MOLECULAR ANALYSIS OF PHENANTHRENE DEGRADATION BY *COMAMONAS TESTOSTERONI* GZ38 AND *ACIDOVORAX* SP. GZ39

By

KAREN PESCE

A dissertation submitted to the

Graduate School-New Brunswick

Rutgers, The State University of New Jersey

in partial fulfillment of the requirements

for the Degree of

Doctor of Philosophy

Graduate Program in Environmental Science

written under the direction of

Professor Gerben J. Zylstra

and approved by

_____

_____

_____

_____

_____

New Brunswick, New Jersey

October, 2008

ABSTRACT OF THE DISSERTATION

Molecular Analysis of Phenanthrene Degradation by *Comamonas testosteroni* GZ38 and *Acidovorax* sp. GZ39

by KAREN PESCE

Dissertation Director:

Professor Gerben J. Zylstra

Polycyclic aromatic hydrocarbons (PAHs) are toxic pollutants that are present in a variety of environments. Some bacterial species have the ability to metabolize these compounds and use them as energy sources for growth. *Comamonas testosteroni* GZ38 and *Acidovorax* sp. GZ39 were isolated from the Passaic River, NJ, based on their ability to grow on phenanthrene. These isolates are unusual in that they can grow on phenanthrene but not on naphthalene. The genes that code for phenanthrene degradation in GZ38 and GZ39 were identified. The DNA sequence of the genes is unique with respect to other sequences in the GenBank database with a single match to a putative phenanthrene degradation gene cluster from the phenanthrene degrader *Alcaligenes faecalis* AFK2 (GenBank accession number AB024945). BLAST matches to the putative genes in both operons suggest that phenanthrene degradation proceeds through phthalate. The putative phthalate degradation gene clusters from GZ38 and GZ39 have also been identified and sequenced. RT-PCR analysis of the phenanthrene degradation genes of GZ39 showed that the genes are expressed as one continuous transcript beginning with a

transposase that is upstream from the start of the phenanthrene operon. The phenanthrene degradation genes of GZ38 are also expressed as one continuous transcript but it does not include a transposase located upstream from the operon. For both GZ38 and GZ39, expression of the genes was observed with both phenanthrene and succinate as growth substrates. RT-qPCR results for GZ38 showed that the phenanthrene degradation genes were equally expressed on phenanthrene and succinate. A *lacZ* assay demonstrated that the promoter region upstream of the *phnAb* gene in GZ39 is constitutively active. A 5' RACE assay was performed in order to map the transcription start site of the phenanthrene operon in GZ39. Inactivation of the phthalate dioxygenase in GZ39 resulted in the accumulation of phthalate when grown on phenanthrene and provides strong evidence that phenanthrene degradation proceeds through phthalate. Although the phenanthrene degradation genes of GZ38, GZ39, and AFK2 are nearly identical, this is the first report of functional information relating to these genes.

ACKNOWLEDGMENTS

I would like to thank my advisor, Dr. Gerben J. Zylstra for changing the way I think about science. His excellent advice has helped to bring this project to completion. I would also like to thank the members of my thesis committee, Drs. Elisabetta Bini, Donna Fennell, Lee Kerkhof, and Peter Strom for their helpful advice.

I would also like to thank my colleagues from whom I have also learned a lot. In particular I would like to recognize Sinéad Ní Chadhain who was integral to all aspects of my research process as well as to the writing of this document. Sinéad, thank you for being patient and for teaching me how to be a scientist.

Hey Dad, now you're not the only PhD in the family.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Introduction

1. Polyaromatic compounds

Polyaromatic compounds (PAHs) are chemical compounds that consist of two or more fused carbon rings. Examples include naphthalene, anthracene, phenanthrene, chrysene, pyrene, and benzo(a)pyrene (Figure 1). PAHs are resistant to nucleophilic attack because there are dense clouds of Π electrons around their ring structures. In addition, they have low aqueous solubility which decreases with increasing molecular mass (Johnsen, Wick et al. 2005). These properties tend to increase the persistence and reduce the bioavailability of PAHs in the environment (Habe and Omori 2003).

The prevalence of PAHs in the environment is a result of natural biochemical processes as well as human activities (Kanaly and Harayama 2000). PAHs are by-products of combustion during forest fires, emitted in volcanic eruptions, and are present in crude oil reserves. Anthropogenic sources include the processing of fossil fuels, production of petroleum products, coal gasification, power plant emissions, incomplete combustion of automobile fuels, in charbroiled meat, and wood preservatives such as creosote (Bos 1984; Mumtaz 1995; Marr 1999; Samanta, Singh et al. 2002).

During the 20^{th} century there was a substantial increase in soil contamination by PAHs due to widespread industrial coal gasification. Much of this industry was located in close proximity to urban areas, which increased public exposure and health risk

(Johnsen, Wick et al. 2005; Liu, Niu et al. 2008). Another important mode of entry into the environment is petroleum spills. Notable recent cases include the *Exxon-Valdez* in Prince William Sound, Alaska and the *Prestige* along the coast of Spain.

PAHs can cycle through the environment in air, water, soil, and within food chains (Mumtaz 1995; Howsam, Jones et al. 2001). There are many ways that these compounds can enter the air including automobile exhaust and power plant emissions. These compounds can then travel through the air and be deposited in soil or water (Hylland 2006; Moon, Kannan et al. 2006; Bozlaker, Muezzinoglu et al. 2008). They can be released directly into the water at industrial sites and into soil at hazardous waste sites where there are leaky storage containers (State of New Jersey 2003). Soil contamination can also lead to groundwater contamination. Once in the environment, PAHs can undergo photodegradation (Plata, Sharpless et al. 2008), volatilization (Kriipsalu, Marques et al. 2008), or bind to particulates (Liang, Tse et al. 2007). However, abiotic degradation of PAHs is a minor mode of removal due to their relative stability (Grant, Muckian et al. 2007).

Toxic effects have been demonstrated in both aquatic life and land-dwelling organisms. Fish exposed to PAHs in sediment have developed DNA adducts and neoplastic lesions (Hylland 2006). Because they are highly stable lipophilic compounds, PAHs can bioaccumulate within living organisms (Clements 1994; Meador, Stein et al. 1995; Klaassen 2001). Several varieties of plants including vegetables have been shown to bioaccumulate PAHs from soil (Tao, Cui et al. 2004). Biomagnification

of PAHs through the food chain has been linked to reproductive impairment of many species of predatory birds worldwide (De Luca-Abbott, Wong et al. 2001). Human exposure to certain PAHs has been correlated with carcinogenicity particularly in people who work with these chemicals and also in those who smoke cigarettes (Popp, Vahrenholz et al. 1997; Shimada 2006). The EPA has classified 16 PAH compounds as priority pollutants due to their abundance and toxicity (Keith 1979). Since these compounds pose deleterious environmental and human health effects, it is imperative that they be removed from contaminated sites.

The most important manner of removal of PAHs from the environment is biodegradation by microorganisms because they possess a wide variety of enzymes that can transform PAHs into less toxic compounds (Mumtaz 1995; Peng, Xiong et al. 2008). There are numerous bacterial species throughout the world that have the ability to biodegrade various PAHs. Naphthalene-degrading organisms described to date include *Alcaligenes denitrificans, Mycobacterium* spp., *Pseudomonas putida, Pseudomonas fluorescens, Pseudomonas paucimobilis, Pseudomonas vesicularis, Burkholderia cepacia, Comamonas testosteroni, Rhodococcus* spp., *Corynebacterium venale, Bacillus cereus, Moraxella* spp., *Streptomyces* spp., *Vibrio* spp. and *Cyclotrophicus* spp. (Mumtaz 1995). Phenanthrene-degrading organisms include *Aeromonas spp., Alcaligenes faecalis, A. denitrificans, Arthrobacter polychromogenes, Beijerinckia* spp., *Micrococcus* spp., *Mycobacterium* spp., *Pseudomonas putida, Pseudomonas. paucimobilis, Rhodococcus* spp., *Vibrio* spp., *Nocardia* spp., *Flavobacterium* spp., *Streptomyces* spp. and *Bacillus*

spp. (Samanta, Singh et al. 2002). In addition, there are several bacterial species including *Spingomonads* that can degrade multiple PAHs.

Since PAH-degrading bacteria are often prevalent at contaminated sites, studying these bacteria as well as the genes that are responsible for PAH degradation is necessary to aid in the development of effective bioremediation strategies. Many studies are focusing on analyzing the bacterial communities from PAH-contaminated sites. Monitoring changes in the bacterial community by mapping gene profiles is essential to the development of effective bioremediation strategies (Ní Chadhain, Norman et al. 2006; Grant, Muckian et al. 2007; Muckian, Grant et al. 2007). Bioremediation has been an effective mode of cleanup of the *Exxon Valdez* oil spill in Prince William Sound, Alaska (Lindstrom, Prince et al. 1991) and the Prestige oil spill off the coast of Spain (Jimenez, Vinas et al. 2007), as well as of reduction in the PAH content of soil from a former creosote plant (Guerin 1999). The study of bacterial species that possess PAH degradation genes is vital for the development of optimal bioremediation strategies for PAH contaminated environments.

2. Overview of biodegradative pathways for polycyclic aromatic hydrocarbons

There are many organisms that can metabolize PAHs. Metabolic pathways have been characterized in mammals, fungi, and in both anaerobic and aerobic bacteria involving a variety of enzymes.

In mammals, the general mode of metabolism of PAHs is by way of the cytochrome P450 1A1, 1A2, and 1B1 enzymes (Figure 2). The action of these enzymes converts PAHs to more water soluble molecules which are more easily excreted. However, along the metabolic pathway some reactions produce reactive intermediates which can be cytotoxic and carcinogenic (Klaassen 2001).

There are a number of fungal species that can degrade PAHs (Peng, Xiong et al. 2008). Fungi possess several enzymes that can attack these compounds. It is believed that some fungi can use various peroxidase enzymes such as lignin peroxidase and manganese peroxidase to detoxify PAHs (Harayama 1997). Specifically, fungi that degrade wood such as *Phanerochaete* sp. attack PAHs using a peroxidase enzyme in conjunction with self-produced hydrogen peroxide (Suthersan 1999). Some fungi such as *Pleurotus ostreatus* possess cytochrome P450 monoxygenases that are involved in PAH degradation (Bezalel, Hadar et al. 1996). *Phanerochaete chrysosporium*, a model white rot fungus, has three P450 genes that are inducible by both naphthalene and phenanthrene (Doddapaneni, Subramanian et al. 2005). However, PAH degradation by fungi occurs at a slow rate, which reduces its practical application for bioremediation (Suthersan 1999).

PAHs can accumulate in sediment or soil under anoxic conditions. There are anaerobic bacteria that can degrade these compounds. Figure 3 describes a general model for naphthalene catabolism by anaerobic bacteria (Annweiler, Michaelis et al. 2002). Pure cultures of sulfate-reducing, iron-reducing, and nitrate-reducing bacteria that are capable

of degrading PAHs have been characterized (Coates, Woodward et al. 1997). In addition, PAHs can be degraded under manganese-reducing and methanogenic conditions (Chang, Shiung et al. 2002).

The first step of aerobic bacterial degradation of PAHs is catalyzed by a dioxygenase enzyme. This enzyme contains four subunits; a ferredoxin (Haigler and Gibson 1990), ferredoxin reductase (Haigler and Gibson 1990), large subunit, and small subunit (Ensley and Gibson 1983). PAHs are oxidized to *cis*-dihydrodiols via the incorporation of both atoms of an oxygen molecule (Gibson and Parales 2000; Habe and Omori 2003). In general, bacteria produce many similar intermediates when comparing the respective pathways for naphthalene and phenanthrene degradation.

2.1 Aerobic bacterial naphthalene degradation

The degradation of naphthalene first involves the addition of an oxygen molecule at the 1,2 position by naphthalene dioxygenase to produce 1,2-dihydroxynaphthalene (Jeffrey, Yeh et al. 1975) Figure 4). Then *cis*-naphthalene dihydrodiol dehydrogenase converts this compound to 1,2-dihydroxynaphthalene (Patel and Gibson 1974) which is then cleaved by 1,2-dihydroxynaphthalene dioxygenase (Eaton and Chapman 1992). This product spontaneously forms 2-hydroxy-2H-chromene-2-carboxylic acid which is acted upon by 2-hydroxy-2H-chromene-2-carboxylate isomerase to form *trans-o*-hydroxybenzylidene-pyruvic acid. Next, *trans-o*-hydroxybenzylidenepyruvate hydratase-

aldolase converts this compound to salicylaldehyde. Salicylaldehyde dehydrogenase transforms salicylaldehyde to salicylate (Habe and Omori 2003). This part of the pathway is termed the upper pathway. Salicylate can then be converted to either catechol or gentisate and then to TCA cycle intermediates (Yen and Serdar 1988). This is considered the lower pathway (Goyal and Zylstra 1997).

2.2 Aerobic bacterial phenanthrene degradation

The initial metabolism of phenanthrene follows a similar pathway as naphthalene. The initial oxygenation is at the 3,4 position with the same metabolic scheme until phenanthrene is degraded to 1-hydroxy-2-naphthoate (Figure 5). There are two lower pathways that can follow. In the first pathway, 1-hydroxy-2-naphthoate can be hydroxylated by 1-hydroxy-2-naphthoate hydroxylase to form 1,2-dihydroxynaphthalene. This compound can then enter the lower naphthalene degradation pathway where it can be metabolized to catechol (Davies and Evans 1964). In the second pathway, which is often termed the phthalate pathway, 1-hydroxy-2-naphthoate is attacked by 1-hydroxy-2-napthoate dioxygenase resulting in the formation of -2-carboxybenzalpyruvic acid (Barnsley 1983). This compound is converted to 2-carboxy-benzaldehyde by 2-carboxy-benzalpyruvate hydratase-aldolase which is then acted upon by 2-carboxybenzaldehyde dehydrogenase resulting in the formation of *o*-phthalic acid. Further metabolism goes through protocatechuate to TCA cycle intermediates (Kiyohara 1976; Kiyohara, Nagao et al. 1982; Goyal and Zylstra 1997).

3. PAH catabolic genes of Gram negative bacteria

3.1 *nah* genes

The degradation of naphthalene by bacteria has been extensively examined with research extending as far back as 1928 when Gray and Thornton documented bacterial growth on naphthalene. In the 1950's and 1960's metabolites in the degradation of naphthalene by *Pseudomonas* species were identified (Davies and Evans 1964). Within the last 25 years, much has been discovered about naphthalene degradation genes in several Pseudomonad species (Connors and Barnsley 1982).

The upper pathway is comprised of genes that catalyze the degradation of naphthalene to salicylate. This operon is comprised of the four component dioxygenase: *nahAa* (ferredoxin), *nahAb* (ferredoxin reductase), *nahAc* (oxygenase alpha subunit), *nahAd* (oxygenase beta subunit), as well as *nahB* (cis-1,2-dihydroxy-1,2-dihydronaphthalene dehydrogenase): *nahF* (salicylaldehyde dehydrogenase), *nahC* (1,2-dihydroxynaphthalene dioxygenase), *nahQ* (function unknown*), nahD* (2-hydroxychromene-2-carboxylate isomerase), and *nahE* (*trans-o*-hydroxybenzylidenepyruvate hydratase-aldolase) (Schell 1983; Eaton and Chapman 1992).

Naphthalene degradation genes have been identified in several *Pseudomonas* species. In some species, these genes are located on large catabolic plasmids.

Pseudomonas putida G7 has these genes on the NAH7 plasmid (Simon, Osslund et al. 1993), *Pseudomonas putida* 9816 on the pWW60 plasmid, and *Pseudomonas putida* NCBI 9816-4 on the pDTG1 plasmid (Simon, Osslund et al. 1993). When compared, the genes of the upper pathway on these plasmids bear a high nucleotide sequence identity. The genes on the NAH7 plasmid and the pDTG1 plasmid are in the same order and have a close sequence similarity (93%) when compared to the corresponding genes in *Pseudomonas putida* OUS82. There are also significant similarities between these genes and those of *Pseudomonas* strain c18 and of the plasmid pDTG1 (Eaton 1994; Habe and Omori 2003).

Pseudomonas aeruginosa PaK1 also possesses naphthalene degradation genes which bear 83-93% nucleotide sequence similarity to the corresponding genes in *Pseudomonas putida* OUS82 and are almost identical to genes in *Pseudomonas* strain c18 (Takizawa, Iida et al. 1999).

The organization of the genes of the upper catabolic pathway for *Pseudomonas putida* 9816, *Pseudomonas putida* 9816-4, *Pseudomonas putida* G7, *Pseudomonas* sp. c18, *Pseudomonas aeruginosa* PaK1, *Pseudomonas putida* OUS82, *Pseudomonas stutzeri* AN10, and *Pseudomonas putida* BS202 is *nahAaAbAcAdBFCQED* for all except *Pseudomonas stutzeri* AN10 which has a deletion of *nahQ* (Figure 6). The sequence identity among all of these species is approximately 90%. These genes are usually referred to as “classical *nah*-like genes” (Habe and Omori 2003).

Within the last fifteen years, PAH degradation genes have been documented in a number of different Gram negative and Gram positive naphthalene and phenanthrene degraders and more has been discovered about the diversity and distribution of these genes.

3.2 *nag* genes-

While the *nah* genes have been found in many organisms, many naphthalene degrading microbes have been isolated which contain different genes coding for naphthalene degradation. One such example is the naphthalene catabolic genes in *Ralstonia* sp. strain U2 (Zhou, Fuenmayor et al. 2001). These genes are referred to as *nag* genes as they encode enzymes that convert naphthalene to gentisate. They are similar in order to the *nah* genes of the upper pathway (*nagAaGHAbAcAdBFCQED*) except for the insertion of two genes, *nagGH*, which encode salicylate hydroxylase. These genes are similar to genes found in *Comamonas testosteroni* GZ42, a phenanthrene and naphthalene degrader. The genes in GZ42 are ordered *nahAaAbAcAdAc2Ad2BFCQED*. *Ac2* and *Ad2* of GZ42 are similar to *nagGH* of *Ralstonia* sp. strain U2.

3.3 *phn* genes of *Burkholderia* sp. RP007

Burkholderia sp. RP007 was isolated in New Zealand and is able to degrade both naphthalene and phenanthrene (Laurie and Lloyd-Jones 1999). Both compounds are

metabolized by way of the aforementioned upper pathway with naphthalene being degraded to salicylate and phenanthrene being degraded to 1-hydroxy-2-napthoic acid and then salicylate. However, the operon containing the phenanthrene degradation genes is not only ordered differently than the *nah* genes (*phnRSFECDAcAdB*) but is missing the ferredoxin and reductase components of the initial dioxygenase. When comparing the putative amino acid sequences of the *phn* genes of RP007 with the *nah* genes the identity ranges from 36-73%.

3.4 *phn* genes of *Alcaligenes faecalis* AFK2

In 1982, Kiyohara et al (Kiyohara, Nagao et al. 1982) isolated a phenanthrene-degrading bacterium, *Alcaligenes faecalis* AFK2, that could not grow on naphthalene. This isolate could degrade 1-hydroxy-2-naphthoate, 2-carboxybenzaldehyde, and phthalate, intermediates in the phthalate pathway, to protocatechuate. The phenanthrene degradation genes of *Alcaligenes faecalis* AFK2 have since been sequenced in their entirety and are unique with regard to both the gene organization and sequence similarity of genes when compared to other published sequences.

4. Phenanthrene degradation by Gram negative isolates

4.1 *Burkholderia* sp. TNFYE

A *Burkholderia* strain, TNFYE-5 was isolated from a crude-oil contaminated site in Tennessee for its ability to grow on phenanthrene (Kang, Hwang et al. 2003). Since this isolate was able to grow on 1-hydroxy-2-naphthoate and phthalate but not salicylate, it also appears to follow the phthalate pathway. This is further supported by the finding that when grown on naphthalene, TNFYE-5 accumulated salicylate.

4.2 *Sinorhizobium* sp. C4

Sinorhizobium sp. C4 was isolated from a PAH-contaminated site in Hawaii for its ability to grow on phenanthrene as a sole carbon source. Both 1-hydroxy-2-naphthoic acid and phthalate were identified as metabolites by GC-MS. This suggests that this organism may degrade phenanthrene through the phthalate pathway. One aspect of their study that was novel is that 1-hydroxy-2-naphthoic acid was metabolized to 2-carboxycinnamic acid instead of 2-carboxybenzalpyruvate or naphthalene-1,2-diol, which are metabolites produced by enzymes in the classical naphthalene pathway and in the phthalate pathway of phenanthrene metabolism, respectively (Keum, Seo et al. 2006).

4.3 *Burkholderia* sp. C3

An isolate from a PAH-contaminated site in Hawaii, *Burkholderia* sp. C3, was shown to have metabolites of phenanthrene degradation from two upper pathways and two lower pathways. Gas Chromatography- Mass Spectrometry (GC-MS) results showed metabolites that are consistent with both an initial dioxygenation at the 1,2-C position and at the 3,4-C position although the latter reaction occurs faster. Both pathways converge at 1,2-dihydroxynaphthalene and then branch again going to both salicylate and phthalate (Seo, Keum et al. 2007).

5. PAH catabolic genes of Gram positive bacteria

5.1 *phn* genes of *Nocardioides* sp. strain KP7

Nocardioides sp. strain KP7 was isolated from Kuwait following an oil spill and grows on phenanthrene but not on naphthalene. This isolate is able to degrade phenanthrene via 1-hydroxy-2-napthoic acid and salicylate. The amino acid sequence for each of the four subunits of the dioxygenase enzyme ranges from 40 to 60% homology to known sequences. In addition, the organization of the genes *phnE* (dihydrodiol dehydrogenase), *phnF* (extradiol oxygenase), *phnA* (alpha subunit of the oxygenase component), *phnB* (beta subunit of the oxygenase component), *phnG* (hydratase-

aldolase), *phnH* (aldehyde dehydrogenase), *phnC* (ferredoxin), *phnD* (ferredoxin reductase) is different than that of the *nah* genes. The enzyme *trans*-2'-carboxybenzalpyruvate hydratase-aldolase was purified from this isolate and characterized and has a deduced amino acid sequence similarity of 40% when compared to *Pseudomonas putida* G7 and *Pseudomonas* sp. strain c18 (Saito, Iwabuchi et al. 2000).

5.2 *nar* genes of *Rhodococcus sp*. strain NCIMB12038

A naphthalene dioxygenase has been characterized from *Rhodococcus sp*. strain NCIMB12038. An operon has been identified that contains *nar1* (regulator), *nar2* (regulator), *rub1* (rubredoxin), *rub2* (rubredoxin), *narAa* (oxygenase alpha subunit), *narAb* (oxygenase beta subunit), and *narB* (*cis*-naphthalene dihydrodiol dehydrogenase). The *narAa* and *narAb* genes are analogous to *nahAc* and *nahAd* genes but have only approximately 30 % amino acid sequence similarity to the *nahAcAd* genes of *Pseudomonas putida* NCIB 9816-4 (Larkin, Allen et al. 1999).

6. PAH degradation by Gram positive bacteria

6.1 *Bacillus* sp.

A *Bacillus* species isolated from soil utilized phenanthrene, naphthalene, 1-hydroxy-2-naphthoic acid, 2-carboxybenzaldehyde, and protocatechuate as sole carbon sources but not salicylate (Doddamani and Ninnekar 2000). Oxygen uptake studies showed that this isolate was able to oxidize 1-hydroxy-2-naphthoic acid, 2-carboxybenzaldehyde, *o*-phthalate, and protocatechuate but did not oxidize salicylate and catechol. In addition, cell-free extracts of *Bacillus* sp. grown on phenanthrene showed high activity of 1-hydroxy-2-naphthoate dioxygenase, 2-carboxybenzaldehyde dehydrogenase, and protocatechuate-4,5-dioxygenase. Phenanthrene appeared to be degraded via the phthalate pathway while naphthalene is degraded via the salicylate pathway. Specific genes responsible for these activities have not been identified.

6.2 *Mycobacterium* sp. BG1

Mycobacterium sp. BG1, isolated from the Great Bay Estuary in New Hampshire was shown to degrade phenanthrene via the phthalate pathway (Guerin and Jones 1988). High Pressure Liquid Chromatography (HPLC) demonstrated that the intermediates 1-hydroxy-2-napthoic acid, protocatechuate, phthalate, and o-carboxybenzaldehyde were detected during and then decreased in intensity during a sixteen day period.

7. Bacterial isolates that degrade multiple PAHs

7.1 *Mycobacterium vanbaalenii* PYR-1

Isolates have also been characterized that degrade multiple PAHs with different dioxygenases enzymes. One such isolate, *Mycobacterium vanbaalenii* PYR-1, degrades several PAHs including naphthalene, pyrene and phenanthrene (Stingley, Khan et al. 2004). Both pyrene and phenanthrene dioxygenase genes were identified in this organism. In fact, these pathways may intersect with a pyrene metabolite entering the phenanthrene degradation pathway. Identification of phenanthrene metabolites 2-carboxybenzalpyruvic acid, 2-carboxybenzaldehyde, and phthalic acid, supports the notion that this isolate metabolizes phenanthrene via the phthalate pathway.

7.2 Sphingomonads

There are a number of *Sphingomonas* species that are capable of degrading a variety of PAHs including naphthalene and phenanthrene. The catabolic genes of the Sphingomonads are arranged very differently when compared to the *nah* genes and can be distributed on different operons. *Novosphingobium aromaticivorans* F199 has 79 genes for catabolism or transport of aromatic compounds located on a 184 kb plasmid. These genes are scattered among 15 gene clusters. Some of these genes are similar in order and sequence similarity to corresponding genes in *Sphingobium yanoikuyae* B1.

Sphingobium sp. P2 has genes that are related to both F199 and B1. Biotransformation assays have showed that P2 degrades phenanthrene through salicylate (Pinyakong, Habe et al. 2003).

8. Overview of GZ38 and GZ39

The work presented in this dissertation will describe the identification and characterization of the phenanthrene degradation pathways in two phenanthrene degrading isolates, *Comamonas* sp. GZ38 and *Acidovorax* sp. GZ39, which were isolated from the Passaic River in NJ, a historically polluted area (State of New Jersey 2003). The Passaic River originates in Mendham, NJ and is over 80 miles in length. During the eighteenth and nineteenth centuries, it was a focal point of the industrial revolution and was vital to the urbanization of Newark and Paterson. The Passaic River was once connected via canals to the Delaware River and also served as a source of hydroelectric power for Paterson. In the nineteenth and twentieth centuries, numerous industries discharged a variety of toxic chemicals into the soil surrounding the river as well as directly into the water. Industries directly responsible for PAH contamination include Diamond Alkali Company (also dioxins), Pitt-Consol Chemical Corporation (cresol), Public Service Electric and Gas Company (PAHs), and Getty Petroleum Corporation (PCBs, lead, PAHs) (State of New Jersey 2003). In 1984, the EPA designated the lower Passaic River as a Superfund Site. However, it still to this day has high levels of many contaminants.

Isolates GZ38 and GZ39 are two of the numerous bacterial species that are present in the sediment of the Passaic River that can degrade the assortment of toxic contaminants. GZ38 has been identified as *Comamonas testosteroni* and GZ39 has been identified as an *Acidovorax* species based on their respective 16s rRNA sequences. Both isolates were found to have the ability to grow on phenanthrene but not on naphthalene. This observation led to the hypothesis that these isolates possess genes that are different from those of the classical *nah* degradation pathway. By cultivating and studying isolates with these capabilities, it is possible to better understand the mechanisms that are involved in the degradation of toxic compounds. Analysis of the genes responsible for phenanthrene degradation may provide new insight into this catabolic pathway as well as molecular tools to aid possible future bioremediation applications.

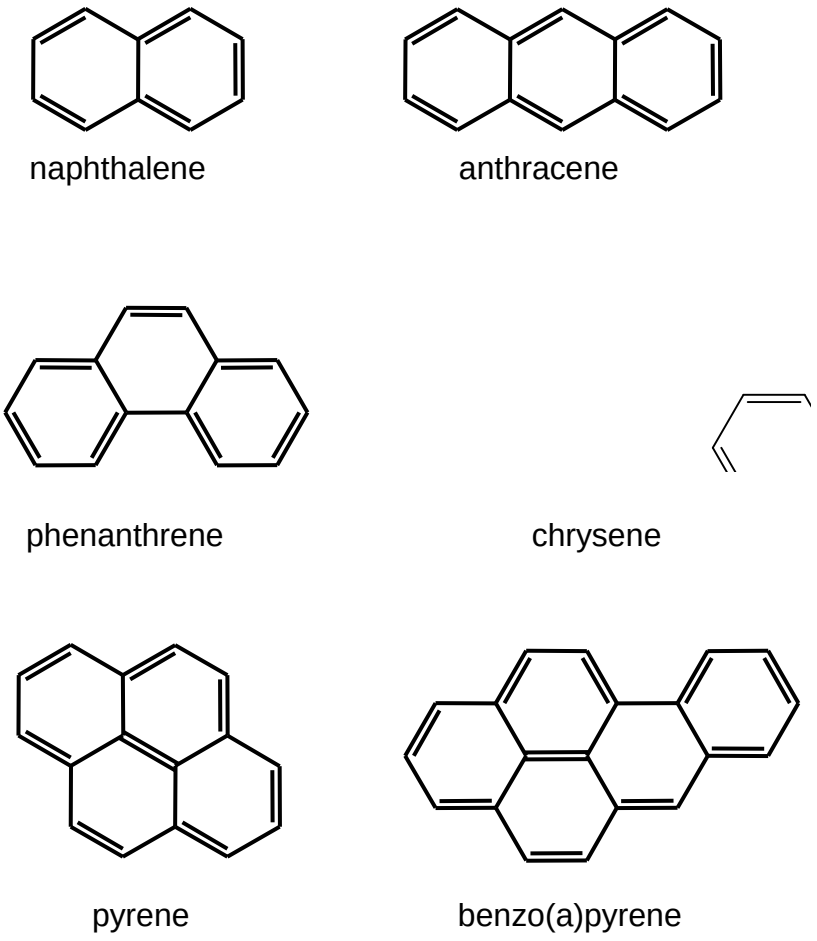


Figure 1. Structures of selected polycyclic aromatic hydrocarbons.

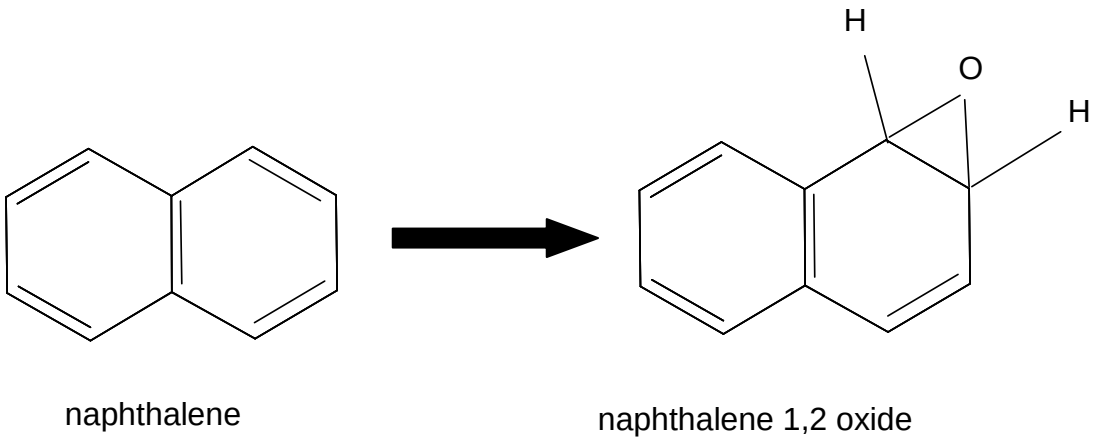


Figure 2. An example of P450 catalysis of PAHs. Naphthalene is converted to naphthalene 1,2-oxide, an epoxide. Epoxides can bind to proteins and nucleic acids which can result in cellular toxicity and genetic mutations.

