GENOMIC ANALYSIS OF POLYAROMATIC HYDROCARBON DEGRADING *PAENIBACILLUS* SPECIES

by

AHMED AL-SHITI

A thesis submitted to the

Graduate School-New Brunswick

Rutgers, the State University of New Jersey

In partial fulfillment of the requirements

for the degree of

Master of Science

Graduate Program in Microbial Biology

written under the direction of

Dr. Gerben J. Zylstra

and approved by

_____

_____

_____

New Brunswick, New Jersey

May 2017

ABSTRACT OF THE THESIS

GENOMIC ANALYSIS OF POLYAROMATIC HYDROCARBON DEGRADING *PAENIBACILLUS* SPECIES

By AHMED AL-SHITI

Thesis Director:

Dr. Gerben J. Zylstra

Polyaromatic hydrocarbons (PAHs) are a group of complex organic compounds that contain at least two aromatic rings. Some of the PAHs, which have high molecular weight, are known carcinogens. PAHs are one of the most prevalent pollutants in the environment and have two main sources: petroleum products and the incomplete combustion of wood. Several microorganisms are known for the ability to degrade PAHs, either aerobically or anaerobically, such as algae, bacteria, and fungi. The initial step of the aerobic degradation of PAHs is catalyzed by dioxygenase enzymes, which introduces oxygen atoms into the aromatic rings. The first aim of this study was to isolate and identify bacterial species that may have unique enzymes for PAH degradation. Two strains, *Paenibacillus validus* RD5 and RD6, were isolated from sediment samples obtained from the Logan River, Utah, USA, by aerobic enrichment using biphenyl as the sole carbon source. The second aim was to sequence the RD5 and RD6 genomes and examine transcription of several predicted dioxygenase genes in *P. validus* RD5 and *P. naphthalenovorans* PRN1 during growth on naphthalene, biphenyl, dibenzofuran, cinnamate, and benzoate using the reverse transcription polymerase chain reaction (RT-PCR) technique. Based on the RT-PCR results and comparison of the dioxygenase sequences to other enzyme sequences in GenBank hypotheses were formulated about the role of three of the enzymes in biphenyl, naphthalene, and dibenzofuran degradation in the two *Paenibacillus* strains.

ACKNOWLEDGEMENTS

I would like to thank my advisor Dr. Gerben J. Zylstra for his guidance, patience and support. None of this work would have been done without his help and motivation. Also, my sincere thanks to Dr. Hung-Kuang Chang and Dr. Dockyu Kim for their assistance and advice. My gratitude to my thesis committee members, Dr. Lee Kerkhof and Dr. Max Haggblom, for their advice.

I would like to thank Dr. Haggblom for providing us the strain *Paenibacillus naphthalenovorans* PRN1. Also, I want to thank Rachel Dean for isolating *P. validus* RD5 and RD6.

I would like to thank all the lab members, Rayan, Thamer, Suha, Aakansha, Rachel, and Igor for their help and support. Deep thanks for helping me solve technical problems and for all the advice and instruction that they gave me during my research.

There are no words can describe my extent thanks and gratitude to my best friend Rayan who was next to me all the time. He helped me to fulfill my dream to finish the masters degree in the United States. My profound thanks for everything you have done for me.

I would like to thank the Higher Committee for Education Development in Iraq (HCEDiraq) and Rutgers University for financial support.

Finally, I want to thank my parents and my brothers for their infinite sacrifice, encouragement and support.

Table of Contents

List of Tables

List of Figures

Chapter I

INTRODUCTION

Polyaromatic hydrocarbons (PAHs) are a group of complex organic compounds, containing only carbon and hydrogen, which are formed by the fusion of at least two aromatic rings (Harvey 1991). These compounds are some of the most prevalent pollutants in the environment and many of them are known carcinogens. Petroleum products and the incomplete combustion of wood are the main sources of PAHs. Also, these pollutants can be found in coal tar, oil shale, creosote, and fossil sources (Harvey 1991; Ravindra et al. 2008; Feng et al. 2009). PAHs may have formed at the beginning of the universe as PAHs have been found in comets and meteorites (Cook et al. 1996).

The simplest PAH forms are naphthalene, phenanthrene, and anthracene while the most common are those with five and six rings. PAHs are divided into two groups based on their size: the "small" PAHs, which contain up to six rings, and the "large" PAHs, which contain more than six rings (Feng et al. 2009). These compounds are biochemically stable, which can resist nucleophilic attack, and the reason for this is the π-electron clouds on the aromatic rings. Most PAHs are insoluble in water, where the solubility decreases with increasing molecular weight. These physical properties are one of the reasons for the difficulty of microorganisms to utilize PAHs as carbon sources, which promote the accumulation of these pollutants in the environment (Johnsen et al. 2005).

The traditional method used to reduce the PAHs hazards in the environment are limited to the physical removal or isolation of the pollutant. These methods are

expensive, time consuming, and not as effective as bioremediation. In contrast, bioremediation can reduce the hazards of these pollutants in a short time with less input of energy (Providenti et al. 1993; Ward et al. 2003). Algae, bacteria, and fungi are microorganisms known to degrade PAHs and use them as carbon and energy source. PAH degradation can be aerobic or anaerobic based on the microorganisms involved in the process. There are several factors that influence the rate of PAH degradation via microorganisms: (i) ecological factors such as pH, temperature, and oxygen; (ii) type and number of the microorganisms; and (iii) the chemical structure of the PAHs (Haritash & Kaushik 2009).

Well studied bacteria known for aerobic degradation of PAHs include pseudomonads, sphingomonads, actinobacteria, and firmicutes. These organisms use oxygen not only as an electron acceptor, but also as substrate for dioxygenase enzymes that introduce oxygen atoms into the aromatic rings in the first step of PAH degradation (Providenti et al. 1993; Haritash & Kaushik 2009; Lyu et al. 2014). Since dioxygenase enzymes are located inside the cell, PAHs must be taken up first by the microorganism before degradation can start. PAHs are then metabolized through dihydroxy compounds and ortho- or meta-cleavage pathways (Cerniglia 1984; Johnsen et al. 2005).

In 1993, the genus *Paenibacillus* was proposed based on phenotypic characteristics and 16s rRNA gene sequences data (Ash et al. 1991; Ash et al. 1993). The species belonging to the genus *Paenibacillus* are gram positive bacteria and are widely distributed in the environment and are found in water, soil, forage, rhizosphere, and insect larvae (McSpadden 2004; Lal & Tabacchioni 2009). *Paenibacillus* species have the ability to produce a special kind of endospore, the ellipsoidal endospore (Ash et al. 1993). There are distinctive characteristics that can

be used to distinguish *Paenibacillus* species: (i) high concentration of anteiso fatty acids (C15:0) in their cellular membrane; (ii) a DNA G+C ratio ranging between 45-54 mol%; and (iii) ability to degrade different types of biopolymer such as alginate, chondroitin, chitin, curdlan, and starch (Ash et al. 1993; Montes et al. 2004; Kanzawa et al. 1995; Nakamura 1987).

In this thesis, we analyzed the genomes of two different *Paenibacillus* species, *P. naphthalenovorans* and *P. validus*, focusing on the dioxygenases that initiate the degradation of different PAHs.

The first objective of this project was to identify the dioxygenase genes that have novel sequences which may code for unique enzymes. The second objective is to test the transcription of these genes under different growth conditions using reverse transcription polymerase chain reaction (RT-PCR) technique.

Chapter II

LITERATURE REVIEW

Characteristics and taxonomy of the genus *Paenibacillus*

In 1835, *Bacillus subtilis* was first described by Ehrenberg under the name of "Vibrio subtilis," which was then renamed by Cohn in 1872. The genus *Bacillus* is part of the Bacillaceae family which contains a large number of varied species. Members of the Bacillaceae family have the feature of producing different kinds of endospores (Claus 1986**)**. Bacterial species that belong to genus *Bacillus* are rod-shape gram-positive bacteria that are found in different environments such as freshwater, seawater, and hot springs. This genus contains aerobic and facultatively anaerobic as well as acidophilic and alkalophilic bacteria (Claus 1986; Nazina et al. 2001).

In 1991, Ash et al. classified the genus *Bacillus* into five phyletic lines, based on 16S rRNA gene sequencing data. After two years, he reclassified the *Paenibacillus*, which was located in the group 3 rRNA bacilli, using both phenotypic features and 16S rRNA gene sequences to establish a new genus (Ash et al. 1993; Ash et al.1991). There are currently more than 30 species under the genus *Paenibacillus* that have similar characteristics to the genus *Bacillus*. The name *"Paenibacillus"* reflects the fact of the similarity with the genus *Bacillus*, where paene in latin means almost (Ash et al. 1991, Lal & Tabacchioni 2009).

The diversity of the *Paenibacillus* species

There are several examples that show the diversity of *Paenibacillus* species. First, *Paenibacillus polymyxa*, which is known for its ability to produce antibacterial

compounds. *Paenibacillus polymyxa* also plays an important role in agriculture and horticulture. This species has the ability to fix nitrogen as well as producing growth regulating substances, hydrolytic enzymes, and 2,3-butandiol, which is used in the production of different chemical compounds. Second, *Paenibacillus koreensis,* which has the ability to produce antifungal compounds (Lal and Tabacchioni 2009; Chung et al. 2000). Third, the species that have the ability to cause human infections such as *Paenibacillus thiaminolyticus*, which was isolated from a patient with renal failure (Ouyang et al. 2008). Fourth, the species that have a significant role in the industrial field such as *P. amylolyticus*, which is known for producing an alpha amylase enzyme that has important role in biofuel production (Ikram-Ul-Haq et al. 2012). Fifth, the species that have the ability to degrade polyaromatic compounds such as *Paenibacillus naphthalenovorans* and *Paenibacillus validus,* which will be studied in the thesis.

Paenibacillus naphthalenovorans

The species name *"naphthalenovorans"* originally came from Persian which means "naphthalene-devouring" bacteria. *P. naphthalenovorans* was one of the *Paenibacillus* species isolated by Daane et al. (2002) from a petroleum contaminated area and salt marsh rhizosphere in Piles Creek, Linden, NJ, USA. Different tests were used in order to classify *P. naphthalenovorans* as a new species such as phenotypic and molecular tests, fatty acid methyl ester (FAME) profiles, and PAH-degradation profiles. In addition, naphthalene was used as the sole aerobic enrichment carbon source to isolate *P. naphthalenovorans* (Daane et al. 2001).

P. naphthalenovorans is a gram positive, strictly aerobic, rod-shaped bacterium that can move using peritrichous flagella. The approximate dimensions are

0.8 µm wide and 2.8-4.0 µm long and the mean of the GC content is 49 mol%. Like the other members of genus *Paenibacillus,* this species has the ability to form ellipsoidal spores in swelling sporangia and contains high concentration of (15:0) anteiso fatty acids. *P. naphthalenovorans* grows perfectly with temperatures that range between 30°C to 37°C and forms mucoid, translucent, white colonies on solid media (Daane et al. 2002).

Paenibacillus validus

Paenibacillus validus has a long taxonomic history. In 1935, *P. validus* was first described by Bredemann and Heigener as *Bacillus validus*. Later, the strains belonging to *B. validus* were reclassified by Smith et al. (1952) and Gordon et al. (1973) as *B. cirulans* strans (Heyndrickx et al. 1995). In 1983, Nakamura and Swezey usedDNA GC content of their *B. validus* strains in order to consider this species unlinked to *B. circulans*. They found that the mean GC content of *B. validus* (54 mol%) was higher than the GC content of *B. circulans*, which ranged from 37-39 mol%. Therefore, *B. validus* was returned to its original classification after further phenotypic and genotypic studies conducted by Nakamura (Nakamura & Swezey 1983, Nakamura 1984). Eventually, *B. validus* was moved with other species to the new genus *Paenibacillus* by Ash et al. in 1993 (Ash et al. 1993).

Unlike *P. naphthalenovorans, P. validus* forms non-mucoid, mottled colonies on solid media. Besides the colony morphology, PAH-degradation patterns, FAME profiles, and 16S rRNA sequences data were used in order to distinguish *P. validus* from other *Paenibacillus* species (Daane et al 2002).

Formation of PAHs

PAHs normally form under low oxygen conditions and via two main mechanisms, pyrosynthesis and pyrolysis. The first step involves formation of free radicals as a result of C-H and C-C bond breakage due to temperatures exceeding 500°C. In the second step, these radicals link together yielding acetylene, which then forms aromatic ring structures that can resist the high temperatures. Higher molecular weight PAHs can be formed by adding the hydrocarbon radicals to simple PAHs (Ravindra et al. 2008).

Sources of PAHs

Petroleum products and the incomplete combustion of wood are the main sources of PAHs. PAHs are also found in fossil sources, crude oil, and coal tar. In addition to that, PAHs have five emission sources; agricultural, industrial, mobile, domestic, and natural. The PAH emission factors are used in estimating air pollution levels which helps in setting environmental protection policies (Ravindra et al. 2008).

