MICROBIAL ECOLOGY OF AROMATIC COMPOUND DEGRADATION IN A COAL TAR WASTE CONTAMINATED AQUIFER

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Jane Miyuki Yagi

May 2009

MICROBIAL ECOLOGY OF AROMATIC COMPOUND DEGRADATION IN A COAL TAR WASTE CONTAMINATED AQUIFER

Jane Miyuki Yagi, Ph.D.

Cornell University 2009

Microbial ecology links community structure, phenotypic potential, community function and interactions of microorganisms within communities and with their environments. Native microbial populations are responsible for biodegradation of naphthalene and other organic pollutants in a shallow coal tar waste-contaminated aquifer in South Glens Falls, NY. Though many details of the biochemistry and genetics of bacterial naphthalene degradation are known from studies in pure culture, molecular methods of analysis applied *in situ* can significantly advance understanding of microbial degradation processes in an ecological context.

PCR-based molecular characterization of 16S and 18S rRNA genes was used to significantly extend previous work characterizing the microbial communities in site groundwater, and to establish the temporally dynamic nature of native microbial communities. Long term natural attenuation of pollutants was documented, and the presence of members of a potentially intricate microbial food web was linked to organic contamination in the subsurface waters. A broad-range PCR assay was used to uncover a diverse suite of Rieske dioxygenase genes, including 32 previously uncharacterized clone groups, in 2 contaminated wells within the aquifer. A quantitative competitive PCR assay detecting *nah* and *nag* genes, encoding naphthalene dioxygenase, showed that these biochemically divergent pathways (associated with aerobic naphthalene con-

version to either catechol or gentisate, via salicylate) were prevalent at the site. RT-PCR results showed that both genes were transcribed *in situ*, at micromolar concentrations of ambient naphthalene, in proportions corresponding with structural gene abundance. Aerobic but not anaerobic naphthalene metabolism was observed in site water incubation experiments. However, metabolites of anaerobic naphthalene metabolism and expressed anaerobic degradation pathway mRNA transcripts (*bssA*) were detected *in situ* showing that anerobic metabolism of contaminants occurred on site. Cytochrome *c* nitrite reductase and ammonia monooxygenase gene expression *in situ* provided evidence of a complete nitrogen cycle via dissimilatory nitrate reduction to ammonia (DNRA) and nitrification carried out by native microbial communities. These data suggest that conditions in the contaminated aquifer had progressed from heterotrophic (carbon-oxidizing) to accumulation of reduced metabolic end products (ammonia, sulfide, methane) supporting lithotrophic and otherwise absent microbial populations.

Genome analysis of the aromatic hydrocarbon-degrading, facultatively chemolithotrophic, contaminated sediment-dwelling β-proteobacterium, *Polaromonas naphthalenivorans* strain CJ2, revealed a mosaic chromosome and 8 extrachromosomal elements. Comparisons with the closest sequenced relative, *Polaromonas* JS666, and related β-proteobacteria revealed both substantial homology and fluidity of genetic content. The genomic context and lateral transfer origins of the chromosomally encoded *nag*-like naphthalene catabolic operon were analyzed. Mobile genetic elements, signal transduction pathways, central and peripheral metabolic pathways, energy metabolism, transport systems, inorganic nutrient scavenging mechanisms, carbon and energy storage functions, and oxidative stress tolerance mechanisms were investigated in detail.

BIOGRAPHICAL SKETCH

Jane Yagi was born in Flushing, New York. She grew up in New Jersey and earned a B. S. in Biochemistry and History/Political Science from Rutgers College, New Brunswick, NJ in 2001.

To Matt

ACKNOWLEDGEMENTS

I am grateful to Dr. Eugene Madsen for his guidance, creativity, positive energy and his support for the big picture, as both a scientific mentor and as a friend. Chris DeRito, Graham Pumphrey and Buck Hanson helped keep days in the lab lively and productive. Life in Ithaca would not have been complete without Friday nights and the wonderful friends I have made in graduate school. My family, Noboru, Michiko, and Lisa helped provide the foundation that made this endeavor possible. Finally, I would like to thank Matthew Gaubatz for his patience and support through the ups and downs of life, big and small.

This scientific work was supported by grants from the National Institute of Environmental Health Sciences, the NSF IGERT Program in Biogeochemistry and Environmental Complexity, and a National Institute of Environmental Health Sciences Training grant.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

1.1 Introduction

The study of microbial ecology addresses four major questions: (i) how are microbial communities structured? (*i.e.*, who is there?); (ii) what is the phenotypic potential of microbial communites? (*i.e.*, what are the catalytic capabilities of the individual members and the community as a whole?); (iii) what is the community function? (*i.e.*, what are the microorganisms actually doing in the environment?); and (iv) how do the microbes interact with each other and their habitat? (Rittmann *et al.*, 2006). Microbes play singularly important roles in global biogeochemical cycles, and understanding microbial ecology is key to ensuring sustainable human interactions with the environment (Schlesinger, 1997).

Polycyclic aromatic hydrocarbons (PAHs), derived from both natural and anthropogenic sources, are toxic pollutants formed during the incomplete combustion of organic matter (Samanta *et al.*, 2002). Phylogenetically and physiologically diverse microbial populations contribute to biodegradation of PAHs in soil, sediment and groundwater (Habe and Omori, 2003; Peng *et al.*, 2008). Decades of detailed investigation in the laboratory have advanced fundamental knowledge about the biochemistry, genetics and physiology of aromatic compound metabolism in bacteria (Yen and Serdar, 1988; Harayama *et al.*, 1992; Moser and Stahl, 2001; Habe and Omori, 2003; Vaillancourt *et al.*, 2006; Peng *et al.*, 2008). Insights gained from the study of isolated metabolic pathways in pure cultures have provided a vital framework for understanding microbially mediated catalytic processes in the environment.

1.2 Aerobic naphthalene degradation in bacteria

The most widely studied bacterial degradation pathways for naphthalene, the simplest and most soluble PAH and a model compound for PAH metabolism, are initiated by the catalytic activity of an evolutionarily conserved naphthalene dioxygenase (NDO; Habe and Omori, 2003; Peng *et al.*, 2008) system. The archetypal strains harboring NDO systems are the soil isolates *Pseudomonas putida* G7 and *P. putida* NCIB 9816-4. Naphthalene-degrading pseudomonads typically oxidize salicylate to catechol, followed by *meta*-cleavage to a 2-hydroxymuconic semialdehyde intermediate, or *ortho*-cleavage (β-ketoadipate pathway) to *cis, cis*-muconate. However, an alternative NDO-mediated degradation pathway has been described for certain members of the family *Comamonadaceae*, including *Ralstonia* sp. strain U2 (Fuenmayor *et al.*, 1998; Zhou *et al.*, 2001) and *Polaromonas naphthalenivorans* strain CJ2 (Jeon *et al.*, 2006; Pumphrey and Madsen, 2007). This pathway is initiated by an evolutionarily conserved initial NDO system, but in biochemical contrast, follows with conversion of salicylate to gentisate (2,5-dihydroxybenzoate; see also Habe and Omori, 2003; Peng *et al.*, 2008). Both pathways ultimately yield Krebs cycle intermediates (pyruvate and fumarate, or succinate and acetyl-CoA, respectively).

Diverse bacteria harboring equally divergent NDO systems include Gram-negative species [e.g., *Pseudomonas* spp., *nah* genes (Dennis and Zylstra, 2004); *Burkholderia* spp., *phn* genes (Laurie and Jones, 1999); *Ralstonia* and *Comamonas* spp., *nag* genes (Zhou *et al.*, 2001; Moser and Stahl, 2001)] and several Gram-positive species [e.g., *Mycobacteria* spp., *nid* genes (Khan *et al.*, 2001); *Rhodococcus* spp., *nar* genes (Kulakov *et al.*, 2005)].

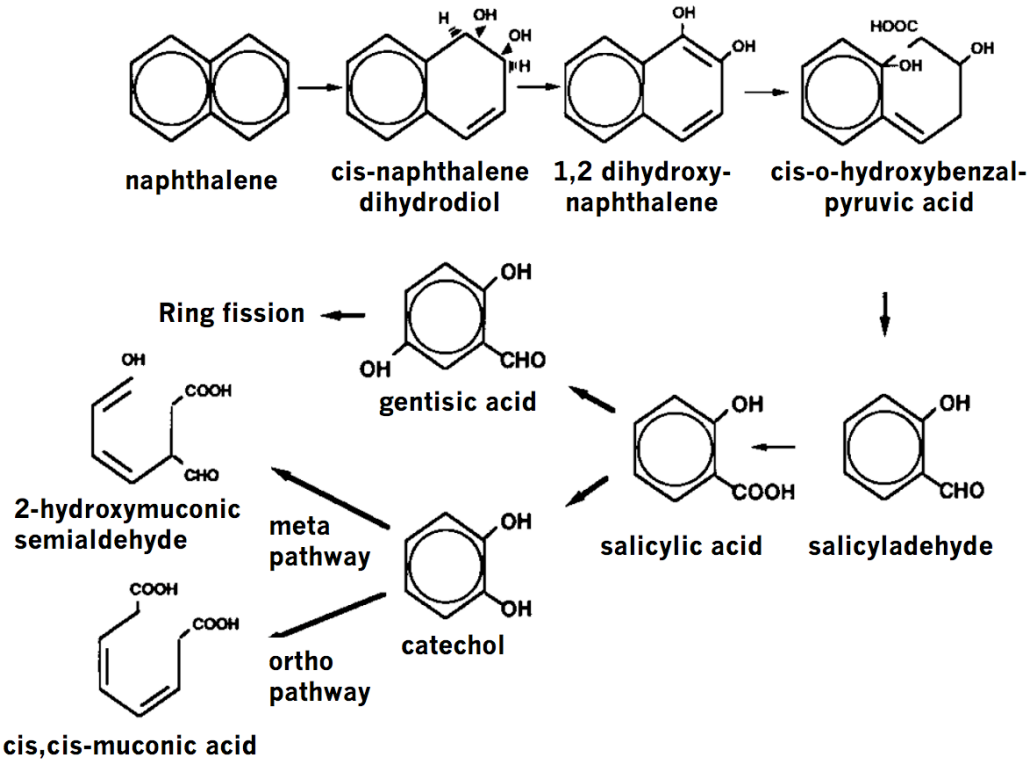


Figure 1.1: Archetypal NDO-mediated naphthalene degradation pathways

The influence of environmental factors on microbial hosts of environmentally important catabolic pathways *in situ* and global regulation of aromatic catabolism is complex and poorly understood (Bott and Kaplan, 2002; Dos Santos *et al.*, 2004; Rediers *et al.*, 2005). Naphthalene degradation increases cellular demand for oxygen, iron and sulfur (Cases and de Lorenzo, 2005). It is known that biodegradative gene regulation may be highly integrated with general cellular metabolism and stress responses (e.g., TOL pathway; Velázquez *et al.*, 2006). Expression of naphthalene degradation pathways may depend on global signals of physiological state in addition to concentrations of substrates and metabolites, as observed for benzoate degradation pathways in *Burkholderia xenovorans* strain LB400 (Denef *et al.*, 2005).

1.3 Microbial ecology of contaminated environments

Microorganisms play key roles in global biogeochemical cycles. For instance, microbial processes important in regulating nitrogen flux in the environment include nitrogen fixation (conversion of atmospheric N_2 into NH_3, ammonia oxidation (conversion of NH_4^+ to NO_3^-), assimilatory nitrate reduction (conversion of NO_3^- to NH_4^+), nitrogen mineralization (conversion of organic N to inorganic N) and denitrification (conversion of NO_3^- to NO, N_2O and N_2).

The ability of microbial communities to degrade contaminants in the environment is dependent on the complex appropriation of nutrients, electron acceptors and other conditions including pH, redox potential, moisture and temperature. Metabolic cooperation is often an important feature of microbial communities (Pelz *et al.*, 1999). It is, therefore, exceedingly difficult to capture relevant field conditions in laboratory settings (Thompson *et al.*, 2005)

The large majority of microbes in the environment are uncultivated (Rappé and Giovannoni, 2003). Environmental genomic DNA sequencing, pioneered by Norman Pace and colleagues (Pace *et al.*, 1985; Olsen *et al.*, 1986; Pace, 1997), has therefore been vital to the characterization of natural microbial communities and identification of microorganisms involved in key processes. Furthermore, microbial sequence dependent technologies for characterization of complex microbial communities include molecular fingerprinting techniques (e.g., T-RFLP, DGGE), fluorescence in situ hybridization (FISH), DNA microarrays, and DNA/RNA stable isotope probing. Burgeoning technologies in metagenomic, metatranscriptomic and various other "omic" analyses promise a future of faster, more comprehensive characterization of complex microbial systems

(Wilmes *et al.*, 2009). In spite of significant advances in large scale sequencing technologies, insightful use of these methods in many natural environments, particularly those with more than a few dominant species, remains limited by the high variability of sequences in largely uncultivated microorganisms, and the inherent incompleteness of genomic data obtained from these environments due to the sheer complexity of communities (Vieites *et al.*, 2009). In this light, standard phylogenetic and functional gene targeted PCR based molecular ecology techniques are critically important for characterization of complex natural environments.

1.4 History and description of the field study site: Site 24

Prior to the development of efficient long-distance natural gas distribution systems, energy needs (e.g., lighting and heating) were locally supported by gas produced from coal, coke, or oil at manufactured gas plant (MGP) facilities (Taylor *et al.*, 1996). MGP activity throughout the 1800s to 1950s resulted in the production and accumulation of large volumes of process by-products, particularly coal-tar waste, throughout the United States (Taylor *et al.*, 1996). Site 24, in South Glens Falls, NY, houses a shallow aquifer contaminated by coal tar waste generated by a nearby MGP facility operated between 1850 and 1940 (Taylor *et al.*, 1996). Between 4,000 and 16,000 gallons of MGP tar were buried at Site 24 in the early 1960s. The site was left undisturbed for 2 decades before Niagara Mowhawk Power Corporation (NMPC) initiated preliminary evaluation work in 1983 (Taylor *et al.*, 1996). Comprehensive research of contaminant fate at Site 24 was organized by Electric Power Research Institute (EPRI) and NMPC in 1987, and the area has since been the subject of extensive, continuous field-scale

study (Taylor *et al.*, 1996). Coal-tar waste PAHs are the major pollutant, with naphthalene as the primary component of contaminants. Active biodegradation processes are evidenced at this field site by direct isolation from groundwater of both an intermediary metabolite of naphthalene metabolism and mRNA transcripts encoding enzymes implicated in napthhalene degradation (Wilson and Madsen, 1996; Wilson *et al.*, 1999). Molecular methods targeting the diversity of microbial naphthalene catabolic systems suggest the coexistence and habitat-specific selection of a variety of different naphthalene degradation genotypes *in situ* (Wilson *et al.*, 2003). Stable isotope probing analysis led to the identification of an active, indigenous naphthalene degrading microorganism, *Polaromonas naphthalenivorans* str. CJ2 (Jeon *et al.*, 2003).

1.5 Genomics for environmental microbiology

Whole-genome sequencing was pioneered by Sanger *et al*. (1977), with sequencing of the bacteriophage ϕX174. Eighteen years later, Fleischmann *et al*. published the first complete microbial genome sequence, for *Haemophilus influenzae* (Fleischmann *et al.*, 1995). Today, with large scale sequencing technologies advancing at an unprecedented rate (Hughes Martiny and Field, 2005), complete genome sequences are publicly available for 786 bacterial strains, and nearly 1500 additional bacterial genome sequencing projects are in current progress (http://www.ncbi.nlm.nih.gov/genomes). Moreover, 56 fully sequenced archaeal genomes are available (with 47 more in progress), and more than 500 eukaryote genome projects have been completed or are in progress.

In addition to yielding valuable knowledge pertaining to human and ecosystem health, sequencing of whole genomes provides detailed information on

the phenotypic potential of biocatalytically and biogeochemically important microorganisms. However, fewer than a third of publicly available bacterial genome sequences (705 out of 2249 total) come from non-pathogenic species. Key genome-based insights on the catalytic capabilities, environmental adaptations and evolutionary mechanisms found in in non-pathogenic environmental microbes attest to the significance of genomic analyses. For example, complete genome analysis of the Gram-negative, polychlorinated biphenyl (PCB)-degrader, *Burkholderia xenovorans* LB400, revealed high genomic plasticity and recent lateral gene transfer within the genus *Burkholderia*, evolution through gene duplication and repeated gene acquisition, and tremendous variety in its potential to metabolize aromatic compounds (Chain *et al.*, 2006). In contrast, the genome of another PCB-degrader, the Gram-positive strain *Rhodococcus* sp. RHA1, suggested infrequent gene duplication and horizontal gene transfer in the evolution of its diverse metabolic capacity. The complete genome sequence of *Cupriavidus necator* JMP134, a model biodegrader of chloroaromatic compounds, enabled a detailed and systematic metabolic reconstruction of catabolic pathways in this metabolically versatile microbe (Perez-Pantoja *et al.*, 2008). Genome analysis of the marine hydrocarbon degrader, *Alcanivorax borkumensis*, revealed details on nutrient-scavenging strategies and stress response mechanisms relevant to life in an oligotrophic habitat (Schneiker *et al.*, 2006; Reva *et al.*, 2008). Complete genome sequencing paves the way for functional genomic studies (e.g., Reva *et al.*, 2006) and systems-level analyses (e.g., Fredrickson *et al.*, 2008) of key microorganisms.

Recent advances in environmental genomics include the development of public resources for comparative analysis and annotation of publicly available sequenced genomes. The current release of the DOE-Joint Genome Institute's

IMG system (available at http://img.jgi.doe.gov/; IMG 2.7) contains a total of 1155 finished and draft bacterial genomes, 56 archaeal, 40 eukaryotic genomes, 932 plasmid sequences and 2387 viruses (Markowitz *et al.*, 2008). The IMG system features an intuitive user interface and offers tools for standardized analysis and annotation of genomes, genes and functions, both individually and in comparisons with other genomes (Markowitz *et al.*, 2008).

1.6 Scope of dissertation

The goal of this work was to use molecular tools, geochemical analyses, and genomic investigation to gain insight on microbial community structure, phenotypic potential, function, and interaction with environmental factors in a coal tar waste-contaminated aquifer.

Chapter 2 is a study of the long-term dynamic nature of a coal-tar waste contaminated site and its microbial community. Sixteen years of chemical monitoring data tracking the natural attenuation of organic contamination (naphthalene, xylenes, toluene, 2-methyl naphthalene, acenaphthylene) are reported. Small-subunit rRNA gene clone libraries were constructed for *Bacteria*, *Archaea* and *Eukarya*. The study presents evidence for (i) long term biologically mediated contaminant degradation; (ii) temporal variability in microbial community composition; and (iii) support for a eukarytic food chain in response to the contamination.

Chapter 3 covers (i) the characterization of Rieske dioxygenase gene diversity in the contaminated aquifer at Site 24; (ii) the assessment, using quantative competitive PCR, of the abundance and temporal variation in two key naphtha-

lene degradation genotypes, designated *nah* and *nag* in native microbial communities; and (iii) the detection of *in situ* expression of *nah* and *nag* genes corresponding well with structural gene abundances.

Chapter 4 describes the measurement of biomarkers (metabolites of anaerobic naphthalene metabolism and expressed mRNA transcripts) extracted from groundwater to document *in situ* aerobic and anaerobic transformations of ammonia, nitrate and aromatic contaminants at Site 24.

Chapter 5 presents the analysis of the complete genome sequence of *P. naphthalenivorans*. Genomic comparisons with related organisms, particularly *Polaromonas* JS666, were used to gain insight on the evolution of genomic potential in strain CJ2. A revised version of Chapter 5 has been accepted for publication in *Environmental Microbiology*.

The concluding chapter comprises previous chapter summaries and conclusions.

CHAPTER 2

SUBSURFACE ECOSYSTEM RESILIENCE: LONG-TERM ATTENUATION OF SUBSURFACE CONTAMINANTS SUPPORTS AN ELABORATE EUKARYOTIC FOOD CHAIN AND A DYNAMIC MICROBIAL COMMUNITY[1]

[1]Jane M. Yagi, Edward F. Neuhauser, John A. Ripp, David M. Mauro and Eugene L. Madsen.

2.1 Abstract

The propensity for groundwater ecosystems to recover from contamination by organic chemicals (in this case coal-tar waste) is of vital concern for scientists and engineers managing polluted sites. Microbially mediated cleanup processes are also of interest to ecologists because they are an important mechanism for the resilience of ecosystems. Here we establish the long-term dynamic nature of a coal-tar waste-contaminated site and its microbial community. We present 16 years of chemical monitoring data tracking responses of a groundwater ecosystem to organic contamination (naphthalene, xylenes, toluene, 2-methyl naphthalene, acenaphthylene) associated with coal-tar waste. In addition, we analyzed small-subunit rRNA genes from two contaminated wells at multiple time points over a 2-year period. Principle component analyses of community rRNA fingerprints (T-RFLP) showed that the composition of native microbial communities varied temporally, yet remained distinctive well-to-well. After screening and analysis of 1372 cloned small-subunit rRNA genes from *Bacteria*, *Archaea*, and *Eukarya*, we discovered that the site supports a robust variety of eukaryotes [e.g., Alveolates (especially anaerobic and predatory ciliates), Stramenopiles, Fungi, even the small metazoan flatworm, *Suomina*] absent from an uncontaminated control well. The study links the dynamic microbial composition of a contaminated site to long-term attenuation of its subsurface contaminants.

2.2 Introduction

The growing impact of human activities on ecosystems is well recognized (Palmer *et al.*, 2004; Liu *et al.*, 2007; Dietz *et al.*, 2003). Resilience has been identified as a property of ecosystems useful in predicting their response to anthropogenic change (Dietz *et al.*, 2003; Folke *et al.*, 2004). Over the last 4 decades, ecosystem resilience has been variously defined as the magnitude of disturbance that a system can experience before it shifts into a different state (Holling, 1973) and the capacity of a system to absorb disturbance and reorganize (Folke *et al.*, 2004). Yet the microbiological processes that routinely underlie ecosystem resilience are often ill defined and unrecognized (Falkowski *et al.*, 2008; Madsen, 2008). In the classic view of aquatic ecosystem resilience (reviewed in Folke *et al.*, 2004), an excess of anthropogenically released materials (for example, phosphorus) may trigger a series of events that convert a clear-water lake (oligotrophic, high oxygen, low plant biomass, dominated by game fish) to a turbid eutrophic water body (low oxygen, high plant biomass, with game fish absent). The capacity of a lake system to receive phosphorus inputs and remain in the clear-water state is its resilience. The pair of key biogeochemical processes forcing the transition from oligotrophic to eutrophic states are photosynthesis (by plants and phototropic microorganisms, no longer limited by phosphorus) and aerobic respiration (by heterotrophic microorganisms that cause anoxia by using deceased phototrophic biomass as a physiological electron donor and O_2 as electron acceptor). In the case of phosphorus release to lakes, microbiological processes overwhelm system resilience, leading to an oxygen-depleted condition, and to accompanying shifts in habitat colonization by higher life forms (Folke *et al.*, 2004). Unlike lakes, the majority of groundwater ecosystems are inhabited ex-

clusively by microorganisms - especially, *Bacteria*, *Archaea*, and unicellular *Eukarya* (Madsen and Ghiorse, 1993; Chapelle, 2003). Thus, habitat colonization by higher life forms (e.g., fish) is moot. When groundwater habitats are impacted by organic contaminants, microbiological reactions can directly destroy contaminants, thereby contributing directly to system resilience.

Ecosystem disturbance by industrial chemical waste is a widespread problem (National Research Council, 2000; Schwarzenbach *et al.*, 2006) and the response of naturally occurring microbial communities to inputs of organic compounds remains an active area of research (Alexander, 1999; Diaz, 2008). The capabilities of naturally occurring microbial communities to enzymatically attack (biodegrade) and grow upon many contaminant compounds (Alexander, 1999; Madsen, 2008) is the basis for a site clean-up technology, termed "Intrinsic Bioremediation" and "Monitored Natural Attenuation" (MNA) by the USEPA and the US National Academy of Sciences (National Research Council, 2000; U.S. Environmental Protection Agency, 1997). Natural attenuation is a manifestation of microbially mediated ecosystem resilience. Although many field studies of MNA have been reported in the past for a variety of chemical pollutants (e.g., Bekins *et al.*, 2001; Kleikemper *et al.*, 2002; Griebler *et al.*, 2004; Hunkeler *et al.*, 2005; Dojka *et al.*, 1998; Ulrich *et al.*, 2003; Suarez and Rifai, 2002; Cozzarelli *et al.*, 2001), to our knowledge, none have comprehensively integrated (i) the long-term field evidence for contaminant elimination; (ii) temporal variability in microbial community composition; and (iii) the composition of an elaborate eukaryotic microbial food chain responding to the contamination. A companion manuscript (Yagi *et al.*, 2009b) used a combination of biomarkers (metabolites and mRNA of transcribed functional genes) to establish that the subsurface microbial community native to the contaminated study site was engaged in an-

aerobic metabolism of aromatic hydrocarbons and complete biogeochemical N cycling between nitrate and ammonia.

2.2.1 Site timeline and prior investigations

In the 1960s, coal-tar waste from a manufactured gas plant was buried in South Glens Falls, NY, resulting in a contaminant plume of polycyclic aromatic hydrocarbons and other monocyclic constituents, dispersed by groundwater flow (Fig. 2.1; Electric Power Research Institute, 1996; Madsen *et al.*, 1991; Nelson *et al.*, 1996; Murarka *et al.*, 1992; Neuhauser *et al.*, 2008). *In situ* biodegradation of contaminants was demonstrated by the enhanced metabolism of naphthalene

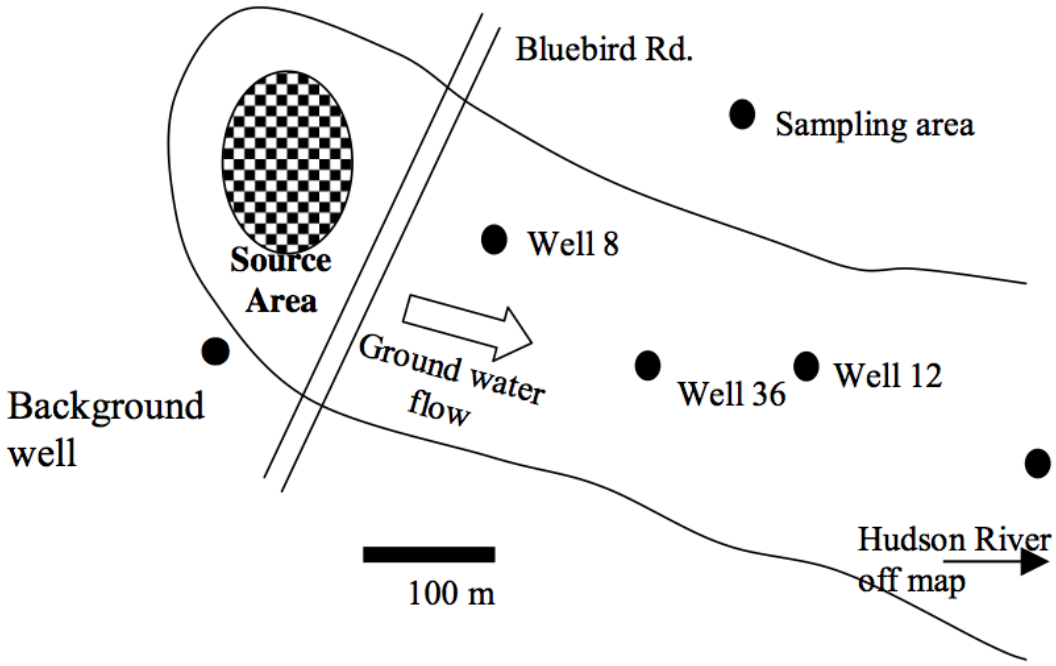


Figure 2.1: Plan view of the coal tar waste-contaminated site in Glens Falls, N.Y., showing locations of the monitoring wells.

and phenanthrene in sediment samples from inside the plume, but not outside (Madsen *et al.*, 1991). Furthermore, the high biomass of protozoa that occurred only in the contaminated zone proved that bacterial prey were actively growing on contaminants (Madsen *et al.*, 1991). In 1991, ~7000 m^3 of sediment containing the main mass of coal tar waste was excavated from the site and the cavity was back filled with clean sand (Murarka *et al.*, 1992), followed by natural attenuation (via solubilization, transport, sorption, dilution and microbial metabolism) of aromatic hydrocarbons that had been transported to down-gradient sediments (U.S. Environmental Protection Agency, 1997; National Research Council, 2000). Detection of both the transient intermediary metabolite, 1,2-dihydroxy-1-2-dihydronaphthalene (Wilson and Madsen, 1996) and mRNA transcripts of the naphthalene dioxygenase gene (*nahAc*) (Wilson *et al.*, 1999; Bakermans and Madsen, 2002a), as well as stable-isotope probing analyses (Jeon *et al.*, 2003) confirmed on-going aerobic microbial metabolism of naphthalene, the major soluble component of the coal-tar waste (Nelson *et al.*, 1996). However, oxygen depletion in the contaminated zone and clear evidence for methanogenesis and sulfate reduction (strictly anaerobic processes incompatible with aerobic naphthalene degradation; see Bakermans *et al.*, 2002; Bakermans and Madsen, 2002b) signify that on-site metabolism is limited by the influx of geochemical oxidants, especially naturally occurring nitrate, sulfate and dissolved oxygen delivered largely by flow from uncontaminated groundwater upgradient into the system (Bakermans *et al.*, 2002). A 16S rRNA-based clone library (100 clones from each of 4 wells across the sites contamination gradient) provided an initial characterization of the on-site bacterial community (Bakermans and Madsen, 2002b).

2.3 Materials and Methods

2.3.1 Site and groundwater sampling

The contaminated site is a rural area in South Glens Falls, NY, where coal tar waste was buried in the early 1960s (Fig. 2.1; Madsen *et al.*, 1991; Murarka *et al.*, 1992; Wilson *et al.*, 1999). Groundwater samples were collected from two monitoring wells located along the ground water flow path and a background control well upgradient of the source contamination. Two commercial firms (GEI, Inc. (Glastonbury, CT) and META Environmental Inc. (Watertown, MA) conducted annual sampling and analysis of the well waters for oxygen and organic contaminants from 1989 to 2005 as previously described (Murarka *et al.*, 1992).

For molecular and microbiological analyses, groundwater was collected between 2005 and 2007, as described by Bakermans and Madsen (2002b). Briefly, water from monitoring wells 12, 36, and the background well was gathered at a flow rate of 300 ml/min using a Geopump-2 (Geotech Environmental, Denver, CO). For each sample, a minimum of 3 well volumes were pumped prior to filtering cells from ~3 liters of water through 0.22 μm pore size, 142 mm diameter Durapore membranes (Millipore Corp., Bedford MA) that were then immediately frozen in sterile Whirlpak® bags (NASCO, Modesto CA) on site (dry ice/ethanol). Filters were stored at -80 °C preceding the analyses described below.

2.3.2 Nucleic acid extraction

For DNA and RNA extraction from groundwater biomass from frozen filters, previously described protocols (Wilson *et al.*, 1999; Bakermans and Madsen, 2002b) were followed, with minor modifications. Briefly, for DNA extraction, filters were crushed and extracted two times each by boiling (5 min) with 5 ml pre-heated extraction buffer (1% SDS, 0.1 M NaCl, 10 mM Tris, 1 mM EDTA, pH 8.0). Decanted extracts were added to an equal volume of phenol (pH 8.0) and mixed vigorously followed by centrifugation at 4 °C (15 min at 10,000 x g). The upper aqueous layer was extracted with an equal volume of phenol:chloroform (1:1), followed by centrifugation as above and a final extraction with an equal volume of chloroform:isoamyl alcohol (24:1) and centrifugation. DNA was precipitated from the final aqueous layer at -80 °C with sodium acetate (0.3 M) and an equal volume of isopropanol. After centrifugation (30 min at 15,000 x g), DNA pellets were washed in 70% ethanol and reprecipitated. Pellets were then air-dried and carefully resuspended in 50 μl TE (pH 8.0).

For RNA processing, all reagents and labware were either certified RNase free, baked overnight at 200 °C, and treated with RNase ZAP® (Ambion, Austin, TX) or diethyl pyrocarbonate-treated (DEPC) for inactivation of ribonuclease activity as appropriate. RNA was extracted from frozen filters using a modified version of the DNA extraction protocol described above. Briefly, crushed filter pieces were processed using acid extraction buffer (pH 5.1), and acidic phenol (pH 4) was used for phenol and phenol:chloroform:isoamyl alcohol (125:25:1) extraction. RNA was precipitated from aqueous extracts with glycogen as co-precipitant (2 μg/ml) at -80 °C using sodium acetate (2.5 M) and 2 volumes of ethanol. Purified RNA pellets were recovered by centrifugation (30 min at

15,000 x g), air-dried, and resuspended in 50 μl RNase-free water. DNA was removed from RNA extracts by DNase I treatment (Invitrogen), and RNA was reverse-transcribed to cDNA using SuperScript III™ reverse transcriptase (RT; Invitrogen) and random hexamers according to the manufacturers instructions. No-RT control PCR reactions were performed using primers and PCR conditions described below to ensure complete removal of contaminating DNA.

2.3.3 rRNA gene clone libraries

Representative DNA clone libraries were generated from selected groundwater extracts (see Results). PCR amplification of prokaryotic and eukaryotic small subunit rRNA was carried out using the following primer pairs: 27f/1492r targeting *Bacteria* (Lane, 1991), 21Fa/1492R targeting *Archaea* (DeLong, 1992), and both 3Fphp/1749Rphp (Richards *et al.*, 2005) and 360Fe/1391Re (Dawson and Pace, 2002) targeting *Eukarya*. PCR amplification was performed using ThermoStart™ DNA polymerase (ABgene®) on a PTC-200 DNA Engine thermocycler (MJ Research). Duplicate reactions were performed for each clone library using 0.5-25 ng DNA template in 25 μl volumes and previously described cycling conditions (Liou *et al.*, 2008; DeLong, 1992; Richards *et al.*, 2005; Dawson and Pace, 2002). Pooled PCR products were examined by agarose gel electrophoresis, purified using the QIAquick® Gel Extraction Kit (Qiagen) and ligated into the vector pCR2.1 (TOPO® TA cloning, Invitrogen) or pSC-A (StrataClone™ PCR cloning, Stratagene) following the manufacturers recommended protocols. Following transformation of plasmids into host cells and blue/white screening, inserts in 80-130 randomly picked colonies were verified by PCR with vector-specific primers that flanked the cloning regions. The amplicons were digested

with HaeIII and HhaI (New England Biolabs) and sorted by restriction fragment length polymorphism (RFLP) patterns on 3% MetaPhor® agarose gels (BioWhittaker; Molecular Applications, Rockland, Maine). Clones containing unique RFLP patterns were selected for sequencing, grown overnight in 3 ml of Luria-Bertani broth with kanamycin (50 μg/μl), pelleted, and plasmids were purified (QiaPrep® spin miniprep kit; Qiagen, Santa Clarita, Calif.). Sequencing (Cornell University Life Sciences Core Laboratories Center) was routinely conducted with M13 forward and reverse primers. Sequences were checked for quality using 4Peaks software (http://www.mekentosj.com/4peaks). SSU rRNA sequences were checked for chimeras using Bellerophon with the Huber-Hugenholtz correction and a 300-bp window size (Huber *et al.*, 2004) and Pintail (Ashelford *et al.*, 2005). Chimeras, vector sequences, and sequences of poor quality were excluded from further phylogenetic analyses. 16S and 18S gene sequences were aligned using the ARB FastAligner tool to the most recent GreenGenes (DeSantis *et al.*, 2006) and SILVA (Pruesse *et al.*, 2007) databases, respectively, running locally in the ARB package (Ludwig *et al.*, 2004). A 50% base frequency filter was used to omit highly variable regions of the SSU rRNA. Alignments were manually edited with ARB Edit4. Sequences comparisons against the GenBank nucleotide database library were carried out via BLAST on-line searches (http://ncbi.nlm.nih.gov/BLAST) and closest BLAST matches were included in phylogenetic comparisons. Phylogenetic trees were constructed in ARB using a neighbor-joining algorithm and a Olsen-corrected distance matrix.

2.3.4 Terminal restriction fragment length polymorphism analysis of community rRNA genes

Small subunit rRNA genes samples were amplified from groundwater DNA or cDNA as described above, except that PCR primer 27F-FAM (5′ end-labeled with phosphoramidite fluorochrome 5-carboxy-fluorescein) with reverse primer 1492R was used to target 16S rRNA genes and similarly, 3′ end-labeled 3FphpF-FAM with reverse primer 1749Rphp was used to amplify 18S rRNA genes. Fluorescently labeled PCR products were digested overnight with MspI (New England Biolabs), purified using the QIAquick® PCR Purification Kit (Qiagen®, USA), and electrophoretically separated in an Applied Biosystems 3730 DNA analyzer (Applied Biosystems) as described previously (Liou *et al.*, 2008). Electrophoretic data was analyzed using Genemapper and/or PeakScanner software (Applied Biosystems) using a threshold value of 50 relative fluorescence units. Representative duplicate DNA extractions from filters collected on the same day were processed to verify reproducibility. T-RF peak heights were used in principal component analysis (PCA) to compare and group community profiles (i.e., T-RFLP patterns) along axes (principal components) on the basis of the fragment patterns alone (Jongman *et al.*, 1995). PCA plots were generated using R statistical software version 2.4.1 (http://www.r-project.org/) and the vegan package version 1.8-8 (http://cc.oulu.fi/~jarioska/; Dixon, 2003). For comparison of T-RF profiles with dominant TRFs derived from environmental groundwater samples, T-RFLP analysis was performed as above on PCR amplicons from clones representing dominant phylotypes from the 16S rRNA gene clone libraries (described above).

2.3.5 Nucleotide sequence accession numbers

The nucleotide sequence data reported here have been submitted to GenBank under accession no. FJ810524 to FJ810621.

2.4 Results

2.4.1 A 16-year record of contaminant attenuation

Well-water concentrations (1989-2005) of 5 coal-tar-waste constituents (naphthalene, xylenes, toluene, 2-methylnaphthalene, acenaphthylene) and oxygen are depicted in Figure 2.2. Changes in well-water contamination since removal of the source of coal-tar waste reflect both excavation-related physical mobilization of chemicals and natural attenuation processes (U.S. Environmental Protection Agency, 1997; National Research Council, 2000). Well 12, farthest downgradient from the source material (Fig. 2.2), featured the lowest contaminant concentrations. Well 36, though further downgradient thafn well 8 (Fig. 2.2), had the highest contaminant concentrations - reflecting the presence of clay lenses that bind coal-tar-waste constituents and resist both groundwater flushing and oxygen infiltration (Murarka *et al.*, 1992; Bakermans *et al.*, 2002). Thus, subsurface matrix effects, especially sorptive surfaces and hydraulic conductivity, appear to play a major role in persistence of the groundwater contaminants. Naphthalene concentrations rose sharply in all 3 wells following source excavation (to ~1600 ppb in 1991-1992) and then gradually diminished through 2005 (Fig. 2.2). Rates of naphthalene attenuation were most rapid for well 12 and slowest for well 36, where the 2005 residual level was still approximately 380 ppb. Patterns

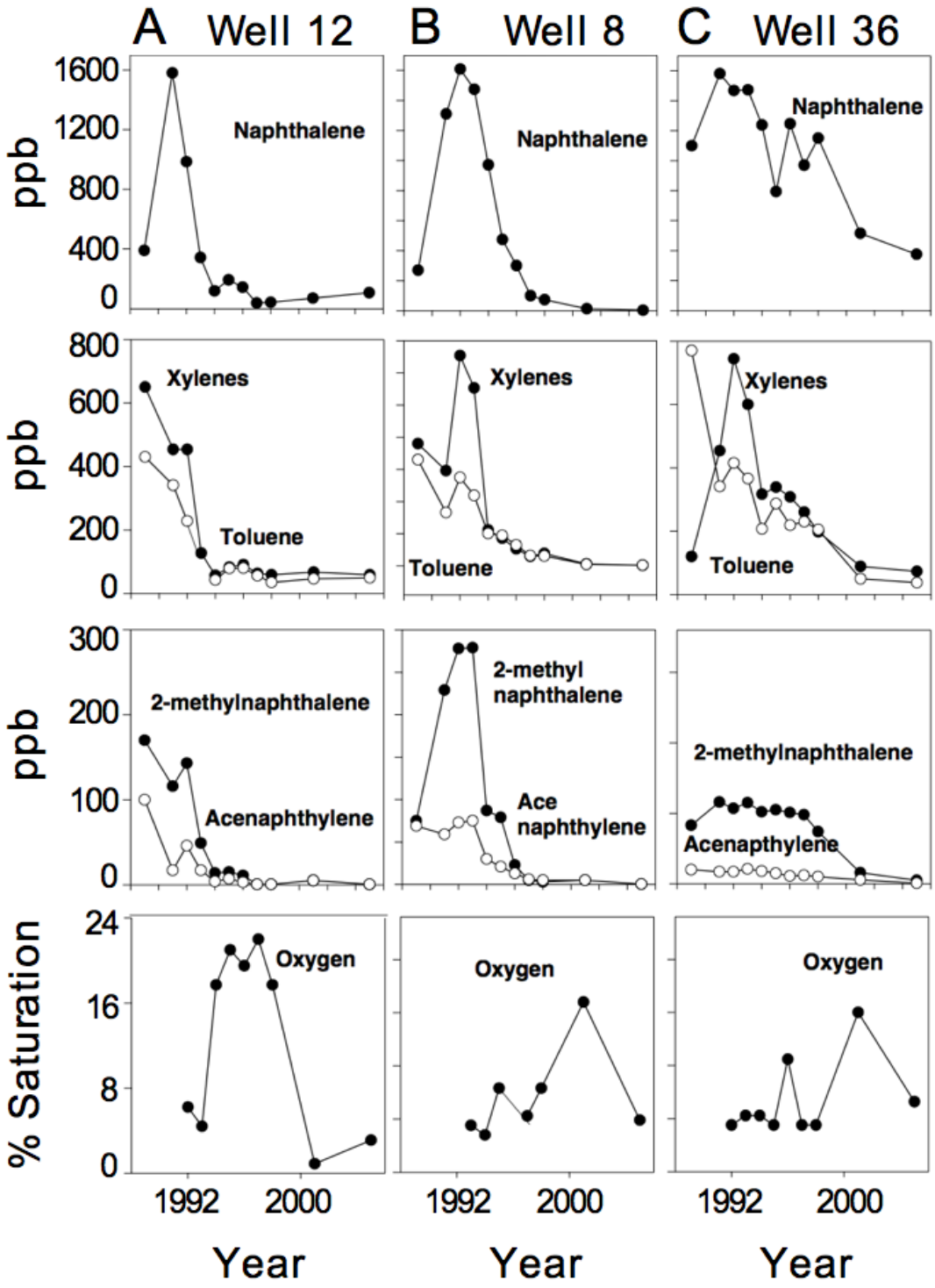


Figure 2.2: Historical trends of five site contaminants (naphthalene, xylenes, toluene, 2-methyl naphthalene, acenaphthylene), and oxygen in the groundwater of Well 12 (A), Well 8 (B), and Well 36 (C) from 1989 to 2005.

in the attenuation of xylenes, toluene, 2-methylnaphthalene, and acenaphthylene (middle panels, Fig. 2.2) were similar to those of naphthalene: diminishing concentrations with time, rapid attenuation in well 12, and slowest attenuation in well 36.

Oxygen depletion associated with contaminant-induced microbial biodegradation processes is widely recognized (National Research Council, 2000). In 1988 and 1989 the average (n=4) level of O_2 in uncontaminated water from an upgradient control well was 30% of saturated levels (~3.3 mg O_2/liter at 10 °C). Routine monitoring of well-water oxygen beginning in 1992-1993 showed that for all 3 contaminated wells, oxygen was far below the level of the background well (≤6% saturation; Fig. 2.2). Such low values are often indistinguishable from zero, given likely intrusion of atmospheric oxygen during on-site pumping and analysis. Fluctuating oxygen concentrations in all wells through to 2005 suggest that the system is geochemically highly dynamic. Replenishment of atmospheric O_2 into the groundwater occurred most markedly in well 12 (lowest contamination, during 1994 to 1998) and least markedly in well 36 (highest contamination).

2.4.2 Dynamic native communities

The on-site fluctuations in chemical groundwater constituents (Fig. 2.2; see also Yagi *et al.*, 2009b) led us to hypothesize that the native microbial communities were causing and/or responding to the geochemical changes. We expected that this would be reflected by dynamic fluctuations in microbial populations. Bacterial community profiles from wells 12 and 36 generated with terminal restriction fragment length polymorphism (T-RFLP) analysis of DNA (total commu-

nity; panels A, C, E of Fig. 2.3) demonstrate time-dependent individual fragments (e.g., 186, 201, 222, 276 only in August 2006) and periodicity (e.g., 162, 430, and 488 shared by November 2005 and 2006, but absent in August 2006). This trend was reinforced by examining reverse-transcribed RNA (metabolically active bacterial community, Panels B and D of Fig. 2.3) that showed many

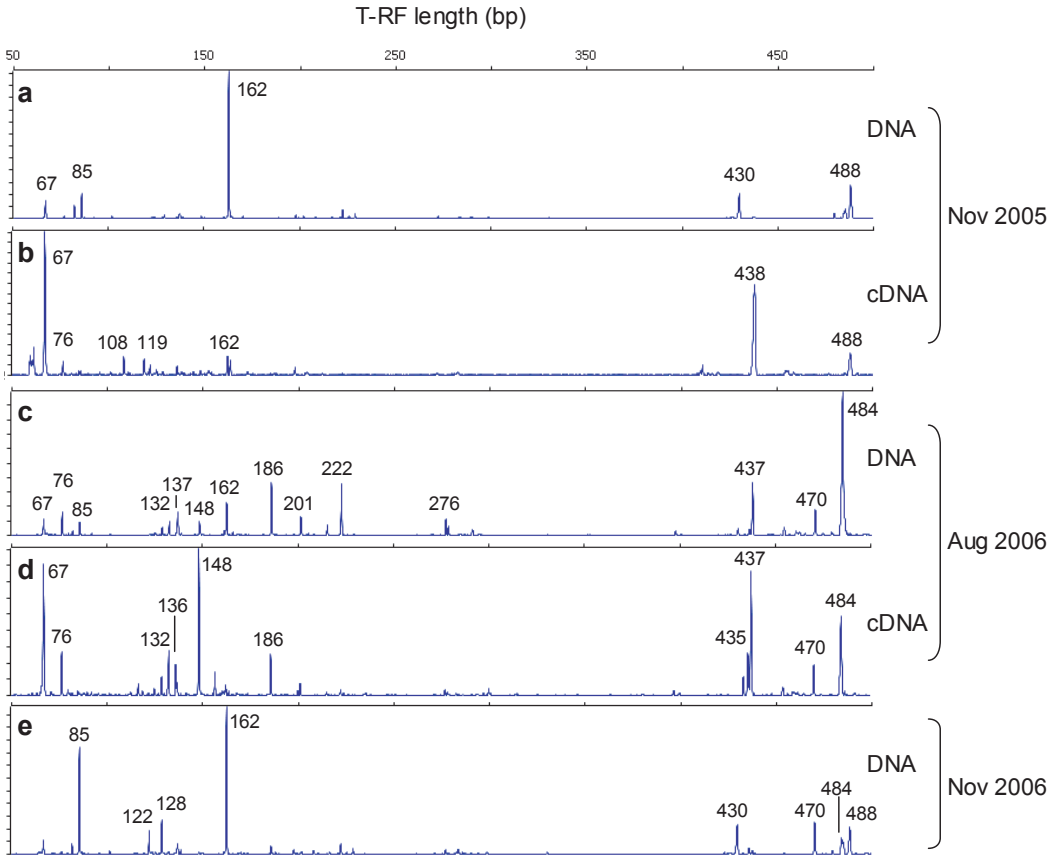


Figure 2.3: Dynamic changes in native bacterial community profiles in Well 36. DNA fingerprints (terminal restriction fragment lengths, T-RFs) generated from rRNA gene amplicons for groundwater microbial communities illustrate shifts in native bacterial communities in Well 36. The x-axis and numbers above the peaks indicate major fragment sizes (phylotypes) in DNA (A) and cDNA (B) from November 2005, DNA (C) and cDNA (D) from August 2006, and DNA (E) from samples taken in November 2006. The y-axis shows the intensity of the peaks in relative fluorescence units.

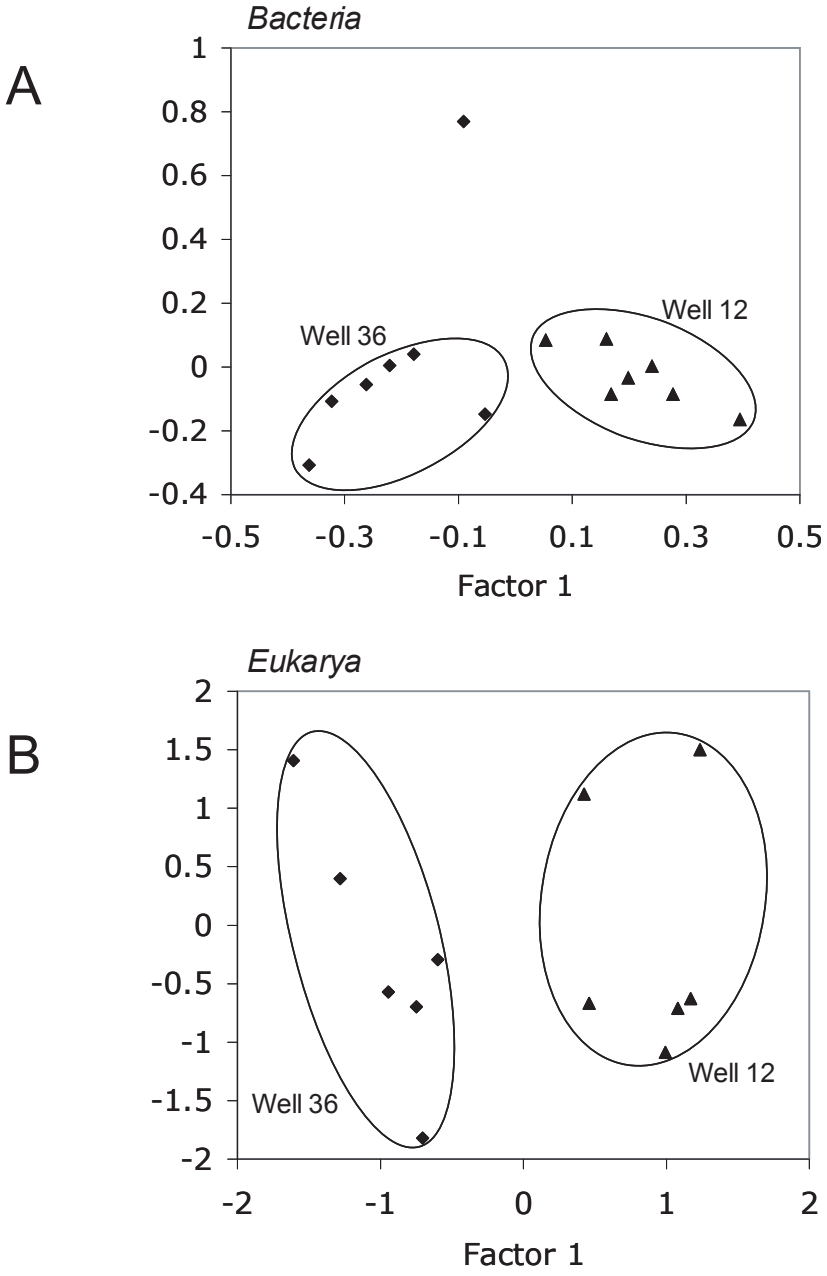


Figure 2.4: Principal components analysis showing the spatial differentiation of microbial community composition (T-RFLP fingerprints of small subunit RNA genes). Analyses for Bacteria (A) and Eukarya (B) were obtained from wells 36 and 12 (September 2005; August, November and December 2006; and May 2007). Ordination points for well 36 profiles are indicated by diamonds, and the triangles represent samples from well 12.

fragments (e.g., 108, 119, 132, 136, 148, 162, 435, 437, 438, 470, 484) that were not shared by the November 2005 and August 2006 profiles. These data clearly demonstrate the dynamic nature of the groundwater microbial communities, given that independently processed replicate samples gathered on a single day did not exhibit such variations (data not shown).