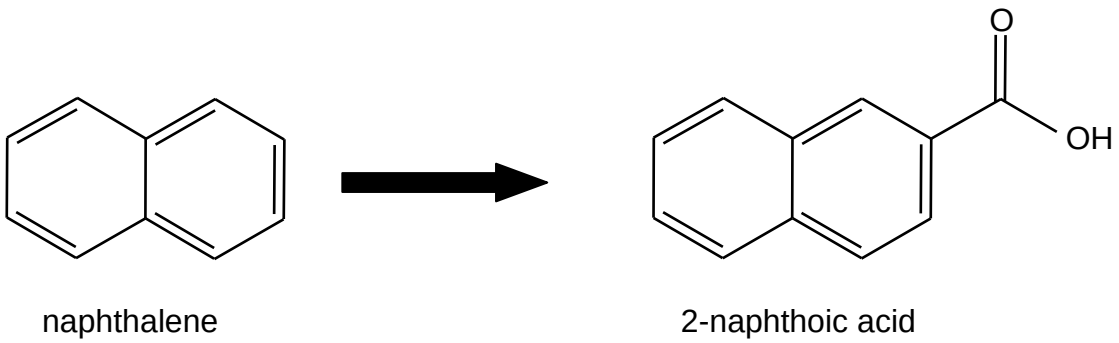


Figure 3. Anaerobic PAH degradation of naphthalene.

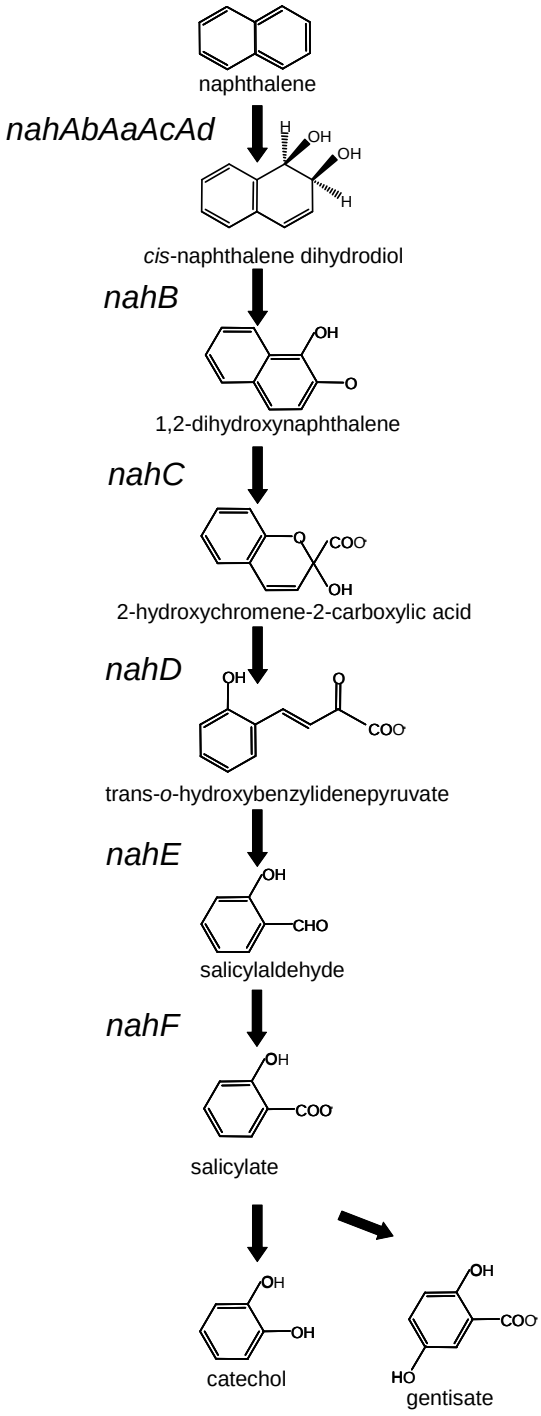


Figure 4. Aerobic bacterial naphthalene degradation pathway.

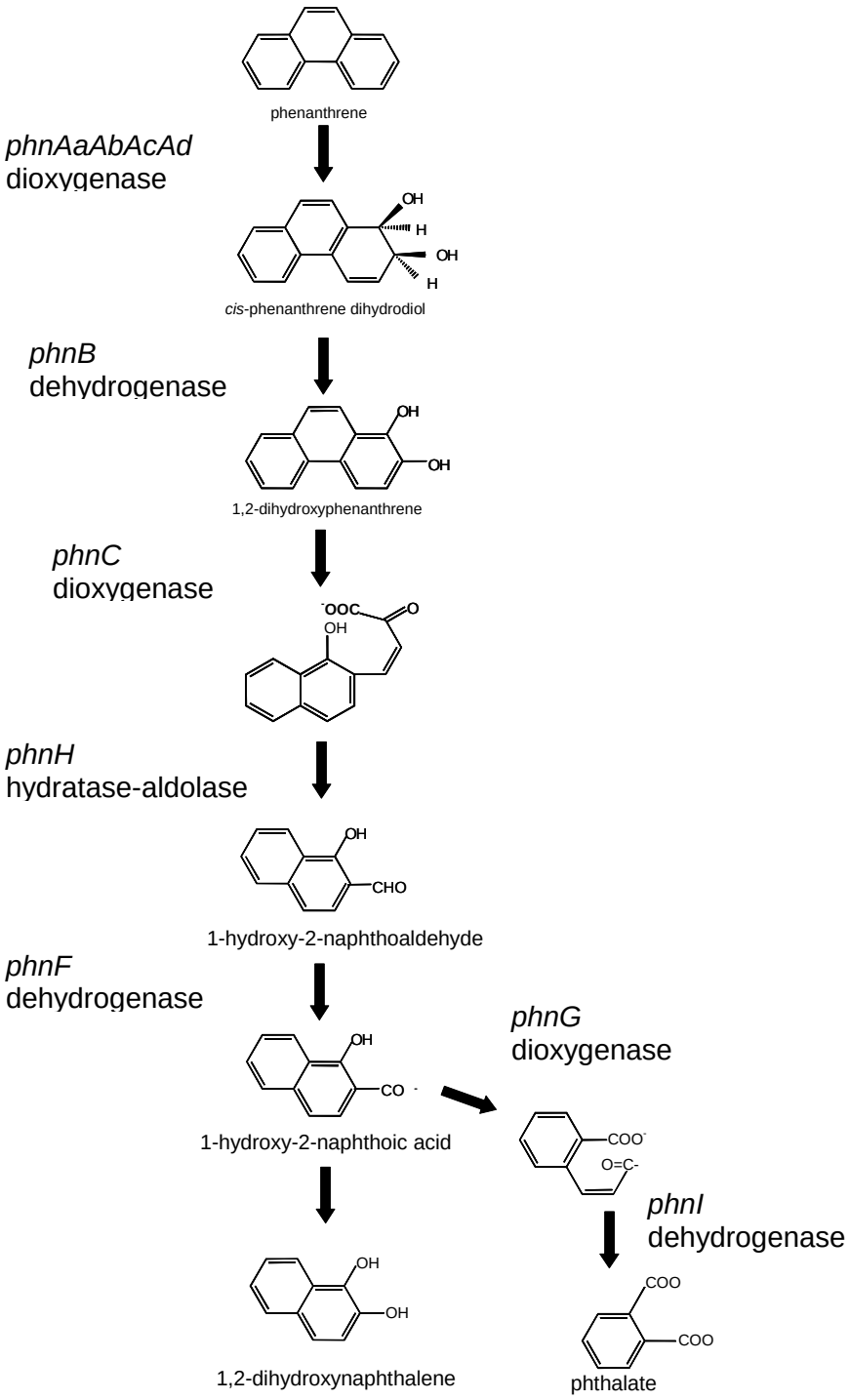


Figure 5. Aerobic bacterial phenanthrene degradation pathway.

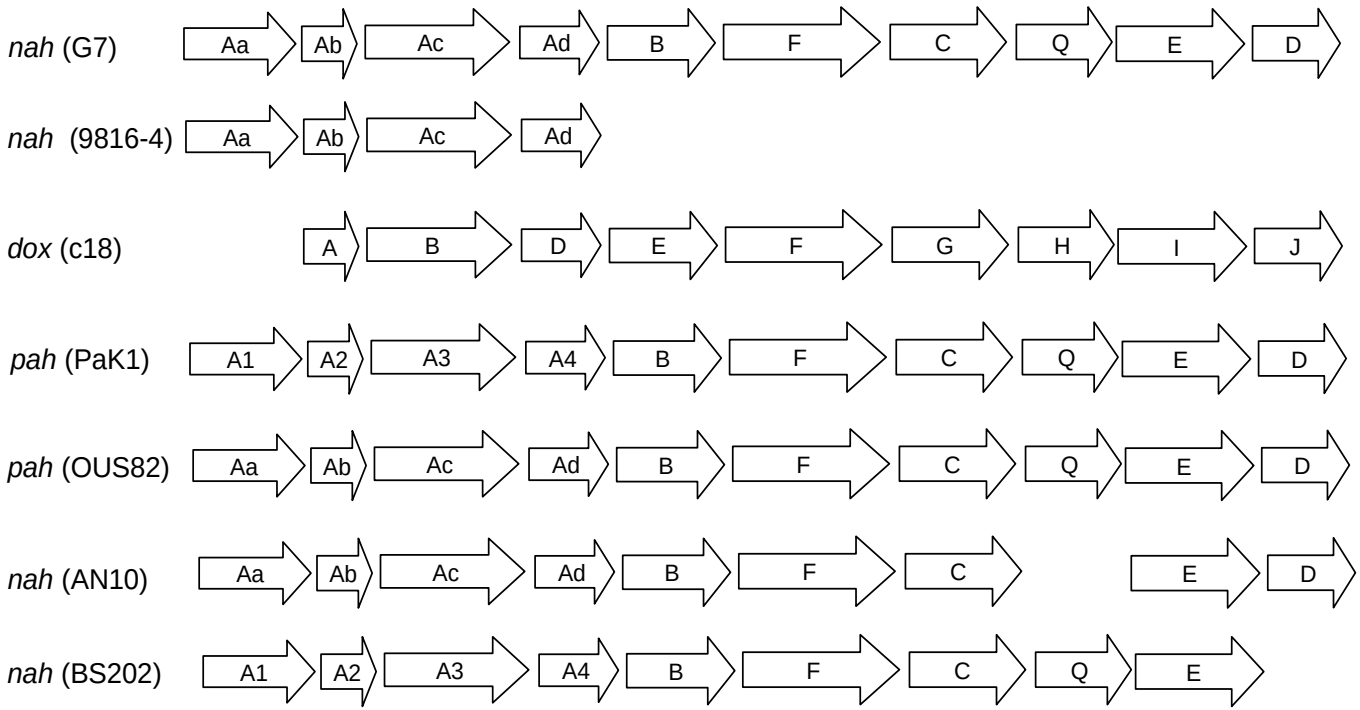


Figure 6. Gene organization of upper pathway *nah* genes of *Pseudomonas* isolates. These sequences are greater than 90% homologous to one another. The strain names and GenBank accession numbers are as follows: *Pseudomonas putida* G7 (U09057), *Pseudomonas putida* 9816-4 (U49496), *Pseudomonas* sp. c18 M60405), *Pseudomonas aeruginosa* PaK1 D84146), *Pseudomonas putida* OUS82 (AB004059), *Pseudomonas stutzeri* AN10(AF039533), and *Pseudomonas putida* BS202 (AF010471).

Materials and Methods

1. Growth of GZ38 and GZ39

GZ38 and GZ39 were isolated by enrichment on phenanthrene from the Passaic River, NJ near the Interstate 78 overpass and identified as *Comamonas testosteroni* and *Acidovorax* sp., respectively, by comparing their 16S ribosomal gene sequence identity with known sequences in the Genbank database. PCR and sequencing of 16S rRNA genes were done using 27f and 1522r primers. Mineral salts basal (MSB) medium (Stanier 1966) was used for carbon source studies. In order to determine growth on anthracene and phenanthrene, a 0.5% ethereal spray of each carbon source was sprayed onto the surface of MSB plates using a Crown Spra-Tool aerosol can (Aervoe Industries, Gardnerville, NV). GZ38 and GZ39 grew on MSB plates that were sprayed with a 0.5% phenanthrene ethereal spray within 24 hours. In both cases, the cells turn a yellow-orange color. GZ39 also grew on MSB plates that were sprayed with a 0.5% anthracene ethereal spray within 24 hours. GZ38 did not grow on anthracene-sprayed plates. Growth of GZ38 and GZ39 on naphthalene was measured by placing naphthalene crystals on the lid of MSB plates. Neither GZ38 nor GZ39 grew in the presence of naphthalene crystals. In addition, neither GZ38 nor GZ39 grow on MSB plates amended with phthalate (20 mM). GZ38 and GZ39 grew on MSB plates containing succinate (20 mM). Additionally, neither GZ38 nor GZ39 grow on MSB plates with salicylate (20 mM). Luria Bertani (LB) medium (Becton Dickinson and Company, Sparks, MD) was used to determine

antibiotic resistance. The antibiotics tested were ampicillin (50 µg/ml), kanamycin (50 µg/ml), and tetracycline (10 µg/ml). Each antibiotic was added to autoclaved LB medium at 50^{o}C and then allowed to solidify. Each isolate was streaked onto plates with antibiotic and monitored for seven days. Growth of both strains was observed on the ampicillin plates within 72 hours. GZ38 and GZ39 were unable to grow on kanamycin and tetracycline. GZ38 and GZ39 were grown at 30^{o}C.

2. Identification of phenanthrene degradation genes in GZ38 and GZ39

The sequence of the initial six genes in the phenanthrene degradation pathway (*phnAbAaAcAdBGST*) of GZ39 were located on a cosmid clone that was part of a cosmid library constructed with genomic DNA from GZ39 (Goyal 1996). The sequence for these genes in GZ38 was obtained by designing PCR primers based on the GZ39 sequence and performing PCR reactions on GZ38 genomic DNA. This resulted in the amplification of the corresponding genes from GZ38, which were then sequenced by primer walking. When these sequences were queried against both nucleotide and protein sequences in the Genbank database the best match was to genes from *Alcaligenes faecalis* AFK2, a phenanthrene degrading organism (Genbank nucleotide accession number AB024945). This sequence contains 13 genes of the phenanthrene "upper" pathway which code for enzymes that catalyze reactions beginning with phenanthrene and ending with 1-hydroxy-2-napthoate. Primers for the remaining seven genes from both GZ38 and GZ39 were designed by using the nucleotide sequence of the phenanthrene degradation genes from *Alcaligenes faecalis* AFK2. The gene sequence of the latter part of the *phn* operon in

GZ38 and GZ39 were obtained by PCR and sequenced directly. Primers were designed to amplify overlapping 1.5 kb-2 kb fragments across the operon. PCR reactions were prepared in 25 µl containing 2 ng/µl of DNA, 1X PCR buffer, 2.5 mM $MgCl_2$, 1 µM each forward and reverse primers, and 1U *Taq* polymerase (Sigma, St. Louis, MO) and amplified with the following program: an initial denaturation step of 94^{o}C for 5 min; 35 cycles of 94^{o}C for 1 min, 56^{o}C for 1 min, and 72^{o}C for 2 min; followed by a final extension of 72^{o}C for 10 minutes. Amplicons were visualized using agarose gel electrophoresis. DNA was purified using the Nucleotrap PCR purification kit (Clontech, Mountain View, CA).

3. Nucleotide sequence determination

PCR products and plasmids were sequenced using an ABI Prism 3100 Genetic Analyzer using Big Dye v3.1 chemistry (Applied Biosystems, Foster City, CA). Sequencing reactions were set up using 5 ng of template DNA, 0.2 µM primer, 2µl 5x buffer, 1 µl Big Dye and 5 µl water and amplified under the following conditions: an initial denaturation at 96^{o}C for 15 seconds; 30 cycles of 96^{o}C for 15 seconds, 50^{o}C for 10 seconds, and 60^{o}C for four minutes. Reactions were precipitated using ethanol and then sequenced. Sequence analysis was performed using the Lasergene DNASTAR software package. Database searches were done with the GenBank database using the BLAST algorithm (Altschul 1990).

4. Nucleotide sequence upstream of the phenanthrene degradation operon from GZ38

A Southern blot assay was performed to obtain nucleotide sequence upstream of the phenanthrene operon. A probe was created using primers for the large subunit of the dioxygenase gene (*phnAc*) with genomic DNA from GZ38 and the PCR DIG Probe Synthesis kit (Roche, Indianapolis, IN). The PCR reaction was prepared in 32 µl water, 5 µl PCR buffer with $MgCl_2$ 10X, 5 µl PCR DIG labeling mix, 3 µl primer 1, 3 µl primer 2, 1 µl Enzyme mix, 1 µl template (20 ng) and amplified with the following program: an initial denaturation step of 94^{o}C for 5 min; 30 cycles of 94^{o}C for 30 seconds, 54^{o}C for 1 minute, and 72^{o}C for 1 minute; and a final extension of 72^{o}C for 7 minutes. Southern hybridizations were performed using standard protocols. Genomic DNA digests were separated by agarose gel electrophoresis performed in 1X Tris-acetate EDTA buffer. Transfer of DNA from agarose gels to the nylon membrane (Bio-Rad Laboratories, Rockville Center, NY) was done using a vacuum blot apparatus (Pharmacia Vacu Gene Pump). Transferred DNA was crosslinked using a Spectrolinker XL-100 UV Crosslinker (Spectronics Corporation). The blot was incubated for 60 minutes at 65^{o}C in hybridization buffer prior to adding 10 µl of the DIG-labeled probe (denatured at 95^{o}C for 2 minutes). The hybridization was incubated overnight at 65^{o}C. The probe was detected with alkaline phosphate-conjugated anti-DIG antibody (Roche, Indianapolis, IN). A positive ApaI to SalI restriction fragment of approximately 4kb in size was gel purified (Qiagen) and cloned into pGEM5z vector (Promega, Madison, WI). Colonies were pooled and screened for the phenanthrene dioxygenase gene (*phnAa*) sequence by PCR. A single clone containing the sequence of interest could not be isolated so a positive pool of DNA containing ten clones was amplified by PCR using the vector

primer T7P and one of the primers for the *phnAc* gene sequence. A 2.5kb PCR product was gel purified, cloned into pGEM-T vector (Promega, Madison, WI), and sequenced.

5. PCR of phthalate degradation genes from GZ38

Primers for the genes in the phthalate degradation operon were designed based on sequence data for *Comamonas testosteroni* YZW-B, which was isolated from the Passaic River. Most of the gene sequence of this operon was obtained by standard PCR. Reactions were prepared in 25 µl containing 2 ng/µl of DNA, 1X PCR buffer, 2.5 mM $MgCl_2$, 1 µM each forward and reverse primers, and 1U *Taq* polymerase (Sigma, St. Louis, MO) and amplified with the following program: an initial denaturation step of 94^{o}C for 5 min; 35 cycles of 94^{o}C for 30 seconds, 55^{o}C for 1 minute, and 72^{o}C for 1 minute; and a final extension of 72^{o}C for 10 minutes and sequenced directly. PCR products were visualized using agarose gel electrophoresis and purified using the Nucleotrap PCR purification kit (Clontech, Mountain View, CA). Approximately 300 bases of the first gene in the pathway, *ophA1*, were obtained by using inverse PCR and sequenced directly. For inverse PCR, twelve single digests of 100 ng of genomic DNA were prepared with BamHI, BglII, EcoRV, HpaI, MluI, NcoI, SmaI, PstI, NotI, SstI, and SalI in 25 µl and incubated overnight at 37^{o}C. Ligation reactions were set up using 20 µl of the digests, 20 µl of 5X ligation buffer, 9 µl H_20 and 1 µl T4 DNA ligase and incubated at 16^{o}C overnight. The following day the reactions were precipitated by adding 5 µl of 3M sodium acetate and 500 µl of 95% ethanol, mixed, and incubated at 70^{o}C for 10 minutes. Samples were centrifuged for five minutes, washed with 1ml of

70% ethanol, and resuspended in 30 µl of TE buffer. PCR reactions were prepared in 30 µl containing 5 µl of DNA template, 1X PCR buffer, 2.5 mM $MgCl_2$, 1 µM each forward and reverse primers, and 1U *Taq* polymerase (Sigma, St. Louis, MO) and amplified with the following program: an initial denaturation step of 94°C for 5 min; 35 cycles of 94°C for 30 seconds, 64°C for 1 minute, and 72°C for 4 minutes; and a final extension of 72°C for 10 minutes. A single band of approximately 650bp in size was observed on a 1% agarose gel from the SalI digest and sequenced directly.

6. Confirmation of phthalate degradation genes in GZ39

Since GZ39 does not grow on phthalate a Southern blot hybridization assay was performed to determine whether GZ39 possessed phthalate genes bearing homology to the phthalate genes of GZ38. A probe was designed to the phthalate dioxygenase of GZ38 and the same procedure was followed that was previously mentioned. The probe hybridized to genomic DNA from GZ39.

7. Identification of phthalate degradation genes in GZ39

Ten 96 well plates from the GZ39 cosmid library were pooled by column and row and screened by PCR for a 650bp fragment of the phthalate dioxygenase gene using degenerate primers (Ní Chadhain 2007). A positive pool was selected and individual clones were screened for the same gene fragment. A single positive clone was identified

from plate 5, column 10, row E. This clone was subcloned by digesting 600 ng of the plasmid DNA with EcoRI and ligating into pZero-Backround–kan vector (Invitrogen, Carlsbad, CA). The ligation was transformed into *E. coli* TOP10 cells and 20 colonies were plasmid prepped and screened using PCR for the phthalate dioxgenase gene. A positive subclone containing a 9kb fragment was sequenced by primer walking out from the ends of the vector and the phthalate dioxygenase gene.

8. Reverse Transcription PCR

Cells were grown overnight at 30^{o}C with shaking at 200 rpm in 25 ml of LB medium with either phenanthrene added at 1 mg/ml dissolved in 100 µl of dimethylformamide or succinate [20 mM]. At 20 hours the OD_{600} was 0.20 and cells were isolated by centrifugation. RNA was isolated using the Qiagen RNeasy kit (Qiagen, Valencia, CA) and then treated with DNase (Qiagen, Valencia, CA) following the column protocol. Reverse transcriptase PCR was done using the Qiagen kit according to the manufacturer's protocols. Reactions were prepared in 25 µl containing 5 µl of 5x buffer, 1 µl dNTP mix, 1 uM each forward and reverse primers, 1 µl Enzyme mix, 10 ng template, and RNase-free water and amplified with the following program: a reverse transcription step of 50^{o}C for 30 min, an initial denaturation step of 95^{o}C for 15 min; 35 cycles of 94^{o}C for 1 min, 55^{o}C for 1 min, and 72^{o}C for 2 min; and a final extension of 72^{o}C for 10 minutes. Overlapping primers were designed to amplify the entire phenanthrene and phthalate operons from GZ38 and GZ39. Positive and negative controls were included as well. Negative controls consisted of performing PCR on the

RNA template without RT using 16S rRNA primers. Samples were separated on a 1% agarose gel, stained in (1X) SYBR green I in TAE buffer, and visualized using a Kodak EDAS 290 digital camera.

9. Reverse Transcription qPCR

Standard curves were constructed using PCR products of full-length genes from GZ38 as template DNA. Five data points were obtained for each primer set. DNA was serially diluted to prepare samples. Starting template concentrations were 1 x 10^5 copies/µl, 1 x 10^4 copies/µl, 1 x 10^3 copies/µl, 1 x 10^2copies/µl, and 1 x 10^1copies/µl. Copy number was calculated using the formula:

$$Copy\ number = \frac{DNA\ concentration\ (g/ul) \times A_b}{length\ of\ PCR\ product\ (bp) \times 660_a}$$

660^a = average molecular weight of one base pair

A^b= Avogadro's number = 6.02 x 10^{23} molecules/mole

A no template control reaction was also performed for each primer set. Primers were designed using ABI software to the 16S rRNA and 23S rRNA genes as housekeeping genes and to the phenanthrene dioxygenase large subunit (*phnAc*) and to the 1-hydroxy-2-naphthoate dioxygenase (*phnG*) of GZ38. Cells were grown in 25 ml of MSB with either succinate (10 mM) or phenanthrene dissolved in 100 µl of dimethylformamide (1 mg/ml) and grown at 30^0C with shaking overnight. Three flasks were set up for each carbon source and three replicates were done for each flask. RNA was isolated using the Qiagen RNeasy kit (Maryland). Reverse transcription reactions were set up using the Qiagen One Step RT-PCR kit. Reactions were prepared in 10 µl containing 1 µl 10x buffer, 2.2 µl 25 mM $MgCl_2$, 2 µl 2.5 mM dNTP, 0.5 µl random hexamers, 0.2 µl RNase inhibitor, 0.25 µl reverse transcriptase, 1 µl RNA template (1 ng/µl), and 2.85 µl RNase-free water. The reaction parameters were 25^oC for 10 minutes, 48^oC for 30 minutes, and 95^oC for 5 minutes. Q-PCR reactions were prepared in 25 µl containing 12.5 µl SYBR, 0.1µM each forward and reverse primers, 1 µl cDNA template, and 6.5 µl RNase-free water and amplified with an initial denaturation step of 95^oC for 15 minutes followed by 40 cycles of 95^oC for 15 seconds and 60^oC for 1 minute. A denaturation curve was done for each run. Data was analyzed using Lasergene software.

10. Phenanthrene Dioxygenase Expression

In order to measure the substrate range of the phenanthrene dioxygenase from GZ39, expression assays were performed using a clone that was previously constructed using the expression vector pTRC99A in *E. coli* DH5α cells (Goyal and Zylstra 1996).

The clone was constructed with a NcoI-PvuI-NsiI fragment that contained each of the four components of the dioxygenase. The clone was first grown overnight in LB liquid medium amended with ampicillin at 37^{o}C with shaking at 200 rpm. Two percent of the overnight culture was inoculated into fresh LB liquid media with ampicillin at 50 mg/ml and grown to an OD_{600} of 0.5. The culture was then induced by adding filter sterilized IPTG to 1 mM and grown to an OD_{600} of 1.0. The culture was centrifuged at 8000 rpm for 10 minutes and 50 ml of cells were washed in 25 ml of 50mM phosphate buffer (pH 7.25). Cells were centrifuged again at 8000 rpm for 10 minutes and resuspended in 25 ml of 50 mM phosphate buffer (pH 7.25). Substrate was added to a final concentration of 100 µM and cells were incubated overnight at 30^{o}C and shaken at 200 rpm. The substrates tested were anthracene, biphenyl, chrysene, dibenzothiophene, naphthalene, phenanthrene, and pyrene. Entire cultures were extracted twice with an equal volume of ethyl acetate, concentrated using a rotary evaporator, and resuspended in 1 ml of methanol except for chrysene, which was insoluble in methanol and therefore resuspended in 1 ml of ethyl acetate. Analysis using (GC-MS) was performed using a HP 5890 Series II gas chromatograph coupled to an HP 5971 mass selective detector. One microliter was injected using split mode with a ten minute solvent delay. The initial oven temperature was 60^{o}C rising 10^{o}C per minute to a final temperature of 300^{o}C. The samples were run on a HP-5MS column (HP part number 19091S-433, 30 x 0.25mm, 0.25 uM film thickness) with helium as the carrier gas.

11. Generation of mutants for promoter analysis of the phenanthrene degradation operon in GZ39

The nucleotide sequence upstream of both *phnAb* and a putative transposase in GZ39 was analyzed using BPROM software (Softberry.com) to identify possible promoter regions. The putative promoter regions upstream of the *phnAb* gene and of the transposase from GZ39 were amplified using primers with SalI and XbaI restriction sites at their 5' ends for directional cloning into the pKRZ1 vector (Rothmel, Shinabarger et al. 1991). The constructs were initially transformed into *E. coli* TOP10 cells for sequence verification. Both constructs were introduced into GZ39 by triparental mating. Each component was initially grown overnight on the appropriate media. The helper construct, pRK2013 in *E. coli* HB101 cells, was grown on an LB plate containing kanamycin at 50 µg/ml, GZ39 was grown on an LB plate, and the respective constructs either PAb:pKRZ in *E. coli* TOP10 cells or Ptnp:pKRZ in *E. coli* TOP10 cells or pKRZ in *E. coli* TOP10 cells were grown on LB plates containing kanamycin at 50 µg/ml. One half of a loopful of each of the three bacteria was mixed in 300 µl of LB liquid of which 5 µl was pipetted onto a nitrocellulose disc filter (Osmonics) and incubated on an LB plate overnight. The cells were then extracted from the filter, resuspended in 200 µl of MSB, and plated onto MSB plates containing kanamycin at 50 µg/ml. The plates were sprayed with a 0.5% phenanthrene ethereal spray and incubated at 30^{o}C for 36 hours.

12. Measurement of promoter activity

The pKRZ1 vector contains a promoterless *lacZ* gene. An active promoter will transcribe this gene which will be translated into β-galactosidase. *o*-nitrophenol- β-D-galactoside (ONPG) is converted by β-galactosidase to galactose and *o*-nitrophenol. When this reaction occurs, a yellow color is observed and *o*-nitrophenol can be quantified by measuring its absorption at 420 nm. The activity of all three pKRZ constructs in GZ39 was analyzed by measuring *lacZ* reporter activity. Cells were grown overnight in 25 ml of LB broth containing kanamycin at 50 ug/ml at 37^{o}C with shaking. Then 500 µl of each culture was subcultured into 25 ml of fresh LB containing kanamycin at 50 µg/ml and grown at 37^{o}C with shaking. The assay was performed when cells reached an OD_{600} of 0.4 to 0.5. Three replicates were assayed for each construct. Cells were cooled on ice for 30 minutes and then 100 µl of cell culture was aliquoted to microfuge tubes along with 900 µl of Z-buffer ($Na_2HPO_4 \cdot 7H_20$ [0.06M], $NaH_2PO_4 \cdot H_20$ [0.04M], KCl [0.01M], $MgSO_4 \cdot 7H_20$ [0.001M], and 15 µl of β-mercaptoethanol [0.05M]). To this, 10 µl of chloroform and 15 µl of 0.1% SDS were added, tubes were vortexed briefly, and incubated at 28^{o}C for five minutes. Next, 100 µl of ONPG (4 mg/ml) in 0.1M phosphate buffer (5.77 ml of 1M Na_2HPO_4 and 4.23 ml of 1M NaH_2PO_4 in 100 ml) was added and mixed. Tubes were incubated at 28^{o}C until a yellow color developed. The reactions were then stopped by adding 250 µl of 1M Na_2CO_3 and tubes were centrifuged for fifteen minutes at 12,000 x g. The OD_{420} was measured using 800 µl of the supernatant from each tube. Miller units were calculated using the following formula (Miller 1972):

$$Units = \frac{1000 \times OD_{420}}{t \times v \times OD_{600}}$$

t = time of reaction in minutes

v = volume of culture used in the assay in ml

13. Identification of the transcription start site of the phenanthrene degradation genes in GZ39

To map the transcription start site, RNA was isolated from the same GZ39 PAb:pKRZ in *E. coli* TOP10 cells construct that was used for the *LacZ* promoter assay. Cells were grown overnight in 10 ml of LB containing 50 mg/ml of kanamycin at 37^{o}C with shaking. One percent of this culture was subcultured into 10 ml of fresh LB with 50 mg/ml of kanamycin and grown at 37^{o}C with shaking to an OD_{600} of 0.5. RNA was isolated using the Qiagen RNeasy kit as previously described and stored at -70^{o}C until assayed. The assay was done using the 5' RACE system kit (Invitrogen, Carlsbad, Ca) according to the manufacturer's protocol. A primer was designed to convert mRNA into cDNA. The cDNA was tailed with dCTP and PCR amplified with a second nested primer and an abridged anchor primer. The PCR reaction was re-amplified and visualized on a gel. Samples were sequenced as already described and analyzed using Lasergene software.

