UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

ECOLOGICAL INVESTIGATIONS OF HYDROCARBONOCLASTIC MICROBIAL COMMUNITIES ASSOCIATED WITH THE BUILT-ENVIRONMENT

A DISSERTATION

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

Degree of

DOCTOR OF PHILOSOPHY

By

CHRISTOPHER R. MARKS
Norman, Oklahoma
2016

ECOLOGICAL INVESTIGATIONS OF HYDROCARBONOCLASTIC MICROBIAL COMMUNITIES ASSOCIATED WITH THE BUILT-ENVIRONMENT

A DISSERTATION APPROVED FOR THE
DEPARTMENT OF MICROBIOLOGY AND PLANT BIOLOGY

BY

_____
Dr. Joseph M. Suflita, Chair

_____
Dr. Amy V. Callaghan

_____
Dr. Kathleen E. Duncan

_____
Dr. Michael J. McInerney

_____
Dr. Mark A. Nanny

This dissertation is dedicated to my parents, Andrew and Sherry Marks, for their unwavering support and love.

Acknowledgements

I would like to thank my advisor, Dr. Joseph Suflita, for giving me the incredible opportunities that I have been afforded throughout my doctoral program. As his student, I have had the chance to pursue my research interests across many different areas through multi-disciplinary collaborations and develop my skills, not just as a researcher, but also a leader in the field. I would also like to thank the members of my doctoral committee for their guidance through this program and the opportunity to conduct research with each member. It has been a privilege to work and publish with Drs. Callaghan, Duncan, McInerney, and Nanny. Special acknowledgements must go to Drs. Amy Callaghan and Boris Wawrik, who first gave me the chance to collaborate as a Master's student, and who have truly been excellent mentors through many research projects and valued friends.

I would also like to thank all of the fantastic people who I have been fortunate enough to call friends during graduate school. None of my accomplishments would have been possible without Drs. Bryan Crable, Tori Parisi, Chris Lyles, and Cat Isom. Their time and guidance was invaluable in helping me develop as a scientist and I am grateful they have become my closest friends. Finally, I would like to thank my contemporaries with whom I have shared all of the ups and downs that come with finishing a doctorate. I am proud to call Nate Losey, Huynh Le, Josh Cooper, Brian Harriman, Jamie Johnson, Daniel Jones, and Zach Myers colleagues and friends.

Table of Contents

List of Tables

List of Figures

Abstract

Evidence for the microbial metabolism of hydrocarbons is routinely identified in diverse habitats, but particularly those associated with the production, processing, storage and use of petroleum. In these environments, microbial activity can have enormous environmental and financial consequences including oil reservoir souring, biocorrosion of the steel infrastructure, and the accidental release of hydrocarbons to undesired locations. My studies were designed to specifically examine the ecological role of microorganisms in a major oil processing facility experiencing aggressive corrosion, in seawater compensated fuel ballast tanks aboard naval surface warfare vessels and to contribute to fundamental knowledge on how anaerobic microbes are able to metabolize large molecular weight paraffin molecules. An interdisciplinary approach that combined geochemical analyses, molecular microbial ecology methods, and mass spectral-based metabolomics was employed in each of these investigations. Molecular surveys of the oil processing facility revealed a systemic colonization of both dead-leg and bulk fluids by anaerobic taxa primarily affiliated with *Halanaerobiales*. A desalter bulk fluid was a notable exception with members of the *Epsilonproteobacteria*, putative microaerophiles, representing the predominant community members. Geochemical and mass spectral analyses showed steep gradients in salinity, pH, acetate and sulfate concentrations, as well as distinct low molecular weight organic constituent profiles within stratified dead-leg fluids. These gradients contrasted with the highly similar chemical properties observed in the bulk fluids that are often recirculated between the three processing modules. The presence of alkylated monoaromatic-dihydrodiols and *Epsilonproetobacteria* in the desalter

bulk fluid, and the lack of signature anaerobic hydrocarbon biodegradation metabolites confirms that oxygen must be introduced to the resident microflora, most likely at or near the desalter unit. The findings further suggest that anaerobic microbial communities exacerbate localized corrosion by linking the metabolism of partially oxidized aerobic crude oil intermediates from the desalter module to the reduction of oxidized sulfur species. Only by examining both the dead-legs and the bulk fluids was it possible to identify the distinct microbial assemblages across small spatial distances within each module and understand their interactions that ultimately form the basis of the proposed mechanism for microbially-influenced corrosion within this facility.

The selection of hydrocarbonoclastic marine microorganisms under defined ecological conditions was also pertinent to the investigation of seawater-compensated fuel ballast tanks relative to the harbor water used to augment the tanks. The examination of ships containing ballast water of different ages, revealed a pattern of succession that ranged from predominantly aerobic to largely anaerobic microbial taxa with a concomitant decrease in available dissolved oxygen and sulfate reserves. Mass spectral analysis showed the presence of signature metabolites associated with the aerobic or anaerobic activation of mono- and polynuclear aromatic hydrocarbons within the ballast tanks of ships that retained their ballast for 1 week and 32 weeks, respectively. The presence of supersaturating concentrations of dissolved alloying metals in all of the samples, along with paired metagenomic and metabolomic data, revealed that marine microorganisms typically catalyzed the biodegradation of diesel fuel components and exacerbated the biocorrosion of the carbon steel infrastructure within seawater-compensated fuel ballast tanks aboard naval vessels.

While there are thousands of chemicals in petroleum mixtures, the way individual components are metabolized in the absence of oxygen is often enigmatic. Such is the case for large molecular weight alkane molecules that are solid at room temperatures. An anaerobic microbial consortium capable of the methanogenic mineralization of long-chain *n*-paraffins (C_{28}-C_{50}) was investigated using a combined metagenomic and targeted transcriptomic analyses. Experiments were desgined to determine the mechanism(s) of paraffin activation under anaerobic conditions and to elucidate the type of interactions occurring between consortial members. Several draft genomes were binned and assembled from members of the predominant orders *Syntrophobacterales* and *Methanomicrobiales*. Five genotypes of alkylsuccinate synthase A were identified within the metagenome and transcription of each was observed during cultivation of the consortium in the presence of *n*-octacosane as a model substrate. Based on the metabolic reconstruction of the numerically dominant draft genomes, it was proposed that high molecular weight paraffins are activated by addition to fumarate by "*Smithella* sp. SDB" and fermented to acetate through a syntrophic interaction with hydrogenotrophic methanogens. The subsequent mineralization of acetate was proposed to occur via syntrophic acetate oxidation and/or acetoclastic methanogenesis based on additional recovered draft genomes. This is the first report to elucidate the metabolic pathway for such a high molecular weight hydrocarbon and it biochemically implicates a *Smithella* as the responsible organism initiating the anaerobic attack.

The surveillance of the chemical and biological components of artificial habitats associated with petroleum production and consumption revealed that the dynamics of

the resident microbial assemblages were governed by the same principles documented in other natural habitats. More specifically, hydrocarbons are susceptible to biological deterioration by the resident microorganisms when in contact with marine waters under both oxic and anoxic conditions. Distinct micro-environments arise across sometimes small spatial scales within artificial habitats and ultimately select for communities of microorganisms that vary widely in membership and/or metabolic capability. These communities can interact with others through connections based on fluid movement associated with industrial processes and the resulting effects of their activities can be manifested locally or distributed throughout the system. The nature of these effects can largely be assessed *a priori* by examining the dominant electron-accepting processes occurring within each distinct micro-habitat. Ultimately, paired-"omics" investigations can help elucidate contributions of specific taxa to environmental processes and services occurring within such engineered systems.

Preface

Evidence for oxygen-independent microbial metabolism of hydrocarbons is routinely identified in diverse habitats associated with the production, processing, storage and use of petroleum products. These microbial activities drive reservoir souring, biocorrosion of steel infrastructure, and other deleterious phenomena, resulting in environmental deterioration and, often, substantial financial losses for oil industry concerns. The work presented in this dissertation is the culmination of several collaborations, and the interdisciplinary approach taken would not have been possible the contributions of several colleagues.

The study presented in Chapter 1 was conducted by members of the University of Oklahoma Biocorrosion Center with samples supplied by the industrial sponsors of the research consortium. Potential differences in chemistry and microbial populations were investigated between stagnant areas and bulk process fluids from three modules within the oil processing facility through the pairing of microbial community surveys, fluid chemical composition analyses, and mass spectrum-based metabolite profiling. Marked differences in chemical composition (decreased [sulfate] and pH, increased [acetate], and salinity gradients) were observed between bulk and stagnant fluids at each sampling point.

Molecular surveys revealed the microbial communities were predominantly comprised of anaerobic taxa typically associated with petroleum reservoirs, and that species diversity was greater in the stagnant fluids within dead-leg samples than the bulk fluid for each respective processing module. Mass spectral analysis of the production fluids

revealed the presence of metabolites indicative of aerobic hydrocarbon biodegradation, though compounds associated with anaerobic hydrocarbon metabolism were not detected. Evidence from community surveys and metabolite profiling indicated a systemic colonization of the facility by anaerobic halophiles that are metabolizing aerobic hydrocarbon transformation products produced in the desalter module and transported throughout the asset. My specific contributions to this research were the handling of all samples upon arrival, chromatographic analysis of brines, design and supervision of all cultivation efforts, statistical analyses of all relevant data, and crafting of the manuscript and figures presented here. This work has also been accepted for publication in the forthcoming book "Microbiologically Influenced Corrosion in the Upstream Oil and Gas Industry" and is written in the style required for publication by CRC Press.

Chapter 2 was conducted and supported by the Multi-disciplinary University Research Initiative sponsored by the United States Office of Naval Research. Corrosion of metallic infrastructure and biodeterioration of diesel fuels are chronic problem encountered within seawater-compensated fuel ballast tanks aboard naval surface warfare vessels. In order to assess the extent of microbial contribution to diesel fuel biodegradation and infrastructure corrosion, samples were obtained from the ballast expansion tanks of several vessels corresponding to a series of increasing water residence times aboard ship and subjected to metagenomic, metabolomic, and chemical analyses. These samples were then compared to coastal seawater originally used in the tanks. Analyses revealed aerobic hydrocarbonoclastic microorganisms minimally initiate aromatic fuel component biodegradation, consume the available

dissolved oxygen, and are succeeded by anaerobes that utilize fuel or metabolites of aerobic fuel decomposition as electron donors for sulfidogenesis. These metabolic processes result in the biological deterioration of refined diesel quality and the sulfide-mediated biocorrosion of shipboard infrastructure. I was personally involved in collecting all samples for this project, as well as the analysis of all metagenomic data, figure creation, and writing of the chapter presented herein.

In contrast to the negative impacts associated with microbiologically-influenced corrosion, metabolic activities of anaerobic hydrocarbon-degrading microorganisms may be leveraged for industrial benefit. Paraffins, which occur naturally in petroleum reservoirs, are frequently deposited on the interior walls of wellbores, clogging systems and resulting in costly interruptions in production.

Chapter 3 details the study of an anaerobic consortium capable of the methanogenic mineralization of long-chain *n*-paraffins (C_{28}-C_{50}) enriched originally to evaluate the potential for a bioaugmentation-based approach to oilfield paraffin deposition. Metagenomic sequencing and targeted transcriptomic analyses revealed the consortium to be primarily comprised of members of the *Syntrophobacterales* and *Methanomicrobiales*. Putative genes encoding the catalytic subunit of alkylsuccinate synthase (assA) along with several draft genomes were assembled from the metagenome. Expression of the detected assA genotypes in the consortium, including that of the "*Smithella* sp. SDB" draft genome, was demonstrated through reverse transcription-polymerase chain reaction (RT-PCR) analysis during growth on *n*-octacosane. Based on metabolic reconstruction of draft genome assemblies and the expression of *assA*, long-chain *n*-paraffins are proposed to be activated via 'fumarate

addition' and mineralized through the obligate syntrophic cooperation between *Smithella* spp. and hydrogenotrophic methanogens. Chapter 3 and Appendix II were written in the style of the journal Environmental Microbiology. These materials are published in their entirety under the same title in Environmental Microbiology journal volume 18, pages 2604-2619. The copyright release agreement for the reproduction of this material as a component of this dissertation is located at the end of Appendix II. My contributions to this work were the extraction and sequencing of the Illumina metagenome dataset, and working closely with Drs. Callaghan and Wawrik to analyze/annotate all of the metagenomic data and draft genomes, crafting of all figures and most of the text.

Chapter 1: Integrated Methodology to Characterize Microbial Populations and Functions Across Small Spatial Scales in an Oil Production Facility

Abstract

Petroleum reservoir microorganisms directly influence the economics of energy recovery operations through deleterious activities such as souring and biocorrosion. We investigated potential differences in chemistry and microbial populations between stagnant areas and bulk process fluids from 3 processing modules (Low Pressure Separator, Desalter and Water Separator) within an oil processing facility. Samples were collected before and after the flushing of a sample valve at each site. Microbial community surveys were conducted through 16S rRNA gene sequencing and paired with analyses of fluid chemical composition through chromatographic methods and mass spectrum-based metabolite profiling. The production fluids were high saline brines (1.5 – 4.0 M [Cl^-]) with a pH range of 4.5 – 6.2, varying both within and between sampling sites. Sulfate (0.1 – 8.0 mM) and acetate (0.5 – 9.3 mM) were present in all samples, but nitrate was below detection levels. Marked chemical differences were observed between bulk fluids and stagnant dead-leg fluids at each sampling point. Stagnant dead-leg samples exhibited chemical stratification with a chloride gradient of ~3 M in the Desalter Module as well as sulfate and acetate gradients in the Water Separator Module. Molecular surveys revealed that the microbial communities were predominantly comprised of three taxonomic groups: *Halanaerobiales*, *Campylobacterales*, and *Desulfovibrionales*. An increase in species diversity was found in the stagnant fluids from each dead-leg relative to the respective bulk fluid. MPN determinations of heterotrophic fermenting, sulfate-reducing, and

thiosulfate-reducing organisms were all greater than $1x10^4$ cells mL^{-1}. Targeted mass spectral analysis of the production fluids revealed the presence of catechols, phenols, and dihydrodiols, indicative of aerobic hydrocarbon biodegradation. Metabolites associated with anaerobic hydrocarbon biodegradation were not detected. Untargeted metabolomic screening revealed over 1000 identified compounds that mapped primarily to known lipid, carbohydrate and amino acid metabolic pathways. Through the pairing of molecular microbial surveys and advanced metabolite profile analyses, these results suggest a systemic colonization of the facility by anaerobic halophiles, commonly associated with petroleum reservoirs. These organisms were likely cross-fed by organic electron donors produced during the transformation of hydrocarbons by aerobic microorganisms. Furthermore, these results highlight the importance of interrogating the small volumes of stagnated fluids within dead-legs for an accurate assessment of the chemical and biological processes occurring within these problematic sites.

Introduction

Microorganisms colonize diverse habitats throughout the natural and engineered environments associated with the oil and gas industry. The metabolic processes of these organisms can have substantive economic and environmental impacts on the production and processing of petroleum through reservoir souring, product biofouling and infrastructure corrosion. The latter process, also termed microbial influenced corrosion (MIC), can result in the hazardous release of hydrocarbons and brines to the surrounding environment. The association of corrosive sulfide formation and viable sulfate-reducing bacteria (SRB) in oil production fluids was established almost a

century ago through the pioneering work of Bastin et al. 1926. Increased appreciation for the metabolic activities of the oilfield microflora led to the promulgation of formal standards for monitoring populations through cultivation-based serial dilution technique (API 1965). Methods for the selective enumeration of microorganisms exhibiting specific metabolic traits (e.g. sulfate-reduction, acid production, etc.) have been expanded upon and remain an industry monitoring standard (NACE International 2014). Despite their widespread use, microbial population monitoring through growth in serial dilutions of selective cultivation media tends to under-represent the size of a physiological group within a sample for two critical reasons: (i) only organisms capable of growth with the nutrients provided are assayed and (ii) commercially available test formulations often do not adequately represent the sample under investigation. Enumeration media for acid-producing bacteria (APBs), thiosulfate-reducing bacteria (TRBs) and sulfate-reducing bacteria (SRBs) are typically selected on the basis of sample salinity and potentially other factors routinely determined by bulk fluid analyses. However, our study demonstrates that the chemical and biological characteristics of process fluids can vary dramatically over small spatial scales both within and between process modules. Thus, results obtained from cultivation-based monitoring of microbial populations in bulk fluids may not be representative of the communities in chronic problem areas such as pipeline dead-legs. These are piping segments that may be continuously or intermittently exposed to bulk process fluids, but generally receive limited or only intermittent flow. These areas are known to be sites of numerous incidents each year with localized corrosion as one of the primary causes of dead-leg integrity failures (Sloley 2011; Murata, Benaquisto, and Storey

2015).

In an effort to overcome such inherent limitations, cultivation-independent approaches for microbial monitoring including quantitative polymerase chain reaction (qPCR) assays targeting diagnostic genes, microarray analyses, and 16S rRNA gene library sequencing have become more widely accepted as part of an overall asset integrity assurance program (Eckert and Skovhus 2011; Maxwell, Hoffman, and Divine 2007). These methods utilize genomic DNA from field samples for a snapshot of the genetic potential and membership of microbial communities within the asset. Among the major advantages of these techniques are their universal applicability across different sample matrices, eliminating the need for multiple media formulations and a generally shorter analysis time relative to growth-based assays (Maxwell, Hoffman, and Divine 2007). As with any assay, molecular methods also have their particular interpretational limits. The sample collection is from the total community - including both viable and actively metabolizing cells as well as inactive members. In addition, interrogation for specific genes is also limited by the specificity of the probe or primer utilized. Such probes are continuously improved as the discipline evolves, they are often initially designed based on reference molecular sequence data available in curated databases and may not possess the coverage or specificity required to make accurate determinations of the desired *in situ* microbial populations.

In addition to genomic data, fundamental information on the predominant pathways associated with aerobic and anaerobic hydrocarbon metabolism is consistently emerging. Methods for the survey and detection of metabolites are being developed and tested in many petroliferous environments. Metabolites can often be conclusively

identified using mass spectral methods (e.g. GC-MS, HPLC-MS/MS, HPLC-qTOF) that can target "signature metabolites" indicative of specific bioconversions or global surveys of all small organic molecules within a sample matrix (Beller 2000; Lisa M. Gieg and Suflita 2005; Bonifay et al. 2013; Bian et al. 2015). As with the molecular methods, results are interpreted relative to curated databases (e.g. Kyoto Encyclopedia of Genes and Genomes; for additional listing see the Metabolomic Society) that link information on genes and transformation products within metabolic pathways. Such information is rapidly growing as the techniques are applied in an increasing number of studies and environments (Kanehisa et al. 2012). Recently, mass spectral-based metabolite profiling was used in conjunction with other techniques to investigate microbial functioning and corrosion processes in a North Slope pipeline system by interrogating resident communities that were planktonic (Duncan et al. 2009), differentially dislodged during pigging operations (Stevenson et al. 2011) and pipeline section cut outs (e.g. "cookie" samples) (Lenhart et al. 2014). Collectively, the resulting information implicated anaerobic hydrocarbon biodegradation in supporting corrosive biofilms driving MIC within a high-temperature oil production pipeline. In consistent fashion, oilfield-associated microbial assemblages were found to metabolize crude oil under the conditions prevalent in the same reservoirs (Gieg et al. 2010). The integration of field and laboratory samples through the strategic combination of traditional, molecular, and metabolite analyses can, in our opinion, lead to substantive insights into the predominant mechanisms influencing important ecological processes in the oil and gas sector.

Our investigation sought to evaluate the potential for the involvement of

microorganisms in the corrosion of carbon steel infrastructure within an oil processing facility through the integration of geochemical analyses, molecular surveys and metabolite profiling of production brines. The study site is a facility handling production from a sour oil field experiencing aggressive corrosion and pipe failures at dead-legs in separator oily water waste lines (Gunaltun and Kumar 2014). The asset receives high salinity fluids (2.8 – 4.2 M Cl^-; 100 – 150 g L^{-1} Cl^-) as well as H_2S and CO_2. Asset managers estimated corrosion rates to be ~2 – 40 mpy (0.05 – 1 mmpy) based on fluid composition and operational parameters, though rates as high as 315 mpy (8 mmpy) were documented in this system (personal communication). Despite the relatively high salinity of the produced fluids, sulfate-reducing and thiosulfate-reducing microorganisms were historically and repeatedly detected in the bulk separator fluid waste lines through cultivation-based enumerations. Through the combination of chemical and microbiological analyses, the role of partial or complete stagnation in microbial community composition and function were assessed and a potential model for the involvement of microorganisms in the failures of infrastructure integrity is proposed. The information emanating from the data collection and analyses exemplified here may ultimately lead to the generation of models for the *a priori* evaluation of mitigation strategies to control deleterious microorganisms.

Experimental Procedures

Site Description and Sampling

Fluid samples were obtained from near stagnant areas of dead-legs sampling points at the 6 o'clock position of 16" (40.64 cm) oily water disposal lines from three processing modules within the facility (Figure 1). These included an oil-water

separator, a low pressure separator and a desalter (Fig. 1). Average brine composition throughout the facility was reported by the asset manager as (mg L^{-1}): 144718 Cl^-, 493 SO_4^{2-}, 159 HCO_3^-, 0 CO_3^{2-}, 86428 Na^+, 5577 K^+, 2270 Ca^{2+}, 2087 Mg^{2+}, 2 Fe^{2+}, 563 Sr^{2+} and 15 Ba^{2+}. The bulk fluid and dead-leg temperatures were reported as 50°C and 30°C, respectively. System pressure ranges between 1 and 3 bar_g with H_2S and CO_2 concentrations of 4% and 3% mole-basis, respectively. Near stagnant fluids were obtained by collecting three successive aliquots of 30 mL samples into separate screw cap vials from each oily water line. The bulk fluid sample (1 L) was obtained from the same sampling points after flushing several liters of production water through the sampling valve. All samples were stored at 4°C during shipping and until use.

Chemical Analyses

Sample fluids were analyzed for various chemical characteristics including pH, alkalinity and common anion (chloride, nitrate, sulfate and acetate) concentration. The pH of bulk fluid samples was determined with a pH meter (Accumet® Basic AB15; Fisher Scientific, Pittsburgh, PA) at 18°C. The limited volume of the dead-leg sample necessitated the addition of a small aliquot (~20 µL) to a pH test strip at 18°C. Alkalinity was determined by titration for the bulk fluid samples using a Hach Alkalinity Test Kit (Hach Company; Loveland, CO) as described by the manufacturer. Inorganic anions (chloride, nitrate and sulfate) were quantified by suppressed ion-chromatography on a Dionex ICS-3000 system (Dionex; Sunnyvale, CA) equipped with Dionex AS4A-SC guard (4 mm x 50 mm) and analytical (4 mm x 250 mm) columns. Separations were performed isocratically at mobile phase of carbonate buffer (1.7 mM HCO_3^-; 1.8 mM CO_3^{2-}; 2.0 ml min^{-1}) and a suppression current of 35 mA. Chloride was determined on 1:10,000 dilution of sample fluids. Samples for

nitrate and sulfate analysis were pretreated with Dionex OnGuardII Ag and Na cartridges to remove interfering halides as instructed by the manufacturer. Acetate was quantified on a high pressure liquid chromatograph system (Beckman System Gold; Solvent Delivery Module 126, Detector Module 166, and NEC PC-8300 controller) equipped with an organic acid column (Prevail; 5µ; 4.6 mm x 250 mm; Grace Davidson Discovery Science; Deerfield, IL) and operated isocratically with a potassium phosphate buffer mobile phase (25 mM; pH 2.5; 1.0 mL min^{-1}) and detection by UV absorbance at 214 nm.

DNA Extraction, Amplification, and Analyses

Bulk fluid samples were filtered (250 mL; 0.45 micron pore size polyethersulfone filter) in the field. The filters were immediately treated with 1 mL DNAzol (Molecular Research Center, Inc., Cincinnati, OH, USA) to aid cell lysis and prevent nucleic acid degradation, shipped to the laboratory and frozen at -80^{o}C until required. Prior to DNA extraction, the filters were briefly thawed, incubated with proteinase K (Qiagen, Venlo, Limburg) for 15 minutes, rinsed by vortexing successively with nuclease-free water, RNA lysis buffer and RNA dilution buffer from the Maxwell®16 (Promega, Madison, WI) Tissue LEV Total RNA Purification Kit (AS1220). The rinses were pooled and transferred to AS1220 cartridge for DNA extraction. Modifications to the manufacturers standard protocol were used to extract DNA as previously described (Oldham et al. 2012). Amicon 30K Ultra-15 centrifugal filter units (EMD Millipore, Billerica, MA, USA) were used to concentrate 10 mL of each dead-leg sample and to decrease the sample salinity. The eluted DNA was cleaned to remove potential PCR inhibitors using the MO BIO PowerDNA Cleanup Kit (MO BIO Laboratories, Inc., Carlsbad, CA, USA).

The population densities of total bacteria/archaea and selected physiological groups in the bulk fluid samples were estimated using qPCR analysis. Bacterial and archaeal 16S rRNA genes were targeted using the primer pairs Bac27F/338R (Stevenson et al. 2011) and Arc333F/958R (Reysenbach and Pace 1995), respectively. The genetic potential for sulfate reduction and dissimilatory sulfide production were assessed by the quantification of the adenoside-5'-phosphosulfate reductase gene (apsA) by RH1apsF/RH2apsR and the dissimilatory sulfite reductase gene (dsrA) by RH1-dsr-F/RH3-dsr-R, respectively (Ben-Dov, Brenner, and Kushmaro 2007). The StepOnePlus™ Real-Time PCR System was used for thermal cycling and StepOne Software v2.1 (Life Technologies Carlsbad, CA) for data acquisition and analysis. Thermal cycling conditions were as described in the references for the primers. A 1:10 dilution series of a control DNA plasmid containing a reference gene insert for each assay was used to generate a 5–7 point standard curve. Standards were assayed in duplicate and the samples in triplicate.

Bioinformatic & Statistical Methods

Samples were sequenced on the Illumina MiSeq using V2 PE250 chemistry (Illumina, Inc., San Diego, CA). Sequences were joined using the computer program PEAR 0.9.51 (Zhang et al. 2014), prior to being processed using QIIME 1.9.1 (Caporaso et al. 2010) and UPARSE v7.0.1090 (Edgar 2013). Taxonomy was assigned using mothur 1.36.0 (Schloss et al. 2009) and the SILVA r119 SSU RNA gene database (E Pruesse et al. 2007) formatted for use in QIIME. Operational Taxonomic Units (OTUs), defined at 97% similarity, that represented greater than 5 percent of any library were further aligned and classified within SINA (Elmar Pruesse, Peplies, and Glöckner 2012) using the SILVA r123 database and ARB 6.0-rc3 (Ludwig et al. 2004). Alpha

diversity measures included Shannon's diversity index, Shannon's evenness, and the number of observed OTUs. Beta diversity was calculated as a weighted Unifrac distance matrix (Lozupone and Knight 2005).

Correlation analyses were calculated as Spearman's Ranked Correlation Coefficients using the vegan package in R (Oksanen et al. 2016). Redundancy Analysis ordinations (RDA) were computed for both community compositional and untargeted metabolite profile data through the vegan R package within the GUSTA ME web-based application (Oksanen et al. 2016; Buttigieg and Ramette 2014). Analyses of community compositions were performed on relative OTU abundance for bulk fluid and dead-leg #2 samples from each module. Comparisons of untargeted metabolite profiles were conducted on calculated relative peak areas for each resolved feature in the bulk fluid and all dead-leg samples from each module. RDA ordinations were based on the Bray-Curtis distance metric and were constrained by Z-score transformed values of explanatory variables measured for each respective sample (pH, salinity, [sulfate] and [acetate]) with Type I scaling.

Metabolite Extraction and Analyses

The organic constituents in sample fluids were investigated by mass spectrometry using both targeted and untargeted metabolite profiling. The targeted assessment of "signature metabolites" diagnostic of selected microbial metabolic processes was conducted on the bulk fluid samples. A subsample of each brine (1 L) was acidified in the field ($pH \leq 2$; 6N HCl), extracted with three aliquots of ethyl acetate (200 mL) and the organic fractions were separated and retained. The pooled organic fractions were dried over anhydrous sodium sulfate and concentrated to a volume of 100 µL under

flowing N_2. Concentrated organic extracts were reacted with *N,O*-bis(trimethylsilyl)trifluoracetamide to yield trimethylsilyl-derivatized compounds. Metabolites were separated and analyzed by GC-MS and putative identifications were made by comparisons to authentic derivatized standards and the NIST library (Gieg and Suflita 2005; Gieg and Suflita 2002; Elshahed et al. 2001; Duncan et al. 2009).

A more global survey of organic components of both bulk fluid and dead-leg brines was conducted through an untargeted mass spectral analysis. An aliquot from each sample (5 mL) was acidified (pH ≤2), extracted with an equal volume of ethyl acetate, the solvent was evaporated to dryness under N_2, reconstituted with isopropanol (100 µL) and analyzed by HPLC (Agilent 1290) interfaced to a high-resolution MS (Agilent G6538A quadrupole time-of-flight (QToF) MS). A 5 µL injection was made onto a zic-HILIC analytical column for the separation when in positive ion detection mode and onto a reverse phase C_{18}-column when analyzing in negative ion mode. Analysis using this combination of columns has, in our hands, proven effective at increasing the array of compounds detected at usable and reliable abundances.

Mass spectrometry data were turned into feature information by processing through the IDEOM v.19 workflow (Creek et al. 2012), using XCMS centWave (Tautenhahn, Böttcher, and Neumann 2008) for peak detection and mzMatch.R (Scheltema et al. 2011) for alignment of samples. Detected features were matched against the Kyoto Encyclopedia of Genes and Genomes (KEGG) metabolite database (Kanehisa et al. 2012). Statistical analyses applied to global putative metabolite profiles were performed by a metabolomics package in R (De Livera and Bowne 2015). Putatively identified metabolites were mapped to known metabolic pathways using Pathos

(Leader et al. 2011).

