UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

AN ASSESSMENT OF MICROBIAL COMMUNITIES AND THEIR POTENTIAL ACTIVITIES ASSOCIATED WITH OIL PRODUCING ENVIRONMENTS

A DISSERTATION

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

Degree of

DOCTOR OF PHILOSOPHY

By

HEATHER SUE NUNN
Norman, Oklahoma
2015

AN ASSESSMENT OF MICROBIAL COMMUNITIES AND THEIR POTENTIAL ACTIVITIES ASSOCIATED WITH OIL PRODUCING ENVIRONMENTS

A DISSERTATION APPROVED FOR THE
DEPARTMENT OF MICROBIOLOGY AND PLANT BIOLOGY

BY

_____
Dr. Bradley S. Stevenson, Chair

_____
Dr. Paul A. Lawson, Co-Chair

_____
Dr. Lee R. Krumholz

_____
Dr. Joseph M. Suflita

_____
Dr. Andrew S. Madden

Dedication

To my parents, Joe and Linda Drilling, for their unconditional love and never ending support.

Acknowledgements

I am grateful for the indispensible guidance and advice of my mentor, Bradley S. Stevenson, because it was essential to my development as a scientist. I would like to thank the rest of my graduate committee, Drs. Paul A. Lawson, Lee R. Krumholz, Joseph S. Suflita, and Andrew S. Madden, for their direction and insight in the classroom and the laboratory. To the members of the Stevenson lab past and present, Lauren Cameron, Dr. Michael Ukpong, Blake Stamps, Brian Bill, James Floyd, and Oderay Andrade, thank for your assistance, discussions, suggestions, humor and friendship. My time as a graduate student was made better because of all of you. I am thankful to my parents, Joe and Linda, my sister, Holly, and the rest of my family for all of their love and support they have always given me. Finally, I am thankful to my husband, Anthony, for being my constant source of love, friendship, encouragement and strength.

Table of Contents

List of Tables

List of Figures

Abstract

Microbial populations have been found in oil-associated environments as early as the 1920s. The proliferation and metabolic activities of these microorganisms can have profound deleterious effects on the infrastructure associated with oil reservoirs, production, transport and storage. Biodegradation of hydrocarbons by reservoir microorganisms can lead to the formation of 'heavy oil' that is of lower economic value and is more difficult to recover. Some members of reservoir microbial communities also participate in microbial influenced corrosion. By applying modern sequencing technologies, much can be learned about the microorganisms present and their metabolic capabilities. The focus of this dissertation was to provide a comprehensive characterization of microbial communities in two oil production facilities and define their metabolic activity by profiling metabolites of hydrocarbons and sequencing their metagenomes.

The most common samples available from oil production facilities are fluids collected at valve openings. These samples are chemically and biologically representative of the bulk fluids at any given location within an oil facility (e.g. pipelines). Microorganisms commonly attach to surfaces and form biofilms that can provide the microbial inhabitants protection from the external environment, allow for localized changes in chemistry, and represent sites of corrosion. Common maintenance of pipelines includes the use of "pigs" which physically disrupt and remove biofilms, corrosion products, and other solids associated with the inner surfaces of a pipeline. Libraries of partial 16S rRNA gene sequences were used to compare the microbial

communities in bulk fluids from several locations throughout an oil production facility with the community associated with a "pig envelope", the fluids enriched with solids removed by a pig. The microbial communities in bulk fluids and biofilms of the oil production facility contained only a few taxa. All samples had similar compositions, but different structure (relative abundances of taxa). An estimation of population density based on qPCR of 16S rRNA gene copy number showed that there was a five-fold increase in the number of bacteria in the pig envelope. The numerically abundant taxa were members of the genera *Thermoanaerobacter, Thermacetogenium* and *Thermovirga*, which should be studied further to determine their ability to degrade hydrocarbons and influence corrosion.

The community structure, genomic potential, and function of microbial assemblages from two oilfields under different management practices were characterized to measure their potential for hydrocarbon biodegradation. High throughput sequencing of 16S rRNA genes was combined with shotgun metagenomic sequencing and a targeted environmental metabolomics survey to interrogate two oil production facilities. The genomic potential for the abundant taxa was thoroughly interrogated for currently known pathways for hydrocarbon metabolism. Several sequences were identified that are closely related to known hydrocarbon degradation genes; however, there is no conclusive evidence that directly links these taxa and the hydrocarbon metabolites that were identified.

The presence of microorganisms and putative signature metabolites in oil-associated environments suggests hydrocarbon degradation is occurring. Hydrocarbon degradation causes souring and 'heavy oil' which is harder to extract and of less value.

Additionally, when microorganisms are identified in close association with corroded surfaces, they are potentially implicated as participating in surface corrosion. In order to directly associate a particular microorganism with a specific activity, there is still a need for controlled experiments. A better understanding of the microorganisms and their activities in oil production facilities will lead to improved monitoring and mitigation for the future.

Chapter 1

Summary

Microorganisms were first recognized to reside in oil reservoirs after they were cultivated from oilfield fluids (Bastin *et al.* 1926). These cultivated bacteria were anaerobic sulfate-reducers, producing hydrogen sulfide that reacted with iron salts in the medium to form iron sulfide. Bastin *et al.* cultivated other bacteria from some of the oilfield fluids that were not sulfate-reducing bacteria (SRB), but their nature and functions were not determined (Bastin *et al.* 1926). Since these initial studies, a wide range of microorganisms have been isolated from petroleum systems including: methanogenic archaea (Ollivier *et al.* 1998, Orphan *et al.* 2000), iron reducing bacteria (Semple and Westlake 1987), acid producing bacteria (Ferris *et al.* 1992), sulfide oxidizing bacteria (Voordouw *et al.* 1996), and aerobic bacteria (Nazina *et al.* 2001, Grabowski *et al.* 2005). The source of microbial diversity found in petroleum systems is still unclear; leading to the question of whether these microorganisms were indigenous to the oil reservoir or had they been introduced with drilling fluids, and water or gas injection. The origin of isolated microorganisms and 16S rRNA gene sequences found in oil production fluids must be interpreted carefully, with the knowledge that the organisms may not be indigenous to the oil reservoir (Magot *et al.* 2000, Head *et al.* 2003, Youssef *et al.* 2009).

Microorganisms, regardless of their origin, are able to exploit new habitat that is exposed through drilling activities and exchange of fluids with the surface that occurs during oil exploration and production. As oil production activities continue at a

location, vast amounts of water and gas are (re)injected into the reservoir to maintain pressure within the formation. These fluids/gases transport electron acceptors and other nutrients to the reservoir, bring hydrocarbons to the surface, and present new combinations of environmental parameters (temperature, pH, salinity, pressure) that may support the growth of microorganisms. These selective forces and the resultant microbial ecology dictate the succession of microbial populations, metabolic activities, and their consequences. Microbial metabolism can preferentially remove lighter chain hydrocarbons from crude oil, leaving only the heavier, more viscous components. By-products of metabolism can include organic acids and sulfides. These products can cause corrosion and sour the oil being recovered, ultimately increasing the difficulty and expense of oil production.

Initially it was assumed that biodegradation of hydrocarbons occurred through aerobic microbial metabolism with oxygen dissolved in meteoric water delivered to the deep subsurface (Palmer 1993). Bacteria capable of anaerobic hydrocarbon degradation, however, were isolated soon thereafter (Rueter *et al.* 1994). Many, but not all, of the mechanisms of hydrocarbon degradation have been determined in pure cultures or consortia enriched from the environment (Widdel and Rabus 2001, Heider 2007, Fuchs *et al.* 2011). Little is known about the ecology of oil production facility ecosystems from a community standpoint, specifically which members of microbial populations are responsible for certain activities such as hydrocarbon degradation. Knowing which microbial groups are present, how they change with location and over time, and how they respond to oil facility management practices will help us to better understand the ecological forces that shape these microbial assemblages. In turn, we posit that this

knowledge will lead to better prevention and monitoring of microbial populations and their detrimental activities. Better resolution of the relevant microbes and their physiologies will also ultimately lead to more directed and effective mitigation strategies.

The research in this dissertation began with an investigation of the structure and function of microbial communities associated with oil production facilities on the North Slope of Alaska. This was a unique sampling opportunity because access to oil production facilities is highly restricted and they are often located in remote locations. A multi-faceted approach to investigate the microbial assemblages and their metabolic capabilities is outlined in Figure 1. A survey of the bacterial assemblages from oil production facilities was conducted to determine how the populations changed across samples from different locations. The composition and structure of these microbial assemblages was determined by sequencing 16S rRNA gene libraries, which were compared to presumed patterns of hydrocarbon metabolism revealed through an assessment of the metabolite profiles. Finally, metagenomic sequencing of selected samples was used in an attempt to link specific microbial taxa with hydrocarbon biodegradation. The integration of

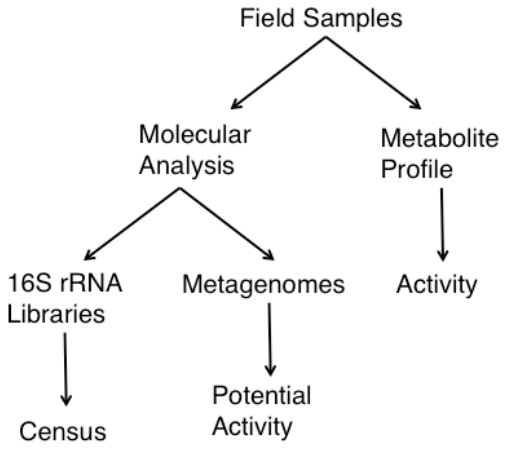


Figure 1. Diagram of the approach taken to investigate field samples to identify the structure of microbial assemblages and their activities in the environment.

these techniques allows for the correlations between community structure and function, and observations in the field, such as fouling or corrosion.

In Chapter 2, the microbial assemblage in produced water and a sample from the 'envelope' of a pipeline inspection gauge (PIG) were compared to correlate specific populations or community structures associated with biocorrosion. The goal of this study was to determine which microorganisms were present and how the communities changed based on the sample type and location. The hypothesis was that the community structure of samples taken from bulk fluids would be different from the 'PIG envelope', and the particular taxa found in the 'PIG envelope' represented those participating in corrosion activities due of their close association with the pipeline surface.

Molecular methods were used to create a census of the microbial communities from multiple locations within an Alaskan North Slope oil production facility, including a primary separator, a seawater line, a seawater/produced water line, an injection well and the 'PIG envelope' sample from a produced water line. The 'PIG' is a tool used for physical mitigation of pipeline corrosion and fouling by scraping the inside pipeline surface. The 'envelope' sample, therefore, represents an enrichment of the solids associated with the pipeline surface, which include paraffins, minerals (corrosion products), biofilms, and microorganisms.

Bacteria were found to be numerically dominant compared to archaea throughout the oil production facility. Thermophilic members of the phyla Firmicutes (*Thermoanaerobacter* and *Thermacetogenium*) and Synergistes (*Thermovirga*) were the most abundant in samples across the field. The *Thermoanaerobacter (Fardeau et al. 1993)*, *Thermacetogenium (Hattori et al. 2000),* and *Thermovirga* (Dahle and Birkeland

2006) are all thermophilic anaerobes that can produce sulfides. These organisms were enriched in material associated with the walls of the pipeline where corrosion is occurring; this direct contact with the corroded surface and ability to produce corrosive sulfides implicates these organisms in microbially influenced corrosion (MIC). The microbial communities of bulk fluid and biofilm samples were similar in composition, containing only a few abundant taxa that were also shared across all the samples. The similarity across the samples was concluded to be due to the homogenizing effect of 30+ years of oil production and fluid recycling that takes place at the facility. The structure of the PIG community, however, was distinct from the bulk fluids due to the increased relative abundance of the genera *Thermacetogenium* and *Thermovirga*. This suggested that these two genera were major components of the biofilm community associated with the surface of the pipeline, implicating them as potential links to MIC.

My contribution in Chapter 2 was to develop a method to extract DNA from filtered bulk fluid, carry out the DNA extractions, amplify the 16S rRNA gene, prepare the 16S rRNA gene libraries for sequencing, analyze the sequence data and generate the community analysis figures. The manuscript was published in Environmental Microbiology (Stevenson *et al.* 2011).

In Chapter 3, a comparative investigation of two North Slope oil production facilities was undertaken by characterizing the resident microbial communities, associated metagenomes and the presence of putative hydrocarbon metabolites in production fluids. The goal was to establish an ecological perspective in the production facilities in order to link the structure of the microbial community to their functions in the facility environment. The two facilities varied in management practices. In facility

A, seawater was injected to maintain reservoir pressure and the temperature of sampled fluids were mesophilic overall (16-35 °C). Facility B produced more water than oil, recycled the produced waters, and had higher temperatures than Field A overall (35-60 °C). The management practices and temperature were found to play an important role in the microbial community structure.

The microbial assemblages in samples from facility A contained only a few abundant taxa, which were mesophilic, facultatively-anaerobic members of the Proteobacteria. Each sample contained either Deltaproteobacteria or Gammaproteobacteria as the most abundant taxa, however, the identity of these taxa differed across samples. Strictly anaerobic, thermophilic members of the Synergistes, and Firmicutes, specifically *Thermovirga, Thermacetogenium,* and *Thermoanaerobacter*, were the most abundant taxa in facility B. Unlike the microbial assemblages at facility A, the same taxa were present and abundant across samples at facility B. It was concluded that this was likely an effect of the practice of recycling production fluids over a 30-year period.

Almost all of the samples contained metabolites indicative of aerobic and anaerobic degradation of hydrocarbons. Samples from facility A contained fewer different metabolites than those from facility B, which could be due to the composition of the crude oil, the microorganisms present in the formation and facility, and the metabolic activity. More aerobic hydrocarbon degradation genes were identified in facility A, potentially because of dissolved oxygen carried into the system through seawater. Conclusive evidence does not exist, however, for aerobic hydrocarbon degradation in either facility. The genomic evidence for anaerobic hydrocarbon

degradation is somewhat limited. Very few genes associated with anaerobic hydrocarbon degradation have been identified and described in any system, making it difficult to collect genomic evidence for this metabolism among the abundant taxa in either facility. Sequences from the metagenomic sequence associated with the genomes of abundant taxa from each facility were binned. One of these sequences from the *Thermovirga* binned genome, however, contained sequence annotated as a putative glycyl radical enzyme that is closely associated with known alkyl succinate synthases. Additional studies to elucidate the anaerobic hydrocarbon biodegradation pathways and a direct measure of gene expression, instead of genomic potential, are needed to help clarify this activity for the *Thermovirga* and the metabolic activates found in other taxa in oil systems.

The majority of my dissertation research is represented in Chapter 3. The samples collected from these oil facilities represent a comprehensive set of samples collected from active oil fields, which was a rare opportunity to directly investigate and compare the microbial ecology of two oil production facilities. I generated and analyzed the 16S rRNA community profiles, participated in compiling a database of associated hydrocarbon degradation genes for metagenomic analysis, and synthesized the link between community, metabolite, and metagenomic profiles. This chapter is also written in the style of the journal Environmental Microbiology.

This dissertation sets the foundation for future investigations to gain further understanding of the role of microbial populations in oil producing environments. The multifaceted approach used here has led to correlations between the microorganisms and their activities, but direct connections between microorganisms, the enzymes they

produce and their activities are still needed. The data generated here suggests that *Thermovirga*, *Thermacetogenium,* and *Thermoanaerobacter* should be investigated further to explore their capacity to degrade hydrocarbon compounds under these conditions. Experiments to determine genetic expression, metabolic profiles, and corrosion studies are desperately needed to enhance what is known about these taxa and their activities in oil production facilities. Additionally, the metagenomic sequence data generated for this dissertation could be used in the future to gain more understanding of the microbial activities in oilfields as more genes associated with hydrocarbon degradation are elucidated. This type of multifaceted research will continue to enable integration of interactions of microbes with each other and their environment, which will lead to better detection, and mitigation of adverse microbial activities in oil production facilities.

References

Bastin, E. S., F. E. Greer, C. A. Merritt and G. Moulton (1926). The presence of sulphate reducing bacteria in oil field waters. *Science* **63**: 21-24.

Dahle, H. and N. K. Birkeland (2006). *Thermovirga lienii* gen. nov., sp nov., a novel moderately thermophilic, anaerobic, amino-acid-degrading bacterium isolated from a North Sea oil well. *Int J Syst Evol Micr* **56**: 1539-1545.

Fardeau, M. L., J. L. Cayol, M. Magot and B. Ollivier (1993). H-2 Oxidation in the Presence of Thiosulfate, by a *Thermoanaerobacter* Strain Isolated from an Oil-Producing Well. *FEMS Microbiol Lett* **113**(3): 327-332.

Ferris, F. G., T. R. Jack and B. J. Bramhill (1992). Corrosion products associated with attached bacteria at an oil field water injection plant. *Canadian Journal of Microbiology* **38**(12): 1320-1324.

Fuchs, G., M. Boll and J. Heider (2011). Microbial degradation of aromatic compounds - from one strategy to four. *Nat Rev Microbiol* **9**(11): 803-816.

Grabowski, A., O. Nercessian, F. Fayolle, D. Blanchet and C. Jeanthon (2005). Microbial diversity in production waters of a low-temperature biodegraded oil reservoir. *FEMS Microbiol Ecol* **54**(3): 427-443.

Hattori, S., Y. Kamagata, S. Hanada and H. Shoun (2000). *Thermacetogenium phaeum* gen. nov., sp. nov., a strictly anaerobic, thermophilic, syntrophic acetate-oxidizing bacterium. *Int J Syst Evol Micr* **50 Pt 4**: 1601-1609.

Head, I. M., D. M. Jones and S. R. Larter (2003). Biological activity in the deep subsurface and the origin of heavy oil. *Nature* **426**(6964): 344-352.

Heider, J. (2007). Adding handles to unhandy substrates: anaerobic hydrocarbon activation mechanisms. *Curr Opin Chem Biol* **11**(2): 188-194.

Magot, M., B. Ollivier and B. K. C. Patel (2000). Microbiology of petroleum reservoirs. *Anton Leeuw Int J G* **77**(2): 103-116.

Nazina, T. N., T. P. Tourova, A. B. Poltaraus, E. V. Novikova, A. A. Grigoryan, A. E. Ivanova, . . . M. V. Ivanov (2001). Taxonomic study of aerobic thermophilic bacilli: descriptions of *Geobacillus subterraneus* gen. nov., sp nov and *Geobacillus uzenensis* sp nov from petroleum reservoirs and transfer of *Bacillus stearothermophilus, Bacillus thermocatenulatus, Bacillus thermoleovorans, Bacillus kaustophilus, Bacillus thermoglucosidasius* and *Bacillus thermodenitrificans* to *Geobacillus* as the new combinations *G. stearothermophilus, G. thermocatenulatus, G. thermoleovorans, G. kaustophilus, G. thermoglucosidasius* and *G. thermodenitrificans*. *Int J Syst Evol Micr* **51**: 433-446.

Ollivier, B., M. L. Fardeau, J. L. Cayol, M. Magot, B. K. C. Patel, G. Prensier and J. L. Garcia (1998). *Methanocalculus halotolerans* gen. nov., sp. nov., isolated from an oil-producing well. *Int J Syst Bacteriol* **48**: 821-828.

Orphan, V. J., L. T. Taylor, D. Hafenbradl and E. F. Delong (2000). Culture-Dependent and Culture-Independent Characterization of Microbial Assemblages Associated with High-Temperature Petroleum Reservoirs. *Appl Environ Microb* **66**(2): 700-711.

Palmer, S. E. (1993). Effects of biodegradation and water washing on crude oil composition. *Organic Geochemistry*. S. M. MH Engel. New York, Plenum Press**:** 511-533.

Rueter, P., R. Rabus, H. Wilkes, F. Aeckersberg, F. A. Rainey, H. W. Jannasch and F. Widdel (1994). Anaerobic oxidation of hydrocarbons in crude oil by new types of sulphate-reducing bacteria. *Nature* **372**(6505): 455-458.

Semple, K. M. and D. W. S. Westlake (1987). Characterization of Iron-Reducing *Alteromonas putrefaciens* Strains from Oil-Field Fluids. *Canadian Journal of Microbiology* **33**(5): 366-371.

Stevenson, B. S., H. S. Drilling, P. A. Lawson, K. E. Duncan, V. A. Parisi and J. M. Suflita (2011). Microbial communities in bulk fluids and biofilms of an oil facility have similar composition but different structure. *Environ Microbiol* **13**(4): 13.

Voordouw, G., S. M. Armstrong, M. F. Reimer, B. Fouts, A. J. Telang, Y. Shen and D. Gevertz (1996). Characterization of 16S rRNA genes from oil field microbial communities indicates the presence of a variety of sulfate-reducing, fermentative, and sulfide-oxidizing bacteria. *Appl Environ Microb* **62**(5): 1623-1629.

Widdel, F. and R. Rabus (2001). Anaerobic biodegradation of saturated and aromatic hydrocarbons. *Curr Opin Biotechnol* **12**(3): 259-276.

Youssef, N., M. S. Elshahed and M. J. McInerney (2009). Microbial processes in oil fields: culprits, problems, and opportunities. *Adv Appl Microbiol* **66**: 141-251.

Chapter 2[1]

Microbial Communities in Bulk Fluids and Biofilms of an Oil Facility Have Similar Composition but Different Structure

Abstract

The oil–water–gas environments of oil production facilities harbor abundant and diverse microbial communities that can participate in deleterious processes such as biocorrosion. Several molecular methods, including pyrosequencing of 16S rRNA libraries, were used to characterize the microbial communities from an oil production facility on the Alaskan North Slope. The communities in produced water and a sample from a 'pig envelope' were compared in order to identify specific populations or communities associated with biocorrosion. The 'pigs' are used for physical mitigation of pipeline corrosion and fouling and the samples are enriched in surface-associated solids (i.e. paraffins, minerals and biofilm) and coincidentally, microorganisms (over 10^5-fold). Throughout the oil production facility, bacteria were more abundant (10- to 150-fold) than archaea, with thermophilic members of the phyla *Firmicutes* (*Thermoanaerobacter* and *Thermacetogenium*) and *Synergistes* (*Thermovirga*) dominating the community. However, the structure (relative abundances of taxa) of the microbial community in the pig envelope was distinct due to the increased relative abundances of the genera *Thermacetogenium* and *Thermovirga*. The data presented here suggest that bulk fluid is representative of the biofilm communities associated with

[1] Stevenson et al,. 2011. Environ Microbiol. **13**(4), 1078-1090

biocorrosion but that certain populations are more abundant in biofilms, which should be the focus of monitoring and mitigation strategies.

Introduction

Petroleum reservoirs and the oil–water–gas environments of production facilities contain abundant and diverse microbial communities. Cultivation-based analyses of these anoxic ecosystems have revealed a wide variety of microbial phenotypes including fermentative organisms, manganese and iron reducers, acetogens, sulfate reducers, methanogens, aerobes and nitrate reducers (reviewed in Magot *et al*., 2000). Molecular-based surveys of microbial diversity and composition offer a more thorough characterization of microbial communities of petroleum reservoirs without the need for *a priori* cultivation. A number of microbial taxa are common among petroleum reservoirs and production facilities, suggesting common evolutionary ecologies. However, differences in the composition (taxa present) and structure (relative abundance of taxa present) of these communities can be correlated with reservoir geochemical and physical properties (e.g. temperature, salinity, pH, sulfate concentration) or production practices of the facility (e.g. water flooding, nitrate amendment) (Grabowski *et al*., 2005; Dahle *et al*., 2008; Duncan *et al*., 2009; Pham *et al*., 2009; van der Kraan *et al*., 2010).

The activities of resident microbial populations can have a negative impact on oil quality and the infrastructure of production facilities. Biodegradation of hydrocarbons by reservoir microorganisms can lead to the formation of ‘heavy oil’ that

is of lower economic value and is more difficult to recover (Head *et al*., 2003). Some members of reservoir microbial communities participate in biocorrosion, which poses a significant risk to oil production and the environment (Duncan *et al*., 2009). Standard industry mitigation strategies include biocide treatments and physical removal of accumulated material from the pipeline. Such approaches represent considerable cost or disruption in production, and their effectiveness is limited. The mechanisms of biocorrosion are not well understood (Little *et al*., 2006). It is not clear, therefore, which microbial populations participate in biocorrosion or what ecological forces promote or deter this activity (Hamilton, 2003; Beech and Sunner, 2004; Beech *et al*., 2005). As a result, efforts to monitor, predict and mediate biocorrosion have had limited success.

Biofilms play an important role in biocorrosion activity (Beech and Sunner, 2004; Beech *et al*., 2005) and therefore, samples of surface-associated microbial communities should contain populations that play a direct role in this process. There are few opportunities, however, to sample biofilm communities from a pipeline in operation. Metal coupons are oftentimes inserted into the flow path of a pipeline to monitor the rate of corrosion as the value of mass loss over time. More rarely, small parts of a pipeline can be sampled, with the resultant ‘cookies’ (circular cores) analyzed. The representative nature of these tools can certainly be questioned, as these methods require a breach in pipeline integrity.

Another means of gaining access to surface-associated biofilm material is to collect the envelope of solids produced by pipeline cleaning ‘pigs’. Pigs are devices originally developed by the hydrocarbon industries that are inserted into pipelines and driven by the flow of fluids (Quarini and Shire, 2007). Cleaning or maintenance pigs are

designed to remove biofilm, mineral and paraffin deposits mechanically. 'Intelligent' inspection pigs can detect dents or changes in internal diameter of a pipeline, can be tracked by global positioning systems, and can measure pipeline integrity via technologies such as magnetic flux detection, ultrasonic transducers or visualization (Goedecke, 2003; Clark, 2005).

This research began with an opportunity to conduct sampling throughout an oil production facility on Alaska's North Slope (ANS) in January 2008. The environmental conditions of ANS fields have been reported previously (Masterson *et al*., 2001) and were shown to harbor a thriving microbial community linked to anaerobic hydrocarbon biodegradation and biocorrosion (Duncan *et al*., 2009; Pham *et al*., 2009). North Slope facilities have been in production for 30+ years, producing oil from several reservoirs ranging in temperature from 27°C to over 70°C. The samples collected for the work described here included produced water from processing facilities (54°C, 'SEP') and a water injection well (44°C, 'INJ'), pipelines carrying treated seawater (10°C, 'SEA') and a mixture of produced water and seawater (54°C, 'SPW'), and a sample from the 'envelope' produced when a pig was used as a physical means to remove solids from a pipe- line transporting produced water from the processing facility to injection wells (50°C, 'PIG') (Figure 1).

The overall goal of this effort was to compare the microbial communities in produced water throughout the oil production facility with those in pipeline-associated biofilms in order to identify specific populations or community structures associated with biocorrosion. Therefore, the pig envelope sample was of great interest. Comparisons between microbial communities of bulk fluids and pipeline-associated

biofilms should provide insight into the conditions driving detrimental microbial activities like hydrocarbon biodegradation and biocorrosion.

Experimental procedures

Description of oil production facility

The oil, gas and water produced at the facility come from a hot anaerobic (average temperature 68°C) reservoir. The produced fluids and gas are processed and recovered oil is sent to the oil export line (Figure 1). Production water (combined with biocide-treated seawater as needed) and low-molecular-weight hydrocarbons (methane and C_2–C_4 *n*-alkanes) are separated and re-injected into the reservoir to maintain the pressure needed to recover additional oil. Rates of biocorrosion are reported to be highest in the processing facility and at the injection wells (personal communication from operators). Several samples of fluids were taken from a processing facility and include a separation stage (SEP), biocide-treated seawater (SEA) and a mixture of produced water and sea- water (SPW). Produced water with no seawater amendment was also sampled from an injection well (INJ). Lastly, a sample was taken from the envelope of a pig (PIG) that was sent between the processing facility and manifold building in a pipeline that delivers produced water (never amended with seawater) from the central facility to an injection well. The pig sample was enriched in solids formerly associated with the inner pipeline surface.

Sample collection

Each sample was collected through access valves into sterile bottles with no headspace. Biomass was collected in the field within 4 h of sampling by vacuum onto a 47 mm cellulose nitrate filter (Thermo Fisher Scientific, Rochester, NY) with a 0.45 mm nominal pore size until negligible flow rates were achieved, which was generally between 150 and 300 ml. Filters were placed in sterile 50 ml screw cap tubes and preserved in the field by the addition of DNAzol® Direct (Molecular Research Center, Cincinnati, OH). Upon arrival in the laboratory, samples were stored at -80°C until cell lysis and DNA extraction were conducted. The PIG sample was collected into sterilized bottles, immediately capped, then returned to the laboratory facility where the headspace was flushed repeatedly with sterile N_2 gas and stored either at or below 20°C. Upon receipt at the University of Oklahoma (approximately 14 days), 15 ml aliquots of the solid–fluid mixture were aseptically transferred to 50 ml Falcon tubes, then frozen at -80°C until DNA extraction as described below.

Sulfate-reducing incubations

The number of sulfate-reducing microorganisms in the PIG envelope was estimated using standard industry protocols [NACE Standard TM0194-2004 (NACE, 2004)]. Six serum bottles containing a modified version of Postgate's Medium B for Sulfate Reducers (C&S Laboratories, Broken Arrow, OK) were used to generate a 10-fold dilution series (10^{-1} to 10^{-6}) of the PIG sample. This proprietary medium contains sodium lactate, sodium acetate and yeast extract as carbon sources, with 2% total salinity. Two series of bottles were inoculated in parallel, incubated at 40°C or 75°C for

14 days, and then shipped at ambient temperatures to the University of Oklahoma for subsequent molecular analysis.

Community analysis of sulfate-reducing incubations

The microbial communities in the sulfate-reducing enrichment cultures described above were analysed using DGGE of amplified bacterial and archaeal 16S rRNA genes. The PCR products for DGGE analysis were produced with the GM5F and 907R primers (approximately 550 bp bacterial 16S rRNA gene sequence) and a touch-down PCR amplification protocol (described in Muyzer *et al*., 1998; Santegoeds *et al*., 1999) or with the Arc333F-GC and 958R primers (described in Struchtemeyer *et al*., 2005) (approximately 600 bp archaeal 16S gene sequence). The DNA extracted from the pig envelope and the two series of enrichments was used directly in bacterial and archaeal DGGE amplification reactions. A nested PCR procedure was used to generate DGGE products from clone library cell suspensions. The first PCR reaction used the vector-specific M13F and M13R primers with reaction conditions described above, followed by amplification using 1 ml of this PCR product as template DNA in a second reaction with the bacterial DGGE primers and amplification conditions. The DGGE analysis was performed using the DCode Universal Mutation Detection System (Bio-Rad Laboratories, Hercules, CA) with a 6% polyacrylamide gel, run at a temperature of 60°C and a constant voltage of 65 V for 16 h. Bacterial and archaeal DGGE gels had a denaturant gradient of urea and formamide from 40% to 80% and 25% to 100% respectively.

DNA extraction

For samples SEPa, SPWa, INJ and SEA, sterile glass beads (2.0 g of 0.5 mm diameter and 3.5 g of 0.1 mm diameter), 10 ml of buffer containing 10 mM Tris-Cl (pH 8.0), 1 mM ethylenediaminetetraacetic acid (EDTA), 0.05% (w/v) Nonidet P40 (Roche Diagnostics GmbH, Mannheim, Germany) and 10 ml of phenol:chloroform:isoamyl alcohol (P:Cl; 25:24:1, pH 8.0) were added to each 50 ml Falcon tube containing a filter. Cell lysis was achieved by agitating each tube horizon- tally on a vortexer (Vortex Genie II, Scientific Industries, Bohemia, NY) for 5 min at maximum speed. For the PIGa sample, cell biomass and solids from 100 ml were first collected by centrifugation at 5000 *g* for 20 min at 4°C and then subjected to lysis as described above. After cell lysis and centrifugation (5000 *g* for 5 min at 4°C), the resulting aqueous phase was extracted sequentially with equal volumes of P:Cl until clear, followed by two extractions with equal volumes of chloroform:isoamyl alcohol (24:1). Glycogen (Fermentas, Glen Bernie, MD) was added as a carrier molecule to the DNA-containing solution at a final concentration of 0.1 mg ml^{-1}, and nucleic acids were precipitated with ethanol as described in Sambrook (2001). Precipitated nucleic acids were resuspended in up to 100 ml of DNA, RNA and nuclease-free water and quantified by UV spectroscopy at 260 nm (NanoPhotometer; Implen, Westlake Village, CA).

The PIGb sample consisted of a 15 ml aliquot of the solid– fluid mixture in a 50 ml Falcon tube frozen at -80°C. DNA was extracted using a commercial kit (PowerSoil Mega DNA Isolation Kit, MOBIO Laboratories, Carlsbad, CA). The manufacturer's instructions were modified with the addition of Lysing Matrix A beads (MP Biomedical, Solon, OH), 5 min bead beating using a vortex adapter (MOBIO

Laboratories, Carlsbad, CA), and a rapid freeze–thaw (15 min at -80°C followed by 5 min in a 65°C waterbath) initially in order to facilitate cell lysis.

DNA was extracted from filter samples SEPb and SPWb using the PowerSoil® DNA Isolation kit (MoBio Laboratories, Carlsbad, CA) with modifications. Filters and liquid were transferred to a Lysing Matrix E tube (MP Biomedical) to which 50 ml of PowerSoil® bead solution, 100 ml of solution C1 (from PowerSoil® kit) and 250 ml of PCR-grade water were added followed by bead beating using a Mini-Beadbeater-1® (BioSpec Products, Bartlesville, OK) at 25 000 r.p.m. for 90 s. Following centrifugation at 10 000 *g* for 30 s, supernatants were divided between two microcentrifuge tubes and the remaining extraction followed the manufacturer's protocols. DNA solutions were kept at -20°C until needed.

Cell pellets were obtained from sulfate-reducing enrichment cultures by centrifugation at 4330 *g* at room temperature. These cell pellets were washed twice and resuspended with sterile 0.85% NaCl, and stored at -20°C. DNA was extracted using the PowerSoil® DNA Isolation Kit following the manufacturer's protocol (MoBio Laboratories).

PCR amplification, cloning and sequencing of 16S rRNA genes

Nearly full-length bacterial 16S rRNA gene fragments were amplified from DNA extracted from the 10^{-6} sulfate-reducing cultures incubated at 40°C and 75°C (primers 27F and 1492R, Table S1) and the PIGa sample (primers 27F and 1392R, Table S1). PCR conditions used for the sulfate-reducing culture DNAs were as specified in Herrick and colleagues (1993). The PCR using PIGa DNA consisted of 1×

Taq buffer with $(NH_4)_2SO_4$ (Fermentas), 1.5 mM $MgCl_2$, 0.2mM each dNTP, 20mM of the forward and reverse primer, 0.625 U of *Taq* DNA Polymerase (Fermentas), and between 10 and 50 ng of DNA in a final volume of 25 ml. Thermal cycling was carried out in a Techne TC-512 thermal cycler (Techne, Burlington, NJ) using the following conditions: initial denaturation for 3 min at 95°C; 30 cycles of 20 s at 95°C, 20 s at 52°C and 1 min at 65°C; and a final extension of 5 min at 65°C. Archaeal 16S rRNA gene fragments were amplified from DNA extracted from the 10^{-1} and 10^{-3} sulfate reducer enrichments at 75°C and the PIGb sample using the primers Arc333F and 958R (Table S1) according to the protocols described by Gieg and colleagues (2008).

