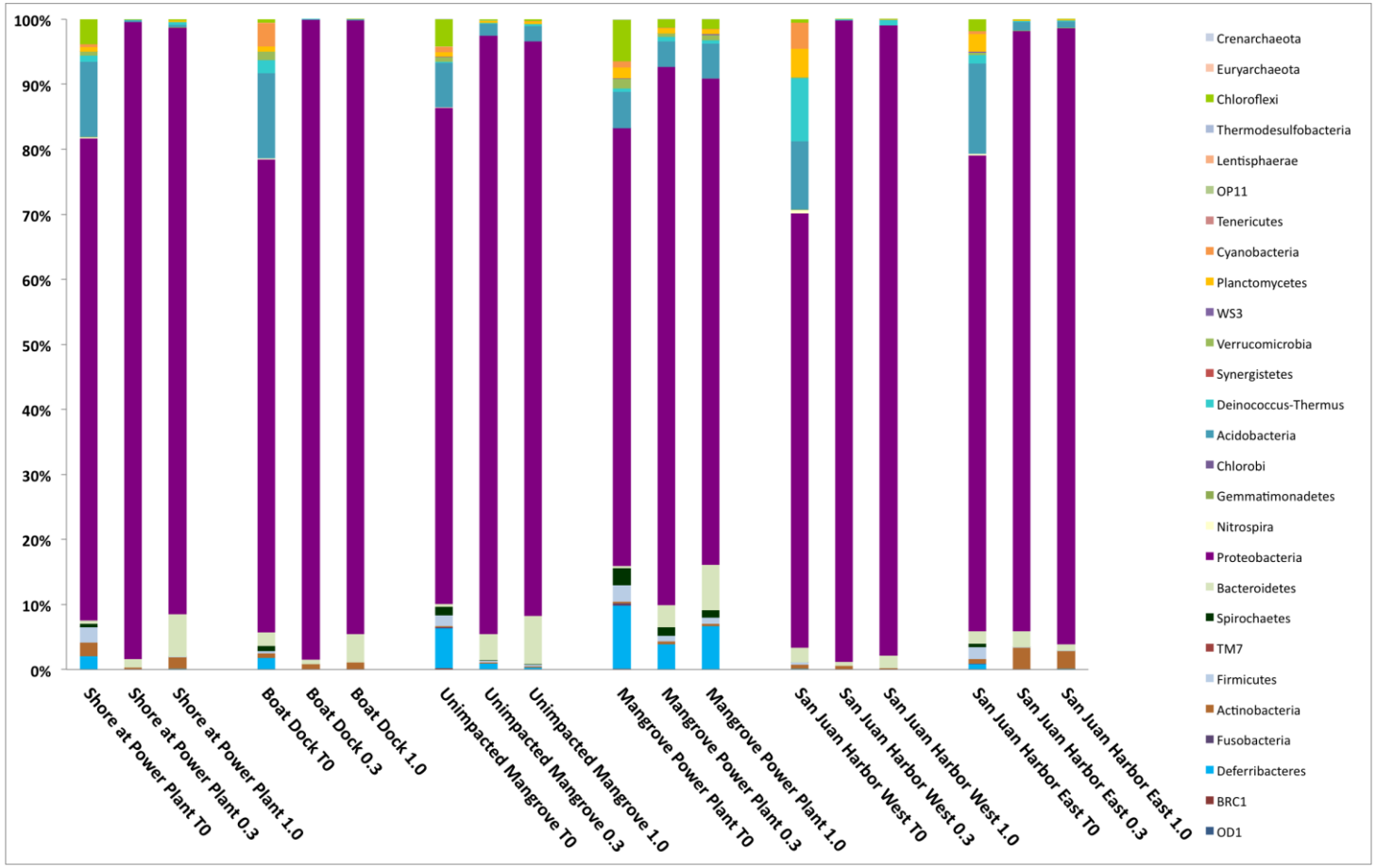


Figure 2. 16S rRNA phyla: The x-axis indicates the site followed by enrichment conditions (T0, before enrichment; 0.3, 33% of the salt concentration in seawater; 1.0, 100% of the salt concentration in seawater). The y-axis indicates the percent of the total sequences at the phyla level.