Results

Chemical Characteristics of Sample Fluids

The physicochemical properties of bulk and dead-leg fluids from production waters of three oil processing modules (Fig. 1) were interrogated by chromatographic and titration methods. All samples were found to be high salinity brines with varying pH and total chloride concentrations between bulk and dead-leg samples within and between modules (Table 1). Bulk fluid pH and acetate concentrations were very similar between processing modules with an average of 6.03 ± 0.20 and 1.04 ± 0.03 mM, respectively. The pH values agree with those predicted by asset managers of 5.91 – 6.36 based on bulk fluid chemical modeling estimations (personal communication). The salinity and sulfate concentrations of the low pressure separator (4.1 M Cl^-; 4.81 mM SO_4^{2-}) and water separator (4.5 M Cl^-; 4.50 mM SO_4^{2-}) bulk fluids were also very similar, though were dramatically higher than that observed in the desalter fluids (1.5 M Cl^-; 1.03 mM SO_4^{2-}). Nitrate was below detection in all samples retrieved from the oil production platform.

In contrast with the general similarities in the circulating bulk fluids, the relatively stagnant samples in the dead-legs exhibited apparent gradients in their characteristics. The desalter dead-leg sample was highly stratified with respect to salinity with a gradient ranging from 4.0 M (closest to the bulk fluids) to 1.7 M Cl^- near the sampling port. Corresponding gradients in sulfate (3.17 to 1.18 mM), acetate (3.29 to 6.87 mM) and pH (~5.0 to 6.0) were also evident in the desalter dead-leg fluids (Table 1). The stagnant fluids in the low pressure separator dead-legs are starkly different in

physicochemical properties than the bulk fluids, though no gradation is apparent. The pH of the dead-leg fluid is lower than the bulk brine (pH = 4.5 vs 5.7) coincident with the increase in acetate concentration (~ 7-9 fold) greater than that measured in the bulk production fluid from that module. Finally, the water separator samples did not show a gradation in pH or salinity within the dead-leg and are representative of the bulk fluids. No chemical data could be determined for water separator dead-leg sample #3 due to the high oil content. Within this dead-leg, there was also an inverse correlation between sulfate loss and acetate production in samples #1 and #2. Interestingly, sample #2 was very similar in chemical properties to the bulk fluid and may represent mixing of the bulk process fluids during the sampling event.

Relationships between brine properties were investigated by calculating Spearman correlation coefficients for each parameter across all samples (Fig. 2). Several parameters exhibited strong positive or negative correlations including: salinity and sulfate (0.78), pH and acetate (-0.79) and between acetate and sulfate (-0.81). A weaker negative correlation was observed between flow and acetate (-0.58). The relationships between flow, acetate and sulfate suggests that the partial stagnation in the dead-legs may be coincident with acetogenesis potentially linked to sulfate utilization.

Microbial Community Characterization

The microbial community composition was investigated by quantifying genetic markers by qPCR. The trends in population structure and membership were assessed via high-throughput sequencing of 16S rRNA genes. Amplicons were successfully produced from all bulk fluid samples and several of the dead-leg samples, despite the

relatively small sample volume (10 ml) available for analysis. The total microbial population in each of the bulk fluid samples was quantified based on qPCR assays targeting the bacterial and archaeal 16S rRNA genes. Microbial population density ranged from $10^5 – 10^8$ copies 16S rRNA gene mL^{-1} process brine, with the largest populations found in the desalter bulk fluid (Table 2). No archaeal 16S rRNA gene or *mcrA* gene amplicons and no mitochondrial 16S rRNA genes (fungal or other) were obtained from any of the samples assayed, suggesting that microbial populations were composed exclusively of Bacteria. The genetic potential for sulfate reduction was quantified by *aprA* abundance and found to be present in all bulk fluid samples ranging from $10^1 – 10^4$ copies *aprA* gene mL^{-1} brine with the greatest abundance in the low pressure separator. The potential for dissimilatory sulfide production was quantified by *dsrA* genes that were found to be present in all bulk fluid samples ranging from $10^4 – 10^6$ copies *dsrA* gene mL^{-1} brine with the greatest abundance also in the low pressure separator. The gene abundance for *dsrA* was several orders of magnitude greater than *aprA* within each of the bulk fluids suggesting that sulfoxy-anions utilization other than sulfate may potentially support more biological sulfide production.

The diversity and composition of each of the microbial communities in the facility were interrogated through the analysis of 16S rRNA gene amplicon libraries for the bulk fluid and dead-leg sample #2 for each processing module. Metrics of community diversity for each library are shown in Table 2. In each module, the dead-leg community had greater species richness (number of OTUs observed at the rarified sequencing depth), evenness and total diversity relative to the corresponding bulk fluid

microbial assemblages. The low pressure separator dead-leg harbored the most diverse microbial population with respect to both number of taxa detected and the evenness of their abundance distribution despite having the lowest measured pH value of 4.5.

Analysis of entire microbial communities based on the taxa observed in each sample as well as their relative abundance in each library allowed for the comparison of compositional differences. Pairwise dissimilarities calculated between each of the bulk fluid and dead-leg #2 sample are shown in Figure 3. The order-level taxonomic composition of each microbial population revealed that members of the *Halanaerobiales* accounted for ~50% of all reads obtained in this study and, along with the *Campylobacterales* and *Desulfovibrionales*, were present at substantive levels (>1.5% relative abundance) in all samples (Fig 3). The low pressure separator bulk fluid community was predominantly composed of *Halanaerobiales* (73.4%), *Desulfovibrionales* (9.4%), and *Campylobacterales* (3.6%) while the more evenly distributed and species rich dead-leg was populated by the *Coriobacteriales* (10.8%), *Halanaerobiales* (9.7%), *Flavobacteriales* (7.6%), *Campylobacterales* (5.2%) and *Desulfovibrionales* (1.8%). Similarly, *Halanaerobiales* (66.6%), *Campylobacterales* (6.6%), and *Desulfovibrionales* (4.0%) were the most abundant taxa in the water separator bulk fluid assemblage. As in the low pressure separator, but unlike the respective bulk fluids, the water separator dead-leg community is highly enriched in members of the *Gammaproteobacteria* with unclassified *Gammaproteobacteria* (28.0%), *Bacillales* (12.0%), *Halanaerobiales* (9.1%), and *Pasteurellales* (7.9%) among the most abundant orders. Despite the strong similarities in bulk fluid

community structure and membership between the low pressure and water separator communities (95.1% pairwise similarity), partial stagnation did not enrich for similar microbial populations within the dead-legs of these modules. In contrast, the desalter bulk fluid community harbored relatively few taxa, with the *Campylobacterales* (71.5%) as the dominant taxonomic group along with the *Halanaerobiales* (16.1%) and the *Desulfovibrionales* (1.6%). This was distinctly different than all other analyzed assemblages with pairwise dissimilarities exceeding 19.7% in all cases. For comparision, the dead-leg community at the desalter is highly similar to the bulk fluids of the low pressure and water separators (pairwise dissimilarities of 2.9% and 4.6%, respectively) and was again primarily populated by *Halanaerobiales* (67.6%), *Desulfovibrionales* (11.1%) and *Campylobacterales* (5.6%).

Further phylogenetic analysis was performed on select ubiquitous OTUs present in each sample to provide increased taxonomic resolution. Two closely related OTUs (0.03% pair-wise distance) classified within the *Halanaerobiales* (OTU 2 and OTU 296) accounted for the majority of sequences affiliated with this order observed in this study. *Halanaerobium congolense* is the closest described type strain to OTU 2 (Fig. 4A) and relative abundances ranged from 5.1% in the water separator dead-leg to 53.8% in water separator bulk fluid. In contrast, OTU 296 was most abundant in the desalter dead-leg (33.2%) and low pressure separator bulk fluid (32.7%) with the lowest abundance found in the low pressure separator dead-leg (0.28%) and water separator dead-leg (0.31%). This OTU is equivalently related to *H. saccharolyticum* and *H. lacusrosei* (Fig. 4A). The *Campylobacterales* are represented by OTU 1, affiliated with *Sulfurospirillum arcachonense* (Fig. 4B), that accounts for 3.4% to

71.4% relative abundance in the low pressure separator bulk fluid and desalter bulk fluid, respectively. Finally, OTUs 3 and 6 are the most abundant representatives of the *Desulfovibrionales* and are most closely related to members of the *Desulfohalobiaceae*, *Desulfonatronovibrio thiodismutans* and *Desulfohalobium utahense*, respectively (Fig. 4C). Both OTU 3 and 6 follow similar enrichment patterns between samples having the highest abundance in the desalter dead-leg and low pressure separator bulk fluids and the lowest in the water separator dead-leg brines.

Redundancy analysis was performed to examine the relationship between the geochemical parameters measured for each sample and the community composition, based on OTU relative abundance (Fig. 5A). The explanatory variables pH, salinity, acetate and sulfate concentrations accounted for 61.66% of the total community dataset variance, and components RDA 1 and RDA 2 together explained 59.3%. Salinity is strongly correlated to RDA 1 and pH to RDA 2, with biplot scores of 0.91 and -0.82, respectively. The ordination shows that highly similar brine chemistries have enriched very similar microbial populations within the desalter dead-leg #2 and the low pressure and water separator bulk fluids, despite the difference in fluid flow.

Metabolite Analyses

Metabolite profiling was performed on each of the brine samples by HPLC-QToF mass spectrometry to survey the low molecular weight organic compounds (presumably putative metabolites) present in each sample, while targeted analyses for compounds indicative of selected microbial catabolic processes was by GC-MS. The untargeted survey resolved 10,115 chemical features and tentatively identified 1,112

compounds based on the comparison of exact masses to a reference database. The brines contain a suite of primary aliphatic acids, ranging from C_{10} to greater than C_{22}, and numerous organo-nitrogen compounds (e.g. amino acids, amines and amides). Features annotated as $C_{12}H_{17}N$ and $C_{14}H_{21}N$ were detected in most of the sample fluids, but were prominent components of the water separator bulk and dead-leg fluids. These features were not able to be identified as specific compounds, but may be secondary or tertiary amines. Two compounds associated with carbohydrate metabolism (tartaric acid and L-arabinonate) were found in high abundance in the dead-leg samples, but were not detected in any of the bulk fluid samples nor water separator dead-leg #2.

Redundancy analysis was conducted on metabolite profile feature relative peak areas for each sample and constrained by the corresponding geochemical parameters: pH, salinity, acetate and sulfate concentrations (Fig. 5B). The explanatory variables represented for 61.74% of the total community dataset variance, and components RDA 1 and RDA 2 accounted for 42.1%. Acetate concentration showed the strongest correlation with RDA 1 and pH with RDA 2, with biplot scores of 0.79 and 0.95, respectively. RDA indicated essentially identical metabolite profiles were observed for the low pressure separator and water separator bulk fluids, and the water separator dead-leg #2 was similar to these bulk brines. Low pressure separator dead-leg samples contain similar metabolite composition and have similar geochemical characteristics, despite a gradient in acetate concentrations. The metabolite profiles for the desalter dead-leg brines are highly variable and are strongly dissimilar between each other (in agreement with strong chemical gradation), and most other samples interrogated.

Overall, ordination analysis revealed that despite some geochemical differences, the bulk fluids within these processing modules had a largely similar metabolite profile in contrast to their microbial assemblages. Conversely, dead-leg fluid metabolite compositions are highly variable within and between processing units.

Targeted metabolite profiling was used to screen for hydrocarbon biotransformation products in the bulk fluid samples. All bulk fluid samples contained alkylated monoaromatic hydrocarbons, aromatic alcohols (e.g. cresols, phenols) and a suite of alkanoic acids. Alkyl-, benzyl-, and naphthylmethyl-succinic acids indicative of anaerobic hydrocarbon activation by fumarate addition were not detected in any sample. Alkylated catechols and other benzenedihydrodiols were identified in the desalter bulk fluid indicating aerobic biodegradation of monoaromatic hydrocarbons at or near this site.

Discussion

This case study represents an integration of traditional and modern investigative tools to assess the contribution of microbial activity to carbon steel corrosion in dead-leg areas within several processing modules in a sour oilfield facility. Analysis of brine chemistry, molecular surveys of microbial community composition and functional potential, in conjunction with metabolite profiling efforts revealed compositional differences, both chemical and biological, within and between the flowing bulk fluids and partially stagnated dead-leg areas in each of the oil processing modules. The interconnectedness of the various processing modules (Fig. 1) was verified by the chemical properties of the bulk fluids from each module (Table 1). That is, the bulk oily processing waters were generally very similar in their chemical characteristics,

with the expected exception of decreased salinity and sulfate content in the desalter. Despite the similarity in bulk fluid properties, the partially stagnated brines in the dead-legs had significantly different and distinct chemical characteristics (Table 1; Figure 3). The inverse correlation between sulfate loss, acetate formation and pH decrease in the low pressure separator and desalter dead-legs suggested that biological sulfate reduction may have occurred over small spatial scales.

Molecular surveys identified microbial assemblages inhabiting the brine in both the bulk and dead-leg fluids in all modules. Despite the similar chemical characteristics in the low pressure separator and the water separator, partial stagnation did not promote the enrichment of a common microbial community in the respective dead-legs on those modules. The dead-leg communities were found to be more diverse than the respective bulk fluid in each module (Table 2). This increase in species richness and evenness may be promoted by stagnation, removing shear forces at the pipe wall as a selective pressure, and/or the decrease in temperature of 20°C in the dead-legs relative to the bulk process fluids. Throughout the oilfield processing facility, dead-legs primarily served as monthly sampling points for system monitoring, but sampling efforts on individual modules need not be coordinated (personal communication). Given the potential differences in fluid residence time between these dead-legs, it is possible that these samples represent different points along a shared successional trend toward a common microbial community. Though this trend is speculative at this time, it constitutes an area deserving of further research and may yield valuable information regarding the rise of deleterious communities within dead-legs.

Regardless of structure and diversity differences between oilfield module

communities, members of a few select taxonomic orders, *Halanaerobiales*, *Campylobacterales* and *Desulfovibrionales*, were ubiquitous in all samples. Members of the *Halanaerobiales* accounted for a substantive fraction, and often the largest portion of each community interrogated, except the desalter bulk fluids. These anaerobic halophiles are commonly detected and isolated from petroleum-impacted environments including oil reservoirs and production infrastructure (Neria-González et al. 2006; Sette et al. 2007; Gales et al. 2011; Bhupathiraju et al. 1999; Bhupathiraju et al. 1994; Ravot et al. 1997; Liang et al., 2016). Members of the *Halanaerobiales* are well known halophiles that thrive in highly saline habitats and typically produce acetate as a major fermentation end-product of complex organic substrates including carbohydrates and yeast extract. To date, no species has been shown to utilize sulfate as an electron acceptor though many have the capacity to reduce thiosulfate to sulfide during heterotrophic growth (Whitman et al. 2015). Despite their common association with petroleum-laden environments, they are not known to metabolize hydrocarbons and the ecological role of these organisms in these habitats remains unknown. A previous study of corrosive biofilms on steel coupons in a Mexican oilfield pipeline revealed the presence of taxa related to the genus *Halanaerobium* in conjunction with the more predominant *Enterobacteriaceae* at a brine salinity of 60 – 80 g L^{-1} NaCl (Neria-González et al. 2006). More recently, Liang et al. found highly saline (77 – 200 g L^{-1} NaCl) production fluids from hydraulically fractured natural gas wells harbored a microbial community dominated by *Halanaerobiales* (Liang et al., 2016). Subsequent corrosion studies conducted on the isolate *Halanaerobium* sp. DL-01 implicated these taxa in the corrosion of carbon steel infrastructure through the production of acetate

and sulfide from the metabolic degradation of guar gum coupled to the reduction of thiosulfate within the formation. Murali Mohan et al., (2013) examined the changes in chemistry and microbial populations in production fluids with time after well completion in the Marcellus Shale formation. The salinity of produced brines increased from ~0.2 – 2.6 M NaCl (10 – 150 g L^{-1} NaCl) and microbial community composition changes mirrored the salinity dynamics with the *Halanaerobium* spp. initially being relatively minor populations but increased in relative abundance to represent >99% of observed taxa at ~150 g L^{-1} NaCl (Murali Mohan et al. 2013).

In contrast, the bulk fluids from the desalter module harbored the only community predominantly populated by *Epsilonproteobacteria* rather than *Halanaerobiales*. This broad taxonomic group is frequently observed in production fluids from petroleum reservoirs (Grabowski et al. 2005; Voordouw et al. 1996; Hubert et al. 2012) and members of the genera *Sulfurospirillum* (Kodama, Ha, and Watanabe 2007), *Sulfurimonas* (Gevertz et al. 2000), *Acrobacter* (Gevertz et al. 2000), and *Sulfuricurvum* (Kodama 2004) have been cultivated from these habitats. Many of these organisms are capable of the oxidation of reduced sulfur species coupled with the reduction of nitrate. Such metabolic capabilities spur interest in the use of nitrate for reservoir souring control purposes. Sulfate-reducing prokaryotes are known to be inhibited by the metabolic activities of nitrate-reducing sulfur-oxidizing bacteria of the *Epsilonproteobacteria* (Hubert and Voordouw 2007; Telang, Jenneman, and Voordouw 2011; Voordouw et al. 2013). *Sulfurimonas denitrificans* CVO, formerly *Thiomicrospira denitrificans* (Takai 2006), is known to oxidize sulfide via the sulfur-oxidation (Sox) pathway to elemental sulfur or sulfate under both denitrifying and

microaerophilic conditions (Sievert et al. 2008; Gevertz et al. 2000). Though *S. denitrificans* CVO, *Arcobacter* sp. FWKO B, *Sulfurospirillum cavolei*, and *Sulfuricurvum kujiense* were all isolated from fluids associated with subsurface petroleum habitats, growth of each organism is inhibited at >1% salt. Clearly, our results attest that other members of the *Epsilonproteobacteria* are able to tolerate more highly saline conditions (~1.5 M Cl^-) in the desalter production fluids. In this regard, *Epsilonproteobacteria* increased in relative abundance from <2% - 17% when produced brine salinity reached ~10% NaCl (~1.8 M Cl^-) (Murali Mohan et al. 2013). Since nitrate was below detection in all production samples, this likely suggests a local oxygen ingress point, the presence of a microaerophilic population of halotolerant sulfide-oxidizing bacteria and potentially an area of active sulfur cycling within the facility. There is a pump immediately upstream of the desalter to increase the low pressure separator outflow. A weak or faulty seal in this unit could conceivably account for localized available oxygen in the desalter bulk fluids.

Metabolites and putative metabolic activities

Untargeted metabolite profiling showed the brines to contain a plethora of small organic compounds of numerous chemical classes. The observation of acidic fermentation intermediates such as tartaric acid and L-arabinonate exclusively in the dead-leg samples and at high concentrations is of particular note. Some species of Gram negative soil taxa have been shown to produce these carboxylic acids to solubilize phosphate minerals (Yi, Huang, and Ge 2007). These organic acids were also shown to be highly effective agents for bioleaching iron, nickel and cobalt from mineral ores under acidic conditions (McKenzie, Denys, and Buchanan 1987).

Targeted metabolite analysis revealed the presence of alkanoic and aromatic acids, BTEX hydrocarbons, ethylene glycols, and hydroxylated aromatic compounds (e.g. cresols and phenol) within the bulk fluids from each of the modules. Interestingly, catechols and alkylated aromatic dihydrodiols were identified solely in the desalter bulk fluid. These molecules are produced by the incorporation of molecular oxygen into the aromatic ring of BTEX hydrocarbons by mono- and dioxygenase enzymes, and represent signature metabolites indicative of the aerobic oxidation of aromatic hydrocarbons.

Taken together with the identification of taxa phylogenetically associated with known sulfur-oxidizing microaerophiles at this site in contrast to the strictly anaerobic taxa found within all other brines from each module, this suggests that oxygen ingress is occurring within the desalter module. These metabolites and potentially other oxygen-bearing organic molecules (e.g. aromatic acids, cresols, phenols, and di/tri-ethylene glycols) that were detected in each of the processing modules may serve as electron donors to the resident microflora driving localized production of acetate and sulfide throughout the facility. Similarly, Liang et al., proposed a scenario where *Halanaerobium* spp. produced acetate and sulfide during metabolism of guar gum in a hydraulically fractured natural gas reservoir, and ultimately resulting in the pitting corrosion of transport pipelines downstream of the wellhead and far removed from the catalytic organism(s) (Liang et al., 2016).

Comparison of the community and metabolite profile ordinations revealed interesting trends between microbial population membership and metabolic activity (Fig. 5). The water separator surge drum receives oily water from both the desalter and the low

pressure separator, and some fluids are specifically recirculated between the water and low pressure separators (Fig. 1). Thus, it is not surprising that the bulk fluids of the low pressure and water separators share very similar chemical characteristics (Table 1), and have enriched for essentially identical populations of microorganisms (Fig. 3A). Further, the metabolite profiles of these biologically equivalent bulk brines indicate identical metabolic processes occurring within these spatially segregated communities and a shared ecology exists between certain processing modules (Fig. 5). The decreased salinity and potential oxygen ingress at the desalter would account for the dramatically different taxa and metabolite profiles observed in the bulk fluid of this module compared to the low pressure and water separators despite the brine circulation pattern. Interestingly, significantly different metabolite profiles were recovered from the desalter dead-leg samples, despite possessing an identical microflora to the low pressure and water separator bulk fluids (Fig. 5). Thus, community membership alone should not be regarded as a surrogate for metabolic activity in MIC monitoring regimes.

Proposed mitigation strategies and hypothesized impact on communities

Through the combination of geochemical, molecular and metabolomic investigations, a model for microbial involvement in corrosion activities within this asset can be proposed. Asset managers had previously predicted corrosion rates if no mitigation actions were taken to be 0.5 to 1 mmpy, but these models cannot account for microbial activity or the biological formation of iron sulfides within the system and frequently under-represent observed pitting rates in the field (Bonis and MacDonald 2015). As fluids stagnate within the dead-legs, *Halanaerobiaceae* spp. and *Desulfohalobiaceae* spp. may oxidize organic substrates circulating in the bulk fluids to acetate coupled to

sulfidogenesis from thiosulfate and sulfate. Bonis and MacDonald (2015) examined cases of recurrent corrosion within fields producing H_2S and CO_2 to elucidate factors determining the risk to infrastructure degradation. Their analysis indicated that impedances in the kinetics of FeS precipitation, iron sulfide deposition, and galvanic effects induced by local chemical differences promoted aggressive localized corrosion within dead-legs with little to no impact to the surrounding infrastructure (Bonis and MacDonald 2015). Based on the proposed activities of the indigenous microflora, metabolic production of sulfide would drive the formation of inorganic sulfide particles and a decrease in local fluid pH associated with the excretion of acidic metabolic end products. The incomplete oxidation of organic electron donors results in the production of organic acids (e.g. acetic acid, L-arabinonic acid, tartaric acid, etc.), or their corresponding conjugate bases depending upon environmental pH, within the stagnated zones. The consequences of these metabolic activities could, in concert with local chemical changes within the dead-legs, act to impede the precipitation kinetics of FeS at the metal-fluid interface. As mentioned previously, dead-legs experience a cooling of approximately 20°C relative to the bulk fluids. In addition, the process brines contain high concentrations of divalent cations (e.g. Ca^{2+} and Mg^{2+}). Cooling and precipitation with alternative cations tend to impede the formation of protective FeS layers (Crolet and Bonis 2010; Bonis and MacDonald 2015). These factors and locally decreased pH values, arising from both biotic and abiotic reactions, would favor the formation of less protective scales (i.e. mackinawite) rather than pyrite. Furthermore, the composite nature of mixed inorganic anions, variable cation mineral molecular volumes, limited diffusion of inhibitors from bulk fluids and the

physical presence of microbial cells may likely result in the formation of a highly porous scale layer, thus allowing for continued under-deposit dissolution of the pipe surface. This suggests that microorganisms indirectly exacerbate corrosion within the dead-legs of this asset by creating differential chemical environments, preventing the formation of a protective FeS layer and probably promoting galvanic coupling with the surrounding FeS. Ultimately, the combination of biotic and abiotic factors promotes aggressive localized anodic corrosion beyond predicted rates.

Surveys of microbial communities and/or metabolites can be correlated with field metadata (e.g. brine chemistry, operational parameters, documented corrosion rates, etc.) through recursive portioning and/or constrained ordination analyses such as canonical correspondence analysis (CCA), principal components analysis (PCA) and redundancy analysis (RDA). These statistical analyses can offer initial insights into relationships between community composition and environmental variables and are reviewed elsewhere (Legendre and Legendre 2012). The observed correlations can then serve as the basis for hypothesis generation and in conjunction with multiple linear regressions and generalized linear models targeting response variables (specific taxa, community structures, etc.) of interest. These methods provide models for the extrapolation of the selected variable in response to environmental parameter changes *a priori*. In the future, this approach could allow asset managers to evaluate the potential impacts of operational changes upon resident microflora and potentially be used to further refine corrosion risk models to incorporate activities driven by microbial processes.

The combined application of molecular methods for microbial community surveys,

geochemical analyses, and mass spectral-metabolite profiling provides deep insights into the microbial processes occurring within an environment. This case study utilized these paired datatypes to investigate biological processes occurring within an engineered environment that ultimately led to the proposal of a mechanism for microbiologically-influenced corrosion within an oil-processing facility experiencing aggressive localized corrosion of carbon steel. Continued research involving expansive surveys of brines throughout field assets is necessary to further develop and test hypotheses linking microbial populations to deleterious activities. This study was conducted exclusively on planktonic fluid communities. Biofilms are also widely recognized as catalysts for biocorrosion and this same investigative approach can be used to explore the ecology and physiology of such assemblages and their impact on infrastructure integrity. Several specific questions remain, such as whether dead-legs select for common microbial communities through a predictable succession pattern and can specific taxa or chemicals serve as indicators of corrosive conditions *in situ*. Ultimately, an increased understanding of the microbial ecology of these engineered habitats will offer valuable data for asset protection planning and production assurance.

References

API, American Petroleum Institute. 1965. “API Recommended Practice for the Biological Analysis of Subsurface Injection Waters.” In , 2nd ed. New York: American Petroleum Institute.

Bastin, Edson S., Frank E. Greer, C. A. Merritt, and Gail Moulton. 1926. “The Presence of Sulphate-Reducing Bacteria in Oil Field Waters.” *Science (New York, N.Y.)* 63 (1618): 21–24. doi:10.1126/science.63.1618.21.

Beller, Harry R. 2000. “Metabolic Indicators for Detecting in Situ Anaerobic Alkylbenzene Degradation.” *Biodegradation* 11 (2-3): 125–39. doi: 10.1023/A:1011109800916.

Ben-Dov, Eitan, Asher Brenner, and Ariel Kushmaro. 2007. “Quantification of Sulfate-Reducing Bacteria in Industrial Wastewater, by Real-Time Polymerase Chain Reaction (PCR) Using *dsrA* and *apsA* Genes.” *Microbial Ecology* 54 (3): 439–51. doi:10.1007/s00248-007-9233-2.

Bhupathiraju, VK K., Aharon Oren, P. K. Sharma, Ralph S. Tanner, Carl R. Woese, and Michael J. McInerney. 1994. “*Haloanaerobium salsugo* Sp. Nov., a Moderately Halophilic, Anaerobic Bacterium from a Subterranean Brine.” *International Journal of Systematic Bacteriology* 44 (3): 565–72. doi:10.1099/00207713-44-3-565.

Bhupathiraju, VK K., Michael J. McInerney, Carl R. Woese, and Ralph S. Tanner. 1999. “*Haloanaerobium kushneri* Sp. Nov., an Obligately Halophilic, Anaerobic Bacterium from an Oil Brine.” *International Journal of Systematic Bacteriolology* 49 (3): 953–60.

Bian, Xin-Yu, Serge M. Mbadinga, Yi-Fan Liu, Shi-Zhong Yang, Jin-Feng Liu, Ru-Qiang Ye, Ji-Dong Gu, and Bo-Zhong Mu. 2015. “Insights into the Anaerobic Biodegradation Pathway of N-Alkanes in Oil Reservoirs by Detection of Signature Metabolites.” *Scientific Reports* 5 (January). Nature Publishing Group: 9801. doi:10.1038/srep09801.

Bonifay, Vincent, Jan A. Sunner, Whitney Smith, and Iwona B. Beech. 2013. “Metabolomics Study of Corroding Microbial Systems.” In *ASMS Conference on Mass Spectrometry and Allied Topics*. Minneapolis.

Bonis, Michel, and Reg MacDonald. 2015. “$H_2S + CO_2$ Corrosion: Additional Learnings from Field Experience,” NACE International. No. 5718: 1–16.

Buttigieg, Pier L., and Alban Ramette. 2014. “A Guide to Statistical Analysis in Microbial Ecology: A Community-Focused, Living Review of Multivariate Data Analyses.” *FEMS Microbiology Ecology* 90 (3). The Oxford University Press: 543–50. doi:10.1111/1574-6941.12437.

Caporaso, J. Gregory, Justin Kuczynski, Jesse Stombaugh, Kyle Bittinger, Frederic D. Bushman, Elizabeth K Costello, Noah Fierer, et al. 2010. “QIIME Allows Analysis of High-Throughput Community Sequencing Data.” *Nature Methods* 7 (5): 335–36. doi:10.1038/nmeth.f.303.

Creek, Darren J., Andris Jankevics, Karl E. V. Burgess, Rainer Breitling, and Michael P. Barrett. 2012. “IDEOM: An Excel Interface for Analysis of LC-MS-Based Metabolomics Data.” *Bioinformatics (Oxford, England)* 28 (7): 1048–49. doi:10.1093/bioinformatics/bts069.

Crolet, J-L., and Michel R. Bonis. 2010. “Algorithm Of The Protectiveness Of Corrosion Layers 1 - Protectiveness Mechanisms And CO_2 Corrosion Prediction.” NACE International.

De Livera, Alysha M. and Jarius B. Bowne. 2015. Analysis of Metabolomics Data. R package version 0.1.4.