Principal component analysis (PCA) of T-RFLP patterns for bacterial and eukaryal domains reveal another aspect of the variability of site microbial communities (Fig. 2.4). During multiple site visits (spanning November 2005, August 2006, November 2006, December 2006, and May, 2007), 7 16S- and 6 18S-rRNA analyses were completed for wells 12 and 36. Over this period, considerable on-site fluctuations in geochemical (Fig. 2.2; see also Yagi *et al.*, 2009b) and population (Fig. 2.3) parameters occurred. Despite these compositional variations in time, well microbial populations remained coherent relative to one another (Fig. 2.4). Thus, despite evidence for temporal variability (Fig. 2.3), site-specific habitat characteristics associated with wells 12 and 36 appeared to be dominant factors that determined overall community composition.

2.4.3 Microbial community composition: Bacteria, Archaea, and an enriched eukaryotic food chain

To identify representatives of all three domains (*Bacteria, Eukarya and Archaea*) present in site well waters, we prepared clone libraries from genes encoding small subunit (SSU) rRNA. We screened 1372 cloned DNA sequences encoding SSU rRNA genes (295 bacterial, 442 archaeal, and 635 eukaryal) and analyzed 116 representative sequences.

Bacteria The bacterial rRNA sequences fell into 11 established phylogenetic classes (β-, δ-, ϵ-, γ-*proteobacteria*, *Acidobacteria, Actinobacteria*, D-BACT lineage, *Firmicutes, Bacteroidetes/Chlorobi*, Candidate Division OP3, and *Chloroflexi*; Fig. 2.5A). This array of *Bacteria* represents a diverse sampling of microbial life, including close relatives of cultured microorganisms and clones derived from subsurface, marine, and lake habitats. *Chlorobia* (28%), *Syntrophus* (26%) and β-*proteobacteria* (9%) dominated the clone library from well 36. The dominant well-characterized taxa in the library from well 12 were β-*proteobacteria* (18%), δ-*proteobacteria* (in particular, *Geobacter psychrophilus*-like sequences; 20%), and *Firmicutes* (18%). Two clades particularly relevant to site processes were observed: *Chlorobia*, recently associated with oxidation of elemental sulfur and sulfide to sulfate, a process likely to occur in the study site when aerobic waters infiltrate zones of prior sulfate reduction (Engel *et al.*, 2003; Hose *et al.*, 2000); and an unclassified group aligning with taxa responsible for anaerobic oxidation of methane using nitrate as an electron acceptor (lineage associated with uncultured bacterium clone D-BACT in Fig. 2.5A; Raghoebarsing *et al.*, 2006).

Archaea Both *Euryarchaeota* and *Crenarchaeota* were represented in the archaeal clone libraries (Fig. 2.5B). Group I.1a *Crenarchaeota* (most closely related to environmental clones derived from subsurface and freshwater sediments) dominated the clones from well 12 (60%), while comprising only 15% of the well 36 library. There were two prominent characteristics of the archaeal sequences: (i) the presence of sequences closely related to the marine ammonia oxidizer, *Nitrosopumilus* (Konneke *et al.*, 2005) and (ii) the absence of sequences related to the common methanogenic orders, *Methanobacteriales* and *Methanomicrobiales*.

Figure 2.5: Phylogenetic analyses of bacterial 16S (A), archaeal 16S (B) and eukaryal 18S (C) rRNA clone libraries derived from Well 36 and Well 12. Sequences from this study are marked in bold type while relevant reference sequences are not. In panel A, clone nomenclature (in bold) specifies source of gene (B, Bacteria), source well [12, 36, or both wells (C)], an arbitrary clone number, and (in parentheses) the number of clones screened by RFLP represented by the sequence. In panel B, clone nomenclature (in bold) specifies source of gene (A, Archaea), source well [12, 36, or both wells (C)], and arbitrary clone number. In panel C, clone nomenclature (in bold) specifies source of gene (Ephp, Eukarya amplified using primer pair 3Fphp/149Rphp or EfeRe, Eukarya amplified using primer pair 360Fe/1391Re), source well [12, 36, or both wells (C)], and arbitrary clone number. Notable bacterial and eukaryal lineages are described in the text. Archaeal sequences clustered with Group I.1a Crenarchaeota and the Deep-sea Hydrothermal Vent Euryarchaeota (DHVE) group II lineages DHVE4 and DHVE5. Trees were determined by neighbor-joining analysis with Olsen correction implemented in the ARB analysis package using a 50% base frequency filter to omit highly variable regions. Bootstrap values $\leq$50% from 100 replicates are indicated next to the nodes. GenBank accession numbers for reference sequences follow the clone or organism name. The scale bar represents the expected number of changes per nucleotide position.

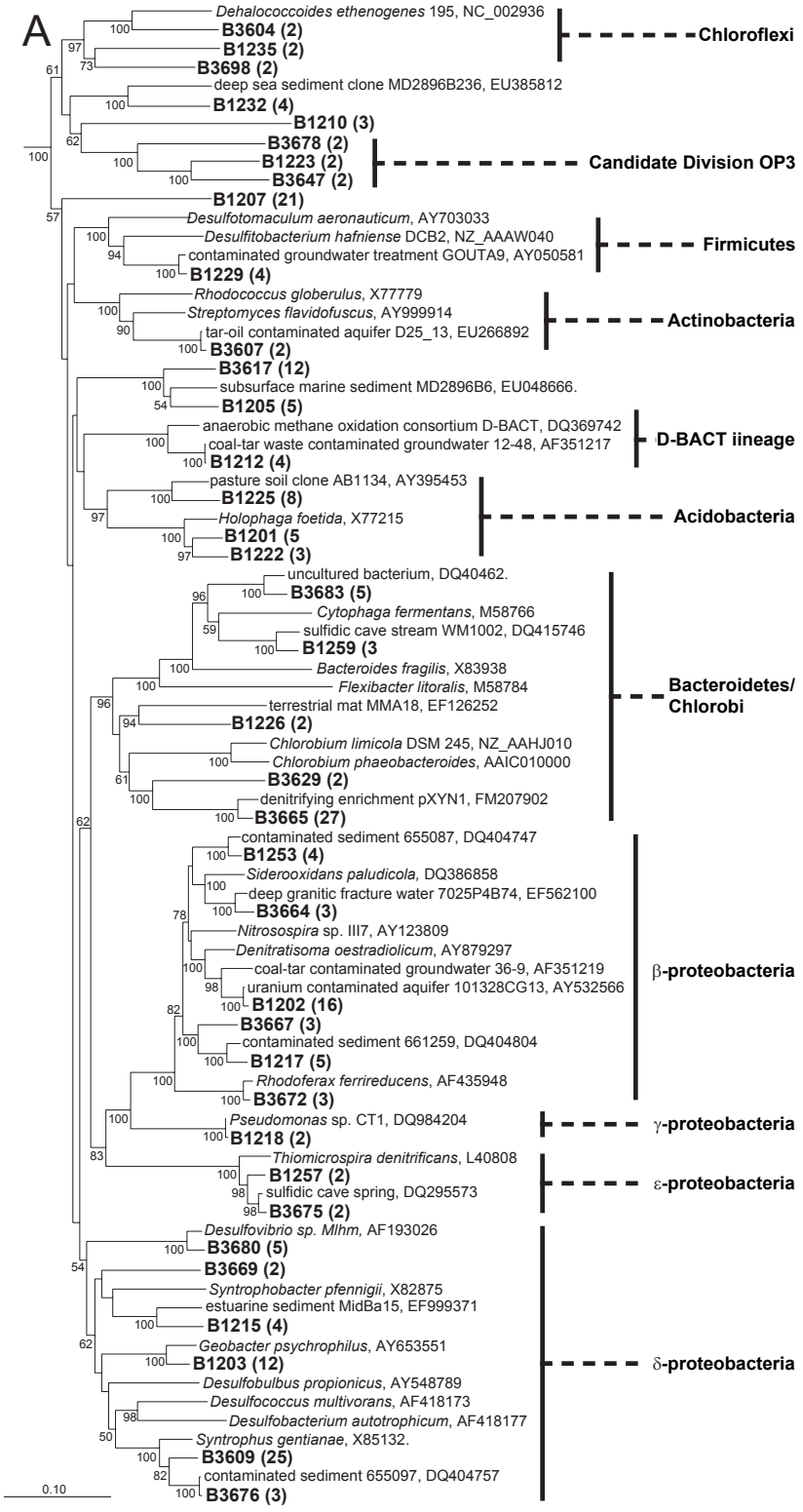
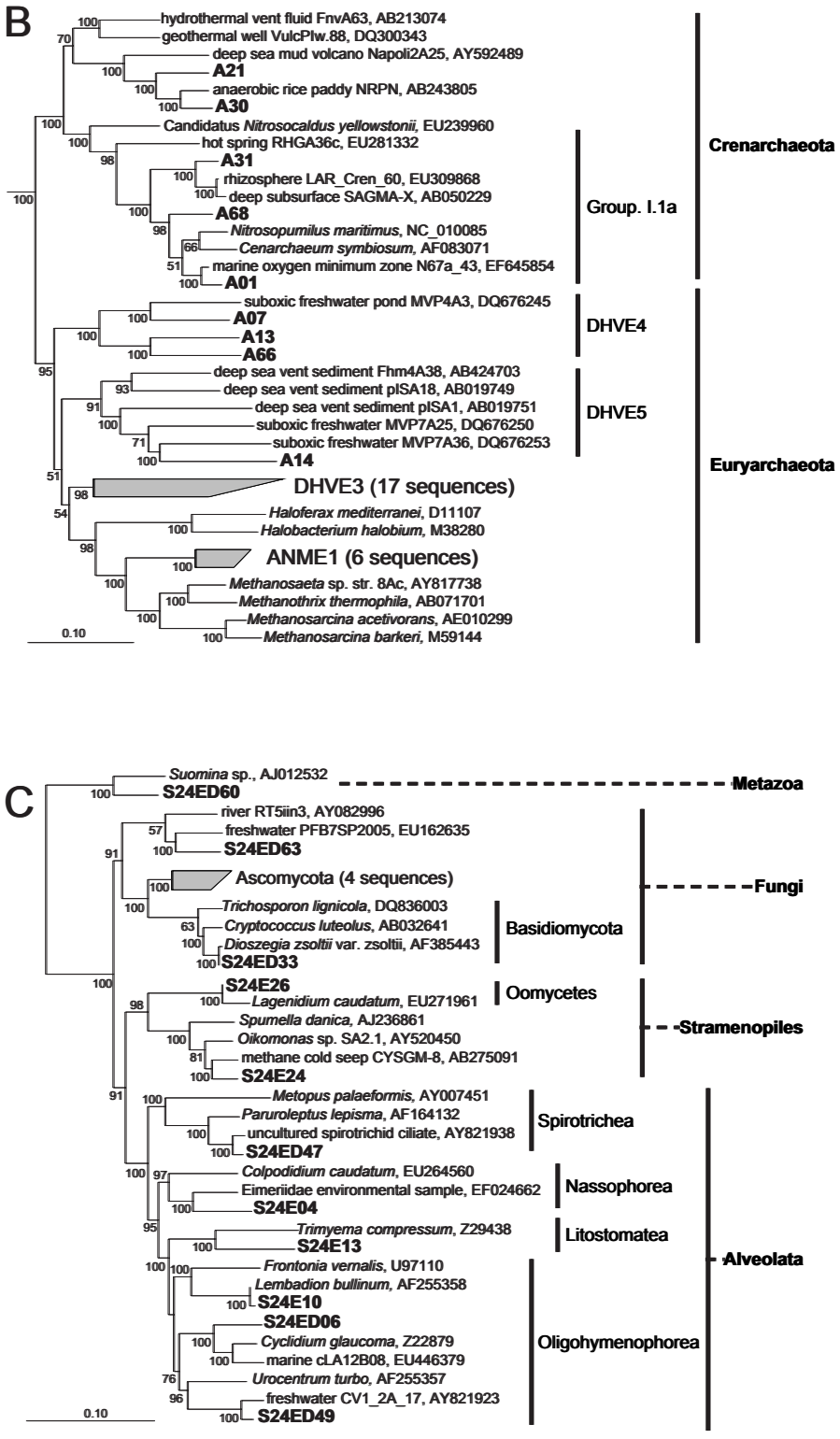

A
Dehalococcoides ethenogenes 195, NC_002936
B3604 (2)
B1235 (2)
B3698 (2)
Chloroflexi
deep sea sediment clone MD2896B236, EU385812
B1232 (4)
B1210 (3)
B3678 (2)
B1223 (2)
B3647 (2)
Candidate Division OP3
B1207 (21)
Desulfotomaculum aeronauticum, AY703033
Desulfitobacterium hafniense DCB2, NZ_AAAW040
contaminated groundwater treatment GOUTA9, AY050581
B1229 (4)
Firmicutes
Rhodococcus globerulus, X77779
Streptomyces flavidofuscus, AY999914
tar-oil contaminated aquifer D25_13, EU266892
B3607 (2)
Actinobacteria
B3617 (12)
subsurface marine sediment MD2896B6, EU048666.
B1205 (5)
anaerobic methane oxidation consortium D-BACT, DQ369742
coal-tar waste contaminated groundwater 12-48, AF351217
B1212 (4)
D-BACT iineage
pasture soil clone AB1134, AY395453
B1225 (8)
Holophaga foetida, X77215
B1201 (5
B1222 (3)
Acidobacteria
uncultured bacterium, DQ40462.
B3683 (5)
Cytophaga fermentans, M58766
sulfidic cave stream WM1002, DQ415746
B1259 (3
Bacteroides fragilis, X83938
Flexibacter litoralis, M58784
terrestrial mat MMA18, EF126252
B1226 (2)
Chlorobium limicola DSM 245, NZ_AAHJ010
Chlorobium phaeobacteroides, AAIC010000
B3629 (2)
denitrifying enrichment pXYN1, FM207902
B3665 (27)
Bacteroidetes/
Chlorobi
contaminated sediment 655087, DQ404747
B1253 (4)
Siderooxidans paludicola, DQ386858
deep granitic fracture water 7025P4B74, EF562100
B3664 (3)
Nitrosospira sp. III7, AY123809
Denitratisoma oestradiolicum, AY879297
coal-tar contaminated groundwater 36-9, AF351219
uranium contaminated aquifer 101328CG13, AY532566
B1202 (16)
B3667 (3)
contaminated sediment 661259, DQ404804
B1217 (5)
Rhodoferax ferrireducens, AF435948
B3672 (3)
β-proteobacteria
Pseudomonas sp. CT1, DQ984204
B1218 (2)
γ-proteobacteria
Thiomicrospira denitrificans, L40808
B1257 (2)
sulfidic cave spring, DQ295573
B3675 (2)
ε-proteobacteria
Desulfovibrio sp. Mlhm, AF193026
B3680 (5)
B3669 (2)
Syntrophobacter pfennigii, X82875
estuarine sediment MidBa15, EF999371
B1215 (4)
Geobacter psychrophilus, AY653551
B1203 (12)
Desulfobulbus propionicus, AY548789
Desulfococcus multivorans, AF418173
Desulfobacterium autotrophicum, AF418177
Syntrophus gentianae, X85132.
B3609 (25)
contaminated sediment 655097, DQ404757
B3676 (3)
δ-proteobacteria
0.10

Figure 2.5 (Continued)



Eukarya While preparing our clone libraries for *Eukarya*, it was intriguing that no PCR amplicon of the 18S rRNA gene was obtained from the background control well. This indication of very low eukaryal biomass outside of the contaminated zone is consistent with our prior report of >300-fold enrichment of predatory protozoa in the contaminated zone (Madsen *et al.*, 1991). We used two different sets of PCR primers to obtain eukaryal sequences leading to the recovery of a broad diversity spanning *Alveolates, Stramenopiles, Fungi*, and small *Metazoa* (Fig. 2.5C) comparable to prior studies (Dawson and Pace, 2002). Particularly striking trends in the on-site eukaryal phylogeny are the prevalence of sequences closely related to: (i) the anaerobic ciliates (e.g., *Paruroleptus*); (ii) predatory ciliates (e.g., *Lembadion*); and (iii) the fresh interstitial-water flatworm, *Suomina*, a Metazoan.

2.5 Discussion

Prior to interpreting results of this study, we must declare limitations implicit in the methodologies. First, though widely accepted as indicative of the degree of contamination, aqueous-phase sampling of well waters within aquifer materials does not completely account for contaminant constituents strongly associated with aquifer sediments. Similarly, planktonic microbial cells (drawn from a ~3-liter aquifer volume) are not necessarily fully representative of the microbial community occupying the mineral matrix. Another key methodological consideration is that the clone-library approach we utilized to characterize microbial populations has potential biases (von Wintzingerode *et al.*, 1997) and provides an incomplete census for any microbial community (e.g., Tringe *et al.*, 2005; Elshahed *et al.*, 2008). Despite the above caveats, the procedures employed

allowed us to successfully tie together information on long-term natural attenuation of organic chemicals with microbial community dynamics in the subsurface contaminated study site (see also our companion report that addresses N cycling, Yagi *et al.*, 2009b).

A variety of studies have made progress demonstrating relationships between the composition of subsurface microbial communities and both site goechemistry and the presence of specific pollutant compounds (e.g., Haack and Bekins, 2000; Haack *et al.*, 2004; Pickup *et al.*, 2001; Roling *et al.*, 2000). In our prior study, Bakermans and Madsen (2002b) used sequencing of 16S rRNA clone libraries to describe bacterial communities in 4 site wells. Clones representing ARDRA (amplified ribosomal DNA restriction analysis) patterns found in the highest abundance were sequenced (31 total). Sequences related to aerobic bacteria (e.g., *Nitrospira, Methylomonas*,and *Gallionella*) predominated among those retrieved from the uncontaminated area of the site, whereas, sequences related to facultatively aerobic and anaerobic bacteria (e.g., *Azoarcus, Syntrophus*, and *Desulfomaculum*) predominated among those retrieved from the contaminated areas of the site (wells 8, 36 and 12). In the present study (~5 years later), we selected two of the previously characterized wells (12 and 36) and focused on temporal variations and both broader (all 3 domains of life) and deeper, more extensive characterization via the cloning and sequencing of small subunit ribosomal RNA genes. In agreement with the prior study, bacterial phylotypes grouping with aerobic nitrite-oxidizing *Nitrospira* and with the γ-proteobacteria comprised a relatively small fraction of sequenced bacterial populations. However, in contrast to the prior study (Bakermans and Madsen, 2002b) where the most abundant ARDRA patterns from well 36 grouped with the low G+C Gram positive organisms and with the β-proteobacteria, predominant phyloypes for

well 36 in this study grouped with *Bacteriodetes/Chlorobi*, δ-proteobacteria, and unclassified lineages (Fig. 2.5A). For well 12, the majority of clones in the current study grouped with the δ-proteobacteria, β-proteobacteria, and unclassified lineages (Fig. 2.5A). The likely reason for differences between results of the two studies is geochemical habitat alterations (hence change in selective pressures) over the intervening years – although incomplete clone coverage in both studies and/or inconsistencies in PCR-amplification outcomes may also have contributed to the apparent discrepancies.

Other notable findings reported here include confirmation of Bakermans and Madsen's (2002b) discovery of sequences representing the D-BACT lineage (Fig. 2.5A) that subsequently were shown by Raghoebarsing *et al.* (2006) to anaerobically oxidize methane using nitrate as an electron acceptor. In addition, sequences closely aligned with *Nitrosopumilus* within the archeal Group I.1a (Fig. 2.5B), known to carry out ammonia oxidation (Konneke *et al.*, 2005) provide support for the likely on-site redox cycling of ammonia, as verified by data in the accompanying companion manuscript (Yagi *et al.*, 2009b).

Studies on the temporal dynamics of naturally occurring microbial communities have been completed for a variety of habitats, often utilizing molecular fingerprinting procedures ranging from DGGE to T-RFLP to ARISA . Key habitats examined for temporal microbial dynamics have included the mid-Atlantic bight (Nelson *et al.*, 2008), the Cariaco Basin (Lin *et al.*, 2008), epilithic biofilms in a freshwater stream (Anderson-Glenna *et al.*, 2008), landfill leachate (Sundberg *et al.*, 2007), an estuarine harbor (Kan *et al.*, 2006), the Mediterranean Sea (Ghiglione *et al.*, 2005), and coastal California seawater (Fuhrman *et al.*, 2006). Each community fingerprinting approach has associated strengths and weak-

nesses. We chose to utilize T-RFLP in this study because it can be consistently applied to pools of DNA and RNA and reveal spatial and temporal variations in community composition (Schutte *et al.*, 2008). Under the best scenario, community composition surveys can be correlated to information about habitat characteristics – leading to predictable, mechanistic relationships between site geochemical conditions and microbial community composition (e.g., Fuhrman *et al.*, 2006). The T-RFLP analyses completed in the present study sought to explore spatial and temporal variations in site well waters. We found that microbial communities in the subsurface site were, in fact, dynamic in time (Fig. 2.3), but also constrained by local geochemical conditions (Fig. 2.4). Slight temporal variability in subsurface microbial community composition was reported by Hendrickx et al. (2005) using DGGE profiles to monitor colonization of aquifer material placed in contaminated and uncontaminated wells over a 122-day period. More pronounced changes in subsurface sediment community composition were found to be associated with aquifer recharge events over a 9-month period using community ARDRA, DGGE, and multivariate statistical analyses (Haack et al. 2004). Similarly, Simon et al. (2001) found in a karst aquifer that bacterial cell density, hydrolytic activity, and respiration were influenced by water-flow events.

The pioneering molecular phylogenetic survey of eukaryotic rRNA gene sequences reported by Dawson and Pace (2002) revealed unanticipated diversity in three anoxic sediments, including seven previously unknown kingdom-level lineages. Since that time, additional nonculture-based studies describing eukaryotic diversity in a variety of ecological settings have emerged. Among these habitats are: a multipond saltern (Casamayor *et al.*, 2002), rice soil (Murase *et al.*, 2006), an anoxic Norwegian fjord (Behnke *et al.*, 2006), Mediterranean seawater

(Diez *et al.*, 2001a), a sulfide-rich freshwater spring (Luo *et al.*, 2005), a seafloor hydrothermal vent field (Lopez-Garcia *et al.*, 2007), and an anaerobic aquifer polluted with landfill leachate (Brad *et al.*, 2008). The latter study, most relevant to data presented here, utilized *Eukarya*-specific PCR primers (Diez *et al.*, 2001b) to amplify 18S rRNA genes and analyze them via DGGE profiling and sequence analysis. Brad *et al.* (2008) were surprised to find that the microeukaryotic community in their anaerobic sandy landfill aquifer site was dominated by fungi (*Basidiomycota* yeasts), and also included both Chlorophyta and the bacterial predatory nanoflagellate, *Heteromita globosa* (within the Cercozoa); groundwater mesofauna were not found. Similar to Brad *et al.* (2008), we found evidence for predatory protozoa only the in contaminated zones in our site (Fig. 2.5C) – this makes it clear that a microbial food chain based on bacterial metabolism of pollutants exists at the site. Although Madsen *et al.* (1991) established at this coal-tar-waste contaminated site that that high numbers of culturable predatory protozoa were an important indicator for *in situ* pollutant metabolism, the extensive diversity of the food chain had not been examined until now. Not only did we find evidence for a variety of fungi, Stramenopiles, and Alveolates, but we also detected sequences for the free-living turbellarian flatworm, *Suomina*, known to consume bacteria and protozoa (Kolasa, 2000). Thus, the contaminant-stimulated trophic community at our study site is both elaborate and has endured for a decade and a half.

The role of microeukaryotes (protists) in microbially-based food chains has been emphasized in several previous literature reviews (Corliss, 2002; Finlay and Esteban, 1998; Novarino *et al.*, 1997; Euringer and Lueders, 2008; Foissner, 2006). Both by feeding on growing prokaryotic biomass and/or by adapting a heterotrophic lifestyle, protists play roles in ecosystems that include reminer-

alization of nutrients, cycling of nutrients to higher trophic levels, excretion of growth factors, and alteration of hydrologic flow and contaminant bioavailability (e.g., Mattison and Harayama, 2001; Mattison *et al.*, 2002). Kinner *et al*. (2002) showed that, in a carbon-limited aquifer, contaminant biodegradation may be enhanced by protistan predation and that the rate of carbon uptake per unit bacterial biomass can be increased. Consistent with the findings of Kinner *et al.* (2002) and Madsen *et al.* (1991), both Sinclair *et al.* (1993) and Zarda *et al.* (1998) also found high subsurface protozoan biomass associated with aromatic-hydrocarbon-contaminated subsurface sediments.

Data describing loss of contaminants from site groundwater (Fig. 2.2) unequivocally documented that the site is recovering via the process of natural attenuation. Prior and current reports (Madsen *et al.*, 1991; Wilson *et al.*, 1999; Jeon *et al.*, 2003; Yagi *et al.*, 2009b) prove that the key attenuating agents are the native microbial populations that actively metabolize the organic contaminant compounds in situ. Human activities increasingly interfere with biosphere function (Liu *et al.*, 2007; Poliakoff *et al.*, 2002). Resilience has been identified as an ecosystem property useful in predicting responses to anthropogenic change (Holling, 1973; Folke *et al.*, 2004; Botton *et al.*, 2006). Microbial biogeochemical processes, such as those described here and in our companion report (Yagi *et al.*, 2009b), underlie ecosystem resilience and increasingly need to be rigorously defined, understood, and managed.

CHAPTER 3

DIVERSITY AND SPATIAL-TEMPORAL VARIATION OF MICROBIAL OXYGENASE ABUNDANCE AND EXPRESSION IN A SHALLOW CONTAMINATED AQUIFER[2]

[2]Jane M. Yagi and Eugene L. Madsen.

3.1 Abstract

The diversity of Rieske dioxygenase genes catalyzing degradation of non-polar polycyclic aromatic hydrocarbons was investigated in a shallow aquifer contaminated with naphthalene and other coal-tar waste-derived compounds. We used the PCR-based approach of Ní Chadhain *et al*. (2006) to create 4 clone libraries from DNA extracted from 2 contaminated wells (each sampled twice). A broad suite of genes was detected, ranging from dioxygenase sequences associated with *Rhodococcus* and *Sphingomonas* to 32 previously uncharacterized Rieske clone groups. The *nag* genes appeared frequently (24% of the total) in the 2 monitoring wells, characterized by low (micromolar) ambient concentrations of naphthalene. A quantitative competitive PCR assay showed that abundances of *nag* genes (and archetypal *nah* genes associated with previously isolated naphthalene degraders from the site) fluctuated substantially over a 10-month period; causes of the temporal fluctuations were uncertain. RT-PCR was employed to detect specific transcripts of expressed naphthalene dioxygenases in total RNA extracts from the 2 contaminated wells. The mRNA sequences revealed that *nah*- and *nag*-type genes were expressed *in situ*, corresponding well with structural gene abundances, and suggesting coexistence of these alternate degradation pathways at low ambient concentrations of naphthalene.

3.2 Introduction

The advancement of genetic information about microbial community processes is increasingly central to microbial ecology (e.g., Frias-Lopez *et al.*, 2008). Depending on biogeochemical context, microbial activities of interest may range

from C-fixation to N-cycling, to biodegradation of contaminants threatening human and ecosystem health. Polycyclic aromatic hydrocarbons (PAHs), derived from both natural and anthropogenic sources, are toxic pollutants formed during the incomplete combustion of organic matter (Samanta *et al.*, 2002). Phylogenetically and physiologically diverse microbial populations contribute to biodegradation of PAHs in soil, sediment and groundwater (Habe and Omori, 2003; Peng *et al.*, 2008). The most widely studied bacterial degradation pathways for naphthalene, the simplest and most soluble PAH and a model compound for PAH metabolism, are initiated by the catalytic activity of an evolutionarily conserved naphthalene dioxygenase (NDO; Habe and Omori, 2003; Peng *et al.*, 2008) system. The 2 primary NDO-mediated degradation pathways are distinguished by conversion of naphthalene, via salicylate, to either catechol (e.g., *nah* genes) or gentisate (e.g., *nag* genes; Habe and Omori, 2003). Many bacteria harbor well characterized NDO systems, including Gram-negative species (e.g., *Pseudomonas* spp., *nah* genes; *Burkholderia* spp., *phn* genes; *Ralstonia* and *Comamonas* spp., *nag* genes) and Gram-positive species (e.g., *Mycobacteria* spp., *nid* genes; *Rhodococcus* spp., *nar* genes) (Dennis and Zylstra, 2004; Laurie and Jones, 1999; Zhou *et al.*, 2001; Moser and Stahl, 2001; Khan *et al.*, 2001; Kulakov *et al.*, 2005).

Naturally occurring microbial communities are exceedingly diverse, potentially harboring as yet uncharacterized genes encoding novel assemblages of often functionally redundant catabolic pathways (Peel and Wyndham, 1999; Huang *et al.*, 2009). This challenges our ability to reliably relate abundance and activity of specific genes to environmental variables (DeBruyn *et al.*, 2007; Heiss-Blanquet *et al.*, 2005; Lloyd-Jones *et al.*, 1999; Laurie and Lloyd-Jones, 2000; Peel and Wyndham, 1999; Wilson *et al.*, 2003; Stach and Burns, 2002; Whyte *et al.*,

2002; Gomes *et al.*, 2007). A variety of culture independent, biomarker-based approaches for characterizing microbial assemblages and their activity in naturally occurring communities have been applied in PAH contaminated environments. Analyses have ranged from metagenomic libraries (Suenaga *et al.*, 2007), to fluorescent in situ hybridization (Rogers *et al.*, 2007), to functional gene arrays (Rhee *et al.*, 2004; Nyyssönen *et al.*, 2008), to PCR targeting both DNA (Da Silva *et al.*, 2006; Nyyssönen *et al.*, 2006; Baldwin *et al.*, 2003, 2008; Cébron *et al.*, 2008; Sipilä *et al.*, 2008) and expressed RNA transcripts (Alfreider *et al.*, 2003; O'Neil *et al.*, 2008; Bordenave *et al.*, 2008). The catabolic importance of archetypal NDO systems (encoded by *nah* genes) in many PAH contaminated environments has been successfully demonstrated by functional-gene-targeted molecular approaches (DeBruyn *et al.*, 2007; Dionisi *et al.*, 2004; Bordenave *et al.*, 2008; Gomes *et al.*, 2007; Witzig *et al.*, 2006; Taylor and Janssen, 2005; Stapleton *et al.*, 2000; Sipilä *et al.*, 2008; Salminen *et al.*, 2008; Tuomi *et al.*, 2004). Furthermore, broad-specificity PCR primers have enlarged the target range of molecular methods for environmental gene detection, and uncovered previously unknown dioxygenase diversity in environmental samples (Yeates *et al.*, 2000; Ní Chadhain *et al.*, 2006; Lozada *et al.*, 2008).

Natural PAH degrading microbial populations typically contend with low and fluctuating bioavailability of complex mixtures of substrates (Andreoni and Gianfreda, 2007). Microorganisms may utilize different degradation mechanisms depending on low or high available substrate concentrations (Guerin and Boyd, 1995). Furthermore, substrate utilization capacity must be optimized in the context of myriad biotic and abiotic environmental stresses (Ramos *et al.*, 2001; Velázquez *et al.*, 2006). Especially at very low concentrations of substrate, microbes may simultaneously use multiple carbon sources, reducing threshold

concentrations for metabolism of individual substrates (Munster, 1993; Lendenmann *et al.*, 1996; Kovarova-Kovar and Egli, 1998).

Previous work at a coal tar waste-contaminated aquifer (South Glens Falls, NY) has provided extensive insight on naphthalene degrading microbial communities *in situ* (Madsen *et al.*, 1991; Wilson *et al.*, 1999; Bakermans and Madsen, 2002b; Bakermans *et al.*, 2002; Jeon *et al.*, 2003; Wilson *et al.*, 2003). Horizontal gene transfer of NDO-harboring plasmids was described, key functional genes were associated with the presence of contaminants in groundwater, and *in situ* expression of naphthalene dioxygenase genes was documented (Wilson *et al.*, 2003; Bakermans and Madsen, 2002a; Wilson *et al.*, 1999). PAHs derived from coal-tar waste support a rich food web in the contaminated groundwater at this site (Madsen *et al.*, 1991; Yagi *et al.*, 2009a). Over more than a decade of monitored natural attenuation at the site, naphthalene concentrations within the contamination plume have diminished from high levels of approximately 1.6 parts per million (ppm) to current low parts per billion (ppb) concentrations (Yagi *et al.*, 2009a). The proven on-site microbially mediated degradation of contaminants and persistence of low levels of PAHs warranted an investigation of the current diversity of aromatic oxygenase genes at the site. Examining catabolic gene abundance and expression in this historically contaminated site can provide insight on the progression of long term natural attenuation and microbes responsible for degradation of low concentrations of PAHs.

This study describes the terminal dioxygenase diversity specific to degradation of non-polar PAHs in a historically coal tar waste-contaminated aquifer. We also used a newly developed quantitative PCR assay to reveal spatial and temporal variation of the most abundant genes recovered in the survey (*nag*

genes) and compared these gene levels to archetypal *nah* gene abundance. We detected expression of both *nag* and *nah* genes confirming *in situ* activity of these naphthalene-degrading populations, despite low ambient concentrations of naphthalene. The results demonstrated a diverse community of aromatic oxygenase harboring organisms, and coexistence of alternate degradation pathways at low concentrations of naphthalene.

3.3 Materials and Methods

3.3.1 Groundwater sampling, biodegradation assay set-up, and DNA/RNA extraction

Groundwater monitoring wells were located in a rural coal tar waste contaminated site in South Glens Falls, NY (Madsen *et al.*, 1991; Murarka *et al.*, 1992; Wilson *et al.*, 1999). Methods for sampling groundwater and measuring its geochemical characteristics have been described previously (Bakermans and Madsen, 2002b; Yagi *et al.*, 2009a,b). Samples were collected from a control well (well 60) upgradient of the source contamination, and from two wells (well 36 and well 12) located within the contaminant plume. Dissolved oxygen and temperature were measured using a YSI model 85 probe (Yellow Springs, OH) inserted into the flow of the emerging groundwater. pH of fresh samples was measured with a model SA250 portable pH meter (Orion, Boston MA). Groundwater naphthalene concentrations were determined in May 2005 by a commercial environmental analysis company (GEI, Inc., Glastonbury, CT) as described previously (Murarka et al., 1992). Subsequent naphthalene concentrations were determined using solvent extraction and GC/MS. Groundwater

from well 12 and well 36 was sampled for naphthalene biodegradation assays (August 2006) by pumping water directly into 40-ml sterile glass I-Chem vials (I-CHEM, Rochester, NY). Vials were completely filled, sealed immediately with Teflon-lined butyl rubber stoppers, and placed on ice for transport to the laboratory. Groundwater microbes in samples from August, November and December 2006 were quantified by acridine orange direct counts (Kemper and Pratt, 1994). Samples were fixed on site by the addition of 2% formaldehyde. Groundwater biomass was collected for molecular analyses (November 2005, August 2006, November 2006, December 2006, and May 2007) on Durapore filters (Millipore Corp,. Bedford MA). Total DNA and RNA were extracted separately from filter-collected groundwater biomass using SDS-boiling lysis procedures with phenol-chloroform extraction as described previously (Wilson *et al.*, 1999; Yagi *et al.*, 2009a,b).

3.3.2 Biodegradation assays

Laboratory naphthalene biodegradation assays were set up within 24 hours of sample collection using replicate, sealed 40-ml serum bottles filled with groundwater on site (see above). Killed control bottles were established by addition of 500 μl poison (5% HCl/0.25M $HgCl_2$) whereas live microcosms received an equivalent volume of sterile water. To ensure that naphthalene concentrations in all of the microcosms were in the range of detection by GC/MS, 50 μl of a concentrated stock of naphthalene dissolved in dichloromethane was added to each bottle, increasing the ambient concentration by 50 μg/L. Vials were incubated at 10 °C in the dark without shaking. At various times (0, 1, 3, 6, 14 days), two live samples and one killed control were sacrificed for solvent extraction

and GC/MS analysis of naphthalene concentrations. Five ml groundwater was removed from each vial using a glass syringe, followed by addition of exactly 2 ml dichloromethane and 6 g NaCl. The vials were shaken vigorously for 10 seconds to achieve complete dissolution of salt, and stored 2 days upside-down at 4 °C to allow phases to settle. Using a glass pipet, about 1.5 ml of the organic layer was transferred from each vial to a tared 2 mL vial with Teflon-lined rubber septum and screw cap, being careful to exclude water from the syringe. A 0.2 g portion of anhydrous sodium sulfate was then added, and each vial was capped and shaken vigorously for 2 min. One ml of the dried extract was transferred to a fresh, tared 2 ml vial with Teflon-lined rubber septa and screw caps. The extracts were concentrated under a stream of nitrogen gas, weighed (to allow quantification of naphthalene present in each water sample), and stored at -20 °C until analysis. Naphthalene concentrations were determined by GC/MS using a Hewlett-Packard Model 6890 Series II gas chromatograph (Bakermans and Madsen, 2002b).

3.3.3 Rieske gene amplification and clone library construction

Four Rieske gene fragment clone libraries were generated. Each was derived from groundwater DNA sampled from either well 36 (November 2005 and August 2006) or well 12 (August 2006 and November 2006). Rieske dioxygenase genes in groundwater DNA were amplified using the primers Rieske_f and Rieske_R (Table 3.1; Ní Chadhain *et al.*, 2006). ThermoStart DNA polymerase (ABgene) was used for PCR amplification on a PTC-200 DNA Engine thermocycler (MJ Research). Duplicate reactions were performed for each clone library using 0.5-25 ng DNA template in 25 ml volumes and previously described cy-

cling conditions (Ní Chadhain *et al.*, 2006). The pooled 78-bp Rieske amplicons were purified by electrophoresis and gel extraction using the QIAquick Gel Extraction Kit (Qiagen) and ligated into the vector pCR2.1 (TOPO-TA cloning, Invitrogen) following the manufacturers' protocols. Inserts were verified in 80-130 randomly picked colonies per library by PCR using M13 primers, and inserts were sequenced using an ABI 3730 automated sequencer (Life Sciences Core Laboratories Center, Cornell University). Rieske fragment sequences were manually checked for quality and edited to exclude to primer binding sites using 4Peaks software (available at http://www.mekentosj.com/4peaks/), and grouped into clone families based on 95% sequence identity. Neighbor-joining trees were constructed using CLUSTALX.

3.3.4 Competitive PCR method development and quantitation of naphthalene degradation genotypes

The bacterial strains and plasmids used for competitive PCR method development and assay are listed in Table 3.1. All strains were cultivated in standard R2A media at 23 °C except for *Escherichia coli* strains, which were grown in Luria Bertani broth at 37 °C. Two primer pairs were designed to specifically amplify the *nah* and *nag* gene clusters by targeting sequences upstream of genes encoding the naphthalene dioxygenase large subunit: known *nag* gene clusters, but not *nah* clusters, encode genes for salicylate-5-hydroxylase in this region (Moser and Stahl, 2001). The 750-bp amplicon from *nah* gene clusters spanned the region from *nahAa* to *nahAc*, and the 888-bp amplicon from *nag* gene clusters spanned the region from *nagH* to *nagAc* (Dennis and Zylstra, 2004; Zhou *et al.*, 2001). Deletion derivatives of the respective gene clusters carried on the

Table 3.1: Bacterial strains, plasmids and PCR primers used in this study

Strain or plasmid	**Description**	**Reference or source**
Polaromonas naphthalenivorans str. CJ2	Host of chromosomal *nag* genes	Jeon *et al.* 2004
Pseudomonas putida NCIB 9816-4	Host of *nah* genes carried on plasmid pDTG1	Dennis and Zylstra 2004
P. putida G7	Host of *nah* genes carried on plasmid NAH7	Yen and Serdar 1988
Ralstonia sp. U2	Host of *nag* genes carried on plasmid pWWU2	Zhou *et al.* 2001
phnAc strain Hg8	Host of *phn* genes	Wilson *et al.* 2003
Escherichia coli TOP10	TA cloning	Invitrogen
Comamonas testosteroni GZ39	Host of *nag* genes	Goyal and Zylstra 1996
Rhodococcus sp. str. NCIMB 12038	Host of *nar* genes	Larkin *et al.* 1999
Plasmid or vector	Description	
pGEM-T	TA cloning vector	Promega
pUC18	*E. coli* cloning vector, Ap^R	
pJMY3	pUC18 derivative containing *nag* deletion fragment	This study
pJMY4	pUC18 derivative containing *nah* deletion fragment	This study
pJMY5	pUC18 derivative containing *phn* deletion fragment	This study
Primer	Sequence (5′-3′)	
Rieske gene amplification		
Rieske_f	TGYMGNCAYMGNGG	Ní Chadhain *et al.* 2006
Rieske_r	CCANCCRTGRTANSWRCA	Ní Chadhain *et al.* 2006
Construction of competitive PCR standards		
Aa.15485.2.T7.f	TAATACGACTCACTATAGGATATTTATGCCGATGCCTTCA	This study
Ac.16075.2.RIR.R	ATCCCCATTTTTGCGGTAACATATCAGACCAGATTCACTTAC	This study
CJ2.5219.T7.f	TAATACGACTCACTATAGGTACCACGACCCTTACTACCA	This study
CJ2.5910.RIR.r	CGGAGCCATCGTTCTGACGTCAGCCCTGCTTTACTCAC	This study
Competitive PCR analysis		
Aa.15485.2.f	ATATTTATGCCGATGCCTTCTA (*nah*)	This study
Ac.16210.2.r	ATCCCCATTTTTGCGGTAACATA (*nah*)	This study
CJ2.5219f	TACCACGACCCTTACTACCA (*nag*)	This study
CJ2.6088r	CGGAGCCATCGTTCTGAC (*nag*)	This study
Naphthalene dioxygenase transcript analysis		
Ac114F.1	GGAACTGGCTTTTTYTSACYC	This study
Ac596R.1	TTTCCGMGGGYGCCTTC	This study

pUC18 plasmid were used as internal standards for competitive PCR reactions (Jin *et al.*, 1994). Shortened amplicons generated from PCR of these deletion derivatives during co-amplification with site-derived target sequences were distinguished from wild-type amplicons by gel electrophoresis, allowing quantification of specific genotypes in field-sampled nucleic acid extracts (Jin *et al.*, 1994). Standard curves for quantitative competitive PCR were constructed with a constant amount of target and serial dilutions of the competitor carried on a plasmid. Primer pairs for competitive PCR were tested for group-specificity on isolated strains of naphthalene degrading bacteria known to harbor particular naphthalene degradation operons. The *nah* gene amplicon but not the *nag* gene amplicon was observed from PCR amplification of total DNA from pure cultures of *P. putida* NCIB 9816-4 and *P. putida* G7, and only the *nag* gene amplicon was amplified from *Ralstonia* sp. U2 and *P. naphthalenivorans* strain CJ2 (data not shown). The competitive PCR assay was validated using DNA extracted from pure cultures. When equimolar quantities of target and standard were added to PCR reactions, similar amplification efficiencies were observed over 3 orders of magnitude tested (data not shown). For instance, when either 50 or 500 copies of *nag* genes were amplified, the PCR plateau was attained after 34 or 36 cycles, respecitively; quantitation was thus carried out after 32 or 34 cycles to fall within the exponential phase of amplification.

3.3.5 Naphthalene dioxygenase transcript analyses

Total RNA extracts from well 36 and well 12 (August 2006) were treated with DNase I (Invitrogen) and converted to cDNA using random hexamer-primed reverse transcription with SuperScript III reverse transcriptase (Invitrogen), ac-

cording to the manufacturer's instructions. Twenty-five μl PCR reactions were set up as described above, using PCR primers Ac114F.1 and Ac596R.1 that target both *nahAc* and *nagAc* (Table 3.1), and cDNA as template, according to the reverse transcriptase manufacturer's recommendations. Gel-purified PCR amplicons were used to generate clone libraries as described above. Vector specific primers were used to screen plasmids for inserts and DNA amplicons for 18 and 16 clones from the well 12 clone library and the well 36 clone library, respectively were digested with the restriction endonucleases HaeIII and HhaI (New England Biolabs). The restriction products were separated on agarose gels, and 11 representative sequences were obtained, as described above. CLUSTALX was used for alignment of nucleic acid and deduced amino acid sequences and construction of neighbor joining trees.

3.3.6 Phylogenetic analyses

Nucleic acid sequences and deduced amino acid sequences were subjected to BLASTN and BLASTP searches, respectively, at the National Center for Biotechnology Information (NCBI) website.

3.3.7 Statistical analyses

Statistical analyses were carried out using the programs Analytic Rarefaction 1.3 (Holland, 2001) and EstimateS (Colwell, 2005). Good's estimate was used to calculate library coverage using the formula [1-(n/N)] x 100, where n is the number of sequences represented by a single clone group only and N is the total number of sequences analysed (Good, 1953; Kemp and Aller, 2004).

3.3.8 Accession numbers

The nucleotide sequence data reported here have been submitted to GenBank under accession no. FJ820292 to FJ820329.

3.4 Results and Discussion

3.4.1 Geochemistry and naphthalene biodegradation potential in groundwater

The South Glens Falls study site harbors a groundwater contamination plume of polycyclic aromatic hydrocarbons, derived from buried coal-tar waste, undergoing natural attenuation (Bakermans *et al.*, 2002). We have previously reported extensive field evidence for in situ aerobic and anaerobic biodegradation of naphthalene, the major soluble component of coal-tar waste (Madsen *et al.*, 1991; Bakermans *et al.*, 2002; Jeon *et al.*, 2003; Yagi *et al.*, 2009b). The high diversity of prokaryotic and eukaryotic microbial communities in site groundwater has also been characterized (Bakermans and Madsen, 2002b; Yagi *et al.*, 2009a). In the current study, groundwater samples were collected over a 2-year sampling period from 2 monitoring wells within the contaminant plume (well 36 and well 12; Table 3.2) and from an upgradient control well (data not shown). Both contaminated wells displayed fluctuating oxygen concentrations far below saturation (about 9 mg l^{-1}; Table 3.2) suggesting that ambient conditions likely featured a dynamic interplay in situ between microaerobic and anaerobic physiological reactions. Given unavoidable diffusion of atmospheric O_2 into the groundwater sampling apparatus, O_2 readings $\leq$0.2 mg l^{-1} were considered

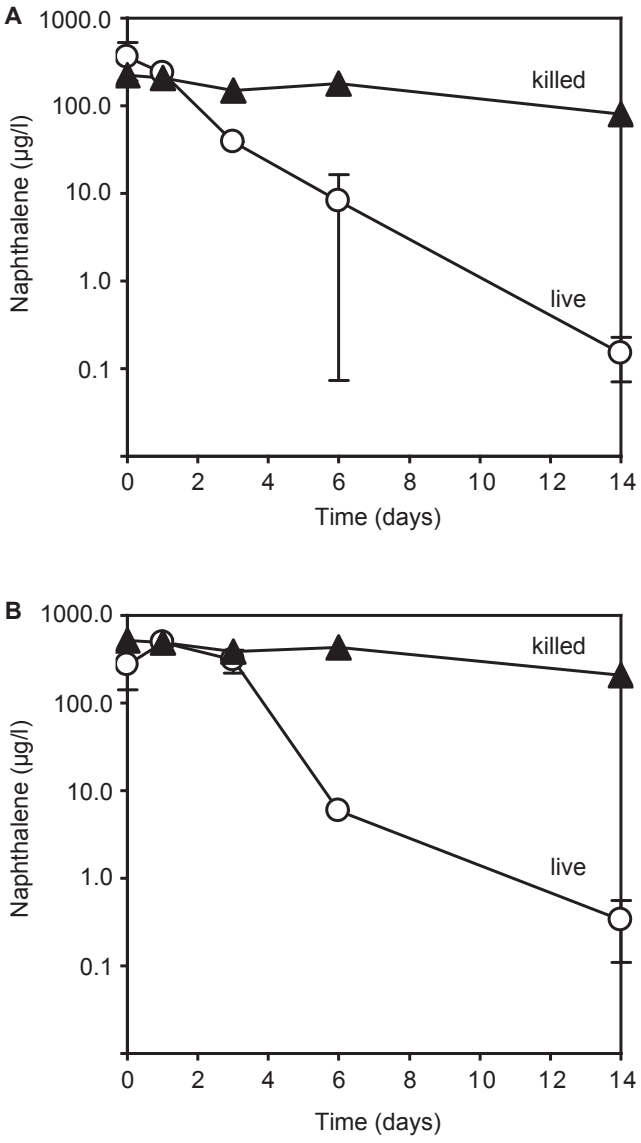


Figure 3.1: Biodegradation of naphthalene in groundwater collected from (A) Well 12 and (B) Well 36. Groundwater samples (○ - open circles) and killed controls (▲ - closed triangles) were incubated at 10°C without shaking. Error bars represent the range of duplicate biodegradation assays from each well.

Table 3.2: Characteristics of groundwater samples: sampling locations, sampling dates, temperature, dissolved oxygen concentration, pH, total cell counts and DNA yields

Well	Sampling date	Temperature (°C)	DO[a] (mg/L)	pH	Total cell count[c] (cells/mL)	DNA yield (μg/L)
36	Nov-05	10.9	0.2	8.1	NA[b]	1.3
	Aug-06	12.7	0.2	7	65 300±10 600	0.87
	Nov-06	12.4	0.2	7	24 900± 3 540	0.77
	Dec-06	10.9	0.7	7.1	16 000±2 200	0.05
	May-07	11.3	1.6	7.5	NA[b]	0.42
12	Aug-06	15.8	0.1	7.3	30 200±5 450	0.5
	Nov-06	12.1	0.2	7	28 200±2 480	0.82
	Dec-06	8.2	0.6	6.8	13 000±5 250	0.22
	May-07	12	0.7	7.2	NA[b]	0.13

[a] Abbreviations: DO, dissolved oxygen

[b] NA denotes data not analyzed

[c] Direct cell counts determined by staining with acridine orange and epifluorescence microscopy

equivalent to zero. The total cell density in well waters fluctuated mildly in the range of ~10^4 cells ml^{-1} (Table 3.2). Well 12, located further downgradient from the source contamination, has historically featured lower PAH concentrations (Yagi *et al.*, 2009a).

To compare potential naphthalene biodegradation activity among microbial populations native to the site, we prepared laboratory-incubated microcosms using waters from wells 36 and well 12 (Fig. 3.1). Samples from August 2006 were used in the biodegradation assays. Because recent measurements indicated decreasing, low ambient naphthalene concentrations in the more contaminated well (380 μg l^{-1} in well 36 in June 2005 compared to 1110 μg l^{-1} in 1999-2000; see Bakermans *et al.*, 2002), bottles were amended with 50 μg l^{-1} naphthalene in order to ensure starting concentrations within the quantitative range of detection by GC/MS. No chemical or nutrient amendments were otherwise made to the microcosms. Ambient concentrations of naphthalene were indeed low (unamended concentrations of 160 μg l^{-1} [or 1.3 μM] in well 36, 75 μg l^{-1} [or 0.6 μM] in well 12), but rapid biodegradation occurred in samples from both wells at 10 °C (Fig. 3.1). Within 14 days of incubation, more than 99% of naphthalene initially present was degraded by native microorganisms in viable treatments from both wells (Fig. 3.1).