1. **Agricultural and Industrial sources**

The agricultural sources normally involve open burning of biomass which produces a significant amount of PAHs. There are several examples of agricultural sources such as open burning of stubble, straw, and sugarcane. Certain PAH emissions from sugarcane burning was studied by Godoi et al. (2004) and Fang et al. (1999). The results showed that the level of these emissions ranged from 7 to 46 ng/m^3 which is similar to the emissions level of some urban centers (Freeman & Cattell 1990).

Most of the industrial sources come from the manufacture of different products such as aluminum, coke, creosote, cement, asphalt, and rubber tires. Chen et

al. (2007) measured PAHs produced by the pyrolysis process of tire scrapping where they found that the mean was 77 μg/g with benzo(a)pyrene and dibenz[a,h]anthracene as the most dominate PAH. Beside these sources, Yang et al. (1998) demonstrated that industrial waste incinerators and heavy oil plants have a high emission factor of PAHs and benzo(a)pyrene (Chen et al. 2007; Ravindra et al. 2008).

2. Mobile sources

These sources involve emissions from aircrafts, vehicles, and ships. The amount of PAH emission vary based on the engine type, fuel type, and the quality of the fuel. Several studies indicate that in urban areas, diesel and gasoline vehicles were the major contributors of PAH emission. These emissions were found to be higher in diesel fueled vehicles (Marchand 2004; Marr et al. 2006; Ravindra et al. 2008). Also, PAH emission from gasoline powered vehicles was found to be affected by the air/fuel ratio. Jones et al. (2004) found that a fast reduction in high molecular weight PAHs were recorded as the air/fuel ratio increased.

Mobile sources of PAH emission can also be generated from rail transportation such as diesel, electric, and coal-fired trains. In addition to that, a significant amount of PAH emission comes from aircraft exhaust as the worldwide estimated jet fuel consumption is 227 billion liter every year. Furthermore, these are other PAH emissions that are hard to quantify such as scraping of rubber tires and brake linings. The most prevailing PAHs found in these emissions were coronene, benzo[ghi]perylene, and pyrene (Chen et al. 2006; Ravindra et al. 2008).

3. Domestic sources

Burning of garbage, oil, wood, coal, and gas is the most dominant source for the domestic emission of PAH. Other sources were also indicated, especially in developing countries, such as using crop waste and dried dung for cooking. In 2002, WHO reported that approximately 75% of the people in China and India use solid fuels for cooking (Mutangadura 2004). The PAH emission factor of burning dried dung was estimated between 3.1 to 5.5 mg/kg (Venkataraman et al. 2002). Two studies were conducted by Boström et al. (2002) and Schauer and Cass (2000) in order to measure the benzo[a]pyrene emission in Sweden and Los Angeles respectively. These studies indicated that the estimated emissions of benzo[a]pyrene from wood burning was higher than the emissions from both gasoline and diesel vehicles. In domestic heating, several factors can affect the PAH emissions rate such as fuel type and burning conditions, which include oxygen availability and moisture (Standley and Simoneit, 1987; Jenkins et al., 1996) where Jenkins et al. (1996) found that PAH emissions from flaming combustion are 4-5 times less than the smoldering combustion.

4. Natural sources of PAHs

The origins of natural sources can be terrestrial or cosmic. Terrestrial sources involve the burning of forests due to lightning strikes as well as volcanic eruptions. However, these sources were found to form a small fraction of PAH emission in comparison to other sources (Wild & Jones 1995; Ravindra et al. 2008). On the other hand, cosmic sources include carbonaceous chondrites, which arise in the asteroid belt. The production of PAHs in the chondrites can be either by high-temperature synthesis or thermal changes of aliphatic compounds. The thermal change requires the existence of magnetite and hydrated phyllosilicates which have low-temperature mineral stages (Halasinski et al. 2005; Ravindra et al. 2008).

PAHs and Human Carcinogenesis

A- Historical significance

The key observation, which led to thinking about PAHs as human carcinogens, was in 1775 by the British surgeon Sir Percival Pott who found that exposure to soot can cause scrotal cancer in chimney sweeps. The second observation was in 1875 by von Volkmann who noticed an elevation of skin cancer cases among coal tar workers (Boström et al. 2002). In 1915, PAHs were demonstrated as carcinogens, when several studies on ears of rabbits showed that exposure to PAH-containing compounds can cause tumors. These results were further proven a few years later when it was shown that the same compounds were tumorigenic in mice. (Phillips 1983). In 1929, dibenz[a,h]anthracene was the first pure carcinogenic PAH to be synthesized. A few years later, benzo[a]pyrene, which was proved to cause tumors in rodents, was isolated from tar (Dipple 1985; Phillips 1983). As a result of that, benzo[a]pyrene is used frequently as a model for tumorigenic PAHs. However, benzo[a]pyrene is not the only PAH that can cause tumors. In 1998, Club et al. conducted an experiment by mixing the feed of mice with coal tar and benzo[a]pyrene separately. After two years, they noticed that coal tar induced tumors in the forestomach, lung, and liver while benzo[a]pyrene induced tumors in forestomach, tongue, and esophagus. These results showed that beside benzo[a]pyrene, other genotoxic components were responsible for the liver and lung tumors (Club et al. 1998).

B- Mechanisms of carcinogenesis

The mutagenic PAHs normally have more than four rings in addition to a "bay region," a structural pocket which can facilitate PAH binding to metabolizing

enzymes. The metabolic alteration of PAHs inside an organism results in polar compounds such as diols, radicals, and quinines. These compounds can covalently bind to DNA and form bulky PAH-DNA adducts. PAH-DNA adducts can be stable, which leads to mutations, or unstable which causes depurination that also leads to mutations if unrepaired.DNA damage may also occur because reactive oxygen species are generated as a result of quinines formation. There are three metabolic pathways that describe the chemical carcinogenesis of PAH: the first pathway that involves dihydrodiol epoxide formation, which was dominant until the late 20th century. The second pathway that involves radical cation formation as a result of P450 peroxidase activity and was proposed in the mid 1990s. The third pathway involves *o*-quinone formation as a result of dihydrodiol dehydrogenase activity and was postulated in late 1990s (Ramesh et al. 2004; Xue &Warshawsky 2005). The enzymes that are involved in PAH metabolism to diol epoxides are cytochrome P450 1A1, 1A2, and 1B1. The genes that code for these enzymes are regulated by the aryl hydrocarbon receptor (AHR) that binds to PAHs. AHR works as a transcription factor that up-regulates the expression of these genes in the presence of PAHs, which ultimately increases enzyme production and thus facilitate PAH metabolism (Nebert et al. 2004).

Microbial degradation of PAHs

The biodegradation of PAHs is one of the major methods used in order to totally remove these pollutants from the environment or at least reduce their effects. Several species of bacteria, fungi, and algae are known for their ability to degrade PAHs such as *Paenibacillus naphthalenovorans, Cunninghamella echinulata, and Selenastrum capricornutum* respectively. These microorganisms use different enzymes in order to initiate the degradation process of PAHs such as dioxygenases, monooxygenases, and lignin peroxidases (Fig. 1). The biodegradation rate of the

PAHs depends on the molecular weight of PAHs as well as the microorganism type that is involved in the degradation. The degradation of PAHs with low molecular weight, two or three rings, is easier and faster than the degradation of PAHs with high molecule weight, more than three rings. The reason for that is mainly the high water solubility of PAHs with low molecular weight, which increases the accessibility of these compounds to the degrading microorganisms. There are other factors that affect PAH biodegradation in soil such as temperature, moisture, soil type, pH, oxygen availability, and PAH concentration (Cerniglia1993; Haritash & Kaushik 2009).

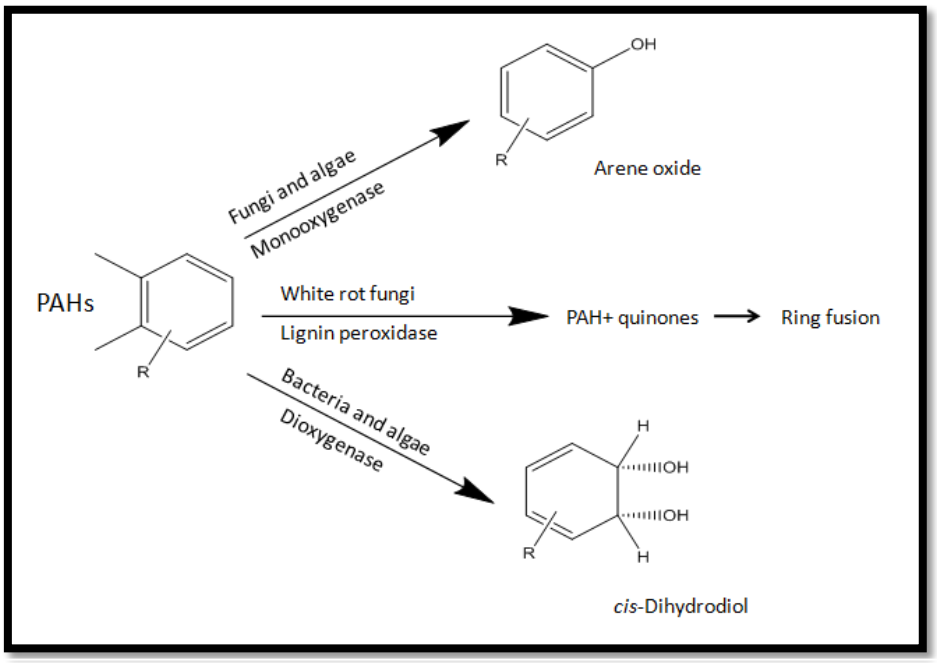


Figure 1. Initial steps of PAHs oxidation via different microorganisms (Adapted from Cerniglia 1993)

There are several research areas in the PAH biodegradation field such as the isolation and detection of microorganisms via innovative methods, genetic studies of PAH-degrading microorganisms, and new techniques for enhancing PAH biodegradation (Cerniglia, 1993). In this thesis, we studied the genomes of two bacterial species that have the ability to degrade three different PAHs with low molecular weight: naphthalene, dibenzofuran, and biphenyl.

1. The bacterial degradation of naphthalene

Naphthalene is the simplest PAH, with two rings, and is often used as a model compound to test the bacterial ability to degrade PAHs as well as understand the degradation pathway of complex PAHs. Several different genera of bacteria are known for their ability to degrade naphthalene and use it as a carbon and energy source.

Naphthalene degradation is catalyzed by a sequence of different enzymes: (i) naphthalene dioxygenase which initiate the degradation process by introducing two oxygen atoms to the aromatic ring forming *cis*-1,2-dihydroxy-1,2-dihydronaphthalene, (ii) *cis*-dihydrodiol dehydrogenase that dehydrogenates the products of the naphthalene dioxygenase to 1,2-dihydroxynaphthalene which is then metabolized to salicylate through several steps (Fig. 2), (iii) a hydroxylase enzyme that converts salicylate to catechol (common pathway), which is further degraded via *meta*- or *ortho*-pathways, (iv) salicylate-5-hydroxylase that converts salicylate to gentisate, as described by Fuenmayor et al. (1998) (Seo et al. 2009).

2. The bacterial degradation of dibenzofuran

The degradation of dibenzofuran, an O heterocyclic aromatic compound, is normally used as model for the biodegradation of biaryl ethers and chlorinated dibenzofurans. There are many bacterial strains that have the ability to degrade dibenzofuran such as *Sphingomonas* sp. strain RW1 and *Terrabacter* sp. strain DBF63. Also, these strains have the potential ability to degrade polychlorinated dibenzo-*p*-dioxins (PCDDs) and polychlorinated dibenzofurans, which are highly toxic compounds that result from the excessive use of pesticides and herbicides (Wittich et al. 1991; Kasuga et al. 2001; Xu et al. 2006).