Duncan, Kathleen E., Lisa M. Gieg, Victoria A. Parisi, Ralph S. Tanner, Susannah G. Tringe, J. Bristow, and Joseph M. Suflita. 2009. “Biocorrosive Thermophilic Microbial Communities in Alaskan North Slope Oil Facilities.” *Environmental Science & Technology* 43 (20): 7977–84. doi:10.1021/es9013932.

Eckert, Richard, and Torben L. Skovhus. 2011. “Methods to Investigate MIC in the Oil and Gas Industry.” *Materials Performance* 50 (8): 2–6.

Edgar, Robert C. 2013. “UPARSE: Highly Accurate OTU Sequences from Microbial Amplicon Reads.” *Nature Methods* 10 (10): 996–98. doi:10.1038/nmeth.2604.

Elshahed, Mostafa S., Lisa M. Gieg, Michael J. McInerney, and Joseph M. Suflita. 2001. “Signature Metabolites Attesting to the in Situ Attenuation of Alkylbenzenes in Anaerobic Environments.” *Environmental Science & Technology* 35 (4): 682–89. doi:10.1021/Es001571u.

Gales, G., N. Chehider, C. Joulian, F. Battaglia-Brunet, J. L. Cayol, A. Postec, J. Borgomano, et al. 2011. “Characterization of *Halanaerocella petrolearia* Gen. Nov., Sp. Nov., a New Anaerobic Moderately Halophilic Fermentative Bacterium Isolated from a Deep Subsurface Hypersaline Oil Reservoir.” *Extremophiles* 15 (5): 565–71. doi:10.1007/s00792-011-0387-y.

Gevertz, D., A. J. Telang, Gerrit Voordouw, and Gary E. Jenneman. 2000. “Isolation and Characterization of Strains CVO and FWKO B, Two Novel Nitrate-Reducing, Sulfide-Oxidizing Bacteria Isolated from Oil Field Brine.” *Applied and Environmental Microbiology* 66 (6): 2491–2501.

Gieg, Lisa M., and Joseph M. Suflita. 2002. “Detection of Anaerobic Metabolites of Saturated and Aromatic Hydrocarbons in Petroleum-Contaminated Aquifers.” *Environmental Science & Technology* 36 (17): 3755–62. doi:10.1021/Es0205333.

Gieg, Lisa M., Irene A. Davidova, Kathleen E. Duncan, and Joseph M. Suflita. 2010. “Methanogenesis, Sulfate Reduction and Crude Oil Biodegradation in Hot Alaskan Oilfields.” *Environmental Microbiology* 12 (11): 3074–86. doi:10.1111/j.1462-2920.2010.02282.x.

Gieg, Lisa M., and Joseph M. Suflita. 2005. *Petroleum Microbiology*. Edited by Bernard Ollivier and Michel Magot. American Society of Microbiology. doi:10.1128/9781555817589.

Grabowski, Agnès, Olivier Nercessian, Françoise Fayolle, Denis Blanchet, and Christian Jeanthon. 2005. “Microbial Diversity in Production Waters of a Low-Temperature Biodegraded Oil Reservoir.” *FEMS Microbiology Ecology* 54 (3): 427–43. doi:10.1016/j.femsec.2005.05.007.

Gunaltun, Yves, and Kamesh Kumar. 2014. “Changes in Water Chemistry in Dead Ends Potentially Influenced by Bacterial Activity.” *NACE Corrosion*, no. 3755: 1–7.

Hubert, Casey R. J., Thomas B. P. Oldenburg, Milovan Fustic, Neil D. Gray, Stephen R. Larter, Kevin Penn, Arlene K. Rowan, et al. 2012. “Massive Dominance of *Epsilonproteobacteria* in Formation Waters from a Canadian Oil Sands Reservoir Containing Severely Biodegraded Oil.” *Environmental Microbiology* 14 (2): 387–404. doi:10.1111/j.1462-2920.2011.02521.x.

Hubert, Casey, and Gerrit Voordouw. 2007. “Oil Field Souring Control by Nitrate-Reducing *Sulfurospirillum* Spp. That Outcompete Sulfate-Reducing Bacteria for Organic Electron Donors.” *Applied and Environmental Microbiology* 73 (8): 2644–52. doi:10.1128/AEM.02332-06.

Kanehisa, Minoru, Susumu Goto, Yoko Sato, Miho Furumichi, and Mao Tanabe. 2012. “KEGG for Integration and Interpretation of Large-Scale Molecular Data Sets.” *Nucleic Acids Research* 40 (D1): D109–14. doi:10.1093/nar/gkr988.

Kodama, Yumiko. 2004. “*Sulfuricurvum kujiense* Gen. Nov., Sp. Nov., a Facultatively Anaerobic, Chemolithoautotrophic, Sulfur-Oxidizing Bacterium Isolated from an Underground Crude-Oil Storage Cavity.” *International Journal of Systematic and Evolutionary Microbiology* 54 (6): 2297–2300. doi:10.1099/ijs.0.63243-0.

Kodama, Yumiko, Le Thu Ha, and Kazuya Watanabe. 2007. “*Sulfurospirillum cavolei* Sp. Nov., a Facultatively Anaerobic Sulfur-Reducing Bacterium Isolated from an Underground Crude Oil Storage Cavity.” *International Journal of Systematic and Evolutionary Microbiology* 57 (Pt 4): 827–31. doi:10.1099/ijs.0.64823-0.

Leader, David P., Karl Burgess, Darren Creek, and Michael P. Barrett. 2011. “Pathos: A Web Facility That Uses Metabolic Maps to Display Experimental Changes in Metabolites Identified by Mass Spectrometry.” *Rapid Communications in Mass Spectrometry : RCM* 25 (22): 3422–26. doi:10.1002/rcm.5245.

Legendre, Pierre, and Louis Legendre. 2012. *Numerical Ecology*. *Developments in Environmental Modelling*. Vol. 24. Developments in Environmental Modelling. Elsevier. doi:10.1016/B978-0-444-53868-0.50009-5.

Lenhart, Tiffany R., Kathleen E. Duncan, Iwona B. Beech, Jan A. Sunner, Whitney Smith, Vincent Bonifay, Bernadette Biri, and Joseph M. Suflita. 2014. “Identification and Characterization of Microbial Biofilm Communities Associated with Corroded Oil

Pipeline Surfaces." *Biofouling* 30 (7). Taylor & Francis: 823–35. doi:10.1080/08927014.2014.931379.

Liang, Renxing, Irene A. Davidova, Christopher R. Marks, Blake W. Stamps, Brian H. Harriman, Bradley S. Stevenson, Kathleen E. Duncan, and Joseph M. Suflita. 2016. "Metabolic Capability of a Predominant *Halanaerobium* sp. in Hydraulically Fractured Gas Wells and its Implication in Pipeline Corrosion". Submitted.

Lozupone, Catherine, and Rob Knight. 2005. "UniFrac: A New Phylogenetic Method for Comparing Microbial Communities." *Applied and Environmental Microbiology* 71 (12): 8228–35. doi:10.1128/AEM.71.12.8228-8235.2005.

Ludwig, Wolfgang, Oliver Strunk, Ralf Westram, Lothar Richter, Harald Meier, Yadhukumar, Arno Buchner, et al. 2004. "ARB: A Software Environment for Sequence Data." *Nucleic Acids Research* 32 (4): 1363–71.

Maxwell, Stephen, Heike Hoffman, and Carol Divine. 2007. "Application of Molecular Microbiology Techniques as Tools for Monitoring Oilfield Bacteria." In *NACE Corrosion Conference & Expo*.

McKenzie, Denis I., Lydia Denys, and Alan Buchanan. 1987. "The Solubilization of Nickel, Cobalt and Iron from Laterites by Means of Organic Chelating Acids at Low pH." *International Journal of Mineral Processing* 21 (3-4): 275–92. doi:10.1016/0301-7516(87)90059-7.

Murali Mohan, Arvind, Angela Hartsock, Kyle J. Bibby, Richard W. Hammack, Radisav D. Vidic, and Kelvin B. Gregory. 2013. "Microbial Community Changes in Hydraulic Fracturing Fluids and Produced Water from Shale Gas Extraction." *Environmental Science & Technology* 47 (22): 13141–50. doi:10.1021/es402928b.

Murata, René, Janet Benaquisto, and Cecily Storey. 2015. "A Methodology for Identifying and Addressing Dead-Legs and Corrosion Issues in a Process Hazard Analysis (PHA)." *Journal of Loss Prevention in the Process Industries* 35 (May): 387–92. doi:10.1016/j.jlp.2015.02.004.

NACE International, TM0194-2014. 2014. "Field Monitoring of Bacterial Growth in Oil and Gas Systems." NACE International.

Neria-González, Isabel, En Tao Wang, Florina Ramírez, Juan M. Romero, and César Hernández-Rodríguez. 2006. "Characterization of Bacterial Community Associated to Biofilms of Corroded Oil Pipelines from the Southeast of Mexico." *Anaerobe* 12 (3): 122–33. doi:10.1016/j.anaerobe.2006.02.001.

Oksanen, Jari, F Guillaume Blanchet, Roeland Kindt, Pierre Legendre, Peter R Minchin, R. B. O'Hara, Gavin L. Simpson, Peter Solymos, M. Henry H. Stevens, and

Helene Wagner. 2016. “Vegan: Community Ecology Package.” http://cran.r-project.org/package=vegan.

Oldham, Athenia L., Heather S. Drilling, Blake W. Stamps, Bradley S. Stevenson, and Kathleen E. Duncan. 2012. “Automated DNA Extraction Platforms Offer Solutions to Challenges of Assessing Microbial Biofouling in Oil Production Facilities.” *AMB Express* 2 (1): 60. doi:10.1186/2191-0855-2-60.

Pruesse, Elmar, C. Quast, K. Knittel, B. Fuchs, and W. Ludwig. 2007. “SILVA: A Comprehensive Online Resource for Quality Checked and Aligned Ribosomal RNA Sequence Data” *Nucleic Acids Research* 35 (21): 7188–96.

Pruesse, Elmar, Jörg Peplies, and Frank O. Glöckner. 2012. “SINA: Accurate High-Throughput Multiple Sequence Alignment of Ribosomal RNA Genes.” *Bioinformatics* 28 (14): 1823–29. doi:10.1093/bioinformatics/bts252.

Ravot, G., Michel Magot, B. Ollivier, B. K C Patel, E. Ageron, P. A. D. Grimont, P. Thomas, and J. L. Garcia. 1997. “*Haloanaerobium congolense* sp. nov., an Anaerobic, Moderately Halophilic, Thiosulfate- and Sulfur-Reducing Bacterium from an African Oil Field.” *FEMS Microbiology Letters* 147 (1): 81–88. doi:10.1016/S0378-1097(96)00508-3.

Reysenbach, Anna-Louise, and Norman R. Pace. 1995. *Reliable Amplification of Hyperthermophilic Archaeal 16S rRNA Genes by the Polymerase Chain Reaction. In Archaea: A Laboratory Manual – Thermophiles.* Edited by F. T. Robb and A. R. Place. Cold Spring Harbor.

Scheltema, Richard A., Andris Jankevics, Ritsert C. Jansen, Morris A. Swertz, and Rainer Breitling. 2011. “PeakML/mzMatch: A File Format, Java Library, R Library, and Tool-Chain for Mass Spectrometry Data Analysis.” *Analytical Chemistry* 83 (7). American Chemical Society: 2786–93. doi:10.1021/ac2000994.

Schloss, Patrick D., Sarah L. Westcott, Thomas Ryabin, Justine R. Hall, Martin Hartmann, Emily B. Hollister, Ryan A. Lesniewski, et al. 2009. “Introducing Mothur: Open-Source, Platform-Independent, Community-Supported Software for Describing and Comparing Microbial Communities.” *Applied and Environmental Microbiology* 75 (23): 7537–41. doi:10.1128/aem.01541-09.

Sette, Lara D., Karen C. M. Simioni, Suzan P. Vasconcellos, Lucia J. Dussan, Eugênio V. S. Neto, and Valéria M. Oliveira. 2007. “Analysis of the Composition of Bacterial Communities in Oil Reservoirs from a Southern Offshore Brazilian Basin.” *Antonie van Leeuwenhoek, International Journal of General and Molecular Microbiology* 91: 253–66. doi:10.1007/s10482-006-9115-5.

Sievert, Stefan M., Kathleen M. Scott, Martin G. Klotz, Patrick S. G. Chain, Loren J. Hauser, James Hemp, Michael Hügler, et al. 2008. “Genome of the

Epsilonproteobacterial Chemolithoautotroph *Sulfurimonas denitrificans*." *Applied and Environmental Microbiology* 74 (4): 1145–56. doi:10.1128/AEM.01844-07.

Sloley, Andrew. 2011. "Watch Out for Dead-Legs." *Chemical Processing*, September. http://www.chemicalprocessing.com/articles/2011/watch-out-for-dead-legs/?show=all.

Stevenson, Bradley S., Heather S. Drilling, Paul A. Lawson, Kathleen E. Duncan, Victoria A. Parisi, and Joseph M. Suflita. 2011. "Microbial Communities in Bulk Fluids and Biofilms of an Oil Facility Have Similar Composition but Different Structure." *Environmental Microbiology* 13 (4): 1078–90. doi:10.1111/j.1462-2920.2010.02413.x.

Takai, K. 2006. "*Sulfurimonas paralvinellae* sp. nov., a Novel Mesophilic, Hydrogen- and Sulfur-Oxidizing Chemolithoautotroph within the Epsilonproteobacteria Isolated from a Deep-Sea Hydrothermal Vent Polychaete Nest, Reclassification of *Thiomicrospira denitrificans* as *Sulfurimonas denitrificans* comb. nov. and emended description of the genus *Sulfurimonas*." *International Journal of Systematic and Evolutionary Microbiology* 56 (8): 1725–33. doi:10.1099/ijs.0.64255-0.

Tamura, Koichiro, Glen Stecher, Daniel Peterson, Alan Filipski, and Sudhir Kumar. 2013. "MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0." *Molecular Biology and Evolution* 30 (12): 2725–29. doi:10.1093/molbev/mst197.

Tautenhahn, Ralf, Christoph Böttcher, and Steffen Neumann. 2008. "Highly Sensitive Feature Detection for High Resolution LC/MS." *BMC Bioinformatics* 9 (1): 504. doi:10.1186/1471-2105-9-504.

Telang, Anita J., Gary E. Jenneman, and Gerrit Voordouw. 2011. "Sulfur Cycling in Mixed Cultures of Sulfide-Oxidizing and Sulfate- or Sulfur-Reducing Oil Field Bacteria." *Canadian Journal of Microbiology*, February. NRC Research Press Ottawa, Canada.

Voordouw, Gerrit, S. M. Armstrong, M. F. Reimer, B. Fouts, Anita 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." *Applied and Environmental Microbiology* 62 (5): 1623–29.

Voordouw, Gerrit, Brenton Buziak, Shiping Lin, Alexander Grigoriyan, Krista M. Kaster, Gary E. Jenneman, and Joseph J. Arensdorf. 2013. "Use of Nitrate or Nitrite for the Management of the Sulfur Cycle in Oil and Gas Fields." In *International Symposium on Oilfield Chemistry*. Society of Petroleum Engineers. doi:10.2118/106288-MS.

Whitman, William B., Fred Rainey, Peter Kämpfer, Martha Trujillo, Jonsik Chun, Paul DeVos, Brian Hedlund, and Svetlana Dedysh, eds. 2015. *Bergey's Manual of Systematics of Archaea and Bacteria*. Chichester, UK: John Wiley & Sons, Ltd.

Yi, Yanmei, Weiyi Huang, and Ying Ge. 2007. “Exopolysaccharide: A Novel Important Factor in the Microbial Dissolution of Tricalcium Phosphate.” *World Journal of Microbiology and Biotechnology* 24 (7): 1059–65. doi:10.1007/s11274-007-9575-4.

Zhang, Jiajie, Kassian Kobert, Tomáš Flouri, and Alexandros Stamatakis. 2014. “PEAR: A Fast and Accurate Illumina Paired-End reAd mergeR.” *Bioinformatics* 30 (5): 614–20. doi:10.1093/bioinformatics/btt593.

Table 1. Characterization of the physicochemical properties of bulk and dead-leg production fluid samples. [a]Values not determined (ND) for Water Separator Dead-leg #3 due to high oil fraction.

	pH	Salinity [Cl^-] (M)	[SO_4^{2-}] (mM)	[Acetate] (mM)
Desalter				
Bulk	6.2	1.5	1.03	1.01
Dead-leg #1	5.2	4.0	3.17	3.29
Dead-leg #2	6.0	3.4	2.10	6.63
Dead-leg #3	6.0	1.7	1.18	6.87
Low Pressure Separator				
Bulk	5.8	4.1	4.81	1.02
Dead-leg #1	4.5	3.1	0.20	7.58
Dead-leg #2	4.5	3.1	0.11	9.27
Dead-leg #3	4.5	2.9	0.16	9.03
Water Separator				
Bulk	6.2	4.5	4.50	1.08
Dead-leg #1	6.1	4.0	7.98	0.59
Dead-leg #2	6.1	4.0	4.15	1.23
Dead-leg #3[a]	ND	ND	ND	ND

Table 2. Summary of community diversity metrics and population sizes for bulk fluid and dead-leg #2 brines from each module.

	Desalter		Low Pressure Separator		Water Separator	
	Bulk	**Dead-leg #2**	**Bulk**	**Dead-leg #2**	**Bulk**	**Dead-leg #2**
Diversity Metrics						
$OTU_{0.03}$[a]	73	91	83	134	87	98
Shannon	0.34	0.54	0.52	0.80	0.52	0.68
Evenness						
Shannon Diversity	2.11	3.53	3.33	5.62	3.33	4.51

qPCR Analyses[b]	**Desalter Bulk**	**Low Pressure Sep. Bulk**	**Water Sep. Bulk**
Bacterial 16S	$3.70x10^8$ ($2.31x10^8$)	$5.98x10^5$ ($5.22x10^4$)	$1.17x10^5$ ($7.78x10^4$)
Archaeal 16S	ND	ND	ND
aprA	$1.27x10^4$ ($1.15x10^4$)	$1.79x10^3$ ($3.26x10^2$)	$8.86x10^1$ ($4.07x10^1$)
dsrA	$1.20x10^6$ ($1.41x10^5$)	$4.43x10^4$ ($5.61x10^3$)	$1.65x10^4$ ($7.14x10^3$)
mcrA	ND	ND	ND

[a]Number of OTUs observed within each sample. Libraries were subsampled without replacement to 1400 sequences. [b]Quantitative PCR values reported as copies of target gene mL^{-1} brine. Target genes are: Bacterial 16S rRNA gene (Bacterial 16S); Archaeal 16S rRNA gene (Archaeal 16S); Adenosine-5'-phosphosulfate reductase subunit A (*aprA*); Dissimilatory sulfite reductase subunit A (*dsrA*); Methylcoenzyme-M reductase (*mcrA*). Values reported are averages of 3 technical replicate measures with standard deviation in parentheses. ND, not detected.

Figure 1. Schematic diagram depicting simplified connectivity of selected processing modules in an oilfield production facility. Sampling points are designated by black triangles.

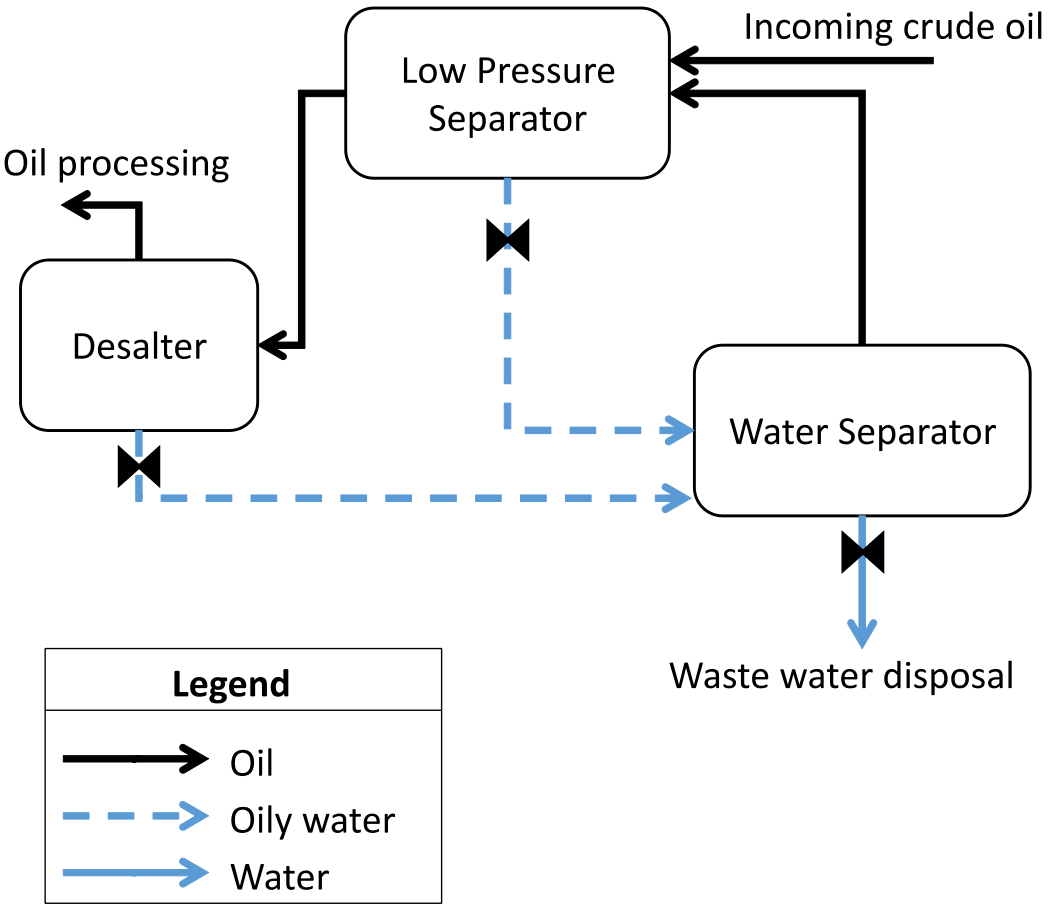


Figure 2. Brine chemical characteristic correlation analysis showing Spearman ranked correlation coefficients between measured fluid properties. Increasingly positive correlations are shown in green, negative shown in red and no observed correlation is indicated in yellow.

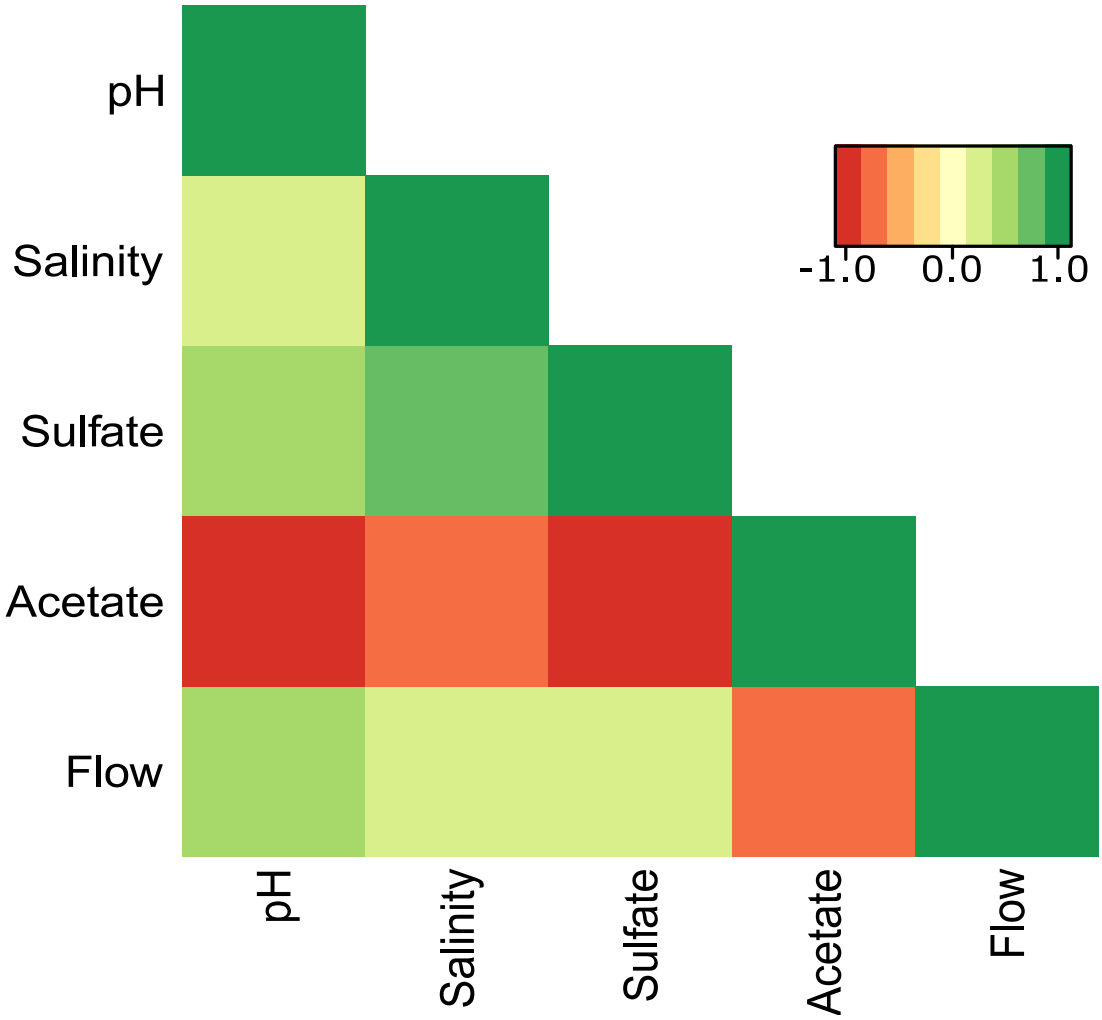


Figure 3. Pairwise dissimilarity matrix heatmap depicting weighted-Unifrac pairwise distances between microbial communities in each sample. Greater pair-wise dissimilarity is shown by increasing red shading.

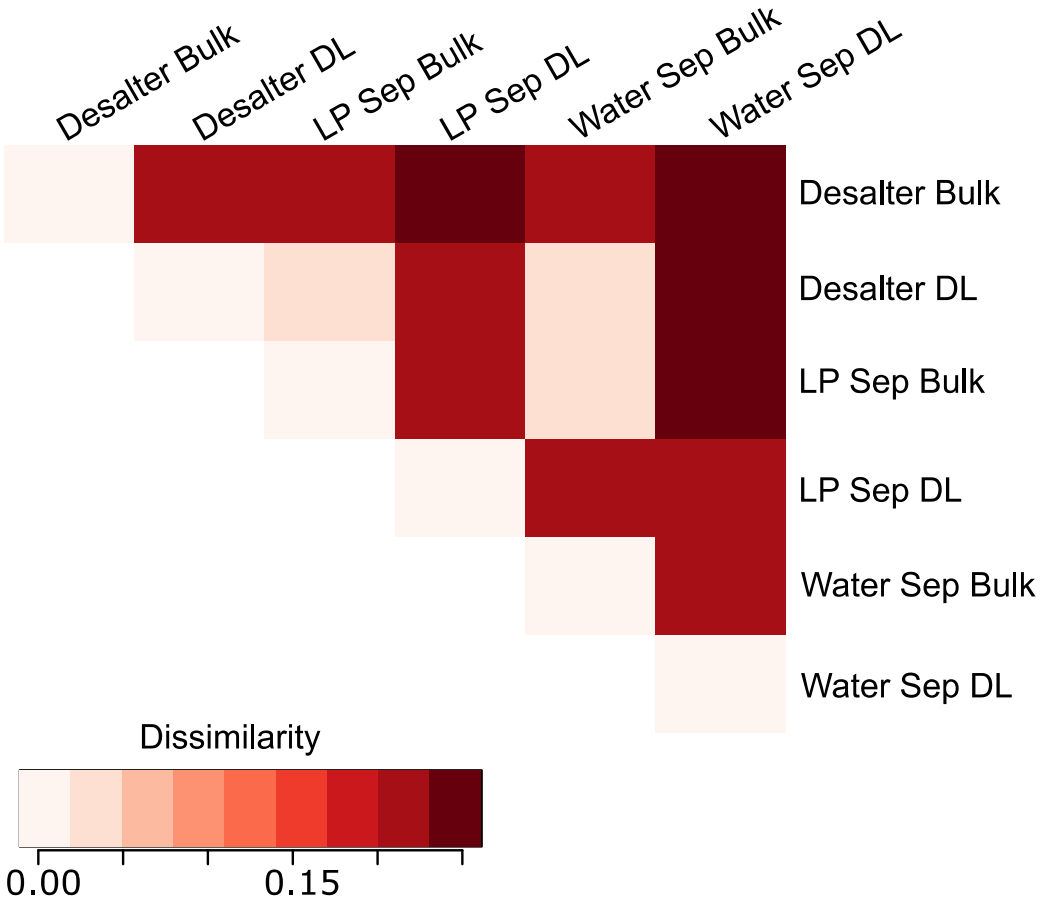


Figure 4. Relative abundance of order-level phylogenetic composition of brine communities.