In agreement with prior reports of biodegradation potential at the site, a lag phase of at least 3 days was observed prior to disappearance of naphthalene in the bottles from well 36 but not from well 12 (Fig. 3.1; Bakermans *et al.*, 2002). The initial conditions for the biodegradation assays included ambient dissolved oxygen (0.1-0.2 mg l^{-1}; Table 3.2) and, most likely, low levels of oxygen introduced during field sampling (approximately 0.75 mg l^{-1}; Bakermans

et al., 2002). Given that the starting concentrations of naphthalene were < 0.5 mg l^{-1}, it is reasonable to expect that the microcosms contained sufficient oxygen to support the observed biodegradation activity by aerobic processes (Bakermans *et al.*, 2002). Though direct cell counts indicated lower bacterial density in well 12 (Table 3.2), the samples from this less contaminated well displayed an earlier onset of biodegradation activity (Fig. 3.1). The relative delay in biodegradation activity in well 36 samples suggests induction or enrichment of latent aerobic microbial activity; both microaerobic and anaerobic physiological conditions have continued to dominate waters from well 36 (Fig. 3.1, see also Yagi *et al.*, 2009a). The absence of a noticeable lag phase in biodegradation assays for well 12 samples (Fig. 3.1) is consistent with the possibility microsites *in situ* harbored active aerobic and microaerobic naphthalene degrading microbial populations.

3.4.2 Rieske dioxygenase genes in groundwater samples

The application of culture independent methods in environmental samples has revealed novel populations responsible for *in situ* naphthalene degradation (Jeon *et al.*, 2003; Huang *et al.*, 2009). While high contaminant concentrations, through toxicity and inhibitory effects, may select for highly specialized microbial populations (Ferguson *et al.*, 2007), it has been reported that naphthalene concentrations < 1 mg l^{-1} are not necessarily inhibitory to native microbial activity (Hudak and Fuhrman, 1988). The ambient concentrations of PAHs in our aquifer (above) were well below this range, and although naphthalene concentrations in pure cultures as low as 0.01 μM have been reported to induce DNA repair mechanisms in *Escherichia coli* cells (Kim *et al.*, 2007), we hypothesized that PAH degrading communities in the groundwater would be highly diverse.

Table 3.3: Characteristics and diversity estimates of Rieske clone libraries analyzed in this study[a]

Well	Sampling date	Number of clones[b]	Number of clone groups[c]	H'[d]	Percent library coverage[e]	Chao-1[f]
36	Nov-05	28	9	1.73	0.86	14.8
	Aug-06	28	7	1.52	0.89	28.6
12	Aug-06	27	13	2.36	0.81	37
	Nov-05	26	14	2.41	0.7	45.3

[a] Libraries were generated using DNA extracted from contaminated groundwater sampled from 2 wells (Well 36 and Well 12) at 3 different times (November 2005; August and November 2006).

[b] Number of clones sequenced from each library.

[c] Clone groups based on Rieske gene sequences (>95% identity).

[d] Shannon-Wiener index calculated using http://biome.sdsu.edu/fastgroup/cal_tools.htm; higher number denotes higher diversity.

[e] Good's estimate used to calculate percent library coverage.

[f] Chao-1 value calculated using EstimateS software.

A previously described degenerate PCR primer pair was used to broadly characterize the complement of both known and unknown aromatic oxygenase genes in groundwater by amplifying sequences encoding the conserved Rieske domain of dioxygenase homologs associated with attack of nonpolar aromatic substrates (Ní Chadhain *et al.*, 2006). It was expected that the generic nature of these primers would enable detection of NDO systems including *nah, nag, phn, nid*, and *nar* genes, as well as many other diverse sequences (Ní Chadhain *et al.*, 2006).

Four Rieske gene fragment clone libraries were generated from the corresponding 78-bp PCR amplicons from well 36 (November 2005 and August 2006) and well 12 (August 2006 and November 2006) (Table 3.3). No Rieske amplification products were obtained from DNA extracted from the background control well. A total of 110 Rieske gene fragments were sequenced (28, 28, 27 and 27 from each library, respectively; Table 3). We classified sequences with > 95% nucleotide identity as individual clone groups, designated by source well (e.g., well 36 = MW36) and an arbitrarily assigned clone number (Fig. 3.2).

Figure 3.2: Unrooted phylogenetic tree determined by neighbor-joining analysis showing diversity of genes encoding aromatic oxygenase enzymes amplified as described by Ní Chadhain *et al.* (2006) from extracted groundwater DNA in 2 contaminated wells (well 36 was sampled November 2005 and August 2006; well 12 was sampled August 2006 and November 2006). Sequences found in this study are indicated in bold type. Abundances of each clone group within libraries are quantified by multiple (or single) symbols as follows: November 2005 (Well 36, ● - filled circles), August 2006 (Well 36, ○ - open circles; Well 12, □ - open squares) and November 2006 (Well 12, ■ - filled squares). GenBank accession numbers for reference sequences representative of the most similar BLAST matches and genes from characterized PAH degradation pathways are shown in boldface type. The 78 bp amplicon occurs within conserved Rieske domain of dioxygenase homologs associated with attack of nonpolar aromatic substrates. CLUSTALX was used for nucleotide sequence alignment and tree construction. Bootstrap values >50% from 1000 replicate analyses are indicated as numbers at the nodes of respective branches. The scale bar represents the expected number of changes per nucleotide position.


nahAc, Pseudomonas putida NCIB9816-4 pDTG1 [AF439407]
bphA1, Sphingomonas yanoikuyae str. B1 [EF151283]
Novosphingobium aromaticivorans DSM 12444 pNL1 [CP000076]
akbA1b, Rhodococcus sp. DK17 pDK2 [AY502075]
Dechloromona saromatica RCB [CP000089]
MW36-15
MW36-13
Mycobacterium sp. KMS pMKMS02 [CP000520]
Mycobacterium gilvum PYR-GCK [CP000656]
MW12-16
UN_PH_16 [DQ325409]
narAa, Rhodococcus sp. NCIMB12038 [AF082663]
MW12-06
MW12-09
MW12-14
dxnA1, Sphingomonas wittichii RW1 pSWIT02 [CP000701]
dxnA1, Sphingomonas sp. [X72850]
bphA1f, Sphingomonas aromaticivorans pNL1 [F079317]
UP_S_4 [DQ325366]
UP_S_25 [DQ325383]
MW12-02
MW36-11
MW12-19
MW12-17
UP_S_8 [DQ325372]
MW36-04
UP_S_24 [DQ325382]
UP_S_16 [DQ325374]
UP_S_15 [DQ325373]
MW12-10
MW12-01
MW12-03
MW12-07
MW12-05
MW36-01
MW36-10
MW36-08
MW12-21
MW12-04
MW12-18
MW12-13
MW12-20
MW36-07
MW12-08
MW12-11
MW36-05
MW36-03
MW12-12
dbtAc, Burkholderia sp. DBT1 [AF380367]
phnAc, Acidovorax sp. NA3 [EU910094]
phnAc, Alcaligenes faecalis [AB024945]
arhA1, Sphingomonas sp. EH1-1 [AY367787]
phnAc, Ralstonia sp. PJ531 [AB066446]
phnAc, Burkholderia sp. RP007 [AF061751]
Burkholderia xenovorans [CP000272]
Ac-ORO4-B65 [AM930904]
bphA, Hyphomonas sp. OC5 [AB429286]
MW36-12
bphA, Pandoraea pnomenusa [CTU47637]
bphA, Burkholderia sp. JB1 [AJ010057]
tcbAa, Pseudomonas nitroreducens J5-1 [EF635855]
Polaromonas naphthalenivorans pPNAP01 [CP000530]
MW36-06
bphA1, P. pseudoalcaligenes [M83673]
bphA, Pseudomonas sp .Cam-1 [AY027651]
bphA1, Pseudomonas sp. B4 [PSU95054]
bphA3, Rhodoccus sp .RHA1 [D32142]
bphA1, Rhodococcus globerulus [X80041]
bphA1, Rhodococcus sp. M5 [U27591]
Cycloclasticus oligotrophus XYLX [U51165]
nagAc, Ralstonia sp. U2 [AF036940]
MW36-14
MW36-09
nahAc, Pseudomonas putida G7 [AB237655]
nagAc, Polaromonas naphthalenivorans CJ2 [CP000529]
0.1

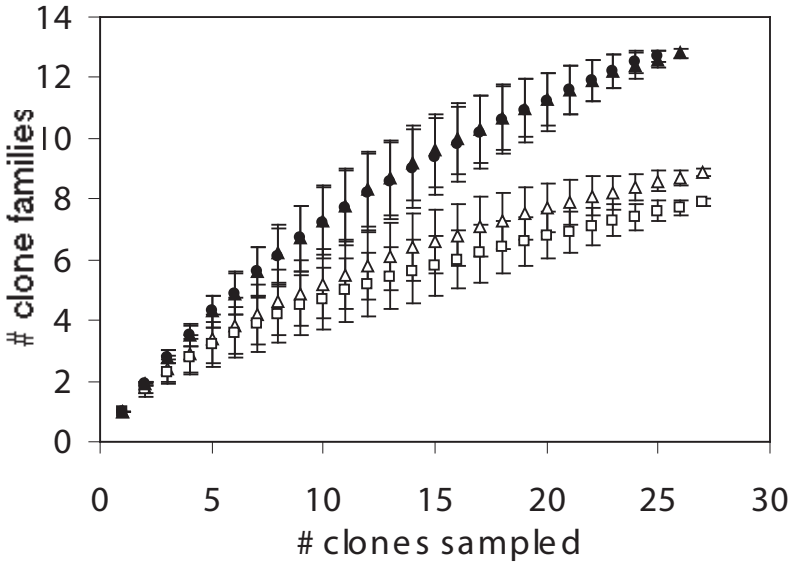


Figure 3.3: Rarefaction curves for Rieske gene fragment clone libraries for well 36 in November 2005 (△ - open triangles) and August 2006 (□ - open squares), and well 12 in August 2006 (● - closed circles) and November 2006 (▲ - closed triangles)

As expected, Shannon-Wiener indices and Chao-1 indicators for the clone libraries suggest that clone group diversity and richness were lower for Rieske gene assemblages amplified from the more contaminated well 36 than for well 12, regardless of sampling date (Table 3.3). This could be attributed to inhibitory aspects of either anaerobic physiological conditions (see above) or higher naphthalene concentrations in well 36 on NDO associated microbial populations. Classical estimates of library coverage (Good's estimate) and rarefaction curves for each library indicated that additional sampling would uncover greater Rieske gene diversity in the groundwater communities (Table 3.3, Fig. 3.3).

In total, 35 clone groups were identified in the 4 clone libraries, with limited overlap between libraries from well 36 and well 12 regardless of sampling time.

Twenty-one clone groups were found exclusively in well 12 samples, and 12 clone groups were identified in well 36 only (Fig. 3.2). The clone library representing well 36 in November 2005 (28 total sequences) contained 9 clone groups (Table 3.3). The dominant clone group, MW36-09 (46% of library total), was identical in sequence to *nagAc* from *Polaromonas naphthalenivorans* str. CJ2 (Fig. 3.2). Clone group MW36-06 (14% of library total) clustered with biphenyl dioxygenases (*bphA*). Clone group MW36-11 (11% of library total) showed no close BLAST matches but clustered with S25 from the N Chadhain study (2006). Clone group MW36-12 clustered with an uncharacterized ring hydroxylating dioxygenase gene amplified from intertidal sediment (Lozada *et al.*, 2008). Clone group MW36-13 was the only other group in the library similar to known genes, affiliating with genes from *Mycobacterium* spp.

The two clone libraries from August 2006 revealed differences in the distribution of aromatic oxygenase genes between well 36 and well 12, suggesting genotypic differences in the functionally important aromatic degrading microbial populations (Table 3.3, Fig. 3.2). Furthermore, the relative abundances of clone groups in well 36 from August 2006 were significantly different from those observed for November 2005, indicating temporal variability in the aromatic degrading populations (Fig. 3.2). The August 2006 sample from well 36 (28 clones total) contained 6 clone groups and was dominated by clone group MW36-15. Six clones fell into clone group MW36-14, which includes the *nagAc* gene from *Ralstonia* sp. U2. Three clones grouped with MW36-04, which had no close matches in the GenBank database but clustered weakly with clones (S24, S16, S15) from the Ní Chadhain study (2006). The well 12 library from August

Table 3.4: Results of database search for matches with > 60% nucleotide identity to groundwater Rieske gene fragments

Clone group	Best match	% identity (nucleotide)	% identity (translated)	Accession number
MW36-09	*nagAc Polaromonas naphthalenivorans* str. CJ2″	100	100	CP000529
MW36-14	*nagAc Ralstonia* sp. U2″	100	100	AF036940
MW36-15	*Stappia aggregata* IAM 12614	63	47	ZP_01549987
MW12-07	Uncultured bacterium clone UP_S_4 PAH dioxygenase large subunit gene	70	73	DQ325366
MW12-16	*Mycobacterium gilvum* PYR-GCK	100	100	CP000656
MW36-06	*bphA Hyphomonas* sp. OC5″	80	80	AB429286
MW36-12	Uncultured bacterium Ac-ORO4-B65	87	100	AM930904

2006 (27 total) contained 14 clone groups. Clone groups MW12-14, MW12-09 and MW36-14 dominated the library (22%, 15%, 11%). Clone group MW12-16, represented by 1 clone, showed high sequence identity with a gene from *Mycobacterium gilvum*. The remaining clone groups from August 2006 had no close matches in the GenBank database.

The library from the November sample from well 12 (2006; 27 total; Table 3.3, Fig 3.2) contained 13 distinct clone groups, and was dominated by clone group MW12-07 (26%). None of the clone groups represented in the library showed significant sequence similarity to known genes, but substantial overlap (13/27 clones) with the August sample from the same well was observed (Fig.. 3.2).

Seven clone groups were more than 60% identical to nucleotide sequences in the GenBank database (Table 3.4). However, the majority of the putative Rieske sequences uncovered from groundwater DNA extracts were novel. Thirty-six of the 110 clones (33%, or 19 clone groups [55% of total]) formed an independent unclassified lineage (Fig. 3.2). These sequences were most closely related to each other and showed no close matches in the Genbank database. To validate the functional significance of this lineage, 13 deduced amino acid sequences were aligned and compared to Rieske sequences from known aromatic hydrocarbon degraders (Fig. 3.4). In comparisons of the deduced amino acid sequences, MW12-03 and MW12-18; MW12-10, MW12-11, 12-12, MW36-08, MW12-20, MW12-04 encoded identical peptides. Despite the short length of the amplicon, the alignment of the deduced amino acid residues from the amplified region of the Rieske-type iron sulfur shows that residues conserved in Rieske centers from known PAH degraders were conserved in the unclassified clones (Fig. 3.4). Even though the aligned region does not contain known catalytically important

residues, the high level of sequence conservation compared to known dioxygenases, combined with inferred conservation of key amino acids in the flanking primer binding regions, suggests a functionally important novel lineage. The best match for clone group MW12-07, the most abundant sequence from this gene cluster found in the clone libraries, was to clone family S4 identified by Ní Chadhain *et al.* (Table 3.4; 2006). The closest matching characterized sequence was from *Sphingomonas wittichii* RW1 (67% amino acid identity). Several members of the *Sphingomonas* genus degrade PAHs (Leys *et al.*, 2004).

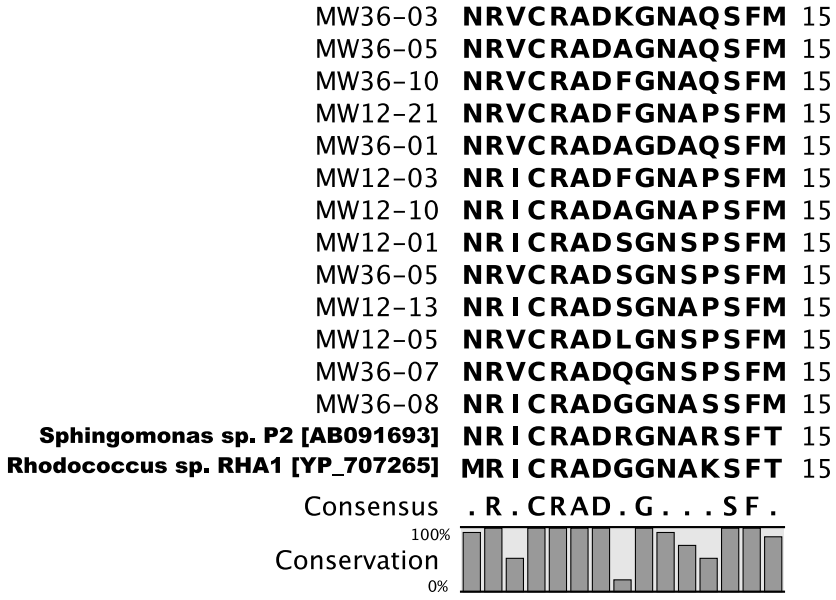


Figure 3.4: Alignment of deduced amino acid sequences representing 19 unclassified sequences found in clone libraries prepared from Well 36 and Well 12. Two reference sequences are shown in bold. Residues with 100% consensus are shown below the alignment. The bars indicate the percent conservation at each position in the alignment. Sequence nomenclature: MW=monitoring well, number=clone number.

Known *nag* gene sequences were abundant in 3 of the 4 Rieske clone libraries (Fig. 3.2). In total, 9 clones (MW36-14) matched *nag* gene sequences from *Comamonas* and *Ralstonia* and 13 clones (MW36-09) matched *nagAc* from *Polaromonas*

naphthalenivorans strain CJ2 (Fig. 3.2). Only 1 clone group, the unclassified clone group MW36-15, was represented by more clones than the latter lineage (16 total). In addition, 7 clone types (35 clones, 32% of the total) appear to be associated with a Gram-positive lineage including *nar* and *nid* genes (MW12-15, MW36-15, MW36-13, MW12-16, MW12-06, MW12-09, MW12-14; Fig. 3.2). Sequences from this cluster dominated the libraries from August, in particular.

Isolates directly obtained from groundwater at the South Glens Falls site have consistently included *Pseudomonas* hosts of *nah* genes (e.g., Bakermans and Madsen, 2002b). Surprisingly, the archetypal *nah* genes were not detected in this Rieske gene survey. In previous 16S rRNA gene clone library analyses, *Pseudomonas* species, and indeed γ-proteobacteria overall represented a minor proportion of the community in contaminated waters (Bakermans and Madsen, 2002b; Yagi *et al.*, 2009a). The extent and nature of PCR bias introduced by the highly degenerate primer set is unknown. However, recent functional assessments suggest that *nah* genes may occur at low frequency in contaminated soil and sediment habitats (Rhee *et al.*, 2004; DeBruyn *et al.*, 2007; Gomes *et al.*, 2007).

These results indicate: (i) spatial and temporal variability in the relative abundance of various oxygenase genes in the groundwater microbial communities; (ii) the prevalence of *nag* genes compared to other known NDO systems at the site; and (iii) a surprising absence of *nah*-type NDOs in the analyzed samples, possibly due to PCR bias. The high diversity and abundance of novel sequences with apparent similarity to known dioxygenase genes suggest that unrecognized PAH degradation pathways may be important in the functioning of the aromatic hydrocarbon degrading microbial communities at the South Glens Falls site.

3.4.3 Quantification of naphthalene dioxygenase genes in contaminated groundwater

Only a small subset of genes sampled by the Rieske fragment survey (Fig. 3.2, Table 3.4) are available in current databases as full sequences; of these, *nag* genes were the most frequently detected sequences (Fig. 3.2). We developed a set of quantitative competitive PCR assays to assess the apparent temporal variability in *nag* gene abundance, and to compare the abundance of *nag* genes and the expected very low abundance archetypal *nah* genes (see above) in the groundwater microbial communities. The internal standard used in the cPCR method enables comparison of multiple samples in the presence of varying levels of PCR inhibitors and primer titration effects of nonspecific competitive DNA binding sites typically associated with complex environmental DNA extracts.

PCR products of the expected size were amplified from total DNA extracted from samples collected in August, November and December 2006, and May 2007. The numbers of *nah* and *nag* gene copies were determined and expressed as copies l^{-1} groundwater. The *nag* gene abundance ranged from below detection to 3×10^5 copies l^{-1} (Fig. 3.5). Because *nah* gene sequences were not obtained in the Rieske clone libraries, we expected to see relatively low copy numbers in the groundwater samples. Surprisingly, the *nah* genes were quite abundant, though copy numbers ranged from below detection to maximum copy numbers about 10-fold higher than observed for *nag* genes (3×10^6 copies l^{-1}; Fig. 3.5). Furthermore, the *nah* genes were present in all of the samples from MW36. In MW12, *nag* genes dominated in August and May and *nah* genes dominated in November and December (Fig. 3.5).

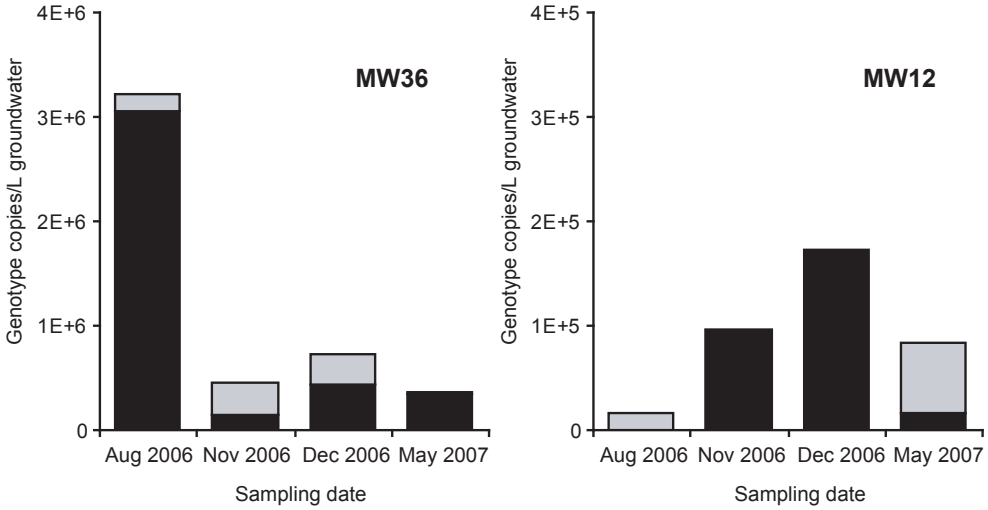


Figure 3.5: Abundance of naphthalene degradation genotypes and proportion of each gene type contribution to total bar height (*nah*-type, black; *nag*-type, gray) in well 36 and well 12 on 4 sampling dates. Copies of target per liter of groundwater (shown as average of duplicate assays) were determined by quantitative competitive PCR for samples collected on four dates (August 2006, November 2006, December 2006 and May 2007). The *nag* and *nah* distinction is based on both intragene sequence difference and adjacent genes. The *nah* type targeted a 750 bp region from *nahAa* to *nahAc*. The *nag* type targeted a 888 bp region spanning *nagH* and *nagAc*.

The *nah* and *nag* gene abundance was substantially higher in well 36 than in well 12, suggesting enrichment of particular naphthalene degrading microbial populations. The naphthalene dioxygenase gene abundance fluctuated in both well 36 and well 12. Real-time PCR has been reported to demonstrate a positive correlation between PAH concentration and normalized PAH degradation gene copy numbers (Cébron *et al.*, 2008). However, DeBruyn *et al.* (2007), using real-time quantitative PCR, demonstrated substantial uncorrelated variation in *nahAc* gene abundance in contaminated sediment. We sought but could not find obvious geochemical parameters (e.g., concentrations of O_2 or contaminants) that might be controlling the observed variation in *nag* and *nah* gene abundance. Aromatic hydrocarbons are typically present, not singly, but as

complex mixtures in the environment. Furthermore, PAH bioavailability in groundwater may be decreased significantly by, for instance, the presence of dissolved organic matter (Kan and Tomson, 1990). Mixed substrate growth, including simultaneous utilization of mixtures of PAHs, is expected in natural environments (Lendenmann *et al.*, 1996; Ihssen and Egli, 2005; Wick *et al.*, 2003; Dean-Ross *et al.*, 2002). Importantly, dissolved and particulate organic carbon are potential alternate substrates for groundwater microbes (Munster, 1993). Under carbon-limiting conditions, *Ralstonia pickettii* has been shown to utilize benzene in the presence of more easily utilizable substrate succinate (Bucheli-Witschel *et al.*, 2008), and *Pseudomonas putida* was shown to simultaneously assimilate glucose and toluene (del Castillo and Ramos, 2007). We speculate that in the presence of extremely low bioavailable PAH concentrations, bacteria are co-utilizing several carbon compounds (Johnsen *et al.*, 2005; López *et al.*, 2008). This could explain variation in NDO-harboring microbial population densities in the absence of obvious changes in naphthalene concentration.

3.4.4 Detection of expressed naphthalene dioxygenase genes

Based on quantified structural gene abundances in the August 2006 samples of wells 12 and 36 (Fig. 3.5) we predicted that cDNA in the well 36 microbial community would include both *nah* and *nag* sequences, and that well 12 cDNA would be dominated by *nag* sequences. Therefore, we used degenerate primers targeting the evolutionarily conserved large NDO (*nah/nag*)subunit gene to examine well 12 and well 36 cDNA from August 2006. RT-PCR amplicons of the expected size were obtained from both samples. From two clone libraries, a total of 34 clones were screened by RFLP analysis. We sequenced

11 clones representing the 3 observed RFLP types (MWnah01, MWnah08, MWnah10; Fig. 3.6). Phylogenetic analysis of the cloned sequences showed that they were, as expected, *nag* and *nah* genes associated with hosts from Gram-negative β- and γ-proteobacteria (e.g., *Comamonas, Pseudomonas*; Fig 3.6). Consistent with results of the structural gene analysis (Fig. 3.5), all of the 18 clones screened from well 12 represented MWnah01, a sequence closely related (98% amino acid sequence identity) to *pahAc* from *Comamamonas testosteroni* strain H. This sequence has been previously detected in expressed mRNA pools at the study site

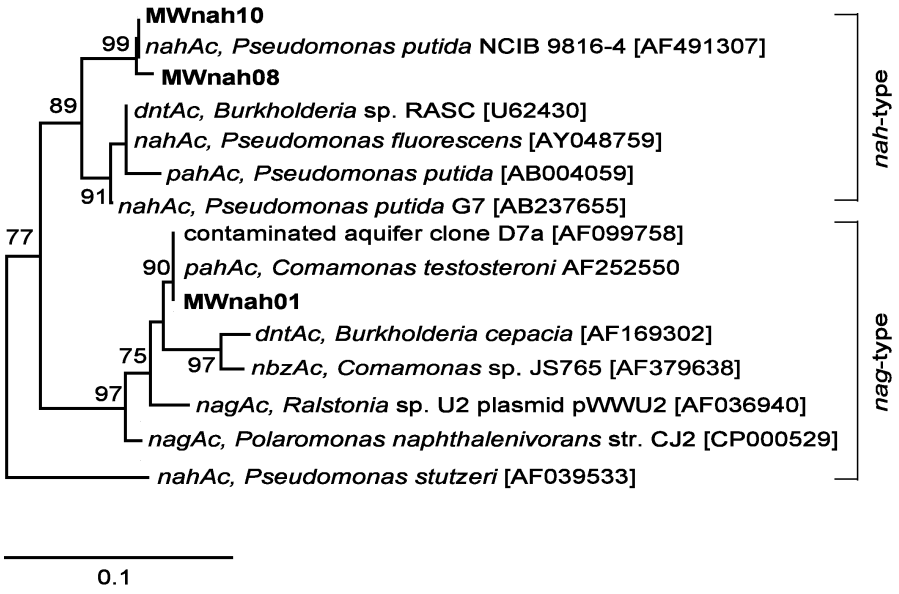


Figure 3.6: Phylogenetic analysis of mRNA transcripts found in Well 12 and Well 36 for aerobic naphthalene degradation (*nahA, nagA*; A). For 2 clone libraries generated from cDNA from Well 12 and Well 36 in August 2006, a total of 34 clones were screened by RFLP analysis. Representatives of 3 resulting RFLP types were sequenced and 138 deduced amino acid sites were used in the alignment. Neighbor-joining analysis was performed using CLUSTAL X. Sequences from this study are marked in bold type. Bootstrap values ≥70% from 100 replicates are indicated at the nodes. GenBank accession numbers for reference sequences follow the clone or organism name. The primers targeted a 446 bp region within both *nahAc* and *nagAc*; thus, capturing sequence variation between the two genotypes.

(Wilson *et al.*, 1999). Eight of the 16 cDNA clones screened from well 36 were also related to *pahAc* from *Comamonas testosteroni* H. The remaining well 36 clones grouped with *nahAc* from *Pseudomonas putida* NCIB 9816-4 (99% amino acid sequence identity).

These results indicate that *ndo* genes were actively expressed in groundwater microbial communities with low ambient levels of naphthalene and oxygen. It is not possible to make a quantitative assessment of relative mRNA transcript frequencies based on the RT-PCR data. Although a naphthalene bioreporter was reported to show induction of nah genes at a naphthalene concentration of 0.35 μM (Heitzer *et al.*, 1992), it has been hypothesized that relatively high concentrations of PAH are necessary to induce transcription of *nah* type genes, whereas prior reports suggest that *nag*-type genes may be actively transcribed in response to lower levels of substrate (Sanseverino *et al.*, 1993; DeBruyn *et al.*, 2007; Huang *et al.*, 2009). Indeed, *nag* genes were more abundant and expressed in well 12, with lower naphthalene concentrations. A recent report showed that *nah* genes were expressed at detectable levels only at naphthalene concentrations of 30 μM and higher (Huang *et al.*, 2009). At our study site, *nah* gene expression was detected in well 36, despite a relatively low ambient naphthalene concentration of 1.3 μM. We detected expression of *nag* genes in groundwater with ambient naphthalene concentrations as low as 0.6 μM. The mRNA transcript profile for well 36 suggests coexisting microbial populations harboring both *nah* and *nag* genes for naphthalene degradation.

Induction of aromatic compound degradation pathways usually occurs at micromolar ranges of substrates (Tropel and van der Meer, 2004). Constitutive expression of catabolic genes is considered an advantage in oligotrophic envi-

ronments because it increases capacity to react quickly to transiently available nutrients (Ihssen and Egli, 2005). It is known that *nah* genes are expressed constitutively in pure cultures (Guerin and Boyd, 1995; Burlage *et al.*, 1990). This study, however, does not enable a distinction between constitutive and substrate induced expression of microbial *nah* and *nag* genes.

3.5 Concluding remarks

We have shown that genes encoding diverse, uncharacterized naphthalene degrading enzymes are present in our contaminated aquifer study site (Fig. 3.2). Some of the dominant sequences are not associated with known hosts; thus, current probes targeting archetypal degradation genes may not accurately reflect biodegradation potential. Based on the occurrence of structural genes (Fig. 3.2), microorganisms related to *Rhodococcus* and *Sphingomonas* may be important functional players in site biodegradation processes. Surprisingly, the broad-specificity Rieske dioxygenase gene-detection assay indicated the prevalence of *nag* gene-encoded NDO systems and seemed to under-detect *nah*-type genes. Substantial temporal variations in *nah* and *nag* gene abundances were assessed over a 10-month period using quantitative competitive PCR (Fig. 3.4). In situ expression of *nag* and *nah* naphthalene dioxygenase genes was demonstrated in these aquifer samples featuring low ambient levels of naphthalene. The results of this study confirm that *nag* and *nah* genes (long studied from cultivated hosts) have environmental relevance. The persistent co-existence of these functionally redundant alleles may eventually provide insights into the selective pressures that operate on their hosts that mediate natural attenuation processes.

CHAPTER 4

SUBSURFACE MICROBIAL COMMUNITIES INVOLVED IN *IN SITU* NAPHTHALENE CATABOLISM CARRY OUT A COMPLETE N CYCLE VIA DISSIMILATORY NITRATE REDUCTION TO AMMONIA AND NITRIFICATION[3]

[3]Jane M. Yagi, C.M. DeRito, J.M. Suflita, L. Geig, C.-O. Jeon, and E.L. Madsen.

4.1 Abstract

Microbial processes are crucial for ecosystem maintenance, yet their documentation in complex open field sites is a challenge. Integration of multiple factors that cause and respond to biogeochemical change (e.g., dynamics of site chemistry, key functional genes, related mRNA transcripts expressed *in situ*, and associated metabolites) is rare. In a companion paper, we presented 16 years of chemical monitoring data tracking responses of a groundwater ecosystem to organic contamination (aromatic compounds) associated with coal-tar waste and we analyzed small-subunit rRNA (from *Bacteria, Archaea, and Eukarya*) genes in the native microbial community which revealed that diverse food webs had developed in response to the contamination and were dynamic in time. Here we report monitoring over 10 months in 3 site wells that document fluctuations in geochemical characteristics that include nitrate, ammonia, sulfate, sulfide, and methane. We also completed laboratory incubations of site water and sediment designed to assess naphthalene biodegradation under methanogenic, sulfate-reducing, iron-reducing, manganese-reducing, nitrate-reducing, and aerobic conditions. Naphthalene metabolism was observed only under aerobic conditions. To address the discrepancy between the laboratory results and the companion report's field monitoring showing disappearance of aromatic pollutants in this anaerobic field site, we measured biomarkers extracted from site waters (metabolites of anaerobic naphthalene metabolism and expressed mRNA transcripts) selected to document aerobic and anaerobic transformations of ammonia, nitrate, and aromatic contaminants. GC/MS detection of 2 carboxylated naphthalene metabolites and sequences of transcribed benzylsuccinate synthase, cytochrome *c* nitrite reductase, and ammonia monooxy-

genase genes show that anaerobic metabolism of aromatic compounds and that both dissimilatory nitrate reduction to ammonia (DNRA) and nitrification were in progress *in situ*. To our knowledge this is the first report of DNRA occurring in a subsurface habitat and also the first report to simultaneously link both ammonia formation (via DNRA) and destruction (via nitrification) to achieve full cycling of nitrogen in any habitat. In our view this site has progressed from a heterotrophic (carbon-oxidizing) microbial ecosystem to one that accumulates reduced metabolic end products (ammonia, sulfide, methane) and these, in turn, support lithotrophic and other microbial populations that would otherwise be absent from the site.

4.2 Introduction

All biosphere habitats (e.g, soils, sediments and both freshwaters and oceans) can be viewed as complex chemical mixtures in thermodynamic disequlibrium (Stumm and Morgan, 1996). Non-photosynthetic microorganisms (particularly members of the *Archaea* and *Bacteria*) colonizing these habitats generate metabolic energy by linking the transfer of electrons from reduced substrates (electron donors; e.g., ammonia, methane, sulfide, carbohydrates, hydrocarbons) to oxidized substrates (electron acceptors; e.g., O_2, nitrate, Fe^{3+}, sulfate; see White, 2000; Madsen, 2008). These microbially-mediated biogeochemical reactions have been formally recognized as a variety of metabolic processes (e.g., respiration, nitrification, denitrification, sulfate reduction, anaerobic methane oxidation, anaerobic ammonia oxidation, biodegradation of organic pollutants; Falkowski *et al.*, 2008; Stumm and Morgan, 1996) essential for biosphere maintenance, particularly the cycling of carbon, nitrogen, sulfur, and other elements.

Documenting the occurrence, ecological impact, and dynamic relationships between these biogeochemical processes poses many challenges (e.g., Konneke *et al.*, 2005; Raghoebarsing *et al.*, 2006; Lam *et al.*, 2007; Michaelis *et al.*, 2002).

Whenever possible, strategies for documenting biogeochemical change involve mass-balance approaches that quantitatively link materials subject to a given metabolic process (e.g., consumption of carbon substrate) to formation of metabolic byproducts (e.g., CO_2). However, owing to the open nature of many natural systems (ocean waters, rivers, soils) convergent lines of evidence from a variety of approaches (e.g., analytical chemistry of metabolites, molecular biology of genes and messenger RNA) are often needed to understand site biogeochemistry (Michaelis *et al.*, 2002; Madsen, 2008; Weiss and Cozzarelli, 2008). Direct detection of mRNA in environmental samples has increasingly become an effective approach for documenting the biogeochemical activity of microbial communities in field sites (McGrath *et al.*, 2008). This approach has been manifest in at least 3 ways: (i) RT-PCR-based targeting of specific functional genes – e.g. naphthalene dioxygenase (Wilson *et al.*, 1999); Fe(II) uptake protein (O'Neil *et al.*, 2008), RuBisCo (Wawrik *et al.*, 2002), or the anammox and denitrification processes (Lam *et al.*, 2007); (ii) creation and analysis of large community cDNA libraries of expressed genes (e.g., Botero *et al.*, 2005; Poretsky *et al.*, 2005; Bailly *et al.*, 2007; Frias-Lopez *et al.*, 2008); and (iii) direct pyrosequencing of total community derived cDNA (Gilbert *et al.*, 2008; Urich *et al.*, 2008).

A companion manuscript (Yagi *et al.*, 2009a) established that the contaminated subsurface study site examined here has a 16-year record of diminishing contaminant concentrations and that site conditions support dynamic microbial communities, including an extensive eukaryotic food chain. Data reported

here: (i) further establish that the physiological setting for contaminant biodegradation fluctuates geochemically; (ii) show in laboratory-based incubations that naphthalene is biodegraded under only aerobic, not 5 types of anaerobic, physiological conditions; and (iii) reveal, via a survey of expressed mRNA, that microbial processes consume aromatics and carry our a complete nitrogen cycle via DNRA and nitrification.

4.3 Materials and Methods

4.3.1 Site and groundwater sampling

The contaminated site is a rural area in South Glens Falls, NY, where coal tar waste was buried in the early 1960s (For additional site description and sampling procedures, see Yagi *et al.*, 2009a).

4.3.2 Geochemical analyses

Alkalinity, dissolved oxygen, temperature, methane, total organic carbon, and concentrations of nitrate, ammonia, sulfate, sulfide and Fe^{2+} were measured using the methods of (Bakermans *et al.*, 2002).

4.3.3 Microcosms examining naphthalene biodegradation

Groundwater from well 36 was sampled in October 2001 for biodegradation assays. To prevent contact between water samples and air (Yager *et al.*, 1997;

Hohnstock-Ashe *et al.*, 2001) receiving glass canning jars (2 L for water samples, 0.5 L for sediments) were flushed on-site with nitrogen gas and, after being filled, sealed without air bubbles. Freshly-gathered, oxygen-free subsurface material (5 m depth) adjacent to well 36 was obtained with a commercially operated Geoprobe hydraulic coring machine. The sediment was immersed in anaerobic groundwater prior to being sealed in a canning jar with no headspace air. The groundwater and sediments were placed on ice and maintained at 4 °C until being dispensed to serum bottles in an anaerobic hood (Coy Laboratory Products, Grass Lake, MI; N_2 headspace) within 2 days, as previously described (Yager *et al.*, 1997; Hohnstock-Ashe *et al.*, 2001). Triplicate sterile 125 ml serums received 72 ml of well 36 groundwater +10% (w/v) well 36 sediment from 2 m below the water table. Poisoned controls were prepared by adding 1 ml of poison (5% HCl, 0.25 M $HgCl_2$) to serum bottles prior to sample addition. To promote aerobic or anaerobic respiration (nitrate-, sulfate-, manganese-, iron-reduction, or methanogenesis) the serum bottle microcosms were amended with air, or anaerobic preparations of 1.6 mL of 1 M KNO_3, 1.6mL of 1 M $NaSO_4$ plus 0.8 mL of 5% Na_2S (reducing agent), 1.6 mL of 1 M MnO_2 slurry, 0.8 mL of 2 M FeOOH slurry, or 0.8 mL of 5% Na_2S, respectively (Yager *et al.*, 1997; Hohnstock-Ashe *et al.*, 2001). The serum bottles were incubated at ambient ground temperature (10 °C) in the dark with no shaking. At various times, the slurries or water were sampled (1 ml) with a syringe. After extraction with 1 ml ethyl acetate (Bakermans and Madsen, 2002b), the samples were analyzed by GC/MS (Bakermans and Madsen, 2002b; Wilson and Madsen, 1996) with a Hewlett-Packard Model 6890 Series II gas chromatograph equipped with a 30 m x 0.25 mm x 0.25 μm film thickness, HP-5MS (5% phenylmethyl siloxane; Hewlett-Packard) fused silica capillary column connected to a Hewlett-Packard

Model 5973 quadrupole mass-selective detector operated at an electron energy of 70 eV and a detector voltage of 1700. A splitless injection was used, with a 1 min delay before septum purge. The carrier gas was helium (linear velocity of 30 cm/s). Injector and detector temperatures were 250 °C and 300 °C respectively.

4.3.4 Metabolite sampling and analysis

The aerobic metabolite 1,2-dihydoxy-1,2-dihydronaphthalene (naphthalene *cis*-dihydrodiol; Wilson and Madsen, 1996) and the anaerobic metabolite of naphthalene, 2-carboxynaphthalene (Gieg and Suflita, 2002) were sought in freshly collected well waters using previously successful procedures (Gieg and Suflita, 2002). For the *cis*-dihydrodiol metabolite, ENVI-Chrom P solid-phase extraction tubes (6 ml; Supelco, Bellefonte, PA) were used on site to concentrate the analyte from 2L of well water. For carboxylated anaerobic metabolites, 2L water was extracted four times in 100 ml ethyl acetate; pooled extracts were concentrated by rotary evaporation under N_2 to 50 ml. Extracts were concentrated under nitrogen, dehydrated, derivatized with BSTFA, and analyzed by GC/MS, as above. Our assays used single ion monitoring mode, with m/z 244, 229, and 127 for the *cis*-dihydrodiol and we analyzed the carboxylated metabolites in scanning mode.

4.3.5 Nucleic acid sampling and analysis

For DNA and mRNA transcript analyses, cells were concentrated from groundwater, frozen on-site, and nucleic acids were extracted and analyzed as pre-

Table 4.1: PCR primers used in this study

Targeted domain or process	Target gene	Primers	Sequence (5′-3′)	Reference
Ammonia-oxidation (bacteria)	*amoA*	amoA-1F* amoA-2R	GGGGHTTYTACTGGTGGT CCCCTCKGSAAAGCCTTCTTC	(Rotthauwe *et al.*, 1997)
Ammonia-oxidation (archaea)	*amoA*	Arch-amoAF Arch-amoAR	STAATGGTCTGGCTTAGACG GCGGCCATCCATCTGTATGT	(Beman and Francis, 2006)
Denitrification	*narG*	narG1960f narG2659r	TAYGTSGGCARGARAA-3 TTYTCRTACCABGTBGC	(Smith *et al.*, 2007)
Methane oxidation	*mxaF*	mxa1003f mxa1561r	GCGGCACCAACTGGGGCTGGT GGGCAGCATGAAGGGCTCCC	(Inagaki *et al.*, 2004)
Methanogenesis	*mcrA* *mcrA*	MLf MLr ME1f ME2r	GGTGGTGTMGGATTCACACART-AYGCWACAGC TTCATTGCRTAGTTWGGRTAGTT CGMATGCARATHGGWATGTC TCATKGCRTAGTTDGGRTAGT	(Luton *et al.*, 2002) (Inagaki *et al.*, 2004)
Anaerobic hydrocarbon degradation	*bssA*	7772f 8546r	GACATGACCGACGCSATYCT TCGTCGTCRTTGCCCCAYTT	(Winderl *et al.*, 2007)
Dissimilatory nitrate reduction to ammonia	*nrfA*	nrfAF1 nrfA7R1	GCNTGYTGGWSNTGYAA TWNGGCATRTGRCARTC	(Smith *et al.*, 2007)

viously described (Yagi *et al.*, 2009a). Groundwater microorganisms were gathered and frozen on site for nucleic acid extraction, as previously described (Wilson *et al.*, 1999; Yagi *et al.*, 2009a). DNA fragments encoding *amoA, narG, mxaF, mcrA, nrfA* and *bssA* genes of archaea and bacteria were PCR amplified using primer pairs listed in Table 4.1. PCR amplification was performed using ThermoStart™ DNA polymerase (ABgene®) on a PTC-200 DNA Engine thermocycler (MJ Research). Replicate reactions consisted of 0.5-25 ng DNA template in 25 μl volumes and previously described cycling conditions (references in Table 4.1). Pooled PCR products were resolved by agarose gel electrophoresis and purified using the QIAquick® Gel Extraction Kit (Qiagen). Gene amplicons showing the expected size were cloned with either the TOPO® TA Cloning Kit (Invitrogen) or the StrataClone™ PCR Cloning Kit (Stratagene) as described by the manufacturers. Total RNA was treated with DNase I (Invitrogen) and converted to cDNA using random hexamer-primed reverse transcription with

SuperScript™ III reverse transcriptase (Invitrogen), according to the manufacturer's instructions. Functional biomarker gene cDNA clone libraries were generated for representative samples from Well 36 and Well 12 collected in 2006: three *bssA* cDNA clone libraries for samples collected in August and November; one archaeal and one bacterial *amoA* gene cDNA clone library for Well 36 in August; two *nrfA* clone libraries from August samples. Clone libraries were screened, and representative inserts were sequenced and analyzed as described above for SSU rRNA libraries. For *mxaF* and *narG* genes, amplified fragments from DNA from well 36 were cloned as described above. Eight and nine clones, respectively, contained inserts of the expected size. These clones were screened by RFLP and 4 representative sequences were obtained.

4.3.6 Nucleotide sequence accession numbers

The nucleotide sequence data reported here have been submitted to GenBank under accession no. FJ810622 to FJ810667.

4.4 Results

4.4.1 Contaminants are reductants that fuel microbial reactions

Data gathered from 3 on-site wells (well 8, well 36, well 12) on 3 occasions spanning 10 months (2006 to 2007) document fluctuating alkalinity, and concentrations of sulfate, nitrate, and reduced substrates that include ammonia, methane, sulfide, Fe^{2+}, and total organic carbon (Table 4.2). The key observation is that

Table 4.2: Time-dependent geochemical[a] characteristics of site well waters.

Characteristic	Well 12			Well 8		Well 36		
	Sampling times			Sampling times		Sampling times		
	8/06	12/06	5/07	8/06	5/07	8/06	12/06	5/07
Nitrate	2.1	2.7	1.8	1.8	1.7	2.3	1.8	1.6
Ammonia	0.13	0.12	0.15	bd	0.002	0.2	0.12	0.09
Sulfate	100	99	480	110	590	150	51	67
Sulfide	0.01	0.003	0.001	0.01	0.003	0.1	0.02	0.014
Methane	nd	2	nd	nd	nd	nd	21	nd
Fe^{2+}	0.14	0.28	0.06	0.09	0.02	0.45	0.14	0.06
Alkalinity (as mg $CaCO_3$)	140	460	nd	130	nd	76	140	460
Total organic carbon	5.9	29	nd	nd	nd	7.3	40	nd

[a] Average of duplicate determinations. All units μg ml^{-1} except methane (μg l^{-1}); bd = below detection; nd = not determined.

nearly all of these analytes fluctuate substantially over time within a given well. A previous report showed that from 1999 to 2001 on-site concentrations of methane, sulfide, hydrogen gas, Fe^{2+}, Mn^{2+}, and alkalinity were elevated in the contaminated wells in comparison to an upgradient control well (Bakermans and Madsen, 2002a). The new data (Table 4.2) confirm the prior findings (presence of methane, sulfide, Fe^{2+}). However, concentrations of sulfate, nitrate, and ammonia have risen approximately 10-fold during the intervening 6 years and nearly all groundwater constituents were temporally variable at different sampling times (Table 4.2). Accompanying depletion of O_2 (Yagi *et al.*, 2009a) and nitrate (Bakermans and Madsen, 2002b) provided evidence of *in situ* aerobic and anaerobic metabolism fueled by the coal tar-waste constituents. Besides influx via water flow, possible endogenous sources of sulfate, ammonia, and nitrate for the groundwater include sulfide oxidation, dissimilatory nitrate reduction to ammonia, and nitrification, respectively. Physiological electron donors for on-site microbial reactions include coal-tar-derived organic compounds, native organics in the dissolved organic carbon pool, methane, ammonia, and sulfide. The complementary set of potential physiological electron acceptors includes O_2, nitrate, sulfate, oxides of Fe and Mn (on aquifer solids), and CO_2. We hy-

pothesize that microbial metabolism of aromatic pollutants has resulted in the observed accumulation of reduced metabolic end products.

4.4.2 Laboratory-based biodegradation occurs via aerobic processes

Given the prevailing oxygen-deprived conditions (reported previously; Yagi *et al.*, 2009a; Bakermans and Madsen, 2002b), we implemented a series of laboratory based biodegradation assays designed to assess the propensity of the endogenous microbial community to metabolize native contaminants under a wide variety of physiological conditions. Our focus was on slurries of freshly gathered water and sediment from well 36 because it had the highest contamination levels and was anaerobic at depth (Yagi *et al.*, 2009a). In the aerobic serum-bottle incubations, ambient naphthalene was fully metabolized within 1 day (Fig. 4.1, panel A). No significant biodegradation activity was observed over the 110-day incubation period in treatments fostering 5 alternate terminal electron-accepting processes (denitrification, Fe-oxide reduction, Mn-oxide reduction, sulfate reduction, and methanogenesis (Fig. 4.1, panels B-F). These data indicate that oxygen-based metabolism of aromatic compounds is the most robust mechanism by which the native groundwater microbial community destroys contaminants, and contrast with field observations of anaerobic loss of contaminants (Yagi *et al.*, 2009a). Acknowledging that field processes may depend on conditions unattained in laboratory incubations, we utilized field-based metabolite and mRNA assays to resolve this discrepancy.