14. Measurement of salicylate accumulation by GZ38 and GZ39

In order to determine if the inability of GZ38 and GZ39 to grow on naphthalene was caused by a missing or non-functional salicylate hydroxylase enzyme an assay was performed to measure the accumulation of salicylate in GZ38 and GZ39 when grown on naphthalene. GZ38 and GZ39 were inoculated in 50 ml of MSB medium containing 20 mM succinate and incubated overnight at 30^{o}C with shaking. Of the overnight culture, 2% was subcultured into 50 mL of fresh media containing 5 mM succinate and 0.5 mM naphthalene. Cells were incubated for 24 hours and removed by centrifugation at 10,000 x g for 30 minutes. The supernatant was extracted with an equal volume of ethyl acetate twice and concentrated using a rotary evaporator. The precipitate was resuspended in 1 ml of methanol and 50 µl of each sample was analyzed by HPLC. The separation was carried out on Beckman Coulter Ultrasphere C18 (250 × 4.6 mm) column (Fullerton, CA) at a flow rate of 1.0 ml min^{-1}. Elution solvent A was 1% acetic acid in water mixture. Elution solvent B was 1% acetic acid in methanol. Solvent A was filtered through a nylon 0.45 µm membrane (Millipore, Bedford, MA) and Solvent B was filtered through a 0.22 µm membrane (Millipore, Bedford, MA). Both solvents were degassed under vacuum before use. The initial gradient conditions were 100% A decreasing to 50% A and 0% B increasing to 50% B. The total analytical time was 60 minutes. A diode array detector was used to measure absorbance of the peak at 254 nm and 280 nm. Results were compared to salicylate standards of 0.10 mM and 0.25 mM in MSB as well as an MSB control.

15. Generation of a phthalate dioxygenase knockout mutant for GZ39

A subclone was previously generated from the cosmid clone 5E10 that contained the entire phthalate operon (*ophA1BRCA2*). There are two SacII restriction sites within the phthalate dioxygenase gene. A kanamycin resistance gene was PCR amplified, tagged with SacII sites at both ends, and cloned into the pGEM-T Easy vector, which is ampicillin resistant. Clones were screened for both ampicillin and kanamycin resistance as well as insert size by restriction enzyme digestion. A pGEM-T Easy clone containing the correct insert was then digested with SalI, gel purified, and ligated into the GZ39 subclone, which was also digested with SalI. Clones were screened by restriction enzyme digestion. A clone containing the kanamycin resistance gene was digested with SalI and a 4 kb fragment containing *ophCkmA2* was gel purified and ligated into the suicide vector pARO180 at the SalI site (Figure 7). Clones were screened for ampicillin and kanamycin resistance as well as by restriction enzyme digestion. This construct was introduced into GZ39 by triparental mating using the same methods previously mentioned. Each component was initially grown overnight on the appropriate media. The helper construct, pRK2013 in *E. coli* HB101 cells, was grown on an LB plate containing kanamycin at 50 µg/ml, GZ39 was grown on an LB plate, and the construct in *E. coli* TOP10 cells was grown on an LB plate containing kanamycin at 50 µg/ml. One half of a loopful of each of the three bacteria was mixed in 300 µl of LB medium of which 5 µl was pipetted onto a nitrocellulose disc filter (Osmonics) and incubated on an LB plate overnight. The cells were then extracted from the filter, resuspended in 200 µl of MSB, and plated onto MSB plates containing succinate [10 mM] and kanamycin at 50 µg/ml.

The plates were incubated at 30^{o}C for 36 hours. The colonies obtained were screened by PCR with primes to determine whether a single or double crossover event had occurred.

16. Measurement of phthalate accumulation by the GZ39 phthalate dioxygenase mutant

A phthalate dioxygenase knockout mutant was first screened for its ability to grow on phenanthrene by streaking it on to MSB plates with kanamycin (50 µg/ml) and sprayed with phenanthrene as previously described. The mutant was able to clear the phenanthrene from the plate but was unable to grow. An additional experiment was performed in order to measure the accumulation of phthalate by this mutant. The mutant was grown overnight in LB with kanamycin (50 µg/ml), subcultured in fresh LB with kanamycin (50 µg/ml), and grown for approximately 20 hours to an OD_{600} of 0.7 to 0.8. Cells were harvested, washed with 50 mM phosphate buffer pH 7.25, and resuspended in 25 ml of the same buffer which concentrated the OD_{600} to 2.0. Triplicate flasks of cells were prepared and incubated for 0, 12, 24, and 36 hours. Culture supernatants were extracted twice with an equal volume of ethyl acetate, concentrated using a rotary evaporator, and resuspended in 1.5 ml of methanol. An identical experiment was performed on the wild type with the exception of the growth media, which did not contain kanamycin. Samples were analyzed by high pressure liquid chromatography as previously mentioned and compared to phthalate standards.

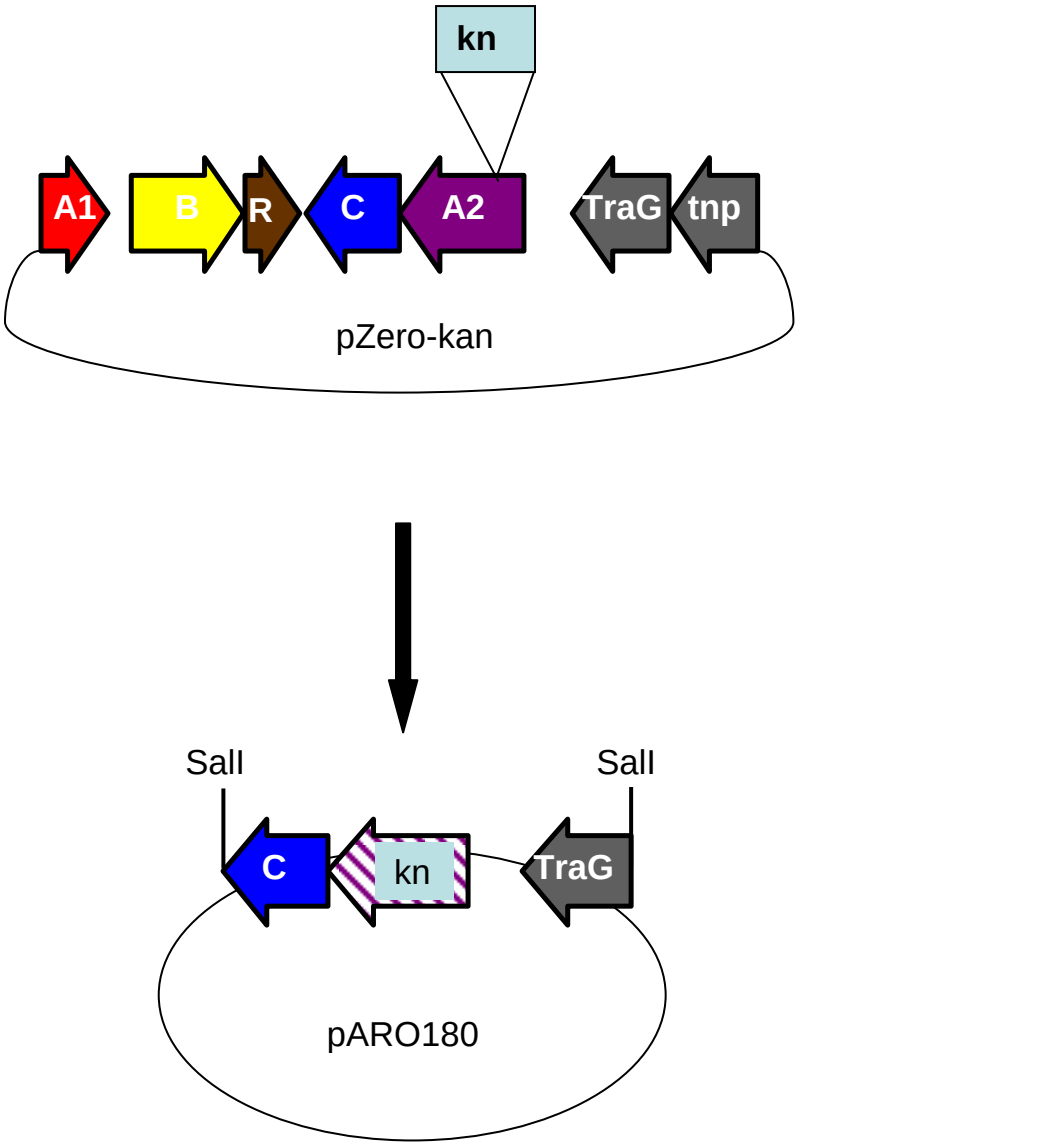


Figure 7. Construction of the phthalate dioxygenase mutant in GZ39.
A kanamycin resistance cassette was inserted into GZ39 cosmid 5E10 subclone 2.
A SalI digest was cloned into pARO180 and inserted into GZ39 by triparental mating.

Chapter 1

Identification and sequence analysis of phenanthrene degradation genes in GZ38 and GZ39

PAH contamination is prevalent in many different environments throughout the world. Several *Pseudomonas* isolates have been studied as model organisms for naphthalene and phenanthrene degradation. This has resulted in the discovery of the classical *nah* genes and elucidation of the corresponding metabolic pathway (Eaton 1994). With the wide diversity of bacteria in the environment, it seems likely that there would be multiple pathways for PAH degradation. It has been established that there are two pathways for phenanthrene degradation, the aforementioned classical pathway and the phthalate pathway (Kiyohara 1976; Kiyohara 1978). Bacterial isolates have been characterized that degrade phenanthrene via phthalate based on information from metabolite analysis. However, gene sequence data and functional data characterizing enzyme activity are lacking.

GZ38 and GZ39 were isolated based on their ability to grow on phenanthrene as the sole carbon source. Because these isolates could not grow on naphthalene and their DNA did not hybridize to the *nah* genes it was hypothesized that these strains possess novel phenanthrene degradation genes.

1. Nucleotide sequence of phenanthrene degradation genes from GZ38 and GZ39

A cosmid library was previously constructed (Goyal and Zylstra 1997) using GZ39 genomic DNA. Clones were screened for the formation of indigo from indole, which is indicative of dioxygenase activity (Ensley and Ratzkin 1983). Within this library, a clone was found to have the initial seven genes in the phenanthrene operon. When compared to sequences in the BLAST database, there was only one nucleotide sequence that was similar to GZ39. The match was to a putative phenanthrene dioxygenase gene cluster from *Alcaligenes faecalis* AFK2 (accession number AB024945) with an identity of 98%. The AFK2 sequence consists of 13 genes (*phnAb, Aa, B, Ac, Ad, D, GST, H, G, I, C, F, E*) and is 24.4 kb in length. Based on the high identity of GZ39 and AFK2, it was hypothesized that the genes from GZ38 shared similar identity. Therefore, primers were designed to the *phn* genes from AFK2 and GZ39 and used to amplify the initial *phn* genes from GZ38. Additional primers were designed to amplify the rest of the phenanthrene degradation genes from GZ38 and GZ39 using the *Alcaligenes faecalis* gene sequence due to the high degree of identity to the initial genes in the operon.

The length of the phenanthrene operon in GZ38 is 20.8 kb and the operon contains eighteen open reading frames (ORFs), all of which are transcribed in the same direction. The length of the *phn* operon in GZ39 is 21.5 kb and it contains nineteen open reading frames also transcribed in the same direction. In both GZ38 and GZ39, each ORF is initiated by an ATG start codon and preceded by a ribosome binding site. When GZ38 and GZ39 are compared to the *Alcaligenes faecalis* (AFK2) sequence, the gene

order is identical and the gene lengths of most identical (Figure 8). The functions of these ORFs were inferred based on sequence homology to previously described genes and their corresponding amino acid sequences (Tables 1 and 2). The best BLASTN and BLASTP matches for both nucleotide and amino acid sequences in the phenanthrene operon for both GZ38 and GZ39 are to AFK2.

The first gene in the operon is the *phnAb* gene which codes for the ferredoxin subunit of the phenanthrene dioxygenase enzyme. The open reading frame is 342 bases in length in GZ38, GZ39, and AFK2. The nucleotide sequence identity when comparing all three genes is 98% and the amino acid identity is 100%. The best characterized protein match for GZ38 and GZ39 is to a ferredoxin of *Burkholderia* sp. DBT1 (Genbank accession number AAK96190), which is involved in dibenzothiophene degradation. This sequence is 46% identical to the ferredoxin of GZ38 and 49% identical to the ferrodoxin of GZ39.

The second gene in the operon is *phnAa*, which codes for the ferredoxin reductase subunit of the phenanthrene dioxygenase enzyme. This open reading frame is 1029 bases in length in GZ38, GZ39, and AFK2 and both the nucleotide sequence identity and protein identity between the three isolates is 99%. The best characterized protein match for GZ38 and GZ39 is to a ferredoxin reductase from *Cycloclasticus* sp. P1 (Genbank accession number ABF56512) at 47% identity.

The third ORF is homologous to a dihydrodiol dehydrogenase, *phnB*. It is predicted to be 834 bases long in all three isolates and the genes share 98% nucleotide sequence identity. For GZ38, GZ39 and AFK2 the deduced amino acid sequences are

100% identical. The best characterized protein match for both isolates is to a *cis*-2,3-dihydroxy-2,3-dihydrobiphenyl dehydrogenase from *Pseudomonas* sp. Cam-1 with 57% identity (Genbank accession number AAK14786.1).

The large subunit of the dioxygenase follows the dihydrodiol dehydrogenase. The ORF is 1359 nucleotides long in GZ38, GZ39, and AFK2. The nucleotide sequences of GZ38 and GZ39 are 96% and 98% identical to AFK2, respectively. The protein sequences of GZ38 and GZ39 also match AFK2 at 96% and 98% respectively. The best characterized match to both GZ38 and GZ39 is to the dioxygenase large subunit of *Burkholderia* sp. DBT1 (Genbank accession number AAK62353.1) at 73% and 71% identity. There are four other matches to the GZ isolates that are putative dioxygenases. The sequences were derived from a dioxygenase clone library and are all fragments that comprise the middle of the protein sequence and are missing about 50 amino acids at the N-terminal and about 100 amino acids at the C-terminal of the GZ isolates. The amino acid sequence identities when comparing these *Burkholderia* sequences to the GZ isolates are approximately 95%.

The next gene is the small subunit of the dioxygenase (*phnAd*), which is 519 nucleotides long and 99% identical in all three bacteria. The amino acid sequence of GZ38 is 96% identical to AFK2 while the amino acid sequence of GZ39 is 100% identical to AFK2. The best characterized match is to the dioxygenase small subunit of *Burkholderia* sp. DBT1 (Genbank accession number AAK62354.1) at 67% identity.

PhnAd is followed by *phnD*, which codes for 2-hydroxychromene-2-carboxylate isomerase. The ORF is 591 nucleotides long in GZ38, GZ39, and AFK2 and GZ38 is 96% identical to AFK2 while GZ39 is 99% identical. In GZ38, the amino acid sequence is 100% identical to AFK2 while GZ39 is 96% identical to AFK2. The best characterized protein match is to an isomerase from *Burkholderia* sp. RP007 (Genbank accession number AAD09871) at 49% identity.

The next ORF codes for glutathione-S-transferase (gst). The sequence is 612 nucleotides long in AFK2 and 615 nucleotides long in GZ38 and GZ39 with 98% nucleotide sequence identity in all three. The amino acid sequence of GZ38 and GZ39 is 87% identical to AFK2. The best characterized protein match at 55% identity to glutathione-S-transferase from *Deftia acidovorans* SPH-1 (Genbank accession number ABX32691).

The *phnH* gene is downstream of *gst* and is 885 nucleotides long in AFK2 and 963 bases long in both GZ38 and GZ39 with a 98% nucleotide sequence similarity. The deduced amino acid sequence of GZ38 and GZ39 is 98% identical to AFK2. This *phnH* sequence from GZ38 and GZ39 matches a *trans*-2'-carboxybenzalpyruvate hydratase-aldolase from *Nocardioides* sp. KP7 (Genbank accession number BAA24966) with 45% identity. This is the best match to a characterized protein.

What follows is 1-hydroxy-2-naphthoate dioxygenase, *phnG*, which is 1050 bases long for all and the nucleotide sequence is 99% identical. When compared to AFK2, the predicted amino acid sequences of GZ38 and GZ39 are 99% identical. The next closest

match is at 45% to 1-hydroxy-2-naphthoate dioxygenase from *Bordetella avium* 197N (Genbank accession number CAJ48218.1).

The next gene, *phnI*, codes for 2-carboxybenzaldehyde dehydrogenase. It begins 179 bases before the end of *phnG* in all three bacteria but differs in length at 1644 bases in GZ38 and GZ39 and 1497 bases in AFK2. GZ38 and GZ39 match AFK2 with 99% sequence identity from bases 1 to 1497. The protein match is also 99% from amino acids 1 to 467 for both GZ38 and GZ39. The next closest match is to *Burkholderia multivorans* ATCC 17616 (Genbank accession number ZP_01568766) at 55%.

The sequence varies somewhat after this gene. In AFK2, there are 2219 nucleotides between *phnI* and the next gene in the pathway, *phnC*. In GZ38, the gap between *phnI* and *phnC* spans 881 nucleotides and in GZ39 there are 885 nucleotides in this same region. When the AFK2 sequence is queried against the Genbank database, a 969 bp ORF located 182bp downstream of *phnI* bears a 100% match to a putative transposase gene from *Acidovorax* sp. JS42 (Genbank accession number YP_984373). For GZ38 and GZ39 the nucleotide sequence in this region is 99% identical with 513 bases (GZ38) and 519 bases (GZ39) of this sequence comprising an open reading frame that has BLASTX matches to transcriptional regulators. The best match for GZ38 and GZ39 is to a putative MarR transcriptional regulator from *Sphingopyxis alaskensis* (Genbank accession number YP_616903). The 881 and 885 bases of sequence from GZ38 and GZ39, respectively, align with part of the 2219 bases of AFK2 sequence. The first 630 bases of the GZ isolates immediately after *phnI* match AFK2 beginning 1338 bases after *phnI* with 99% identity. The GZ isolates then have a series of nine Gs that do

not match AFK2. The GZ isolates again match AFK2 with 98% nucleotide sequence identity for the remainder of the sequence up to *phnC*.

The next gene in the sequence, *phnC* , codes for 3,4-dihydroxyphenanthrene dioxygenase and is 909 nucleotides long in all three isolates and the genes share 99% nucleotide identity. The protein sequence of GZ38 is 100% identical to AFK2 and the protein sequence of GZ39 is 99% identical to AFK2. The best match to a characterized protein is to a hydratase aldolase from *Sphingobium* sp. CHY-1 (Genbank accession number CAG17579) at 62%.

PhnF, which codes for 1-hydroxy-2-naphtoaldehyde dehydrogenase is the next gene. It is 1536 nucleotides long in all three isolates and 100% identical when comparing nucleotide sequence and protein sequence. The best match to a characterized protein is to an aldehyde dehydrogenase from *Sphinogbium yanoikuyae* B1 (Genbank accession number ABM79801) at 62%.

The region from *phnF* to *phnE*, the final gene in the operon, varies in length and sequence for each isolate. In AFK2, this sequence is 6670 bases in length and contains one open reading frame. It is 1560 bases downstream of *phnF* and is 1278 bases long. The best BLASTX match is to a hypothetical protein. In GZ38, this region is 6329 long and contains 3 ORFs. The first is located 102 bases downstream of *phnF* and is 1674 bases long. The only BLASTN match for this sequence is to AFK2. The best BLASTX match is to a hypothetical protein from *Pseudomonas putida* KT2440 (Genbank accession number NP_745344.1).

The second ORF immediately follows and is 1191 bases in length. The only BLASTN match is to AFK2. The only BLASTX match is to a protein of unknown function from *Marimonas* sp. MWYL1 (Genbank accession number YP_001341990.1).

The third open reading frame is 1009 bases downstream from the second and is 1608 bases long. The only BLASTN match is again to AFK2. The best BLASTX match is to a putative transporter from *Pseudomonas putida* KT2440 (Genbank accession number NP_745346.1).

In GZ39, the region from *phnF* to *phnE* is 6990 bases long and contains four open reading frames. The first open reading frame is located 109 bases down stream of *phnF* and is 1674 bases long. It matches the corresponding ORF in GZ38 with 100% nucleotide sequence identity. A BLASTN search with this sequence matches only AFK2 sequence.

The second ORF is 1362 bases long and matches AFK2 with 99% sequence identity. The best BLASTX match is to a protein of unknown function from *Marimonas* sp. MWYL1 (Genbank accession number YP_001341990.1).

The third ORF is 1089 bases long and has 99% nucleotide sequence identity to AFK2 from base 19478 to 20567. The best BLASTX match is to a glycosol hydrolase from *Polaramonas* sp. JS666 (Genbank accession number YP_546995.1).

The fourth open reading frame is 868 bases downstream from the third open reading frame and is 1539 bases in length. It matches AFK2 with 99% identity. The best BLASTX match is to a putative transporter from *Pseudomonas putida* KT2440 (Genbank accession number NP_745344.1).

The last gene in the operon, *phnE*, codes for a hydratase-aldolase and is 999 bases long in AFK2 and GZ39. These isolates share 100% identity in both DNA and protein sequence. In GZ38 this gene is 957 bases long. There are 15 amino acids missing in GZ38 that are present in GZ39 and AFK2 from position 123 to 137 in the protein sequence. The rest of the protein sequence is 100% identical. The best match to a characterized protein is to a hydratase-aldolase from *Sphingobium yanoikuyae* B1 (Genbank accession number ABM79813.1) at 64% amino acid identity for GZ38 and 67% amino acid identity for GZ39.

2. Nucleotide sequence upstream of the phenanthrene degradation operon of GZ38

Sequencing of the cosmid that contained the phenanthrene genes from GZ39 revealed a putative transposase located 1294 bases upstream from the first gene of the phenanthrene operon (*phnAb*) that is 1200 nucleotides in length. In AFK2, there are 1718 bases upstream from the start of *phnAb* which match GZ39 with 99% nucleotide sequence identity. Primers that were designed to this region did not amplify a corresponding sequence from GZ38 suggesting this region differs substantially in GZ38. Therefore, a southern blot experiment was performed in order to obtain the gene

sequence upstream of *phnAb* using genomic DNA from GZ38. An ApaI to SalI fragment containing phenanthrene dioxygenase sequence was gel purified and cloned into pGEM5z. When a single clone containing partial *phnAb* gene sequence could not be isolated, the positive pool of DNA was used as template for PCR reactions with the vector primers and primers for the *phnAb* gene. A 4kb band was gel purified and sequenced. An additional 2182 bases of DNA sequence upstream of *phnAb* in GZ38 was obtained as a result.

When compared to GZ39, 1050 nucleotides upstream of *phnAb* in GZ38 match with 99% sequence identity. The remaining 1132 bases of GZ38 do not match GZ39. Within this sequence there is a 582 base pair open reading frame that bears 79% identity with the amino acid sequence of an integrase from *Burkholderia cenocepacia* MC0-3 (accession number ZP_01563775.1).

3. Discussion-

PAH-degradation genes have been well characterized in numerous *Pseudomonas* isolates including *Pseudomonas putida* NAH7, *Pseudomonas* sp. C18 and *Pseudomonas* sp. OUS82 that share high sequence homology with the *nah* genes (Denome, Stanley et al. 1993; Kiyohara, Torigoe et al. 1994). In *Burkholderia* sp. RP007, genes have also been identified that encode proteins that degrade both naphthalene and phenanthrene (Laurie and Lloyd-Jones 1999). These genes are different from the *nah* genes.

GZ38 and GZ39 are unusual not only because they do not grow on naphthalene but also because the nucleotide sequence of their phenanthrene degradation genes stand apart from all previously reported sequence with the exception of AFK2. What makes these genes even more unusual is that according to their putative functions, they encode proteins that degrade phenanthrene through phthalate rather than through salicylate.

There are reports of other bacteria that can grow on phenanthrene along with evidence of degradation through the phthalate pathway. A *Bacillus* sp. is able to oxidize metabolites along the phenanthrene pathway (Doddamani and Ninnekar 2000). Metabolites in the phenanthrene degradation pathway were detected by HPLC from *Mycobacterium* sp. BG1 (Guerin and Jones 1988). Although the aforementioned studies have established that metabolites consistent with the degradation of phenanthrene through phthalate are produced by these isolates, there is no sequence data for any of the genes in the respective pathways.

In this study, genes for the phenanthrene pathway were identified and sequence alignments of the products of the phenanthrene degradation genes of GZ38 and GZ39, while highly similar (98-100% identity) to the corresponding genes in AFK2, share limited sequence homology with other sequences in the Genbank database. Comparison of each of the fourteen putative protein sequences in the phenanthrene degradation operon of GZ38 and GZ39 to sequences in Genbank revealed that they at most 70% homologous to analogous proteins from other PAH degradation operons.

Closer comparison of the dioxygenase protein, which tends to be conserved, further underscores its uniqueness (Figure 9). Phylogenetic analysis showed that the dioxygenase large subunits of both GZ38 and GZ39 are related to the dibenzothiophene dioxygenase large subunit from *Burkholderia* sp. DBT1 at 67% amino acid similarity. They also cluster with uncharacterized dioxygenase fragments from *Burkholderia* sp. Ch3-5 and Eh1-1 that are part of a PCR library of partial gene sequences. This cluster groups away from PAH dioxygenases from *Pseudomonas stutzeri*, *Pseudomonas putida* NAH7, and *Pseudomonas putida* pDTG1, which are model organisms that contain the *nah* genes. Further analysis of the predicted amino acid sequences for the enzymes encoded by the genes in GZ38 and GZ39 confirmed the presence of the Rieske-type [2Fe-2S] binding site motif, which is a highly conserved region (Gibson and Parales 2000).

Phylogenetic trees were constructed for the other components of the dioxygenase enzyme. When assembled into a phylogenetic tree (Figure 10), the dioxygenase small subunits of GZ38, GZ39, AFK2, and *Burkholderia* sp. DBT1 group together while all other sequences are highly dissimilar. The *nah* sequences of the Pseudomonads bear less than 25% sequence similarity to GZ38 and GZ39 while the more closely related *Burkholderia* sp. DBT1 is 70% identical. Proteins from the other isolates on the tree collectively share approximately 30% sequence identity with the GZ isolates.

Trees containing the protein sequences of the ferredoxin and ferredoxin reductase (Figures 11 and 12) further underscore the uniqueness of the dioxygenase from GZ38 and GZ39. The ferredoxin protein sequences of GZ38, GZ39 and AFK2 have no matches

above 46% with *Sphingopyxis macrogoltabida*, *Sphingomonas* sp. A4, and *Burkholderia* sp. DBT1 as the only isolates with greater than 40% homology. The ferredoxin of NAH7 is the most distantly related at 30%. A similar trend is observed for the ferredoxin reductase. When GZ38, GZ39, and AFK2 are compared to other sequences in the GenBank database, there is no match above 47%. The most closely related proteins are from *Sphingopyxis macrogoltabida* and *Sphingomonas* sp. A4 while the ferredoxin reductase from NAH7 is only 21% homologous.

For GZ38 and GZ39 as well as the *Alcaligenes faecalis* isolate all of the phenanthrene degradation genes are ordered the same way and transcribed in the same direction. The entire operon is 20.8 kb in length for GZ38 and 21.5 kb in length for GZ39. All three sequences share 98-99% nucleotide sequence identity when compared to one another. Additionally, when compared to other phenanthrene degradation genes, this sequence is highly dissimiliar.

The organization of the neighboring genes from the organisms with the most closely related protein matches to the phenanthrene dioxygenase large subunits of GZ38 and GZ39 were also compared (Figure 13). The most closely related match other than *Alcaligenes faecalis* is to *Burkholderia* sp. DBT1, a dibenzothiophene degrader. There is an isomerase immediately upstream from the large subunit and a small subunit followed by a dehydrogenase immediately downstream. The only match to a phenanthrene degrading isolate is to *Burkholderia* sp. RP007. The available sequence read from left to right contains two regulators, an ORF that is similar to salicylate dehydrogenase, a hydratase-aldolase, phenanthrene dioxygenase large subunit, phenanthrene dioxygenase

small subunit, and an isomerase. There is no ferredoxin or ferredoxin reductase in this sequence. Of these genes, only the dioxygenase large subunit and the isomerase bear any similarity to the respective ORFs of GZ38 and GZ39. However, RP007 degrades phenanthrene through salicylate not phthalate. The gene organization of the other operons containing matches to the large subunit of the dioxygenase also varies when compared to one another and are different from GZ38 and GZ39. There are no discernable patterns or similarities. GZ38 and GZ39 are unique with regard to both nucleotide sequence and gene organization.

When the putative protein sequence of the large subunit of the phenanthrene dioxygenase is queried against completed microbial genomes, the matches with the highest percent identity of around 50% are to a naphthalene/biphenyl dioxygenase from *Novosphingobium aromaticivorans* F199 plasmid pNL1 (accession number NP_049062), a naphthalene dioxygenase from *Pseudomonas* sp. ND6 plasmid ND6-1 (accession number NP_943188), and a naphthalene dioxygenase from *Pseudomonas putida* plasmid NAH7 (accession number YP_534822). Other than a dioxygenase beta subunit and a dehydrogenase which are immediately downstream from the large subunit, there are no other genes either upstream or downstream that are involved in PAH degradation on pNL1. The matches to the other two plasmids are to naphthalene dioxygenase large subunits that are part of the *nah* gene pathway.

Because the nucleotide sequences of the phenanthrene degradation genes of GZ38 and GZ39 over the entire operon are 98-100% homologous with respect to one another as well as to AFK2 it can be inferred that they most likely have a common origin. The

presence of two different transposons upstream of each operon in GZ38 and GZ39 is indicative that these genes were most likely exchanged between bacteria.

In polluted environments, bacteria that have the ability to metabolize xenobiotics have an advantage. The exchange of catabolic genes between bacteria is essential to their adaptation to these environments. This includes the transfer of entire operons, rearranging pre-existing genes, or merging genes or gene fragments from multiple microorganisms (Top and Springael 2003). There are three accepted mechanisms by which genes can be transferred; transformation, conjugation, and transduction. All three have been documented in various environmental microcosms (Davison 1999). However, the movement of catabolic genes is often associated with mobile genetic elements such as plasmids and transposons.

There are many bacterial plasmids in the environment that contain genes for xenobiotic degradation including aromatic hydrocarbons. Some examples are NAH7 (naphthalene), pDTG1 (naphthalene), pWWO (toluene), and pPS12-1(1,2,4,5-tetrachlorbenzene).

Transposons, some of which are located on plasmids, can also mediate gene transfer between bacteria. There are several xenobiotics for which their respective catabolic genes are associated with transposons. Specific examples include naphthalene (NAH7 and pDTG1), toluene (pWWO), benzene, and chlorobenzene (Tan 1999, Top 2003). In addition, close variants of the classic naphthalene degradation genes (*nah*) have also been found in the genomic DNA of isolates. Studies have also found that

closely related catabolic operons can be present in bacteria from different geographic locations (Top and Springael 2003).

The fact that the phenanthrene degradation genes in GZ38, GZ39, and AFK2 have 98-100% sequence identity and identical gene arrangement supports the hypothesis that they have a common origin. These highly conserved genes are associated with three different putative tranposons in each of the three isolates. GZ38 and GZ39 have two different transposons upstream of the phenanthrene operon. Based on the available sequence data for AFK2, it most likely has the same transposon upstream of the phenanthrene operon as GZ39. Nucleotides 1 through 424 of the AFK2 sequence match the last 424 nucleotides of the transposon from GZ39 with 100% identity. Additionally, AFK2 has a transposon that is located between *phnI* and *phnC*. The presence of these transposons provides insight as to how these genes may have traveled. GZ38 and GZ39 were isolated from the same environment while AFK2 is from a different geographic location.

Horizontal gene transfer of catabolic genes for phenanthrene may have led to the adaptation of GZ38 and GZ39 to the heavily polluted environment of the Passaic River, an environment rich in PAHs such as phenanthrene, anthracene, pyrene, chrysene, and fluoranthene, but in which naphthalene was not detected (Norman et al in preparation). The incorporation of these novel phenanthrene degradation genes by GZ38 and GZ39 may provide these isolates with a metabolic advantage. In PAH-contaminated environments with mixed populations of microorganisms these genes could play an

important role. The transfer of genes is integral to the development of new catabolic pathways and can accelerate the rate of bioremediation in polluted environments.

The phenanthrene operons of GZ38 and GZ39 are substantially different from other known PAH degradative operons in both arrangement and composition. Every putative protein is highly divergent when compared to proteins in the GenBank database. The discovery of these genes further expands our understanding of PAH degradation.