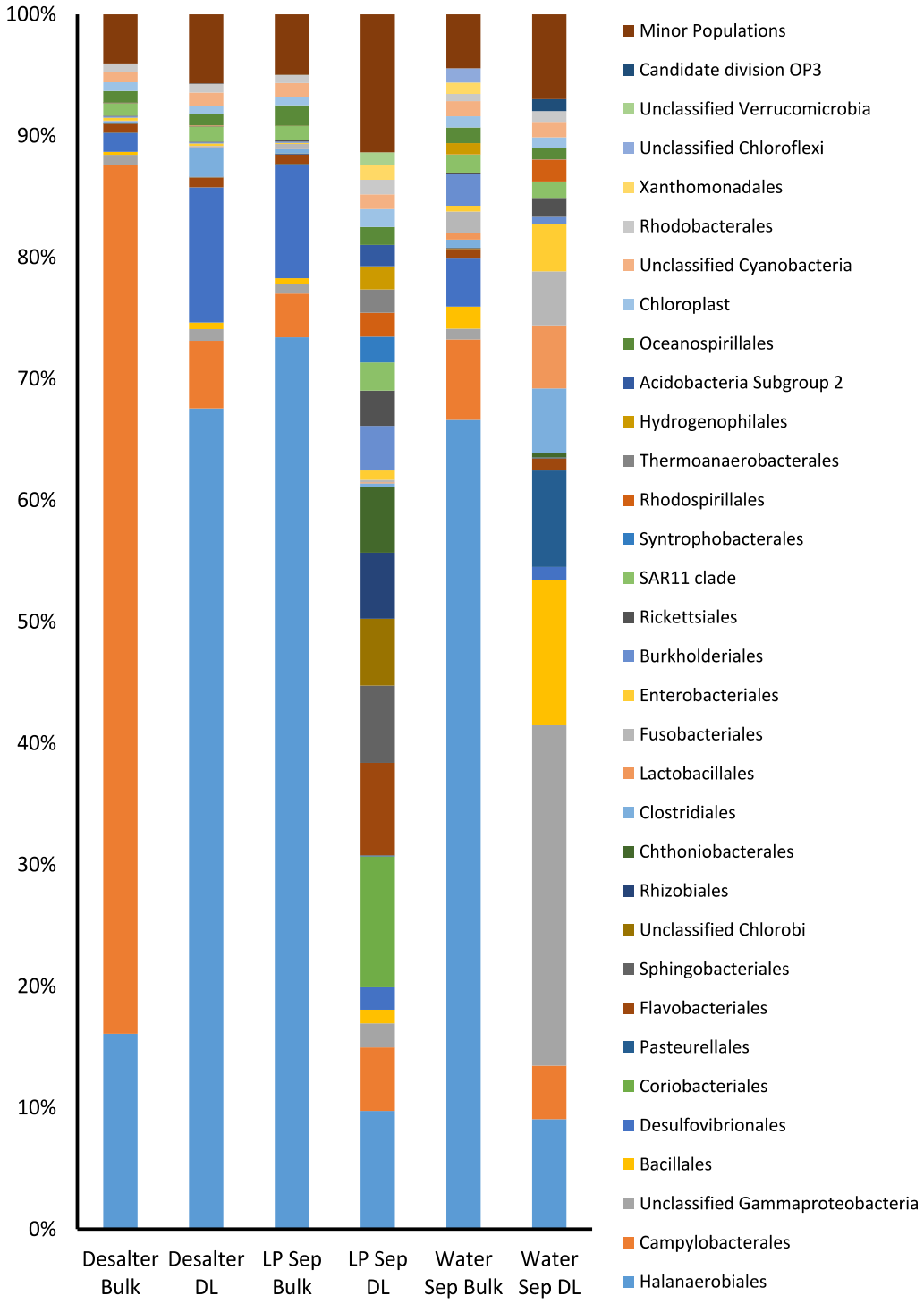


Figure 5. Phylogenetic analyses of selected abundant OTUs from the (A) *Halanaerobium* spp., (B) *Sulfurospirillum* spp., and (C) *Desulfohalobiaceae* spp. Neighbor-joining dendrograms were constructed from alignments of the OTU representative sequences with full length type strains using the Tajima-Nei model with pairwise deletion in MEGA6 (Tamura et al. 2013). Genbank accession numbers are designated in parentheses and sequences obtained in this study are designated in bold. Bootstrap values (n = 500 replicates) greater than 65 are shown.

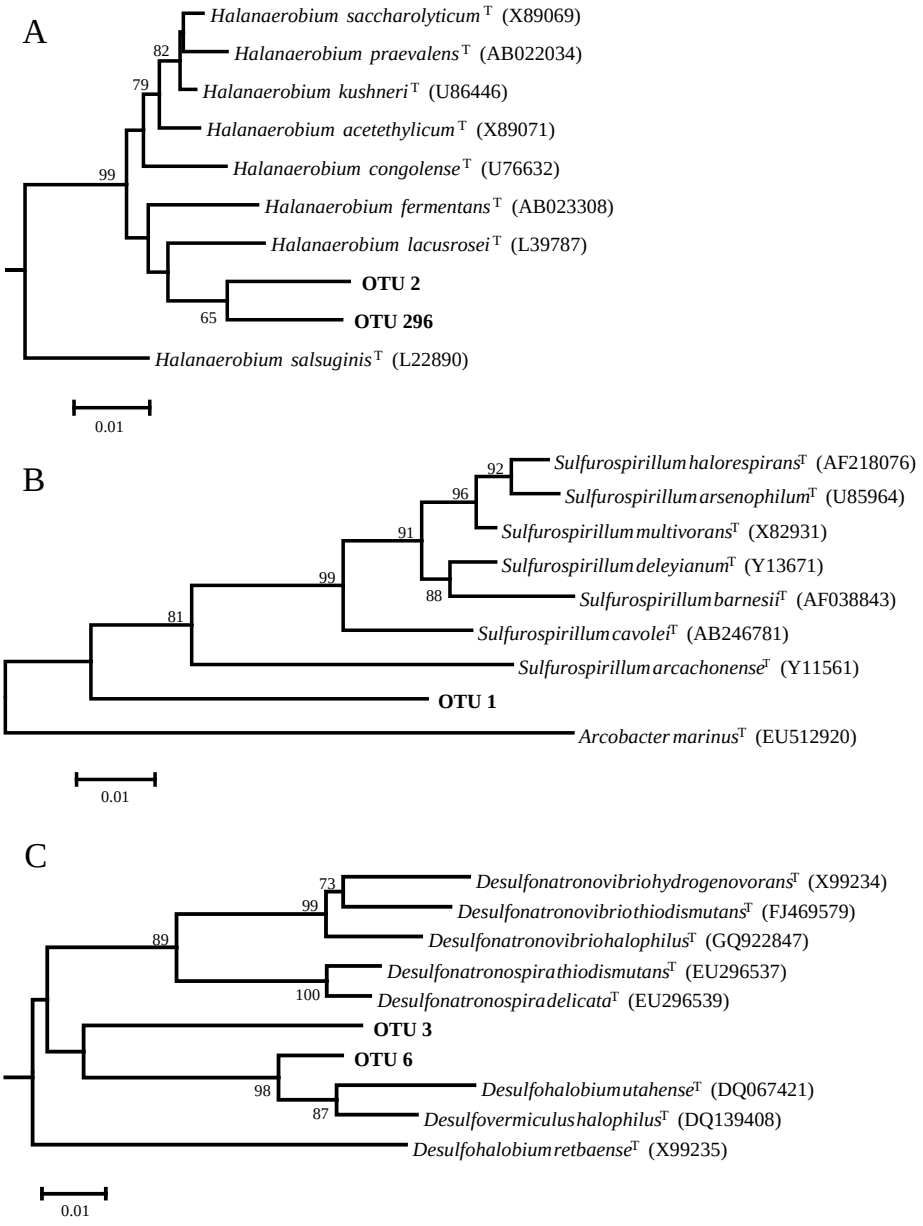
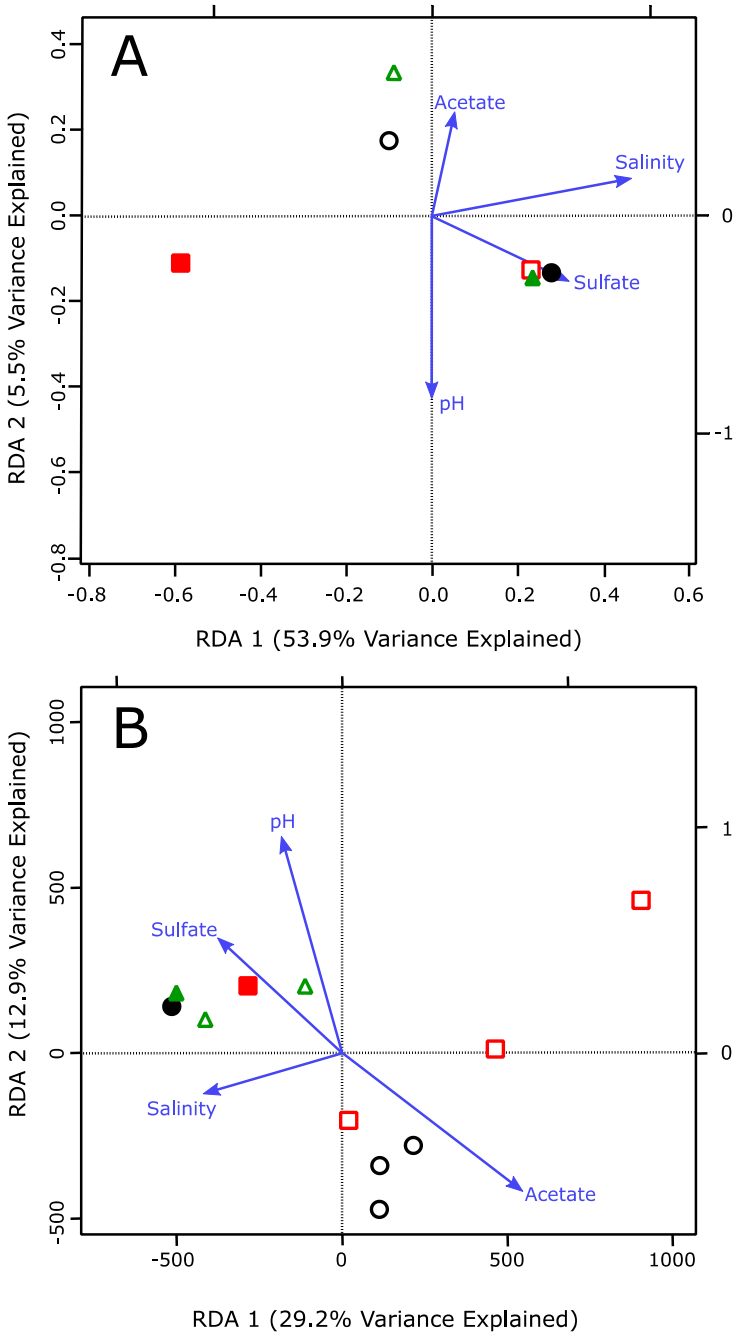


Figure 6. Redundancy Analyses (RDA) of (A) Community composition and (B) Untargeted metabolite profiles. Ordinations are constrained by environmental parameters (Table 1) for each sample, shown in blue vectors. Bulk fluid and dead-leg samples are indicated by closed and open markers, respectively. Desalter, low pressure separator, and water separator samples are shown in red squares, black circles and green triangles, respectively.



Chapter 2: Successional Ecology of Hydrocarbonoclastic Microbial Communities within Naval Fuel-Compensated Ballast Tanks

Abstract

The microbial metabolism of hydrocarbons can be associated with the biocorrosion of the mostly carbon steel energy infrastructure and result in major economic losses and potentially devastating environmental impacts. These deleterious activities are frequently observed when steel alloys interact with mixtures of hydrocarbons and seawater. We sought to determine if marine microorganisms in ballast tank compensation water were responsible for metal and fuel deterioration. Water samples were obtained from the ballast expansion tanks aboard 3 vessels with residence times of 1 week, ~20 weeks and 32 weeks, respectively. These samples were subjected to metagenomic, metabolomic, and chemical analyses and compared to coastal seawater originally used in the tanks. Universal 16S rRNA gene amplicon library analyses revealed an enrichment in *Gammaproteobacteria* from 18% (seawater), 64% (1 week), 32% (~20 weeks) and 22% (32 weeks), respectively. The *Deltaproteobacteria* portion of the community increased with residence time from 3-4% (seawater and 1 week), 19% (~20 weeks) and 42% (32 weeks) to become the most abundant Class observed in the eight-month sample. Shotgun metagenome analyses revealed little change in the abundance of oxidative phosphorylation genes (*ccoN/coxA*) in the 1 week and 20 week samples, but a 2-fold decrease in the 32-week sample relative to the harbor water. Conversely, sulfate and thiosulfate reduction genes remained relatively unchanged in the 1 week and 20 week samples but exhibited a strong enrichment after 32-weeks, with a 2.5-fold and 29-fold increase in abundance for

aprA and *phsA*, respectively. Aerobic BTEX activation genes were strongly enriched after 1 week (*etbAa*, 365-fold; *xylM*, 93-fold), but declined in the other samples. In contrast, anaerobic community was highly enriched in fumarate-addition genes at 32-weeks (*assA*, 5.7%; *bssA/hbsA/ibsA*, 4.7%; *nmsA*, 20.5%), specifically related to aromatic substrates, despite being undetected in the harbor water sample. Environmental metabolomic investigations revealed the presence of substituted catechols in the 1-week old ballast fluids while phenylethanol and benzylsuccinates were detected in the 20 and 32 week samples. All fluid samples contained elevated concentrations of Fe, Mn, Cu, and Ni, with the 8 month-old fluid being over-saturated with CuS. The high metal concentrations relative to the harbor water are indicative of steel alloy corrosion, specifically of the Cu-Ni sluice pipes connecting the ballast tanks. These findings define the ecological succession patterns within fuel ballast tanks, wherein aerobic hydrocarbonoclastic microorganisms initiate fuel biodegradation, followed by the proliferation of an anaerobic microbial community that can use fuel or aerobic fuel metabolites as electron donors coupled to sulfate reduction. The deterioration of hydrocarbon quality and production of corrosive sulfides are major consequences associated with the interaction between petroleum products and water within steel fuel infrastructure.

Introduction

The corrosion of metal ballast tanks containing seawater has been a commonly observed phenomenon since the late 19th century with the widespread growth of metallic vessels. Beginning in the 1960s, the United States Navy became aware of the role of microorganisms in the biocorrosion of shipboard seawater-displaced fuel ballast

tanks (Hazlett, 1966, 1967). Aircraft carriers storing aviation fuel (JP-5), diesel, and gasolines experienced aggressive fuel souring and corrosion of their fuel ballast tanks linked to sulfate-reducing marine microorganisms (Klemme and Neihof, 1969). Research led to the use of sodium chromate or thiopyridine derivatives as biological control treatments (Klemme and Neihof, 1969; Klemme and Leonard, 1971). Gasoline was found to be particularly susceptible to biodeterioration, and reports of ballast tank corrosion decreased substantially after the Navy moved away from its use in fleet operations (Neihof, 1988).

Since then, similar corrosion-based ship wasting in maritime shipping vessels has prompted renewed research in the extent and mechanisms of this activity. A previous study examined winged ballast tanks within a cargo vessel and found the system experienced two distinct modes of corrosion resulting in dramatically different corrosion rates (Cleland, 1995). Vertical metal surfaces near the air-water interface lost 0.15 mmpy, while a much more severe corrosion rate of 6 mmpy was observed on submerged horizontal surfaces and tank bottoms. Further examination of corroded areas in the submerged zone revealed the corrosion layer to contain sulfides and cultures of the surrounding biofilm yielded a mixed community of aerobic and sulfate-reducing bacteria (SRB) (Cleland, 1995). Huang and colleagues, also reported aggressive corrosion (2 mmpy) of horizontal surfaces and uncoated tank bottoms in oil cargo holds of oil tanker ships, and enumerated populations of SRBs at $10^5 – 10^7$ cells mL^{-1} through cultivation methods (Huang et al., 1997).

More recently, studies have been undertaken to further examine the link between microbial fuel biodeterioration and sulfidogenic biocorrosion of mild steels commonly

employed in compensated fuel ballast tanks. Alternative fuel blends containing fatty-acid methylesters (FAMEs) from biologically derived feedstocks supported high rates of sulfate reduction and severe pitting corrosion when mixed with seawater inoculum from Key West, FL and the ballast tank of the USS Gettysburg (Aktas et al., 2010; Jason S Lee et al., 2012). Petroleum-derived F-76 diesel used by the USS Gettysburg was also susceptible to microbial degradation by the ballast tank community and stimulated sulfidogenesis in batch incubations (Lyles et al., 2013). This investigation also revealed that refined petroleum-derived diesel fuels supported heterotrophic sulfate reduction rates independent of fuel sulfur content, but less than that of the F-76 formulation utilized by the vessel (Lyles et al., 2013). These studies highlight the general susceptibility of hydrocarbons to microbial attack and degradation under anoxic conditions and demonstrate the direct link between fuel oxidation and biological production of corrosive sulfides.

In spite of these lessons, some warship classes currently operated by the United States Navy were designed and produced with seawater-compensated fuel ballast tanks for the storage of F-76 diesel fuel. During operations, these ships use seawater to displace diesel from these storage tanks into the service tanks immediately upstream of the engines, as well as to maintain ballast as fuel is consumed. Ships of these classes have reported a high rate of fuel filter fouling and degradation of ballast tank components at a rate exceeding those expected for abiotic marine corrosion (Personal communication). In response to these observations, this study was undertaken to examine the ecology of marine microflora within ballast tanks and identify mechanisms of diesel fuel biodeterioration of and the associated biocorrosion of shipboard

infrastructure. As the United States Navy continues to implement the use of green fuels, the findings of this research on petroleum-derived diesel may support *apriori* estimates on fouling and corrosion risks.

Materials & Methods

Sample Collection

Ballast tank water samples were obtained by collecting fluid from ballast bank expansion tank sampling ports (Fig. 1) into sterile polypropylene bottles. Three separate samples were collected from each sampling point and immediately treated as described below for downstream analyses. Ballast tank water (250 mL) was preserved for DNA extraction by addition of 5 mL of DNAzol® (Molecular Research Center Inc.; Cincinnati, OH) immediately upon collection. A 1 L sample was acidified (pH<2) by the addition of 6 N HCl for metabolite analyses. Finally, a 250 mL polypropylene bottle was top-filled without preservative for chemical and metals analyses. A fuel sample (25 mL) was also obtained from each vessel directly from the service tank, immediately upstream of the engine, in a glass vial with PTFE lined cap for mass spectral characterization of fuel components. In addition to ballast tank samples, several liters of surface water were collected from San Diego Harbor where each shipped was docked to serve as a control. All samples were shipped overnight on ice to the lab and stored at 4°C until further use.

Chemical Analyses

Nitrate and sulfate were quantified by suppressed ion-chromatography on a Dionex ICS-3000 system (Dionex; Sunnyvale, CA) equipped with Dionex AS4A-SC

guard (4 mm x 50 mm) and analytical (4 mm x 250 mm) columns. Isocratic separations were performed with a mobile phase of carbonate buffer (1.7 mM HCO_3^-; 1.8 mM CO_3^{2-}; 2.0 ml min^{-1}) and a suppression current of 27 mA. Dissolved oxygen was measured in the field using a FirestingO_2 optical oxygen meter (Pyro Science GmbH; Aachen, Germany) as described by the manufacturer.

Upon receipt, 25 mL of sample was filtered with 0.45 μm polyether sulfone filter (Whatman Puradisc), and diluted to a metal concentration range between 0 to 40 ppm using 0.1% (v/v) nitric acid (Fluka, Ultrapure). Dissolved metal concentrations were measured using a Varian AA 240FS Atomic Absorption Spectrometer with a graphite tube furnace (Varian GTA120). Sample volume injected into the graphite furnace was 10 μL with argon as purging gas. Temperature gradient started at 85 ºC and was held for 5 s, then ramped to 95 ºC and held for 40 s, then ramped to 700 ºC for 7 s before going to 2400 ºC for 4.8 s. At this final temperature, absorption was measured using a background correction and a spectral bandwidth/slit width of 0.2. A six point calibration curve ranging over 0 to 40 ppm was used for each metal examined. All samples were measured in triplicate, and each time after measuring five samples, two known calibration standards and the blank were rerun with recalibration of the calibration slope if necessary.

Cell Counts

Cell counts were determined using unpreserved sample fluids by epifluorescence microscopy under blue excitation using a micrometer grid on an Olympus BX-61 microscope. Cells were fixed using a 1/10 volume of 37%

formaldehyde and a dilution of the fixed samples was stained with 5 mL of 5 ng mL^{-1} 4,6-diamidino-2-phenylindole (DAPI) in the dark for 45 min at room temperature. Fixed cells were then filtered onto 0.22 mm, 25mm diameter black stained polycarbonate membrane filters (Sterlitech, Kent WA). At least 30 fields were counted and averaged for each slide.

DNA Extraction and Amplicon Analyses

Water samples that were preserved in the field for DNA analysis were filtered (250 mL; 0.45 µm pore size PES filter) upon sample receipt in the lab and stored at -80°C until use. Bulk genomic DNA was extracted from each filter as previously described (Oldham et al., 2012). Quantitative PCR was employed to enumerate the population densities of bacteria, archaea and picoeukaryotes within each sample. Bacterial and archaeal 16S rRNA genes were targeted and amplified with Eubac8F/338R and A8F/344R primer pairs (Stevenson et al., 2011). Marine picoeucaryote 18S rRNA genes were amplified using the 345F/499R primer pair and methods previously described (Zhu et al., 2005). Dissimilatory sulfate reduction pathway genes, adenosine-5'-phosphosulfate reductase A gene (*aprA*) and dissimilatory sulfite reductase A (*dsrA*), were assessed with the RH1apsF/RH2apsR and RH1-dsr-F/RH3-dsr-R primer pairs, respectively (Ben-Dov et al., 2007).

Metagenome Analysis

Shotgun metagenome libraries were sequenced on an Illumina MiSeq using PE 300 chemistry by Oklahoma Medical Research Foundation (Oklahoma City, OK). Raw sequences were processed for quality by removing sequencing adapters using Cutadapt

(Martin, 2011), trimming each read for a minimum average PHRED quality score of 30 and mate pairing with Trimmomatic (Bolger et al., 2014). Unpaired reads were not retained for downstream analyses.

Community composition was assessed from unassembled mate-paired reads with significant homology to known small subunit (SSU) rRNA genes. Reads were searched against the Silva SSU Reference database release 108 (Pruesse et al., 2007) pre-clustered at 97% sequence similarity using USEARCH (Edgar, 2010). Hits with a minimum of 70% homology to reference SSU gene sequences were extracted for further analysis. Putative SSU reads were analyzed using QIIME 1.9.1 (Caporaso et al., 2010) through closed-reference OTU clustering (97% similarity) and taxonomic assignment against the Silva 111 SSU reference sequence and taxonomy databases. Alpha diversity metrics were computed for each sample based upon metagenome reads containing SSU fragments rarefied to 1800 sequences for each library.

Functional gene annotation was performed by translated homology searching of unassembled mate-paired reads against the KEGG database (Kanehisa et al., 2012) using DIAMOND (Buchfink et al., 2015) in the sensitive mode retaining only the top hit ($E < 10^{-5}$) for each query. Putative anaerobic hydrocarbon activation genes were identified by a translated homology search in USEARCH against the AnHyDeg database (Callaghan and Wawrik, 2016). Acceptance criteria were 60% identity over alignments with at least 35 amino acids. Functional gene abundances were calculated by dividing the number of hits to specific reference gene by the number of hits to the single copy marker gene *rpoB* to normalize for differential library population size (Howard et al., 2008). A modified approach was used to estimate the total

hydrocarbon-activating proportion of each community from that previously used (Biers et al., 2009) by a summation for selected aerobic and anaerobic specific activation genes of the following: (hits to functional $gene_i$ /length of functional $gene_i$)/hits to *rpoB*. Aerobic activation genes included: *alkB, EtbAa, tmoA, todC1, dmpB, nahAc*, and 1-methylnaphthalene hydroxylase-like genes. Anaerobic activation genes were *ebdA, assA, bssA, nmsA*.

Metabolite Analyses

Mass spectral analysis of signature metabolites associated with oxygen-dependent and independent activation and catabolism of diesel fuel components was performed on ballast tank samples acidified (pH <2) immediately upon sampling. Methods used for the targeted (GC-MS) and global survey of organic intermediates were as reported in Chapter 1.

Results

Water Sample Properties

Harbor and ballast tank water samples were analyzed for chemical composition and dissolved metal concentrations. Table 1 summarizes the residence times of ballast tank waters aboard each ship and concentrations of selected chemical constituents. Ship 1 ballast water contained ~50% less dissolved oxygen than that estimated for the harbor water where the ship was docked. Sulfate concentrations were unchanged between the harbor water and Ships 1 & 2, but fluids from Ship 3 exhibited a 5 mM (17%) loss after 8 months within the ballast tank. Dissolved metal concentrations for manganese, copper and nickel were quantified for the ballast waters of each ship and compared to

the harbor water as a baseline of metal intake into each system. All of the ballast tanks had significantly higher concentrations of each metal as compared to the surrounding harbor water. After just one-week residence within the ballast tank of Ship #1 there was a minimum increase of 65-fold or more for each element quantified. Copper and nickel levels were considerably higher than manganese in both Ships #1 and #2. This trend was not observed in Ship #3 with manganese being 10-fold higher than either copper or nickel. Overall there is a positive correlation between ballast tank water residence time and total dissolved metal concentrations, as well as the sulfide-containing particulates for each species. Comparison of the dissolved metals present within each ballast tank relative to the harbor water indicates that the intake fluids accounted for a negligible proportion of the observed metals and implicates infrastructure dissolution as a potential cause.

Marine Populations & Community Composition

The magnitude of microbial populations within each water sample was quantified by several means including: direct cell counting via epifluorescence microscopy, domain-specific qPCR, and single copy marker gene quantification from shotgun metagenomics sequencing (Table 2). Direct cell counting showed the microbial population sizes to be very similar between all ships sampled at approximately 10^5 cells mL^{-1}, regardless of residence time, and all were between 2 and10-fold less than the harbor surface water community. Quantitative PCR enumeration of bacterial and archaeal 16S rRNA genes ranged from 10^5 to 10^7 and 10^2 to 10^3 gene copies mL^{-1}, respectively. The Bacteria bloomed after one week aboard Ship #1 relative to the harbor community, and then declined with increased residence

time as seen for Ships #2 & #3. The Archaea were several orders of magnitude less abundant than Bacteria in all samples. The largest population of picoeukaryotic organisms was found in the harbor surface water, with sharply decreased in numbers in all ballast tank samples.

Microbial community diversity was assessed by the Shannon diversity index, observed species, and Simpson's evenness metric and are summarized in Table 3. All of the communities showed a bias towards some taxa (e.g. more uneven distribution of species) and show a trend of increasing dominance with increasing residence time. There is also a limited change in the species richness (e.g. observed species) of only 8% between the richest community in Ship #1 and the least rich in Ship #2.

The composition of microbial populations within each sample were determined by comparison of each 16S rRNA gene amplicon library against the Silva SSU reference database (Fig. 2). Taxa belonging to the *Gammaproteobacteria* (19.7%, 64.6%, 32.3%, 23.1%), *Alphaproteobacteria* (24.2%, 10.8%, 11.4%, 7.6%), *Flavobacteria* (20.5%, 7.1%, 12.1%, 1.8%), and the *Deltaproteobacteria* (3.1%, 4.5%, 21.1%, 43.3%) accounted for substantive proportions of the communities within the Harbor water and Ships #1, #2 and #3, respectively. The *Rhodobacterales* (8.9%) and members of the SAR 11 clade (7.7%) were the *Alphaproteobacteria* accounting for greatest proportion of the total Harbor community. Within the Harbor water gammaproteobacterial population, sequences were predominantly affiliated with the *Oceanospirillales* and *Alteromonadales* at 8.9% and 4.6% total relative abundance, respectively. The San Diego Harbor water sample contained several groups of marine *Eukaryota*, and were dominated by the *Alveolata* (11.0%) and *Stramenopiles* (4.0%).

Despite accounting for 18.8% of the total marine community in the Harbor water sample, members of the *Eukaryota* did not comprise more than 0.5% of the ballast tank assemblages. The sharp decline from 1 week aboard ship (Harbor vs. Ship #1) continued and no eukaryotic rRNA gene sequences were detected in Ship #3. Sequences classified as *Archaea* never exceeded 0.3% of any community sampled.

The assemblage within the ballast tank of Ship #1 showed a stark shift with the *Gammaproteobacteria* enriched to become the dominant taxonomic Class, with the concomitant decline of the *Alphaproteobacteria* and the *Flavobacteria*. The *Alteromonadales, Oceanospirillales,* and *Thiotrichales* increased to account for 42%, 35%, and 19% of all observed Gammaproteobacteria, respectively. The *Epsilonproteobacteria* were also enriched by 10-fold (2.8%) after 1 week within the ballast tank of Ship #1 represented almost exclusively by members of the *Campylobacterales* (97% of *Epsilonproteobacteria*).

The *Deltaproteobacteria* were the dominant Class observed in the Ship #2 community at 21.0%, a ~3.5-fold increase relative to the Ship #1 sample. Members of the *Desulfuromonadales* and *Desulfobacterales* accounted for the majority of the observed *Deltaproteobacteria* at 80% and 10%, respectively. Though the total proportion of *Gammaproteobacteria* was smaller in the Ship #2 community as compared to Ship #1, the *Alteromonodales* were enriched to represent 71% of these taxa. Taxa affiliated with the *Clostridia* (7.2%) and *Mollicutes* (4.6%) accounted for substantive portions of the total ballast tank community in Ship #2, with increases of ~3-fold and ~5-fold relative to Ship #1, and their highest abundances observed in this

study. The *Alphaproteobacteria*, *Flavobacteria*, and *Epsilonproteobacteria* were all slightly higher in relative abundance in the Ship #2 community as compared to Ship #1.

Finally, the ballast tank community of Ship #3, with a residence time of 32-weeks, was the most different from the Harbor water assemblage (Fig. 2). The *Deltaproteobacteria* (43.3%) were the most abundant taxa, with members belonging to the *Desulfobacterales*, *Desulfarculales*, and *Desulfovibrionales* at 66%, 27%, and 6% of this group, respectively. This community also contained the largest population of *Bacteroidetes* (8.2%), represented predominantly by unclassified members of the *Bacteroidales* (57% of the *Bacteroidetes*), and the smallest segment of *Flavobacteriia* observed in any sample. *Gammaproteobacteria*, *Alphaproteobacteria*, and *Epsilonproteobacteria* accounted for smaller community abudance in Ship #3 relative to Ship #2. Ship #3 had the smallest proportion of *Alphaproteobacteria* found in this study.