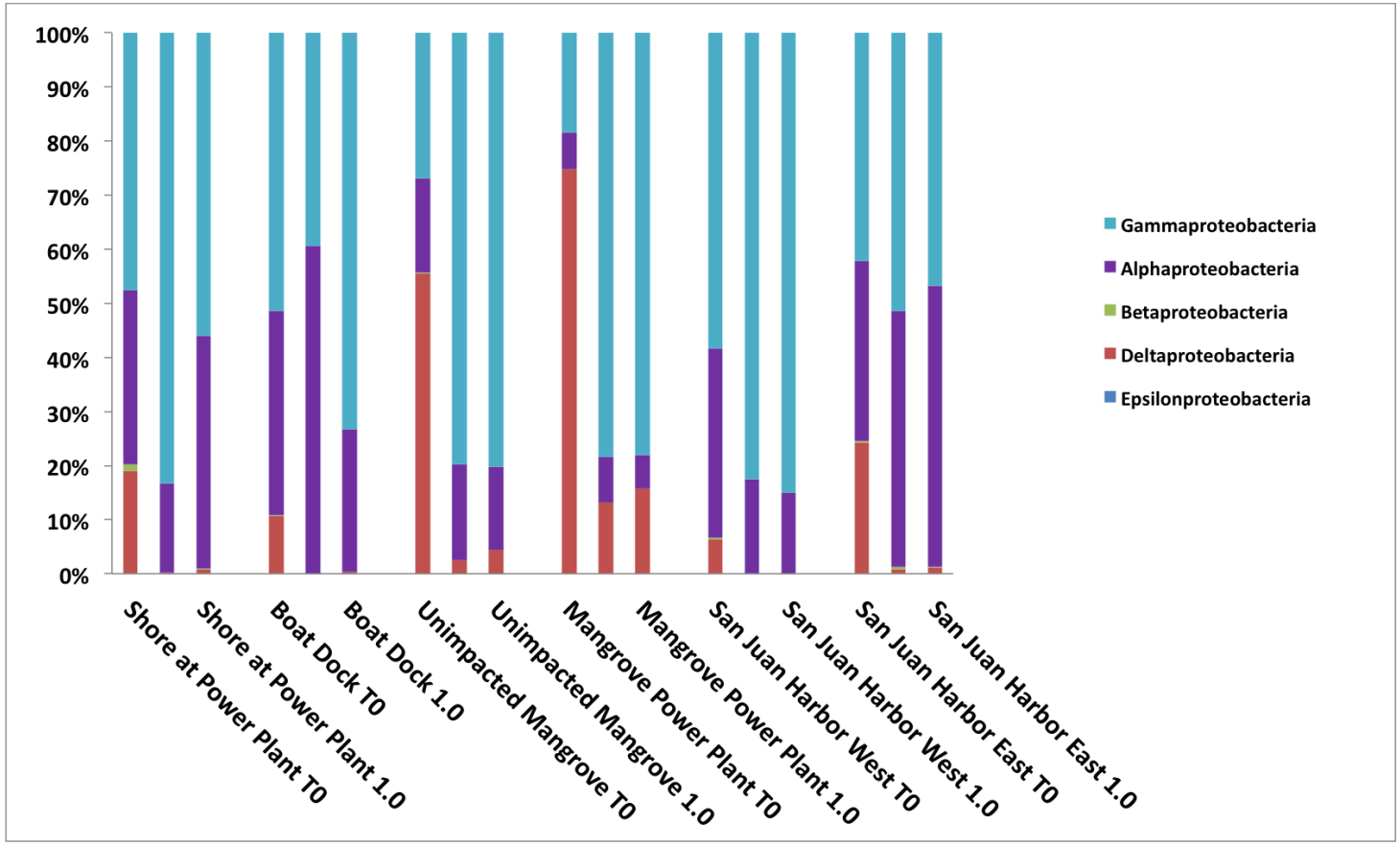


Figure 3. 16S rRNA Proteobacterial classes: The x-axis indicates the site followed by enrichment conditions (T0, before enrichment; 0.3, 33% of the salt concentration in seawater; 1.0, 100% of the salt concentration in seawater). The y-axis indicates the percent of the total Proteobacterial sequences at the class level.