Table 1. Best BLASTP matches in the GenBank database to the GZ38 phenanthrene degradation sequence

Start	End	Accession Number	Identity (%)	Name	Organism
678	1019	BAA76320.1	100	Ferredoxin	*Alcaligenes faecalis*
		AAK96190	49	Ferredoxin	*Burkholderia* sp. DBT1
		BAE93940	46	Ferredoxin	*Sphingomonas sp.* A4
1043	2071	BAA76321	99	ferredoxin reductase	*Alcaligenes faecalis*
		ABF56512	47	ferredoxin reductase	*Cycloclasticus* sp. P1
		BAC81546	47	ferredoxin reductase	*Cycloclasticus* sp. A5
2170	3003	BAA76322.1	91	dihydrodiol dehydrogenase	*Alcaligenes faecalis*
		AAK14786.1	57	cis-2,3-dihydroxy-2,3-dihydrobiphenyl dehydrogenase	*Pseudomonas* sp. Cam-1
3060	4418	BAA76323.1	96	dioxygenase large subunit	*Alcaligenes faecalis*
		AAK62353.1	73	dioxygenase large subunit	*Burkholderia* sp. DBT1
4448	4966	BAA76324.1	96	dioxygenase small subunit	*Alcaligenes faecalis*
		AAK62354.1	67	dioxygenase small subunit	*Burkholderia* sp. DBT1
5125	5715	BAA76325.1	100	Isomerase	*Alcaligenes faecalis*
		AAD09871	49	Isomerase	*Burkholderia* sp. RP007
5781	6599	YP_001412814.1	27	transcriptional regulator	*Parvibaculum lavamentivorans*
6678	7292	BAA76326.1	87	glutathione-S-transferase	*Alcaligenes faecalis*
		ABX32691	55	glutathione-S-transferase	*Deftia acidovorans* SPH-1
7370	8332	BAA76327.1	98	trans-2'-carboxybenzalpyruvate hydratase-aldolase	*Alcaligenes faecalis*
		BAA24966	45	trans-2'-carboxybenzalpyruvate hydratase-aldolase	*Nocardioides* sp. KP7
8372	9421	BAA76328	99	1-hydroxy-2-naphthoate dioxygenase	*Alcaligenes faecalis*
		CAJ48218.1	45	1-hydroxy-2-naphthoate dioxygenase	*Bordetella avium* 197N
9243	10886	BAA76329.1	99	2-carboxybenzaldehyde dehydrogenase	*Alcaligenes faecalis*
		ZP_01568766	55	betaine-aldehyde dehydrogenase	*B. multivorans* ATCC 17616
11768	12676	BAA76330.1	100	3,4-dihydroxyphenanthrene dioxygenase	*Alcaligenes faecalis*
		CAG17579	63	extradiol dioxygenase	*Sphingomonas sp.* CHY-1
12752	14287	BAA76331.1	100	1-hydroxy-2-naphtoaldehyde dehydrogenase	*Alcaligenes faecalis*
		ABM79801	62	aldehyde dehydrogenase	*Sphingobium yanoikuyae* B1
20611	21567	BAA76332.1	95	hydratase-aldolase	*Alcaligenes faecalis*
		ABM79813.1	64	hydratase-aldolase	*Sphingobium yanoikuyae* B1

Table 2. Best BLASTP matches in the GenBank database to the GZ39 phenanthrene degradation sequence

Start	End	Accession #	Identity (%)	Name	Organism
1780	2979	YP_552075.1	74	Transposase	*Pseudomonas* sp. JS666
4274	4615	BAA76320.1	100	Ferredoxin	*Alcaligenes faecalis*
		AAK96190	49	Ferredoxin	*Burkholderia* sp. DBT1
		BAE93940	46	Ferredoxin	*Sphingomonas* sp. A4
4639	5667	BAA76321	99	ferredoxin reductase	*Alcaligenes faecalis*
		ABF56512	47	ferredoxin reductase	*Cycloclasticus* sp. P1
		BAC81546	47	ferredoxin reductase	*Cycloclasticus* sp. A5
5766	6599	BAA76322.1	100	dihydrodiol dehydrogenase	*Alcaligenes faecalis*
		AAK14786.1	52	cis-2,3-dihydroxy-2,3-dihydrobiphenyl dehydrogenase	*Pseudomonas* sp. Cam-1
6656	8014	BAA76323.1	98	dioxygenase large subunit	*Alcaligenes faecalis*
		AAK62353.1	71	dioxygenase large subunit	*Burkholderia* sp. DBT1
8044	8562	BAA76324.1	100	dioxygenase small subunit	*Alcaligenes faecalis*
		AAK62354.1	68	dioxygenase small subunit	*Burkholderia* sp. DBT1
8721	9311	BAA76325.1	96	Isomerase	*Alcaligenes faecalis*
		AAD09871	49	Isomerase	*Burkholderia* sp. RP007
10276	10890	BAA76326.1	87	glutathione-S-transferase	*Alcaligenes faecalis*
		ABX32691	54	glutathione-S-transferase	*Deftia acidovorans* SPH-1
10968	11930	BAA76327.1	98	trans-2'-carboxybenzalpyruvate hydratase-aldolase	*Alcaligenes faecalis*
		BAA24966	45	trans-2'-carboxybenzalpyruvate hydratase-aldolase	*Nocardioides* sp. KP7
11970	13019	BAA76328	99	1-hydroxy-2-naphthoate dioxygenase	*Alcaligenes faecalis*
		CAJ48218.1	45	1-hydroxy-2-naphthoate dioxygenase	*Bordetella avium* 197N
12841	14484	BAA76329.1	99	2-carboxybenzaldehyde dehydrogenase	*Alcaligenes faecalis*
		ZP_01568766	55	betaine-aldehyde dehydrogenase	*B. multivorans* ATCC 17616
15378	16286	BAA76330.1	99	3,4-dihydroxyphenanthrene dioxygenase	*Alcaligenes faecalis*
		CAG17579	62	extradiol dioxygenase	*Sphingomonas* sp. CHY-1
16362	17890	BAA76331.1	100	1-hydroxy-2-naphtoaldehyde dehydrogenase	*Alcaligenes faecalis*
		ABM79801	62	aldehyde dehydrogenase	*Sphingobium yanoikuyae* B1
24884	25882	BAA76332.1	100	hydratase-aldolase	*Alcaligenes faecalis*
		ABM79813.1	67	hydratase-aldolase	*Sphingobium yanoikuyae* B1

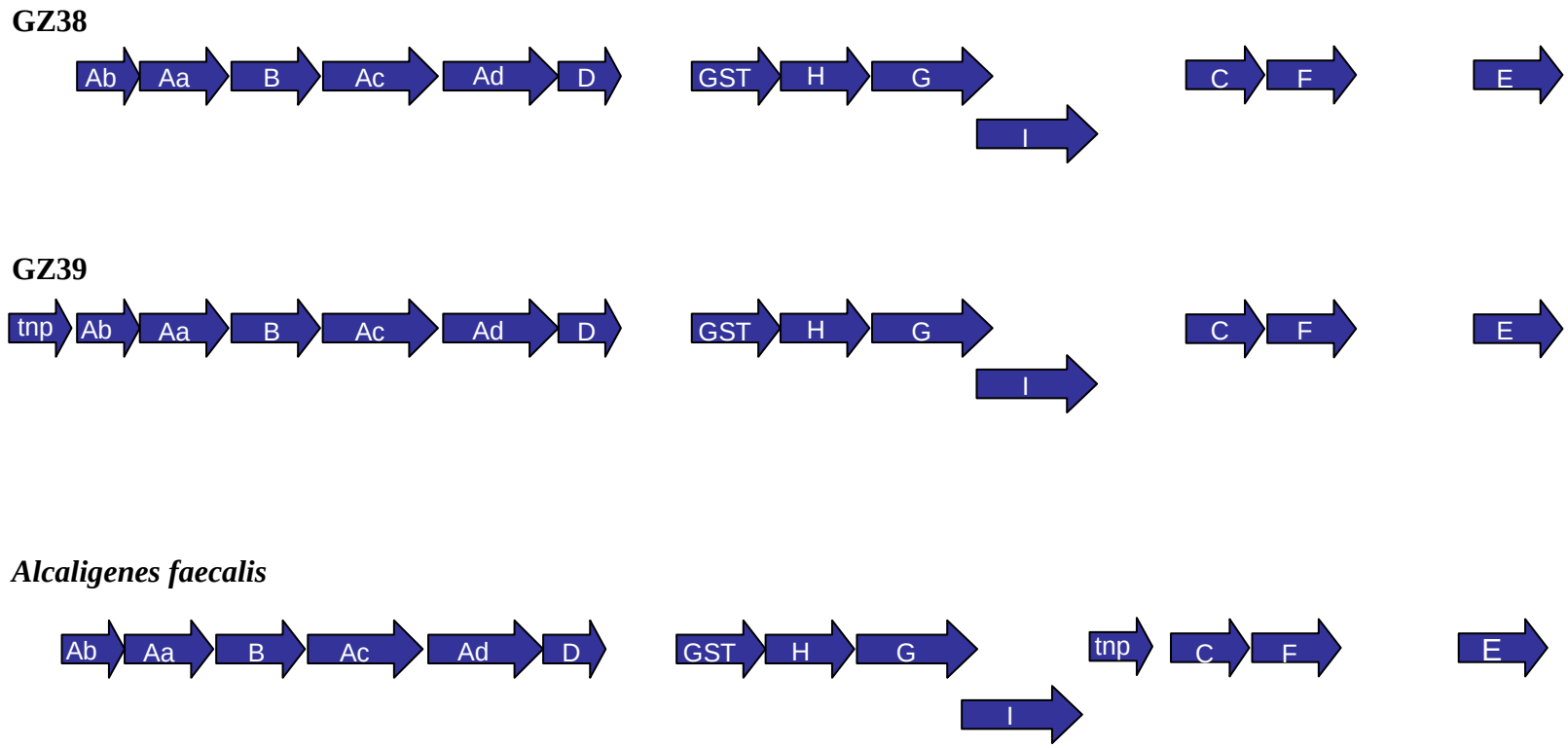


Figure 8. Gene organization of GZ38, GZ39, and AFK2. Open reading frames (ORFs) are indicated with arrows with abbreviated gene names inside each ORF.

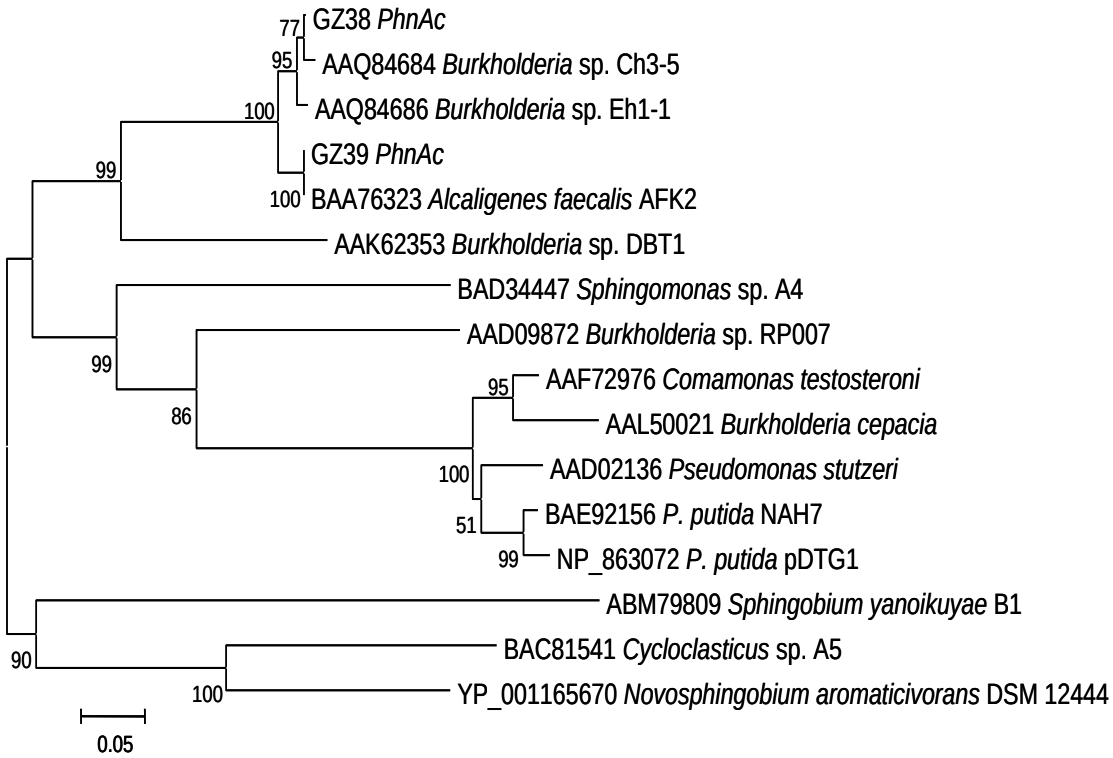


Figure 9. Phylogenetic tree of GZ38 and GZ39 phenanthrene dioxygenase large subunit and related sequences. The dendrogram was constructed from a ClustalW alignment of amino acid sequences by neighbor-joining analysis using MEGA 4.0. Bootstrap values greater than 50 are indicated at branch nodes.

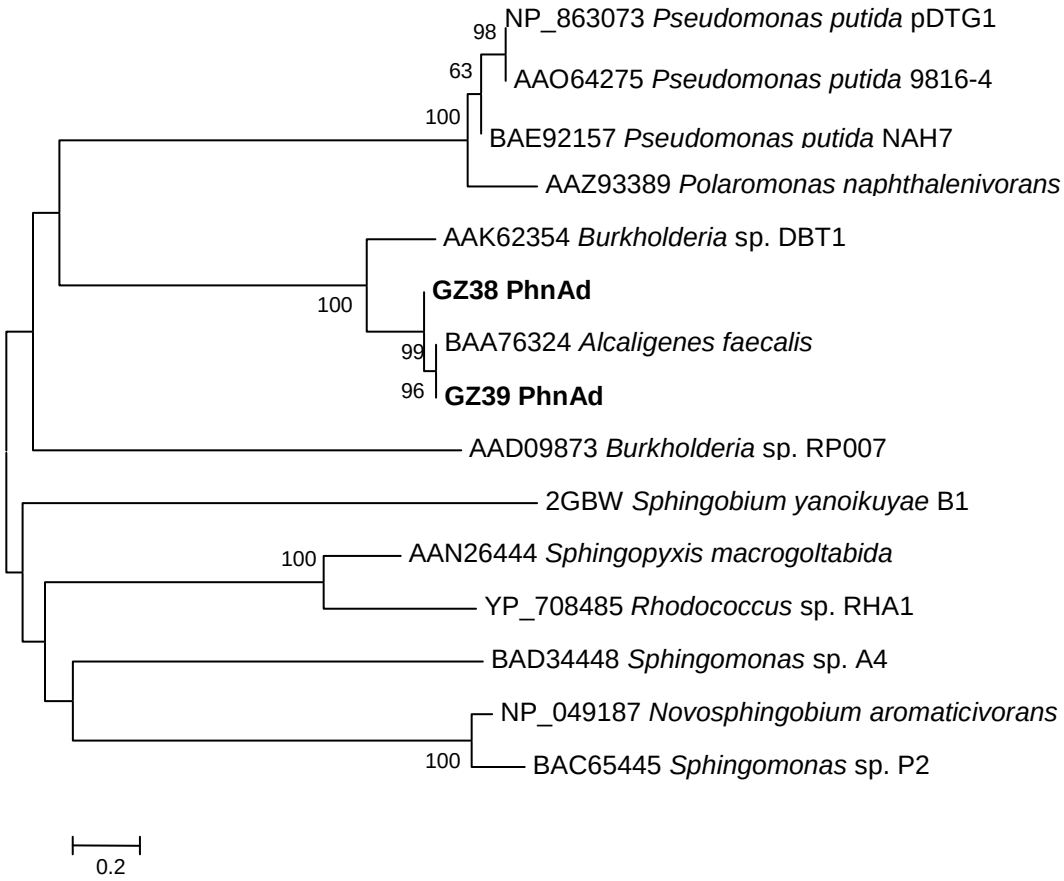


Figure 10. Phylogenetic tree of GZ38 and GZ39 phenanthrene dioxygenase small subunit and related sequences. The dendrogram was constructed from a ClustalW alignment of amino acid sequences by neighbor-joining analysis using MEGA 4.0. Bootstrap values greater than 50 are indicated at branch nodes.

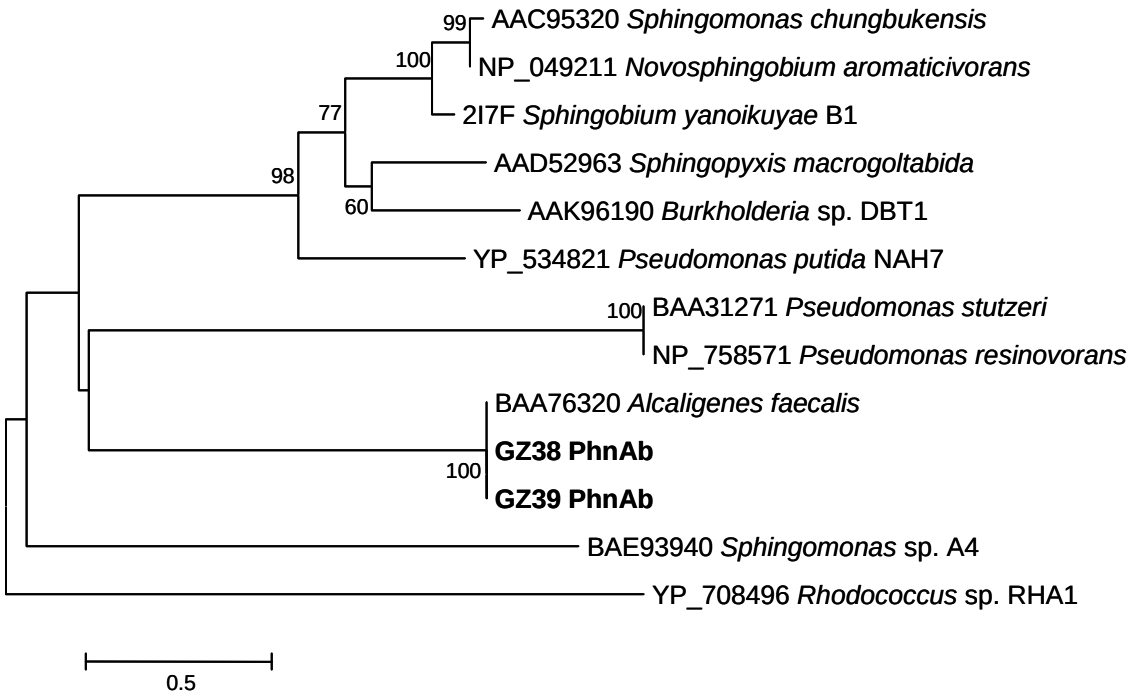


Figure 11. Phylogenetic tree of GZ38 and GZ39 phenanthrene dioxygenase ferredoxin and related sequences. The dendrogram was constructed from a ClustalW alignment of amino acid sequences by neighbor-joining analysis using MEGA 4.0. Bootstrap values greater than 50 are indicated at branch nodes.

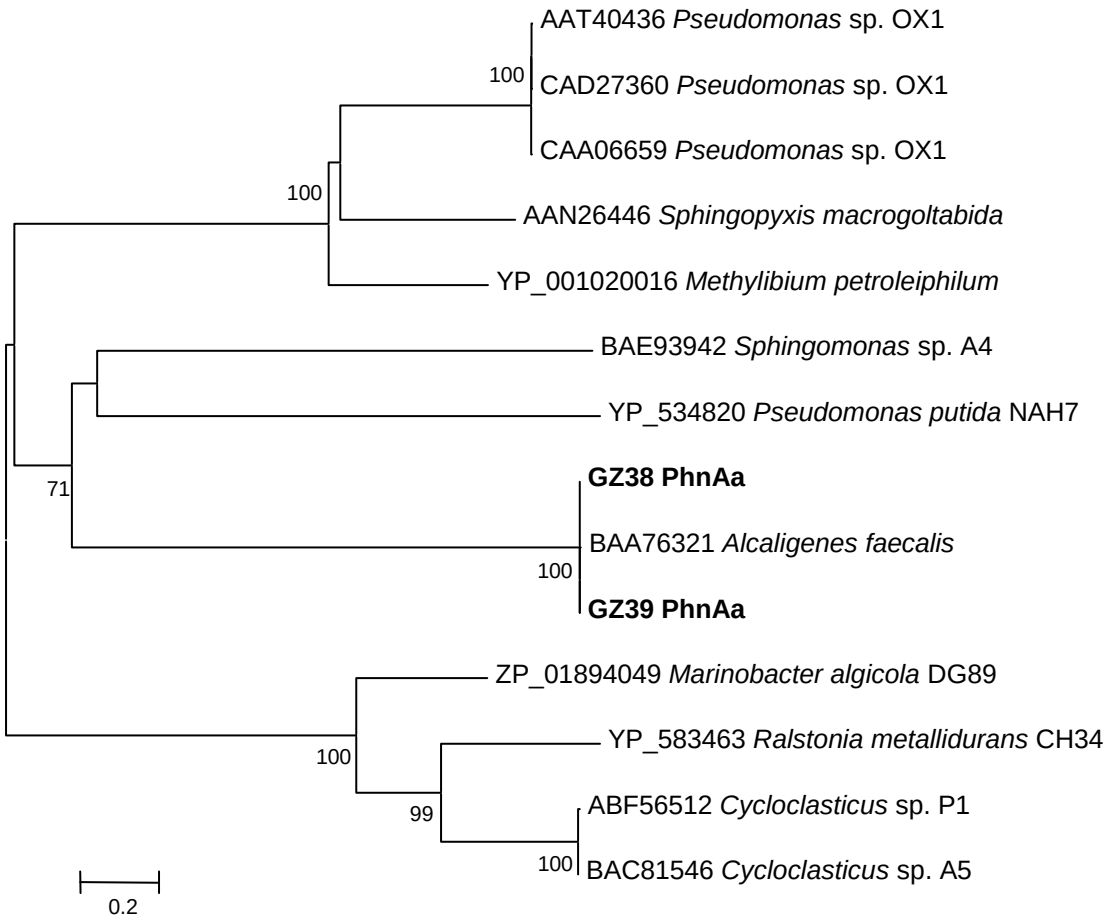


Figure 12. Phylogenetic tree of GZ38 and GZ39 phenanthrene dioxygenase ferredoxin reductase and related sequences. The dendrogram was constructed from a ClustalW alignment of amino acid sequences by neighbor-joining analysis using MEGA 4.0. Bootstrap values greater than 50 are indicated at branch nodes.

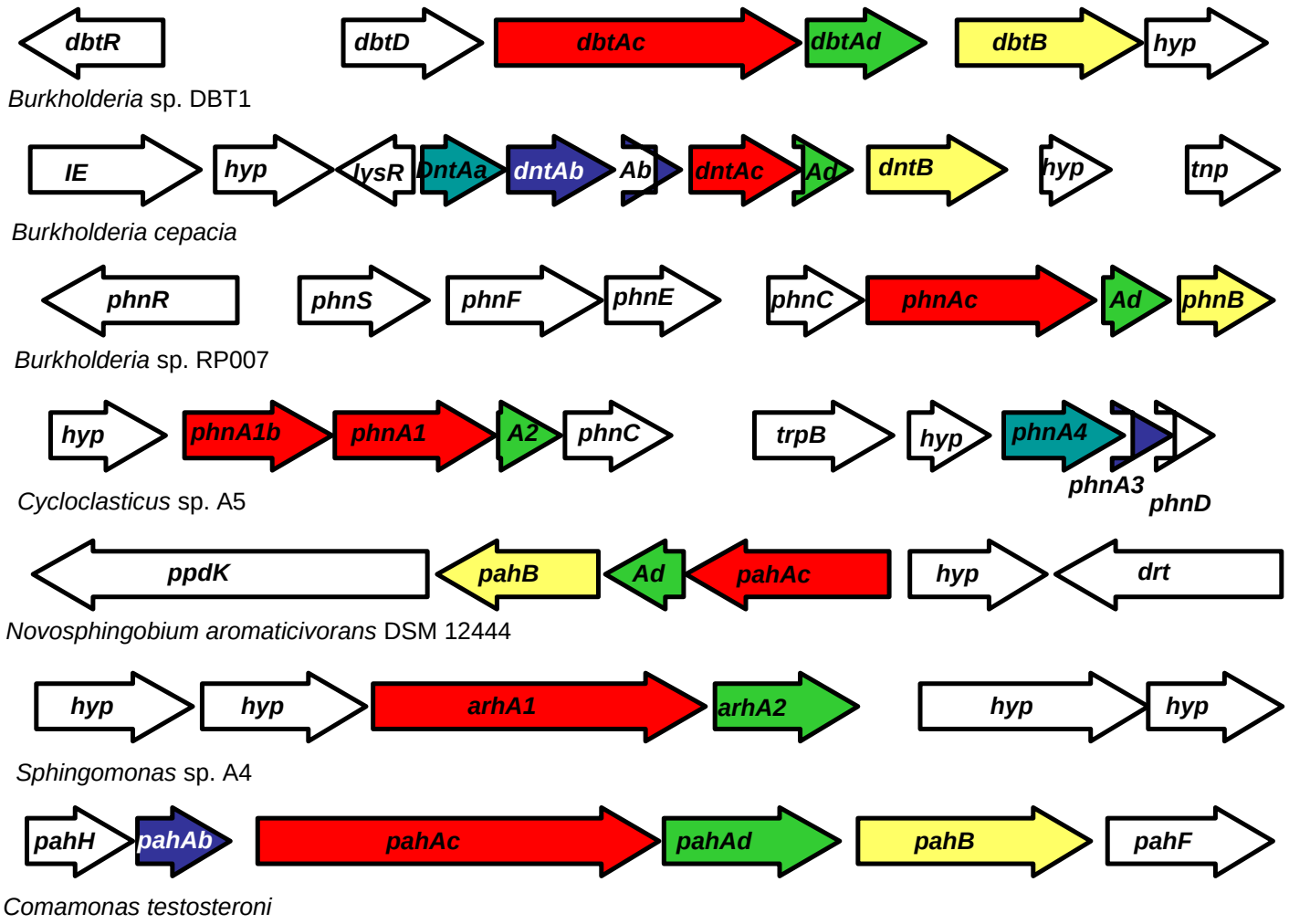


Figure 13. Phenanthrene degradation operons of organisms with matches to GZ38 and GZ39

Chapter 2

Identification and sequence of phthalate degradation genes in GZ38 and GZ39

1. Nucleotide sequence of the phthalate degradation genes of GZ38

The nucleotide sequence data from the phenanthrene degradation operon in GZ38 suggests that phenanthrene degradation proceeds through phthalate. Specifically, the presence of the gene encoding 1-hydroxy-2-naphthoate dioxygenase, led to the hypothesis that GZ38 might also possess genes for phthalate degradation.

In Gram negative bacteria, phthalate is dihydroxylated to 4, 5-dihydro-dihydroxyphthalate by a two component enzyme that consists of a phthalate dioxygenase and a phthalate dioxygenase reductase. A dehydrogenase then removes two electrons and two hydrogens resulting in 4, 5-dihydroxyphthalate and NADH. Next 4, 5-dihydroxyphthalate decarboxylase removes a carboxyl group to form protocatechuate which can be metabolized to tricarboxylic acid cycle intermediates (figure 14).

It was initially unclear if GZ38 had phthalate degradation genes because it did not grow on MSB plates with phthalate. Because GZ38 was one of several *Comamonas* isolates that originated from the Passaic River and some of those isolates possess phthalate degradation genes (Wang, Zhou et al. 1995), PCR reactions using primers designed based on sequence from other Passaic River *Comamonas testosteroni* phthalate

degraders were attempted in order to assess whether or not these genes were present in GZ38.

Primers were designed to amplify phthalate gene sequences GZ38 which were derived from phthalate gene sequences of other *Comamonas* isolates YZW-B and YZW-E from the Passaic River. These primers amplified a DNA sequence from GZ38 that included all of the genes in the pathway but only amplified a 500 bp region of the reductase gene. The last 300 nucleotides of the reductase gene were obtained through inverse PCR. The gene sequence amplified from GZ38 is 7.5 kb in length and contains five complete open reading frames; *ophA1*, *ophC, ophR, ophB,* and *ophA2* (figure 15). Additionally, there is nucleotide sequence between *ophA1* and *ophC* that aligns with two fragments of a phthalate transporter gene. Each orf in the operon begins with an ATG start codon and is preceded by a ribosome binding site. The functions of these orfs were deduced based on sequence homology to previously described genes and corresponding amino acid sequences. They code for phthalate dioxygenase reductase, 4,5-dihydroxyphthalate decarboxylase, *cis*-phthalate dihydrodiol dehydrogenase, and phthalate dioxygenase respectively.

BLAST searches with the predicted protein sequences were performed for the five complete orfs and the single partial orf (Table 3). The first orf is 963 nucleotides in length. The best predicted protein match is to what is labeled a ferredoxin from *Ralstonia picketti* 12D (accession number ZP_02008300) at 85%. However, when this sequence is queried against the conserved domain database, the N-terminal end matches pfam00970, FAD_binding_6, Oxidoreductase FAD-binding domain and the C-terminal end aligns

with cd00207, fer2, 2Fe-2S iron-sulfur cluster binding domain suggesting that this sequence may be mis-annotated (Marchler-Bauer A 2004). This is confirmed by examining the best match to a characterized protein, which is to phthalate dioxygenase reductase from *Burkholderia cepacia* DB01 (accession number 2PIA_A) at 80% sequence identity.

Based on a BLASTX search of the sequence upstream of the *ophA1* gene, there is partial gene sequence of a phthalate transporter gene. It begins 278 nucleotides upstream of *ophA1* with an ATG start codon. In YZW-B, this open reading frame is 1344 nucleotides in length while for GZ38 there are 452 nucleotides. The first 274 nucleotides of this region in GZ38 match the first 274 nucleotides of the *ophD* gene in YZW-B with 97% identity and nucleotides 275-452 of GZ38 match nucleotides 1166 to 1344 of YZW-B with 96% identity. When this sequence is queried against the Genbank database, all of the matches are also to the beginning and the end of *ophD* genes. The best match is to *Ralstonia picketti* 12D (accession number ZP_02008299) at 83% for the first 91 amino acids and 84% for the last 59 amino acids.

The second orf begins 43 bases downstream of the end of the partial transporter sequence and is 993 nucleotides long. The predicted amino acid sequence is 99% identical to that of a 4,5-dihydroxyphthalate decarboxylase from *Comamonas testosteroni* (accession number BAA03974).

The next orf is 594 nucleotides long and the putative amino acid sequence is 85% identical to a MarR transcriptional regulator from *Ralstonia pickettii* 12D (accession number ZP_02008298).

After this orf there are five nucleotides which are followed by a 22 base long inverted repeat (CTCCAGGAAAGAAAAAACCTCG) that is separated by seven nucleotides. Following the second repeat, there are seven nucleotides leading up to the next orf.

This orf is 1176 nucleotides long and the amino acid sequence matches an oxidoreductase from *Ralstonia pickettii* 12D (accession number ZP_02008294) with 86% identity. The best characterized protein match at 84% is to a phthalate dihydrodiol dehydrogenase from *Burkholderia cepacia* DB01 (accession number AAD03557).

The final orf is 1330 nucleotides long and is 19 bases away from *ophB*. The N-terminal region of the protein sequence aligns with cd03479, N-terminal Rieske domain of the oxygenase alpha subunit. The predicted protein sequence is 90% identical to a rieske domain protein from *Ralstonia pickettii* 12D (accession number ZP_02008293). The best characterized match is to a phthalate dioxygenase from *Burkholderia cepacia* DB01 at 86% identity (accession number AAD03558). When the protein sequence of GZ38 is compared to other phthalate dioxygenase protein sequences it branches together with the aforementioned *Ralstonia* and *Burkholderia* sequences along with a phthalate dioxygenase from *Burkholderia vietanamiensis* G4 (accession number YP_001116718). It is more distantly related (72-76% identity) to the amino acid sequences of *Polaromonas* sp. JS666 (accession number YP_548741), *Pseudomonas putida* (accession

number QO5183), *Polaromonas naphthalenivorans* (accession number YP_973790), and *Comamonas testosteroni* KF-1 (accession number ZP_01521408).

