UNIVERSITY OF CALIFORNIA
RIVERSIDE

Application of Metagenomics for Identification of Novel Petroleum Hydrocarbon Degrading Enzymes in Natural Asphalts from the Rancho La Brea Tar Pits

A Dissertation submitted in partial satisfaction
of the requirements for the degree of

Doctor of Philosophy

in

Environmental Toxicology

by

Jean-Paul Mendoza Baquiran

December 2010

Dissertation Committee:
Dr. David E. Crowley, Chairperson
Dr. James Borneman
Dr. Sharon Walker

The Dissertation of Jean-Paul Mendoza Baquiran is approved:

Chairperson

University of California, Riverside

“Think for yourself, question authority”

Timothy Leary

ACKNOWLEDGEMENTS

I would first like to thank David Crowley, who has not only been a great advisor and mentor, but a great friend as well. He taught me more than just how to become a scientist, but also life values in which I have greatly benefited from. He always gave me the freedom and support to do just about anything I wanted to, inside and outside of the laboratory. Not only is he a great scientist, but a great person that I have had the pleasure to work with.

I would also like to thank Sharon Walker and James Borneman for their support and advice throughout my time in graduate school. They both have been very helpful and accessible whenever I needed them most. I truly believe that I had the best dissertation committee a graduate student could ever have. In addition, I would like to thank everyone in the Environmental Toxicology program and Environmental Sciences department, whom have guided and helped me throughout this chapter in my life. You all will be missed.

And of course, I would like to thank my family and friends whom brought me inspiration in everything I do. Every single one of you has played an important role in my development as a scientist and person. I have been blessed to have a beautiful family, whom has supported me through everything I have done. I am also lucky to have a wild and crazy group of friends that have made my life at UCR a great one. I love you all and I'm bringing the Jameson. Peace!

ABSTRACT OF THE DISSERTATION

Application of Metagenomics for Identification of Novel Petroleum Hydrocarbon Degrading Enzymes in Natural Asphalts from the Rancho La Brea Tar Pits.

by

Jean-Paul Mendoza Baquiran

Doctor of Philosophy, Graduate Program in Environmental Toxicology
University of California, Riverside, December 2010
Dr. David E. Crowley, Chairperson

Recent studies on the biodiversity of asphalt deposits at the Rancho La Brea Tar Pits in Los Angeles, California have revealed the existence of several hundred new species of bacteria and gene sequences encoding putative novel degradative enzymes. The presence of fossilized extinct animal remains in the La Brea Tar Pits has led to estimations that these natural asphalt seeps have existed for at least 40,000 year. These deposits consist of petroleum that has been degraded to the extent that the remaining material is comprised mainly of asphalts and heavy oils, which have saturated into the soil matrix. Since petroleum hydrocarbons are both a target and a product of microbial metabolism, the role of microorganisms that inhabit this and other similar environments is directly relevant to development of technology for bioremediation, biotransformation of petroleum hydrocarbons, and microbial enhanced oil recovery for extracting and refining heavy oil.

In this research, both culture-dependent and culture-independent methods were used to characterize microorganisms and consortia from the La Brea Tar Pits, which are able to degrade a variety of polycyclic aromatic hydrocarbons (PAHs) and BTEX (benzene, toluene, ethyl

benzene and xylene). Initial studies applied PCR-DGGE methods to identify the microbial consortia that degrade selected petroleum hydrocarbons during enrichment cultures on pure compounds or mixtures of substances found in petroleum. This study revealed a single *Pseudomonas sp.* that may be able to degrade multiple PAHs and biphenyl. In addition, PCR based techniques identified a naphthalene dioxygenase from *Pseudomonas stutzeri* that appears to function as a major degradative enzyme in this system. DNA microarray analysis further revealed the presence of a wide variety of petroleum hydrocarbon degrading genes and has verified that naphthalene dioxygenase is the most commonly found degradative enzyme in this microbial community.

To further investigate the true microbial diversity of the La Brea Tar Pits, the microbial community associated with the heavy oil at Rancho La Brea was studied using a metagenomics approach. A fosmid clone library was constructed using 38 Kb fragments of DNA extracted from the asphalt-soil samples of Pit 101. This library of about 3,000 clones was then screened for DNA inserts, which contained specific genes that could be targeted using PCR based methods. One selected clone contained the 16S rRNA gene of an unclassified Rhodospirallacaea and a putative 2-nitropropane dioxygenase, which suggest that this new organism degrades persistent nitroalkanes that are common in asphalt.

The current investigation of the La Brea Tar pits concluded using Illumina technologies (Illumina, Inc.) to deep sequence the metagenomic library of over 3,000 clones using a high throughput sequencer. Over 75 MB of DNA has been sequenced, from which over 650 contigs with an average length of 500 bp were assembled. Bioinformatics analysis indicated the presence of genes encoding three types of dioxygenases, one of which encoded a naphthalene

dioxygenase from *Pseudomonas sp.* along with two other genes that are most similar to previously reported genes encoding biphenyl and toluene dioxygenases. However, sequence analysis revealed that these genes were not significantly similar to these known dioxygenases and may thus be novel. The results of this research provide a foundation for further studies on the evolution and assembly of metabolic pathways in bacteria that have undergone long term adaptation to survival in natural asphalts.

TABLE OF CONTENTS

CHAPTER ONE: INTRODUCTION 1

BIBLIOGRAPHY 30

CHAPTER TWO: CHARACTERIZATION OF MICROORGANISMS AND ENZYMES ABLE TO DEGRADE PETROLEUM HYDROCARBONS FROM THE RANCHO LA BREA TAR PITS

ABSTRACT 39

INTRODUCTION 41

MATERIALS AND METHODS 46

RESULTS 50

DISCUSSION 56

BIBLIOGRAPHY 67

CHAPTER THREE: LINKING PHYLOGENY AND FUNCTION OF A PREVIOUSLY UNKNOWN ORGANISM AND A NOVEL 2-NITROPROPANE DIOXYGENASE USING METAGENOMICS

ABSTRACT 71

INTRODUCTION 73

MATERIALS AND METHODS 77

RESULTS 80

DISCUSSION 92

BIBLIOGRAPHY 95

CHAPTER FOUR: APPLYING ILLUMINA SEQUENCING COUPLED WITH A METAGENOMIC CLONE LIBRARY TO IDENTIFY NOVEL PETROLEUM DEGRADING GENES

ABSTRACT 98

INTRODUCTION 100

MATERIALS AND METHODS 103

RESULTS 105

DISCUSSION 121

BIBLIOGRAPHY 125

CHAPTER FIVE: GENERAL CONCLUSIONS 128

LIST OF FIGURES

CHAPTER ONE:

1.1 Phylogenetic tree of halophillic Archaea found in the La Brea Tar Pits 7

1.2 Phylogenetic tree of *Pseudomonas stutzeri* found in the La Brea Tar Pits 9

1.3 Phylogenetic tree of dioxygenase sequences found in the La Brea Tar Pits 11

1.4 Initial oxidation of Naphthalene to cis-1,2-Dihydroxy-1,2-Dihydronaphthalene by naphthalene dioxygenase 21

1.5 Flowchart illustrating metagenomics techniques and approaches 24

CHAPTER TWO:

2.1 DGGE community analysis of La Brea Tar Pit microbial communities Exposed to different petroleum hydrocarbons 52

2.2 Phylogenetic tree of DGGE band sequences from Pit 101 bacteria exposed to different PAHs 59

2.3 Phylogenetic tree of naphthalene dioxygenase genes from Pit 101 61

2.4 Phylogenetic tree of catechol 2,3-dioxygenase genes from Pit 101 62

2.5 Phylogenetic tree of PAH degrading bacteria isolated from enrichment cultures using phenanthrene as a sole carbon source 66

CHAPTER THREE:

3.1 Phylogenetic tree of 16S rDNA from clone IH-22 along with representative organisms taken from the NCBI database 83

3.2 Phylogenetic tree of 23S rDNA from clone IH-22 along with representative organisms taken from the NCBI database 85

3.3 Phylogenetic tree of 5S rDNA from clone IH-22 along with representative organisms taken from the NCBI database 86

3.4 Phylogenetic tree of 23S-16S rDNA intergenic spacer from clone IH-22 along with representative organisms taken from the NCBI database 87

3.5 Phylogenetic tree of putative 2-nitropropane dioxygenase found on clone IH-22 along with other representative genes taken from NCBI database 88

3.6 Graph of top 15% represented functional genes found in the Rancho La Brea Tar Pits 90

CHAPTER FOUR:

4.1 Phylogenetic analysis of dioxygenases detected on contigs with other known dioxygenase sequences from the NCBI database 117

4.2 Distribution of organisms from phylogenetic classification of contigs classified by class 118

4.3 Relative distribution organisms from phylogenetic assessment of contigs 119

LIST OF TABLES

CHAPTER TWO:

2.1 Table of BLAST results for selected DGGE bands from microbial consortia exposed to different PAHs 52

2.2 BLAST results for the bacterial isolates from phenanthrene enrichment cultures 64

CHAPTER THREE:

3.1 List of contigs and open reading frames detected after assembly of DNA insert from clone IH-22 81

3.2 Table of microarray analysis results of the top 15% represented functional genes 89

CHAPTER FOUR:

4.1 List of contigs greater than 400 bp in length along with their gene product, percent identity and organism of metagenome from which they originated 106

4.2 Distribution of contigs into different protein categories 114

4.3 Distribution of unknown contigs classified by similar sequences from different metagenomes 120

CHAPTER ONE: INTRODUCTION

PETROLEUM MICROBIOLOGY

The role of microorganisms in petroleum environments has been an area of interest for many biotechnology based applications. Petroleum is a complex mixture of heavy to light hydrocarbons and many other organic compounds, including organometallo constituents. Since petroleum hydrocarbons are recognized as substrates supporting microbial growth, they are both a target and a product of microbial metabolism (Hamme et al., 2003). Biodegraded oils represent a significant fraction of the petroleum in conventional oil reserves and interest in exploiting petroleum-degrading organisms for environmental cleanup has become central to petroleum microbiology (Atlas, 1981, Hamme et al., 2003).

In 1926, Edison Bastin and coworkers submitted the first report of active microbial communities in a petroleum reservoir (Bastin et al., 1926) and since then a wide variety of bacteria have been identified using both culture-dependent and culture-independent methods. Culture dependent methods include isolation and cultivation of bacteria on solid media, most-probable-number (MPN) liquid assays and Biolog substrate utilization plates. Microorganisms isolated from these sites have been extensively tested for bioremediation applications as they have the ability to degrade petroleum hydrocarbons in both the laboratory and environment (Ollivier and Magot, 2005). Alternatively, culture independent methods include phospholipid fatty acid analysis, fluorescence in situ hybridization and a variety of PCR-based approaches like denaturing gradient gel electrophoresis (DGGE) and 16S rRNA gene microbial surveys. Initial reports of underground reservoirs using molecular approaches suggest that the majority of

microorganisms inhabiting these environments are new species that represent a rich pool of novel genetic diversity with potential importance for industrial and petroleum microbiology (Hamme et al., 2003).

To date, studies of petroleum reservoirs and petroleum contaminated sites using culture based techniques have revealed physiologically diverse assemblages of thermophilic and hyperthermophilic anaerobic microorganisms along with bacteria that are able to live on complex petroleum hydrocarbons. Previously isolated anaerobic microbes include sulfate reducers (Stetter et al., 1993, TardyJacquenod et al., 1996), sulfidogens (Lharidon et al., 1995), fermentative bacteria (Davey et al., 1993, Grassia et al., 1996), manganese and iron reducers (Greene et al., 1997), methanogens (Ng et al., 1989, Rozanova et al., 1997) and acetogens (Davydovacharakhchyan et al., 1992). Classes of hydrocarbons that bacteria isolated from petroleum sites have shown to metabolize include n-alkanes (Kato et al., 2001), BTEX compounds (Lu et al., 2006) and polycyclic aromatic hydrocarbons (PAHs) (Alquati et al., 2005, Coral and Karagoz, 2005). Although these culture based approaches are useful for understanding the physiological properties of the isolated organisms, they do not provide comprehensive information on the composition of microbial communities, because only less than 1% of bacteria and archaea can be cultured in the laboratory (Amann et al., 1995). Therefore, it is difficult to assess the importance of these isolated organisms within their natural environment.

Due to the disparity in diversity between isolated bacteria and *in situ* microbial communities, culture-independent methods have been established in order to circumvent the culturing bias. Diversity studies on petroleum impacted sites have focused on PCR based

methods that examine the extracted DNA of the microbial communities. Community fingerprint analyses using PCR coupled with denaturing gradient gel electrophoresis (DGGE) have identified predominant bacteria in petroleum samples. Many phylogenetic groups have been represented from these studies, which include *Pseudomonas sp., Ochrobactrum sp., Staphylococcus sp., Sphingomonas sp*. and *Burkholderia sp.*(Jun et al., 2008, Yoshida et al., 2005). In addition, 16S rDNA clone libraries have been constructed to identify individual members of the microbial community present in petroleum based sites, which include *Arcobacter nitrofigili, Clostridium ramosum, Desulfobacula toluolicaa, Pandoraea pnomenusa* and *Pseudomonas stutzeri* (Grabowski et al., 2005, Li et al., 2006, Orphan et al., 2000, Voordouw et al., 1996). These methods have also revealed that diverse communities consisting of many novel and previously uncultured microorganisms exist in both petroleum reservoirs and petroleum contaminated sites. However, because only the 16S rRNA gene sequences are being examined, the actual functions of the community members within their environment are unknown. The inability to link functionality to species represents a major limitation in these studies.

CRUDE OIL, ASPHALTS AND HEAVY OIL

Crude oil is an extremely complex and variable mixture of organic compounds. These subsurface petroleum reservoirs have accumulated underground as a result of the anaerobic degradation of marine organisms over a very long time. Under conditions of high temperature and pressure, the organic material has been converted to natural gas, liquid crude oil, shale oil, tars and bitumen (Scragg, 2005). The majority of the compounds in crude oil and natural asphalts are hydrocarbons, which can range in molecular weight from methane to the high molecular weight tars and bitumen. These hydrocarbons also come in a diverse range of

structures; straight and branched chains, single or condensed rings and aromatic rings. The two major groups of aromatic hydrocarbons are the monocyclic hydrocarbons (eg. benzene, toluene, ethylbenzene and xylene) and polycyclic aromatic hydrocarbons (PAHs) such as naphthalene, anthracene and phenanthrene. In addition to hydrocarbons, crude oil contains between 0.05 and 3.0% heterocyclic compounds, which contain sulfur, nitrogen, oxygen and various heavy metals. The proportion of each individual compound can vary greatly between petroleum sources, which affect the properties of the oil. Light oils are oils with a high proportion of low molecular weight materials and heavy oils consist of a high proportion of high molecular weight compounds (Scragg, 2005).

Heavy oils such as asphalts, arise from the biodegradation of crude oils leaving a residue characterized by a high content of asphaltenes, composed mainly of polyaromatic carbon (i.e. polycondensed aromatic benzene units with oxygen, nitrogen and sulfur), combined with minor amounts of heavy metals, particularly vanadium and nickel that occur in porphyrin structures (Curiale et al., 1983). These biodegraded heavy oils dominate the world petroleum inventory and represent a significant fraction of the petroleum in conventional oil reserves (Head et al., 2003). The effects of biodegradation on the composition and physical properties of crude oil are well known. As microbial biodegradation removes the lighter fraction of the oil, which includes n-alkanes, isoprenoids and light aromatic compounds, heavier compounds are left, which include n-cyclohexanes, n-alkylbenzenes, hopanes, steranes and other asphaltines up to 30 carbons or more. This biological degradation of petroleum results in a deterioration of crude oil quality due to the enrichment of resins and asphaltenes during the subsequent removal of saturated hydrocarbons.

THE LA BREA TAR PITS

The Rancho La Brea Tar Pits are Pleistocene natural asphalt deposits located in Hancock Park of downtown Los Angeles, California. The presence of fossilized extinct animal remains at Rancho La Brea has led to estimations that these asphalt seeps have been in existence for at least 40,000 years (Akersten et al., 1983). These deposits consist of petroleum that has been degraded to the extent that a significant fraction of asphalts and heavy oils remain (Kim and Crowley, 2007). There have been many studies which examine petroleum degrading microorganisms in recently contaminated sites, or oil reservoirs which consist of the lighter fractions of crude oil. However little is known about the microbial diversity in heavy oils and natural asphalts.

In 1999, Kadavy et al. conducted the first microbial study based on the Rancho La Brea Tar Pits. They successfully cultured 14 isolates from the larvae of the oil fly *Helaeomyia petrolei*, which is found around natural oil seeps and pools of waste oil (Kadavy et al., 1999). 9 of the 14 isolates were identified as *Providencia rettgeri* and 3 were likely *Acinetobacter* isolates. Although there was no evidence that linked the isolates to petroleum hydrocarbon degradation, they were likely to exhibit pronounced solvent tolerance and may be a future source of industrially useful, solvent-tolerant enzymes (Kadavy et al., 1999).

The only other microbial study from the Rancho La Brea Tar Pits came in 2007, when Kim and Crowley conducted a survey of this site, which include detailed sequence information

on both culturable and non-culturable organisms and also revealed several hundred new species of bacteria and archaea (Kim and Crowley, 2007). In addition, DNA sequencing of petroleum hydrocarbon degrading genes from the tar pits show at least three new groups of enzymes that have not been previously described (Kim and Crowley, 2007). The discovery of new microorganisms and genes from the La Brea Tar Pits suggests that this site represents a rich pool of novel genetic diversity with potential applications for biotechnology.

MICROBIAL DIVERSITY IN NATURAL ASPHALTS OF THE RANCHO LA BREA TAR PITS *(KIM AND CROWLEY, 2007)*

As previously mentioned, in 2007 Kim and Crowley conducted a microbial survey, in which 235 individual 16S rRNA genes were sequenced to identify the predominant phylogenetic groups within the microbial community of the Rancho La Brea Tar Pits. Of the dominant phylogenetic groups, sequence analysis revealed the presence of halophillic Archaea. From 29 clones, there were two clusters of closely related species from two unclassified genera as shown in Figure 1.1. Sixteen clones were similar to *Natronorubrum* spp, and another 11 were identified as *Natrialba* spp. Their overwhelming presence within the survey implies that both are dominant Archaeal species. Of the 29 Archaea sequences that were obtained, only 2 were outside of the clusters representing the new genera. The occurrence of the closely related Archaea imply an initial selection of closely related bacteria that share specific traits that enable them to survive in this environment, such as selection for more efficient petroleum degrading enzymes (Kim and Crowley, 2007).

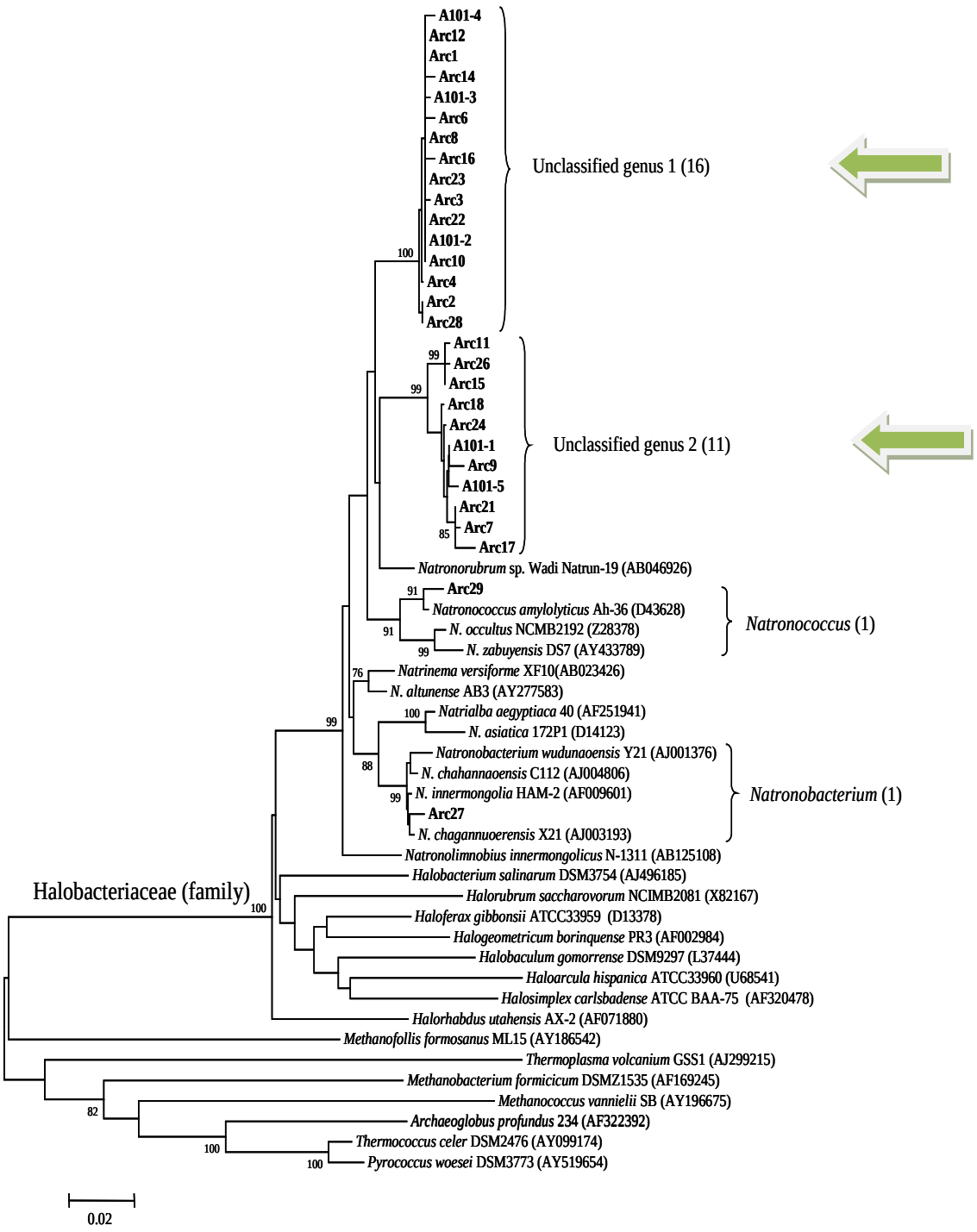

Figure 1.1. Phylogenetic tree showing partial list of Archaea, which reveals two sets of closely related species representing unclassified genera, indicated by arrows (Kim and Crowley, 2007).

In addition to Archaea, the predominant Bacteria in the tar pits were *Gammaproteobacteria,* (purple sulfur bacteria) (Kim and Crowley, 2007). Of 60 clones, there were 3 unclassified families in the order *Chromatiales*. Other families included X*anthomanadaceae* (5 clones) and *Pseudomonadaceae* (5 clones). There were also 5 additional phyla that included *Planctomycetes*, *Gemmatimonadetes*, BRC1, *Nitrospira* and *Verrucomicrobia*. Clones also revealed representative species from Alpha- and Betaprotobacteria, along with various representatives from *Acidobacteria, Actinobacteria, Bacteroidetes* and *Clostridia.* There were also a significant proportion of unclassified clones present in the survey, which imply that novel microorganisms exist within the tar pits. In total, one new order and 7 new families of bacteria were found. One unique finding was the identification of 16 clones representing the family Rubroacteraceae, which are of interest for their ability to survive in high-ionizing radiation environments (Ferreira et al., 1999). This finding suggests that the La Brea Tar Pits have selected for bacteria that can withstand mutagenic environments.

The La Brea Tar Pits also yielded large numbers of *Pseudomonas* sp. Pseudomonads are among the most catabolically diverse bacteria in nature. Of special interest is their ability to degrade petroleum hydrocarbons, halogenated solvents and many other types of xenobiotics in pure culture (Chaudhry and Chapalamadugu, 1991). Using a *Pseudomonas*-selective primer set, 14 unique sequences that represent 8 new genomovars of *P. stutzeri* were found (Figure 1.2). Each genomovar with sequences of 95-98% similarity suggest a radiation of new species that have adapted to growth on natural asphalts and heavy oil compounds. A major question that

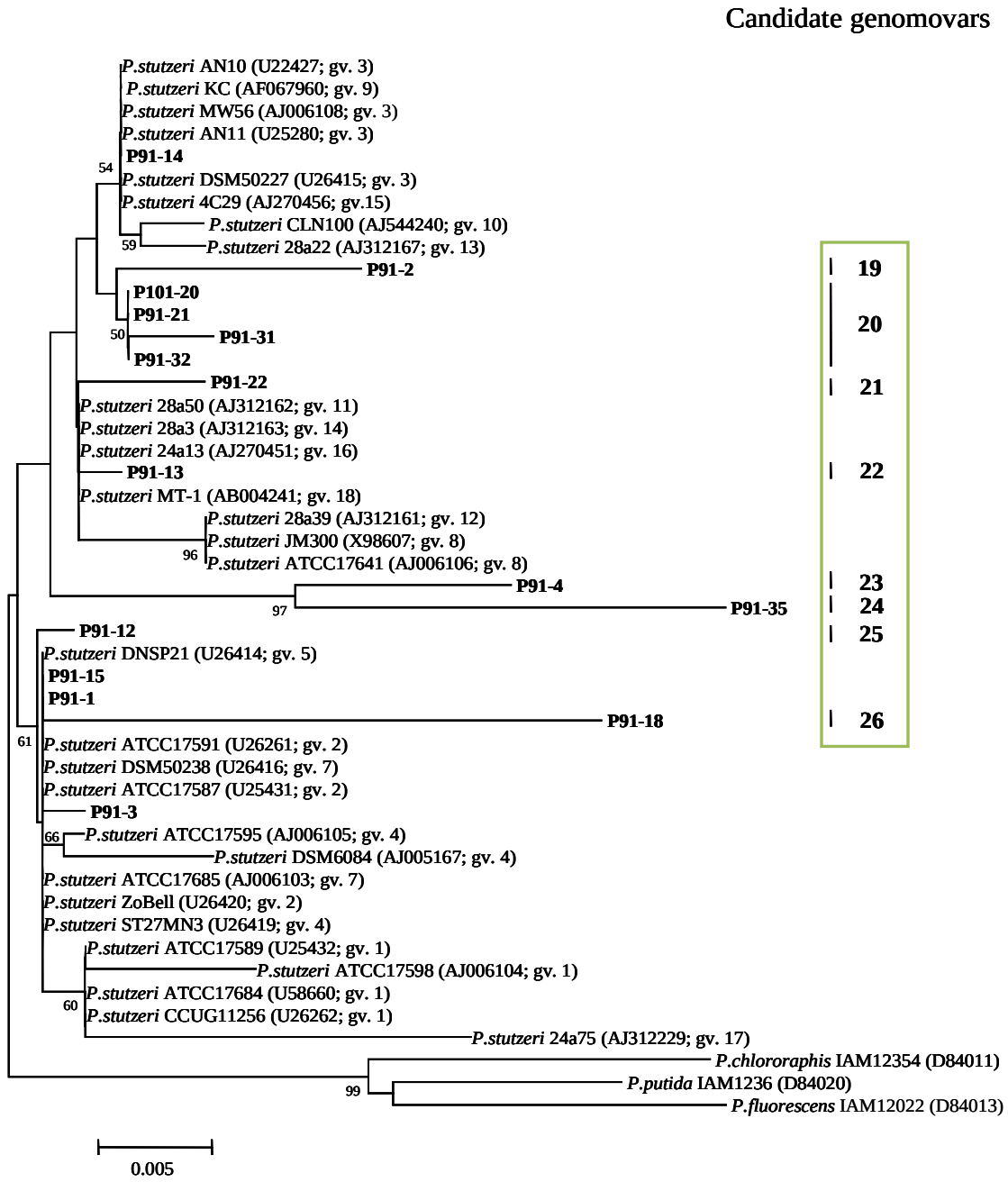

Figure 1.2. Phylogenetic tree of *Pseudomonas stutzeri* found in the La Brea Tar Pits. Candidate genomovars are enclosed in box (Kim and Crowley, 2007).

will be addressed is whether these genomovars are catabolically distinct, each using different types of petroleum hydrocarbons, or whether they share common catabolic pathways.

In conjunction with the species diversity for Archaea and Bacteria, Kim and Crowley also surveyed key functional genes encoding dioxygenase enzymes. Many of the dioxygenase sequences are predominantly associated with proteobacterial biphenyl dioxygenases with similarities ranging from 79 to 95%. In addition, detailed sequence analysis revealed at least three new types of ring hydroxylating enzymes (Figure 1.3). Although these enzymes were most closely associated with sequences encoding benzene, toluene and biphenyl dioxygenases, they were sufficiently distant that it is not possible to infer their function.

In addition to phylogenetic studies, microorganisms were cultured and tested for their ability to grow using asphalt a sole carbon source. Isolates included *Pseudomonas* spp., *Bacillus* spp., and *Citrobacter* spp. Using 20% salt medium yielded three isolates of haloalkalophilic *Bacillus* spp. All of the isolates were able to grow on M9 medium using asphalt as the sole carbon source implying their ability to metabolize petroleum compounds. To confirm the presence of bacteria associated to the asphalt particle, fluorescent in situ hybridization (FISH) was used to study the physical relationships between the various groups of microorganisms and the asphalt particles. Asphalt samples were microscopically examined using FISH with DNA probes targeted for *Bacteria*, *Archaea* and *Pseudomonas* spp. Cells occurred in dense clusters of colonies, which suggests that this close association may be consortia that metabolize the asphalt cooperatively or that may be conducive for horizontal gene transfer of mobile genetic elements which spread catabolic genes for degrading petroleum hydrocarbons throughout a microbial community (Kim and Crowley, 2007).

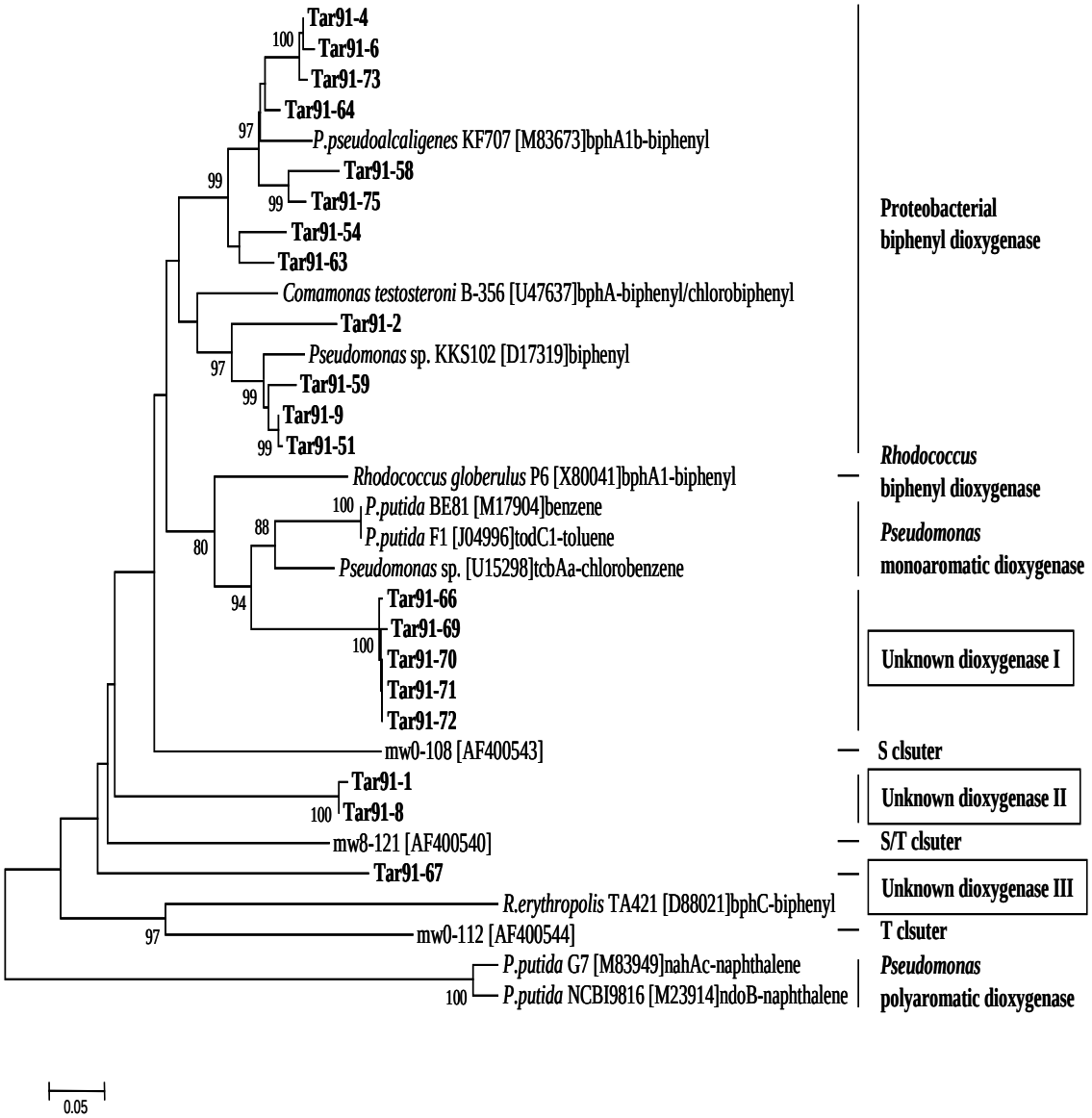

Figure 1.3. Diversity of dioxygenase sequences obtained from the La Brea Tar Pits using an aldolase primer set. Sequences from the tar pits (in bold) suggest three new classes of dioxygenases (Kim and Crowley, 2007).