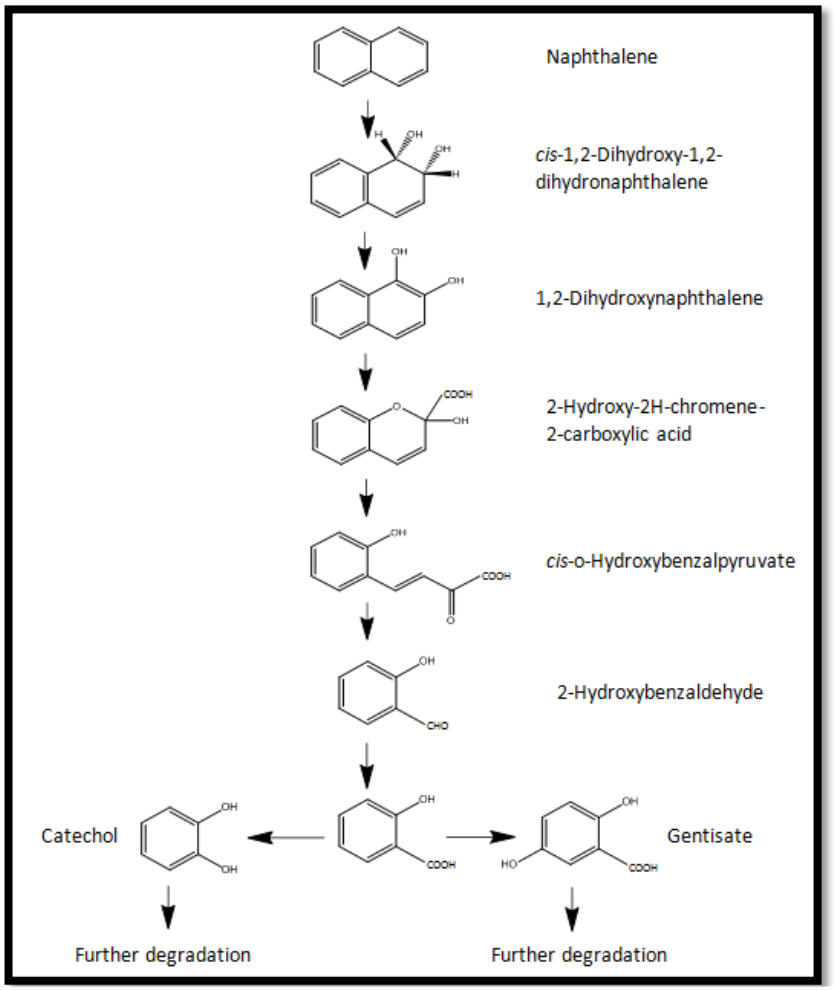


Figure 2. The proposed pathway of naphthalene degradation via bacteria (Adapted from Seo et al. 2009)

The first step in dibenzofuran degradation involves an angular attack by dibenzofuran 4,4a-dioxygenase, which forms an unstable intermediate that then spontaneously forms 2,2',3-trihydroxybiphenyl (Fig.3). The second step, catalyzed by an extradiol dioxygenase, involves ring cleavage of 2,2',3-trihydroxy-biphenyl forming 2-hydroxy-6-(2-hydroxyphenyl)-6-oxo-2,4-hexadienoic acid, which is then cleaved to salicylic acid and 2-hydroxypenta-2,4-dienoic acid by a hydrolase. Ultimately, salicylic acid can be further degraded via either the catechol pathway or the gentisic acid pathway (Fig. 3)(Wittich et al. 1991; Kasuga et al. 2001).

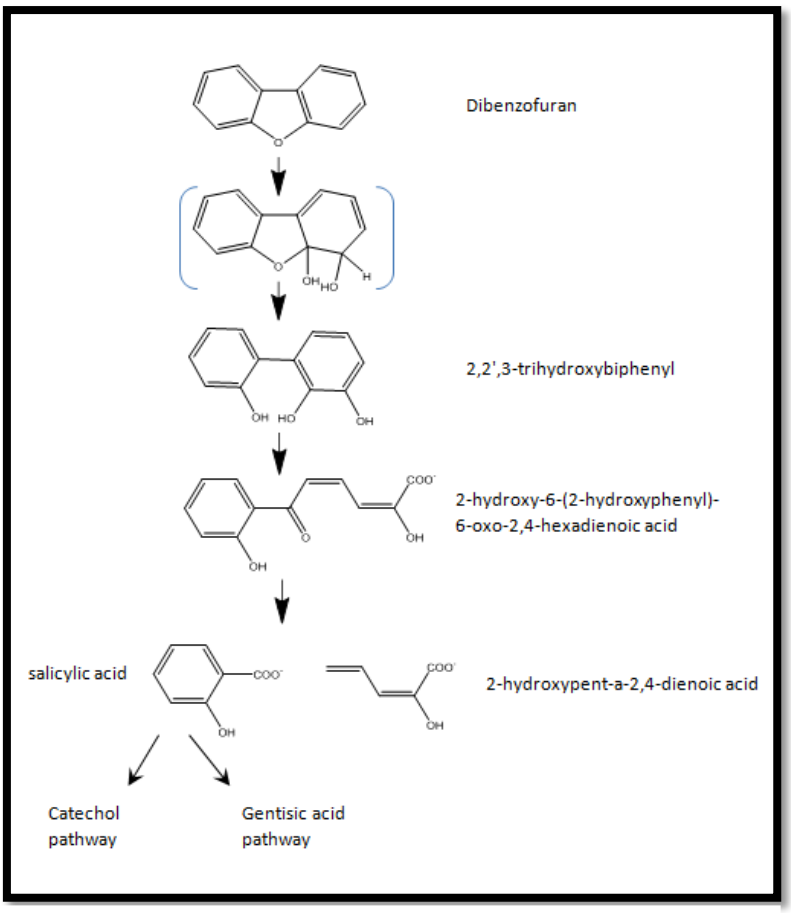


Figure 3. The degradation pathway of dibenzofuran (Adapted from Kasuga et al. 2001).

3. The bacterial degradation of biphenyl

Crude oil and coal tar are natural sources of biphenyl in the environment (Adams & Richardson 1953). In the late 1930s, biphenyl was used for many years as a fungicide in order to control the postharvest decay of citrus fruit (Eckert 1990). Biphenyl forms the nucleus of polychlorinated biphenyls (PCBs) which are widely distributed pollutants that can cause health effects. The microbial degradation of PCBs is one of the major methods used to remove these pollutants from the environment. There are many bacterial strains that are identified as PCB degraders such as *Alcaligenes eutrophus* H850 and *Burkholderia xenovorans* LB400 (Bedard et al. 1987; Fortin et al. 2006).

The degradation pathway of biphenyl starts by a dioxygenase attack, which is similar to the strategy used in the aerobic degradation of polyaromatic compounds, yielding a *cis*-2,3-dihydrodiol. The next step involves the conversion of the dihydrodiol to 2,3-dihydroxybiphenyl which is then cleaved via an extradiol dioxygenase yielding 2-hydroxy-6-oxo-6-phenyl-2,4-dienoate (HOPDA). Ultimately, a hydrolase cleaves HOPDA to benzoate and 2-hydroxypenta-2,4-dienoate which are then further degraded in the lower pathway of biphenyl (Fig.4) (Fortin et al. 2006).

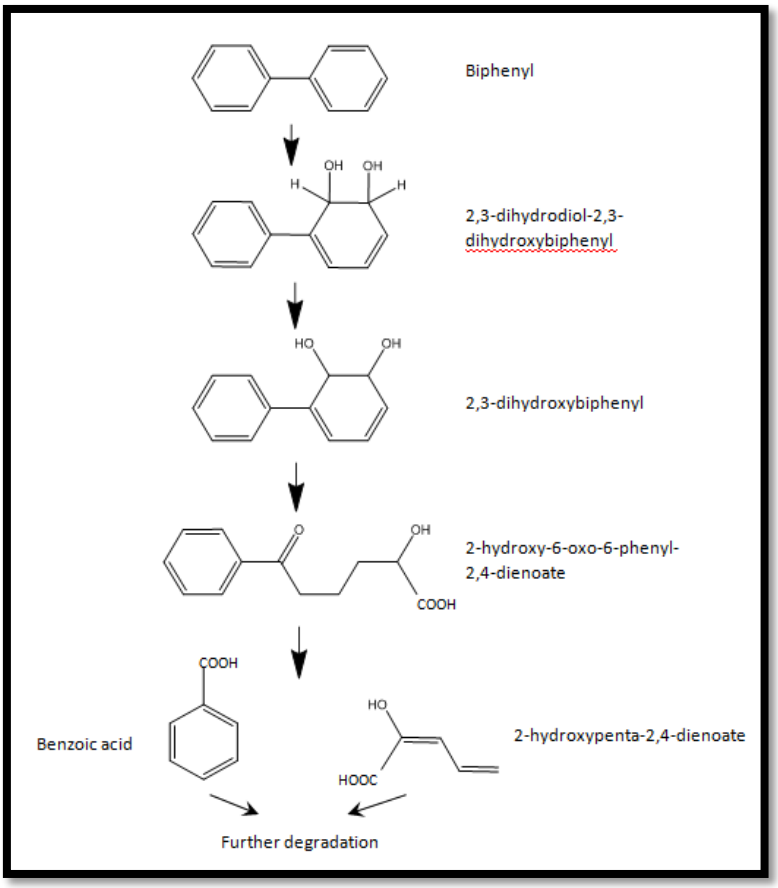


Figure 4. Biphenyl metabolic pathway (Adapted from Fortin et al. 2006)

Chapter III

STUDY OBJECTIVES

The first objective was to isolate and identify new *Paenibacillus* strains that have the ability to degrade PAHs as the sole carbon source. The second objective was to sequence their genomes and search for dioxygenase genes which may code for unique PAH dioxygenases. The third objective was to test the expression of the identified genes and measure their transcription under different growth conditions using reverse transcription polymerase chain reaction (RT-PCR) technique. Objectives two and three were also used to identify and analyze dioxygenase genes in the genome sequence of *P. naphthalenovorans* PRN1.

Chapter IV

MATERIALS AND METHODS

Enrichment media preparation

Mineral salts basal (MSB) medium (1L) (Stanier et al. 1966) was prepared by adding 40 ml solution A, 20 ml solution B, and 10 ml solution C to 930ml dH2O. One liter of solution A (pH 6.8) contained 138g Na_2HPO_4•$7H_2O$ and 68g KH_2PO_4. Two liters of solution B (pH 6.8) contained 20g nitrilotriacetic acid, 14.6g KOH, 59.1g $MgSO_4$, 6.7g $CaCl_2$, 18.5mg (NH_4) Mo_7O_{24}• -$4H_2O$, 198mg $FeSO_4$•$7H_2O$, and 100ml Trace metals solution, which contained 250mg Na_2EDTA•$2H_2O$, 1.1g $ZnSO_4$•$7H_2O$, 0.5g $FeSO_4$•$7H_2O$, 154mg $MnSO_4$•H2O, 40mg $CuSO_4$•5H2O, 25mg $Co(NO_3)_2$•$6H_2O$, and 18mg $Na_2B_4O_7$•$10H_2O$. One liter of solution C contained 100g $(NH_4)2SO_4$ only. Luria-Bertani (LB) broth medium was prepared by adding 10g tryptone, 5g yeast extract, and 5g sodium chloride to 1L dH_2O.

Isolation of PAH degrading bacteria from soil

The PAH degrading strains RD5 and RD6 were isolated from sediment samples obtained from the Logan River, Logan, Utah, USA, by aerobic enrichment using biphenyl as the sole carbon source. The isolation process involved adding 1 g sediment and 50 mg biphenyl (~6.5mM) to 50 ml MSB medium. The flasks were incubated at 30°C and 180 rpm for three days. The bacterial growth was monitored by taking samples from the enrichment media and the bacteria counted using light microscopy. After several dilutions and regrowth of the enrichment culture several purification steps were conducted by growing the bacteria on MSB biphenyl plates or on LB plates. Biphenyl was added to the plate lid so that the bacteria could grow on

the vapors. The strains were stored in 20% (v/v) glycerol in LB at -70 °C. The two strains were also tested on a variety of different polyaromatic compounds.

Extraction of genomic DNA

Genomic DNA was extracted from cells grown on LB agar using the MO BIO UltraClean Microbial DNA Isolation Kit. The first step of DNA extraction involved cell lysis using SDS, followed by centrifugation that separate the DNA from large cellular components. The second step involved adding a reagent that precipitates the proteins and other cell molecules, which may affect the purity of the DNA. The third step involved adding a salt solution which helps the DNA to bind to the spin filter. Ethanol was used to clean the DNA bound to the silica filter membrane, followed by centrifugation to remove the ethanol wash solution. Finally, elution buffer was added which helps in releasing the DNA that is bound to the filter. The concentration and the purity of the DNA were measured using the Nano drop 2000 spectrophotometer (Thermo Scientific).

PCR amplification and sequencing of 16S rRNA genes

The universal primers 27F, 517F, and 1522R were used in the PCR amplification of the 16S rRNA genes (Hendrickson et al. 2002; Johnson 1994). The PCR mixture of 20 µl contained 0.8 µl 27F or 517F primer (5 µM), 0.8 µl 1522R primer (5 µM), 10 µl ReadyMix (Sigma), 20 ng DNA template, and dH2O up to 20 µl. The amplification was performed using an Applied Biosystems Veriti 96 well thermal cycler programmed as: 3 min of initial denaturation at 95°C; followed by 30 cycles of 95°C for 30 sec, 55°C for 30 sec, and 72°C for 1 min; 10 min of final extension at 72°C. The PCR products were mixed with bromophenol blue (DNA loading dye) and run on 0.8% agarose gels. The agarose gel was made by adding 0.32 g agarose to 40

ml TAE buffer (1X), which is a mixture of Tris-acetate (0.04 M) and EDTA (0.001 M). Then, the PCR products were stained with ethidium bromide and visualized via UV light. The size of the bands was determined using a 1 kb plus DNA ladder.