In addition, the functional genes within the metagenomes associated with the utilization of common inorganic electron donors and acceptors were characterized by translated Blast analysis. Figure 3 shows the normalized gene abundances within each sample. Nitrification genes ammonia monooxygenase (*amoA*) and hydroxylamine dehydrogenase (*hao*) did not exceed 0.3% normalized abundance within any sample library. Genes associated with the oxidation of reduced inorganic sulfur species were present within all of the samples analyzed. Sulfur-oxidizing system (SOX) protein B (*soxB*) declined from 5.9% in the Harbor water metagenome to ~4.2% of the functional capacity of Ships #1 and #2. This gene further declined to barely detectable levels of 0.5% in the Ship #3 ballast water sample. Nitrogen fixation, represented by nitrogenase

gene *nifH*, increased with residence time from a low level in the Harbor water (0.9%) to 5.5%, 7.1%, and 8.8% in the libraries of Ships #1, #2, and #3, respectively.

Genes affiliated with the reduction of oxygen and nitrate exhibited a marked increase between the Harbor water and Ships #1 and #2. Cytochrome c oxidase subunit I genes (*ccoN* and *coxA*) represented aerobic respiration and had the highest levels of any inorganic redox genes assayed in the Harbor and ballast tank waters of Ships #1 and #2. There was a 50% decrease in cytochrome c oxidase subunit I gene abundance after 8 months aboard Ship #3. Nitrate reductase genes (cytoplasmic *narG* and periplasmic *napA*) had a combined abundance that markedly increased in ballast tank waters of Ships #1 (22.1%) and #2 (38.0%) relative to the Harbor water (9.0%) and then declined sharply to 11.3% in the Ship #3 library. The potential for sulfate and thiosulfate reduction were investigated through adenylylsulfate reductase (*aprA*) and thiosulfate reductase (*phsA/prsA*) genes, respectively. Total sulfidogenesis potential was marked by the abundance of the dissimilatory sulfite reductase gene, *dsrA*. All genes associated with the reduction of sulfoxyanions remained at relatively low levels in all samples excluding Ship #3. Between Ship #2 and Ship #3, there were increases of 4-fold in *aprA*, 5-fold in *phsA*, and 5-fold in *dsrA* genes. These were the highest levels observed for each of these genes within this study and represented increases relative to the Harbor water library ranging between 3.6-fold for adenylylsulfate reductase to 30-fold for thiosulfate reductase.

Fuel Biodegradation Genes & Metabolites

Normalized gene abundances and observed selected metabolic intermediates associated with the activation and biodegradation of diesel fuel components are shown in Figures 4 & 5. Oxygen-dependent mechanisms for the activation of diesel components were assayed by quantifying the proportion of genes annotated as alkane 1-monooxygenase (*alkB*) for aliphatic substrates, ethylbenzene dioxygenase (*etbAa*), toluene monooxygenase (*tmoA*), benzene/toluene dioxygenase (*todC1*), and xylene monooxygenase (*xylM*) for BTEX hydrocarbons, phenol hydroxylase (*dmpL*) and phenol 2-monooxygenase for phenolic compounds, and naphthalene 1,2-dioxygenase (*nahAc*) and 1-methylnaphthalene hydroxylase for polyaromatic hydrocarbons. Catechol 1,2-dioxygenase (*catA*) and catechol 2,3-dioxygenase (*catE*), along with salicaldehyde dehydrogenase and salicylate hydroxylase were used as metrics for genetic potential for downstream aerobic metabolism of mono- and polyaromatic hydrocarbons. The alkyl-, benzyl-, and naphthylmethyl-succinate synthases (*assA, bssA, nmsA*) served as indicator genes for activation of aliphatic, mono-, and polyaromatic hydrocarbons by addition to fumarate. The ethylbenzene dehydrogenase gene *ebdA* served as a marker for the potential for the anaerobic activation of ethylbenzene by dehydrogenation and subsequent hydroxylation of the ethyl-substituent.

Alkane monooxygenase (12.4%) was the only hydrocarbon activation gene (aerobic or anaerobic) that constituted >2.0% of the total Harbor water community potential. No genes associated with the anoxic activation of hydrocarbon substrates were identified. None of the signature aerobic or anaerobic hydrocarbon biodegradation metabolites (Fig. 5) were detected in the San Diego seawater sample.

The aerobic BTEX activating portion of the Ship #1 community bloomed significantly relative to the Harbor community with ethylbenzene dioxygenase, xylene monooxygenase, and phenol hydroxylase increasing by 366-fold, 92.6-fold, and 8.0-fold respectively. The ring-cleaving enzyme catechol 2,3-dioxygenase also increased significantly in Ship #1 by 11.4-fold to 25.3% of the total community, while catechol 1,2-dioxygenase increased by 3-fold to a proportion of only 2.5% of the assemblage. Metabolite analysis identified the presence of phenylethanol, 3-methylcatechol, 4-methyl catechol, and methylnaphthalene-diol, indicative of the oxygen-dependent biological activation of mono- and polyaromatic hydrocarbons. There was a slight increase of 12% in the proportion flora possessing *alkB,* and genes related to the aerobic activation of naphthalenes remained at less than 1% of the community potential. Genes for the alkyl- (2.1%), "mono-aromatic"- (0.3%), and naphthylmethyl (1.2%)-succinate synthases all showed slight increases relative to the harbor water metagenome, but these values are likely inflated by the presence of other common glycyl-radical enzymes (e.g. pyruvate-formate lyase) that share significant protein sequence homology. Despite these gene enrichments, no succinate-derivative metabolites were observed in the Ship #1 sample.

The community within Ship #2 saw a further 11.6% enrichment of *alkB* possessing taxa over that of Ship #1. In contrast, the genes for ethylbenzene dioxygenase, xylene monooxygenase, and phenol hydroxylase declined sharply to account for less than 2.0% of the total community each. The anaerobic gene *nmsA* increased by 440.1% relative to Ship #1 to represent 7.7% of the Ship #2 population. Overall, the total proportion of hydrocarbonoclastic taxa in Ship #2 decreased as

compared to Ship #1, though there was an enrichment in oxygen-independent activation genes (Fig. 6).

The ballast tank water of Ship #3 contained the largest hydrocarbon-activating community of any sample investigated in this study. Alkane 1-monooxygenase (*6.1%*) and catechol dioxygenase (*catA/catE*) genes (0.4%, 2.0%) decreased to their smallest proportion observed in any sample. In contrast, the community possessing *assA* (5.7%), *bssA/hbsA/ibsA* (4.7%) and *nmsA* (20.5%) were found in the highest respective abundances. Mono- and poly-aromatic succinate metabolites were present corresponding to alkylated aromatic hydrocarbon substrates. No catechols or aromatic dihydrodiol intermediates were observed in the Ship #3 water sample. Naphthoic and alkanoic acids, aliphatic alcohols, and phenylethanol were also putatively identified in the Ship #3 water sample, as well as in Ship #1.

Discussion

This study investigated the role of marine microorganisms in diesel fuel biodegradation and infrastructure biofouling/biocorrosion aboard naval surface warfare vessels. The temporal dynamics of community membership and metabolic activities within seawater-compensated fuel ballast tanks were characterized through metagenomics and metabolomics paired with geochemical analyses of ballast tank water samples. All ships sampled in this study were stationed in San Diego Bay, CA USA and routinely take seawater into their ballast tanks. The Harbor water sample obtained from San Diego Bay, CA USA served as the baseline control for comparison of microbiome successional patterns as a function of increasing residence time aboard ship.

Ribosomal RNA gene analysis revealed the surface waters of San Diego Harbor to be primarily colonized by *Bacteria* belonging to the *Alphaproteobacteria, Flavobacteria,* and *Gammaproteobacteria* and marine eukaryotes of the *Alveolata* and *Stramenopiles* (Fig2.). This community differs from the surface water community of the North Pacific Tara Station 133, ~635 miles NW, that was dominated by the *Actinobacteria* (35%), with much smaller proportions of the *Gammaproteobacteria* (11%), *Alphaproteobacteria* (7%), and *Flavobacteria* (2%) (Sunagawa et al., 2015). San Diego Bay receives terrestrial discharges from the major urban, industrial, and military instillations within the San Diego, CA metropolitan area. These inputs and repeated chemical releases have resulted in the chronic introduction of pesticides, polyhalogenated aromatic compounds, and fuel hydrocarbons into the bay. The presence these organic pollutants may play a role in enriching for different resident marine heterotrophic taxa within this littoral habitat.

The marine microbial community was significantly altered after a one-week residence time within a ballast tank (Ship #1) as compared to the initial harbor community. Despite the total cell population decrease of ~100-fold (Table 2), there was only a slight decrease in community diversity and an increase in observed taxa (Table 3). *Gammaproteobacteria* were enriched to be the dominant community members, predominantly driven by increases in the *Altermonodales, Oceanospiralles,* and *Thiotrichales* (Fig. 2). These bacterial Orders contain numerous well documented hydrocarbonoclastic aerobic taxa of the belonging to the genera *Marinobacter*, *Alcanivorax,* and *Cycloclasticus,* respectively. Several previous studies have noted the enrichment of each of these groups in marine waters impacted by oil spills across the

globe (Geiselbrecht et al., 1998; Kasai et al., 2002; Maruyama et al., 2003; Dubinsky et al., 2013; King et al., 2015). *Marinobacter hydrocarbonoclasticus* has previously been shown to preferentially degrade short to moderate length n-alkanes (C_8-C_{14}) and selected mono-aromatic hydrocarbons, specifically ethylbenzene when incubated with F-76 diesel fuel. This organism also showed limited transformation capacity for alkylbenzenes, but no activity towards naphthalene (Striebich et al., 2014). Pure culture studies of *M. vinifirmus* and *M. hydrocarbonoclasticus* further demonstrated the complete oxidation of BTEX hydrocarbons under aerobic conditions (Berlendis et al., 2010). *Alcanivorax* spp. have repeatedly been documented to degrade a wide range of aliphatic hydrocarbons under aerobic conditions in marine environments (Yakimov et al., 1998; Harayama et al., 2004; Liu and Shao, 2005). To date, no studies have linked *Alcanivorax* spp. to the biodegradation of aromatic hydrocarbons and the genomes of *A. dieselolei* B-5 and *A. borkumensis* SK2 lack the genes for the catechol dioxygenases (*catA*/*xylM*) that are essential for the downstream metabolism of BTEX and PAHs (Schneiker et al., 2006; Lai et al., 2012). *Cycloclasticus pugetti* is an obligate aerobe with the capacity to degrade mono- and polyaromatic hydrocarbons (Dykesterhouse et al., 1995). *Cycloclasticus* strains isolated from coastal waters in the Gulf of Mexico (Geiselbrecht et al., 1998) and Kamaishi Bay, Japan (Kasai et al., 2002), were also associated with the aerobic degradation of a variety of alkylated PAHs and dibenzothiophenes. Genetic studies revealed the presence of distinct naphthalene and biphenyl cluster dioxygenases in several *Cycloclasticus* isolates (Geiselbrecht et al., 1998). Inspection of the publically available genome sequence for *C. pugetti* PS-1 (IMG database) also revealed the presence of multiple copies of ethylbenzene

dioxygenase and toluate dioxygenase genes. Thus, the increased proportions of *Alteromonodales* and *Thiotrichales* likely account for the enrichment of aromatic hydrocarbon degradation genes and presence of signature aerobic metabolites (e.g. alkylcatechols & dihydromethylnaphthalene) observed in Ship #1 (Figs. 4 & 5). Based on phylogenetic, genomic, and metabolic evidence suggests that the ballast tank community of Ship #1 is composed of taxa responsible for the aerobic oxidation of the more water-soluble aromatic hydrocarbon components of F-76 diesel fuel.

The microbiome of Ship #2 was also predominantly colonized by members of the *Alteromonadales*, but also had a sharp increase in the Deltaproteobacteria, specifically the *Desulfuromonadales* (Fig. 2). In concert with the decreased relative abundance of the *Gammaproteobacteria*, there was also a decrease in the aerobic hydrocarbon-activation genes relative to Ship #1 (Fig. 4). Despite the bloom in *Desulfuromonadales*, there was surprisingly little enrichment in genes identified in sulfidogenesis (Fig. 3). Anaerobic hydrocarbon-activation genes were also observed in a limited capacity, exhibiting only a slight increase in *nmsA* abundance relative to Ship #1 (Fig. 4). The *Desulfuromonadaceae* are strictly anaerobic heterotrophs, with some aerotolerance capacity, that typically respire sulfur compounds to sulfide, with the exception of some *Pelobacter* spp. that exhibit strictly fermentative metabolism. Members of this Family typically utilize fatty acids and alcohols as substrates, but to date none have been demonstrated to activate or oxidize hydrocarbons (Kuever et al., 2005). *D. acetoxidans* and *D. palmitatis* were shown to reduce ferric iron and several other species have been documented to reduce Mn(IV) as an electron acceptor (Roden and Lovley, 1993; Finster et al., 1994; Coates et al., 1995; Vandieken et al., 2006). The

oxygen tolerance and ability to utilize metals as electron acceptors, suggests these taxa bloomed in response to environmental conditions with oxygen becoming increasingly scarce in the ballast tank, and potentially establishing localized anoxic zones in Ship #2. As no metabolic or genomic evidence suggested a role for these organisms in fuel biodegradation, it is possible that aerobic catabolites produced by the cohabitating *Alteromonadales* serve as electron donors and carbon sources.

After 32-weeks, the ballast tank community of Ship #3 was predominantly colonized by anaerobic taxa affiliated with the *Deltaproteobacteria* (42%), specifically the *Desulfobacteraceae* (27%) and *Desulfarculaceae* (11%) (Fig. 2), with a concomitant increase in sulfoxyanion reductive respiration genes to the highest levels observed in this study (Fig. 3). Conversely, the Ship #3 community possessed the smallest population of c-type cytochrome oxidase containing taxa found in any sample (Fig. 3). Metagenome analysis estimated that of all observed genomes, 27.0% encoded an adenylylsulfate reductase and 24.8% a thiosulfate reductase (Fig. 3). This community also possessed the highest levels of anaerobic hydrocarbon activation genes: alkylsuccinate synthase (5.7%), mono-aromatic succinate synthases (4.7%), and naphthylmethylsuccinate synthases (20.5%). Several species within the *Desulfobacterales* are documented hydrocarbon-degrading anaerobes targeting a variety of aliphatic and aromatic substrates. *Desulfobacula toluolica* Tol2 (Rabus et al., 1993; Rabus and Widdel, 1995; Wöhlbrand et al., 2013), *Desulfosarcina cetonica*, formerly *Desulfobacterium cetonicum* (Galushko and Rozanova, 1991), *Desulfobacteraceae* strain mXS1 (Harms et al., 1999) and *Desulfosarcina ovata* oXyS1 (Harms et al., 1999; Keuver et al., 2005) utilize toluene as well as various other alkylbenzenes and cresol

isomers as electron donors coupled to the reduction of sulfate or thiosulfate to sulfide. *D. tolulica* and *D. cetonica* activate toluene and cresols to the corresponding monoarylsuccinates via benzylsuccinate synthase (Bss) and hydroxybenzylsuccinate synthase (Hbs), (Müller et al., 1999; Muller et al., 2001; Wöhlbrand et al., 2013). Recently, a more nutritionally restricted member of the *Desulfosarcina*, *D*. sp. PP31 was described to completely oxidize *p*-xylene, but was unable to utilize any other BTEX substrate or xylene isomers (Higashioka et al., 2012). Currently, the activation mechanism utilized by this organism remains unknown. Metabolite analysis of culture fluids of the currently unclassified marine *Desulfobacteraceae* strain EbS7 revealed ethylbenzene is initially activated by addition of the substituent to fumarate by a Bss reaction, as mentioned for related *Deltaproteobacteria*. A small group of unclassified *Desulfobacteraceae* isolates, NaphS2 (Galushko et al., 1999) and NaphS3 and NaphS6 (Musat et al., 2009), are the best described PAH utilizing sulfate-reducing taxa. These three isolates have been shown to activate naphthalene and 2-methylnaphthalene by parallel pathways of direct carboxylation and addition to fumarate resulting in the corresponding napthoate and 2-(naphthylmethyl)succinate, respectively (Musat et al., 2009; DiDonato et al., 2010; Mouttaki et al., 2012). Beyond the aromatic hydrocarbons, several species have been characterized utilizing aliphatic compounds. *Desulfatibacillum alkenivorans* AK-01 (Callaghan et al., 2006), the well-established model organism for fumarate addition activation of aliphatic hydrocarbons, as well as *D. aliphaticivorans* CV2803 (Cravo-Laureau et al., 2004) have been shown to catabolize a wide range of *n*-alkanes and *n*-alkenes under anoxic conditions (So and Young, 1999; Callaghan et al., 2012). In contrast, *Desulfococcus oleovorans* Hxd3 has

a similar substrate utilization profile, but has been shown to activate these hydrocarbons via a sub-terminal oxidation mechanism catalyzed by the putative oxygen-independent Alkane C2-Methylene Hydroxylase Complex, followed by carboxylation and mineralization via β-oxidation (Heider et al., 2016). Thus, given the diversity of hydrocarbonoclastic taxa within this group with the capacity for the catabolism of both aromatic and aliphatic substrates, it is likely that the organisms within the Ship #3 assemblage affiliated with *Desulfobacterales* harbor the succinate synthase A genes identified in the metagenome and are primarily for the fuel biodeterioration within this ballast tank. This is further supported by the concomitant identification of mono- and polyaromatic succinate metabolites in the fluids of Ship #3 (and only in this sample) along with the sharp increase in genomes encoding *bssA/hbsA/ibsA* and *nmsA* alleles (Figs. 4 and 5).

From an ecological perspective, the fuel compensated ballast tank habitat is a very interesting ecosystem. After a very short exposure (7 days) to a habitat with large quantities of toxic refined petroleum and an absence of light, a typical coastal marine community based upon phototrophic primary production shifted to one driven by heterotrophic consumption of diesel hydrocarbons and essentially a complete loss of the eukaryotic component of the ecosystem. The resulting hydrocarbonoclastic community of Ship #1 is very similar to the response of open-ocean surface water ecosystem to the sudden enormous influx of petroleum compounds from the Deep Water Horizon Oil spill. Here too, surface water assemblages in the Gulf of Mexico underwent dynamic shifts within the first week and experienced strong enrichments of hydrocarbonoclastic aerobes of the *Oceanospiralles, Alteromonadales,* and *Thiotrichales* (Dubinsky et al.,

2013). Over time, as the efforts to stem the unmitigated flow of crude oil and gas led to partial capture and ultimately the successful shut in, the rapidly growing *Oceanospiralles* declined with staggered enrichment of *Altermononadaceae* and *Flavobacteriaceae* with increasing time (Dubinsky et al., 2013). The change in pre-spill surface ratio of *Alpha-*/*Gammaproteobacteria* from >1.7 to >1.0 is also observed in this study with a shift from 1.4 to ~0.2 between the San Diego Harbor and Ship #1 samples (Hazen et al., 2010; King et al., 2015). Thus, the community response dynamics within an artificial habitat devoid of light and surrounded by metallic infrastructure closely mimicked the successional pattern observed in the natural surface marine habitat. Figure 6 shows the shifts in community proportions encoding hydrocarbon activation genes, as well as the relative abundance of genes from oxygen-dependent and independent pathways. After 7 days, the hydrocarbonoclastic community is enriched to 2.4 times that observed in San Diego Harbor. At 20-weeks and beyond, the abundance decreased to approximately only 1.5 times that of the Harbor assemblage. Within Ship #1, 95% of the activation genes were associated with aerobic mechanisms associated with well documented aerobic hydrocarbonoclastic taxa, aerobic respiration genes, signature aerobic catabolic intermediates, and a decreased dissolved oxygen content (Figs. 2,3,5, and 6; Table 1). As residence time increased, the fuel-degrading activation gene subset was increasingly affiliated with anoxic mechanisms after the initial bloom. At 32-weeks, the hydrocarbon-degrading community predominantly employed oxygen-independent mechanisms (68%) with key metabolites associated with the fumarate addition-mediated degradation of BTEX and PAH fuel components (Figs. 5 and 6). Thus, the Ship #2 sample appeared to

approximate a transition between two distinct hydrocarbonoclastic marine microbiomes within the fuel compensated ballast tanks. Based upon phylogenetic, functional genomic and metabolic evidence, these biomes are supported primarily by the oxidation of more water-soluble aromatic fuel components with succession driven by electron acceptor availability.

As demonstrated in this study and others involving marine waters and fuel mixtures, microorganisms endemic to seawater habitats in both littoral and oceanic zones are capable of degrading hydrocarbons under oxic and hypoxic conditions. Previous work also clearly established the link between microbial biodeterioration of hydrocarbons and the biocorrosion of carbon steel. Analysis of dissolved metal concentrations present within each sample served as potential indicators of ballast tank corrosion (Table 1). The levels of dissolved metals in all of the ballast tank fluids are considerably greater than the surrounding harbor waters, and thus likely are the result of release from the ballast tank infrastructure. Total dissolved manganese, copper, and nickel appeared to positively correlate with the increase in sulfidogenic taxa between Ships #1 and #2 (Fig. 6). There was also a strong increase in the abundance of sulfide-bearing particulates of each metal species associated with the strongly sulfidogenic community of Ship #3. A thermodynamic model based upon the premise of microbial sulfidogenesis in the presence of various metals/alloys predicts that sulfides will react with metal oxides to produce metallosulfide particles, thus potentially exacerbating the depletion of the protective oxide layers (McNeil and Odom, 1994). Copper-nickel alloys are commonly utilized in marine infrastructure, but are highly susceptible to corrosion and leaching by sulfides and organic acids and other products of microbial

metabolism (Pope, 1987; Wagner and Little, 1993). Previous research on localized corrosion of low carbon steels in marine systems showed that pit initiation begins at MnS inclusions acting as cathodic sites to the surrounding iron matrix and that pit development releases soluble manganese into the surrounding fluid matrix (Melchers et al., 2016). Analysis of metal concentrations and particulates suggests that the CuNi-alloy sluice pipes transporting water and fuel between ballast tanks were the primary sites of active biocorrosion in Ships #1 and #2 and that aggressive pitting corrosion of the ballast tank carbon steel was occurring in Ship #3.

This study demonstrated key findings in regards to the operation and maintenance of fuel compensated ballast tanks of the United States Navy: petroleum-derived diesel fuels are susceptible to biological deterioration when in contact with seawater, regardless of oxygenation status or residence time; and that shipboard infrastructure is negatively impacted by microbiologically influenced corrosion and fouling driven by the degradation of fuels and production of sulfides. Studies conducted on simulated ballast tank systems highlighted the inherent stratified nature of these habitats with zones of variable oxygenation (Lee et al., 2004; Heyer et al., 2013). Furthermore, the deleterious impacts of microbial activity were not homogenously distributed within the ballast tank systems, rather horizontal surfaces (e.g. tank bottoms) were sites of aggressive general and localized corrosion (Cleland, 1995; Lee et al., 2004; J. S. Lee et al., 2012; Heyer et al., 2013). Guidelines issued for practices aimed at reducing the wastage of ballast tanks aboard naval vessels have suggested the strict deoxygenation of displacement water and coating of internal surfaces (Askheim Erik, 1997; Goldie, 2012). Ship operators have estimated that a vessel on standard operations

consumes ~1000 gallons of fuel per day, necessitating the introduction of new seawater into ballast tanks on a regular basis (personal communication). This scenario of cyclic oxygenation/deoxygenation has repeatedly been observed to promote the highest rates of generalized and pitting corrosion (J. S. Lee et al., 2012). Furthermore, this condition promotes the persistence of biofilms harboring both aerobic and anaerobic taxa with the capacity for sulfur cycling and the formation of corrosive sulfur granules (Cleland, 1995). As treatment of incoming seawater for the removal of oxygen and sulfate is not readily feasible, a critical consideration for ship operators is to eliminate the introduction of sediments into ballast tanks and to ensure all internal surfaces are coated. Sediments settle onto horizontal surfaces and promote further stratification within the ballast tank and likely serve as a source of sulfide-producing prokaryotes (e.g. *Desulfobacterales*). Given the stratified nature of these systems, further research is warranted to determine how well the sampling of expansion tanks represents the various ballast tank zones. Ultimately, this study identified successional trends in the ecology of ballast tank ecosystems that are critical for understanding the deleterious impact of microorganisms on infrastructure and will hopefully serve to inform future studies to lessen the biological wastage of naval resources.

References

Aeckersberg, F., Bak, F., and Widdel, F. (1991) Anaerobic oxidation of saturated hydrocarbons to CO_2 by a new type of sulfate-reducing bacterium. *Arch. Microbiol.* **156**: 5–14.

Aeckersberg, F., Rainey, F.A., and Widdel, F. (1998) Growth, natural relationships, cellular fatty acids and metabolic adaptation of sulfate-reducing bacteria that utilize long-chain alkanes under anoxic conditions. *Arch. Microbiol.* **170**: 361–369.

Aktas, D.F., Lee, J.S., Little, B.J., Ray, R.I., Davidova, I. a., Lyles, C.N., and Suflita, J.M. (2010) Anaerobic metabolism of biodiesel and its impact on metal corrosion. *Energy and Fuels* **24**: 2924–2928.

Askheim Erik (1997) Ballast tanks and cargo holds in DNV`s guidelines for corrosion protection of ships. *PCE, June* **2**: 26–35.

Ben-Dov, E., Brenner, A., and Kushmaro, A. (2007) Quantification of sulfate-reducing bacteria in industrial wastewater, by real-time polymerase chain reaction (PCR) using dsrA and apsA genes. *Microb. Ecol.* **54**: 439–451.

Berlendis, S., Cayol, J.-L., Verhé, F., Laveau, S., Tholozan, J.-L., Ollivier, B., and Auria, R. (2010) First evidence of aerobic biodegradation of BTEX compounds by pure cultures of *Marinobacter*. *Appl. Biochem. Biotechnol.* **160**: 1992–1999.

Biers, E.J., Sun, S., and Howard, E.C. (2009) Prokaryotic genomes and diversity in surface ocean waters: Interrogating the global ocean sampling metagenome. *Appl. Environ. Microbiol.* **75**: 2221–2229.

Bolger, A.M., Lohse, M., and Usadel, B. (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinforma.* **30** : 2114–2120.

Buchfink, B., Xie, C., and Huson, D.H. (2015) Fast and sensitive protein alignment using DIAMOND. *Nat Meth* **12**: 59–60.

Callaghan, A. V, Morris, B.E.L., Pereira, I.A.C., McInerney, M.J., Austin, R.N., Groves, J.T., et al. (2012) The genome sequence of *Desulfatibacillum alkenivorans* AK-01: a blueprint for anaerobic alkane oxidation. *Environ. Microbiol.* **14**: 101–113.

Callaghan, A.V. and Wawrik, B. (2016) AnHyDeg: a curated database of anaerobic hydrocarbon degradation genes (https://github.com/AnaerobesRock/AnHyDeg). DOI: 10.5281/zenodo.61278

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

Cleland, J.H. (1995) Corrosion Risks in Ships ' Ballast Tanks and the IMO Pathogen Guidelines. *Eng. Fail. Anal.* **2**: 79–84.

Coates, J.D., Lonergan, D.J., Philips, E.J.P., Jenter, H., and Lovley, D.R. (1995) *Desulfuromonas palmitatis* sp. nov., a marine dissimilatory Fe(III)-reducer that can oxidize long-chain fatty acids. *Arch. Microbiol.* **164**: 406–413.

Cravo-Laureau, C., Matheron, R., Cayol, J.-L., Joulian, C., and Hirschler-Réa, A. (2004) *Desulfatibacillum aliphaticivorans* gen. nov., sp. nov., an *n*-alkane- and *n*-alkene-degrading, sulfate-reducing bacterium. *Int. J. Syst. Evol. Microbiol.* **54**: 77–83.

Creek, D.J., Jankevics, A., Burgess, K.E. V, Breitling, R., and Barrett, M.P. (2012) IDEOM: an Excel interface for analysis of LC-MS-based metabolomics data. *Bioinformatics* **28**: 1048–9.

DiDonato, R.J., Young, N.D., Butler, J.E., Chin, K.-J., Hixson, K.K., Mouser, P., et al. (2010) Genome sequence of the deltaproteobacterial strain NaphS2 and analysis of differential gene

expression during anaerobic growth on naphthalene. *PLoS One* **5**: e14072.

Dubinsky, E.A., Conrad, M.E., Chakraborty, R., Bill, M., Borglin, S.E., Hollibaugh, J.T., et al. (2013) Succession of hydrocarbon-degrading bacteria in the aftermath of the deepwater horizon oil spill in the gulf of Mexico. *Environ. Sci. Technol.* **47**: 10860–10867.

Dykesterhouse, S.E., Gray, J.P., Herwig, R.P., Lara, J.C., and Staley, J.T. (1995) *Cycloclasticus pugetii* gen. nov., sp. nov., an aromatic hydrocarbon-degrading bacterium from marine sediments. *Int. J. Syst. Bacteriol.* **45**: 116–123.

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

Elshahed, M.S., Gieg, L.M., McInerney, M.J., and Suflita, J.M. (2001) Signature metabolites attesting to the in situ attenuation of alkylbenzenes in anaerobic environments. *Environ. Sci. Technol.* **35**: 682–689.

Finster, K., Bak, F., and Pfennig, N. (1994) *Desulfuromonas acetexigens* sp. nov., a dissimilatory sulfur-reducing eubacterium from anoxic freshwater sediments. *Arch. Microbiol.* **161**: 328–332.

Galushko, Minz, Schink, and Widdel (1999) Anaerobic degradation of naphthalene by a pure culture of a novel type of marine sulphate-reducing bacterium. *Environ. Microbiol.* **1**: 415–420.