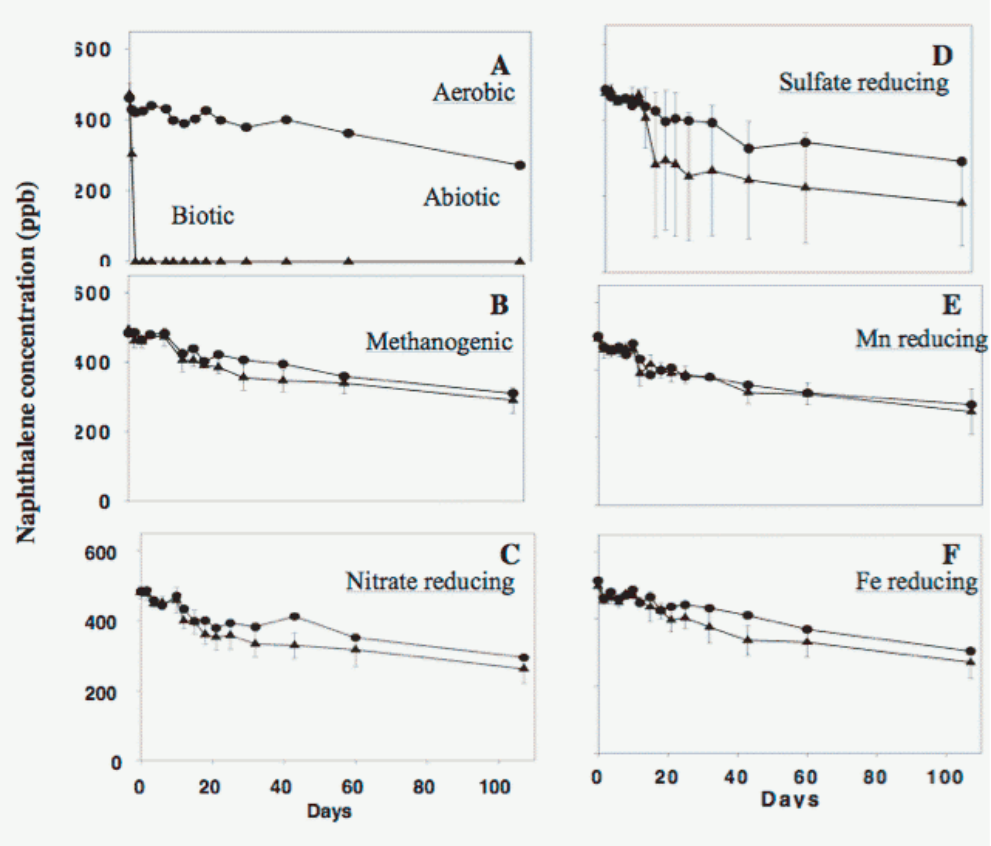


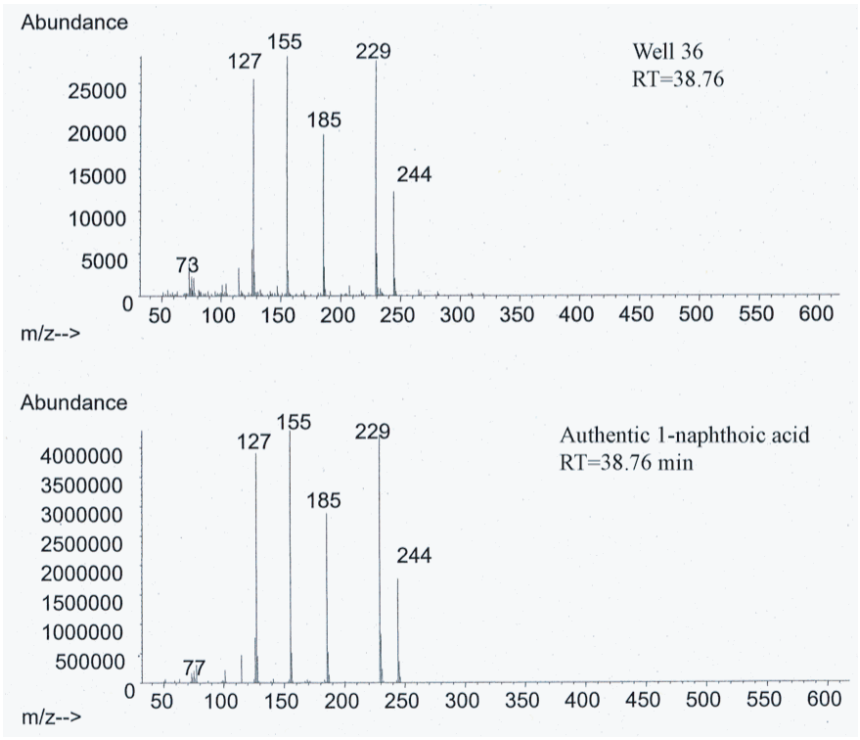
Figure 4.1: Naphthalene biodegradation in microcosms containing Well 36 groundwater and subsurface sediment adjacent to Well 36 in treatments (10 °C) designed to favor aerobic metabolism (A), methanogenesis (B), nitrate reduction (C), sulfate reduction (D), manganese-oxide reduction (E), iron-oxide reduction (F). Closed triangles (▲) represent data from viable samples, while closed circles (●) represent data from poisoned controls. Data points represent the average of triplicate samples; error bars represent standard deviations.

4.4.3 Detection of metabolites

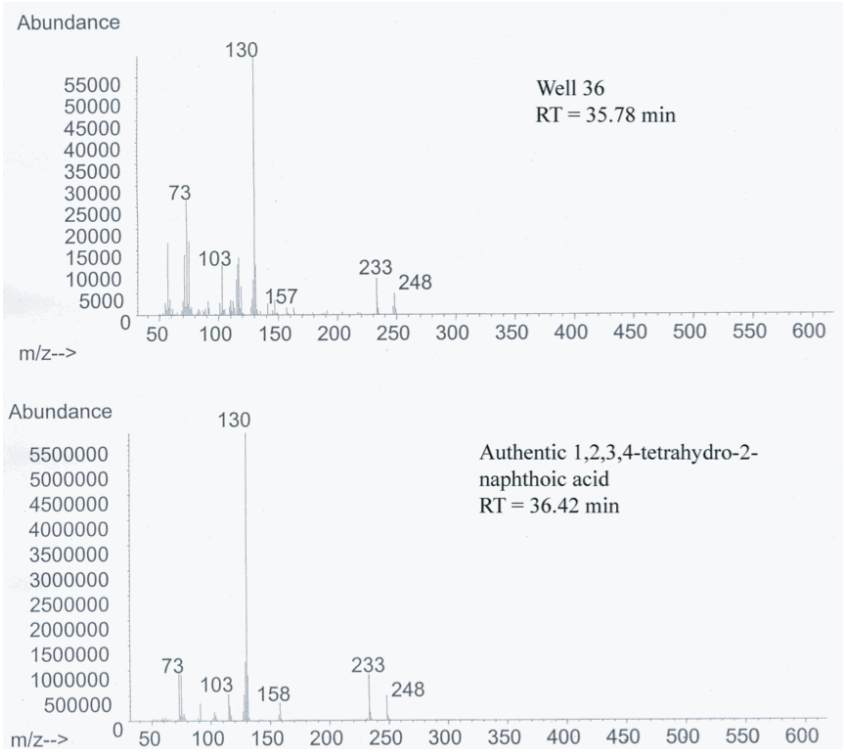
Detection of metabolites is an insightful, independent means of assessing in situ metabolic activity (Griebler *et al.*, 2004; Young and Phelps, 2005; Gieg and Suflita, 2002). In 1995, our research group successfully documented the occurrence of a transient, unique metabolite of aerobic naphthalene metabolism (1,2-dihydoxy-1,2-dihydronaphthalene) at concentrations ranging from 0.06 ppb

Figure 4.2: Intermediary metabolites of naphthalene biodegradation found in site well waters. A. Mass spectra of authentic 1-napthoic acid and 1-napthoic acid extracted Well 36. Matching mass spectra and retention times between authentic standard and metabolite in well water (absent in control well) proves that anaerobic metabolism of naphthalene was in progress in situ. RT, retention time. B. Mass spectra of another well water metabolite (retention time 35.78 min) and authentic 1,2,3,4-teterahydro-2-napthoic acid (retention time 36.42 min) extracted from Well 36. The matching mass spectra and differing retention times strongly suggest that the metabolite is 1,2,3,4-tetrahydro-1-napthoic acid, an isomer of the standard which is a signature compound for anaerobic naphthalene metabolism (Gieg and Suflita, 2002). RT, retention time.

A


Abundance
127
155
229
185
244
73
Well 36
RT=38.76
m/z-->
Abundance
127
155
229
185
244
77
Authentic 1-naphthoic acid
RT=38.76 min
m/z-->

B


Abundance
130
73
103
157
233
248
Well 36
RT = 35.78 min
m/z-->
Abundance
130
73
103
158
233
248
Authentic 1,2,3,4-tetrahydro-2-naphthoic acid
RT = 36.42 min
m/z-->

to 45 ppm in 7 on-site monitoring wells (Wilson and Madsen, 1996). To corroborate our earlier observations and extend them toward anaerobic processes, we used liquid- and solid-phase extraction combined with GC/MS to concentrate and document the presence of the *cis*-dihydrodiol-type aerobic metabolite (above) and key carboxylated indicators of anaerobic naphthalene metabolism (Young and Phelps, 2005; Gieg and Suflita, 2002; Griebler *et al.*, 2004) in waters from the 3 contaminated wells (8, 12, and 36) and an uncontaminated well (background control; for site map, see Yagi *et al.*, 2009a) in January 2001 and November 2005. Our assay, using authentic standards derivatized with BSTFA, was sensitive to ~1 ng/l. In single ion monitoring mode using m/z 228, 127, plus 116 and 244, 229, and 127 as indicator ions (for the diol and carboxylates, respectively), we were unable to detect evidence for any metabolites in samples analyzed in January, 2001. However, when the assay for the carboxylated compounds was carried out in November, 2005, we successfully found both 1-naphthoic acid and very strong evidence for 1,2,3,4-tetrahydro-1-naphthoic acid in well 36 but not in the uncontaminated background control well (Fig 4.2). Thus, while transient intermediary metabolites for both aerobic and anaerobic metabolism of naphthalene have been found on site, their patterns of occurrence vary temporally. This observation is fully consistent with the dynamic state of the site indicated by both geochemical (Table 4.2; Yagi *et al.*, 2009a) and community composition (Yagi *et al.*, 2009a) data.

4.4.4 Expressed genes prove *in situ* microbial metabolism of aromatic hydrocarbons and ammonia

In situ aerobic metabolism of naphthalene has previously been demonstrated via detection and sequencing of transcripts of the naphthalene dioxygenase gene, *nahAc*, in site groundwater (Wilson *et al.*, 1999; Bakermans and Madsen, 2002a). Hypothesizing their prominent roles in the biogeochemical repertoire of the study site, we explored six additional key processes: anaerobic catabolism of aromatic hydrocarbons, aerobic oxidation of ammonia (nitrification), anaerobic reduction of nitrate (the first step in denitrification), dissimilatory reduction of nitrate to ammonia (DNRA), methanogenesis, and aerobic methane oxidation. DNA and mRNA pools extracted from site waters were screened for individual functional genes (*bssA* for anaerobic catabolism of aromatic hydrocarbons; *amoA* (Archaea) and *amoA* (Bacteria) for ammonia monooxygenase; *narG* for the first step in denitrification; *nrfA* for the second step in DNRA, *mcrA* for methyl-coenzyme-M-reductase in methanogenesis, and *mxaF* for the second step in aerobic methane oxidation). PCR amplification from extracted DNA was successful for all targeted genes except *mcrA*. Reverse transcriptase-based detection of expressed genes was successful for *bssA, nrfA* and both bacterial and archaeal *amoA*. Because our sampling locations and times were limited, an absence of gene amplicons in pools of DNA and mRNA did not necessarily mean a given process is absent from the site.

DNA sequences *narG*, encoding a membrane-bound nitrate reductase, is one of nearly a dozen biomarkers widely used to assess denitrification in anaerobic habitats (Smith *et al.*, 2007). The 3 representative partial DNA sequences we

obtained from well 36 (Table 2) were 82-85% identical to genes from the well-characterized denitrifiers, including *Geobacter metallireducens*. Site waters also harbored partial sequences 98% identical to the *mxaF* gene of *Methylocystis* (Table 4.3); *mxaF* is one of several established biomarkers for aerobic methane oxidation (Inagaki *et al.*, 2004).

Table 4.3: Closest relatives in GenBank for partial *narG* (652 bp) and *mxfA* (554 bp) DNA sequences obtained from site waters

Targeted gene	Sequence name	Best BLAST matches and closest match in cultured organisms	Maximum identity	Accession number
narG	JMYnarG02	Uncultured bacterium clone RT-250_16 *Geobacter metallireducens* GS-15	84% 82%	DQ481115 CP000148
narG	JMYnarG03	Uncultured bacterium partial *narG* *Geobacter metallireducens* GS-16	88% 84%	AM408519 CP000148
narG	JMYnarG05	Uncultured bacterium clone GRAMO27 *Methylobacterium* sp. 4-46	85% 81%	AY955194 CP000943
*mxf*A	JMYmxaF01	*Methylocystis* sp. 5FB2	98%	EF212330

Sequences of expressed genes The recognition of benzylsuccinate synthase (BssA) as a pivotal enzyme in anaerobic metabolism of aromatic hydrocarbons, especially toluene, has driven recent investigations of the diversity of *bssA*-related genes in environmental samples (Beller *et al.*, 2002; Winderl *et al.*, 2007). We successfully detected *bssA* transcripts in the mRNA pool from wells 12 and 36 (November 2005 and August 2006). The expressed environmental *bssA* sequences (total of 95 clones and 12 RFLP types; 24 sequences in Fig. 4.3A fell into 5 clades, including sequences grouping with genes hosted by β- and δ-proteobacteria, and a cluster associated with the unidentified *bssA* F2 group previously reported (Winderl *et al.*, 2007).

Members of the Bacteria and Archaea both carry out nitrification reactions,

Figure 4.3: Phylogenetic analysis of mRNA transcripts found in site waters for 3 key site metabolic processes: anaerobic aromatic hydrocarbon degradation (*bssA*; A), aerobic metabolism of ammonia (*amoA*, B) and anaerobic dissimilatory reduction of nitrate to ammonia (*nrfA*; C). The alignment is based upon 250, 137, and 209 deduced amino acid sites, respectively. Neighbor-joining analyses were performed on CLUSTALX alignments using the ARB analysis package. Sequences from this study are marked in bold type, while reference sequences are not. Bootstrap values ≥50% from 100 replicates are indicated next to the nodes. GenBank accession numbers for reference sequences follow the clone or organism name. The scale bar represents the expected number of changes per nucleotide position.

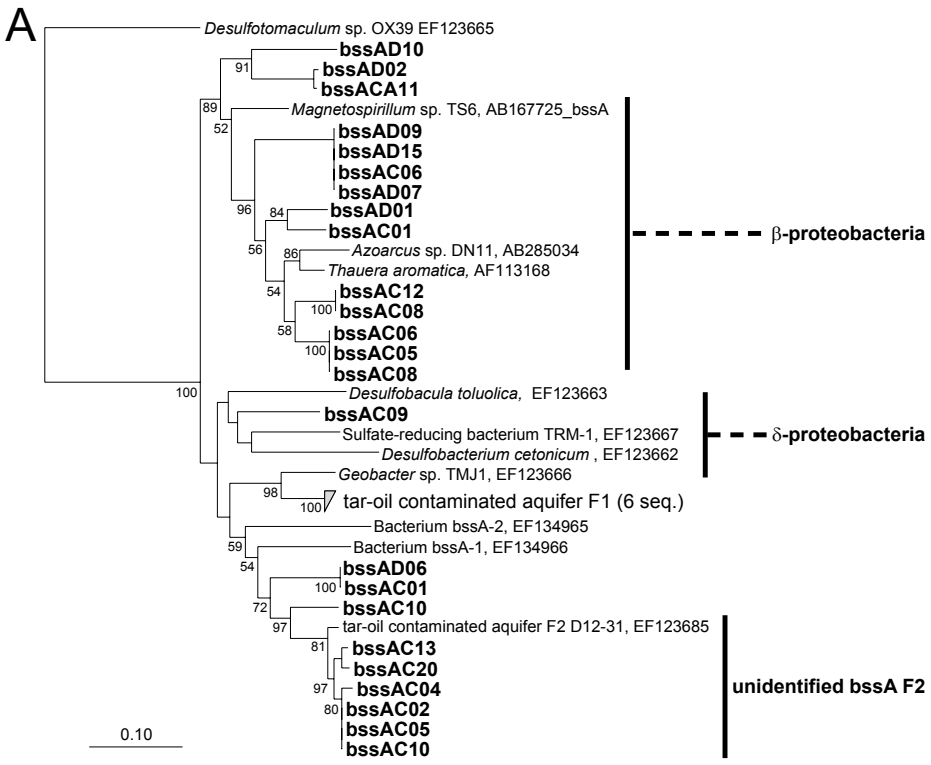

A
Desulfotomaculum sp. OX39 EF123665
bssAD10
bssAD02
bssACA11
Magnetospirillum sp. TS6, AB167725_bssA
bssAD09
bssAD15
bssAC06
bssAD07
bssAD01
bssAC01
Azoarcus sp. DN11, AB285034
Thauera aromatica, AF113168
bssAC12
bssAC08
bssAC06
bssAC05
bssAC08
β-proteobacteria
Desulfobacula toluolica, EF123663
bssAC09
Sulfate-reducing bacterium TRM-1, EF123667
Desulfobacterium cetonicum , EF123662
δ-proteobacteria
Geobacter sp. TMJ1, EF123666
tar-oil contaminated aquifer F1 (6 seq.)
Bacterium bssA-2, EF134965
Bacterium bssA-1, EF134966
bssAD06
bssAC01
bssAC10
tar-oil contaminated aquifer F2 D12-31, EF123685
bssAC13
bssAC20
bssAC04
bssAC02
bssAC05
bssAC10
unidentified bssA F2
0.10

Figure 4.3 (Continued)

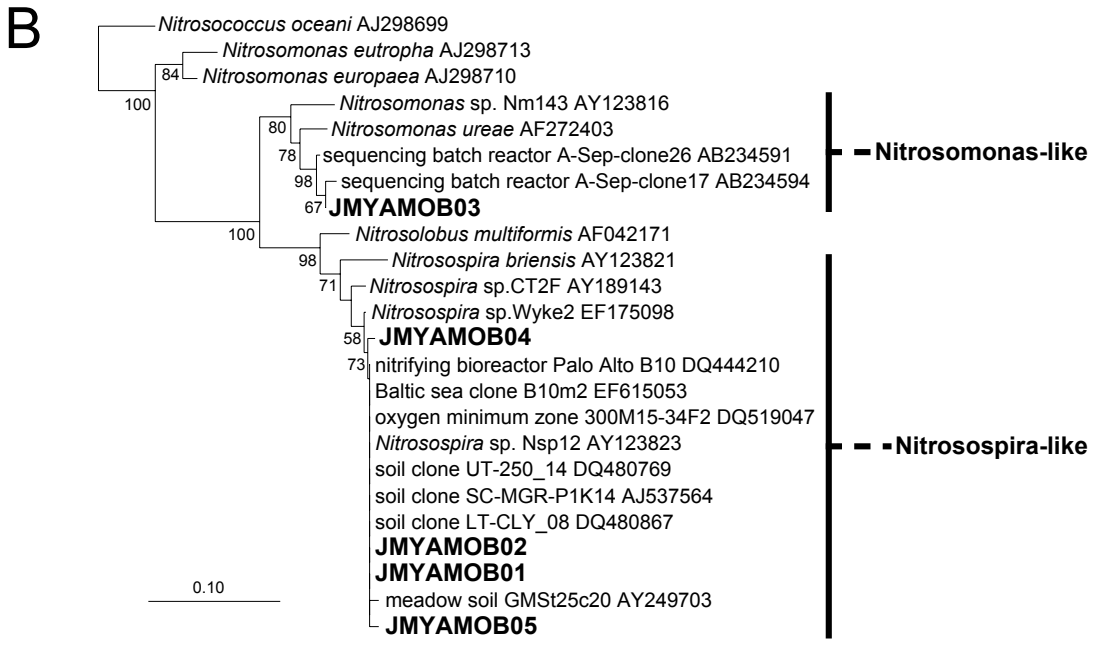
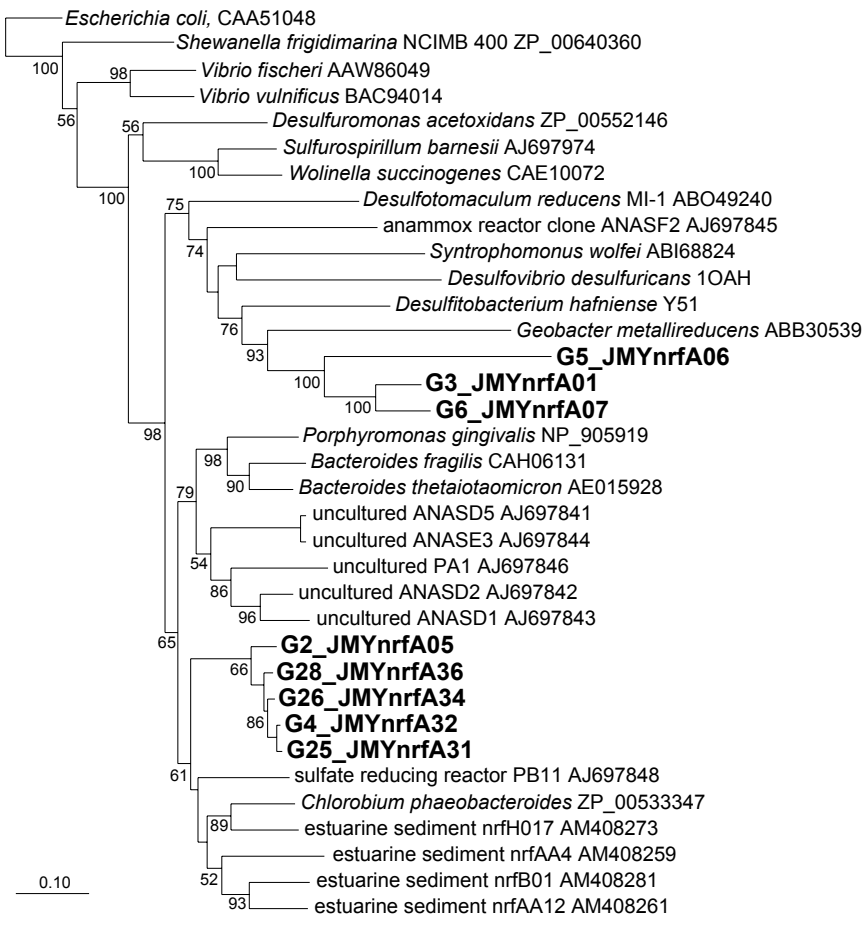


Figure 4.3 (Continued)

C



though the catalytic subunits (AmoA) of the ammonia monooxygenase enzymes they utilize are distinctive (Beman and Francis, 2006). We detected expression of both types of *amoA* in water from well 36 (August 2006). After a total of 22 archaeal *amoA* genes were cloned, only a single RFLP type was detected and its sequence matched (99% sequence similarity) that of uncultured environmental clones. From a pool of 49 bacterial *amoA* clones, we found 2 RFLP types; 5 representative clone sequences (Fig. 3B) were associated with established nitrifying genera, *Nitrosospira* and *Nitrosomonas*.

nrfA encodes an enzyme (cytochrome *c* nitrite reductase) responsible for a 6-electron transfer that reduces nitrite to ammonia in the DNRA process (Smith *et al.*, 2007; Simon, 2002). *nrfA* was expressed (total of 48 clones, 7 RFLP types, and 8 representative sequences) in water from wells 12 and 36 in August, 2006. These environmental transcripts were related to *nrfA* genes previously associated with *Chlorobium phaeobacteroides* and *Geobacter* (Fig. 4.3C; see Smith *et al.*, 2007).

4.5 Discussion

Prior to interpreting results of this study, we must declare limitations implicit in the methodologies. As mentioned by Yagi *et al.*, (2009a), aqueous-phase sampling of aquifer materials does not account for constituents strongly associated with aquifer sediments. For example, the potentially important solid-phase electron donor, sulfide (which forms metal precipitates) and solid-phase electron acceptors (Fe and Mn oxides) could not be rigorously determined. Furthermore, the heterogeneous mixture of microsites sampled in a volume of water drawn from the aquifer imposes an averaging effect, which explains how a sin-

gle sample can deliver seemingly incompatible mixtures such as oxygen and sulfide or mRNA transcripts of both aerobic and anaerobic processes. Multi-level monitoring well sampling approaches have shown that gradients at contaminant plume fringes are important to natural attenuation processes (Wilson *et al.*, 2004; Anneser *et al.*, 2008). Another key methodological consideration is that the clone-library approach used to characterize genes and mRNA transcripts has potential biases (von Wintzingerode *et al.*, 1997) and yields an incomplete census for the assayed genes (Elshahed *et al.*, 2008; Tringe *et al.*, 2005). Finally, the spatial and temporal variability in this groundwater study system, documented here and by Yagi *et al.*, (2009a): (i) largely preclude seeking high-resolution maps of microbial processes; and (ii) warrant giving high weight to detecting the presence of genes, transcripts, and metabolites and low weight to their absence. Despite the above caveats, the procedures employed allowed us to successfully use insightful biomarkers to explore on-site biogeochemical cycles, complementing information presented in the companion report on microbial community composition and long-term change in the aquifers contaminant concentrations (Yagi *et al.*, 2009a).

Expression of *bssA* gene transcripts and naphthalene-specific anaerobic metabolites (Figs. 4.2, 4.3) in groundwater unequivocally prove that the microbial community native to the study site metabolizes the aromatic contaminant compounds *in situ*. Additionally, this study has uncovered a variety of insights about a suite of biogeochemical processes. The pools of ammonia and nitrate fluctuate (Table 4.2). Given the prevalence of anaerobic conditions and transcripts for the *nrfA* gene in groundwater (Fig. 4.3C), we conclude that dissimilatory reduction of nitrate to ammonia (DNRA; Smith *et al.*, 2007; Simon, 2002) is a key mechanism of in situ ammonia production. Furthermore, the presence

of mRNA transcripts of both bacterial and archaeal *amoA* genes (Fig. 4.3B) provides clear evidence for *in situ* ammonia oxidation. Therefore, complete redox cycling of N occurs on site and it stems indirectly from coal tar waste, whose oxidation creates ammonia (via DNRA) at elevated concentrations in site waters. The presence in clone libraries of sequences representing microbial taxa known to carry out ammonia oxidation (e.g., those related to *Nitrosopumilus*; see Konneke *et al.*, 2005; Yagi *et al.*, 2009a) provides additional support for redox cycling of ammonia. The concept that emerges from the above observations is that the dynamic geochemistry of this subsurface site fosters the development of microbial populations catalyzing biogeochemical redox reactions, separated spatially and/or temporally, that are complementary to one another.

4.5.1 Prior studies on subsurface geochemical dynamics and dissimilatory nitrate reduction to ammonia

A variety of reports characterizing subsurface habitats have emphasized their dynamic geochemistry (e.g., Christensen *et al.*, 2000; McGuire *et al.*, 2000, 2002; Vroblesky and Chapelle, 1994) in space and time. It logically follows that the composition of microbial communities are also variable in space and time (Yagi *et al.*, 2009a; Bennett *et al.*, 2000; Haack *et al.*, 2004; Simon *et al.*, 2001; Hendrickx *et al.*, 2005). Although the implications of dynamic subsurface geochemistry have been reported for arsenic (e.g., Oremland and Stolz, 2003; Islam *et al.*, 2004), very little attention has been given to the biogeochemical impacts of such dynamic conditions for broader cycling of carbon, nitrogen, sulfur, and other elements. In an early report on modeling subsurface geochemical reactions, Hunter *et al.* (1998) astutely raised the issue of secondary reactions that influ-

ence biogeochemical processes in groundwater contaminant plumes. Hunter *et al.* (1998) and later Christensen *et al.*, (2000) and van Breukelen and Griffoen (2004) have advanced transport modeling of subsurface contaminants, explicitly accounting for the accrual of reduced byproducts of microbial metabolism (methane, sulfide, Fe (II), Mn(II), ammonia) and showing that these reduced substrates have the potential to foster secondary redox reactions that include aerobic methanotrophy, nitrification, anaerobic ammonia oxidation, and anaerobic methane oxidation. Evidence for the latter was reported by van Breukelen and Griffoen (2004), who found enrichment of residual $^{13}CH_4$ in the anaerobic portion of a landfill-leachate groundwater plume. To our knowledge, the present study is the first to simultaneously use both empirical field geochemical measurements and a molecular (mRNA)-based strategy to document that the secondary redox reaction, nitrification, can be coupled to DNRA, thus constituting a complete subsurface N cycle.

Dissimilatory reduction of nitrate to ammonia is the 8-electron-accepting conversion of NO_3^- to NH_3 (Tiedje, 1988; Simon, 2002). Long considered a physiological peculiarity restricted to enteric bacteria (Cole, 1996) or sulfur reducers (Schumacher and Kroneck, 1992), the ecological significance of DNRA has been underexplored and under-emphasized for decades. Early investigations by Tiedje and colleagues (Tiedje, 1988; Tiedje *et al.*, 1982; Caskey and Tiedje, 1979; Tiedje *et al.*, 1981) laid important foundations for understanding DNRA and suggested conditions in which DNRA might out compete the better understood process of denitrification (Zumft, 1997). Both are characteristic of anaerobic habitats. Denitrification (the stepwise 5-electron-accepting conversion of nitrate to N_2 gas) features a greater free energy yield per electron than DNRA. The settings in which dentrification is favored (Tiedje, 1988) may include: (i)

habitats experiencing temporary anoxia, and (ii) carbon-poor, electron-acceptor-rich conditions (a low ratio of carbon to electron acceptor). In contrast, DNRA features a slightly greater free energy yield per nitrate molecule reduced. DNRA has been hypothesized to be favored (Tiedje, 1988): (i) when electron-acceptor limitation restricts metabolism, and (ii) in carbon-rich, electron-acceptor-poor conditions (a high ratio of carbon to electron acceptor).

DNRA has been reported to occur in a variety of habitats that include anaerobic bioreactors (Mazéas *et al.*, 2008)(Mazeas et al., 2008); landfills (Berge *et al.*, 2005); freshwater sediment (Laverman *et al.*, 2007); a glacial meltwater stream (McKnight *et al.*, 2004); marine sediments (Welsh *et al.*, 2001; Laverman *et al.*, 2006; Sorensen, 1978); and anaerobic river sediments (Kelso *et al.*, 1997; Brunet and Garcia-Gil, 1996). In recent years, there has been a resurgence of efforts aimed at documenting and understanding the occurrence of DNRA in tropical forest soils (Pett-Ridge *et al.*, 2006; Silver *et al.*, 2001, 2005; Templer *et al.*, 2008) by measuring the physiological conversion of $^{15}NO_3^-$ to $^{15}NH_3$ by soil samples. These studies have shown that DNRA occurs in upland tropical forest soils where DNRA may be limited by the supply of nitrate (Silver *et al.*, 2001). Furthermore, forest management practices, while influencing gross nitrification, did not seem to alter rates of DNRA (Silver *et al.*, 2005). Also, denitrification was stronger sink for nitrate than DNRA in consistently anoxic soils, compared to variable redox soils (Pett-Ridge *et al.*, 2006). Because DNRA produces ammonia (not N_2 gas), DNRA leads to N retention in ecosystems. A great deal more needs to be learned about DNRA, especially the conditions that allow it to occur and the microbial populations that carry out the process.

4.5.2 Proposal for biogeochemical succession

Based on clear geochemical fluctuations (Table 4.2; Yagi *et al.*, 2009a) and community composition (described in the companion report, Yagi *et al.*, 2009a) and dynamic patterns of metabolic biomarkers (Figs. 2, 3; Wilson *et al.*, 1999; Bakermans and Madsen, 2002a), we hypothesize that diverse cyclic oxidation-reduction reactions for carbon, nitrogen and sulfur play prominent roles in the biogeochemistry of this subsurface contaminated site. It is certain that the coal-tar waste deposited in the 1960s provided a source of energy-rich and electron-rich organic compounds that markedly altered site geochemistry and the composition of the naturally occurring microbial community (Yagi *et al.*, 2009a; Bakermans and Madsen, 2002b). As the aromatic hydrocarbons were microbially oxidized *in situ*, metabolic byproducts (reduced forms of the electron acceptors including nitrogen gas from denitrification, ammonia from dissimilatory reduction of nitrate to ammonia, methane, and sulfide) were produced. We have presented evidence for complete *in situ* redox cycling of nitrogen: ammonia is endogenously produced and oxidized on site.

In our prior study, methane of microbial origin was found repeatedly in site waters at concentrations exceeding 50 ppm, exhibiting $\delta^{13}C$ values indicative of aceticlastic reactions (Bakermans et al., 2002); moreover, methane is still generated on site (Table 4.2). We speculate that, like ammonia, endogenous methane is likely a physiological resource that selects for populations that generate ATP by transferring electrons to O_2, nitrate, or sulfate, available in this site's dynamic geochemical setting. Consistent with the *in situ* oxidation of methane is the presence of the *mxaF* gene in the community (Table 4.3), and the prevalence in well waters of taxa strongly associated with processes linking oxidation of

methane to the reduction nitrate (Raghoebarsing *et al.*, 2006; Yagi *et al.*, 2009a; Bakermans and Madsen, 2002b).

We hypothesize that, at this study site, ammonia, methane, and sulfide reservoirs all undergo cyclic patterns of endogenous production and oxidation as the redox conditions fluctuate in time and space. We speculate that all three pools of reduced byproducts (ammonia, methane, and sulfide, created indirectly via microbial oxidation of coal-tar waste) have played hidden and perhaps important roles in site biogeochemistry by serving as electron donors sustaining otherwise unsupported groups of microbial populations, including sulfide oxidizers, nitrifyers, and both aerobic and anaerobic methane oxidizers. Therefore, we propose a succession of biogeochemical stages: from pristine aerobic heterotrophy, to a contaminated state of strong heterotrophy with accrual of reduced metabolites, to a physiologically fluctuating state that supports both heterotrophic microbial populations and chemolithotrophs that oxidize ammonia and other reduced compounds.

CHAPTER 5

THE GENOME OF *POLAROMONAS NAPHTHALENIVORANS* STRAIN CJ2, ISOLATED FROM COAL TAR-CONTAMINATED SEDIMENT, REVEALS PHYSIOLOGICAL AND METABOLIC VERSATILITY AND EVOLUTION THROUGH HORIZONTAL GENE TRANSFER[4]

[4]Jane M. Yagi, D. Sims, T. Brettin, D. Bruce and E.L. Madsen. Manuscript submitted to *Environmental Microbiology*.

5.1 Abstract

We analyzed the genome of the aromatic hydrocarbon-degrading, facultatively chemolithotrophic β-proteobacterium, *Polaromonas naphthalenivorans* strain CJ2. Recent work has increasingly shown that *Polaromonas* species are prevalent in a variety of pristine, oligotrophic environments, as well as polluted habitats. Besides a circular chromosome of 4.4 Mb, strain CJ2 carries 8 plasmids ranging from 353 to 6.4 kb in size. Overall, the genome is predicted to encode 4929 proteins. Comparisons of DNA sequences at the individual gene, gene cluster, and whole-genome scales revealed strong trends in shared heredity between strain CJ2 and other members of the *Comamonadaceae* and *Burkholderiaceae*. BLASTP analyses of protein coding sequences across strain CJ2's genome showed that genetic commonalities with other β-proteobacteria diminished significantly in strain CJ2's plasmids compared to the chromosome, especially for the smallest ones. Broad trends in nucleotide characteristics (G+C content, GC skew, Karlin signature difference) showed at least 6 anomalous regions in the chromosome, indicating alteration of genome architecture via horizontal gene transfer. Detailed analysis of one of these anomalous regions (96 kb in size, containing the *nag*-like naphthalene catabolic operon) indicates that the fragment's insertion site was within a putative MiaB-like tRNA modifying enzyme coding sequence. The mosaic nature of strain CJ2's genome was further emphasized by the presence of 309 mobile genetic elements scattered throughout the genome, including 131 predicted transposase genes, 178 phage-related genes, and representatives of 12 families of insertion elements. A total of three different terminal oxidase genes were found (a putative cytochrome aa_3-type oxidase, cytochrome cbb_3-type oxidase, and cytochrome *bd*-type quinol oxidase), suggesting adapta-

tion by strain CJ2 to variable aerobic and microaerobic conditions. Sequence-suggested abilities of strain CJ2 to carry out nitrogen fixation and grow on the aromatic compounds, biphenyl and benzoate, were experimentally verified. These new phenotypes and genotypes set the stage for gaining additional insights into the physiology and biochemistry contributing to strain CJ2's fitness in its native habitat, contaminated sediment.

5.2 Introduction

P. naphthalenivorans strain CJ2 is one of nine members of the family *Comamonadaceae* (Wen *et al.*, 1999) with complete genomes sequenced to date (http://img.jgi.doe.gov/). In addition to strain CJ2, five polaromonad species have been described in detail: *Polaromonas* sp. strain JS666 (Mattes *et al.*, 2008), *P. vacuolata* (Irgens *et al.*, 1996), *P. aquatica* (Kampfer *et al.*, 2006), *P. hydrogenivorans* (Sizova and Panikov, 2007), and *P. jejuensis* (Weon *et al.*, 2008). *Polaromonas* species are most frequently associated with cold and/or nutrient poor habitats, including glacial meltwaters and sediment (Foght *et al.*, 2004; Irgens *et al.*, 1996; Liu *et al.*, 2006; Newsham *et al.*, 2004; Panikov and Sizova, 2007), cave environments (Barton, 2006), and oligotrophic freshwater lakes and aquifers (Ball and Crawford, 2006; Loy *et al.*, 2005; Page *et al.*, 2004). It has long been recognized that microbial growth in nature is defined by chronic nutrient limitation (Morita, 1988), and understanding the features of polaromonads conferring success in nutrient poor environments may provide useful insight into the nature of bacterial adaptation to oligotrophy.

Polaromonas naphthalenivorans strain CJ2, isolated from coal-tar waste contaminated sediment, carries out field biodegradation of naphthalene, the sim-

plest member of the polycyclic aromatic hydrocarbon (PAH) class of compounds (Jeon *et al.*, 2003, 2004). Recent studies suggest that many polaromonads serve important roles in pollutant degradation (Aburto *et al.*, 2009; Alfreider *et al.*, 2002; Coleman *et al.*, 2002; Connon *et al.*, 2005; Devers *et al.*, 2007; Eriksson *et al.*, 2002; Jeon *et al.*, 2003; Mattes *et al.*, 2008). The biochemical details of aerobic naphthalene catabolism have been studied in diverse Gram-negative and Gram-positive bacteria, with the archetypal *nah* genes of naphthalene degradation most commonly associated with *Pseudomonas* species (Peng *et al.*, 2008). The naphthalene catabolic gene cluster in strain CJ2 is evolutionarily related to the *nag* catabolic gene cluster in the comamonad, *Ralstonia* sp. strain U2 (Fuenmayor *et al.*, 1998; Zhou *et al.*, 2001). Although the majority of isolated naphthalene-degrading bacteria are pseudomonads (Yen and Serdar, 1988), recent functional gene-targeted molecular analyses have shown that diverse naphthalene degradation alleles, including the *nag* genes carried by *P. naphthalenivorans* strain CJ2, are prevalent in many PAH contaminated environments (Bordenave *et al.*, 2008; DeBruyn *et al.*, 2007; Dionisi *et al.*, 2004; Gomes *et al.*, 2007; Jeon *et al.*, 2003; Ní Chadhain *et al.*, 2006) and expressed (as mRNA) at a coal tar waste-contaminated site (Bakermans and Madsen, 2002a; Wilson *et al.*, 1999). Some of strain CJ2's physiological and biochemical traits, particularly those related to aromatic compound metabolism and physiological response to naphthalene, have been elucidated (Jeon *et al.*, 2006; Pumphrey and Madsen, 2007; Park *et al.*, 2007a,b). Additional genetic features characteristic of organisms capable of metabolizing naphthalene, and responsible for their survival and in situ activity in field sites remain to be explored.

We present the analysis of the whole genome sequence of *Polaromonas naphthalenivorans* strain CJ2. The genome was examined for features (e.g., oxidative

stress tolerance, horizontal gene transfer, autotrophic and heterotrophic metabolism, energy metabolism, gene regulation, signal transduction, transport, and aromatic biodegradation pathways) providing insight into the ecological fitness traits associated with microbial survival and contaminant biodegradation processes in freshwater sediment.

5.3 Materials and Methods

5.3.1 Bacterial strains

P. naphthalenivorans strain CJ2 was isolated from coal-tar waste contaminated surface sediment from South Glens Falls, NY (Jeon *et al.*, 2003, 2004). Cultures were grown in Stanier's minimal basal salts (MSB) media using 0.2% pyruvate as carbon source (Stanier *et al.*, 1966). Genomic DNA was isolated using standard protocols for bacteria (Wilson, 1987).

5.3.2 Sequencing, gene prediction and annotation

Sequencing was carried out at the Joint Genome Institute (JGI) using a combination of 3 kb, 8 kb, and fosmid DNA libraries. All general aspects of library construction and sequencing performed at the JGI can be found at http://www.jgi.doe.gov/. The Phred/Phrap/Consed software package (www.phrap.com) was used for sequence assembly and quality assessment (Ewing and Green, 1998; Ewing *et al.*, 1998; Gordon *et al.*, 1998). After the shotgun stage, reads were assembled with parallel phrap (High Performance Software, LLC). Possible misassemblies were corrected with Dupfinisher (Han and Chain, 2006) or transpo-

son bombing of bridging clones (Epicentre Biotechnologies, Madison, WI). Gaps between contigs were closed by editing in Consed, custom primer walk or PCR amplification (Roche Applied Science, Indianapolis, IN). A total of 1673 additional reactions were necessary to close gaps and to raise the quality of the finished sequence. The completed genome sequences of *Polaromonas napthalenivorans* strain CJ2 contains 47,802 reads in the chromosome and 16,182 reads (combined) for the plasmids, achieving an average of 11.0x coverage per base with an error rate of 0.00 in 100,000.

5.3.3 Comparative genomics

Comparative analyses were carried out using tools on the JGI Integrated Microbial Genomes server (http://img.jgi.doe.gov/pub/main.cgi). The IMG annotations are available at URL http://genome.ornl.gov/microbial/pnap/. The IMG Function Category comparisons tool was used for ortholog analysis. In addition, BLASTP analysis was used with cutoff values of $<10^{-5}$ (E value) and 30% identity for orthologs and $<10^{-2}$ (E value) and 20% identity for unique CDSs. The translated sequences of the predicted protein-coding genes were used for BLASTP analyses against the NCBI database. Results were parsed through a script to get best hits. Whole genome comparisons were conducted using MUMmer (Kurtz *et al.*, 2004). Cumulative GC skew analysis data were retrieved from the Comparative Genometrics website (Roten *et al.*, 2002). IS families were defined according to BLAST against the IS Finder database (http://www-is.biotoul.fr/). Transcriptional regulators and two component regulators were found using those terms in IMG gene product searches. The IslandPath program was used to identify chromosomal deviations in GC con-

tent (http://www.pathogenomics.sfu.ca/islandpath) (Hsiao *et al.*, 2003). Transporters were identified using the transporter database (http://www.tcdb.org; Saier *et al.*, 2006). Restriction/modification system characterization was based on data in REBASE (http://rebase.neb.com/ rebase/rebase.html; Roberts *et al.*, 2007). Nitrogen fixation was tested with the classical acetylene reduction assay (Burris, 1972).

5.3.4 Accession numbers

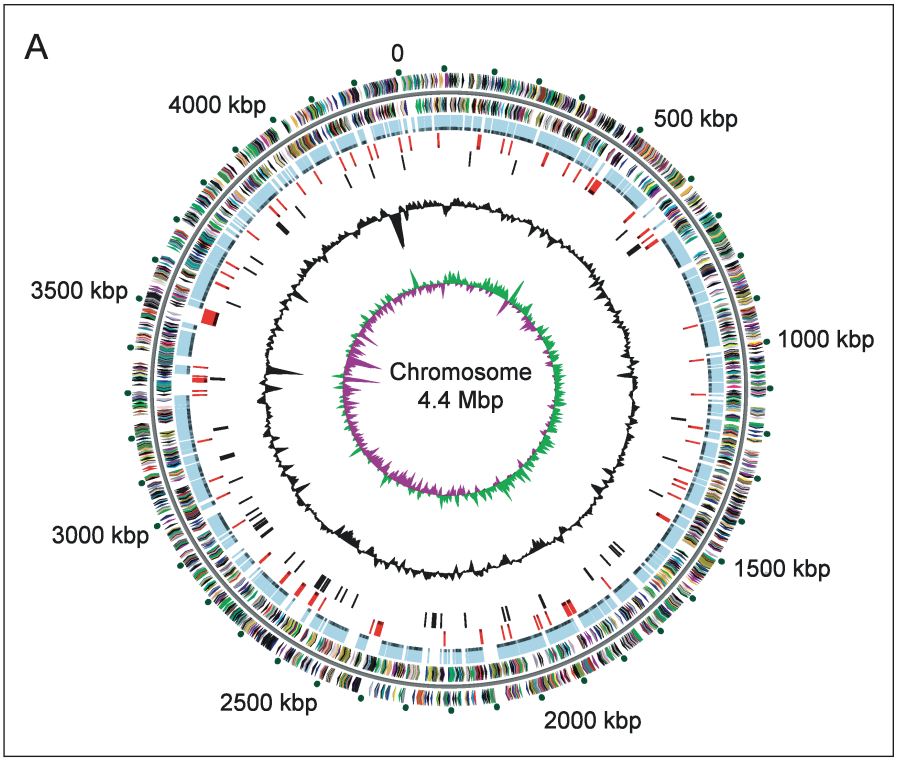
The complete nucleotide sequence of *P. naphthalenivorans* has been deposited in GenBank under accession numbers - NC_008781 (chromosome), NC_008757 (pPNAP01), NC_008758 (pPNAP02), NC_008759 (pPNAP03), NC_008760 (pPNAP04), NC_008761 (pPNAP05), NC_008762 (pPNAP06), NC_008763 (pPNAP07), NC_008764 (pPNAP08).

5.4 Results and Discussion

5.4.1 General features of the genome

The 5,366,143 base pairs (bp) genome of *Polaromonas naphthalenivorans* strain CJ2 (Phylum, *Proteobacteria*; Class, *β-proteobacteria*; Order, *Burkholderiales*; Family, *Comamonadaceae*) comprises nine circular replicons (Fig. 5.1): a single chromosome of 4,410,291 bp (G+C content of 63%) and 8 plasmids ranging from 6,459 bp to 353,291 bp in length (Table 5.1). Sequence analysis shows 88.6% of the DNA encoding proteins or stable RNAs. The entire genome encodes 4,929 candidate protein coding genes of which 3511 (71.2%) have functional predictions

Figure 5.1: Circular representation of the chromosome and 8 plasmids of *Polaromonas naphthalenivorans* strain CJ2. (A) Chromosome. Shown are (starting from the outermost ring): (i) predicted forward strand gene products and (ii) predicted reverse-strand gene products, color coded by COG classifications; (iii) CDSs with highest BLAST match to *P. naphthalenivorans* JS666; (iv) genes associated with mobile genetic elements; (v) predicted protein coding sequences associated with catabolic pathways; (vi) G+C content; (vii) GC skew (B). Eight plasmids. For each plasmid, shown are (starting from the outermost ring): (i) predicted forward strand gene products and (ii) predicted reverse-strand gene products, color coded by COG classifications; (iii) G+C content; (iv) GC skew. Circles 1-8 represent pPNAP01 through pPNAP08, respectively. The five largest plasmids are shown at 5X scale relative to the chromosome size; pPNAP06 is shown at 20X scale; pNAP07 and pPNAP08 are shown at 40X scale.


A
0
500 kbp
1000 kbp
1500 kbp
2000 kbp
2500 kbp
3000 kbp
3500 kbp
4000 kbp
Chromosome
4.4 Mbp

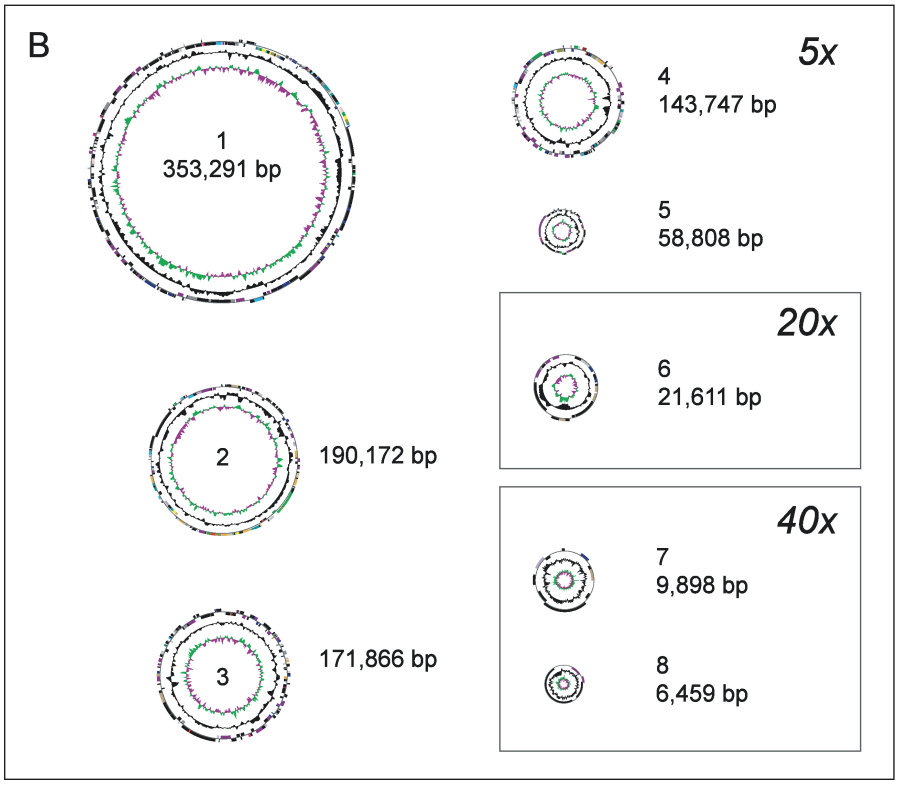

B
1
353,291 bp
2
190,172 bp
3
171,866 bp
5x
4
143,747 bp
5
58,808 bp
20x
6
21,611 bp
40x
7
9,898 bp
8
6,459 bp

Table 5.1: General features of the genome of *Polaromonas naphthalenivorans*

	Chrom.	Plasmids							
		1	2	3	4	5	6	7	8
Size (base pairs)	4,410,291	353,291	190,172	171,866	143,747	58,808	21,611	9,898	6,459
G+C content	0.63	0.58	0.58	0.57	0.59	0.6	0.56	0.57	0.52
Coding density (%)	91	84	78	77	76	81	76	63	77
tRNA	48	0	1	1	1	0	0	0	0
rRNA operons	2	0	0	0	0	0	0	0	0
Total number of CDSs	4141	305	162	156	137	52	20	11	5
Average CDS size (bp)	974	972	926	852	798	916	879	563	998
CDSs with function prediction (%)	3196 (77.2)	131 (43.0)	77 (47.5)	45 (28.8)	83 (60.6)	22 (42.3)	9 (45.0)	4 (36.4)	4 (80.0)
CDSs without function prediction	945	174	85	111	54	30	11	7	1
CDS assigned to COG (%)	3219 (77.7)	127 (41.6)	80 (49.4)	48 (30.8)	79 (57.7)	22 (42.3)	9 (45.0)	2 (18.2)	1 (20.0)
Predicted transposase genes	86	9	13	3	21	2	0	0	0

based on the COG, Pfam and InterPro databases. Of the remaining predicted proteins, 1305 (26.5%) matched only proteins of unknown function, and 113 (2.3%) predicted proteins had no database matches. The presence of at least 5 plasmids was previously known (Jeon *et al.*, 2006), and genome sequencing has revealed the existence of an additional 3 extrachromosomal elements. Eight hundred forty eight protein coding sequences (CDSs) are encoded on the plasmids, which comprise 18% of the genome (Table 5.1). The plasmids feature lower G+C content and coding density than the chromosome (Table 5.1). The US DOE JGI Integrated Microbial Genomes (IMG) annotation identified 75 pseudogenes in *P. naphthalenivorans* strain CJ2. The genome encodes two complete sets of rRNA genes and 51 tRNA genes (Table 5.1). Throughout this article, we will use locus tag notation appearing on the IMG website (http://img.jgi.doe.gov) to describe the positions of various genes under discussion.

5.4.2 Genomic comparisons

As expected, in comparisons with other genomes catalogued in the IMG database, the gene content of strain CJ2 overlaps most extensively with that of its closest sequenced relative, *Polaromonas* JS666 (Mattes *et al.*, 2008), and to a high degree with other members of the family *Comamonadaceae*, including *Rhodoferax*, *Comamonas*, *Delftia* and *Acidovorax* (Table 5.2). However, in the comparison between strain CJ2 and *Polaromonas* JS666, 1567 (31.8% of the total) genes from strain CJ2 were not found in strain JS666 at a threshold of 30% minimum identity ($1e^{-05}$ significance level). Among the 3362 genes specific to strain CJ2, 1,400 genes were assigned to well-defined functional categories (COGs, clusters of orthologous groups of proteins); of the genes assigned to COG clusters, 1103 (78.8%) were chromosomally located (Table 5.1). Genes that distinguish strain CJ2 from strain JS666 include loci involved in nutrient acquisition and storage, such as genes for transport of aliphatic sulfonates and ammonium (Pnap_0037-0040; Pnap_0178), poly-3-hydroxybutyrate depolymerase (Pnap_0127, Pnap_3566, Pnap_2756), ferritin (Pnap_0325), alkanesulfonate monooxygenase (Pnap_0560, Pnap_2418), nitrogen fixation genes, hydrogenase genes, a molybdenum transporter (Pnap_2037), a naphthalene degradation locus, sulfide dehydrogenase (Pnap_1666), and sulfite dehydrogenase (Pnap_2692).