2. Nucleotide sequence of the phthalate degradation genes of GZ39

Since GZ39 also possesses the same putative gene that encodes 1-hydroxy-2-naphthoate dioxygenase, it was expected to have the genes for phthalate degradation. Interestingly, it too did not grow on MSB plates with phthalate and it also did not produce PCR products with the phthalate primers that were successful on GZ38.

A southern blot experiment was then performed to determine whether GZ39 possessed genes encoding enzymes for phthalate degradation. A probe was designed to the large subunit of the phthalate dioxygenase gene of GZ38 and genomic DNA from GZ38 was used as template. The probe hybridized to genomic DNA from GZ39 digested with EcoRI and PstI and to DNA digested with PstI and SalI. Since these fragments were less than 2kb in length, a different approach was taken in order to obtain the phthalate operon sequence. Clones from the previously constructed cosmid library (Goyal and Zylstra 1996) were pooled by row and column and screened by PCR using primers for the reductase gene. This produced many false positives and led to the use of a new set of degenerate primers (Ní Chadhain 2007) that amplify a 650bp fragment of the dioxygenase gene. A positive cosmid clone was identified and subcloned using the pZERO Background-kan vector (Invitrogen). A positive subclone containing an 8.9 kb insert was identified using the same primers and sequenced by primer walking across the

insert beginning with m13_f and m13_r primers located on either end and out from both ends of the dioxygenase gene.

Seven open reading frames are contained within this sequence. BLAST searches with this sequence revealed the presence of a phthalate degradation operon. Based on GenBank database searches, these genes code for phthalate dioxygenase reductase (*ophA1*), 4, 5- dihydroxyphthalate dehydrogenase (*ophB*), a regulator, 4, 5-dihydroxyphthalate decarboxylase (*ophC*), phthalate dioxygenase (*ophA2*), *traG*, and a transposase (Figure 16). Each orf begins with an ATG start codon and all but *ophB* are preceded by a ribosome binding site.

The gene organization and the putative amino acid sequences for all of the genes in the operon differ in both placement and orientation from that of GZ38, YZW-B, and YZW-E. BLAST searches with the predicted protein sequences were done for the five open reading frames that correspond to phthalate gene sequences (Table 4). The first orf is 717 nucleotides long. The predicted amino acid sequence matches a phthalate dioxygenase reductase from *Burkholderia cepacia* DB01 (accession number P33164) with 57% sequence identity. When the protein sequences of *ophA1* from GZ39 and GZ38 are compared, they share only 57% identity to one another.

The second orf is 1200 nucleotides long and begins 245 nucleotides downstream of *ophA1*. The best protein match is to a 4,5-dihydroxyphthalate dehydrogenase from *Polaromonas naphthalenivorans* CJ2 (accession number YP_973786) at 58% sequence similarity. This protein sequence bears only 53% sequence identity to the corresponding sequence from GZ38.

A putative transcriptional regulator is located thirteen nucleotides downstream of the dehydrogenase. This orf is 579 nucleotides long and the best amino acid match is to a MarR transcriptional regulator from *Polaromonas* JS666 (accession number YP_548738) at 65%. When compared to the protein sequence of the regulator in the phthalate operon of GZ38, the regulator from GZ39 is 64% identical.

The next orf matches 4,5-dihydroxyphthalate decarboxylase from *Polaromonas naphthalenivorans* CJ2 (accession number YP_973789) with 79% amino acid identity. The orf is 990 nucleotides long and is located 59 nucleotides downstream of the preceding orf. The phthalate decarboxylases from GZ38 and GZ39 share only 62% identity.

Fifteen nucleotides separate the decarboxylase from the fifth orf, which codes for phthalate dioxygenase. This orf is 1317 nucleotides long and it matches a phthalate dioxygenase from *Comamonas testosteroni* KF1 (accession number ZP_01521408) with 79% sequence identity. The phthalate dioxygenase from GZ39 matches the phthalate dioxygenase from GZ38 with 77% amino acid identity.

3. Discussion-

Phylogenetic trees were constructed in order to compare the dehydrogenase, decarboxylase, reductase, and dioxygenase amino acid sequences of the phthalate

operons of GZ38 and GZ39 to isolates with related proteins. The protein sequence of the dehydrogenase gene of GZ38 is most closely related to dehydrogenases from YZW-B and *Ralstonia pickettii* 12D and also groups with dehydrogenases from *Burkholderia vietnamiensis* G4 and *Burkholderia cepacia* DB01. The dehydrogenase from GZ39 falls into a separate clade with dehydrogenases from *Rhodobacterales bacterium* H and *Polaromonas* sp. JS666 (Figure 17). In the protein tree containing decarboxylase amino acid sequences (Figure 18), GZ38 and GZ39 again fall into separate clades. GZ38 closely matches decarboxylases from a *Comamonas testosteroni* isolate and YZW-B. GZ39 has no close relatives and groups with *Polaromonas naphthalenivorans*, *Verminephrobacter eiseniae* EF, and *Comamonas testosteroni* KF1. In the reductase tree (Figure 19), GZ38 groups closely with YZW-B and *Ralstonia picketti* 12D while GZ39 forms a clade with *Polaramonas* sp. JS666 and a *Pseudomonas putida* isolate. The phthalate dioxygenase protein sequence tree (Figure 20) shows GZ38 and GZ39 in separate clades. GZ38 groups with phthalate dioxygenases from *Ralstonia pickettii* 12D, *Burkholderia cepacia* DB01 and *Burkholderia vietnamiensis* G4 while GZ39 groups with *Pseudomonas putida* and *Polaromonas* sp. JS666.

While the reductase, regulator, decarboxylase, dehydrogenase, and dioxygenase protein sequences from GZ38 bear around 85% homology to the corresponding protein sequences from *Ralstonia pickettii* 12D, the decarboxylase protein from GZ38 is 98% identical to a decarboxylase from *Comamonas testosteroni*.

The phthalate genes of GZ39 are less similar to other known phthalate genes. While the amino acid matches are to *Polaromonas* sp. JS666 and *Polaromonas naphthalenivorans* CJ2 for the reductase, dehydrogenase, regulator, and decarboxylase, the best dioxygenase match is to phthalate dioxygenases from GZ38 and *Comamonas testosteroni* KF1.

The gene organization of phthalate genes of isolates with close BLAST matches to the phthalate genes of GZ38 and GZ39 was also compared (Figures 21 and 22). With regard to gene order, GZ38, YZW-B, *Burkholderia vietnamiensis* G4, and *Burkholderia cepacia* DB01 have some commonalities. In all, the reductase is the initial gene in the sequence and is transcribed from right to left. For GZ38 and DB01, a transporter gene fragment follows. YZW-B also has a transporter gene. GZ38 appears to be missing about 1kb of nucleotide sequence in the middle of the transporter gene. DB01 has a frameshift in the *ophD* gene sequence that renders it inactive but it is still able to grow on phthalate. It is interesting that both of these isolates have anomalies in their phthalate transporter nucleotide sequences. This gene was determined to be unnecessary for phthalate uptake and growth in DB01 suggesting that there is an alternative mode of phthalate transport into the cell (Chang and Zylstra 1998). The first 91 amino acids in the transporter of YZW-B match the first 91 amino acids in the partial orf in GZ38 with 98% sequence identity. GZ38 is missing nucleotide sequence that would code for the next 300 amino acids but has 177 nucleotides that match the last 178 nucleotides of YZW-B with 96% identity. This large gap in nucleotide sequence most likely renders the protein inactive and may explain the organisms inability to use phthalate as a sole carbon source. GZ38, YZW-B, and DB01 have a decarboxylase and a regulator next that is transcribed from

left to right. DB01 has two orfs that are not present in GZ38; a gene that codes for quinolinate phosphoribosyl transferase (*ophE*) which serves to increase growth on phthalate, and a transposase. In transferring these genes between species, there may be some propensity toward losing the middle portion of the sequence, which contains *ophD*, *ophE*, and a transposase, as both YZW-B and GZ38 lack part of these genes. Finally, GZ38, YZW-B, and DB01 have a dehydrogenase and a dioxygenase that are both transcribed from right to left. GZ38 has a putative transposase upstream of the dioxygenase.

There are also similarities between the phthalate operons of GZ38 and *Ralstonia pickettii* 12D. The phthalate operon of *Ralstonia pickettii* 12D is ordered *ophA2-ophB*-hyp-*marR*-porin-*ophC-ophD-ophA1*. While 12D has an unknown orf and a porin that are not present in GZ38, The organization of the remaining genes is a mirror image to that of GZ38. Due to the high percent identity (80-85%) across the GZ38 phthalate operon to phthalate operons from other species, it is possible that the genes may have been transferred as a cassette.

The organization of the phthalate genes of GZ39 is highly dissimilar to other known phthalate operons (Figure 22). The gene organization of other isolates such as *Polaromonas* sp. JS666 and *Comamonas testosteroni* KF1, which have protein matches to GZ39, is very different. In *Polaromonas* sp. JS666, the operon arrangement is *marR-ophB-ophA1-ophA2*-hyp-*ophC*-transposase. In KF1, the operon arrangement is *ophA2-ophC*-hyp-*ophA1-ophB* with all of these orfs in the same direction. Furthermore, the best BLAST matches are to genes from a variety of organisms. This coupled with the unique

organization suggests the phthalate genes in GZ39 may have been acquired from a number of organisms and occurred during multiple acquisition events.

Comparative sequence alignments of the putative protein sequences for the phthalate genes of GZ38 and GZ39 show that they are divergent from one another. That they appear to be unrelated is interesting given that the phenanthrene degradation genes of these two isolates are virtually identical. If degradation of phenanthrene proceeds through phthalate, it most likely occurs with genes that are very different from one another but code for the same functions.

Given that neither isolate can grow on phthalate, the phthalate pathways in both isolates may be used solely for the purpose of phenanthrene degradation. GZ38 has an incomplete phthalate transporter gene and GZ39 does not have a phthalate transporter present in the operon. Since the phthalate genes of GZ38 are related to *Burkholderia cepacia* DB01, *Burkholderia vietnamiensis* G4, and YZW-B, isolates that contain phthalate transporter genes, it is possible that GZ38 may have also had a functioning transporter at one time. A phthalate transporter would not be necessary for phenanthrene metabolism because if phthalate is a metabolite of the phenanthrene degradation pathway then it would already be inside the cell.

Just as with both phenanthrene operons, there are transposases associated with both phthalate operons. It is possible that GZ38 and GZ39 could have acquired the phthalate degradation genes either before or after acquiring the phenanthrene degradation genes.

The best protein matches for the genes of GZ38 are to genes from different isolates but are similar in organization to *Burkholderia cepacia* DB01, *Burkholderia vietnamiensis* G4, and YZW-B. All of these isolates have transposases associated with the phthalate operon. Since there is a conservation of this phthalate operon, it is possible that for GZ38 the phthalate genes are derived from some common ancestor and were acquired all at once. Sequence analysis revealed the presence of a 17 nucleotide long inverted repeat that is separated by six nucleotides that occurs in GZ38 between *ophR* and *ophB*. This sequence is not present in YZW-B or DB01. However, DB01 has *ophE* and a transposase in this region. Based on the presence of this repeat it can be inferred that additional sequence including a transposon may have once been in this region of the sequence. Based on the sequence similarities and gene organization of the respective phthalate operons of GZ38 and DB01 it is possible that *ophE* and the transposon were deleted from the sequence at some point before it reached GZ38. This supports the theory that these genes have been exchanged between isolates via transposition.

The possible recruitment of phthalate operons in both GZ38 and GZ39 could have been a response to environmental pressure. Precipitating events would include a loss of function in an existing pathway and an increase in phenanthrene in the environment. It is also possible that the phthalate genes provide an advantage such as more efficient degradation of phenanthrene. While the phthalate operon of GZ38 most likely was acquired all at once, the story of the phthalate operon of GZ39 is less clear. Since there are no other known isolates with similar gene organization and the protein matches are

low (ranging from 53-79%), it is unclear whether these genes were acquired simultaneously from one organism or were recruited at different times from multiple organisms.

Table 3. Best BLASTP matches in the GenBank database to the GZ38 phthalate degradation sequence

Start	End	Accession #	Identity	Name	Organism
1091	129	ZP_02008300	85	ferredoxin	*Ralstonia pickettii* 12D
		2PIA_A	80	phthalate dioxygenase reductase	*Burkholderia cepacia* DB01
1865	2857	BAA03974	99	4,5-dihydroxyphthalate decarboxylase	*Comamonas testosteroni*
		ZP_02008298	93	4,5-dihydroxyphthalate decarboxylase	*Ralstonia pickettii* 12D
3149	3742	ZP_02008296	85	transcriptional regulator, MarR family	*Ralstonia pickettii* 12D
		YP_001116714	86	transcriptional regulator, MarR family	*Burkholderia vietnamiensis* G4
4592	3805	ZP_02008294	86	oxidoreductase	*Ralstonia pickettii* 12D
		AAD03557	84	phthalate dihydrodiol dehydrogenase	*Burkholderia cepacia* DB01
6329	5066	ZP_02008293	90	rieske (2Fe-2S) domain protein	*Ralstonia pickettii* 12D
		AAD03558	86	phthalate dioxygenase	*Burkholderia cepacia* DB01

Table 4. Best BLASTP matches in the GenBank database to the GZ39 phthalate degradation sequence

Start	End	Accession #	Identity	Name	Organism
661	1377	2PIA_A	57	dioxygenase reductase	*Burkholderia cepacia* DB01
		P33164	57	phthalate dioxygenase reductase	*Burkholderia cepacia*
1623	2822	YP_973786	58	4,5-dihydroxyphthalate dehydrogenase	*Polaromonas naphthalenivorans* CJ2
		YP_548739	57	4,5-dihydroxyphthalate dehydrogenase	*Polaromonas* sp. JS666
2836	3414	YP_001628850	60	transcriptional regulator, MarR family	*Bordetella petrii*
		YP_548738	65	transcriptional regulator, MarR family	*Polaromonas* sp. JS666
4463	3474	YP_973789	79	4,5-dihydroxyphthalate decarboxylase	*Polaromonas naphthalenivorans* CJ2
		YP_997681	75	4,5-dihydroxyphthalate decarboxylase	*Verminephrobacter eiseniae* EF01-2
5795	4479	ZP_01521408	79	rieske (2Fe-2S) domain protein	*Comamonas testosteroni* KF-1
		YP_548741	76	phthalate dioxygenase	*Polaromonas* sp. JS666

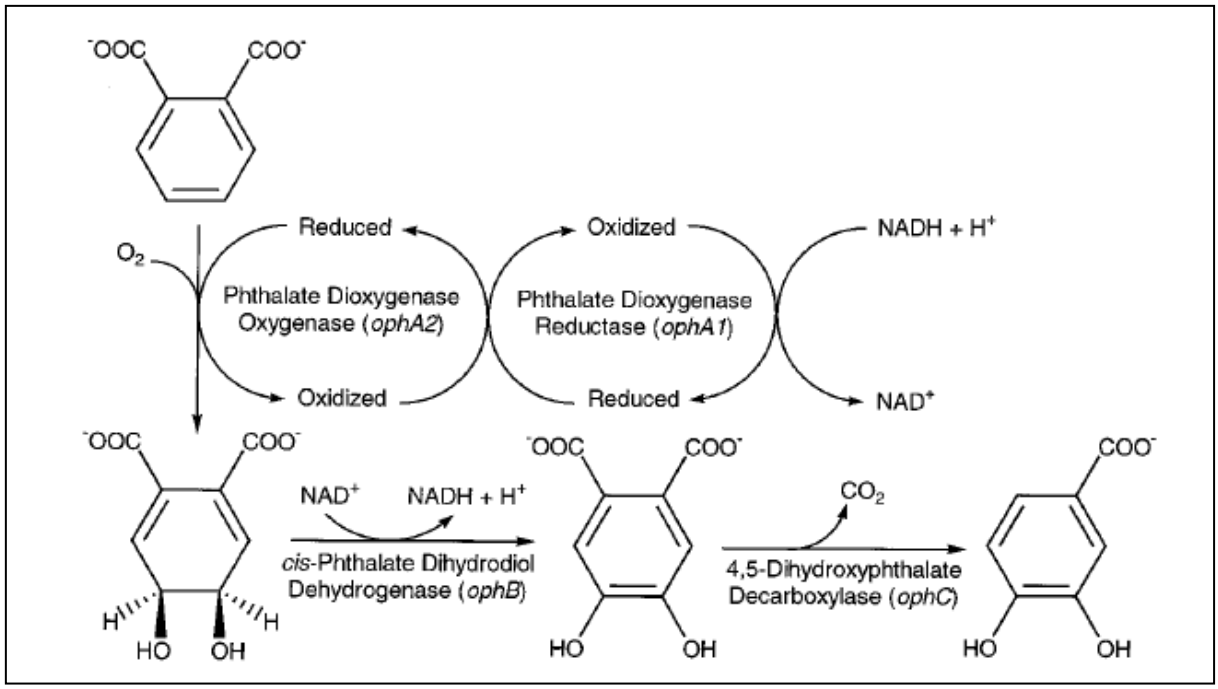


Figure 14. Pathway of phthalate degradation in Gram negative bacteria

Figure 15. Gene organization of the phthalate degradation operon of GZ38

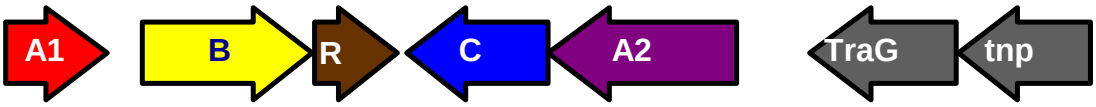


Figure 16. Gene organization of the phthalate degradation operon of GZ39

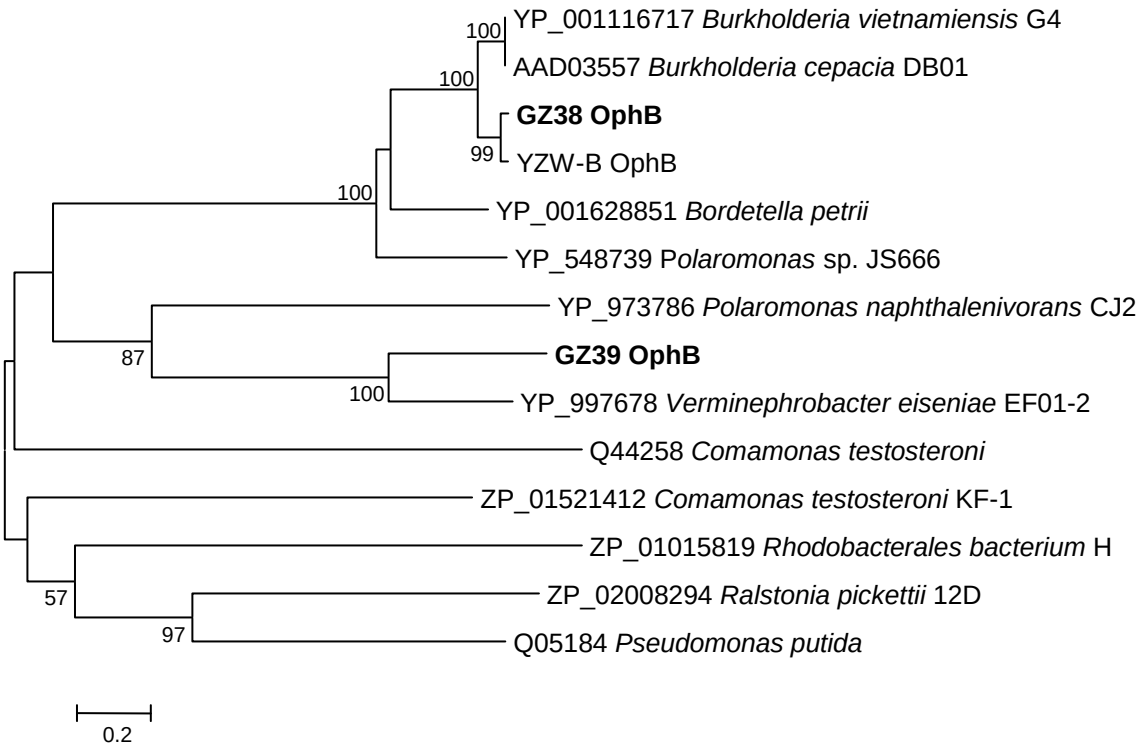


Figure 17. Phylogenetic tree of GZ38 and GZ39 phthalate dehydrogenase and related sequences. The dendrogram was constructed from a ClustalW alignment of amino acid sequences by neighbor-joining analysis using MEGA 4.0. Bootstrap values greater than 50 are indicated at branch nodes.

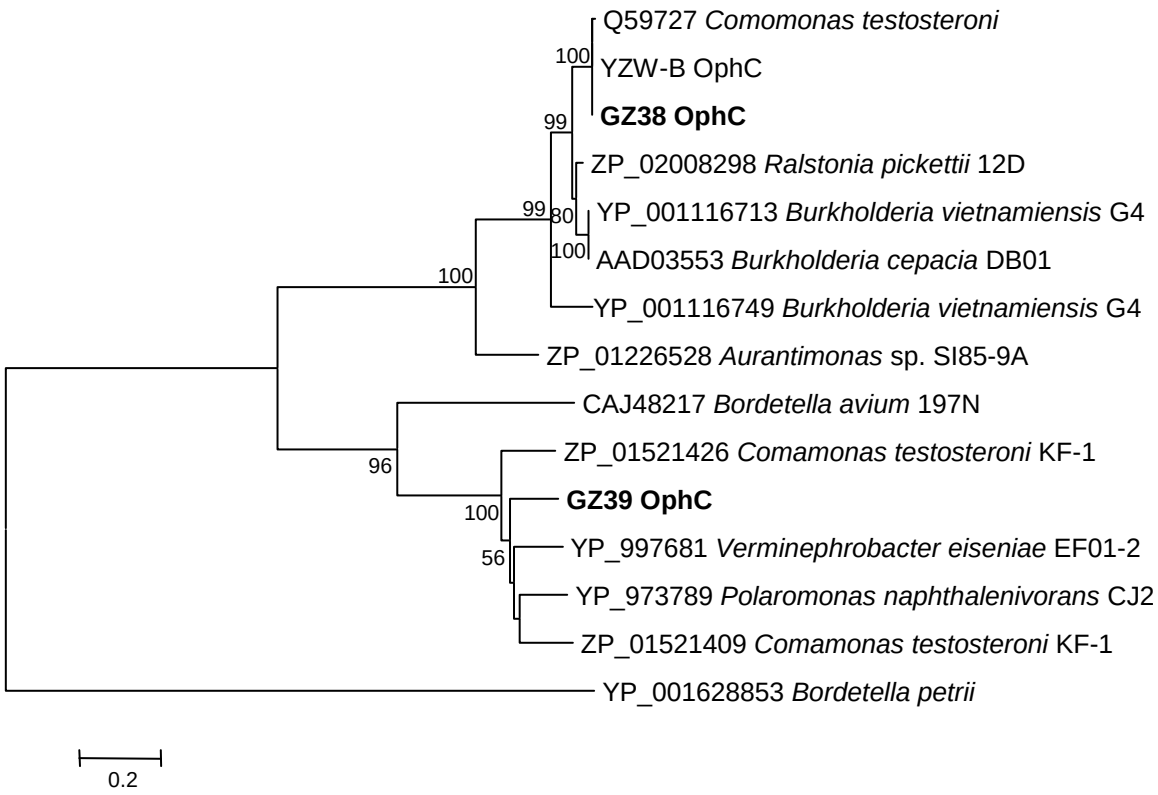


Figure 18. Phylogenetic tree of GZ38 and GZ39 phthalate decarboxylase and related sequences. The dendrogram was constructed from a ClustalW alignment of amino acid sequences by neighbor-joining analysis using MEGA 4.0. Bootstrap values greater than 50 are indicated at branch nodes.

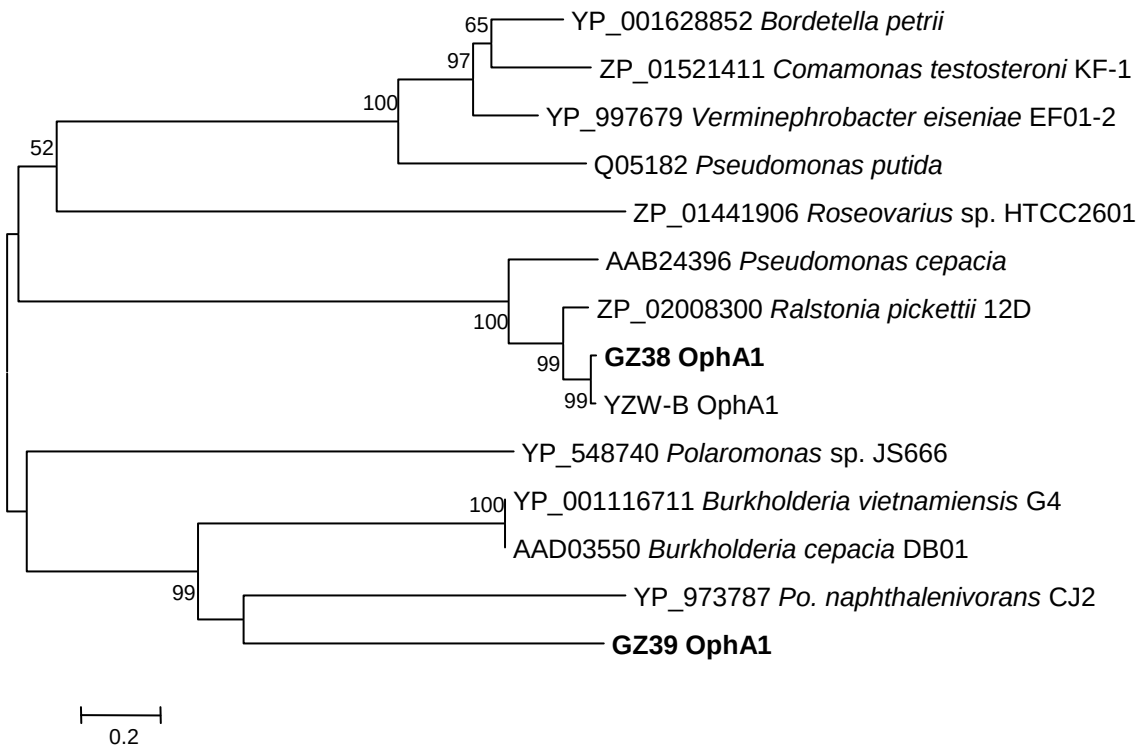


Figure 19. Phylogenetic tree of GZ38 and GZ39 phthalate reductase and related sequences. The dendrogram was constructed from a ClustalW alignment of amino acid sequences by neighbor-joining analysis using MEGA 4.0. Bootstrap values greater than 50 are indicated at branch nodes.

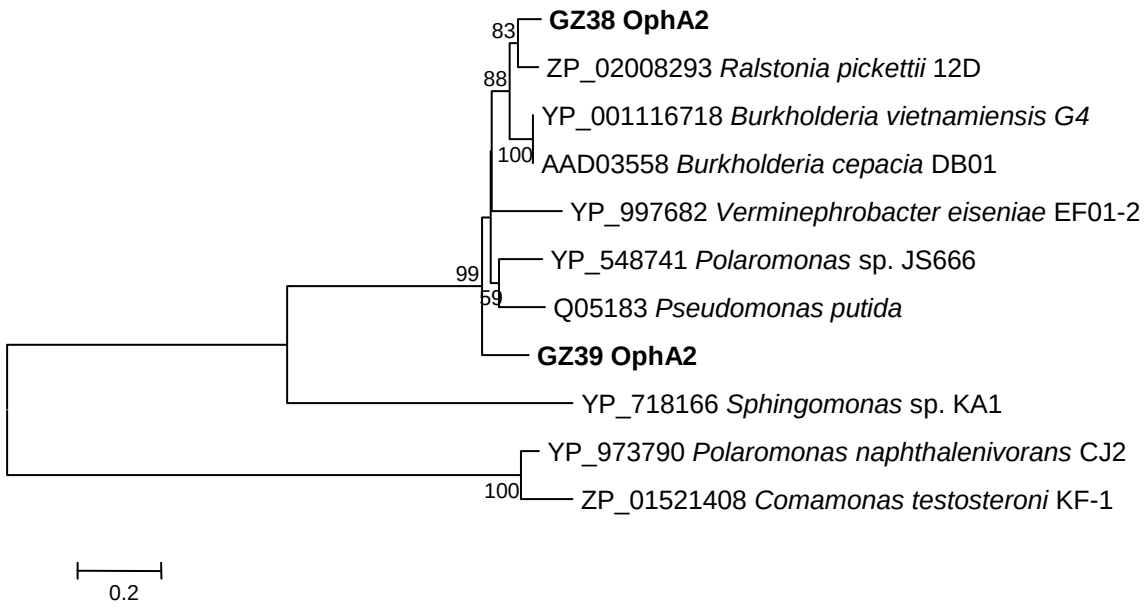


Figure 20. Phylogenetic tree of GZ38 and GZ39 phthalate dioxygenase and related sequences. The dendrogram was constructed from a ClustalW alignment of amino acid sequences by neighbor-joining analysis using MEGA 4.0. Bootstrap values greater than 50 are indicated at branch nodes.