Bacterial and archaeal 16S rRNA clone libraries were created using the TOPO TA Cloning Kit for Sequencing (Invitrogen Corp., Carlsbad, CA). Transformants were transferred into a 96-well microtitre plate containing Luria–Bertani broth containing 10% (v/v) glycerol and 50 mg ml^{-1} ampicillin, grown overnight at 37°C and stored at -80°C until needed. Plasmid DNA was purified from the transformed cells and sequencing performed on an ABI model 3730 capillary sequencer using the M13 flanking regions as sequencing primer sites.

Analysis of clone sequencing data

Cloned 16S rRNA sequences were aligned using the greengenes NAST-aligner (DeSantis *et al*., 2006a) and screened for chimeras using Bellerophon (Huber *et al*., 2004). Distance matrices were calculated for clone sequence libraries using tools available at the greengenes website [http://greengenes. lbl.gov (DeSantis *et al*., 2006b)] and used to cluster sequences into OTUs based on a 3% dissimilarity cut-off using the

program DOTUR (Schloss and Handelsman, 2005). Representative sequences for each OTU were assigned a phylogenetically consistent taxonomy based on a naïve Bayesian rRNA classifier available through the Ribosomal Database Project (Wang *et al*., 2007). The percent sequence identity of representative sequences to sequences for isolated organisms in the RDP Release 10.20 was determined using the SeqMatch tool. The SSU rRNA clone sequences have been deposited in GenBank under Accession No. HM994865–HM994881.

Pyrosequencing of bacterial 16S rRNA gene libraries

A region of the bacterial 16S rRNA gene (*Escherichia coli* positions 27–338) was PCR-amplified from each sample DNA solution. The primers (TiA-8nt-27F and TiB-338R) were modeled after those from Hamady and colleagues (2008) and contained the Titanium Fusion A or B primer sequence (454 Life Sciences, Branford, CT) followed by an 8 nt unique barcode (forward primer only, listed in TableS2), and sequence from the general bacterial primers 27F and 338R (Table S1). Each PCR reaction was conducted as described above. Each sample was amplified in four replicate PCRs, pooled and purified with the Wizard® PCR Preps DNA Purification System as directed by the manufacturer (Promega, Madison, WI). Pooled, purified samples were then quantified by UV spectroscopy at 260 nm (NanoPhotometer; Implen) and comparison with DNA markers following agarose gel electrophoresis. Equimolar amounts of each uniquely bar- coded PCR product were combined and sequenced using the Genome Sequencer FLX instrument with the GS FLX Titanium series reagents (454 Life Sciences) at the Advanced Center for Genome Technology at the University of

Oklahoma. All pyrosequences have been deposited in the short read archive of GenBank under Accession No. SRA023443.2. Individual libraries are under Accession No. SRS114954.2–SRS114959.2.

Analysis of pyrosequencing data

Sequences from pyrosequencing of bacterial 16S rRNA gene libraries were analysed using mothur, an open-source software package for describing and comparing microbial communities (Schloss *et al*., 2009). All sequence reads were screened to remove those that contain any errors in the forward primer or barcode regions, ambiguities, homopolymers greater than six nucleotides in length, or an average quality score <25. The reverse complement of each sequence was then binned according to its unique barcode and unique sequences were aligned against the SILVA reference database (Pruesse *et al*., 2007) using the NAST aligner in the mothur software package (Schloss, 2009). Sequences within one nucleotide of one another were pre-clustered to reduce the number of sequences that may be unique simply due to a sequencing error. A distance matrix was generated, and used to cluster sequences in OTUs with sequence dissimilarity cut-offs of 3%, 5% and 10%. Each OTU at 3% was assigned a phylogenetically consistent taxonomy based on a naïve Bayesian rRNA classifier available through the Ribosomal Database Project (Wang *et al*., 2007). The relationship between libraries based on community structure of the PIGa, SEPa, SEPb, SPWa, SPWb, INJ and SEA libraries was determined based on determination of the Yue and Clayton measure of similarity Θ_{YC} (Yue and Clayton, 2005). The libshuff method was used to indicate the probability (a = 0.05) that communities have the same structure by

chance, applying Bonferroni's correction for multiple comparisons ($n = 7$) so that significance required $P < 0.007$. The abundance- based Coverage Estimator of species richness [ACE (Chao *et al*., 2000)], Chao1 richness estimator (Chao, 1984) and the Shannon diversity index (Magurran, 2004) were all calculated using the mothur software package on a subset of the libraries PIGa, SEPa, SPWa, INJ and SEA that was randomly sampled to the depth of the smallest library ($n = 629$ reads in SEA).

Enumeration of bacteria and archaea using qPCR

The number of bacterial and archaeal 16S rRNA gene copies in each sample was estimated using qPCR. SYBR Green PCRs contained 0.75× Power SYBR Green PCR master Mix (Applied Biosystems/Life Technologies Corp., Carlsbad, CA), 250 nM Eub27F and 125 nM Eub338R for bacteria, 500 nM Arc8F and 1 mM Arc344R for archaea and 2 ml of template DNA in a total volume of 42 ml. Real-time thermal cycling was performed in an Applied Biosystems 7300 Real time PCR System (Life Technologies Corp., Carlsbad, CA) using the thermal profile of 95°C for 10 min, followed by 40 cycles of 92°C for 30 s, 55°C for 30 s and 72°C for 30 s for bacteria. For archaea, the thermal profile was 95°C for 10 min, followed by 40 cycles of 94°C for 30 s, 55°C for 45 s and 72°C for 45 s. For each qPCR, a dilution series of control DNA was run in duplicate with triplicate reactions of unknown samples. Data acquisition and analyses were performed using the 7300 System SDS software.

Results

Sulfate-reducing incubations

Blackening of the culture medium was observed in all dilutions of the sulfate-reducing incubations after 14 days, although the darkening of the cultures incubated at 40°C preceded those incubated at 75°C. A comparison of bacterial denaturing gradient gel electrophoresis (DGGE) profiles amplified from the sulfate-reducing enrichments and the PIG sample showed that different incubation temperatures resulted in community shifts (Figure 2). The cloned sequence of an operational taxonomic unit (OTU) (e.g. C440-6, E475-6, E575-6, E675-6) that consistently appears in all libraries of sulfate-reducing incubations and the PIG sample is > 99% similar to the 16S rRNA gene sequence of *Thermoanaerobacter pseudethanolicus* [CP000924 (Rainey *et al*., 1993)]. These sequences were also detected in samples from the same oil production facility collected in 2006 (Duncan *et al*., 2009). Additional bacterial clone sequences from the 10^{-6} sulfate-reducing culture incubated at 40°C were 99.6–100% similar to that of the *Deltaproteobacterium Desulfomicrobium thermophilum* (Thevenieau *etal*., 2007) (sequences C240-6, C540-6, C840-6) or 99.7–99.8% similar to the *Synergistetes Thermovirga lienii* (Dahle and Birkeland, 2006) (sequences C140-6 and C640-6), isolated from a thermophilic oil reservoir. These sequences were also detected in the 2006 ANS samples (Duncan *et al*., 2009).

Archaeal DGGE community profiles from the sulfate-reducing cultures incubated at 40°C showed a single dominant band in the same relative position throughout (data not shown). Cloned PCR products amplified with archaeal DGGE

primers from DNA extracted from the 10^{-1}, 10^{-3} and 10^{-6} incubations at 40°C were found to be 99.7–100% similar to *Methanothermobacter thermautotrophicus* strain delta H (Wright and Pimm, 2003). This is also the dominant archaeal OTU (A1Fa10) obtained directly from the PIG envelope clone library (described below). In contrast, three cloned sequences from the archaeal clone library from the 10^{-1} and 10^{-3} cultures incubated at 75°C (AF375-1, AF475-1 and AF575-3) were 99% similar to *Thermococcus alcaliphilus* (Keller *et al*., 1995). Similar sequences were abundant in samples taken in 2006 from different locations within the same oil production facility (Duncan *et al*., 2009). No amplification was obtained with the primers Arc333GC and 958R using DNA extracted from the 10^{-4}, 10^{-5} and 10^{-6} bottles in the dilution series incubated at 75°C.

Clone sequence libraries of the PIG sample

The diversity of bacteria far exceeded that of the archaea in the PIG sample 16S rRNA clone library (Figure 3A and B). At a dissimilarity cut-off of 3%, the bacterial 16S rRNA gene sequences from the pig envelope (n = 87) formed 22 OTUs, whereas only two OTUs were detected from archaeal 16S rRNA gene sequences (n = 84).

For the bacterial clone library of the PIG sample (Figure 3A), the *Firmicutes* (*Thermacetogenium*, 34.5%; *Thermoanaerobacter*, 3.4%; *Sedimentibacter*, 1.1%; and *Robinsoniella*, 1.1%) and *Synergistetes* (*Thermovirga*, 29.9%) composed 70% of the sequenced clones. This clone library also contained OTUs that were members of the *Thermodesulfobacterium* (*Thermodesulfobacter*, 12.6%), *Bacteroidetes* (*Proteiniphilum*, 9.2%), *Thermotogae* (*Kosmotoga*, 2.3% and *Thermotoga*, 1%) and

Deltaproteobacteria (*Desulfomicrobium*, 2.3%). Seven OTUs represented 34.5% of clones, which were 93.4–99.6% similar to *Themacetogenium phaeum* strain PB (Hattori *et al*., 2000). Three OTUs representing 29.9% of the clones were 93.8–99.7% similar to *T.lienii*, strain Cas60314 [DQ071273 (Dahle and Birkeland, 2006)].

Clones identified as members of the archaeal genera *Methanothermobacter* (75%) and *Methermicoccus* (25%) dominated the archaeal clone library from the PIG sample (Figure 3B). The most abundant archaeal OTU (A1Fa01, 75% of clones) was highly similar (> 99%) to *M. thermautotrophicus* strain delta H (Wright and Pimm, 2003) and to sequences previously detected in samples collected in 2006 from the same ANS field (Duncan *et al*., 2009). It is 96% similar to the sole archaeal OTU detected in a crude oil-degrading methanogenic enrichment cultured at 55°C from an inoculum collected from the ANS field in May 2007 (Gieg *et al*., 2010). The second OTU (A1Fa04, 25% of clones) was 99% similar to that of *Methermicoccus shengliensis* (Cheng *et al*., 2007) and to a minority of sequences previously detected in 2006 (Duncan *et al*., 2009), all recovered from oil field ecosystems.

Characterization and comparison of bacterial communities

Pyrosequencing of 16S rRNA gene libraries was used to survey the bacterial communities from the separator (SEPa, 2328 reads; SEPb, 277 reads), biocide-treated seawater (SEA, 629 reads) and seawater/produced water mix (SPWa, 1328 reads; SPWb, 4339 reads) pipelines of the processing facility, injection well (INJ, 1625 reads) and the envelope of solids produced from a pig (PIGa, 1635 reads). The 12,161 sequences from these seven libraries were clustered into 2331 OTUs (1652 singletons)

at a sequence dissimilarity cut-off of 3%. Table1 shows the observed number of OTUs, the abundance-based coverage estimator (ACE), the Chao1 estimator of species diversity and a non-parametric estimate of the Shannon diversity index (H') for each library randomly sampled based on the smallest library (SEA, n = 629; SEPb was not included). Although the diversity of the separator community is higher than the other bacterial communities, all of the values overlap between 95% confidence intervals. Rarefaction analysis of each library indicates that the separator and pig communities are indeed more diverse than the communities of the seawater/produced water, injection well and seawater samples (Figure 4).

The relative abundances of taxa from bacterial 16S rRNA pyrosequencing for each sample can be seen in Figure 5. Additionally, the dendrogram (Figure 5A) illustrates the similarity between communities using an overlap index (Θ_{YC}) developed by Yue and Clayton, which is based on the proportions of both shared and non-shared species between two communities (Yue and Clayton, 2005). As with the bacterial clone libraries, the pyrosequenced libraries (with the exception of the seawater library) are dominated by the anaerobic thermophilic *Firmicutes* (*Thermoanaerobacter* and *Thermacetogenium*) and *Synergistete Thermovirga* (Figure 5C). With the greater depth of sequencing afforded by pyrosequencing, many more OTUs that are present at lower abundances were also detected.

A comparison of OTUs for the most abundant taxa shows that the PIG, SEP, SPW and INJ libraries are dominated by the same few OTUs (Figure 5C). Differences between communities are found in the proportions of OTUs and presence/absence of rare OTUs. Of the most abundant OTUs, the PIG sample differs mostly in the relative

proportion of *Thermacetogenium* versus *Thermoanaerobacter* compared with the other samples.

The RDP's LibCompare tool (Cole *et al*., 2009) was used to identify significant differences ($P < 0.001$) in the relative abundance of several taxa between the bacterial communities represented by the SEPa ($n = 2328$ sequences) and PIGa ($n = 1635$) libraries (Table 2). The PIG bacterial community had significantly higher abundances for the *Bacteroidetes* genus *Proteiniphilum* (49.2-fold) and the *Firmicutes* genera *Thermacetogenium* (1.5-fold) and *Sedimentibacter* (9.9-fold). Of particular note are the 39 OTUs (120 sequences, 7.3%) for the *Bacteroidete* genus *Proteiniphilum* (94.0–96.8% similar to *Proteiniphilum acetatigenes*, strain TB107, AY742226) that are uniquely found in the PIG sample. The SEP bacterial community had many taxa with greater relative abundances including the *Firmicutes* genera *Thermoanaerobacter* (15.2-fold) and *Thermolithobacter* (23.9- fold). The *Proteobacteria* in the SEP community included the *Betaproteobacterium* genus *Variovorax* (9.2-fold), *Deltaproteobacterium* genus *Desulfomicrobium* (18.0- fold) and the *Gammaproteobacterium* genera *Marinobacterium* (219.0-fold) and *Pseudomonas* (21.4-fold), which were all more abundant compared with the pig community. Lastly, the *Thermotogae* genera *Kosmotoga* (19.2- fold), *Thermosipho* (17.7-fold) and *Petrotoga* (13.5-fold) were more abundant in the SEP community.

Enumeration of bacteria and archaea using quantitative PCR

Using quantitative PCR (qPCR), the number of bacterial and archaeal 16S rRNA gene copies was estimated per ml of each sample (Figure 6). The number of bacterial

16S rRNA gene copies was calculated to be 3.3×10^3 (INJ), 2.3×10^3 (SEPb), 7.7×10^4 (SPWb), 1.5×10^5 (SEA) and 9.4×10^9 (PIG) per ml of sample. The archaea were less abundant by 10-fold in samples from the INJ (3.4×10^2 ml^{-1}), 50-fold in SPWb (1.4×10^3 ml^{-1}), 170-fold in SEA (8.7×10^2 ml^{-1}) and 150-fold in PIG (6.4×10^7 ml^{-1}) but more abundant than the bacteria in the SEPb sample (2.1×10^5 ml^{-1}). Based on qPCR data, the pig sample represented a 1.2×10^5-fold enrichment in the number of bacterial 16S rRNA gene copies compared with a bulk fluid sample from a similar pipeline (SPWb).

Discussion

The objective of this study was to compare microbial communities in produced water and biofilms associated with pipelines in order to document populations associated with biocorrosion. As is common practice, the water, gas and oil produced from the sampled reservoir is transported to a processing facility, where each phase is separated (Fig. 1). Smaller hydrocarbons are recovered as gas and returned to the formation using gas injection wells. Produced water is also returned to the formation using water injection wells. Biocide-treated water (seawater in the facility described here) is used when necessary to maintain pressures and facilitate further oil production. The addition of non-produced water, called water flooding, has the potential drawback of introducing substrates that promote microbial growth such as electron acceptors (e.g. sulfate, thiosulfate, oxygen and nitrate).

Bacterial community composition and structure of produced water samples

A survey of the bacterial communities across the oil production facility was conducted using pyrosequencing of bacterial 16S rRNA gene libraries. A random, equal sampling (n = 629 reads) of these libraries was used to determine diversity indices unbiased by sampling effort (Table 1). These indices and rarefaction curves of OTUs show that the SEP and PIG bacterial communities were more diverse than the other samples, and that additional sampling effort will be needed to approach complete coverage. Identification and grouping of OTUs using a phylogenetically based taxonomic hierarchy indicated that the bacterial communities surveyed here are similar to those found in other high-temperature (> 70°C) anaerobic oil fields of the ANS (Duncan *et al*., 2009), China (Nazina *et al*., 2006) and the North Sea (Dahle *et al*., 2008). Bacterial communities described here and mesophilic (27°C) fields of the ANS shared an abundance of strictly anaerobic, Gram-positive bacteria (*Thermacetogenium, Thermoanaerobacter* and *Thermovirga*) but the mesophilic produced waters had a higher abundance of proteobacteria and uncultivated candidate divisions JS1, WS6 and OP11 (Pham *et al*., 2009).

Most of the diversity of the sampled bacterial communities was found in the rare OTUs (1652 singletons), containing a single read in one or more libraries. The rarity of these detected organisms makes it difficult to speculate what role they play in the ecology of this ecosystem. At the least, they represent a pool of genomic diversity as a resource of adaptive innovation to respond to environmental and ecological shifts over time (Kunin and Gaston, 1993). The molecular surveys conducted here were also used to identify populations associated with locations, conditions and practices of oil

production with a particular focus on the microorganisms implicated in pipeline biocorrosion.

The overall bacterial community composition across the oil production sites was remarkably similar (based on shared OTUs) despite the fact that some samples represent produced water only (INJ), receive biocide-treated seawater as an amendment (SPW), are heated to separate produced water from hydrocarbons (SEP), or are enriched in microbial biofilms attached to the inner surface of a pipeline (PIG) (Fig. 5). The similar community composition was not unexpected since these samples are all exposed to oil production fluids from the same reservoir, have similar environmental conditions (e.g. thermophilic, anoxic, hydrocarbon-enriched), and have been affected by the homogenizing effect of oil production over a 30+ year period.

Comparisons of the community structure for each bacterial library (relative abundances of shared and unique OTUs) indicated the SEA and PIG communities were outliers to those from the SEP, SPW and INJ samples (Figure 5). Differences among the community structures of almost all samples were determined to be significant (α = 0.05) based on the Cramer-von Mises test statistic implemented in ∫-LIBSHUFF with a Bonferroni correction for multiple comparisons (n = 7) (Schloss *et al*., 2004). Bacterial community structure did not differ ($P < 0.007$) between replicate samples (i.e. SEPa versus SEPb, SPWa versus SPWb) despite different DNA extraction methods. Differences in the community structure of samples SPWa and SEPb were not significant, which was interpreted as an indication of their similar origin and composition (produced water amended with seawater).

'Pig envelope' contains microorganisms associated with biocorrosion

The pig envelope represents an enrichment of surface-associated solids, including accumulated biofilms from a produced water pipeline. This, to our knowledge, is the first time that the microbial diversity in such a sample has been surveyed by molecular methods. The population density of cultivable sulfate-reducing microorganisms in the pig envelope was estimated to be $> 10^6$ cells ml^{-1} using a standard industry protocol (NACE, 2004). This may be an underestimate, as qPCR results indicate there are 9.4×10^9 bacterial and 6.4×10^7 archaeal 16S rRNA gene copies per ml of the PIG envelope (Figure 6). The pig envelope contains between four to six orders of magnitude more bacteria and two to five orders of magnitude more archaea than the produced water samples, evidence that the pig sample is enriched in biofilm-associated microorganisms. It should be noted that extracellular DNA (eDNA) is a major component of the extracellular polymeric substances excreted by bacteria in biofilms (Steinberger and Holden, 2005; Tani and Nasu, 2010). Therefore, it is possible that the direct survey of biofilms with molecular methods may in fact be biased towards those microorganisms that either are more abundant or produce more eDNA than other biofilm cohabitants.

The bacteria detected in the sulfate-reducing incubations at 40°C include close relatives of the *Firmicute T. pseudethanolicus*, the *Deltaproteobacterium D. thermophilum* and the *Synergistete T. lienii*. The same *T. pseudethanolicus*-like sequences were detected in dilutions incubated at 75°C, indicating the ability to grow at temperatures common throughout the oil production facility (40–50°C) as well as reservoir temperatures (> 70°C). These findings are consistent with direct molecular

surveys of the pig envelope, as these sequences and those closely related to members of the genera *Thermacetogenium, Thermodesulfobacterium, Thermotoga*, *Kosmotoga* and *Proteiniphilum* make up over 90% of the bacterial community (Figures 2 and 5). Described species from all of the genera listed above except for *Proteiniphilum* are able to reduce one or more sulfur compounds (e.g. S°, $S_2O_3^{-2}$ and SO_4^{-2}) as an electron acceptor and produce hydrogen sulfide, which is highly corrosive to steel (Little and Wagner, 1992). Sulfate-reducing bacteria have been implicated in biocorrosion of metal surfaces from many environments (Little and Wagner, 1992; Lee *et al*., 1995; Little *et al*., 2006) but few studies have implicated these high-temperature bacteria (Duncan *et al*., 2009).

The most abundant archaea detected in the 40°C enrichments were close relatives of the autotrophic methanogen *M. thermautotrophicus*. These findings are also consistent with the molecular survey of archaea in the pig sample, as *Methanothermobacter* and the less abundant methylotrophic methanogen *Methermicoccus* were the only archaeal sequences detected. Sequences similar to *T. alcaliphilus* were detected in the 75°C enrichments but were only detected in the 10^{-1} to 10^{-3} dilutions (data not shown), suggesting a much lower abundance that would also explain its absence from the limited clone library ($n = 84$).

Ecological insights based on phylogenetically inferred physiologies

The bacterial communities elucidated from pyrosequencing of 16S rRNA gene libraries were dominated by a few OTUs that represent the genera *Thermovirga*, *Thermoanaerobacter* and *Thermacetogenium* (≥ 90% 16S rRNA sequence similarity).

Thermovira lienii (Dahle and Birkeland, 2006) and *T. pseudethanolicus* (Onyenwoke *et al*., 2007), the nearest characterized representatives of these genera, are thermophilic, anaerobic, Gram-positive bacteria capable of utilizing carbohydrates, proteinous compounds, amino acids and organic acids fermentatively or with elemental sulfur as an electron acceptor (*T. lienii*). *Thermoanaerobacter pseudethanolicus* can reduce thiosulfate to H_2S non-energetically. Several of the products formed by *T. lienii* (e.g. ethanol, acetate, H_2 and CO_2) could serve as substrates for the closest characterized representative of the genus *Thermacetogenium, T. phaeum* (Hattori *et al*., 2000). *Thermacetogenium phaeum* is a thermophilic acetogen originally isolated from a thermophilic anaerobic bioreactor that can also oxidize acetate with sulfate or thiosulfate. Furthermore, *T. phaeum* can grow syntrophically with an autotrophic methanogen, which has been postulated to be important in high-temperature, methanogenic oil fields (Hattori *et al*., 2000; Nazina *et al*., 2006; Hattori, 2008). In the microbial communities surveyed here, however, the archaea are 50- to 150-fold less abundant than the bacteria in all but the SEP community, suggesting that only a small percentage of these bacteria might be existing syntrophically with an archaeon in these environments. The role of these dominant bacteria in the biodegradation of oil has not been established but closely related sequences to the *Thermacetogenium* OTUs were obtained from thermophilic crude oil-degrading, methanogenic consortia enriched from samples obtained from different locations within the same oil production facility (Gieg *et al*., 2010).

The pig envelope, representative of the biofilm community, is dominated by the genera *Thermacetogenium* and *Thermovirga*. Samples of bulk fluids, however, contain

the same representative populations, but at different relative abundances. Bulk fluid samples from these pipelines are therefore capable of capturing the same OTUs. The envelopes of solids produced by pigging activities should be an important component to identifying the microbes associated with biocorrosion. Deep sequencing of samples from the oil facility allows for a more comprehensive analysis of shifts in community composition associated with sites, treatments and temporal variation throughout the oil production facility. Since relatively few OTUs dominated the microbial community, chip-based technologies may eventually prove appropriate for monitoring shifts in microbial communities as oil production activities occur (Hazen *et al*., 2010). This research presents a relatively short list of microbial taxa that are currently being targeted for cultivation and further study for their ability to degrade hydrocarbons and enhance rates of corrosion either directly or through their metabolic activity.

Acknowledgements

Support from the National Science Foundation (Award No. 0647712) is gratefully acknowledged. We also thank two anonymous reviewers who provided insightful comments, enhancing the final manuscript.

References

Beech, I.B., Sunner, J.A., and Hiraoka, K. (2005) Microbe–surface interactions in biofouling and biocorrosion processes. *Int Microbiol* **8:** 157–168.

Beech, I.B., and Sunner, J. (2004) Biocorrosion: towards understanding interactions between biofilms and metals. *Curr Opin Biotechnol* **15:** 181–186.

Chao, A. (1984) Nonparametric estimation of the number of classes in a population. *Scand J Stat* **11:** 265–270.

Chao, A., Hwang, W.-H., Chen, Y.-C., and Kuo, C.-Y. (2000) Estimating the number of shared species in two communities. *Stat Sinica* **10:** 227–246.

Cheng, L., Qiu, T.L., Yin, X.B., Wu, X.L., Hu, G.Q., Deng, Y., *et al*. (2007) *Methermicoccus shengliensis* gen. nov., sp nov., a thermophilic, methylotrophic methanogen isolated from oil-production water, and proposal of *Methermicoccaceae* fam. nov. *Int J Syst Evol Microbiol* **57:** 2964–2969.

Clark, B. (2005) Pigging continues on strong course world-wide. *Pipeline Gas J* **232:** 26–36.

Cole, J.R., Wang, Q., Cardenas, E., Fish, J., Chai, B., Farris, R.J., *etal*. (2009) The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. *Nucleic Acids Res* **37:** D141–D145.

Dahle, H., and Birkeland, N.K. (2006) *Thermovirga lienii* gen. nov., sp nov., a novel moderately thermophilic, anaerobic, amino-acid-degrading bacterium isolated from a North Sea oil well. *Int J Syst Evol Microbiol* **56:** 1539–1545.

Dahle, H., Garshol, F., Madsen, M., and Birkeland, N.K. (2008) Microbial community structure analysis of produced water from a high-temperature North Sea oil-field. *Antonie Van Leeuwenhoek* **93:** 37–49.

DeSantis, T.Z., Hugenholtz, P., Keller, K., Brodie, E.L., Larsen, N., Piceno, Y.M., *et al*. (2006a) NAST: a multiple sequence alignment server for comparative analysis of 16S rRNA genes. *Nucleic Acids Res* **34:** W394–W399.

DeSantis, T.Z., Hugenholtz, P., Larsen, N., Rojas, M., Brodie, E.L., Keller, K., *et al*. (2006b) Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl Environ Microbiol* **72:** 5069–5072.

Duncan, K.E., Gieg, L.M., Parisi, V.A., Tanner, R.S., Tringe, S.G., Bristow, J., *et al*. (2009) Biocorrosive thermophilic microbial communities in Alaskan North Slope oil facilities. *Environ Sci Technol* **43:** 7977–7984.

Gieg, L.M., Duncan, K.E., and Suflita, J.M. (2008) Bioenergy production via microbial conversion of residual oil to natural gas. *Appl Environ Microbiol* **74:** 3022–3029.

Gieg, L.M., Davidova, I.A., Duncan, K.E., and Suflita, J.M. (2010) Methanogenesis, sulfate reduction and crude oil biodegradation in hot Alaskan oilfields. *Environ Microbiol* **12:** 3074–3086.

Goedecke, H. (2003) Ultrasonic or MFL inspection: which technology is better for you? *Pipeline Gas J* **230:** 34–37.

Grabowski, A., Nercessian, O., Fayolle, F., Blanchet, D., and Jeanthon, C. (2005) Microbial diversity in production waters of a low-temperature biodegraded oil reservoir. *FEMS Microbiol Ecol* **54:** 427–443.

Hamady, M., Walker, J.J., Harris, J.K., Gold, N.J., and Knight, R. (2008) Error-correcting barcoded primers for pyrosequencing hundreds of samples in multiplex. *Nat Methods* **5:** 235–237.

Hamilton, W.A. (2003) Microbially influenced corrosion as a model system for the study of metal microbe interactions: a unifying electron transfer hypothesis. *Biofouling* **19:** 65– 76.

Hattori, S. (2008) Syntrophic acetate-oxidizing microbes in methanogenic environments. *Microbes Environ* **23:** 118– 127.

Hattori, S., Kamagata, Y., Hanada, S., and Shoun, H. (2000) *Thermacetogenium phaeum* gen. nov., sp nov., a strictly anaerobic, thermophilic, syntrophic acetate-oxidizing bacterium. *Int J Syst Evol Microbiol* **50:** 1601–1609.

Hazen, T.C., Dubinsky, E.A., DeSantis, T.Z., Andersen, G.L., Piceno, Y.M., Singh, N., *et al*. (2010) Deep-sea oil plume enriches indigenous oil-degrading bacteria. *Science* **330:** 204–208.

Head, I.M., Jones, D.M., and Larter, S.R. (2003) Biological activity in the deep subsurface and the origin of heavy oil. *Nature* **426:** 344–352.

Herrick, J.B., Madsen, E.L., Batt, C.A., and Ghiorse, W.C. (1993) Polymerase chain reaction amplification of naphthalene catabolic and 16S ribosomal RNA gene sequences from indigenous sediment bacteria. *Appl Environ Microbiol* **59:** 687–694.

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

Keller, M., Braun, F.J., Dirmeier, R., Hafenbradl, D., Burggraf, S., Rachel, R., *et al*. (1995) *Thermococcus alcaliphilus* sp nov, a new hyperthermophilic archaeum growing on polysulfide at alkaline pH. *Arch Microbiol* **164:** 390– 395.

van der Kraan, G.M., Bruining, J., Lomans, B.P., van Loosdrecht, M.C.M., and Muyzer, G. (2010) Microbial diversity of an oil-water processing site and its associated oil field: the possible role of microorganisms as information carriers from oil-associated environments. *FEMS Microbiol Ecol* **71:** 428–443.

Kunin, W.E., and Gaston, K.J. (1993) The biology of rarity: patterns, causes and consequences. *Trends Ecol Evol* **8:** 298–301.

Lee, W., Lewandowski, Z., Nielsen, P.H., and Hamilton, W.A. (1995) Role of sulfate-reducing bacteria in corrosion of mild steel – a review. *Biofouling* **8:** 165.

Little, B., and Wagner, P. (1992) *Quantifying Sulfate Reducing Bacteria in Microbiologically Influenced Corrosion*. Technical Report. Monterey, CA, USA: Naval Research Lab.

Little, B.J., Lee, J.S., and Ray, R.I. (2006) Diagnosing microbiologically influenced corrosion: a state-of-the-art review. *Corrosion* **62:** 1006–1017.

Magot, M., Ollivier, B., and Patel, B.K.C. (2000) Microbiology of petroleum reservoirs. *Antonie Van Leeuwenhoek* **77:** 103–116.

Magurran, A.E. (2004) *Measuring Biological Diversity*. Malden, MA, USA: Blackwell.

Masterson, W.D., Dzou, L.I.P., Holba, A.G., Fincannon, A.L., and Ellis, L. (2001) Evidence for biodegradation and evaporative fractionation in West Sak, Kuparuk and Prudhoe Bay field areas, North Slope, Alaska. *Org Geochem* **32:** 411–441.

Muyzer, G., Brinkoff, T., Nubel, U., Santegoeds, C., Schafer, H., and Wawer, C. (1998) Denaturing gradient gel electro- phoresis (DGGE) in microbial ecology. In *Molecular Microbial Ecology Manual*, Vol. 3.4.4. Akkermans, A.D.L., van Elsas, J.D., and deBruijn, F.J. (eds). Dordrecht, the Netherlands: Kluwer, pp. 1–27.

NACE (2004) *Field Monitoring of Bacterial Growth in Oil and Gas Systems*. Houston, TX, USA: NACE, TM0194- 2004.

Nazina, T.N., Shestakova, N.M., Grigor'yan, A.A., Mikhailova, E.M., Tourova, T.P., Poltaraus, A.B., *et al*. (2006) Phyloge- netic diversity and activity of anaerobic microorganisms of high-temperature horizons of the Dagang oil field (P.R. China). *Microbiology* **75:** 55–65.

Onyenwoke, R.U., Kevbrin, V.V., Lysenko, A.M., and Wiegel, J. (2007) *Thermoanaerobacter pseudethanolicus* sp nov., a thermophilic heterotrophic anaerobe from Yellowstone National Park. *Int J Syst Evol Microbiol* **57:** 2191– 2193.

Pham, V.D., Hnatow, L.L., Zhang, S.P., Fallon, R.D., Jackson, S.C., Tomb, J.F., *et al*. (2009) Characterizing microbial diversity in production water from an Alaskan mesothermic petroleum reservoir with two independent molecular methods. *Environ Microbiol* **11:** 176–187.

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

Quarini, J., and Shire, S. (2007) A review of fluid-driven pipeline pigs and their applications. *Proc IME E J Process Mech Eng* **221:** 1–10.

Rainey, F.A., Ward, N.L., Morgan, H.W., Toalster, R., and Stackebrandt, E. (1993) Phylogenetic analysis of anaerobic thermophilic bacteria – aid for their reclassification. *J Bacteriol* **175:** 4772–4779.

Sambrook, J. (2001) *Molecular Cloning: A Laboratory Manual (3-Volume Set)*. Cold Spring Harbor, NY, USA: Cold Spring Harbor Laboratory Press.

Santegoeds, C.M., Damgaard, L.R., Hesselink, C., Zopfi, J., Lens, P., Muyzer, G., *et al*. (1999) Distribution of sulfate-reducing and methanogenic bacteria in anaerobic aggregates determined by microsensor and molecular analyses. *Appl Environ Microbiol* **65:** 4618–4629.

Schloss, P.D. (2009) A high-throughput DNA sequence aligner for microbial ecology studies. *PLoS ONE* **4:** e8230. Schloss, P.D., and Handelsman, J. (2005) Introducing DOTUR, a computer program for defining operational taxonomic units and estimating species richness. *Appl Environ Microbiol* **71:** 1501–1506.

Schloss, P.D., Larget, B.R., and Handelsman, J. (2004) Integration of microbial ecology and statistics: a test to compare gene libraries. *Appl Environ Microbiol* **70:** 5485–5492.

Schloss, P.D., Westcott, S.L., Ryabin, T., Hall, J.R., Hartmann, M., Hollister, E.B., *et al*. (2009) Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl Environ Microbiol* **75:** 7537–7541.