BIOREMEDIATION

Since it is well known that microbes are able to use petroleum as a carbon and energy source, the most practical use of the La Brea Tar Pits' microorganisms and their hydrocarbon degrading enzymes is for bioremediation applications. Bioremediation is the process by which living organisms degrade or transform hazardous organic compounds to a relatively non toxic state (Chapelle, 1999, Dua et al., 2002, Paul et al., 2005, Salanitro, 2001). In bacterial mineralization, bioremediation results in complete degradation of the pollutant, leaving only carbon dioxide and water (Heitzer and Sayler, 1993). Bioremediation is often considered a cost-effective and environmentally friendly alternative approach in comparison to the conventional methods of remediation, such as excavation and incineration, which is very costly and can result in the generation of toxic air emissions (Kuiper et al., 2004). In addition, bioremediation techniques can be applied in situ without the removal of the contaminated soil, thus not disturbing the environment. The application of bioremediation has been growing, due to its success in the cleanup of petroleum hydrocarbons, as seen in the Exxon Valdez oil spill (Bragg et al., 1994, Pritchard et al., 1992). Currently, microorganisms with the capability to degrade various pollutants (e.g. polycyclic aromatic hydrocarbons, nitroaromatics, polychlorinated biphenyls and oil components) have been isolated in hopes of exploiting their metabolic potential for the remediation of contaminated sites (Spain, 1995, Dua et al., 2002).

There are different types of bioremediation technologies and strategies that are continuing to be developed. They aim to assist the microorganisms' growth and increase microbial populations by creating optimum environmental conditions for them to detoxify or mineralize the contaminants present. The specific bioremediation technology that is to be

applied is determined by several factors. These are (1) the nature of the pollutants, (2) the quantity and concentration of the pollutant, (3) the soil structure and conditions, which determine the movement of the pollutants and (4) the type of indigenous microorganisms present and their nutritional state (Balckburn and Hafker, 1993, Dua et al., 2002, Kuiper et al., 2004, Long, 1993). Since there are different microorganisms that can degrade different types of compounds and survive under different conditions, each site may require a specific bioremediation approach or a combination thereof. There are several types of bioremediation technologies and strategies that are being used today, which include natural attenuation, biostimulation, rhizoremediation, phytoremediation and bioaugmentation.

Natural attenuation relies on the natural processes of the indigenous microorganisms to remediate or attenuate the contaminated soil or groundwater. When a site first becomes polluted, the indigenous microbial population will adapt to the newly added contaminants and the microorganisms' ability to metabolize the xenobiotics results in the bioremediation of the compounds. Natural attenuation occurs at most polluted sites, as it is considered to be the simplest form of bioremediation. Contaminant levels are consistently being monitored until the compounds of interest are degraded or detoxified to a safe level. This strategy is usually applied on sites where no other restoration techniques are applicable or at sites with "low" public or environmental value, because restoration can take up to decades (Kuiper et al., 2004). Natural attenuation has shown to be successful in BTEX contaminated groundwater (Wickramanayake et al., 2000, Yong and Mulligan, 2004). However, one of the main limitations to this technology is that the microbial community present on the site may not possess the required catabolic genes

to completely degrade the xenobiotics present, which can also result in the formation of more toxic intermediate compounds (Heitzer and Sayler, 1993, Kuiper et al., 2004).

Biostimulation involves the direct addition of nutrients to the contaminated site, such as phosphorus, nitrogen or oxygen, in order to stimulate and increase the biodegradation rates of the indigenous microorganisms. This is the most widely used bioremediation procedure as the primary advantage to this strategy is that biodegradation will take place using naturally occurring bacteria, that are indigenous to the contaminated site; therefore the microorganisms will have already adapted to their environment. Biostimulation has had much success in the restoration of oil contaminated sites (Margesin and Schinner, 2001, Margesin and Schinner, 1999). In the case of the Exxon Valdez oil spill, the effectiveness of the natural biodegradation of the oil was limited by the availability of nitrogen and phosphorus. Biodegradation of the oil was stimulated and enhanced by the addition of an oleophilic fertilizer by up to two fold (Pritchard et al., 1992). However, like natural attenuation, the degradative ability of the microbial community may be limited unless all the catabolic genes required to fully degrade the various components of the petroleum are not present.

Phytoremediation is defined as the use of plants to extract, sequester or detoxify pollutants. This strategy is environmentally friendly and visually attractive, while keeping the structure of the soil highly maintained (Khan et al., 2000). There are two main types of pollutants in which phytoremediation target, elemental (e.g. heavy metals and radionuclides) and organic (e.g. polychlorinated biphenyls and polycyclic aromatic hydrocarbons) pollutants (Meagher, 2000, Kuiper et al., 2004). In the phytoremediation of elemental pollutants such as metals and metaloids, the pollutant is extracted from soil by the plant root system and then

converted into a less toxic state. This can been demonstrated for a number of different plant species (e.g. tobacco, sunflower, mustard and maize) (Kuiper et al., 2004). In the case of organic pollutants, it is known that certain compounds can be mineralized completely by using poplar trees, willow, alfalfa and various grass species (Kuiper et al., 2004). However, much less is known about the phytoremediation of organic compounds compared to bacterial degradation. In addition, there are several limitations to this technology, 1) the slow growth of plants and the time for sufficient remediation, 2) the limited depth of the root system, 3) the selectivity of certain pollutants, 4) the probability of being in a food chain and 5) dependence on the climate (Kuiper et al., 2004, Khan et al., 2000).

Rhizoremediation is another bioremediation technology that takes advantage of both the plant enzymes involved in phytoremediation and the degradation abilities of the plant-associated microbial community in the rhizosphere (Kuiper et al., 2004, Anderson et al., 1993). The plant can be considered to be a biological pump and treatment system, where the roots attract water, accumulating water-soluble pollutants in the rhizosphere, and concludes with microbial degradation or translocation of the contaminant into the plant tissues (Erickson, 1997). Compounds that have been shown to degrade in the rhizosphere include trichloroethylene (Walton and Anderson, 1990), polycyclic aromatic hydrocarbons (Radwan et al., 1995), polychlorinated biphenyls (Brazil et al., 1995), and certain herbicides and pesticides (Hoagland et al., 1994, Jacobsen, 1997). These studies suggest that the plants are protected by the degrading bacteria, which suggest that they have the ability to harbor large numbers of degrader bacteria on their root systems. In addition, the plant roots were found to incorporate nutrients (e.g. sugars, organic acids and amino acids) while improving aeration, which results in

an increase in the microbial population numbers and metabolic activity in the rhizosphere, thus stimulating biodegradation of the pollutants (Vancura and Hovadik, 1965, Kuiper et al., 2004).

Bioaugmention is a method to improve degradation and enhance the transformation rate of xenobiotics by the inoculation (i.e. seeding) of specific catabolically-relevant microbes into the contaminated area to degrade the pollutants of interest (Kuiper et al., 2004, Thompson et al., 2005). Many different microbes have been isolated that are genetically equipped to mineralize recalcitrant pollutants such as PAHs, chloriniated aliphatics and aromatics, nitroaromatics, long-chain alkanes and BTEX compounds (Grosser et al., 1991, Kato et al., 2001, Spain, 1995, Cerniglia, 1993, Lu et al., 2006). In addition to wild-type isolates, microorganisms can also be genetically modified, so they contain the required catabolic genes needed for biodegradation. These degradative genes can be incorporated into the chromosome directly or can be sustained in the genome as catabolic plasmids or other mobile genetic elements (MGEs) (Chen et al., 1999, Deutschbauer et al., 2006, Pieper and Reineke, 2000, Singh et al., 2008, Timmis and Pieper, 1999). The construction of genetically engineered microbes (GEMs) is relevant for the degradation of pollutants that have no microbial catabolic pathways, such as PCBs and chloroethylenes (Megharaj et al., 1997, Brazil et al., 1995). The GEMs that have been constructed are equipped with new metabolic pathways, in which parts of known pathways are combined, or by optimizing known pathways, via over expression of certain genes or operons (Chen et al., 1999, Paul et al., 2005, Pieper and Reineke, 2000, Singh et al., 2008). The use of bioaugmentation for the remediation of soil and groundwater has yielded ambiguous results (Goldstein et al., 1985). Goldstein et al. (1985) reported five possible reasons for the failures of this strategy. First, the concentration of the contaminants at the site can be too low to support

growth of the inoculated bacteria. Also, contaminants with low bioavailability may not be able to support microbial growth. Second, the presence of other compounds may inhibit growth of the inoculum. Third, protozoan-grazing rates on the inoculum can be higher than the growth rates of the bacteria. Fourth, the inoculum can prefer to use other carbon sources that are present in the soil, instead of the contaminant. Fifth, the inocula may fail because of the inability of the microbes to spread through the soil and reach the pollutant.

PETROLEUM HYDROCARBON CONTAMINATION

Contamination caused by petroleum hydrocarbons has gathered much attention due to concerns stemming from large scale releases into the environment (Chapelle, 1999). Polycyclic aromatic hydrocarbons (PAHs) and BTEX (benzene, toluene, ethylbenzene and xylene) compounds are environmental pollutants commonly found in petroleum contaminated sites. PAHs have carcinogenic and mutagenic properties and it is the heavier fractions that pose the greatest long term recalcitrance and threat to the environment (Talley, 2006). The lighter compounds (i.e. BTEX) are able to leach into groundwater basins and contaminate freshwater supplies (Testa and Winegardner, 1991). Each year, approximately 1,680,000 gallons of crude oil are spilled on land due to pipeline failures and more than 200,000 underground storage tanks in the US have leaked gasoline and other fuels into the soil, sediments and groundwater aquifers (Bedient et al., 1994), making petroleum and its derived fuels the most ubiquitous organic pollutants around the world (Salanitro, 2001). Due to the amount of petroleum hydrocarbons contaminating the environment, improved bioremediation technologies are necessary. Considering the amount of biodegradation that has taken place within the deposits of the La

Brea Tar Pits, its microbial community may be a source of novel and more efficient enzymes for the remediation of petroleum hydrocarbons and other emerging xenobiotics.

Polycyclic aromatic hydrocarbon contamination has gathered much attention due to their carcinogenic and mutagenic properties and ubiquitous nature. On the basis of their abundance and toxicity, the U.S. Environmental Protection agency has classified 16 PAHs spanning from 2-6 condensed aromatic rings as priority pollutants for environmental cleanup (Achten and Hofmann, 2009, Keith and Telliard, 1979). Their chemical properties and environmental fate are dependent upon both the molecular size (i.e. number of aromatic rings) and the pattern of the ring linkage. PAH molecule stability and hydrophobicity are two primary factors which contribute to the persistence of high molecular weight PAHs in the environment (Kanaly and Harayama, 2000). The relationship between PAH environmental persistence and increasing numbers of aromatic rings is consistent with the results of various studies correlating environmental biodegradation rates and PAH molecules size (Bossert and Bartha, 1986, Heitkamp and Cerniglia, 1987, Herbes and Schwall, 1978). PAHs are present as natural constituents in fossil fuels, are formed during combustion of organic material. Pont sources of PAHs may originate from oil spills, tanker leakage, contaminated industrial sites (e.g. gas plants, refineries, etc.), seepage from natural oil reservoirs, aluminum production or steel works (Achten and Hofmann, 2009, Morgan and Watkinson, 1989). In the next couple of decades, billions of dollars will be spent to remediate all of the sites polluted with polycyclic aromatic hydrocarbons (Rosenberg, 1993).

PETROLEUM HYDROCARBON DEGRADATION

As mentioned previously, PAHs represent a large group of soil pollutants contained in petroleum and a number of them have been prioritized for environmental cleanup. Once these pollutants enter the soil, they are trapped into soil pores and immobilized by adsorption to the soil matrix (Aprill and Sims, 1990). Biological degradation then remains as the most significant means of decontamination as the ability of microbes to degrade PAHs has been noted for many years. The first studies on the metabolic pathways and the enzymatic reactions involved in the mineralization of naphthalene were by Davies and Evans (1964). The degradation of PAHs by microbes has since been studied extensively and has been shown for a wide variety of bacteria, fungi, and algae (Cerniglia, 1993, Davies and Evans, 1964, Kuiper et al., 2004). The bacterial PAH catabolic genes often are located on large plasmids, together with regulatory genes. These catabolic plasmids and other mobile genetic elements contribute to the adaptation of the indigenous microbes toward PAH pollutants and other xenobiotics, through horizontal gene transfer (Springael and Top, 2004, Stuart-Keil et al., 1998, Top and Springael, 2003, Vandermeer et al., 1992, Vandermeer, 1994).

The biodegradation of high-molecular-weight (HMW) PAHs with three of more fused rings is less known. However, the number of organisms found to degrade HMW PAHs is increasing (Mueller et al., 1989, Heitkamp and Cerniglia, 1989, Kanaly and Harayama, 2000). In 1988, Heitkamp and Cerniglia published the first study on the isolation of a bacterium from the environment that could extensively degrade PAHs containing four aromatic rings. In 1989, Mueller et al. demonstrated for the first time that the utilization of a PAH containing four or more aromatic rings as a sole source of carbon and energy by bacteria is possible. In PAHs with

more than three fused rings, co-metabolism often serves as the main route for biodegradation. Two- or three-ringed PAHs and PAHs degradative intermediates (e.g. salicylic acid) are able to function as a carbon source during the co-metabolic degradation of HMW PAHs and also serve for induction of enzymes used for degradation of PAHs such as naphthalene (Kuiper et al., 2004).

The biochemical pathways for the biodegradation of aromatic compounds have been well described (Cerniglia, 1993, Johnsen et al., 2005). It is understood that the initial step in the aerobic catabolism of a PAH molecule occurs via oxidation of the PAH to a dihydrodiol by a multicomponent enzyme system (Kanaly and Harayama, 2000). The dihydrodiols may then be processed through either an ortho cleavage pathway or a meta cleavage pathway, leading to central intermediates such as protocatechuates and catechols. These compounds are further converted to tricarboxylic acid cycle intermediates (Kuiper et al., 2004). In the case of a number of PAHs, dioxygenase enzymes are usually the first biocatalysts used in aerobic petroleum hydrocarbon degradation (Hamme et al., 2003). Naphthalene dioxygenases introduce an oxygen molecule at the 1,2-position of the aromatic nucleus, producing dihydrodiol compounds (Figure 1.4), which facilitates further degradation of the compound by other enzymes (Habe et al., 2003). It is known that naphthalene dioxygenases are able to attack a variety of petroleum hydrocarbons (e.g. biphenyl, phenanthrene and anthracene), and thus may prove to be very useful enzymes for bioremediation (Habe et al., 2003, Hamme et al., 2003).

In addition, the novel 16S rRNA gene sequences also suggest that there may be new petroleum degrader species, which are presently metabolizing the asphalts. The occurrence of the new species and novel degradative genes implies that there may be previously undescribed catabolic pathways for the degradation of petroleum hydrocarbons. Although, we cannot

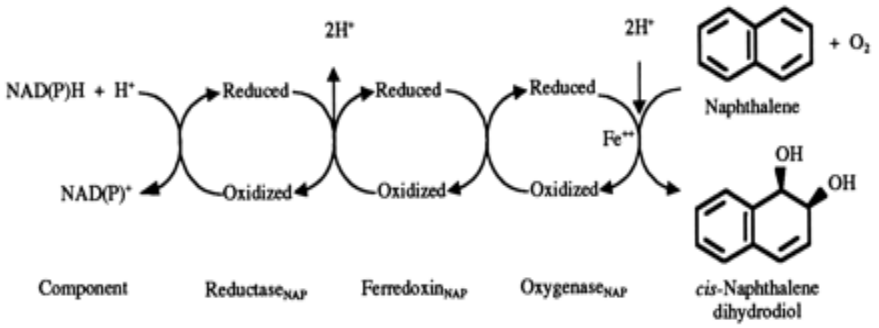

Figure 1.4. Initial oxidation of Naphthalene to cis-1,2-Dihydroxy-1,2-dihydronaphthalene by naphthalene dioxygenase (Habe et al., 2003).

assume function of the newly discovered gene sequences, we also cannot reject the possibility that these enzymes and others found in the tar pits may be more efficient for petroleum biodegradation than those that have been previously studied for culturable bacteria.

METAGENOMICS

Studies of microorganisms in pure laboratory culture for over a century have led to significant advances into microbial genetics and physiology, biotechnology and molecular biology. However, the majority of the bacteria have not or perhaps cannot be cultured under laboratory conditions. Microorganisms live in very complex microbial communities (e.g. soil, asphalt, sewage, contaminated sites) and thus it is extremely difficult to recreate those conditions on a Petri dish. In the last decade, the most recent innovation that circumvents this problem is the study of the collective DNA contained in microbial communities. This technology is termed “metagenomics”.

Metagenomics describes the functional and sequence-based analysis of the collective microbial genomes contained in an environmental sample. In metagenomics, the power of genomic analysis is applied to entire communities of microbes, bypassing the need to isolate and culture individual bacterial community members (NRC, 2007). The main principle involves to the extracting and sequencing of DNA from the microbial community by extracting bulk DNA from environmental samples, which can then be used to discover new genes and enzymes or to assemble catabolic pathways or even entire genomes. There are several different strategies when analyzing an environment using metagenomic approaches. However, every analysis begins with the extraction of DNA directly from its environment. Depending on the strategy, the DNA fragments can range in size from as small as 1 kb to over 100 kb. These fragments are then

ligated into suitable vectors (e.g. cosmids, fosmids, BACs) then cloned into appropriate hosts (e.g. *E. coli*), thus yielding a library of DNA sequences that can be amplified by culture of cells containing the DNA inserts. Once the library is created, the DNA inserts can be randomly sequenced or screened using PCR primers or by monitoring expression of functional genes using an appropriate assay procedure for selected functions such utilization of a particular substrate (Figure 1.5).

Using the random sequencing approach, also known as shotgun sequencing, large numbers of sequences with overlapping fragments are analyzed with the aim of assembling entire genes, pathways, or even an entire organism's genome. This strategy is useful for very selective and simple environments, which contain few dominant organisms. For sequence based and functional based screening, selected metagenomic clones are sequenced, instead of the entire library. This strategy is mainly used for more diverse communities, which contain many different organisms. In this case, the DNA inserts are larger (i.e. 20 kb to over 100 kb), which makes it easier to obtain entire gene or operon sequences and identify the particular taxa with which the genes are associated. In a sequenced based analysis, clones are selected by searching for genes of interest via polymerase chain reaction or colony blotting. Clones that contain the target gene are then selected for analysis. In contrast, using a functional based analysis, genes from the DNA insert are expressed in an expression vector and the clones with the desired functions are selected, after which the entire insert is sequenced and analyzed.

Metagenomics has led to several advances in microbiology and biotechnology by examining unculturable microbes. Metagenomics has been used successfully to isolate genes encoding antibiotics (Gillespie et al., 2002), discover unknown genes (Venter et al., 2004),

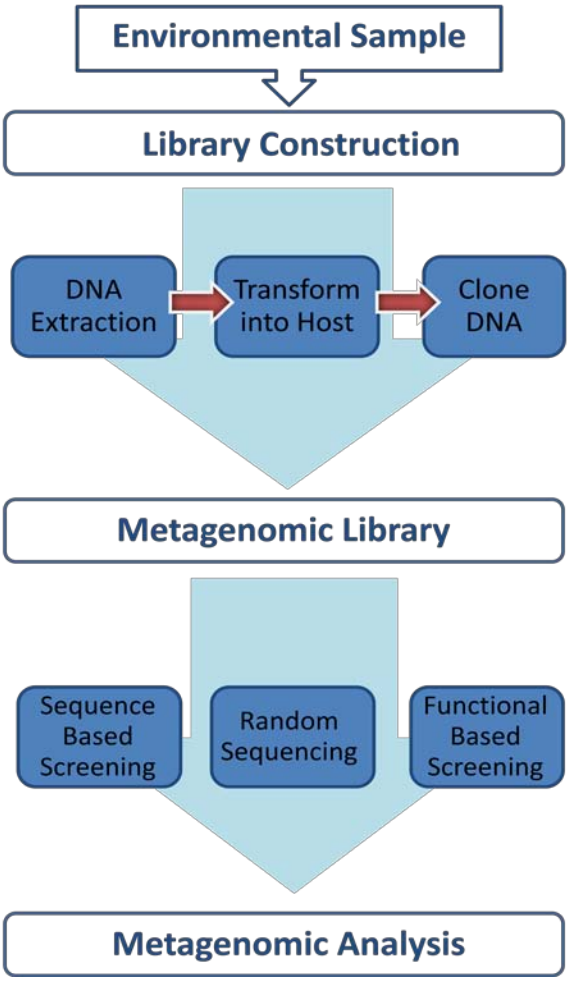

Figure 1.5. Metagenomics involves constructing a DNA library from an environment's microbial population and then analyzing the functions and sequences in the library.

express novel biocatalysts (Voget et al., 2003, Wexler et al., 2005), find new gene functions (Beja et al., 2000) and even sequence near complete genomes of so far unculturable microbes (Martin et al., 2006, Tyson et al., 2004). Metagenomic approaches also have lead to the discovery of petroleum degrading genes. In 2007, Suenaga et al. was able to find novel genes for the degradation of aromatic compounds from coke plant wastewater and Ono et al. characterized naphthalene-catabolic genes from oil contaminated soil.

NEXT-GENERATION SEQUENCING

There is growing awareness of the power of high-throughput DNA sequencing to explore the microbial diversity in complex communities. Interest first started when the extent of the microbial diversity that is unknown was revealed through the development of polymerase chain reaction-based technologies to amplify and sequence 16S rDNA signatures from unculturable and uncharacterized organisms (Schmidt et al., 1991, Kuiper et al., 2004). By extending the cloning and sequencing approaches that employ high-throughput shotgun sequencing to entire complex communities, Venter and colleagues revealed the diversity of protein-coding genes in ocean-dwelling microbes (Venter et al., 2004). The majority of these genes are unknown, but can be tentatively assigned to different functions and taxa using bioinformatic approaches that compare the similiarities of these sequences to those studied in other culturable microorganisms.

DNA sequencing has been dominated for three decades by Sanger dideoxy chain-termination methods. This sequencing technology is still being used in many laboratories worldwide, providing more than 1 kb of sequencing per sample and has driven the first revolution in whole-genome and metagenomic sequencing. This technology has currently

reached a plateau with the development of single-tube chemistry using fluorescently labeled termination bases, heat stable polymerases, and automated capillary electrophoresis (Snyder et al., 2009).

While shotgun-sequencing is still used for the finishing of whole-genome sequencing and whole community sequencing, it is technically challenging, expensive, and labor intensive. More recently, next-generation sequencing technologies have and are continuing to be developed, and promise to deliver sequence data much faster and cheaper than Sanger methods. These new technologies produce data faster and cheaper due to two fundamental differences in the methods that are employed. Advances include elimination of the requirement for in vivo amplification of DNA fragments and the subsequent creation of a clone library. Second, the ability of developing new sequencing chemistries enable the samples to be sequenced in real time.

Since first introduced to the market in 2005, next-generation sequencing has had a tremendous impact on genomic research. Currently, the most advanced technologies that are commercially available today include the 454 FLX Pyrosequencing instrument (Roche Applied Science), the Illumina IIG analyzer (Illumina, Inc.) and the SOLiD instrument from Applied Biosystems. These next-generation sequencing technologies offer dramatic increases in cost-effective sequence throughput, but at the expense of read lengths. The throughput of these sequencers is about the order of multiple gigabases per week.

The 454 technology, which was the first next-generation sequencing technology to be released, takes advantage of a highly efficient in vitro DNA amplification method known as emulsion PCR. In emulsion PCR, individual DNA fragment-carrying streptavidin beads, obtained

through shearing the DNA and attaching the fragments to the beads using adapters, are captured in emulsion droplets. The droplets act as individual amplification reactors, producing ~10^7 clonal copies of a unique DNA template per bead. The beads are transferred into the wells of a picotiter plate and the clonally related templates are analyzed using a pyrosequencing reaction. The use of the picotiter plate allows hundreds of thousands of pyrosequencing reaction to be carried out in parallel, which increases the throughput.

The Illumina IIG analyzer achieves clone-free DNA amplification by attaching single-stranded DNA fragments to a solid surface known as a flow cell, and conducts solid-phase bridge amplification of a single molecule DNA template. In this process, one end of a single DNA molecule is attached to a solid surface using an adapter, which produces the template for the synthesis of complementary strands. After the amplification step, individual flow cells with over 40 million clusters are produced, with each cluster composed of approximately 1000 clonal copies of a single template molecule. The templates are then sequenced in a parallel fashion using a DNA sequencing-by-synthesis approach.

The ABI/SOLiD technology uses massively parallel sequencing by hybridization-ligation, implemented in the supported oligonucleotide ligation and detection system (SOLiD). Construction of sequencing libraries for analysis on the SOLiD instrument begins with an emulsion PCR single-molecule amplification step similar to that used in the 454 technique. The amplification products are transferred onto a glass surface where sequencing occurs by sequential rounds of hybridization and ligation with 16 dinucleotide combinations that are labeled with four different fluorescent dyes (each dye used to label four dinucleotides).

Due to their much improved cost and time effectiveness, next generation sequencing technologies have found broad applicability not only for phylogenetics, but also for functional genomics, gene expression profiling, genome annotation, small ncRNA discovery and profiling and detection of aberrant transcription. Although next generation sequencers are already being widely used, there are other sequencing methods in the pipeline, such as nanopore sequencing, which is being developed to decrease the sequencing cost and enhance throughput even further.

FROM METAGENOMICS TO BIOREMEDIATION APPLICATIONS

Metagenomics has great potential for both fundamental and industrial applications that range from the understanding of microbial adaptation and evolution to the discovery of new enzymes for their direct use (Ferrer et al., 2007, Ferrer et al., 2005, Lorenz and Eck, 2005, Deutschbauer et al., 2006). Sites polluted with toxic chemicals and industrial wastes have transformed environmental biotechnology because these habitats include niches for microorganisms that have the necessary enzymes to use these compounds as their carbon and energy source. The genetic diversity found in these environments includes genes encoding degradative enzymes and pathways for recalcitrant chemicals, which are potentially useful for the remediation of environmental pollution and as sources of novel catalytic activities with applications for green chemistry and biotechnology (Galvao et al., 2005, Schmid et al., 2001). There are a wide range of studies that examine the biotransformation, biodegradation and bioremediation of petroleum hydrocarbons (Atlas and Cerniglia, 1995, Atlas, 1995, Bogardt and Hemmingsen, 1992, Rosenberg, 1993). However, surveys of the enzymes that are available in nature have only begun.

In the case of the La Brea Tar Pits, the discovery of novel enzymes and pathways has opened the door in the search for new bacteria, genes, and catabolic pathways that can be used for biotechnology. Selection of bacterial communities for degrading petroleum substances occurs rapidly after even short-term exposures of soil to petroleum hydrocarbons following oil spills (Vandermeer et al., 1992, Vandermeer, 1994). Over time spans encompassing millennia, bacteria that can tolerate this environment would be expected to undergo genetic adaptations that may lead to the evolution of new ecotypes and species and enzymes for growth on petroleum hydrocarbons. During adaptation of communities, genes for petroleum hydrocarbon-degrading enzymes that are carried on plasmids or transposons may be exchanged between species. In turn, new catabolic pathways eventually may be assembled and modified for efficient regulation (Rabus et al., 2005).Therefore, some of these genes and pathways may be used to construct genetically modified organisms. The combination of enzymes and pathways from different organisms in one recipient strain is a useful strategy for designing bacteria with enhanced degradation capabilities. Several GMOs already have been successfully constructed to increase degradative capability and utility for bioremediation (Dua et al., 2002, Chen et al., 1999, Paul et al., 2005, Timmis and Pieper, 1999). These natural asphalts that originated 40,000 years ago likely contain efficient enzymes and catabolic pathways, which can possibly improve these designer strains for the biodegradation of petroleum hydrocarbons and other organic compounds. The initial steps to achieving this future goal would be to first find and characterize the genes that encode these biocatalysts.

BIBLIOGRAPHY

ACHTEN, C. & HOFMANN, T. 2009. Native polycyclic aromatic hydrocarbons (PAH) in coals - A hardly recognized source of environmental contamination. *Science of the Total Environment,* 407**,** 2461-2473.

AKERSTEN, W. A., SHAW, C. A. & JEFFERSON, G. T. 1983. RANCHO-LA-BREA - STATUS AND FUTURE. *Paleobiology,* 9**,** 211-217.

ALQUATI, C., PAPACCHINI, M., RICCARDI, C., SPICAGLIA, S. & BESTETTI, G. 2005. Diversity of naphthalene-degrading bacteria from a petroleum contaminated soil. *Annals of Microbiology,* 55**,** 237-242.

AMANN, R. I., LUDWIG, W. & SCHLEIFER, K. H. 1995. PHYLOGENETIC IDENTIFICATION AND IN-SITU DETECTION OF INDIVIDUAL MICROBIAL-CELLS WITHOUT CULTIVATION. *Microbiological Reviews,* 59**,** 143-169.

ANDERSON, T. A., GUTHRIE, E. A. & WALTON, B. T. 1993. BIOREMEDIATION IN THE RHIZOSPHERE. *Environmental Science & Technology,* 27**,** 2630-2636.

APRILL, W. & SIMS, R. C. 1990. EVALUATION OF THE USE OF PRAIRIE GRASSES FOR STIMULATING POLYCYCLIC AROMATIC HYDROCARBON TREATMENT IN SOIL. *Chemosphere,* 20**,** 253-265.

ATLAS, R. M. 1981. MICROBIAL-DEGRADATION OF PETROLEUM-HYDROCARBONS - AN ENVIRONMENTAL PERSPECTIVE. *Microbiological Reviews,* 45**,** 180-209.

ATLAS, R. M. 1995. Petroleum biodegradation and oil spill bioremediation. *Marine Pollution Bulletin,* 31**,** 178-182.

ATLAS, R. M. & CERNIGLIA, C. E. 1995. BIOREMEDIATION OF PETROLEUM POLLUTANTS - DIVERSITY AND ENVIRONMENTAL ASPECTS OF HYDROCARBON BIODEGRADATION. *Bioscience,* 45**,** 332-338.

BALCKBURN, J. W. & HAFKER, W. R. 1993. The impact of biochemistry, bioavailability and bioactivity on the selection of bioremediation techniques. *Trends in Biotechnology,* 11**,** 328-333.

BASTIN, E. S., GREER, F. E., MERRITT, C. A. & MOULTON, G. 1926. The presence of sulphate reducing bacteria in oil field waters. *Science,* 63**,** 21-24.

BEDIENT, P. B., RIFAI, H. S. & NEWELL, C. J. 1994. Ground water contamination: transport and remediation. *Ground water contamination: transport and remediation.***,** xvi + 541 pp.

BEJA, O., ARAVIND, L., KOONIN, E. V., SUZUKI, M. T., HADD, A., NGUYEN, L. P., JOVANOVICH, S., GATES, C. M., FELDMAN, R. A., SPUDICH, J. L., SPUDICH, E. N. & DELONG, E. F. 2000.

Bacterial rhodopsin: Evidence for a new type of phototrophy in the sea. *Science,* 289**,** 1902-1906.

BOGARDT, A. H. & HEMMINGSEN, B. B. 1992. ENUMERATION OF PHENANTHRENE-DEGRADING BACTERIA BY AN OVERLAYER TECHNIQUE AND ITS USE IN EVALUATION OF PETROLEUM-CONTAMINATED SITES. *Applied and Environmental Microbiology,* 58**,** 2579-2582.