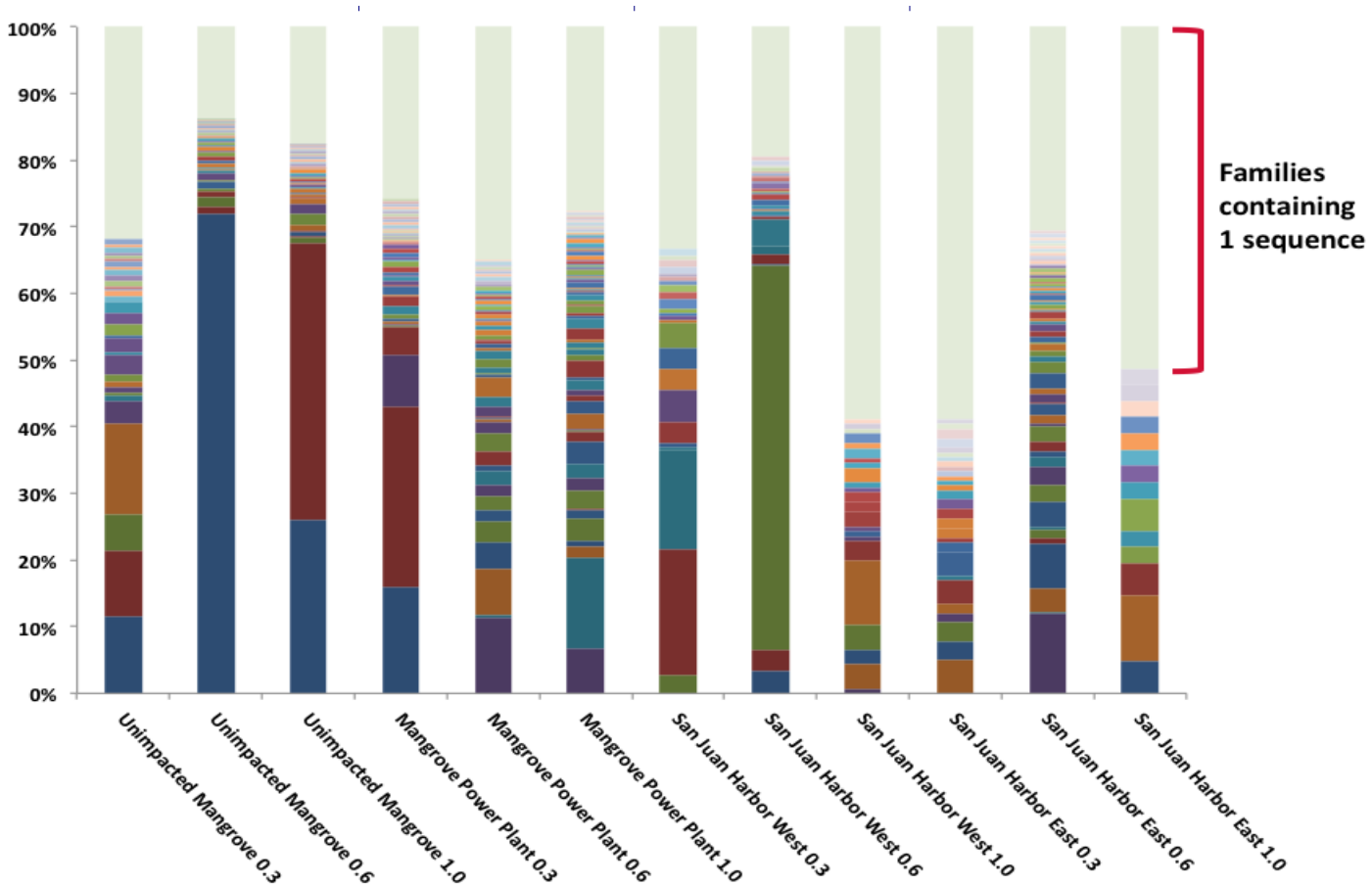


Figure 4. AlkB diversity: The x-axis indicates the site followed by enrichment conditions (T0, before enrichment; 0.3, 33% of the salt concentration in seawater; 0.6, 66% of the salt concentration in seawater, 1.0, 100% of the salt concentration in seawater). The y-

axis indicates the percent of the total AlkB sequences. Each AlkB sequence was compared to a set of known AlkB protein sequences, and the best match is reported as a percent of the total AlkB sequences. Families containing only one sequence are indicated with a bracket.