Steinberger, R.E., and Holden, P.A. (2005) Extracellular DNA in single- and multiple-species unsaturated biofilms. *Appl Environ Microbiol* **71:** 5404–5410.

Struchtemeyer, C.G., Elshahed, M.S., Duncan, K.E., and McInerney, M.J. (2005) Evidence for aceticlastic methanogenesis in the presence of sulfate in a gas condensate-contaminated aquifer. *Appl Environ Microbiol* **71:** 5348– 5353.

Tani, K., and Nasu, M. (2010) Roles of extracellular DNA in bacterial ecosystem. In *Extracellular Nucleic Acids*. Kikuchi, Y., and Rykova, E.Y. (eds). Berlin, Germany: Springer-Verlag, pp. 25–37.

Thevenieau, F., Fardeau, M.L., Ollivier, B., Joulian, C., and Baena, S. (2007) *Desulfomicrobium thermophilum* sp nov., a novel thermophilic sulphate-reducing bacterium isolated from a terrestrial hot spring in Colombia. *Extremophiles* **11:** 295–303.

Wang, Q., Garrity, G.M., Tiedje, J.M., and Cole, J.R. (2007) Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl Environ Microbiol* **73:** 5261–5267.

Wright, A.-D.G., and Pimm, C. (2003) Improved strategy for presumptive identification of methanogens using 16S ribo- printing. *J Microbiol Methods* **55:** 337–349.

Yue, J.C., and Clayton, M.K. (2005) A similarity measure based on species proportions. *Commun Stat Theory* **34:** 2123–2131.

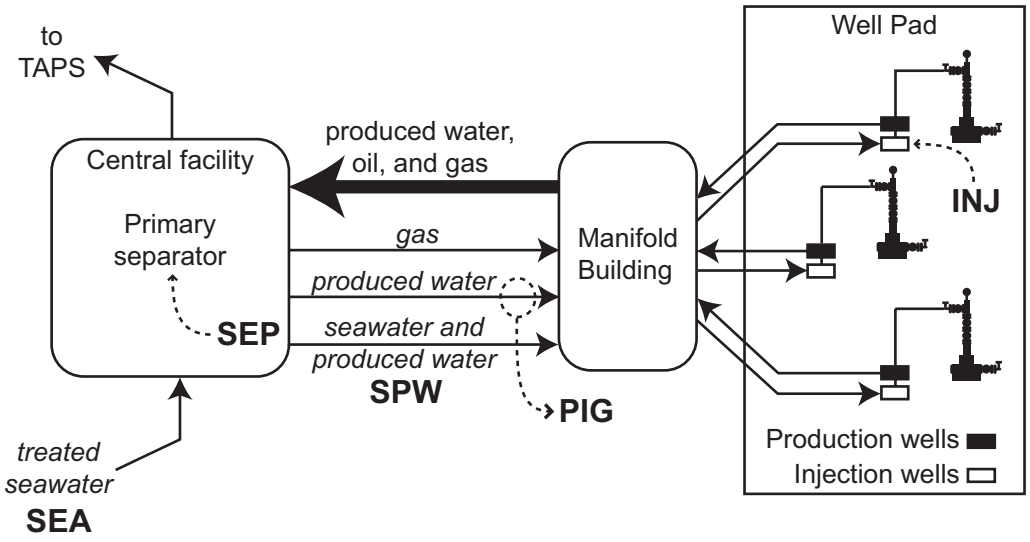


Figure 1. **Schematic diagram of the sampled oil production facility. Arrows indicate the direction of flow in pipelines. Locations of samples taken are indicated in bold.**

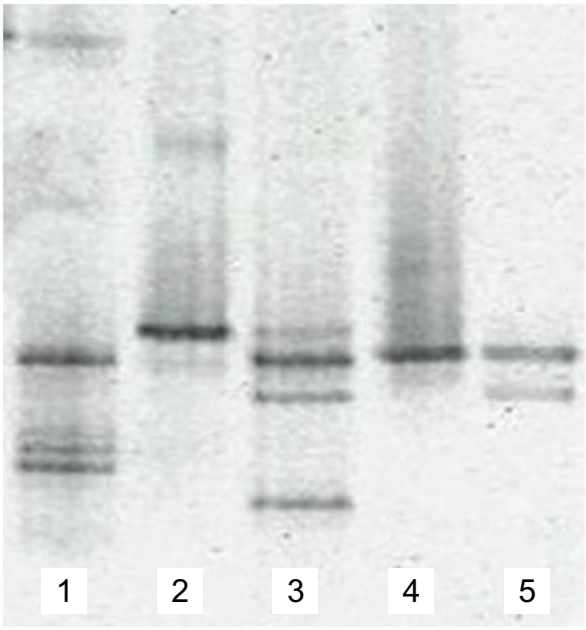


Figure 2. **DGGE of bacterial 16S rDNA PCR fragments from the PIGb (1), cloned sequence (C540-6, 98% similarity to *Desulfomicrobium thermophilum*, AY464939) from 10^{-6} enrichment culture at 40°C (2), 10^{-6} enrichment culture at 40°C (3), cloned sequence (E675-6, 99% similarity to *Thermoanaerobacter pseudethanolicus*, CP000924) from 10^{-6} enrichment culture at 75°C (4), 10^{-6} enrichment culture at 75°C (5).**

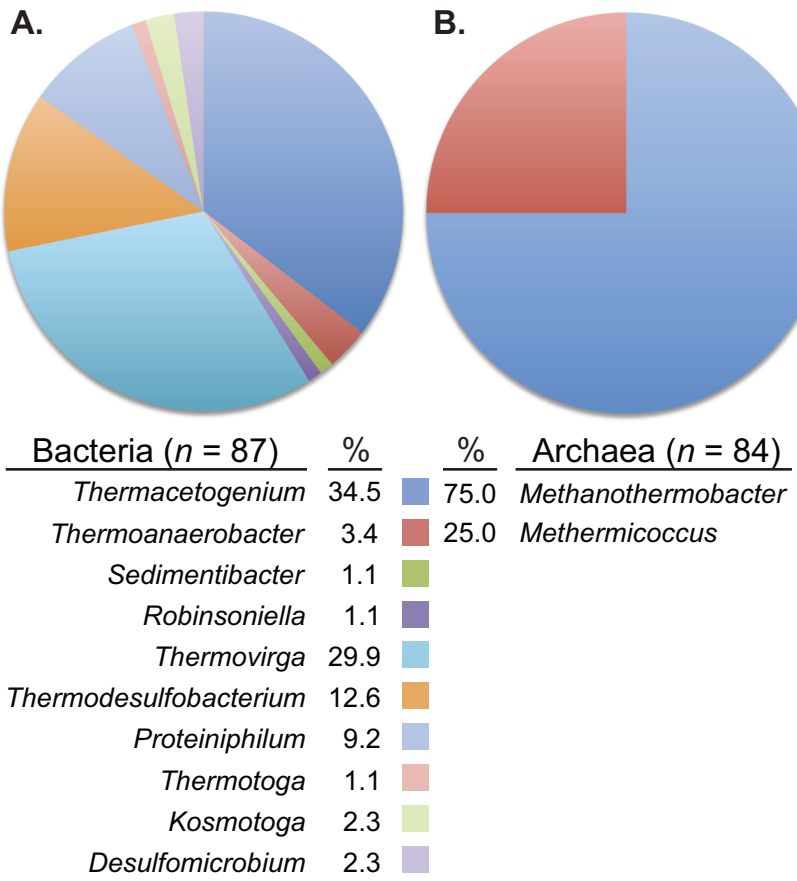


Figure 3. Relative abundance (%) of genera in (A) bacterial and (B) archaeal 16S rRNA clone libraries from the pig sample.

Table 1. **Number of observed OTUs, non-parametric Shannon (*H*^), ACE, and Chao1 indices at a dissimilarity cutoff of 3% for each randomly subsampled library (n=629)**

Library	OTUs	Shannon (*H*^)[a]	ACE[b]	Chao1[b]
PIGa	198	4.249	2091 (1753, 2501)	887 (599, 1383)
SEPa	217	4.613	3622 (3109, 4226)	1515 (926, 2598)
SPW	196	3.920	2022 (1683, 2440)	700 (501, 1028)
INJ	188	4.000	2271 (1832, 2828)	793 (545, 1212)
SEA	196	3.978	2127 (1777, 2555)	816 (563, 1245)

[a] Non-parametric estimate of traditional Shannon diversity index

[b] Mean values are given with 95% low and high confidence levels in parentheses

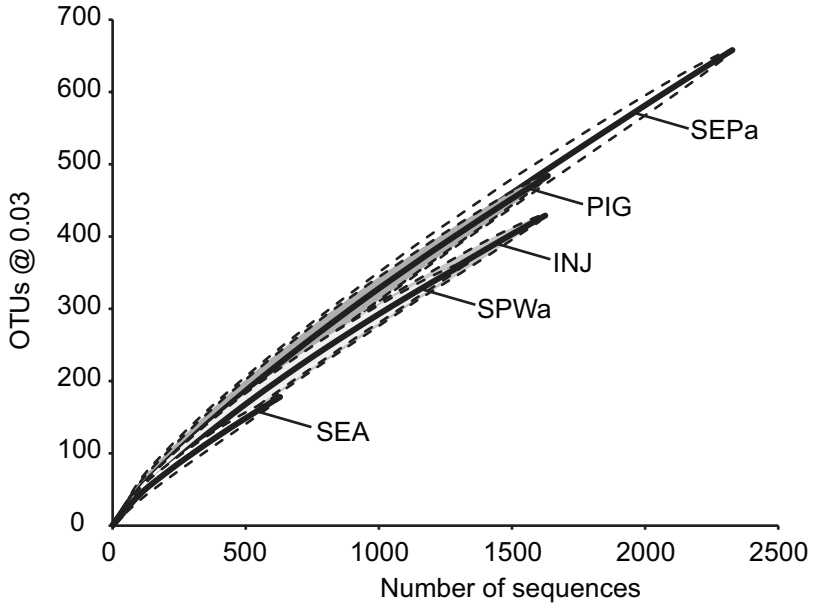


Figure 4. **Rarefaction analysis of OTUs (3% dissimilarity cut-off) detected from pyrosequencing of bacterial 16S rRNA gene libraries. Dotted lines and shaded areas represent 95% confidence intervals.**

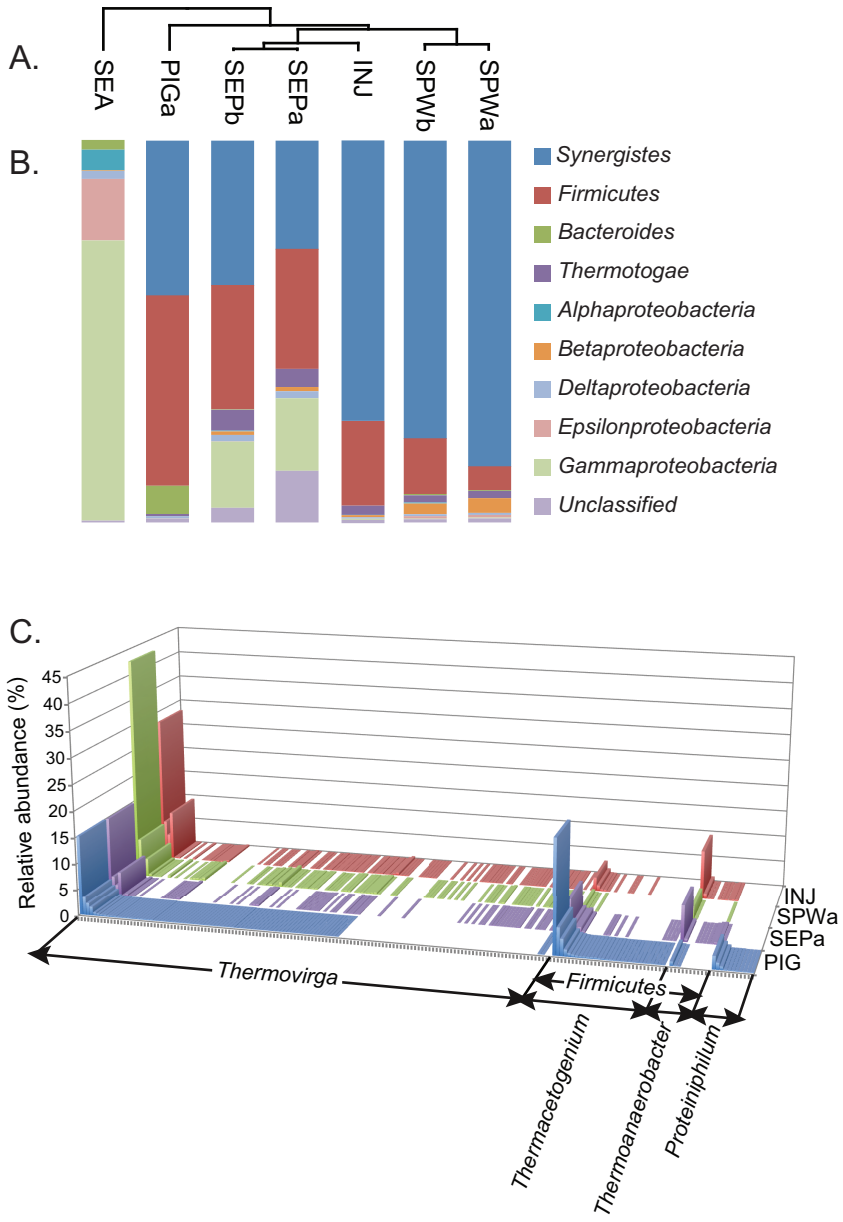


Figure 5. Pyrosequencing analysis of 16S rRNA gene libraries showing (A) a dendogram based on comparisons of community structure Θ_{YC} values, (B) relative abundances of represented phyla (class level for Proteobacteria) and (C) a comparison of OTUs for major taxa.

Table 2. Significantly different relative abundances[a] for taxa between the microbial communities of the separator (SEP) and PIG in a pairwise comparison (implementation of the RDP LibCompare statistic, $P < 0.001$)

Taxonomic rank	Name	SEP	PIG	Fold increase	Fold decrease
Phylum	*Bacteroidetes*	0.3	5.2	**16.9**	
Genus	*Proteiniphilum*	0.1	3.0	**49.2**	
Phylum	*Firmicutes*	45.2	34.7		1.3
Class	*Clostridia*	42.6	34.7		1.2
Family	*Thermoanaerobacteraceae*	37.1	32.4		1.1
Genus	*Thermacetogenium*	20.2	31.3	1.5	
Genus	*Thermoanaerobacter*	16.3	1.1		**15.2**
Family	*Incertae Sedis XI*	0.1	1.2	**19.7**	
Genus	*Sedimentibacter*	**0.1**	1.0	**9.9**	
Class	*Thermolithobacteria*	2.4	**0.1**		**23.9**
Genus	*Thermolithobacter*	2.4	**0.1**		**23.9**
Phylum	*Proteobacteria*	29.1	0.4		**67.6**
Class	*Betaproteobacteria*	1.3	**0.1**		**12.8**
Genus	*Variovorax*	0.9	**0.1**		**9.2**
Class	*Deltaproteobacteria*	3.1	0.4		**7.9**
Genus	*Desulfomicrobium*	2.3	0.1		**18.0**
Class	*Gammaproteobacteria*	24.3	**0.1**		**243.4**
Order	*Alteromonadales*	22.0	**0.1**		**220.2**
Genus	*Marinobacterium*	21.9	**0.1**		**219.0**
Order	*Pseudomonadales*	2.2	**0.1**		**22.0**
Genus	*Pseudomonas*	2.1	**0.1**		**21.4**
Phylum	*Thermotogae*	7.6	0.4		**19.8**
Genus	*Kosmotoga*	3.3	0.2		**19.2**
Genus	*Thermosipho*	1.8	**0.1**		**17.7**
Genus	*Petrotoga*	1.3	**0.1**		**13.5**

[a] values show relative abundance (% of library) for each taxonomic level. Only taxa comprising > 0.5% of the community in one of the two samples being compared are listed. Bold indicates changes greater than 4. Bold underline indicates fold calculations were based on change of abundance from 0 to 0.1 in sample.

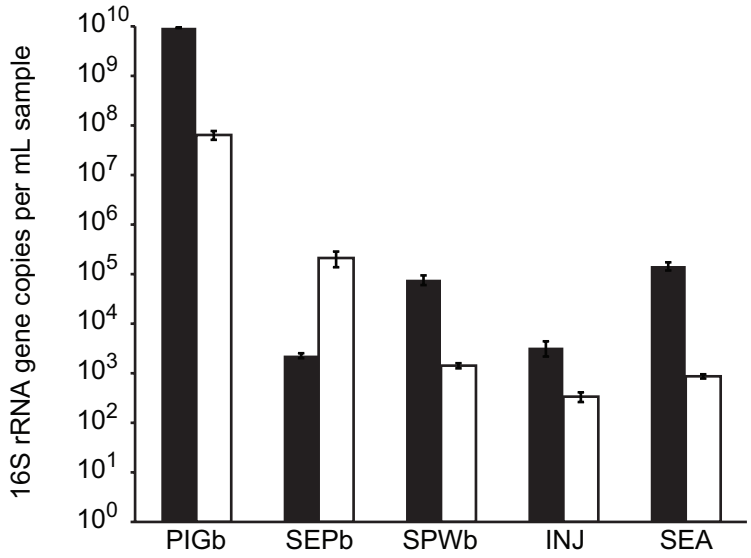


Figure 6. Copy number of 16S rRNA genes per mL of sample for bacteria (black) and archaea (white) based on qPCR analysis. Error bars indicate standard deviation of means (n=3). Values not determined are designated as "n.d.".

Chapter 3

New Insights into the Ecology of Oil Production Facilities Using Microbial Communities, Metagenomes and Hydrocarbon Metabolites

Abstract

The proliferation of microorganisms and their activities pose significant challenges to the oil and gas industry, such as reservoir souring and corrosion. There is a critical need to better understand the role of microorganisms in these detrimental and damaging activities. A comprehensive study of two oil production facilities on the North Slope of Alaska was conducted to link the microbial populations, their metabolic potential, and their detected activities. The resident microbial communities were characterized by sequencing 16S rRNA gene libraries, their metabolic potential was determined through metagenome sequencing, and their metabolic activity was assessed by identifying putative hydrocarbon metabolites. The younger of the two facilities (facility A), used seawater for secondary recovery operations and harbored produced fluids that were 16-35 °C. In contrast, the nearby, older facility (facility B) recycled production fluids as a routine part of their oil recovery operations and had produced fluids that were higher in temperature (35-60 °C). The microbial communities at facility A consisted of mesophilic, facultatively anaerobic members of the Delta- and Gammaproteobacteria. The composition of microbial communities from facility B were mainly strictly anaerobic, thermophilic members of the Synergistes, and Firmicutes.

Almost all samples from both facilities contained putative metabolites consistent with hydrocarbon biodegradation. Metagenomes representative of the two oil production facilities were sequenced to identify genes involved in hydrocarbon degradation pathways and provide a link between microbial community structure and hydrocarbon degradation. Genes associated with aerobic hydrocarbon degradation were identified in both facilities, but were more abundant in facility A. No known genes for anaerobic hydrocarbon degradation were identified from either facility. However, putative glycyl radical enzymes (GREs) related to, but separate from, AssA and BssA were detected. These sequences share the conserved glycine motif and cysteine residue of other known GREs. The metagenomic and metabolomic analyses implicate environmental conditions and management practices in selecting the microbial communities and their activities. They also suggest a tentative link between the abundant taxa and hydrocarbon degradation that hinge on the confirmation of novel GREs.

Introduction

Anaerobic bacteria were isolated from oil production fluids as early as 1926, providing the first support for the presence of microorganisms in subsurface oil reservoirs (Bastin *et al.* 1926). It is not entirely clear whether these microorganisms were indigenous to the reservoir or were introduced as a result of production activities (Magot 2005). Regardless of their provenance, microorganisms in oil reservoirs can mineralize hydrocarbons, increasing oil viscosity, sulfur content, and total acid number (Masterson *et al.* 2001, Head *et al.* 2003, Larter *et al.* 2003). The recovery of such bio-

deteriorated oil increases production costs and can be economically unfavorable (Roling *et al.* 2003). Additionally, hydrocarbon consumption by microorganisms produces polar acidic intermediates and end products such as naphthenic acids, low molecular weight organic acids, and carbon dioxide (Duncan *et al.* 2009). The presence of polar compounds in oil processing facilities and transport pipelines increases their vulnerability to corrosion and requires additional prevention and maintenance costs (Suflita *et al.* 2008).

In this study, the community structure, genomic potential, and function of microbial assemblages was investigated to understand the likelihood for bacterial hydrocarbon degradation within two oil fields on the North Slope of Alaska. These two fields experience different management practices. The mature field, facility B, has been producing oil from multiple reservoirs since 1981, whereas the younger operation, facility A, is associated with a single reservoir and has been in production since about 2000 (Carman and Hardwick 1983, Masterson and Eggert 1992, Collett 1993, Gingrich *et al.* 2001, Masterson *et al.* 2001, Holba *et al.* 2004, Verma and Bird 2005, Houseknecht and Bird 2006, Peters *et al.* 2006). At facility B, production water and low molecular weight condensate are recycled to pressurize formations; however, fluid used for secondary recovery operations from facility A was largely mesothermic seawater (18-35 °C). The produced fluids sampled at facility B had higher temperatures overall (35-60 °C) compared to those sampled at facility A (24-47 °C). The viscosity of the oils and sulfur content shows that the geology and petroleum geochemistry differs *within* each field as well as *between* the two fields (Gingrich *et al.* 2001, Masterson *et al.* 2001,

Holba *et al.* 2004). A brief characterization of the origin of each sample and the recorded temperatures are presented in Table S1.

The bacterial assemblages from both fields were characterized using 454 pyrosequencing of 16S rRNA gene libraries. Comparison of the diversity and composition of microbial assemblages across different locations provided insight to the environmental conditions selecting for particular taxa across both oil processing facilities. The *in situ* hydrocarbon biodegradation activity of the microorganisms was assessed through characterization of polar organic compounds differentially associated with the microbial metabolism of hydrocarbons. The presence of these putative metabolites provides evidence of microbial hydrocarbon biodegradation, and identifies hydrocarbons that are actively being metabolized (Beller *et al.* 1995, Beller 2000, Gieg and Suflita 2002). Aerobic hydrocarbon biodegradation pathways and the organisms capable of catalyzing those bioconversions are reasonably well understood (Pérez-Pantoja *et al.* 2010). Anaerobic hydrocarbon biodegradation pathways are less well described but several mechanisms of hydrocarbon activation are known (Callaghan 2013b). Shotgun metagenome sequencing was used to characterize the genomic potential of populations that were numerically dominant in the fluids sampled from the oil production facilities. Additionally, metagenomes were used to link potential hydrocarbon biodegradation pathways to the putative hydrocarbon metabolites. The integration of molecular microbial ecology with analytical chemistry has led to increased insights into the microbial ecology of oil reservoirs and production facilities and the impact of production management practices.

Experimental Procedures

Microbial Community Characterization

Fluid samples from facility B (n = 9) and facility A (n = 6) were obtained via access valves, collected in sterile 1 L glass bottles, and kept under a nitrogen headspace. Biomass was collected from 150 to 300 ml volumes by vacuum filtration onto a 47 mm diameter cellulose nitrate filter with a 0.45 μm pore-size (Thermo Fisher Scientific, Rochester, NY) less than 4 h after sampling. The filters containing biomass were immersed in 1 ml DNAzol® Direct (Molecular Research Center, Cincinnati, OH) and stored in sterile 50 ml screw cap tubes to preserve samples for shipping. Upon arrival, the filters were stored at -80°C until needed. DNA was extracted from each sample using the bead beating protocol described in Stevenson et al. (Stevenson *et al.* 2011). Briefly, sterile glass beads, 10 mL of buffer [10 mM Tris-Cl (pH 8.0), 1 mM ethylenediaminetetraacetic acid and 0.05% Nonidet P40 (Roche Diagnostics GmbH, Mannheim, Germany)] and 10 mL of phenol:chlorophorm:isoamyl alcohol (P:Cl; 25:24:1) were added to each 50 mL tube containing a filter. Tubes were shaken horizontally on a vortex mixer at full speed for 5 min to lyse cells. Following centrifugation at 5000 x *g* for 5 min at 4°C, the resulting aqueous phase was extracted with equal volumes of P:Cl until clear, followed by two extractions with equal volumes of chlorophorm:isoamyl alcohol (24:1). Glycogen was added as a carrier molecule at a final concentration of 0.1 μg ml^{-1} prior to precipitation with ethanol as described by Sambrook et al. (Sambrook 2001). Precipitated nucleic acids were resuspended in up to

100 µl of DNA, RNA and nuclease-free water and quantified by UV spectroscopy at 260 nm (NanoPhotometer; Implen, Westlake Village, CA).

Pyrosequencing of 16S rRNA gene libraries

Extracted DNA was prepared for pyrosequencing following the protocol described previously (Stevenson *et al.* 2011). The primers TiA-8nt-27F and TiB-338R (Table S2) were used to amplify bacterial 16S rRNA genes. These primers contained the Titanium Fusion A or B primer sequence (454 Life Sciences, Branford, CT) followed by an 8 nt unique barcode (forward primer only, Table S3) and sequence from the general bacterial primers Eub27F and Eub338R. Each PCR reaction contained 1x Taq buffer with $(NH_4)_2SO_4$ (Fermentas, Glen Burnie, Maryland), 1.5 mM $MgCl_2$, 0.2 mM each dNTP, 0.2 µM of the forward and reverse primer, 0.625 U of Taq DNA polymerase (Fermentas) and between 10 and 50 ng of template DNA in a final volume of 25 µL. Thermal cycling was carried out in a Techne TC-512 thermal cycler (Techne, Burlington, NJ) using the following conditions: initial denaturation for 5 min at 95°C; 30 cycles of 30 s at 96°C, 45 s at 55°C and 45s at 72°C; and a final extension of 10 min at 72°C. Each sample was amplified in four replicate reactions, pooled and purified with the Wizard® PCR Preps DNA Purification System (Promega, Madison, WI). Pooled and purified samples were then quantified by UV spectroscopy at 260 nm (NanoPhotometer; Implen). PCR product size fidelity was assessed by comparison with DNA markers following agarose gel electrophoresis. Equimolar amounts of each uniquely barcoded PCR product were combined and sequenced using the Genome Sequencer FLX instrument with the GS FLX Titanium series reagents (454 Life

Sciences).

The primers 8AF and A344R (Connelly *et al.* 2014) were used for Archaeal 16S rRNA gene amplification, however no amplification was observed.

Sequence Processing and Analysis

Bacterial 16S rRNA gene library sequences were analyzed using QIIME*,* an open-source bioinformatics software package for comparison and analysis of microbial communities (Caporaso *et al.* 2010b). The libraries were de-multiplexed, and trimmed to remove any sequences that contained errors in the forward primer or barcode regions, ambiguities, or an average quality score <25. Sequences were denoised to reduce the error profile inherent in 454 pyrosequencing (Reeder and Knight 2010), assigned to OTUs at a cutoff of 97 % similarity and chimeras were identified and removed (Edgar 2010). Sequences were aligned using pyNAST (Caporaso *et al.* 2010a). Phylogeny was assigned using the RDP naïve Bayesian rRNA classifier (Wang *et al.* 2007).

Metabolite Profiling

Fluids (1 L) were collected from production wells, injection wells, and processing facilities from the elder, facility B (n=34), and younger, facility A (n=16), and immediately acidified to pH < 2 with 50% HCl to protonate acidic intermediates and halt microbial activity. Acidified samples were kept at room temperature in the dark until extraction with ethyl acetate. When thick oil/water emulsions were encountered, field samples were either treated with a commercial emulsion-breaking reagent (Emulstron®, Champion Technologies, Inc., Houston, TX, USA), water extracted, or

both methods were employed. For water extractions, hot water (65°C, 1 L) was added to oily samples and shaken overnight at room temperature. The water and emulsion breaker were each subject to organic extraction to rule out the introduction of contaminants from either method to the sample. Extracts were dried over anhydrous Na_2SO_4 and concentrated by rotary evaporation and under a stream of N_2 to a volume of 100 µL (10,000-fold concentration). All concentrated extracts were derivatized with *N,O*- bis(trimethylsilyl)trifluoroacetamide (BSTFA) (Sigma Aldrich, St. Louis, MO) prior to analysis. Derivatized extracts were analyzed with an Agilent 6890 model gas chromatograph (GC) coupled with a model 5973 mass spectrometer (MS) and separated with a DB-5ms capillary column (30 m x 0.25 mm i.d. x 1.0 µm film, J&W Scientific, Folsom, CA, USA). The starting oven temperature was 45°C (held 5 min), increased 4°C/min to 270°C and held for 10 min before mass spectral analysis (Duncan *et al.* 2009). All putative metabolite identifications were made by comparison to authentic derivatized standards (purchased commercially or synthesized) or with previously reported MS profiles of nearly 100 compounds implicated in anaerobic hydrocarbon metabolism belonging to the groups carboxylic acids, alkanoic acids, naphthalene putative monoaromatics, benzoate-associated metabolites, and benzylsuccinates. The chromatographic peaks of positively identified metabolites were analyzed with respect to their integrated area using the MS ChemStation (G1701DA D.01.00) Software (Agilent Technologies, Santa Clara CA). Standard curves were prepared with select representative compounds from the metabolites. For most compounds, the limit of detection for the GC-MS analysis was approximately 20 µM (Duncan *et al.* 2009).

A hierarchical clustering analysis based on Euclidean distances and average linkages was used to describe the relationship among all samples. The open source software packages CLUSTER 3.0 (http://bonsai.ims.utokyo.ac.jp/~mdehoon/software/cluster/software.htm) (de Hoon *et al.* 2004) and Java Tree View (http://jtreeview.sourceforge.net) (Saldanha 2004), were used to generate a dendrogram and heat map to visualize these relationships.

Shotgun Metagenome Sequencing

One sample from facility B (B37) and two samples from the younger facility A (A01 and A02) were selected for metagenomic sequencing. Overall, samples were chosen based on temperature and location. The pig envelope sample (B37) was selected because it was a surface-associated sample. Samples A01 and A02 were both obtained from separators at facility A. Sample A01 was taken from a primary separator, an apparatus used to physically separate the crude oil from produced fluids and it represented low temperature collection (35 °C). Sample A02 was taken from a secondary separator, which is downstream of the primary separator and heated (66 °C) to aid in further fluid separation. Small insert (300 bp) libraries were prepared with the Agilent SureSelect kit (Agilent Technologies, Irvine, CA) and sequencing was performed using the Illumina MiSeq v2 PE250 kit.

Sequences were uploaded to the MetaGenome Rapid Annotation Subsystems Technology (MG-RAST) server, which provided quality control and an automated annotation (Glass *et al.* 2010). The metagenome sequence data can be accessed within MG-RAST under the accessions 4537264.3 (K37), 4556546.3 (A01), and 4556547.3

(A02). Additionally, the raw sequence reads were assembled and scaffolded after quality trimming in CLC Genomics Workbench 7.0 (QIAGEN Inc., Boston, MA, USA). Scaffolded contiguous sequences (contigs) were binned by tetranucleotide frequency into genomes using Metawatt (Strous *et al.* 2012). After assembly and binning, metagenomes and binned genomes were annotated with Prokka (Seemann 2014). Prokka is an automated genome annotation pipeline, allowing for the determination of genetic features including RNAs, repeats, and coding DNA sequence (CDS) regions rapidly and accurately in both isolated genomes and metagenomes. AromaDeg, a novel database for phylogenomics of aerobic bacterial degradation of aromatics, was used to search the annotated CDS regions, translate to protein sequence, and identify oxygenases associated with aerobic degradation of hydrocarbons (Duarte *et al.* 2014). Anaerobic glycyl radical enzymes (GREs) were identified by annotation using Prokka, as well as by local BLAST searches for genes associated with anaerobic hydrocarbon degradation using a custom, curated list (Table S4, personal communication, Dr. Amy Callaghan). A phylogenetic tree was produced to more accurately identify succinate-synthase genes using the Maximum Likelihood method based on the Tamura-Nei model (Tamura and Nei 1993) within MEGA6 (Tamura *et al.* 2013). Sequences were aligned using MAFFT (Katoh and Standley 2013) and all positions containing gaps and missing data were eliminated. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using the Maximum Composite Likelihood (MCL) approach, and then selecting the topology with superior log likelihood value. Bins were assessed for completeness and quality using CheckM (Parks

et al. 2014). In addition, each bin was compared to the genome of a close relative and the contigs of each bin were reordered against the reference genome using CAR, a tool for contig assembly of prokaryotic draft genomes (Lu *et al.* 2014). Expectation-Maximization Iterative Reconstruction of Genes from the Environment (EMIRGE) was also used to reconstruct full-length ribosomal genes from the short read metagenome sequencing data (Miller *et al.* 2011, Miller 2013).

Results

Characterization of 16S rRNA Gene Libraries

Libraries of 16S rRNA genes were used to characterize the bacterial assemblages from samples across both oil fields (Figure 1). The 46,007 sequences obtained clustered into 160 OTUs (379 singletons) at a sequence similarity cutoff of 97 %. Comparisons of the community structure for each bacterial library (relative abundances of shared and unique OTUs) indicated that overall samples from the same field showed similar assemblages, but each field had outliers. Members of the phyla Synergistetes and Firmicutes were relatively more abundant in the high temperature samples obtained from the elder oilfield. These sequences represented the genera *Thermoanaerobacter, Thermacetogenium* and *Thermovirga*. *Thermoanaerobacter* dominated two samples from facility B (B17 and B18) and sample A02 (processing facility) from facility A (Figure 1). The exception among samples from facility B was B35 where the Gammaproteobacteria were most abundant within the community profile of this low temperature seawater from a primary separator. In this respect, the sample

was more similar to those obtained from the seawater-pressurized (and lower temperature) system at facility A. The microbiome of facility B was comprised of sulfur/thiosulfate reducing bacteria, sulfate-reducing archaea, syntrophic bacteria, peptide- and amino acid-fermenting bacteria and hydrogen-utilizing thermophilic methanogens (Duncan *et al.* 2009, Stevenson *et al.* 2011). In contrast, members of the Proteobacteria dominated the lower temperature samples from facility A. *Shewanella* and *Pseudomonas* accounted for 50% and 43% of sample A01 and A04, respectively. The Firmicutes (67%) made up the majority in the high temperature sample A02 and 50% were members of the genus *Thermoanaerobacter*. Sample A13, a surface water sample, was different from all other samples in that the Deltaproteobacteria (67%) were more numerous relative to the other organisms detected.

Characterization of Putative Metabolites

Metabolite profiling was used to detect nearly 100 polar organic compounds, some of which are associated with hydrocarbon biodegradation in the 50 samples collected from both fields (Figure 2). Virtually all metabolites detected at the facility A were also found to be present at facility B. The metabolites at facility A are a subset of the metabolites found at facility B. Metabolites were not assayed directly from the pig envelope sample, B37, however, samples B01-B06 were from the same pipeline that B37 was taken from and therefore metabolites from those samples were assumed to be representative of B37.