The bands were cut from the gel and purified using the Gene Clean III Kit according to the manufacturer's instructions. Sequencing samples were sent to Genewiz (New Brunswick, NJ) and contained 5 µl 27F or 517F primer (5 µM), 40 ng purified 16s rRNA gene, and dH_2O up to 15 µl. The National Center for Biotechnology Information (NCBI) database was used to identify the 16S rRNA gene sequences via BLAST (Altschul et al. 1990).

Isolation of the dioxygenase genes

The dioxygenase genes from RD5 and RD6 were amplified using the universal primers DOα-2 and DOα-3 (Iida et al. 2002). The PCR conditions were set as: 3 min initial denaturation at 94°C; followed by 30 cycles of 94°C for 30sec, 45°C for 1 min, and 72°C for 1 min; 10 min of final extension at 72°C. The PCR products were analyzed on an (0.8%) agarose gel and then purified as described above. The purified dioxygenase genes from RD5 were sent directly to Genewiz for sequencing as described above while the genes isolated from RD6 were cloned into the pGEM-T easy vector (Promega) before sending them for sequencing.

The pGEM-T cloning mixture (10 µl) contained 5 µl (2X) ligation buffer, inserted DNA (3:1) ratio, 1 µl T-easy vector, 1 µl T4 ligase, and dH_2O up to 10 µl. After 24 h incubation at 4°C, the vector plus dioxygenase gene ligated mixture was transformed into *E. coli* DH5α. The transformation involved (i) adding 2 µl of the cloning mixture to 25 µl of DH5α and incubating on ice for 30min, (ii) heat shock for 45 sec at 40°C, (iii) ice for 5 min, (iv) adding 100 µl of LB broth to the tubes and

incubation for 1.5 h at 37°C. Then, DH5α were moved to LB agar, which contains X-gal (80 μg/ml) and ampicillin (100 μg/ml), and incubated for 24 h at 37°C. White colonies were selected for plasmid isolation, which was done according to the manufacturer's instructions (Macherey-Nagel).

Plasmids were digested by mixing 2 μl cut smart buffer, 500 ng pGEM-T+ dioxygenase gene, 1 μl *Eco*RI (Biolabs), and dH_2O up to 20 μl. The digestion mixture was incubated at 37°C for 1 h and then run on an 0.8% agarose gel. The pGEM-T+ dioxygenase genes (500 ng) were sent to Genewiz for sequencing using the T7 primer. The sequences were identified by BLAST using the NCBI database.

DNA extraction for full genome sequence analysis

The DNA was extracted from RD5 and RD6 according to the Qiagen genomic DNA isolation protocol. The bacteria were first grown in LB broth for 24 h. Then, the cells (2 ml) were harvested by centrifugation at 5000 x g for 10 min (4°C). After discarding the supernatant, the bacterial pellets were resuspended in 1 ml of Qiagen buffer B1 plus RNase (0.2 mg). The cells were lysed using 20 μl lysozyme (100 mg/ml) and the proteins were degraded using 45 μl of proteinase K (20 mg/ml) with incubation at 37°C for 30 min. For efficient deproteinization, 0.35 ml of buffer B2 was added and mixed by vortexing for a few seconds followed by incubation at 50°C for 30 min.

The Qiagen genomic-tip 20/G was equilibrated using 2 ml of buffer QBT and the tips allowed to empty by gravity flow. Then, the lysis mixture was applied to the equilibrated Qiagen Genomic-tip and allowed to enter the resin by gravity flow. Next, the Qiagen Genomic-tip was washed by 3 ml of Buffer QC. After that, the genomic DNA was eluted by adding 2 ml of Buffer QF. The eluted DNA was precipitated by

adding 1.4 ml of isopropanol followed by centrifugation at 5000 x g for 15 min at 4°C. After removing the supernatant, the DNA pellets were washed with 1 ml of cold 70% ethanol followed by a brief vortex and centrifugation at 5000 x g for 10 min at 4°C. The supernatant was carefully removed and the DNA pellets were air-dried for 10 min and then resuspended in 100 µl of Tris (10 mM), pH 8.5. Eventually, the genomic DNA was dissolved in the Tris buffer at 55°C for 2 h.

Genomic DNA was analyzed by agarose (0.8%) gel electrophoresis for 16 h at 25V and the band sizes determined using the 1 kb Extended DNA ladder. Finally, the genomic DNA was sequenced by MicrobesNG (University of Birmingham, UK).

Table 1. The Qiagen buffers composition

Buffer	Composition
Buffer B1 (Bacterial Lysis Buffer)	50 mM Tris.Cl, pH 8.0; 50 mM EDTA, pH 8.0; 0.5% Tween-20; 0.5% Triton X-100
Buffer B2 (Bacterial Lysis Buffer)	3 M guanidine HCl; 20% Tween-20
Buffer QBT (Equilibration Buffer)	750 mM NaCl; 50 mM MOPS, pH 7.0; 15% isopropanol, 0.15% Triton X-100
Buffer QC (Wash Buffer)	1.0 M NaCl; 50 mM MOPS, pH 7.0; 15% isopropanol
Buffer QF (Elution Buffer)	1.25 M NaCl; 50 mM Tris·Cl, pH 8.5; 15% isopropanol

Genome sequence analysis

The draft genome sequence of PRN1 was downloaded from GenBank and the draft genome sequences of RD5 and RD6 were received from MicrobesNG. The Blast+ program was downloaded from NCBI and installed on a Windows 10 PC.

Local databases of the amino acid and nucleotide sequences of PRN1, RD5, and RD6 were constructed with the makeblastdb program. The three databases were searched using a seed database of known dioxygenase amino acid sequences and the predicted amino acid sequences of the RD6 dioxygenase PCR products using the standalone programs blastp and tblastn. The independent sequence hits were extracted and re-run against the three databases to ensure capturing all of the dioxygenase genes and proteins in the three strains. Using the DNAStar Lasergene SeqBuilder program the amino acid sequences for each dioxygenase hit were extracted from the assembly contigs. Searching for related sequences in the GenBank database was performed with both the standalone Blast+ programs and the web-based Blast program. Dendrograms depicting the relationship of the different dioxygenase amino acid sequences to each other were constructed using the DNAStar Lasergene MegAlign program using clustalV.

PCR primer design

Seven different primers were designed (Sigma-Aldrich) based on the full genome sequence of *P. naphthalenovorans*. These primers were used to test transcription of the predicted dioxygenase genes in the strain PRN1 under different conditions (Table 2). The designed primers were first tested on the genomic DNA of the strain PRN1 as a positive control. The experiment was conducted by adding every primer separately, with final concentration of 0.2 μM, to the genomic DNA with final concentration of 1 ng/μl. The PCR conditions were set as follow: 3 min of initial denaturation at 95°C; followed by 30 cycles of 95°C for 30 sec, variable annealing temperatures for 30 sec, and 72°C for 1 min; 10 min of final extension at 72°C. The annealing temperatures of the seven primes were set as follow: 44°C for PRN1_108

and PRN1_115; 48°C for PRN1_114 and PRN1_109; 50°C for PRN1_106; 52°C for PRN1_111; 54°C for PRN1_123. The PCR products were run on a 1.5% agarose gel.

RNA isolation

The PRN1 strain was first grown on LB medium for 24 h and then moved to different flasks that contain 50 ml MSB broth plus naphthalene crystals (4 mM), DBF crystals (3 mM), benzoate (5 mM), cinnamate (20 mM), and succinate (10 mM) as the negative control . The cells were harvested after reaching mid-log phase by centrifugation at 5000 x g for 10 min at 4°C, and then resuspended using 1 ml TRIzol Reagent (Invitrogen by Thermo Fisher Scientific). Next, 0.2 ml of chloroform was added and the mixture was shaken vigorously and incubated for 3 min at room temperature followed by centrifugation at 12,000 x g for 15 min at 4°C. After that, the aqueous phase (~0.5 ml) was moved to a new tube followed by adding 0.5 ml isopropanol (100%). The mixture was incubated for 10 min at room temperature followed by centrifugation at 12,000 x g for 10 min at 4°C. Then the supernatant was discarded and the RNA pellets were washed using 1 ml of 75% ethanol followed by another wash using 1 ml of 95% ethanol. In the two washing steps, the tubes were briefly vortexed and then centrifuged at 7500 x g for 5 min at 4°C followed by discarding the wash. Eventually, the RNA pellets were air-dried for 5 min and then resuspended using 20 µl of RNase-free water with incubation at 55°C for 15 min.

The isolated RNA was treated with 1 µl TURBO DNase plus 2 µl 10X TURBO DNase buffer (Invitrogen by Thermo Fisher Scientific) and the mixture was incubated at 37°C for 25 min. Then, 2 µl of DNase inactivation reagent was added followed by incubation at room temperature for 5 min. Finally, the mixture was

centrifuged at 10,000 x g for 1.5 min and the RNA was transferred to fresh tubes and stored at -70°C.

RT-PCR

RT-PCR was conducted using a High Capacity cDNA Reverse Transcription Kit (Applied Biosystems). The RT-PCR mixture of 20 µl contained 2 µl 10X RT buffer, 0.8 µl 25X dNTP mix (100 mM), 2 µl 10X RT Random Primers, 1 µl multiscribe reverse transcriptase, 100 ng RNA, and nuclease-free H_2O up to 20 µl. The PCR conditions were set as follow: 25°C for 10 min, 37°C for 120 min, and 85°C for 5 min. The RT-PCR products (cDNA) was stored at -20°C. Finally, the designed primers were tested separately on the cDNA using the same concentrations and PCR conditions as used with the genomic DNA described above.

The three steps above were repeated with the RD5 strain, which was isolated and sequenced in this study, using a different set of designed primers (Table 2). The PCR conditions were kept the same except the annealing temperatures which were set as follow: 44°C for RD5_7 and RD5_27 primers; 45°C for RD5_3 primer; 46°C for RD5_5 primer; 47°C for RD5_30 and RD5_140 primers; 49°C for RD5_1 primer; 51°C for RD5_302 primers. In addition to the substrates mentioned above for PRN1, RD5 was also grown on biphenyl (3 mM) but not dibenzofuran.

Table 2: Primers used in this study

Gene name	Primer name	Primer sequence (5'- 3')	Product size	Reference
16S rRNA	27F 517F 1522R (Universal primers)	AGAGTTTGATCMTGGCTCAG GCCAGCAGCCGCGGTA AAGGAGGTGATCCANCCRCA	1- 1.5 Kb	Suzuki & Giovannoni (1996), Hendrickson et al. (2002).
Dioxygenase	DOα 2 DOα 3 (Universal primers)	TGYHSNTAYCAYGGNTGG TCNRCNGCNARYTTCCART T	300bp	Iida et al. 2002
PRN1_111	PRN1_111-F PRN1_111-R	GATTTACCGCGGACCAACCT TATCCCGGACGATTCCTCTT	~1 kb	This study
PRN1_123	PRN1_123-F PRN1_123-R	TCGTCACGAGAAATGTCGCG TCTGACTCTTCTCGCTTGCG	~900bp	This study

PRN1_108	PRN1_108-F PRN1_108-R	GAAACGGGTTCAATTAGTCG TAAGTCTTTGGTCAATCTGT	800bp	This study
PRN1_114	PRN1_114-F PRN1_114-R	GACGGGACAAAGATAAGAGA ACGCGAACTTCGGATCCCCA	~700bp	This study
PRN1_109	PRN1_109-F PRN1_109-R	TGCTGAAACAGGAGTTAGAC CACTACCCATCGCTGCGGGA	600bp	This study
PRN1_115	PRN1_115-F PRN1_115-R	CTTTGGAAGTGGATAGGATA TACAACCCAGCGCTGAGGAC	~500bp	This study
PRN1_106	PRN1_106-F PRN1_106-R	CGTCCCATTCCAAGAGTAGA GACTTGGGACTGCTTGTGAT	350bp	This study

RD5_1	RD5_1F RD5_1R	GCTGAACTTGGATAAGGCAA GCCATACCGCTTCACAACGT	1.1kb	This study
RD5_3	RD5_3F RD5_3R	ATTCCAGATACTTGCCATAA ACCTCACATTTTCGCTGTCC	1kb	This study
RD5_5	RD5_5F RD5_5R	CAAATCACGTGGCCGACTGA CCGTTTCGCACATAAAATAC	~900bp	This study
RD5_7	RD5_7F RD5_7R	AACGGTAACTTCAGCGTGGA ACACTTTCAATGCTTGTATC	~800bp	This study
RD5_27	RD5_27F RD5_27R	CAAATGTTCAATTGGTTAGC CCACAGCGTTGGACCGTAAA	~700bp	This study

RD5_30	RD5_30F RD5_30R	GTTCCAGCCAATACTCGTGC TATATGCAGGATATGGCCAT	~600bp	This study
RD5_140	RD5_140F RD5_140R	TCTTGCAGCATCTTCAGCCA TGAACGTAATTTGAATCCGG	~500bp	This study
RD5_302	RD5_302 RD5_302R	CCAACTGGAAGCTGACATCG CACACACGCATGTTCAAGTA	370bp	This study

Chapter V

RESULTS

Isolation and identification of PAH-degrading bacteria

Paenibacillus naphthalenovorans PRN1 was received from Dr. Max Haggblom and is the type strain of the species. Two strains, RD5 and RD6, were isolated by aerobic enrichment of sediment samples from the Logan River (Utah) using biphenyl as the sole carbon source. The turbidity of the enrichment media and direct count of the bacteria were used as an indicator of bacterial growth. The two strains formed white colonies when grown on LB agar and on MSB plus biphenyl. Both strains appeared yellow when grown on naphthalene possibly due to the formation of a *meta* cleavage product in the catabolic pathway. The three strains (PRN1, RD5, and RD6) were tested for growth on minimal medium with different aromatic and nonaromatic carbon sources (Table 3). The major difference between the three strains is that PRN1 grows on dibenzofuran and RD5 and RD6 grow on biphenyl. In addition, strain RD6 grows noticeably faster on solid minimal medium with naphthalene than strain RD5.