BOSSERT, I. D. & BARTHA, R. 1986. STRUCTURE-BIODEGRADABILITY RELATIONSHIPS OF POLYCYCLIC AROMATIC-HYDROCARBONS IN SOIL. *Bulletin of Environmental Contamination and Toxicology,* 37**,** 490-495.

BRAGG, J. R., PRINCE, R. C., HARNER, E. J. & ATLAS, R. M. 1994. EFFECTIVENESS OF BIOREMEDIATION FOR THE EXXON-VALDEZ OIL-SPILL. *Nature,* 368**,** 413-418.

BRAZIL, G. M., KENEFICK, L., CALLANAN, M., HARO, A., DELORENZO, V., DOWLING, D. N. & OGARA, F. 1995. CONSTRUCTION OF A RHIZOSPHERE PSEUDOMONAD WITH POTENTIAL TO DEGRADE POLYCHLORINATED-BIPHENYLS AND DETECTION OF BPH GENE-EXPRESSION IN THE RHIZOSPHERE. *Applied and Environmental Microbiology,* 61**,** 1946-1952.

CERNIGLIA, C. E. 1993. Biodegradation of polycyclic aromatic hydrocarbons. *Current Opinion in Biotechnology,* 4**,** 331-338.

CHAPELLE, F. H. 1999. Bioremediation of petroleum hydrocarbon-contaminated ground water: The perspectives of history and hydrology. *Ground Water,* 37**,** 122-132.

CHAUDHRY, G. R. & CHAPALAMADUGU, S. 1991. BIODEGRADATION OF HALOGENATED ORGANIC-COMPOUNDS. *Microbiological Reviews,* 55**,** 59-79.

CHEN, W., BRUHLMANN, F., RICHINS, R. D. & MULCHANDANI, A. 1999. Engineering of improved microbes and enzymes for bioremediation. *Current Opinion in Biotechnology,* 10**,** 137-141.

CORAL, G. & KARAGOZ, S. 2005. Isolation and characterization of phenanthrene-degrading bacteria from a petroleum refinery soil. *Annals of Microbiology,* 55**,** 255-259.

CURIALE, J. A., HARRISON, W. E. & SMITH, G. 1983. STERANE DISTRIBUTION OF SOLID BITUMEN PYROLYZATES - CHANGES WITH BIODEGRADATION OF CRUDE-OIL IN THE OUACHITA MOUNTAINS, OKLAHOMA. *Geochimica Et Cosmochimica Acta,* 47**,** 517-523.

DAVEY, M. E., WOOD, W. A., KEY, R., NAKAMURA, K. & STAHL, D. A. 1993. ISOLATION OF 3 SPECIES OF GEOTOGA AND PETROTOGA - 2 NEW GENERA, REPRESENTING A NEW LINEAGE IN THE BACTERIAL LINE OF DESCENT DISTANTLY RELATED TO THE THERMOTOGALES. *Systematic and Applied Microbiology,* 16**,** 191-200.

DAVIES, J. I. & EVANS, W. C. 1964. OXIDATIVE METABOLISM OF NAPHTHALENE BY SOIL PSEUDOMONADS - RING-FISSION MECHANISM. *Biochemical Journal,* 91**,** 251-&.

DAVYDOVACHARAKHCHYAN, I. A., MILEEVA, A. N., MITYUSHINA, L. L. & BELYAEV, S. S. 1992. ACETOGENIC BACTERIA FROM OIL-FIELDS OF TATARIA AND WESTERN SIBERIA. *Microbiology,* 61**,** 208-216.

DEUTSCHBAUER, A. M., CHIVIAN, D. & ARKIN, A. P. 2006. Genomics for environmental microbiology. *Current Opinion in Biotechnology,* 17**,** 229-235.

DUA, M., SINGH, A., SETHUNATHAN, N. & JOHRI, A. K. 2002. Biotechnology and bioremediation: successes and limitations. *Applied Microbiology and Biotechnology,* 59**,** 143-152.

ERICKSON, L. E. 1997. An overview of research on the beneficial effects of vegetation in contaminated soil. *Bioremediation of Surface and Subsurface Contamination,* 829**,** 30-35.

FERREIRA, A. C., NOBRE, M. F., MOORE, E., RAINEY, F. A., BATTISTA, J. R. & DA COSTA, M. S. 1999. Characterization and radiation resistance of new isolates of Rubrobacter radiotolerans and Rubrobacter xylanophilus. *Extremophiles,* 3**,** 235-238.

FERRER, M., GOLYSHINA, O., BELOQUI, A. & GOLYSHIN, P. N. 2007. Mining enzymes from extreme environments. *Current Opinion in Microbiology,* 10**,** 207-214.

FERRER, M., MARTINEZ-ABARCA, F. & GOLYSHIN, P. N. 2005. Mining genomes and 'metagenomes' for novel catalysts. *Current Opinion in Biotechnology,* 16**,** 588-593.

GALVAO, T. C., MOHN, W. W. & DE LORENZO, V. 2005. Exploring the microbial biodegradation and biotransformation gene pool. *Trends in Biotechnology,* 23**,** 497-506.

GILLESPIE, D. E., BRADY, S. F., BETTERMANN, A. D., CIANCIOTTO, N. P., LILES, M. R., RONDON, M. R., CLARDY, J., GOODMAN, R. M. & HANDELSMAN, J. 2002. Isolation of antibiotics turbomycin A and B from a metagenomic library of soil microbial DNA. *Applied and Environmental Microbiology,* 68**,** 4301-4306.

GOLDSTEIN, R. M., MALLORY, L. M. & ALEXANDER, M. 1985. REASONS FOR POSSIBLE FAILURE OF INOCULATION TO ENHANCE BIODEGRADATION. *Applied and Environmental Microbiology,* 50**,** 977-983.

GRABOWSKI, A., NERCESSIAN, O., FAYOLLE, F., BLANCHET, D. & JEANTHON, C. 2005. Microbial diversity in production waters of a low-temperature biodegraded oil reservoir. *Fems Microbiology Ecology,* 54**,** 427-443.

GRASSIA, G. S., MCLEAN, K. M., GLENAT, P., BAULD, J. & SHEEHY, A. J. 1996. A systematic survey for thermophilic fermentative bacteria and archaea in high temperature petroleum reservoirs. *Fems Microbiology Ecology,* 21**,** 47-58.

GREENE, A. C., PATEL, B. K. C. & SHEEHY, A. J. 1997. Deferribacter thermophilus gen nov, sp nov, a novel thermophilic manganese- and iron-reducing bacterium isolated from a petroleum reservoir. *International Journal of Systematic Bacteriology,* 47**,** 505-509.

GROSSER, R. J., WARSHAWSKY, D. & VESTAL, J. R. 1991. INDIGENOUS AND ENHANCED MINERALIZATION OF PYRENE, BENZO A PYRENE, AND CARBAZOLE IN SOILS. *Applied and Environmental Microbiology,* 57**,** 3462-3469.

HABE, H., OMORI, T. & RM 2003. Genetics of polycyclic aromatic hydrocarbon metabolism in diverse aerobic bacteria. *Bioscience Biotechnology and Biochemistry,* 67**,** 225-243.

HAMME, J. D. V., AJAY, S. & WARD, O. P. 2003. Recent advances in petroleum microbiology. *Microbiology and Molecular Biology Reviews,* 67**,** 503-549.

HEAD, I. M., JONES, D. M. & LARTER, S. R. 2003. Biological activity in the deep subsurface and the origin of heavy oil. *Nature,* 426**,** 344-352.

HEITKAMP, M. A. & CERNIGLIA, C. E. 1987. EFFECTS OF CHEMICAL-STRUCTURE AND EXPOSURE ON THE MICROBIAL-DEGRADATION OF POLYCYCLIC AROMATIC-HYDROCARBONS IN FRESH-WATER AND ESTUARINE ECOSYSTEMS. *Environmental Toxicology and Chemistry,* 6**,** 535-546.

HEITKAMP, M. A. & CERNIGLIA, C. E. 1989. POLYCYCLIC AROMATIC HYDROCARBON DEGRADATION BY A MYCOBACTERIUM SP IN MICROCOSMS CONTAINING SEDIMENT AND WATER FROM A PRISTINE ECOSYSTEM. *Applied and Environmental Microbiology,* 55**,** 1968-1973.

HEITZER, A. & SAYLER, G. S. 1993. MONITORING THE EFFICACY OF BIOREMEDIATION. *Trends in Biotechnology,* 11**,** 334-343.

HERBES, S. E. & SCHWALL, L. R. 1978. MICROBIAL TRANSFORMATION OF POLYCYCLIC AROMATIC-HYDROCARBONS IN PRISTINE AND PETROLEUM-CONTAMINATED SEDIMENTS. *Applied and Environmental Microbiology,* 35**,** 306-316.

HOAGLAND, R. E., ZABLOTOWICZ, R. M. & LOCKE, M. A. 1994. PROPANIL METABOLISM BY RHIZOSPHERE MICROFLORA. *Bioremediation through Rhizosphere Technology,* 563**,** 160-183.

JACOBSEN, C. S. 1997. Plant protection and rhizosphere colonization of barley by seed inoculated herbicide degrading Burkholderia (Pseudomonas) cepacia DBO1(pRO101) in 2,4-D contaminated soil. *Plant and Soil,* 189**,** 139-144.

JOHNSEN, A. R., WICK, L. Y., HARMS, H. & CU 2005. Principles of microbial PAH-degradation in soil. *Environmental Pollution,* 133**,** 71-84.

JUN, W., MA, T., ZHAO, L. X., LV, J. H., LI, G. Q., LIANG, F. L. & LIU, R. L. 2008. PCR-DGGE method for analyzing the bacterial community in a high temperature petroleum reservoir. *World Journal of Microbiology & Biotechnology,* 24**,** 1981-1987.

KADAVY, D. R., PLANTZ, B., SHAW, C. A., MYATT, J., KOKJOHN, T. A. & NICKERSON, K. W. 1999. Microbiology of the oil fly, Helaeomyia petrolei. *Applied and Environmental Microbiology,* 65**,** 1477-1482.

KANALY, R. A. & HARAYAMA, S. 2000. Biodegradation of high-molecular-weight polycyclic aromatic hydrocarbons by bacteria. *Journal of Bacteriology,* 182**,** 2059-2067.

KATO, T., HARUKI, M., IMANAKA, T., MORIKAWA, M. & KANAYA, S. 2001. Isolation and characterization of long-chain-alkane degrading Bacillus thermoleovorans from deep subterranean petroleum reservoirs. *Journal of Bioscience and Bioengineering,* 91**,** 64-70.

KEITH, L. H. & TELLIARD, W. A. 1979. PRIORITY POLLUTANTS I-A PERSPECTIVE VIEW. *Environmental Science & Technology,* 13**,** 416-423.

KHAN, A. G., KUEK, C., CHAUDHRY, T. M., KHOO, C. S. & HAYES, W. J. 2000. Role of plants, mycorrhizae and phytochelators in heavy metal contaminated land remediation. *Chemosphere,* 41**,** 197-207.

KIM, J. S. & CROWLEY, D. E. 2007. Microbial diversity in natural asphalts of the rancho la brea tar pits. *Applied and Environmental Microbiology,* 73**,** 4579-4591.

KUIPER, I., LAGENDIJK, E. L., BLOEMBERG, G. V. & LUGTENBERG, B. J. J. 2004. Rhizoremediation: A beneficial plant-microbe interaction. *Molecular Plant-Microbe Interactions,* 17**,** 6-15.

LHARIDON, S., REYSENBACH, A. L., GLENAT, P., PRIEUR, D. & JEANTHON, C. 1995. HOT SUBTERRANEAN BIOSPHERE IN A CONTINENTAL OIL-RESERVOIR. *Nature,* 377**,** 223-224.

LI, H., YANG, S. Z., MU, B. Z., RONG, Z. F. & ZHANG, J. 2006. Molecular analysis of the bacterial community in a continental high-temperature and water-flooded petroleum reservoir. *Fems Microbiology Letters,* 257**,** 92-98.

LONG, G. M. 1993. CLEAN UP HYDROCARBON CONTAMINATION EFFECTIVELY. *Chemical Engineering Progress,* 89**,** 58-67.

LORENZ, P. & ECK, J. 2005. Metagenomics and industrial applications. *Nature Reviews Microbiology,* 3**,** 510-516.

LU, S. J., WANG, H. Q. & YAO, Z. H. 2006. Isolation and characterization of gasoline-degrading bacteria from gas station leaking-contaminated soils. *Journal of Environmental Sciences-China,* 18**,** 969-972.

MARGESIN, R. & SCHINNER, F. 1999. Biological decontamination of oil spills in cold environments. *Journal of Chemical Technology and Biotechnology,* 74**,** 381-389.

MARGESIN, R. & SCHINNER, F. 2001. Bioremediation (natural attenuation and biostimulation) of diesel-oil-contaminated soil in an alpine glacier skiing area. *Applied and Environmental Microbiology,* 67**,** 3127-3133.

MARTIN, H. G., IVANOVA, N., KUNIN, V., WARNECKE, F., BARRY, K. W., MCHARDY, A. C., YEATES, C., HE, S. M., SALAMOV, A. A., SZETO, E., DALIN, E., PUTNAM, N. H., SHAPIRO, H. J., PANGILINAN, J. L., RIGOUTSOS, I., KYRPIDES, N. C., BLACKALL, L. L., MCMAHON, K. D. & HUGENHOLTZ, P. 2006. Metagenomic analysis of two enhanced biological phosphorus removal (EBPR) sludge communities. *Nature Biotechnology,* 24**,** 1263-1269.

MEAGHER, R. B. 2000. Phytoremediation of toxic elemental and organic pollutants (vol 3, pg 153, 2000). *Current Opinion in Plant Biology,* 3**,** 435-435.

MEGHARAJ, M., JOVCIC, A., BOUL, H. L. & THIELE, J. H. 1997. Recalcitrance of 1,1-dichloro-2,2-bis(p-chlorophenyl)ethylene (DDE) to cometabolic degradation by pure cultures of aerobic and anaerobic bacteria. *Archives of Environmental Contamination and Toxicology,* 33**,** 141-146.

MORGAN, P. & WATKINSON, R. J. 1989. HYDROCARBON DEGRADATION IN SOILS AND METHODS FOR SOIL BIOTREATMENT. *Critical Reviews in Biotechnology,* 8**,** 305-333.

MUELLER, J. G., CHAPMAN, P. J. & PRITCHARD, P. H. 1989. ACTION OF A FLUORANTHENE-UTILIZING BACTERIAL COMMUNITY ON POLYCYCLIC AROMATIC HYDROCARBON COMPONENTS OF CREOSOTE. *Applied and Environmental Microbiology,* 55**,** 3085-3090.

NG, T. K., WEIMER, P. J. & GAWEL, L. J. 1989. POSSIBLE NONANTHROPOGENIC ORIGIN OF 2 METHANOGENIC ISOLATES FROM OIL-PRODUCING WELLS IN THE SAN-MIGUELITO FIELD, VENTURA COUNTY, CALIFORNIA. *Geomicrobiology Journal,* 7**,** 185-192.

OLLIVIER, B. & MAGOT, M. 2005. Petroleum microbiology. *Petroleum microbiology***,** xi + 365 pp.

ORPHAN, V. J., TAYLOR, L. T., HAFENBRADL, D. & DELONG, E. F. 2000. Culture-dependent and culture-independent characterization of microbial assemblages associated with high-temperature petroleum reservoirs. *Applied and Environmental Microbiology,* 66**,** 700-711.

PAUL, D., PANDEY, G., PANDEY, J. & JAIN, R. K. 2005. Accessing microbial diversity for bioremediation and environmental restoration. *Trends in Biotechnology,* 23**,** 135-142.

PIEPER, D. H. & REINEKE, W. 2000. Engineering bacteria for bioremediation. *Current Opinion in Biotechnology,* 11**,** 262-270.

PRITCHARD, P. H., MUELLER, J. G., ROGERS, J. C., KREMER, F. V. & GLASER, J. A. 1992. Oil spill bioremediation: Experiences, lessons and results from the Exxon Valdez oil spill in Alaska. *Biodegradation,* 3**,** 315-335.

RABUS, R., KUBE, M., HEIDER, J., BECK, A., HEITMANN, K., WIDDEL, F. & REINHARDT, R. 2005. The genome sequence of an anaerobic aromatic-degrading denitrifying bacterium, strain EbN1. *Archives of Microbiology,* 183**,** 27-36.

RADWAN, S., SORKHOH, N. & ELNEMR, I. 1995. OIL BIODEGRADATION AROUND ROOTS. *Nature,* 376**,** 302-302.

ROSENBERG, E. 1993. EXPLOITING MICROBIAL-GROWTH ON HYDROCARBONS - NEW MARKETS. *Trends in Biotechnology,* 11**,** 419-424.

ROZANOVA, E. P., SAVVICHEV, A. S., MILLER, Y. M. & IVANOV, M. V. 1997. Microbial processes in a West Siberian oil field flooded with waters containing a complex of organic compounds. *Microbiology,* 66**,** 718-725.

SALANITRO, J. P. 2001. Bioremediation of petroleum hydrocarbons in soil. *Advances in Agronomy, Vol 72,* 72**,** 53-105.

SCHMID, A., DORDICK, J. S., HAUER, B., KIENER, A., WUBBOLTS, M. & WITHOLT, B. 2001. Industrial biocatalysis today and tomorrow. *Nature,* 409**,** 258-268.

SCHMIDT, T. M., DELONG, E. F. & PACE, N. R. 1991. ANALYSIS OF A MARINE PICOPLANKTON COMMUNITY BY 16S RIBOSOMAL-RNA GENE CLONING AND SEQUENCING. *Journal of Bacteriology,* 173**,** 4371-4378.

SCRAGG, A. 2005. Environmental Biotechnology, Second Edition. *Environmental Biotechnology, Second Edition*.

SINGH, S., KANG, S. H., MULCHANDANI, A. & CHEN, W. 2008. Bioremediation: environmental clean-up through pathway engineering. *Current Opinion in Biotechnology,* 19**,** 437-444.

SNYDER, L. A. S., LOMAN, N., PALLEN, M. J. & PENN, C. W. 2009. Next-Generation Sequencing-the Promise and Perils of Charting the Great Microbial Unknown. *Microbial Ecology,* 57**,** 1-3.

SPAIN, J. C. 1995. BIODEGRADATION OF NITROAROMATIC COMPOUNDS. *Annual Review of Microbiology,* 49**,** 523-555.

SPRINGAEL, D. & TOP, E. M. 2004. Horizontal gene transfer and microbial adaptation to xenobiotics: new types of mobile genetic elements and lessons from ecological studies. *Trends in Microbiology,* 12**,** 53-58.

STETTER, K. O., HUBER, R., BLOCHL, E., KURR, M., EDEN, R. D., FIELDER, M., CASH, H. & VANCE, I. 1993. HYPERTHERMOPHILIC ARCHAEA ARE THRIVING IN DEEP NORTH-SEA AND ALASKAN OIL-RESERVOIRS. *Nature,* 365**,** 743-745.

STUART-KEIL, K. G., HOHNSTOCK, A. M., DREES, K. P., HERRICK, J. B., MADSEN, E. L. & LP 1998. Plasmids responsible for horizontal transfer of naphthalene catabolism genes between bacteria at a coal tar contaminated site are homologous to pDTG1 from Pseudomonas putida NCIB 9816-4. *Applied and Environmental Microbiology,* 64**,** 3633-3640.

TALLEY, J. W. 2006. Bioremediation of Relcalcitrant Compounds. *Bioremediation of Relcalcitrant Compounds*.

TARDYJACQUENOD, C., CAUMETTE, P., MATHERON, R., LANAU, C., ARNAULD, O. & MAGOT, M. 1996. Characterization of sulfate-reducing bacteria isolated from oil-field waters. *Canadian Journal of Microbiology,* 42**,** 259-266.

TESTA, S. M. & WINEGARDNER, D. L. 1991. RESTORATION OF PETROLEUM-CONTAMINATED AQUIFERS. *Testa, S. M. and D. L. Winegardner. Restoration of Petroleum-Contaminated Aquifers. Xi+269p. Lewis Publishers, Inc.: Chelsea, Michigan, USA. Illus. Maps***,** XI+269P.

THOMPSON, I. P., VAN DER GAST, C. J., CIRIC, L. & SINGER, A. C. 2005. Bioaugmentation for bioremediation: the challenge of strain selection. *Environmental Microbiology,* 7**,** 909-915.

TIMMIS, K. N. & PIEPER, D. H. 1999. Bacteria designed for bioremediation. *Trends in Biotechnology,* 17**,** 201-204.

TOP, E. M. & SPRINGAEL, D. 2003. The role of mobile genetic elements in bacterial adaptation to xenobiotic organic compounds. *Current Opinion in Biotechnology,* 14**,** 262-269.

TYSON, G. W., CHAPMAN, J., HUGENHOLTZ, P., ALLEN, E. E., RAM, R. J., RICHARDSON, P. M., SOLOVYEV, V. V., RUBIN, E. M., ROKHSAR, D. S. & BANFIELD, J. F. 2004. Community structure and metabolism through reconstruction of microbial genomes from the environment. *Nature,* 428**,** 37-43.

VANCURA, V. & HOVADIK, A. 1965. Root exudates of plants II. Composition of root exudates of some vegetables. *Plant and Soil,* 22**,** 21-32.

VANDERMEER, J. R. 1994. GENETIC ADAPTATION OF BACTERIA TO CHLORINATED AROMATIC-COMPOUNDS. *Fems Microbiology Reviews,* 15**,** 239-249.

VANDERMEER, J. R., DEVOS, W. M., HARAYAMA, S. & ZEHNDER, A. J. B. 1992. MOLECULAR MECHANISMS OF GENETIC ADAPTATION TO XENOBIOTIC COMPOUNDS. *Microbiological Reviews,* 56**,** 677-694.

VENTER, J. C., REMINGTON, K., HEIDELBERG, J. F., HALPERN, A. L., RUSCH, D., EISEN, J. A., WU, D. Y., PAULSEN, I., NELSON, K. E., NELSON, W., FOUTS, D. E., LEVY, S., KNAP, A. H., LOMAS, M. W., NEALSON, K., WHITE, O., PETERSON, J., HOFFMAN, J., PARSONS, R., BADEN-TILLSON, H., PFANNKOCH, C., ROGERS, Y. H. & SMITH, H. O. 2004. Environmental genome shotgun sequencing of the Sargasso Sea. *Science,* 304**,** 66-74.

VOGET, S., LEGGEWIE, C., UESBECK, A., RAASCH, C., JAEGER, K. E. & STREIT, W. R. 2003. Prospecting for novel biocatalysts in a soil metagenome. *Applied and Environmental Microbiology,* 69**,** 6235-6242.

VOORDOUW, G., ARMSTRONG, S. M., REIMER, M. F., FOUTS, B., TELANG, A. J., SHEN, Y. & GEVERTZ, D. 1996. Characterization of 16S rRNA genes from oil field microbial communities indicates the presence of a variety of sulfate-reducing, fermentative, and sulfide-oxidizing bacteria. *Applied and Environmental Microbiology,* 62**,** 1623-1629.

WALTON, B. T. & ANDERSON, T. A. 1990. MICROBIAL-DEGRADATION OF TRICHLOROETHYLENE IN THE RHIZOSPHERE - POTENTIAL APPLICATION TO BIOLOGICAL REMEDIATION OF WASTE SITES. *Applied and Environmental Microbiology,* 56**,** 1012-1016.

WEXLER, M., BOND, P. L., RICHARDSON, D. J. & JOHNSTON, A. W. B. 2005. A wide host-range metagenomic library from a waste water treatment plant yields a novel alcohol/aldehyde dehydrogenase. *Environmental Microbiology,* 7**,** 1917-1926.

WICKRAMANAYAKE, G. B., GAVASKAR, A. R. & KELLEY, M. E. 2000. *Natural Attenuation Considerations and Case Studies: Remediation of Chlorinated and Recalcitrant Compounds,* Columbus, OH, U.S.A., Batelle Press.

YONG, R. N. & MULLIGAN, C. N. 2004. *Natural attenuation of contaminants in soil,* Boca Raton, FL, Lewis Publishers.

YOSHIDA, N., YAGI, K., SATO, D., WATANABE, N., KUROISHI, T., NISHIMOTO, K., YANAGIDA, A., KATSURAGI, T., KANAGAWA, T., KURANE, R. & TANI, Y. 2005. Bacterial communities in petroleum oil in stockpiles. *Journal of Bioscience and Bioengineering,* 99**,** 143-149.

CHAPTER TWO:

CHARACTERIZATION OF MICROORGANISMS AND ENZYMES ABLE TO DEGRADE PETROLEUM HYDROCARBONS FROM THE RANCHO LA BREA TAR PITS

ABSTRACT

Recent studies on the biodiversity of asphalt deposits at the Rancho La Brea Tar Pits in Los Angeles, California have revealed the existence of several hundred new species of bacteria and gene sequences encoding putative novel degradative enzymes. The presence of fossilized extinct animal remains in the La Brea Tar Pits has led to estimations that these natural asphalt seeps have existed for at least 40,000 years. These deposits consist of petroleum that has been degraded to the extent that the remaining material is comprised mainly of asphalts and heavy oils, which have saturated into the soil matrix. Since petroleum hydrocarbons are both a target and a product of microbial metabolism, the role of microorganisms that inhabit this and other similar environments is directly relevant to development of technology for bioremediation, biotransformation of petroleum hydrocarbons, and microbial enhanced oil recovery for extracting and refining heavy oil.

In this research, both culture-dependent and culture-independent methods were used to characterize microorganisms and consortia from the La Brea Tar Pits, which are able to degrade a variety of polycyclic aromatic hydrocarbons (PAHs) and BTEX (benzene, toluene, ethyl benzene and xylene). Initial studies applied PCR-DGGE methods to identify the microbial consortia that degrade selected petroleum hydrocarbons during enrichment cultures on pure compounds or mixtures of substances found in petroleum. This study revealed a single

Pseudomonas sp. that may be able to degrade multiple PAHs and biphenyl. In addition, PCR based techniques identified naphthalene dioxygenase and catechol 2,3-dioxygenase from *Pseudomonas stutzeri* that appears to function as a major degradative enzyme in this system and verifies the presence of this organism. In addition, sequencing of the dominant bands from the DGGE analysis revealed the presence of both known petroleum degraders and other novel organisms.

INTRODUCTION

The role of microorganisms in petroleum environments has been an area of interest for many biotechnology based applications. Petroleum is a complex mixture of heavy to light hydrocarbons and many other organic compounds, including organometallo constituents, which represent the end products from anaerobic degradation of organic materials over a very long time. Under conditions of high temperature and pressure, the degraded organic material has been converted to natural gas, liquid crude oil, shale oil, tars and bitumen (Scragg, 2005). The majority of the compounds in crude oil and natural asphalts are hydrocarbons, which can range in molecular weight from methane to the high molecular weight bitumen. Since petroleum hydrocarbons are recognized as substrates supporting microbial growth, they are both a target and a product of microbial metabolism (Hamme et al., 2003). Biodegraded oils represent a significant fraction of the petroleum in conventional oil reserves and the use of petroleum-degrading organisms for environmental cleanup and production of oil based products are the central themes in petroleum microbiology (Atlas, 1981, Hamme et al., 2003).

In 1926, Edison Bastin and coworkers submitted the first report of active microbial communities in a petroleum reservoir (Bastin et al., 1926) and since then a wide variety of bacteria have been identified using both culture-dependent and culture-independent methods. Culture dependent methods include isolation and cultivation of bacteria on solid media, most-probable-number (MPN) liquid assays and Biolog substrate utilization plates. Microorganisms isolated from these sites have been extensively tested for bioremediation applications as they have the ability to degrade petroleum hydrocarbons in both the laboratory and the environment (Ollivier and Magot, 2005). Alternatively, culture independent methods include phospholipid

fatty acid analysis, fluorescence in situ hybridization and a variety of PCR-based approaches such as denaturing gradient gel electrophoresis (DGGE) and 16S rRNA gene based surveys. Initial reports of underground reservoirs using molecular approaches suggest that the majority of microorganisms inhabiting these environments are new species that represent a rich pool of novel genetic diversity (Hamme et al., 2003).

While much research has examined the microbiology of lighter grade petroleum and the microbial communities that are associated with oil spills on land, to date there has been very little research on heavy oils and asphalts. Recently, our research group has studied the biodiversity of the Rancho La Brea Tar Pits, which are Pleistocene natural asphalt deposits located in Hancock Park of downtown Los Angeles, California. The presence of fossilized remains of extinct animals at Rancho La Brea has led to estimations that these asphalt seeps are at least 40,000 years old (Akersten et al., 1983). These deposits consist of petroleum that has been degraded to the extent that they are comprised mostly of asphalt and heavy oil (Kim and Crowley, 2007). In 2007, Kim and Crowley conducted a 16S rRNA gene survey of this site, which included detailed sequence information on both culturable and non-culturable organisms. This study revealed several hundred new species of bacteria and archaea (Kim and Crowley, 2007). In addition, DNA sequencing of petroleum hydrocarbon degrading genes from the tar pits show at least three new groups of enzymes that have not been previously described (Kim and Crowley, 2007).

As microbes are able to use petroleum as a carbon and energy source, one of the most practical applications of the La Brea Tar Pits' microorganisms and their hydrocarbon degrading enzymes is for bioremediation, in which living organisms are used to degrade or transform

hazardous organic compounds to a relatively non toxic state (Chapelle, 1999, Dua et al., 2002, Paul et al., 2005, Salanitro, 2001). Bioremediation is often considered a cost-effective and environmentally friendly alternative approach in comparison to the conventional methods of remediation, such as excavation and incineration, which are very costly and can result in the generation of toxic air emissions (Kuiper et al., 2004). In addition, bioremediation techniques can be applied in situ without the removal of the contaminated soil, thus not disturbing the environment. The use of bioremediation for environmental decontamination has been growing, and has attracted public interest with its first major success demonstrated during the cleanup of petroleum hydrocarbons from the Exxon Valdez oil spill (Bragg et al., 1994, Pritchard et al., 1992). Currently, microorganisms with the ability to degrade various pollutants (e.g. polycyclic aromatic hydrocarbons, nitroaromatics, polychlorinated biphenyls and oil components) have been isolated in hopes of exploiting their metabolic potential for the remediation of contaminated sites (Spain, 1995, Dua et al., 2002).

Contamination caused by petroleum hydrocarbons has gathered much attention due to concerns stemming from large scale releases into the environment (Chapelle, 1999). Polycyclic aromatic hydrocarbons (PAHs) and BTEX (benzene, toluene, ethylbenzene and xylene) compounds are environmental pollutants commonly found in petroleum contaminated sites. PAHs have carcinogenic and mutagenic properties and it is the heavier fractions that pose the greatest long term recalcitrance (Talley, 2006). The lighter compounds (i.e. BTEX) are able to leach into groundwater basins and contaminate freshwater supplies (Testa and Winegardner, 1991). Each year, an average of approximately 1,680,000 gallons of crude oil are spilled on land due to pipeline failures and more than 200,000 underground storage tanks in the US have

leaked gasoline and other fuels into the soil, sediments and groundwater aquifers (Bedient et al., 1994), making petroleum and its derived fuels the most ubiquitous organic pollutants around the world (Salanitro, 2001). Due to the amount of petroleum hydrocarbons contaminating the environment, improved bioremediation technologies are necessary.

PAHs represent a large group of soil pollutants and a number of them have been prioritized for environmental cleanup. Once these pollutants enter the soil, they are trapped into soil pores and immobilized by adsorption to the soil particles and diffusion into the soil matrix (Aprill and Sims, 1990). Biological degradation remains as the most significant means of decomposition as the ability of microbes to degrade PAHs has been noted for many years. The first demonstration of the metabolic pathway and the enzymatic reactions that are involved in the mineralization of naphthalene was by Davies and Evans (1964). Degradation of PAHs by microbes has since been studied extensively for a wide variety of bacteria, fungi, and algae (Cerniglia, 1993, Davies and Evans, 1964, Kuiper et al., 2004). The bacterial PAH catabolic genes often are located on large plasmids, together with regulatory genes. These catabolic plasmids and other mobile genetic elements contribute to the adaptation of the indigenous microbes toward PAH pollutants and other xenobiotics, through horizontal gene transfer (Springael and Top, 2004, Stuart-Keil et al., 1998, Top and Springael, 2003, Vandermeer et al., 1992, Vandermeer, 1994).