Of the many putative metabolites found in samples from both fields, benzoate is a central intermediate of both aerobic and anaerobic degradation of aromatic

hydrocarbons including benzene and toluene (Beller 2000, Pérez-Pantoja *et al.* 2010). Benzoate and the associated downstream metabolites (cyclohexane-1-carboxylate, cyclohex-1-ene carboxylate, glutarate, and pimelate) (Elshahed *et al.* 2001, Elshahed and McInerney 2001) were found in 40 of the 50 samples investigated. Toluic acids (*o*-, *m*-, and *p*-) were detected across both fields suggesting xylene degradation is occurring aerobically and/or anaerobically in both fields (Griebler *et al.* 2004, Koutinas *et al.* 2010, Wang *et al.* 2015). Also, cresols (*o*-, *m*-, and *p*-) were detected in nearly every sample indicating the aerobic degradation of toluene is a common process at both facilities (Shields *et al.* 1989, Kukor and Olsen 1991, Whited and Gibson 1991). Several putative naphthalene or alkylnaphthalene intermediates were found in 8 samples from facility B providing evidence for the anaerobic microbial biodegradation of the polycyclic aromatic hydrocarbon 2-methylnaphthalene (Safinowski and Meckenstock 2004). Alkylsuccinic acids were detected in 5 samples from facility A and throughout facility B. The most abundant alkylsuccinates were in the range of *n*-C_1-C_4, suggesting the anaerobic oxidation of methane, ethane, propane and butane. Alkanoic acids can be formed as intermediates during aerobic and/or anaerobic degradation of alkanes (Griebler *et al.* 2004, Agrawal and Gieg 2013, Callaghan 2013a, Ji *et al.* 2013). The most abundant alkanoic acids detected in both fields were butyrate and pentanoate. Acetate was detected in a few samples from both fields, as were several short-chained branched fatty acids, however higher molecular weight fatty acids (C_{14}-C_{22} in chain length) and dicarboxylic acids were rarely detected in either field.

Metagenomes

One metagenome from facility B (B37) and two from facility A (A01 and A02) were sequenced producing 934 million, 1.69 billion, and 597 million reads respectively (Table S5). Bacteria account for more than 90% of the metagenomes from facility A, whereas facility B was comprised of 87% bacteria and 12% archaea. There were 2-fold more identified protein features and identified functional categories in A01 than the other two metagenomes, which may be because the database is more complete for well characterized organisms like the Proteobacteria found in A01.

All genomes binned by Metawatt were near complete and lacking in contamination, with the exception of *Thermoanaerobacter* from B37 (63% complete) (Table S6). Reordered contigs also showed a high level of similarity to known genomes, with some appearing to be novel in their structure or distantly related to known sequenced relatives (Table S6). The bacterial 16S rRNA genes detected in the metagenomes matched well to the 16S rRNA genes from the amplicon sequencing results using EMIRGE (Figure S1). Although archaeal 16S genes were not found in the 16S library, one metagenome (B37) was found to contain 12%. For each sample, the binned genomes from the metagenome were also the dominant taxa in the 16S rRNA survey: [Sample, genus, (size of binned genome), (relative abundance)] B37, *Thermacetogenium*, (2.9Mb), (41.8%); B37, *Thermovirga,* (1.9Mb), (35.1%); B37, *Thermoanaerobacter,* (1.4Mb), (1.1%); B37, *Methanothermobacter,* (1.9Mb), (12.3%); A01, *Shewanella* (3.7Mb) (50.4%); A01, *Pseudomonas,* (3.5Mb), (19.3%); and A02, *Thermoanaerobacter,* (2.3Mb), (45%).

The metagenomes were investigated for specific genes coding for enzymes involved in both aerobic and anaerobic hydrocarbon degradation. In aerobic bacterial hydrocarbon degradation, aromatic compounds are activated by mono- or dioxygenases (Harayama *et al.* 1992) to eventually form dihydroxylated aromatic compounds as intermediates that then are subject to ring cleavage reactions (Butler and Mason 1996). Oxygenases were detected through a local BLAST of AromaDeg (Duarte *et al.* 2014) in the binned genomes of *Pseudomonas* (A01), *Thermacetogenium* (B37), *Thermoanaerobacter* (A02), and *Thermovirga* (B37). The possible substrates inferred for the oxygenases were benzoate and 2-chlorobenzoate (*Pseudomonas*), unknown belonging to the homoprotocatechate family (*Thermacetogenium* and *Thermoanaerobacter*), and 2-aminophenol (*Thermovirga*). Additionally, the *Pseudomonas* bin (A01) also contained a halobenzoate 1,2 dioxygenase. Phylogenetic trees were constructed to visualize the results (Figures S2-S6).

In anaerobic bacterial hydrocarbon degradation one of the most understood pathways for alkane activation and degradation is the addition of non-methane alkanes across the double bond of fumarate to form alkyl-substituted succinates (Callaghan 2013a). Alkane activation is catalyzed by the GRE alkylsuccinate synthase (ASS) (Callaghan *et al.* 2008). Alkylsuccinate synthases were searched for in the metagenomes because alkylsuccinate metabolites were present in many samples from both fields. However, very few sequences were identified with homology to known *ass* genes, and a phylogenetic tree was constructed to visualize the results (Figure 3). Additionally, sequences were aligned with other known GREs to look for the conserved glycine motif and cysteine residues (Figure 4).

Discussion

Oil Facility Characteristics and Practices

Although facility A and facility B are both located on Alaska's North Slope, the two fields had differing characteristics and management practices. Facility B has been producing oil longer than facility A and also produces larger percentages of water than oil (water cuts) when compared to facility A. The produced water and gas from wells at facility B is re-injected into the formation to maintain pressure for oil production, and because water conducts heat better than oil, temperatures of sampled processed fluids were between 35-60 °C. In contrast, seawater is commonly used to for secondary oil recovery purposes at facility A. New wells are drilled when the water cuts become too high and there is very little recycling of produced fluids. As a result, the temperatures of the sampled production fluids were between 24-47 °C, overall lower than those at facility B. These facility characteristics and facility management practices have enriched for the organisms that are capable of survival and proliferation under the ambient conditions.

Microbial Community Characterization

The bacterial assemblages from oilfield samples were comprised of only a few abundant taxa (Figure 1 and Figure S1). The particular taxa that were present varied depending on the facility and the sample. Members of the anaerobic, high-temperature Firmicutes and Synergistetes were numerically dominant in most of the samples from facility B. The one low temperature sample from facility B, B35, came from a pipeline

carrying treated seawater and, coincidentally, the microbial community was very similar to those from facility A. The organisms detected at facility A were mostly members of the Proteobacteria, which are common in seawater samples. The one high temperature sample from facility A (A02) came from a secondary separator that is heated to facilitate the separation of water from the oil. Coincidentally, the microbial community from A02 was more similar to those from the higher temperature samples from facility B. It can, therefore, be postulated that the elevated temperature of the separator enriches the community for high temperature organisms.

Two genomes belonging to the genera *Shewanella* and *Pseudomonas* were binned from the metagenome sequence from sample A01. Members of both genera have been isolated from hydrocarbon-contaminated environments previously (Martin-Gil *et al.* 2004, Gerdes *et al.* 2005). Both *Shewanella* and *Pseudomonas* binned genomes contained respiratory nitrate reductases, suggesting that both organisms were capable of respiring nitrate. Interestingly, genes associated with sulfate reduction were present within the annotated metagenomes, but not in any of the bins that represent the dominant taxa within each community.

A single genome, belonging to the genus *Thermoanaerobacter* was binned from the metagenome from A02 and the metagenome from B37. Species of *Thermoanaerobacter* have been isolated from subsurface oil wells and production waters (Fardeau *et al.* 1993, Grassia *et al.* 1996), and are known to oxidize H_2 by reducing thiosulfate to sulfide (Fardeau *et al.* 1993). Both *Thermoanaerobacter* binned genomes contained a complete anaerobic sulfite reductase, but there was no evidence of any genes associated with sulfate reduction. The genomic data suggested that the

Thermoanaerobacter taxa were capable of respiring sulfite to sulfide, but not capable of reducing sulfate, which is consistent with what is currently known (Fardeau *et al.* 1993).

A genome belonging to the genus *Thermacetogenium* was binned from the B37 sample. Sequences closely related to the 16S rRNA gene from members of the genus *Thermacetogenium* have been obtained from thermophilic crude oil degrading, methanogenic consortia enriched from samples obtained from different locations at facility B (Gieg *et al.* 2010). Members of the genus *Thermacetogenium* are known to grow acetogenically on several alcohols, methoxylated aromatic compounds, pyruvate, cysteine, formate and hydrogen/CO_2 (Hattori *et al.* 2000). Members of the *Thermoacetogenium* can also oxidize acetate to CO_2 syntrophically with a hydrogenotrophic methanogen (Hattori *et al.* 2000, Oehler *et al.* 2012). Although a member of the genus *Thermacetogenium* was described as being able to oxidize acetate with the reduction of sulfate or thiosulfate as the electron acceptor, Oehler, et al. (2012), could not reproduce those results and found little evidence for sulfate or thiosulfate reduction in the genome of this organism. Similarly, evidence of sulfate metabolism was not found in the assembled genome, but a subunit of a sulfite reductase was present.

Another binned genome from sample B37 belonged to the genus *Thermovirga.* *Thermovirga lienii* was isolated from an oil field in the North Sea (Dahle and Birkeland 2006), and is a thermophilic strict anaerobe, able to ferment proteinaceous substrates, some single amino acids, some organic acids, and can reduce cysteine and elemental sulfur to hydrogen sulfide (Dahle and Birkeland 2006).

The metagenomes and binned genomes were thoroughly searched for genes specifically associated with the degradation of hydrocarbons and other compound in crude oil. Genes involved in aerobic hydrocarbon degradation were identified in the binned genomes from both facilities (Figures S2-S6). This was a curious finding, as there is no precedent to suggest anaerobic organisms are producing mono- or dioxygenase enzymes for hydrocarbon degradation but these genes were also found in the published genomes of representative isolates. Interestingly metabolites associated with aerobic hydrocarbon degradation were present in these samples (discussed below).

No sequences of known anaerobic hydrocarbon degradation genes were found in the metagenomes. Sequences of potential glycyl radical enzymes (GREs), related to known alkylsuccinate synthases, benzylsuccinate synthases and other catalytic subunits of GREs were aligned. The dominant features of the catalytic subunits of GREs are the conserved glycine motif located near the C-terminus and the conserved cysteine residue(s) located in the middle of the polypeptide chain (Becker *et al.* 1999, Logan *et al.* 1999, Callaghan *et al.* 2008). The *Thermoacetogenium* binned genome contained a single gene annotated as a formate acetyltransferase (Figure 3), which is a type of GRE but not involved in hydrocarbon degradation. The binned genome of *Thermovirga* also contained a putative GRE sequence (Figure 3), as did the sequenced genome from the type strain *Thermoviga lienii*. The putative GRE sequences identified in this study contain the single cysteine residue like the known Ass and Bss sequences (Figure 4), whereas the putative pyruvate formate lyase (PFL) sequences that were identified contain two consecutive cysteine residues (Figure 4). All of the potential GREs, including PFLs, contain the conserved glycine motif (Figure 4), which suggests that

they are truly GREs that may be involved in the bacterial metabolisms including hydrocarbon degradation occurring in the oil facilities.

The *Thermoanaerobacter*, *Thermacetogenium* and *Thermovirga* make up between 60.9 - 93.4 % of the microbial communities found in the samples of facility B. Their abundance would implicate them in hydrocarbon degradation and corrosion. Each of these organisms is known to produce sulfides, but they do so using compounds other than sulfate. The sulfides can react with the iron of the pipelines to produce iron sulfides and exacerbate corrosion. Currently there is no direct evidence to link these organisms to hydrocarbon degradation although this study suggests they may contain genes capable of hydrocarbon activation. As such, they should be further investigated for their ability to degrade hydrocarbons.

Putative Metabolites of hydrocarbon degradation

Metabolites of aerobic and anaerobic hydrocarbon degradation were found at both oil fields. The metabolites present at facility A were largely a subset of those found at facility B, with the exception of *m*-methylbenzylsuccinic acid, *p*-methylbenzylsuccinic acid, and an alkane with 21 carbon atoms. The greater diversity of metabolites discovered at facility B could be due to the facility producing oil from multiple formations that are in differing, more advanced stages of biodegradation. Also, more aerobic hydrocarbon degradation genes and metabolites were found at A01 than the other two sites. One explanation could be that the seawater provides a source of oxygen, especially at the lower temperatures recorded at facility A sites. It is also possible that these genes were identified because the Proteobacteria have been much

better characterized, as have the genes associated with aerobic hydrocarbon degradation.

Conclusion

Temperature and management practices have influenced the microbial communities at each oil field. The younger facility A contains fluids at lower temperatures than the older facility B. Oxygen intrusion may be greater at facility A, with the seawater additions being the likely source. The conditions in facility A have selected for Proteobacteria with the genetic potential to degrade hydrocarbons through multiple aerobic pathways. The fluid samples at facility B were comprised of anaerobic thermophiles. Although fewer known hydrocarbon degradation genes were recognized in the metagenome from facility B, putative hydrocarbon metabolites were identified from samples across the facility. Several explanations for the paucity of identified hydrocarbon degradation genes identified from facility B could be that: 1) the microorganisms present in these samples are incapable of anaerobic hydrocarbon degradation, 2) the genes responsible for these activities in these systems are different from known genes associated with anaerobic hydrocarbon degradation, 3) there are undiscovered pathways for anaerobic hydrocarbon degradation for which the requisite genes are unknown, or 4) the organisms capable of these activities are present in the reservoirs while the water soluble metabolites are circulated throughout the production fluids.

A major goal of this study was to use shotgun metagenomic sequencing to link the abundant members of the observed microbial communities with detected putative metabolites of hydrocarbon degradation. The genes encoding multiple enzymes involved in aerobic degradation and the presence of cresols (*o*-, *m*-, and *p*-) across a wide number of samples in both facilities provided evidence of aerobic toluene degradation (Shields *et al.* 1989, Kukor and Olsen 1991, Whited and Gibson 1991). Additionally, phthalic acids (*o*-, *m*-, and *p*-) were also found at both fields indicating the aerobic degradation of polycyclic aromatic hydrocarbons (PAHs) (Eaton 2001). Known genes in the pathways of anaerobic hydrocarbon degradation, however, were not found. The putative GRE gene sequences that were found in binned genomes formed their own phylogenetic group when compared to known genes (Figure 3). These genes were closely related to *ass* and *bss* but did not group with either. The putative GRE sequences identified here also contain the signature cysteine residue in the middle of the polypeptide chain like *ass* and *bss*. These sequences, therefore, might represent a new class of GREs used for hydrocarbon activation in high temperature organisms (Figure 4). However, further investigation of these genes will be necessary to identify their functional role.

References

Agrawal, A. and L. M. Gieg (2013). In situ detection of anaerobic alkane metabolites in subsurface environments. *Front Microbiol* **4**: 140.

Bastin, E. S., F. E. Greer, C. A. Merritt and G. Moulton (1926). The presence of sulphate reducing bacteria in oil field waters. *Science* **63**: 21-24.

Becker, A., K. Fritz-Wolf, W. Kabsch, J. Knappe, S. Schultz and A. F. Volker Wagner (1999). Structure and mechanism of the glycyl radical enzyme pyruvate formate-lyase. *Nat Struct Mol Biol* **6**(10): 969-975.

Beller, H. R. (2000). Metabolic indicators for detecting in situ anaerobic alkylbenzene degradation. *Biodegradation* **11**(2-3): 125-139.

Beller, H. R., W. H. Ding and M. Reinhard (1995). Byproducts of anaerobic alkylbenzene metabolism useful as indicators of in situ bioremediation. *Environ Sci Technol* **29**(11): 2864-2870.

Butler, C. S. and J. R. Mason (1996). Structure-function Analysis of the Bacterial Aromatic Ring-hydroxylating Dioxygenases. *Advances in Microbial Physiology*. R. K. Poole, Academic Press. Volume 38**:** 47-84.

Callaghan, A. V. (2013a). Enzymes involved in the anaerobic oxidation of n-alkanes: from methane to long-chain paraffins. *Front Microbiol* **4**.

Callaghan, A. V. (2013b). Metabolomic investigations of anaerobic hydrocarbon-impacted environments. *Curr Opin Biotech* **24**(3): 506-515.

Callaghan, A. V., B. Wawrik, S. M. N. Chadhain, L. Y. Young and G. J. Zylstra (2008). Anaerobic alkane-degrading strain AK-01 contains two alkylsuccinate synthase genes. *Biochem Bioph Res Co* **366**(1): 142-148.

Caporaso, J. G., K. Bittinger, F. D. Bushman, T. Z. DeSantis, G. L. Andersen and R. Knight (2010a). PyNAST: a flexible tool for aligning sequences to a template alignment. *Bioinformatics* **26**(2): 266-267.

Caporaso, J. G., J. Kuczynski, J. Stombaugh, K. Bittinger, F. D. Bushman, E. K. Costello, . . . R. Knight (2010b). QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* **7**(5): 335-336.

Carman, G. J. and P. Hardwick (1983). Geology and Regional Setting of Kuparuk Oil-Field, Alaska. *Aapg Bulletin-American Association of Petroleum Geologists* **67**(6): 1014-1031.

Collett, T. S. (1993). Natural-Gas Hydrates of the Prudhoe Bay and Kuparuk River Area, North Slope, Alaska. *Aapg Bulletin-American Association of Petroleum Geologists* **77**(5): 793-812.

Connelly, T. L., S. E. Baer, J. T. Cooper, D. A. Bronk and B. Wawrik (2014). Urea Uptake and Carbon Fixation by Marine Pelagic Bacteria and Archaea during the Arctic Summer and Winter Seasons. *Appl Environ Microb* **80**(19): 6013-6022.

Dahle, H. and N. K. Birkeland (2006). *Thermovirga lienii* gen. nov., sp nov., a novel moderately thermophilic, anaerobic, amino-acid-degrading bacterium isolated from a North Sea oil well. *Int J Syst Evol Micr* **56**: 1539-1545.

de Hoon, M. J. L., S. Imoto, J. Nolan and S. Miyano (2004). Open source clustering software. *Bioinformatics* **20**(9): 1453-1454.

Duarte, M., R. Jauregui, R. Vilchez-Vargas, H. Junca and D. H. Pieper (2014). AromaDeg, a novel database for phylogenomics of aerobic bacterial degradation of aromatics. *Database* **2014**: bau118.

Duncan, K. E., L. M. Gieg, V. A. Parisi, R. S. Tanner, S. G. Tringe, J. Bristow and J. M. Suflita (2009). Biocorrosive Thermophilic Microbial Communities in Alaskan North Slope Oil Facilities. *Environ Sci Technol* **43**(20): 7977-7984.

Eaton, R. W. (2001). Plasmid-encoded phthalate catabolic pathway in *Arthrobacter keyseri* 12B. *J Bacteriol* **183**(12): 3689-3703.

Edgar, R. C. (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**(19): 2460-2461.

Elshahed, M. S., V. K. Bhupathiraju, N. Q. Wofford, M. A. Nanny and M. J. McInerney (2001). Metabolism of benzoate, cyclohex-1-ene carboxylate, and cyclohexane carboxylate by *Syntrophus aciditrophicus* strain SB in syntrophic association with H2-using microorganisms. *Appl Environ Microb* **67**(4): 1728-1738.

Elshahed, M. S. and M. J. McInerney (2001). Benzoate Fermentation by the Anaerobic Bacterium *Syntrophus aciditrophicus* in the Absence of Hydrogen-Using Microorganisms. *Appl Environ Microb* **67**(12): 5520-5525.

Fardeau, M. L., J. L. Cayol, M. Magot and B. Ollivier (1993). H-2 Oxidation in the Presence of Thiosulfate, by a *Thermoanaerobacter* Strain Isolated from an Oil-Producing Well. *FEMS Microbiol Lett* **113**(3): 327-332.

Gerdes, B., R. Brinkmeyer, G. Dieckmann and E. Helmke (2005). Influence of crude oil on changes of bacterial communities in Arctic sea-ice. *FEMS Microbiol Ecol* **53**(1): 129-139.

Gieg, L. M., I. A. Davidova, K. E. Duncan and J. M. Suflita (2010). Methanogenesis, sulfate reduction and crude oil biodegradation in hot Alaskan oilfields. *Environ Microbiol* **12**(11): 3074-3086.

Gieg, L. M. and J. M. Suflita (2002). Detection of anaerobic metabolites of saturated and aromatic hydrocarbons in petroleum-contaminated aquifers. *Environ Sci Technol* **36**(17): 3755-3762.

Gingrich, D., D. Knock and R. Masters (2001). Geophysical interpretation methods applied at Alpine Oil Field: North Slope, Alaska. *The Leading Edge* **20**(7): 730-738.

Glass, E. M., J. Wilkening, A. Wilke, D. Antonopoulos and F. Meyer (2010). Using the metagenomics RAST server (MG-RAST) for analyzing shotgun metagenomes. *Cold Spring Harb Protoc* **2010**(1): pdb prot5368.

Grassia, G. S., K. M. McLean, P. Glenat, J. Bauld and A. J. Sheehy (1996). A systematic survey for thermophilic fermentative bacteria and archaea in high temperature petroleum reservoirs. *FEMS Microbiol Ecol* **21**(1): 47-58.

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**(2): 617-631.

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

Hattori, S., Y. Kamagata, S. Hanada and H. Shoun (2000). *Thermacetogenium phaeum* gen. nov., sp. nov., a strictly anaerobic, thermophilic, syntrophic acetate-oxidizing bacterium. *Int J Syst Evol Micr* **50 Pt 4**: 1601-1609.

Head, I. M., D. M. Jones and S. R. Larter (2003). Biological activity in the deep subsurface and the origin of heavy oil. *Nature* **426**(6964): 344-352.

Holba, A. G., L. Wright, R. Levinson, B. Huizinga and M. Scheihing (2004). Effects and impact of early-stage anaerobic biodegradation on Kuparuk River Field, Alaska. *Geological Society, London, Special Publications* **237**(1): 53-88.

Houseknecht, D. and K. Bird (2006). Oil and gas resources of the Arctic Alaska petroleum province. *US Geological Survey Professional Paper*.

Ji, Y., G. Mao, Y. Wang and M. Bartlam (2013). Structural insights into diversity and n-alkane biodegradation mechanisms of alkane hydroxylases. *Front Microbiol* **4**.

Katoh, K. and D. M. Standley (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* **30**(4): 772-780.

Koutinas, M., M.-C. Lam, A. Kiparissides, R. Silva-Rocha, M. Godinho, A. G. Livingston, . . . A. Mantalaris (2010). The regulatory logic of m-xylene biodegradation by *Pseudomonas putida* mt-2 exposed by dynamic modelling of the principal node Ps/Pr of the TOL plasmid. *Environ Microbiol* **12**(6): 1705-1718.

Kukor, J. J. and R. H. Olsen (1991). Genetic organization and regulation of a meta cleavage pathway for catechols produced from catabolism of toluene, benzene, phenol, and cresols by *Pseudomonas pickettii* PKO1. *J Bacteriol* **173**(15): 4587-4594.

Larter, S., A. Wilhelins, I. Head, M. Koopmans, A. Aplin, R. Di Primio, . . . N. Telnaes (2003). The controls on the composition of biodegraded oils in the deep subsurface - part 1: biodegradation rates in petroleum reservoirs. *Organic Geochemistry* **34**(4): 601-613.

Logan, D. T., J. Andersson, B.-M. Sjöberg and P. Nordlund (1999). A glycyl radical site in the crystal structure of a class III ribonucleotide reductase. *Science* **283**(5407): 1499-1504.

Lu, C. L., K. T. Chen, S. Y. Huang and H. T. Chiu (2014). CAR: contig assembly of prokaryotic draft genomes using rearrangements. *BMC Bioinformatics* **15**(1): 381.

Magot, M. (2005). Indigenous Microbial Communities in Oil Fields. *Petroleum Microbiology*. B. Ollivier and M. Magot. Washington, DC, ASM Press.

Martin-Gil, J., M. C. Ramos-Sanchez and F. J. Martin-Gil (2004). *Shewanella putrefaciens* in a fuel-in-water emulsion from the Prestige oil spill. *Anton Leeuw Int J G* **86**(3): 283-285.

Masterson, W. and J. Eggert (1992). Kuparuk River Field-USA, North Slope, Alaska. *AAPG Stratigraphic Traps III*: 257-284.

Masterson, W. D., L. I. P. Dzou, A. G. Holba, A. L. Fincannon and L. Ellis (2001). Evidence for biodegradation and evaporative fractionation in West Sak, Kuparuk and Prudhoe Bay field areas, North Slope, Alaska. *Organic Geochemistry* **32**(3): 411-441.

Miller, C. S. (2013). Assembling full-length rRNA genes from short-read metagenomic sequence datasets using EMIRGE. *Methods Enzymol* **531**: 333-352.

Miller, C. S., B. J. Baker, B. C. Thomas, S. W. Singer and J. F. Banfield (2011). EMIRGE: reconstruction of full-length ribosomal genes from microbial community short read sequencing data. *Genome Biol* **12**(5): R44.

Oehler, D., A. Poehlein, A. Leimbach, N. Muller, R. Daniel, G. Gottschalk and B. Schink (2012). Genome-guided analysis of physiological and morphological traits of the fermentative acetate oxidizer *Thermacetogenium phaeum*. *Bmc Genomics* **13**: 723.

Parks, D. H., M. Imelfort, C. T. Skennerton, P. Hugenholtz and G. W. Tyson (2014). CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *PeerJ PrePrints* **2**: e554v551.

Pérez-Pantoja, D., B. González and D. Pieper (2010). Aerobic degradation of aromatic hydrocarbons. *Handbook of Hydrocarbon and Lipid Microbiology*, Springer**:** 799-837.

Peters, K. E., L. B. Magoon, K. J. Bird, Z. C. Valin and M. A. Keller (2006). North Slope, Alaska: Source rock distribution, richness, thermal maturity, and petroleum charge. *Aapg Bulletin* **90**(2): 261-292.

Reeder, J. and R. Knight (2010). Rapid denoising of pyrosequencing amplicon data: exploiting the rank-abundance distribution. *Nat Methods* **7**(9): 668-669.

Roling, W. F. M., I. M. Head and S. R. Larter (2003). The microbiology of hydrocarbon degradation in subsurface petroleum reservoirs: perspectives and prospects. *Research in Microbiology* **154**(5): 321-328.

Safinowski, M. and R. U. Meckenstock (2004). Enzymatic reactions in anaerobic 2-methylnaphthalene degradation by the sulphate-reducing enrichment culture N 47. *FEMS Microbiol Lett* **240**(1): 99-104.

Saldanha, A. J. (2004). Java Treeview--extensible visualization of microarray data. *Bioinformatics* **20**(17): 3246-3248.

Sambrook, J. (2001). *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor, NY, Cold Spring Harbor Laboratory Press.

Seemann, T. (2014). Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30**(14): 2068-2069.

Shields, M. S., S. O. Montgomery, P. J. Chapman, S. M. Cuskey and P. H. Pritchard (1989). Novel Pathway of Toluene Catabolism in the Trichloroethylene-Degrading Bacterium-G4. *Appl Environ Microb* **55**(6): 1624-1629.

Stevenson, B. S., H. S. Drilling, P. A. Lawson, K. E. Duncan, V. A. Parisi and J. M. Suflita (2011). Microbial communities in bulk fluids and biofilms of an oil facility have similar composition but different structure. *Environ Microbiol* **13**(4): 13.

Strous, M., B. Kraft, R. Bisdorf and H. E. Tegetmeyer (2012). The binning of metagenomic contigs for microbial physiology of mixed cultures. *Front Microbiol* **3**: 410.

Suflita, J. M., T. J. Phelps and B. Little (2008). *Carbon Dioxide Corrosion and Acetate : A Hypothesis on the Influence of Microorganisms*. Houston, TX, USA, NACE International.

Tamura, K. and M. Nei (1993). Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Mol Biol Evol* **10**(3): 512-526.

Tamura, K., G. Stecher, D. Peterson, A. Filipski and S. Kumar (2013). MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol Biol Evol* **30**(12): 2725-2729.

Verma, M. K. and K. J. Bird (2005). Role of reservoir engineering in the assessment of undiscovered oil and gas resources in the National Petroleum Reserve, Alaska. *Aapg Bulletin* **89**(8): 1091-1111.

Wang, Q., G. M. Garrity, J. M. Tiedje and J. R. Cole (2007). Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl Environ Microb* **73**(16): 5261-5267.

Wang, X., Q. Wang, S. Li and W. Li (2015). Degradation pathway and kinetic analysis for p-xylene removal by a novel *Pandoraea* sp. strain WL1 and its application in a biotrickling filter. *J Hazard Mater* **288**(0): 17-24.

Whited, G. M. and D. T. Gibson (1991). Toluene-4-monooxygenase, a three-component enzyme system that catalyzes the oxidation of toluene to p-cresol in *Pseudomonas mendocina* KR1. *J Bacteriol* **173**(9): 3010-3016.

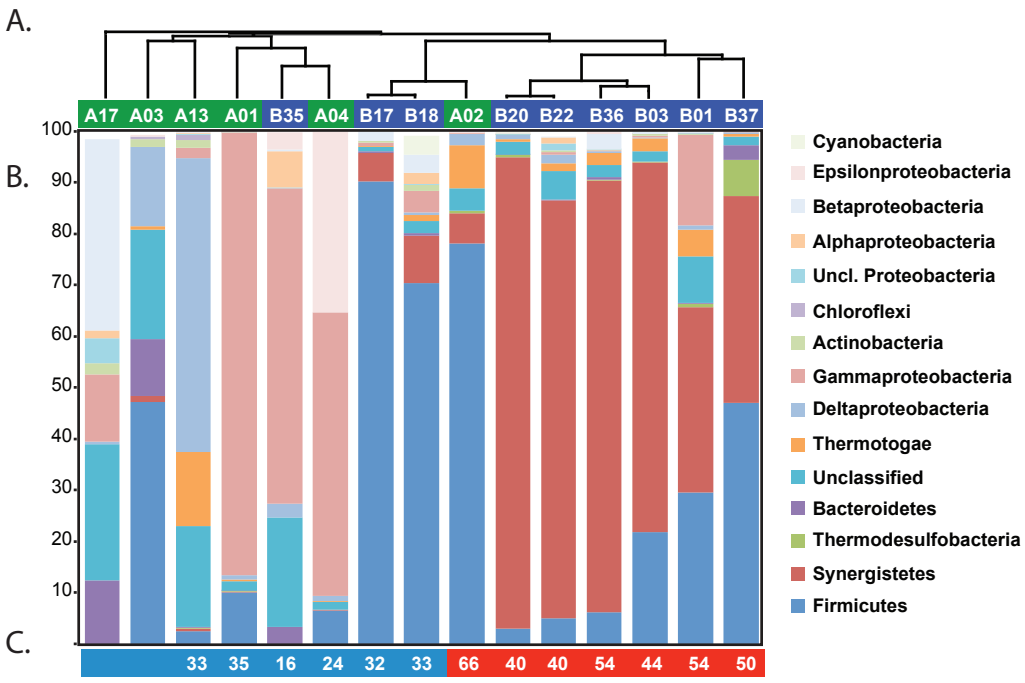


Figure 1. Comparison of 16S rRNA gene libraries. A dendogram (A) groups the libraries based on comparisons of community structure θ_{YC} values, (facility A is in green and facility B is in blue), relative abundance (B) is shown as the % of total represented Phyla (class level for Proteobacteria), and (C) temperatures (in °C) at the time of sampling are given (temperatures <40 °C are in blue, ≥40 °C are in red).

Figure 2. A dendogram (A) showing the relatedness of samples based on presence/absence of metabolites, and a heatmap (B) of putative metabolites detected (black boxes) or not detected (white boxes) are shown.