STRAIN	DD	DBF	NAP	BPH	CIN	BEN	SUC
PRN1	—	+	+	—	+	+	+
RD5	—	—	+	+	+	+	+
RD6	—	—	+	+	+	+	+

Table 3. Growth of *Paenibacillus* strains on various carbon sources.
Abbreviations: DD: dibenzo-*p*-dioxin, DBF: dibenzofuran, NAP: naphthalene, BPH: biphenyl, CIN: cinnamate, BEN: benzoate, SUC: succinate.

In order to identify strains RD5 and RD6 to the species level their genomic DNAs were isolated and the 16S rRNA genes PCR amplified and directly sequenced with internal primers. The sequences were compared to 16S rRNA gene sequences from type strains of *Paenibacillus*. The closest match is *P. validus* which is closely related to *P. naphthalenovorans* PRN1 (Figure 5).

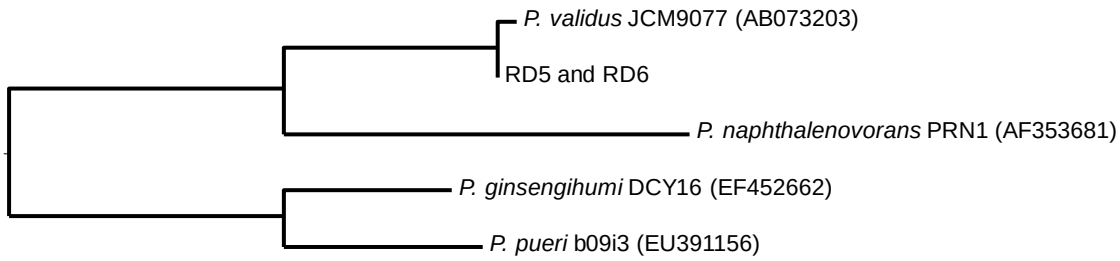


Figure 5. Dendrogram showing the relationship between the 16S rRNA gene sequences of RD5 and RD6 and those from the closest matching *Paenibacillus* type strains.

Preliminary analysis of dioxygenase genes in RD5 and RD6

Strains RD5 and RD6 were characterized further by analyzing the types of dioxygenase genes that can be PCR amplified from their genomes. The universal primers DOα-2 and DOα-3 previously used to characterize dioxygenase genes in *Paenibacillus* sp. strain YK5 (Iida et al. 2002) were used to isolate a fragment of the dioxygenase genes from strains RD5 and RD6. The expected 300 bp band was observed after running the PCR products on a 0.8% gel. After direct sequencing, the RD5 PCR product showed a single nucleotide sequence while the RD6 PCR product showed a mixture of sequences. The RD6 PCR products were then cloned into the pGEM-T vector and six clones were sequenced. The sequences of the seven sequenced PCR products were compared with sequences in the GenBank (NCBI) database. A dendrogram (Figure 6) of the RD5 and RD6 sequences along with the

closest matches in GenBank shows that the dioxygenase gene sequence fragments from RD5 and RD6 are unique. The RD5 and RD6 dioxygenase gene fragment sequences clustered away from genes from other closely related bacteria. The six RD6 sequences fell into three groups and the single RD5 sequence matched one of the RD6 sequences (Figure 6).

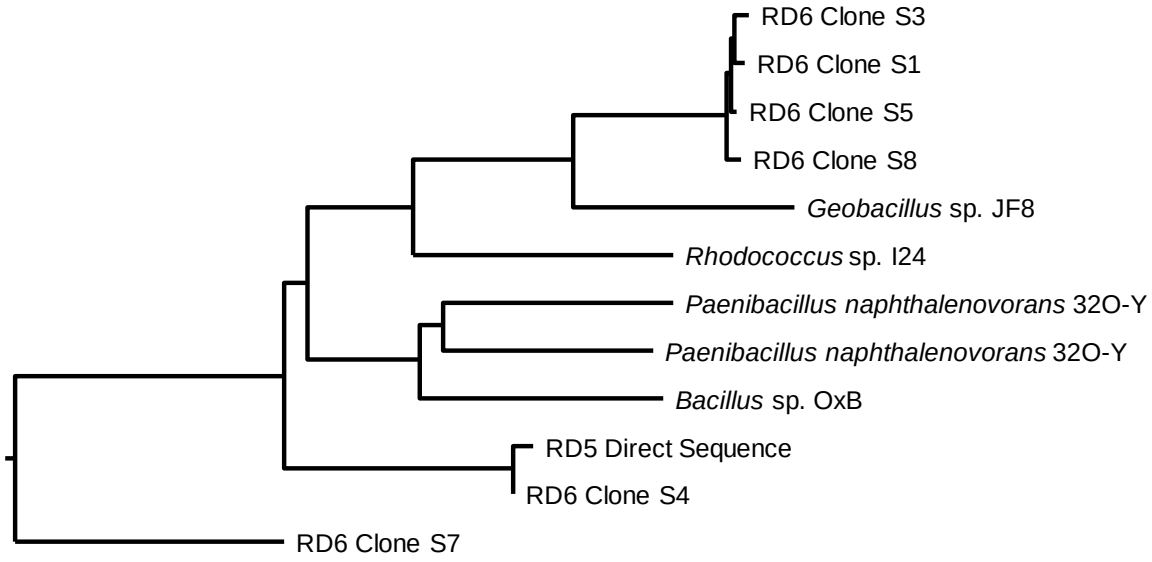


Figure 6. Phylogenic tree of the RD5 and RD6 dioxygenase gene fragment nucleotide sequences compared to the closest nucleotide sequences in GenBank.

Genome Sequence of PRN1, RD5, and RD6

The genome sequence of *P. naphthalenovorans* was completed in 2016 in a collaboration between the Haggblom laboratory and the DOE Joint Genome Institute. The GenBank project accession is FNDY00000000. Total genomic DNA of RD5 and RD6 was isolated and the quality verified by agarose gel electrophoresis (large high molecular weight only with no smearing). Sequencing and assembly of the RD5 and RD6 genomes was performed by MicrobesNG at the University of Birmingham (UK). The sequencing technology utilized 2x250 bp paired-end reads and Illumina MySeq

or HiSeq instruments. As shown in Table 4 the three draft genomes are similar in size (>5 megabases) and quality (N50 and L50 values).

Strain	Total Length	Contigs	Largest Contig	GC (%)	N50	L50
RD5	5,843,697	174	474,937	51.76	192,561	10
RD6	5,358,074	48	804,754	52.38	271,878	5
PRN1	5,289,735	99	501,793	50.00	145,459	11

Table 4. Summary of the draft genome sequences of RD5, RD6, and PRN1. Abbreviations: N50: is the length for which the collection of all contigs of that length or longer covers at least half an assembly, L50: is the number of contigs equal to or longer than N50. L50 is thus the minimal number of large contigs that cover half the genome.

Identification of genes encoding dioxygenases in PRN1, RD5, and RD6

The PRN1, RD5, and RD6 genomes were examined for genes coding for dioxygenases of the alpha:beta two subunit oxygenase component type. This dioxygenase family initiates the metabolism of the target aromatic compounds in this study (polyaromatic hydrocarbons as well as benzoate and cinnamate). A local database of each genome was separately scanned with the Blast+ tblastn program using seed amino acids sequences of known dioxygenase alpha subunits. Protein hits from each of the three genomes were extracted, placed in a single fasta file, and were used again to scan the three genomes using the Blast+ tblastn program. This process was iteratively repeated until no new dioxygenase amino acid sequence were found. Because the Rieske motif found in dioxygenase alpha subunit sequences is quite common the collection of amino acid sequences was manually curated to remove single component dioxygenase amino acid sequences and to remove non-aromatic

dioxygenase enzyme sequences (nitrite reductase, for example). In total, six dioxygenases were found in RD6, seven in PRN1, and eight in RD5.

Comparison of the discovered dioxygenases from the three strains to each other and to the recently completed *P. naphthalenovorans* 32O-Y genome sequence (Butler et al., 2016) reveals some interesting information (Figure 7). The dioxygenases found in *P. naphthalenovorans* 32O-Y and PRN1 are practically identical. There is only one cluster of closely related sequences between the four strains (near 32O-Y ALS23877 in Figure 7). Two of the RD5 sequences have no close relative among the three other genomes but the other six RD5 sequences are practically identical to the six RD6 sequences. Several of the sequences, especially in the bottom portion of the dendrogram in Figure 7 are distantly related to each other suggesting that the *P. naphthalenovorans* and *P. validus* strains may have dioxygenase enzymes with different functions.

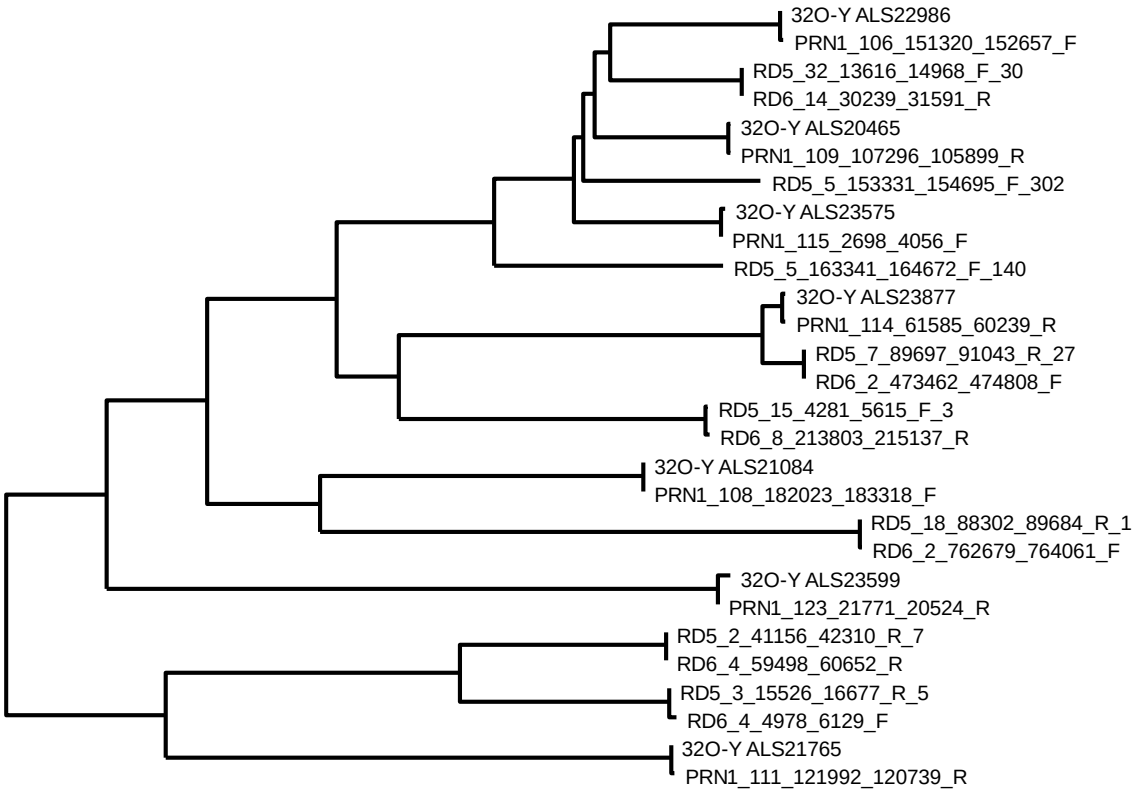


Figure 7. Relationship between different dioxygenase alpha subunit amino acid sequences in *P. naphthalenovorans* PRN1 and 32O-Y and *P. validus* RD5 and RD6. Nomenclature: The PRN1, RD5, and RD6 sequences are labelled S*trainName_Contig#_BeginPosition_EndPosition_Direction*. The 32O-Y sequenced are labelled with the GenBank accession number.