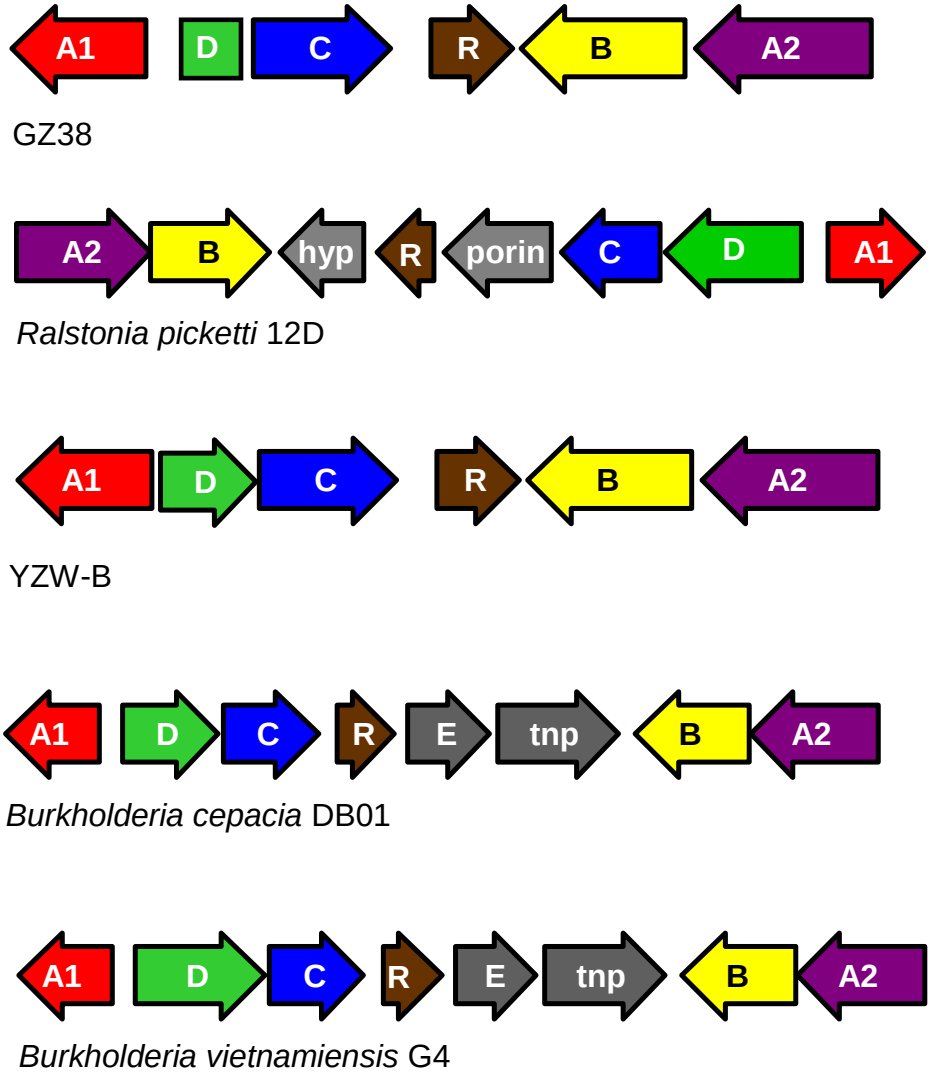


Figure 21. Gene organization of the phthalate operon of GZ38 and related operons

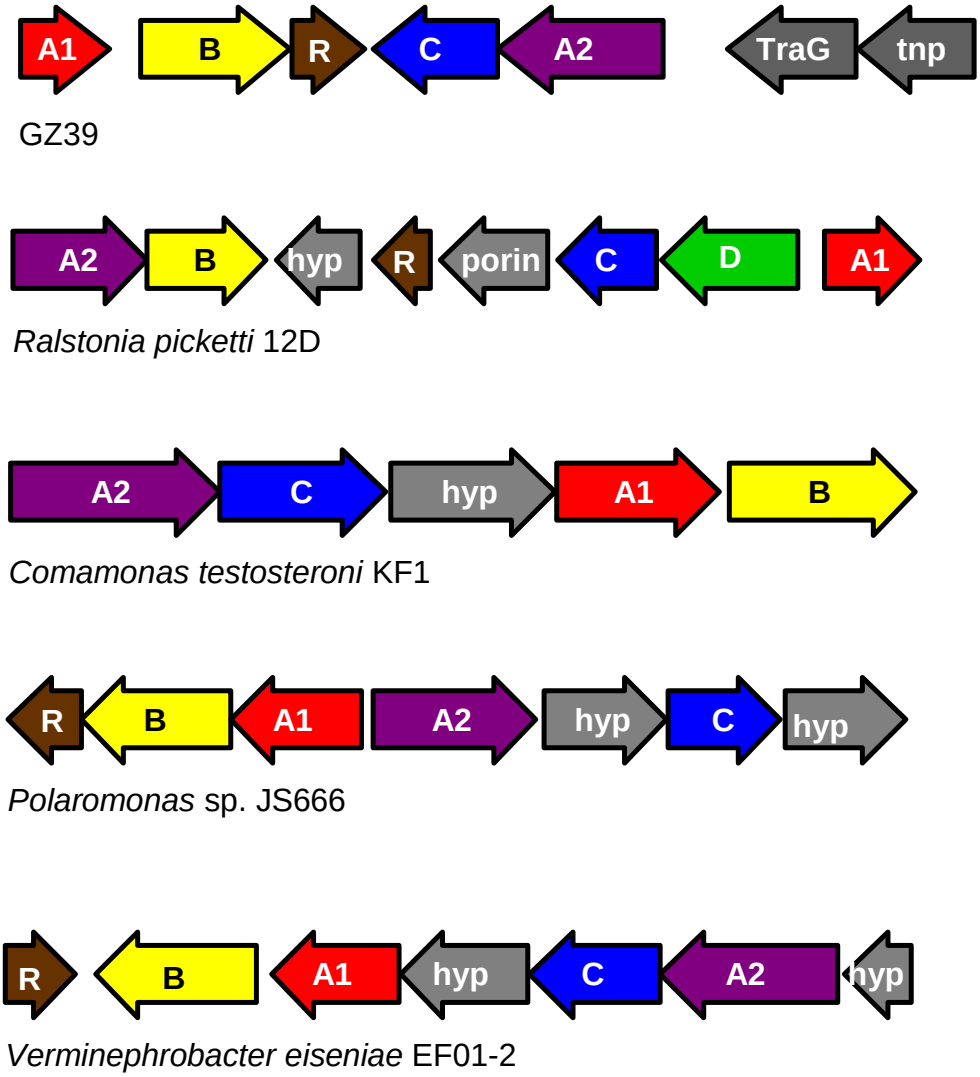


Figure 22. Gene organization the phthalate degradation operon of GZ39 and related operons

Chapter 3

Functional analysis of the phenanthrene and phthalate degradation genes in GZ38 and GZ39

1. Expression of the phenanthrene dioxygenase in GZ39

Analysis of the nucleotide sequence of the phenanthrene degradation genes of GZ38 and GZ39 has revealed that these genes are novel. Although a similar sequence was found in the GenBank database for *Alcaligenes faecalis* AFK2, there are no functional data to prove that AFK2's putative dioxygenase is in fact capable of degrading phenanthrene. To determine if the phenanthrene degradation genes are functional enzyme expression assays were performed using the GZ39 phenanthrene dioxygenase genes and to determine the substrate range of the dioxygenase enzyme. The dioxygenase contains four components that sandwich a dehydrogenase in the operon as follows: *phnAb-phnAa-phnB-phnAc-phnA* as described in Chapter 1. A clone was constructed from a NcoI-PvuI-NsiI fragment, which contained the components of the dioxygenase and that fragment was cloned into pTRC99A (Goyal 1996) and transformed into *E. coli* DH5α cells.

The clone was grown to an OD_{600} of 0.5, concentrated to an OD_{600} of 1.0, and a resting cell assay was performed using the following substrates: anthracene, biphenyl, chrysene, dibenzothiophene, naphthalene, phenanthrene, and pyrene. For comparison, a

dioxygenase clone from *Sphingobium yanoikuye* B1, a known PAH degrader, was used as a positive control for naphthalene and phenanthrene degradation. GC-MS analysis of extracted samples showed that the GZ39 phenanthrene dioxygenase clone was active against naphthalene and phenanthrene only (Table 5). Both substrates were completely consumed. For naphthalene, the substrate peak would be expected at a retention time of 8.24 minutes. The GC total ion chromatograph (TIC) shows a single peak with a retention time of 12.47 minutes with a mass of 144, which corresponds to naphthalene dihydrodiol (Figure 23). For phenanthrene, the substrate peak would be expected at 15.79 minutes but no peak was detected. The product peak occurred at 19.22 minutes and had a mass (194) that was consistent with phenanthrene dihydrodiol (Figure 24). This is further supported by the library match of the mass spectrum to phenanthrene dihydrodiol. All of these results matched those of the positive control. Based on these results, this dioxygenase enzyme is indeed capable of phenanthrene oxidation.

2. Accumulation of salicylate by GZ39

The oxidation of naphthalene by the expression clone indicated that although GZ39 does not grow on naphthalene, it may be able to metabolize it before reaching a stoppage point in the pathway. Kang et. al. (2003) demonstrated that *Burkholderia* sp. TNFYE-5, an isolate that was unable to grow on naphthalene, accumulated salicylate when grown on naphthalene (Kang, Hwang et al. 2003). This could occur if a missing or nonfunctional salicylate hydroxylase enzyme was present. A similar approach was taken to determine if the inability of GZ39 to grow on naphthalene could also be attributed to a

missing or nonfunctional salicylate hydroxylase. An initial PCR screening for the salicylate hydroxylase gene was done using PCR primers that were designed to the salicylate hydroxylase in *Pseudomonas putida* NCBI 9816. No PCR product was observed and therefore an experiment was designed to measure salicylate accumulation. Cells were grown in MSB with succinate and 0.5 mM naphthalene for 24 hours and samples were extracted and analyzed by HPLC with salicylate standards. A peak was observed for the salicylate standards at 27.28 minutes. A similar peak was observed for the GZ39 sample at 27.08 minutes. Based on these results, GZ39 accumulated salicylate when grown on naphthalene. It is therefore, possible that growth of GZ39 does not occur on naphthalene because of a nonfunctional or nonexistent salicylate hydroxylase enzyme.

3. Operonic structure of the phenanthrene degradation genes of GZ38

The phenanthrene degradation operon of GZ38 contains 18 open reading frames and is 20.8 kb long. In order to determine the operonic structure of the phenanthrene genes, RT- PCR was performed. Cells were grown in MSB liquid and phenanthrene. RNA was isolated after 20 hours and 20 primer sets were designed that encompassed the entire operon along with a 2 kb region upstream of the initial phenanthrene degradation gene that included a putative transposase. The first primer set was located inside the transposase and the rest of the primers overlapped one another until the end of the operon. With the exception of the third primer set which resulted in a 650 bp product, the rest of the primer sets amplified 1-1.5 kb products. Amplified products were visualized using agarose gel electrophoresis.

All primers within the operon produced a cDNA product of the correct size suggesting that there is a single continuous transcript for the entire operon (Figure 25). No product was observed for the transposase indicating that it is not part of the transcript. Negative controls that were performed on the RNA sample using Taq polymerase yielded no amplicons.

Additionally, RT-PCR was done using RNA from cells that were grown on succinate. Primers were used that were designed to amplify the large and small subunit of the phenanthrene dioxygenase gene sequence. An RT-PCR product of the correct size was observed (Figure 26).

4. Operonic structure of the phthalate degradation genes of GZ38

The operonic structure of the phthalate genes was also determined by RT-PCR for GZ38. RNA from the cells grown on phenanthrene was used because GZ38 does not grow on phthalate. Six primer sets were designed for the phthalate operon. The first primer set encompassed the reductase gene and the expected size of the amplicon was 635 bp. The second primer set included the beginning of the reductase gene, the intergenic region between the reductase and the partial transporter gene and the beginning of the partial transporter gene. The product was expected to be 862 bp in length. The next sequence targeted was the partial transporter sequence and the beginning of the decarboxylase gene. These primers were expected to produce a product that was 1132 bp long. The next set of primers was used to amplify a 951 bp region spanning the

decarboxylase and the regulator. Next, primers were designed to amplify a 538 bp region that consisted of part of the regulator, the intergenic region between the regulator and the dehydrogenase, and part of the dehydrogenase. The final sequence of interest was a 1051 bp region that included the dehydrogenase and the dioxygenase genes.

For this operon, three transcripts were observed in RNA isolated from cells grown on phenanthrene (Figure 27). The reductase, (*ophA1*), which is transcribed in the opposite direction from the adjacent transporter fragment, comprises one transcript. The second transcript contains the transporter fragment, the decarboxylase, and the regulator (*ophD-ophC-ophR*). The third transcript contains the dioxygenase and the dehydrogenase (*ophA2-ophB*), which are transcribed in the opposite direction from the genes in the second transcript. The intergenic regions between the reductase and the transporter fragment and between the regulator and the dehydrogenase gene did not amplify. PCR done without RT resulted in no PCR product.

5. RT-qPCR of the phenanthrene degradation genes of GZ38

Since RT-PCR products for the phenanthrene genes were observed with both phenanthrene and succinate as substrates, it was unclear whether or not these genes were induced by phenanthrene. RT-qPCR experiments were done in order to establish whether or not the phenanthrene degradation genes are in fact induced by phenanthrene. Standard curves were constructed as previously described (Figure 28). RNA was isolated from GZ38 cells grown in MSB with phenanthrene as previously described when cells reached

an OD_{600} of 0.1. For comparison, cells were also grown in MSB liquid with succinate. For both the phenanthrene and succinate grown cells, three samples were taken from three separate flasks for a total of nine samples for each substrate.

RNA was amplified with primer sets for the following housekeeping genes: 16S rRNA subunit and 23S rRNA subunit. Primers were also designed to amplify part of the large subunit of the phenanthrene dioxygenase (*phnAc*) and 1-hydroxy-2-naphthoate dioxygenase (*phnG*). The average Ct value of the 16S rRNA amplification for cells grown on succinate was 14.72. This is comparable to the average Ct value of 15.28 for cells grown on phenanthrene. The average Ct value of the 23S rRNA amplification for cells grown on succinate is 14.39. This is also comparable to the average Ct value for cells grown on phenanthrene which is 14.95. This establishes that the RNA extracted from all flasks is of the same quality. The Ct values for *phnAc* and *phnG* would be expected to be considerably higher for the phenanthrene-grown cells when compared to Ct values for the succinate-grown cells if the phenanthrene genes are inducible. For *phnAc*, the average Ct values were 24.26 for succinate-grown cells and 26.42 for phenanthrene-grown cells. The average Ct value for *phnG* for cells grown on succinate was 25.51 and for cells grown on phenanthrene was 26.41 (Table 6).

Ct values for both phenanthrene and succinate grown cells were nearly identical indicating that growth substrate does not appear to have an effect on gene expression. Based on these results, the phenanthrene degradation genes are not inducible by phenanthrene and are constitutively transcribed.

6. Operonic structure of the phenanthrene degradation genes of GZ39

The phenanthrene operon of GZ39 contains 19 open reading frames and is 21.5 kb long. RT-PCR was performed to determine how these genes are transcribed. Cells were grown in MSB liquid and phenanthrene for 20 hours. RNA was isolated and 21 primer sets were designed that began upstream of the transposase gene before the first gene, *phnAb,* and spanned the entire operon. Of the 21 primer sets, 18 were also used for RT-PCR on GZ38.

All primers produced a product of the correct size except for a primer set located upstream of the transposase, which did not amplify a product from cDNA (Figure 29). Negative controls resulted in no PCR product. These results demonstrate that the putative transposase and the sequence upstream of the *phnAb* gene are transcribed along with the entire phenanthrene operon. Additional RT-PCR results for RNA isolated from GZ39 showed that the phenanthrene dioxygenase gene was also transcribed in cells that were grown on succinate (Figure 30).

7. Operonic structure of the phthalate genes of GZ39

RT-PCR was performed using RNA isolated from cells grown on phenanthrene in order to determine the operonic structure of the phthalate genes of GZ39. Primers were designed to amplify the intergenic regions of the five orfs. The first primer set encompassed the end of the reductase gene and the beginning of the dehydrogenase. The expected size of the amplicon was 1009 bp. The second primer set included the end of

the dehydrogenase gene and the beginning of the regulator. The product was expected to be 941 bp in length. Two primer sets were designed to the end of the regulator and the beginning of the decarboxylase. These primers were expected to produce products that were 1100 bp and 524 bp long. The last set of primers was used to amplify a 959 bp region spanning the decarboxylase and the dioxygenase.

Based on the gene organization, it would be expected that there would be a minimum of two transcripts with the regulator and decarboxylase genes as points of stoppage since the orfs are in opposite directions. However, amplification occurred for all primer sets (Figure 31). PCR done without RT resulted in no PCR product. Based on this data, transcription is proceeding through one or both stop signals. It is assumed that this is not adversely affecting protein synthesis since cells are able to grow on phenanthrene.

8. Promoter activity for the phenanthrene degradation genes of GZ39

The RT-PCR results for GZ39 indicate that the putative transposase located upstream of the phenanthrene operon is transcribed along with the genes for phenanthrene degradation. Furthermore, these genes are constitutively transcribed as they were expressed equally on both phenanthrene and succinate. This led to additional questions about gene regulation. Specifically, does the transposase have a promoter that is rendering the genes to be continuously on or is there a promoter upstream of the first gene in the operon, *phnAb*, that is affecting gene regulation?

In order to answer these questions, a *lacZ* promoter assay was performed. Putative promoters upstream of the transposase and *phnAb* were inserted into the PKRZ1 vector which contains a *lacZ* gene but lacks a promoter. Since this vector does not have a promoter, transcription will only occur if there is an active promoter upstream of the gene. Promoter activity is measured by adding *o*-nitrophenol- β-D-galactoside (ONPG), which is converted by β-galactosidase to galactose and *o*-nitrophenol. When this reaction occurs, a yellow color is observed and *o*-nitrophenol can be quantified by measuring its absorption at 420 nm. These two constructs along with the empty vector as a negative control were introduced into GZ39 by triparental mating. Three replicate assays were performed for each construct with and without phenanthrene added to the growth media.

Of the three constructs, pKRZ, Ptnp:pKRZ, and PAb:pKRZ, activity was observed only for the latter (Figure 32). The vector-only construct and the construct containing the region upstream of the transposase did not turn a yellow color in either the phenanthrene or non-phenanthrene grown cultures. The average calculated Miller units for the vector-only construct was 9.4 with phenanthrene and 11.3 with succinate. The construct with the sequence upstream of the transposase also was inactive under both inducing and non-inducing conditions with average values of 11.0 with phenanthrene and 12.7 with succinate. In contrast, the construct containing the region upstream of the *phnAb* had high β-galactosidase activity, turning yellow and registering high absorbance values at 420 nm for both the phenanthrene and succinate cultures. The average calculated Miller units for this construct was 224.7 with phenanthrene and 223.9 without

phenanthrene, indicating that the promoter was active under both growth conditions (Table 7).

The promoter region upstream of the transposase was never on while the promoter region upstream of *phnAb* was active irregardless of whether or not phenanthrene was present. If this promoter is always on then these results further support the theory that the phenanthrene degradation genes are constitutively transcribed.

9. Mapping the transcription start site for the phenanthrene genes in GZ39

Based on the results of the LacZ assay, the transcription start site was presumed to be upstream of the *phnAb* gene. In order to map the transcription start site of the phenanthrene degradation genes, a 5' RACE assay was performed using RNA from GZ39. RNA was isolated from the GZ39 PAb:pKRZ1 construct in *E. coli* TOP10 cells and assayed using the 5'RACE kit (Invitrogen). The mRNA was converted to cDNA using a gene specific primer and reverse transcriptase. The product was tailed and amplified using a second gene specific primer and an anchor primer. Samples were then sequenced and compared to known sequence data for GZ39 in this region.

Since the DNA is tailed with a series of C's, the start of the tail is the transcription start site. Based on the chromatograms of two samples (Figure 33), the transcription start site is located 112 bases upstream of the *phnAb* gene start site. The DNA sequence upstream of the start site was examined and possible promoter sequences were determined to be TAACAT at the -10 region and TGGAAA at the -35 region. These

sequences are similar to the consensus sequence of the *E. coli* promoter sequences TATAAT (-10 region) and TTGACA (-35 region). DNA sequence from GZ38 in this region is 100% identical to that of GZ39. It is likely, therefore, that the transcription start site and promoter regions in GZ38 are identical to GZ39. This would account for the constitutive expression of the phenanthrene genes in GZ38.

10. Evidence for the involvement of phthalate in phenanthrene metabolism in GZ39

Sequence data for the phenanthrene and phthalate genes pointed towards metabolism of phenanthrene through phthalate. In order to confirm that this is in fact the case, a mutant strain of GZ39 was constructed in which a gene encoding kanamycin resistance was inserted into the phthalate dioxygenase gene thereby disrupting its function and blocking the phthalate and phenanthrene degradation pathways. A kanamycin resistance gene was inserted into the phthalate dioxygenase gene of one of the subclones from the GZ39 cosmid library (5E10 subclone 2) and introduced into GZ39 by triparental mating. Mutants were initially screened by PCR to verify that a double crossover had occurred. Mutants that had integrated the construct into the genome turned green. It is unclear why this occurred. A green mutant was then streaked onto an LB plate with kanamycin and sprayed with phenanthrene. It was able to clear the phenanthrene from the plate but did not grow. In order to confirm that this phenotypic change was due to the loss of phthalate dioxygenase activity, an additional experiment was performed. Without a functional phthalate dioxygenase, it would be expected that the mutant would accumulate phthalate. Cells were grown on LB with ampicillin to an

OD_{600} of 1.0, washed and concentrated to an OD_{600} of 2.0, and incubated with phenanthrene. Triplicate samples were collected at 0, 12, 24, and 36 hour time points, extracted, and analyzed using HPLC. An identical experiment was done using GZ39 wild type cells. A single peak was observed that was compared to phthalate standards and quantified based on the phthalate standard curve (Figure 34). In GZ39 mutant cells treated with phenanthrene a small quantity of phthalate accumulated over time. Wild type cells did not accumulate phthalate (Figure 35).

11. Discussion

GZ38 and GZ39 have unique phenanthrene degradation operons with respect to other known bacterial isolates with the exception of AFK2. Sequence analysis of the phenanthrene dioxygenases from GZ38 and GZ39 revealed that they differ considerably from other known dioxygenases. The expression of the dioxygenase from GZ39 provides proof that this enzyme is capable of degrading both naphthalene and phenanthrene. Previous studies have been done that have characterized PAH dioxygenases from other bacterial isolates. The *nah* dioxygenases from *Pseudomonas putida* OUS82 and *Pseudomonas aeruginosa* PaK1 also were active on both naphthalene and phenanthrene (Takizawa, Iida et al. 1999). The phenanthrene dioxygenase from the distantly related *Burkholderia* sp. RP007 has also been expressed. Since this *phn* operon is missing the ferredoxin and reductase components of the dioxygenase it is presumed that this clone was functional due to a non-specific ferredoxin and reductase from the *E. coli* DH5α cells. This clone was able to oxidize both naphthalene and phenanthrene (Laurie and

Lloyd-Jones 1999). On the other hand, the dioxygenase from the more distantly related *Sphingobium yanoikuyae* B1 has a wide substrate range as is a trademark of many Sphingomonads. This enzyme is active on multiple compounds including the PAHs anthracene, naphthalene, and phenanthrene (Kim and Zylstra 1999; Ní Chadhain, Moritz et al. 2007).

RT-PCR was performed in order to determine the phenanthrene operonic structure for both GZ38 and GZ39. The results indicate that the *phn* genes of GZ38 are transcribed as one continuous message beginning with *phnAb* and proceeding to the last gene in the operon, *phnE*. In GZ39, transcription begins with the transposase that is upstream of *phnAb*, the first gene in the pathway, and is continuous through the last gene in the operon, *phnE*. Although these transcripts are large in size, there are other examples of long transcriptional units stemming from sizable operons (Dal, Trautwein et al. 2005; Schubbe, Wurdemann et al. 2006).

RT-PCR also revealed that the *phn* genes were constitutively expressed as transcription of the phenanthrene genes occurred with either phenanthrene or succinate as substrates in both GZ38 and GZ39. This observation was confirmed by RT-qPCR experiments in which the Ct values for succinate and phenanthrene samples were not significantly different, indicating similar transcript levels under both growth conditions.

The unusual length of the transcript along with the inclusion of the upstream transposase in the GZ39 transcript led to the questions of where the phenanthrene promoter was located and if there were multiple promoters involved with this operon. The *lacZ* promoter assay was performed in order to determine the location and number of

promoters mediating transcription. The LacZ assay showed that there is a single active promoter upstream of the *phnAb* gene. However, this promoter was not induced by phenanthrene but was instead constitutively active.

Additionally, the 5'RACE assay pinpointed a precise transcription start site located upstream of *phnAb* along with possible -10 and -35 regions. One possible explanation for the apparent loss of transcriptional regulation is that GZ39 may have two transcripts; one that originates at the start of the transposase and a second that begins just upstream of the first gene in the phenanthrene operon. Another is that the loss of regulation is occurring at the promoter region. The σ factor from the RNA polymerase should bind to the -10 and -35 sequences to initiate transcription and should be released after transcription begins. In this case, there may be something that is preventing the σ factor from dissociating such as interference from the upstream transposase. It is unusual for catabolic genes to be constitutively expressed because of the large energy cost to the cell.

It is not known if the activity of the promoter upstream of *phn*Ab of GZ38 mimics that of GZ39 but the nucleotide sequence is 100% identical in this region for both isolates.

The phthalate operon of GZ39 was induced by phenanthrene but could not be divided into operonic units. Transcription proceeds through a region where two orfs oppose one another. Two basic types of transcription termination sites have been described in bacteria: factor independent and factor dependent. Factor-dependent transcription terminators attach to specific regions of the RNA polymerase enzyme.

Transcription proceeds until the enzyme encounters an RNA polymerase termination site along the DNA. This causes dissociation of the RNA and RNA polymerase from the DNA. Factor-dependent termination is not well understood and the specific sequences for termination sites are not well understood. This made it impossible to determine if the presence of two overlapping transcripts is due to a faulty factor dependent transcription termination system. The second type of transcription termination is factor independent. A model termination sequence consists of a GC rich inverted repeat followed by a short run of A's. The inverted repeat causes a hairpin loop to form, which in turn causes RNA polymerase to become less stable. The series of A's causes the RNA polymerase to dissociate from the DNA (Snyder 1997). In the phthalate operon of GZ39 there are 59 nucleotides between the regulator and the decarboxylase. Contained within this sequence, there is an inverted repeat consisting of the sequence GCCACTT that occurs in the 5' to 3' direction that is separated by five nucleotides from the same sequence that occurs in the opposing direction on the opposite strand. However, the string of A's that would cause termination of transcription is not present. In this case, it is possible that without the series of A's, transcription does not stop as would be expected. This apparent irregularity in transcription does not seem to affect gene function. If these genes were acquired in a piecemeal fashion, it is plausible that multiple transposition events could have resulted in the deletion of part of the transcription termination signal sequence

The phthalate operon of GZ38 was also induced by phenanthrene and was transcribed as three operonic units. A factor independent termination site could not be pinpointed for the reductase transcript. This could be due to a limitation in sequence data as only 128 nucleotides downstream of the reductase have been identified.

Transcription may not end in this region or may not be under factor independent control. The other point of stoppage is between the regulator and the dehydrogenase. In this region, there is an inverted repeat that is 17 nucleotides in length, GGAAAGAAAAAACCTCG. This sequence may be involved in transcription termination or it could be a relic of other transposition events as in both *Burkholderia cepacia* DBO1 and *Burkholderia cepacia* G4, isolates with similar operon organization, there are two open reading frames present that are not in GZ38.

There is evidence of metabolism of phenanthrene through phthalate in other bacterial isolates based on GC-MS analysis of metabolites (Guerin and Jones 1988; Doddamani and Ninnekar 2000; Keum, Seo et al. 2006; Seo, Keum et al. 2007). For GZ38 and GZ39, both phenanthrene and phthalate operons have been identified. The accumulation of phthalate by the GZ39 knockout mutant links these two pathways and provides strong evidence that phenanthrene metabolism proceeds through phthalate. The measurable amount of phthalate is in the micromolar range despite the initial concentration of 2 mM phenanthrene. Based on the results, phenanthrene metabolism seems to begin almost immediately and then considerably slows down at 12 hours. It is possible that phthalate is exerting some toxic effect on the cell, which is halting the metabolism of phenanthrene and therefore limiting the amount of phthalate produced.

Table 5. PAH biotransformation by *E. coli* DH5α cells expressing the phenanthrene dioxygenase from GZ39

Substrate	Product	Molecular Mass	Retention Time (min)	Relative amount (%)
Anthracene	ND[a]			
Chrysene	ND			
Dibenzothiophene	ND			
Naphthalene	1,2-dihydroxynaphthalene	144	12.47	100
Phenanthrene	1,2-dihydroxyphenanthrene	194	19.22	100
Pyrene	ND			

[a] Not Detected

Table 6. RT-qPCR results for GZ38. Cells grown on phenanthrene were compared to cells grown on succinate. Average Ct values were calculated for 16S rRNA and 23S rRNA housekeeping genes and *phnAc* and *phnG* genes target genes.

Sample	**Gene**	**Average Ct**
GZ38 succinate	16S rRNA	14.72
GZ38 phenanthrene		15.28
GZ38 succinate	23S rRNA	14.39
GZ38 phenanthrene		14.95
GZ38 succinate	*phnAc*	24.26
GZ38 phenanthrene		26.42
GZ38 succinate	*phnG*	25.51
GZ38 phenanthrene		26.41

Table 7. Calculated Miller Units for the LacZ assay performed on GZ39

Construct	**Replicate 1**	**Replicate 2**	**Replicate 3**
39Ab:pKRZ-phe	258.8	215.6	197.5
39Ab:pKRZ+phe	220.0	227.7	226.4
39tnp:pKRZ-phe	14.5	12.2	11.6
39tnp:pKRZ+phe	11.7	10.6	10.9
39:pKRZ-phe	15.0	9.1	10.0
39:pKRZ+phe	10.4	6.6	11.8

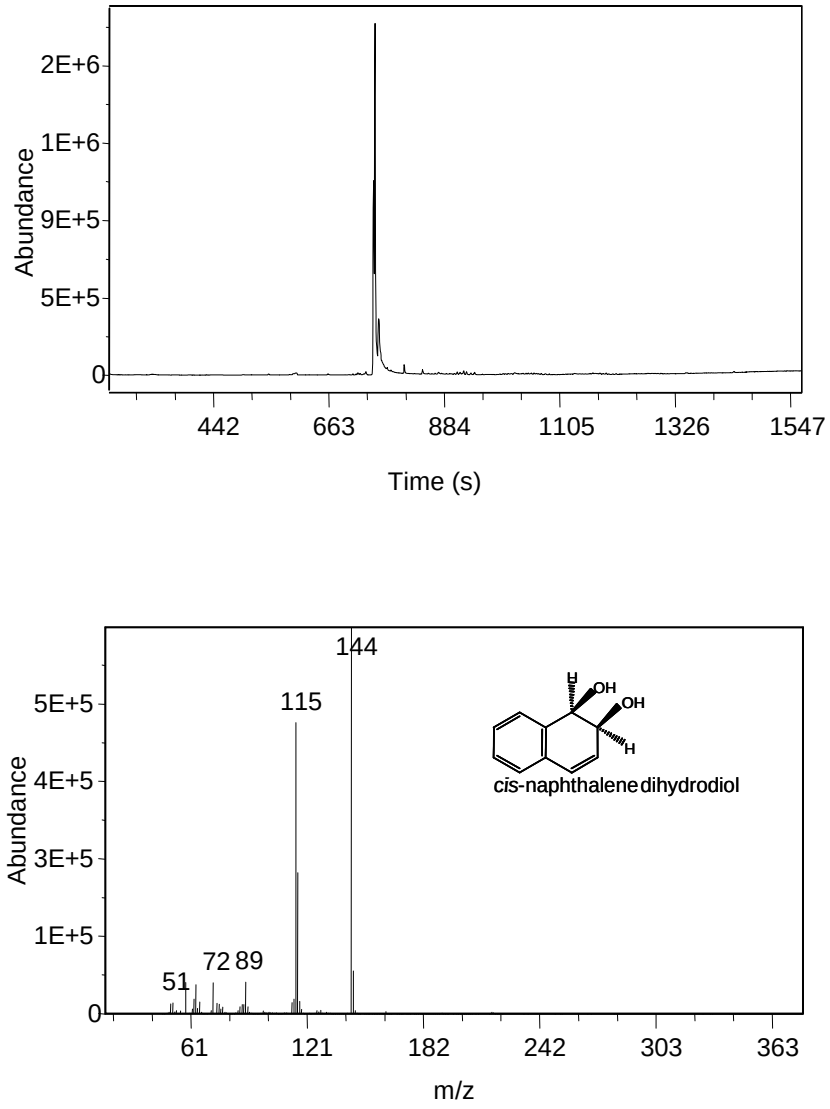


Figure 23. GC-MS trace of the products of naphthalene biotransformation by *E. coli* DH5α cells containing the phenanthrene dioxygenase from GZ39. Naphthalene was converted to a *cis*-dihydrodiol.