The biochemical pathways for the biodegradation of aromatic compounds have been well described (Cerniglia, 1993, Johnsen et al., 2005). Under aerobic conditions, the initial step for destruction of a PAHs, involves oxidation of one of the aromatic rings to a dihydrodiol by a multi-component enzyme system (Kanaly and Harayama, 2000). The dihydrodiols may then be

processed through either an ortho cleavage pathway or a meta cleavage pathway, leading to central intermediates such as protocatechuates and catechols, which are further converted to tricarboxylic acid cycle intermediates (Kuiper et al., 2004). As in the case of a number of PAHs, dioxygenase enzymes are usually the first biocatalysts used in aerobic petroleum hydrocarbon degradation (Hamme et al., 2003). Naphthalene dioxygenases introduce an oxygen molecule at the 1,2-position of the aromatic nucleus, producing dihydrodiol compounds, which then facilitates further degradation of the compound by subsequent enzymes in the pathway (Habe et al., 2003). Naphthalene dioxygenases are able to attack a variety of petroleum hydrocarbons (e.g. biphenyl, phenanthrene and anthracene), and thus are one of the key classes of enzymes involved in the degradation of these compounds (Habe et al., 2003, Hamme et al., 2003).

In this study, both culture-based and culture-independent techniques were used to investigate the petroleum degrading community of the La Brea Tar Pits. We exposed the La Brea Tar Pit Bacteria to a variety of PAHs and BTEX compounds and used PCR-DGGE analysis to identify the microbial consortia that degrade petroleum hydrocarbons. The petroleum degrading community was further investigated by PCR and sequencing of naphthalene dioxygenase genes and another gene, catechol 2,3-dioxygenase, that is involved in the lower degradation pathway. In addition, we used culture based techniques to examine and compare the culturable and unculturable diversity of the petroleum degrading community.

MATERIALS AND METHODS

Exposure of La Brea Tar Pit bacteria to petroleum hydrocarbons. Samples from Pit 101 of the Rancho La Brea Tar Pits were collected as previously described (Kim and Crowley, 2007). Five grams of asphalt-soil mixtures were added to 50 ml of mineral salts medium (MSM: 1.361 g/l of KH_2PO_4, 0.661 g/l of $(NH_4)_2SO_4$, 0.426 g/l of Na_2HPO_4, 0.123g/l of $MgSO_4 \cdot 7H_2O$ and 10ml/l of a trace mineral solution consisting of 252.9 mg/l of $CaSO_4 \cdot 2H_2O$, 200 mg/l of $FeSO_4 \cdot 7H_2O$, 20 mg/l of $MnSO_4 \cdot H_2O$, 10 mg/l of $Na_2MoO_4 \cdot 2H_2O$, 31.3 mg/l of $CuSO_4 \cdot 5H_2O$, 10 mg/l of $CoSO_4 \cdot 7H_2O$ and 5 mg/l H_3BO_3) supplemented with petroleum hydrocarbons in 250-ml Erlenmeyer flasks. Enrichment cultures for PAHs were produced using individual PAHs as sole carbon sources. PAHs were dissolved in acetone to prepare stock solutions, and were added to 50 ml MSM at the following concentrations (mg l^{-1}): 100 biphenyl (BPH), 200 naphthalene (NAH), 100 phenanthrene (PHE), 100 chrysene (CHR), and 100 pyrene (PYR). An additional treatment consisted of a mixture of PAHs containing (mg l^{-1}) 50 NAH, 25 PHE, 25 CHR and 25 PYR. The acetone carrier was evaporated from the samples for 24 h, after which the media were inoculated with approximately 100 mg of asphalt-soil taken from Pit 91. Also, 200 PPM benzene (BEN) and a mixture of 50 PPM of benzene, toluene, ethyl benzene and xylenes were also added separately to 50 ml of MSM in an Erlenmeyer flask. In addition, a control flask, which contained only the 5 ml sample with no other carbon source, was prepared. All 8 treatments including the control were incubated at 28°C with shaking (i.e. 150 RPM). The treatments containing BEN, BPH, NAH and the control were incubated for one week, while the treatments containing PHE, CHR, PYR and the PAH combination were incubated for two weeks. After each respective

incubation period, the DNA was extracted from each flask using the Fast DNA Soil Spin Kit (MP Bio).

Comparison and identification of microbial consortia exposed to different petroleum hydrocarbons using Denaturing Gradient Gel Electrophoresis (DGGE). After the DNA extraction, each DNA sample served as a PCR template. Targeted DNA sequences were amplified using a PTC-200 Thermal cycler (MJ Research). PCR amplification of the Eubacterial V3 region of the 16S rRNA gene was performed using primers PRBA338F, with a 5' GC clamp and PRUN518R. The PCR reaction conditions were as follows; 1X concentration of GoTaq Green PCR Master Mix (Promega), 0.5 µm of both PRBA338FGC and PRUN518R primers and at least 20 ng of template DNA in a 50 µl reaction, using the following method: 95˚C for 5 min then 30 cycles of 94˚C for 30 sec, 55˚C for 30 sec, 72˚C for 1 min, followed by a final extension step for 10 min at 72˚C. The PCR products were separated on an acrylamide gel containing a 40 to 60% linear denaturing gradient, using a DCode TM universal mutation detection system (Bio-Rad Laboratories, Hercules, Calif). The gels were then stained with ethidium bromide and imaged using a computerized digital image capture and analysis system (Biorad Quantity One).

To identify individual members of each microbial consortium, bands from the DGGE gel were cut and DNA was extracted from the bands by adding 20 µl of sterile water and subjecting the gel to several freeze thaw cycles. The extracted DNA was then amplified using the same PCR conditions as above. The PCR products were purified using a PCR purification kit (Quiagen) and directly sequenced at the UC Riverside genomics core facility. Individual organisms were identified and taxonomically classified using the BLAST and RDP databases. Sequence alignments were performed using Clustal X and phylogenetic tree analysis was performed using MEGA 4.

PCR amplification of naphthalene dioxygenase and catechol 2,3-dioxygenase genes from microbial consortia exposed to polycyclic aromatic hydrocarbons. DNA samples from microbial consortia that were exposed to NAH, PHE, CHR, PYR and a combination thereof was used as a DNA template for PCR amplification. Naphthalene dioxygenase genes were amplified using the primers NAH-F and NAH-R as described in Baldwin et al (Baldwin et al., 2003). Each PCR reaction contained the following: 1X concentration of GoTaq Green PCR Master Mix (Promega), 0.5 µm of both NAH-F and NAH-R primers and at least 20 ng of template DNA in a 50 µl reaction, using the following method: 95˚C for 5 min then 30 cycles of 94˚C for 30 sec, 47˚C for 30 sec, 72˚C for 1 min, followed by a final extension step for 10 min at 72˚C. Catechol 2,3-dioxygenase genes were amplified using the primers 23DOF and 23DOR as described in Wikstrom et al. (Wikstrom et al., 1996). Each PCR reaction contained the following: 1X concentration of GoTaq Green PCR Master Mix (Promega), 0.5 µm of both 23DOF and 23DOR primers and at least 20 ng of template DNA in a 50 µl reaction, using the following method: 95˚C for 5 min then 30 cycles of 94˚C for 30 sec, 50˚C for 30 sec, 72˚C for 1 min, followed by a final extension step for 10 min at 72˚C. The resulting PCR products generated from each primer set were purified using a PCR purification kit (Quiagen) and then ligated into P-GemT vectors (Promega). The vectors and ligated inserts were cloned into JM109 *E. coli* competent cells. The inserts were sequenced at the UC Riverside Genomics Core and identified by using the BLAST database. Phylogenetic analysis was performed using Clustal X and MEGA 4.

Isolation and Identification of PAH degrading bacteria. Erlenmeyer flasks with 1 g of the asphalt-soil mixtures from Pit 101 were added to 50 ml of mineral salts medium (MSM: 1.361 g/l of KH_2PO_4, 0.661 g/l of $(NH_4)_2SO_4$, 0.426 g/l of Na_2HPO_4, 0.123g/l of $MgSO_4 \cdot 7H_2O$ and 10ml/l of

a trace mineral solution consisting of 252.9 mg/l of $CaSO_4 \cdot 2H_2O$, 200 mg/l of $FeSO_4 \cdot 7H_2O$, 20 mg/l of $MnSO_4 \cdot H_2O$, 10 mg/l of $Na_2MoO_4 \cdot 2H_2O$, 31.3 mg/l of $CuSO_4 \cdot 5H_2O$, 10 mg/l of $CoSO_4 \cdot 7H_2O$ and 5 mg/l H_3BO_3) supplemented with 100 PPM of phenanthrene and was incubated at 28 ˚C for 10 days with shaking at 150 RPM. After 10 days, 0.5 ml of the culture was added to 50 ml of fresh PHE MSM media and incubated for another 10 days. After two successive transfers, the cultures were plated on MSM media and 1.5% agar with phenanthrene as the sole carbon source, using a plate sublimination method (Alley and Brown, 2000). Colonies that grew on phenanthrene, as evidenced by the presence of clearing zones, were picked and re-grown in PHE MSM media. Cultures were grown for 1-3 days and the cells were harvested for DNA extraction using a simple boil lysis method. Genomic isolate DNA was used as a PCR template and primers 27F and 1492R was used to amplify a 1.5 Kb fragment of the 16S rDNA. The PCR products were purified using a PCR purification kit (Quiagen) and was directly sequenced at the UC Riverside Genomics Core. Each isolate was identified and classified using the BLAST and RDP databases.

RESULTS

Denaturing gradient gel electrophoresis analysis of La Brea Tar Pit consortia exposed to various petroleum hydrocarbons. Denaturing gradient gel electrophoresis was used to investigate the microbial consortia that degrade different petroleum hydrocarbons. 16S rDNA PCR products (177 bp) amplified from the Pit 101 asphalt-soil mixtures that were exposed to benzene, biphenyl, naphthalene, phenanthrene, chrysene, pyrene, a combination of the PAHs, and a combination of BTEX compounds was analyzed using a DGGE gel analysis (Figure 2.1) coupled with band sequencing of selected organisms and phylogenetic analysis (Figure 2.2 and Table 2.1). The gel also included a control (lane 1), in which the microbial diversity was shown to have a different community fingerprint when compared to samples exposed to petroleum hydrocarbons. In contrast, there were distinct similarities and differences between the samples that were exposed to both the PAHs and BTEX compounds, which revealed that there were shifts in the species composition of the degrader communities that reflected specific enrichment on the different test substrates. The presence of some common bands across all lanes representing the different communities further suggested that some similar organisms were present in all of the degrader communities. This would suggest that such microorganisms either have the ability to degrade all of the hydrocarbons tested, or used a common metabolite produced as an intermediate following catabolism of the primary substrates.

In total, 25 bands were sequenced from the DGGE gel, of which 16 sequences were most closely related to organisms that have previously been shown to be able to degrade petroleum compounds (Table 2.1). Thirteen of the sequences had low homology to other known sequences (97% or lower) and 5 sequences had high percent similarities (97% or higher) to

those of uncultured organisms, indicating that most of these species are previously undescribed. The control lane contains two sequences most closely related to 16S rDNA gene sequences belonging to *Bacillus sp.* (bands 1 and 2).

Prior research has shown that various strains of this genus are able to degrade crude oil and can be isolated from alkaliphilic and halophilic hydrocarbon contaminated environments (Table 2.1). The most common bacteria represented in the gel were pseudomonads. A previously undescribed *Pseudomonas sp.* was found in the benzene exposed sample. In addition, the band located at the same lane position in the DGGE gel lanes from the biphenyl, naphthalene, phenanthrene, chrysene, pyrene and PAH treatments, was identified as a pseudomonad that is closely related to *Pseudomonas stutzeri*, a well characterized petroleum degrader.

Other microorganisms that occurred in more than one degrader community enriched on the different substrates included one similar to an uncultured *Oxalobacter sp.,* which was present in the chrysene and pyrene treatments, an uncultured *Skermanella sp.* present in phenanthrene and pyrene treatments and a bacterium similar to a previously reported uncultured Comamonadacea, which was present in the 16S rRNA gene profiles generated from the phenanthrene, chrysene and pyrene treatments. All of these aforementioned organisms are closely related to previously described species that are able to degrade petroleum hydrocarbons. Other organisms that were identified here belonged to diverse genera, and included Xanthomonas, Naxibacter, Luteimonas, Massilia and Paracoccus. Half of the latter taxa contain previously described species that were isolated from soil contaminated with petroleum hydrocarbons (table 2.1).

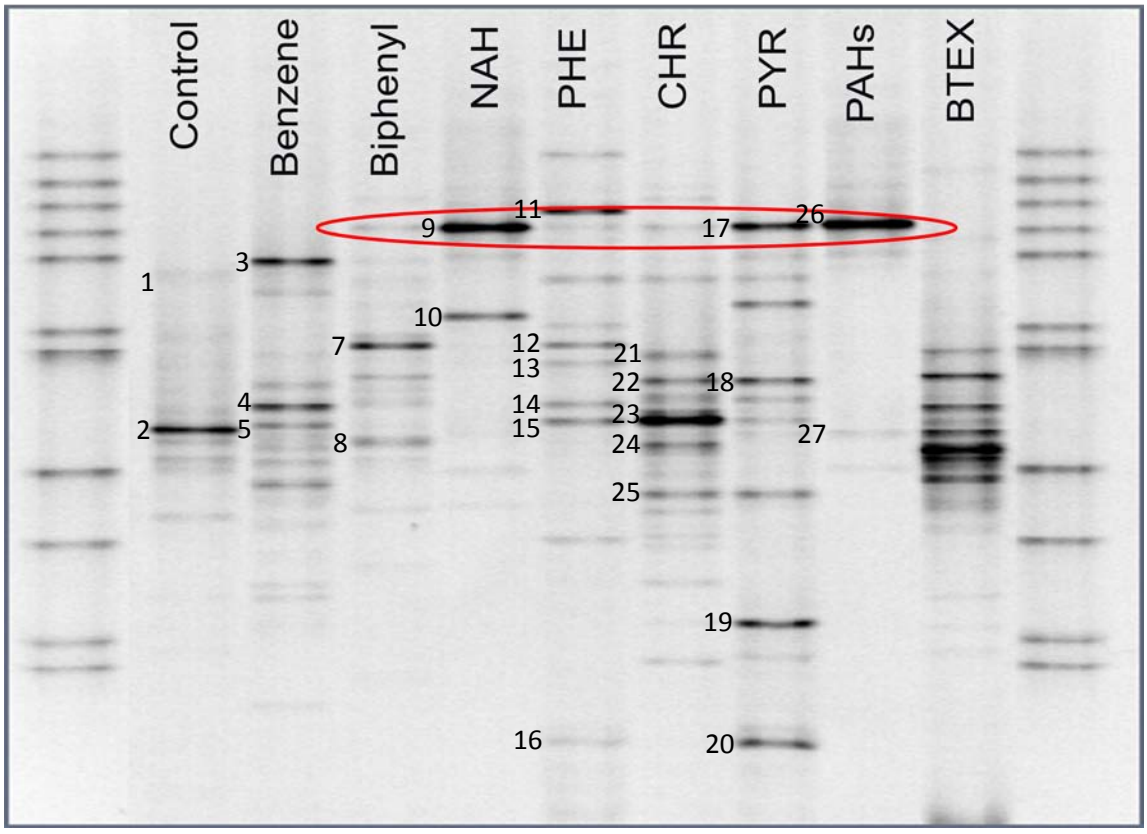

Figure 2.1. DGGE community analysis of La Brea Tar Pit microbial communities exposed to different petroleum hydrocarbons. Lanes from left to right illustrate the control, benzene, biphenyl, naphthalene, phenanthrene, chrysene, pyrene, and a combination of PAHs and BTEX compound exposures. There is an evident community shift when exposed to the different petroleum hydrocarbons. Oval indicates same band present in 6 different treatments. Numbers refer to bands that were sequenced and reported in table 2.1.

Bands	BLAST Results	% Homology	Types of Hydrocarbons Degraded	Reference
|CON-DGGE01|	*Bacillus sp.*	95%	Crude oil	(Meintanis et al., 2006)
|CON-DGGE02|	*Bacillus sp.*	97%	Alkaliphilic and halophilic hydrocarbons	(Al-Awadhi et al., 2007)
|BEN-DGGE03|	*Pseudomonas sp.*	97%	MTBE	(Lalevic et al., unpublished)
|BEN-DGGE04|	Uncultured bacterium	95%	N/A	(Ishii et al., 2009)
|BEN-DGGE05|	Uncultured bacterium	95%	N/A	(Qin et al., unpublished)
|BPH-DGGE07|	*Xanthomonas sp.*	95%	Petroleum hydrocarbons	(Milton et al., unpublished)
|BPH-DGGE08|	Uncultured bacterium	98%	Crude oil	(Paixao et al., unpublished)
|NAH-DGGE09|	*Pseudomonas stutzeri*	93%	Crude oil	(Gao et al., unpublished)
|NAH-DGGE10|	Uncultured bacterium	98%	N/A	(Ding et al., unpublished)
|PHE-DGGE11|	*Naxibacter sp.*	95%	N/A	(Weon et al., unpublished)
|PHE-DGGE12|	Uncultured *Luteimonas sp.*	86%	Petroleum hydrocarbons	(Milton et al., unpublished)
|PHE-DGGE13|	Uncultured *Massilia sp.*	85%	N/A	(Jangid et al., 2008)
|PHE-DGGE14|	Uncultured alpha proteobacterium	98%	Phenanthrene	(Ibarrolaza et al., unpublished)
|PHE-DGGE15|	Uncultured Comamondaceae bacterium	91%	Crude oil	(Pereira et al., unpublished)
|PHE-DGGE16|	Uncultured *Skermanella sp.*	100%	Polycyclic aromatic hydrocarbons	(Llado et al., 2009)
|PYR-DGGE17|	*Pseudomonas stutzeri*	97%	Crude oil	(Gao et al., unpublished)
|PYR-DGGE18|	Uncultured *Oxalobacter sp.*	97%	Toluene	(Winderl et al., 2008)
|PYR-DGGE19|	*Burkholderiales sp.*	98%	N/A	(Brooks et al., unpublished)
|PYR-DGGE20|	Uncultured *Skermanella sp.*	98%	N/A	(Ceha-Navarro et al., unpublished)
|CHR-DGGE22|	Uncultured Oxalobacteraceae	91%	N/A	(Green et al., 2007)
|CHR-DGGE23|	Uncultured Comamonadacea bacterium	93%	Crude oil	(Pereira et al., unpublished)
|CHR-DGGE24|	*Bacillus drentensis*	94%	MTBE, ETBE and TAME	(Purswani et al., unpublished)
|CHR-DGGE25|	Uncultured *Paracoccus sp.*	97%	acetochlor	(Ni et al., unpublished)
|PAH-DGGE26|	*Pseudomonas stutzeri*	99%	Crude oil	(Gao et al., unpublished)
|PAH-DGGE27|	Uncultured bacterium	95%	N/A	(Larouche et al., unpublished)

Table 2.1. Table of BLAST results for selected DGGE bands from microbial consortia exposed to different PAHs. Many of the identified organisms were closely related to those that are able to degrade other petroleum hydrocarbons or were isolated from contaminated sites. Band names are denoted by which hydrocarbon the organism was found, (CON=no compound, NAH=naphthalene, PHE=phenanthrene, PYR=pyrene, CHR=chrysene and PAH=combination) and the band number as reported in figure 2.1.

Characterization of petroleum hydrocarbon degrading genes found in La Brea Tar Pit microorganisms. Petroleum degrading genes from the La Brea Tar Pits were analyzed by sequencing PCR products, which were amplified using primers that target naphthalene dioxygenase (377 bp) and catechol 2,3- dioxygenase genes (721 bp). A total of 26 naphthalene dioxygenase genes were sequenced from the NAH, PHE, PYR and PAH treatments. Of these, 7 of the sequences were found to be very closely related, all of which came from samples enriched on pyrene. The other 19 sequences were also closely related to each other, but were distinctly grouped from those represented in the pyrene treatment. Both groups of naphthalene dioxygenases have previously been shown to be carried by isolates of *Pseudomonas stutzeri* and *Pseudomonas balearica*.

A total of 16 clones carrying 2,3-catechol dioxygenase genes were sequenced from the PHE, CHR, PYR and PAH treatments. All but one sequence (PHE-2) were essentially identical and most similar to a gene carried by a previously described *Pseudomonas sp*. The remaining 2,3-catechol dioxygenase clone has been previously described for *Sphingomonas sp.* and is phylogenetically distinct from the other gene sequences.

Isolation and identification of petroleum degrading microorganisms from the La Brea Tar Pits. Eighteen organisms were isolated from an enrichment culture using phenanthrene as the sole carbon source. All isolates were confirmed to degrade phenanthrene by generating clearing zones on plates containing sublimed phenanthrene crystals on the agar surface. All 18 isolates were identified from their 16S rDNA gene sequences by comparisons to both the NCBI and RDP databases. Of these, 7 of the 18 had low homology to other organisms in the NCBI database or

are as of yet unidentified. Another 13 of the 18 isolates were most closely related to other organisms that have previously been shown to be able to degrade petroleum hydrocarbons.

Sequence analysis indicated 5 of the 18 isolates were unique and the other 13 represented the same species. The latter were most closely related to *Ralstonia sp.*, Cupriavidus sp., and Wautersia sp., all of which are bacterial genera that have previously been shown to include species that degrade petroleum hydrocarbon. Isolates PHD-ISO07 and PHE-ISO06 appeared to have no close relatives and maybe represent new genera of families of bacteria. Isolates PHE-ISO02 and PHE-ISO09 were also distinct species, but could be assigned as closely related to *Mesorhizobium sp.*, and *Bordetella sp.*.

DISCUSSION

As illustrated by the DGGE analysis, selective enrichment of microbial communities from the La Brea Tar Pit samples on selected petroleum hydrocarbons resulted in distinct 16S rRNA gene profiles. The community fingerprints of all of the exposed samples showed a dramatic shift in the community structure when compared to the control lane, which contained only the original asphalt-soil mixture. The control sample had only one relatively dominant species (control lane, band 1), which was identified as a *Bacillus sp.*, most similar to a previously isolated *Bacillus* from an alkaliphilic and halophilic, hydrocarbon contaminated site (Al-Awadhi et al., 2007). This organism did not appear to be present in any of the petroleum hydrocarbon degrading consortia, but because Pit 101 is slightly alkaline, along with its relatively high salt content, this *Bacillus sp.* may be an important petroleum hydrocarbon degrader in the original tar sample.

The DGGE analysis also revealed many shared bands in the community fingerprints of the microbial consortia that degrade structurally related petroleum hydrocarbons. The fingerprint of the benzene exposed samples showed a similar community structure to that of the community grown on the BTEX substrate, of which benzene is one of the component compounds. In addition, the samples exposed to phenanthrene, chrysene and pyrene also appear to have similar community fingerprints, but differed from the community enriched on naphthalene. This suggested similar communities were selected on PAH containing three or more fused aromatic rings. Interestingly, the sample exposed to the combination of PAHs including naphthalene showed a very different 16S rRNA gene profile. It has previously been shown that some bacteria are able to grow on high molecular weight PAH using enzymes that

cometabolize PAH following their induction during growth on naphthalene (Safinowski et al., 2006).

Sequence identification of the bands representing the degrader communities on different substrates revealed some individual species that were common to the several of the enrichment cultures. The most noteworthy from this analysis was the presence of a DGGE band in the biphenyl, naphthalene, chrysene, pyrene and the PAH combination. Bands 9, 17 and 26, which were located in the same position throughout all of the previously mentioned treatments, were all sequenced. Band 26 and 17 displayed a 99% and 97% similarity to a *Pseudomonas stutzeri*, a well known petroleum hydrocarbon degrader (Table 2.1). Band 9 was found to be only 93% similar to a *Pseudomonas sp.*, which may be due to the poor sequencing quality of that particular band, but which was still most closely related to the Pseudomonas genus. Genomovars of *P. stutzeri* are versatile degraders of many organic compounds and its ability to degrade petroleum hydrocarbons is widely known (Alquati et al., 2005, Brazil et al., 1995, Chaudhry and Chapalamadugu, 1991, Kang et al., 2006, Ma et al., 2006, Taylor and Janssen, 2005). More importantly, the presence of the band in five of the different petroleum hydrocarbon exposures suggests that this organism can degrade each compound or a common metabolite and may be considered for bioremediation technologies.

Although *Pseudomonas stutzeri* was present in 5 of the different enrichment cultures, its dominance differs in the different consortia, based on DGGE band intensity. *P. stutzeri* appear to be the main degrader in the naphthalene, pyrene and the PAH combination treatments, which may be due to the structural similarities of pyrene and naphthalene. The dominating presence in the combination treatment suggests that the *Pseudomonas stutzeri* may

be degrading the naphthalene first, before the other compounds. Other bands represented species common to those that grow on PAHs containing three or four aromatic ring PAHs. The fact that they are present together in these treatments suggests that they may work together to degrade heavier PAHs and the metabolites that arise. There was a band present in the phenanthrene, chrysene and pyrene treatments only, which was identified as an unknown member of the Comamondaceae family (Table 2.1). This organism may be novel and appears to have the ability to degrade three and four ringed PAHs. Bands 18 and 22, found in chrysene and pyrene exposures respectively, suggest that this organism degrades four ringed PAHs. The band's sequence identifies the degrader as an unidentified Oxalobacteraceae member. Band 18 specifically, is most similar to an isolate of *Oxalobacter sp.*, previously shown to degrade toluene (Winderl et al., 2008). Bands 16 and 20, appeared to represent an unknown *Skermanella sp., a species previously shown to* degrade crude oil and PAHs (Llado et al., 2009). The fact that this organism is present in the pyrene and phenanthrene enrichment cultures, but not in the chrysene enrichment suggests that this particular organism may degrade a common metabolite that is produced during phenanthrene and pyrene degradation, but not chrysene. Overall, the distribution of organisms that degrade PAHs suggest that there are specific members that may act cooperatively within the consortia to help degrade the heavier PAHs by utilizing certain metabolites throughout the degradation pathway. In terms of bioremediation, strategies which use an entire consortium to clean up different compounds may be more successful than applying few organisms with wide degradative capabilities.

Figure 2.2. Phylogenetic tree of DGGE band sequences from Pit 101 bacteria exposed to different PAHs **(in bold)**. Sequences of related organisms were taken from GenBank and shown along with accession number. The evolutionary history was inferred using the Neighbor-Joining method. The optimal tree with the sum of branch length = 2.99393627 is shown. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) is shown next to the branches.

To further characterize the petroleum hydrocarbon degrading community, we amplified several degradative genes from the PAH exposed samples to help assess the diversity of the petroleum degrading enzymes. Many primer sets that targeted different petroleum hydrocarbons were tested (data not shown), however only two (naphthalene 1,2 dioxygenase and catechol 2,3-dioxygenase) provided legitimate PCR products. A total of 26 naphthalene dioxygenase clones were sequenced and phylogenetic analysis revealed two distinct groups of the gene. Nineteen of the clones, which came from the naphthalene, phenanthrene and PAH combination treatments all grouped together and are most closely related to a naphthalene dioxygenase gene from *Pseudomonas stutzeri* found in bilge waste (Olivera et al., 2003) and *Pseudmonas balearica* (Figure 2.3). Another 7 clones grouped together, but separately from the other 19. These clones came from the pyrene treatment only and were highly similar to a naphthalene dioxygenase from *Pseudomonas stutzeri* (figure 2.3). The presence of the two groups of genes implies that there may be two different *Pseudomonas spp.*, which contain different naphthalene dioxygenases and depending on the compound being degraded, that organism will be selected. However, these two different types of dioxygenases also have been shown to occur in the same organism (Ferrero et al., 2002).

Clones that were sequenced using the catechol 2,3-dioxygenase primer set showed relatively similar results in which most of the clones appeared to represent a couple of enzymes. From the 16 clones that were sequenced, 15 were very similar to a catechol 2,3 dioxygenase gene from members of the *Pseudomonas sp.* The sequences show that they are similar to a gene from *Pseudomonas aeruginosa* (figure 2.4), which is able to degrade PAHs and also another unknown *Pseudomonas sp.* that was found in a polluted environment (Junca and Pieper, 2004).

Figure 2.3. Phylogenetic tree of naphthalene dioxygenase genes from Pit 101 **(in bold)** taken from different PAH sample exposures and other similar sequences taken from GenBank along with accession numbers. Clones from different PAH exposures are denoted in clone name, NAH=naphthalene, PHE=phenanthrene, PYR=pyrene and PAH=combination of PAHs. The evolutionary history was inferred using the Neighbor-Joining method. The optimal tree with the sum of branch length = 1.66683257 is shown. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree.

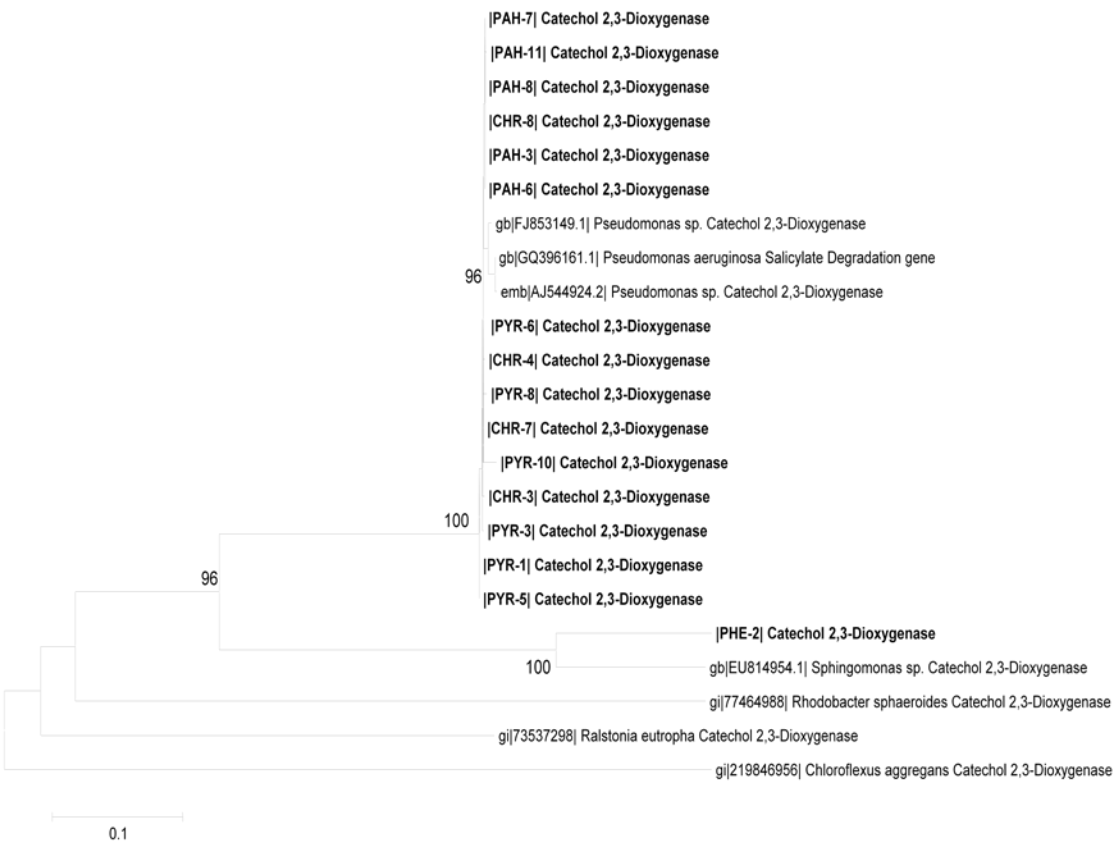

Figure 2.4. Phylogenetic tree of catechol 2,3-dioxygenase genes from Pit 101 **(in bold)** taken from different PAH sample exposures and other similar sequences taken from GenBank along with accession numbers. Clones from different PAH exposures are denoted in clone name, PHE=phenanthrene, PYR=pyrene, CHR=chrysene and PAH=combination of PAHs. The evolutionary history was inferred using the Neighbor-Joining method. The optimal tree with the sum of branch length = 2.26571718 is shown. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree.

There was one clone that did not belong to a *Pseudomonas sp.* and was instead similar to a catechol 2,3-dioxygenase gene from a *Sphingomonas sp.*, which was isolated from TNT contaminated soil and can also degrade PAHs (figure 2.4). This gene was found only in the samples exposed to phenanthrene and not in any of the other treatments. This could imply that there are more organisms able to degrade phenanthrene and its metabolites, so there is less of a selective pressure allowing a more diverse set of organisms to survive. Also, the results showed that the main degraders in the La Brea Tar Pit asphalts are from the genus *Pseudomonas* and more specifically *Pseudomonas stutzeri*. This bacterium and other members of the *Pseudomonas* genus are widely known PAH degraders and are frequently isolated from petroleum contaminated sites, so their presence in the La Brea Tar Pits was to be expected.