Substantial homology was also observed in comparisons of strain CJ2 with *Burkholderia xenovorans* LB400 and *Ralstonia* spp., members of the family *Burkholderiaceae* (Table 5.2). By contrast, overlap in gene content with *Escherichia coli* K-12 was found to be relatively low, with a slightly higher proportion of

Table 5.2: Homology comparisons between the genome of *Polaromonas naphthalenivorans* strain CJ2 and 12 other bacteria

Compared organism	Taxonomic lineage (class, family)	No. of genes with homologs[a]	% of genes in strain CJ2	No. of CJ2 homologs located in the chromosome (%)	Total number of genes in compared genome
β-proteobacteria					
Polaromonas JS666	*Burkholderiales, Comamonadaceae*	3362	68.21	3069 (91.3)	5569
Rhodoferax ferrireducens T118	*Burkholderiales,Comamonadaceae*	3004	60.95	2801 (93.2)	4555
Comamonas testosteroni KF-1	*Burkholderiales, Comamonadaceae*	2992	60.7	2761 (92.3)	5454
Delftia acidovorans SPH-1	*Burkholderiales, Comamonadaceae*	2990	60.66	2776 (92.8)	6055
Acidovorax JS42	*Burkholderiales, Comamonadaceae*	2982	60.5	2704 (90.6)	4320
Methylibium petroleophilum PM1	*Burkholderiales, Genera incertae sedis*	2849	57.8	2660 (93.4)	4475
Verminephrobacter eiseniae EF01-2	*Burkholderiales, Comamonadaceae*	2715	55.08	2506 (92.3)	5050
Azoarcus sp. EBN1	*Rhodocyclales, Rhodocyclaceae*	2594	52.63	2361 (91.0)	4599
Burkholderia xenovorans LB400	*Burkholderiales, Burkholderiaceae*	3079	62.47	2814 (91.4)	8951
Ralstonia eutropha H16	*Burkholderiales, Burkholderiaceae*	3009	61.05	2772 (92.1)	6629
Ralstonia eutropha JMP134	*Burkholderiales, Burkholderiaceae*	3015	61.17	2784 (92.3)	6529
γ-proteobacteria					
Escherichia coli K-12	*Enterobacteriales, Enterobacteriaceae*	1957	39.7	1852 (94.6)	4391

[a] Common genes (given as number of genes with homologs in the organisms compared) between strain CJ2 and selected organisms, calculated using a maximum E-value of 10^{-5} and minimum identity of 30%.

homologs contained within the chromosome (Table 5.2). This chromosomal bias may be a manifestation of reduced extrachromosomal commonality, perhaps reflecting higher fluidity of plasmid genetic content between strain CJ2 and more closely related organisms (see below).

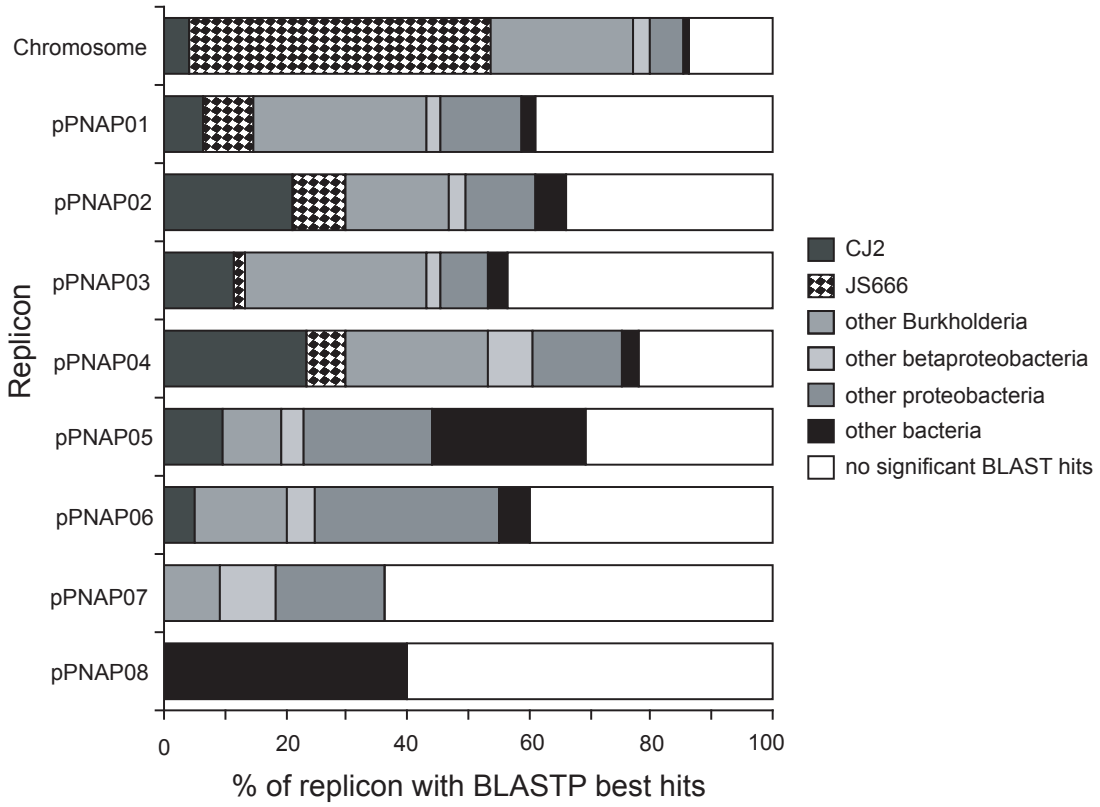


Figure 5.2: Taxonomic distribution of top BLASTP hits of translated CDSs from *P. naphthalenivorans* str. CJ2. The proportion of hits from each of the eight replicons to predicted protein coding sequences from (i) strain CJ2; (ii) the closest sequenced relative, *Polaromonas* JS666 (iii) other species within the order *Burkholderiales* (iv) other β-proteobacteria; (iv) other proteobacteria; and (vi) other bacteria are shown as shaded bars. The proportion of total translated CDSs with no significant BLASTP hits are shown as white bars.

In BLASTP analysis of translated CDSs using the NCBI nonredundant database, 4119 (83.6%) CDSs had a significant BLAST match (defined as an E value $< 2E^{-20}$, >30% sequence identity and alignment of more than 70% query sequence length; Fig. 5.2). As expected, the greatest proportion (2101; 51%)

of translated CDSs in strain CJ2's 9 replicons are shared with *Polaromonas* strainJS666 (Fig. 5.2). Overall, the majority of BLAST matches for strain CJ2's chromosome were to genes from the order *Burkholderiales* (3200 best hits). Best BLAST matches to non-*Burkholderia* protein sequences for the plasmids ranged from 0% (pPNAP08) to 53% (pPNAP04; Fig. 5.2) of predicted CDSs. In particular, *Methylibium petroleiphilum* strain PM1 and *Leptothrix cholodnii*, members of the *Sphaerotilus-Leptothrix* subcluster within the class β-proteobacteria (Nakatsu *et al.*, 2006), harbored 212 of the top hits. Overall, 548 (11% of total) translated CDSs best matched a sequence associated with an organism outside of the order *Burkholderiales*. Furthermore, 349 protein best matches to non-*Polaromonas* proteins had a high identity (at least 80%), perhaps suggestive of recent lateral gene transfer; these matches included genes encoding the biphenyl dioxygenase genes in *Burkholderia xenovorans* LB400 (Pnap_4137-4153). Other matches (Pnap_2313-2345) encode nitrogenase genes, shared in *Burkholderia* species. In addition, a gene cluster (Pnap_1954-1960) encoding hydrogenase genes is conserved in *Methylibium petroleiphilum* PM1.

Whole genome MUMmer alignment of both nucleotide and protein coding sequences for strain CJ2 and strain JS666 (Fig. 5.3) revealed chromosomal gene synteny with extensive genome-wide rearrangements (cf. Stein *et al.*, 2007), perhaps mediated by mobile genetic elements (see below). Nucleotide composition analyses (G+C content, GC skew, Karlin Signature difference, cumulative GC skew) revealed several chromosomal regions with anomalous signatures (shaded Regions I-VI, Fig. 5.3). These regions were analyzed for functional gene content and signatures of recent lateral gene acquisition by *P. naphthalenivorans*. Region I (Pnap_0537-0611) encoded several transporter gene clusters, in-

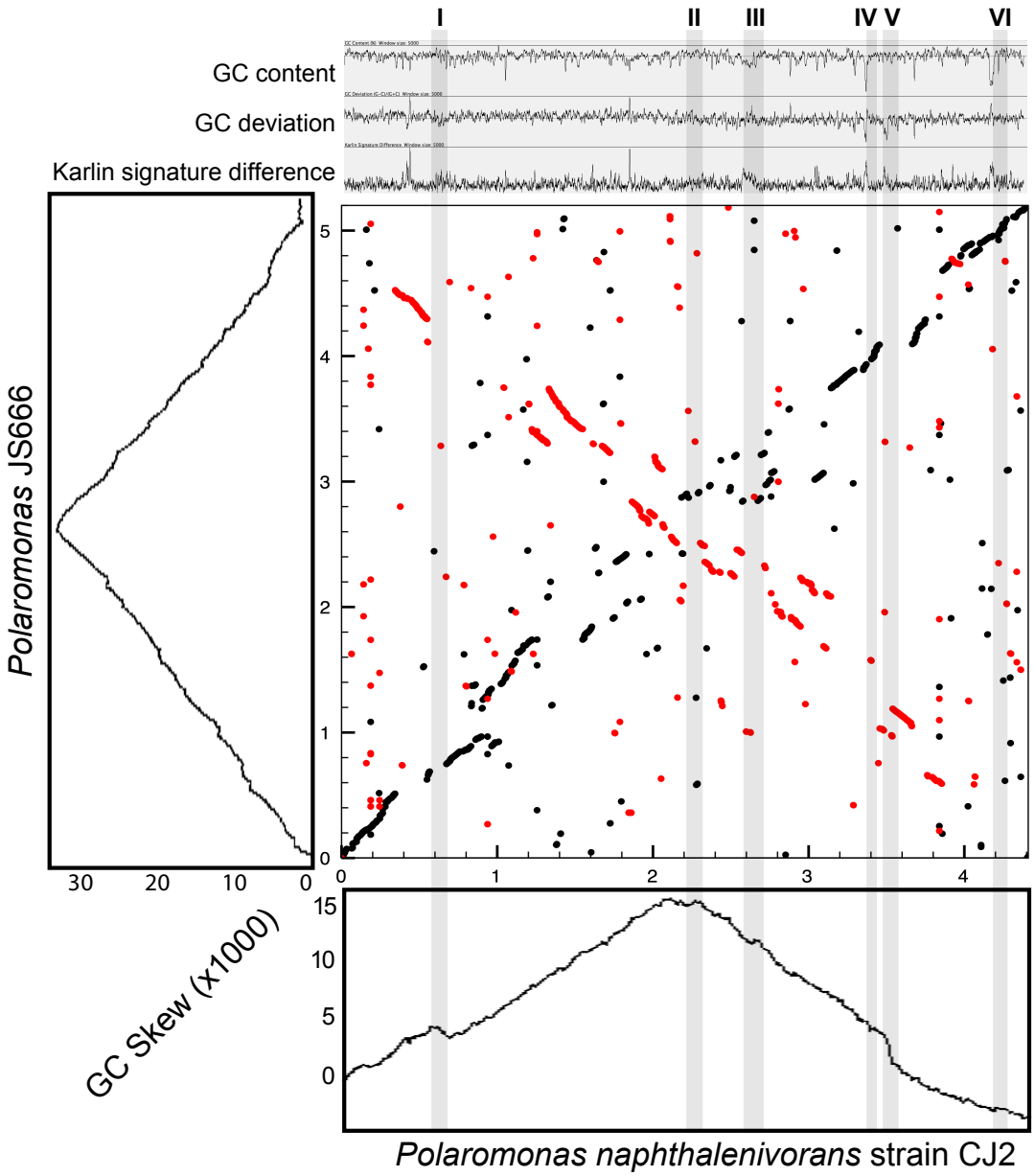


Figure 5.3: Synteny between the genomes of *P. naphthalenivorans* str. CJ2 and *Polaromonas* JS666 is shown in a dotplot comparing the chromosomes of strain CJ2 and strain JS666. The red and black circles indicated reciprocal best hits in the forward and reverse strands for amino acid regions with > 75% identity. These data suggest extensive chromosomal rearrangement in strain CJ2 in the form of reciprocal inversions. Plots of cumulative GC skew for strain CJ2 and strain JS666 are shown next to the axes. G+C content, GC deviation, and the Karlin Signature difference diagrams generated using Artemis are shown above the dotplot. At least 6 chromosomal insertions (shaded regions I-VI) in strain CJ2 are evident in these analyses (see text).

cluding genes most closely relately to homologs in the family *Pseudomonadaceae* and possibly involved with assimilation of sulfonate and sulfate esters (e.g., Pnap_0547-0554; 66-91% amino acid identity to genes in *Pseudomonas* spp. and *Azotobacter* spp.). Region II (Pnap_2075-2151) included gene clusters for degradation of benzoate and catechol. The catabolic gene arrangement appears to comprise a mosaic of homologs to genes found in γ-proteobacteria like *Pseudomonas* and *Azotobacter*, and β-proteobacteria including *Delftia acidovorans* and *Leptothrix cholodnii* (e.g., Pnap_2107-2120). Region III (Pnap_2439-2536) included a complete naphthalene catabolic operon and resembled a genomic island, with G+C content and nucleotide composition suggestive of a mosaic structure (see below). Region IV (Pnap_3159-3204) included genes annotated for exopolysaccharide biosynthesis functions. Regions V and VI encoded several phage-associated genes.

To better understand the genetic context of the naphthalene degradation gene cluster in strain CJ2, we examined chromosomal region III in detail (Fig. 5.4, Appendix A). The ORFs homologous to the flanking regions of a conspicuous 96 kb region are continuous in strain JS666 but interrupted in strain CJ2. This configuration suggests that strain CJ2 acquired genes for naphthalene catabolism via insertion of a mobile genetic element into the chromosome. Consistent with this hypothesis is the presence of *trb* (conjugal transfer) genes within the same region (Fig. 5.4). tRNA genes ($tRNA^{Leu}$ and $tRNA^{Val}$) flank each end of the region. A putative MiaB-like tRNA modifying enzyme coding sequence (Pnap_2439) adjacent to $tRNA^{Leu}$ is disrupted in strain CJ2, with truncated regions of the gene encoded on opposite ends of the putative DNA insertion site (Fig. 5.5). BLASTN analysis of the corresponding regions in strain CJ2 and strain JS666 showed significant nucleotide alignment, including the intergenic region

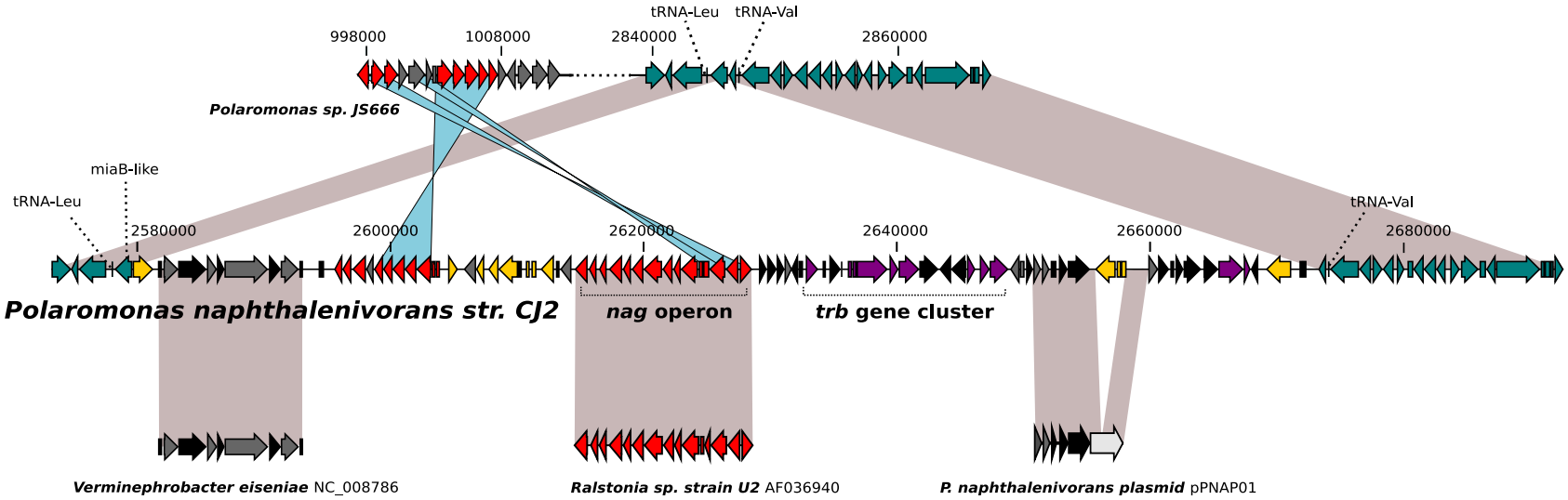


Figure 5.4: Genetic arrangement in *P. naphthalenivorans* str. CJ2 of the putative 96 kb chromosomal insertion including the naphthalene catabolic operon. Orthologous DNA regions are connected by grey boxes. Blue boxes denote orthologous regions in opposite coding directions. The scales indicate coding positions in the respective genomes. Transposases, integrases and recombinase genes are shown in yellow. Genes associated with Type IV secretion are shown in purple. Genes associated with naphthalene degradation are shown in red. Hypothetical genes are shaded black. Three highly conserved gene clusters, and their respective homologs (from *Verminephrobacter eiseniae, Ralstonia* sp. strain U2, and *P. naphthalenivorans* plasmid pPNAP01, are indicated.

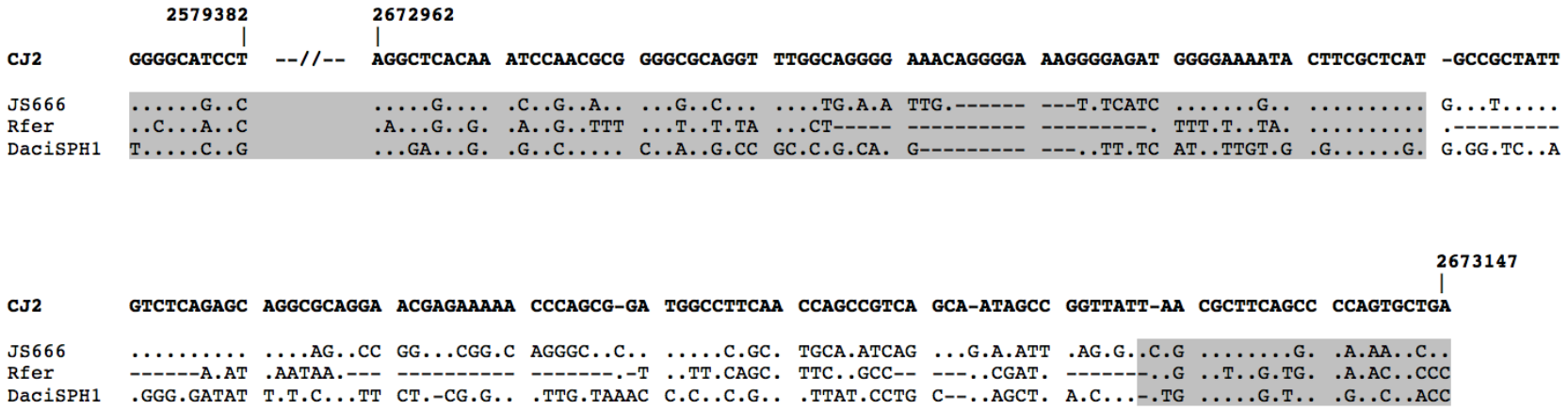


Figure 5.5: Nucleotide alignment of the truncated *miaB*-like gene sequences flanking the putative 96-kb inserted region in *P. naphthalenivorans* strain CJ2. The nucleotide sequence from *P. naphthalenivorans* str. CJ2 was compared to corresponding genomic regions in *Polaromonas* JS666 (JS666), *Rhodoferax ferrireducens* (Rfer) and *Delftia acidovorans* (DaciSPH1) using CLUSTALX alignment. The shaded residues are predicted CDSs in relatives of strain CJ2. Alignment gaps are indicated as dashes, and positions with identity to the CJ2 sequence are indicated as dots.

(Fig. 5.5). A putative phage recombinase (Pnap_2440) is encoded adjacent to the truncated 3' end of the *miaB*-like gene (Fig. 5.4). This recombinase shares 60% amino acid identity with a phage recombinase in *Acidovorax avenae* subsp. citrulli AAC00-1 (Aave_1644) that is also adjacent to a gene encoding a MiaB-like tRNA modifying enzyme. Numerous putative transposases, and integrases, and an additional recombinase are encoded within the region (Fig. 5.4; Appendix A). In addition to the naphthalene catabolic operon, two gene clusters in the region are likely candidates for recent recombination or lateral gene transfer (Pnap_2442 to 2449, with homology to *Verminephrobacter eiseniae*; and Pnap_2515 to 2520, with homology to a gene cluster in plasmid pPNAP01; Fig. 5.4). Furthermore, the right end (~59.3 kb) of the putative inserted region is flanked by a direct DNA repeat of ~250 bp (located at base positions 2,612,832 - 2,613,123 and 2,671,883 - 2,672,124). The high conservation of intergenic sequences in repeated regions suggests that relatively recent recombination has occurred. Of the remaining CDSs, 28 are annotated as hypothetical protein-coding genes (Appendix A).

5.4.3 Restriction/modification systems

According to REBASE, *P. naphthalenivorans* strain CJ2 harbors at least one type I restriction modification system, and two type IV endonucleases (Roberts *et al.*, 2007). A putative type IV Mrr endonuclease (Pnap_3732) and McrBP (Pnap_4260) methyl-specific restriction systems are also encoded, and may control entry and expression of foreign DNA (Table 5.3). Methyl-specific restriction systems mediate the introduction of methylated, hydromethylated and glucosyl-hydroxymethylated DNA. The restriction modification systems

Table 5.3: Restriction/modification systems in *P. naphthalenivorans* strain CJ2 according to REBASE

Locus tag	Restriction system	Type	Description of neighbor	Recognition seq.
Pnap_2899	PnaORF2899P	Type II	*Azoarcus* sp. EbN1	-
Pnap_3732	PnaMrrP	Type IV	*Delftia acidovorans* SPH-1	-
Pnap_4260	PnaMcrBP	Type IV	Environmental sample Sargasso Sea	-
Pnap_4283	PnaORF4283P	Type II	*Burkholderia thailandensis* E264	-
Pnap_4777 Pnap_4778 Pnap_4782	PnaORF4777P (3)	Type I	*Burkholderia xenovorans* LB400 *Polaromonas* sp. JS666 *Shewanella putrefaciens* CN-32	-
Pnap_4916	M.PnaORF4916P (2)	Type I		-
Pnap_4924	PnaORF4924P	Type I	*Desulfovibrio vulgaris* subsp. vulgaris DP4	-
Pnap_4973	M.PnaORF4973P	Type II		GTCGAC
Pnap_4987	PnaORF4987P	Type II	*Rhodoferax ferrireducens*	-
Pnap_4910	M.PnaORF4910P	Type II	Environmental sample Sargasso Sea *Burkholderia cenocepacia* HI2424 *Burkholderia cenocepacia* AU 1054	CTGCAG

may be responsible for low transformation efficiencies observed for strain CJ2 (Jeon *et al.*, 2006).

5.4.4 Plasmid analysis

COG analysis revealed functional specialization among strain CJ2's eight plasmids (Fig. 5.6). The proportion of plasmid genes in four COG cluster categories clearly varies: 3 of the 8 plasmids lack genes involved in metabolism/energy production, while plasmid pPNAP04 features a particularly high proportion of unknown hypothetical genes (Fig. 5.6). The largest plasmid is biased toward intracellular trafficking, secretion, and vesicular transport, mainly due to genes for conjugal transfer machinery (plasmid pPNAP01, Table 5.1; COG group U). These genes are found in a cluster (Pnap_4191-4204) absent in strain JS666, but similar to genes found in several organisms, including a chromosomal locus in *Burkholderia multivorans* ATCC 17616. Plasmid pPNAP01 also features a puta-

tive biphenyl degradation gene cluster and a putative phthalate degradation gene cluster similar in gene order to a cluster in the unfinished genome of *Comamonas testosteroni* KF-1 (plasmid pPNAP04) (see below, Catabolism of aromatic compounds). Plasmid pPNAP02 is biased towards inorganic ion transport/metabolism with several genes playing roles in heavy metal transport and

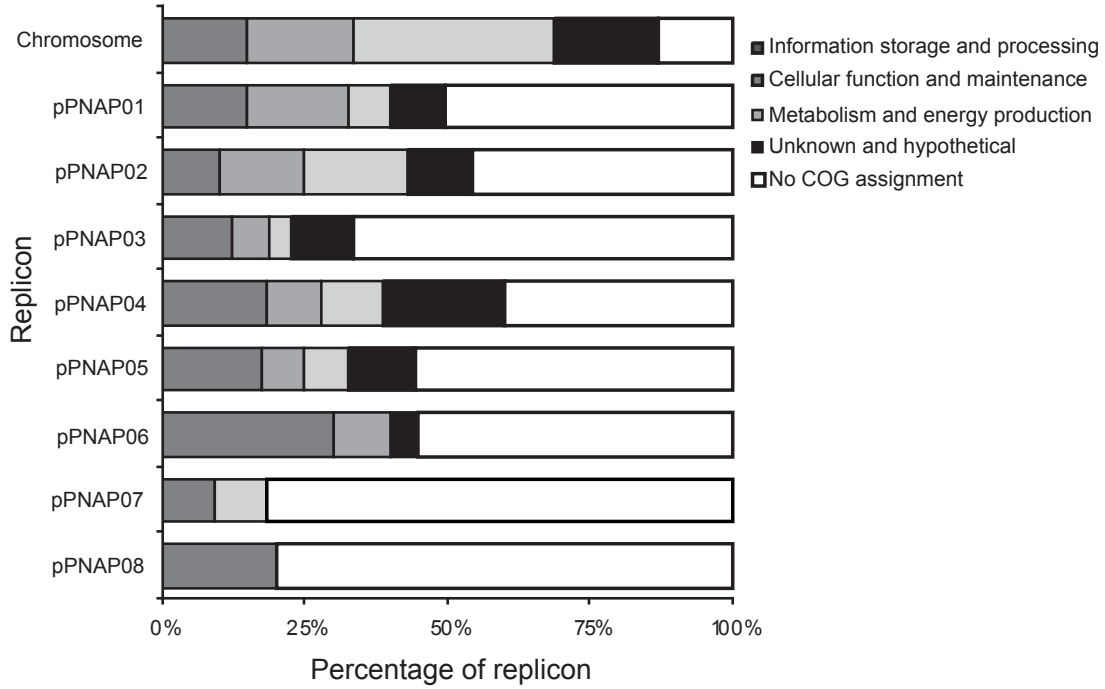


Figure 5.6: Functional distribution of genes among the nine replicons assigned to COG groups for strain CJ2. The COG assignments were grouped into four clusters: 1) Information storage and processing (J, K, L, A); 2) Cellular function and maintenance (D, V, T, M, N, U, O); 3) Metabolism and energy production (C, G, E, F, H, I P, Q); 4) Unknown and hypothetical (R, S). Genes may be assigned to more than one COG group.

detoxification, including those encoding 2 putative P-type ATPases (Pnap_4491, 4524) and 2 CzcA family heavy metal efflux pumps (Pnap_4513, 4569; COG group P; Fig. 5.6; Nies, 2003). pPNAP03 has a bias toward defense mechanisms (COG group V). Plasmid pPNAP06 encodes a putative bacterial toxin-

antitoxin (TA) system, possibly responsible for plasmid maintenance (Gerdes *et al.*, 2005). The *parA* gene encoded on plasmid pPNAP06 is highly similar to one in pXAC33 (82% ID) from several organisms including *Xanthomonas axonopodis* pv. citri str. 306.

Eight plasmids comprise a comparatively large number of independent replicons in a bacterium (Bentley and Parkhill, 2004). Phylogenetic analysis of *parA* genes has been used to gain insight into the evolution of plasmids (Bentley and Parkhill, 2004; McLeod *et al.*, 2006; Yamaichi and Niki, 2000). The chromosome and four largest plasmids (pPNAP01, pPNAP02, pPNAP03 and pPNAP04) encode distinct *parAB* loci probably responsible for DNA segregation and distribution to daughter cells during cell division (Pnap_4247/4248; Pnap_4471/4472; Pnap_4933/4934; Pnap_4692/4693). Additional homologs of the plasmid partition ATPase gene, *parA*, are encoded on three of the four remaining replicons (Pnap_4776; Pnap_4980; Pnap_4996; plasmid pPNAP07 seems to harbor a truncated *parB* gene). The clustering of individual *parA* sequences from strain CJ2 with homologs from various *Proteobacteria* suggests diverse origins and/or frequent lateral transfer of the plasmids between species (Fig. 5.7). The clustering of individual *parA* sequences from strain CJ2 with homologs from various *Proteobacteria* suggests diverse origins and/or frequent lateral transfer of the plasmids between species (Fig. 5.7). The *parA* gene of plasmid pPNAP01 (Pnap_4248) clusters with *Burkholderia* genes (75 - 76% nucleotide identity; best match to *parA* from plasmid pBMUL01 in *Burkholderia multivorans*). The *parA* gene from plasmid pPNAP02 (Pnap_4472) is most closely related to the homolog from plasmid 2 in strain JS666 (Bpro_5515; 53% nucleotide identity). No closely related homologs were found for the *parA* gene encoded on plasmid pPNAP03 (at a significance level of 30% nucleotide identity). The

parA gene from plasmid pPNAP04 clusters with homologs in plasmid RPME01 from *Methylibium petroleiphilum* PM1 (46% nucleotide identity) and plasmid 2 in *Azoarcus* EbN1 (42% nucleotide identity; Fig. 5.7). The chromosomal *parA* homolog clustered with other chromosome associated *parA* genes. The absence of *parB* and *rep* elements in some of the plasmids suggests interdependence for stable maintenance of the replicons.

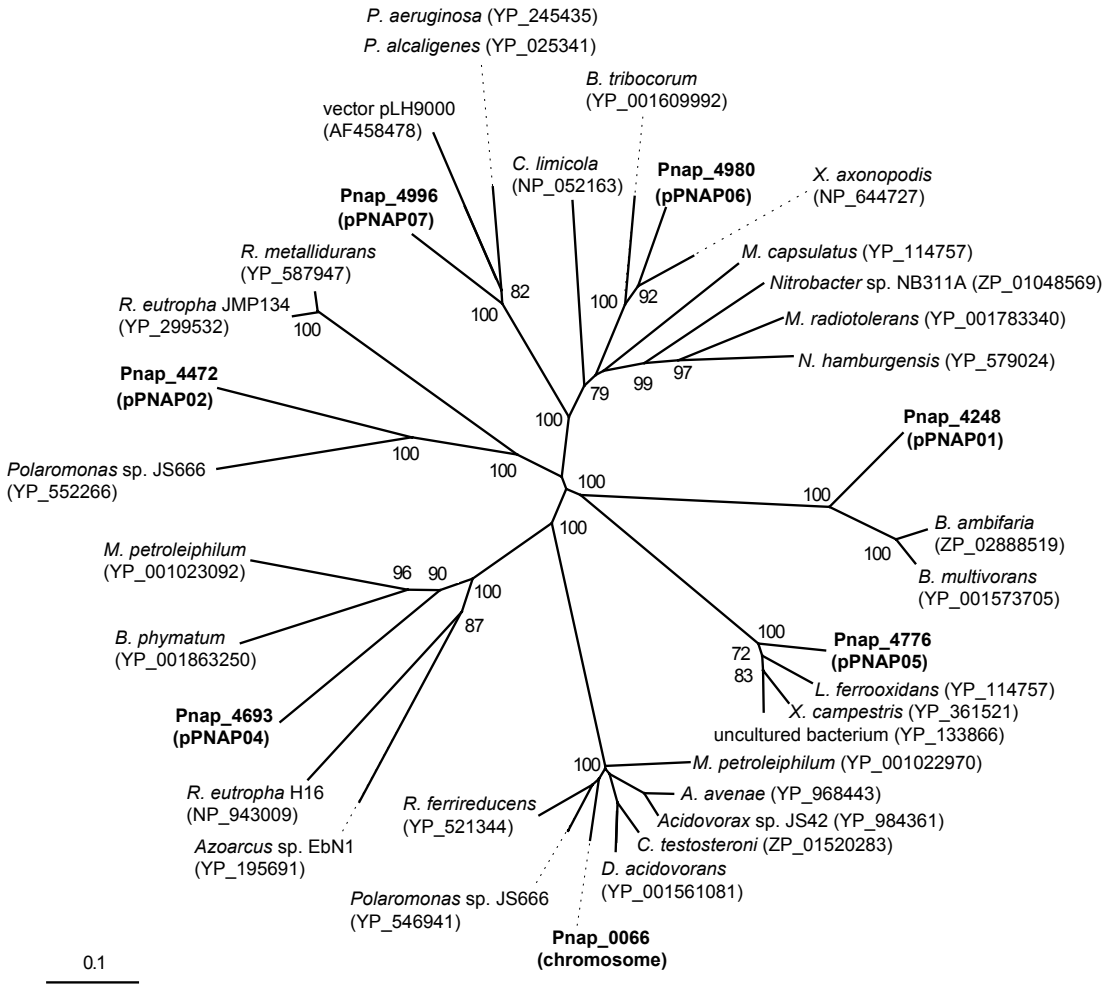


Figure 5.7: Phylogenetic relationship of 7 distinctive parA sequences (bold) from strain CJ2 plasmids. Reference sequences for parA (not bold) are included to display genetic relatedness to plasmid-partitioning mechanisms in other taxa. Trees were determined by neighbor-joining analysis using CLUSTAL X. Bootstrap values >70% are indicated at the nodes.

5.4.5 Mobile genetic elements

Mobile genetic elements are important in adaptation and evolution in bacterial genomes (de la Cruz and Davies, 2000; Top and Springael, 2003; Thomas and Nielsen, 2005; Frost *et al.*, 2005; van der Meer and Sentchilo, 2003; Nojiri *et al.*, 2004). According to the JGI annotation, strain CJ2 has 131 predicted transposase genes (86 chromosomally located, 48 on plasmids; Table 5.1). When phage-related genes are added to the tally, strain CJ2 is host to 309 mobile elements, 204 of which reside on the chromosome (Appendix B). Of the 20 insertion sequence families catalogued in ISFinder, 12 families are represented in CJ2. Members of the unclassified ISNCY are also present. Individual insertion sequence elements were repeated up to 8 times in the genome. The large number of complete or partial insertion sequences identified in the genome suggest recent genetic flux in *P. naphthalenivorans* strain CJ2. The largest family represented in strain CJ2 is IS630. Clearly evolution of strain CJ2 has been strongly influenced by genetic rearrangements mediated by mobile genetic elements (see also Table 5.1, Fig. 5.3, and the above section on genomic comparisons).

5.4.6 Signal transduction pathways and regulation

COG analysis for strain CJ2 predicts 289 (5.9%) genes involved in signal transduction (COG category K). Soil and sediment bacteria often encode a complex array of two-component systems for sensing and responding to dynamic environmental conditions. Strain CJ2 has at least 61 predicted histidine kinase associated CDSs (HPKs), 53 response regulator receiver proteins (RRs) and 4 hybrid proteins with both histidine kinase and response regulatory domains. At least

35 HPKs were paired with RRs. At least ten CDSs were grouped with PAS/PAC-domain proteins, characterized as sensors of oxygen and redox (Zhulin *et al.*, 1997). One PAS/PAC sensor signal transduction histidine kinase on the chromosome of strain CJ2 (Pnap_1953) is part of the nickel-dependent hydrogenase gene cluster (see below).

The genome of strain CJ2 encodes the major sigma factor RpoD (Pnap_1604), as well as RpoN (Pnap_3710), RpoH (Pnap_0889), and 7 extracytoplasmic function (ECF) family homologs (compared to 9 in *Polaromonas* JS666). RpoN may be involved nitrogen assimilation and environmental acclimation (Jones *et al.*, 2007), and RpoH may be important under heat stress (Yura and Nakahigashi, 1999). A homolog of the *fur* ferric uptake regulator was also identified (Pnap_3791, 89% ID JS666; see below). Quorum sensing homologs were not found in strain CJ2.

5.4.7 Central metabolism

The genome of strain CJ2 encodes the complete Embden-Meyerhof-Parnas (EMP) glycolytic pathway, the pentose phosphate (PP) pathway, the Entner-Doudoroff (ED) pathway, and the complete citric acid cycle (CAC).

5.4.7.1 Glycolysis

Genes for the complete Embden-Meyerhof-Parnas (EMP) glycolytic pathway for transformation of glucose to pyruvate are present. Glucokinase, catalyzing the first step in the EMP pathway (EC: 2.7.1.2; Pnap_3695), is encoded directly

adjacent to the pentose phosphate (PP) pathway gene for glucose-6-phosphate-1-dehydrogenase (EC: 1.1.1.49; Pnap_3694). This paired gene arrangement, and a proximally located *lysR* transcriptional regulator (Pnap_3697) suggest coordinated regulation of the EMP and PP catabolic pathways in strain CJ2. Furthermore, the gene for glucose-6-phosphate isomerase (EC: 5.3.1.9; Pnap_0655), catalyzing the second step in the EMP pathway, is clustered with the PP pathway genes for glucose-6-phosphate-1-dehydrogenase (EC: 1.1.1.49; Pnap_0653) and phosphoglyceraldehyde transaldolase (EC: 2.2.1.2; Pnap_0654), also with an adjacently situated *rpiR* transcriptional regulator. This gene arrangement, also observed in *Polaromonas* sp. JS666, *R. ferrireducens*, and two *Acidovorax species*, reinforces the hypothesis of coordination between the EMP and PP pathways in strain CJ2. The sequence of the EMP pathway-specific 6-phosphofructokinase (EC: 2.7.1.11; Pnap_3557) gene is most similar to the corresponding gene in *Rhodoferax ferrireducens* T118. A pair of genes essential for the EMP pathway, phosphoglycerate kinase (EC: 2.7.2.3; Pnap_3874) and pyruvate kinase (EC: 2.7.1.40; Pnap_3873), are adjacent to one another and share both sequence identity and gene order with *Polaromanas* sp. JS666. Furthermore, two of the genes required for both the EMP pathway and CO_2 fixation (fructose-1,6-bisphosphate aldolase and 3-phosphoglycerate kinase) occur in 2 copies, and one of each of these copies (Pnap_1988 and Pnap_1987, respectively) are likely function in an operon involved in CO_2 fixation (see below).

5.4.7.2 Pentose phosphate pathway

The pentose phosphate (PP) pathway provides the cell with crucial metabolic precursors (e.g., pentose phosphates for nucleic acids and erythrose for aromatic

amino acids), NADPH for biosynthesis, and several catalytic reactions shared with CO_2 fixation pathways. Strain CJ2 harbors two genes encoding glucose-6-phosphate-1-dehydrogenase (EC: 1.1.1.49), catalyzing the initiation of the PP pathway. One copy of this gene (Pnap_0653) is proximal to, but divergently transcribed from a putative PP pathway transaldolase (EC: 2.2.1.2; Pnap_0654) gene and the gene for glucose-6-phosphate isomerase (Pnap_0655), catalyzing the second step in the EMP pathway (see also above). The second gene copy (Pnap_3694) is adjacent to a glucokinase (Pnap_3695), which catalyzes the first step in the EMP pathway (see also above). There are 2 gene copies encoding ribulose-5-phosphate-3-epimerase (EC: 5.1.3.1) in strain CJ2: one copy is associated with autotrophic CO_2 fixation (see below); the other (Pnap_3660) occurs in a highly conserved gene cluster (with genes encoding an ApaG domain protein and a putative site-specific recombinase transmembrane protein) shared with *Polaromanas* JS666, *R. ferrireducens, Delftia acidovorans*, and two *Acidovorax* species. Overall, the *P. naphthalenivorans* genome annotation clearly includes seven of the eight genes encoding the PP pathway: it is likely that the eighth gene, encoding 6-phosphogluconolactonase, is present in the genome but absent from the annotation, as has been commonly reported for other genome sequences (Cordwell, 1999).

5.4.7.3 Entner-Doudoroff pathway

Genes are present encoding all 9 steps in the Entner-Doudoroff (ED) pathway, used in aerobic prokaryotes to convert glucose to pyruvate. The two genes unique to the ED pathway, 6-phosphogluconate dehydratase (Pnap_2986) and 2-keto-3-deoxy-phosphogluconate aldolase (Pnap_2985) are adjacent to one an-

other; this arrangement and the sequences themselves show close identity to those of *Polaromonas* sp. JS666, *R. ferrireducens, D. acidovorans*, and two *Acidovorax* species. The JGI annotation revealed 3 glycoside hydrolase genes likely to assist in carbohydrate degradation; of these, one is phage related (Pnap 2239).

5.4.7.4 Citric acid cycle

Strain CJ2 encodes enzymes catalyzing all ten steps of the complete citric acid cycle (CAC). In an operon arrangement resembling that of *Polaromonas* sp. JS666, the four succinate dehydrogenase subunits (Pnap_3032-3035) and a citrate synthase (Pnap_3037) are adjacent to a GntR-type transcriptional regulator gene (Pnap_3031). Aconitase (Pnap_3027) and malate dehydrogenase (Pnap_3030) are divergently transcribed from *gntR*. Key genes in strain CJ2 for funneling pyruvate into the citric acid cycle (CAC) include pyruvate dehydrogenase complex dihydrolipoamide acetyl transferase (EC: 2.3.1.61; Pnap_1782), dihydrolipoamide dehydrogenase (EC: 1.8.1.4; Pnap_1784), and oxoacid E1 subunit (EC: 1.2.4.1; Pnap_1781). This highly conserved three-gene cluster is shared with *Polaromonas* sp. JS666 and *R. ferrireducens* T118. The two genes (encoding isocitrate lyase and malate synthase; Pnap_2855 and Pnap_3683) required for the glyoxylate bypass pathway, within the citric acid cycle, are present in strain CJ2's genome.

5.4.7.5 CO_2 fixation

Sizova and Panikov (2007) experimentally demonstrated autotrophic fixation of CO_2 by strain CJ2. Two adjacent, uniquely arranged clusters of genes are likely responsible for the Calvin-Benson-Bassham (CBB) CO_2 fixation cycle. The genes

in the first cluster (encoding RuBisCo activation proteins *cbbO*, *cbbQ*, the *cbbM*-type Form II ribulose bisphosphate carboxylase large subunit, and a regulatory gene; Pnap_1976-1979) match in sequence and gene order to a corresponding cluster in *Thiobacillus denitrificans* (Badger and Bek, 2008; Beller *et al.*, 2006). *cbbQ* and *cbbO* encode enzymes involved in posttranslational activation. The *cbbMQO* genes, in *T. denitrificans*, are encoded directly downstream from and divergently cotranscribed from a *cbbR* regulatory gene (Beller *et al.*, 2006; Hayashi *et al.*, 1997). In strain CJ2, the *cbbMQO* genes are adjacent to a *cbbR*-like LysR regulatory gene (Pnap_1979) whose sequence is most similar to a gene in *R. ferrireducens* (Rfer_1394). The second cluster of CBB metabolic genes (Pnap_1980-1988) includes ribulose-5-phosphate-3-epimerase (EC: 5.1.3.1), fructose-1,6-bisphosphatase (EC: 3.1.3.11), phosphoribulose kinase (EC: 2.7.1.19), transketolase (EC: 2.2.1.1), 3-phosphoglycerate kinase/isomerase (EC: 2.7.2.3), and fructose-bisphosphate aldolase (EC: 4.1.2.13) – these show similar gene order to *Ralstonia eutropha* H16 Chromosome 2 and to *R. ferrireducens*; both are known autotrophs. Several of the CBB genes are found in 2 or 3 copies (e.g., 3-phosphoglycerate kinase (Pnap_3874), transketolase (Pnap_4056), sedoheptulose-1,7-bisphosphate aldolase (Pnap_3869), ribulose-5-phosphate-3-epimerase(Pnap_3660); remarkably, one copy of each of these replicated, non-operon associated CBB genes match (in decreasing order) corresponding genes in *Polaromonas JS666, R. ferrireducens, Acidovorax, Delftia*, and *Methylibium petroleiphilum* PM1. Identification of genes encoding 11 of the 12 steps in the CBB cycle was successful; however, locating sedoheptulose-1,7-bisphosphatase was a challenge until its overlapping functionality with fructose-1,6-bis-phosphatase (Pnap_1067) was recognized. Consistent with autotrophy in strain CJ2, two genes for carbonic anhydrase were also found (Pnap_0456, Pnap_3756). In *T. denitrificans*, an operon

near the CBB gene cluster encodes carboxysome shell proteins; no similar genes were found in strain CJ2. Integrase genes are located directly adjacent to the CBB genes in strain CJ2 and suggest a role for lateral transfer in acquisition of CBB cycle genes.

Form II Rubisco enzymes are adapted to function in low-O_2 and high CO_2 environments (Badger and Bek, 2008). Interestingly, it has been hypothesized that facultative autotrophs (capable of switching between CBB-cycle CO_2 fixation and heterotrophy), closely regulate expression of the CBB cycle, and characteristically harbor the Form II RuBisCo large subunit gene in a single gene cluster with *cbb* genes encoding metabolic enzymes of the CBB cycle (Badger and Bek, 2008). The Form II RuBisCo and CBB gene clustering found in strain CJ2 are consistent with strain CJ2's lifestyle as a heterotroph dwelling in freshwater sediment, where tight regulation in response to organic and inorganic substrate availability may confer an advantage in the face of fluctuating growth conditions.

5.4.7.6 H_2 oxidation

H_2-based lithothrophy in strain CJ2 has been demonstrated by Sizova and Panikov (2007). Genes enabling H_2 utilization reside in a large operon encoding NiFe-hydrogenase metabolism that is located adjacent to the CBB gene cluster (Pnap_1953-1975). The structure of the operon and sequence comparisons show that strain CJ2's hydrogenase genes are most closely related to those in *Methylibium petroleiphilum* PM1 and in the megaplasmid pHG1 of *Cupriavidus necator*. Notably, no homologs for this hydrogenase cluster are found in the genome of strain CJ2's close taxonomic relative, *Polaromonas* sp. JS666. Faculta-

tive chemoautolithotrophs are characterized by versatile metabolism (Friedrich and Schwartz, 1993). H_2 may be utilized under starvation conditions (Friedrich and Schwartz, 1993). Alternatively, H_2 oxidation may be used to recycle H_2 produced by nitrogenase enzyme activity to increase the efficiency of N_2 fixation (Friedrich and Schwartz, 1993). For example, in N_2-fixing *Alcaligenes eutrophus*, hydrogenases were expressed during electron donor limitation, while RuBisCo was expressed under carbon limitation with energy excess (Friedrich, 1982).

5.4.8 Energy metabolism

The following sections summarize genes for aerobic respiration and denitrification in strain CJ2.

5.4.8.1 Aerobic respiration

Strain CJ2 was isolated under aerobic conditions on minimal media saturated with naphthalene vapor at 10 °C. Genes required for formation of complexes I, II, and III of oxidative phosphorylation are present: NADH:quinone oxidoreductase (EC 1.6.5.3; Pnap_1424-1437), succinate dehydrogenase (EC: 1.3.99.1; Pnap_3032-3035) and the cytochrome bc_1 complex, or quinol:cytochrome *c* oxidoreductase (EC:1.10.2.2; Pnap_0713-0715). In addition to using the proton-translocating NADH dehydrogenase of complex I in energy production, strain CJ2 may employ a Type II NADH dehydrogenase (EC: 1.6.99.3; Pnap_1280) to optimize the (NADH)/(NAD^+) balance under changing environmental conditions (Melo *et al.*, 2004; Pohlmann *et al.*, 2006). Strain CJ2 encodes several terminal oxidase enzymes, including a cytochrome aa_3-type cytochrome *c* oxi-

dase (EC:1.9.3.1; Pnap_0878-0882), a cytochrome cbb_3-type oxidase (EC: 1.9.3.1; Pnap_3779-3783) and a cytochrome *bd*-type quinol oxidase (Pnap_0494 and Pnap_0495). It is likely that under high oxygen tensions the aa_3-type oxidase is used, and that the cbb_3-type oxidase is important under microaerophilic conditions (Pitcher and Watmough, 2004). The quinol oxidase *bd* may also be a high-affinity terminal oxidase for growth at low O_2 tension. Strikingly, the cbb_3-type oxidase genes are of low G+C content; the closest sequenced relative of strain CJ2, *Polaromonas* strain JS666 also has genes encoding the cbb_3-type oxidase, but the gene order in strain CJ2 is more similar to those of *Rhodoferax*.

5.4.8.2 Denitrification

Many key genes for respiratory denitrification are found in strain CJ2 (Zumft, 1997). The structural enzymes are encoded in a conserved gene cluster including putative nitrite/nitrate transport genes (Pnap_3756-3765). A dissimilatory nitrate reductase (EC: 1.7.99.4; Pnap_3761), a copper-containing NO-forming nitrite reductase (EC 1.7.1.4; Pnap_1326) and a putative nitric oxide reductase (Pnap_1328) are present. No complete orthologs for nitrous oxide reductase were found, but Pnap_2785 is similar to the N-terminal domain (EC: 1.7.99.6). The nitrate reductase gene is conserved in strain JS666 (90% nucleotide identity, Bpro_4595) and evolutionarily related to Nar-type dissimilatory nitrate reductases (*Acidovorax avenae*, Aave_0661; and *Rhodoferax ferrireducens*, Rfer_2792). The nitrite reductase is homologous to the copper-containing nitrite reductase (*nirK*) in the ammonia oxidizing bacterium *Nitrosospira multiformis* (no homolog in strain JS666; 63% nucleotide identity, NmulA_1998). According to Cantera and Stein (2007), the *nirK* promoter region contains conserved recognition mo-

tifs for NsrR and DNR (dissimilative nitrate respiration regulator, see Cantera and Stein, 2007; Ellis *et al.*, 2007; Norton *et al.*, 2008). The nitric oxide reductase (Pnap_1328) is similar to *norZ* or *norB2*.