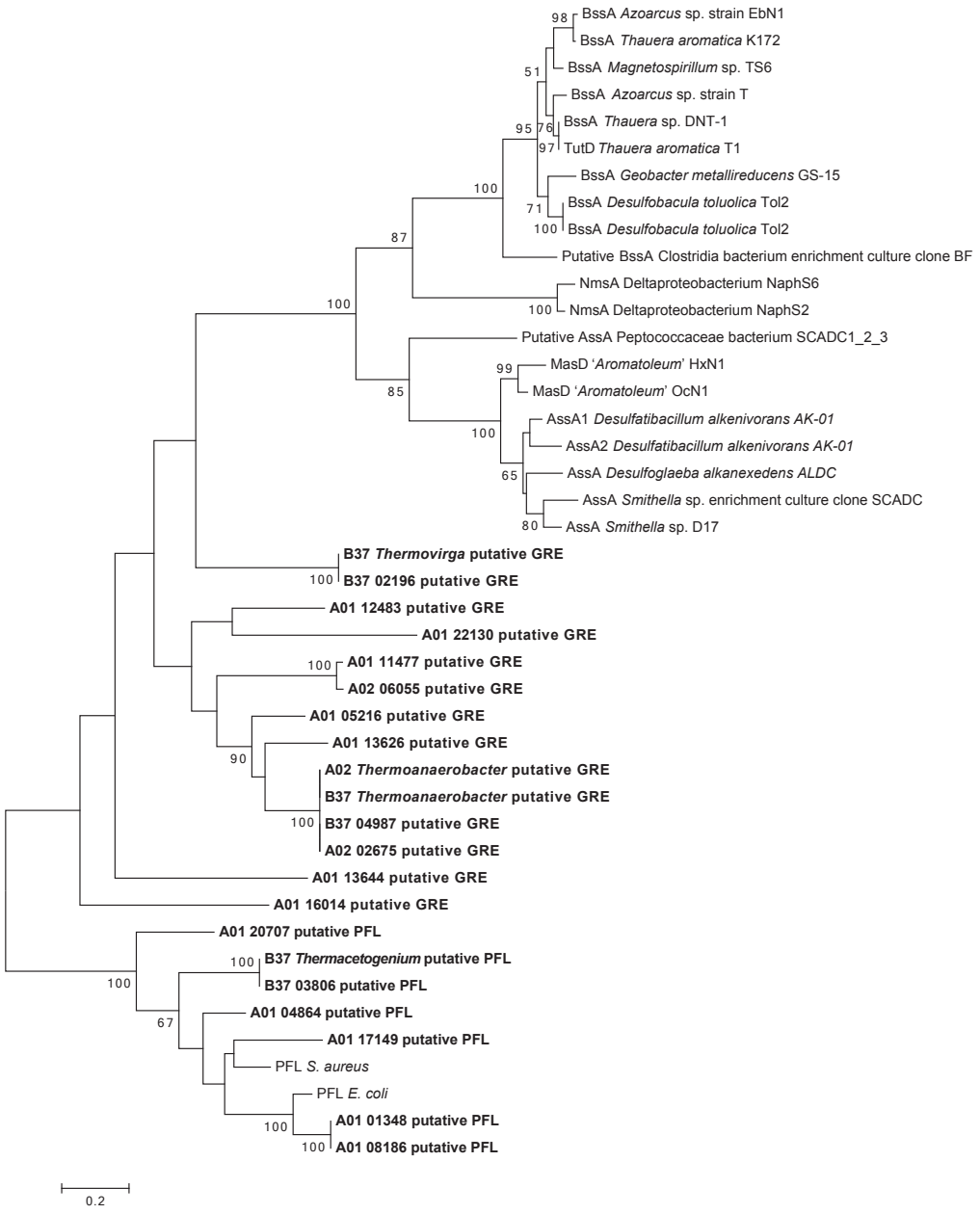
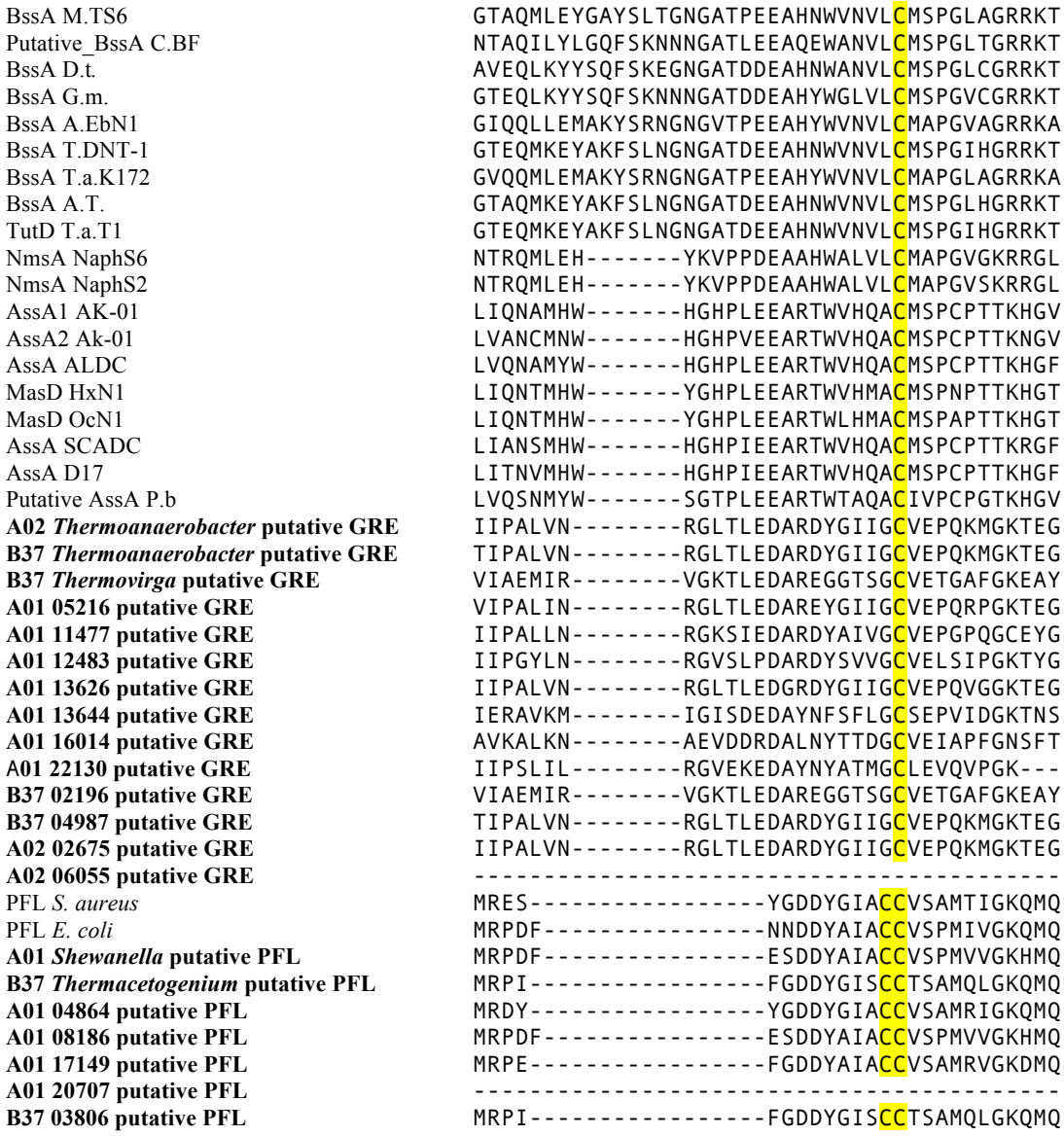


Figure 3. A neighbor-joining tree that shows the phylogenetic relationship of AssA and BssA protein sequences from reference strains and those from binned genomes (in bold). Branch lengths are scaled to the number of substitutions per site and bootstrap values over 50% (n = 5000) are given next to each conserved node. The tree was rooted with representative pyruvate formate lyase genes (PFL) as the outgroup. Abbreviations include Bss (benzylsuccinare synthase), Tut ("toluene-utilizing" benzylsuccinate synthase), Nms (naphthylmethylsuccinate synthase), and Ass (alkulsuccinate synthase).

A.


BssA M.TS6 GTAQMLEYGAYSLTGNGATPEEAHNWVNVLCMSPGLAGRRKT
Putative_BssA C.BF NTAQILYLGQFSKNNNGATLEEAQEWANVLCMSPGLTGRRKT
BssA D.t. AVEQLKYYSQFSKEGNGATDDEAHNWANVLCMSPGLCGRRKT
BssA G.m. GTEQLKYYSQFSKNNNGATDDEAHYWGLVLCMSPGVCGRRKT
BssA A.EbN1 GIQQLLEMAKYSRNGNGVTPEEAHYWVNVLCMAPGVAGRRKA
BssA T.DNT-1 GTEQMKEYAKFSLNGNGATDEEAHNWVNVLCMSPGIHGRRKT
BssA T.a.K172 GVQQMLEMAKYSRNGNGATPEEAHYWVNVLCMAPGLAGRRKA
BssA A.T. GTAQMKEYAKFSLNGNGATDEEAHNWVNVLCMSPGLHGRRKT
TutD T.a.T1 GTEQMKEYAKFSLNGNGATDEEAHNWVNVLCMSPGIHGRRKT
NmsA NaphS6 NTRQMLEH-------YKVPPDEAAHWALVLCMAPGVGKRRGL
NmsA NaphS2 NTRQMLEH-------YKVPPDEAAHWALVLCMAPGVSKRRGL
AssA1 AK-01 LIQNAMHW-------HGHPLEEARTWVHQACMSPCPTTKHGV
AssA2 Ak-01 LVANCMNW-------HGHPVEEARTWVHQACMSPCPTTKNGV
AssA ALDC LVQNAMYW-------HGHPLEEARTWVHQACMSPCPTTKHGF
MasD HxN1 LIQNTMHW-------YGHPLEEARTWVHMACMSPNPTTKHGT
MasD OcN1 LIQNTMHW-------YGHPLEEARTWLHMACMSPAPTTKHGT
AssA SCADC LIANSMHW-------HGHPIEEARTWVHQACMSPCPTTKRGF
AssA D17 LITNVMHW-------HGHPIEEARTWVHQACMSPCPTTKHGF
Putative AssA P.b LVQSNMYW-------SGTPLEEARTWTAQACIVPCPGTKHGV
A02 Thermoanaerobacter putative GRE IIPALVN--------RGLTLEDARDYGIIGCVEPQKMGKTEG
B37 Thermoanaerobacter putative GRE TIPALVN--------RGLTLEDARDYGIIGCVEPQKMGKTEG
B37 Thermovirga putative GRE VIAEMIR--------VGKTLEDAREGGTSGCVETGAFGKEAY
A01 05216 putative GRE VIPALIN--------RGLTLEDAREYGIIGCVEPQRPGKTEG
A01 11477 putative GRE IIPALLN--------RGKSIEDARDYAIVGCVEPGPQGCEYG
A01 12483 putative GRE IIPGYLN--------RGVSLPDARDYSVVGCVELSIPGKTYG
A01 13626 putative GRE IIPALVN--------RGLTLEDGRDYGIIGCVEPQVGGKTEG
A01 13644 putative GRE IERAVKM--------IGISDEDAYNFSFLGCSEPVIDGKTNS
A01 16014 putative GRE AVKALKN--------AEVDDRDALNYTTDGCVEIAPFGNSFT
A01 22130 putative GRE IIPSLIL--------RGVEKEDAYNYATMGCLEVQVPGK---
B37 02196 putative GRE VIAEMIR--------VGKTLEDAREGGTSGCVETGAFGKEAY
B37 04987 putative GRE TIPALVN--------RGLTLEDARDYGIIGCVEPQKMGKTEG
A02 02675 putative GRE IIPALVN--------RGLTLEDARDYGIIGCVEPQKMGKTEG
A02 06055 putative GRE ----------
PFL S. aureus MRES----------YGDDYGIACCVSAMTIGKQMQ
PFL E. coli MRPDF----------NNDDYAIACCVSPMIVGKQMQ
A01 Shewanella putative PFL MRPDF----------ESDDYAIACCVSPMVVGKHMQ
B37 Thermacetogenium putative PFL MRPI----------FGDDYGISCCTSAMQLGKQMQ
A01 04864 putative PFL MRDY----------YGDDYGIACCVSAMRIGKQMQ
A01 08186 putative PFL MRPDF----------ESDDYAIACCVSPMVVGKHMQ
A01 17149 putative PFL MRPE----------FGDDYAIACCVSAMRVGKDMQ
A01 20707 putative PFL ----------
B37 03806 putative PFL MRPI----------FGDDYGISCCTSAMQLGKQMQ

B.

BssA M.TS6	VQFNVVSTEEMRAAQREPEKHHDLIVRVSGYSARFVDIPTYG
Putative BssA C.BF	VQFNVVSAEQMQAAQKEPEKHGDLIVRVSGYSARFTDISKYA
BssA D.t.	IQFNVVSTAEMKAAQKEPEKHQDLIVRVSGFSSRFVDIPTYG
BssA G.m.	VQFNCVSTAEMKAAQKEPEKHQDLIVRVSGFSARFVDIPTYG
BssA A.EbN1	VQFNVVSTEEMKAAQREPEKHQDLIVRVSGFSARFVDIPTYG
BssA T.DNT-1	VQFNVVSTDEMRAAQREPEKHHDLIVRVSGYSARFVDIPTYG
BssA T.a.K172	VQFNVVSTEEMKAAQREPEKHQDLIVRVSGFSARFVDIPTYG
BssA A.T.	VQFNVVSTDEMRAAQREPEKHSDLIVRVSGYSARFVDLPTYG
TutD T.a.T1	VQFNVVSTDEMRAAQREPEKHHDLIVRVSGYSARFVDIPTYG
NmsA NaphS6	VQFNVVDTKDMLEAQKEPEKWQSMIVRIAGYSARFVSLPRNA
NmsA NaphS2	VQFNVVETKDMLEAQKEPEKWESLIVRIAGYSARFVSLPKNA
AssA1 AK-01	VQFNMVSDETLRAAQKDPEKYSEVIVRVAGYSAHFVDISRKT
AssA2 AK-01	IQFNMVSDKTLRAAQKDPEKYQEVIVRVAGYSAHFVDISRKT
AssA ALDC	VQFNMVDDATLRAAQKDPEKYQEVIVRVAGYSAHFVDISRKT
MasD HxN1	IQFNVISDKVLRAAQNDPEGYQEVIVRVAGYSAHFIDISRKT
MasD OcN1	IQFNMVSDKVLRSAQKDPEGYQEVIVRVAGYSAHFIDISRKT
AssA SCADC	IQFNCVSDETLRSAQREPEKYQEVIVRVAGYSAHFVDISRKT
AssA D17	IQFNVIDDTTLRSAQREPEKYQEVIVRVAGYSAHFVDISRKT
Putative AssA P.b	VQFNMVDNETLYAAQKEPEKYSELMVRVAGYSAHFTGLNKKT
A02 *Thermoanaerobacter* putative GRE	VQFNVVSKEMLLDAQKNPEKYRTLVVRVAGYSAYFTALDKAI
B37 *Thermoanaerobacter* putative GRE	VQFNVVSKEMLLDAQKNPEKYRTLVVRVAGYSAYFTALDKAI
B37 *Thermovirga* putative GRE	IQFNVVDEATLRDAQKHPEKYRDLIVRVAGYSDYFCDLGKEL
A01 05216 putative GRE	VQFNVIDRNTLLAAQKEPEKYRDLVVRVAGYSAQFVSLDKSV
A01 11477 putative GRE	VQYNVVDSKRLRDAQKNPDKYKDLVVRVAGYSAFFVGLDPDL
A01 12483 putative GRE	IQFNVVSADTLRAAQEDPAKYRNLVIRVAGYSAIFVELNKSI
A01 13626 putative GRE	VQYNVIGRETLVDAQKHPDKYKNLVVRVAGYSAHFIYLDKSL
A01 13644 putative GRE	TQFNVVTKEDLIRAQEHPDEYRSLVVRVAGYSAFFTVLGKSV
A01 16014 putative GRE	VQFNILKEDLLRKAQQEPEKYRWLLVRVAGWSAYFVELSRPV
A01 22130 putative GRE	IQFNVVDEATLRDAQKHPEKYRDLIVRVAGYSDYFCDLGKEL
B37 04987 putative GRE	VQFNVVSKEMLLDAQKNPEKYRTLVVRVAGYSAYFTALDKAI
A02 02675 putative GRE	VQFNVVSKEMLLDAQKNPEKYRTLVVRVAGYSAYFTALDKAI
A02 06055 putative GRE	VQYNVVDSKRLRDAQKNPDKYKDLVVRVAGYSAFFVGLDPDL
PFL *S. aureus*	LNINVFNRETLIDAMEHPEEYPQLTIRVSGYAVNFIKLTREQ
PFL *E. coli*	LNVNVMNREMLLDAMENPEKYPQLTIRVSGYAVRFNSLTKEQ
A01 *Shewanella* putative PFL	LNVNVMNREMLEDAVVNPDKYPQLTIRVSGYAVRFNSLTPEQ
B37 *Thermacetogenium* putative PFL	INVNVFDRELLVDAMHHPEKYPQLTIRVSGYAVNFVKLSREQ
A01 04864 putative PFL	INVNVFEKETLLDAMEHPEKYPQLTIRVSGYAVNFIKLTRKQ
A01 08186 putative PFL	LNVNVMNREMLEDAVVNPDKYPQLTIRVSGYAVRFNSLTPEQ
A01 17149 putative PFL	LNVNVLNRETLKEAYENPDMHPNLTIRVSGYAVNFHKLSKQQ
A01 20707 putative PFL	LNVNVLSRELLEDAMEHPDKYPDLTIRVSGYAVHFNRLTRKQ
B37 03806 putative PFL	INVNVFDRELLVDAMHHPEKYPQLTIRVSGYAVNFVKLSREQ

Figure 4. Partial amino acid sequence alignments of BssA, AssA, and other predicted catalytic subunits of glycyl radical enzymes. (A) The alignment of the region containing the conserved cysteine residue located in the middle of the polypeptide. (B) The alignment of the conserved glycine motif located near the C-terminus of the polypeptide. Conserved regions are highlighted in yellow. Sequences from the binned genomes or metagenomes are designated in bold. The abbreviated names are as follows: M.TS6, *Magnetospirillum* sp. TS6 (BAD42366); C.BF, Clostridia bacterium enrichment culture clone BF (ADJ93876.1); D.t., *Desulfobacula toluolica* Tol2 (YP_006759014.1); G.m., *Geobacter metallireducens* GS-15 (YP_006720507.1); A.EbN1, *Azoarcus* sp. strain EbN1 (YP_158060.1); T.DNT-1, *Thauera* sp. DNT-1 (BAC05501); T.a.K172, *Thauera aromatica* K172 (CAA05052.1); A.T, *Azoarcus* sp. strain T (AAK50372.1); T.a.T1, *Thauera aromatica* T1 (AAC38454.1); NaphS6, Deltaproteobacterium NaphS6 (CAO72222.1); NaphS2, Deltaproteobacterium NaphS2 (CAO72219.1); AK-01, *Desulfatibacillum alkenivorans* AK-01 (ACL03428.1 and ACL03892.1); ALDC, *Desulfoglaeba alkanexedens* ALDC (ADJ51097.1); HxN1, '*Aromatoleum'* HxN1 (CAO03074.1); OcN1, '*Aromatoleum'* OcN1 *(*CBK27727.1); SCADC, *Smithella* sp. enrichment culture clone SCADC (AHI85732.1); D17, *Smithella* sp. D17 (KFZ44314.1); P.b, Peptococcaceae bacterium SCADC1_2_3 (KFI38250); *S. aureus, Staphylococcus aureus* subsp. *aureus* Mu50, formate acetyltransferase (Q99WZ7.1); *E. coli, Escherichia coli* K-12, formate acetyltransferase (P09373.2).

Appendix A. Supplementary Information

Chapter 2

Table S1. PCR and DGGE primers used in this study

Primer name	Position[a]	Sequence (5'-3')	Target	Reference
27F	8-27	AGA GTT TGA TCC TGG CTC AG	Bacteria	(Wilson et al., 1990; Nakatsu and Marsh 2007)
1492R	1492-1510	TAC CTT GTT ACG ACT T	Bacteria, Archaea	(Wilson et al., 1990; Nakatsu and Marsh 2007)
1392R		ACG GGC GGT GTG TRC'	Universal[c]	(Nakatsu and Marsh 2007)
Arc333F	334-348	TCC AGG CCC TAC GGG	Archaea	(Reysenbach and Pace 1995)
GM5F[b]	341-357	CCT ACG GGA GGC AGC AG	Bacteria	(Santegoeds et al., 1999)
907R	907-928	CCG TCA ATT CCT TTG AGT TT	Universal[c]	(Santegoeds et al., 1999)
Arc333F-GC[d]	334-348	TCC AGG CCC TAC GGG	Archaea	(Reysenbach and Pace 1995)
958R		YCC GGC GTT GAM TCC AAT T	Archaea	(Reysenbach and Pace 1995)
TiA-8nt-27F[e]	8-27	CCA TCT CAT CCC TGC GTG TCT CCG ACT CAG NNN NNN NN**C** **A**AG AGT TTG ATC CTG GCT CAG	Bacteria	This study
TiB-338R[e]		CCT ATC CCC TGT GTG CCT TGG CAG TCT CAG **CA**T GCT GCC TCC CGT AGG AGT	Bacteria	This study

[a] Position in the 16S rRNA of *E. coli* (Brosius et al., 1981)

[b] This primer has the following GC-clamp at the 5' end: 5'-CGCCCGCCGCGCCCCGCGCCCGTCCCGCCGCCCCCGCCCG-3'

[c] All organisms (bacteria, brchaea, and eukaryotes)

[d] This primer has the following GC-clamp at the 5' end: 5'-CGCCCGCCGCGCGCGGCGGGCGGG GCGGGGCGACGCGGGG-3'

[e] Life Sciences 454 Titanium "A" and "B" primers are underlined, N's designate location of unique barcode (see Table S2), and spacer nucleotides are in bold.

Table S2. List of 8 nucleotide barcodes used for parallel pyrosequencing of multiple libraries

Barcode (5'-3')	Library
CAACAGCT	PIGa
CAACACGT	SEPa
CAACACCA	SEPb
CAACCAAC	SPWa
CAACAGGA	SPWb
CAACCTAG	INJ
CAACCATG	SEA

Supplementary References

Brosius, J., Dull, T.J., Sleeter, D.D. and Noller, H.F. (1981) Gene organization and primary structure of a ribosomal RNA operon from *Escherichia coli*. *J Mol Biol* **148**: 107-127.

Nakatsu, C.H. and Marsh, T.L. (2007) Analysis of microbial communities with danaturing gradient gel electrophoresis and terminal restriction fragment polymorphism. "In" Methods for general and molecular microbiology. Reddy, C.A., Beveridge, T.J., Breznak, J.A.et al. Washington, D.C., ASM Press**:** 909-923.

Reysenbach, A.-L. and Pace, N.R. (1995) Reliable amplification of hyperthermophilic Archaeal 16S rRNA genes by the polymerase chain reaction. Archaea: Thermophiles: A laboratory manual. pp. 101-105.

Santegoeds, C.M., Damgaard, L.R., Hesselink, C., Zopfi, J., Lens, P., Muyzer, G., et al. (1999) Distribution of sulfate-reducing and methanogenic bacteria in anaerobic aggregates determined by microsensor and molecular analyses. *Appl Environ Microbiol* **65**: 4618-4629.

Wilson, K.H., Blitchington, R.B. and Greene, R.C. (1990) Amplification of bacterial 16S ribosomal DNA with polymerase chain reaction. *J Clin Microbiol* **28**: 1942-1946.

Chapter 3

Table S1. Sample and site characteristics

Sample	Type of sample	Temperature (°C)
A01	Processing: inlet separator	35.00
A02	Processing: low-pressure separator	65.56
A03	"Dead leg"	35.00
A04	Processing facility	23.72
A05	Processing facility	23.72
A06	Processing facility	23.33
A07	Processing facility	16.67
A08	Producer	36.67
A09	Producer	40.56
A10	Producer	43.33
A11	Producer	41.11
A12	Producer	37.22
A13	Producer	32.78
A14	Producer	24.44
A15	Producer	36.11
A16	Producer	46.67
A17	Surface lake water	ND
B01	Processing: primary separator	54.44
B02	Gas-lifted producer	30.00
B03	Water injector	43.89
B04	Water injector	50.06
B05	Gas-lifted producer	39.44
B06	Gas-lifted producer	48.67
B07	Processing: primary separator	54.44
B08	Producer	54.44
B09	Producer	51.11
B10	Producer	44.44
B11	Producer	52.22
B12	Producer	61.67
B13	Producer	59.44
B14	Producer	53.89
B15	Producer	51.11
B16	Producer	21.11
B17	Producer	31.67
B18	Producer	33.33
B19	Producer	22.22
B20	Producer	40.00
B21	Producer	41.67
B22	Producer	40.00
B23	Producer	50.06
B24	Producer	40.11

B25	Primary separator	51.67
B26	Producer	51.72
B27	Water injector	38.89
B28	Water injector	49.44
B29	Producer	40.00
B30	Producer	56.06
B31	Water injector	57.56
B32	Producer	38.33
B33	Producer	15.33
B34	Producer	50.61
B35	Primary separator sea water	10.00
B36	Primary separator	54.44
B37	Pig envelope	50.00

Table S2. PCR primers used in this study

Primer name	Position [a]	Sequence (5'-3')	Target	Reference
27F	8-27	AGA GTT TGA TCC TGG CTC AG	Bacteria	(Wilson *et al.* 1990)
TiA-8nt-27F*	8-27	CCA TCT CAT CCC TGC GTG TCT CCG ACT CAG NNN NNN NN**C** **A**AG AGT TTG ATC CTG GCT CAG	Bacteria	(Stevenson *et al.* 2011)
TiB-338R*		CCT ATC CCC TGT GTG CCT TGG CAG TCT CAG **C**A**T** GCT GCC TCC CGT AGG AGT	Bacteria	(Stevenson *et al.* 2011)

[a] Position in the 16S rRNA of *E. coli* (Brosius *et al.* 1981)

* Life Sciences 454 Titanium "A" and "B" primers are underlined, N's designate location of unique barcode (see Table S4) and spacer nucleotides are in bold.

Supplementary References

Stevenson, B. S., H. S. Drilling, P. A. Lawson, K. E. Duncan, V. A. Parisi and J. M. Suflita (2011). Microbial communities in bulk fluids and biofilms of an oil facility have similar composition but different structure. *Environ Microbiol* **13**(4): 13.

Wilson, K.H., Blitchington, R.B. and Greene, R.C. (1990) Amplification of bacterial 16S ribosomal DNA with polymerase chain reaction. *J Clin Microbiol* **28**: 1942-1946.

Table S3. List of 8 nucleotide barcodes used for parallel pyrosequencing of multiple libraries

Barcode (5'-3')	Library
CAACACGT	B01
CAACCTAG	B03
CAACGATC	B17
CAACGTAC	B18
CAACGTTG	B20
CAACTCCT	B22
CAACCATG	B35
CAACCAAC	B36
CAACAGCT	B37
CAAGCATC	A01
CAAGCAAG	A02
CAAGGATG	A03
CAAGGAAC	A04
CAAGGTAG	A13
CAAGCTTG	A17

Table S4. Proteins involved in anaerobic hydrocarbon degradation. Sequences were obtained from cultured and uncultured strains. NOTE: The sequences below reflect the sequences available at the time of analysis. This custom, curated list of genes was obtained from Dr. Amy Callaghan.

Gene	Organism	Accession Number	Protein Sequence
Ethylbenzene dehydrogenase, alpha subunit	*'Aromatoleum' aromaticum* EbN1	YP_158333.1	MTRDEMISVEPEAAELQDQHRRDFLKRSGAAVLSL SLSSLATGVVPGFLKDAQAGTKAPGYASWEDIYRK EWKWDKVNWGSHLNICWPQGSCKFYVYVRNGIV WREEQAAQTPACNVDYVDYNPLGCQKGSAFNNN LYGDERVKYPLKRVGKRGEGKWKRVSWDEAAGD IADSIIDSFEAQGSDGFILDAPHVHAGSIAWGAGFR MTYLMDGVSPDINVDIGDTYMGAFHTFGKMHMG YSADNLLDAELIFMTCSNWSYTYPSSYHFLSEARY KGAEVVVIAPDFNPTTPAADLHVPVRVGSDAAFW LGLSQVMIDEKLFDRQFVCEQTDLPLLVRMDTGKF LSAEDVDGGEAKQFYFFDEKAGSVRKASRGTLKL DFMPALEGTFSARLKNGKTIQVRTVFEGLREHLKD YTPEKASAKCGVPVSLIRELGRKVAKKRTCSYIGFS SAKSYHGDLMERSLFLAMALSGNWGKPGTGAFA WAYSDDNMVYLGVMSKPTAQGGMDELHQMAEG FNKRTLEADPTSTDEMGNIEFMKVVTSAVGLVPPA MWLYYHVGYDQLWNNKAWTDPALKKSFGAYLD EAKEKGWWTNDHIRPAPDKTPQVYMLLSQNPMRR KRSGAKMFPDVLFPKLKMIFALETRMSSSAMYADI VLPCAWYYEKHEMTTPCSGNPFFTFVDRSVAPPGE CREEWDAIALILKKVGERAAARGLTEFNDHNGRK RRYDELYKKFTMDGHLLTNEDCLKEMVDINRAVG VFAKDYTYEKFKKEGQTRFLSMGTGVSRYAHANE VDVTKPIYPMRWHFDDKKVFPTHTRRAQFYLDHD WYLEAGESLPTHKDTPMVGGDHPFKITGGHPRVSI HSTHLTNSHLSRLHRGQPVVHMNSKDAAELGIKD GDMAKLFNDFADCEIMVRTAPNVQPKQCIVYFWD AHQYKGWKPYDILLIGMPKPLHLAGGYEQFRYYF MNGSPAPVTDRGVRVSIKKA
Ethylbenzene dehydrogenase, beta subunit	*'Aromatoleum' aromaticum* EbN1	YP_158332.1	MTYVQDGNKSELRKAKRQLVTVIDLNKCLGCQTC TVACKNIWTKRPGTEHMRWNNVTTYPGKGYPRD YERKGGGFLRGEPQPGVLPTLIDSGDDFQFNHKEV FYEGKGQTVHFHPTSKSTGKDPAWGYNWDEDQG GGKWPNPFFFYLARMCNHCTNPACLAACPTGAIY KREDNGIVLVDQERCKGHRHCVEACPYKAIYFNPV SQTSEKCILCYPRIEKGIANACNRQCPGRVRAFGYL DDTTSHVHKLVKKWKVALPLHAEYGTGPNIYYVP PMGARGFGEDGEITDKTRIPLDVLEGLFGPEVKRVL AVLHTERENMRAGRGSELMDLLISKKWSDRFGGF TNDPLTQS
Ethylbenzene dehydrogenase, gamma subunit	*'Aromatoleum' aromaticum* EbN1	CAI07430.1	MKAKRVPGGKELLLDLDAPIWAGAESTTFEMFPTP LVMVKEVSPFLALSEGHGVIKRLDVAALHNGSMIA LRLKWASEKHDKIVDLNSFVDGVGAMFPVARGAQ AVTMGATGRPVNAWYWKANANEPMEIVAEGFSA VRRMKDKAGSDLKAVAQHRNGEWNVILCRSMAT GDGLAKLQAGGSSKIAFAVWSGGNAERSGRKSYS GEFVDFEILK
Ethylbenzene dehydrogenase, delta subunit	*'Aromatoleum' aromaticum* EbN1	YP_158330.1	MSEKLEVELSQHNRATAARSRCYRILSGMFGYPVD SLPEDFMLQARRELRESLDGLPFTVPAIESLANIGVS RGVSSENLAVTYSRIFDNCSGRPAVSLHEKDYSRD ETKFIWEELIRFYEHFGLNYDLGACKEWPDHICIQL EFLHYLTFLEAGAPVDMVDIYATAEADFLEKHVAE WIPKFSEKLRSVAEGTPYADLARVVAQFVEGDAEF NRRRRTIQ

Ethylbenzene dehydrogenase, alpha subunit	*Azoarcus* sp. EB1	AAK76387.1	MTRDEMISVEPEAAELQDQDRRDFLKRSGAAVLSL SLSSLATGVVPGFLKDAQAGTKAPGYASWEDIYRK EWKWDKVNWGSHLNICWPQGSCKFYVYVRNGIV WREEQAAQTPACNVDYVDYNPLGCQKGSAFNNN LYGDERVKYPLKRVGKRGEGKWKRVSWDEAAGD IADSIIDSFEAQGSDGFILDAPHVHAGSIAWGAGFR MTYLMDGVSPDINVDIGDTYMGAFHTFGKMHMG YSADNLLDAELIFMTCSNWSYTYPSSYHFLSEARY KGAEVVVIAPDFNPTTPAADLHVPVRVGSDAAFW LGLSQVMIDEKLFDRQFVCEQTDLPLLVRMDTGKF LSAEDVDGGEAKQFYFFDEKAGSVRKASRATLKL DFMPALEGTFSARLKNGKTIQVRTVFEGLREHLKD YTPEKASAKCGVPVSLIRELGRKVAKKRTCSYIGFS SAKSYHGDLMERSLFLAMALSGNWGKPGTGAFA WAYSDDNMVYLGVMSKPTAQGGMDELHQMAEG FNKRTLEADPTSTDEMGNIEFMKVLTSAVGLVPPA MWLYYHVGYDQLWNNKAWTDPALKKSFGAYLD EAKEKGWWTNDHIRPAPDKTPQVYMLLSQNPMRR KRSGAKMFPDVLFPKLKMIFALETRMSSSAMYADI VLPCAWYYEKHEMTTPCSGNPFFTFVDRSVAPPGE CREEWDAIALILKKVGERAAARGLTEFNDHNGRK RRYDELYKKFTMDGHLLTNEDCLKEMVDINRAVG VFAKDYTYEKFKKEGQTRFLSMGTGASRYAHANE VDVTKPIYPMRWHFDDKKVFPTHTRRAQFYLDHD WYLEAGESLPTHKDTPMVGGDHPFKITGGHPRVSI HSTHLTNSHLSRLHRGQPVVHMNSKDAAELGIKD GDMAKLFNDFADCEIMVRTAPNVQPKQCIVYFWD AHQYKGWKPYDILLIGMPKPLHLAGGYEQFRYYF MNGSPAPVTDRGVRVSIK
Ethylbenzene dehydrogenase, beta subunit	*Azoarcus* sp. EB1	AAK76388.1	MTYVQDGNKSELRKAKRQLVTVIDLNKCLGCQTC TVACKNIWTKRPGTEHMRWNNVTTYPGKGYPRD YERKGGGFLRGEPQPGVLPTLIDSGDDFQFNHKEV FYEGKGQTVHFHPTSKSTGKDPAWGYNWDEDQG GGKWPNPFFFYLARMCNHCTNPACLAACPTGAIY KREDNGIVLVDQERCKGHRHCVEACPYKAIYFNPV SQTSEKCILCYPRIEKGIANACNRQCPGRVRAFGYL DDTTSHVHKLVKKWKVALPLHAEYGTGPNIYYVP PMGARGFGEDGEITDKTRIPLDVLEGLFGPEVKRVL AVLHTERENMRAGRGSELMDLLISKKWSDRFGGL TNDPLTQS
Ethylbenzene dehydrogenase, gamma subunit	*Azoarcus* sp. EB1	AAK76389.1	MKAKRVPGGKELLLDLDAPIWAGAESTTFEMFPTP LVMVKEVSPFLALSEGHGVIKRLDVAALHNGSMIA LRLKWASEKHDKIVDLNSFVDGVGAMFPVARGAQ AVTMGATGRPVNAWYWKANANEPMEIVAEGFSA VRRMKDKAGSDLKAVAQHRNGEWNVILCRSMAT GDGLAKLQAGGSSKIAFAVWSGGNAERSGRKSYS GEFVDFEILK
Ethylbenzene dehydrogenase, delta subunit	*Azoarcus* sp.EB1		Not sure if it has the delta subunit
Putative ethylbenzene dehydrogenase, subunit A (Annotated as molybdopterin oxidoredutase)	*Desulfococcus oleovorans Hxd3*	YP_0015280 81	MKEVKISRRTFLKGTSATVALLSLNSLGFLGGNTIA NATEKIFEDWKYAGWENLHREEWTWDKVTYGTH LVDCYPGNCLWRVYSKDGVVFREEQAAKYPVIDP SGPDFNPRGCQKGASYSLQMYNPDRLKYPMKQVG GRGSGKWKRVSWDQCLAEIAEGIVDGLEAQGPESI IFESGPGNGGYVHVMAVHRLMVSLGATVLDLDSTI GDFNRGIYETFGKFMFMDSVDGWYFGKLLLIWHM NPVYTRIPSYHFISEARYNGAEIISIAPDYNPSCMHA DEYIPVEMGSDAALGLAVCQVLMNKKWVDYPFV KEQSDLPLLVRKDTDRFLSAADIEKGARDDQFCFW DSKN