In conclusion, using culture based methods, I was able to isolate and culture organisms that degrade phenanthrene as a sole carbon source. A total of 18 isolates were confirmed to degrade PAH based on the formation of clearing zones on PHE amended plates and were identified by sequencing the 16S rDNA genes of each organism. The diversity of these culturable organisms was very different from that of the culture-independent techniques. None of the organisms isolated were found on the DGGE analysis and in the 16S rDNA microbial survey by Kim and Crowley. In addition, no *Pseudomonas sp.* was isolated and the majority of the isolates were found to belong to the α- or β-proteobacteria class instead of γ, of which members of Pseudomonas belong to. Phylogenetic analysis revealed that 14 of the 18 isolates were closely related to *Ralstonia sp. Cupriavidus sp.* and *Wautersia sp.* and are all representatives of species that have been previously shown to degrade petroleum hydrocarbons (Table 2.2). There were also 2 isolates PHE-ISO06 and PHE-ISO07 that were not closely related to any known species and

Isolate	BLAST Results	% Homology	Compounds Degraded or Isolated From	Reference
|PHE-ISO01|	Uncultured soil bacterium clone	99%	Polycyclic aromatic hydrocarbons	(Powell et al., 2008)
|PHE-ISO02|	*Mesorhizobium sp.*	99%	N/A	(Samba et al., unpublished)
|PHE-ISO03|	Uncultured soil bacterium clone	98%	Polycyclic aromatic hydrocarbons	(Singleton et al., 2005)
|PHE-ISO04|	*Wautersia sp.*	99%	Phenol	(Dong et al., unpublished)
|PHE-ISO05|	Uncultured bacterium clone	98%	Trichloroethylene and toluene	(Han et al., unpublished)
|PHE-ISO06|	Soil Phyllobacteriaceae bacterium	92%	N/A	(Ben-Dov, E. and A. Kushmaro, 2009)
|PHE-ISO07|	*Ralstonia sp.*	95%	2,4-Dichlorophenoxyacetic acid	(Huong et al., 2007)
|PHE-ISO08|	*Ralstonia sp.*	98%	Benzene, toluene, ethylbenzene, and o-xylene	(Lee S.K. and S.B. Lee, unpublished)
|PHE-ISO09|	Uncultured bacterium clone	97%	N/A	(Field et al., 2010)
|PHE-ISO10|	*Ralstonia sp.*	99%	Benzo[a]pyrene	(Kannaly et al., 2002)
|PHE-ISO11|	Burkholderiaceae bacterium A1Y15	98%	Benzene, toluene, ethylbenzene, and o-xylene	(Hendrikx et al., 2006)
|PHE-ISO12|	*Ralstonia sp.*	98%	N/A	(Hong et al., unpublished)
|PHE-ISO13|	*Cupravidus necator*	98%	Trichloroethylene	(Hanada et al., 1998)
|PHE-ISO14|	*Wautersia numazuensis*	98%	Chlorinated aliphatic hydrocarbons	(Kageyama et al., 2005)
|PHE-ISO15|	*Cupravidus necator*	98%	Phenol	(Futamata et al., 2005)
|PHE-ISO16|	Uncultured soil bacterium clone	98%	Polycyclic aromatic hydrocarbons	(Powell et al., 2008)
|PHE-ISO17|	*Cupriavidus respiraculi*	98%	N/A	(Chopade et al., unpublished)
|PHE-ISO18|	*Ralstonia sp.*	98%	2,4-Dichrolophenoxyacetic Acid	(Huong et al., 2007)

Table 2.2. BLAST results for the bacterial isolates from phenanthrene enrichment cultures. Many of the identified organisms were closely related to those that are able to degrade other petroleum hydrocarbons or were isolated from contaminated sites. Phylogenetic analysis is reported in figure 2.5.

may be novel (Figure2.5) To address this issue, a metagenomics approach may be needed to overcome both the culturing and PCR biases and resolve the true genetic diversity of the La Brea Tar Pits.

Figure 2.5. Phylogenetic tree of PAH degrading bacteria isolated from enrichment cultures **(in bold)** using phenanthrene as a sole carbon source along with related organisms taken from GenBank along with accession numbers. The evolutionary history was inferred using the Neighbor-Joining method. The optimal tree with the sum of branch length = 1.32613366 is shown. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree.

BIBLIOGRAPHY

AKERSTEN, W. A., SHAW, C. A. & JEFFERSON, G. T. 1983. RANCHO-LA-BREA - STATUS AND FUTURE. Paleobiology, 9**,** 211-217.

AL-AWADHI, H., SULAIMAN, R. H. D., MAHMOUD, H. M. & RADWAN, S. S. 2007. Alkaliphilic and halophilic hydrocarbon-utilizing bacteria from Kuwaiti coasts of the Arabian Gulf. Applied Microbiology and Biotechnology, 77**,** 183-186.

ALLEY, J. F. & BROWN, L. R. 2000. Use of sublimation to prepare solid microbial media with water-insoluble substrates. Applied and Environmental Microbiology, 66**,** 439-442.

ALQUATI, C., PAPACCHINI, M., RICCARDI, C., SPICAGLIA, S. & BESTETTI, G. 2005. Diversity of naphthalene-degrading bacteria from a petroleum contaminated soil. Annals of Microbiology, 55**,** 237-242.

APRILL, W. & SIMS, R. C. 1990. EVALUATION OF THE USE OF PRAIRIE GRASSES FOR STIMULATING POLYCYCLIC AROMATIC HYDROCARBON TREATMENT IN SOIL. Chemosphere, 20**,** 253-265.

ATLAS, R. M. 1981. MICROBIAL-DEGRADATION OF PETROLEUM-HYDROCARBONS - AN ENVIRONMENTAL PERSPECTIVE. Microbiological Reviews, 45**,** 180-209.

BALDWIN, B. R., NAKATSU, C. H. & NIES, L. 2003. Detection and enumeration of aromatic oxygenase genes by multiplex and real-time PCR. Applied and Environmental Microbiology, 69**,** 3350-3358.

BASTIN, E. S., GREER, F. E., MERRITT, C. A. & MOULTON, G. 1926. The presence of sulphate reducing bacteria in oil field waters. Science, 63**,** 21-24.

BEDIENT, P. B., RIFAI, H. S. & NEWELL, C. J. 1994. Ground water contamination: transport and remediation. Ground water contamination: transport and remediation.**,** xvi + 541 pp.

BRAGG, J. R., PRINCE, R. C., HARNER, E. J. & ATLAS, R. M. 1994. EFFECTIVENESS OF BIOREMEDIATION FOR THE EXXON-VALDEZ OIL-SPILL. Nature, 368**,** 413-418.

BRAZIL, G. M., KENEFICK, L., CALLANAN, M., HARO, A., DELORENZO, V., DOWLING, D. N. & OGARA, F. 1995. CONSTRUCTION OF A RHIZOSPHERE PSEUDOMONAD WITH POTENTIAL TO DEGRADE POLYCHLORINATED-BIPHENYLS AND DETECTION OF BPH GENE-EXPRESSION IN THE RHIZOSPHERE. Applied and Environmental Microbiology, 61**,** 1946-1952.

CERNIGLIA, C. E. 1993. Biodegradation of polycyclic aromatic hydrocarbons. Current Opinion in Biotechnology, 4**,** 331-338.

CHAPELLE, F. H. 1999. Bioremediation of petroleum hydrocarbon-contaminated ground water: The perspectives of history and hydrology. Ground Water, 37**,** 122-132.

CHAUDHRY, G. R. & CHAPALAMADUGU, S. 1991. BIODEGRADATION OF HALOGENATED ORGANIC-COMPOUNDS. Microbiological Reviews, 55**,** 59-79.

DAVIES, J. I. & EVANS, W. C. 1964. OXIDATIVE METABOLISM OF NAPHTHALENE BY SOIL PSEUDOMONADS - RING-FISSION MECHANISM. Biochemical Journal, 91**,** 251-&.

DUA, M., SINGH, A., SETHUNATHAN, N. & JOHRI, A. K. 2002. Biotechnology and bioremediation: successes and limitations. Applied Microbiology and Biotechnology, 59**,** 143-152.

FERRERO, M., LLOBET-BROSSA, E., LALUCAT, J., GARCIA-VALDES, E., ROSSELLO-MORA, R. & BOSCH, R. 2002. Coexistence of two distinct copies of naphthalene degradation genes in Pseudomonas strains isolated from the western Mediterranean region. Applied and Environmental Microbiology, 68**,** 957-962.

HABE, H., OMORI, T. & RM 2003. Genetics of polycyclic aromatic hydrocarbon metabolism in diverse aerobic bacteria. Bioscience Biotechnology and Biochemistry, 67**,** 225-243.

HAMME, J. D. V., AJAY, S. & WARD, O. P. 2003. Recent advances in petroleum microbiology. Microbiology and Molecular Biology Reviews, 67**,** 503-549.

JOHNSEN, A. R., WICK, L. Y., HARMS, H. & CU 2005. Principles of microbial PAH-degradation in soil. Environmental Pollution, 133**,** 71-84.

JUNCA, H. & PIEPER, D. H. 2004. Functional gene diversity analysis in BTEX contaminated soils by means of PCR-SSCP DNA fingerprinting: comparative diversity assessment against bacterial isolates and PCR-DNA clone libraries. Environmental Microbiology, 6**,** 95-110.

KANALY, R. A. & HARAYAMA, S. 2000. Biodegradation of high-molecular-weight polycyclic aromatic hydrocarbons by bacteria. Journal of Bacteriology, 182**,** 2059-2067.

KANG, Y. S., KIM, Y. J., JEON, C. O. & PARK, W. 2006. Characterization of naphthalene-degrading Pseudomonas species isolated from pollutant-contaminated sites: Oxidative stress during their growth on naphthalene. Journal of Microbiology and Biotechnology, 16**,** 1819-1825.

KIM, J. S. & CROWLEY, D. E. 2007. Microbial diversity in natural asphalts of the rancho la brea tar pits. Applied and Environmental Microbiology, 73**,** 4579-4591.

KUIPER, I., LAGENDIJK, E. L., BLOEMBERG, G. V. & LUGTENBERG, B. J. J. 2004. Rhizoremediation: A beneficial plant-microbe interaction. Molecular Plant-Microbe Interactions, 17**,** 6-15.

LLADO, S., JIMENEZ, N., VINAS, M. & SOLANAS, A. M. 2009. Microbial populations related to PAH biodegradation in an aged biostimulated creosote-contaminated soil. Biodegradation, 20**,** 593-601.

MA, Y. F., WANG, L., SHAO, Z. Z. & ID 2006. Pseudomonas, the dominant polycyclic aromatic hydrocarbon-degrading bacteria isolated from Antarctic soils and the role of large plasmids in horizontal gene transfer. Environmental Microbiology, 8**,** 455-465.

OLIVERA, N. L., COMMENDATORE, M. G., DELGADO, O. & ESTEVES, J. L. 2003. Microbial characterization and hydrocarbon biodegradation potential of natural bilge waste microflora. Journal of Industrial Microbiology & Biotechnology, 30**,** 542-548.

OLLIVIER, B. & MAGOT, M. 2005. Petroleum microbiology. Petroleum microbiology**,** xi + 365 pp.

PAUL, D., PANDEY, G., PANDEY, J. & JAIN, R. K. 2005. Accessing microbial diversity for bioremediation and environmental restoration. Trends in Biotechnology, 23**,** 135-142.

PRITCHARD, P. H., MUELLER, J. G., ROGERS, J. C., KREMER, F. V. & GLASER, J. A. 1992. Oil spill bioremediation: Experiences, lessons and results from the Exxon Valdez oil spill in Alaska. Biodegradation, 3**,** 315-335.

Safinowski, M. C. Griebler, and R.U. Meckenstock. 2006. Anaerobic cometabolic transformation of polycyclic and heterocyclic aromatic hydrocarbons: evidence from laboratory and field studies. Environ. Sci. Technol. 40:4165-4173.

SALANITRO, J. P. 2001. Bioremediation of petroleum hydrocarbons in soil. Advances in Agronomy, Vol 72, 72**,** 53-105.

SCRAGG, A. 2005. Environmental Biotechnology, Second Edition. Environmental Biotechnology, Second Edition.

SPAIN, J. C. 1995. BIODEGRADATION OF NITROAROMATIC COMPOUNDS. Annual Review of Microbiology, 49**,** 523-555.

SPRINGAEL, D. & TOP, E. M. 2004. Horizontal gene transfer and microbial adaptation to xenobiotics: new types of mobile genetic elements and lessons from ecological studies. Trends in Microbiology, 12**,** 53-58.

STUART-KEIL, K. G., HOHNSTOCK, A. M., DREES, K. P., HERRICK, J. B., MADSEN, E. L. & LP 1998. Plasmids responsible for horizontal transfer of naphthalene catabolism genes between bacteria at a coal tar contaminated site are homologous to pDTG1 from Pseudomonas putida NCIB 9816-4. Applied and Environmental Microbiology, 64**,** 3633-3640.

TALLEY, J. W. 2006. Bioremediation of Relcalcitrant Compounds. Bioremediation of Relcalcitrant Compounds.

TAYLOR, P. M. & JANSSEN, P. H. 2005. Variations in the abundance and identity of class II aromatic ring-hydroxylating dioxygenase genes in groundwater at an aromatic hydrocarbon-contaminated site. Environmental Microbiology, 7**,** 140-146.

TESTA, S. M. & WINEGARDNER, D. L. 1991. RESTORATION OF PETROLEUM-CONTAMINATED AQUIFERS. Testa, S. M. and D. L. Winegardner. Restoration of Petroleum-Contaminated Aquifers. Xi+269p. Lewis Publishers, Inc.: Chelsea, Michigan, USA. Illus. Maps**,** XI+269P.

TOP, E. M. & SPRINGAEL, D. 2003. The role of mobile genetic elements in bacterial adaptation to xenobiotic organic compounds. Current Opinion in Biotechnology, 14**,** 262-269.

VANDERMEER, J. R. 1994. GENETIC ADAPTATION OF BACTERIA TO CHLORINATED AROMATIC-COMPOUNDS. Fems Microbiology Reviews, 15**,** 239-249.

VANDERMEER, J. R., DEVOS, W. M., HARAYAMA, S. & ZEHNDER, A. J. B. 1992. MOLECULAR MECHANISMS OF GENETIC ADAPTATION TO XENOBIOTIC COMPOUNDS. Microbiological Reviews, 56**,** 677-694.

WIKSTROM, P., WIKLUND, A., ANDERSSON, A. C. & FORSMAN, M. 1996. DNA recovery and PCR quantification of catechol 2,3-dioxygmase genes from different soil types. Journal of Biotechnology, 52**,** 107-120.

WINDERL, C., ANNESER, B., GRIEBLER, C., MECKENSTOCK, R. U. & LUEDERS, T. 2008. Depth-resolved quantification of anaerobic toluene degraders and aquifer microbial community patterns in distinct redox zones of a tar oil contaminant plume. Applied and Environmental Microbiology, 74**,** 792-801.

CHAPTER THREE:

LINKING PHYLOGENY AND FUNCTION OF A PREVIOUSLY UNKNOWN ORGANISM AND A NOVEL 2-NITROPROPANE DIOXYGENASE USING METAGENOMICS

ABSTRACT

Recent studies on the biodiversity of asphalt deposits at the Rancho La Brea Tar Pits in Los Angeles, California have revealed the existence of several hundred new species of bacteria and gene sequences encoding putative novel degradative enzymes. The presence of fossilized extinct animal remains in the La Brea Tar Pits has led to estimations that these natural asphalt seeps have existed for at least 40,000 year. These deposits consist of petroleum that has been degraded to the extent that the remaining material is comprised mainly of asphalts and heavy oils, which have saturated into the soil matrix. Since petroleum hydrocarbons are both a target and a product of microbial metabolism, the role of microorganisms that inhabit this and other similar environments is directly relevant to development of technology for bioremediation, biotransformation of petroleum hydrocarbons, and microbial enhanced oil recovery for extracting and refining heavy oil.

To investigate the true microbial diversity of the La Brea Tar Pits, the microbial community associated with the heavy oil at Rancho La Brea was studied using a metagenomics approach. A fosmid clone library was constructed using 38 Kb fragments of DNA extracted from the asphalt-soil samples of Pit 101. This library of about 3,000 clones was then screened for DNA inserts, which contained specific genes that could be targeted using PCR based methods. One

selected clone contained the 16S rRNA gene of an unclassified Rhodospirallacaea and a putative 2-nitropropane dioxygenase, which suggest that this new organism degrades persistent nitroalkanes that are common in asphalt. In addition a DNA-DNA microarray was used to verify the presence of 2-NPD and many other petroleum degrading genes.

INTRODUCTION

Studies of microorganisms in pure laboratory culture for over a century have led to significant advances into microbial genetics and physiology, biotechnology and molecular biology. However, the majority of the bacteria has not or cannot be cultured under laboratory conditions. Microorganisms live in very complex microbial communities (e.g. soil, asphalt, sewage, contaminated sites) and thus it is extremely difficult to recreate those conditions on a Petri dish. Due to the limitations and biases that arise from culturing microorganisms to examine their diversity and their community role, a relatively new type of methodology was developed to help circumvent these problems, metagenomics.

"Metatgenomics" describes the functional and sequence-based analysis of the collective microbial genomes contained in an environmental sample. In metagenomics, the power of genomic analysis is applied to entire communities of microbes, bypassing the need to isolate and culture individual bacterial community members (NRC, 2007). This allows you to sequence the DNA of the microbial community directly from its environment, which can then be used to discover new genes and enzymes or to analyze entire genomes. There are several different strategies when analyzing an environment using metagenomic approaches. However, every analysis begins with the extraction of DNA directly from its environment. Depending on the strategy analysis, the DNA fragments can range in size from as small as 1 kb to over 100 kb. These fragments are then ligated into suitable vectors (e.g. cosmids, fosmids, BACs) then cloned into appropriate hosts (e.g. E. coli), thus resulting in the metagenomic library. Once the library is created, the DNA inserts can be randomly sequenced or analyzed using a sequenced or functional based approach.

In random sequencing, the DNA inserts are shotgun sequenced (i.e. high throughput sequencing) and then assembled, attempting to recreate an entire organism's genome, metabolic pathway or certain target genes. This strategy is useful for very selective and simple environments, in which they contain few dominant organisms. However, for complex and diverse communities, this task will be extremely difficult because of the sequencing coverage required for the respective metagenome. In these diverse and complex communities, where many different organisms and biological processes are present, directed sequencing of target organisms or genes is a widely used strategy. In sequence based and functional based screening of a metagenomic library, selected clones are sequenced, instead of the entire library or community. In this case, the DNA inserts are larger (i.e. 20 kb to over 100 kb) in order to gather more information for that particular species. In a sequenced based analysis, clones are selected by searching for genes of interest via polymerase chain reaction or colony blotting. Clones that contain the target gene are then selected for analysis. In a functional based analysis, genes from the DNA insert are expressed in an expression vector and the metagenomic clones with the desired function are selected. Once a clone with the gene or protein of interest is found, the entire DNA insert is sequenced along with functional gene analysis.

Metagenomics has led to several advances in microbiology and biotechnology by examining unculturable majority of the microbial diversity. These approaches have been used successfully to isolate antibiotics (Gillespie et al., 2002), discover unknown genes (Venter et al., 2004), express novel biocatalysts (Voget et al., 2003, Wexler et al., 2005), find new gene functions (Beja et al., 2000) and even sequence near complete genomes of so far unculturable microbes (Martin et al., 2006, Tyson et al., 2004). Metagenomic approaches have also lead to

the finding of novel petroleum degrading genes and gene families. In 2007, Suenaga et al. was able to find novel genes for the degradation of aromatic compounds from coke plant wastewater by using a functional screen for extradiol dioxygenases. These clones were tested for their substrate specificity for various catecholic compounds, in which they found a novel gene family of extradiol dioxygenases. Also in 2007, Ono et al. found and characterized naphthalene-catabolic genes and plasmids from oil contaminated soil by screening for specific transposons, which are known to carry petroleum hydrocarbon degrading genes for horizontal gene transfer. By using this approach they found cosmid clones, which contained a naphthalene upper pathway operon for the conversion of naphthalene to salicylate. In almost all cases, these metagenomic approaches were used to retrieve novel genes and catabolic pathways as well as to gain insights into the gene evolution of degradative enzymes (Suenaga et al., 2007, Ono et al., 2007).

In the case of the La Brea Tar Pits, the discovery of novel enzymes and pathways is promising. Selection of bacterial communities for petroleum substances occurs rapidly after even short-term exposures of soil to petroleum hydrocarbons following oil spills (Vandermeer et al., 1992, Vandermeer, 1994). Over time spans encompassing millennia, such as the La Brea Tar Pits, the bacteria that can tolerate this environment would be expected to undergo genetic adaptations that may lead to the evolution of new ecotypes and species and enzymes for growth on petroleum hydrocarbons. During adaptation of communities, genes for petroleum hydrocarbon-degrading enzymes that are carried on plasmids or transposons may be exchanged between species. In turn, new catabolic pathways eventually may be assembled and modified for efficient degradation of petroleum hydrocarbons (Rabus et al., 2005). Here, we constructed

a metagenomic library of roughly 3,000 clones using a fosmid vector (Epicentre), in which 38 kb DNA fragments extracted from Pit 101 of the Rancho La Brea Tar Pits were successfully ligated and cloned. The metagenomic library was screened using PCR sequence analysis targeting specific phylogenetic markers of novel organisms of interest. In our search, we sequenced a DNA insert which belongs to a thus far unculturable novel organism from the Rhodospirallacea family. This DNA fragment also carried a putative 2-nitropropane dioxygenase (2-NPD), which is used to degrade nitroalkanes found in petroleum environments. In addition, an environmental microarray analysis, which was used to detect key microbial processes, such as the biodegradation of petroleum hydrocarbons confirmed the presence of 2-NPD, which implies that the degradation of nitroalkanes is an important function within the La Brea Tar Pits.

MATERIALS AND METHODS

Pit 101 sampling and DNA extraction. Previously unexposed samples were removed from approximately 10 cm under the surface of Pit 101 of the Rancho La Brea Tar Pits in Los Angeles, CA as previously described (Kim and Crowley, 2007). Samples were removed from the asphalt pits with sterile, autoclaved spatulas and were transferred into sterile sampling bags (Fisherbrand) for processing. The soil-tar samples were used to extract high molecular weight DNA about 38 Kb in length. DNA was extracted using an indirect cell lysis technique, which first separated the cells from the soil-tar matrix, then chemically lysed for subsequent DNA extraction (Gabor et al., 2003). The DNA was then purified by multiple low melting point agarose gel extractions using an agarase enzyme (Epicentre).

Fosmid library construction. The metagenomic library was constructed using the CopyControl™ Fosmid Library Production Kit (Epicentre) and manufactures' protocols were followed. Briefly, DNA fragments of over 38 kb were selected and purified. The resultant DNA was then blunt-end repaired and ligated to fosmid vector pCC1FOS™ and pCC2FOS™. The fosmid and insert DNA was packaged by Lambda Phage and then used to infect *E. coli* strain EP1300™. A library of over 3000 clones was picked and grown individually using LB medium supplemented with chloramphemicol in 384 well plates. The DNA inserts of roughly 38 Kb were verified by digesting a random selection of fosmids using NOT I restriction enzyme (New England Biolabs).

Metagenomic library screening. 10 µl aliquots from each clone were collected and pooled by rows from each plate. 50 µl aliquots from each pool of rows of the same plate were then combined and re-grown in a 5 ml culture overnight. The fosmids were induced to high copy number and subsequently extracted using a FOSMIDMAX™ Purification Kit (Epicentre).

DNA from the pooled plates were screened via PCR analysis (Erkel et al., 2005). The primers used include Pseudomonas-selective Ps289F - 5' GGT CTG AGA GGA TGA TCA GT 3' and Ps1285R - 5' AGC TCC ACC TCG CGG C -3', which amplify a 750 bp fragment of the 16S rDNA specific for *Pseudomonas sp.* (Widmer et al., 1998). Plates that were found positive for our target gene were examined further by subjecting pooled fosmid DNA from each row to another round of PCR analysis and rows that were positive were examined further until single clones were selected and isolated.

Sequencing fosmid DNA and analysis. The positive clone, IH-22 was streaked and purified, then re-grown overnight in a 5 ml LB culture supplemented with chloramphenicol. After incubation, the cells were harvested and fosmid DNA was purified using the FOSMIDMAX^{TM} Purification Kit (Epicentre). Roughly 5 µg of fosmid DNA was collected and sheared in to fragments of .75-1.5 kb using a pressurized nebulizer. The sheared DNA was cloned in a TOPO TA shotgun sequencing vector (Invitrogen) and approximately 288 clones were randomly shotgun sequenced at the UC Riverside Genomics Core. The sequencing reads were assembled using the Straden Assembly Software. Glimmer was used to predict open reading frames (ORFs) and all of the ORFs were annotated using the NCBI database and the PFAM or KEGG databases where applicable.

Microarray probe construction and analysis. A total of 474 gene sequences were compiled and used to design probes, which would detect genes that encode key enzymes used for petroleum hydrocarbon degradation. Microarray probes were designed from the sequences using Oligoarray 2.1 and a total of 376 probes were constructed (Rouillard et al., 2003). The probes were designed with a predicted melting temperature of 80-85˚C; a length of 25-50 bases; a minimum temperature of 65˚C for cross -hybridization and secondary structure predictions, and

to be located within 1000 bases of the 3' end of the input sequence. The microarray probes were synthesized *in situ* on a Combimatrix 12K microarray and each of the probes were replicated four times across the microarray. Three micrograms of purified DNA extracted from Pit 101 of the La Brea Tar Pits was fragmented, labeled, applied and hybridized to the microarray in accordance with CombiMatrix's suggested protocol (CombiMatrix Corporation, Mukilteo, WA, USA; www.combimatrix.com). Slides were scanned on an Axon 4000B scanner and feature intensity data measured from the scanned images using CombiMatrix software.

RESULTS

De novo assembly of fosmid insert DNA and ORF prediction from clone IH-22. Roughly 288 shotgun sequences were used to assemble the DNA insert of clone IH-22. Vector sequences and *E.* coli host sequences were screened and the final assembly yielded five contigs with an average length of 7.58 kb, and range between 1.1 kb and 22.8 kb. The total length of DNA sequences was roughly 37.9 KB, which was very close to the estimated 38 kb insert. In total, 24 ORFs were predicted with an average length of 1.2 kb and 6 additional partial ORFs were found with an average length of 672.2 bp. Of the total ORFs predicted, only 7 of them had an 85% or greater identity match to other similar genes in the NCBI database and 15 ORFs did not have any match from the PFAM or KEGG database (Table 3.1).

Phylogenetic analysis of clone IH-22 and functional gene predtion. There were several phylogenetic markers found in contig JPB003, which has a total length 22.8 kb. The full length 16S rRNA gene was detected and analyzed using the NCBI database. The gene sequence was found to be 97% similar to an uncultured Rhodospirallacea 16S rDNA clone and several other organisms that were previously found by others in oil-contaminated sites (Liu et al., 2009). This sequence was also classified as an unclassified Rhodospirallacea member using the RDP database, which implies that this organism is thus far an unculturable, previously undescribed organism (Figure 3.1).

Also found on contig JPB003 was the full length 23S large subunit ribosomal RNA gene and the 5S small subunit gene. There were no significant matches found in the NCBI databases as the genes were 89% similar to *Rhodospirillum rubrum* and 91% similar to *Rhodospirillum centenum*, respectively. Phylogenetic analysis illustrated very low similarities to other known

Contig	ORF	BLAST Gene Product	% Identity	Organism (NCBI Database)
>JPB001	orf1-00002	2-nitropropane dioxygenase	93	Rhodococcus jostii RHA1
	orf1-00004	methylcrotonoyl-CoA carboxylase	78	Variovorax paradoxus S110
	orf1-00006	conserved hypothetical protein	81	Anaeromyxobacter dehalogenans 2CP-1
	orf1-00010	type I secretion outer membrane protein, TolC	81	Pseudomonas putida F1
	orf1-00011	methyl-accepting chemotaxis protein	74	Methylibium petroleiphilum PM0
>JPB002	**orf2-00001**	outer membrane efflux protein	81	Anaeromyxobacter dehalogenans 2CP-C
	orf2-00002	type I secretion outer membrane protein, TolC	81	Pseudomonas putida F1
	orf2-00004	conserved hypothetical protein	81	Anaeromyxobacter dehalogenans 2CP-0
>JPB003	orf3-00001	enoyl-CoA hydratase	69	Rhodospirillum centenum SW
	orf3-00004	glutathione-regulated potassium-efflux system protein KefB	67	Rhodospirillum centenum SW
	orf3-00005	Carbamoyl-phosphate synthase L chain ATP-binding	97	Parvibaculum lavamentivorans DS-1
	orf3-00006	3-methylcrotonoyl-CoA carboxylase	73	Rhodopseudomonas palustris BisB5
	orf3-00007	pyruvate carboxyltransferase	78	Xanthobacter autotrophicus Py2
	orf3-00009	Uncharacterized conserved protein	69	Rhodopseudomonas palustris BisA53
	orf3-00010	16S rRNA gene	97	Uncultured Rhodospirillaceae bacterium clone
	orf3-00011	16S-23S ITS Region	86	Azospirillum sp. B510
	orf3-00012	23S rRNA gene	89	Rhodospirillum rubrum
	orf3-00013	5S rRNA gene	91	Rhodospirillum centenum SW
	orf3-00014	IS3 family transposase orfA	98	Rhodobacter sphaeroides ATCC 17029
	orf3-00015	transcriptional regulator protein	68	Agrobacterium radiobacter K84
	orf3-00016	(2Fe-2S)-binding domain protein	77	Methylobacterium radiotolerans JCM 2831
	orf3-00017	putative dehydrogenase	73	Burkholderia cenocepacia J2315
	orf3-00019	conserved hypothetical protein	71	Starkeya novella DSM 506
	orf3-00021	glucose-methanol-choline oxidoreductase	83	Verminephrobacter eiseniae EF01-2
	orf3-00022	binding-protein-dependent transport system	77	Verminephrobacter eiseniae EF01-2
	orf3-00026	ABC transporter ATP-binding protein	100	Mesorhizobium loti MAFF303099
	orf3-00028	extracellular solute-binding protein	67	Verminephrobacter eiseniae EF01-2
	orf3-00029	transcriptional regulator, GntR family	70	Verminephrobacter eiseniae EF01-2
>JPB004	orf4-00001	transcriptional regulator, GntR family protein	78	Rhodospirillum centenum SW
>JPB005	orf5-00001	mandelate racemase/muconate lactonizing enzyme	76	Ruegeria pomeroyi DSS-3
	orf5-00002	succinate-semialdehyde dehydrogenase I	81	Ralstonia solanacearum PSI07

Table 3.1. List of contigs and open reading frames detected after assembly of DNA insert from clone IH-22. BLAST results of each ORF along with percent identity of the closest matched gene from NCBI database are shown. ORF greater than 1 kb are denoted in bold.

Contig	ORF	PFAM	KEGG
>JPB001	orf1-00002	N/A	N/A
	orf1-00004	carboxyl transferase	propionyl-CoA carboxylase
	orf1-00006	N/A	AnaeK_0592 hypothetical protein
	orf1-00010	outer membrane efflux protein	type I secretion outer membrane protein, TolC
	orf1-00011	N/A	N/A
>JPB002	**orf2-00001**	N/A	N/A
	orf2-00002	outer membrane efflux protein	type I secretion outer membrane protein, TolC
	orf2-00004	N/A	AnaeK_0592 hypothetical protein
>JPB003	orf3-00001	N/A	N/A
	orf3-00004	N/A	N/A
	orf3-00005	Carbamoyl-phosphate synthetase large chain domain protein	carbamoyl-phosphate synthase L chain
	orf3-00006	N/A	N/A
	orf3-00007	pyruvate carboxyltransferase	pyruvate carboxyltransferase
	orf3-00009	Uncharacterised conserved protein	uncharacterised conserved protein
	orf3-00010	16S rRNA gene	16S rRNA gene
	orf3-00011	N/A	N/A
	orf3-00012	23S rRNA gene	23S rRNA gene
	orf3-00013	5S rRNA gene	5S rRNA gene
	orf3-00014	IS3 family transposase orfA	SPOA0091 IS3 family transposase orfA
	orf3-00015	N/A	N/A
	orf3-00016	(2Fe-2S)-binding domain protein	(2Fe-2S)-binding domain protein
	orf3-00017	N/A	N/A
	orf3-00019	N/A	N/A
	orf3-00021	glucose-methanol-choline oxidoreductase	glucose-methanol-choline oxidoreductase
	orf3-00022	binding-protein-dependent transport system	binding-protein-dependent transport system
	orf3-00026	N/A	N/A
	orf3-00028	N/A	N/A
	orf3-00029	regulatory protein GntR	transcriptional regulator, GntR family
>JPB004	orf4-00001	N/A	N/A
>JPB005	orf5-00001	N/A	N/A
	orf5-00002	N/A	N/A

Table 3.1. (cont) Table of contigs and open reading frames predicted using GLIMMER after assembly of DNA insert from clone IH-22. ORFs are annotated using PFAM and KEGG databases where applicable. ORFs greater than 1 kb are denoted in bold.