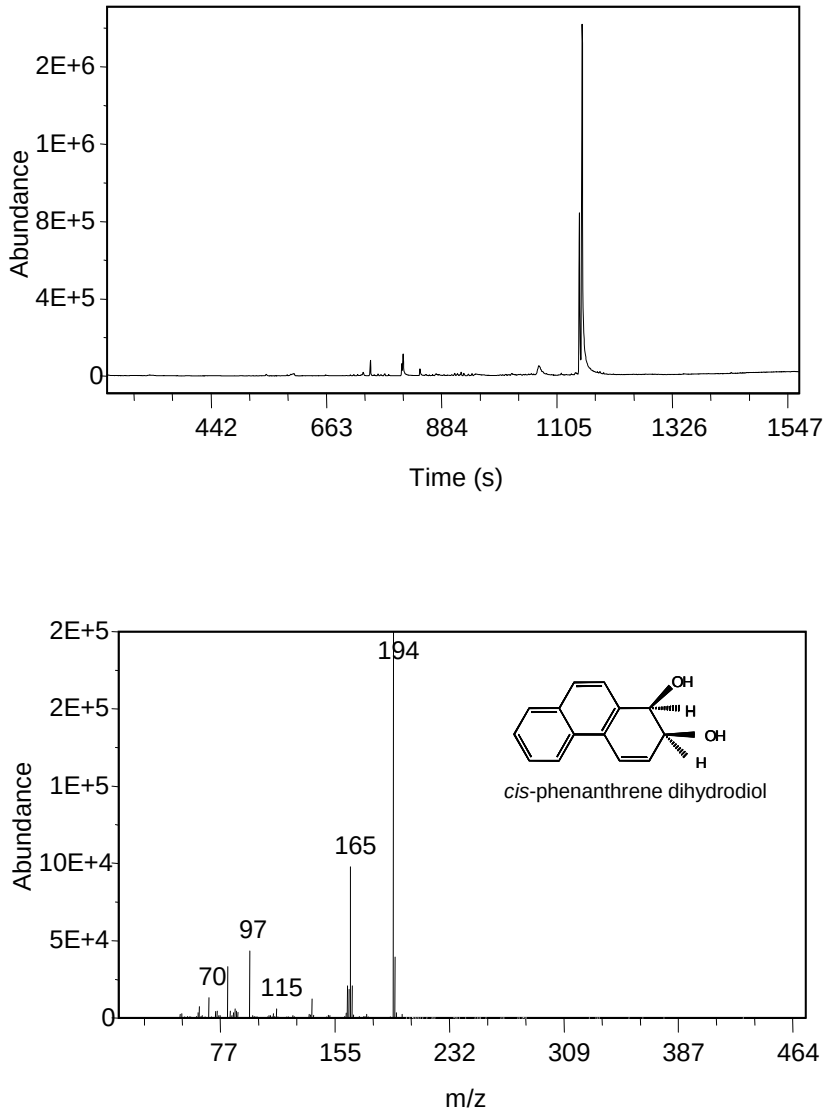


Figure 24. GC-MS trace of the product of phenanthrene biotransformation by *E. coli* DH5α cells containing the phenanthrene dioxygenase from GZ39. Phenanthrene was converted to a *cis*-dihydrodiol.

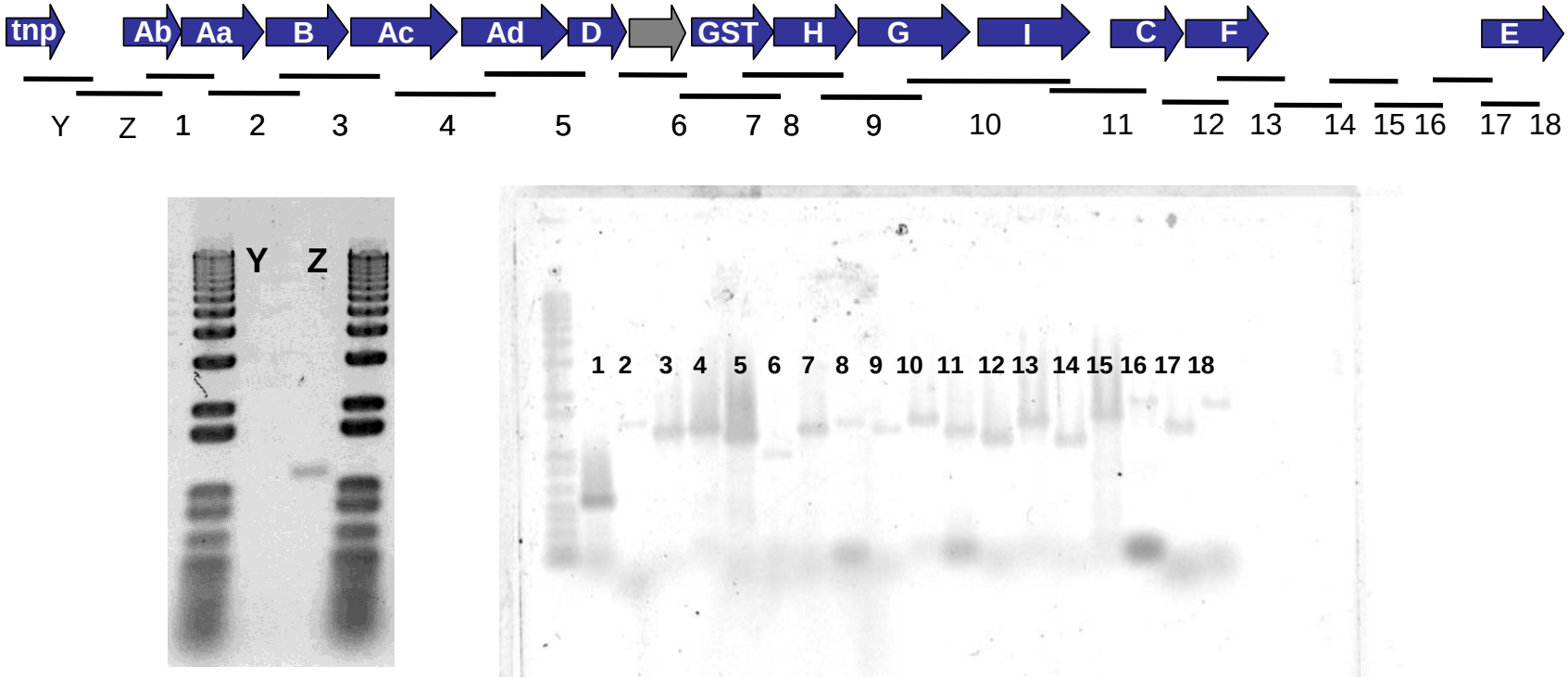


Figure 25. RT-PCR analysis of the GZ38 phenanthrene operon. PCR primers were designed to overlap adjacent genes and used to amplify cDNA from GZ38 cells that were grown on phenanthrene. The results indicate that a single message is transcribed from the region upstream of *phnAb* to *phnE*.

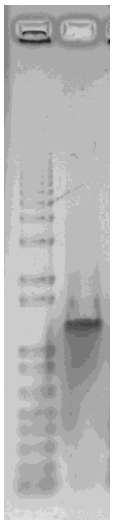


Figure 26. RT-PCR analysis of the phenanthrene dioxygenase from GZ38 cells that were grown on succinate. The results indicate that the dioxygenase gene is transcribed in the presence of succinate.

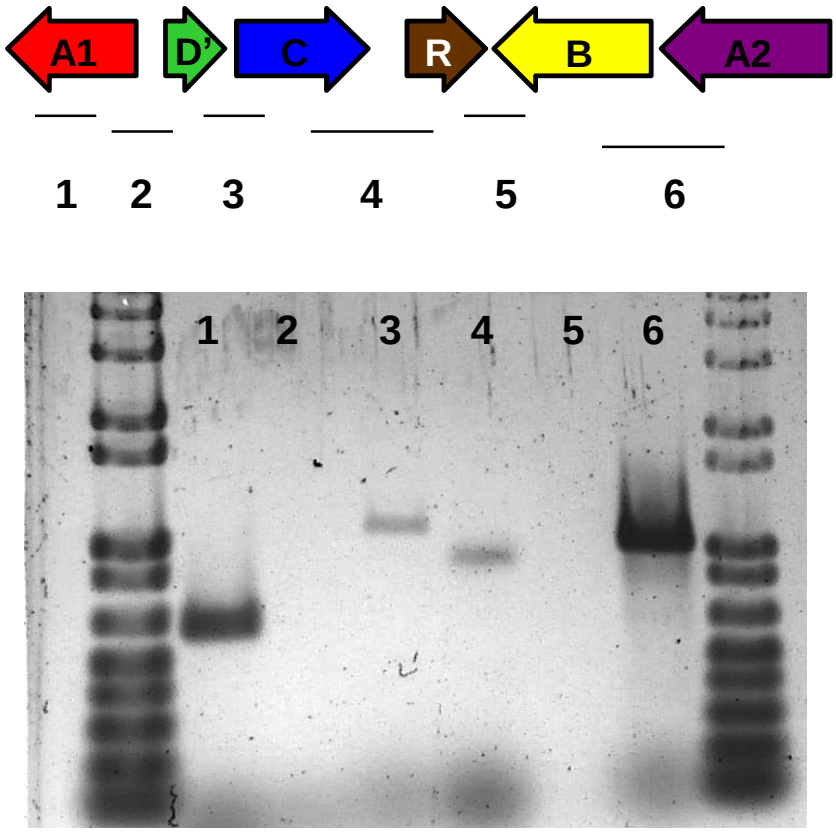
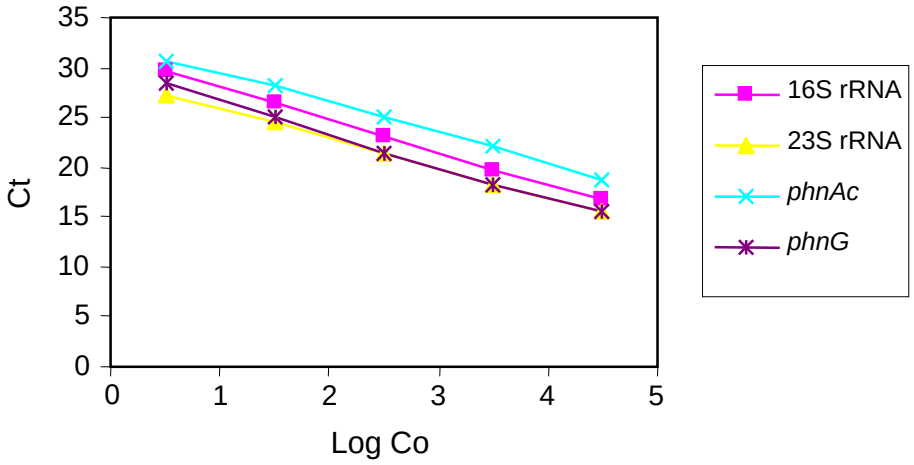


Figure 27. RT-PCR analysis of the GZ38 phthalate operon. PCR primers were designed to overlap adjacent genes and used to amplify cDNA from GZ38 cells that were grown on phenanthrene. The results indicate that three messages are transcribed.



Gene	Equation	r^2
16S rRNA	Y=-3.24 + 32.86	0.9996
23S rRNA	Y= -2.98 + 30.30	0.9992
phnAc	Y=-3.01 + 33.91	0.9946
phnG	Y=-3.27 + 31.95	0.9978

Figure 28. RT-qPCR standard curve. Standard curves were constructed by performing PCR on serially diluted samples of PCR products of full length genes from GZ38 as template DNA.

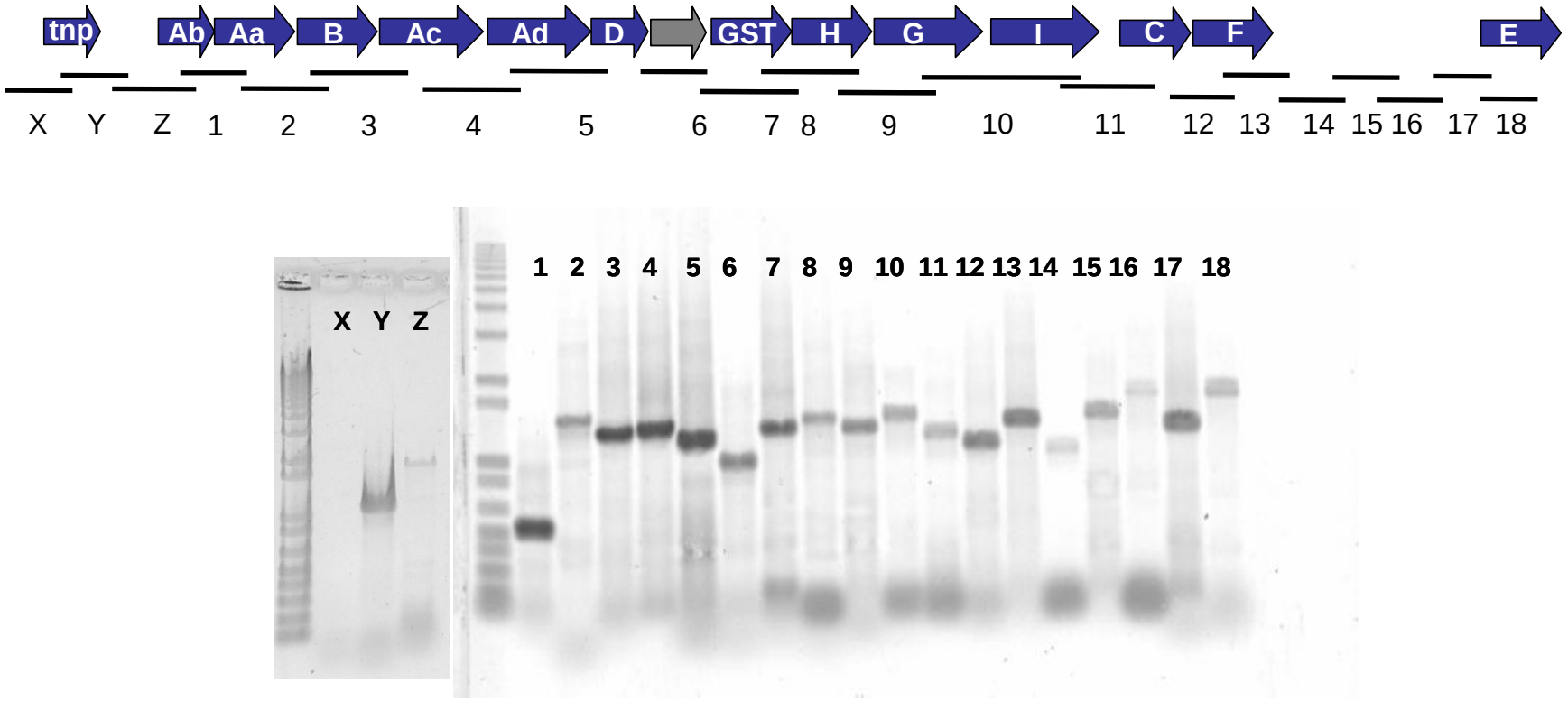


Figure 29. RT-PCR analysis of the GZ39 phenanthrene operon. PCR primers were designed to overlap adjacent genes and used to amplify cDNA from GZ39 cells that were grown on phenanthrene. The results indicate that a single message is transcribed from the transposase to *phnE*.

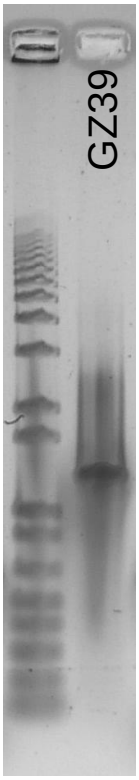


Figure 30. RT-PCR analysis of the phenanthrene dioxygenase from GZ39 cells that were grown on succinate. The results indicate that the dioxygenase gene is transcribed in the presence of succinate.

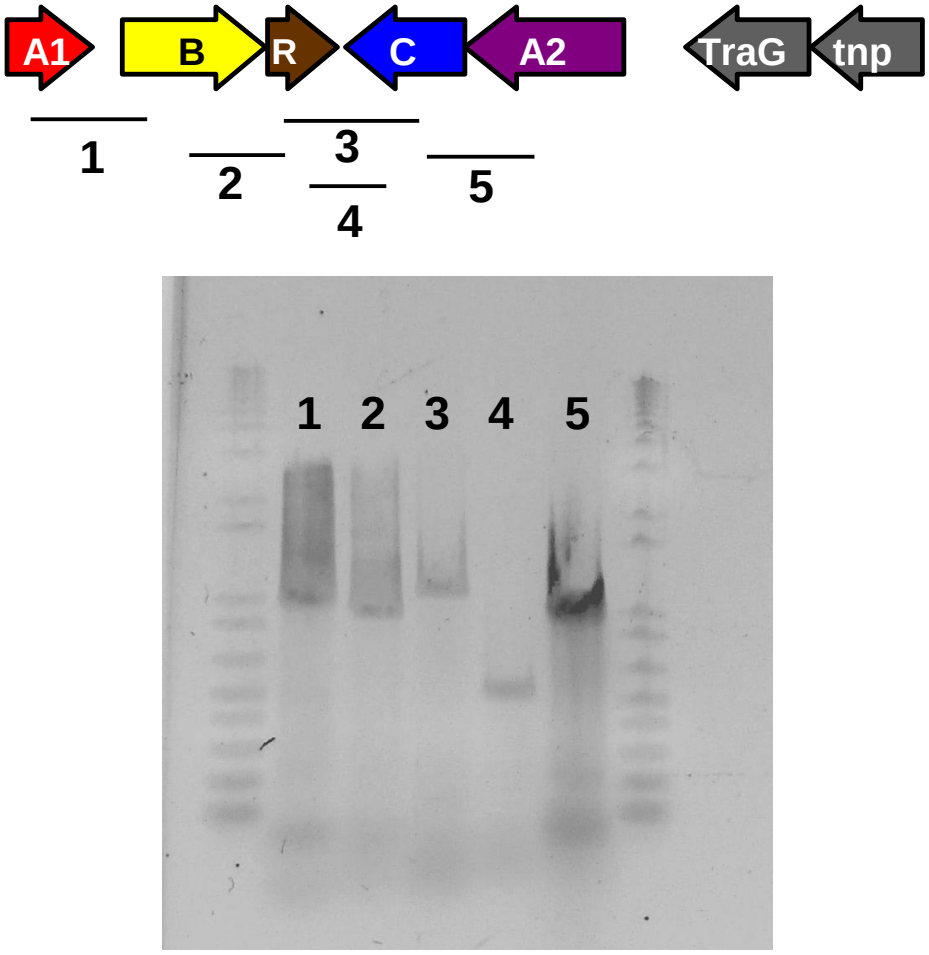


Figure 31. RT-PCR analysis of the GZ39 phthalate operon. PCR primers were designed to overlap adjacent genes and used to amplify cDNA from GZ39 cells that were grown on phenanthrene. The results indicate that there is no stoppage in transcription.

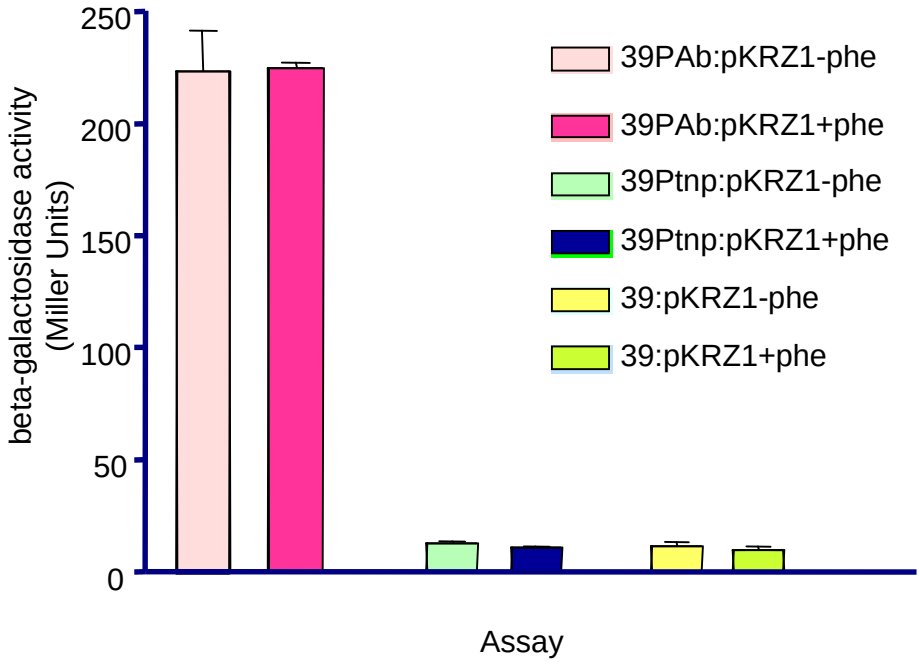


Figure 32. A LacZ assay was performed in order to measure promoter activity. Two possible promoter regions were targeted after analysis with BPROM software from Softberry.com. These sequences were amplified and cloned into pKRZ, a promoterless vector containing the *lacZ* gene. The constructs were introduced into GZ39 by triparental mating. Promoter activity was measured by adding *o*-nitrophenol-β -D-galactoside (ONPG) which is converted by β–galactosidase to galactose and *o*-nitrophenol. When this reaction occurs, a yellow color is observed and *o*-nitrophenol can be quantified by measuring its absorption at 420nm.

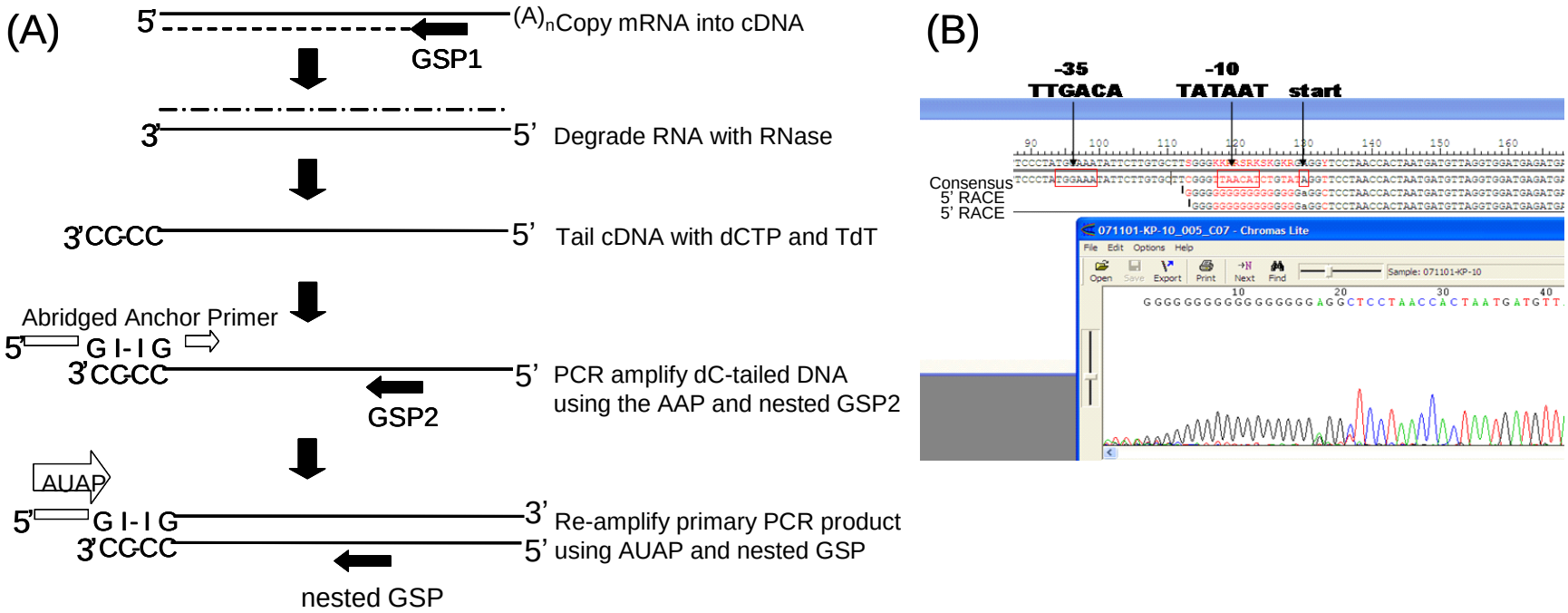


Figure 33. (A). A 5'RACE assay was performed in order to identify the transcription start site of the phenanthrene degradation genes. Reverse transcription of RNA and tailing of cDNA was followed by a series of PCR reactions.
(B). The final PCR product was sequenced and analyzed using Lasergene software.

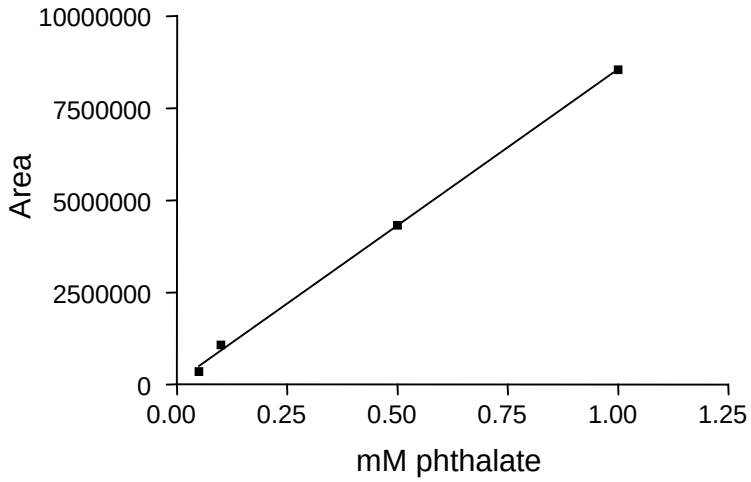


Figure 34. Phthalate standard curve as determined by HPLC analysis.

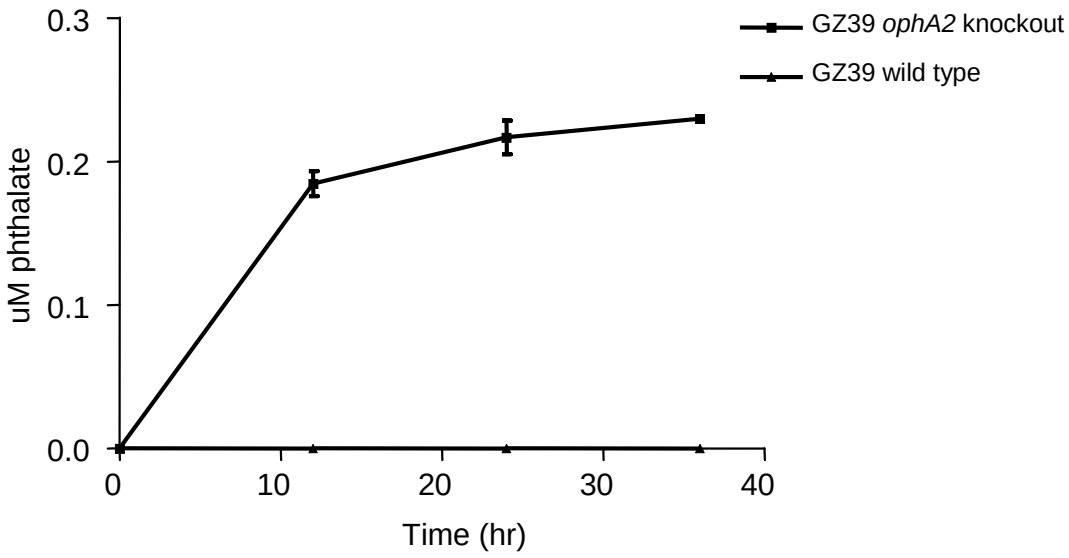


Figure 35. Accumulation of phthalate by the GZ39 phthalate dioxygenase knockout mutant measured using HPLC. Cells were incubated with phenanthrene and samples were taken at 0, 12, 24, and 36 hours and compared to phthalate standards.

Conclusions and Future Directions-

The genes responsible for phenanthrene degradation in GZ38 and GZ39 were identified and characterized. These genes, which comprise a 20.8 kb region in GZ38 and a 21.5 kb region in GZ39, were nearly identical to one another but considerably different in both sequence and organization from other known isolates with the exception of *Alcaligenes faecalis* AFK2, for which there is only sequence data. According to the putative functions for these genes, phenanthrene degradation appeared to proceed through phthalate. The genes responsible for phthalate degradation were also identified and characterized for both GZ38 and GZ39. The two phthalate operons are highly dissimilar when comparing sequence identity and gene organization. Expression in *E. coli*, RT-PCR experiments, a lacZ promoter assay, a 5'RACE assay, insertional inactivation of the phthalate dioxygenase gene, and extensive analysis of the phenanthrene and phthalate gene sequences revealed the following:

1. The phenanthrene degradation genes of GZ38 and GZ39, while 98-100% identical to AFK2, bear only 45-73% amino acid identity to other PAH degradation genes.

2. The phenanthrene dioxygenase from GZ39 was expressed in *E. coli* cells and was tested for activity against several substrates. The dioxygenase had a narrow substrate range displaying activity against naphthalene and phenanthrene only.

3. Although the dioxygenase is capable of the biotransformation of naphthalene, GZ39 is unable to grow on naphthalene. HPLC analysis revealed that cells incubated with naphthalene accumulated salicylate leading to the hypothesis that there is a missing or nonfunctional salicylate hydroxylase enzyme. A breakdown in the pathway could prevent growth of GZ39 on naphthalene.

4. RT-PCR of the phenanthrene genes of GZ38 and GZ39 indicate that they are both transcribed as one continuous transcript. Both isolates have a different transposase upstream of the initial gene in the operon, *phnAb*. For GZ38 the transposase is not part of the transcript while for GZ39 the transposase is included in the transcript. Expression of the genes occurred on both phenanthrene and succinate.

5. A lacZ promoter assay revealed that the promoter region upstream of the initial gene in the operon, *phnAb*, was always active in GZ39.

6. A 5'RACE assay mapped the transcription start site upstream of *phnAb* along with possible -10 and -35 regions in GZ39.

7. GZ38 and GZ39 are unable to grow on phthalate and are both missing functional transporter genes in the region sequenced. The phthalate genes may have been acquired by these isolates in order to degrade phenanthrene.

8. A mutant was created in GZ39 by inserting a kanamycin cassette into the phthalate dioxygenase gene. This resulted in the loss of growth on phenanthrene and accumulation of phthalate by cells grown on phenanthrene.

Polycyclic aromatic hydrocarbons are ubiquitous pollutants in the environment. Because of their stability, these compounds can be persistent for years. The remediation of PAH-contaminated areas is imperative because of the significant health risks that these compounds pose for wildlife, plants, and people. Restoring PAH-contaminated sites reduces the environmental burden as well as the risks associated with human exposure to toxic chemicals. The removal of these compounds is most efficiently accomplished by bacteria. This research characterized novel phenanthrene and phthalate degradation operons in GZ38 and GZ39.

GZ38 and GZ39 were identified as *Comamonas testosteroni* and *Acidovorax* sp. respectively. It is interesting that two different species would have nearly identical genes for phenanthrene degradation. The constitutive expression of these genes in both organisms is also unusual but this could be the result of selective pressure as the Passaic River is a historically polluted environment. Long term exposure to PAHs has most likely had an effect on the microbial community in the sediment. Specifically for GZ38 and GZ39, it is possible that the concentrations of PAHs have been sustained and at a high enough level that the response of GZ38 and GZ39 has gone beyond enzyme induction. Constitutive expression of the phenanthrene degradation genes may be advantageous to these organisms because it would allow for the continuous removal of PAHs and possibly serve as a protective mechanism.

Little is known about the genome of both of these organisms beyond the phenanthrene and phthalate operons. It would be interesting to determine if these organisms have catabolic genes for other xenobiotics. This is plausible because the Passaic River contains a mixture of pollutants. Further characterization of GZ38 and GZ39 may help to further determine their bioremediation potential and their importance to the microbial community. Cosmid libraries could be constructed in order to characterize the metabolic capabilities of GZ38 and GZ39. The libraries could be screened using PCR as well as function-based assays in order to identify novel genes and metabolic processes.

Of the PAHs tested, GZ38 could only grow on phenanthrene and GZ39 was able to grow on anthracene and phenanthrene. It would be interesting to determine if these

isolates interact with other microbes in the environment in ways (i.e. competition or co-metabolism) that would facilitate degradation of other PAHs besides anthracene and phenanthrene.