Galushko, A. and Rozanova, E. (1991) *Desulfobacterium cetonicum* sp. nov.: a sulfate-reducing bacterium which oxidizes fatty acids and ketones. *Microbiologiya* **60**: 102–107.

Geiselbrecht, A.D., Hedlund, B.P., Tichi, M.A., and Staley, J.T. (1998) Isolation of marine polycyclic aromatic hydrocarbon (PAH)-degrading *Cycloclasticus* strains from the Gulf of Mexico and comparison of their PAH degradation ability with that of Puget Sound *Cycloclasticus* strains. *Appl. Environ. Microbiol.* **64**: 4703–10.

Gieg, L.M. and Suflita, J.M. (2005) Petroleum Microbiology Ollivier,B. and Magot,M. (eds) American Society of Microbiology.

Goldie, B. (2012) Corrosion protection of cargo tanks in crude oil carriers: the new IMO regulations. *J. Prot. Coatings Linings* **29**: 24–33.

Harayama, S., Kasai, Y., and Hara, A. (2004) Microbial communities in oil-contaminated seawater. *Curr. Opin. Biotechnol.* **15**: 205–214.

Harms, G., Zengler, K., Rabus, R., Aeckersberg, F., Minz, D., Rosselló-Mora, R., and Widdel, F. (1999) Anaerobicoxidation of *o*-xylene, *m*-xylene, and homologous alkylbenzenes by new types of sulfate-reducing bacteria. *Appl. Environ. Microbiol.* **65**: 999–1004.

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**: 204–8.

Hazlett, R.N. (1966) Control of sulfate-reducing bacteria in navy fuel systems Washington, DC.

Hazlett, R.N. (1967) Examination of corrosion deposits from uss iwo jima aviation gasoline storage tanks Washington, DC.

Heider, J., Szaleniec, M., Sünwoldt, K., and Boll, M. (2016) Ethylbenzene degydrogenase and related molybdenum enzymes involved in oxygen-independent alkyl chain hydroxylation. *J Mol Biotechnol.* **26**: 45-62.

Heyer, A., D'Souza, F., Morales, C.F.L., Ferrari, G., Mol, J.M.C., and de Wit, J.H.W. (2013) Ship ballast tanks a review from microbial corrosion and electrochemical point of view. *Ocean Eng.* **70**: 188–200.

Higashioka, Y., Kojima, H., and Fukui, M. (2012) Isolation and characterization of novel sulfate-reducing bacterium capable of anaerobic degradation of *p*-xylene. *Microbes Environ.* **27**: 273–277.

Howard, E.C., Sun, S., Biers, E.J., and Moran, M.A. (2008) Abundant and diverse bacteria involved in DMSP degradation in marine surface waters. *Environ. Microbiol.* **10**: 2397–2410.

Huang, R.T., McFarland, B.L., and Hodgman, R.Z. (1997) Microbial influenced corrosion in cargo oil tanks of crude oil tankers. In, *Corrosion*. NACE International, Houston, TX.

Kanehisa, M., Goto, S., Sato, Y., Furumichi, M., and Tanabe, M. (2012) KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Res.* **40** : D109–D114.

Kasai, Y., Kishira, H., and Harayama, S. (2002) Bacteria belonging to the genus *Cycloclasticus* play a primary role in the degradation of aromatic hydrocarbons released in a marine environment. *Appl. Environ. Microbiol.* **68**: 5625–5633.

Keuver, J., Rainey, F.A., and Widdel, F. (2005) Genus X. *Desulfosarcina* Widdel 1981, 382VP. In, Brenner,D.J., Krieg,N.R., Staley,J.T., and Garrity,G.M. (eds), *Bergey's Manual of Systematic Bacteriology*. Springer, New York, pp. 981–984.

King, G.M., Kostka, J.E., Hazen, T.C., and Sobecky, P.A. (2015) Microbial responses to the deepwater horizon oil spill: from coastal wetlands to the deep sea. *Ann. Rev. Mar. Sci.* **7**: 377–401.

Klemme, D.E. and Leonard, J.M. (1971) Inhibitors for marine sulfate-reducing bacteria in shipboard fuel storage tanks Washington, DC.

Klemme, D.E. and Neihof, R.A. (1969) Control of marine sulfate-reducing bacteria in water-displaced shipboard fuel storage tanks Washington, DC.

Kuever, J., Rainey, F.A., and Widdle, F. (2005) Family I. *Desulfuromonadaceae*. In, Brenner,D.J., Krieg,N.R., Staley,J.T., and Garrity,G.M. (eds), *Bergey's Manual of Systematic Bacteriology*. Springer, New York, p. 1006.

Lai, Q., Li, W., and Shao, Z. (2012) Complete genome sequence of *Alcanivorax dieselolei* type strain B5. *J. Bacteriol.* **194**: 6674–6674.

Leader, D.P., Burgess, K., Creek, D., and Barrett, M.P. (2011) Pathos: a web facility that uses metabolic maps to display experimental changes in metabolites identified by mass spectrometry. *Rapid Commun. Mass Spectrom.* **25**: 3422–6.

Lee, J.S., Ray, R.I., Lemieux, E.J., Falster, A.U., and Little, B.J. (2004) An evaluation of carbon steel corrosion under stagnant seawater conditions. *Biofouling* **20**: 237–47.

Lee, J.S., Ray, R.I., Little, B.J., Duncan, K.E., Oldham, A.L., Davidova, I.A., and Suflita, J.M. (2012) Sulphide production and corrosion in seawaters during exposure to FAME diesel. *Biofouling* **28**: 465–78.

Lee, J.S., Ray, R.I., Little, B.J., and Lemieux, E.J. (2012) Evaluation of deoxygenation as a corrosion control measure for ballast tanks. *http://dx.doi.org/10.5006/1.3278153*.

Liu, C. and Shao, Z. (2005) *Alcanivorax dieselolei* sp. nov., a novel alkane-degrading bacterium isolated from sea water and deep-sea sediment. *Int. J. Syst. Evol. Microbiol.* **55**: 1181–1186.

Lyles, C.N., Aktas, D.F., Duncan, K.E., Callaghan, A. V, Stevenson, B.S., and Suflita, J.M. (2013) Impact of organosulfur content on diesel fuel stability and implications for carbon steel corrosion. *Environ. Sci. Technol.* **47**: 6052–6062.

Martin, M. (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17**: 10–12.

Maruyama, A., Ishiwata, H., Kitamura, K., Sunamura, M., Fujita, T., Matsuo, M., and Higashihara, T. (2003) Dynamics of microbial populations and strong selection for cycloclasticus pugetii following the nakhodka oil spill. *Microb. Ecol.* **46**: 442–453.

McNeil, M. and Odom, A. (1994) Thermodynamic prediction of microbiologically influenced corrosion (MIC) by sulfate-reducing bacteria (SRB). In, *Microbiologically Influenced Corrosion Testing*. ASTM International, 100 Barr Harbor Drive, PO Box C700, West Conshohocken, PA 19428-2959, pp. 173–173–7.

Melchers, R., Chaves, I., and Jeffrey, R. (2016) A conceptual model for the interaction between carbon content and manganese sulphide inclusions in the short-term seawater corrosion of low carbon steel. *Metals (Basel).* **6**: 132.

Mouttaki, H., Johannes, J., and Meckenstock, R.U. (2012) Identification of naphthalene carboxylase as a prototype for the anaerobic activation of non-substituted aromatic hydrocarbons. *Environ. Microbiol.* **14**: 2770–4.

Muller, J.A., Galushko, A.S., Kappler, A., and Schink, B. (2001) Initiation of anaerobic degradation of p-cresol by formation of 4-hydroxybenzylsuccinate in *Desulfobacterium cetonicum*. *J. Bacteriol.* **183**: 752–757.

Müller, J.A., Galushko, A.S., Kappler, A., and Schink, B. (1999) Anaerobic degradation of *m* -cresol by *Desulfobacterium cetonicum* is initiated by formation of 3-hydroxybenzylsuccinate. *Arch. Microbiol.* **172**: 287–294.

Musat, F., Galushko, A., Jacob, J., Widdel, F., Kube, M., Reinhardt, R., et al. (2009) Anaerobic degradation of naphthalene and 2-methylnaphthalene by strains of marine sulfate-reducing bacteria. *Environ. Microbiol.* **11**: 209–219.

Neihof, R.A. (1988) Microbes in Fuel: An Overview with a Naval Perspective. In, Chesneau,H.L. and Dorris,M.M. (eds), *Distillate Fuel: Contamination, Storage, and Handling, ASTM STP 1005*. American Society for Testing and Materials, Philadelphia, pp. 6–14.

Oldham, A.L., Drilling, H.S., Stamps, B.W., Stevenson, B.S., and Duncan, K.E. (2012) Automated DNA extraction platforms offer solutions to challenges of assessing microbial biofouling in oil production facilities. *AMB Express* **2**: 60.

Pope, D.H. (1987) Microbial corrosion in fossil-fired power plants - a study of microbiologically influenced corrosion and a practical guide for its treatment and prevention Palo Alto, CA.

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

Rabus, R., Nordhaus, R., Ludwig, W., and Widdel, F. (1993) Complete oxidation of toluene under strictly anoxic conditions by a new sulfate-reducing bacterium. *Appl. Environ. Microbiol.* **59**: 1444–51.

Rabus, R. and Widdel, F. (1995) Conversion studies with substrate analogues of toluene in a sulfate-reducing bacterium, strain Tol2. *Arch. Microbiol.* **164**: 448–51.

Roden, E.E. and Lovley, D.R. (1993) Dissimilatory Fe(III) reduction by the marine microorganism *Desulfuromonas acetoxidans*. *Appl. Environ. Microbiol.* **59**: 734–42.

Scheltema, R.A., Jankevics, A., Jansen, R.C., Swertz, M.A., and Breitling, R. (2011) PeakML/mzMatch: a file format, Java library, R library, and tool-chain for mass spectrometry data analysis. *Anal. Chem.* **83**: 2786–93.

Schneiker, S., Martins dos Santos, V.A.P., Bartels, D., Bekel, T., Brecht, M., Buhrmester, J., et al. (2006) Genome sequence of the ubiquitous hydrocarbon-degrading marine bacterium *Alcanivorax borkumensis*. *Nat. Biotechnol.* **24**: 997–1004.

So, C.M., Phelps, C.D., and Young, L.Y. (2003) Anaerobic transformation of alkanes to fatty acids by a sulfate-reducing bacterium, strain Hxd3. *Appl. Environ. Microbiol.* **69**: 3892–900.

So, C.M. and Young, L.Y. (1999) Isolation and characterization of a sulfate-reducing bacterium that anaerobically degrades alkanes. *Appl. Environ. Microbiol.* **65**: 2969–76.

Stevenson, B.S., Drilling, H.S., Lawson, P. a, Duncan, K.E., Parisi, V. a, and Suflita, J.M. (2011) Microbial communities in bulk fluids and biofilms of an oil facility have similar composition but different structure. *Environ. Microbiol.* **13**: 1078–90.

Striebich, R.C., Smart, C.E., Gunasekera, T.S., Mueller, S.S., Strobel, E.M., McNichols, B.W., and Ruiz, O.N. (2014) Characterization of the F-76 diesel and Jet-A aviation fuel hydrocarbon

degradation profiles of Pseudomonas aeruginosa and *Marinobacter hydrocarbonoclasticus*. *Int. Biodeterior. Biodegradation* **93**: 33–43.

Sunagawa, S., Coelho, L.P., Chaffron, S., Kultima, J.R., Labadie, K., Salazar, G., et al. (2015) Structure and function of the global ocean microbiome. *Science (80-.).* **348**: 1–10.

Tautenhahn, R., Böttcher, C., and Neumann, S. (2008) Highly sensitive feature detection for high resolution LC/MS. *BMC Bioinformatics* **9**: 504.

Vandieken, V., Mussmann, M., Niemann, H., and Jørgensen, B.B. (2006) *Desulfuromonas svalbardensis* sp. nov. and *Desulfuromusa ferrireducens* sp. nov., psychrophilic, Fe(III)-reducing bacteria isolated from Arctic sediments, Svalbard. *Int. J. Syst. Evol. Microbiol.* **56**: 1133–9.

Wagner, P. and Little, B.J. (1993) Impact of alloying on microbiologically influenced corrosion - a review. *Mater. Perform.* **32**: 65–68.

Wöhlbrand, L., Jacob, J.H., Kube, M., Mussmann, M., Jarling, R., Beck, A., et al. (2013) Complete genome, catabolic sub-proteomes and key-metabolites of *Desulfobacula toluolica* Tol2, a marine, aromatic compound-degrading, sulfate-reducing bacterium. *Environ. Microbiol.* **15**: 1334–1355.

Yakimov, M.M., Golyshin, P.N., Lang, S., Moore, E.R.B., Abraham, W.-R., Lunsdorf, H., and Timmis, K.N. (1998) *Alcanivorax borkumensis* gen. nov., sp. nov., a new, hydrocarbon-degrading and surfactant-producing marine bacterium. *Int. J. Syst. Bacteriol.* **48**: 339–348.

Zhu, F., Massana, R., Not, F., Marie, D., and Vaulot, D. (2005) Mapping of picoeucaryotes in marine ecosystems with quantitative PCR of the 18S rRNA gene. *FEMS Microbiol. Ecol.* **52**: 79–92.

Table 1. Summary of water sample residence times and selected chemical properties. Ship #2 ballast water residence time was reported from naval personnel between 16 and 24 weeks. *, Estimated from NOAA data for San Diego Harbor, San Diego, CA USA. Relative frequency of each metal observed in metal-sulfide particles analyzed by EDX is shown in parentheses. Not detected, ND; Not reported, NR.

	Harbor	**Ship #1**	**Ship #2**	**Ship #3**
Residence Time (weeks)	0	1	~20	32
Dissolved O_2 (ppm)	8*	3.7	ND	ND
Sulfate (mM)	29	29	29	25
Manganese (ppb)	0.16 (ND)	10.67 (NR)	125.07 (NR)	1119.28 (NR)
Copper (ppb)	0.40 (ND)	206.28 (47%)	864.51 (42%)	101.19 (85%)
Nickel (ppb)	0.60 (ND)	95.75 (26%)	1564.29 (11%)	23.90 (39%)

Table 2. Marine microbial population estimates and functional gene qPCR enumeration. [a]Cell counts reported as cells mL^{-1}. [b]Reported as gene copies mL^{-1}. Bacteria, Archaea and Picoeukaryote represent respective 16S and 18S rRNA genes. Adenosine-5'-phosphosulfate reductase A gene (*apsA*). Dissimilatory sulfite reductase A gene (*dsrA*). Below detection limit (BDL). Standard deviations are shown in parantheses for qPCR analyses.

	Harbor	**Ship #1**	**Ship #2**	**Ship #3**
Cell Counts[a]	$2.13x10^6$	$1.11x10^5$	$7.54x10^5$	$9.42x10^5$
Bacteria[b]	$2.14x10^6$ ($3.52x10^4$)	$6.44x10^7$ ($4.64x10^5$)	$1.05x10^5$ ($1.95x10^3$)	$7.35x10^5$ ($2.25x10^4$)
Archaea[b]	$1.81x10^2$ ($3.76x10^1$)	$1.93x10^3$ ($7.79x10^1$)	$2.28x10^2$ ($1.56x10^2$)	$1.05x10^2$ ($1.58x10^1$)
Picoeukaryotes[b]	$1.87x10^4$ ($5.78x10^3$)	$9.78x10^2$ ($3.02x10^1$)	$3.67x10^1$ (5.45)	$2.34x10^1$ (9.69)
apsA[b]	BDL	$9.69x10^2$ ($3.10x10^1$)	$5.84x10^2$ ($3.72x10^2$)	$2.07x10^4$ ($1.58x10^3$)
dsrA[b]	$6.90x10^1$ (7.03)	$1.39x10^4$ ($1.43x10^3$)	$3.38x10^2$ ($5.88x10^1$)	$2.20x10^4$ ($1.83x10^3$)

Table 3. Summary of metagenome library statistics and extracted community diversity metrics. Hits to [a]small subunit rRNA genes (SSU) and [b]RNA polymerase B (*rpoB*) fragments defined as any read with minimum homology scores defined in methods. [c]Diversity metrics were calculated based on extracted SSU reads extracted from each library rarified to a uniform sequencing depth of 1,800.

	Harbor	**Ship #1**	**Ship #2**	**Ship #3**
Metagenome Statistics				
Library size (reads)	5,598,474	8,603,258	5,344,908	6,479,122
Hits to SSU fragments[a]	1993	8320	2827	2368
Hits to *rpoB* fragments[b]	4602	9588	5221	5511
Diversity Metrics[c]				
Shannon Diversity	7.855	7.513	6.561	6.734
Observed OTUs	1091	1138	1052	1133
Simpson Evenness	0.055	0.041	0.019	0.023

Figure 1. Cartoon schematic of ballast tank system.

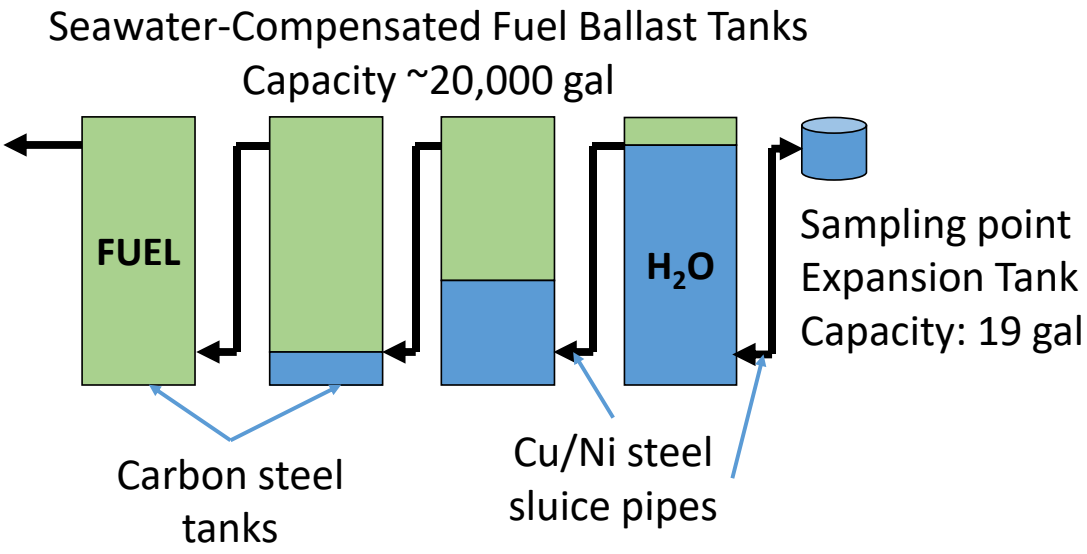


Figure 2. Metagenome analysis of community composition. Order-level taxa based on extracted 16S and 18S rRNA genes.

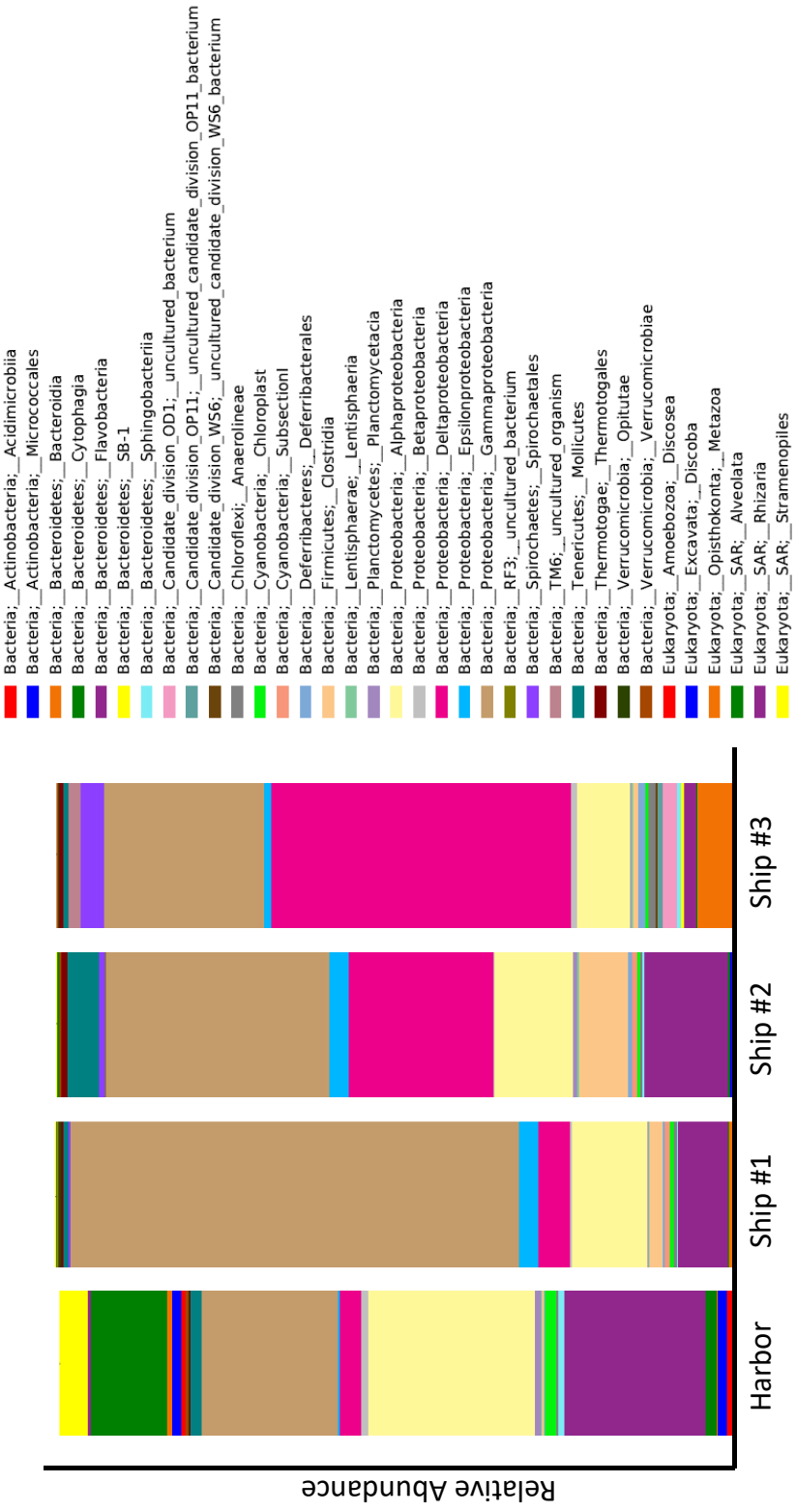


Figure 3. Metagenome analysis of selected energy conservation genes. All gene abundances are normalized to single copy marker gene (*rpoB*) abundance for each sample. Colors are scaled based on values in discrete rows. Gene abbreviations: *amoA*, ammonia monooxygenase; *hao*, hydroxylamine oxidoreductase; *mtoA/pioA*, decaheme c-type cytochrome involved in neutrophilic iron oxidation; *soxB*, sulfur oxidation gene; *sqr*, sulfide:quinone oxidoreductase; *ccoN* cbb3-type cytochrome C oxidase; *coxA*, aa3-type cytochrome C oxidase; *aprA*, adenylylsulfate reductase; *phsA*, thiosulfate reductase; *dsrA*, dissimilatory sulfite reductase.

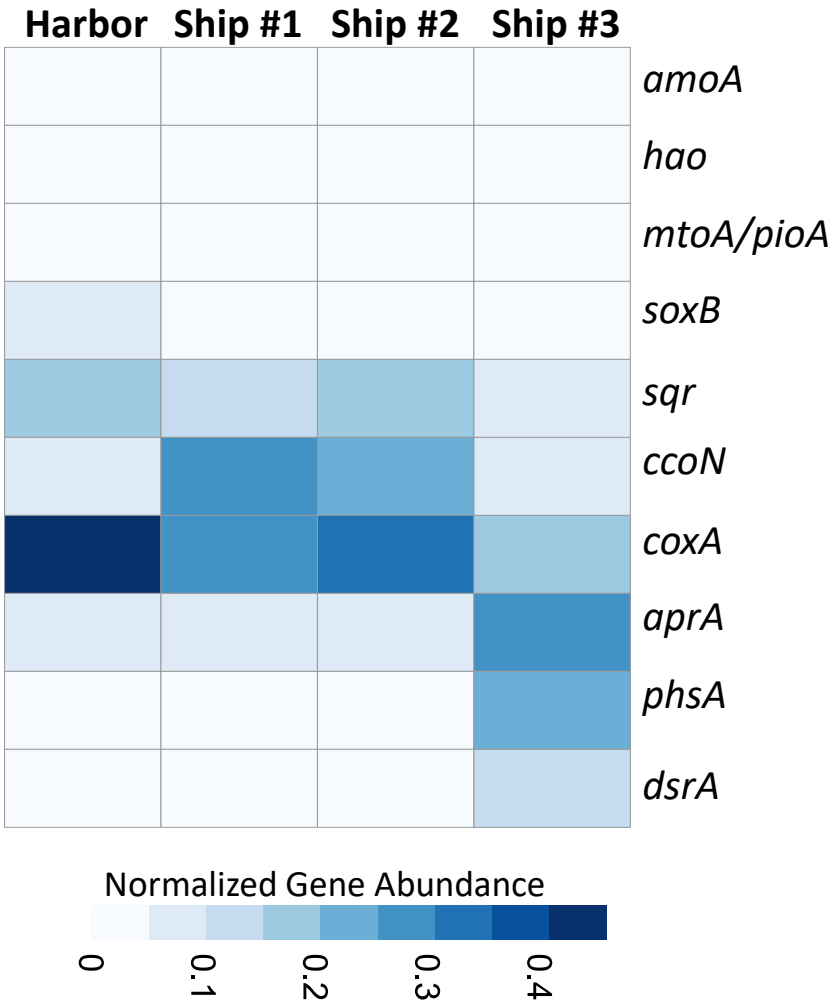


Figure 4. Metagenome analysis of selected hydrocarbon catabolic genes. All gene abundances are normalized to single copy marker gene (*rpoB*) abundance for each sample. Numbers correspond to genes associated with metabolic reactions indicated in figure 5.

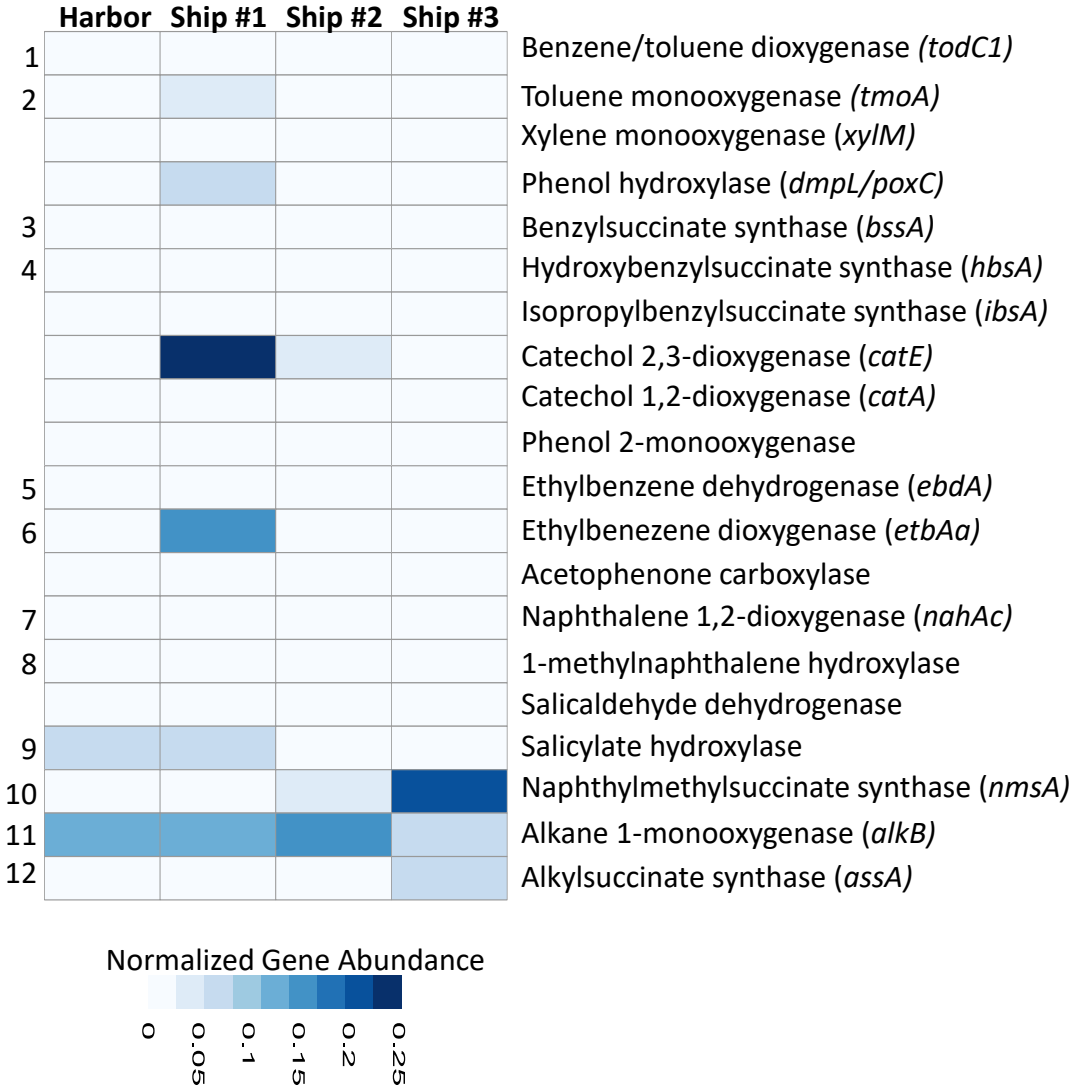


Figure 5. Analysis of fuel metabolites within Ship #1 and Ship #3 ballast tank fluids. Blue boxes indicate compounds identified in Ship #1 only, red in Ship #3 only, and black were observed in both samples. Numbers above arrows correspond to transformations catalyzed by metabolic genes listed in figure 4.