			NKVVKAPLETLKLPCDPALEGVYKATLLDGKTVE VEPVFNKLKALLDSEYTPEQASEMCRTNPDTIRRM AEKCYKASGGIQVMVGWNSPKYYHGDLIERAMCL VLTLTGSLGKKGCGIRGWNESLFEGAFTQIFKQKIG LMNLTRQLPLMRRVWESFKKEDPTISDEMASIKQE RVIDRNMVNYTIPAFLYYNHSNYKKAWDNKAWH CPTMKREFKEYYDEAVNSGWWDGYVKPEKDQTP TVYCWMGTNPARKNRGWEKNILGSLWDKYKTIFG FETRWTTTLLYGDYALPCAGFYEKLDTRFPTPHVP WLTLTDEAVKPIGECKTEWEICLMLAKKIEEAAKK RGKTNFVPRTPVQEEYMPFLNQFSVRDAKGNVEIS KLHDWQLMGSSGKDFDEIMEDALQSSVMLGNLPP GTNLKYMREHGIVRFVDVSIFDPVTMNLATDIKPD EPIIPLTWHTGEKKVPYPTYNRRIQFYIDHPWFLEA KEELPVHKDNPKIGGNYPLRLTSGHQRWSIHSIWV SNEQLLRTHQGRPFMFMSPEDAEKRGVEDGDLVR VFNDFASFKIHVKVTPAARPEKGAAKPGQVIIYHA WEPFMFPEWKSYDIAIPGMIKWLDLVNNYGHLHY WRWNWCAQPIDRGISVNVEKA
Putative ethylbenzene dehydrogenase, subunit B (Annotated 4Fe-4S ferredoxin iron-sulfur binding domain-containing protein)	*Desulfococcus oleovorans Hxd3*	YP_0015280 83	MAVDEKVLTDREKELAAAFEETKEMLRKSNRQIG MVMDLNKCIGCQLCSMACKTLWTSKEGREYMW WNKVNTMPGKGSPKDWEKMGGGYKVKFGGKVM EPVQGKHPTRKEFGDMWDYNWNDVVNNKAGAV HLQAKNQTDGSVPDWSMNWDEDQGAGQYPNGFY FYLPRICNHCAYPACVDACPRNAIYKRDEDGVVLI DQDRCKGYRFCLEACPYKVIYFNFLTDTSMKCIFC YPRLDEKVANACARQCTARVRWIGYTEEPESIIYK LTKVWKVALPLHAEYGTNPNVFYVPPLAPHRLAP DGKIDPSKPRIPREYLRYLFGPEVDASLDIMQKELD KVRAGGRSELMEILKAKEWKSMFGEFTKDPAVMD RKPPKGVTFVRDGIITK
Putative ethylbenzene dehydrogenase, subunit C (annotated as gamma-subunit of ethylbenzene dehydrogenase)	*Desulfococcus oleovorans* Hxd3	YP_0015280 82	MEKVICKKIATATKELLNCDSSVWQSGKAVLETAA TPLANQPSPYIKGVYDETKIGAVKKITIKAVHNGKD IVFYCEWESGKPNKEIGDINVFPDGVALLFPFKDID KTPINEMGTQDYPTNAWYWRPDFEEKPKNQVSHG LSTSLYTEKSSISSYSRWANGKWCVVMARPMKAT QPGEETVDLMPGKAIGFGIAVWEGSNGERGGIKAF SKEWRELVIEA
Putative ethylbenzene dehydrogenase, subunit D (annotated as delta-subunit of ethylbenzene dehydrogenase	*Desulfococcus oleovorans* Hxd3	YP_0015280 84	MNQAAVDMANGYDDAQMAAFRSAELTSDADKQ AAARSDMYALMADLFRYPDKEFQAFVRNGELRDA LVGITENLPFACALSDGETEKLQFSPLLDEDGVEAG FIRLFEAGPGDPPCPLIEGKYVKDSNRRAIFEDLIRF YNHFGLSYAEGAHEDRPDHVIYEMEFMHYICFLTL RAGQQEKSIEDLLLAQRDFLKHHLLKWAGKLAER VAEVVNDIPEDYAGTFYTNVARLLARFIEADYAYL NEAQTR
Phenylphosphat e carboxylase, alpha subunit	*'Aromatoleum' aromaticum* EbN1	CAI07883.1	MAKISAPKNNREFIEACVRSGDAVRIRQEVDWDNE AGAIVRRVCELGEAAPFMENIKDYPGFSYFGAPLST YRRMAISLGMDPASTLPQIGAEYLKRTNSEPIEPVVI DRRDAPCKENILLGDDVDLTKLPVPLVHDGDGGR YVGTWHAVITRHPVRGDVNWGMYRQMMWDGRT MSGAVFPFSDLGKALTDYYLPRGEGCPFATAIGLSP LAAMAACAPSPIPEPELCGMLSGEAVRLVKCETND LEVPADAEIIIEGMILPDYKVEEGPFGEYTGYRTSPR DFRVTFRVDAITYRNNATMTISNMGVPQDEGQLLR

			SFSLGLELEKLLKSQGIPVTGVYMHPRSTHHMMIV GVKPTYAGIAMQIAQLAFGSKLGPWFHMVMVVD DQTDIFNWDEVYHAFCTRCNPERGIHVFKNTTGTA LYPHATPHDRKYSIGSQVLFDCLWPVDWDKTNDV PSVVSFKKVYPQDIQDKVTNSWTDYGFKPVK
Phenylphosphate carboxylase, beta subunit	*'Aromatoleum' aromaticum* EbN1	CAI07885.1	MDLRYFINQCAEAHELKRVTTEVDWNLEISHVSKL TEEKKGPALLFESIKGYDTPVFTGAFATTKRLAIML GLPHHLTLCESAQQWMKKTITSEGLIKAKEVKDGP VLENVLSGGKVDLNMFPVPKFFPLDGGRYIGTMVS VVLRDPETGEVNLGTYRMQMLDDKRCGVQILPGK RGERIMKKYAKLGKKMPAAAIIGCDPLIFMAGTLM HKGASDFDITGTVRGQQAEFLMAPLTGLPVPAGAE IVLEGEIDPNAFLPEGPFAEYTGYYTDELHKPIPKPV LEVQQILHRNNPILWATGQGRPVTDVHMLLAFTRT ATLWTELEQMRIPGVQSVCVLPESTGRFWSVVSLK QAYPGHSRQVADAVMGSNTGSYGMKGIITVDEDI AADDLQRVFWALSCRYDPMRGTELIKRGRSTPLDP ALDPDGDKLTTSRILMDACIPYEWKQKPVEARMD EEMLAKIRARWHEYGID
Phenylphosphate carboxylase, gamma subunit	*'Aromatoleum' aromaticum* EbN1	CAI07884.1	MEQAKNIKLVILDVDGVMTDGRIVINDEGIESRNFD IKDGMGVIVLQLCGIDVAIITSKKSGAVRHRAEELK IKRFHEGIKKKTEPYAQMLEEMNISDAEVCYVGDD LVDLSMMKRVGLAVAVGDAVADVKEAATYVTTA RGGHGAVREVAELILKAQGKWDAVLAKIH
Phenylphosphate carboxylase, delta subunit	*'Aromatoleum' aromaticum* EbN1	CAI07882.1	MNQWEVFVMDLAELPEGTELELSVRTLNPGLKKY TYQRVKAELSNALDKFPDRLQVRFGRGQLCSQQFS IRIIEQVQRMPAKYL
Putative anaerobic benzene carboxylase, AbcD, partial (Phenylphosphate carboxylase gamma subunit)	Clostridia bacterium enrichment culture clone BF	ADJ94001.1	MYDKPSNRLEVHKGFTFGPDPDELKKQEGREVYD VIILSDSIDELTSQQNQEITLAIRLLTPKDRKMRYCT RTVSAMISSDPTKYRDRLHVRFQRGLLHPKPWSIEI VKELEGLMSLEADHTLARS
Putative anaerobic benzene carboxylase AbcA, partial (3-octaprenyl-4-hydroxybenzoate carboxy-lyase)	Clostridia bacterium enrichment culture clone BF	ADJ94002.1	MYRDLRGYIDALEKNNQLVRTNKEVDWDLEAGAI ARLGCQLNSKAMLMENIKDYPGFRYLGGVFTNWQ RAAIAIGLNPNASRREIQDTYRERLDHPIKPIIIEDAP CQENVITDEDVNLFDLPAPMVHDGDGGRYPTTWS FVATPSYNTDWINWGMYRQMIHNERTLGGLVLPS QDIGIQYYGGWEPAGEKMPFAAVIGGEVLTTAVA AVPYGVGRSEVDFAGGLLLEPVELVKCVSVDLLVP ANAEIIIEGYVSPKERCYEGPFGEYTGYRTSPRAPR VAYHVTAITYRDNAIMPFANMGVPTDECDMIANT LWRADLMRVLEENGIPVLDVSMPTEMVMHSAFISI KKPYSSIPNKVAKLIFGEKNTGLFLHQIYVCDEDVD VQDLSQVMHVFATKWHPKRGTHFFDKTVGVPLHP FLDMEERTWSKASKVCFDCTWPSEWSPTAEVPVR SSFWDIYPERVIEKVCSTWKDFGQDQDMLSALADY NRIGRWPGSDTFNKPVVIKKK
BssA	*Magnetospirillum* sp. TS6	BAD42366	MTANVLEYRGKVLNFTPENLAEVNIPAEELHEHLQ NPSTERTKRLKARCRWKHASAGEFIEKGVTAGIER MRHITEAHKASVGKPEVIRRALGLKNILEKGTLCL QPDEFIVGYHAEDPNMFPLYPELSYMAVQDYLLSD YAPQPTSEAAEINDYWKPYSLQAKCSPYFDPTDLG RMFQVSSMEAPSFASGYNSIVPPYETVLEDGLLAR VKLAKSHIEQAQAEMGAFPWNGSKGLEWIEKIDN

			WQAMIIACEAVISWARRHARMCKIVAEKFEADPK RRAELLEIADICQRVPAEPCKGLKDAFQAKWFTYLI CHAIDRYASATAHKEDTLLWPYYKASVVDKSFQP MTYENAVEWVEMERLKISEHGAGKSRAYREIFPGS NDLFILTVGGTLGDGSDACTDMTDAILEGAKRIRTT EPSIVFRYSKKSRAKTLRWVFECVRDGLGYPSIKHD EIGTAQMLEYGAYSLTGNGATPEEAHNWVNVLCM SPGLAGRRKTQKTRSEGGGSIFPAKLLEITLNDGYD WSYADMQLGPKTGQATQFKTFEDLWEAFRKQYQ YAIALCIRAKDVSRYFEGRILQMPFVSAIDDGCMEL GVDANVLSEQPNGWHNPITTIVAGNSLVAIKKLIFD DKKYTMEQLVEALKANWEGHEEMRLDFKNAPKW GNDDAYCDEIIKNFYEDIVGGEMSKITNYSGGPVRP TGQAVGLYMEVGSRTGPTPDGRLGGEAADDGGIS PYMGTDKKGPTAVLRSVSKVQKNQKANLLNQRLS VPIMRSAHGFDIWHAYMNTWHDLNIDHVQFNVVS TEEMRAAQREPEKHHDLIVRVSGYSARFVDIPTYG QNTIIARHEQDFSANDLEFLNCDL
Putative benzylsuccinate synthase BssA	Clostridia bacterium enrichment culture clone BF	ADJ93876.1	MSVLHADLEYKGKVLKFSLENTAEGDIPQEEIHQH LARPSTPRTRRLKENCRWKHTAAGEFVDEGIKAG MERMRLITDAHKKSADKSYVMRRAAGLDNILRNC TILLMEDELIVGYHAENPKWIPMFPELSYMQIQDYL QSNYAPEPADEAREIAEYWKPYSLQTACEPYFTQE ELKVAYHPALIESPGFSNGYNSIVPPYETVLEDGLL QRITLAEEHIKEAREQLAELPWNAEKLELLDKIDN WQAMVVADKAVIAWARRHARLCRIVAENFETDS KRKEELLMIADICQRVPAEPCKGLRDAMQAKWFT YLICHAIERYASGYAQLEDKMLWPYYEASVVEKN FQPMTHADAVELFECERLKISNHGAGKSRQYREIFP GANDLFILSVGGLNRHGEDACNDCTDAILEAARNI RTTEPSIVFRWHPLCRQKTKRLVFECIRDGLGYPSI KHQALNTAQILYLGQFSKNNNGATLEEAQEWANV LCMSPGLTGRRKTQKTRSEGGSSLFPEKILELTLFD GFDWSYANDQYGLHTGDPRDFKTFEELWEAYRKQ IQYFMSLAIRAKDTSRMMEGRLLQMPFVSSIDDGC MEYGIDAMVLSEQPNPWHNPLCNVVALNSLVAIK KLIYDDKKYTMDQLIEALQNNWEGYEKIHKDFLA APKWGNDDSYCDAIGKAFFEDIISGEMGKIINYSGA PVMPGGQAVGLYMTSGSRMGPTPDGRYGGDPGG DGGISPYQGTDHNGPTAVLKSVSKVNAATQKAELL NQRLSPSIMRSKHGFKIWLSYMNTWYDLNIDHVQF NVVSAEQMQAAQKEPEKHGDLIVRVSGYSARFTDI SKYAQDTIIARTEQDFGVKDLEFLNTEI
BssA	*Desulfobacula toluolica* Tol2	YP_0067590 14.1	MATIAEAVEYRGKIIEFPLEHQEENDIPDERLHEHL ARPSTQRTKRLKARCRWKHASAGEFVEKGVTSGIQ RMRYITEAHKKSRGKPEVIRRALGLANILNKYTLV LQEDEFIIGYHAEDPNMFPLYPELAYMAVADYLVS DYAPQPAEEAKEIMEYWKPYSMQGKCEGYFDPED LMRMYQVSTMEAPGFATGYNSIVPPYETILEDGLL KRIEMAEANIKAAKEDLKKTPWDATKGLKWIPMI DNWEAMIIADKAVIAWARRHARLAKFVAENIETDP KRKEELLEIADINQRIPAEPCKGLKDAFQAKWYTFL ICHAIDRYASGFAQTEDTMLYPYYKASVIDKTFQP MTHSEAVEMVEMERLKISEHGAGKSRAYREIFPGS NDLFILSIGGTKPDGSDASNEMTNAILEATRNIRTTE PSIVFKYSDKSNDKTKRLVFECIRDGLGYPSIKHNEI AVEQLKYYSQFSKEGNGATDDEAHNWANVLCMS PGLCGRRKTQKTRSEGGGSIFPAKILEVTLNDGYD WSYADMQLGPQTGKAEDFKTYEELYEAFKVQYQ YAVSLVIKCKDTMRYFEGKFLQMPFASSLDDGCM ELGRDGCELSEQPNGWHNPITTIVAANSMVAIKKLI YDDKKYTMKQLLDALKANWEGYEEMHKDFKAAP

			KWGNDNEYADATIKDFYEDIIGGEMGRITNYSGGP VLPVGQAVGLYMEIGSRTGPTPDGRLGGEAADDG GISPYMGTDHKGPTAVLKSVSKVQKNQKANLLNQ RLSVPIMRSKHGFDIWNAYMKTWHDLKIDHIQFNV VSTAEMKAAQKEPEKHQDLIVRVSGFSSRFVDIPTY GQNTIIARNEQQFGAEDFEYLNLDI
BssA	*Geobacter metallireducens* GS-15	YP_006720507.1	MSTVAASIKYNDKVVDFPLANQEENAIADEVLHEN LARPTTERTKRLKARCRWKHAAAGEFVDAEVRAG IERMRFITEAHKASAGQPEVIRRALGLANILNKSTL VLQQDEFIIGYHAEDPEMFPLYPELAYMAVQDYLM SDYSPQPKEEAAEINEYWKKYSMQAKGERYFTQE ELLQMYQVSTMEAPGFATGYNSISPPYETVLQDGL LKRIEMAHEKIEHAKREMQKIPWDATTGLDWIAKI DVWKAMIIADEAVINWARRHARLAKIVAENFETNP ARKEELLEIAEISHRVPAEPCKGLKDAFQAKWYTY LICHAIDRYASGYAQKEDEMLEPYYNISVKEKSFQP MTHTDVVEMVEMERLKISEHGAGKSRAYREIFPGS NDLFILTIGGTKPGYVDACSDMTDAILEGARNIRTT EPSIVFRWHPVGREKTKRLVFECIRDGLGYPSIKHD VIGTEQLKYYSQFSKNNNGATDDEAHYWGLVLCM SPGVCGRRKTHKTRSEGGGSIFPAKMMEIVLADGF DWSYSGMQLGPHTGDPTTFKTFEQLWEAFRSQYA YATSKVIRAKDIMRYYESKFLQMPFVSSIDDGCME LGIDSMELSEQPNGWHNPITTVVAANSLVAIKKLIY DEKKYTMAQLVTALRANWHGYEDMRQDFLNAPK WGNDDEYADTIIKKFYEDIIGGEMAKITNYSGGPVL PVGQAVGLYMEIGSRTGPTPDGRFGGDAGDDGGIS PYMGTDKKGPTAVLKSVSKVQKNQKANLLNQRLS VPIMRSVHGFTIWKSYMDAWEKLNIDHVQFNCVS TAEMKAAQKEPEKHQDLIVRVSGFSARFVDIPTYG QNTIIARNEQAFGADDLEYLNTQL
BssA	*'Aromatoleum' aromaticum* EbN1	YP_158060.1	MTGAQTMEYKGKVLQFTPENPAEADIPADELHEH LQNPSTERTRRLKARCRWKHASAGEFCEKGVTAGI ERMRLLTESHWDTRGKPEPIRRALGLKNILDKCTL VLQPDEFIVGYHAEDPNMFPLYPELSYMAVQDYLK SKYSPQPAKEAQEIVDYWKPFSLQARCEPYFDPVD LRRGYQVSTIEGPVFASGYNSVIPPYETILEDGLLAR IALAEKNIEHARAEMEKFPWNAPTGLEWIDKIDNW EAMVIACKAVIAWARRHARLCKIVAERFETDPKRK AELLEIADICQRVPAEPARGLKDAMQAKWFTFLIC HAIERYASGYAQKEDSLLWPYYKASVIDKTFQPME HKDAVELIEMERLKVSEHGAGKSRAYREIFPGSND LFILTLGGTNGDGSDACNDMTDAILEAAKRIRTTEP SIVFRYSKKNRAKTLRWVFECIRDGLGYPSIKNDDL GIQQLLEMAKYSRNGNGVTPEEAHYWVNVLCMA PGVAGRRKAQKTRSEGGSAIFPAKLLEITLSNGYD WSYADMQMGPETGHAKDFATFDQLWEAFRKQYQ YAIALAIRCKDVSRTMECRFLQMPFVSALDDGCME LGMDANALSEQPNGWHNPITTIVAGNSLVAIKKLI YDEKKYTMEQLMDALKANWEGYEEMRRDFKNAP KWGNDDDAADTLISRFYEEILGGEMMKNINYSGGP VKPVGQAVGLYMEVGSRTGPTPDGRFGGEAADDG GISPYSGTDKKGPTAVLRSVSKVQKNQKANLLNQR LSVPIMRSKHGFDIWHAYMDTWHELNIDHVQFNV VSTEEMKAAQREPEKHQDLIVRVSGFSARFVDIPTY GQNTIIARNEQDFNAQDLEFLNAEL
BssA	*Thauera* sp.DNT-1	BAC05501	MNDIVSAKILEYKGKTLNFTPEDPAEAKIPADELHE HLQKPSTARTKRLKERCRWKHASAGEFIEKSVTAG IERMRYLTEAHKASEGKPEAIRRALGLANVLNKST LVLQEDEFIVGYHAEDPNMFPLYPELSHMAVQDYL RSDYSPQPADEAAAINEYWKPHSLQSKCQPYFDPA DLGRMYQVSSMEAPSFASGYNSIVPPYETVLEDGL

			LARIKLAEKHIAEAQGDMSTFPWNGTKGLDNIAKI DNWKAMVIACKAVISWARRQARLCRIVAENFETD PKRQAELLEIADICHRIPAEPCKGLKDAMQAKFFTF LICHAIERYASGYAQKEDTLLWPYYKASVIDKKFQ PMDHMGAVELVEMERLKISEHGAGKSRAYREIFPG SNDLFILTVGGTNARGEDACNDMTDAILEAAKRIR TAEPSIVFRYSKKNREKTLRWVFECIRDGLGYPSIK HDEIGTEQMKEYAKFSLNGNGATDEEAHNWVNVL CMSPGIHGRRKTQKTRSEGGGSIFPAKLLEITLNDG YDWSYADMQLGPKTGDLTSLKTFEDVWEAFREQY QYAINLCICTKDVSRYFEQRFLQMPFVSAIDDGCM ELGMDACALSEQPNGWHNPITTIVAANSLVAIKKL VFEEKKYTLEQLSQALKANWEGFEEMRVDFKRAP KWGNDDDYADGIITRFYEEIIGGEMRKITNYSGGPV MPTGQAVGLYMEVGSRTGPTPDGRFGGEAADDGG ISPYMGTDKKGPTAVLRSVSKVQKNQKGNLLNQR LSVPIMRSKHGFEIWNSYMKTWHDLNIDHVQFNV VSTDEMRAAQREPEKHHDLIVRVSGYSARFVDIPT YGQNTIIARQEQDFSASDLEFLNVEI
TutD	*Thauera aromatica* T1	AAC38454.1	MNDIVSAKVLEYKGKKLNFTPEDPAEETIPADELH EHLQKPSTARTKRLKERCRWKHASAGEFIEKSVTA GIERMRYLTEAHKASEGKPEAIRRALGLANVLNKS TLVLQEDEFIVGYHAEDPNMFPLYPELSHMAVQDY LRSDYSPQPADEAAAINEYWKPHSLQSKCQPYFDP ADLGRMYQVSSMEAPSFASGYNSIVPPYETVLEDG LLARIKLAEKHIAEAQADMSTFPWNGTKGLDNIAK IDNWKAMVIACKAVISWARRQGRLCKIVAENFETD PKRQAELLEIADICQRIPAEPCKGLKDAMQAKFFTF LICHAIERYASGYAQKEDTLLWPYYKASVVDKKFQ PMSHMDAVELVEMERLKISEHGAGKSRAYREIFPG SNDLFILTVGGTNAKGEDACNDMTDAILEAAKRIR TAEPSIVFRYSKKNREKTLRWVFECIRDGLGYPSIK HDEIGTEQMKEYAKFSLNGNGATDEEAHNWVNVL CMSPGIHGRRKTQKTRSEGGGSIFPAKLLEISLNDG YDWSYADMQLGPKTGDLSSLKSFEDVWEAFRKQY QYAINLCISTKDVSRYFEQRFLQMPFVSAIDDGCME LGMDACALSEQPNGWHNPITTIVAANSLVAIKKLV FEEKKYTLEQLSQALKANWEGFEEMRVDFKRAPK WGNDDDYADGIITRFYEEIIGGEMRKITNYSGGPV MPTGQAVGLYMEVGSRTGPTPDGRFGGEAADDGG ISPYMGTDKKGPTAVLRSVSKVQKNQKGNLLNQR LSVPIMRSKHGFEIWNSYMKTWHDLNIDHVQFNV VSTDEMRAAQREPEKHHDLIVRVSGYSARFVDIPT YGQNTIIARQEQDFSASDLEFLNVEI
BssA	*Thauera aromatica* K172	CAA05052.1	MSDVQTLEYKGKVVQFAPENPREAEIPADELHEHL QNPSTERTRRLKARCRWKHAAAGEFCEKGVTAGI ERMRLLTESHWATRGEPEPIRRAHGLKNILDKSTL VLQTDEFIVGYHAEDPNMFPLYPELSYMAVQDYL KSKYSPQPAKEAQEIVDYWKPFSLQARCEPYFDPV DLHRGYQVSTIEGPVFATGYNSVIPPYETVLEDGLQ ARIALAEEKIEHARAEMEKFPWHAPSGLEWIDKID NWKAMVIACKAVIAWARRHARLCKIVAEHFETDP KRKAELLEIADICQRMPAEPARGLKDAMQSKWFTF LICHAIERYASGFAQKEDSLLWPYYKASVIDKTFQP MEHKDAVELIEMERLKVSEHGAGKSRAYREIFPGS NDLFILTLGGTNGDGSDACNDMTDAILEATKRIRTT EPSIVFRYSKKNRAKTLRWVFECIRDGLGYPSIKHN ELGVQQMLEMAKYSRNGNGATPEEAHYWVNVLC MAPGLAGRRKAQKTRSEGGSAIFPAKLLEITLNNG YDWSYADMQMGPETGYAKDFATFDQLWEAFRKQ YQYAIALAIRCKDVSRTMECRFLQMPFVSALDDGC MELGMDANALSEQPNGWHNPITSIVAGNSLVAIKK

			LIYDEKKYTMAQLMDALQANWEGYEEMRRDFKN APKWGNDDDDADVLISRFYEEILGGEMMKNINYS GGPVKPTGQAVGLYMEVGSRTGPTPDGRFGGEAA DDGGISPYSGTDKKGPTAVLRSVSKVQKNQKANLL NQRLSVPIMRSKHGFDIWHAYMDTWHDLNIDHVQ FNVVSTEEMKAAQREPEKHQDLIVRVSGFSARFVD IPTYGQNTIIARNEQNFNAQDLEFLNVEL
BssA	*Azoarcus* sp. T	AAK50372.1	MNDIASAKVLEYKGKTLNFTPEDPAEAKIPSDELHE HLQKPSTARTKRLKERCRWKHASAGEFIEKSVTAG IERMRYLTEAHKASEGQPEVIRRALGLANVLNKST LVLQEDEFIVGYHAEDPNMFPLYPELSHMAVQDYL RSDYSPQPADEAAAINDYWKPHSLQSKCQPYFDPA DLGRMYQVSSMEAPPFASGYNSIVPPYETVLEDGL LARIKLAEKHIAEAQADMSTFPWNGTKGLDNIAKI DHWKAMVIACKAVISWARRQGRLCRIVAENFETD PKRQAELLEVADICHRVPAEPCKGLKDAMQAKFFT FLICHAIERYASGYAQKEDTLLWPYYKASVIDKKF QPMDHMGAVELVEMERLKISEHGAGKSRAYREIFP GSNDLFILTVGGTNAKGEDACNDMTDAILEATKRI RTAEPSIVFRYSKKSREKTLRWVFECIRDGLGYPSI KHDEIGTAQMKEYAKFSLNGNGATDEEAHNWVN VLCMSPGLHGRRKTQKTRSEGGSSVFPAKVLEITL NDGYDWSYADMQLGPKTGELSSLKTFEDIWEAFR KQ YQYAINLGISTKDVSRYFEQRYLQLPFVSAIDDGCM EFGMDACALSEQPNAWHNEVSTVVAANSLVAIKK LVFEEKKYTLEQLSQALKANWEGFEEMRVDFKRA PKWGNDDEYADSIVSRFYEEVIGGELRKITNYSGAP VLPTGQAVGLYMEVGSRMGPTPDGRFGGEAADDG GISPYMGTDKKGPTAVLRSVSKVQKNQKANLLNQ RLSVPIMRSKHGFEIWNAYMKTWHELNIDHVQFN VVSTDEMRAAQREPEKHSDLIVRVSGYSARFVDLP TYGQNTIIARQEQDFSASDLEFLNVEI
BssA	*Desulfobacula toluolica* Tol2	CCK78310	MATIAEAVEYRGKIIEFPLEHQEENDIPDERLHEHL ARPSTQRTKRLKARCRWKHASAGEFVEKGVTSGIQ RMRYITEAHKKSRGKPEVIRRALGLANILNKYTLV LQEDEFIIGYHAEDPNMFPLYPELAYMAVADYLVS DYAPQPAEEAKEIMEYWKPYSMQGKCEGYFDPED LMRMYQVSTMEAPGFATGYNSIVPPYETILEDGLL KRIEMAEANIKAAKEDLKKTPWDATKGLKWIPMI DNWEAMIIADKAVIAWARRHARLAKFVAENIETDP KRKEELLEIADINQRIPAEPCKGLKDAFQAKWYTFL ICHAIDRYASGFAQTEDTMLYPYYKASVIDKTFQP MTHSEAVEMVEMERLKISEHGAGKSRAYREIFPGS NDLFILSIGGTKPDGSDASNEMTNAILEATRNIRTTE PSIVFKYSDKSNDKTKRLVFECIRDGLGYPSIKHNEI AVEQLKYYSQFSKEGNGATDDEAHNWANVLCMS PGLCGRRKTQKTRSEGGGSIFPAKILEVTLNDGYD WSYADMQLGPQTGKAEDFKTYEELYEAFKVQYQ YAVSLVIKCKDTMRYFEGKFLQMPFASSLDDGCM ELGRDGCELSEQPNGWHNPITTIVAANSMVAIKKLI YDDKKYTMKQLLDALKANWEGYEEMHKDFKAAP KWGNDNEYADATIKDFYEDIIGGEMGRITNYSGGP VLPVGQAVGLYMEIGSRTGPTPDGRLGGEAADDG GISPYMGTDHKGPTAVLKSVSKVQKNQKANLLNQ RLSVPIMRSKHGFDIWNAYMKTWHDLKIDHIQFNV VSTAEMKAAQKEPEKHQDLIVRVSGFSSRFVDIPTY GQNTIIARNEQQFGAEDFEYLNLDI
TutG	*Thauera aromatica* T1	AAC38455	MEGSNMETGQNLQNQPHTEVGTARPCRSCKWQTP DPTDPHRGQCTANRHAMGGVWKRWLRDVENTTC SRHEEGKLSFRDHV
BssB	*Thauera*	BAC05502	MEGYNMEAEKNLQNQPHAEVGTARPCRSCKWQT

	sp.DNT-1		PDPTDPHRGQCTANRHAMGGVWKRWLRDVENTT CSRHEEGKLSFRDHV
BssB	*Thauera aromatica* K172	O87944	MSATPHTQVHWEENTARPCRKCKWQTPDPTDPLR GQCTVNRHAMGGVWKRWIRDVEHMTCSRHEEGE LSFRDHV
BssB	*Azoarcus* sp. T	AAK50373	MEAEKSLQNQPYTEVGTAKPCRICKWQTPDPTDPH RGQCTANRHAMGGVWKRWLRDVENTTCSRHEEG KLSFRDHV
BssB	*Desulfobacula toluolica* Tol2	CCK78309	MSVVKTQLYMEKDTKKPCRSCKWEIADPTNPDQG QCTVNRTSAGAVWKRWVRDVHNMTCSRHEEGKL SFREHV
BssB	*Geobacter metallireducens* GS-15	YP_006720506.1	MSPRGPGQEATVSATSNGNMMHEEPGTKKPCQSC KWQIADPTNPLRGQCTVNRNAMGGVWKRWVTDV NRMTCGKHEVGKLSFREHV
BssB	*'Aromatoleum aromaticum'* EbN1	YP_158061.1	MSATPHTQVHWEENTPSPCRKCKWQTPDPTDPLR GQCTVNRHAMGGVWKRWIRDVEHMTCSRHEEGE LSFRDHV
Putative BssB	*Magnetospirillum* sp. TS6	BAD42367.1	MTMTQASTSAPANAQPAVQMHREEGTSRPCSICK WQTPDPTDPIRGQCTVNRHANGGVWKRWIRDALN TTCGRHEEGKLSFRDHV
TutF	*Thauera aromatica* T1	AAC38453	MGTTTCKQCANFFPVPKDADDYEAGKADCVREKE DEKGKYWLSKPIFENSAQCEAFQTKR
BssC	*Thauera* sp.DNT-1	BAC05500	MGTTTCKQCANFFPVPKDADDYEAGKADCVREKE DEKGKYWLSKPVFENSTQCEAFYAKR
BssC	*Thauera aromatica* K172	O87942	MTTCKDCAFFFSIPEDADDFEKSKGDCVTQKDDEK GRYWLSKPVFENDQCCGAFHKR
BssC	*Azoarcus* sp. T	AAK50371	METTTCKQCANFFPLPKDADDYEAGKADCVREKE DDKGKYWLSKPIFENSTRCEAFHAKR
BssC	*'Aromatoleum' aromaticum* EbN1	CAI07158.1	MTTCKDCEFFFSVPEDADDFEKSKGDCVTQKEDEK GKYWLTRPVFENDQCCGTFHKR
BssC	*Desulfobacula toluolica* Tol2	CCK78311	MSKCESCKFFIRVPNDADDFEPGKGDCITEVKDEK GKYWLSTPVFQDTDACPNFNKAI
Putative BssC	*Magnetospirillum* sp. TS6	BAD42365.1	MTSCNNCEFFNPVPKDADDYEAGKGDCVTEKADE KGRYWLSRPVFEGSASCKSYSKR
TutE	*Thauera aromatica* T1	AAC38452	MVVVDRPPPPDLGGGVIVRNECSHSEPIDTGRMTFP TCDSGADVKIPLVTEIQRFSLQDGPGFRTTVFLKGC PLRCPWCHNPETQKVGKEYYYNRDRCVSCGRCAT VCPTGASQLLDGPGASQVLKLDRSKCINCMRCVA VCLTGSRDSVGMEMTLDEILREVLSDEPFYRNSGG GVTISGGDPLFHPAFTLELARKIKERGVHVAIETSCF PKKWATIQPLLKLVDLFIVDLKSLNRKKHEETVGW PLQPILDNIEHLIQAKANIRIHIPVIPGFNDSPMDFED YIAYLGRHAAQLDGVDILNYHVYGEGKYRSLGRE NEYQYFGVEENPPEKVVPLAKGLKLAGITSVTIGGL VGITADRHKSSRDAGTGCIA
BssD	*Thauera* sp.DNT-1	BAC05499	MESRENHFLRLGRSGLQTGYAGTLDQAERTCARQ PIETGRMKFTTRDAGADVKTPLVTEIQRFSLQDGPG FRTTVFLKGCPLKCPWCHNPETQKVTKEYYYNRD RCVSCGRCAAVCPTGASQLVDGPRGTQTLKLDRD KCTNCMRCVAVCLTGSRESVGVEMTLDEILQEVLS DEPFYRNSGGGVTISGGDPLFHPAFTLELARRIKER GVHVAIETSCFPKKWGTIQPLLEFVDLFIVDLKSLN PKKHEEVVGWPLQPILENIEHLLQAKADVRIHIPVIP GFNDSQTDFEDYITYLSRHAAQLDGVDILNYHVYG EGKYRSLGREKEYQYLGVEENPPEKVVPLAKGLKL AGIKSVTIGGLVGITADRHTSSRDAGTGCIA
BssD	*Thauera*	O87941	MKIPLITEIQRFSLQDGPGIRTTIFLKGCPLRCPWCH