The characterization of the phenanthrene degradative genes of GZ38 and GZ39 adds to a growing catalog of PAH degradation genes from various species of soil bacteria that have been isolated from contaminated soil. With recent characterizations of bacterial isolates with PAH dioxygenase genes that have low homology to *nah*-like genes *Ralstonia* sp. U2 (Fuenmayor and Rodriguez Lemoine 1992), *Burkholderia* sp. RP007 (Laurie and Lloyd-Jones 1999), *Nocardioides* sp. KP7 (Saito, Iwabuchi et al. 2000), *Alcaligenes faecalis* AFK2 (Kiyohara, Nagao et al. 1982), GZ38, and GZ39 there is a need for alternative tools for detection and monitoring of the PAH-degrading bacterial community. These discoveries create opportunities to expand our knowledge of PAH degradation. In the past, studies have focused on the *nah* genes because they were present in many contaminated sites. Studies that have used molecular techniques including DNA hybridization and PCR to locate PAH catabolic genes have utilized DNA probes and primers designed to nucleotide sequence of the nah genes (Goyal and Zylstra 1997; Ahn, Sanseverino et al. 1999). Although the *nah* genes are present in many contaminated sites, their importance in the environment is unclear. This additional sequence data could be used to develop new probes for PAH degradation genes in contaminated environments. Studies have been done using degenerate primers to the large subunit of the PAH dioxygenase. One study screened 19 PAH degrading bacteria from contaminated sites and found that while 6 of these isolates had dioxygenase sequences that were similar to that of *Burkholderia* sp. RP007 and 4 others had dioxygenase sequences that were

homologous to *Ralstonia* sp. U2, PCR products could not be obtained from the remaining isolates. None of the isolates hybridized to *nah* gene sequence (Widada, Nojiri et al. 2002). Another study used degenerate primers to the Rieske center of the large subunit of the naphthalene, biphenyl, benzene, and toluene dioxygenases of the Pseudomonads to detect 10 different *Mycobacterium, Gordona, Sphingomonas, Pseudomonas, Rhodococcus* and *Xanthomonas* strains (Hamann, Hegemann et al. 1999). Using new sequence data to develop DNA probes could lead to the discovery of novel dioxygenase genes.

The discovery of new gene sequences will provide the basis for effective monitoring of PAH degradation by microbial communities. Studies have been done using TRFLP (Grant, Muckian et al. 2007) and DGGE profiles (Kanaly, Bartha et al. 2000) that show that there is both diversity and fluctuation in microbial communities that have been exposed to PAHs. Community profiles vary depending on the environment and the predominant PAHs that are present (Muckian, Grant et al. 2007). Identifying the key players in a contaminated environment is integral to developing effective bioremediation strategies.

Given that the *nah*-like genes of many Pseudomonads share greater than 90% sequence similarity and almost identical gene organization, it is likely that these genes have been transferred between these organisms. In many of these organisms, the genes are located on plasmids which can serve as the mode of gene transfer. For GZ38 and GZ39 it is unknown whether the phenanthrene degradation genes are located on a plasmid or within the chromosomal DNA. In *Alcaligenes faecalis* AFK2, the

phenanthrene degradation genes are located on a large plasmid (Kiyohara, Nagao et al. 1982). All three organisms are from the same order, Burkholderiales, and GZ38 and GZ39 are derived from the same family, Comamonadaceae. Transfer of genes between these and other related organisms may be possible because they are closely related. The remarkable similarities of the phenanthrene degradation genes in GZ38, GZ39, and AFK2 suggest that these genes may have traveled as a unit. This is further supported by the presence of transposons in all three isolates upstream of the operon, although the transposon sequence of GZ38 differs greatly from that of GZ39 and AFK2 which are nearly identical.

Phenanthrene degradation is known to occur by two pathways. In the first pathway, phenanthrene is metabolized to 1-hydroxy-2-napthoic acid and then to 1-hydroxy-2-naphthalene. Degradation proceeds to naphthalene and then to gentisate or salicylate. Genes that encode enzymes for this pathway have been identified in several isolates. In the second pathway, phenanthrene is metabolized to 1-hydroxy-2-napthoic acid and a ring cleavage dioxygenase converts it to 2-carboxybenzaldehyde. Further degradation proceeds through phthalate.

The degradation of phenanthrene through phthalate by gram negative soil bacteria has been established. The inability of a particular isolate to grow on naphthalene, the identification of 2-carboxybenzaldeyde as a product of the oxidation of 1-hydroxy-2-naphthoic acid, and the identification of other metabolites consistent with the conversion of phenanthrene to phthalate and protocatechuate established this pathway (Barnsley 1983). *Alcaligenes faecalis* AFK2 is also able to oxidize intermediates along the

phenanthrene to phthalate pathway (Kiyohara, Nagao et al. 1982). Other studies done on phenanthrene degrading isolates have found metabolites consistent with phenanthrene degradation through phthalate (Samanta, Chakraborti et al. 1999). There is also sequence data available for gram positive isolates that degrade phenanthrene through phthalate (Iwabuchi and Harayama 1998; Doddamani and Ninnekar 2000; Saito, Iwabuchi et al. 2000) . However, sequence data for genes along this pathway in gram negative bacteria is lacking. Phthalate degradation operons were identified in both GZ38 and GZ39. Although these isolates have virtually identical phenanthrene degradation genes, sequence analysis of the phthalate operons of both isolates revealed that they are highly dissimilar to one another. The operonic structure of the phthalate genes of GZ38 is similar to isolates *Burkholderia cepacia* DB01, *Burkholderia vietanamiensis* G4, and YZW-B. Each isolate has also transposons associated with the phthalate operon. The phthalate operon of GZ39 is unlike any known phthalate operon in both nucleotide sequence and organization. This implies that there may have been multiple transposition events. These significant differences in the phthalate operons of GZ38 and GZ39 do not appear to affect their ability to metabolize phenanthrene as the rate of growth on phenanthrene is the same for both isolates.

Although the phthalate genes of GZ38 and GZ39 are highly divergent from one another, both isolates are unable to grow on phthalate and appear to be unable to transport phthalate into the cell. GZ38 has a putative partial transporter gene and GZ39 does not have a transporter gene in the region that was sequenced. A non-functional phthalate transporter gene has been documented in *Burkholderia cepacia* DB01 (Chang and Zylstra 1999) due to a frameshift mutation. It is interesting then, that both isolates would have a

non-functioning phthalate transporter. Phthalate degradation genes have been identified in other isolates from the Passaic River, all of which are able to grow on phthalate and have functioning transporters (Wang, Zhou et al. 1995). After acquisition of the phthalate degradation genes of GZ38, additional modification of the transporter gene may have occurred as it is missing approximately 1kb of nucleotide sequence that is present in the middle of both *Burkholderia cepacia* DB01 and YZW-B. This deletion may serve to protect the cell from toxic concentrations of phthalate. It is also possible that this operon was transmitted between bacteria as a unit and that the loss of function occurred during horizontal gene transfer events. Gene transfer followed by adaptive mutations has been previously documented between bacterial isolates (Jeenes, Reineke et al. 1982; Reineke, Jeenes et al. 1982). There are several studies that document the transfer of catabolic plasmids in soil microcosms (Hill KE 1998; Dejonghe, Goris et al. 2000) and in a freshwater mesocosm (Fulthorpe and Wyndham 1991). Further research could be done on the phthalate transporter of GZ38. Since the phthalate degradation genes of GZ38 have a high degree of homology to *Burkhoderia cepacia* DB01, the phthalate transporter from DB01 could be cloned and inserted into GZ38 and phthalate uptake experiments could be performed.

Because the phthalate genes of GZ39 are more mosaic, a transporter gene may not have been recruited. If the purpose of the phthalate genes is to aid in the metabolism of phenanthrene, then a transporter would not be necessary.

The significance of the phenanthrene degradation genes of GZ38 and GZ39 in their natural environment has yet to be determined. Quantitative PCR experiments using

several different primer sets designed to different dioxygenase gene families including that of GZ38 and GZ39 would provide a more accurate snapshot of the PAH dioxygenase gene diversity in the Passaic River sediment. The ability to determine and compare copy numbers of specific dioxygenase gene families would pinpoint the key players in the microbial community in a particular environment and aid in the development of effective bioremediation strategies. For example, one study used Q-PCR to compare the nah genes to the phn genes in contaminated soils and found that there were higher copy numbers of the phenanthrene dioxygenase related to *Burkholderia* sp. RP007 than to the *nah* dioxygenase (Laurie and Lloyd-Jones 2000). A more complete understanding of the phenanthrene and phthalate degradation genes of GZ38 and GZ39 as well as of their respective roles in their natural environment could provide a basis for the development of future bioremediation strategies.

References

Ahn, Y., J. Sanseverino, et al. (1999). "Analyses of polycyclic aromatic hydrocarbon-degrading bacteria isolated from contaminated soils." Biodegradation **10**(2): 149-57.

Altschul, S. F., Gish, W., Miller, W., Myers, E.W. & Lipman, D.J. (1990). ""Basic local alignment search tool."" J. Mol. Biol. **215**: 403-410.

Annweiler, E., W. Michaelis, et al. (2002). "Identical ring cleavage products during anaerobic degradation of naphthalene, 2-methylnaphthalene, and tetralin indicate a new metabolic pathway." Appl Environ Microbiol **68**(2): 852-8.

Barnsley, E. A. (1983). "Phthalate pathway of phenanthrene metabolism: formation of 2'-carboxybenzalpyruvate." J Bacteriol **154**(1): 113-7.

Bezalel, L., Y. Hadar, et al. (1996). "Metabolism of phenanthrene by the white rot fungus *Pleurotus ostreatus*." Appl Environ Microbiol **62**(7): 2547-53.

Bos, R. P., J. L. G. Theuws, C.-M. Leijdekkers, and P. T. Henderson. (1984). "The presence of the mutagenic polycyclic aromatic hydrocarbons benzo[a]pyrene and benz[a]anthracene in creosote P1." Mutat. Res. **130**: 153-158.

Bozlaker, A., A. Muezzinoglu, et al. (2008). "Atmospheric concentrations, dry deposition and air-soil exchange of polycyclic aromatic hydrocarbons (PAHs) in an industrial region in Turkey." J Hazard Mater **153**(3): 1093-102.

Chang, B. V., L. C. Shiung, et al. (2002). "Anaerobic biodegradation of polycyclic aromatic hydrocarbon in soil." Chemosphere **48**(7): 717-24.

Chang, H. K. and G. J. Zylstra (1998). "Novel organization of the genes for phthalate degradation from *Burkholderia cepacia* DBO1." J Bacteriol **180**(24): 6529-37.

Chang, H. K. and G. J. Zylstra (1999). "Characterization of the phthalate permease *ophD* from *Burkholderia cepacia* ATCC 17616." J Bacteriol **181**(19): 6197-9.

Clements, W. H., J. T. Oris, and T. E. Wissing. (1994). " Accumulation and food chain transfer of fluoranthene and benzo[a]pyrene in *Chironomus riparius* and *Lepomis macrochirus*. ." Arch. Environ. Contam. Toxicol. **26**: 261-266.

Coates, J. D., J. Woodward, et al. (1997). "Anaerobic degradation of polycyclic aromatic hydrocarbons and alkanes in petroleum-contaminated marine harbor sediments." Appl Environ Microbiol **63**(9): 3589-93.

Connors, M. A. and E. A. Barnsley (1982). "Naphthalene plasmids in *pseudomonads*." J Bacteriol **149**(3): 1096-101.

Dal, S., G. Trautwein, et al. (2005). "Transcriptional organization of genes for protocatechuate and quinate degradation from *Acinetobacter* sp. strain ADP1." Appl Environ Microbiol **71**(2): 1025-34.

Davies, J. I. and W. C. Evans (1964). "Oxidative metabolism of naphthalene by soil *pseudomonads*. The ring-fission mechanism." Biochem J **91**(2): 251-61.

Davison, J. (1999). "Genetic exchange between bacteria in the environment." Plasmid **42**(2): 73-91.

De Luca-Abbott, S. B., B. S. Wong, et al. (2001). "Review of effects of water pollution on the breeding success of waterbirds, with particular reference to ardeids in Hong Kong." Ecotoxicology **10**(6): 327-49.

Dejonghe, W., J. Goris, et al. (2000). "Effect of dissemination of 2,4-dichlorophenoxyacetic acid (2,4-D) degradation plasmids on 2,4-D degradation and on bacterial community structure in two different soil horizons." Appl Environ Microbiol **66**(8): 3297-304.

Denome, S. A., D. C. Stanley, et al. (1993). "Metabolism of dibenzothiophene and naphthalene in Pseudomonas strains: complete DNA sequence of an upper naphthalene catabolic pathway." J Bacteriol **175**(21): 6890-901.

Doddamani, H. P. and H. Z. Ninnekar (2000). "Biodegradation of phenanthrene by a Bacillus species." Curr Microbiol **41**(1): 11-4.

Doddapaneni, H., V. Subramanian, et al. (2005). "Physiological regulation, xenobiotic induction, and heterologous expression of P450 monooxygenase gene pc-3 (CYP63A3), a new member of the CYP63 gene cluster in the white-rot fungus Phanerochaete chrysosporium." Curr Microbiol **50**(6): 292-8.

Eaton, R. W. (1994). "Organization and evolution of naphthalene catabolic pathways: sequence of the DNA encoding 2-hydroxychromene-2-carboxylate isomerase and trans-o-hydroxybenzylidenepyruvate hydratase-aldolase from the NAH7 plasmid." J Bacteriol **176**(24): 7757-62.

Eaton, R. W. and P. J. Chapman (1992). "Bacterial metabolism of naphthalene: construction and use of recombinant bacteria to study ring cleavage of 1,2-dihydroxynaphthalene and subsequent reactions." J Bacteriol **174**(23): 7542-54.

Ensley, B. D., B. J. Ratzkin, T. D. Osslund, M. J. Simon, L. P. Wackett, and and D. T. G. 1983. (1983). "Expression of naphthalene oxidation genes in *Escherichia coli* results in the biosynthesis of indigo. ." Science **222**: 167-169.

Ensley, B. D. and D. T. Gibson (1983). "Naphthalene dioxygenase: purification and properties of a terminal oxygenase component." J Bacteriol **155**(2): 505-11.

Fuenmayor, S. L. and V. Rodriguez Lemoine (1992). "Characterization of polycyclic aromatic hydrocarbons degradative soil *Pseudomonas*." Acta Cient Venez **43**(6): 349-54.

Fulthorpe, R. R. and R. C. Wyndham (1991). "Transfer and Expression of the Catabolic Plasmid pBRC60 in Wild Bacterial Recipients in a Freshwater Ecosystem." Appl Environ Microbiol **57**(5): 1546-1553.

Gibson, D. T. and R. E. Parales (2000). "Aromatic hydrocarbon dioxygenases in environmental biotechnology." Curr Opin Biotechnol **11**(3): 236-43.

Goyal, A. K. and G. J. Zylstra (1996). "Molecular cloning of novel genes for polycyclic aromatic hydrocarbon degradation from *Comamonas testosteron*i GZ39." Appl Environ Microbiol **62**(1): 230-6.

Goyal, A. K. and G. J. Zylstra (1997). "Genetics of naphthalene and phenanthrene degradation by *Comamonas testosteroni*." J Ind Microbiol Biotechnol **19**(5-6): 401-7.

Grant, R. J., L. M. Muckian, et al. (2007). "Microbial community changes during the bioremediation of creosote-contaminated soil." Lett Appl Microbiol **44**(3): 293-300.

Guerin, T. F. (1999). "Bioremediation of phenols and polycyclic aromatic hydrocarbons in creosote contaminated soil using ex-situ landtreatment." J Hazard Mater **65**(3): 305-15.

Guerin, W. F. and G. E. Jones (1988). "Mineralization of phenanthrene by a *Mycobacterium* sp." Appl Environ Microbiol **54**(4): 937-44.

Habe, H. and T. Omori (2003). "Genetics of polycyclic aromatic hydrocarbon metabolism in diverse aerobic bacteria." Biosci Biotechnol Biochem **67**(2): 225-43.

Haigler, B. E. and D. T. Gibson (1990). "Purification and properties of ferredoxinNAP, a component of naphthalene dioxygenase from *Pseudomona*s sp. strain NCIB 9816." J Bacteriol **172**(1): 465-8.

Haigler, B. E. and D. T. Gibson (1990). "Purification and properties of NADH-ferredoxinNAP reductase, a component of naphthalene dioxygenase from *Pseudomonas* sp. strain NCIB 9816." J Bacteriol **172**(1): 457-64.

Hamann, C., J. Hegemann, et al. (1999). "Detection of polycyclic aromatic hydrocarbon degradation genes in different soil bacteria by polymerase chain reaction and DNA hybridization." FEMS Microbiol Lett **173**(1): 255-63.

Harayama, S. (1997). "Polycyclic aromatic hydrocarbon bioremediation design." Curr Opin Biotechnol **8**(3): 268-73.

Hill KE, T. E. (1998). "Gene transfer in soil systems using microcosms." FEMS Microbiol Ecol **25**: 319-329.

Howsam, M., K. C. Jones, et al. (2001). "Dynamics of PAH deposition, cycling and storage in a mixed-deciduous (Quercus-Fraxinus) woodland ecosystem." Environ Pollut **113**(2): 163-76.

Hylland, K. (2006). "Polycyclic aromatic hydrocarbon (PAH) ecotoxicology in marine ecosystems." J Toxicol Environ Health A **69**(1-2): 109-23.

Iwabuchi, T. and S. Harayama (1998). "Biochemical and molecular characterization of 1-hydroxy-2-naphthoate dioxygenase from *Nocardioides* sp. KP7." J Biol Chem **273**(14): 8332-6.

Jeenes, D. J., W. Reineke, et al. (1982). "TOL plasmid pWW0 in constructed halobenzoate-degrading *Pseudomonas* strains: enzyme regulation and DNA structure." J Bacteriol **150**(1): 180-7.

Jeffrey, A. M., H. J. Yeh, et al. (1975). "Initial reactions in the oxidation of naphthalene by *Pseudomonas putida*." Biochemistry **14**(3): 575-84.

Jimenez, N., M. Vinas, et al. (2007). "The Prestige oil spill: bacterial community dynamics during a field biostimulation assay." Appl Microbiol Biotechnol.

Johnsen, A. R., L. Y. Wick, et al. (2005). "Principles of microbial PAH-degradation in soil." Environ Pollut **133**(1): 71-84.

Kanaly, R. A., R. Bartha, et al. (2000). "Rapid mineralization of benzo[a]pyrene by a microbial consortium growing on diesel fuel." Appl Environ Microbiol **66**(10): 4205-11.

Kanaly, R. A. and S. Harayama (2000). "Biodegradation of high-molecular-weight polycyclic aromatic hydrocarbons by bacteria." J Bacteriol **182**(8): 2059-67.

Kang, H., S. Y. Hwang, et al. (2003). "Degradation of phenanthrene and naphthalene by a *Burkholderia* species strain." Can J Microbiol **49**(2): 139-44.

Keith, L. H. a. T., W.A. (1979). "Priority Pollutants I. A perspective view." Environ. Sci. Technol. **13**: 416-423.

Keum, Y. S., J. S. Seo, et al. (2006). "Degradation pathways of phenanthrene by *Sinorhizobium* sp. C4." Appl Microbiol Biotechnol **71**(6): 935-41.

Kim, E. and G. J. Zylstra (1999). "Functional analysis of genes involved in biphenyl, naphthalene, phenanthrene, and m-xylene degradation by *Sphingomonas yanoikuyae* B1." J Ind Microbiol Biotechnol **23**(4-5): 294-302.

Kiyohara, H., and K. Nagao. (1978). "The catabolism of phenanthrene and naphthalene by bacteria." J. Gen. Microbiol. **105**: 69-75.

Kiyohara, H., K. Nagao, and R. Nomi. (1976). "Degradation of phenanthrene through o-phthalate by an *Aeromonas* sp." Agric. Biol. Chem. **40**: 1075-1082.

Kiyohara, H., K. Nagao, et al. (1982). "Phenanthrene-degrading phenotype of *Alcaligenes faecalis* AFK2." Appl Environ Microbiol **43**(2): 458-61.

Kiyohara, H., S. Torigoe, et al. (1994). "Cloning and characterization of a chromosomal gene cluster, *pah*, that encodes the upper pathway for phenanthrene and naphthalene utilization by *Pseudomonas putida* OUS82." J Bacteriol **176**(8): 2439-43.

Klaassen, C. D. (2001). Casarett and Doull's Toxicology: The Basic Science of Poisions. New York, McGraw-Hill.

Kriipsalu, M., M. Marques, et al. (2008). "Fate of polycyclic aromatic hydrocarbons during composting of oily sludge." Environ Technol **29**(1): 43-53.

Larkin, M. J., C. C. Allen, et al. (1999). "Purification and characterization of a novel naphthalene dioxygenase from *Rhodococcus* sp. strain NCIMB12038." J Bacteriol **181**(19): 6200-4.

Laurie, A. D. and G. Lloyd-Jones (1999). "The *phn* genes of *Burkholderia* sp. strain RP007 constitute a divergent gene cluster for polycyclic aromatic hydrocarbon catabolism." J Bacteriol **181**(2): 531-40.

Laurie, A. D. and G. Lloyd-Jones (2000). "Quantification of *phnAc* and *nahAc* in contaminated new zealand soils by competitive PCR." Appl Environ Microbiol **66**(5): 1814-7.

Liang, Y., M. F. Tse, et al. (2007). "Distribution patterns of polycyclic aromatic hydrocarbons (PAHs) in the sediments and fish at Mai Po Marshes Nature Reserve, Hong Kong." Water Res **41**(6): 1303-11.

Lindstrom, J. E., R. C. Prince, et al. (1991). "Microbial populations and hydrocarbon biodegradation potentials in fertilized shoreline sediments affected by the T/V Exxon Valdez oil spill." Appl Environ Microbiol **57**(9): 2514-22.

Liu, G., Z. Niu, et al. (2008). "Polycyclic aromatic hydrocarbons (PAHs) from coal combustion: emissions, analysis, and toxicology." Rev Environ Contam Toxicol **192**: 1-28.

Marchler-Bauer A, B. S. (2004). "CD-Search: protein domain annotations on the fly." Nucleic Acids Res. **32**: 327-331.

Marr, L. C., T. W. Kirchstetter, R. A. Harley, A. H. Miguel, S. V. Hering, and S. K. Hammond. (1999). "Characterization of polycyclic aromatic hydrocarbons in motor vehicle fuels and exhaust emissions." Environ. Sci. Technol. **33**: 3091-3099.

Meador, J. P., J. E. Stein, et al. (1995). "Bioaccumulation of polycyclic aromatic hydrocarbons by marine organisms." Rev Environ Contam Toxicol **143**: 79-165.

Miller, J. H. (1972). Experiments in Molecular Genetics. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

Moon, H. B., K. Kannan, et al. (2006). "Atmospheric deposition of polycyclic aromatic hydrocarbons in an urban and a suburban area of Korea from 2002 to 2004." Arch Environ Contam Toxicol **51**(4): 494-502.

Muckian, L., R. Grant, et al. (2007). "Bacterial community structure in soils contaminated by polycyclic aromatic hydrocarbons." Chemosphere **68**(8): 1535-41.

Mumtaz, M. a. G., J. (1995). Toxicological Profile for Polycyclic Aromatic Hydrocarbons. Atlanta, GA, U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES, Agency for Toxic Substances and Disease Registry.

Ní Chadhain, S. M., E. M. Moritz, et al. (2007). "Identification, cloning, and characterization of a multicomponent biphenyl dioxygenase from *Sphingobium yanoikuyae* B1." J Ind Microbiol Biotechnol **34**(9): 605-13.

Ní Chadhain, S. M., L. Tillotson, and G. J. Zylstra. (2007). Molecular characterization of functional gene diversity during biodegradation of phthalates and phthalate esters. American Society for Microbiology,107th General Meeting, Toronto, ON.

Ni Chadhain, S. M., R. S. Norman, et al. (2006). "Microbial dioxygenase gene population shifts during polycyclic aromatic hydrocarbon biodegradation." Appl Environ Microbiol **72**(6): 4078-87.

Patel, T. R. and D. T. Gibson (1974). "Purification and propeties of (plus)-cis-naphthalene dihydrodiol dehydrogenase of *Pseudomonas putida*." J Bacteriol **119**(3): 879-88.

Peng, R. H., A. S. Xiong, et al. (2008). "Microbial biodegradation of polyaromatic hydrocarbons." FEMS Microbiol Rev.

Pinyakong, O., H. Habe, et al. (2003). "The unique aromatic catabolic genes in *sphingomonads* degrading polycyclic aromatic hydrocarbons (PAHs)." J Gen Appl Microbiol **49**(1): 1-19.

Plata, D. L., C. M. Sharpless, et al. (2008). "Photochemical degradation of polycyclic aromatic hydrocarbons in oil films." Environ Sci Technol **42**(7): 2432-8.

Popp, W., C. Vahrenholz, et al. (1997). "DNA single strand breakage, DNA adducts, and sister chromatid exchange in lymphocytes and phenanthrene and pyrene metabolites in urine of coke oven workers." Occup Environ Med **54**(3): 176-83.

Reineke, W., D. J. Jeenes, et al. (1982). "TOL plasmid pWW0 in constructed halobenzoate-degrading *Pseudomonas* strains: prevention of meta pathway." J Bacteriol **150**(1): 195-201.

Rothmel, R. K., D. L. Shinabarger, et al. (1991). "Functional analysis of the *Pseudomonas putida* regulatory protein CatR: transcriptional studies and determination of the CatR DNA-binding site by hydroxyl-radical footprinting." J Bacteriol **173**(15): 4717-24.

Saito, A., T. Iwabuchi, et al. (2000). "A novel phenanthrene dioxygenase from *Nocardioides* sp. Strain KP7: expression in Escherichia coli." J Bacteriol **182**(8): 2134-41.

Samanta, S. K., A. K. Chakraborti, et al. (1999). "Degradation of phenanthrene by different bacteria: evidence for novel transformation sequences involving the formation of 1-naphthol." Appl Microbiol Biotechnol **53**(1): 98-107.

Samanta, S. K., O. V. Singh, et al. (2002). "Polycyclic aromatic hydrocarbons: environmental pollution and bioremediation." Trends Biotechnol **20**(6): 243-8.

Schell, M. A. (1983). "Cloning and expression in *Escherichia coli* of the naphthalene degradation genes from plasmid NAH7." J Bacteriol **153**(2): 822-9.

Schubbe, S., C. Wurdemann, et al. (2006). "Transcriptional organization and regulation of magnetosome operons in *Magnetospirillum gryphiswaldense*." Appl Environ Microbiol **72**(9): 5757-65.

Seo, J. S., Y. S. Keum, et al. (2007). "Degradation of phenanthrene by *Burkholderia* sp. C3: initial 1,2- and 3,4-dioxygenation and meta- and ortho-cleavage of naphthalene-1,2-diol." Biodegradation **18**(1): 123-31.

Shimada, T. (2006). "Xenobiotic-metabolizing enzymes involved in activation and detoxification of carcinogenic polycyclic aromatic hydrocarbons." Drug Metab Pharmacokinet **21**(4): 257-76.

Simon, M. J., T. D. Osslund, et al. (1993). "Sequences of genes encoding naphthalene dioxygenase in *Pseudomonas putida* strains G7 and NCIB 9816-4." Gene **127**(1): 31-7.

Snyder, L. a. C., W. (1997). Molecular Genetics of Bacteria, American Society for Microbiology.

Stanier, R. Y., Palleroni, N.J., and Doudoroff, M. (1966). "The aerobic pseudomonads: a taxonomic study." J. Gen. Microbiol. **43**: 159-271.

State of New Jersey, D. o. E. P. (2003). NJ Passaic River Directive. Trenton, NJ 59.

Stingley, R. L., A. A. Khan, et al. (2004). "Molecular characterization of a phenanthrene degradation pathway in *Mycobacterium vanbaalenii* PYR-1." Biochem Biophys Res Commun **322**(1): 133-46.

Suthersan, S. S. (1999). Remediation engineering : design concepts, Boca Raton: CRC Press LLC.

Takizawa, N., T. Iida, et al. (1999). "Nucleotide sequences and characterization of genes encoding naphthalene upper pathway of *Pseudomonas aeruginosa* PaK1 and *Pseudomonas putida* OUS82." J Biosci Bioeng **87**(6): 721-31.

Tao, S., Y. H. Cui, et al. (2004). "Polycyclic aromatic hydrocarbons (PAHs) in agricultural soil and vegetables from Tianjin." Sci Total Environ **320**(1): 11-24.

Top, E. M. and D. Springael (2003). "The role of mobile genetic elements in bacterial adaptation to xenobiotic organic compounds." Curr Opin Biotechnol **14**(3): 262-9.

Wang, Y. Z., Y. Zhou, et al. (1995). "Molecular analysis of isophthalate and terephthalate degradation by *Comamonas testosteroni* YZW-D." Environ Health Perspect **103 Suppl 5**: 9-12.

Widada, J., H. Nojiri, et al. (2002). "Molecular detection and diversity of polycyclic aromatic hydrocarbon-degrading bacteria isolated from geographically diverse sites." Appl Microbiol Biotechnol **58**(2): 202-9.

Yen, K. M. and C. M. Serdar (1988). "Genetics of naphthalene catabolism in pseudomonads." Crit Rev Microbiol **15**(3): 247-68.

Zhou, N. Y., S. L. Fuenmayor, et al. (2001). "*nag* genes of *Ralstonia* (formerly *Pseudomonas*) sp. strain U2 encoding enzymes for gentisate catabolism." J Bacteriol **183**(2): 700-8.

Curriculum vita

Karen V. Pesce

EDUCATION-

2008, Ph.D. Environmental Science, Rutgers University
1998, M.A. Education, Seton Hall University, *Summa cum laude*
1996, B.S. Biology, Seton Hall University, *Magna cum laude*
1994, B.A. English, Seton Hall University, *Magna cum laude*

PROFESSIONAL EXPERIENCE-

2001-present — *Graduate Student, Environmental Toxicology Program Rutgers University*

- Molecular Analysis of Phenanthrene Degradation *by Comamonas testosteroni* GZ38 and *Acidovorax* sp. GZ39
- Teaching Assistant for 11:126:483 Nucleotide Sequence Analysis and 11:126:486 Biotech Robotics
- Mentored Hassan Syed, a high school student during the summer of 2006

2006 — *Adjunct Faculty, Monmouth University Science Department*

- Taught BY101 Issues and Methods of Biology

2000-2001 — *Teacher of Science, Brick Memorial High School, Brick, NJ*

- Developed innovative lessons plans for chemistry and integrated science courses
- Monitored and assessed the progress of about 100 students

1999-2000 — *Teacher of Science, Jackson High School, Jackson, NJ*

- Adapted the geophysical science curriculum to include various activities to meet the needs of a block schedule
- Monitored and assessed the progress of about 140 students

1996-1999 — *Teacher of Science, Union Catholic High School, Scotch Plains,NJ*

- Designed and implemented the Regular and Remedial Chemistry curriculum as well as the Environmental Science curriculum including all labs
- Created two technology based modules that are used in the classroom

1998 — *Adjunct Faculty, Seton Hall University Math Department*

- Taught Developmental Algebra

1996 *Adjunct Faculty, Seton Hall University School of Business*
- Taught Business Statistics for the EOP Graduate Access Program

TECHNICAL PROFICIENCES-

- Cultivation and isolation of bacteria
- DNA and RNA isolation, amplification, sequencing, and cloning from bacterial cultures, soil, and sediment
- PCR, Inverse PCR, RT-PCR, QPCR
- DNA sequence analysis (Lasergene, MEGA)
- Mutagenesis
- Enzyme expression
- Southern Blotting
- Gas Chromatography-Mass Spectrometry, High Performance Liquid Chromatography

PUBLICATIONS and PRESENTATIONS-

Pesce, K.V., Ní Chadhain, S M., and Zylstra, G. J. Genetic Analysis of Phenanthrene Degradation in *Comamonas testosteroni* GZ39. 2nd Annual Symposium on Microbiology, Rutgers University, 2008.

Ní Chadhain, S. M., R. S. Norman, **K. V. Pesce**, J. J. Kukor, and G. J. Zylstra. 2006. Microbial dioxygenase gene population shifts during polycyclic aromatic hydrocarbon biodegradation. *Applied and Environmental Microbiology.* 72:4078-4087.

Ní Chadhain, S. M., **K. V. Pesce**, R. S. Norman, J. J. Kukor, and G. J. Zylstra. Shifts in Dioxygenase Gene Populations During Polycyclic Aromatic Hydrocarbon Biodegradation. American Society for Microbiology 105th General Meeting, Atlanta, GA, June 2005.

DISTINCTIONS-

- GANN Fellowship, Rutgers University
- Completed three marathons and placed in multiple triathlons
- Seton Hall University Athletic Scholarship
- SHU-1994 Big East All Conference Team - Softball
- SHU-1994 GTE District II Academic All American
- SHU-1993-1994 All Academic Team, Big East Conference