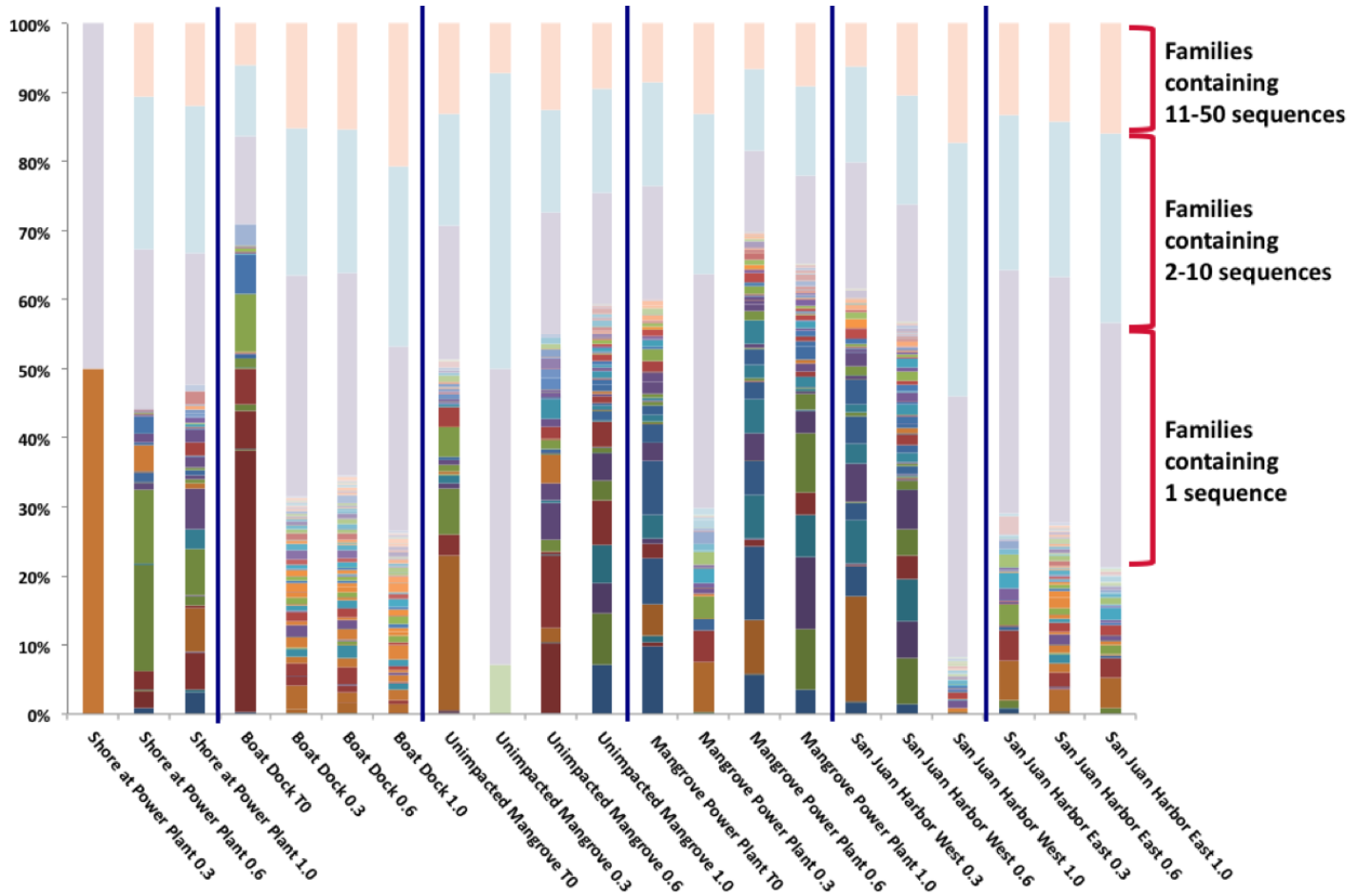


Figure 5. P450 diversity: The x-axis indicates the site followed by enrichment conditions (T0, before enrichment; 0.3, 33% of the salt concentration in seawater; 0.6, 66% of the salt concentration in seawater, 1.0, 100% of the salt concentration in seawater). The y-

axis indicates the percent of the total P450 sequences. Each P450 sequence was compared to a set of known P450 protein sequences, and the best match is reported as a percent of the total P450 sequences. Families containing only one sequence, 2-10, and 11-50 sequences are indicated with brackets.