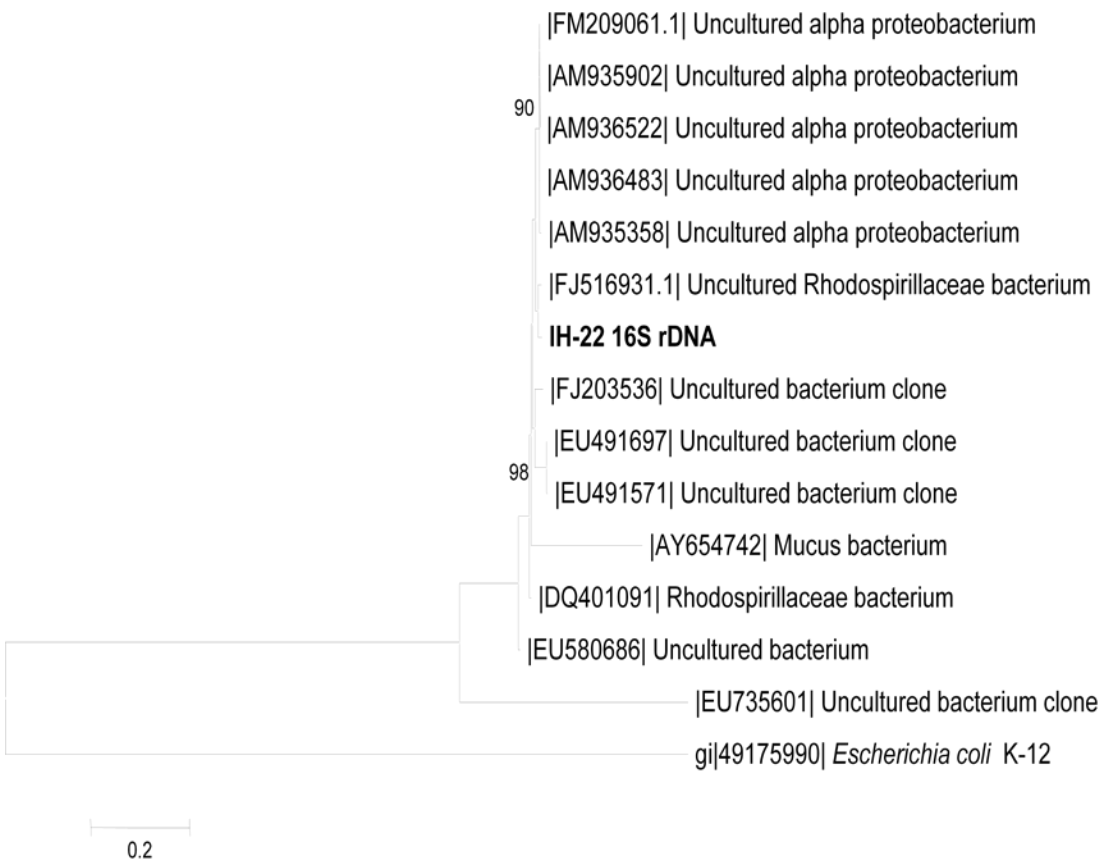

Figure 3.1. Phylogenetic tree of 16S rDNA from clone IH-22 **(in bold)** along with representative organisms taken from the NCBI database. The evolutionary history was inferred using the Neighbor-Joining method. The optimal tree with the sum of branch length = 3.22734020 is shown. The percentage of replicate trees in which the associated taxa clustered together (>85) in the bootstrap test (1000 replicates) are shown next to the branches. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree

organisms, which further implied that this DNA insert belongs to a novel member of the Rhodospirallacea family (Figures 3.2 and 3.3). Another marker that was used for phylogenetic analysis was the 16S-23S ribosomal RNA intergenic spacer. As with the other phylogenetic markers, there were no significant matches in the NCBI database and the closest match was found to be 86% similar to an *Azospirillum sp*., which is also a genus from the Rhodospirallacea family (Figure 3.4).

There was one partial ORF that had a specific degradative function. A 556 bp fragment was found at the 5' end of of contig JPB0001 (Table 3.1). This partial ORF was found to be 93% similar to a 2-nitropropane dioxygenase from the organism *Rhodococcus jostii* (Figure 3.5). There were also other ORFs that were predicted, which appear to also have low similarity to housekeeping genes found in the NCBI, KEGG, PFAM databases (Table 3.1).

Confirmation of 2-NPD in the Rancho La Brea Tar Pits using a microarray analysis and the detection of petroleum hydrocarbon degrading genes. Microarray analysis was performed to detect the presence of specific functional genes used to degrade petroleum hydrocarbons. There were a variety of petroleum degrading genes that were detected, including 2-nitropropane dioxygenase. 2-NPD was found at a relatively high abundance in comparison to the other dioxygenases detected, which confirmed the presence of this gene in the Rancho La Brea Tar Pits (Figure 3.6).

Other petroleum degrading genes were detected from the environmental microarray analysis (Table 3.2). Of the top 15% most represented genes in the microarray, the most common gene was methane monoxygenase, which was detected at a very high abundance with a frequency of 76. The second most found petroleum degrading gene was naphthalene

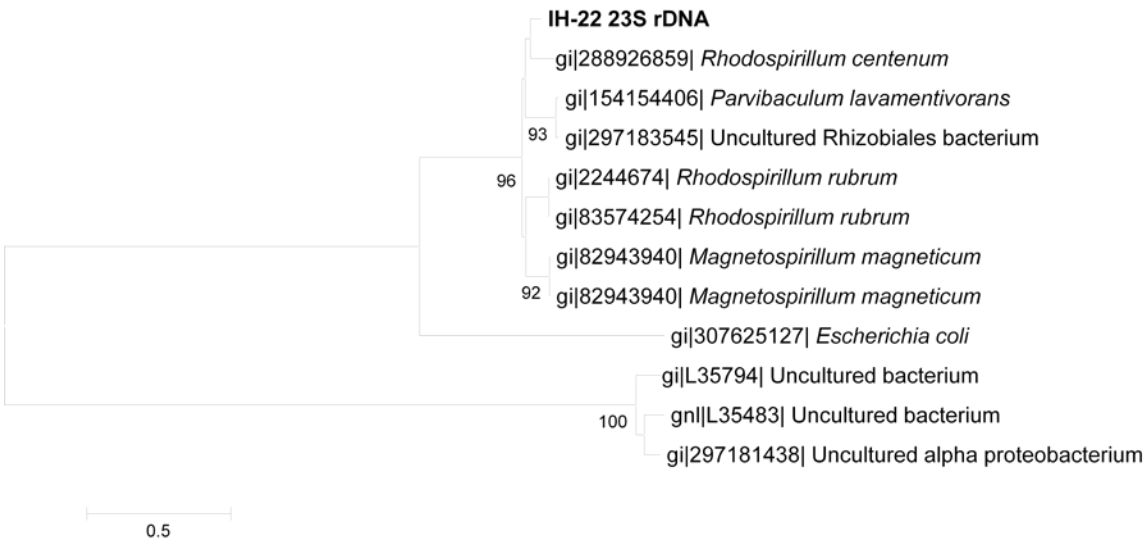

Figure 3.2. Phylogenetic tree of 23S rDNA from clone IH-22 **(in bold)** along with representative organisms taken from the NCBI database. The evolutionary history was inferred using the Neighbor-Joining method . The optimal tree with the sum of branch length = 5.52868653 is shown. The percentage of replicate trees in which the associated taxa clustered together (>85) in the bootstrap test (1000 replicates) are shown next to the branches. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree.

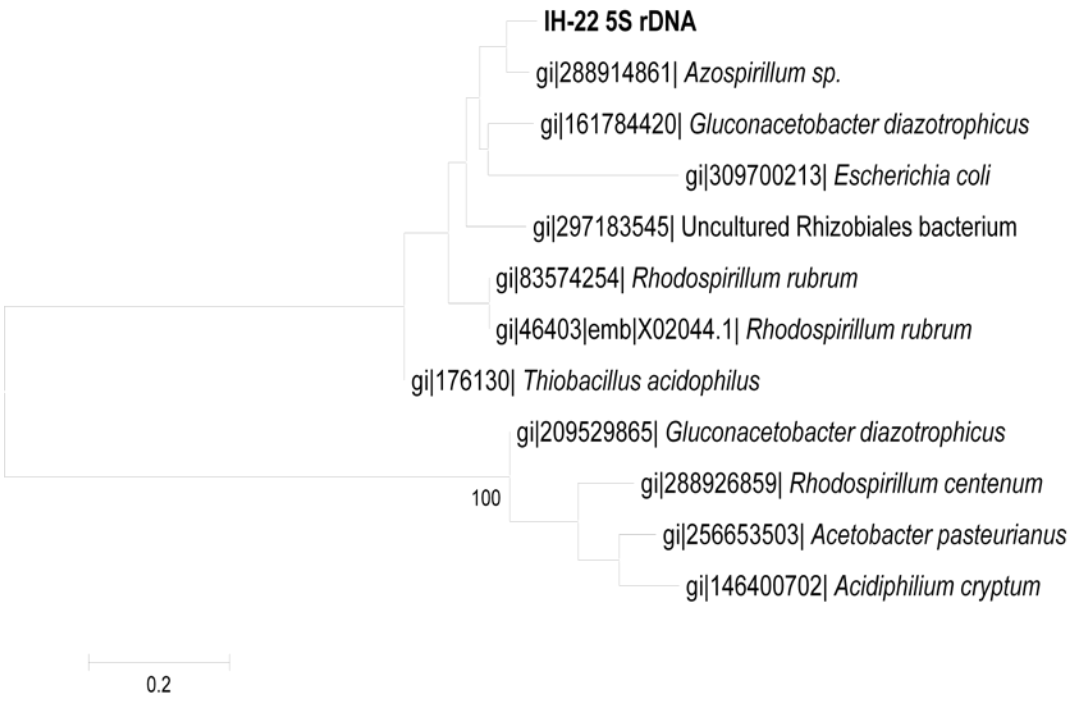

Figure 3.3. Phylogenetic tree of 5S rDNA from clone IH-22 **(in bold)** along with representative organisms taken from the NCBI database. The evolutionary history was inferred using the Neighbor-Joining method. The optimal tree with the sum of branch length = 2.28686511 is shown. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches (> 85). The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree.

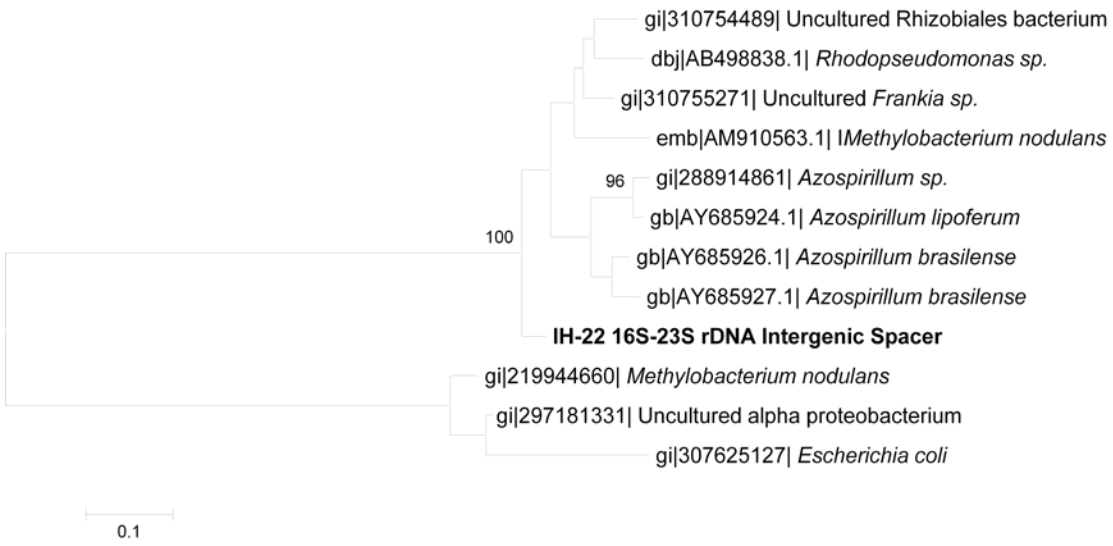

Figure 3.4. Phylogenetic tree of 23S-16S rDNA intergenic spacer from clone IH-22 **(in bold)** along with representative organisms taken from the NCBI database. The evolutionary history was inferred using the Neighbor-Joining method. The optimal tree with the sum of branch length = 1.91609559 is shown. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches (> 85). The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree.

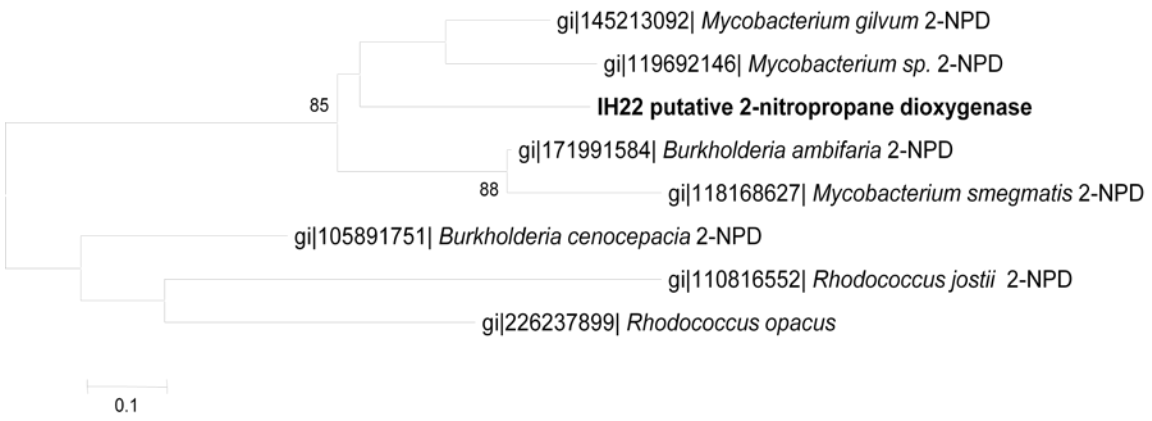

Figure 3.5. Phylogenetic tree of putative 2-nitropropane dioxygenase found on clone IH-22 **(in bold)** along with other representative genes taken from NCBI database. The evolutionary history was inferred using the Neighbor-Joining method. The optimal tree with the sum of branch length = 3.06078971 is shown. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches (> 85). The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree.

Functional Gene	**Frequency of representation in top 15%**
Methane Monoxygenase	76
Naphthalene Dioxygenase	10
Alkaline Phosphatase	9
Benzoyl-CoA Reductase	4
Biphenyl Dioxygenase	3
Chlorobenzoate Dioxygenase	3
Hydroxyphenol Dioxygenase	3
Nitropropane Dioxygenase	3
Benzene Dioxygenase	2
Benzoate Dioxygenase	2
Halobenzoate Dioxygenase	2
Protocatechuate Dioxygenase	2
Reductive Dehalogenase	2
Aromatic Ring Dioxygenase	1
Benzylsuccinate Synthase	1
Benzaldehyde Dehydrogenase	1
Benzoyl Oxygenase	1
Benzyl Succinate Dehydrogenase	1
Catechol Dioxygenase	1
Extradiol Ring Dioxygenase	1
Phenylpriopionate Dioxygenase	1
Toluate Dioxygenase	1
Toluene Dioxygenase	1
Xylene Monooxygenase	1

Table 3.2. Table of microarray analysis results of the top 15% represented functional genes.

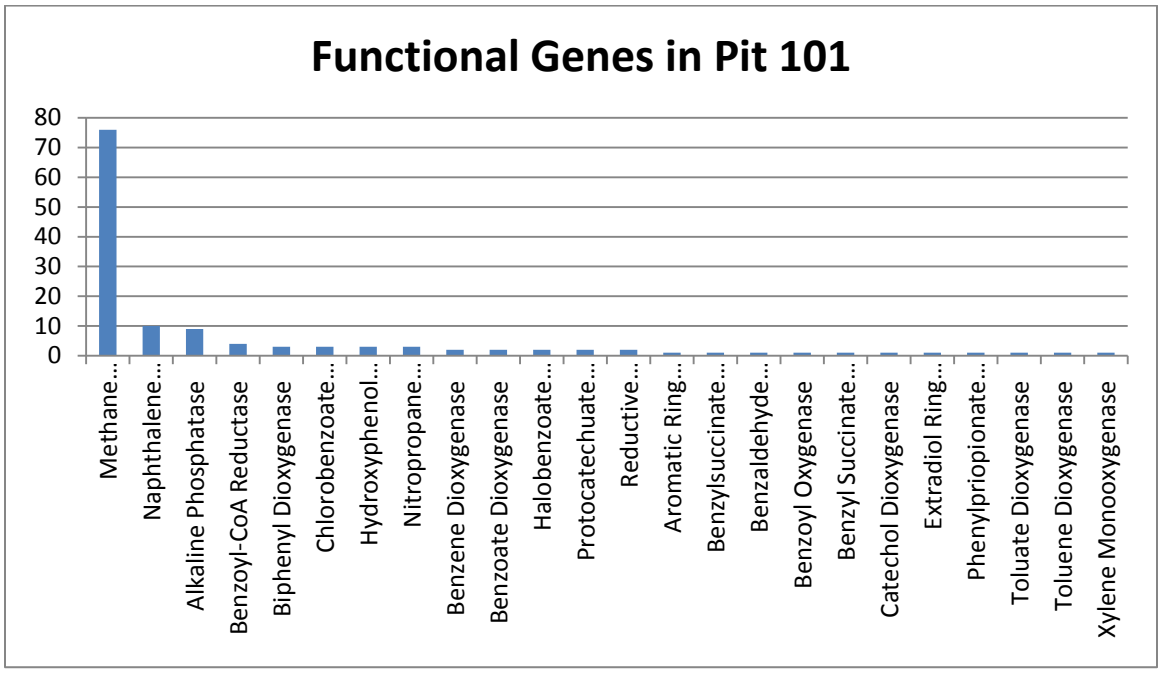

Figure 3.6. Graph of top 15% represented functional genes found in the Rancho La Brea Tar Pits.

dioxygenase, which had a frequency of 10 and has already been found as an important enzyme in the La Brea Tar Pits. Biphenyl dioxygenase and 2-NPD were both represented at a frequency of 3, benzene dioxygenase was represented at a frequency of 2 and several others (aromatic ring dioxygenase, catechol dioxygenase, extradiol ring dioxygenase, toluene dioxygenase and xylene dioxygenase) were found at a frequency of 1. In addition, benzylsuccinate sythase, benzaldehyde dehydrogenase and benzyl succinate dehydrogenase were found at a frequency of 1; these particular enzymes are used for anaerobic degradation of petroleum hydrocarbons.

DISCUSSION

By applying metagenomic techniques to the La Brea Tar Pits, we were able to link function and phylogeny of a thus far previously undescribed, unculturable organism. Phylogenetic analysis of the DNA insert of clone IH-22, illustrated that this fragment belongs to an uncultured novel member of the Rhodospirallacea family. There were no culturable organisms, which were significantly similar to the 16S rDNA found on this fragment. In addition, the other phylogenetic markers which were analyzed also had no significant matches in the NCBI database. Neither the 5S rDNA, 23S rDNA nor the 23S-16S intergenic spacer region showed to have much similarity with any known organism. The fact that there are no significant matches in the NCBI or RDP databases implied that there have been no similar organisms that have been cultured or examined, thus indicating that this organism is indeed novel.

Gene prediction and functional annotation of this DNA fragment resulted in the discovery of a putative petroleum degradative gene. 2-nitropropane dioxygenase is an enzyme used to oxidize nitroalkanes into their corresponding carbonyl compounds and nitrile and is an important function in petroleum contaminated environments (Francis and Gadda, 2008, Gadda et al., 2007, Mijatovic and Gadda, 2008). Although this was a partial ORF of about 556 bp, this sequence was found to be 93% similar to a 2-nitropropane dioxygenase gene found in *Rhodococcus jostii*. This partial ORF was found on the 5' end of JPB001, and according to the assembly, it is assumed to be located at the 5' end of the DNA insert, which suggests that the 5' end of this gene may have been cut in the extraction process.

Phylogenetic analysis of the putative 2-NPD suggests that this gene may be novel as with the organism it belongs to. The microarray analysis which detected the presence of this

gene also implied that the degradation of nitroalkanes is an important process and therefore this organism may play an important role in the La Brea Tar Pits. If so, studies which include the full sequencing of the 2-NPD gene should be performed, along with the expression of this gene. The 93% similarity is relatively significant enough to predict this gene's function, however if this gene is novel, there may be new characteristics that are associated with this enzyme. In addition, qPCR analysis may be a good way to quantify this organism's presence in the La Brea Tar Pits.

The other ORFs that were predicted on this DNA insert also show low similarity to other functional genes in the NCBI, PFAM and KEGG databases. This suggests that these genes, along with the other phylogenetic markers and 2-NPD have not been examined so far. However, the gene functions that are predicted seem to be housekeeping genes that encode enzymes for functions that are common to many bacteria. A polished insert sequence, with 5-6X coverage may aid in the gene prediction of the additional ORFs, however to warrant any additional in-depth sequencing and analysis, functions other than housekeeping and metabolism should be present.

The microarray analysis identified the presence of other important degradative genes in the La Brea Tar Pits. The most represented petroleum degrading gene by a wide margin is methane monooxygenase, which oxidizes the C-H bond in methane and other alkanes. This gene was represented roughly 7.5 times higher than the next most represented gene in naphthalene dioxygenase, which catalyzes the first reaction in the degradation of aromatic hydrocarbons. The presence of these genes in high numbers is not that surprising, especially in a petroleum contaminated environment. It is well known that methane monooxygenases and naphthalene

dioxygenases are readily found in sites where petroleum degradation takes place. However, the more important finding from the microarray analysis is the relatively low number of genes involved in the anaerobic degradation of petroleum. Asphalts are thought to be an anaerobic system, but the overwhelming presence of monooxygenases and dioxygenases found in the La Brea Tar Pits could suggest that there may be oxygen being transported within the asphalts.

Selected sequencing of large insert DNA metagenomic libraries proved to be a valuable technique in examining the unculturable majority of microbial communities. This approach allowed us to examine a thus novel organism from the Rhodospirallacea family, which potentially plays a major role in the degradation of nitroalkanes in the Rancho La Brea Tar Pits. The microarray analysis confirmed the presence of petroleum degrading genes and also provides a useful tool for further examination of the metagenomic library. The important processes that were found in the microarray analysis should serve as future target genes for the metagenomic library. By targeting the genes found in the microarray analysis, one may find DNA inserts, which contain full length degradative gene sequences and even whole metabolic pathways of specific petroleum hydrocarbons. DNA fragments which contain entire metabolic pathways, along with phylogenetic markers may provide valuable information regarding the metabolism of specific compounds and insights into the gene evolution of degradative enzymes.

BIBLIOGRAPHY

BEJA, O., ARAVIND, L., KOONIN, E. V., SUZUKI, M. T., HADD, A., NGUYEN, L. P., JOVANOVICH, S., GATES, C. M., FELDMAN, R. A., SPUDICH, J. L., SPUDICH, E. N. & DELONG, E. F. 2000. Bacterial rhodopsin: Evidence for a new type of phototrophy in the sea. *Science,* 289**,** 1902-1906.

ERKEL, C., KEMNITZ, D., KUBE, M., RICKE, P., CHIN, K. J., DEDYSH, S., REINHARDT, R., CONRAD, R. & LIESACK, W. 2005. Retrieval of first genome data for rice cluster I methanogens by a combination of cultivation and molecular techniques. *Fems Microbiology Ecology,* 53**,** 187-204.

FRANCIS, K. & GADDA, G. 2008. The nonoxidative conversion of nitroethane to ethylnitronate in Neurospora crassa 2-nitropropane dioxygenase is catalyzed by histidine 196. *Biochemistry,* 47**,** 9136-9144.

GABOR, E. M., DE VRIES, E. J. & JANSSEN, D. B. 2003. Efficient recovery of environmental DNA for expression cloning by indirect extraction methods. *Fems Microbiology Ecology,* 44**,** 153-163.

GADDA, G., FRANCIS, K. & BELAINEH, M. 2007. The mechanism of 2-nitropropane dioxygenase: A model for flavin semiquinone intermediates in enzymatic catalysis. *Faseb Journal,* 21**,** A272-A272.

GILLESPIE, D. E., BRADY, S. F., BETTERMANN, A. D., CIANCIOTTO, N. P., LILES, M. R., RONDON, M. R., CLARDY, J., GOODMAN, R. M. & HANDELSMAN, J. 2002. Isolation of antibiotics turbomycin A and B from a metagenomic library of soil microbial DNA. *Applied and Environmental Microbiology,* 68**,** 4301-4306.

KIM, J. S. & CROWLEY, D. E. 2007. Microbial diversity in natural asphalts of the rancho la brea tar pits. *Applied and Environmental Microbiology,* 73**,** 4579-4591.

LIU, R. Y., ZHANG, Y., DING, R., LI, D., GAO, Y. X. & YANG, M. 2009. Comparison of archaeal and bacterial community structures in heavily oil-contaminated and pristine soils. *Journal of Bioscience and Bioengineering,* 108**,** 400-407.

MARTIN, H. G., IVANOVA, N., KUNIN, V., WARNECKE, F., BARRY, K. W., MCHARDY, A. C., YEATES, C., HE, S. M., SALAMOV, A. A., SZETO, E., DALIN, E., PUTNAM, N. H., SHAPIRO, H. J., PANGILINAN, J. L., RIGOUTSOS, I., KYRPIDES, N. C., BLACKALL, L. L., MCMAHON, K. D. & HUGENHOLTZ, P. 2006. Metagenomic analysis of two enhanced biological phosphorus removal (EBPR) sludge communities. *Nature Biotechnology,* 24**,** 1263-1269.

MIJATOVIC, S. & GADDA, G. 2008. Oxidation of alkyl nitronates catalyzed by 2-nitropropane dioxygenase from Hansenula mrakii. *Archives of Biochemistry and Biophysics,* 473**,** 61-68.

ONO, A., MIYAZAKI, R., SOTA, M., OHTSUBO, Y., NAGATA, Y. & TSUDA, M. 2007. Isolation and characterization of naphthalene-catabolic genes and plasmids from oil-contaminated soil by using two cultivation-independent approaches. *Applied Microbiology and Biotechnology,* 74**,** 501-510.

RABUS, R., KUBE, M., HEIDER, J., BECK, A., HEITMANN, K., WIDDEL, F. & REINHARDT, R. 2005. The genome sequence of an anaerobic aromatic-degrading denitrifying bacterium, strain EbN1. *Archives of Microbiology,* 183**,** 27-36.

ROUILLARD, J. M., ZUKER, M. & GULARI, E. 2003. OligoArray 2.0: design of oligonucleotide probes for DNA microarrays using a thermodynamic approach. *Nucleic Acids Research,* 31**,** 3057-3062.

SUENAGA, H., OHNUKI, T., MIYAZAKI, K. & AC 2007. Functional screening of a metagenomic library for genes involved in microbial degradation of aromatic compounds. *Environmental Microbiology,* 9**,** 2289-2297.

TYSON, G. W., CHAPMAN, J., HUGENHOLTZ, P., ALLEN, E. E., RAM, R. J., RICHARDSON, P. M., SOLOVYEV, V. V., RUBIN, E. M., ROKHSAR, D. S. & BANFIELD, J. F. 2004. Community structure and metabolism through reconstruction of microbial genomes from the environment. *Nature,* 428**,** 37-43.

VANDERMEER, J. R. 1994. GENETIC ADAPTATION OF BACTERIA TO CHLORINATED AROMATIC-COMPOUNDS. *Fems Microbiology Reviews,* 15**,** 239-249.

VANDERMEER, J. R., DEVOS, W. M., HARAYAMA, S. & ZEHNDER, A. J. B. 1992. MOLECULAR MECHANISMS OF GENETIC ADAPTATION TO XENOBIOTIC COMPOUNDS. *Microbiological Reviews,* 56**,** 677-694.

VENTER, J. C., REMINGTON, K., HEIDELBERG, J. F., HALPERN, A. L., RUSCH, D., EISEN, J. A., WU, D. Y., PAULSEN, I., NELSON, K. E., NELSON, W., FOUTS, D. E., LEVY, S., KNAP, A. H., LOMAS, M. W., NEALSON, K., WHITE, O., PETERSON, J., HOFFMAN, J., PARSONS, R., BADEN-TILLSON, H., PFANNKOCH, C., ROGERS, Y. H. & SMITH, H. O. 2004. Environmental genome shotgun sequencing of the Sargasso Sea. *Science,* 304**,** 66-74.

VOGET, S., LEGGEWIE, C., UESBECK, A., RAASCH, C., JAEGER, K. E. & STREIT, W. R. 2003. Prospecting for novel biocatalysts in a soil metagenome. *Applied and Environmental Microbiology,* 69**,** 6235-6242.

WEXLER, M., BOND, P. L., RICHARDSON, D. J. & JOHNSTON, A. W. B. 2005. A wide host-range metagenomic library from a waste water treatment plant yields a novel alcohol/aldehyde dehydrogenase. *Environmental Microbiology,* 7**,** 1917-1926.

WIDMER, F., SEIDLER, R. J., GILLEVET, P. M., WATRUD, L. S. & DI GIOVANNI, G. D. 1998. A highly selective PCR protocol for detecting 16S rRNA genes of the genus Pseudomonas (sensu stricto) in environmental samples. *Applied and Environmental Microbiology,* 64**,** 2545-2553.

CHAPTER FOUR:

APPLYING ILLUMINA SEQUENCING COUPLED WITH A METAGENOMIC CLONE LIBRARY TO IDENTIFY NOVEL PETROLEUM DEGRADING GENES

ABSTRACT

Recent studies on the biodiversity of asphalt deposits at the Rancho La Brea Tar Pits in Los Angeles, California have revealed the existence of several hundred new species of bacteria and gene sequences encoding putative novel degradative enzymes. The presence of fossilized extinct animal remains in the La Brea Tar Pits has led to estimations that these natural asphalt seeps have existed for at least 40,000 year. These deposits consist of petroleum that has been degraded to the extent that the remaining material is comprised mainly of asphalts and heavy oils, which have saturated into the soil matrix. Since petroleum hydrocarbons are both a target and a product of microbial metabolism, the role of microorganisms that inhabit this and other similar environments is directly relevant to development of technology for bioremediation, biotransformation of petroleum hydrocarbons, and microbial enhanced oil recovery for extracting and refining heavy oil.

The current investigation of the La Brea Tar pits concluded using Illumina technologies (Illumina, Inc.) to deep sequence the metagenomic library of over 3,000 clones using a high throughput sequencer. Over 75 MB of DNA has been sequenced, from which over 650 contigs with an average length of 500 bp were assembled. Bioinformatics analysis indicated the presence of genes encoding three types of dioxygenases, one of which encoded a naphthalene dioxygenase from *Pseudomonas sp.* along with two other genes that are most similar to

previously reported genes encoding biphenyl and toluene dioxygenases. However, sequence analysis revealed that these genes were not significantly similar to these known dioxygenases and may thus be novel. The results of this research provide a foundation for further studies on the evolution and assembly of metabolic pathways in bacteria that have undergone long term adaptation to survival in natural asphalts.