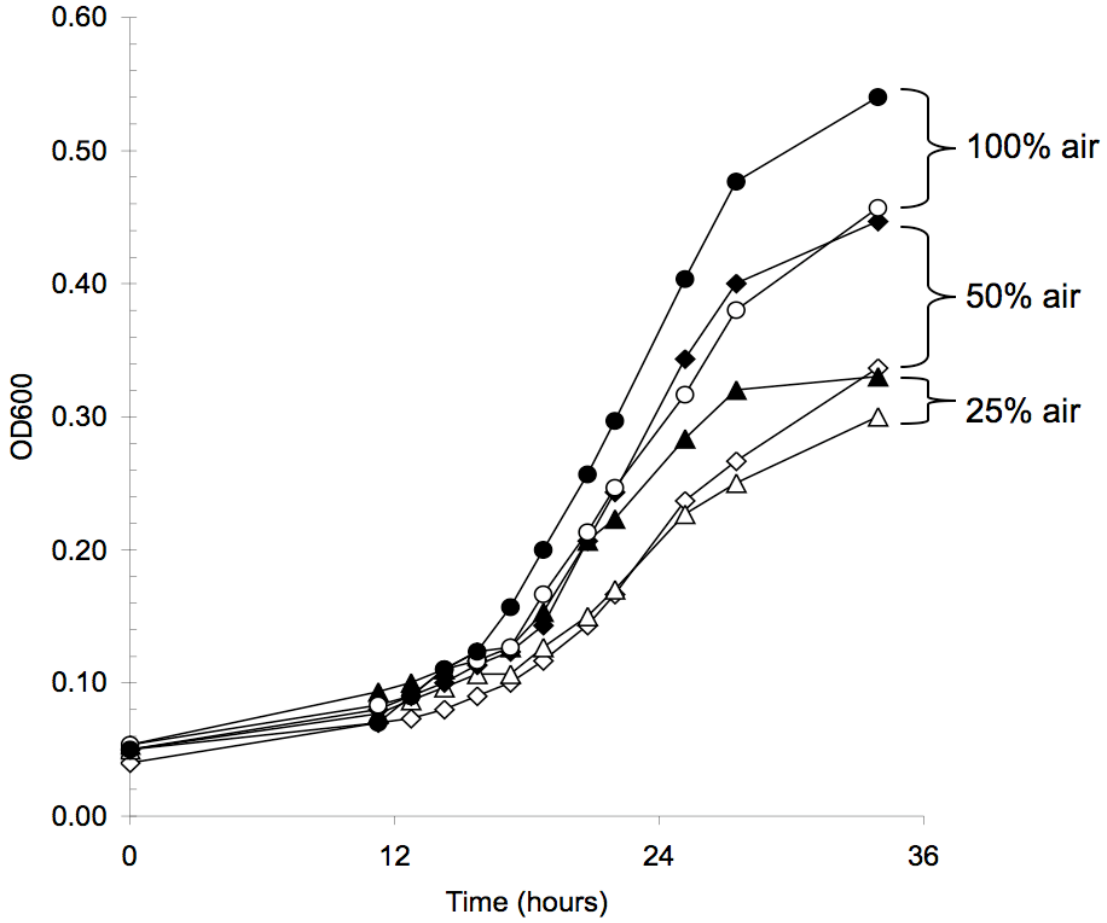


Figure 5.8: Influence of added nitrate (10 mM) on growth of P. naphthalenivorans strain CJ2 under aerobic and microaerobic conditions. To test aerobic and microaerobic growth with (open symbols) or without (closed symbols) added nitrate, yeast extract (0.5%) and pyruvate (0.2%) supplemented minimal basal salts media was prepared anaerobically in crimp-seal Balch tubes with 75% headspace volume. To establish aerobic and microaerobic growth conditions, 100% (circles; aerobic), 50% (diamonds; microaerobic) or 25% (triangles; microaerobic) of the headspace volume was replaced with sterile air. No growth was observed in the absence of both oxygen and nitrate (data not shown). Data points represent the average OD600 for triplicate cultures. Error bars have been omitted for clarity.

Despite the presence of denitrification genes, we were unable to grow strain CJ2 anaerobically in the presence of nitrate. Nitrate respiration has been shown to supplement aerobic growth under microaerobic conditions (Lengeler, 1999), but strain CJ2 showed no nitrate-dependent increase in growth efficiency under oxygen-limiting conditions (Fig. 5.8). In fact, added nitrate (10 mM KNO_3) inhibited growth of strain CJ2 in yeast extract and pyruvate supplemented minimal media under both aerobic and microaerobic conditions (Fig. 5.8).

5.4.9 Transport systems

The automated JGI annotation lists at least 351 genes involved in transport. This is about 7% of the genome, compared to 5-6% for the average bacterial genome (Larimer *et al.*, 2004; Paulsen *et al.*, 2000; Ren and Paulsen, 2005). Thorough manual inspection revealed 551 transport-related genes (11.2% of protein coding sequences), which is lower than *Ralstonia eutropha* (12.3%) but higher than the average found in Eubacteria (9.2%) (Pohlmann *et al.*, 2006; Ren and Paulsen, 2005). Transporters in strain CJ2 were classified according to the TC numbering system (Paulsen *et al.*, 2000). Primary transport systems (powered directly by ATP hydrolysis) include at least 180 genes associated with ATP-binding cassette (ABC) systems, 12 P-type ATPases, 8 type II secretion genes, and 13 Type IV secretion genes. Secondary transport systems include 28 major facilitator superfamily (MFS) members, 25 resistance-nodulation-cell division (RND) pumps, and 22 tripartite ATP-independent transporters. Putative uptake systems for iron, phosphorus, sulfur and nitrogen compounds are described below. The RND family exporters play a role in efflux of toxic organic compounds.

5.4.10 Assimilation of iron, phosphorus, sulfur and nitrogen

The following sections summarize genes for iron, phosphorus and sulfur assimilation in strain CJ2.

5.4.10.1 Iron

A key component of the iron-sulfur cluster dioxygenase enzymes involved in metabolism of aromatic compounds, iron is particularly important to the fitness of strain CJ2. The four TonB-dependent iron siderophore receptors in strain CJ2 are found adjacent to a gene encoding bacterioferritin (Pnap_2558), adjacent to a putative ABC-type transport system for uptake of hydroxamate-type Fe^{3+}-siderophores (Pnap_2368), or unassociated with known iron acquisition genes (Pnap_3005, 3599). Strain CJ2 also encodes a a Dps-type ferritin-like protein (Pnap_0325) and a putative high affinity iron permease (Pnap_3854). The iron-regulated genes in strain CJ2 are probably controlled by a Fur (ferric uptake regulation) protein (Pnap_3791).

5.4.10.2 Phosphorus

Strain CJ2 may acquire inorganic phosphate by at least two mechanisms: (i) using a low-affinity Pit-type (P_i) transporter (Pnap_1152); and (ii) via a Pst-type high affinity ABC-transporter system (Pnap_2202-2206) regulated by a *phoBR* response regulator (Pnap_2200-2201) (Monds *et al.*, 2006). Alkaline phosphatase (Pnap_1209) may be used to hydrolyze organic phosphate esters under (P_i)-limiting conditions (see Carbon and energy storage below).

5.4.10.3 Sulfur

Organic sulfur in soils is comprised mostly of sulfonates and sulfate esters; hence, many soil bacteria carry genes for utilization of alkanesulfonates and alkyl or aromatic sulfate esters (Kertesz, 2000). Members of the *Comamonadaceae*, including polaromonads, have been reported to play a key role in mineralization of carbon bound sulfur in rhizosphere soils (Schmalenberger *et al.*, 2008). Genes for alkanesulfonate utilization were found in strain CJ2 (Pnap_2415, homology to *ssuF*; Pnap_2418, homology to *ssuD*; Pnap_2421, homology to *ssuE*; Pnap_0560). A putative sulfite reductase is present, as well as a *yedY*-like sulfite oxidase gene (Pnap_2427 and Pnap_0667, respectively).

5.4.10.4 Nitrogen

Three predicted transporters may mediate uptake of ammonia from the environment (Pnap_4104, Pnap_0178, Pnap_0121). Strain CJ2 lacks glutamate dehydrogenase, and probably incorporates ammonia via the ATP-dependent activity of glutamine synthetase and glutamate synthase (Pnap_0683 and Pnap_0684). Strain CJ2 may reduce nitrate to ammonia using an assimilatory nitrate reductase (EC: 1.7.1.1; Pnap_1404), found in a cluster with *nirB* and *nirD* genes encoding an assimilatory nitrite reductase (EC: 1.7.1.4; Pnap_1405, 1407). An adjacent gene cluster encodes a nitrate-specific ABC transport system (Pnap_1411-1414). A gene homologous to *narK* (Pnap_3762) may be responsible for nitrate uptake and nitrite excretion (Hartig *et al.*, 1999). Structural and accessory genes for urease (EC: 3.5.1.5; Pnap_0971-0977) are present, suggesting that strain CJ2 assimilates urea via transformation to ammonia. Strain CJ2 also encodes an ABC-

transport gene cluster with high homology to urea transporters (Pnap_0997-1000).

Strain CJ2 encodes a molybdenum dependent nitrogenase (EC:1.18.6.1) and associated enzymes for reduction of nitrogen gas to ammonia (Pnap_2343-2351; Pnap_1574-1587). Expression of nitrogenase activity is probably regulated by *nifA* (Pnap_1570). The capacity to fix nitrogen may be advantageous to strain CJ2 for growth in carbon-rich habitats such as coal-tar waste contaminated sediment.

5.4.11 Carbon and energy storage

Intracellular polyhydroxyalkanoate, polyphosphate and polyglucose granules have been reported in cells of *P. naphthalenivorans* (Jeon *et al.*, 2004). Consistent with these observations, several genes associated with biosynthesis and degradation of carbon and energy storage inclusions were found. Strain CJ2 may synthesize glycogen (glycogen synthase, Pnap_1107) and utilize stored glycogen through the activities of glycogen debranching enzyme (Pnap_1103) and glycogen phosphorylase (Pnap_1105). Furthermore, the genome encodes the capacity to synthesize and utilize polyhydroxybutyrate (e.g., Pnap_3566, poly-betahydroxybutyrate polymerase; Pnap_3568, phosphate butyryltransferase; Pnap_0127, polyhydroxybutyrate depolymerase; Madison and Huisman, 1999). Finally, strain CJ2 encodes a specialized kinase for synthesis of polyphosphate (Pnap_2212). A putative exopolyphosphatase (Pnap_2209) and a phosphate uptake regulator and ABC transport system (Pnap_2202 thru Pnap_2206, described above) complete the system for storage of energy and phosphorus (van Veen, 1997; Kornberg *et al.*, 1999).

5.4.12 Oxidative stress tolerance

Strain CJ2 was isolated from surface freshwater sediment, a habitat likely to fluctuate between aerobic, anaerobic and microaerobic conditions (Jeon *et al.*, 2004). In experiments designed to test the preference of strain CJ2 for microaerobic conditions, the heaviest growth of strain CJ2 in soft agar (0.2%) was observed about 2 mm below the agar surface, suggestive of a microaerophilic lifestyle. Several genes encoding tolerance to reactive oxygen species such as superoxide and hydroperoxides were found. CJ2 has three predicted catalase genes (Pnap_2662, Pnap_0993, Pnap_1830), one superoxide dismutase gene (Pnap_1927) and two alkyl hydroperoxide reductases (Pnap_2234, Pnap_2401). Additionally we found 7 glutaredoxin genes and 6 thioredoxin genes that may play a role in thiol redox control in strain CJ2 (Holmgren, 1989). A putative nitric oxide dioxygenase (Pnap_1304) is divergently transcribed from a *nifA* transcriptional regulator (Pnap_1305).

5.4.13 Catabolism of aromatic compounds and capacity for biodegradation

Strain CJ2 harbors genes for at least four central pathways and numerous peripheral pathways for aromatic compound metabolism. Growth substrates predicted from genome analysis include benzoate, biphenyl, *p*-cresol, salicylate, 3- and 4-hydroxybenzoate, naphthalene, phthalate, vanillate and ferulate. Naphthalene catabolism proceeds via the gentisate biochemical pathway (Pumphrey and Madsen, 2007) and is catalyzed by *nag* genes (Jeon *et al.*, 2006; Park *et al.*, 2007a) related to those of *Ralstonia* sp. strain U2 (Fuenmayor *et al.*, 1998; Zhou *et al.*, 2001). Overall, the most likely central pathway intermedi-

ates of aromatic compound metabolism in strain CJ2 include benzoyl-CoA, gentisate (Pnap_2452-2454; Pnap_2474; Pnap_2476; Pnap_3145-3147), catechol (via the β-ketoadipate pathway; [Pnap_0864-0869; Pnap_2116]) and protocatechuate (through a meta-cleavage pathway; [Pnap_2027-2028; Pnap_2733-2734]).

5.4.13.1 Benzoate

Based on genome content, benzoate degradation in CJ2 is predicted to occur via hydroxylation (catechol–ketoadipate pathway) or via coenzyme A (CoA) ligation to benzoyl-CoA (cf. Gescher etal., 2006). A carbon source specific regulatory network for selection between these alternative degradation pathways was reported for *Burkholderia xenovorans* LB400, and the pathways in strain CJ2 may be under similar regulatory control (Denef *et al.*, 2004, 2005, 2006). Benzoate 1,2-dioxygenase is encoded (Pnap_2111-2112) near genes for catechol degradation via catechol 1,2-dioxygenase. CoA activation of benzoate may be catalyzed by a putative benzoate-CoA ligase (Pnap_2948), located in a gene cluster with benzoyl-CoA pathway homologs similar to those encoded on chromosome 1 in *B. xenovorans* LB400 (Pnap_2940-2954).

5.4.13.2 Biphenyl

A gene cluster putatively involved in biphenyl degradation (Pnap_4140-4153) is found on plasmid pPNAP01, adjacent to a phage integrase gene (Pnap_4137). The genes show high sequence identity to those of *Burkholderia xenovorans* LB400 (Denef *et al.*, 2004, 2005; Chain *et al.*, 2006), suggesting that biphenyl degradation proceeds via conversion to benzoate, and subsequent activation via benzoate-

CoA ligase. *Burkholderia xenovorans* LB400 also encodes an adjacent phage integrase (Bxe_C1182), with sequence similarity to a homolog in strain CJ2 suggesting recent lateral transfer of the gene cluster.

5.4.13.3 Other aromatic compounds

Putative genes for *p*-cresol degradation (Pnap_3834-3841) are found on the chromosome in strain CJ2. The closest homologs, with similar gene order, were found in *Azoarcus* EBN1. This gene cluster may be regulated by an AraC-type transcriptional regulator (Pnap_3834). Genes for 2-hydroxybenzoate (salicylate), 3-hydroxybenzoate and 4-hydroxybenzoate are also encoded on the chromosome. Park *et al.* (2007a) demonstrated that strain CJ2 utilizes 3-hydroxybenzoate via conversion to gentisate (Pnap_3143-3147). Metabolism of 4-hydroxybenzoate may occur via conversion to protocatechuate (Pnap_1901-1902) by 4-hydroxybenzoate 3-monooxygenase (Pnap_1901; similar to *Rhodoferax ferrireducens*; 81% amino acid identity). In addition, a putative 4-hydroxybenzoyl-CoA thioesterase (Pnap_0799; more closely related to *Rhodoferax ferrireducens* (79% amino acid identity) than to JS666 (74% amino acid identity) is included in a potential ferulate degradation gene cluster (Pnap_0796-0807; see below).

Gram-negative bacteria degrade phthalate via 4,5-dihydroxyphthalate and protocatechuate (Chang and Zylstra, 1998). A gene cluster putatively associated with phthalate degradation is encoded on plasmid pPNAP04. Genes for biodegradation of phthalate included a putative phthalate 4,5-dioxygenase (Pnap_4624, 4627), 4,5-dihydroxyphthalate dehydrogenase (Pnap_4623) and 4,5-dihydroxyphthalate decarboxylase (Pnap_4626. The gene order is most similar to that of *Comamonas testosteroni* KF-1 (unfinished sequence). Vanillate degra-

dation, catalyzed by gene products encoded by *vanAB* homologs (Pnap_2724; Pnap_2725), most likely proceeds via protocatechuate (Priefert *et al.*, 1997; Providenti *et al.*, 2006). The gene arrangement suggests that *vanAB* genes are co-transcribed with a putative ABC transporter (Pnap_2726-2730; cf. Hara *et al.*, 2000). A putative protocatechuate 4,5-dioxygenase, encoded by homologs of *ligAB* (Pnap_2027; Pnap_2028), is arranged in a putative operon including genes for protocatechuate degradation via activity of protocatechuate 4,5-dioxygenase (Pnap_2026-2034).

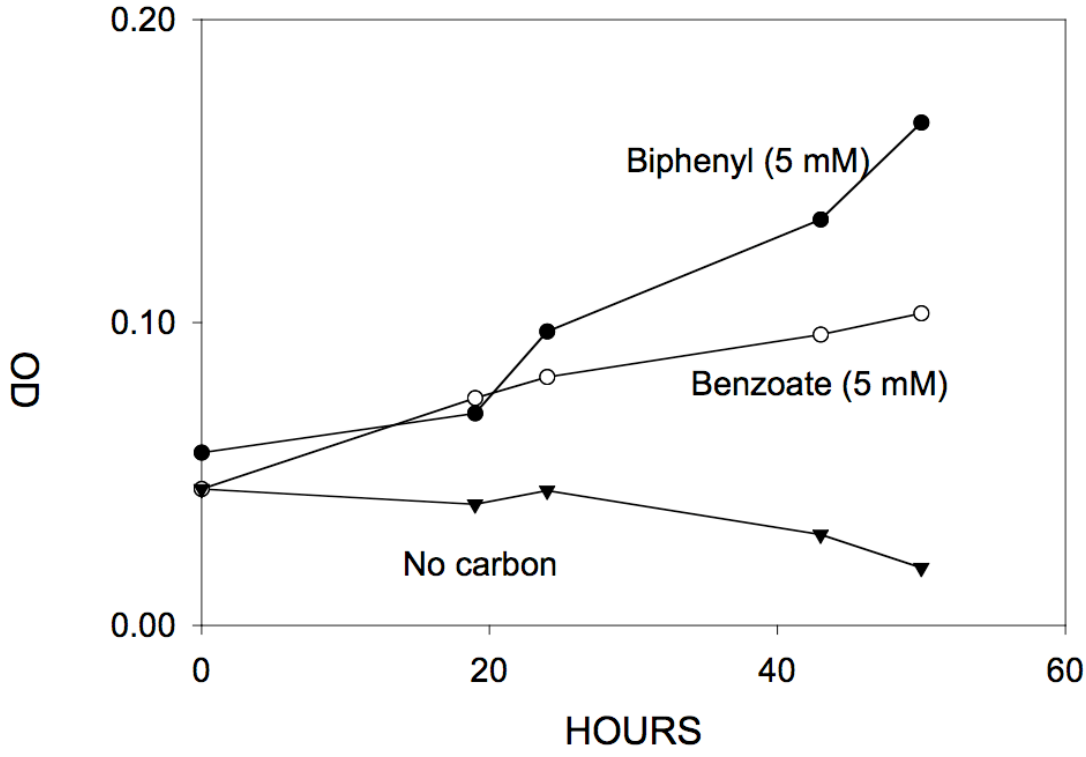


Figure 5.9: Growth of *P. naphthalenivorans* strain CJ2 in MSB broth that was unamended, amended with 5 mM benzoate, or amended with 0.1% (w/v) biphenyl. Data points represent the average OD_{600} for triplicate cultures.

The gene sequence and order for a putative ferulate degradation gene cluster is conserved in JS666 (Bpro_3112-3118), with the exception of an additional ABC transporter (Pnap_0800-0803) in the strain CJ2 gene cluster (Pnap_0796-0807).

Interestingly, *Rhodoferax ferrireducens* T118 carries a similar ABC transporter in the homologous gene cluster (Rfer_0263-0273). The gene cluster also includes a 4-hydroxybenzoyl-CoA thioesterase (Pnap_0799).

P. naphthalenivorans strain CJ2 was isolated for its capacity to utilize naphthalene as sole carbon and energy source (Jeon *et al.*, 2003, 2004). However, Pumphrey and Madsen (2007) showed that inhibitory metabolites accumulated in broth culture when strain CJ2 was grown on naphthalene at concentrations above 55 μM, well below the aqueous naphthalene saturation concentration of 230 μM. This phenotype is consistent with adaptation to oligotrophic conditions in the highly variable sediment environment. Similar metabolic imbalance and/or toxicity (Pumphrey and Madsen, 2007) may explain the weak, though significant, growth of strain CJ2 observed on minimal medium supplemented with either biphenyl and benzoate (Fig. 5.9). The growth of strain CJ2 on biphenyl (Fig. 5.9) was corroborated by formation of colonies on minimal salts medium supplememted with biphenyl vapor.

5.4.14 Conclusions

P. naphthalenivorans strain CJ2 metabolized the toxic pollutant, naphthalene, within a complex field microbial community native to freshwater sediment (Jeon *et al.*, 2003). Knowing the complete genome of strain CJ2 allows us to begin developing an integrated understanding of the physiology, genetics, ecology, and evolution of this facultatively chemolithoautotrophic bacterium. Emerging issues for future exploration include the role of autotrophy, nitrogen fixation, and what appear to be horizontally-acquired traits (e.g., catabolism of aromatic compounds) in determining strain CJ2's ecological fitness in its native habitat.

CHAPTER 6

SUMMARY AND CONCLUSIONS

In Chapter 2, geochemical monitoring data documenting contaminant removal at Site 24 were described. T-RFLP analysis revealed that the microbial communities from two locations within the contaminant plume (well 36, well 12) were dynamic in time but constrained by local biogeochemical conditions. The most abundant bacterial phylotypes in the 16S rRNA gene survey for well 36 grouped with *Bacteriodetes/Chlorobi*, δ-proteobacteria, and unclassified lineages. In contrast, the dominant phylotypes from well 12 were members of the δ-proteobacteria, β-proteobacteria, and unclassified lineages. Bacterial sequences were retrieved from Site 24 groundwater aligning with the D-BACT lineage, previously found on site (Bakermans and Madsen, 2002b), and shown by Raghoebarsing *et al*. (2006) to anaerobically oxidize methane using nitrite as an electron acceptor (see also Ettwig *et al.*, 2008). Sequences related to *Nitrosopumilus*, an ammonia-oxidizing species within the archaeal group I.1a, were also detected. A molecular 18S rRNA gene survey confirmed prior cultivation-based reports that eukaryotes were enriched in the contaminated wells, corroborating the presence of a microbial food chain based on bacterial metabolism of pollutants. The diverse array of eukaryote sequences retrieved included fungi, Stramenophiles, Alveolates, and the free-living turbellarian flatworm, *Suomina*.

Chapter 3 describes the diversity and spatial-temporal variation of archetypal (i.e., *nah* genes) as well as biochemically uncharacterized microbial oxygenase genes in contaminated groundwater at Site 24. Broad-specificity degenerate primers designed by Ní Chadhain *et al.* (2006) were used to generate 4 Rieske dioxygenase gene clone libraries representing well 36 and well 12. Rieske gene sequences were assigned to clone groups based on 95% sequence identity; genes associated with *Rhodococcus*, *Sphingomonas*, and β-proteobacterial *nag* genes were retrieved in the survey. Although *nah* genes were surpisingly not

represented in the Rieske gene clone libraries, they were detected in groundwater DNA extracts using a *nah*-gene specific competitive PCR assay. Competitive PCR was also used to show that *nah* and *nag* gene abundance fluctuated substantially over a 10-month period. *In situ* expression of both *nah* and *nag* genes was observed using RT-PCR, qualitatively corresponding with structural gene abundance. This study suggested the coexistence and *in situ* expression of *nah* and *nag* gene encoded biodegradation pathways at Site 24 even in the presence of low (micromolar) ambient concentrations of naphthalene.

Chapter 4 describes the fluctuations in geochemical characteristics in contaminated groundwater at Site 24, including concentrations of nitrate, ammonia, sulfate, sulfide and methane. Microcosms containing groundwater and sediment from Site 24 demonstrated that biodegradation of naphthalene occurred under aerobic, but not methanogenic, sulfate-reducing, iron-reducing, manganese-reducing or nitrate reducing conditions. In spite of the absence of anaerobic naphthalene biodegradation in laboratory incubations, metabolites of anaerobic naphthalene metabolism (1-naphthoic acid and 1,2,3,4-tetrahydro-1-naphthoic acid) and mRNA transcripts (*bssA*) associated with anaerobic aromatic compound degradation were detected in extracts from groundwater fixed on-site. The presence of these biomarkers was in agreement with field monitoring data showing disappearance of aromatic pollutants under anaerobic conditions (see Yagi *et al.*, 2009a). The potential for denitrification and aerobic methane oxidation was demonstrated by PCR-based detection of *narG* encoding nitrate reductase and *mxaF* genes. Efforts to amplify *mcrA*, encoding methyl-coenzyme-M-reductase and involved in methanogenesis, were unsuccessful. RT-PCR was used to show that cytochrome *c* nitrite reductase and ammonia monooxygenase transcripts were expressed within contaminated groundwater

microbial communities, supporting a complete N cycle involving dissimilatory nitrate reduction to ammonium (DNRA) in this subsurface habitat. Linking the data from biogeochemical and molecular biomarker analyses, accumulation of reduced metabolic end products (ammonia, sulfide, methane) from the activity heterotrophic microbial communities in this coal tar contaminated ecosystem supports lithotrophic and other microbial populations that would otherwise be absent from the site.

Chapter 5 presents the genome analysis of *Polaromonas naphthalenivorans* strain CJ2, isolated from Site 24 sediments. The complete genome sequence revealed 3 previously unknown plasmids for a total of 8 extrachromosomal elements. Nearly a third (31.8%) of the genes in strain CJ2 were not found in *Polaromonas* JS666, its closest sequenced relative. Six chromosomal regions were hypothesized to have originated via horizontal gene transfer, based on nucleotide characteristics (G+C content, GC skew, Karlin signature difference). The *nag*-like naphthalene catabolic operon is located within one of these regions, and was associated with genetic signatures of recent mobility. Genes for strain CJ2's previously documented capacity to fix CO_2 autotrophically (Sizova and Panikov, 2007) were found, as well as a likely genetic basis for H_2 oxidation. Many key genes for respiratory denitrification were present, but attempts to grow strain CJ2 anaerobically in the presence of nitrate were unsuccessful. Nitrogen fixation by strain CJ2 and growth on the aromatic compounds biphenyl and benzoate were experimentally verified.

APPENDIX A
PROTEIN CODING SEQUENCES IN THE PUTATIVE 96 KB DNA INSERTION IN *POLAROMONAS NAPHTHALENIVORANS* STRAIN CJ2

Locus Tag	COG	GC%	IMG Description	Description of best BLASTP hit against NR (July 2008)	% identity
Pnap_2439	J	0.58	MiaB-like tRNA modifying enzyme YliG	hypothetical protein Bpro_2701 [Polaromonas sp. JS666]	92.72
Pnap_2440		0.62	Recombinase	recombinase [Acidovorax avenae subsp. citrulli AAC00-1]	60.75
Pnap_2441		0.62	hypothetical protein		
Pnap_2442	O	0.56	AAA ATPase, central domain protein	ATPase central domain-containing protein [Burkholderia ambifaria AMMD]	88.14
Pnap_2443	O	0.58	hypothetical protein	hypothetical protein Bamb_1970 [Burkholderia ambifaria AMMD]	87.13
Pnap_2444		0.49	RNA polymerase, sigma-24 subunit, ECF subfamily	ECF subfamily RNA polymerase sigma-24 factor [Verminephrobacter eiseniae EF01-2]	52.82
Pnap_2445		0.53	hypothetical protein	hypothetical protein Veis_3741 [Verminephrobacter eiseniae EF01-2]	54.5
Pnap_2446	L	0.59	UvrD/REP helicase	UvrD/REP helicase [Verminephrobacter eiseniae EF01-2]	69.45
Pnap_2447		0.59	hypothetical protein	hypothetical protein Veis_3743 [Verminephrobacter eiseniae EF01-2]	77.67
Pnap_2448	H	0.61	UBA/THIF-type NAD/FAD binding protein	UBA/THIF-type NAD/FAD binding protein [Verminephrobacter eiseniae EF01-2]	77.56
Pnap_2449		0.54	hypothetical protein	hypothetical protein Veis_3745 [Verminephrobacter eiseniae EF01-2]	72.53
Pnap_2450		0.53	hypothetical protein	hypothetical protein Rfer_2538 [Rhodoferax ferrireducens T118]	71.01
Pnap_2451			interesting		
Pnap_2452	O	0.62	maleylacetoacetate isomerase	maleylacetoacetate isomerase [Leptothrix cholodnii SP-6]	73.11
Pnap_2453	Q	0.59	fumarylacetoacetate (FAA) hydrolase	fumarylacetoacetate (FAA) hydrolase [Delftia acidovorans SPH-1]	82.25
Pnap_2454	Q	0.61	Cupin 2, conserved barrel domain protein	gentisate 1,2-dioxygenase [Polaromonas naphthalenivorans CJ2]	73.22
Pnap_2455	S	0.57	hemerythrin HHE cation binding region	hemerythrin HHE cation binding region [Dechloromonas aromatica RCB]	35.98
Pnap_2456	E	0.57	ABC transporter related	ABC transporter related [Polaromonas sp. JS666]	70.04
Pnap_2457	E	0.59	ABC transporter related	ABC transporter related [Polaromonas sp. JS666]	72.18
Pnap_2458	E	0.59	inner-membrane translocator	inner-membrane translocator [Leptothrix cholodnii SP-6]	60.63
Pnap_2459	E	0.59	inner-membrane translocator	inner-membrane translocator [Leptothrix cholodnii SP-6]	77.12
Pnap_2460	E	0.58	extracellular ligand-binding receptor	extracellular ligand-binding receptor [Polaromonas sp. JS666]	70.29
Pnap_2461	P	0.56	Rieske (2Fe-2S) domain protein	Rieske (2Fe-2S) domain-containing protein [Polaromonas naphthalenivorans CJ2]	90.38
Pnap_2462	Q	0.52	aromatic-ring-hydroxylating dioxygenase, beta subunit	aromatic-ring-hydroxylating dioxygenase, beta subunit [Polaromonas naphthalenivorans CJ2]	96.33
Pnap_2463					
Pnap_2464	L	0.61	IstB domain protein ATP-binding protein	IstB-like ATP-binding protein [Polaromonas sp. JS666]	78.57
Pnap_2465	E	0.54	Proline dehydrogenase	L-proline dehydrogenase [Acidothermus cellulolyticus 11B]	35.11
Pnap_2466	K	0.52	transcriptional regulator, GntR family	GntR family transcriptional regulator [Nostoc punctiforme PCC 73102]	37.31
Pnap_2467	L	0.62	IstB domain protein ATP-binding protein	istB transposition helper protein [Azoarcus sp. EbN1]	66.82
Pnap_2468	L	0.62	Integrase, catalytic region	transposase, IS21 family [Cupriavidus taiwanensis]	59.44
Pnap_2469		0.64	hypothetical protein		
Pnap_2470		0.57	recombinase		
Pnap_2471	L	0.58	transposase IS3/IS911 family protein	transposase IS3/IS911 [Rhodoferax ferrireducens T118]	82.3
Pnap_2472	L	0.54	transposase IS116/IS110/IS902 family protein	transposase IS116/IS110/IS902 family protein [Verminephrobacter eiseniae EF01-2]	83.14
Pnap_2473		0.59	hypothetical protein	hypothetical protein Pnap_2535 [Polaromonas naphthalenivorans CJ2]	88.16
Pnap_2474	Q	0.61	5-carboxymethyl-2-hydroxymuconate delta-isomerase(EC:5.3.3.10)	putative maleylpyruvate isomerase [Pseudomonas alcaligenes]	57.8
Pnap_2475					
Pnap_2476	Q	0.68	Cupin 2, conserved barrel domain protein	gentisate 1,2-dioxygenase [Ralstonia sp. U2]	85.8

Locus Tag	COG	GC%	IMG Description	Description of best BLASTP hit against NR (July 2008)	% identity
Pnap_2477	O	0.56	Glutathione S-transferase, C-terminal domain	glutathione-S-transferase-like protein [Ralstonia sp. U2]	93.53
Pnap_2478	Q	0.53	DSBA oxidoreductase	2-hydroxychromene carboxylate isomerase [Ralstonia sp. U2]	92.89
Pnap_2479	M	0.56	dihydrodipicolinate synthetase	trans-o-hydroxybenzylidenepyruvate hydratase-aldolase [Ralstonia sp. U2]	96.4
Pnap_2480	M	0.56	OmpW family protein	putative aldolase [Polaromonas naphthalenivorans CJ2]	99.53
Pnap_2481		0.56	Glyoxalase/bleomycin resistance protein/dioxygenase(EC:1.13.11.39)	1,2-dihydroxynaphthalene dioxygenase [Ralstonia sp. U2]	94.04
Pnap_2482	C	0.60	aldehyde dehydrogenase(EC:1.2.1.65)	salicylaldehyde dehydrogenase [Polaromonas naphthalenivorans CJ2]	99.79
Pnap_2483	Q	0.56	short-chain dehydrogenase/reductase SDR(EC:1.3.1.60)	cis-naphthalene dihydrodiol dehydrogenase-like protein [Comamonas testosteroni]	93.05
Pnap_2484	Q	0.53	aromatic-ring-hydroxylating dioxygenase, beta subunit	naphthalene 1,2 dioxygenase small oxygenase component [Polaromonas naphthalenivorans CJ2]	97.94
Pnap_2485	P	0.55	ring hydroxylating dioxygenase, alpha subunit(EC:1.14.12.12)	polyaromatic hydrocarbon dioxygenase large subunit [Comamonas testosteroni]	95.75
Pnap_2486	R	0.57	Rieske (2Fe-2S) domain protein	Rieske (2Fe-2S) domain-containing protein [Polaromonas naphthalenivorans CJ2]	90.38
Pnap_2487	Q	0.52	aromatic-ring-hydroxylating dioxygenase, beta subunit	salicylate-5-hydroxylase small oxygenase component [Ralstonia sp. U2]	86.34
Pnap_2488	R	0.56	Rieske (2Fe-2S) domain protein	salicylate-5-hydroxylase large oxygenase component [Polaromonas naphthalenivorans CJ2]	98.1
Pnap_2489	C	0.58	Oxidoreductase FAD-binding domain protein	ferredoxin reductase [Ralstonia sp. U2]	75
Pnap_2490	K	0.54	transcriptional regulator, LysR family	LysR-like regular protein [Ralstonia sp. U2]	83.06
Pnap_2491		0.53	hypothetical protein		
Pnap_2492		0.53	hypothetical protein		
Pnap_2493		0.60	hypothetical protein		
Pnap_2494		0.60	Lytic transglycosylase, catalytic	conjugal transfer protein [Limnobacter sp. MED105]	46.1
Pnap_2495		0.52	hypothetical protein		
Pnap_2496		0.46	hypothetical protein		
Pnap_2497	U	0.52	type II secretion system protein E	probable conjugal transfer protein TrbB [Limnobacter sp. MED105]	57.41
Pnap_2498		0.56	hypothetical protein		
Pnap_2499		0.56	hypothetical protein	hypothetical protein Tbd_0930 [Thiobacillus denitrificans ATCC 25259]	41.09
Pnap_2500	R	0.63	hypothetical protein		
Pnap_2501		0.60	Conjugal transfer protein TrbC		
Pnap_2502	N	0.56	Conjugal transfer TrbD family protein		
Pnap_2503	U	0.58	CagE, TrbE, VirB component of type IV transporter system	hypothetical protein pHCG3_p125 [Oligotropha carboxidovorans OM5]	44.73
Pnap_2504	U	0.52	conjugal transfer protein TrbJ		
Pnap_2505	U	0.58	Type IV secretory pathway TrbL components-like protein		
Pnap_2506		0.57	hypothetical protein		
Pnap_2507		0.58	hypothetical protein	conserved hypothetical protein [bacterium Ellin514]	36.3
Pnap_2508		0.56	hypothetical protein		

Locus Tag	COG	GC%	IMG Description	Description of best BLASTP hit against NR (July 2008)	% identity
Pnap_2509	U	0.57	Conjugal transfer protein	conjugal transfer protein F [Limnobacter sp. MED105]	38.32
Pnap_2510	U	0.55	Conjugal transfer protein TrbG/VirB9/CagX	TrbG mating pair formation protein [Plasmid QKH54]	33.2
Pnap_2511	U	0.56	conjugation TrbI family protein		
Pnap_2512	O	0.57	cytochrome c biogenesis protein, transmembrane region	cytochrome c biogenesis protein, transmembrane protein [Rhodoferax ferrireducens T118]	90.16
Pnap_2513	O	0.53	Thioredoxin domain	hypothetical protein Rfer_1598 [Rhodoferax ferrireducens T118]	85.71
Pnap_2514		0.54	hypothetical protein		
Pnap_2515	K	0.56	transcriptional regulator, TetR family	TetR family transcriptional regulator [Polaromonas sp. JS666]	73.76
Pnap_2516		0.61	late embryogenesis abundant protein		
Pnap_2517		0.59	hypothetical protein	hypothetical protein Bpro_5553 [Polaromonas sp. JS666]	78.23
Pnap_2518		0.60	hypothetical protein	hypothetical protein Pnap_4175 [Polaromonas naphthalenivorans CJ2]	74.67
Pnap_2519	R	0.63	ABC-1 domain protein	hypothetical protein Pnap_4176 [Polaromonas naphthalenivorans CJ2]	89.18
Pnap_2520			matches pPNAP01 and JS666 plasmid 2		
Pnap_2521	L	0.66	transposase IS66	transposase IS66 [Polaromonas naphthalenivorans CJ2]	92.54
Pnap_2522	L	0.65	IS66 Orf2 family protein	IS66 Orf2 family protein [Polaromonas naphthalenivorans CJ2]	93.16
Pnap_2523		0.64	transposase IS3/IS911 family protein	transposase IS3/IS911 family protein [Polaromonas naphthalenivorans CJ2]	87.16
Pnap_2524	R	0.60	beta-lactamase domain protein	carbon-phosphorus lyase complex accessory protein [Methylibium petroleiphilum PM1]	68.4
Pnap_2525		0.59	protein of unknown function DUF6, transmembrane	hypothetical protein Mpe_B0433 [Methylibium petroleiphilum PM1]	62.36
Pnap_2526		0.52	hypothetical protein		
Pnap_2527		0.56	hypothetical protein		
Pnap_2528		0.55	hypothetical protein		
Pnap_2529	L	0.55	domain of unknown function DUF1738	TraC3 protein [uncultured bacterium]	33.96
Pnap_2530	U	0.56	TRAG family protein	PROBABLE TRAG CONJUGAL TRANSFER TRANSMEMBRANE PROTEIN [Mesorhizobium loti]	42.93
Pnap_2531	O	0.54	conserved hypothetical conjugal transfer protein TraF		
Pnap_2532		0.55	hypothetical protein		
Pnap_2533		0.56	Relaxase/mobilization nuclease family protein	hypothetical protein pRA2_25 [Pseudomonas alcaligenes]	
Pnap_2534		0.51	hypothetical protein		
Pnap_2535		0.56	hypothetical protein	hypothetical protein Pnap_2473 [Polaromonas naphthalenivorans CJ2]	88.16
Pnap_2536	S	0.52	polyhydroxyalkonate synthesis repressor, PhaR	polyhydroxyalkanoate synthesis repressor PhaR [Polaromonas sp. JS666]	79.58

APPENDIX B
GENES ASSOCIATED WITH MOBILE DNA IN *POLAROMONAS NAPHTHALENIVORANS* STRAIN CJ2

Locus Tag	GC%	IMG Description	Organisms with BLASTP hit against NR (July 2008	% identity	IS Family
Pnap_0078	0.62	putative transposase, IS891/IS1136/IS1341	Polaromonas sp. JS666	93.88	IS200/IS605
Pnap_0079	0.63	transposase, IS605 OrfB family	Polaromonas sp. JS666	95.53	IS200/IS605
Pnap_0141	0.68	putative transposase of insertion sequence ISRm10-1	Polaromonas naphthalenivorans CJ2	94.3	IS630
Pnap_0142	0.70	hypothetical protein	Polaromonas naphthalenivorans CJ2	90.68	IS630
Pnap_0165	0.62	putative transposase	Polaromonas naphthalenivorans CJ2	100	ISNCY
Pnap_0276	0.62	transposase, IS4 family protein	Polaromonas naphthalenivorans CJ2	84.95	IS5
Pnap_0279	0.57	transposase, IS605 OrfB family	Polaromonas naphthalenivorans CJ2	87.07	IS200/IS605
Pnap_0347	0.70	hypothetical protein	Polaromonas naphthalenivorans CJ2	90.68	IS630
Pnap_0348	0.68	putative transposase of insertion sequence ISRm10-1	Polaromonas naphthalenivorans CJ2	94.3	IS630
Pnap_0409	0.61	transposase, IS4			IS5
Pnap_0410	0.68	putative transposase of insertion sequence ISRm10-1	Polaromonas naphthalenivorans CJ2	94.3	IS630
Pnap_0411	0.70	hypothetical protein	Polaromonas naphthalenivorans CJ2	90.68	IS630
Pnap_0414	0.66	hypothetical protein			ND
Pnap_0415	0.64	Recombinase	Rhizobium leguminosarum bv. viciae 3841	70.72	ND
Pnap_0416	0.61	transposase IS66	Rhodoferax ferrireducens T118	70.21	IS66
Pnap_0417					IS4
Pnap_0418	0.59	ISXo7 transposase	Polaromonas naphthalenivorans CJ2	100	IS630
Pnap_0419	0.64	transposase, IS4 family	Polaromonas naphthalenivorans CJ2	93.08	IS4
Pnap_0420	0.60	transposase IS116/IS110/IS902 family protein	Polaromonas naphthalenivorans CJ2	99.72	IS110
Pnap_0421	0.62	hypothetical protein	Polaromonas naphthalenivorans CJ2	95.62	IS4
Pnap_0422	0.66	transposase IS66	Burkholderia cenocepacia MC0-3	69.03	IS66
Pnap_0423	0.67	IS66 Orf2 family protein	Rhodoferax ferrireducens T118	83.16	IS66
Pnap_0424	0.65	transposase IS3/IS911 family protein			IS66
Pnap_0425	0.61	transposase, IS4 family protein	Polaromonas naphthalenivorans CJ2	86.62	IS5
Pnap_0426	0.59	transposase IS116/IS110/IS902 family protein	Polaromonas naphthalenivorans CJ2	69.3	IS110
Pnap_0427	0.65	transposase IS116/IS110/IS902 family protein	Azotobacter vinelandii AvOP	58.81	IS110
Pnap_0428	0.60	hypothetical protein	Polaromonas naphthalenivorans CJ2	82.64	ISNCY
Pnap_0429	0.61	Integrase, catalytic region	Marinobacter sp. ELB17	76.56	IS3
Pnap_0430	0.58	transposase IS3/IS911 family protein	Marinobacter sp. ELB17	79.38	IS3
Pnap_0537	0.61	transposase, IS4 family protein			IS5
Pnap_0541	0.62	putative transposase	Polaromonas naphthalenivorans CJ2	100	ISNCY
Pnap_0593	0.60	transposase, IS4 family protein	Polaromonas naphthalenivorans CJ2	100	IS5
Pnap_0594	0.62	putative transposase	Polaromonas naphthalenivorans CJ2	99.78	ISNCY
Pnap_0613	0.57	transposase, IS605 OrfB family	Polaromonas naphthalenivorans CJ2	86.59	IS200/IS605
Pnap_0871	0.59	ISXo7 transposase	Polaromonas naphthalenivorans CJ2	100	IS630
Pnap_1058	0.59	transposase, IS605 OrfB family	Polaromonas naphthalenivorans CJ2	98.59	IS200/IS605
Pnap_1112					IS200/IS605
Pnap_1136	0.52	transposase, IS4 family protein	Polaromonas naphthalenivorans CJ2	80	ISNCY
Pnap_1138	0.65	transposase IS116/IS110/IS902 family protein	Polaromonas naphthalenivorans CJ2	97.34	IS110
Pnap_1239	0.70	hypothetical protein	Polaromonas naphthalenivorans CJ2	90.68	IS630

Locus Tag	GC%	IMG Description	Organisms with BLASTP hit against NR (July 2008	% identity	IS Family
Pnap_1240	0.68	putative transposase of insertion sequence ISRm10-1	Polaromonas naphthalenivorans CJ2	94.3	IS630
Pnap_1278	0.57	transposase, IS605 OrfB family	Polaromonas naphthalenivorans CJ2	85.19	IS200/IS605
Pnap_1279	0.53	transposase IS200-family protein	Dichelobacter nodosus VCS1703A	75.91	IS200/IS605
Pnap_1294	0.57	transposase, IS605 OrfB family	Polaromonas naphthalenivorans CJ2	87.07	IS200/IS605
Pnap_1321	0.60	transposase, IS605 OrfB family	Candidatus Desulforudis audaxviator MP104C	35.07	no hits
Pnap_1393	0.59	ISXo7 transposase	Polaromonas naphthalenivorans CJ2	100	IS630
Pnap_1394	0.64	transposase, IS4 family protein	Polaromonas naphthalenivorans CJ2	91.08	IS5
Pnap_1395	0.59	transposase IS116/IS110/IS902 family protein	Polaromonas naphthalenivorans CJ2	69.3	IS110
Pnap_1396	0.65	transposase IS116/IS110/IS902 family protein	Azotobacter vinelandii AvOP	58.81	IS110
Pnap_1397	0.44	hypothetical protein			ISNCY
Pnap_1637	0.65	transposase IS116/IS110/IS902 family protein	Polaromonas naphthalenivorans CJ2	97.34	IS110
Pnap_1738	0.60	hypothetical protein			IS4
Pnap_1739	0.60	ISXo7 transposase	Polaromonas naphthalenivorans CJ2	75.07	IS630
Pnap_1740	0.55	conserved hypothetical transposase			IS3
Pnap_1741					IS3
Pnap_1742					IS3
Pnap_1743	0.64	transposase	Azoarcus sp. EbN1	76.15	IS5
Pnap_1744	0.62	putative transposase	Polaromonas naphthalenivorans CJ2	100	ISNCY
Pnap_1745	0.59	transposase	Polaromonas naphthalenivorans CJ2	98.68	IS5
Pnap_1746	0.64	Insertion element protein	Nitrosomonas europaea ATCC 19718	49.27	ISNCY
Pnap_1747	0.65	transposase IS116/IS110/IS902 family protein	Azotobacter vinelandii AvOP	58.81	IS110
Pnap_1748	0.59	transposase IS116/IS110/IS902 family protein	Polaromonas naphthalenivorans CJ2	69.3	IS110
Pnap_1749	0.64	transposase, IS4 family protein	Polaromonas naphthalenivorans CJ2	91.08	IS5
Pnap_1750	0.58	Integrase, catalytic region	Polaromonas sp. JS666	78.49	IS3
Pnap_1751	0.60	transposase IS116/IS110/IS902 family protein	Polaromonas naphthalenivorans CJ2	99.72	IS110
Pnap_1752	0.64	hypothetical protein	Polaromonas naphthalenivorans CJ2	74.09	IS4
Pnap_1753	0.55	ISCc1, transposase OrfB	Polaromonas naphthalenivorans CJ2	83.76	IS5
Pnap_1754	0.59	Transposase and inactivated derivatives-like protein			IS5
Pnap_1805	0.65	ISSod10, transposase OrfB	Polaromonas naphthalenivorans CJ2	90.53	IS630
Pnap_1806	0.61	transposase and inactivated derivative	Polaromonas naphthalenivorans CJ2	88.79	IS630
Pnap_1824	0.66	putative transposase of insertion sequence ISRm10-1	Polaromonas naphthalenivorans CJ2	94.3	IS630
Pnap_1825	0.68	hypothetical protein	Polaromonas naphthalenivorans CJ2	86.13	IS630
Pnap_1833	0.59	transposase, IS605 OrfB family	Nitrosococcus oceani ATCC 19707	72.33	IS200/IS605
Pnap_1909	0.66	ISSod10, transposase OrfB	Polaromonas naphthalenivorans CJ2	100	IS630
Pnap_1910	0.62	transposase and inactivated derivative	Polaromonas sp. JS666	90.91	IS630
Pnap_1931	0.56	hypothetical protein			IS3
Pnap_2083	0.66	putative transposase of insertion sequence ISRm10-1	Polaromonas naphthalenivorans CJ2	94.3	IS630
Pnap_2084	0.68	hypothetical protein	Polaromonas naphthalenivorans CJ2	86.13	IS630
Pnap_2471	0.58	transposase IS3/IS911 family protein	Rhodoferax ferrireducens T118	82.3	IS66
Pnap_2472	0.54	transposase IS116/IS110/IS902 family protein	Verminephrobacter eiseniae EF01-2	83.14	IS110

Locus Tag	GC%	IMG Description	Organisms with BLASTP hit against NR (July 2008	% identity	IS Family
Pnap_2514	0.54	hypothetical protein			ISL3
Pnap_2521	0.66	transposase IS66	Polaromonas naphthalenivorans CJ2	92.54	IS66
Pnap_2522	0.65	IS66 Orf2 family protein	Polaromonas naphthalenivorans CJ2	93.16	IS66
Pnap_2523	0.64	transposase IS3/IS911 family protein	Polaromonas naphthalenivorans CJ2	87.16	IS66
Pnap_2566	0.61	hypothetical protein			IS200/IS605
Pnap_2574	0.62	transposase and inactivated derivative	Polaromonas sp. JS666	90.91	IS630
Pnap_2575	0.65	ISSod10, transposase OrfB	Polaromonas naphthalenivorans CJ2	100	IS630
Pnap_2764	0.58	transposase, IS605 OrfB family	Polaromonas naphthalenivorans CJ2	82.48	IS200/IS605
Pnap_2857	0.60	hypothetical protein	Polaromonas naphthalenivorans CJ2	93.59	IS200/IS605
Pnap_2858	0.64	Insertion element protein	Nitrosomonas europaea ATCC 19718	49.27	ISNCY
Pnap_2900	0.60	transposase	Polaromonas naphthalenivorans CJ2	91.08	IS630
Pnap_2901	0.63	transposase	Polaromonas naphthalenivorans CJ2	91.79	IS630
Pnap_2996	0.60	transposase	Polaromonas naphthalenivorans CJ2	91.08	IS630
Pnap_2997	0.63	transposase	Polaromonas naphthalenivorans CJ2	91.79	IS630
Pnap_3108	0.60	transposase IS116/IS110/IS902 family protein	Polaromonas naphthalenivorans CJ2	99.72	IS110
Pnap_3113	0.65	hypothetical protein	Polaromonas naphthalenivorans CJ2	94.54	IS4
Pnap_3114					IS4
Pnap_3151	0.58	transposase, IS605 OrfB family	Polaromonas naphthalenivorans CJ2	98.59	IS200/IS605
Pnap_3173	0.58	transposase, IS605 OrfB	Chlamydia suis	37.97	IS200/IS605
Pnap_3402	0.58	putative transposase, IS891/IS1136/IS1341 family	Polaromonas naphthalenivorans CJ2	90.94	IS200/IS605
Pnap_3403	0.54	transposase, IS4 family protein	Polaromonas naphthalenivorans CJ2	79.26	ISNCY
Pnap_3404	0.56	transposase, IS605 OrfB	Polaromonas naphthalenivorans CJ2	87.5	IS200/IS605
Pnap_3461	0.51	regulatory protein, MerR			ISNCY
Pnap_3462	0.58	hypothetical protein			ISNCY
Pnap_3556	0.58	transposase, IS4 family protein	Polaromonas sp. JS666	74.84	ISNCY
Pnap_3633	0.64	transposase	Azoarcus sp. EbN1	76.15	IS5
Pnap_3634	0.61	transposase	Polaromonas naphthalenivorans CJ2	98.68	IS5
Pnap_3635	0.70	hypothetical protein	Polaromonas naphthalenivorans CJ2	90.68	IS630
Pnap_3636	0.68	putative transposase of insertion sequence ISRm10-1	Polaromonas naphthalenivorans CJ2	94.3	IS630
Pnap_3724	0.59	ISXo7 transposase	Polaromonas naphthalenivorans CJ2	100	IS630
Pnap_3880	0.57	transposase, IS605 OrfB family	Polaromonas naphthalenivorans CJ2	86.59	IS200/IS605
Pnap_4008	0.63	transposase	Polaromonas naphthalenivorans CJ2	91.79	IS630
Pnap_4009	0.60	transposase	Polaromonas naphthalenivorans CJ2	91.08	IS630
Pnap_4034	0.68	putative transposase of insertion sequence ISRm10-1	Polaromonas naphthalenivorans CJ2	94.3	IS630
Pnap_4035	0.70	hypothetical protein	Polaromonas naphthalenivorans CJ2	90.68	IS630
Pnap_4095	0.57	transposase, IS605 OrfB family			IS200/IS605
Pnap_4096	0.58	transposase, IS4 family protein	Nitrosomonas europaea ATCC 19718	58.45	IS982
Pnap_4123	0.63	Resolvase, N-terminal domain	Burkholderiales bacterium MUL2G11	92.09	Tn3
Pnap_4124	0.61	transposase Tn3	Aeromonas punctata	66.29	Tn3
Pnap_4125	0.65	transposase, IS4 family protein	Polaromonas naphthalenivorans CJ2	84.95	IS5