RT-PCR of the predicted dioxygenase genes from PRN1 and RD5

Now that the different dioxygenase genes have been identified in the *Paenibacillus* strains the question is which gene(s) and their associated protein(s) are involved in degrading each of the target aromatic compounds. PCR primers were designed based on the identified dioxygenase gene sequences in *P. naphthalenovorans* PRN1 and *P. validus* RD5. The primers were designed in such a way that each target gene would result in a different size PCR product to eliminate any confusion that may occur when the same size product is generated for each primer

pair (Table 2). The designed primers for PRN1 (7 primer pairs) and RD5 (8 primer pairs) were first tested on the respective genomic DNA as a positive control (Figure 8). As expected, each PCR primer pair amplified a different size fragment from the respective genomes.

The next goal was to determine which of the dioxygenase genes are expressed under different growth conditions using RT-PCR. This involved isolation of total RNA from both strains following growth on naphthalene, dibenzofuran (PRN1 only), biphenyl (RD5 only), benzoate, cinnamate, and succinate (as the negative control). The isolated RNA was treated with DNase, reverse transcribed to cDNA, and subjected to individual PCR reactions with each primer set. The data (Figure 9 and Table 5) shows that three of the seven PRN1 genes and four of the eight RD5 genes are expressed under one or more of the growth conditions. In *P. naphthalenovorans* PRN1_108 is expressed during growth on DBF, PRN1_114 is expressed during growth on naphthalene, cinnamate, and benzoate, and PRN1_109 is expressed during growth on cinnamate and benzoate. In *P. validus* the RD5_2 (7), RD5_3 (5), and RD5_7 (27) genes were all expressed during growth on naphthalene while the RD5_18 (1) gene was expressed only during growth on benzoate. None of the predicted RD5 genes were expressed in the presence of biphenyl or cinnamate.

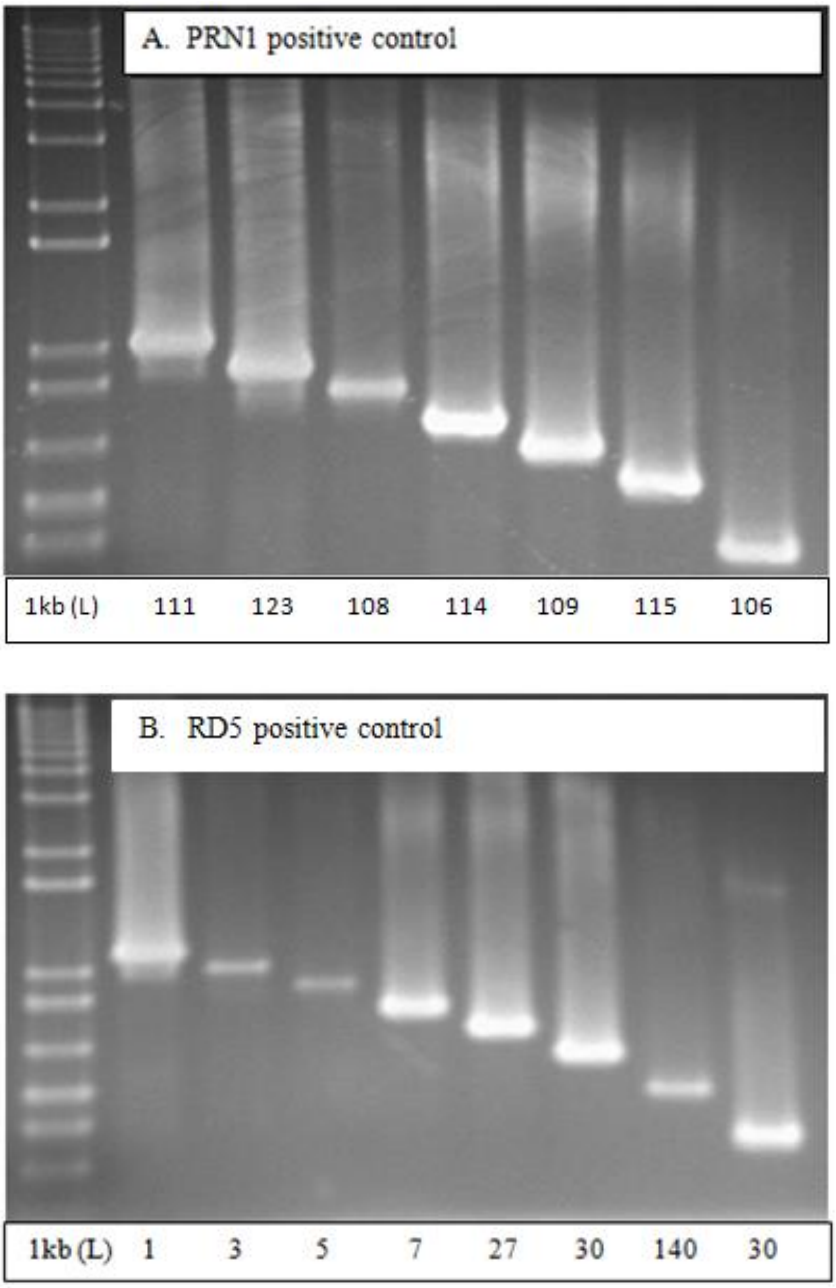
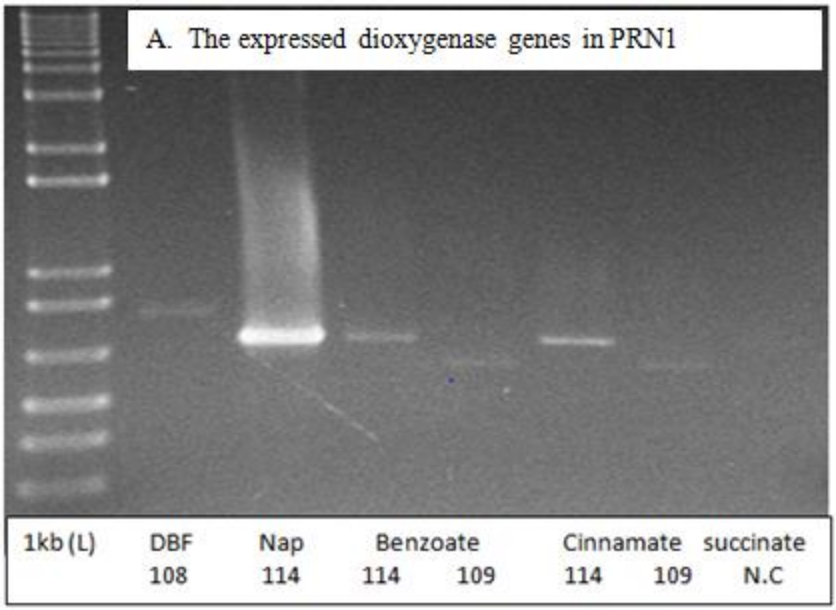


Figure 8. PCR amplification of RD5 or PRN1 genomic DNA with the PRN1 (panel A) and the RD5 (panel B) dioxygenase primer pairs.



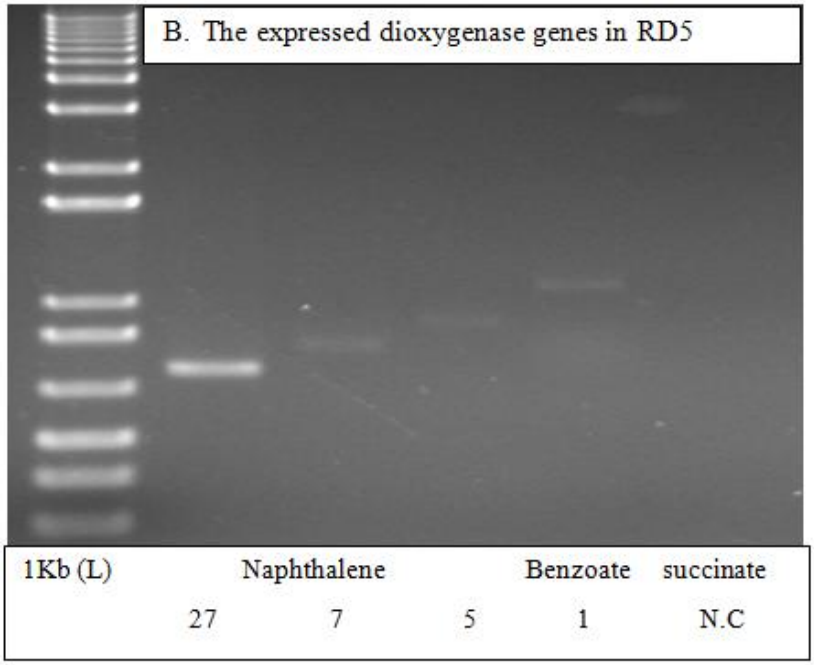


Figure 9. RT-PCR of RNA extracted from either RD5 or PRN1 grown under different conditions with the indicated dioxygenase primer pairs

Gene	DBF	NAP	BPH	CIN	BEN
PRN1_106			NA		
PRN1_108	+		NA		
PRN1_109			NA	+	+
PRN1_111			NA		
PRN1_114		+	NA	+	+
PRN1_115			NA		
PRN1_123			NA		
RD5_2 (7)	NA	+			
RD5_3 (5)	NA	+			
RD5_5 (140)	NA				
RD5_5 (302)	NA				
RD5_7 (27)	NA	+			
RD5_15 (3)	NA				
RD5_18 (1)	NA				+
RD5_32 (30)	NA				

Table 5. Gene induction in during growth on different hydrocarbons as determined by RT-PCR for each gene. A + indicates that a PCR product was detected. A blank indicates that no PCR product was detected. NA: Not applicable.

Comparison of the *Paenibacillus* PRN1 and RD5 dioxygenases with closest GenBank matches

It is possible to glean information about the biological activity of newly discovered dioxygenase enzymes through comparison with other amino acid sequences in the GenBank database. Using the local Blast+ blastP program we queried the GenBank database with the seven PRN1 dioxygenases and the eight RD5 dioxygenases. The blastP program was set with a 0.001 Expect value cutoff and to return only the 10 best matches for each amino acid sequence. The Accession Numbers for all of the GenBank matches were collated, deduplicated, and the corresponding amino acid sequences retrieved from the database. Using the DNAStar MegAlign program and clustalV a dendrogram was constructed with the retrieved sequences along with the PRN1 and RD5 dioxygenase sequences. The dendrogram was visually inspected and poor matches were removed. The dendrogram was redrawn with the best matches and the PRN1 and RD5 sequences (Figure 10). The results of the analysis show that there are not many good matches in the GenBank database for the *Paenibacillus* PRN1 and RD5 dioxygenase sequences. There is an exact match of the PRN1_108 amino acid sequence to the *Paenibacillus* sp. strain YK5 dibenzofuran dioxygenase (DbfA1) (Iida et al. 2006). There is a close match between the JF8 BphA biphenyl dioxygenase and RD5_15 and there is a close match between the suspected JF8 NahA1 naphthalene dioxygenase and PRN1_114 and RD5_7. The *Geobacillus* sp. strain JF8 grows on both biphenyl and naphthalene and the initial biphenyl dioxygenase has been identified (Mukerjee-Dhar et al., 2005; Shimura et al., 1999; Shintani et al., 2014).