	aromatica K172		NPETQDARQEFYFYPDRCVGCGRCVAVCPAETSRL VRNSDGRTIVQIDRTNCQRCMRCVAACLTEARAIV GQHMSVDEILREALSDSAFYRNSGGGVTISGGDPL YFPDFTRQLASELHARGVHVAIETSCFPKQGKVVE SMIGIVDLFIVDLKTLDAHKHLDVIGWPLAPILANL ETLFAAGAKVRIHIPVIPGFNDSHADIDAYAEYLGK HAAAISGIDLLNFHCYGEGKYTFLGRAGSYQYSGV DETPAEKIVPLAQALKARGLAVTIGGIVGIANGKNE LTGDIALEVHH
BssD	*Azoarcus* sp. T	AAK50370	MRIPVVTEIQRFSLQDGPGIRTTIFLKGCPLRCPWCH NPETQDVRQEFYYYPARCVGCGRCMAVCPAGTSR LVHNSDGRTIVELDRTDCQRCMRCVAACLTDART TVGQRMSVEEILREALSDSPFYRNSGGGVTISGGDP LFHPSFTLELARRIKERNVHVAIETSCFPKRWEVIQP LLEFVDLFIVDLKSLNPKKHEEVVGWPLQPILDNIE HLMRSKANIRIHIPVIPGFNDSPQDFDDYISYLSRHT MQLEGVDILNYHVYGEGKYRSLGREKEYQYLGVE ENPPEKVVPLVKGLKLAGIKNVTIGGLVGITADKD ETDRDAATRCIT
BssD	*Desulfobacula toluolica* Tol2	CCK78312.1	MCIKNHPKIPLITEIQRFCLQDGPGFRTTIFLKGCPL HCPWCHNPETQSPKKEFYYYPEKCSNCGRCAKSCP TGASTMKIGTDNTPVLDLDRSKCIACMQCVDACLS SARAVVGQNLTIDTIMEEALADQPFFKNSGGGVTIS GGDPLLFPDFTLELVKRLKKEGLHVAIETSCFQKW DKIRPLLSYVDLFLVDIKSLISKKHKIVVGWP LEPILENIKNLIEKNANLRIHLPIIPNFNDSMDDFKTC EIFLSRIADRLMGVDILPYHVYGEGKYNFLGRHDA YACKDVKQSPSEKIEPLVKALKRLHIKDLSVGGLV GMGGYSGEKSKKGL
Putative BssD	*Magnetospirillum* sp. TS6	BAD42364.1	MKIPLITEIQRFSLQDGPGIRTTIFLKGCPLRCPWCH NPETQDTRREMFYYENRCVGCGRCVAVCSTGAST LVDTGGKSPTLVVNRDKCDRCLRCAAVCLTEARGI SGQAMTVDEILREALSDKPFYKNSGGGVTLSGGDP LMYPEFVLELARRLHDEGVHLAMETSCFPKHWET MEPLVEHTDLFIVDLKCLNAKRHEEVVGWPLQPIL RNLNNLFERNATVRIHIPVIPGFNDSEEDFRDYAEFL AGYADRLNGVDILNYHSYGEGEYAALGRMETYKF AGVAENPAEVVLPLVKALKDKGIPGITVGGLVGIT CNR
BssD	*'Aromatoleum' aromaticum* EbN1	CAI07157.1	MRIPLVTEIQRFSLQDGPGIRTTIFLKGCPLHCPWCH NPETQDARQEFYYYPARCVGCGRCVAVCPAGTSR LVHHSDGRTRIELDRTNCQRCMRCVAACLTEARSI VGQRMSVEDILREALSDSAFYRNSGGGVTISGGEPL YFPEFTRQLAGELHAAGVHVAIETSCFPKNRETVES LLDVVDLFIVDLKSLDPRKHFEVIGWPLAPILANLE MLFAAGANVRIHIPLIPGFNDSPADFDAYVDYLGSR AESITGVDLLSFHSYGEGKYAFLGRSDSYQYSGVE EPPAEKTMPLARALKNKGLAVTVGGIVGIIDSKSEL TDGDILEVHQ
BssD	*Geobacter metallireducens* GS-15	ABB31775.1	MLTPLITEIQRFCLHDGPGIRTTIFVKGCPLHCPWCH NPENINPKQEFYYHASKCSNCGQCLTACPSGVDNP EHDDCIGKTTDRSHCTSCFQCVSACRFGARETVGK FISIESIVQEAVSDRIFYQHSGGGVTVSGGEPLMYPE FTRDLTYRLKVKENVHVAIETSLFAEWHNIEPLLK YVDLFIVDIKTPDPQKHQHVIGGSLHKILSNLERLLE ARATVRTHLPIIPGINDTSQDFEAYAEYLGQFANQL SGVDILPYHSYATGKYVQLGRSYQYLGVPDLPAQQ LTPLVNALRQQGIREITLGGMVGSSPAVENVAGTR SLKPRRDYFSRPVYSPQGRGVVVTQR
NmsA	Delta proteobacteriu	CAO72222.1	MQENVALKIAEETPGSTERIQFLYDRCRWKHVAGG MYMRPEVKVGIARARLLTESYKETRGESEMIRRAK GLDHILENYPIFINDEEFVVGDAAENPDTLAIFPEM

	m NaphS6		GFFPTIDIVEDPELMDDDIRDEAREIAMFWKPLGLQ DKCMPYYDQHEIDIATPWTIVDVPPFIANYMSVCP AYMSVLEDGLLGRIKSSEENIAKAFVKLRAYPWNG PENMPLMDQIDVWRSMIIADKAVIKWARRYGRLA KIIAENFDLSDSVLGAEGRKNELLEISDICYRMPAEP AKGFKDAMQSKWFVYLVCHSLERYSSGYAHLEDR LMWPYYKASVIDKTAQPMTRDEAIELIECERLKVC ERGVAKGRAHREGQPGANDLHIITIGGLDENGNDA TNDLTNAILEASLSVRTPEPSLGFRYSPKINAKTRRL VFENIAAGFGFPSIKHEEKNTRQMLEHYKVPPDEA AHWALVLCMAPGVGKRRGLQKTRTEGGGLIWIDK CCEIAFYDGFDHSFANIQTGPKTGDATKFKTFEELF EAFEKQVEFATALHYRNKDVTRRAEIKFIESPFVAS LDDACMDDGVGAFVDKTYPNPWNNTPGEQTAAD SLAAVKKLVFDDKKYTMEEVVNAMKANFEGHEE MRKDMLAAPKWGNDDPYVDEIGERIFTMVADKL MEQTTYSGMHPLGNPQTVSTFATRAPRIGALPFGK LHGEVLHDGGSSPYVGLDKKGPTAVLKSVAHIPYD RYKGVQFNQRLPVSVMRGDKGFQVWTAYMKAW HDLNIDHVQFNVVDTKDMLEAQKEPEKWQSMIVR IAGYSARFVSLPRNAQDSIIARTEQPVG
NmsA	Delta proteobacterium NaphS2	CAO72219.1	MQENVALKISEETPGSTERIQFLYDRCRWKHVAGG MYMRPEVKVGIARARLLTESYKETRGESEMIRRAK GLDHVLENYPIFINDEEFIVGDAAENPDTLAIFPEM GFFPTIDIVEDPELMDDDIRDEAREIAMFWKPLGLQ DKCMPYYDQHEIDIATPWTIVDVPPFIANYMSVCP AYMSVLEDGLLGRIKSAEENIEKAFVKLRAYPWNG PENMPLMDQIDVWRAMIIADKAVIKWARRYSRLA KIVAENFDLSDSVQGAEERKNELLEISDICYRMPAE PAKGFKDAMQSKWFVYLVCHSLERYSSGYAHLED RLMWPYYKASVIDKTAQPMTRDEAIQLVELERLK VCERGVAKGRAHREGQPGANDLHIITIGGLDENGN DATNDLTNVILEASLNIRTPEPSLGFRYSPKINAKTR RLVFENIAAGFGFPSIKHEEKNTRQMLEHYKVPPDE AAHWALVLCMAPGVSKRRGLQKTRTEGGGLIWID KCCEIAFYDGFDHSFANIQTGPKTGDATKFKTFEEL FEAFEKQVEFATALHYRNKDVTRRAEIKFIESPFVA SLDDACMDDGVGAFVDKTYPNPWNNTPGEQTAA DSLAAVKKLVFDDKKYTMEEVVNAMKANFDGYE EMRKDMLAAPKWGNDDPYVDEIGERVFKMVADK LMEQTTYSGMHPLGNPQTVSTFATRAPRIGALPFG KLHGEVLHDGGSSPYVGLDKKGPTAVLKSVAHIPY DRYKGVQFNQRLPVSIMRGDKGFQVWSAYMKAW HDLNIDHVQFNVVETKDMLEAQKEPEKWESLIVRI AGYSARFVSLPKNAQDAIIARNEQQIG
AssA1	*Desulfatibacillum alkenivorans* AK-01	ACL03428.1	MVAEPAQDQSVQELEDKQEWWWVAEKKRSKRLD YLRKSIWKKGALGGNYAPGIKLDLECATLFTDMW KFWKYDPIMMRRAKAIAHVLDKKTIFITDHAQLVG YFGSLPNTIMWRVDGASMVNEEAYNEPGIMPEPEN ESLQKVAELNDYWAGQTAVDKVARILDPEDAVKF LSGAIGWGAPSSAYGYSGKDYEYLFAGRRGFEDIIE EINAAIEKAEDKTVGVPGPEILDIYDRLQNWDAMIL VLEAGIRHAKRYARLARTMAENMETDEKRREELL KIAETCERVPARAPRNLQESLQYDHFIQIFARTEAH EGAWPARPDYYHGPYYDKDVNVDKTLTKEDALD LVGEFMIRAYEVGGFAPRWAREGLQGITGTWVWT LGGVNKDGSDACNDLTVAFLQAARLVRVSNPTFG FRWHPKVKDEVLRECFECIRHGLGYPSMRNDPLLI QNAMHWHGHPLEEARTWVHQACMSPCPTTKHGV QPMRMASATANCAKMVEYALHNGYDHVVGMQM GPETGDAAKFHDFEDLFQAWVKQMEWLTSLLVRT VNLGRYKDPEFFGRPFLSGMSERSVESGLDVVSPV

			GDRGNCWVTAFTWVENIDSLAAVKKLVFDDKKY TMEQLLTALKANWDGYEEMRLDFVNNAPKWGND DDYVDDIMLRCLRETARHSRVMKCPSGNSWPILPE NVSGNIHYASIVGALPNGRRCGDALYDGGISPGPG LDKKGPSAVLKSCGKIDHVSDGRAFLLNQRLSPTQ LAGEKGYQLWKAYIRTWADLGLDHVQFNMVSDE TLRAAQKDPEKYSEVIVRVAGYSAHFVDISRKTQD NIIQRTVQGI
AssA2	*Desulfatibacillum alkenivorans* AK-01	ACL03892.1	MTAEPAKLDVSLQEHEENQEWWWIAEKKRSKRLD YLRKAVWKKGALGGNYAPGIKVDLEGPKLFTDM WNFWKFDPIIMRRAKALAHVFDNISIFITDHSQIVG YWGSAPHTISWRVDGASIVNEELYNEPGIMPEPEEE SLRKVAEINDYWAGQTAVDKVARILDPEDAVKFLS GAIGWGAPTSAYGYSGKNYEYYLKGERGFEDIIAD IEDHIAEAEEKTIGTPGPDILPIYDRIQNWEAMITVL EAAIRFAKRYARLARTMAEHLETDEKRKEELLRIA ETCERVPAKAPRNLQESFQMDMLIQTMCRFEASEG AWPARPDYYHGPFYEKDVLQDKTLTEEEATDLIGE FMIRAYEVGGFGPRWAREGMQGITGTWVWTIGGV KPDGSDACNALTTAFLRTARLIRVSNPTFAFRWHP KVSDEVMRECFECIRHGLGYPSFRHDPILVANCMN WHGHPVEEARTWVHQACMSPCPTTKNGVQPFRM ASATANCAKMVEYALHNGYDHVVGMQMGPQTG DARTFTDFEQLFDAWTRQMQWLLSLLVRTVNLGR YKDAEFYGRPLLSGITERAVERGIDAVNPEGERGN CWITGFTWVENADSLAAVKKLVFDDKKYTMDQLI TALESNWDGYEQMRLDFVNKAPKWGNDDDYVD DIMLRCLRTLAKHSRVMRCTSNNTWPISPQNVSGN IHYSSVVGALPNGRRLGDALYDGGISPGPGLDKKG PTAVLKSCGKIDHVGDGRAFLLNQRLSPTQMAGEK GYQLWRAYMRTWADLGLDHIQFNMVSDKTLRAA QKDPEKYQEVIVRVAGYSAHFVDISRKTQDNIIQRT VQGI
AssA	*Desulfoglaeba alkanexedens* ALDC	ADJ51097.1	MRLTRTPRSAANCKEVIDILLRRIQWAQAAEARQD IEQLKESQQWWKVAEKLRSPRLDYLRKAVWKKG AIGGAYAPGIKIEIMRNILFTESWKENERDPIMMRK AKALAHVWRNIPIFITDHAQLVGYVGSAPNTLGM WPIEGASMVNEEAYNEEGVIPEPEEESLKIMADLN NYWAGNTAIDQVARLLDPEDAVKFMSGAIGWGVP TSAYGYSGKDYEYILTGKRGFEDIMEEIEARLDEAQ EIVRGTPGPELLPYYEKIQNWEAMLIVLEAAVDWA RRYARLARIIAENFESDPKRKEELLRIAETCERVPA KAPRSLQESLQMDHFIQILARYEAYEGAWPARPDY YHGPFYDKDVNIEKNLTREEALDLVGEFLIRTYEV GGFGPRWGREGLQGITGTWVWTIGGVKPDGTDAC NDLTRAFLQAARLVRVANPTFAFRWHPKVPDDIM RECFECIRHGLGYPSMRNDPILVQNAMYWHGHPLE EARTWVHQACMSPCPTTKHGFQPFRMASATANCA KMIEYALFNGYDPVVNMQMGPKTGDARKFKSFDE LFEAWVKQMEWLTDTLVRTVNLGRVKDPEFYGRP FLSAISERSVEQGTDIVNPEGERGNSWVTGFTWVE NADSLAAVKKLVFDEKKYTMDQLMDALQANWD GYEEMRLDFVRNAPKWGNDDDYVDQIMVRCLEE VARHSKELKDPTGNPWPLLPENVSGNIHYANIVGA LPNGRKRGDALYDGGISPGPGLDKKGPTAVLKSCG KFDHVRLGRAFLLNQRLSPTQLKGEQGYQLWKAY MRTWADLGLDHVQFNMVDDATLRAAQKDPEKYQ EVIVRVAGYSAHFVDISRKTQDNIIQRTIQGL
MasD	'*Aromatoleum*' HxN1	CAO03074.1	MTATSTLSKTDLKNCETVEELRENQQWWWLAERE RSARLDYLRKATWKKGALGGNYFDGIRLDLEYPT LFTEAWKKYPNDPSMLRRAKATAYVLDNISIFITDS AQLVGYVGSAPHTIAWRVDGASTVNSEVYNEPGIH

			AEPEAESLKKVAEINSYWNGQTAVDKVGRLIDPED AVKFFSGAIGWGTPSSAFGYSGKNFEYFMKGDRAF SQIIAEIDEKIDEAEEATIGTPSPHILPLYDKLNNWH AMKLVLEAAIRFAGRYARLARVMAAKETDEQRKK ELLRVAETCERVPANPPRNLQESLQYEHFVQVLAR YEAHEGAWPSRPDYYHGPLYAKDVEVEKNITESE AIDLVGEYMIRCSEYGSFSPRYMREGLQGVTGTFV WTLGGVNQDGTDACNGMTIALLKAARLVRVANP TFGFRWHPKVSNEVLRECFECIRQGLGYPTLRNDP VLIQNTMHWYGHPLEEARTWVHMACMSPNPTTK HGTSPFRMASATMNSAKTIEYVLHNGYDRVVNMQ MGPKTGDAREIKDFEDLFERWTVQLKWLMNLLVR TVNLGRFKDPEFFGRPFLSAITERAVEHGIDAVSPE GERGNAWVTAFTWIENVDSMAAIKKLVFDDKKYT MSQLIDALEAEWDGYEQMRLDFVKNGPKWGNDD DYVDDIMLRCLSVAAEHSRNIQCTSGNCWPILPEN VSGNIHYAN IVGALPNGRRRGDALYDGGVSPGPGLDKAGPTAV LKSVGKIDHVNQGRSFLLNQRLSPTQLAGDKGFQL WNSYVRTWAELGIDHIQFNVISDKVLRAAQNDPEG YQEVIVRVAGYSAHFIDISRKTQDNIIQRTVQGLG
MasD	'*Aromatoleum*' OcN1	CBK27727.1	MTATVTLSKTELESCGTVEELRENQQCCGTVEELR ENQQWWWLAERERSARLDYLRKAIWKKGALGGN YAKGIRLDLEYPTLFTEAWKKYPYDPIMLRRAKST AYVLDNISIFITDNAQLVGYVGSAPNTIAWRVDGTS QVNAEIYNEPGIHAEPEEESLKKVADINSYWAGQT AVDKVARILDPEDAVKFLSGAIGWGAPTSAFGYSG KNYEYFLSGDRAFNQIIEEIDEKIADAEEKTTGTPGP DVMPYYDKLNNWTAMKVVLEAGIRFAGRYANLA RIVAAKETDEKRKAELLRIAETCERVPANPPRNLQE SLQYDHFIQVLARYEAHEGAWPARPDYYHGPLYA KDVEVEKNITEDEALDLIGEFMIRAAEYGSFAPRWS REGLQGVTGTFVWTLGGVKKDGTDACNGMTIALL KAARLVRVANPTFGFRWHPKVSNEVLRECFECIRQ GLGYPTMRNDPVLIQNTMHWYGHPLEEARTWLH MACMSPAPTTKHGTSPFRMASATMNSAKVIEYAL HNGYDRVVNMQMGPKTGDARDIKDFEELFERWTI QLKWLMNLLVRTVNLGRTKDPEFFGRPFLSAITER AVERGIDAVNPEGERGNAWITAFTWVENIDSLAAI KKLVFDDKKYTMTQVVDALEAEWNGYEQMRLDF VSAPKWGNDDQYVDDIMLRCLTIAAEHSHKIKCPS GNNWPILPENVSGNIHYANIVGALPNGRRRGDALY DGGVSPGPGLDRAGPTAVLKSVGKIDHVGQGRAF LLNQRLSPTQLVGEKGFQLWRSYIRTWADLGIDHI QFNMVSDKVLRSAQKDPEGYQEVIVRVAGYSAHFI DISRKTQDNIIQRTVQGLG
AssA	*Smithella* sp. SCADC	AHI85732.1	MAEPAKNISIQELEKKQEWWWAAEKKRSKRLDYL RKAMWKKGAIGGLYPAGLQVDLEECQLDTAKAR ELENAPDPYVVKYAKIFAHYLDNKTIFITDKAQLVS YVGSLPNTIGWNPTTASMVNLEVLNDSSALPEPLD ASLKVINEVAAYWAGKADADRMMPRVDMTDVM KVLSGTIGWGSAVARLGYSGKDYEYIMTGKRGFE DVIAEIDENMDKADAQAHMPAANKEIGALYDKMN TWEAMKITLEAGIRHAQRYARLARIIAENFETDNK RKEELLQIANCCERVPAKPPRTLQESLQYDLFIQNF SRTEAVEGSWPARPDYYHGQYYDKDVNIDKRITK EEAMDLVGEFLIRAAEVSQYKPKWAREGLQGIEGT WVWTLGGVKQDGSDACNDMTIALLQAARLVRVA NPTFSFRWHPKVKEEVLRECFECIRQGLGYPSMRN DPLLIANSMHWHGHPIEEARTWVHQACMSPCPTT KRGFQPMRMASATANCAKIIEYVFTSGFDPAVSMQ IGAETPDAATFTSFEQVYAAWITQMKTIFSILTRAV

			NRARILGAELTPRPFLSAVSERSVESGLDVLTPSISQ GNSWITAFTWVENADSLAAIKKLVFEEKKYTMAE LKKALADDWQGHEEMRLDFVKNAPKWGNDDDY VDKIMLRCLHDCAVFSHELKDPMGNNWPILPENVS GNIHYANVVGALPSGRRRGDALYDGGISPGPGLDK KGPTAVLKSCGKIDHITDGRAFLLNQRLSPTQLAGE KGYSFWKSYMDTWYNLGLDHIQFNCVSDETLRSA QREPEKYQEVIVRVAGYSAHFVDISRKTQDNIIQRT VQGIG
AssA	*Smithella* sp. D17	KFZ44314.1	MIKRDNNGGYIMAEAAVNIKELEKNQEWWRVAE QKRSKRLDYLRKAVWKKGAVGGLYAPGIKVDLER AGLFTEGYFANLNDPGMVRHAKAVANVLDNIPIFI TDQAQLVGYVGSAPHTISWLIEGGSLLNEEVYNEP GIAPEPEAESLKKMAAMMDKWNGYSVQDRFIPYV LSFAAEEGVKLLSGAVGWGLPISISGYNGKDYEYIL TAKRGFEDIIAEIDRRQEDMKDKARKKAGPEYLHV GSRMMNWEGMKIALQAVIRYANRYARLAKIVAEN YETDPKRREELLRIAETCERVPAKPQRNLQESLQFD HFLQVVERFESGGGAWPSRPDYYHGSWYDKDVNI DKRLTKEEALDLVGEFMIRAHEVGSFFPRWTREGL QGITGTWVWTLGGVKQDGADACNDLTVAFLQAA RLVRVSNPTFAFRWHPQVKDEVMRECFECIRQGLG YPSMRNDPVLITNVMHWHGHPIEEARTWVHQAC MSPCPTTKHGFQPMRMASATANMAKVVEYALFN GYDPIVNMQMGPQTGDARKFKSFDELYNAWAEQL KWLMNLLTMSVNFGRVMSPEMCPRSFLSSISERCV ESGQDAASPEGDRGNSWITAFTWVENIDSLAAVKK LVFDDKKYTMDQLITALEANWEGFEQMRLDFVKN APKWGNDDDYVDDIMLKCLREAAAFSRVIKDPSG NNWPILPENVSGNIHYANIVRALPNGRRLGDALYD GGISPGPGLDKKGPTAVLKSCGKINHITDGRAFLLN QRLSPTQLSGEKGYQLWRSYMHTWADLGLDHIQF NVIDDTTLRSAQREPEKYQEVIVRVAGYSAHFVDIS RKTQDNIIQRTVQGIG
AssA	*Smithella* sp. ME-1	AWGX01000974 (NT – protein accession not available); Needs to be properly translated; At a quick glance – BlastX alignment to other assA's either at ~143 nt or ~ 180 nt	CTGCGTTCAAAGAGTGGTTGATCCAAGACAAGCT TACTACAAGGCAAAACCGGTAGGAGCAGATATG GATGCAAACTCGTGCTCTTCATTTGAGAAGAAAT AAAACAGAACATATGCTTCCCCGGTTTAAAAGG GGAGTCAGCTGAAAAGCCGAAAGGTAACAAGAA GGGAGGAAATTTTTATGGCAGAGCCAGCAACAA ATGCAAGTATTAAAGATCTAGAAAAAAAACAAG AATGGTGGTGGGCCGCCGAGAAGAAAAGATCGA AGAGATTGGATTATCTCCGTAAAGCCATCTGGAA AAAAGGAGCGGTCGGTGGACTTTATCCGGCCGG ATTACATGTTGATTTGGAGCAGGGCGTATTGCAT ACACAAGGGGCACGTGAGCTGGAACTTTCACCT GACCCTTATGTCGTCAAGTATGCCAAGATTCTTG CTCATTATCTGGATAACAAGACGATCTTCATCAC GGATAAGGCCCAGCTTGTGGGTTATGTAGGCAGT GCGCCGAACACAATTGCCTGGAATCCAACCCTG GCTACGATCCTGAATGCGGAAGTTATCAATGATT CTACGGCGATTCCTGAACCACTTGATGATAGTAT AAAGGTCATCAATGATGTGACTGCTTACTGGTCG GGGAAGGCCGATGCCGACAAGATGTTGCCACAT GCGGACATGCCTGATATAATGAAGGTTCTCTCCG GCACAATCGGATGGGGTGTTCCAATGGCCAGAG GCGGTTATTCCGGCAAGGATTATGAGTACATCAT GACCGGCAAGCGAGGATTCGCAGACGTTATCGA TGAACTTGACCACAATATTGATGAGGCGGATGCC AAGGCTCATATGCCTGCCGCTAATATGGAAATTA CCCCTCTGTATGCAAAGATGAATAATTGGGAGGC GATGAAGATAACCCTCGAAGCCTGTATACGTCAC

			GGGCAGCGCTTTGCCCGCCTGGCCAGAATCATGG CGGAAAATTTCGAGACCGACAATAAGCGCAAGG AAGAACTTCTGCAGATTGCCGATTGTTGCGAAAG AGTACCGGCTAAAGTGCCGCGGACTTTACAGGA ATCGCTGCAGTATGACCTTTTCATCCAGTTGTTC AGCCGGAGCGAGGCCACCGAAGGAGCCTGGCCT GGACGACCGGACTACTATCATGGCCCTTATTACG ACAAGGACGTCAATATTGACAAGCGTATTACCA AGGAAGAAGCTTTGGATCTGGTCGGCGAGTTTCT GATCAGGGCTGCTGAAGTTTCTCAGTACAAGCCC AAATGGGTCCGTGAAGGCTTGCAGGGCATAGAA GGAACCTGGGTCTGGACCCTGGGCGGGGTCAAT AAGGACGGCAGCGATGCCTGCAACGACTTGACG ATTGCGTTGCTTCAGGCTGCCCGTCTGGTAAGAG TTGCCAACCCGACCTTCAGTTTCCGTTGGCATCC CAAAGTCAAGGAAGAAGTCCTGCGCGAATGTTT CGAATGCATCCGTCAGGGTCTCGGTTATCCTTCG ATGCGTAATGATCCTGTCCTGATCGCTAATTCCA TGCATTGGCACGGACATCCCATTGAGGAAGCCC GCACCTGGGTGCATCAGGCCTGTATGTCACCGTG TCCTACTACCAAGCGCGGTTTCCAGCCGATGCGC ATGGCCAATGCTACCGCCAACTGCGCCAAGATC ATCGAGTACGTCTTTACCAGAGGCTTCGACCCTA TTGTCAACATGCAGATCGGGGCCGAAACGCCTG ATCCTGCCACCTTCTCCAGTTTCGACCAGGTATA CGATGCGTGGATTACTCAGATGAAAACCATTTTC AGTATTCTGGCCCGTATGGTCAATGCAGCCAGAG TATACGCTCCCGAATTTACACCACGGCCTTTCCT GTCGGGCATCTCCGAGCGCTCGGTGGAAAGCGG TCTGGATGTAATGACGCCTTCACTCAGCCGTGGT AATTCCTGGATAACAGCCTTTACCTGGGTGGAAA ATATTGACAGTTTGGCAGCGATCAAAAAACTGGT CTTTGATGAAAAGAAGTACACCATGGCCCAGCTT AAGCAGGCCATCGCCGACGATTGGCAGGGCCAT GAGGAGATGCGTCTTGATTTCGTCAAGAATGCGC CTAAATGGGGCAATGATGATGACTATGTCGACA GCATCATGCTCAAATGCCTGCGTGATATTGCCGT TTTCTCGCATGAACTGAAGGATCCCATGGGGAAC AACTGGCCGATTCTGCCCGAAAACGTCAGCGGC AATATTCACTATGCCAATGTTGTGGGCGCTTTGC CCAGCGGCCGCAGAAGAGGAGACGCTCTTTATG ATGGCGGTATTTCTCCCGGGCCGGGACTCGATAA GAAGGGACCGACCGCTGTTTTGAATTCATGTGCC AAGATAGACCATATCACCGATGGCCGGGCATTC CTCTTGAATCAGAGACTGTCCCCAACTCAGCTTG CCGGAGAAAAGGGATACTCACTCTGGAAAGCGT ATATGGACACCTGGTATCAGATGGGACTGGATC ACATTCAGTTTAACTGTGTTTCCGATGAGACCCT GAGAAGCGCCCAGAGGGAGCCTGAAAAATAT
Putative AssA	*Peptococcaceae* bacterium SCADC1_2_3	KFI38250	MSTAVEEKEKALHEHLEKREYWWWAEKARSSRID YLRKAVWPKAAKGSSYLPGTKICWHGYKIYREVF DTPDAKMDPYILTFAKAQAKVLDTIPIFIIDQSRIVG YTGSAPHKLYMTPQSSFNVAWDMYNDRMDLVDR EEDRPWIKEAIEFMKPYTYEARADRILSRRQRIMCS TSQTFTGSSHRESQTYSTTQWGYFHSGFNNILAQV NENIAEAEKVLYEVGARPDFPAEYDALNKLDNWR SFKIVIEAMLRWAKRYSRLAKIIAENFETDPQRKQE LMEISRICAKVPGEAPESYHEFLMYDNMAQRAKRY EWGEGAWPAKPDTWGWPCYKKDVIDEKTLSQDD ALDLTCEYLIRAFEYGNPRNRQYRELLTGDPGPYV WALGGILPDGSDGANELSDTILKAARLVRVASPTY ALRYHPKIRTETLRQAFECIRHGLGYPNIRHDDVLV

			QSNMYWSGTPLEEARTWTAQACIVPCPGTKHGVM PARYSASSTLGSKCVELALWNGFNPVFKMQIGPKT GDPTKMTFDELFDACIEQFKVIHWEGCKIRNISRW VEEEIGRPMLSSGWEECIETGKNAFQRREYGNNWL TTFIWTDGWDAMAALKKLVYDEKKYTMEQVLEM LKVNWEGYEVERMDFVRAPKWGNDDDYVDEIVV KAYNRTRDEVLLPCKDWGRSSRGIPTAPQNIAAYT VAGVLLGALPNGRRLGDTCYDGGCSPGAGLDKKG PTAVLRSVGKIDHVTSHRANLLNQRLSPTQLVGDK GFELWHNYIKTWHDLRINHVQFNMVDNETLYAAQ KEPEKYSELMVRVAGYSAHFTGLNKKTQDTIIARN VQTL
AssB1	*Desulfatibacillum alkenivorans* AK-01	ACL03426.1	MSRRSQWKQVANGMLGDVSKARSVEERENVPEK ACGFCKNFSENAYASDGRGYCKKLKVGSDIKVNPP VYVMEGQASLTVLFNSDGAGCEYFNRQDLIDTDG NEVNDPSYRRVQRQMEKAIK
AssB2	*Desulfatibacillum alkenivorans* AK-01	ACL03893.1	MSRRSQWKQVANGMLGDVSKARTVEERENVPEK ACGFCKNFSENAYASDGRGYCKKLKVGSDIKAVPP VYVMEGEASLTVLFNSDGGKCEHYVRQALIDTSGS EVNDPNYRRAQRQLKK
Putative AssB	*Smithella* sp. SCADC	KGL06508.1	MARMSRESFAKATGACGRTAEELEKESHSDQVHE KNCGVCKHYLENSYTSDGRGSCSMLKDGSDITSDP PVFVLDGKNGYMLRILTDASRCKYFEKMEFIDHDG TECSDPMYRRSMRQLQDK
Putative AssB	*Smithella* sp. D17	KFZ44313.1	MAENDKWKQALSVSMSASHRSSGVEARENVPQIS CGVCKNFSETAFSSDGRGTCRILKSGSDISLEKPVY VLDGDVSLFIKFNTDAARCERFERMEIIDTDGSECA DPAYRRSQRQMEKIVK
MasE	'*Aromatoleum*' HxN1	CAO03075.1	MKCTECGHEAEVMKFRYHYNPRIDASLSLRQCPEC QAVVTVDELKREVLGRMHNGDDPWGKSAGIENL AEG
Putative uncharacterized protein similar to MasE	*Desulfatibacillum alkenivorans* AK-01	ACL03429.1	MKCTECSHEAGVSSFRYLYNARIDAPITLRQCPQC QAWLAVDEMAGEARQRVDAGEAPWGKSAGIEGL AEDAR
Putative uncharacterized protein similar to MasE	*Desulfatibacillum alkenivorans* AK-01	ACL03890.1	MKCPECSYEAPVPEFRYLYNVRIDAPLTLRQCPQC QTWLSVDELAGEATGKVDAGDAPWGKSAGIERDL EEAV
Putative uncharacterized protein similar to MasE	*Desulfatibacillum alkenivorans* AK-01	ACL03896.1	MELKCSQCGHTGDKVDFRFLNHADSAGVQCYRQC PQCHGAVFCEEMEEPEAYCGNTVWGTSGLRGRVF TRPRPQKEREQEG
Putative uncharacterized protein similar to MasE	*Smithella* sp. SCADC	KGL06507.1	MIKIPRLKQTKEGKDMKCVECKFDGPIDKFRYLYN ARIDSSMTLRQCPNCQAWLAVDELTGAVKQKVGL G EAPWGKSAGIEGLATD
AssC1	*Desulfatibacillum alkenivorans* AK-01	ACL03427.1	MSTCAECRSFFLREDEPGQGDCVRRVVDPRQAFYQ SKPVREDNDASGCESFQKK
AssC2	*Desulfatibacillum alkenivorans* AK-01	ACL03891.1	MSTCSDCKSFFPREDEPGKGDCVRRVVDPRQAYYT TRPKNPEDDASGCGEFQKR
MasC	'*Aromatoleum*' HxN1	CAO03073.1	MSTCKECRNYFPINEEASRGDCVRRISDERQSYYTA RPTTEAAKCEGCSDYLENTRTAKAH
Putative AssC	*Smithella* sp.	KFO69020.1	MPLCKECKKFFPIKEDPNKGDCVQRVVDPRQAYY