Locus Tag	GC%	IMG Description	Organisms with BLASTP hit against NR (July 2008	% identity	IS Family
Pnap_4246	0.58	hypothetical protein	Polaromonas naphthalenivorans CJ2	86.63	IS5
Pnap_4249	0.65	transposase, IS4 family protein	Polaromonas naphthalenivorans CJ2	86.09	IS5
Pnap_4250	0.65	putative transposase	Polaromonas naphthalenivorans CJ2	92.42	IS5
Pnap_4274					IS3
Pnap_4289	0.58	IstB domain protein ATP-binding protein	Polaromonas sp. JS666	87.45	IS21
Pnap_4290	0.60	Integrase, catalytic region	Polaromonas sp. JS666	80.08	IS21
Pnap_4362	0.56	transposase IS116/IS110/IS902 family protein	Polaromonas naphthalenivorans CJ2	83.82	IS110
Pnap_4363	0.61	transposase IS3/IS911 family protein	Verminephrobacter eiseniae EF01-2	87.23	IS3
Pnap_4396	0.60	Integrase, catalytic region	Polaromonas sp. JS666	80.08	IS21
Pnap_4397	0.58	IstB domain protein ATP-binding protein	Polaromonas sp. JS666	87.45	IS21
Pnap_4408	0.52	transposase, IS4 family protein	Polaromonas naphthalenivorans CJ2	80	ISNCY
Pnap_4434	0.62	transposase, IS4 family protein	Polaromonas naphthalenivorans CJ2	84.95	IS5
Pnap_4447	0.63	transposase IS3/IS911 family protein	Polaromonas naphthalenivorans CJ2	45.3	IS66
Pnap_4448	0.63	IS66 Orf2 family protein	Polaromonas naphthalenivorans CJ2	90.27	IS66
Pnap_4449	0.62	transposase IS66	Polaromonas naphthalenivorans CJ2	88.03	IS66
Pnap_4451	0.53	hypothetical protein			IS66
Pnap_4473	0.66	phage integrase family protein	Polaromonas naphthalenivorans CJ2	78.3	IS91
Pnap_4480	0.51	putative transposase			IS630
Pnap_4481	0.68	hypothetical protein	Polaromonas naphthalenivorans CJ2	92.29	IS4
Pnap_4482	0.59	transposase family protein	Leptospirillum sp. Group II UBA	64.04	IS630
Pnap_4484	0.62	transposase, IS4 family protein	Burkholderia vietnamiensis G4	85.59	IS5
Pnap_4485	0.57	transposase IS116/IS110/IS902 family protein	Polaromonas naphthalenivorans CJ2	94.38	IS110
Pnap_4505	0.61	transposase	Polaromonas naphthalenivorans CJ2	90.45	IS630
Pnap_4506	0.61	transposase	Polaromonas naphthalenivorans CJ2	91.08	IS630
Pnap_4531	0.60	hypothetical protein			IS66
Pnap_4532	0.67	transposase IS3/IS911 family protein	Polaromonas naphthalenivorans CJ2	82.43	IS66
Pnap_4533	0.66	IS66 Orf2 family protein	Polaromonas naphthalenivorans CJ2	93.16	IS66
Pnap_4534	0.66	transposase IS66	Polaromonas naphthalenivorans CJ2	92.35	IS66
Pnap_4596	0.60	transposase	uncultured	62.81	IS110
Pnap_4597	0.62	transposase IS3/IS911 family protein			IS3
Pnap_4601	0.45	hypothetical protein			IS630
Pnap_4608	0.65	transposase IS116/IS110/IS902 family protein	Aeromonas salmonicida subsp. salmonicida A449	75.15	IS110
Pnap_4609					IS4
Pnap_4621	0.59	hypothetical protein			no hits
Pnap_4630	0.55	transposase, IS4 family protein	Acinetobacter baumannii ATCC 17978	50.78	ISNCY
Pnap_4636			Delftia acidovorans SPH-1	63.78	ND
Pnap_4638	0.62	transposase IS66	Polaromonas naphthalenivorans CJ2	100	IS66
Pnap_4639					ND
Pnap_4640	0.64	Recombinase	Rhizobium leguminosarum bv. viciae 3841	70.72	ND
Pnap_4641	0.60	transposase IS66	Polaromonas naphthalenivorans CJ2	99.56	IS66

Locus Tag	GC%	IMG Description	Organisms with BLASTP hit against NR (July 2008	% identity	IS Family
Pnap_4642	0.60	transposase IS116/IS110/IS902 family protein	Polaromonas naphthalenivorans CJ2	79.14	IS110
Pnap_4643	0.56	hypothetical protein			IS4
Pnap_4647	0.63	transposase, IS4 family protein	Polaromonas naphthalenivorans CJ2	91.08	IS5
Pnap_4648	0.60	transposase, IS4 family protein	Acidovorax sp. JS42	72.48	IS5
Pnap_4649	0.61	IS1477 transposase	Xanthomonas campestris pv. campestris str. B100	69.78	IS3
Pnap_4650	0.58	transposase IS3/IS911 family protein	Delftia acidovorans SPH-1	76.14	IS3
Pnap_4677	0.55	transposase, IS4 family protein	Verminephrobacter eiseniae EF01-2	47.79	IS5
Pnap_4691	0.71	phage integrase family protein	Polaromonas naphthalenivorans CJ2	78.3	IS91
Pnap_4702	0.62	transposase IS66	Polaromonas naphthalenivorans CJ2	100	IS66
Pnap_4703					ND
Pnap_4704	0.64	Recombinase	Rhizobium leguminosarum bv. viciae 3841	70.72	ND
Pnap_4705	0.61	transposase IS66	Polaromonas naphthalenivorans CJ2	100	IS66
Pnap_4706	0.61	IS66 Orf2 family protein	Polaromonas naphthalenivorans CJ2	90.27	IS66
Pnap_4707	0.65	transposase IS3/IS911 family protein			IS66
Pnap_4708	0.61	IS66 Orf2 family protein	Polaromonas naphthalenivorans CJ2	100	IS66
Pnap_4709	0.61	transposase IS66	Polaromonas naphthalenivorans CJ2	100	IS66
Pnap_4711	0.65	transposase IS116/IS110/IS902 family protein	Polaromonas naphthalenivorans CJ2	97.34	IS110
Pnap_4712	0.56	transposase IS116/IS110/IS902 family protein	Polaromonas naphthalenivorans CJ2	94.38	IS110
Pnap_4713					IS5
Pnap_4723	0.61	transposase			no hits
Pnap_4724	0.63	transposase	Polaromonas naphthalenivorans CJ2	91.79	IS630
Pnap_4725	0.60	transposase	Polaromonas naphthalenivorans CJ2	91.08	IS630
Pnap_4729	0.64	hypothetical protein	Polaromonas naphthalenivorans CJ2	74.09	IS4
Pnap_4740	0.65	putative transposase	Polaromonas naphthalenivorans CJ2	93.84	ISNCY
Pnap_4741	0.66	putative transposase	Acidovorax sp. JS42	68.44	IS91
Pnap_4750	0.54	putative transposase	Acidovorax sp. JS42	66.67	IS630
Pnap_4752	0.65	hypothetical protein	Polaromonas naphthalenivorans CJ2	99.55	IS4
Pnap_4770	0.59	Tn5044 transposase	Pseudomonas aeruginosa PA7	91.16	Tn3
Pnap_4787	0.58	IstB domain protein ATP-binding protein	Polaromonas sp. JS666	87.45	IS21
Pnap_4788	0.60	Integrase, catalytic region	Polaromonas sp. JS666	80.08	IS21
Pnap_4801	0.58	transposase, IS4 family protein	Polaromonas sp. JS666	74.84	ISNCY
Pnap_4819	0.62	transposase (putative)	Polaromonas naphthalenivorans CJ2	82	IS5
Pnap_4820	0.62	hypothetical protein	Polaromonas naphthalenivorans CJ2	86.67	IS5
Pnap_4825	0.66	hypothetical protein	Polaromonas naphthalenivorans CJ2	99.55	IS4
Pnap_4842	0.64	hypothetical protein	Polaromonas naphthalenivorans CJ2	74.09	IS4
Pnap_4937	0.62	transposase, IS4 family protein			IS5
Pnap_4950	0.65	hypothetical protein	Polaromonas naphthalenivorans CJ2	96.15	IS4
Pnap_4951	0.66	hypothetical protein	Polaromonas naphthalenivorans CJ2	84.68	IS4
Pnap_4952	0.62	transposase, IS4 family	Polaromonas naphthalenivorans CJ2	75.93	IS4

APPENDIX C
ISOLATION OF NOVEL *POLAROMONAS* STRAINS FROM SITE 24

Introduction

Polaromonas naphthalenivorans str. CJ2 harbors a unique, chromosomal, *nag*-type naphthalene catabolic operon (Pumphrey and Madsen, 2007; Jeon *et al.*, 2003, 2004; Park *et al.*, 2007a,b). Strain CJ2 was cultivated from Site 24 seep sediment on MSB plates with naphthalene vapor at 10 °C. Naphthalene dioxygenase genes identical or closely related to the *nag* genes of strain CJ2 are prevalent (and expressed *in situ*) within microbial communities at Site 24 (Wilson *et al.*, 1999; Jeon *et al.*, 2003; Yagi and Madsen, 2009). Genome analysis of *P. naphthalenivorans* has revealed strong evidence for acquisition of genetic features via lateral gene transfer, and comparative genome analyses indicate a potential for substantial genomic diversity within the genus *Polaromonas*.

A collection of *Polaromonas* strains closely related to strain CJ2, as well as other hosts of *nag* genes native to Site 24, could be useful for studying the evolution and fitness of naphthalene degraders at the site. Repeated attempts to cultivate hosts of *nag* genes via dilution plating of groundwater from different monitoring wells (e.g., well 36, well 12, well 8, well 46, well 37) collected on various sampling dates between 2005 and 2007 were unsuccessful. This section describes the successful isolation of two potential *Polaromonas* strains closely related to strain CJ2, NSS33 and NSS22, from Site 24 sediment.

Sediment collection and strain isolation

Site 24 seep sediment was collected in November 2006 and stored at 4 °C. In April 2007, 11 g (wet weight) sediment were transfered to a 50 ml conical vial and sterile PBS was added to a final volume of 45 ml. This slurry was homogenized for 90 sec (30 sec on, 30 sec on ice) using a Waring blender. Serial dilutions were prepared in sterile PBS, and 100 μl each from 0.1X and 0.001X dilutions were plated in triplicate on MSB plates. Plates were incubated at 10 °C in a plastic chamber containing naphthalene crystals. After 2 weeks of incubation, 66 colonies with diameter > 1 mm were transferred to fresh MSB plates, and incubated with naphthalene vapor 10 °C. After 1 week of incubation, the transferred strains with robust growth (27 total) were screened by colony PCR. PCR using *nag*-specific primers (Fig. C.1) resulted in an amplicon of the expected size for 5 out of the 27 screened isolates. Restriction fragment length polymorphism analysis, ERIC-PCR and 16S rRNA gene sequence determination were used for further characterization of the isolates (Fig. C.2; Fig. C.3; Fig. C.4). Two isolates, NSS22 and NSS33, were phylogenetically associated with the genus *Polaromonas* (Fig. C.4) and were distinguished from *P. naphthalenivorans* str. CJ2 by their genomic fingerprints (Fig. C.3).

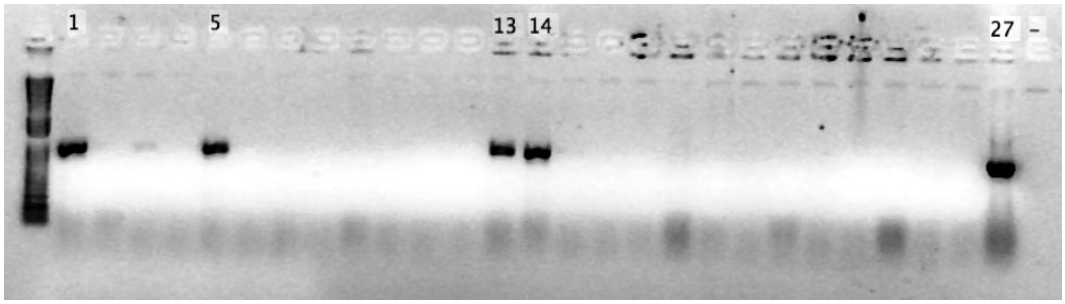


Figure C.1: Detection of *nag*-type genes in 27 uncharacterized Site 24 sediment isolates by PCR. PCR primers CJ2.5219f (5′-TACCACGACCCTTACTACCA-3′) and CJ2.6088r (5′-CGGAGCCATCGTTCTGAC-3′) were used in colony PCR, amplifying an 888-bp region spanning the genes *nagH* to *nagAc* from 5 isolates (shown in lanes 1, 5, 13, 14 and 27 and, hereafter designated NSS16, NSS42, NSS33, NSS22 and JYS62, respectively). The control lane represents equivalent PCR conditions except that 5 μl sterile water was added in place of cells as template. A 1-kb DNA ladder is shown at the far left.

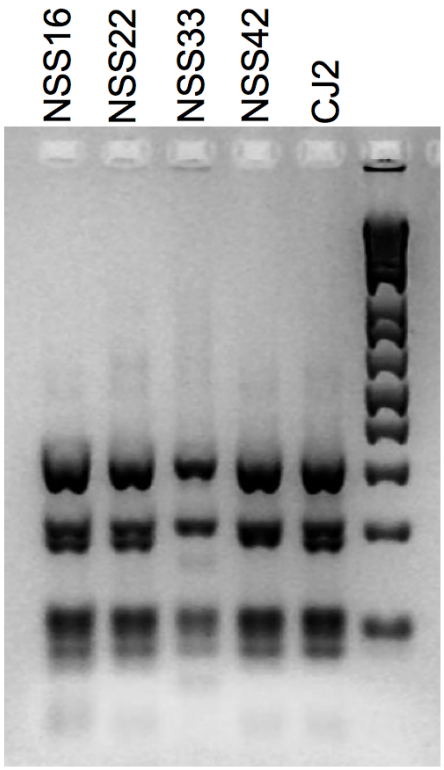


Figure C.2: RFLP fingerprints from PCR amplified Site 24 sediment isolates and *P. naphthalenivorans* str. CJ2 16S rRNA genes digested with restriction enzymes HaeIII and HhaI. Isolates NSS16 and NSS22 showed fingerprint patterns indistinguishable from that of strain CJ2. A 1-kb DNA ladder is shown at the far right.

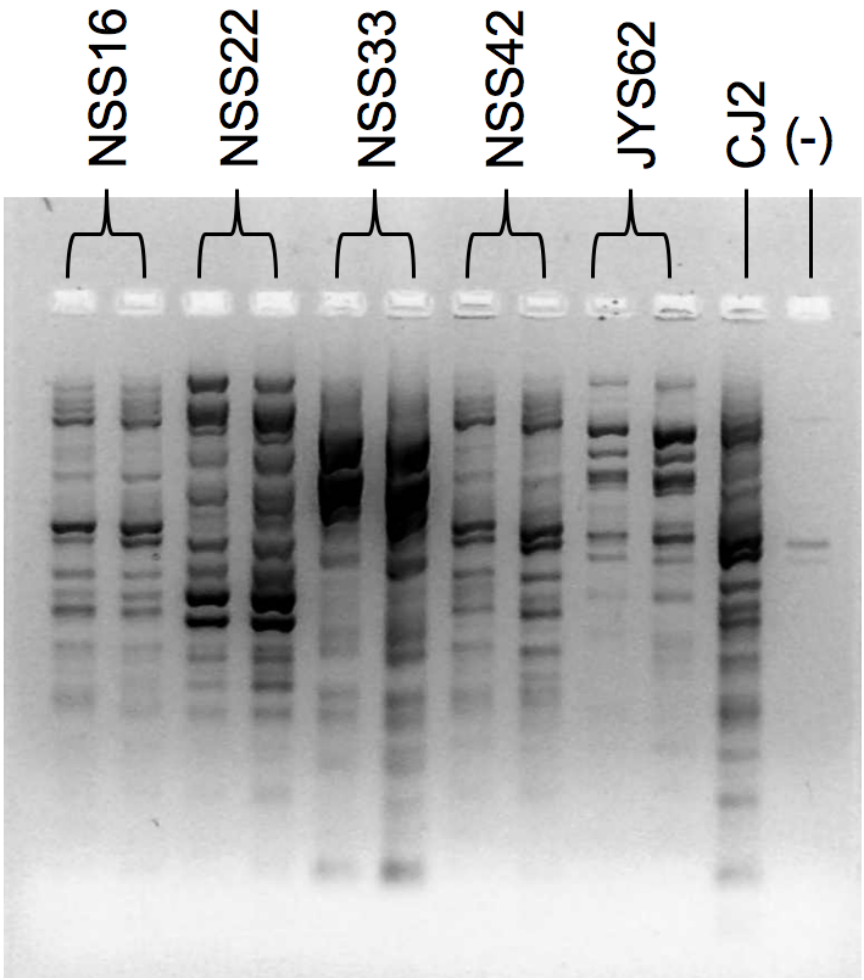


Figure C.3: ERIC-PCR fingerprinting patterns for Site 24 sediment isolates and *P. naphthalenivorans* str. CJ2. ERIC-PCR primers corresponded to conserved motifs in bacterial repetitive elements and have been used to generate genomic fingerprints of diverse environmental strains (e.g., Louws *et al.*, 1994). Colony PCR using PCR primers ERIC-1R (5′-ATGTAAGCTCCTGGGGATTCAC-3′) and ERIC-2 (5′-AAGTAAGTGACTGGGGTGAGCG-3′) was used to generate fingerprints for 5 Site 24 sediment isolates and strain CJ2. For each reaction, 20 μl of each of the ERIC-PCR mixtures was loaded onto a 3% agarose gel. The electrophoretic patterns from isolates NSS16, NSS22, NSS33, NSS42, JYS62 and strain CJ2 are shown. Isolate JYS62 failed to grow upon subsequent transfer and further analysis of this strain was discontinued. The control lane represents equivalent PCR conditions except that 5 μl sterile water was added in place of cells as template.

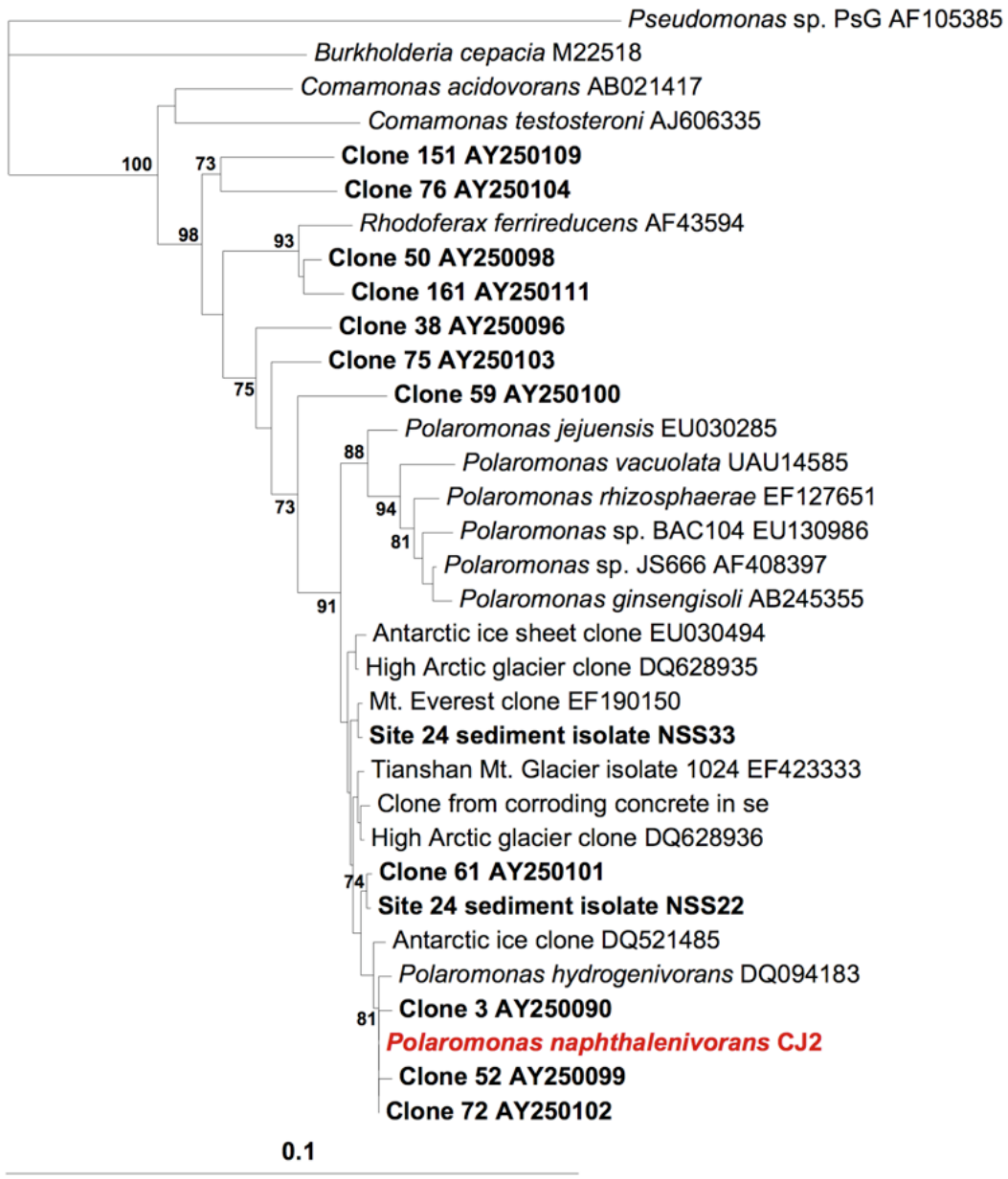


Figure C.4: Phylogenetic analysis of 16S rRNA genes from *Polaromonas* strains catalogued in GenBank and Site 24 sediment isolates. The tree was determined by neighbor-joining analysis implemented in the CLC Free workbench 6.0 package (CLC Bio A/S, Aarhus, Denmark). Bootstrap values from 100 replicates are indicated at the nodes. The scale bar represents the expected number of changes per nucleotide position.

BIBLIOGRAPHY

Aburto, A., A. Fahy, F. Coulon, G. Lethbridge, K. N. Timmis, A. S. Ball, and T. J. McGenity. 2009. Mixed aerobic and anaerobic microbial communities in benzene-contaminated groundwater. J. Appl. Microbiol. **106**:317–328.

Alexander, M. 1999. *Biodegradation and Bioremediation*. 2nd edn. Academic Publ. San Diego, CA.

Alfreider, A., C. Vogt, and W. Babel. 2002. Microbial diversity in an *in situ* reactor system treating monochlorobenzene contaminated groundwater as revealed by 16S ribosomal DNA analysis. Syst. Appl. Microbiol. **25**:232–240.

Alfreider, A., C. Vogt, and W. Babel. 2003. Expression of chlorocatechol 1,2-dioxygenase and chlorocatechol 2,3-dioxygenase genes in chlorobenzene-contaminated subsurface samples. Appl. Environ. Microbiol. **69**:1372–1376.

Anderson-Glenna, M. J., V. Bakkestuen, and N. J. Clipson. 2008. Spatial and temporal variability in epilithic biofilm bacterial communities along an upland river gradient. FEMS Microbiol. Ecol. **64**:407–418.

Andreoni, V. and L. Gianfreda. 2007. Bioremediation and monitoring of aromatic-polluted habitats. Appl. Microbiol. Biotechnol. **76**:287–308.

Anneser, B., F. Einsiedl, R. U. Meckenstock, L. Richters, F. Wisotzky, and C. Griebler. 2008. High-resolution monitoring of biogeochemical gradients in a tar oil-contaminated aquifer. Appl. Geochem. **23**:1715–1730.

Ashelford, K. E., N. A. Chuzhanova, J. C. Fry, A. J. Jones, and A. J. Weightman. 2005. At least 1 in 20 16S rRNA sequence records currently held in public repositories is estimated to contain substantial anomalies. Appl. Environ. Microbiol. **71**:7724–7736.

Badger, M. R. and E. J. Bek. 2008. Multiple Rubisco forms in proteobacteria: their functional significance in relation to CO_2 acquisition by the CBB cycle. J. Exp. Bot. **59**:1525–1541.

Bailly, J., L. Fraissinet-Tachet, M.-C. Verner, J.-C. Debaud, M. Lemaire, M. Wesolowski-Louvel, and R. Marmeisse. 2007. Soil eukaryotic functional diversity, a metatranscriptomic approach. ISME J. **1**:632–642.

Bakermans, C., A. M. Hohnstock-Ashe, S. Padmanabhan, P. Padmanabhan, and E. L. Madsen. 2002. Geochemical and physiological evidence for mixed aerobic and anaerobic field biodegradation of coal tar waste by subsurface microbial communities. Microb. Ecol. **44**:107–117.

Bakermans, C. and E. L. Madsen. 2002a. Detection in coal tar waste-contaminated groundwater of mRNA transcripts related to naphthalene dioxygenase by fluorescent in situ hybridization with tyramide signal amplification. J. Microbiol. Methods **50**:75–84.

Bakermans, C. and E. L. Madsen. 2002b. Diversity of 16S rDNA and naphthalene dioxygenase genes from coal-tar-waste-contaminated aquifer waters. Microb. Ecol. **44**:95–106.

Baldwin, B. R., C. H. Nakatsu, and L. Nies. 2003. Detection and enumeration of aromatic oxygenase genes by multiplex and real-time PCR. Appl. Environ. Microbiol. **69**:3350–3358.

Baldwin, B. R., C. H. Nakatsu, and L. Nies. 2008. Enumeration of aromatic oxygenase genes to evaluate monitored natural attenuation at gasoline-contaminated sites. Water Res. **42**:723–731.

Ball, C. L. and R. L. Crawford. 2006. Bacterial diversity within the planktonic community of an artesian water supply. Can. J. Microbiol. **52**:246–259.

Barton, H. 2006. Introduction to cave microbiology: a review for the non-specialist. J. Cave Karst Stud. **68**:43–54.

Behnke, A., J. Bunge, K. Barger, H.-W. Breiner, V. Alla, and T. Stoeck. 2006. Microeukaryote community patterns along an O_2/H_2S gradient in a supersulfidic anoxic fjord (Framvaren, Norway). Appl. Environ. Microbiol. **72**:3626–3636.

Bekins, B. A., I. M. Cozzarelli, E. M. Godsy, E. Warren, H. I. Essaid, and M. E. Tuccillo. 2001. Progression of natural attenuation processes at a crude oil spill site: II. Controls on spatial distribution of microbial populations. J. Contam. Hydrol. **53**:387–406.

Beller, H., S. Kane, T. Legler, and P. Alvarez. 2002. A real-time polymerase chain reaction method for monitoring anaerobic, hydrocarbon-degrading bacteria based on a catabolic gene. Environ. Sci. Technol. **36**:3977–3984.

Beller, H. R., P. S. G. Chain, T. E. Letain, A. Chakicherla, F. W. Larimer, P. M. Richardson, M. A. Coleman, A. P. Wood, and D. P. Kelly. 2006. The genome sequence of the obligately chemolithoautotrophic, facultatively anaerobic bacterium *Thiobacillus denitrificans*. J. Bacteriol. **188**:1473–1488.

Beman, J. M. and C. A. Francis. 2006. Diversity of ammonia-oxidizing archaea and bacteria in the sediments of a hypernutrified subtropical estuary: Bahía del Tobári, Mexico. Appl. Environ. Microbiol. **72**:7767–7777.

Bennett, P. C., F. K. Hiebert, and J. R. Rogers. 2000. Microbial control of mineral-groundwater equilibria: Macroscale to microscale. Hydrogeol. J. **8**:47–62.

Bentley, S. D. and J. Parkhill. 2004. Comparative genomic structure of prokaryotes. Annu. Rev. Genet. **38**:771–791.

Berge, N., D. Reinhart, and T. Townsend. 2005. The fate of nitrogen in bioreactor landfills. Crit. Rev. Environ. Sci. Technol. **35**:365–399.

Bordenave, S., M. Goñi-urriza, C. Vilette, S. Blanchard, P. Caumette, and R. Duran. 2008. Diversity of ring-hydroxylating dioxygenases in pristine and oil contaminated microbial mats at genomic and transcriptomic levels. Environ. Microbiol. **10**:3201–3211.

Botero, L. M., S. D'Imperio, M. Burr, T. R. McDermott, M. Young, and D. J. Hassett. 2005. Poly(A) polymerase modification and reverse transcriptase PCR amplification of environmental RNA. Appl. Environ. Microbiol. **71**:1267–1275.

Bott, T. L. and L. A. Kaplan. 2002. Autecological properties of 3-chlorobenzoate- degrading bacteria and their population dynamics when introduced into sediments. Microb. Ecol. **43**:199–216.

Botton, S., M. van Heusden, J. R. Parsons, H. Smidt, and N. van Straalen. 2006. Resilience of microbial systems towards disturbances. Crit. Rev. Microbiol. **32**:101–112.

Brad, T., M. Braster, B. M. van Breukelen, B. M., N. M. van Straalen, and W. F. M. Roling. 2008. Eukaryotic diversity in an anaerobic aquifer polluted with landfill leachate. Appl. Environ. Microbiol. **74**:3959–3968.

Brunet, R. C. and L. J. Garcia-Gil. 1996. Sulfide-induced dissimilatory nitrate reduction to ammonia in anaerobic freshwater sediments. FEMS Microbiol. Ecol. **21**:131–138.

Bucheli-Witschel, M., T. Hafner, I. Rüegg, and T. Egli. 2008. Benzene degradation by *Ralstonia pickettii* PKO1 in the presence of the alternative substrate succinate. Biodegradation. Published online 28 November 2008 **doi**:10.1007/s10532–008–9233–z.

Burlage, R. S., L. A. Bemis, A. C. Layton, G. S. Sayler, and F. Larimer. 1990. Comparative genetic organization of incompatibility group P degradative plasmids. J. Bacteriol. **172**:6818–6825.

Burris, R. H. 1972. *Nitrogen fixation-assay methods and techniques. In Methods in enzymology 24B*. Academic Press Inc., New York.

Cantera, J. J. L. and L. Y. Stein. 2007. Molecular diversity of nitrite reductase genes (*nirK*) in nitrifying bacteria. Environ. Microbiol. **9**:765–776.

Casamayor, E. O., R. Massana, S. Benlloch, L. Øvreas, B. Díez, V. J. Goddard, J. M. Gasol, F. Rodríguez-Valera, and C. Pedrós-Alío. 2002. Changes in archaeal, bacterial and eukaryal assemblages along a salinity gradient by comparison of genetic fingerprinting methods in a multipond solar saltern. Environ. Microbiol. **4**:338–348.

Cases, I. and V. de Lorenzo. 2005. Promoters in the environment: Transcriptional regulation in its natural context. Nat. Rev. Microbiol. **3**:105–118.

Caskey, W. H. and J. M. Tiedje. 1979. Evidence for Clostridia as agents of dissimilatory reduction of nitrate to ammonium in soils. Soil Sci. Soc. Am. J. **43**:931–936.

Cébron, A., M.-P. Norini, T. Beguiristain, and C. Leyval. 2008. Real-time PCR quantification of PAH-ring hydroxylating dioxygenase (PAH-RHDα) genes from Gram positive and Gram negative bacteria in soil and sediment samples. J. Microbiol. Methods **73**:148–159.

Chain, P. S. G., V. J. Denef, K. T. Konstantinidis, L. M. Vergez, L. Agullo, V. L. Reyes, L. Hauser, M. Cordova, L. Gomez, M. Gonzalez, M. Land, V. Lao, F. Larimer, J. J. Lipuma, E. Mahenthiralingam, S. A. Malfatti, C. J. Marx, J. J. Parnell, A. Ramette, P. Richardson, M. Seeger, D. Smith, T. Spilker, W. J. Sul, T. V. Tsoi, L. E. Ulrich, I. B. Zhulin, and J. M. Tiedje. 2006. *Burkholderia xenovorans* LB400 harbors a multi-replicon, 9.73-Mbp genome shaped for versatility. Proc. Natl. Acad. Sci. USA **103**:15280–15287.

Chang, H.-K. and G. J. Zylstra. 1998. Novel organization of the genes for phthalate degradation from *Burkholderia cepacia* DBO1. J. Bacteriol. **180**:6529–6537.

Chapelle, F. H. 2003. *Groundwater Microbiology and Geochemistry 2nd edn*. John Wiley and Sons, New York.

Christensen, T. H., P. L. Bjerg, S. A. Banwart, R. Jakobsen, G. Heron, and H.-J. Albrechtsen. 2000. Characterization of redox conditions in groundwater contaminant plumes. J. Contam. Hydrol. **45**:165–241.

Cole, J. 1996. Nitrate reduction to ammonia by enteric bacteria: redundancy, or a strategy for survival during oxygen starvation? FEMS Microbiol. Lett. **136**:1–11.

Coleman, N. V., T. E. Mattes, J. M. Gossett, and J. C. Spain. 2002. Biodegradation of *cis*-dichloroethene as the sole carbon source by a β-proteobacterium. Appl. Environ. Microbiol. **68**:2726–2730.

Colwell, R. K. 2005. EstimateS: Statistical estimation of species richness and shared species from samples. http://purl.oclc.org/estimates.

Connon, S. A., A. Tovanabootr, M. Dolan, K. Vergin, S. J. Giovannoni, and L. Semprini. 2005. Bacterial community composition determined by culture-independent and-dependent methods during propane-stimulated bioremediation in trichloroethene-contaminated groundwater. Environ. Microbiol. **7**:165–178.

Cordwell, S. J. 1999. Microbial genomes and "missing" enzymes: redefining biochemical pathways. Arch. Microbiol. **172**:269–279.

Corliss, J. O. 2002. Biodiversity and biocomplexity of the protists and an overview of their significant roles in maintenance of our biosphere. Acta Protozool. **41**:199–220.

Cozzarelli, I. M., B. A. Bekins, M. J. Baedecker, G. R. Aiken, R. P. Eganhouse, and M. E. Tuccillo. 2001. Progression of natural attenuation processes at a crude-oil spill site: I. Geochemical evolution of the plume. J. Contam. Hydrol. **53**:369–385.

Da Silva, M. L. B., R. Kamath, and P. J. J. Alvarez. 2006. Effect of simulated rhizodeposition on the relative abundance of polynuclear aromatic hydrocarbon catabolic genes in a contaminated soil. Environ. Toxicol. Chem. **25**:386–391.

Dawson, S. C. and N. R. Pace. 2002. Novel kingdom-level eukaryotic diversity in anoxic environments. Proc. Natl. Acad. Sci. USA **99**:8324–8329.

de la Cruz, F. and J. Davies. 2000. Horizontal gene transfer and the origin of species: lessons from bacteria. Trends Microbiol. **8**:128–133.

Dean-Ross, D., J. Moody, and C. Cerniglia. 2002. Utilization of mixtures of polycyclic aromatic hydrocarbons by bacteria isolated from contaminated sediment. FEMS Microbiol. Ecol. **41**:1–7.

DeBruyn, J. M., C. S. Chewning, and G. S. Sayler. 2007. Comparative quantitative prevalence of Mycobacteria and functionally abundant *nidA, nahAc*, and *nagAc* dioxygenase genes in coal tar contaminated sediments. Environ. Sci. Technol. **41**:5426–5432.

del Castillo, T. and J. L. Ramos. 2007. Simultaneous catabolite repression between glucose and toluene metabolism in *Pseudomonas putida* is channeled through different signaling pathways. J. Bacteriol. **189**:6602–6610.

DeLong, E. F. 1992. Archaea in coastal marine environments. Proc. Natl. Acad. Sci. USA **89**:5685–5689.

Denef, V. J., J. A. Klappenbach, M. A. Patrauchan, C. Florizone, J. L. M. Rodrigues, T. V. Tsoi, W. Verstraete, L. D. Eltis, and J. M. Tiedje. 2006. Genetic and genomic insights into the role of benzoate-catabolic pathway redundancy in *Burkholderia xenovorans* LB400. Appl. Environ. Microbiol. **72**:585–595.

Denef, V. J., J. Park, T. V. Tsoi, J. M. Rouillard, H. Zhang, J. A. Wibbenmeyer, W. Verstraete, E. Gulari, S. A. Hashsham, and J. M. Tiedje. 2004. Biphenyl and benzoate metabolism in a genomic context: outlining genome-wide metabolic networks in *Burkholderia xenovorans* LB400. Appl. Environ. Microbiol. **70**:4961–4970.

Denef, V. J., M. A. Patrauchan, C. Florizone, J. Park, T. V. Tsoi, W. Verstraete, J. M. Tiedje, and L. D. Eltis. 2005. Growth substrate- and phase-specific expression of biphenyl, benzoate, and C1 metabolic pathways in *Burkholderia xenovorans* LB400. J. Bacteriol. **187**:7996–8005.

Dennis, J. J. and G. J. Zylstra. 2004. Complete sequence and genetic organization of pDTG1, the 83 kilobase naphthalene degradation plasmid from *Pseudomonas putida* strain NCIB 9816-4. J. Mol. Biol. **341**:753–768.

DeSantis, T. Z., P. Hugenholtz, N. Larsen, M. Rojas, E. L. Brodie, K. Keller, T. Huber, D. Dalevi, P. Hu, and G. L. Andersen. 2006. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. Appl. Environ. Microbiol. **72**:5069–5072.

Devers, M., N. El Azhari, N. U. Kolic, and F. Martin-Laurent. 2007. Detection and organization of atrazine-degrading genetic potential of seventeen bacterial isolates belonging to divergent taxa indicate a recent common origin of their catabolic functions. FEMS Microbiol. Lett. **273**:78–86.

Diaz, E. 2008. *Microbial Biodegradation: Genomics and Molecular Biology*. Caister Academic Press, Norwich UK.

Dietz, T., E. Ostrom, and P. C. Stern. 2003. The struggle to govern the commons. Science **302**:1907–1912.

Diez, B., C. Pedrós-Alío, T. L. Marsh, and R. Massana. 2001a. Application of denaturing gradient gel electrophoresis (DGGE) to study the diversity of marine picoeukaryotic assemblages and comparison of DGGE with other molecular techniques. Appl. Environ. Microbiol. **67**:2942–2951.

Diez, B., C. Pedros-Alio, and R. Massana. 2001b. Study of genetic diversity of eukaryotic picoplankton in different oceanic regions by small-subunit rRNA gene cloning and sequencing. Appl. Environ. Microbiol. **67**:2932–2941.

Dionisi, H. M., C. S. Chewning, K. H. Morgan, F. M. Menn, J. P. Easter, and G. S. Sayler. 2004. Abundance of dioxygenase genes similar to *Ralstonia* sp. strain U2 *nagAc* is correlated with naphthalene concentrations in coal tar-contaminated freshwater sediments. Appl. Environ. Microbiol. **70**:3988–3995.

Dixon, P. 2003. VEGAN, a package of R functions for community ecology. J. Veg. Sci. **14**:927–930.

Dojka, M. A., P. Hugenholtz, S. K. Haack, and N. R. Pace. 1998. Microbial diversity in a hydrocarbon- and chlorinated-solvent-contaminated aquifer undergoing intrinsic bioremediation. Appl. Environ. Microbiol. **64**:3869–3877.

Dos Santos, V. A., S. Heim, E. R. Moore, M. Strtz, and K. N. Timmis. 2004. Insights into the genomic basis of niche specificity of *Pseudomonas putida* KT2440. Environ. Microbiol. **6**:1264–1286.

Electric Power Research Institute 1996. Characterization and monitoring before and after source removal at a former manufactured gas plant (MGP) disposal site. TR-105921 EPRI Palo Alto, CA.

Elhai, J., A. Vepritskiy, A. Muro-Pastor, E. Flores, and C. Wolk. 1997. Reduction of conjugal transfer efficiency by three restriction activities of *Anabaena* sp. strain PCC 7120. J. Bacteriol. **179**:1998–2005.

Ellis, M., J. Grossmann, R. Eady, and S. Hasnain. 2007. Genomic analysis reveals widespread occurrence of new classes of copper nitrite reductases. J. Biol. Inorg. Chem. **12**:1119–1127.

Elshahed, M., N. Youssef, A. Spain, C. Sheik, F. Najar, L. Sukharnikov, B. Roe, J. Davis, P. Schloss, and V. Bailey. 2008. Novelty and uniqueness patterns of rare members of the soil biosphere. Appl. Environ. Microbiol. **74**:5422–5428.

Engel, A. S., N. Lee, M. L. Porter, L. A. Stern, P. C. Bennett, and M. Wagner. 2003. Filamentous *Epsilonproteobacteria* dominate microbial mats from sulfidic cave springs. Appl. Environ. Microbiol. **69**:5503–5511.

Eriksson, M., G. Dalhammar, and W. W. Mohn. 2002. Bacterial growth and biofilm production on pyrene. FEMS Microbiol. Ecol. **40**:21–27.

Ettwig, K., S. Shima, K. v. d. Pas-Schoonen, J. Kahnt, M. Medema, M. Jetten, and M. Strous. 2008. Denitrifying bacteria anaerobically oxidize methane in the absence of Archaea. Environ. Microbiol. **10**.

Euringer, K. and T. Lueders. 2008. An optimised PCR/T-RFLP fingerprinting approach for the investigation of protistan communities in groundwater environments. J. Microbiol. Methods **75**:262–268.

Ewing, B. and P. Green. 1998. Base-calling of automated sequencer traces using Phred. II. Error probabilities.

Ewing, B., L. D. Hillier, M. C. Wendl, and P. Green. 1998. Base-calling of automated sequencer traces using Phred. I. Accuracy assessment.

Falkowski, P. G., T. Fenchel, and E. F. Delong. 2008. The microbial engines that drive earth's biogeochemical cycles. Science **320**:1034–1039.

Ferguson, A. S., W. E. Huang, K. A. Lawson, R. Doherty, O. Gibert, K. W. Dickson, A. S. Whiteley, L. A. Kulakov, I. P. Thompson, R. M. Kalin, and M. J. Larkin. 2007. Microbial analysis of soil and groundwater from a gasworks site and comparison with a sequenced biological reactive barrier remediation process. J. Appl. Microbiol. **102**:1227–1238.

Finlay, B. J. and G. F. Esteban. 1998. Freshwater protozoa: biodiversity and ecological function. Biodiversity Conserv. **7**:1163–1186.

Fleischmann, R., M. Adams, O. White, R. Clayton, E. Kirkness, A. Kerlavage, C. Bult, J. Tomb, B. Dougherty, and J. Merrick. 1995. Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. Science **269**.

Flett, F., V. Mersinias, and C. Smith. 1997. High efficiency intergeneric conjugal transfer of plasmid DNA from *Escherichia coli* to methyl DNA-restricting streptomycetes. FEMS Microbiol. Lett. **155**:223–229.

Foght, J., J. Aislabie, S. Turner, C. E. Brown, J. Ryburn, D. J. Saul, and W. Lawson. 2004. Culturable bacteria in subglacial sediments and ice from two southern hemisphere glaciers. Microb. Ecol. **47**:329–340.

Foissner, W. 2006. Biogeography and dispersal of micro-organisms: A review emphasizing protists. Acta Protozool. **45**:111–136.

Folke, C., S. Carpenter, B. Walker, M. Scheffer, T. Elmqvist, L. Gunderson, and C. S. Holling. 2004. Regime shifts, resilience, and biodiversity in ecosystem management. Annu. Rev. Ecol. Evol. Syst. **35**:557–581.

Fredrickson, J. K., M. F. Romine, A. S. Beliaev, J. M. Auchtung, M. E. Driscoll, T. S. Gardner, K. H. Nealson, A. L. Osterman, G. Pinchuk, and J. L. Reed. 2008. Towards environmental systems biology of *Shewanella*. Nat. Rev. Microbiol. **6**:592–603.

Frias-Lopez, J., Y. Shi, G. W. Tyson, M. L. Coleman, S. C. Schuster, S. W. Chisholm, and E. F. DeLong. 2008. Microbial community gene expression in ocean surface waters. Proc. Natl. Acad. Sci. USA **105**:3805–3810.

Friedrich, B. and E. Schwartz. 1993. Molecular biology of hydrogen utilization in aerobic chemolithotrophs. Annu. Rev. Microbiol. **47**:351–383.

Friedrich, C. G. 1982. Depression of hydrogenase during limitation of electron donors and derepression of ribulosebisphosphate carboxylase during carbon limitation of *Alcaligenes eutrophus*. J. Bacteriol. **149**:203–210.

Frost, L. S., R. Leplae, A. O. Summers, and A. Toussaint. 2005. Mobile genetic elements: The agents of open source evolution. Nat. Rev. Microbiol. **3**:722–732.

Fuenmayor, S. L., M. Wild, A. L. Boyes, and P. A. Williams. 1998. A gene cluster encoding steps in conversion of naphthalene to gentisate in *Pseudomonas* sp. strain U2. J. Bacteriol. **180**:2522–2530.

Fuhrman, J. A., I. Hewson, M. S. Schwalbach, J. A. Steele, M. V. Brown, and S. Naeem. 2006. Annually reoccurring bacterial communities are predictable from ocean conditions. Proc. Natl. Acad. Sci. USA **103**:13104–13109.

Gerdes, K., S. K. Christensen, and A. Lobner-Olesen. 2005. Prokaryotic toxin-antitoxin stress response loci. Nat. Rev. Microbiol. **3**:371–82.

Ghiglione, J. F., M. Larcher, and P. Lebaron. 2005. Spatial and temporal scales of variation in bacterioplankton community structure in the NW mediterranean sea. Aquatic Microb. Ecol. **40**:229–240.

Gieg, L. and J. Suflita. 2002. Detection of anaerobic metabolites of saturated and aromatic hydrocarbons in petroleum-contaminated aquifers. Environ. Sci. Technol. **36**:3755–3762.

Gilbert, J. A., D. Field, Y. Huang, R. Edwards, W. Li, P. Gilna, and I. Joint. 2008. Detection of large numbers of novel sequences in the metatranscriptomes of complex marine microbial communities. PLoS ONE **3**:e3042.

Gomes, N. C. M., L. R. Borges, R. Paranhos, F. N. Pinto, E. Krögerrecklenfort, L. C. S. Mendonça-Hagler, and K. Smalla. 2007. Diversity of *ndo* genes in mangrove sediments exposed to different sources of polycyclic aromatic hydrocarbon pollution. Appl. Environ. Microbiol. **73**:7392–7399.

Good, I. 1953. The population frequencies of species and the estimation of population parameters. Biometrika **40**:237–264.

Gordon, D., C. Abajian, and P. Green. 1998. Consed: A graphical tool for sequence finishing. Genome Res. **8**:195–202.

Griebler, C., M. Safinowski, A. Vieth, H. H. Richnow, and R. U. Meckenstock. 2004. Combined application of stable carbon isotope analysis and specific metabolites determination for assessing in situ degradation of aromatic hydrocarbons in a tar oil-contaminated aquifer. Environ. Sci. Technol. **38**:617–631.

Guerin, W. F. and S. A. Boyd. 1995. Maintenance and induction of naphthalene degradation activity in *Pseudomonas putida* and an *Alcaligenes* sp. under different culture conditions. Appl. Environ. Microbiol. **61**:4061–4068.

Haack, S. K. and B. A. Bekins. 2000. Microbial populations in contaminant plumes. Hydrogeol. J. **8**:63–76.

Haack, S. K., L. R. Fogarty, T. G. West, E. W. Alm, J. T. McGuire, D. T. Long, D. W. Hyndman, and L. J. Forney. 2004. Spatial and temporal changes in microbial community structure associated with recharge-influenced chemical gradients in a contaminated aquifer. Environ. Microbiol. **6**:438–448.

Habe, H. and T. Omori. 2003. Genetics of polycyclic aromatic hydrocarbon metabolism in diverse aerobic bacteria. Biosci. Biotechnol. Biochem. **67**:225–243.

Han, C. and P. Chain. 2006. Finishing repetitive regions automatically with Dupfinisher. In: Arabnia H. R., Valafar, H. eds. Las Vegas, Nevada, USA: CSREA Press.

Hara, H., E. Masai, Y. Katayama, and M. Fukuda. 2000. The 4-oxalomesaconate hydratase gene, involved in the protocatechuate 4,5-cleavage pathway, is essential to vanillate and syringate degradation in *Sphingomonas paucimobilis* SYK-6. J. Bacteriol. **182**:6950–6957.

Harayama, S., M. Kok, and E. L. Neidle. 1992. Functional and evolutionary relationships among diverse oxygenases. Annu. Rev. Microbiol. **46**:565–601.

Hartig, E., U. Schiek, K.-U. Vollack, and W. G. Zumft. 1999. Nitrate and nitrite control of respiratory nitrate reduction in denitrifying *Pseudomonas stutzeri* by a two-component regulatory system homologous to NarXL of *Escherichia coli*. J. Bacteriol. **181**:3658–3665.

Hayashi, N. R., H. Arai, T. Kodama, and Y. Igarashi. 1997. The novel genes, *cbbQ* and *cbbO*, located downstream from the RubisCO genes of *Pseudomonas hydrogenothermophila*, affect the conformational states and activity of RubisCO. Biochem. Biophys. Res. Commun. **241**:565–569.

Heiss-Blanquet, S., Y. Benoit, C. Marechaux, and F. Monot. 2005. Assessing the role of alkane hydroxylase genotypes in environmental samples by competitive PCR. J. Appl. Microbiol. **99**:1392–1403.

Heitzer, A., O. F. Webb, J. E. Thonnard, and G. S. Sayler. 1992. Specific and quantitative assessment of naphthalene and salicylate bioavailability by using a bioluminescent catabolic reporter bacterium. Appl. Environ. Microbiol. **58**:1839–1846.

Hendrickx, B., W. Dejonghe, W. Boënne, M. Brennerova, M. Cernik, T. Lederer, M. Bucheli-Witschel, L. Bastiaens, W. Verstraete, E. M. Top, L. Diels, and D. Springael. 2005. Dynamics of an oligotrophic bacterial aquifer community during contact with a groundwater plume contaminated with benzene toluene, ethylbenzene, and xylenes: an in situ mesocosm study. Appl. Environ. Microbiol. **71**:3815–3825.

Hohnstock-Ashe, A. M., S. E. Bilotta, and E. L. Madsen. 2001. Further biogeochemical characterization of a trichloroethene-contaminated fractured dolomite aquifer: Electron source and microbial communities involved in reductive dechlorination. Environ. Sci. Technol. **35**.

Holland, S. 2001. Analytic rarefaction, version 1.3. University of Georgia, Athens.

Holling, C. S. 1973. Resilience and stability of ecological systems. Annu. Rev. Ecol. Syst. **4**:1–23.

Holmgren, A. 1989. Thioredoxin and glutaredoxin systems. J. Biol. Chem. **264**:13963–13966.

Hose, L. D., A. N. Palmer, M. V. Palmer, D. E. Northup, P. J. Boston, and H. R. DuChene. 2000. Microbiology and geochemistry in a hydrogen-sulphide-rich karst environment. Chem. Geol. **169**:399–423.

Hsiao, W., I. Wan, S. J. Jones, and F. S. L. Brinkman. 2003. IslandPath: aiding detection of genomic islands in prokaryotes. Bioinformatics **19**:418–420.

Huang, W. E., A. Ferguson, A. C. Singer, K. Lawson, I. P. Thompson, R. M. Kalin, M. J. Larkin, M. J. Bailey, and A. S. Whiteley. 2009. Resolving genetic functions within microbial populations: In situ analyses using rRNA and mRNA stable isotope probing coupled with single-cell Raman-fluorescence in situ hybridization. Appl. Environ. Microbiol. **75**:234–241.

Huber, T., G. Faulkner, and P. Hugenholtz. 2004. Bellerophon: a program to detect chimeric sequences in multiple sequence alignments. Bioinformatics **20**:2317–2319.

Hudak, J. and J. Fuhrman. 1988. Effects of four organic pollutants on the growth of natural marine bacterioplankton populations. Mar. Ecol. Prog. Ser. **47**:185–194.