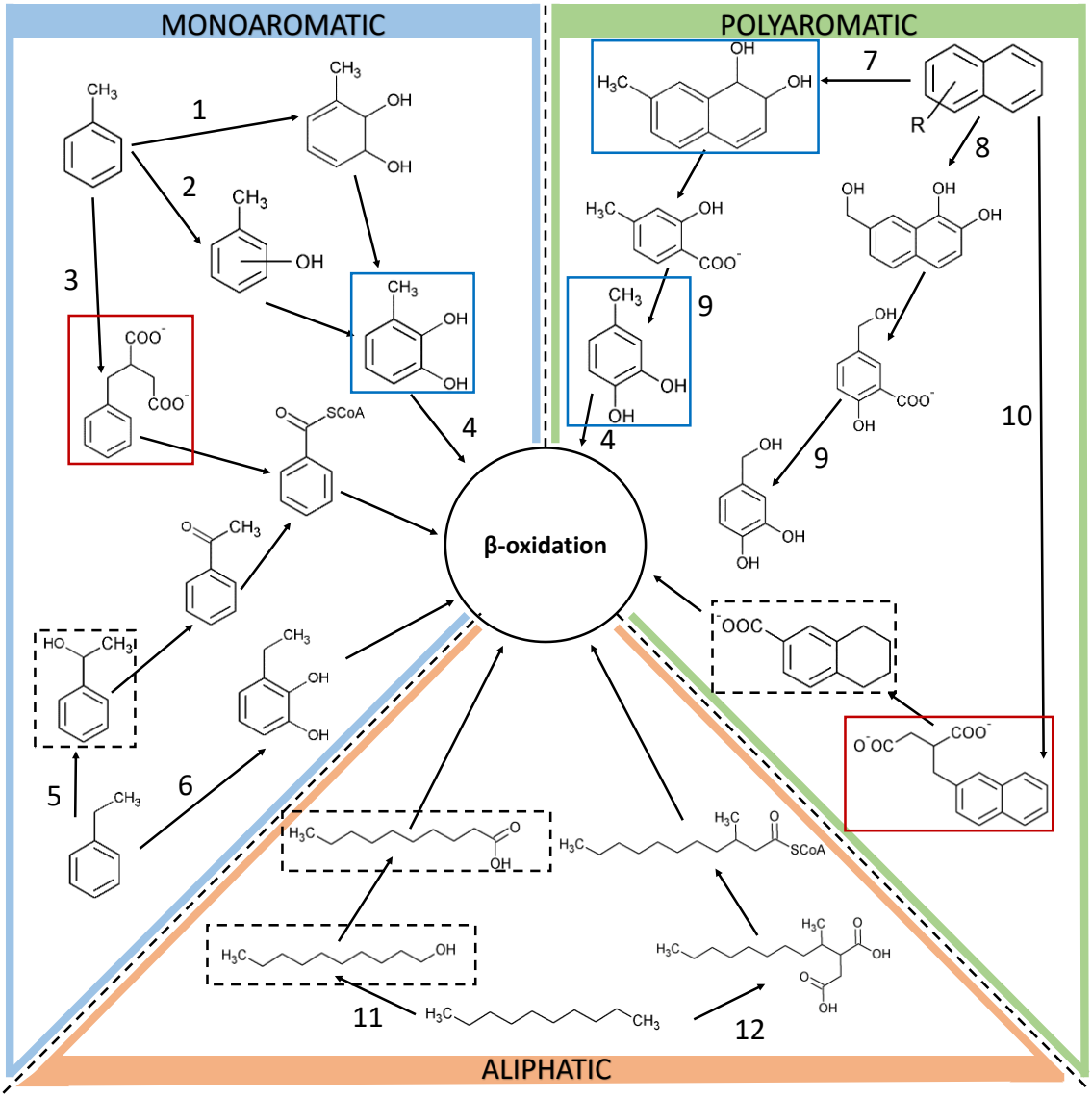
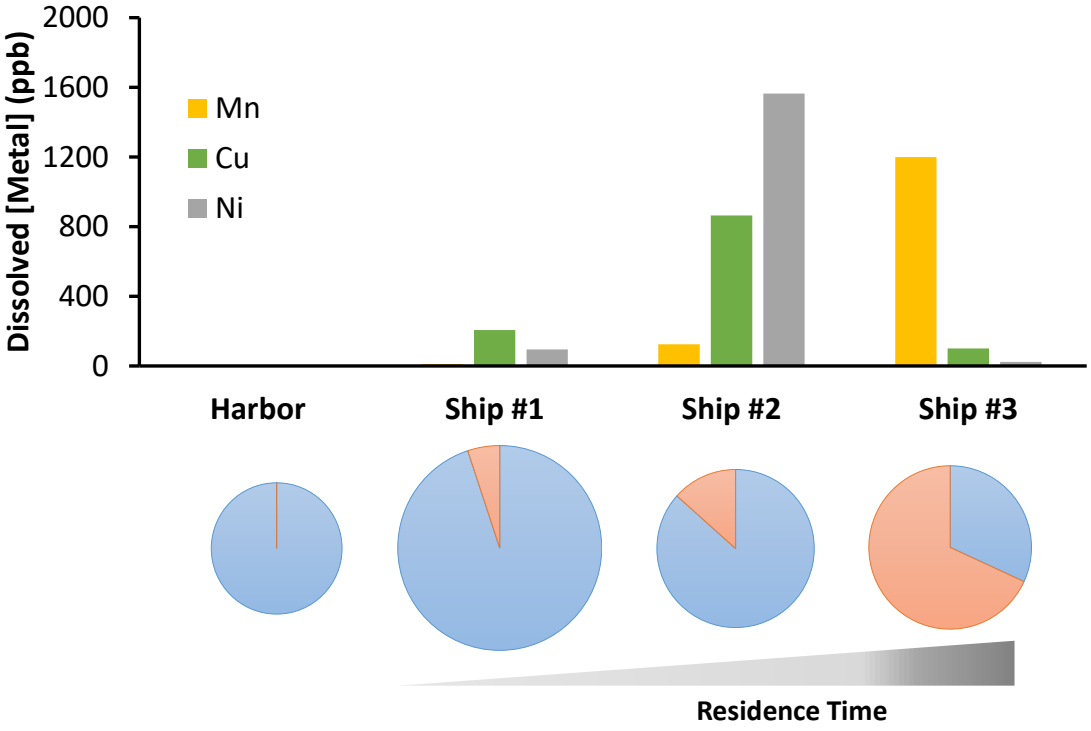


Figure 6. Graphical comparisons of hydrocarbonoclastic community proportions and ballast tank dissolved metal concentrations. Pie chart area is proportional to the total abundance of aerobic and anaerobic hydrocarbon activation genes normalized to *rpoB* for each library. Blue represents the aerobic, and orange represents anaerobic hydrocarbon activation gene subpopulations.



Chapter 3: Methanogenic Paraffin Degradation Proceeds via Alkane Addition to Fumarate by *"Smithella"* spp. Mediated by a Syntrophic Coupling with Hydrogenotrophic Methanogens

Abstract

Anaerobic microbial biodegradation of recalcitrant, water-insoluble substrates, such as paraffins, presents unique metabolic challenges. To elucidate this process, a methanogenic consortium capable of mineralizing long-chain *n*-paraffins (C_{28}-C_{50}) was enriched from San Diego Bay sediment. Analysis of 16S rRNA genes indicated the dominance of Syntrophobacterales (43%) and Methanomicrobiales (26%). Metagenomic sequencing allowed draft genome assembly of dominant uncultivated community members belonging to the bacterial genus *Smithella* and the archaeal genera *Methanoculleus* and *Methanosaeta*. Five contigs encoding homologs of the catalytic subunit of alkylsuccinate synthase (*assA*) were detected. Additionally, mRNA transcripts for these genes, including a homolog binned within the *"Smithella"* sp. SDB genome scaffold, were detected via RT-PCR, implying that paraffins are activated via 'fumarate addition'. Metabolic reconstruction and comparison with genome scaffolds of uncultivated *n*-alkane degrading *"Smithella"* spp. are consistent with the hypothesis that syntrophically growing *"Smithella"* spp. may achieve reverse electron transfer by coupling the reoxidation of ETF_{red} to a membrane-bound FeS oxidoreductase functioning as an ETF:menaquinone oxidoreductase. Subsequent electron transfer could proceed via a periplasmic formate dehydrogenase and/or hydrogenase, allowing energetic coupling to hydrogenotrophic methanogens such as *Methanoculleus*. Ultimately, these data provide fundamental insight into the energy conservation

mechanisms that dictate interspecies interactions salient to methanogenic alkane mineralization.

Introduction

Methanogenic conversion of *n*-alkanes by microbial consortia was first investigated in 1950 (Kuznetsov, 1950), but not definitively demonstrated until Zengler et al. (1999) described an enrichment culture that produced methane when amended with *n*-hexadecane as the sole carbon and energy source. The authors proposed that "microbial alkane cracking" under methanogenic conditions involves thermodynamic coupling of at least three microbial taxa, including syntrophic bacterial species that convert the alkane to hydrogen and acetate, acetoclastic methanogens that cleave the acetate to methane and carbon dioxide, and hydrogenotrophic methanogens that convert the carbon dioxide and hydrogen to methane. Phylogenetic analysis of the requisite culture via 16S rRNA gene clone libraries revealed members of the genera *Syntrophus*, *Desulfovibrio*, *Methanosaeta*, and *Methanospirillum* (Zengler et al., 1999), and subsequent analysis via Illumina sequencing verified the dominance of *Smithella* and *Methanosaeta* under hexadecane-grown conditions (Embree et al., 2014). More recently, metagenomic and metatranscriptomic sequencing of these hexadecane-grown cultures further revealed metabolically active consortium members including *Methanocalculus*, *Methanoculleus*, *Methanosaeta*, *Smithella* and *Desulfovibrio* (Embree et al., 2014; Embree et al., 2015). Since the seminal work in 1999, several subsequent studies have provided evidence that these taxa are important in methanogenic alkane biodegradation (Gieg et al., 2008; Siddique et al., 2011; Wang et al., 2011; Wang et al., 2012; Cheng et al., 2013; Tan et al., 2013; Liang et al., 2015).

The salient feature among these studies focusing on methanogenic alkane biodegradation (Zengler et al., 1999; Gray et al., 2011; Siddique et al., 2011; Tan et al., 2013) and other hydrocarbon-impacted systems (as reported in Gray et al. 2011) is the dominance of bacteria belonging to the Syntrophaceae. Despite this unifying observation, it has been a challenge to directly attribute *n*-alkane activation to *Syntrophus/Smithella* spp. Toward this end, some efforts have been directed toward elucidating the biodegradation pathway as a step toward identifying the key syntrophic players. Anaerobic *n*-alkane activation is thought to proceed via several mechanisms including alkane addition to fumarate (i.e. 'fumarate addition'), intra-aerobic hydroxylation, and possibly anaerobic hydroxylation followed by carboxylation of the alkane chain (for review see Callaghan 2013). Among these, 'fumarate addition' is the best characterized mechanism and is presumably catalyzed by the glycyl radical enzyme alkylsuccinate synthase (also known as methylalkylsuccinate synthase) (ASS/MAS) (Callaghan et al., 2008; Grundmann et al., 2008; Webner, 2012). The large catalytic subunit, encoded by *assA/masD*, has been recognized as a useful biomarker for anaerobic alkane metabolism (Callaghan et al., 2010), and has been targeted in several investigations of alkane-degrading, methanogenic enrichment cultures (Li et al., 2012; Mbadinga et al., 2012; Wang et al., 2012; Aitken et al., 2013; Cheng et al., 2013). Metagenomic and/or metatranscriptomic approaches have also been exploited (Tan et al., 2013; Tan et al., 2015), and more directly, single-cell genome sequencing and metatranscriptome analysis have been used to test the hypothesis that *"Smithella"* spp. are actively involved in alkane metabolism in the methanogenic alkane-degrading consortium originally reported by Zengler (Zengler et

al., 1999; Embree et al., 2014). Although Embree et al. reported that *Smithella* ME-1 cannot activate alkanes via 'fumarate addition' (Embree et al., 2014) and that the mode of activation in *"Smithella*" is unknown (Embree et al., 2015), re-analysis of their data by Tan et al. (Tan et al., 2014a) indicated that the requisite *"Smithella"* sp. ME-1 draft genome does indeed contain *assA* to which transcriptomic reads can be mapped when the culture is grown on hexadecane. Interestingly, 'fumarate addition' activity has also been proposed for a novel member of the Peptococcaceae (Phylum: Firmicutes) under methanogenic conditions (Tan et al., 2014c; Abu Laban et al., 2015). Overall, these data point to a greater diversity of bacteria capable of 'fumarate addition' reactions than has previously been appreciated.

The above studies have made significant contributions toward elucidating the microbial community structure of methanogenic alkane-degrading consortia, providing evidence of 'fumarate addition' and identifying *Smithella*-like organisms as critical members, but few studies in the literature have addressed long-chain paraffin degradation under methanogenic conditions (Townsend et al., 2003; Jones et al., 2008). The study herein is a metagenomic characterization of a methanogenic enrichment culture involved in the degradation of the long-chain paraffin, *n*-octacosane (C_{28}). Previous work demonstrated the presence of *assA* genes in this culture, but the abundance of the requisite organisms containing these genes and *assA* gene expression were not investigated (Callaghan et al., 2010). In the study herein, the detection of several *assA* genes and the corresponding mRNA transcripts are demonstrated, including an *assA* genotype associated with the binned and assembled genomic sequence for a *"Smithella"* sp., the dominant organism in the consortium.

Our data support the hypothesis that *"Smithella"* spp. are the dominant phylotypes that catalyze alkane degradation in the consortium described here. The data further suggest that prior observations for short and medium chain-length alkanes, with respect to the alkane activation mechanisms and the syntrophic relationships with hydrogenotrophic and acetoclastic methanogens, can be extended to the methanogenic conversion of water-insoluble long-chain paraffins.

Results

Culture Enrichment

A sediment-free consortium (hereafter referred to as SDB) was enriched from contaminated San Diego Bay sediment. The culture produces methane in the presence of long-chain alkanes including pentacosane ($C_{25}H_{52}$), octacosane ($C_{28}H_{58}$), dotriacontane ($C_{32}H_{66}$), tetracontane ($C_{40}H_{82}$), and pentacontane ($C_{50}H_{102}$). The highest rates of methane production (6.9 ± 0.2 mmol L^{-1} of methane after 31 weeks of incubation) were measured in cultures that were amended with octacosane ($C_{28}H_{58}$) (Appendix II Figure S1A). Lower, but significant, amounts of methane were also produced in cultures that were amended with shorter ($C_{10\text{-}18}$) or longer (C_{40} and C_{50}) chain-length *n*-alkanes (Appendix II Figure S1B). Given that the highest activities were observed with octacosane ($C_{28}H_{58}$), this substrate was used for the routine propagation and characterization of the SDB culture. Quantitative growth experiments were conducted with limiting amounts of octacosane to determine stoichiometry (Appendix II Table S1). Octacosane loss and methane production correlated with the theoretical stoichiometry based on the equation: $C_{28}H_{58} + 13.5H_2O \rightarrow 6.75CO_2 + 21.25CH_4$. Methane production was approximately 93 ± 19.7% of the predicted methane yield.

Community Composition

The phylogenetic structure of the SDB consortium was investigated by 16S PCR-based techniques and via metagenomic analysis of both 16S rRNA and functional genes. The 16S PCR-based analysis (Appendix II Figure S2) revealed a clear dominance of Deltaproteobacteria, accounting for 74% of bacterial reads. Four additional bacterial phyla accounted for more than 1% of the observed reads, including OP9 (8.2%), Deferribacteres (6.0%), Spirochaetes (4.1%), and Thermotogae (3.3%). Analysis of the archaeal 16S rRNA PCR products revealed dominant contributions of the Methanosarcinales (47%), Methanomicrobiales (36%), and Thermoplasmatales (17%). A similar picture emerged from the analysis of metagenomic sequence data with respect to both 16S rRNA genes (Figure 1A) and the phylogenetic affiliations of the closest database matches of functional gene-containing reads (Figure 1B). The deltaproteobacterial 16S rRNA genes were most closely related to species within the Syntrophobacterales and accounted for >43% of the detected 16S-containing reads. Specifically, 34% of the reads were affiliated with *"Smithella"* spp. Matches to OP9 accounted for 8.9% of reads, and several additional minor components of the metagenome were classified as gamma-proteobacterial lineages (2%), Spirochaetes (1.8%), Thermotogae (1.9%), Chloroflexi (3.8%), and Planctomyces (2.4%). Deferribacteres-like sequences accounted for only 0.2% of the 16S metagenomic reads. Among the archaeal sequences, Methanomicrobiales were dominant and accounted for 26% of the metagenomic 16S-containing reads. Methanosarcinales and Thermoplasmatales represented only 4.0% and 4.1%, respectively.

At the functional gene level (Figure 1B), sequence database search results were similarly dominated by deltaproteobacterial matches (35%), but also included a significant proportion of matches to other proteobacterial genes (alpha- 2%, beta- 2%, and gamma-proteobacteria 4%), genes from Firmicute lineages (10%), and genes from a wide range of other phyla. Functional genes most similar to archaeal genes belonged primarily to organisms classified within the Methanomicrobia, accounting for 15% of metagenomic reads. A significant portion of metagenomic reads (>11%) could not be assigned a phylogenetic affiliation.

Genomic binning

Tetranucleotide frequency based scaffolding yielded nineteen genome bins, four of which were associated with 16S rRNA genes (Appendix II Table S2). Ten of these scaffolds contained more than 1% of metagenomic reads, while the top five genome scaffolds accounted for more than 63% of all reads. Genome-to-genome comparisons via recruitment analysis in RAST (nmpdr.org) initially indicated that the most abundant organism in the SDB consortium has a genome most similar to the draft genome of *Syntrophus aciditrophicus* SB (data not shown). However, phylogenetic analysis of the full length 16S rRNA gene associated with this scaffold (hereafter referred to as *"Smithella"* sp. SDB) indicates a closer relationship to uncultivated *"Smithella"* spp. This is supported by a blastp comparison of *"Smithella"* sp. SDB's predicted proteins with those from *Syntrophus aciditrophicus* SB as well as requisite *"Smithella"* draft genomes (McInerney et al., 2007; Embree et al., 2014; Tan et al., 2014b) (Appendix II Table S3). *"Smithella"* sp. SDB shares 1,384 predicted proteins with *S. aciditrophicus* SB, but it shares 1,843 and 1,993 proteins with *"Smithella"* spp. ME-1 and SCADC,

respectively, which were also recovered from methanogenic, alkane-degrading consortia (Tan et al., 2013; Embree et al., 2014). Together, the *"Smithella"* spp. that contain *assA* form a monophyletic group that is distinct from *Smithella propionica* LYP, *"Smithella"* sp. F21, and other strains within the order Syntrophobacterales (Figure 2). The 16S rRNA gene associated with the *"Smithella"* sp. SDB scaffold is only 93.5% identical to *Smithella propionica*. Similarly, the 16S rRNA genes of *"Smithella"*-like strains implicated in methanogenic degradation of alkanes are only 94.2%, 94.4%, and 96.2% (for SCADC, D17, and F21 respectively; ME-1 does not contain a 16S gene) identical to *S. propionica*. MaxBin analysis also revealed that the second most abundant bacterial taxon is putatively associated with the *Desulfuromonas* reference genome of *D. acetoxidans* (hereafter referred to as *"Desulfuromonas"* sp. SDB) and accounts for 12% of metagenomic reads. *Desulfuromonas acetoxidans* is a freshwater deltaproteobacterial sulfate reducer that is known to utilize acetate, ethanol, and propanol as carbon and energy sources (Pfennig and Biebl, 1976).

Genome scaffolding further supported the interpretation that hydrogenotrophs within the Methanomicrobiales are the dominant methanogenic lineages in the SDB consortium (Appendix II Table S2). Two of the top six scaffolds exhibited greatest recruitment to the genome of *Methanosphaerula palustris* E1-9c and accounted for 13.3% of metagenomic reads. Phylogenetic analysis of the 16S rRNA genes associated with these scaffolds suggests that these organisms are mostly closely related to *Methanoculleus bourgensis* and *Methanolinea tarda* (94.6% and 94.3% identity, respectively; see Appendix II Figure S3). Two additional archaeal scaffolds were obtained, one of which affiliated with a 16S rRNA gene most closely related to the

acetoclastic methanogen *Methanosaeta concilii* (91.3% identity; hereafter referred to as *"Methanosaeta"* sp. SDB).

Hydrocarbon Biodegradation Genes

The SDB consortium was previously investigated for the presence of genes involved in anaerobic hydrocarbon degradation via PCR amplification of *assA* (i.e. the catalytic subunit of alkylsuccinate synthase) (Callaghan et al., 2010). In that study, five distinct *assA* genotypes were observed in the consortium, and both the 454- and Illumina-based metagenomic sequencing herein confirmed the presence of two of these *assA* genotypes (GenBank Numbers GU485661 and GU453662 for OTUs 1 and 4 reported in Callaghan et al. 2010) as well as three additional dominant *assA* OTUs that were not detected via PCR-based approaches (Figure 3A; Appendix II Table S4). Analysis of the 5 contigs containing *assA*-like genes indicated that at least two of them are associated with gene clusters encoding the four presumed subunits of alkylsuccinate synthase (i.e. alpha, beta, gamma, and *masE*-like subunits) as detected in previous studies (Grundmann et al., 2008; Callaghan et el., 2012) and implicated for '*Aromatoleum*' sp. HxN1 (Webner, 2012) (Figure 3B). Although, the gene encoding the putative alkylsuccinate synthase activase, *assD,* was not identified on the same contigs as *assA*, a homolog to *assD* was identified on another scaffold (scaffold 13542). Blastp analysis shows that the closest match is to *assD2* in *Desulfatibacillum alkenivorans* AK-01 (WP_015946970) (61% identity). Metagenomic analysis also included searches for other known genes involved in anaerobic hydrocarbon degradation including the subunits of benzylsuccinate synthase, naphthyl-2-methylsuccinate synthase, ethylbenzene dehydrogenase, phenylphosphate carboxylase, benzene carboxylase,

acetophenone carboxylase, and phenylphosphate synthase. However, the requisite signatures were not detected in either the 454- or Illumina-based data sets.

Given the high frequency of *assA*-like gene sequences in the SDB metagenome (~1:842 reads), the five genotypes detected in the metagenome were targeted for gene expression studies by designing specific PCR primers based on the 454-metagenomic data (Illumina-based data became available only later in the study). RT-PCR amplification of *assA*-mRNA was readily observed using multiple primer sets (Appendix II Table S5). Transcripts were detected for four out of the five *assA* genotypes in cultures amended with octacosane, in the presence and/or absence of the heptamethylnonane used as an inert carrier (Figure 3A). More importantly, *assA* transcripts from the SDB consortium were most similar to the *assA* genes previously ascribed to *"Smithella"* sp. SCADC (contigs 5, 17, 149, and 508) and *"Smithella"* sp. D17 (contig 15), which originated from methanogenic, alkane-degrading consortia (Tan et al., 2014b) (Figure 3A). Specifically, SDB contig 15 was associated with the *"Smithella"* sp. SDB scaffold, which contained a 16S rRNA gene most closely affiliated with *"Smithella"* sp. SCADC and D17 (Figure 2).

Comparative Genome Analysis

To gain insight into the salient functional potential of uncultured *"Smithella"* spp., a comparative genome analysis was conducted for the *"Smithella"* sp. SDB scaffold (Appendix II Table S3) and genome scaffolds previously reported in the literature (Appendix II Table S6). Single copy marker gene analysis indicates that the *"Smithella"* sp. SDB scaffold is near complete (>95%). In addition to the genes involved in alkane activation (i.e. the *ass* gene cluster) observed in *"Smithella"* spp.

SDB, ME-1, SCADC, and D17 (Figure 3B), these genomes contain the genetic machinery for beta-oxidation of the resulting fatty acids, but appear to lack the genes encoding the Wood-Ljungdahl pathway for the complete oxidation of acetyl-CoA. This is consistent with their inferred syntrophic lifestyle and an inability to mineralize alkanes via a respiratory pathway. Likewise, genes related to sulfur, sulfate and nitrate/nitrite respiratory activities are absent in all of the *"Smithella"* spp. genomes analyzed herein. Genome analysis further indicated that the genomes lack complete oxidative and reductive TCA cycles, but they all contain *re*-citrate synthase for 2-oxoglutarate synthesis. The latter is consistent with the genome of *Syntrophus aciditrophicus* SB (Kim et al., 2013). All of the draft genomes contain homologs for acetate-CoA synthetase (ADP-forming), which is used by many acetate-forming archaea to make ATP from acetyl-CoA and ADP (Bräsen et al., 2014). Several of the "*Smithella"* genomes contain acetate kinase (all but *"Smithella"* sp. F21), but only *"Smithella"* sp. SCADC has a gene encoding phosphotransacetylase.

A comparison with the genomes of *Desulfatibacillum alkenivorans* AK-01 (Callaghan et al., 2012) and *Syntrophorhabdus aromaticivorans* (Nobu et al., 2015) was also conducted to explore the presence of putative membrane complexes that are present in these hydrocarbon-degrading syntrophs. Pairwise comparisons indicate that *"Smithella"* spp. bins lack genes for the formation of DsrMKJOP, QrcABCD/MopABCD, Ohc, HmC/9Hc, Rnf, and FixC. However, comparisons suggested the presence of homologs to the electron transfer flavoprotein (Etf) complexes (Appendix II Table S6) observed in *Desulfatibacillum alkenivorans* AK-01 (Callaghan et al., 2012), *Syntrophus aciditrophicus* SB (McInerney et al., 2007),

Syntrophomonas wolfei (Sieber et al., 2010), and *Syntrophorhabdus aromaticivorans* UI (Nobu et al., 2015), in which a membrane FeS oxidoreductase is colocalised with the genes for Etf. In *"Smithella"* sp. SDB, the EtfB is encoded on a different contig than the EtfA subunit and the membrane FeS oxidoreductase, which may be an artifact of the assembly process. Additionally, a closer inspection of *"Smithella"* sp. SDB revealed genes encoding a periplasmic formate dehydrogenase, components for an F-type ATPase, and genes that contain hydrogenase-like motifs (Appendix II Table S6). Finally, RAST annotation did not indicate the presence of genes related to flagellum formation, chemotaxis, or quorum sensing (data not shown). However, *"Smithella"* spp. SDB, ME-1, and SCADC have genes related to the construction of a type IV pilus, which suggests that motility may be possible in these strains.

Discussion

The methanogenic mineralization of linear paraffins by microbial consortia derived from anoxic environments has been well documented. However, significant questions remain regarding the fate of high molecular weight alkanes in these systems. This study sought to characterize the roles of dominant community members in an enrichment culture capable of mineralizing long chain paraffins to methane and carbon dioxide. Several possible routes for the methanogenic biodegradation of alkanes have been evaluated by others, and it has been proposed that oxidation of hydrocarbons to acetate and hydrogen, coupled to syntrophic acetate oxidation and hydrogenotrophic methanogenesis, is the most likely process to occur in petroleum-laden methanogenic environments (Dolfing et al., 2008). With respect to alkanes, known mechanisms of anaerobic alkane degradation lead to the production of fatty acids, which are further

metabolized via β-oxidation (Callaghan, 2013). Thus, the energetic constraints of syntrophic fatty acid oxidation may also govern the syntrophic oxidation of *n*-alkanes. Non-equilibrium thermodynamic considerations indicate that the methanogenic oxidation of linear fatty acids is constrained by the entropic change between reactants and products and is energetically driven by an overall mass flux from the system (e.g. a microbial cell). This implies that anaerobic consortia will have a tendency to select for primary fermenting organisms that incompletely oxidize substrates to acetate rather than CO_2 in order to maximize the chemical efflux, and thus the free energy released per molecule of substrate metabolized (McInerney and Beaty, 1988). In order to drive the necessary chemical flux, partnered organisms are required to consume the acetate and hydrogen or formate to maintain a state of disequilibrium, constituting syntrophic cooperation.

Phylogenetic Analysis

Phylogenetic analysis of the SDB consortium (Figure 1) revealed a highly enriched community, predominantly populated by Syntrophobacterales (43%) in conjunction with both hydrogenotrophic (26%) and acetoclastic archaeal taxa (4%). These observations are consistent with the interpretation that a syntrophic couple is responsible for the mineralization of octacosane. Thus far, only two bacterial isolates (*Desulfatibacillum alkenivorans* AK-01 and *Desulfoglaeba alkanexedens* ALDC) have demonstrated syntrophic *n*-alkane degradation in co-culture with a methanogen (Callaghan et al., 2012; Lyles et al., 2014). The dominant phylotypes observed herein, as well as in previously described methanogenic alkane-degrading consortia, however, are most closely affiliated with yet-to-be cultivated lineages in the genus *Smithella*

(Zengler et al., 1999; Jones et al., 2008; Siddique et al., 2011; Tan et al., 2013). *Smithella* is represented by the sole isolate, *S. propionica*, a fatty acid-oxidizing bacterium shown to convert propionate to acetate, CO_2, and methane in syntrophic co-culture with *Methanospirillum hungateii*. *S. propionica* is also capable of axenic growth on crotonate (Liu et al., 1999). Physiologically, *S. propionica* is similar to the fatty acid-oxidizing *Syntrophobacter wolinii* and *Syntrophobacter pfennigii* (Liu et al., 1999), but it is phylogenetically most similar to *Syntrophus aciditrophicus* (Jackson et al., 1999). While each of these species has been shown to be capable of syntrophic metabolism, coupled with a hydrogenotrophic methanogen, neither *S. propionica* nor *S. aciditrophicus* utilize sulfate as a terminal electron acceptor.