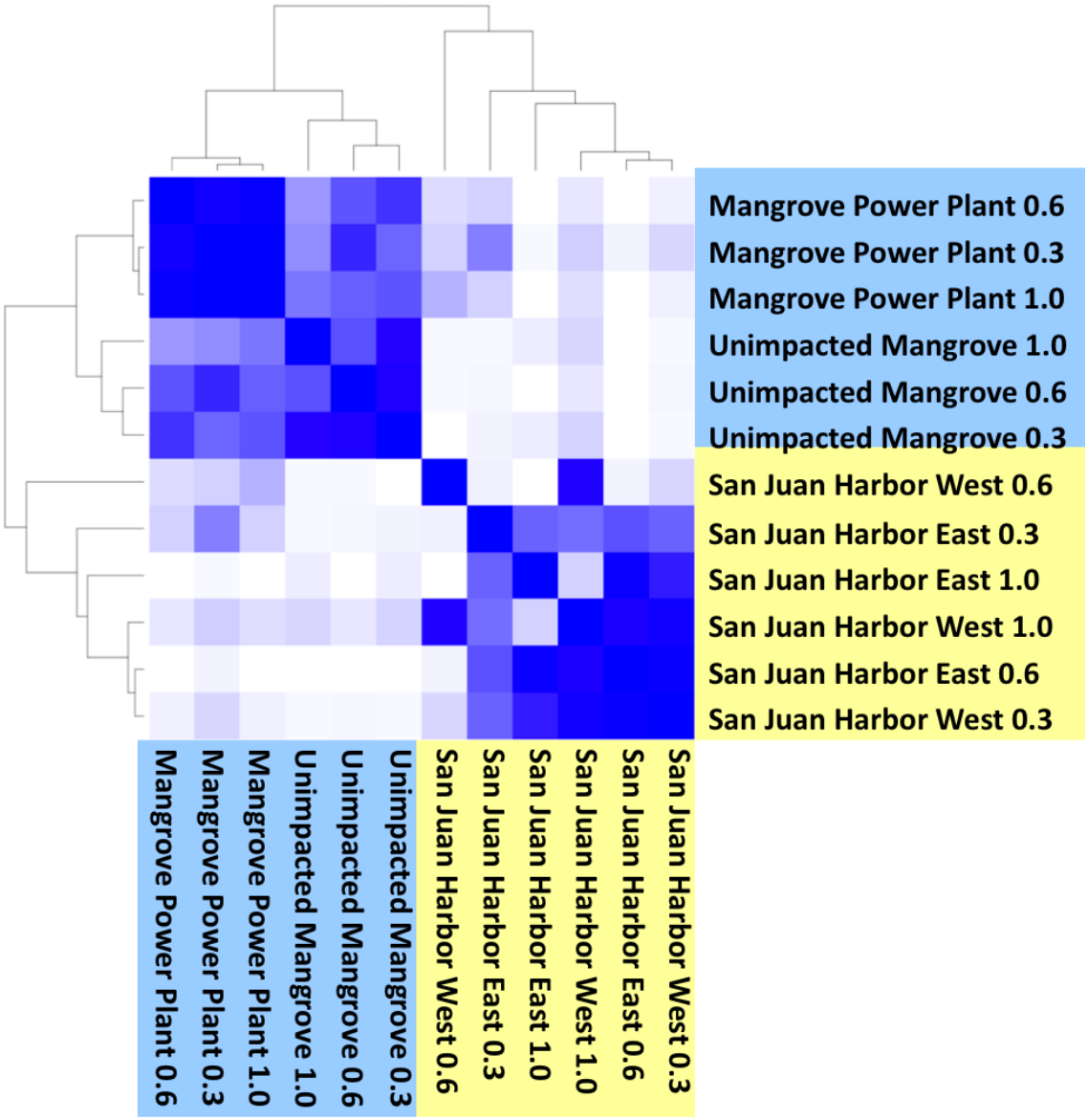


Figure 6. AlkB cluster dendrogram and heatmap: The diagram compares the degree that samples are related to each other. Each sample is labeled by site followed by enrichment condition (T0, before enrichment; 0.3, 33% of the salt concentration in seawater; 0.6, 66% of the salt concentration in seawater, 1.0, 100% of the salt concentration in seawater).