INTRODUCTION

Petroleum reservoirs, natural oil seeps, and oil contaminated environments select for microorganisms that transform and degrade the many different components that comprise petroleum hydrocarbons. Petroleum itself is a product derived from microbial transformations of organic substances that have then aged under pressure over millions of years, during which time it continues to undergo additional transformation by microbes that inhabit the oil. The metabolic versatility of these microorganisms holds great promise for bioremediation, industrial microbiology, development of better methods for oil recovery and processing, and the development of new petroleum based products. Nonetheless, our ability to access the genetic diversity of petroleum degrading communities has been a slow process, and mainly reflects the methods that have been available to microbiologists. The first decades of petroleum microbiology research began with plate cultures and biochemical studies with pure cultures to characterize the primary pathways. During the 1970s, DNA sequencing using Sanger dideoxy chain-termination methods led to many new advances and enabled discovery of new organisms and enzymes. This sequencing technology is still currently being used in many laboratories worldwide, providing more than 1 kb of sequencing per sample and has driven the first revolution in whole-genome sequencing. However, this technology has reached its plateau in the advancement and development of single-tube chemistry with fluorescently marked termination bases, heat stable polymerases, and automated capillary electrophoresis (Snyder et al., 2009). With the quest for ever more powerful methods, new technologies now take sequencing to another level in which it is possible to obtain millions of base reads and to sequence an entire microbial genome. The work of Venter and his colleagues in 2004 revealed

the diversity of protein-coding genes in ocean-dwelling microbes and thus showed the practical application of high-throughput sequencing for gene discovery (Venter et al., 2004). These tools offer similar promise for the study of microbial communities that inhabit petroleum deposits such as the Rancho La Brea Tarpits.

Next-generation sequencing (NGS) technologies produce data faster and more economically due to two fundamental differences in DNA sequencing. First, is elimination of the requirement for in vivo amplification of DNA fragments and the subsequent creation of a clone library, which is prone to clone biases and is very labor intensive. Second, the new sequencing chemistries enable samples to be sequenced in real time, as compared to Sanger sequencing in which the bases are called after the creation of dideoxy chain extension products. Since first introduced to the market in 2005, NGS has had a tremendous impact on genomic research. The technologies that are commercially available today include the 454 FLX Pyrosequencing based instrument (Roche Applied Science), the Illumina IIG analyzer (Illumina, Inc.) and SOLiD instrument from Applied Biosystems. These next-generation sequencing technologies offer dramatic increases in cost-effective sequence throughput, but at the expense of read lengths. The throughput of these sequencers is about the order of multiple gigabases per week.

Metagenomics has great potential for both fundamental and industrial applications that range from the understanding of microbial adaptation and evolution to the discovery of new enzymes (Ferrer et al., 2007, Ferrer et al., 2005, Lorenz and Eck, 2005, Deutschbauer et al., 2006). The application of metagenomics to the study of sites polluted with toxic chemicals and industrial wastes has transformed environmental biotechnology because these habitats include niches for microorganisms that have the necessary enzymes to use these compounds as their

carbon and energy sources. The genetic diversity in these environments includes genes encoding for degradative enzymes and pathways for recalcitrant chemicals, which may be used for the remediation of environmental pollution and as sources of novel catalytic activities useful in green chemistry and biotechnology (Galvao et al., 2005, Schmid et al., 2001). Still other genes of interest include those encoding for stress tolerance, heavy metal resistance, surfactant production, and enzymes involved in biogeochemical cycling of elements such as nitrogen, phosphorus, and sulfur.

In this study, I applied Illumina based next generation sequencing coupled with a fosmid library to discover new genes and catabolic pathways for petroleum degradation. This study represents a hybrid approach to metagenomics in that rather than shotgun sequencing of total DNA from microbial communities in oil, I used a previously constructed fosmid library consisting of 3,000 clones with DNA inserts of about 38 kb. By applying the Illumina technology to the fosmid library, I hypothesized that it should be possible to validate the final assembled reads by returning to the clone library to verify the full length sequences, and eventually express these genes and pathways in culturable organisms. There are currently no known studies in which only Illumina sequencing has been used to characterize a metagenome, which is constrained by the challenge of assembling DNA base reads of 30 to 40 bases into contigs that can be further assembled into gene sequences and pathways. This study thus represents a first application to test the efficacy of these methods for identifying genes encoding novel petroleum degrading enzymes, and insight into the types of genes that are carried by oil inhabiting microorganisms.

MATERIALS AND METHODS

Construction of Illumina sequencing library using Pit 101 metagenomic DNA. Previously unexposed samples of asphalt-soil aggregates were removed from approximately 10 cm under the surface of Pit 101 of the Rancho La Brea Tar Pits in Los Angeles, CA as previously described (Kim and Crowley, 2007). Samples were removed from the asphalt pits with sterile, autoclaved spatulas and were transferred into sterile sampling bags (Fisherbrand) for processing. The soil-tar samples were used to extract high molecular weight DNA about 38 Kb in length. DNA was extracted using an indirect cell lysis technique, which first separated the cells from the soil-tar matrix, which was then chemically lysed to extract DNA (Gabor et al., 2003). The extracted DNA was then purified by multiple low melting point agarose gel extractions using an agarase enzyme (Epicentre).

The metatgenomic library was constructed using a CopyControl™ Fosmid Library Production Kit (Epicentre) following the manufacturers' protocols. Briefly, DNA fragments of over 38 kb were selected and purified by gel electrophoresis. The resultant DNA was then blunt-end repaired and ligated to fosmid vector pCC1FOS™ and pCC2FOS™. The fosmids containing insert DNA were then packaged into Lambda Phage and used to infect *E. coli* strain EP1300™. A library of over 3,000 clones was picked and grown individually using LB medium supplemented with chloramphenicol in 384 well plates. The DNA insert of roughly 38 Kb was verified by digesting a random selection of fosmids using NOT I restriction enzyme (New England Biolabs).

To improve screening efficiency, the entire library was grown up using pooled fosmid samples with 24 clones per well. Each pool of 24 clones was grown in LB supplemented with chloramphenicol and fosmid induction solution to increase the number of fosmids per cell.

Fosmid DNA was extracted from each pool of clones using a FosmidMAX DNA Purification Kit (Epicentre). The fosmid DNA from the entire library was pooled together and sheared into 400 bp fragments for the Illumina sequencing library. The resulting fosmid DNA was used to construct the Illumina sequencing library with all manufacturer's protocols followed as directed.

Data analysis of Illumina sequencing library. All of the Illumina sequencing reads were processed using the UC Riverside Bioinformatics Bioconductor suite on a Linux server. The reads were assembled using Velvet and the contigs were analyzed using the NCBI, PFAM, KEGG and Uniprot databases. Phylogenetic analysis and tree construction of the contigs of interest were analyzed using Mega 4 software.

RESULTS

De novo assembly and gene annotation of Illumina paired-end sequencing data from the Rancho La Brea Tar Pits. Illumina sequencing yielded roughly two million reads from both paired end and single end libraries. Sequencing reads ranged from about 32 to 64 bp in length and the paired end reads were separated by 400 bp spacers. Both fosmid vector sequences and competent *E.* coli EP1300 (Epicentre) host cell sequences were removed prior to analysis. Velvet was used for contig assembly, in which over 75 Mb of DNA were assembled, resulting in 662 contigs with an average length about 500 bp. Gene annotation and data analysis was performed via UCR's bioconductor suite, in which contigs were analyzed using the NCBI, PFAM, KEGG and Uniprot databases where applicable on a Linux platform.

Of the 661 contigs, contigs of over 400 bp were used for statistical analysis. 64% of the contigs had significant protein matches in the PFAM, KEGG, Uniprot or NCBI databases of about 80% or greater (Table 4.1). Roughly 32% of the contigs were either found have no known function, of which 11.3% were predicted to encode a conserved protein of unknown function and 20.4% had no matches in any of the databases (Table 4.2). The majority of the contigs, 33.9% were found to be involved in the biosynthesis and metabolism of endogenous precursors and other organic molecules. Membrane proteins made up 10% of the contigs, while 8.1% were assigned to encode DNA binding proteins, 5.0% represented transporter proteins, 4.5% were involved in heavy metal resistance, another 4.5% were structural proteins and 2.3% contained sequences similar to known mobile genetic elements (Table 4.2).

Contig	Gene Product	Percent Identity	Organism or Metagenome
>NODE_1008_length_2387_cov_68.960617-001	Integrase catalytic region	100	*Pseudomonas aeruginosa*
>NODE_1008_length_2387_cov_68.960617-002	transposase IS3/IS911 family protein	100	*Pseudomonas aeruginosa*
>NODE_10220_length_587_cov_20.371380	NADH-quinone oxidoreductase subunit M	93	*Achromobacter xylosoxidans*
>NODE_10227_length_457_cov_8.901532	RND efflux system, outer membrane lipoprotein NodT family	100	*Cupriavidus taiwanensis*
>NODE_104_length_595_cov_612.329407	phosphatidylethanolamine-binding protein	100	*Uncultured bacterium clone*
>NODE_1042_length_5938_cov_69.700912-001	heavy metal efflux pump, CzcA family	99	*Cupriavidus metallidurans*
>NODE_1042_length_5938_cov_69.700912-002	amino acid permease	99	*Acidovorax citrulli*
>NODE_1075_length_654_cov_19.324160	Unknown	91	Marine
>NODE_1085_length_1620_cov_42.726543	inner membrane protein	100	*Ralstonia pickettii*
>NODE_11608_length_542_cov_12.798893	alkyl hydroperoxide reductase	94	*Achromobacter xylosoxidans*
>NODE_117884_length_525_cov_5.830476	putative mannitol-1-phosphate/altronate dehydrogenase	100	*Cupriavidus metallidurans*
>NODE_12072_length_1025_cov_7.729756	putative transport system permease protein	98	*Pseudomonas entomophila*
>NODE_125_length_6411_cov_66.789116-001	TonB-dependent siderophore receptor	99	*Ralstonia eutropha*
>NODE_125_length_6411_cov_66.789116-002	endonuclease/exonuclease/phosphatase	99	*Xanthomonas campestris*
>NODE_1256_length_433_cov_468.443420	conserved hypothetical protein	99	Uncultured prokaryote clone DeadSea
>NODE_1260_length_697_cov_89.822098-001	putative transposase insL for insertion sequence IS186	100	*Pseudomonas aeruginosa*
>NODE_1260_length_697_cov_89.822098-002	insertion sequence:IS185	100	*Pseudomonas aeruginosa*
>NODE_1260_length_697_cov_89.822098-003	transposase IS4 family protein	100	*Pseudomonas aeruginosa*
>NODE_1274_length_789_cov_41.910011	phosphopantetheinyltransferase	100	*Bordetella parapertussis*
>NODE_1286_length_1022_cov_19.143835	transcription-repair coupling factor	88	*Cupriavidus taiwanensis*
>NODE_12904_length_492_cov_16.756098	NADH-quinone oxidoreductase subunit B	92	*Achromobacter xylosoxidans*
>NODE_1386_length_400_cov_19.242500	DNA repair exonuclease	79	*Ralstonia solanacearum*
>NODE_1441_length_1166_cov_313.321625	ABC transporter related protein	100	*Kineococcus radiotolerans*
>NODE_145_length_390_cov_688.320496	NADH-ubiquinone oxidoreductase	94	*Rhodococcus opacus*
>NODE_1501_length_781_cov_266.678619	adenosylmethionine-8-amino-7-oxononanoate aminotransferase	100	*Cupriavidus taiwanensis*
>NODE_15178_length_470_cov_4.648936	conserved hypothetical protein	77	Uncultured prokaryote clone DeadSea
>NODE_15481_length_422_cov_38.386257	TonB-dependent siderophore receptor	77	*Pseudomonas stutzeri*
>NODE_1570_length_538_cov_18.000000	NADH-ubiquinone oxidoreductase	82	*Rhodococcus opacus*
>NODE_159_length_438_cov_586.815063	alkyl hydroperoxide reductase	95	*Achromobacter xylosoxidans*

Table 4.1. List of contigs that are greater than 400 bp in length along with their gene product, percent identity and organism of metagenome from which they originated. Gene products, percent identity and organisms/metagenomes were determined via NCBI, PFAM, KEGG and uniprot databases.

>NODE_1607_length_381_cov_14.438320	transcription-repair coupling factor	78	*Cupriavidus metallidurans*
>NODE_16298_length_520_cov_8.115385	conserved hypothetical protein	99	Uncultured prokaryote clone DeadSea
>NODE_1638_length_669_cov_27.934231	putative RNA polymerase	75	*Methylibium petroleiphilum*
>NODE_1644_length_771_cov_34.351490	transport system permease protein	99	*Methylibium petroleiphilum*
>NODE_1686_length_1430_cov_46.735664	MbtH domain protein	99	*Cupriavidus taiwanensis*
>NODE_16912_length_572_cov_23.396852	Unknown	85	Epibiont
>NODE_1707_length_763_cov_82.158585	DNA-binding transcriptional regulator	99	*Ralstonia pickettii*
>NODE_1717_length_515_cov_27.388350	Unknown	88	Marine
>NODE_17178_length_560_cov_11.273214	Unknown	81	Marine
>NODE_17219_length_811_cov_13.865598	conserved hypothetical protein	81	*Rhodococcus opacus*
>NODE_1739_length_996_cov_55.507027	two component heavy metal response transcriptional regulator	99	*Variovorax paradoxus*
>NODE_1764_length_769_cov_31.798439	8-amino-7-oxononanoate synthase	99	*Ralstonia pickettii*
>NODE_17877_length_644_cov_8.350932	pectinesterase	100	*Ralstonia solanacearum*
>NODE_1799_length_666_cov_26.527027	NADH-ubiquinone oxidoreductase	89	*Rhodococcus opacus*
>NODE_1858_length_1021_cov_48.278160	transport system permease protein	100	*Methylibium petroleiphilum*
>NODE_1863_length_687_cov_22.163027	MscS Mechanosensitive ion channel	100	*Ralstonia eutropha*
>NODE_1938_length_680_cov_19.130882	Unknown	86	Clay loam
>NODE_1942_length_901_cov_25.855717	aminopeptidase N. Metallo peptidase	77	*Methylobacterium chloromethanicum*
>NODE_20350_length_1807_cov_9.537354	NADH oxidoreductase (quinone), F subunit	95	*Achromobacter xylosoxidans*
>NODE_2040_length_383_cov_20.527414	Unknown	87	Marine
>NODE_2045_length_1473_cov_25.630007	formate dehydrogenase	87	*Methylibium petroleiphilum*
>NODE_20792_length_404_cov_24.938118	protein of unknown function	100	*Methylobacterium populi*
>NODE_2161_length_739_cov_21.713125	Unknown	83	Compost
>NODE_2175_length_926_cov_49.263500	MscS Mechanosensitive ion channel	100	*Ralstonia eutropha*
>NODE_2176_length_901_cov_21.308546	Unknown	71	Compost
>NODE_21873_length_1053_cov_14.842355	NADH-quinone oxidoreductase subunit J	95	*Achromobacter xylosoxidans*
>NODE_2198_length_477_cov_22.060797	NADH-ubiquinone oxidoreductase	85	*Ralstonia eutropha*
>NODE_2279_length_2989_cov_18.645031	metallophosphoesterase	87	*Variovorax paradoxus*
>NODE_2294_length_915_cov_23.209837	Unknown	83	Marine
>NODE_2296_length_2360_cov_16.805508	deaminase reductase	88	*Rhodococcus jostii*

Table 4.1. (cont) List of contigs that are greater than 400 bp in length along with their gene product, percent identity and organism of metagenome from which they originated. Gene products, percent identity and organisms/metagenomes were determined via NCBI, PFAM, KEGG and uniprot databases.

>NODE_2307_length_597_cov_23.534338	Unknown	74	Marine
>NODE_2308_length_533_cov_23.146341	Unknown	85	Freshwater sediment
>NODE_2411_length_1016_cov_21.068897	RelA/SpoT domain protein	85	*Rhodococcus erythropolis*
>NODE_2521_length_450_cov_25.733334	transcription-repair coupling factor	91	*Cupriavidus taiwanensis*
>NODE_2530_length_504_cov_18.684525	major facilitator superfamily MFS_1	100	*Methylibium petroleiphilum*
>NODE_2550_length_594_cov_19.154882	gamma-glutamyltranspeptidase	79	*Cupriavidus metallidurans*
>NODE_257_length_1275_cov_448.960785	methltransferase	99	Uncultured bacterium clone
>NODE_2571_length_721_cov_16.051317	Unknown	83	Fossil
>NODE_25743_length_893_cov_14.828668	NADH dehydrogenase (quinone), D subunit	94	*Achromobacter xylosoxidans*
>NODE_2600_length_583_cov_25.861063	NADH-ubiquinone oxidoreductase	95	*Cupriavidus taiwanensis*
>NODE_2615_length_608_cov_21.340460	transport system permease protein	100	*Methylibium petroleiphilum*
>NODE_26234_length_898_cov_11.040089	conserved hypothetical protein	97	Uncultured prokaryote clone DeadSea
>NODE_2637_length_447_cov_27.342281	transcription-repair coupling factor	92	*Cupriavidus taiwanensis*
>NODE_2642_length_682_cov_18.167156	Unknown	89	Mine drainage
>NODE_2648_length_491_cov_29.327902	Unknown	90	Clay loam
>NODE_2654_length_379_cov_23.865435	Unknown	76	Mine drainage
>NODE_2664_length_453_cov_21.794466	Unknown	85	Marine
>NODE_2701_length_742_cov_20.396227	NADH dehydrogenase I, chain D	93	*Rhodococcus opacus*
>NODE_2717_length_954_cov_17.098532	Unknown	88	Ant fungus
>NODE_2745_length_438_cov_15.956621	outer membrane porin protein	82	*Bordetella parapertussis*
>NODE_27511_length_559_cov_19.103756	gamma-glutamyltranspeptidase	89	*Cupriavidus metallidurans*
>NODE_2756_length_435_cov_31.331034	methyltransferase	100	*Cupravidus necator*
>NODE_27948_length_597_cov_15.237856	Unknown	80	Freshwater sediment
>NODE_27972_length_448_cov_10.569197	metal-dependent hydrolase	99	Uncultured prokaryote clone
>NODE_2813_length_431_cov_22.410673	TonB-dependent siderophore receptor	77	*Pseudomonas stutzeri*
>NODE_28191_length_472_cov_7.529661	protein of unknown function	98	*Methylobacterium populi*
>NODE_2828_length_577_cov_63.911613	phosphatidate phosphatase	99	*Cupriavidus metallidurans*
>NODE_28451_length_387_cov_43.441860	2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase	80	*Pseudomonas putida*
>NODE_2899_length_401_cov_21.024939	deaminase reductase	89	*Rhodococcus jostii*
>NODE_2918_length_4190_cov_19.542244	transcription-repair coupling factor	86	*Cupriavidus taiwanensis*

Table 4.1. (cont) List of contigs that are greater than 400 bp in length along with their gene product, percent identity and organism of metagenome from which they originated. Gene products, percent identity and organisms/metagenomes were determined via NCBI, PFAM, KEGG and uniprot databases.

>NODE_29243_length_682_cov_18.167156	Unknown	89	Fossil
>NODE_2927_length_1009_cov_18.769079	Unknown	88	Mine drainage
>NODE_2934_length_473_cov_19.253700	putative secreted peptidase	85	*Streptomyces coelicolor*
>NODE_2943_length_1447_cov_14.255701	NADH-ubiquinone oxidoreductase	91	*Ralstonia eutropha*
>NODE_295_length_613_cov_19.295269	dethiobiotin synthase	100	*Rhodococcus opacus*
>NODE_2954_length_767_cov_23.736637	conserved hypothetical protein	82	*Rhodococcus opacus*
>NODE_2965_length_840_cov_35.046429	6,7-dihydropteridine reductase	99	*Bordetella parapertussis*
>NODE_2972_length_5919_cov_19.851326	putative membrane protein	88	*Kineococcus radiotolerans*
>NODE_3_length_426_cov_11.354460	Unknown	89	Stromalite
>NODE_3026_length_1291_cov_15.422928	peptidase carboxypeptidase	76	*Bordetella pertussis*
>NODE_30478_length_524_cov_17.979008	NADH dehydrogenase (quinone), D subunit	95	*Achromobacter xylosoxidans*
>NODE_30729_length_632_cov_10.639240	Ribose/xylose/arabinose/galactoside transport systems	99	*Variovorax paradoxus*
>NODE_3122_length_521_cov_18.155470	Unknown	84	*Freshwater sediment*
>NODE_31290_length_427_cov_15.859485	two component heavy metal response transcriptional regulator	99	*Variovorax paradoxus*
>NODE_3157_length_918_cov_17.008715	c4-dicarboxylate transport protein	86	*Rhodococcus jostii*
>NODE_3172_length_422_cov_26.180096	NADH-ubiquinone oxidoreductase	86	*Ralstonia eutropha*
>NODE_31868_length_784_cov_13.899235	transcription-repair coupling factor	93	*Achromobacter xylosoxidans*
>NODE_3213_length_484_cov_21.070248	Unknown	97	Human gut
>NODE_3214_length_462_cov_17.186148	pyridoxamine 5'-phosphate oxidase-related FMN-binding	81	*Conexibacter woesei*
>NODE_32200_length_552_cov_8.278986	altronate oxidoreductase	98	*Methylibium petroleiphilum*
>NODE_33060_length_617_cov_15.233387	Unknown	88	Fungus garden
>NODE_3311_length_1060_cov_11.802830	putative antioxidant protein	90	*Cupriavidus taiwanensis*
>NODE_3362_length_638_cov_15.484326	NADH-ubiquinone oxidoreductase	86	*Ralstonia eutropha*
>NODE_33924_length_658_cov_11.539514	aldehyde dehydrogenase	98	*Rhodopseudomonas palustris*
>NODE_3396_length_551_cov_17.177858	2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase	99	*Pseudomonas putida*
>NODE_34099_length_398_cov_4.364322	protein of unknown function	100	*Methylobacterium populi*
>NODE_34212_length_400_cov_21.040001	copper/silver efflux system protein CusC	100	*Cupriavidus metallidurans*
>NODE_34367_length_377_cov_38.018566	molybdenum cofactor biosynthesis protein	98	*Variovorax paradoxus*
>NODE_34731_length_566_cov_15.637809	Unknown	91	Mine drainage
>NODE_348_length_429_cov_19.734266	Unknown	78	Clay loam

Table 4.1. (cont) List of contigs that are greater than 400 bp in length along with their gene product, percent identity and organism of metagenome from which they originated. Gene products, percent identity and organisms/metagenomes were determined via NCBI, PFAM, KEGG and uniprot databases.

>NODE_34803_length_450_cov_18.862223	molybdopterin biosynthesis MoaE protein	100	*Variovorax paradoxus*
>NODE_3484_length_439_cov_18.794989	protein of unknown function	100	*Methylobacterium populi*
>NODE_3515_length_682_cov_20.485336	transposase insL for insertion sequence IS186	100	*Rhodococcus opacus*
>NODE_354_length_584_cov_624.654114	putative dioxygenase	100	*Cupravidus necator*
>NODE_355_length_617_cov_664.849243	putative dioxygenase	98	*Cupravidus necator*
>NODE_35503_length_477_cov_21.492662	Unknown	68	Compost
>NODE_3553_length_421_cov_15.767220	Unknown	95	Human gut
>NODE_35666_length_378_cov_7.841270	insertion sequence IS185	100	*Pseudomonas aeruginosa*
>NODE_35869_length_522_cov_29.821838	conserved hypothetical protein	98	Uncultured prokaryote clone DeadSea
>NODE_36056_length_701_cov_11.449358	conserved hypothetical protein	99	Uncultured prokaryote clone DeadSea
>NODE_3626_length_3580_cov_21.209217	putative RNA polymerase, sigma-24 subunit	76	*Methylibium petroleiphilum*
>NODE_3648_length_1841_cov_16.704508	NADH-ubiquinone oxidoreductase	89	*Ralstonia eutropha*
>NODE_365_length_1699_cov_14.843437	adenosylmethionine-8-amino-7-oxononanoate transferase	99	*Cupriavidus taiwanensis*
>NODE_36985_length_541_cov_23.194084	conserved hypothetical protein	98	*Achromobacter xylosoxidans*
>NODE_3708_length_894_cov_16.731544	sensor kinase protein	87	*Methylobacterium chloromethanicum*
>NODE_38142_length_501_cov_18.181637	Unknown	80	Stromalite
>NODE_38247_length_904_cov_9.401548	Unknown	79	Stromalite
>NODE_3851_length_383_cov_24.028721	Unknown	95	Human gut
>NODE_39587_length_398_cov_6.447236	Unknown	91	Microbial mat
>NODE_4010_length_4431_cov_20.541412	RelA/SpoT domain protein	85	*Rhodococcus erythropolis*
>NODE_40121_length_451_cov_11.447893	putative dioxygenase	88	*Cupravidus necator*
>NODE_4070_length_1595_cov_16.494671	Unknown	87	Mine drainage
>NODE_40811_length_587_cov_5.381601	conserved hypothetical protein	99	*Achromobacter xylosoxidans*
>NODE_41395_length_389_cov_10.953728	conserved hypothetical protein	81	*Achromobacter xylosoxidans*
>NODE_41759_length_421_cov_4.570071	two component heavy metal response transcriptional regulator	99	*Variovorax paradoxus*
>NODE_420_length_575_cov_29.766956	naphthalene dioxygenase	93	*Pseudomonas aeruginosa*
>NODE_42671_length_638_cov_15.537618	NADH dehydrogenase (quinone), G subunit	92	*Achromobacter xylosoxidans*
>NODE_4372_length_673_cov_20.882616	Unknown	88	Fungus garden
>NODE_4480_length_2350_cov_17.249361	deaminase reductase	87	*Rhodococcus jostii*
>NODE_4520_length_818_cov_24.751833	Unknown	90	Compost

Table 4.1. (cont) List of contigs that are greater than 400 bp in length along with their gene product, percent identity and organism of metagenome from which they originated. Gene products, percent identity and organisms/metagenomes were determined via NCBI, PFAM, KEGG and uniprot databases.

>NODE_4573_length_2753_cov_24.397747	aminopeptidase N. Metallo peptidase	78	*Methylobacterium chloromethanicum*
>NODE_46062_length_443_cov_5.106095	metal-dependent hydrolase	99	Uncultured prokaryote clone
>NODE_4705_length_652_cov_10.337423	molybdopterin converting factor, subunit 1	100	*Variovorax paradoxus*
>NODE_47429_length_766_cov_7.152741	low-specificity L-threonine aldolase	98	*Cupriavidus taiwanensis*
>NODE_4860_length_994_cov_15.377264	phosphatidate phosphatase	80	*Achromobacter xylosoxidans*
>NODE_4950_length_620_cov_10.656451	protein of unknown function	100	*Methylobacterium populi*
>NODE_51010_length_625_cov_5.969600	putative oxidoreductase	98	*Ralstonia pickettii*
>NODE_5108_length_3723_cov_12.189900-001	glyoxalase/bleomycin resistance protein/dioxygenase	94	*Achromobacter xylosoxidans*
>NODE_5108_length_3723_cov_12.189900-002	isochorismatase	94	*Pseudomonas putida*
>NODE_5108_length_3723_cov_12.189900-003	major facilitator superfamily MFS_1	94	*Methylibium petroleiphilum*
>NODE_52_length_843_cov_365.517212	metal-dependent hydrolase	99	Uncultured prokaryote clone
>NODE_5210_length_1100_cov_12.474545	Unknown	93	Mine drainage
>NODE_5273_length_1286_cov_20.330482	TonB-dependent siderophore receptor	81	*Ralstonia eutropha*
>NODE_528_length_1684_cov_15.581947	conserved hypothetical protein	82	*Achromobacter xylosoxidans*
>NODE_5280_length_1224_cov_20.428921	NADH dehydrogenase (quinone), G subunit	92	*Achromobacter xylosoxidans*
>NODE_53304_length_496_cov_6.955645	conserved hypothetical protein	99	*Achromobacter xylosoxidans*
>NODE_55024_length_383_cov_24.028721	Unknown	84	Stromalite
>NODE_5604_length_525_cov_13.320000	Unknown	90	Fungus garden
>NODE_57574_length_379_cov_6.031662	protein of unknown function	100	*Methylobacterium populi*
>NODE_58_length_2530_cov_47.362450-001	two component heavy metal response transcriptional regulator	99	*Variovorax paradoxus*
>NODE_58_length_2530_cov_47.362450-002	heavy metal sensor signal transduction histidine kinase	99	*Cupriavidus metalliduran*s
>NODE_5808_length_3737_cov_19.962805	putative secreted peptidase	85	*Streptomyces coelicolor*
>NODE_5872_length_478_cov_18.271967	biotin biosynthesis	100	*Thiobacillus denitrificans*
>NODE_58996_length_415_cov_3.720482	putative dioxygenase	87	*Cupravidus necator*
>NODE_5967_length_861_cov_15.252032	TonB-dependent receptor	77	*Bordetella pertussis*
>NODE_60131_length_846_cov_9.026005	adenylosuccinate lyase	98	*Pseudomonas stutzeri*
>NODE_6060_length_1568_cov_9.480229	conserved hypothetical protein	97	*Achromobacter xylosoxidans*
>NODE_6063_length_944_cov_16.697035	outer membrane porin protein	81	*Bordetella parapertussis*
>NODE_6067_length_669_cov_7.264574	putative phage terminase, large subunit	97	*Salmonella enterica*
>NODE_60959_length_540_cov_6.155556	dipeptidyl carboxypeptidase	99	*Pseudomonas stutzeri*

Table 4.1. (cont) List of contigs that are greater than 400 bp in length along with their gene product, percent identity and organism of metagenome from which they originated. Gene products, percent identity and organisms/metagenomes were determined via NCBI, PFAM, KEGG and uniprot databases.

>NODE_616_length_518_cov_407.252899	endonuclease/exonuclease/phosphatase	98	*Xanthomonas campestris*
>NODE_6224_length_568_cov_13.825705	putative membrane protein	100	*Kineococcus radiotolerans*
>NODE_6321_length_472_cov_16.389830	metallophosphoesterase	94	*Cupravidus necator*
>NODE_6455_length_2012_cov_8.422465	Unknown	99	Mine drainage
>NODE_649_length_2646_cov_17.698792	Unknown	91	Mine drainage
>NODE_651_length_409_cov_544.929077	transposase insL for insertion sequence IS187	100	*Rhodococcus opacus*
>NODE_6552_length_830_cov_22.986748	excision nuclease subunit B	100	*Kribbella flavida*
>NODE_65631_length_498_cov_9.012048	Unknown	96	Mine drainage
>NODE_6574_length_580_cov_11.993103	transcription-repair coupling factor	89	*Achromobacter xylosoxidans*
>NODE_6815_length_440_cov_8.195455	Unknown	87	Stromalite
>NODE_6825_length_451_cov_22.223948	excision nuclease subunit B	100	*Kribbella flavida*
>NODE_6836_length_864_cov_8.939815	Unknown	90	Compost
>NODE_6936_length_904_cov_14.875000	protein of unknown function	100	*Methylobacterium populi*
>NODE_6943_length_1159_cov_22.192408	TonB-dependent siderophore receptor	78	*Ralstonia eutropha*
>NODE_6985_length_792_cov_15.737373	Unknown	77	Ant fungus
>NODE_70300_length_684_cov_5.489766	conserved hypothetical protein	96	*Achromobacter xylosoxidans*
>NODE_7151_length_429_cov_21.783216	head-tail joining protein	97	*Bordetella parapertussis*
>NODE_7262_length_757_cov_12.742404	putative structural protein	100	*Ralstonia pickettii*
>NODE_743_length_583_cov_261.926239	adenosylmethionine-8-amino-7-oxononanoate transferase	100	*Cupriavidus taiwanensis*
>NODE_749_length_418_cov_9.770335	3-isopropylmalate dehydrogenase	78	*Ralstonia pickettii*
>NODE_7540_length_1668_cov_18.490408	glyoxalase/bleomycin resistance protein/dioxygenase	84	*Thermomonospora curvata*
>NODE_7594_length_623_cov_7.133226	outer membrane porin protein	90	*Achromobacter xylosoxidans*
>NODE_7744_length_603_cov_18.600332	conserved hypothetical protein	83	*Achromobacter xylosoxidans*
>NODE_7783_length_497_cov_19.935614	Unknown	85	Marine
>NODE_784_length_1200_cov_31.331667	RNA helicase	100	*Chromobacterium violaceum*
>NODE_789_length_442_cov_692.773743	molybdopterin converting factor, subunit 1	99	*Variovorax paradoxus*
>NODE_794_length_614_cov_313.040710	naphthalene dioxygenase	96	*Pseudomonas aeruginosa*
>NODE_80_length_382_cov_470.039276	two component heavy metal response transcriptional regulator	99	*Variovorax paradoxus*
>NODE_8084_length_673_cov_11.335810	RisS protein	93	*Achromobacter xylosoxidans*
>NODE_81_length_1479_cov_118.091957	secretion protein HlyD protein	100	Uncultured prokaryote clone

Table 4.1. (cont) List of contigs that are greater than 400 bp in length along with their gene product, percent identity and organism of metagenome from which they originated. Gene products, percent identity and organisms/metagenomes were determined via NCBI, PFAM, KEGG and uniprot databases.