Hughes Martiny, J. and D. Field. 2005. Ecological perspectives on the sequenced genome collection. Ecol. Lett. **8**:1334–1345.

Hunkeler, D., R. Aravena, K. Berry-Spark, and E. Cox. 2005. Assessment of degradation pathways in an aquifer with mixed chlorinated hydrocarbon contamination using stable isotope analysis. Environ. Sci. Technol. **39**:5975–5981.

Ihssen, J. and T. Egli. 2005. Global physiological analysis of carbon-and energy-limited growing escherichia coli confirms a high degree of catabolic flexibility and preparedness for mixed substrate utilization. Environ. Microbiol. **7**:1568–1581.

Inagaki, F., U. Tsunogai, M. Suzuki, A. Kosaka, H. Machiyama, K. Takai, T. Nunoura, K. H. Nealson, and K. Horikoshi. 2004. Characterization of C1-metabolizing prokaryotic communities in methane seep habitats at the Kuroshima Knoll, Southern Ryukyu Arc, by analyzing *pmoA, mmoX, mxaF, mcrA,* and 16S rRNA genes. Appl. Environ. Microbiol. **70**:7445–7455.

Irgens, R. L., J. J. Gosink, and J. T. Staley. 1996. *Polaromonas vacuolata* gen. nov., sp. nov., a psychrophilic, marine, gas vacuolate bacterium from Antarctica. Int. J. Syst. Evol. Microbiol. **46**:822–826.

Islam, F. S., A. G. Gault, C. Boothman, D. A. Polya, J. M. Charnock, D. Chatterjee, and J. R. Lloyd. 2004. Role of metal-reducing bacteria in arsenic release from Bengal delta sediments. Nature **430**:68–71.

Jeon, C. O., P. Padmanabhan, C. DeRito, J. R. Snape, and E. L. Madsen. 2003. Discovery of a bacterium, with distinctive dioxygenase, that is responsible for *in situ* biodegradation in contaminated sediment. Proc. Natl. Acad. Sci. USA **100**:13591–13596.

Jeon, C. O., M. Park, H. S. Ro, W. Park, and E. L. Madsen. 2006. The naphthalene catabolic (*nag*) genes of *Polaromonas naphthalenivorans* CJ2: evolutionary implications for two gene clusters and novel regulatory control. Appl. Environ. Microbiol. **72**:1086–1095.

Jeon, C. O., W. Park, W. C. Ghiorse, and E. L. Madsen. 2004. *Polaromonas naphthalenivorans* sp. nov., a naphthalene-degrading bacterium from naphthalene-contaminated sediment. Int. J. Syst. Evol. Microbiol. **54**:93–97.

Jin, C. F., M. Mata, and D. J. Fink. 1994. Rapid construction of deleted DNA fragments for use as internal standards in competitive PCR. Genome Res. **3**:252–255.

Johnsen, A. R., L. Y. Wick, and H. Harms. 2005. Principles of microbial PAH-degradation in soil. Environ. Pollut. **133**:71–84.

Jones, J., D. J. Studholme, C. G. Knight, and G. M. Preston. 2007. Integrated bioinformatic and phenotypic analysis of RpoN-dependent traits in the plant growth-promoting bacterium *Pseudomonas fluorescens* SBW25. Environ. Microbiol. **9**:3046–3064.

Jongman, R., C. Ter Braak, and O. Van Tongeren. 1995. *Data Analysis in Community and Landscape Ecology*. Cambridge University Press, Cambridge.

Kampfer, P., H.-J. Busse, and E. Falsen. 2006. *Polaromonas aquatica* sp. nov., isolated from tap water. Int. J. Syst. Evol. Microbiol. **56**:605–608.

Kan, A. and M. Tomson. 1990. Ground water transport of hydrophobic organic compounds in the presence of dissolved organic matter. Environ. Toxicol. Chem. **9**:253–263.

Kan, J., K. Wang, and F. Chen. 2006. Temporal variation and detection limit of an estuarine bacterioplankton community analyzed by denaturing gradient gel electrophoresis (DGGE). Aquatic microbial ecology **42**:7–18.

Kelso, B. H. L., R. V. Smith, R. J. Laughlin, and S. D. Lennox. 1997. Dissimilatory nitrate reduction in anaerobic sediments leading to river nitrite accumulation. Appl. Environ. Microbiol. **63**:4679–4685.

Kemp, P. and J. Aller. 2004. Bacterial diversity in aquatic and other environments: what 16S rDNA libraries can tell us. FEMS Microbiol. Ecol. **47**:161–177.

Kertesz, M. A. 2000. Riding the sulfur cycle - metabolism of sulfonates and sulfate esters in gram-negative bacteria. FEMS Microbiol. Rev. **24**:135–175.

Khan, A. A., R.-F. Wang, W.-W. Cao, D. R. Doerge, D. Wennerstrom, and C. E. Cerniglia. 2001. Molecular cloning, nucleotide sequence, and expression of genes encoding a polycyclic aromatic ring dioxygenase from *Mycobacterium* sp. strain PYR-1. Appl. Environ. Microbiol. **67**:3577–3585.

Kim, S., J. Choi, and H. Kahng. 2007. A proteomics strategy for the analysis of bacterial biodegradation pathways. OMICS **11**:280–294.

Kleikemper, J., O. Pelz, M. H. Schroth, and J. Zeyer. 2002. Sulfate-reducing bacterial community response to carbon source amendments in contaminated aquifer microcosms. FEMS Microbiol. Ecol. **42**:109–118.

Kolasa, J. 2000. The biology and ecology of lotic microturbellarians. Freshwater Biol. **44**:5–14.

Konneke, M., A. E. Bernhard, J. R. d. l. Torre, C. B. Walker, J. B. Waterbury, and D. A. Stahl. 2005. Isolation of an autotrophic ammonia-oxidizing marine archaeon. Nature **437**:543–546.

Kornberg, A., N. N. Rao, and D. Ault-Riche. 1999. Inorganic polyphosphate: A molecule of many functions. Annu. Rev. Biochem. **68**:89–125.

Kovach, M. E., P. H. Elzer, D. S. Hill, G. T. Robertson, M. A. Farris, R. M. Roop, and K. M. Peterson. 1995. Four new derivatives of the broad-host-range cloning vector pBBR1MCS, carrying different antibiotic-resistance cassettes. Gene **166**:175–176.

Kovarova-Kovar, K. and T. Egli. 1998. Growth kinetics of suspended microbial cells: From single-substrate-controlled growth to mixed-substrate kinetics. Microbiol. Mol. Biol. Rev. **62**:646–666.

Kulakov, L. A., S. Chen, C. C. R. Allen, and M. J. Larkin. 2005. Web-type evolution of *Rhodococcus* gene clusters associated with utilization of naphthalene. Appl. Environ. Microbiol. **71**:1754–1764.

Kurtz, S., A. Phillippy, A. Delcher, M. Smoot, M. Shumway, C. Antonescu, and S. Salzberg. 2004. Versatile and open software for comparing large genomes. Genome Biol. **5**:R12.

Lam, P., M. M. Jensen, G. Lavik, D. F. McGinnis, B. Muller, C. J. Schubert, R. Amann, B. Thamdrup, and M. M. M. Kuypers. 2007. Linking crenarchaeal and bacterial nitrification to anammox in the Black Sea. Proc. Natl. Acad. Sci. USA **104**:7104–7109.

Lane, D. J. 1991. *Nucleic Acid Techniques in Bacterial Systematics, Stackebrandt, E. and Goodfellow, M. eds.* Wiley, Chichester.

Larimer, F. W., P. Chain, L. Hauser, J. Lamerdin, S. Malfatti, L. Do, M. L. Land, D. A. Pelletier, J. T. Beatty, A. S. Lang, F. R. Tabita, J. L. Gibson, T. E. Hanson, C. Bobst, J. L. Torres y Torres, C. Peres, F. H. Harrison, J. Gibson, and C. S. Harwood. 2004. Complete genome sequence of the metabolically versatile photosynthetic bacterium *Rhodopseudomonas palustris*. Nat. Biotech. **22**:55–61.

Laurie, A. D. and L. G. Jones. 1999. Conserved and hybrid meta-cleavage operons from PAH-degrading *Burkholderia* RP007. Biochem. Biophys. Res. Commun. **262**:308–314.

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**:1814–1817.

Laverman, A., R. Canavan, C. Slomp, and P. Cappellen. 2007. Potential nitrate removal in a coastal freshwater sediment (Haringvliet Lake, The Netherlands) and response to salinization. Water Res. **41**:3061–3068.

Laverman, A. M., P. Van Cappellen, D. van Rotterdam-Los, C. Pallud, and J. Abell. 2006. Potential rates and pathways of microbial nitrate reduction in coastal sediments. FEMS Microbiol. Ecol. **58**:178–192.

Lendenmann, U., M. Snozzi, and T. Egli. 1996. Kinetics of the simultaneous utilization of sugar mixtures by escherichia coli in continuous culture. Appl. Environ. Microbiol. **62**:1493–1499.

Lengeler, J. W. 1999. *Biology of the Prokaryotes, Lengeler, Joseph W. and Drews, Gerhart and Schlegel, Hans Günter*. Blackwell Publishing, New York.

Leys, N. M. E. J., A. Ryngaert, L. Bastiaens, W. Verstraete, E. M. Top, and D. Springael. 2004. Occurrence and phylogenetic diversity of *Sphingomonas* strains in soils contaminated with polycyclic aromatic hydrocarbons. Appl. Environ. Microbiol. **70**:1944–1955.

Lin, X., M. I. Scranton, A. Y. Chistoserdov, R. Varela, and G. T. Taylor. 2008. Spatiotemporal dynamics of bacterial populations in the anoxic Cariaco Basin. Limnol. Oceanogr. **53**:37–51.

Liou, J. S. C., C. M. DeRito, and E. L. Madsen. 2008. Field-based and laboratory stable isotope probing surveys of the identities of both aerobic and anaerobic benzene-metabolizing microorganisms in freshwater sediment. Environ. Microbiol. **10**:1964–1977.

Liu, J., T. Dietz, S. R. Carpenter, M. Alberti, C. Folke, E. Moran, A. N. Pell, P. Deadman, T. Kratz, J. Lubchenco, E. Ostrom, Z. Ouyang, W. Provencher, C. L. Redman, S. H. Schneider, and W. W. Taylor. 2007. Complexity of coupled human and natural systems. Science **317**:1513–1516.

Liu, Y. Q., T. D. Yao, N. Z. Jiao, S. C. Kang, Y. H. Zeng, and S. J. Huang. 2006. Microbial community structure in moraine lakes and glacial meltwaters, Mount Everest. FEMS Microbiol. Lett. **265**:98–105.

Lloyd-Jones, G., A. D. Laurie, D. W. Hunter, and R. Fraser. 1999. Analysis of catabolic genes for naphthalene and phenanthrene degradation in contaminated New Zealand soils. FEMS Microbiol. Ecol. **29**:69–79.

López, Z., J. Vila, J. J. Ortega-Calvo, and M. Grifoll. 2008. Simultaneous biodegradation of creosote-polycyclic aromatic hydrocarbons by a pyrene-degrading mycobacterium. Appl. Microbiol. Biotechnol. **78**:165–172.

Lopez-Garcia, P., A. Vereshchaka, and D. Moreira. 2007. Eukaryotic diversity associated with carbonates and fluid-seawater interface in Lost City hydrothermal field. Environ. Microbiol. **9**:546–554.

Louws, F., D. Fulbright, C. Stephens, and F. De Bruijn. 1994. Specific genomic fingerprints of phytopathogenic *xanthomonas* and *pseudomonas* pathovars and strains generated with repetitive sequences and PCR. Appl. Environ. Microbiol. **60**:2286–2295.

Loy, A., W. Beisker, and H. Meier. 2005. Diversity of bacteria growing in natural mineral water after bottling. Appl. Environ. Microbiol. **71**:3624–3632.

Lozada, M., J. P. R. Mercadal, L. D. Guerrero, W. D. Di Marzio, M. A. Ferrero, and H. M. Dionisi. 2008. Novel aromatic ring-hydroxylating dioxygenase genes from coastal marine sediments of Patagonia. BMC Microbiol. **8**.

Ludwig, W., O. Strunk, R. Westram, L. Richter, H. Meier, Yadhukumar, A. Buchner, T. Lai, S. Steppi, G. Jobb, W. Förster, I. Brettske, S. Gerber, A. W. Ginhart, O. Gross, S. Grumann, S. Hermann, R. Jost, A. König, T. Liss, R. Lüßmann, M. May, B. Nonhoff, B. Reichel, R. Strehlow, A. Stamatakis, N. Stuckmann, A. Vilbig, M. Lenke, T. Ludwig, A. Bode, and K.-H. Schleifer. 2004. ARB: a software environment for sequence data. Nucleic Acids Res. **32**:1363–1371.

Luo, Q. W., L. R. Krumholz, F. Z. Najar, A. D. Peacock, B. A. Roe, D. C. White, and M. S. Elshahed. 2005. Diversity of the microeukaryotic community in sulfide-rich Zodletone spring (Oklahoma). Appl. Environ. Microbiol. **71**:6175–6184.

Luton, P. E., J. M. Wayne, R. J. Sharp, and P. W. Riley. 2002. The *mcrA* gene as an alternative to 16S rRNA in the phylogenetic analysis of methanogen populations in landfill. Microbiology **148**:3521–3530.

Madison, L. L. and G. W. Huisman. 1999. Metabolic engineering of poly(3-hydroxyalkanoates): From DNA to plastic. Microbiol. Mol. Biol. Rev. **63**:21–53.

Madsen, E. L. 2008. *Environmental Microbiology: From Genomes to Biogeochemistry*. Blackwell, Malden MA.

Madsen, E. L. and W. Ghiorse. 1993. *Aquatic Microbiology: An Ecological Approach.* Blackwell Scientific Publications, Boston.

Madsen, E. L., J. L. Sinclair, and W. C. Ghiorse. 1991. In situ biodegradation - microbiological patterns in a contaminated aquifer. Science **252**:830–833.

Marciset, O. and B. Mollet. 1994. Multifactorial experimental design for optimizing transformation: Electroporation of *Streptococcus thermophilus*. Biotechnology and Bioengineering **43**:490–496.

Markowitz, V. M., E. Szeto, K. Palaniappan, Y. Grechkin, K. Chu, I.-M. A. Chen, I. Dubchak, I. Anderson, A. Lykidis, K. Mavromatis, N. N. Ivanova, and N. C. Kyrpides. 2008. The integrated microbial genomes (IMG) system in 2007: data content and analysis tool extensions. Nucleic Acids Res. **36**:D528–D533.

Mattes, T. E., A. K. Alexander, P. M. Richardson, A. C. Munk, C. S. Han, P. Stothard, and N. V. Coleman. 2008. The genome of *Polaromonas*sp. strain JS666: Insights into the evolution of a hydrocarbon- and xenobiotic-degrading bacterium, and features of relevance to biotechnology. Appl. Environ. Microbiol. **74**:6405–6416.

Mattison, R. G. and S. Harayama. 2001. The predatory soil flagellate *Heteromita globosa* stimulates toluene biodegradation by a *Pseudomonas* sp. FEMS Microbiol. Lett. **194**:39–45.

Mattison, R. G., H. Taki, and S. Harayama. 2002. The bacterivorous soil flagellate *Heteromita globosa* reduces bacterial clogging under denitrifying conditions in sand-filled aquifer columns. Appl. Environ. Microbiol. **68**:4539–4545.

Mazéas, L., V. Vigneron, K. Le-Ménach, H. Budzinski, J. Audic, N. Bernet, and T. Bouchez. 2008. Elucidation of nitrate reduction pathways in anaerobic bioreactors using a stable isotope approach. Rapid Commun. Mass Spectrom. **22**:1746–1750.

McGrath, K., S. Thomas-Hall, C. Cheng, L. Leo, A. Alexa, S. Schmidt, and P. Schenk. 2008. Isolation and analysis of mRNA from environmental microbial communities. J. Microbiol. Methods **75**:172–176.

McGuire, J., D. Long, M. Klug, S. Haack, and D. Hyndman. 2002. Evaluating behavior of oxygen, nitrate, and sulfate during recharge and quantifying reduction rates in a contaminated aquifer. Environ. Sci. Technol. **36**:2693–2700.

McGuire, J. T., E. W. Smith, D. T. Long, D. W. Hyndman, S. K. Haack, M. J. Klug, and M. A. Velbel. 2000. Temporal variations in parameters reflecting terminal-electron-accepting processes in an aquifer contaminated with waste fuel and chlorinated solvents. Chem. Geol. **169**:471–485.

McKnight, D. M., R. L. Runkel, C. M. Tate, J. H. Duff, and D. L. Moorhead. 2004. Inorganic N and P dynamics of Antarctic glacial meltwater streams as

controlled by hyporheic exchange and benthic autotrophic communities. J. N. Am. Benthol. Soc. **23**:171–188.

McLeod, M. P., R. L. Warren, W. W. L. Hsiao, N. Araki, M. Myhre, C. Fernandes, D. Miyazawa, W. Wong, A. L. Lillquist, D. Wang, M. Dosanjh, H. Hara, A. Petrescu, R. D. Morin, G. Yang, J. M. Stott, J. E. Schein, H. Shin, D. Smailus, A. S. Siddiqui, M. A. Marra, S. J. M. Jones, R. Holt, F. S. L. Brinkman, K. Miyauchi, M. Fukuda, J. E. Davies, W. W. Mohn, and L. D. Eltis. 2006. The complete genome of *Rhodococcus* sp. RHA1 provides insights into a catabolic powerhouse. Proc. Natl. Acad. Sci. USA **103**:15582–15587.

Melo, A. M. P., T. M. Bandeiras, and M. Teixeira. 2004. New insights into Type II NAD(P)H:quinone oxidoreductases. Microbiol. Mol. Biol. Rev. **68**:603–616.

Michaelis, W., R. Seifert, K. Nauhaus, T. Treude, V. Thiel, M. Blumenberg, K. Knittel, A. Gieseke, K. Peterknecht, T. Pape, A. Boetius, R. Amann, B. B. Jorgensen, F. Widdel, J. Peckmann, N. V. Pimenov, and M. B. Gulin. 2002. Microbial reefs in the Black Sea fueled by anaerobic oxidation of methane. Science **297**:1013–1015.

Monds, R. D., P. D. Newell, J. A. Schwartzman, and G. A. O'Toole. 2006. Conservation of the Pho regulon in *Pseudomonas fluorescens* Pf0-1. Appl. Environ. Microbiol. **72**:1910–1924.

Morita, R. Y. 1988. Bioavailability of energy and its relationship to growth and starvation survival in nature. Can. J. Microbiol. **34**:436–441.

Moser, R. and U. Stahl. 2001. Insights into the genetic diversity of initial dioxygenases from PAH-degrading bacteria. Appl. Microbiol. Biotechnol. **55**:609–618.

Munster, U. 1993. Concentrations and fluxes of organic carbon substrates in the aquatic environment. Antonie van Leeuwenhoek **63**:243–274.

Murarka, I., E. Neuhauser, M. Sherman, B. B. Taylor, D. M. Mauro, J. Ripp, and T. Taylor. 1992. Organic substances in the subsurface: Delineation, migration, and remediation. J. Hazard. Mater. **32**:245–261.

Murase, J., M. Noll, and P. Frenzel. 2006. Impact of protists on the activity and structure of the bacterial community in a rice field soil. Appl. Environ. Microbiol. **72**:5436–5444.

Nakatsu, C. H., K. Hristova, S. Hanada, X.-Y. Meng, J. R. Hanson, K. M. Scow, and Y. Kamagata. 2006. *Methylibium petroleiphilum* gen. nov., sp. nov., a novel methyl tert-butyl ether-degrading methylotroph of the Betaproteobacteria. Int. J. Syst. Evol. Microbiol. **56**:983–989.

National Research Council 2000. *Natural attenuation for Groundwater Remediation*. National Academy Press, Washington DC.

Nelson, E. C., G. S, E. J. C, M. G. X., and R. G. Luthy. 1996. Chemical characterization of coal tar-water interfacial films. Environ. Sci. Technol. **30**:1014–1022.

Nelson, J. D., S. E. Boehme, C. E. Reimers, R. M. Sherrell, and L. J. Kerkhof. 2008. Temporal patterns of microbial community structure in the Mid-Atlantic Bight. FEMS Microbiol. Ecol. **65**:484–493.

Neuhauser, E. F., J. A. Ripp, N. A. Azzolina, E. L. Madsen, D. M. Mauro, and B. Skelly. 2008. MNA of MGP Coal tar MAHs and PAHs in groundwater: a 14-year field study. Ground Water Monit. Rem. (submitted).

Newsham, K. K., J. Rolf, D. A. Pearce, and R. J. Strachan. 2004. Differing preferences of Antarctic soil nematodes for microbial prey. Eur. J. Soil. Biol. **40**:1–8.

Ní Chadhain, S. M., R. S. Norman, K. V. Pesce, J. J. Kukor, and G. J. Zystra. 2006. Microbial dioxygenase gene population shifts during polycyclic aromatic hydrocarbon biodegradation. Appl. Environ. Microbiol. **72**:4078–4087.

Nies, D. 2003. Efflux-mediated heavy metal resistance in prokaryotes. FEMS Microbiol. Rev. **27**:313–339.

Nojiri, H., M. Shintani, and T. Omori. 2004. Divergence of mobile genetic elements involved in the distribution of xenobiotic-catabolic capacity. Appl. Microbiol. Biotechnol. **64**:154–174.

Norton, J. M., M. G. Klotz, L. Y. Stein, D. J. Arp, P. J. Bottomley, P. S. G. Chain, L. J. Hauser, M. L. Land, F. W. Larimer, M. W. Shin, and S. R. Starkenburg. 2008. Complete genome sequence of *Nitrosospira multiformis*, an ammonia-oxidizing bacterium from the soil environment. Appl. Environ. Microbiol. **74**:3559–3572.

Novarino, G., A. Warren, H. Butler, G. Lambourne, A. Boxshall, J. Bateman, N. E. Kinner, R. W. Harvey, R. A. Mosse, and B. Teltsch. 1997. Protistan communities in aquifers: a review. FEMS Microbiol. Rev. **20**:261–275.

Nyyssönen, M., R. Piskonen, and M. Itävaara. 2006. A targeted real-time PCR assay for studying naphthalene degradation in the environment. Microb. Ecol. **52**:533–543.

Nyyssönen, M., R. Piskonen, and M. Itävaara. 2008. Monitoring aromatic hydrocarbon biodegradation by functional marker genes. Environ. Pollut. **154**:192–202.

Olsen, G. J., D. J. Lane, S. J. Giovannoni, N. R. Pace, and D. A. Stahl. 1986. Microbial ecology and evolution: a ribosomal RNA approach. Annu. Rev. Microbiol. **40**:337–365.

O'Neil, R. A., D. E. Holmes, M. V. Coppi, L. A. Adams, M. J. Larrahondo, J. E. Ward, K. P. Nevin, T. L. Woodard, H. A. Vrionis, A. L. N'Guessan, and D. R. Lovley. 2008. Gene transcript analysis of assimilatory iron limitation in Geobacteraceae during groundwater bioremediation. Environ. Microbiol. **10**:1218–1230.

Oremland, R. and J. Stolz. 2003. The ecology of arsenic. Science **300**:939–944.

Pace, N. 1997. A molecular view of microbial diversity and the biosphere. Science **276**:734–740.

Pace, N., D. Stahl, D. Lane, and G. Olsen. 1985. Analyzing natural microbial populations by rRNA sequences. ASM News **51**:4–12.

Page, K. A., S. A. Connon, and S. J. Giovannoni. 2004. Representative freshwater bacterioplankton isolated from Crater Lake, Oregon. Appl. Environ. Microbiol. **70**:6542–6550.

Palmer, M., E. Bernhardt, E. Chornesky, S. Collins, A. Dobson, C. Duke, B. Gold, R. Jacobson, S. Kingsland, R. Kranz, M. Mappin, M. L. Martinez, F. Micheli, J. Morse, M. Pace, M. Pascual, S. Palumbi, O. J. Reichman, A. Simons, A. Townsend, and M. Turner. 2004. Ecology: Ecology for a crowded planet. Science **304**:1251–1252.

Panikov, N. S. and M. V. Sizova. 2007. Growth kinetics of microorganisms isolated from Alaskan soil and permafrost in solid media frozen down to -35 degrees C. FEMS Microbiol. Ecol. **59**:500–512.

Park, M., Y. Jeon, H. H. Jang, H.-S. Ro, W. Park, E. L. Madsen, and C. O. Jeon. 2007a. Molecular and biochemical characterization of 3-hydroxybenzoate 6-hydroxylase from *Polaromonas naphthalenivorans* CJ2. Appl. Environ. Microbiol. **73**:5146–5152.

Park, M., Y. Jeon, E. L. Madsen, and C. O. Jeon. 2007b. Protection of *Polaromonas naphthalenivorans* CJ2 from naphthalene toxicity by extracellular polysaccharide capsules. J. Appl. Biol. Chem. **50**:41–45.

Paulsen, I. T., L. Nguyen, M. K. Sliwinski, R. Rabus, and M. H. Saier. 2000. Microbial genome analyses: comparative transport capabilities in eighteen prokaryotes. J. Mol. Biol **301**:75–100.

Peel, M. C. and R. C. Wyndham. 1999. Selection of *clc*, *cba*, and *fcb* chlorobenzoate-catabolic genotypes from groundwater and surface waters

adjacent to the Hyde Park, Niagara Falls, Chemical Landfill. Appl. Environ. Microbiol. **65**:1627–1635.

Pelz, O., M. Tesar, R. M. Wittich, E. R. B. Moore, K. N. Timmis, and W. R. Abraham. 1999. Towards elucidation of microbial community metabolic pathways: unravelling the network of carbon sharing in a pollutant-degrading bacterial consortium by immunocapture and isotopic ratio mass spectrometry. Environ. Microbiol. **1**:167–174.

Peng, R., A. Xiong, Y. Xue, X. Fu, F. Gao, W. Zhao, Y. Tian, and Q. Yao. 2008. Microbial biodegradation of polyaromatic hydrocarbons. FEMS Microbiol. Rev. **32**:927–955.

Perez-Pantoja, D., R. De la Iglesia, D. H. Pieper, and B. Gonzalez. 2008. Metabolic reconstruction of aromatic compounds degradation from the genome of the amazing pollutant-degrading bacterium *Cupriavidus necator* JMP134. FEMS Microbiol. Rev. **32**:736–794.

Pett-Ridge, J., W. L. Silver, and M. K. Firestone. 2006. Redox fluctuations frame microbial community impacts on N-cycling rates in a humid tropical forest soil. Biogeochemistry **81**:95–110.

Pickup, R. W., G. Rhodes, M. L. Alamillo, H. E. H. Mallinson, S. F. Thornton, and D. N. Lerner. 2001. Microbiological analysis of multi-level borehole samples from a contaminated groundwater system. J. Contam. Hydrol. **53**:269–284.

Pitcher, R. S. and N. J. Watmough. 2004. The bacterial cytochrome cbb_3 oxidases. BBA-Bioenergetics **1655**:388–399.

Pohlmann, A., W. F. Fricke, F. Reinecke, B. Kusian, H. Liesegang, R. Cramm, T. Eitinger, C. Ewering, M. Potter, E. Schwartz, A. Strittmatter, I. Vosz, G. Gottschalk, A. Steinbuchel, B. Friedrich, and B. Bowien. 2006. Genome sequence of the bioplastic-producing "Knallgas" bacterium *Ralstonia eutropha* H16. Nat. Biotech. **24**:1257–1262.

Poliakoff, M., J. M. Fitzpatrick, T. R. Farren, and P. T. Anastas. 2002. Green chemistry: Science and politics of change. Science **297**:807–810.

Poretsky, R. S., N. Bano, A. Buchan, G. Lecleir, J. Kleikemper, M. Pickering, W. M. Pate, M. A. Moran, and J. T. Hollibaugh. 2005. Analysis of microbial gene transcripts in environmental samples. Appl. Environ. Microbiol. **71**:4121–4126.

Priefert, H., J. Rabenhorst, and A. Steinbuchel. 1997. Molecular characterization of genes of *Pseudomonas* sp. strain HR199 involved in bioconversion of vanillin to protocatechuate. J. Bacteriol. **179**:2595–2607.

Providenti, M. A., R. E. Shaye, K. D. Lynes, N. T. McKenna, J. M. O'Brien, S. Rosolen, R. C. Wyndham, and L. B. Lambert. 2006. The locus coding for the 3-nitrobenzoate dioxygenase of *Comamonas* sp. strain JS46 is flanked by IS1071 elements and is subject to deletion and inversion events. Appl. Environ. Microbiol. **72**:2651–2660.

Pruesse, E., C. Quast, K. Knittel, B. M. Fuchs, W. Ludwig, J. Peplies, and F. O. Glöckner. 2007. SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. Nucleic Acids Res. **35**:7188–7196.

Pumphrey, G. M. and E. L. Madsen. 2007. Naphthalene metabolism and growth inhibition by naphthalene in *Polaromonas naphthalenivorans* strain CJ2. Microbiology **153**:3730–3738.

Raghoebarsing, A. A., A. Pol, K. T. van de Pas-Schoonen, A. J. P. Smolders, K. F. Ettwig, W. I. C. Rijpstra, S. Schouten, J. S. S. Damste, H. J. M. Op den Camp, M. S. M. Jetten, and M. Strous. 2006. A microbial consortium couples anaerobic methane oxidation to denitrification. Nature **440**:918–921.

Ramos, J. L., M. T. Gallegos, S. Marques, M. Ramos-Gonzalez, M. Espinosa-Urgel, and A. Segura. 2001. Responses of Gram-negative bacteria to certain environmental stressors. Curr. Opin. Microbiol. **4**:166–171.

Rappé, M. S. and S. J. Giovannoni. 2003. The uncultured microbial majority. Annu. Rev. Microbiol. **57**:369–394.

Rediers, H., P. B. Rainey, J. Vanderleyden, and R. D. Mot. 2005. Unraveling the secret lives of bacteria: Use of in vivo expression technology and differential fluorescence induction promoter traps as tools for exploring niche-specific gene expression. Microbiol. Mol. Biol. Rev. **69**:217–261.

Ren, Q. and I. T. Paulsen. 2005. Comparative analyses of fundamental differences in membrane transport capabilities in prokaryotes and eukaryotes. PLoS Comp. Biol. **1**:e27.

Reva, O. N., P. F. Hallin, H. Willenbrock, T. Sicheritz-Ponten, B. Tummler, and D. W. Ussery. 2008. Global features of the *Alcanivorax borkumensis* SK2 genome. Environ. Microbiol. **10**:614–625.

Reva, O. N., C. Weinel, M. Weinel, K. Bohm, D. Stjepandic, J. D. Hoheisel, and B. Tummler. 2006. Functional genomics of stress response in *Pseudomonas putida* KT2440. J. Bacteriol. **188**:4079–4092.

Rhee, S. K., X. Liu, L. Wu, S. C. Chong, X. Wan, and J. Zhou. 2004. Detection of genes involved in biodegradation and biotransformation in microbial communities by using 50-mer oligonucleotide microarrays. Appl. Environ. Microbiol. **70**:4303–4317.

Richards, T. A., A. A. Vepritskiy, D. E. Gouliamova, and S. A. Nierzwicki-Bauer. 2005. The molecular diversity of freshwater picoeukaryotes from an oligotrophic lake reveals diverse, distinctive and globally dispersed lineages. Environ. Microbiol. **7**:1413–1425.

Rittmann, B. E., M. Hausner, F. Loffler, N. G. Love, G. Muyzer, S. Okabe, D. B. Oerther, J. Peccia, L. Raskin, and M. Wagner. 2006. A vista for microbial ecology and environmental biotechnology. Environ. Sci. Technol. **40**:1096–1103.

Roberts, R. J., T. Vincze, J. Posfai, and D. Macelis. 2007. REBASE–enzymes and genes for DNA restriction and modification. Nucleic Acids Res. **35**.

Rogers, S. W., T. B. Moorman, and S. K. Ong. 2007. Fluorescent in situ hybridization and micro-autoradiography applied to ecophysiology in soil. Soil Sci. Soc. Am. J. **71**:620–631.

Roling, W. F. M., B. M. van Breukelen, M. Braster, M. T. Goeltom, J. Groen, and H. W. van Verseveld. 2000. Analysis of microbial communities in a landfill leachate polluted aquifer using a new method for anaerobic physiological profiling and 16S rDNA based fingerprinting. Microb. Ecol. **40**:177–188.

Roten, C.-A., P. Gamba, J.-L. Barblan, and D. Karamata. 2002. Comparative genometrics (CG): a database dedicated to biometric comparisons of whole genomes. Nucleic Acids Res. **30**:142–144.

Rotthauwe, J.-H., K. P. Witzel, and W. Liesack. 1997. The ammonia monooxygenase structural gene *amoA* as a functional marker: molecular fine-scale analysis of natural ammonia-oxidizing population. Appl. Environ. Microbiol. **63**.

Saier, M. H., C. V. Tran, and R. D. Barabote. 2006. TCDB: the Transporter Classification Database for membrane transport protein analyses and information. Nucleic Acids Res. **34**.

Salminen, J. M., P. M. Tuomi, and K. S. Jorgensen. 2008. Functional gene abundances (*nahAc, alkB, xylE*) in the assessment of the efficacy of bioremediation. Appl. Biochem. Biotechnol. **151**:638–652.

Samanta, S. K., O. V. Singh, and R. K. Jain. 2002. Polycyclic aromatic hydrocarbons: environmental pollution and bioremediation. Trends Biotechnol. **20**:243–248.

Sanger, F., G. M. Air, B. G. Barrell, N. L. Brown, A. R. Coulson, J. C. Fiddes, C. Hutchison 3rd, P. M. Slocombe, and M. Smith. 1977. Nucleotide sequence of bacteriophage phi X174 DNA. Nature **265**:687–695.

Sanseverino, J., B. M. Applegate, J. M. King, and G. S. Sayler. 1993. Plasmid-mediated mineralization of naphthalene, phenanthrene, and anthracene. Appl. Environ. Microbiol. **59**:1931–1937.

Schäfer, A., A. Tauch, W. Jäger, J. Kalinowski, G. Thierbach, and A. Phler. 1994. Small mobilizable multi-purpose cloning vectors derived from the *Escherichia coli* plasmids pK18 and pK19: selection of defined deletions in the chromosome of *Corynebacterium glutamicum.* Gene **145**:69–73.

Schlesinger, W. H. 1997. *Biogeochemistry: An Analysis of Global Change, 2nd edn.* Academic Press, San Diego, CA.

Schmalenberger, A., S. Hodge, A. Bryant, M. J. Hawkesford, B. K. Singh, and M. A. Kertesz. 2008. The role of *Variovorax* and other *Comamonadaceae* in sulfur transformations by microbial wheat rhizosphere communities exposed to different sulfur fertilization regimes. Environ. Microbiol. **10**:1486–1500.

Schneiker, S., V. A. P. M. dos Santos, D. Bartels, T. Bekel, M. Brecht, J. Buhrmester, T. N. Chernikova, R. Denaro, M. Ferrer, C. Gertler, A. Goesmann, O. V. Golyshina, F. Kaminski, A. N. Khachane, S. Lang, B. Linke, A. C. McHardy, F. Meyer, T. Nechitaylo, A. Pühler, D. Regenhardt, O. Rupp, J. S. Sabirova, W. Selbitschka, M. M. Yakimov, K. N. Timmis, F.-J. Vorholter, S. Weidner, O. Kaiser, and P. N. Golyshin. 2006. Genome sequence of the ubiquitous hydrocarbon-degrading marine bacterium *Alcanivorax borkumensis*. Nat. Biotech. **24**:997–1004.

Schumacher, W. and P. M. H. Kroneck. 1992. Anaerobic energy metabolism of the sulfur-reducing bacterium "Spirillum" 5175 during dissimilatory nitrate reduction to ammonia. Arch. Microbiol. **157**:464–470.

Schutte, U. M. E., Z. Abdo, S. J. Bent, C. Shyu, C. J. Williams, J. D. Pierson, and L. J. Forney. 2008. Advances in the use of terminal restriction fragment length polymorphism (T-RFLP) analysis of 16S rRNA genes to characterize microbial communities. Appl. Microbiol. Biotechnol. **80**:365–380.

Schwarzenbach, R. P., B. I. Escher, K. Fenner, T. B. Hofstetter, C. A. Johnson, U. von Gunten, and B. Wehrli. 2006. The challenge of micropollutants in aquatic systems. Science **313**:1072–1077.

Silver, W. L., D. J. Herman, and M. K. Firestone. 2001. Dissimilatory nitrate reduction to ammonium in upland tropical forest soils. Ecology **82**:2410–2416.

Silver, W. L., A. W. Thompson, A. Reich, J. J. Ewel, and M. K. Firestone. 2005. Nitrogen cycling in tropical plantation forests: Potential controls on nitrogen retention. Ecol. Appl. **15**:1604–1614.

Simon, J. 2002. Enzymology and bioenergetics of respiratory nitrite ammonification. FEMS Microbiol. Rev. **26**:285–309.

Simon, K. S., J. Gibert, P. Petitot, and R. Laurent. 2001. Spatial and temporal patterns of bacterial density and metabolic activity in a karst aquifer. Archiv fur Hydrobiologie **151**:67–82.

Simon, R., U. Priefer, and A. Pühler. 1983. A broad host range mobilization system for in vivo genetic engineering: transposon mutagenesis in gram negative bacteria. Bio/technology **1**:784–791.

Sipilä, T. P., A. K. Keskinen, M.-L. Akerman, C. Fortelius, K. Haahtela, and K. Yrjälä. 2008. High aromatic ring-cleavage diversity in birch rhizosphere: PAH treatment-specific changes of I.E.3 group extradiol dioxygenases and 16S rRNA bacterial communities in soil. ISME J. **2**.

Sizova, M. and N. Panikov. 2007. *Polaromonas hydrogenivorans* sp. nov., a psychrotolerant hydrogen-oxidizing bacterium from Alaskan soil. Int. J. Syst. Evol. Microbiol. **57**:616–619.

Smith, C. J., D. B. Nedwell, L. F. Dong, and A. M. Osborn. 2007. Diversity and abundance of nitrate reductase genes (*narG* and *napA*), nitrite reductase genes (*nirS* and *nrfA*), and their transcripts in estuarine sediments. Appl. Environ. Microbiol. **73**:3612–3622.

Sorensen, J. 1978. Capacity for denitrification and reduction of nitrate to ammonia in a coastal marine sediment. Appl. Environ. Microbiol. **35**:301–305.

Stach, J. E. M. and R. G. Burns. 2002. Enrichment versus biofilm culture: a functional and phylogenetic comparison of polycyclic aromatic hydrocarbon-degrading microbial communities. Environ. Microbiol. **4**:169–182.

Stanier, R. Y., N. Palleroni, and M. Douderoff. 1966. The aerobic pseudomonads: a taxonomic study. J. Gen. Microbiol. **43**:159–271.

Stapleton, R. D., G. S. Sayler, J. M. Boggs, E. L. Libelo, T. Stauffer, and W. G. Macintyre. 2000. Changes in subsurface catabolic gene frequencies during natural attenuation of petroleum hydrocarbons. Environ. Sci. Technol. **34**:1991–1999.

Stein, L. Y., D. J. Arp, P. M. Berube, P. S. G. Chain, L. Hauser, M. S. M. Jetten, M. G. Klotz, F. W. Larimer, J. M. Norton, H. J. M. Op den Camp, M. Shin, and X. Wei. 2007. Whole-genome analysis of the ammonia-oxidizing bacterium, *Nitrosomonas eutropha* C91: implications for niche adaptation. Environ. Microbiol. **9**:2993–3007.

Stumm, W. and J. J. Morgan. 1996. *Aquatic Chemistry: Chemical Equilibria and Rates in Natural Waters, 3rd edn.* Wiley, New York.

Suarez, M. and H. Rifai. 2002. Evaluation of BTEX remediation by natural attenuation at a coastal facility. Ground Water Monit. Rem. **22**:62–77.

Suenaga, H., T. Ohnuki, and K. Miyazaki. 2007. Functional screening of a metagenomic library for genes involved in microbial degradation of aromatic compounds. Environ. Microbiol. **9**:2289–2297.

Sundberg, C., J. K. Stendahl, K. Tonderski, and P. Lindgren. 2007. Overland flow systems for treatment of landfill leachates: Potential nitrification and structure of the ammonia-oxidising bacterial community during a growing season. Soil Biol. Biochem. **39**:127–138.

Taylor, B., D. Mauro, and J. Foxwell. 1996. Characterization and monitoring before and after source removal at a former manufactured gas plant (MGP) disposal site. Electric Power Research Institute, Palo Alto, CA, R-105921.

Taylor, P. M. and P. H. Janssen. 2005. Variations in the abundance and identity of class II aromatic ring-hydroxylating dioxygenase genes in groundwater at an aromatic hydrocarbon-contaminated site. Environ. Microbiol. **7**:140–146.

Templer, P., W. Silver, J. Pett-Ridge, K. M. DeAngelis, and M. Firestone. 2008. Plant and microbial controls on nitrogen retention and loss in a humid tropical forest. Ecology **89**:3030–3040.

Thomas, C. M. and K. M. Nielsen. 2005. Mechanisms of, and barriers to, horizontal gene transfer between bacteria. Nat. Rev. Microbiol. **3**:711–721.

Thompson, I., C. van der Gast, L. Ciric, and A. Singer. 2005. Bioaugmentation for bioremediation: the challenge of strain selection. Environ. Microbiol. **7**:909–915.

Tiedje, J. 1988. *Ecology of denitrification and dissimilatory nitrate reduction to ammonium. In Biology of Anaerobic Microorganisms. Zehnder, A. J. B. (ed).* Wiley and Sons, New York.

Tiedje, J., J. Sorensen, and Y. Chang. 1981. Assimilatory and dissimilatory nitrate reduction: perspectives and methodology for simultaneous measurement of several nitrogen cycle processes. Ecological Bulletin (Stockholm) **33**:331–342.

Tiedje, J. M., S. A. J., M. D. D., and R. J. A.. 1982. Denitrification: Ecological niches, competition and survival. Antonie van Leeuwenhoek **48**:569–583.

Top, E. M. and D. Springael. 2003. The role of mobile genetic elements in bacterial adaptation to xenobiotic organic compounds. Curr. Opin. Biotechnol. **14**:262–269.

Tringe, S. G., C. von Mering, A. Kobayashi, A. A. Salamov, K. Chen, H. W. Chang, M. Podar, J. M. Short, and E. J. a. Mathur. 2005. Comparative metagenomics of microbial communities. Science **308**:554–557.

Tropel, D. and J. R. van der Meer. 2004. Bacterial transcriptional regulators for degradation pathways of aromatic compounds. Microbiol. Mol. Biol. Rev. **68**:474–500.

Tuomi, P. M., J. M. Salminen, and K. S. Jorgensen. 2004. The abundance of *nahAc* genes correlates with the ^{14}C naphthalene mineralization potential in petroleum hydrocarbon-contaminated oxic soil layers. FEMS Microbiol. Ecol. **51**:99–107.

Ulrich, G., G. Breit, I. Cozzarelli, and J. Suflita. 2003. Sources of sulfate supporting anaerobic metabolism in a contaminated aquifer. Environ. Sci. Technol. **37**:1093–1099.

Urich, T., A. Lanzen, J. Qi, D. Huson, C. Schleper, and S. C. Schuster. 2008. Simultaneous assessment of soil microbial community structure and function through analysis of the meta-transcriptome. PLoS ONE **3**.

U.S. Environmental Protection Agency 1997. *Use of Monitored Natural Attenuation at Superfund, RCRA Corrective Action, and Underground Storage Tank Sites.* OSWER Monitored Natural Attenuation Policy, OSWER Directive 9200.4-17 EPA, Office of Solid Waste and Emergency Response Washington , DC.

Vaillancourt, F. H., J. T. Bolin, and L. D. Eltis. 2006. The ins and outs of ring-cleaving dioxygenases. Crit. Rev. Biochem. Mol. Biol. **41**:241–267.

van der Meer, J. R. and V. Sentchilo. 2003. Genomic islands and the evolution of catabolic pathways in bacteria. Curr. Opin. Biotechnol. **14**:248–254.

van Veen, H. W. 1997. Phosphate transport in prokaryotes: molecules, mediators and mechanisms. Antonie van Leeuwenhoek **72**:299–315.

Velázquez, F., V. de Lorenzo, and M. Valls. 2006. The *m*-xylene biodegradation capacity of *Pseudomonas putida* mt-2 is submitted to adaptation to abiotic stresses: evidence from expression profiling of *xyl* genes. Environ. Microbiol. **8**:591–602.

Vieites, J., M. Guazzaroni, A. Beloqui, P. Golyshin, and M. Ferrer. 2009. Metagenomics approaches in systems microbiology. FEMS Microbiol. Rev. **33**:236–255.

von Wintzingerode, F., U. B. Gobel, and E. Stackebrandt. 1997. Determination of microbial diversity in environmental samples: pitfalls of PCR-based rRNA analysis. FEMS Microbiol. Rev. **21**:213–229.

Vroblesky, D. and F. Chapelle. 1994. Temporal and spatial changes of terminal electron-accepting processes in a petroleum hydrocarbon-contaminated aquifer and the significance for contaminant biodegradation. Wat. Resour. Res. **30**:1561–1561.

Wawrik, B., J. Paul, and F. Tabita. 2002. Real-time PCR quantification of rbcL (ribulose-1, 5-bisphosphate carboxylase/oxygenase) mRNA in diatoms and pelagophytes. Appl. Environ. Microbiol. **68**:3771–3779.

Weiss, J. V. and I. M. Cozzarelli. 2008. Biodegradation in contaminated aquifers: Incorporating microbial/molecular methods. Groundwater **46**:305–322.

Welsh, D., G. Castadelli, M. Bartoli, D. Poli, M. Careri, R. d. Wit, and P. Viaroli. 2001. Denitrification in an intertidal seagrass meadow, a comparison of ^{15}N-isotope and acetylene-block techniques: dissimilatory nitrate reduction to ammonia as a source of N_2O? Marine Biol. **139**:1029–1036.

Wen, A., M. Fegan, C. Hayward, S. Chakraborty, and L. I. Sly. 1999. Phylogenetic relationships among members of the *Comamonadaceae*, and description of *Delftia acidovorans* (den Dooren de Jong 1926 and Tamaoka *et al.* 1987) gen. nov., comb. nov. Int. J. Syst. Evol. Microbiol. **49**:567–576.

Weon, H.-Y., S.-H. Yoo, S.-B. Hong, S.-W. Kwon, E. Stackebrandt, S.-J. Go, and B.-S. Koo. 2008. *Polaromonas jejuensis* sp. nov., isolated from soil in korea. Int. J. Syst. Evol. Microbiol. **58**:1525–1528.

White, D. 2000. *The Physiology and Biochemistry of Prokaryotes, 2nd edn.* Oxford Univ. Press, New York.

Whyte, L., A. Schultz, J. B. v. Beilen, A. Luz, V. H. Pellizari, D. Labbe, and C. Greer. 2002. Prevalence of alkane monooxygenase genes in Arctic and Antarctic hydrocarbon-contaminated and pristine soils. FEMS Microbiol. Ecol. **41**:141–150.

Wick, L. Y., N. Pasche, S. M. Bernasconi, O. Pelz, and H. Harms. 2003. Characterization of multiple-substrate utilization by anthracene-degrading *Mycobacterium frederiksbergense* LB501T. Appl. Environ. Microbiol. **69**:6133–6142.

Wilmes, P., S. Simmons, V. Denef, and J. Banfield. 2009. The dynamic genetic repertoire of microbial communities. FEMS Microbiol. Rev. **33**:109–132.

Wilson, K. 1987. *Current protocols in molecular biology.* Wiley, New York.

Wilson, M. S., C. Bakermans, and E. L. Madsen. 1999. In situ, real-time catabolic gene expression: extraction and characterization of naphthalene dioxygenase mRNA transcripts from groundwater. Appl. Environ. Microbiol. **65**:80–87.

Wilson, M. S., J. B. Herrick, C. O. Jeon, D. E. Hinman, and E. L. Madsen. 2003. Horizontal transfer of *phnAc* dioxygenase genes within one of two phenotypically and genotypically distinctive naphthalene-degrading guilds from adjacent soil environments. Appl. Environ. Microbiol. **69**:2172–2181.

Wilson, M. S. and E. L. Madsen. 1996. Field extraction of a transient intermediary metabolite indicative of real time *in situ* naphthalene biodegradation. Environ. Sci. Technol. **30**:2099–2103.

Wilson, R. D., S. F. Thornton, and D. M. Mackay. 2004. Challenges in monitoring the natural attenuation of spatially variable plumes. Biodegradation **15**:359–369.

Winderl, C., S. Schaefer, and T. Lueders. 2007. Detection of anaerobic toluene and hydrocarbon degraders in contaminated aquifers using benzylsuccinate synthase (*bssA*) genes as a functional marker. Environ. Microbiol. **9**:1035–1046.

Witzig, R., H. Junca, H.-J. Hecht, and D. H. Pieper. 2006. Assessment of toluene/biphenyl dioxygenase gene diversity in benzene-polluted soils: Links between benzene biodegradation and genes similar to those encoding isopropylbenzene dioxygenases. Appl. Environ. Microbiol. **72**:3504–3514.

Yager, R. M., S. E. Bilotta, C. L. Mann, and E. L. Madsen. 1997. Metabolic adaptation and *in situ* attenuation of chlorinated ethenes by naturally occurring microorganisms in a fractured dolomite aquifer near Niagara Falls, New York. Environ. Sci. Technol. **31**:3138–3147.

Yagi, J. M. and E. L. Madsen. 2009. Diversity and spatial-temporal variation of microbial oxygenase abundance and expression in a shallow contaminated aquifer. Manuscript submitted.

Yagi, J. M., E. F. Neuhauser, J. A. Ripp, D. M. Mauro, , and E. L. Madsen. 2009a. Subsurface ecosystem resilience: Long-term attenuation of subsurface contaminants supports an elaborate eukaryotic food chain and a dynamic microbial community. Manuscript submitted.

Yagi, J. M., J. M. Suflita, L. M. Gieg, C. O. Jeon, and E. L. Madsen. 2009b. Subsurface microbial communities involved in *in situ* naphthalene catabolism carry out a complete N cycle via dissimilatory nitrate reduction to ammonia and denitrification. Manuscript submitted.

Yamaichi, Y. and H. Niki. 2000. Active segregation by the *Bacillus subtilis* partitioning system in *Escherichia coli*. Proc. Natl. Acad. Sci. USA **97**:14656–14661.

Yeates, C., A. Holmes, and M. Gillings. 2000. Novel forms of ring-hydroxylating dioxygenases are widespread in pristine and contaminated soils. Environ. Microbiol. **2**:644–653.

Yen, K. M. and C. M. Serdar. 1988. Genetics of naphthalene catabolism in pseudomonads. Crit. Rev. Microbiol. **15**:247–268.

Young, L. Y. and C. D. Phelps. 2005. Metabolic biomarkers for monitoring *in situ* anaerobic hydrocarbon degradation. Environ. Health. Persp. **113**:62–67.

Yura, T. and K. Nakahigashi. 1999. Regulation of the heat-shock response. Curr. Opin. Microbiol. **2**:153–158.

Zhou, N.-Y., S. L. Fuenmayor, and P. A. Williams. 2001. *nag* genes of *Ralstonia* (formerly *Pseudomonas*) sp. strain U2 encoding enzymes for gentisate catabolism. J. Bacteriol. **183**:700–708.

Zhulin, I. B., B. L. Taylor, and R. Dixon. 1997. Pas domain S-boxes in archaea, bacteria and sensors for oxygen and redox. Trends Biochem. Sci. **22**:331–333.

Zumft, W. 1997. Cell biology and molecular basis of denitrification. Microbiol. Mol. Biol. Rev. **61**:533–616.