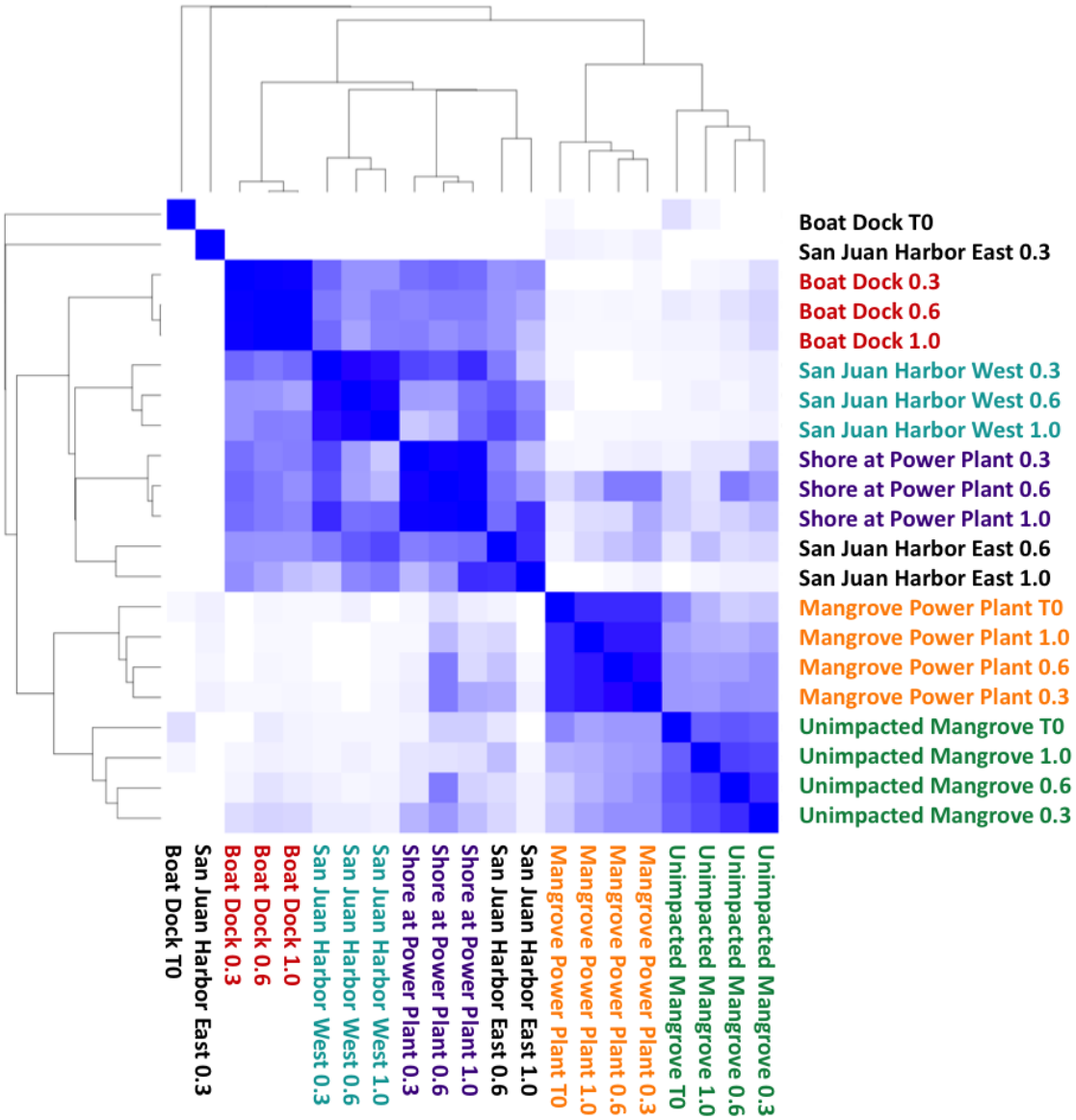


Figure 7. P450 cluster dendrogram and heatmap: The diagram compares the degree that samples are related to each other. Each sample is labeled by site followed by enrichment condition (T0, before enrichment; 0.3, 33% of the salt concentration in seawater; 0.6, 66% of the salt concentration in seawater, 1.0, 100% of the salt concentration in seawater).