	SCADC		KAKPVGADMDANSCSSFEKK
Putative AssC	*Smithella* sp. D17	KFZ44316.1	MPLCKECKKFFPVKEDPNKGDCVQRVVDPRQAYY KAKPVGASMDASSCSSFEKK
MasG	'*Aromatoleum*' HxN1	CAO03077.1	MKPRQPLIADIQKFSVNDGPGFRTSVFLKGCSMRC AWCHNPETIAFESQIFWKSRLCMQCGTCARICPRG AANKPVPVEAAQAEGSTYYKIDHDKCDACMECVS ACPYSAMEITGRTMSVDEIMTIVEQDMLFYKNSGG GFTLSGGDPTAFPDFSETLLREAKRRELHTCMDTN GHCAWSVFERVMPHVDIFLLDLKHLDDGPHREMT GVTNRRVLENMASLVKANAEVWIRIPFIPQFSDQID YHQRVADFLVELPGSIARVDIIPFHNFCQTKYQWL GRDWRYRNVEAIESSSLAPYLELYKKNGLNATIGG SGFETDCKRSD
AssD1	*Desulfatibacillum alkenivorans* AK-01	ACL03425.1	MANACLITEIQRFAINDGPGFRTNVFLKGCPLKCV WCHNPETIDAKAQVFWKKRLCVQCGACMEACPTE AIQPPIDPVLAQSEGVDYYKIDLDRCNHSMQCAAV CPYGALEITGKLLTVKEILDEVESDLPFYKNSGGGM TLSGGEPTAHPDFAEKLLAGAKARGLHTCLDTNGY CSWDILQRLLKYIDIVLFDLKHTDPEKHKQWTGVD NALIMKNLARLTQTGVETWVRIPVIPGFNDSIEDHQ AAVEFLNGLPGKIHRVDLLPYHNWCQDKYGWLGL DWPLGRVEAMEPSLLEIPKEYYEMSGLKTTIGGSG FEDAN
AssD2	*Desulfatibacillum alkenivorans* AK-01	ACL03895.1	MAETCLITEIQRFAVNDGPGIRTNVFLKGCPLKCA WCHNPETISAKPQVFWKKRLCVQCGACMDACPTG AVQPPIDPVLSQAEGTDYYKIDLDRCDYSLSCVDA CPYGALEITGQAMTVEEILDEVESDRPFYENSGGG MTLSGGEPTAHPAFSEKILAGAKARGLHTCLDTNG HCSWSVLESLLKHTDVVLFDLKHTDPEKHKEWTG VDNALIKENLALLVKNRVETWVRIPVIPGFNDAIQD HKAAALFLSELPGRVARVDLLPYHNWCQDKYGW LGQNWPLGRTEAMEPSLLEIPRELYEMSGLEATIGG SGFEGA
AssD2'	*Desulfatibacillum alkenivorans* AK-01	ACL03894.1	MEPGKTIGIVNDIQRMSTNDGPGFRTTVFLKGCLLD CKWCHNPEGRRRFPEVIPFYTNCIGCGDCVEACAA GALSLNGDAKPVIDRALCTDCFQCARTCSHSGLVP WGTIQTAAEVMKEVFSDEPFFRHSGGGLTLSGGEP MAQPGFVLALFTLAKKGAEEGKPIHTALDTCGHAP WEDYARVLPLADLVLLDLKHMDPQPHKAYTGAT NRLILDNAQKMAEAGAVMRIRVPIIPGVNDNKEN WTATAKFAASLGDAVQGVDLLPYHPYAGSKYRAF GMEYDFPAGEGYEDARLEPVIDLFLEHVYEVTIGG
Putative AssD	*Smithella sp. SCADC*	KFO69022.1	MKESVLITDIAKFSVNDGPGFRTNVFLKGCPLKCE WCHNPETIPTYPEVYWKKRLCVQCGACMEACPRD AINPPIEPELANIEGSNYHKIDRAKCDNCLKCIDACK YDALELVGKTMSVSEVIDVVEQDLPFYRTSGGGM TLSGGEPTMHPAFALELLKEAKRRMINTCLDTNGY CNFSILEEIAEYTDIFLFDLKHLDSIKHLEKTGVK NELILKNLELLCKTEFDIWVRIPVIPGFNDDIDFHTK ASSYLASLARNIKRVDLIPFHNYCQDKYSWLGR KWSYSETEAINPSFLDIHADFYRQQGLLTTVGGSGF EEMDTAKTFRNS
Phenylethanol dehydrogenase	*'Aromatoleum' aromaticum* EbN1	Q5P5I4.1	MTQRLKDKLAVITGGANGIGRAIAERFAVEGADIAI ADLVPAPEAEAAIRNLGRRVLTVKCDVSQPGDVEA FGKQVISTFGRCDILVNNAGIYPLIPFDELTFEQWK KTFEINVDSGFLMAKAFVPGMKRNGWGRIINLTST TYWLKIEAYTHYISTKAANIGFTRALASDLGKDGIT VNAIAPSLVRTATTEASALSAMFDVLPNMLQAIPRL QVPLDLTGAAAFLASDDASFITGQTLAVDGGMVR H
Acetophenone	*'Aromatoleum'*	Q5P5G2.1	MYTVDIDTGGTMTDALVSDGEQRHAIKVDTTPHD

carboxylase, alpha subunit	*aromaticum* EbN1		YTVSFNGCLSEAAKRLGYPSTEAFLAKVGMIRWSS TITTNVLGERRGSKVGLLVTEGNEENLYGTVQSPV VGELVDERNIIGLPSNPTAVDILSGVKQLLEGGVRR ICVCLANAFPDNGAEREIKAVIEDQYPDHIIGAVPV LLGSEMAPLRHDQTRVHYSLMNAYTHTQLATSLF KAEDLLRDDHNWTGPLLIGNTNGGVARIGKTKSV DTIESGPVFGTFGGAYMARLYGLKDVVCFDVGGT TTKASIIRDGQPMFQRGGELMEVPVQSSFAMLRSA VVGGGSIARVRDKSVTLGPESMGAAPGPACYGLG GNEATLTDALLALGYLDPNNFLGGRRQLKVDLAR AAIERNVAKPLGVSLEVAALSIRDEAVAMMTELLQ ATLAEAKLTAQDAALFAFGGNGPMFAAFVAERLG VQAAYAFNLGPVFSAFGSAISDVVHVYERGVDLR WNATVKGQLLPTLDALQTQAERDLKGESFDPAKA AYVWELDFGTTEAEVSTVRAELAQSAASTVLDAL TQAVTAAG VASLPLLGARLSSRFVVGAHGMKKRADRVPAEAP ASREMRFNGASEAASPVYRWETMNVGDIAVGPAV VNGSTLTCPIPPRWQLRVDDYGNAELSRAQ
Acetophenone carboxylase, beta subunit	*'Aromatoleum' aromaticum* EbN1	Q5P5G3.1	MYERIRFTEYLDLDLNDEHWYCHDCGTKLISARES YKKGCLVAERRPHEIHNPVIEGEYSFAPDENWVRIL EFYCPGCTRQIETEYLPPGHPITVDIEVDIDSLKARL KKGVIVIKDGKLTKPEAEVLA
Acetophenone carboxylase, gamma subunit	*'Aromatoleum' aromaticum* EbN1	Q5P5G4.1	MSSLTNQDAINSIDIDVGGTFTDFVLTLDGERHIAK CPTTPHDLSIGFLNAVEAGGDKVGLSVEELLPRIDII RYSTTVALNRLLQRQGPRIGLLTTEGHEDAILIGRG AQWTDGQRVAERRNIAVQNKPLPLIERDLILGVRE RIDSSGSVVRPLDEEDVRTKLRMLMDRGARAIVVS LLWSFMNPAHEKRVREIIREEYKEYHIGFVP VVMSHSVVSKIGEYERTMTAVLDAYLQRSMQNDI GATWDKLRAKGYHGAFLMIHNSGGSADIFKTPAS RTFNGGPVAGLMGSAYFANKLGYKNVVAGDVGG TSFDVALVVESSVRNYTFRPVIDKWMVNVTMMQT ISVGSGGGSIAKVDRSGTRLEVGPRSAGSMPGPVC YDLGGTEPTVTDADVVLGYINPDTYYGGRMPLNK AKAEKAIREKIAQPLGIETIEAAALIRYIVDENMASA IKREVHMRGYHPEDFVLFAFGGAGPTHMAGLKGD IPKAVVFPAAPVFCAMGSSIMDIVHMYEQSRRMVF MEPGTEKFVVDYEHFNQTVDTMIERARQELRSEGL EVDDASFGLELDMLYGGQVNLKRMSSPLLHIRTAE DALKVYQAFETEFSEAFSPLVVNKPGGVFLDNFVL RVTVPTWKPPIPEYPLQGTDPSAAFLGKRKAYWPE TKHWADTPTYQFELLQAGNVIDGPAIVEAELTTIV VPPRQRLSIDTHGLAILEAIDPAPPTKRVSAAAAAIV
Acetophenone carboxylase, delta subunit	*'Aromatoleum' aromaticum* EbN1	Q5P5G5.1	MAIPTLEQKLTWLKPAPASSRELDLAAQIDPAQFEI GFQRTNDILDEGMDVFVRSCRCAMGVAGDSLVAI MTADGDIVNGSCGTYLHAVIPPLIIKYILETYGDEIR DGDLWFANDAVYGGVHNPDQMVCMPVYYEGKL VAWTAALVHTTETGAIEPGGMPVSATTRFEEGMN LPPMRIGENFKLREDVVSMFVAFGLRAPSMIAVDL KARCTTADRVRTRIIELCEREGADYVTGLFRKMLQ VAEAGARELIEQWPDGKYRCVTFSDAVGLKQGLV RSCYMTLEKKGDRMLVDLSETGPETPSPYNAHPQA AIAHFSNYIYEYLFHSLPISNGTFANIDFKFGKNTCL SPDPRAATSCSVMISTGVMSAVHNACAKAMFSTSL WKQSGASMGNGGNALVLAGQNQWGSSFADMLA YSINTEGQGARPTEDGMDAFGFPWCVFGRAPNTES VENEFPLLVPLSNHWKDSCGHGKYRGGVGTAQV WVAHHVPELYMMAIADNTKLQTPQPLFGGYAPCT VPGIGIRNANIKELMAEGSDKIKLDVETLLAERTID GKYEIEFQGRSVRPYSNGEVVTFAFSCGGTGYGDP LDRDPKSVEVDLLKGVLTEQTAQNIYKVKWDANL

			RRVDLDETSRLRAAEHDARRKRGVPYEQFEREWL KQRPDDEILKYYGTWPDAKVAQPLLRA
Acetophenone carboxylase, epsilon subunit	*'Aromatoleum' aromaticum* EbN1	Q5P5G6.1	MEAAGALWRRRMQELARGAGKPHAPLFVPLIMGC AAQIEAIPAIDMVRDGTRLRKNLSELRRMLKLDAL TCAVPSCMEAEAVGVEVSQDQWPPRIGTTAQVDV TAEIDADRLAASPRIAAALDAVRQIAVDPGEPVIAA ALTGPAALVAQLRAAGVEAGDEAIYDFAGRILATL ARLYAEAGVNLLSWHEAARPAEEQDDFWKGALG TAGNVARFHRVPPVLVLPASLAAGPWPAQAVPCP ALNHPPLPPVRTHARAWAADPAGWPCLPVEGVAE RLILTDAEVPPETEIATLKAQVERVRGE
Phenylphosphate synthase subunit A, phenol induced	*Thauera aromatica* K172	CAC12685	MKFPVPHDIQAKTIPGTEGWERMYPYHYQFVTDDP QRNQYEKETFWFYDGLHYPEPLYPFDTIWDEAWY LALSQFNNRIFQVPPVRGVDHRIINGYVYISPVPIKD PDEIGKRVPNFMERAGFYYKNWDELEAKWKVKM EATIAELEALEVPRLPDAEDMSVVTEGVGESKAYH LLKNYDDLINLGIKCWQYHFEFLNLGYAAYVFFM DFAQKLFPSIPLQRVTQMVSGIDVIMYRPDDELKEL AKKAVSLEVDEIVTGHREWSDVKAALSAHRHGAE WLEAFEKSRYPWFNISTGTGWFHTDRSWNDNLNIP LDGIQTYIGKLHAGVAIERPMEAVRAERDRITAEYR DLIDSDEDRKQFDELLGCARTVFPYVENHLFYVEH WFHSVFWNKMREVAAIMKEHCMIDDIEDIWYLRR DEIKQALWDLVTAWATGVTPRGTATWPAEIEWRK GVMQKFREWSPPPAIGIAPEVIQEPFTIVLWGVTNS SLSAWAAVQEIDDPDSITELKGFAASPGTVEGKAR VCRSAEDIRDLKEGEILVAPTTSPSWAPAFAKIKAC VTDVGGVMSHAAIVCREYGMPAVVGTGLSTRVVR TGMTLRVDGSSGLITIITD
Phenylphosphate synthase subunit A, phenol induced	*Thauera aromatica* K172	CAC12686	MGSIVSTVALSAATADSTSPKVCPFEACGKDSVPL VGGKCASLGELINAGVRVPPGFALTTSGYAQFMRE AGIQADIGALLEGLDHQDMDKLEEASRAIREMIESR PMPIELEDLIAEAYRKLSVRCYLPAAPVAVRSSATA EDLPGASFAGQQDTYLWIRGVDDLIHHVRRCISSL YTGRAIAYRMKMGFPHEQVAISVGVQMMANAYT AGVMFTIHPGTGDRSVIVIDSNFGFGESVVSGEVTP DNFVVNKVTLDIIERTISTKELCHTVDLKTQKSVAL PVPAERQNIQSITDDEISELAWAAKKIEKHYGRPMD IEWAIDKNLPADGNIFILQARPETIWSNRQKASATT GSTSAMDYIVSSLITGKRLG

Table S5. Metagenome Overview

	A01	A02	B37
Basepair count prior to QC (Mb)	1696	597	934
Sequence count prior to QC (million)	6.6	2.3	2.9
Average sequence length prior to QC (bp)	255	260	318
Basepair count after QC (Mb)	863	356	825
Sequence count after QC (million)	4.5	1.7	2.7
Average sequence length after QC (bp)	193	209	305
Predicted Protein Features	4,100,901	1,529,791	740,519
Predicted rRNA Features	862,812	444,500	118,458
Identified Protein Features	2,486,860	948,670	417,376
Identified rRNA Features	27,454	121,246	2,362
Identified Functional Categories	2,190,436	765,717	338,065
Archaea (%)	2.8	7.4	12
Bacteria (%)	96.6	91.9	87.3

Table S6. CheckM results for binned genomes showing genome completeness, contamination, and strain heterogeneity.

Location	Bin	Completeness (%)	Contamination (%)	Strain Heterogeneity
B37	*Thermovirga* sp.	99.69	1.85	100.00
A02	*Thermoanaerobacter* sp.	99.37	1.92	20.00
B37	*Methanothermobacter* sp.	99.21	7.45	55.00
B37	*Thermoacetogenium* sp.	95.97	2.69	37.50
A01	*Shewanella* sp.	85.22	0.00	0.00
A02	*Pseudomonas* sp.	75.29	0.00	0.00
B37	*Thermoanaerobacter* sp.	63.24	3.04	46.67

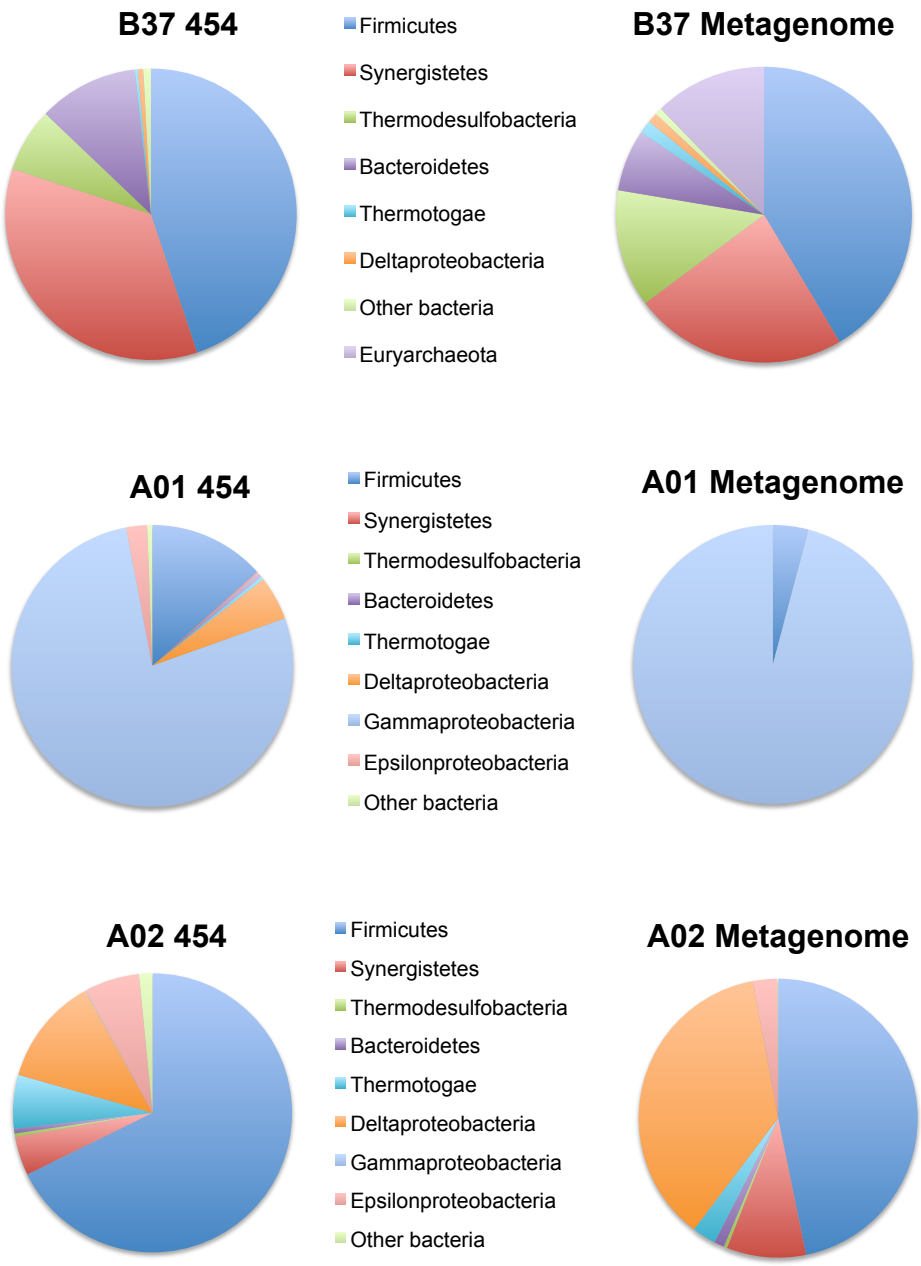


Figure S1. 454 vs Metagenome 16S rRNA Classifications

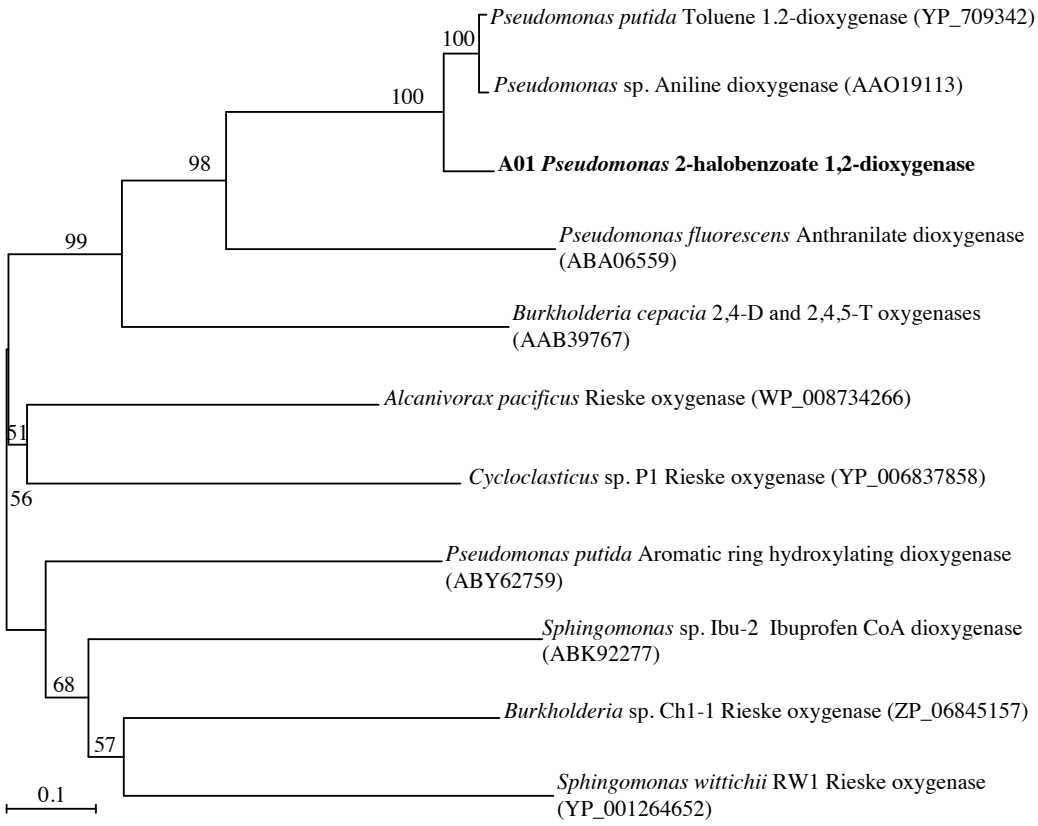


Figure S2. Neighbor-joining tree showing the phylogenetic relationship with representative sequences of α-subunits of the benzoate family of reiske non-heme iron oxygenases from reference strains and those from binned genomes (in bold). Bootstrap values over 50% (n=500) are given next to each conserved node.

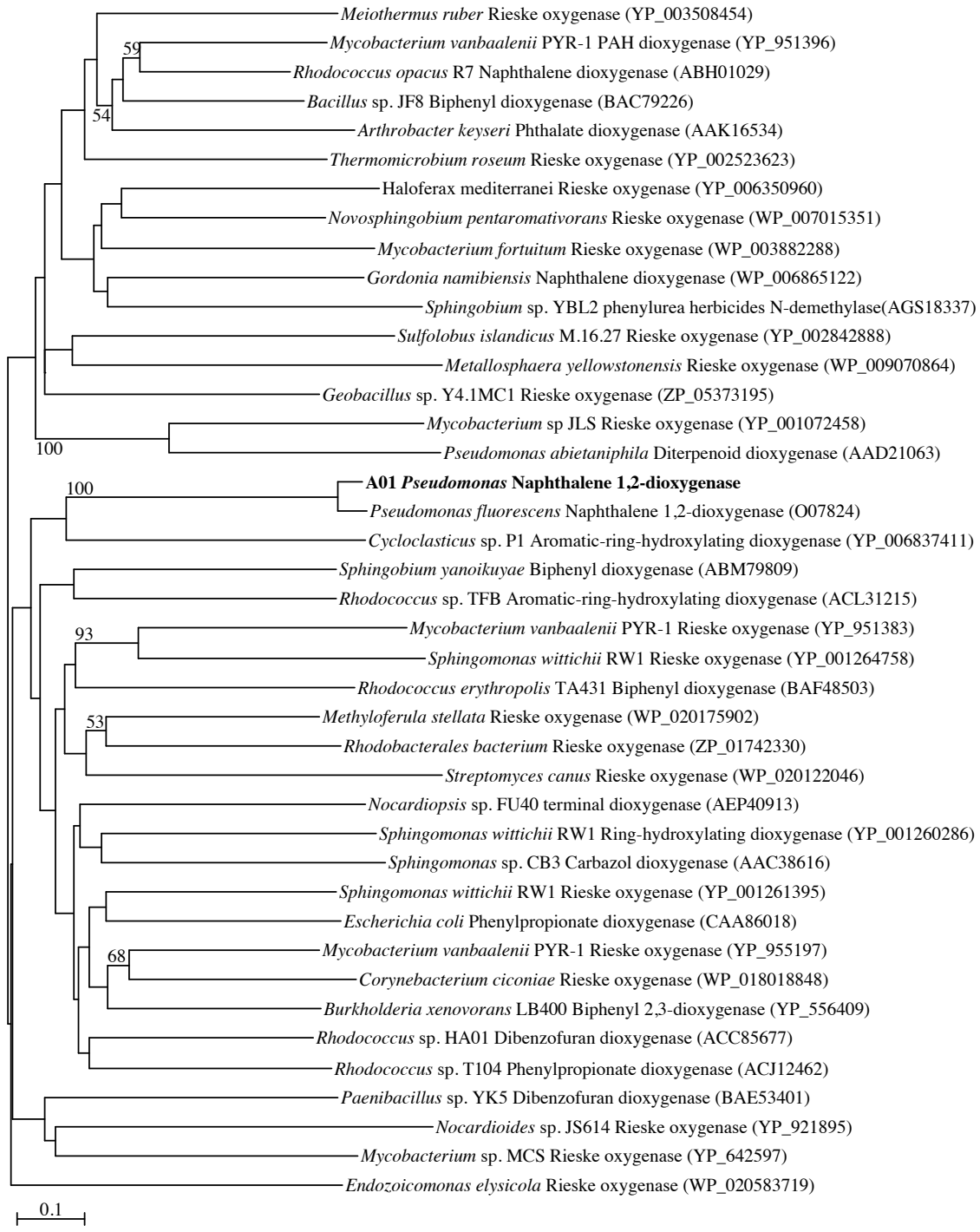


Figure S3. Neighbor-joining tree showing the phylogenetic relationship with representative sequences of biphenyl/naphthalene family of reiske non-heme iron oxygenases from reference strains and those from binned genomes (in bold). Bootstrap values over 50% (n=500) are given next to each conserved node

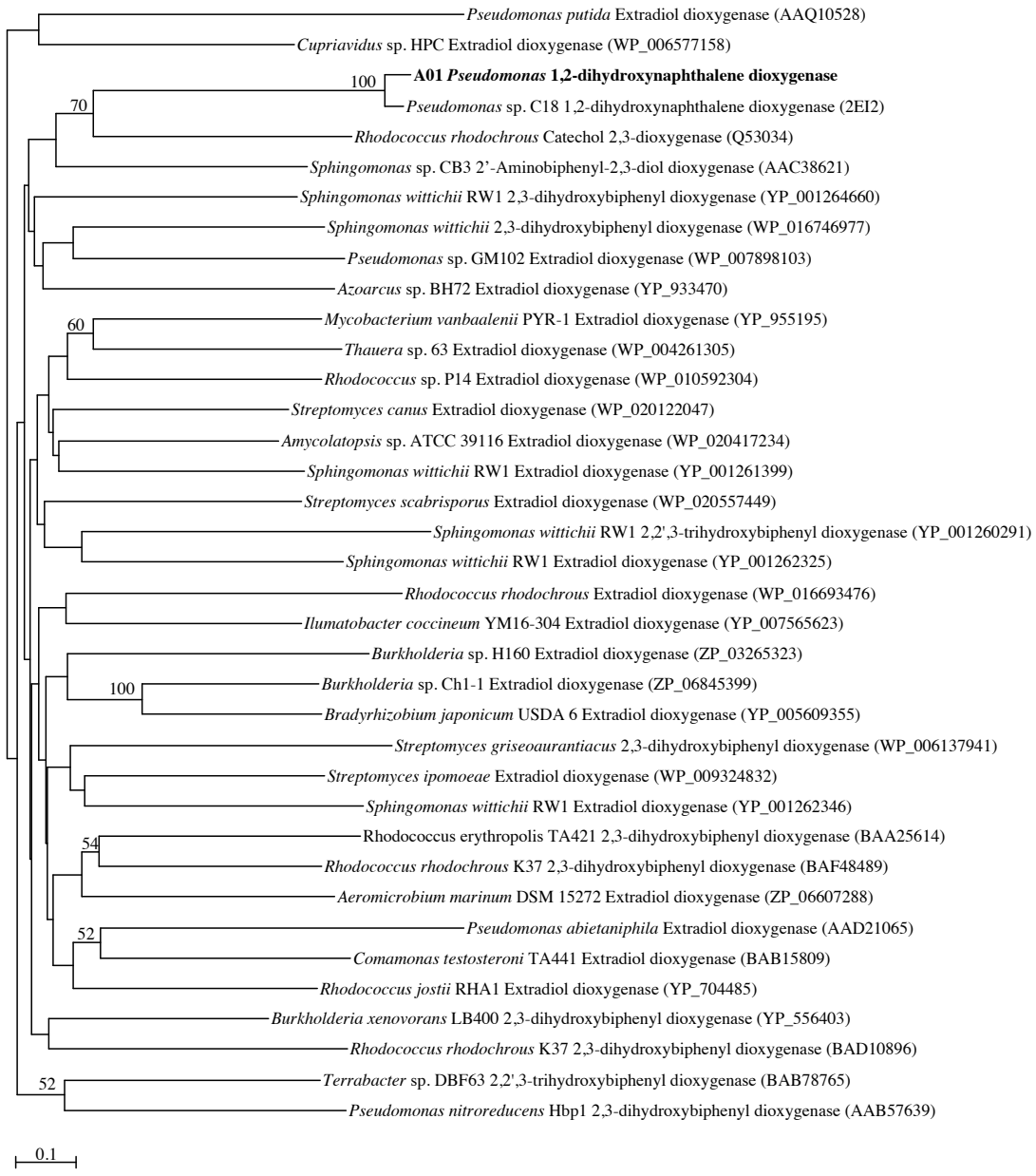


Figure S4. Neighbor-joining tree showing the phylogenetic relationship with representative sequences of extradiol dioxygenases of the vicinal oxygen chelate superfamily, where characterized enzymes typically have a preference for bicyclic substrates, from reference strains and those from binned genomes (in bold). Bootstrap values over 50% (n=500) are given next to each conserved node.

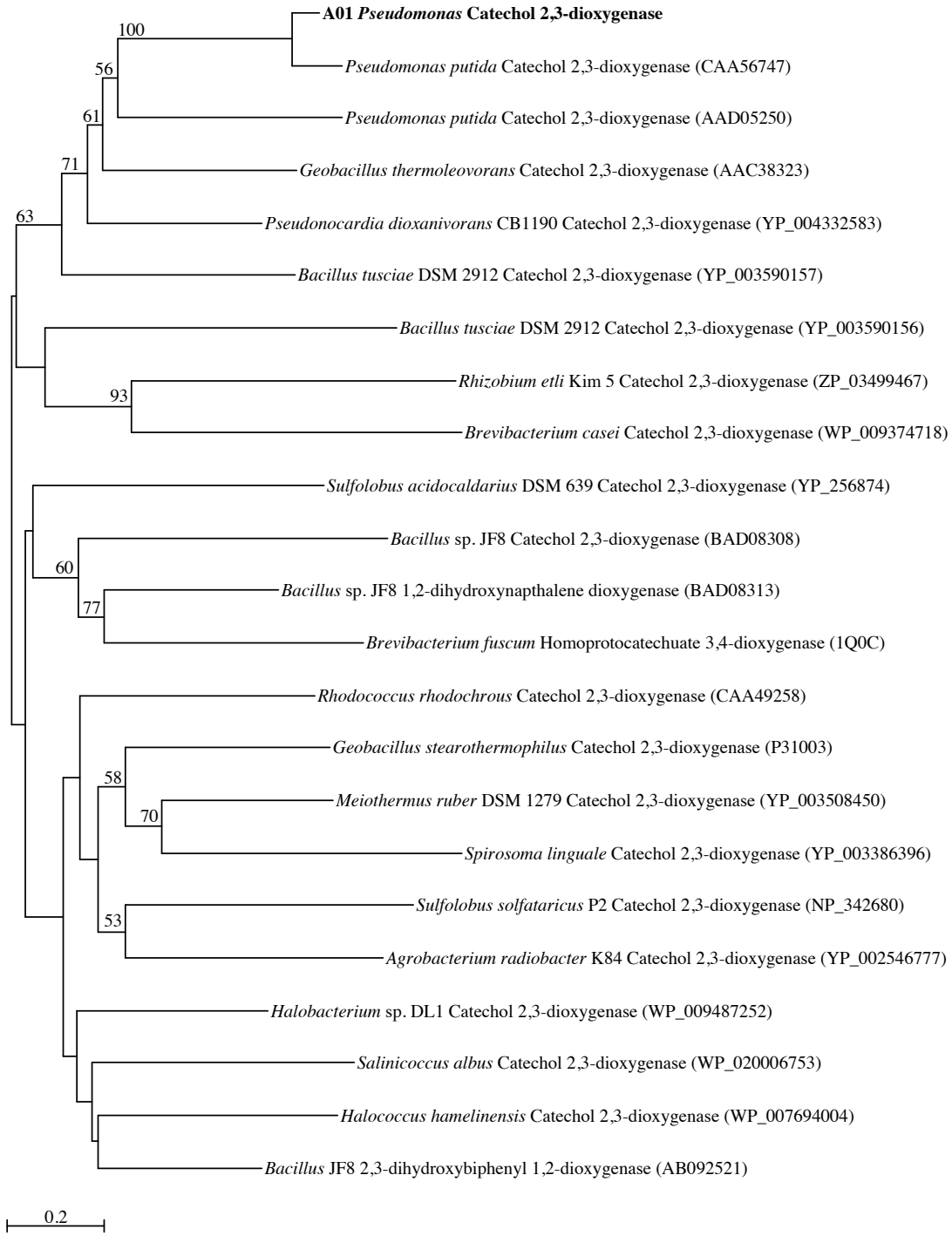


Figure S5. Neighbor-joining tree showing the phylogenetic relationship with representative sequences of extradiol dioxygenases of the viccinal oxygen chelate superfamily, where characterized enzymes typically have a preference for monocyclic substrates, from reference strains and those from binned genomes (in bold). Bootstrap values over 50% (n=500) are given next to each conserved node.

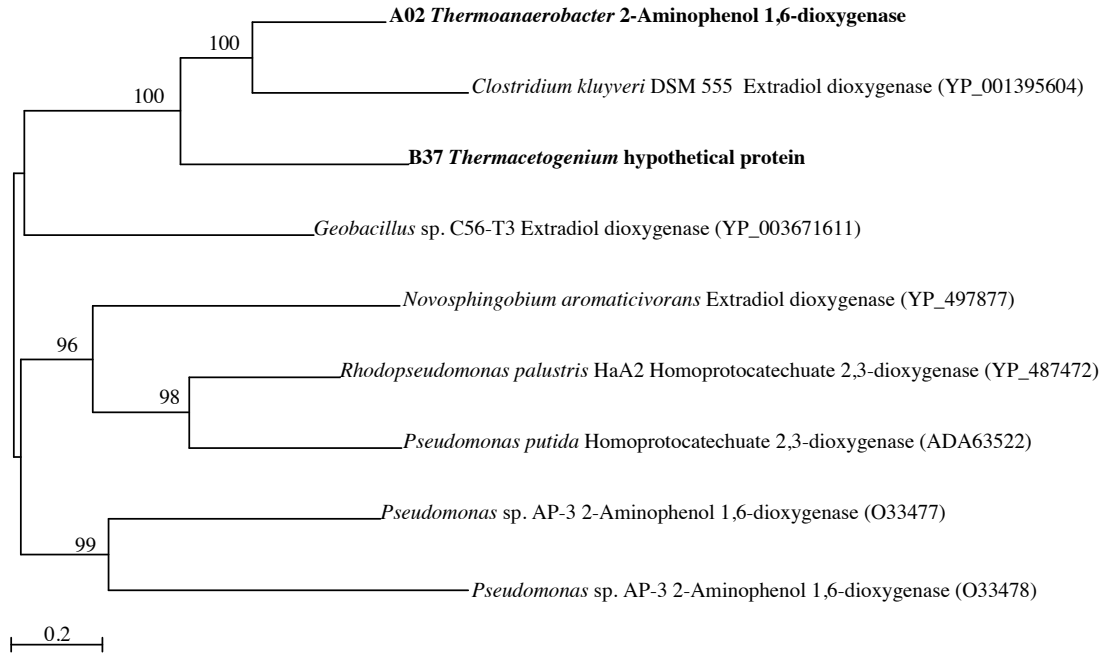


Figure S6. Neighbor-joining tree showing the phylogenetic relationship with representative sequences of the homoprotocatechuate family of LigB superfamily of extradiol dioxygenases from reference strains and those from binned genomes (in bold). Bootstrap values over 50% (n=500) are given next to each conserved node.