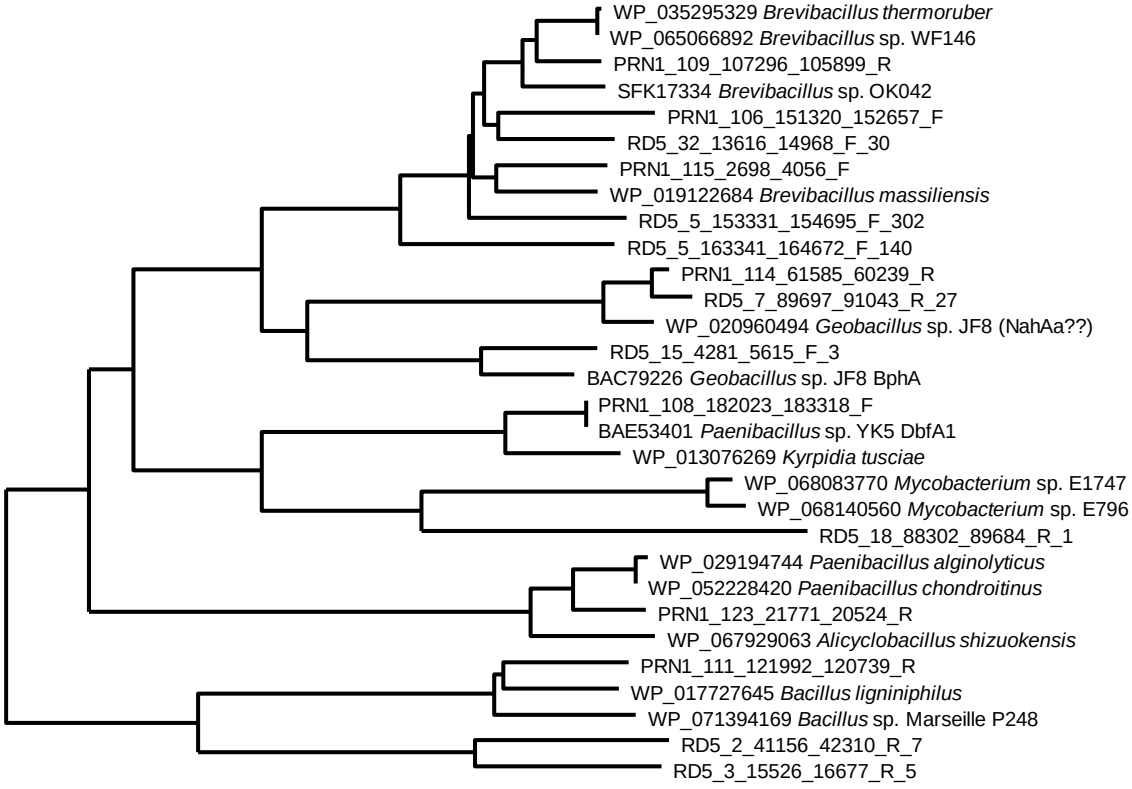


Figure 10. Relationship between different dioxygenase alpha subunit amino acid sequences in *P. naphthalenovorans* PRN1 and *P. validus* RD5 and best matches in the GenBank database.
Nomenclature: The PRN1 and RD5 sequences are labelled S*trainName_Contig#_BeginPosition_EndPosition_Direction*. The database sequences are labelled with the GenBank accession number.

Chapter VI
DISCUSION

The goal of this project was to identify the genes in *Paenibacillus* responsible for polyaromatic hydrocarbon degradation. The genomes of two new *P. validus* strains (RD5 and RD6) were sequenced and analyzed along with one previously sequenced *P. naphthalenovorans* (PRN1) strain. Seven genes encoding dioxygenase alpha subunits were found in the *P. naphthalenovorans* PRN1 strain, eight were found in the *P. validus* RD5 strain, and six were found in the *P. validus* RD6 strain. Using RT-PCR we examined the transcription of 15 predicted dioxygenase alpha subunit genes from *P. naphthalenovorans* PRN1 and *P. validus* RD5 following growth on different carbon sources. The results of the RT-PCR experiments along with the phylogenetic analysis of the dioxygenases is summarized in Figure 11.

Interestingly, some RD5 genes showed unusual results such as RD5_2(7) and RD5_3(5), which were expressed during growth on naphthalene but clustered away from the dioxygenase genes of other naphthalene degrading bacteria. Also, the RD5_18(1) gene, which was expressed during growth on benzoate but clustered closely to the angular dioxygenase genes. A similar phenomenon was seen with some PRN1 genes where PRN1_109 and PRN1_114 were expressed during growth on more than one aromatic substrate. Many aromatic pathways have multiple dioxygenases catalyzing different enzymatic steps in the catabolic pathway. The naphthalene degradation pathway through gentisate, for instance, has both a naphthalene dioxygenase and a gentisate dioxygenase (but naphthalene degradation through the salicylate pathway only has the naphthalene dioxygenase). In addition, the biphenyl degradation pathway has both a biphenyl and a benzoate dioxygenase.

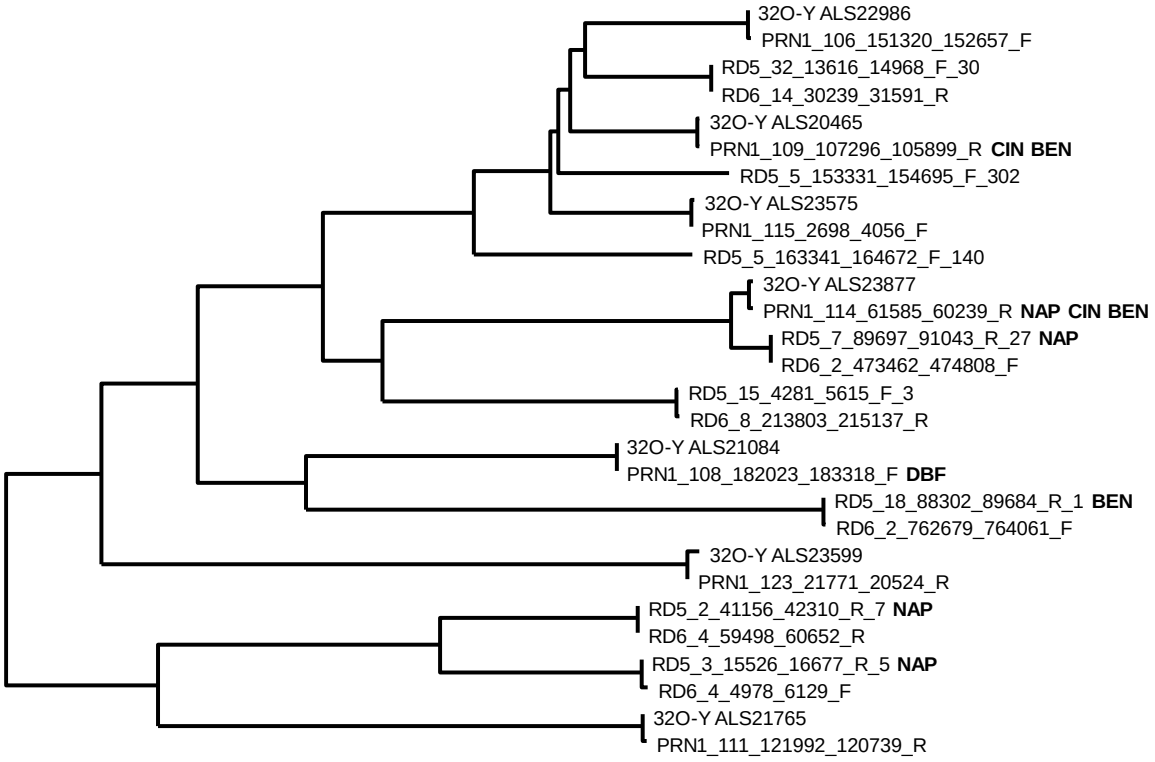


Figure 11. Gene induction analysis with the phylogenetic groupings of the dioxygenase enzymes. Abbreviations: NAP: naphthalene induction, DBF: dibenzofuran induction, CIN: cinnamate induction, BEN: benzoate induction.

The PRN1_108 enzyme, induced during growth on dibenzofuran, is practically identical to the well studied dibenzofuran dioxygenase of *Paenibacillus* sp. YK5 (Iida et al. 2006). The PRTN1_108 dioxygenase thus most certainly catalyzes the first enzymatic step in PRN1 dibenzofuran metabolism.

The RT-PCR results showed that both the RD5_7(27) and PRN1_114 genes were expressed during growth on naphthalene and the amino acid sequences of the encoded proteins are closely related to a dioxygenase in a *Geobacillus* strain capable of growth on naphthalene (Mukerjee-Dhar et al., 2005; Shimura et al., 1999; Shintani et al., 2014). Based on these data we strongly suspect that the RD5_7(27) and

PRN1_114 dioxygenases catalyze the initial dioxygenase step in naphthalene metabolism in both PRN1 and RD5.

While we did not detect expression of any of the *P. validus* RD5 dioxygenase genes during growth on biphenyl the RD5_15(3) dioxygenase is closely related to the known BphA1 biphenyl dioxygenase in *Geobacillus* sp. strain JF8 (Mukerjee-Dhar et al., 2005; Shimura et al., 1999; Shintani et al., 2014). Based on this data we strongly suspect that the RD5_15(3) dioxygenase catalyzes the initial enzymatic step in RD5 biphenyl degradation.

More experiments are needed in order to further study the role of the dioxygenase genes and enzymes in *Paenibacillus* aromatic metabolism. For instance, the genes could be cloned into an expression vector and the enzymatic activity tested in *E. coli*. Also, knocking out these genes in either RD5 or PRN1 can be conducted followed by testing for growth on different aromatic substrates. Both experiments in combination can accurately identify the genes and their encoded proteins.

Chapter VII

REFERENCES CITED

Adams NG, Richardson DM. 1953. Isolation and identification of biphenyls from West Edmond crude oil. Anal Chem **25**:1073-1074.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. J Mol Biol **215**:403-410.

Ash C, Farrow JAE, Wallbanks S, Collins MD. 1991. Phylogenetic heterogeneity of the genus *Bacillus* revealed by comparative analysis of small subunit ribosomal RNA sequences. Lett Appl Microbiol **13**:202-206.

Ash C, Priest FG, Collins MD. 1993. Molecular identification of rRNA group 3 bacilli (Ash, Farrow, Wallbanks and Collins) using a PCR probe test. Proposal for the creation of a new genus *Paenibacillus*. Antonie Van Leeuwenhoek **64**:253-260.

Bedard DL, Wagner RE, Brennan MJ, Haberl ML, Brown JF, Jr. 1987. Extensive degradation of Aroclors and environmentally transformed polychlorinated biphenyls by *Alcaligenes eutrophus* H850. Appl Environ Microbiol **53**:1094-1102.

Bostrom CE, Gerde P, Hanberg A, Jernstrom B, Johansson C, Kyrklund T, Rannug A, Tornqvist M, Victorin K, Westerholm R. 2002. Cancer risk assessment, indicators, and guidelines for polycyclic aromatic hydrocarbons in the ambient air. Environ Health Perspect **110 Suppl 3**:451-488.

Bredemann G, Heigener H. 1935. *Bacillus validus* Bredemann and Heigener. Zentralbl Bakteriol Parasitenkd Infektionskr Hyg Abt. 2 **93**:97-98.

Butler RR, 3rd, Wang J, Stark BC, Pombert JF. 2016. Complete genome sequences of two interactive moderate thermophiles, *Paenibacillus napthalenovorans* 32O-Y and Paenibacillus sp. 32O-W. Genome Announc **4**:e01717-01715.

Cerniglia CE. 1984. Microbial metabolism of polycyclic aromatic hydrocarbons. Adv Appl Microbiol **30**:31-71.

Cerniglia CE. 1993. Biodegradation of polycyclic aromatic hydrocarbons. Curr Opin Biotechnol **4**:331-338.

Chen SJ, Su HB, Chang JE, Lee WJ, Huang KL, Hsieh LT, Huang YC, Lin WY, Lin CC. 2007. Emissions of polycyclic aromatic hydrocarbons (PAHs) from the pyrolysis of scrap tires. Atmos Environ **41**:1209-1220.

Chen YC, Lee WJ, Uang SN, Lee SH, Tsai PJ. 2006. Characteristics of polycyclic aromatic hydrocarbon (PAH) emissions from a UH-1H helicopter engine and its impact on the ambient environment. Atmos Environ **40**:7589-7597.

Chung YR, Kim CH, Hwang I, Chun J. 2000. *Paenibacillus koreensis* sp. nov., a new species that produces an iturin-like antifungal compound. Int J Syst Evol Microbiol **50 Pt 4**:1495-1500.

Claus D, Berkeley R (1986). Genus *Bacillus* Cohn, 1872. Bergey's Manual of Systematic Bacteriology. Sneath P., Mair N., Sharpe M., Holt J. Baltimore, Williams & Wilkins Co. **2:**1105-1139.

Cook DJ, Schlemmer S, Balucani N, Wagner DR, Steiner B, Saykally RJ. 1996. Infrared emission spectra of candidate interstellar aromatic molecules. Nature **380**:227-229.

Culp SJ, Gaylor DW, Sheldon WG, Goldstein LS, Beland FA. 1998. A comparison of the tumors induced by coal tar and benzo[a]pyrene in a 2-year bioassay. Carcinogenesis **19**:117-124.

Daane LL, Harjono I, Barns SM, Launen LA, Palleron NJ, Haggblom MM. 2002. PAH-degradation by *Paenibacillus* spp. and description of *Paenibacillus naphthalenovorans* sp. nov., a naphthalene-degrading bacterium from the rhizosphere of salt marsh plants. Int J Syst Evol Microbiol **52**:131-139.