CHAPTER FIVE

CONCLUSIONS AND SIGNIFICANCE

Alkane hydroxylases are integral in the degradation of alkanes in the environment. The two classes of alkane hydroxylase genes *alkB* and P450 code for a diverse group of enzymes that activate the terminal carbon in *n*-alkanes. At this time there are no known studies that have looked at the diversity of these genes in the environment and in enrichment cultures in as much detail as those presented in this dissertation.

The advent of pyrosequencing and widespread use of the technology allowed an opportunity to explore the diversity of these genes in the environment to further elucidate their ubiquitous nature in various geographical locations and aquatic habitats. Each of the studies that comprise this dissertation addressed a different scientific question while also collectively answering the question: What is the extent of alkane monooxygenase diversity and distribution in the environment?

The first study investigated alkane monooxygenases in Passaic River sediment. The overarching aim of this study was to determine if sampling location along a river in a relatively short distance (1000 meters) showed a difference in alkane monooxygenase community composition in the sediment and following enrichment. This is important to address since many studies pick one sampling location within a larger habitat. As

expected the populations collected nearest to each other in the river at 0 and 100 meters were most similar, and did differ from samples at 1000 meters. Sampling location choice within an environment will influence the population observed. Therefore, to obtain a general idea of the composition of a population in an environment one sample is sufficient. However, to obtain a greater depth of knowledge multiple locations within an environment should be analyzed.

The Central Asia project addressed the issue of alkane monooxygenase diversity and biogeography by comparing communities recovered from sediments from distinct regions within two different countries. Pyrosequencing of the alkane monooxygenase genes revealed distinct populations both within and between countries. The functional gene populations were very divergent between Kyrgyzstan and Tajikistan. So, even while there are some cosmopolitan families of AlkB and P450 proteins, different families are found in disparate locations. By no means can one sample in a singular portion of the world and know the full range of genetic diversity of a protein of interest. Environmental pressures do shape microbial diversity so novel genes can potentially be revealed by investigating untapped geographic regions.

In Puerto Rican sediment, the affect of salinity on the alkane monooxygenase populations was investigated. It was assumed that enrichments at three different salt concentrations would select for a population of microorganisms that are better suited for a certain salinity range. Sediments were collected from mangrove and shore areas of Puerto Rico with similar salt concentrations. It was determined that salt concentration was not the driving factor in the alkane monooxygenase populations observed following enrichment. Instead, initial habitat type (mangrove or non-mangrove) was the main

determinant of alkane monooxygenase community composition. This may be in part because the enrichments only exposed the bacteria present in sediment to a different salinity for a short time. That may not be long enough to confer a greater shift in the population. If the enrichments were done over a much longer time or were subcultured several times a greater difference in the populations may have occurred. Also, the enrichments were done at salt concentrations equal to or less than that of seawater. The lower salt concentrations did not give any one organism an advantage, explaining why the enrichments were so similar to each other.

Pyrosequencing allowed for a thorough investigation of alkane monooxygenase communities and uncovered novel groups of both P450 and AlkB monooxygenases in these studies. As bioinformatic tools develop further, this type of work may even prove to yield more insight into microbial community composition and enzymes involved in degradative and other microbial processes in the environment.

CURRICULUM VITAE

Elyse Anne Rodgers-Vieira

2002-2004	Summer Intern, Science and Engineering Technical Scholar, Lawrence Livermore National Laboratory, CA
2003-2004	Undergraduate Researcher, The Salk Institute, CA
2004-2005	Undergraduate Researcher, Scripps Institution of Oceanography, University of California San Diego, CA
2005	B. S. University of California San Diego, CA
2005-2006	Laboratory Technician, University of California San Diego, CA
2006-2012	Graduate Assistant, Rutgers, the State University of New Jersey, NJ
2008-2009	Teaching Assistant, Rutgers, the State University of New Jersey, NJ
2012	Ph.D. Microbiology and Molecular Genetics, Rutgers, the State University of New Jersey, NJ