>NODE_81359_length_405_cov_5.869136	transposase IS4 family protein	100	*Pseudomonas aeruginosa*
>NODE_8208_length_693_cov_11.484848	3-carboxy-cis,cis-muconate cycloisomerase	84	*Rhodopseudomonas palustris*
>NODE_82227_length_391_cov_3.710997	NADH-ubiquinone oxidoreductase	89	*Ralstonia eutropha*
>NODE_8628_length_644_cov_18.183229	twin-arginine translocation	79	*Beutenbergia cavernae*
>NODE_86431_length_436_cov_7.605505	putative integral membrane protein	98	*Rhodospirillum rubrum*
>NODE_87929_length_390_cov_3.074359	protein of unknown function	98	*Methylobacterium populi*
>NODE_885_length_589_cov_31.568760	adenosylmethionine-8-amino-7-oxononanoate transferase	99	*Cupriavidus taiwanensis*
>NODE_8943_length_673_cov_7.310550	sodium:dicarboxylate symporter	79	*Methylobacterium populi*
>NODE_9023_length_719_cov_33.815022	conserved hypothetical protein	97	*Achromobacter xylosoxidans*
>NODE_911_length_6094_cov_85.537903-001	secretory system II protein	100	Uncultured prokaryote clone DeadSea
>NODE_911_length_6094_cov_85.537903-002	TPR repeat-containing protein	100	*Thiobacillus denitrificans*
>NODE_934_length_570_cov_583.584229	DNA packaging protein	99	*Bordetella parapertussis*
>NODE_95_length_1242_cov_29.065218	dethiobiotin synthase	100	*Rhodococcus opacus*
>NODE_95402_length_409_cov_5.608802	insertion sequence IS185	100	*Pseudomonas aeruginosa*
>NODE_958_length_420_cov_287.842865	ABC transporter related protein	99	*Kineococcus radiotolerans*
>NODE_96_length_1884_cov_137.957001	ABC transporter related protein	99	*Kineococcus radiotolerans*
>NODE_1315_length_1146_cov_21.297556	NADH-ubiquinone oxidoreductase	90	*Rhodococcus jostii*

Table 4.1. (cont) List of contigs that are greater than 400 bp in length along with their gene product, percent identity and organism of metagenome from which they originated. Gene products, percent identity and organisms/metagenomes were determined via NCBI, PFAM, KEGG and uniprot databases.

Protein Type	% Abundance
Biosynthesis and metabolism	33.9%
Conserved proteins	11.3%
Membrane proteins	10.0%
DNA-Binding proteins	8.1%
Transport	5.0%
Heavy metal resistance	4.5%
Structural proteins	4.5%
Mobile genetic elements	2.3%
Unknown	20.4%

Table 4.2. Distribution of contigs into different protein categories.

Identification of petroleum degrading genes. Among the assembled contigs, there were six that encoded dioxygenases (Table 4.1). Phylogenetic analysis indicated that there were two contigs, [NODE 794 length 614 cov 313.040710] and [NODE 420 length 575 cov 29.766956], which were similar to naphthalene dioxygenase genes from *Pseudomonas stutzeri*, *Pseudomonas aeruginosa* and those previously found from the Rancho La Brea Tar Pits (Baquiran and Crowley, unpublished data). There were also two others, [NODE 354 length 584 cov 624.654114] and [NODE 355 length 617 cov 664.84924], which appeared to be most similar to toluene or biphenyl dioxygenases, but no significant similarity was found. The last two contigs [NODE 40121 length 451 cov 11.447893] and [NODE 58996 length 415 cov 3.720482] were most closely related to an aromatic ring hydroxylating dioxygenase from *Escherichia coli*, however there was no statistically significant similarity (Figure 4.1).

Organism and metagenome distribution of the contigs. In addition to gene annotation, the contigs were also assigned to specific organisms. Approximately 39% of the organisms belonged to the class Betaproteobacteria, 25% belonged to Actinobacteria, 17% belonged to Gammaproteobacteria, 11% belonged to Alphaproteobacteria and 8% were unknown (Figure 4.2). The most abundant organism was *Achromobacter xylosoxidans*, which was represented by 13.8% of the contigs. The next most abundant was *Cupriavidus taiwanensis* at 7.2%, then *Variovorax paradoxus* and *Ralstonia eutropha* at 6.1%. *Pseudomonas aeurginosa* and *Rhodoccus opacus* were both represented by 5.5% of the contigs, along with many other organisms, which were represented by less than 5% of the contigs (Figure 4.3).

For contigs that were not assigned a function or organism, a preliminary classification was attempted by comparing them to unknown sequences from metagenomes from other

environments. The majority of the unclassified contigs, 20.0% were most similar to those from the acid mine drainage metagenome. The next highest similar metagenome was the Antarctic Ocean at 17.8%, and the compost and stromalite metagenome were next at 11.8%. There were eight other metagenomes that were represented at 6.7% or less (Table 4.3).

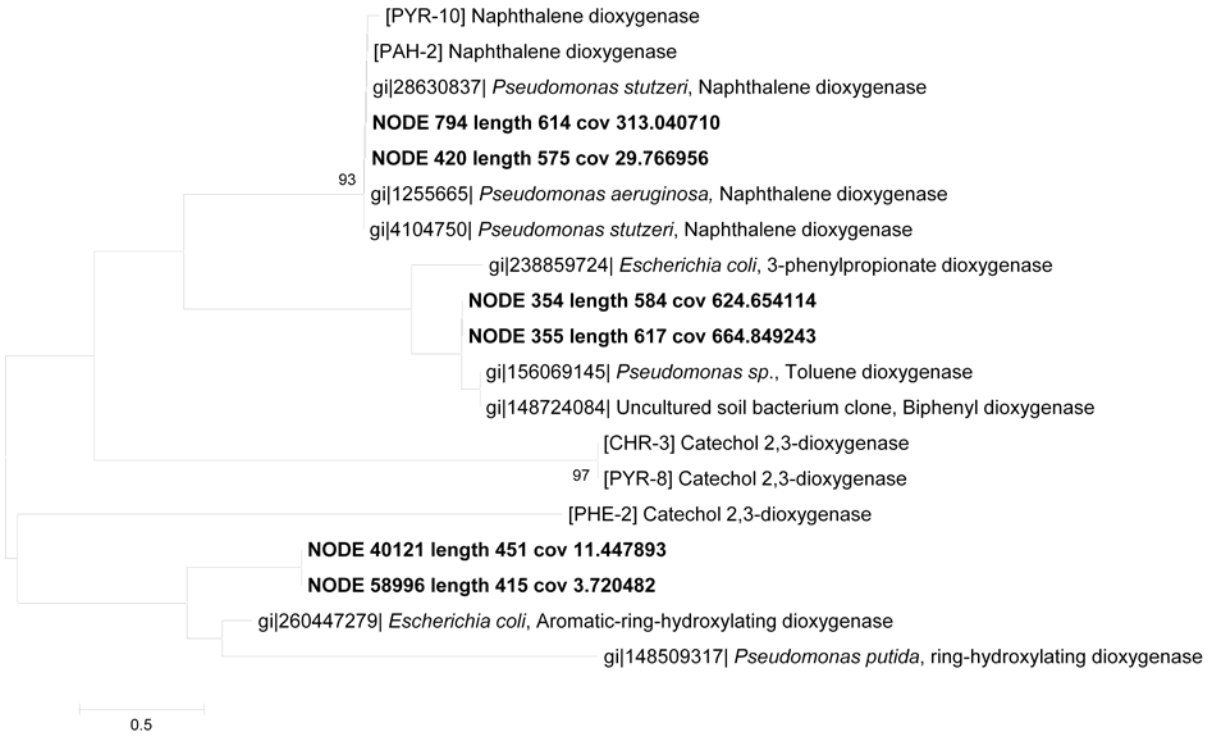

Figure 4.1. Phylogenetic analysis of dioxygenases detected on contigs **(in bold)** with other known dioxygenase sequences from the NCBI database. The evolutionary history was inferred using the Neighbor-Joining method. The optimal tree with the sum of branch length = 10.07987831 is shown. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown (>85) next to the branches. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree.

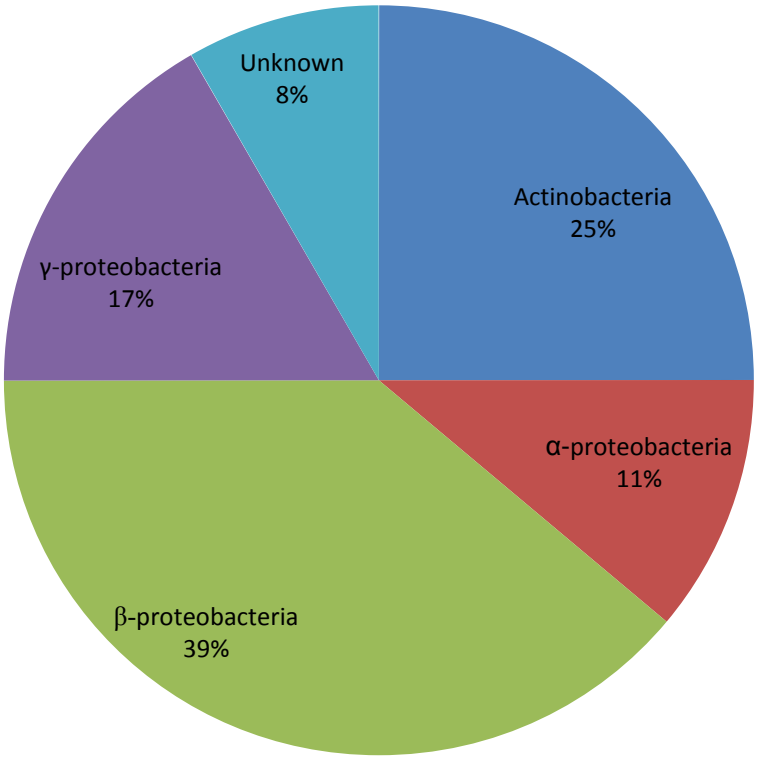

Figure 4.2. Distribution of organisms from phylogenetic classification of contigs classified by class.

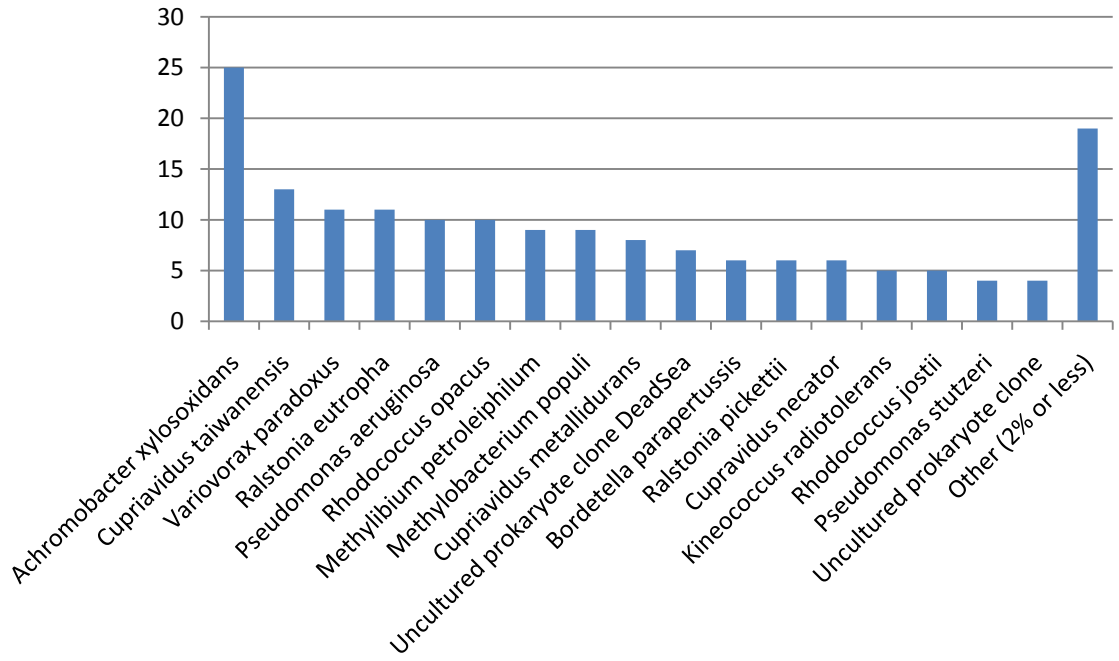

Figure 4.3. Relative distribution organisms from phylogenetic assessment of contigs.

Metagenome	% Abundance
Mine drainage	20.0%
Marine	17.8%
Compost	11.1%
Stromalite	11.1%
Clay loam	6.7%
Freshwater sediment	6.7%
Fungus garden	6.7%
Human gut	6.7%
Ant fungus	4.4%
Fossil	4.4%
Epibiont	2.2%
Microbial mat	2.2%

Table 4.3. Distribution of unknown contigs classified by similar sequences from different metagenomes.

DISCUSSION

The results from this study confirmed the presence of a diverse microbial community in the Rancho La Brea asphalts, which is comprised of many previously uncharacterized microorganisms that inhabit a soil matrix saturated with heavy petroleum hydrocarbons. Roughly 75 Mb of DNA was assembled, resulting in 661 different contigs of an average length of only 500 bp. Many of these contigs could be assigned to particular functional genes, but as expected there were many novel sequences that could not be identified. Using this technology, it was also apparent that much more sequencing would be needed to assemble large scaffolds that are necessary for constructing composite genomes to infer organism-function and evolutionary relationships.

The contigs that were assigned a function via the KEGG, PFAM, NCBI and Uniprot databases suggested a low diversity of genes relative to the amount of DNA sequenced. There are a number of reasons why this may have occurred (e.g. cloning biases of the fosmid library and sequencing biases of the Illumina library). Nevertheless, there were several types of genes that may be of interest with respect to microbial life in natural asphalts. There were only six contigs that encoded petroleum hydrocarbon degrading genes and phylogenetic analysis revealed 3 different types of dioxygenases. The first type was a known naphthalene dioxygenase that belongs to *Pseudomonas aeurginosa*, which is consistent with previous studies on Pit 101 and appears to be a major degradative enzyme in the La Brea Tar Pits. The second type of dioxygenase discovered here was most similar to the toluene and biphenyl type dioxygenases, and the third was most closely related to an aromatic ring hydroxylating dioxygenase. However, because of the relatively low similarity of these new genes to known genes in the databases

used here, their function could not be inferred via sequence data alone. The next approach would be to find these dioxygenases in the fosmid library to get the full gene sequence and even an entire catabolic pathway, and then attempt to express these in a culturable bacterium.

In addition to these degradative genes, mobile genetic elements (MGEs) were found on several contigs. The finding of insertion sequences, transpose regions and integrases imply that horizontal gene transfer occurs within the microbial community of Pit 101. All of the MGEs detected belonged to the same types previously found in *Pseudomonas aeurginosa* and suggest that this organism is at least partially responsible for harboring and disseminating petroleum degrading enzymes within this microbial community. It is also known that organisms from *Pseudomonas sp.* are facilitators of horizontal gene transfer (Nojiri et al., 2004, Stuart-Keil et al., 1998), meaning that these MGEs may be have value for adapting microbial communities to the presence of petroleum following a contamination event in soil without prior history of exposure to petroleum hydrocarbons.

Unlike gene function, the distribution of organisms in which the contigs belong, displayed a wider diversity. Many different organisms that are known to degrade petroleum hydrocarbons including several that have been found in previous studies on Pit 101 were detected. The most abundant organism, i.e. that was represented by the most contigs, was *Achromobacter xylosoxidans*, which can degrade many petroleum hydrocarbons such as polychlorinated biphenyls, polycyclic aromatic hydrocarbons and MTBE (Eixarch and Constanti, 2010, Jencova et al., 2004, Medina-Moreno et al., 2005). The next most represented organism was *Cupriavidus taiwanensis*, which is closely related to *Ralsonia eutropha.* Both of these bacteria are well known for their ability to degrade organic compounds in heavy metal

contaminated environments (Chen et al., 2008, Ganesh-Kumar et al., 2010, Hussain et al., 2009, Perez-Pantoja et al., 2009), and both were also previously isolated from Pit 101 using phenanthrene as a sole carbon source (Baquiran and Crowley, unpublished data). Detecting these two organisms in both the Illumina library and through enrichment cultures, suggests that they are dominant petroleum degraders in the Rancho La Brea Tar Pits. In addition to the above, there were many other organisms detected through Illumina sequencing that are well known petroleum hydrocarbon degraders, however because of the many relatively short contigs, more in-depth sequencing would is needed to determine their metabolic roles within this environment.

As expected, there were also a high amount of contigs that were unknown to any of the databases. The occurrence of unknown genes found in metagenomic studies is a common occurrence as previously unknown species are being sequenced (Venter et al., 2004, Handelsman, 2005). Even though the gene functions are unknown, one way that such data sets are now being analyzed is through comparative metagenomics, in which quantitative gene content analysis reveals habitat-specific fingerprints that reflect known characteristics of the sampled environments (Tringe et al., 2005). The identification of environment-specific genes through a gene-centric comparative analysis allows for interpreting and diagnosing metadata from complex environments. Comparative metagenomics has been successfully used to compare a wide range of environments, such as the human gut microbiota, farm soil and whale remains (Bodaker et al., 2010, Kurokawa et al., 2007, Quaiser et al., 2008, Tringe et al., 2005). In analyzing the unknown contigs from this environment, the majority of the contigs were most closely related to those from an acid mine biofilm. This preliminary analysis suggests that

because the acid mine environment revealed similar pathways for carbon and nitrogen fixation and energy generation (Tyson et al., 2004), both environments may require similar functional capabilities for survival in extreme environments.

Applying Illumina sequencing to characterize the Pit 101 metagenome, made clear the need for longer reads and more sequencing data in general. Two million reads of 36 to 72 bp was not enough to characterize a complex metagenome such as this, with only 75 Mb of DNA being assembled. Short reads from next generation sequencing technologies need roughly 10-12x coverage to fully sequence a single genome, so it is apparent that we have only partially sequenced this environment. Although, we have satisfactory data in finding novel dioxygenases and detecting more petroleum degrading organisms, there is much more to be discovered. We know that there are many more organisms and petroleum degrading genes detected by Kim and Crowley, which were not represented in the assembled contigs. However, by reverse blasting those same organisms and genes to the primary Illumina reads, we found that there are reads that contain the same sequences (data not shown). This suggests that although we need more sequencing to accomplish our primary objective of sequencing whole catabolic pathways, approach used here provides a viable step towards this goal.

BIBLIOGRAPHY

ATLAS, R. M. 1995. Petroleum biodegradation and oil spill bioremediation. *Marine Pollution Bulletin,* 31**,** 178-182.

ATLAS, R. M. & CERNIGLIA, C. E. 1995. BIOREMEDIATION OF PETROLEUM POLLUTANTS - DIVERSITY AND ENVIRONMENTAL ASPECTS OF HYDROCARBON BIODEGRADATION. *Bioscience,* 45**,** 332-338.

BODAKER, I., SHARON, I., SUZUKI, M. T., FEINGERSCH, R., SHMOISH, M., ANDREISHCHEVA, E., SOGIN, M. L., ROSENBERG, M., MAGUIRE, M. E., BELKIN, S., OREN, A. & BEJA, O. 2010. Comparative community genomics in the Dead Sea: an increasingly extreme environment. *Isme Journal,* 4**,** 399-407.

BOGARDT, A. H. & HEMMINGSEN, B. B. 1992. ENUMERATION OF PHENANTHRENE-DEGRADING BACTERIA BY AN OVERLAYER TECHNIQUE AND ITS USE IN EVALUATION OF PETROLEUM-CONTAMINATED SITES. *Applied and Environmental Microbiology,* 58**,** 2579-2582.

CHEN, W. M., WU, C. H., JAMES, E. K. & CHANG, J. S. 2008. Metal biosorption capability of Cupriavidus taiwanensis and its effects on heavy metal removal by nodulated Mimosa pudica. *Journal of Hazardous Materials,* 151**,** 364-371.

DEUTSCHBAUER, A. M., CHIVIAN, D. & ARKIN, A. P. 2006. Genomics for environmental microbiology. *Current Opinion in Biotechnology,* 17**,** 229-235.

EIXARCH, H. & CONSTANTI, M. 2010. Biodegradation of MTBE by Achromobacter xylosoxidans MCM1/1 induces synthesis of proteins that may be related to cell survival. *Process Biochemistry,* 45**,** 794-798.

FERRER, M., GOLYSHINA, O., BELOQUI, A. & GOLYSHIN, P. N. 2007. Mining enzymes from extreme environments. *Current Opinion in Microbiology,* 10**,** 207-214.

FERRER, M., MARTINEZ-ABARCA, F. & GOLYSHIN, P. N. 2005. Mining genomes and 'metagenomes' for novel catalysts. *Current Opinion in Biotechnology,* 16**,** 588-593.

GABOR, E. M., DE VRIES, E. J. & JANSSEN, D. B. 2003. Efficient recovery of environmental DNA for expression cloning by indirect extraction methods. *Fems Microbiology Ecology,* 44**,** 153-163.

GALVAO, T. C., MOHN, W. W. & DE LORENZO, V. 2005. Exploring the microbial biodegradation and biotransformation gene pool. *Trends in Biotechnology,* 23**,** 497-506.

GANESH-KUMAR, S., SOLOMON, R. D. J., KALIMUTHU, K. & VIMALAN, J. 2010. EVIDENCE OF AEROBIC DEGRADATION OF POLYCHLORINATED BIPHENYL AS GROWTH SUBSTRATE BY

NEW BACTERIA STENOTROPHOMONAS SP JSG1 AND CUPRIAVIDUS TAIWANENSIS JSG2. *Carpathian Journal of Earth and Environmental Sciences,* 5**,** 169-176.

HANDELSMAN, J. 2005. Metagenomics or megagenomics? *Nature Reviews Microbiology,* 3**,** 457-458.

HUSSAIN, S., SIDDIQUE, T., ARSHAD, M. & SALEEM, M. 2009. Bioremediation and Phytoremediation of Pesticides: Recent Advances. *Critical Reviews in Environmental Science and Technology,* 39**,** 843-907.

JENCOVA, V., STRNAD, H., CHODORA, Z., ULBRICH, P., HICKEY, W. J. & PACES, V. 2004. Chlorocatechol catabolic enzymes from Achromobacter xylosoxidans A8. *International Biodeterioration & Biodegradation,* 54**,** 175-181.

KIM, J. S. & CROWLEY, D. E. 2007. Microbial diversity in natural asphalts of the rancho la brea tar pits. *Applied and Environmental Microbiology,* 73**,** 4579-4591.

KUIPER, I., LAGENDIJK, E. L., BLOEMBERG, G. V. & LUGTENBERG, B. J. J. 2004. Rhizoremediation: A beneficial plant-microbe interaction. *Molecular Plant-Microbe Interactions,* 17**,** 6-15.

KUROKAWA, K., ITOH, T., KUWAHARA, T., OSHIMA, K., TOH, H., TOYODA, A., TAKAMI, H., MORITA, H., SHARMA, V. K., SRIVASTAVA, T. P., TAYLOR, T. D., NOGUCHI, H., MORI, H., OGURA, Y., EHRLICH, D. S., ITOH, K., TAKAGI, T., SAKAKI, Y., HAYASHI, T. & HATTORI, M. 2007. Comparative metagenomics revealed commonly enriched gene sets in human gut microbiomes. *DNA Research,* 14**,** 169-181.

LORENZ, P. & ECK, J. 2005. Metagenomics and industrial applications. *Nature Reviews Microbiology,* 3**,** 510-516.

MEDINA-MORENO, S. A., HUERTA-OCHOA, S. & GUTIERREZ-ROJAS, M. 2005. Hydrocarbon biodegradation in oxygen-limited sequential batch reactors by consortium from weathered, oil-contaminated soil. *Canadian Journal of Microbiology,* 51**,** 231-239.

NOJIRI, H., SHINTANI, M. & OMORI, T. 2004. Divergence of mobile genetic elements involved in the distribution of xenobiotic-catabolic capacity. *Applied Microbiology and Biotechnology,* 64**,** 154-174.

PAUL, D., PANDEY, G., PANDEY, J. & JAIN, R. K. 2005. Accessing microbial diversity for bioremediation and environmental restoration. *Trends in Biotechnology,* 23**,** 135-142.

PEREZ-PANTOJA, D., DONOSO, R. A., SANCHEZ, M. A. & GONZALEZ, B. 2009. Genuine genetic redundancy in maleylacetate-reductase-encoding genes involved in degradation of haloaromatic compounds by Cupriavidus necator JMP134. *Microbiology-Sgm,* 155**,** 3641-3651.

QUAISER, A., LOPEZ-GARCIA, P., ZIVANOVIC, Y., HENN, M. R., RODRIGUEZ-VALERA, F. & MOREIRA, D. 2008. Comparative analysis of genome fragments of Acidobacteria from deep Mediterranean plankton. *Environmental Microbiology,* 10**,** 2704-2717.

ROSENBERG, E. 1993. EXPLOITING MICROBIAL-GROWTH ON HYDROCARBONS - NEW MARKETS. *Trends in Biotechnology,* 11**,** 419-424.

SCHMID, A., DORDICK, J. S., HAUER, B., KIENER, A., WUBBOLTS, M. & WITHOLT, B. 2001. Industrial biocatalysis today and tomorrow. *Nature,* 409**,** 258-268.

SCHMIDT, T. M., DELONG, E. F. & PACE, N. R. 1991. ANALYSIS OF A MARINE PICOPLANKTON COMMUNITY BY 16S RIBOSOMAL-RNA GENE CLONING AND SEQUENCING. *Journal of Bacteriology,* 173**,** 4371-4378.

SNYDER, L. A. S., LOMAN, N., PALLEN, M. J. & PENN, C. W. 2009. Next-Generation Sequencing-the Promise and Perils of Charting the Great Microbial Unknown. *Microbial Ecology,* 57**,** 1-3.

STUART-KEIL, K. G., HOHNSTOCK, A. M., DREES, K. P., HERRICK, J. B., MADSEN, E. L. & LP 1998. Plasmids responsible for horizontal transfer of naphthalene catabolism genes between bacteria at a coal tar contaminated site are homologous to pDTG1 from Pseudomonas putida NCIB 9816-4. *Applied and Environmental Microbiology,* 64**,** 3633-3640.

TRINGE, S. G., VON MERING, C., KOBAYASHI, A., SALAMOV, A. A., CHEN, K., CHANG, H. W., PODAR, M., SHORT, J. M., MATHUR, E. J., DETTER, J. C., BORK, P., HUGENHOLTZ, P. & RUBIN, E. M. 2005. Comparative metagenomics of microbial communities. *Science,* 308**,** 554-557.

TYSON, G. W., CHAPMAN, J., HUGENHOLTZ, P., ALLEN, E. E., RAM, R. J., RICHARDSON, P. M., SOLOVYEV, V. V., RUBIN, E. M., ROKHSAR, D. S. & BANFIELD, J. F. 2004. Community structure and metabolism through reconstruction of microbial genomes from the environment. *Nature,* 428**,** 37-43.

VENTER, J. C., REMINGTON, K., HEIDELBERG, J. F., HALPERN, A. L., RUSCH, D., EISEN, J. A., WU, D. Y., PAULSEN, I., NELSON, K. E., NELSON, W., FOUTS, D. E., LEVY, S., KNAP, A. H., LOMAS, M. W., NEALSON, K., WHITE, O., PETERSON, J., HOFFMAN, J., PARSONS, R., BADEN-TILLSON, H., PFANNKOCH, C., ROGERS, Y. H. & SMITH, H. O. 2004. Environmental genome shotgun sequencing of the Sargasso Sea. *Science,* 304**,** 66-74.

CHAPTER FIVE: GENERAL CONCLUSIONS

This research on the Rancho La Brea Tar Pits further expanded on the studies initiated by Kim and Crowley, in which they discovered the existence of several hundred new species of bacteria and gene sequences encoding putative novel degradative enzymes. It is well known that petroleum hydrocarbons are substrates supporting microbial growth, so they are both a target and a product of microbial metabolism. There is much interest in exploiting petroleum-degrading organisms for environmental cleanup and the application of molecular biology methods for characterizing these microbes and their catabolic enzymes has become central to petroleum microbiology. Here I investigated the presence of petroleum degrading microorganisms and enzymes using a variety of methods, including denaturing gradient gel electrophoresis and enrichment cultures, microarray analysis, and characterization of a metagenomic clone library using targeted PCR primers, and next generation Illumina sequencing. In all of these studies, we found a diverse petroleum degrading community that provides new insight into the diversity and functional characteristics of this petroleum degrading community that has undergone selection to this environment for about 40,000 years.

In my first study, I identified microbial consortia that degrade a variety of polycyclic aromatic hydrocarbons using denaturing gradient gel electrophoresis. Results showed that when cultured samples of the asphalt soil on selected petroleum hydrocarbons, the enrichment culture selects for distinct consortia of bacteria that are able to degrade these compounds as a sole carbon source. While individual substrates each selected for their own community, I also identified an unclassified *Pseudomonas sp.* that appeared in communities across different substrates that may be able to degrade biphenyl, naphthalene, phenanthrene, chrysene, pyrene

and a combination of PAHs. In addition, I created clone libraries of both upper and lower pathway petroleum degrading genes, which are most similar to those from *Pseudomonas stutzeri*. These results, coupled with the finding from Kim and Crowley, in which they found a new species of *Pseudomonas stutzeri* imply that this organism may be novel and have a wide degradative selectivity. *Pseudomonas stutzeri* has been successfully cultured in the laboratory, so isolation of this organism would verify its degradative capacity and function within the Rancho La Brea Tar Pits.

In order to study the uncultured majority of the Rancho La Brea Tar Pits, I constructed a fosmid clone library from high molecular weight DNA extracted from Pit 101. This metagenomic library consisted of DNA inserts of about 38 kb and consisted of about 3,000 clones. The clone that was sequenced contained phylogenetic markers that identified this organism as an unknown member of the Rhododospirallacea family. In addition, a partial gene sequence encoding a novel 2-nitropropane dioxygenase was found on the same insert. This finding suggests that this novel Rhodospirallacea member degrades persistent nitroalkanes in natural asphalts. Meaning that this metagenomic study was able to link phylogeny and function of an as of yet unknown organism.

The metagenomic library was examined even further using next generation sequencing techniques. Illumina based sequencing was applied to the fosmid library in hopes of assembling whole DNA inserts from the fosmid library. What resulted was 75 Mb of DNA, which assembled into 661 contigs with an average length of 500 bp. This was much less than the expected result, however I was still able to find 3 types of dioxygenases, of which 2 are unknown and the other is most similar to a naphthalene dioxygenase from *Pseudomonas stutzeri*. This data, in addition to

the microarray analysis that detected a wide variety of petroleum degrading genes further illustrated the importance of the naphthalene dioxygenase enzyme in this system. Further sequencing is needed to assemble larger contigs and scaffolds, which may reveal the presence of more degradative genes along with composite genomes of the dominant organisms. In addition, probing the original fosmid library for the genes detected in the microarray analysis and the Illumina sequencing data may uncover whole catabolic pathways of genes found on the same DNA fragment.

There is much more work that can be done on the Rancho La Brea Tar Pits. This includes further shotgun sequencing of the petroleum degrading community, as well as further characterization of specific microorganisms contained in the tar that might have applications for biotechnology, such as microbial enhanced oil recovery and for the biosynthesis of small molecules. The Rancho La Brea Tar Pits represents a rich source of genetic diversity that has only begun to be explored. This research provided new insights into the types of microorganisms and enzymes that may be recovered from this site, and sets the stage for continued research. Of particular interest are genes encoded on mobile genetic elements as these are commonly used for the assembly and evolution of catabolic pathways, and for their transfer between different species. Illumina sequencing indicated the presence of DNA reads that appear to represent transposons and insertion elements. Follow-up studies could focus on PCR targeting of these mobile genetic elements in the fosmid library as a means to identify novel enzymes and catabolic pathways. In addition, expression of these novel enzymes and pathways may prove to be more efficient in petroleum degradation because these organisms have been evolving in the tar pits for millennia. Given the complexity of the community in the asphalts, metagenomic

approaches are useful for identifying potential new genes, but also yields what appear to be extraneous sequences for various housekeeping genes. Combining metagenomics with targeted PCR to access genes in the fosmid library, revealed by shotgun sequencing may be the most efficient approach for new enzyme discovery in this system.