Daane LL, Harjono I, Zylstra GJ, Haggblom MM. 2001. Isolation and characterization of polycyclic aromatic hydrocarbon-degrading bacteria associated with the rhizosphere of salt marsh plants. Appl Environ Microbiol **67**:2683-2691.

Dipple A. 1985. Polycyclic aromatic hydrocarbon carcinogenesis - an introduction. ACS Symposium Series **283**:1-17.

Eckert JW (1990). Impact of fungicide resistance on citrus fruit decay control. Managing Resistance to Agrochemicals, American Chemical Society. **421:**286-302.

Fang M, Zheng M, Wang F, To KL, Jaafar AB, Tong SL. 1999. The solvent-extractable organic compounds in the Indonesia biomass burning aerosols - characterization studies. Atmos Environ **33**:783-795.

Feng XL, Pisula W, Mullen K. 2009. Large polycyclic aromatic hydrocarbons: Synthesis and discotic organization. Pure Appl Chem **81**:2203-2224.

Fortin PD, Horsman GP, Yang HM, Eltis LD. 2006. A glutathione S-transferase catalyzes the dehalogenation of inhibitory metabolites of polychlorinated biphenyls. J Bacteriol **188**:4424-4430.

Freeman DJ, Cattell FCR. 1990. Wood burning as a source of atmospheric polycyclic aromatic hydrocarbons. Environ Sci Technol **24**:1581-1585.

Fuenmayor SL, Wild M, Boyes AL, Williams PA. 1998. A gene cluster encoding steps in conversion of naphthalene to gentisate in *Pseudomonas* sp. strain U2. J Bacteriol **180**:2522-2530.

Gardener BBM. 2004. Ecology of *Bacillus* and *Paenibacillus* spp. in agricultural systems. Phytopathology **94**:1252-1258.

Godoi AFL, Ravindra K, Godoi RHM, Andrade SJ, Santiago-Silva M, Van Vaeck L, Van Grieken R. 2004. Fast chromatographic determination of polycyclic aromatic hydrocarbons in aerosol samples from sugar cane burning. J Chromatogr A **1027**:49-53.

Gordon R, Haynes W, Pang C. 1973. The genus *Bacillus*. Agriculture Handbook no. 427.

Halasinski TM, Salama F, Allamandola LJ. 2005. Investigation of the ultraviolet, visible, and near-infrared absorption spectra of hydrogenated polycyclic aromatic hydrocarbons and their cations. Astrophys J **628**:555-566.

Haritash AK, Kaushik CP. 2009. Biodegradation aspects of polycyclic aromatic hydrocarbons (PAHs): a review. J Hazard Mater **169**:1-15.

Harvey R. 1991. Polycyclic aromatic hydrocarbons: chemistry and carcinogenicity. Cambridge University Press, Cambridge.

Hendrickson ER, Payne JA, Young RM, Starr MG, Perry MP, Fahnestock S, Ellis DE, Ebersole RC. 2002. Molecular analysis of *Dehalococcoides* 16S ribosomal DNA from chloroethene-contaminated sites throughout North America and Europe. Appl Environ Microbiol **68**:485-495.

Heyndrickx M, Vandemeulebroecke K, Scheldeman P, Hoste B, Kersters K, De Vos P, Logan NA, Aziz AM, Ali N, Berkeley RC. 1995. *Paenibacillus* (formerly *Bacillus*) *gordonae* (Pichinoty et al. 1986) Ash et al. 1994 is a later subjective synonym of *Paenibacillus* (formerly *Bacillus*) *validus* (Nakamura 1984) Ash et al. 1994: emended description of *P. validus*. Int J Syst Bacteriol **45**:661-669.

Iida T, Mukouzaka Y, Nakamura K, Kudo T. 2002. Plasmid-borne genes code for an angular dioxygenase involved in dibenzofuran degradation by *Terrabacter* sp. strain YK3. Appl Environ Microbiol **68**:3716-3723.

Iida T, Nakamura K, Izumi A, Mukouzaka Y, Kudo T. 2006. Isolation and characterization of a gene cluster for dibenzofuran degradation in a new dibenzofuran-utilizing bacterium, *Paenibacillus* sp. strain YK5. Arch Microbiol **184**:305-315.

Ikram-ul-Haq, Hameed U, Mahmood Z, Javed MM. 2012. Solid state fermentation for the production of alpha-amylase by *Paenibacillus amylolyticus*. Pak J Bot **44**:341-346.

Jenkins BM, Jones AD, Turn SQ, Williams RB. 1996. Emission factors for polycyclic aromatic hydrocarbons from biomass burning. Environ Sci Technol **30**:2462-2469.

Johnsen AR, Wick LY, Harms H. 2005. Principles of microbial PAH-degradation in soil. Environ Pollut **133**:71-84.

Kanzawa Y, Harada A, Takeuchi M, Yokota A, Harada T. 1995. *Bacillus curdlanolyticus* sp. nov. and *Bacillus kobensis* sp. nov., which hydrolyze resistant curdlan. Int J Syst Evol Microbiol **45**:515-521.

Kasuga K, Habe H, Chung JS, Yoshida T, Nojiri H, Yamane H, Omori T. 2001. Isolation and characterization of the genes encoding a novel oxygenase component of angular dioxygenase from the gram-positive dibenzofuran-degrader *Terrabacter* sp. strain DBF63. Biochem Biophys Res Commun **283**:195-204.

Lal S, Tabacchioni S. 2009. Ecology and biotechnological potential of *Paenibacillus polymyxa*: a minireview. Indian J Microbiol **49**:2-10.

Lyu Y, Zheng W, Zheng T, Tian Y. 2014. Biodegradation of polycyclic aromatic hydrocarbons by *Novosphingobium pentaromativorans* US6-1. PLoS One **9**:e101438.

Marchand N, Besombes JL, Chevron N, Masclet P, Aymoz G, Jaffrezo JL. 2004. Polycyclic aromatic hydrocarbons (PAHs) in the atmospheres of two French alpine valleys: sources and temporal patterns. Atmos Chem Phys **4**:1167-1181.

Marr LC, Dzepina K, Jimenez JL, Reisen F, Bethel HL, Arey J, Gaffney JS, Marley NA, Molina LT, Molina MJ. 2006. Sources and transformations of

particle-bound polycyclic aromatic hydrocarbons in Mexico City. Atmos Chem Phys **6**:1733-1745.

Montes MJ, Mercade E, Bozal N, Guinea J. 2004. *Paenibacillus antarcticus* sp. nov., a novel psychrotolerant organism from the Antarctic environment. Int J Syst Evol Microbiol **54**:1521-1526.

Mukerjee-Dhar G, Shimura M, Miyazawa D, Kimbara K, Hatta T. 2005. *bph* genes of the thermophilic PCB degrader, *Bacillus* sp. JF8: characterization of the divergent ring-hydroxylating dioxygenase and hydrolase genes upstream of the Mn-dependent BphC. Microbiology (Reading, England) **151**:4139-4151.

Mutangadura G. 2002. World health report 2002: reducing risks, promoting healthy life.

Nakamura LK. 1984. *Bacillus Amylolyticus* sp. nov., nom. rev., *Bacillus Lautus* sp. nov., nom. rev., *Bacillus Pabuli* sp. nov., nom. rev., and *Bacillus Validus* sp. nov., nom. rev. Int J Syst Evol Microbiol **34**:224-226.

Nakamura LK. 1987. *Bacillus alginolyticus* sp. nov. and *Bacillus chondroitinus* sp. nov., 2 alginate degrading species. Int J Syst Evol Microbiol **37**:284-286.

Nakamura LK, Swezey J. 1983. Taxonomy of *Bacillus circulans* Jordan 1890 - base composition and reassociation of deoxyribonucleic acid. Int J Syst Evol Microbiol **33**:46-52.

Nazina TN, Tourova TP, Poltaraus AB, Novikova EV, Grigoryan AA, Ivanova AE, Lysenko AM, Petrunyaka VV, Osipov GA, Belyaev SS, Ivanov MV. 2001. Taxonomic study of aerobic thermophilic bacilli: descriptions of *Geobacillus subterraneus* gen. nov., sp. nov. and *Geobacillus uzenensis* sp. nov. from petroleum reservoirs and transfer of *Bacillus stearothermophilus*, *Bacillus thermocatenulatus*, *Bacillus thermoleovorans*, *Bacillus kaustophilus*, *Bacillus thermodenitrificans* to *Geobacillus* as the new combinations *G. stearothermophilus*, *G. th*. Int J Syst Evol Microbiol **51**:433-446.

Nebert DW, Dalton TP, Okey AB, Gonzalez FJ. 2004. Role of aryl hydrocarbon receptor-mediated induction of the CYP1 enzymes in environmental toxicity and cancer. J Biol Chem **279**:23847-23850.

Ouyang J, Pei Z, Lutwick L, Dalal S, Yang L, Cassai N, Sandhu K, Hanna B, Wieczorek RL, Bluth M, Pincus MR. 2008. Case report: *Paenibacillus thiaminolyticus*: a new cause of human infection, inducing bacteremia in a patient on hemodialysis. Ann Clin Lab Sci **38**:393-400.

Phillips DH. 1983. Fifty years of benzo(a)pyrene. Nature **303**:468-472.

Providenti MA, Lee H, Trevors JT. 1993. Selected factors limiting the microbial degradation of recalcitrant compounds. J Ind Microbiol **12**:379-395.

Ramesh A, Walker SA, Hood DB, Guillen MD, Schneider K, Weyand EH. 2004. Bioavailability and risk assessment of orally ingested polycyclic aromatic hydrocarbons. Int J Toxicol **23**:301-333.

Ravindra K, Sokhi R, Van Grieken R. 2008. Atmospheric polycyclic aromatic hydrocarbons: Source attribution, emission factors and regulation. Atmos Environ **42**:2895-2921.

Schauer JJ, Cass GR. 2000. Source apportionment of wintertime gas-phase and particle-phase air pollutants using organic compounds as tracers. Environ Sci Technol **34**:1821-1832.

Seo JS, Keum YS, Li QX. 2009. Bacterial degradation of aromatic compounds. Int J Environ Res Public Health **6**:278-309.

Shimura M, Mukerjee-Dhar G, Kimbara K, Nagato H, Kiyohara H, Hatta T. 1999. Isolation and characterization of a thermophilic *Bacillus* sp. JF8 capable of degrading polychlorinated biphenyls and naphthalene. FEMS Microbiol Lett **178**:87-93.

Shintani M, Ohtsubo Y, Fukuda K, Hosoyama A, Ohji S, Yamazoe A, Fujita N, Nagata Y, Tsuda M, Hatta T, Kimbara K. 2014. Complete genome sequence of the thermophilic polychlorinated biphenyl degrader *Geobacillus* sp. strain JF8 (NBRC 109937). Genome Announc 2:e01213-01213.

Smith N, Gordon R, Clark F. 1952. Aerobic spore forming bacteria. Agriculture Monograph no. 16.

Standley LJ, Simoneit BRT. 1987. Characterization of extractable plant wax, resin, and thermally matured components in smoke particles from prescribed burns. Environ Sci Technol **21**:163-169.

Stanier RY, Palleroni NJ, Doudoroff M. 1966. The aerobic pseudomonads: a taxonomic study. J Gen Microbiol **43**:159-271.

Suzuki MT, Giovannoni SJ. 1996. Bias caused by template annealing in the amplification of mixtures of 16S rRNA genes by PCR. Appl Environ Microbiol **62**:625-630.

Venkataraman C, Negi G, Sardar SB, Rastogi R. 2002. Size distributions of polycyclic aromatic hydrocarbons in aerosol emissions from biofuel combustion. J Aerosol Sci **33**:503-518.

Ward O, Singh A, Van Hamme J. 2003. Accelerated biodegradation of petroleum hydrocarbon waste. J Ind Microbiol Biotechnol **30**:260-270.

Wild SR, Jones KC. 1995. Polynuclear aromatic hydrocarbons in the United Kingdom environment: a preliminary source inventory and budget. Environ Pollut **88**:91-108.

Wittich RM, Wilkes H, Sinnwell V, Francke W, Fortnagel P. 1992. Metabolism of dibenzo-*p*-dioxin by *Sphingomonas* sp. strain RW1. Appl Environ Microbiol **58**:1005-1010.

Xu P, Yu B, Li FL, Cai XF, Ma CQ. 2006. Microbial degradation of sulfur, nitrogen and oxygen heterocycles. Trends Microbiol **14**:398-405.

Xue W, Warshawsky D. 2005. Metabolic activation of polycyclic and heterocyclic aromatic hydrocarbons and DNA damage: a review. Toxicol Appl Pharmacol **206**:73-93.

Yang HH, Lee WJ, Chen SJ, Lai SO. 1998. PAH emission from various industrial stacks. J Hazard Mater **60**:159-174.