The archaeal component of the SDB consortium is numerically dominated by hydrogenotrophic members of the Methanomicrobiales relative to the acetoclastic Methanosarcinales. The dominance of the hydrogenotrophs is further supported by the identification of two abundant draft genomes related to the genera *Methanoculleus* and *Methanolinea* (Appendix II Table S2), together accounting for 13.3% of the community coding potential. The genome bin affiliated with the acetoclastic genus *Methanosaeta* represents only 0.5% of the total metagenome. Similar ratios between hydrogenotrophic and acetoclastic methanogenic taxa have been described in several cultures degrading *n*-alkanes (Gieg et al., 2008; Jones et al., 2008; Gray et al., 2011). *Methanoculleus* spp. and *Methanolinea* spp. exclusively produce methane from the reduction of CO_2 and/or utilization of formate (Romesser et al., 1979; Maestrojuán et al., 1990; Imachi et al., 2008). In contrast, members of the genus *Methanosaeta* utilize acetate as the sole methanogenic substrate and have not been shown to oxidize

hydrogen or formate as a reductant (Patel and Sprott, 1990). Additionally, *Methanosaeta* spp. activate acetate to acetyl-CoA via an AMP-forming acetyl-CoA synthetase resulting in an acetate threshold concentration approximately ten-fold lower than *Methanosarcina* spp. that utilize the acetate kinase and phosphotransacetylase system (Jetten et al., 1992; Smith and Ingram-Smith, 2007). This creates a kinetic niche separation based on acetate affinities that explains the commonly observed predominance of *Methanosaeta* over the more metabolically diverse *Methanosarcina* spp. in hydrocarbon-degrading, methanogenic environments (Penner and Foght, 2010; Siddique et al., 2011; Siddique et al., 2012).

Transcription Experiments and Comparative Genome Analysis

Given the repeated enrichment of *"Smithella"* spp. from varying environments under methanogenic conditions and the lack of genetic evidence for respiratory capacity, it is likely that these organisms are obligately fermentative with respect to the oxidation of *n*-alkanes. Transcriptional evidence of *assA* genotypes obtained in this study, along with the findings of the re-analysis of the *"Smithella"* sp. ME-1 metatranscriptome (Tan et al., 2014a), suggest that *"Smithella"* spp. activate paraffins by alkane addition to fumarate using alkylsuccinate synthase (ASS) and further metabolize the intermediates to acetyl-CoA by β-oxidation. Thus, *"Smithella"* spp. likely employ similar physiological strategies for the fermentation of *n*-alkanes to those proposed for the saturated fatty acid-oxidizing members of the family Syntrophomonadaceae represented by *Syntrophomonas wolfei* (McInerney et al., 1981; Sousa et al., 2007), *Syntrophus aciditrophicus* (Jackson et al., 1999), and the alkane-oxidizing syntroph *Desulfatibacillum alkenivorans* AK-01 (Callaghan et al., 2012).

Reverse Electron Transfer

Each of the above model organisms requires a mechanism for reverse electron transfer to overcome the thermodynamic barrier of coupling the oxidation of acyl-CoA intermediates to their corresponding enoyl-CoAs (E' of approximately -10 mV) (Sato et al., 1999) to the production of hydrogen (E' of -261 mV) or formate (E' of -258 mV) (McInerney et al., 2009). Comparative analysis of binned *"Smithella"* spp. genomes (Appendix II Table S6) suggests that these organisms link electrons derived from the acyl-CoA dehydrogenase to the production of hydrogen/formate. This might be achieved by coupling the reoxidation of ETF_{red} by a membrane-bound FeS oxidoreductase functioning as an ETF:menaquinone oxidoreductase and transferring the electrons to a periplasmic formate dehydrogenase and/or hydrogenase through a reverse menaquinone loop (Figure 4). This has previously been proposed for *Syntrophus aciditrophicus* (McInerney et al, 2007) and *Syntrophomonas wolfei* (Sieber et al., 2010) and is supported by preliminary proteomic studies that revealed FeS oxidoreductase and EtfAB proteins to be highly abundant in syntrophically grown cultures of *Syntrophomonas wolfei* (Schmidt et al., 2013; Sieber et al., 2015). The draft genome of *"Smithella"* sp. SDB encodes an FeS oxidoreductase and EtfAB proteins (Appendix II Table S6) with high amino acid sequence similarity to homologs in *Syntrophus aciditrophicus* (SYN_02636-02638) and *Syntrophomonas wolfei* (Swol_0696-0698) (McInerney et al., 2007; Sieber et al., 2015). A closer inspection of the EtfAB complex in *"Smithella"* sp. SDB did not reveal the proposed signature motif for electron bifurcation that has been identified in other bacterial species (Chowdhury et al., 2015). With regard to *Syntrophomonas wolfei*, proteins of the homologous Fix System were also detected, but at significantly lower levels, leading to the hypothesis that the FeS

oxidoreductase system serves as the primary conduit for electrons derived from the acyl-CoA dehydrogenase (Sieber et al., 2015). The Fix System ETF:quinone oxidoreductase (FixC) was not detected in the *"Smithella"* sp. SDB genome, further supporting the notion that *"Smithella"* spp. may similarly utilize the FeS oxidoreductase system as the primary Etf-linked reverse electron transfer mechanism.

Several potential mechanisms for the reoxidation of NADH are suggested by the draft genome of *"Smithella"* sp. SDB. Although essential components of the Rnf complex were not observed, genes with homology to the subunits of the membrane-bound Ion-translocating Ferredoxin:NADH oxidoreductase (IFO) complex were detected. This complex was initially described for the genome of *Syntrophorhabdus aromaticivorans*, and it has been proposed that IFO acts analogously to the Rnf complex (also absent in *S. aromaticivorans*) by coupling the oxidation of NADH to the reduction of ferredoxin via proton motive force (Nobu et al., 2015). A potential soluble NADH-dependent formate dehydrogenase and putative hydrogenase subunits were also detected in *"Smithella"* sp. SDB (Appendix II Table S6), which could allow cytoplasmic electron confurcation from reduced ferredoxin and NADH to formate or hydrogen. These putative confurcating systems have been annotated in the genomes of several, known syntrophic bacteria across multiple phyla, though it should be noted that there is currently debate as to whether a confurcating mechanism involving ferredoxin would be required, or whether these cytoplasmic hydrogenases and formate dehydrogenases form a soluble NADH-dependent complex in *Syntrophomonas wolfei* (Schmidt et al., 2013).

ATP Biosynthesis

The predominant route for ATP synthesis in fermentative bacteria is typically through substrate-level phosphorylation. During the fermentation of paraffins, *"Smithella"* spp. could generate ATP in the conversion of acetyl-CoA intermediates to acetate. "*Smithella"* sp. SCADC could potentially use the traditional route through acetyl-P to acetate for substrate-level phosphorylation. However, the other "*Smithella"* draft genomes either lack acetate kinase or phosphotransacetylase, suggesting that the two-step process involving acetyl-P may not be employed in those organisms. In all draft genomes, genes for ADP-forming acetyl-CoA synthases were identified. These enzymes catalyze the single-step, concerted activation of acetate to acetyl-CoA and are almost ubiquitously distributed across all domains of life (Lindahl and Chang, 2001; Starai and Escalante-Semerena, 2004).

Generation of Proton Motive Force

Although ATP is likely generated via substrate level phosphorylation, a proton gradient is still essential for the production of hydrogen and/or formate by syntrophic fermenters. Proton motive force might be generated by *"Smithella"* sp. SDB through the reversal of an ATP synthase, consuming ATP generated during substrate-level phosphorylation and thereby pumping protons into the periplasmic space. Additionally, the symport of protons with acetate across the cytoplasmic membrane could produce energy through a chemical gradient that is independent of ATP hydrolysis (Michels et al., 1979). The production of a chemical gradient based on proton motive force has been demonstrated through the simultaneous efflux of protons and lactate in studies of *Streptococcus cremoris* (Otto et al., 1980) and *E. coli* membrane vesicles (Ten Brink and Konings, 1980). It has been proposed that *Syntrophomonas wolfei* may conserve

energy through both substrate level phosphorylation and acetate excretion during syntrophic butyrate oxidation, thereby increasing the overall energy conservation efficiency in such an energetically limited lifestyle (McInerney and Beaty, 1988).

Altogether, the comparative genome analysis indicates that hydrocarbon-utilizing *"Smithell*a*"* spp. likely have several physiological properties common to well-characterized, linear fatty acid-oxidizing syntrophs that enable these organisms to overcome the thermodynamic barriers associated with the fermentation of *n*-paraffins. Specifically, the proposed mechanisms for coupling the acyl-CoA dehydrogenase and NADH reoxidation to the production of hydrogen and/or formate, as well as energy conservation strategies, suggest that these organisms are physiologically very similar to *Syntrophus aciditrophicus*. These similarities likely reflect an evolutionary trajectory based on substrate utilization profiles that ultimately select for a shared metabolic framework.

Proposed Syntrophy Model

The strong co-enrichment of hydrogenotrophic methanogens with the proposed alkane-oxidizing *"Smithella"* sp. SDB implies a syntrophic pairing with *"Methanoculleus"* sp. SDB and/or *"Methanolinea"* sp. SDB through interspecies electron transfer via hydrogen or formate. *"Methanosaeta"* sp. SDB potentially consumes acetate as the primary fermentation product of *"Smithella"* sp. SDB. This could form the basis of syntrophic interactions, but its inability to utilize hydrogen or formate as an electron donor and its relatively lower abundance suggest that *"Methanosaeta"* sp. SDB is likely relegated to a minor role in this process. Additionally, given the methanogenic conditions and the relatively high abundance of

Desulfuromonas sp. SDB (12% of metagenome reads), it is unlikely that *Desulfuromonas* persists in the SDB consortium by reducing small amounts of sulfate present in the medium (<1 mM), but rather by converting the acetate produced by *"Smithella"* sp. SDB to H_2 and CO_2, which are subsequently utilized by hydrogenotrophic methanogens. It is possible that the high abundance of *Desulfuromonas* relative to *"Methanosaeta"* sp. SDB might be governed by acetate uptake kinetic parameters. Although *Desulfuromonas* has not previously been shown to be capable of syntrophic acetate oxidation ($\Delta G'_0 = +94.9$ kJ·mol^{-1}; Table 1), the proposed scenario provides another route for acetate utilization.

Collectively, these observations regarding the anaerobic biodegradation of high molecular weight paraffins are consistent with the multi-organism hypothesis, initially proposed by Zengler et al. (Zengler et al., 1999) for methanogenic hexadecane mineralization. Per Zengler's model, a primary, activating organism ferments the alkane to acetate and H_2 in syntrophic association with a hydrogenotrophic methanogen. The acetate produced is ultimately consumed by an acetotrophic organism(s) through syntrophic acetate oxidation and/or acetoclastic methanogenesis (Zengler et al., 1999).

Conclusion

Phylogenetic, metagenomic, and transcriptional data provide compelling evidence for the interpretation that the dominant contributor to anaerobic paraffin activation in the SDB consortium is a bacterial strain most closely related to *Smithella propionica* and that this organism catalyzes this reaction via alkylsuccinate synthase. Based on comparative genome analysis, *"Smithella"* sp. SDB is similar to *Syntrophus aciditrophicus* with respect to its central metabolic machinery and energy conservation

strategies. To date, alkane-utilizing *"Smithella"*-like syntrophs that contain *assA* in their draft genomes form a monophyletic clade distinct from *S. propionica*, suggesting that these organisms likely represent a novel, uncultivated genus within the order Syntrophobacterales. The cultivation and characterization of a requisite type strain would therefore be of great value and could provide a model for understanding the genetics, biochemistry, and ecology underlying methanogenic mineralization of alkane substrates.

Materials and Methods

Sediment Collection

Samples were collected from the sediments of Paletta Creek in San Diego Bay with technical assistance from the Office of Naval Research. Paletta Creek is a site contaminated with polychlorinated biphenyls (PCBs), polycyclic aromatic hydrocarbons (PAHs), and chlordane (California Environmental Protection Agency, Last Accessed: March 25th, 2015). Grab samples were collected and placed in Mason jars, which were filled to capacity to exclude a head space. Samples were stored at 4˚C and shipped to the lab, where paraffin enrichments were conducted.

Establishment of Methanogenic Enrichment Culture

All sediment manipulations were performed in an anaerobic glove box, and strict anaerobic technique was used for all culture manipulations and media/substrate preparations (Hungate, 1969; Balch and Wolfe, 1976). Enrichments were established in sterile serum bottles (160 ml) with 50 ± 0.5 g of sediment as inocula and 75 ml of reduced, sulfate-free, bicarbonate-buffered, seawater mineral medium (Widdel and Bak, 1992). Primary enrichments were amended with equimolar amounts of several *n*-

alkanes, including decane ($C_{10}H_{22}$), dodecane ($C_{12}H_{26}$), hexadecane ($C_{16}H_{34}$), and octadecane ($C_{18}H_{38}$) (80 µmol total hydrocarbon). Bottles were incubated under an N_2:CO_2 (80:20) headspace at 31ºC. Methane production was used as a proxy for alkane utilization to monitor cultures (see below). Incubations that produced significantly more methane, relative to substrate unamended (i.e. no substrate) and sterile (i.e. autoclaved) controls, were subcultured and amended with 20 mg of octacosane ($C_{28}H_{58}$), 5 mg of tetracontane ($C_{40}H_{82}$), or 5 mg of pentacontane ($C_{50}H_{102}$) as substrates. Substrates were delivered in a 1 ml overlay of sterile 2,2,4,4,6,8,8-heptamethylnonane (HMN). Repeated transfers were conducted for >3 years to obtain a sediment-free culture capable of long-chain alkane degradation.

Methane Measurements

Methane was quantified as previously described (Wawrik et al., 2012b) by analyzing 0.2 ml of the enrichment bottle headspace using a Varian 3300 GC equipped with a flame ionization detector (FID) and packed stainless steel column (Poropak Q, 80/100; Supelco, Bellafont, PA).

Octacosane Quantification

For quantitative experiments, the enrichment cultures were incubated in 25-mL serum bottles containing 12 mL of medium overlain with 2 ml of HMN containing 2 mg of $C_{28}H_{58}$. Sterile (i.e. autoclaved cultures) and uninoculated controls were included. The octacosane concentration was measured by sampling the HMN overlay. Small aliquots (1 µL) were injected directly into an Agilent Technologies 6890 gas chromatograph (GC) equipped with a DB-5ms capillary column (30 m x 0.25 mm, 1 µm film) coupled with an Agilent 5973 mass spectrometer (MS). The oven temperature

was increased from 45°C to 250°C (at a rate of 10^{o} min^{-1}), held for 10 minutes, and then increased to 290°C (at a rate of 10^{o} min^{-1}). The injector temperature was 250°C. Helium was used as a carrier gas at a flow rate of 1.2 mL min^{-1}.

Microbial Community Characterization

For DNA extraction, subsamples of the SDB consortium growing on octacosane were centrifuged at 14,000 x g for 15 min, and cell pellets were extracted as previously described (Callaghan et al., 2010) via a method adapted from Rainey et al. (Rainey et al., 1996). DNA was quantified using a Qubit 2.0 fluorometer and the Quant-iT dsDNA BR Assay Kit (Life Technologies, Carlsbad, CA). Community profiling was conducted by PCR amplification and multiplexed 454-sequencing of 16S rRNA genes. The 16S rRNA gene PCR was performed using the forward primer 27F (5'-AGAGTTTGATCMTGGCTCAG-3') and the reverse primer 338R (5'-TGCTGCCTCCCGTAGGAGT-3') for bacteria, and primers A8F (5'-TCCGGTTGATCCTGCC-3') and A344R (5'-TCGCGCCTGCTGCICCCCGT-3') for archaea (Nakatsu & Marsh, 2007). All primers included 5' Titanium Fusion adapter sequences (forward primer A-tag: CCATCTCATCCCTGCGTGTCTCCGACTCAG; reverse primer B-tag: CCTATCCCCTGTGTGCCTTGGCAGTCTCAG) as well as a unique 8-nucleotide barcode tag in the reverse primer as previously described (Hamady *et al.*, 2008) to allow direct 454-sequencing. Reaction chemistry included 200 nM of each primer in PCR Supermix (Life Technologies, Carlsbad, CA) and 2 μL of template DNA. Thermal cycling was performed as follows: 95°C for 7 min. and 30 cycles of 95°C for 20 sec., 55°C for 20 sec., and 72°C for 40 sec. For amplification of archaeal 16S rRNA genes, the extension step was extended to 60 seconds. The PCR products

were cleaned, DNA was included in a multiplexed 454-sequencing run, and the respective sequences were computationally binned into bacterial and archaeal 16S sequence libraries (Johnson et al., 2015). The resulting reads were quality screened by removing any reads with average Q-scores < 25 and trimming to remove ends with poor quality (Q<25). Sequences were clustered and classified using the QIIME pipeline after de-multiplexing and clustering into Operational Taxonomic Units (OTUs) via UCLUST (Edgar, 2010) at the 97% identity level. From the resulting OTUs, a representative set of sequences was aligned to the SILVA small subunit rRNA reference alignment (www.arb-silva.de) using the PyNAST algorithm (Caporaso et al., 2010) to allow classification.

Metagenomic Analysis

Metagenomic sequencing proceeded in two phases. Initially, genomic DNA was shotgun sequenced by means of the Roche 454 GS FLX Titanium platform. The resulting data (202,122 high quality reads longer than 100bp) aided primarily in the design of *assA* PCR primers for RT-PCR experiments (see below). Subsequently, a full lane of Illumina MiSeq PE 250 was used to generate a metagenome of greater depth. All metagenome analyses described below reflect the larger, Illumina-based sequencing run. Sequence adapters and polyA-tail artifacts were removed from reads using Cutadapt (Martin, 2011) and HomerTools (Heinz et al., 2010). Paired end reads were joined by TRIMMOMATIC (Bolger et al., 2014), and all unpaired reads were discarded. Reads were assembled using Meta-Ray (Boisvert et al., 2010; Boisvert et al., 2012) using a Kmer setting of 31 and discarding all contigs <1kb. Open reading frames (ORFs) were predicted for all contigs using Prodigal (Hyatt et al., 2012), and the

resulting protein-coding sequences were annotated via database comparisons to NCBI Refseq (Pruitt et al., 2007), COG (Tatusov et al., 2000), PFAM (Finn et al., 2014), M5NR (Wilke et al., 2012), and KEGG (Kanehisa and Goto, 2000; Kanehisa et al., 2012). In addition, all ORFs were compared to a database of protein sequences encoding enzymes/subunits involved in anaerobic hydrocarbon degradation including alkylsuccinate synthase, benzylsuccinate synthase, naphthyl-methylsuccinate synthase, ethylbenzene dehydrogenase, naphthalene carboxylase, acetophenone carboxylase, phenylphosphate synthase and carboxylase, and anaerobic benzene carboxylase. Database searches were performed using USEARCH (Edgar, 2010), DIAMOND (Buchfink et al., 2015), and HMMER3 (Finn et al., 2011). The abundances of individual ORFs in the metagenome were determined by mapping unassembled, unpaired reads with the short-read aligner Bowtie2 (Song et al., 2014). Phylogeny of predicted ORFs was assessed by parsing blast outputs via the Lowest Common Ancestor algorithm in MEGAN5. Tetranucleotide frequency based analysis was used to bin assembled DNA sequence data into genome scaffolds using MaxBin (Wu et al., 2014). Given that assemblies of metagenomic Illumina data frequently yield contigs where 16S rRNA genes are underrepresented, the 16S rRNA gene-containing reads were then manually assembled into each genome scaffold via SeqMan (DNASTAR). Scaffold completeness was estimated by comparison with a database of single copy housekeeping genes (Wu et al., 2013) via HMMER search (Finn et al., 2011). Recruitment analysis was conducted in RAST (rast.nmpdr.org), where draft genomes for *"Smithella"* spp. were manually added to the RAST system, or by pairwise blastp analysis using Diamond (Buchfink et al., 2015).

assA Gene Detection and RT-PCR

Genes encoding the catalytic subunit of alkylsuccinate synthase (*assA*) were previously reported for the SDB consortium via sequencing of cloned PCR products (Callaghan et al., 2010). Subsequent 454-based metagenome analysis revealed 248 reads with significant homology to *assA*, which were assembled with the prior PCR-derived reads (Appendix II Table S4). PCR primers for each OTU were then designed with the criterion that at least two mismatches to each of the untargeted OTUs should be present (Appendix II Table S5). Given the lack of overlap among 454-based metagenomic OTUs, the mismatch criterion could not be achieved for contigs 347 and 348. A primer pair was therefore designed to target these two contigs that also included perfect sequence identities to contigs 345 and 350. Annealing temperature gradient PCR was performed for each primer pair. The PCR reactions included 5 ng of SDB community DNA as a template and 1 μM of respective primers in PCR Supermix (Life Technologies, Carlsbad, CA). Thermal cycling conditions were: 40 cycles of 95°C for 30 s, 55-65°C for 60 s, and 72°C for 75 sec. The highest temperatures where no appreciable loss of amplification was observed in agarose gels were chosen as annealing temperatures for RT-PCR experiments. The RNA was extracted as previously described (Wawrik et al., 2012a). In brief, 10-mL samples of SDB culture were filtered onto 0.45 μm Supor filters (Pall). Filters were stored in 2-mL screw cap tubes containing 750 μL of RLT buffer (Qiagen) and ca. 50 mg of 0.1 mm muffled glass beads after flash freezing in liquid nitrogen at -80°C. Samples were later thawed, 7.5 μL of β-mercaptoethanol was added, and tubes were agitated for two minutes using a bead-beater. Supernatants were extracted using a Qiagen RNEasy kit (Qiagen, Valencia, CA) as recommended by the manufacturer. DNA contamination was

removed by digestion with 0.1 U μL^{-1} of RQ1 DNAse (Promega, Madison, WI) at 37°C for 30 minutes. PCR for 16S rRNA genes was conducted as described above to confirm complete removal of DNA contamination from RNA samples. After DNAse inactivation (10 min at 65°C), RNA was reverse transcribed using SuperScript®III (Life Technologies, Carlsbad, CA) reverse transcriptase and random hexamers. The RT-PCR reactions were performed by applying optimized PCR conditions derived from annealing temperature gradient experiments, and positive amplification was assessed by visualizing bands of the correct size via gel electrophoresis. All positive RT-PCR reactions were subsequently cleaned using a Qiagen PCR purification kit (Qiagen, Valencia, CA) and cloned using a TOPO®TA-Cloning Kit for Sequencing (Life Technologies, Carlsbad, CA) as recommended by the manufacturer. Colonies were screened for inserts via PCR using the M13 priming sites on the cloning vector, and twelve PCR products of the correct size were Sanger sequenced for each positive RT-PCR reaction to confirm that the detected mRNA corresponded to targeted *assA* OTUs.

Phylogenetic Analysis

Phylogenetic analysis of 16S rRNA gene sequences was conducted on ClustalW alignments of requisite 16S rRNA genes via neighbor-joining tree analysis using the Tajima-Nei distance method, 5,000 bootstrap replicates, and pairwise deletion using MEGA6 (Tamura et al., 2013). For protein-coding genes (e.g. the catalytic subunit genes for alkylsuccinate synthase, *assA*), predicted amino acid sequences were aligned with a representative set of sequences from the database using ClustalW, and alignments were manually curated. Neighbor-joining trees were generated using

MEGA6 (Tamura et al., 2013) using the Poisson model distance method and conducting 5,000 bootstrap replicates with pairwise deletion.

Comparative "Smithella" Genome Analysis

The dominant, binned *"Smithella"* genome (*Smithella* sp. SDB) from the SDB culture, as well as *"Smithella"*-like genome bins D17, SCADC, ME-1, and F21 from previous reports of methanogenic alkane-degrading cultures (Embree et al., 2014; Tan et al., 2014b), were annotated via RAST (Aziz et al., 2008), and subsystem analysis was used to investigate the presence of metabolic pathways related to alkane metabolism and energy conservation (Appendix II Table S6). To determine the presence of alkylsuccinate synthase-associated genes, the requisite genes in *ass* gene cluster 1 from *Desulfatibacillum alkenivorans* AK-01 (Callaghan et al., 2012) were manually blastp searched against the predicted proteins in each of the putative *"Smithella"* genome bins. A match was considered to be positive at an E-score of <1E-10, an AA identity of > 25%, and if the blast alignment covered the length of the entire protein that was queried. Additionally, a pairwise genome comparison was conducted between *"Smithella"* sp. SDB and all genes found in *Desulfatibacillum alkenivorans* AK-01 and *Syntrophorhabdus aromaticivorans* strain UI (Nobu et al., 2015) via the blastp option of Diamond (Buchfink et al., 2015) to identify homologs to membrane complexes described in their respective genome descriptions (Callaghan et al., 2012; Nobu et al., 2015). All predicted proteins from *"Smithella"*-like genome bins SDB, D17, SCADC, ME-1, and F21were also searched against the KOBAS database via the blastp option of Diamond (Xie et al., 2011) to determine KEGG orthology, and metabolic pathways were reconstructed using KEGG Mapper

(http://www.genome.jp/kegg/tool/map_pathway.html). The KEGG maps were primarily used to confirm observations for RAST annotation and to identify ABC transporters. Signal peptides and transmembrane helices were predicted using SignalP (Petersen et al., 2011) and TMHMM (http://www.cbs.dtu.dk/services/TMHMM/) tools respectively. Twin arginine sites were predicted via TatP (Bendtsen et al., 2005).

In order to identify genes shared among all *"Smithella"*-like genome bins, pairwise searches were conducted among all predicted proteins in genome bins SDB, D17, SCADC, ME-1, and F21 via the blastp option of Diamond (Appendix II Table S3). Results were imported into a SQLite database in order to identify proteins in *"Smithella"* sp. SDB, which had homologs in D17, SCADC, ME-1, and F21 (Appendix II Table S7). Since *"Smithella"* sp. F21 is not known to contain known genes related to alkane activation, the same analysis was also conducted by excluding this genome bin (Appendix II Table S8). Draft genome completeness was estimated as previously described (Wu et al., 2013) by blast searching 124 single copy marker genes shared by all bacteria against all predicted proteins in each of the five *"Smithella"* genome bins.

References

Abu Laban, N., Dao, A., Semple, K., and Foght, J. (2015) Biodegradation of C_7 and C_8 *iso*-alkanes under methanogenic conditions. *Environ Microbiol* **17**: 4898-4915

Aitken, C.M., Jones, D.M., Maguire, M.J., Gray, N.D., Sherry, A., Bowler, B.F.J. et al. (2013) Evidence that crude oil alkane activation proceeds by different mechanisms under sulfate-reducing and methanogenic conditions. *Geochim Cosmochim Ac* **109**: 162-174.

Aziz, R.K., Bartels, D., Best, A.A., DeJongh, M., Disz, T., Edwards, R.A. et al. (2008) The RAST server: Rapid annotations using subsystems technology. *BMC Genomics* **9**: 75.

Balch, W.E., and Wolfe, R.S. (1976) New approach to the cultivation of methanogenic bacteria: 2-mercaptoethanesulfonic acid (HS-CoM)-dependent growth of *Methanobacterium ruminantium* in a pressurized atmosphere. *Appl Environ Microbiol* **32**: 781-791.

Bendtsen, J.D., Nielsen, H., Widdick, D., Palmer, T., and Brunak, S. (2005) Prediction of twin-arginine signal peptides. *BMC Bioinformatics* **6**: 167.

Boisvert, S., Laviolette, F., and Corbeil, J. (2010) Ray: Simultaneous assembly of reads from a mix of high-throughput sequencing technologies. *J Comput Biol* **17**: 1519-1533.

Boisvert, S., Raymond, F., Godzaridis, E., Laviolette, F., and Corbeil, J. (2012) Ray Meta: scalable *de novo* metagenome assembly and profiling. *Genome Biol* **13**: R122.

Bolger, A.M., Lohse, M., and Usadel, B. (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**: 2114-2120.

Bräsen, C., Esser, D., Rauch, B., and Siebers, B. (2014) Carbohydrate metabolism in Archaea: Current insights into unusual enzymes and pathways and their regulation. *Microbiol Mol Biol Rev* **78**: 89–175.

Buchfink, B., Xie, C., and Huson, D.H. (2015) Fast and sensitive protein alignment using DIAMOND. *Nature Meth* **12**: 59-60.

California Environmental Protection Agency (Last Accessed: March 25th, 2015). San Diego Region – Sediment Toxicity TMDLs for the Mouths of Paleta, Chollas, and Switzer Creeks URL http://www.waterboards.ca.gov/sandiego/water_issues/programs/tmdls/sediment_toxicity.shtml

Callaghan, A.V. (2013) Enzymes involved in the anaerobic oxidation of *n*-alkanes: from methane to long-chain paraffins. *Front Microbiol* **4**: Article 89.

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

Callaghan, A.V., Davidova, I.A., Savage-Ashlock, K., Parisi, V.A., Gieg, L.M., Suflita, J.M. et al. (2010) Diversity of benzyl- and alkylsuccinate synthase genes in hydrocarbon-impacted environments and enrichment cultures. *Environ Sci Technol* **44**: 7287-7294.

Callaghan, A.V., Morris, B.E.L., Pereira, I.A.C., McInerney, M.J., Austin, R.N., Groves, J.T. et al. (2012) The genome sequence of *Desulfatibacillum alkenivorans* AK-01: A blueprint for anaerobic alkane oxidation. *Environ Microbiol* **14**: 101-113.

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

Cheng, L., Rui, J., Li, Q., Zhang, H., and Lu, Y. (2013) Enrichment and dynamics of novel syntrophs in a methanogenic hexadecane-degrading culture from a Chinese oilfield. *FEMS Microbiol Ecol* **83**: 757-766.

Chowdhury, N.P., Kahnt, J., and Buckel, W. (2015) Reduction of ferredoxin or oxygen by flavin-based electron bifurcation in *Megashaera elsdenii*. *The FEBS Journal* **282**: 3149-3160.

Dolfing, J., Larter, S.R., and Head, I.M. (2008) Thermodynamic constraints on methanogenic crude oil biodegradation. *ISME J* **2**: 442-452.

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

Embree, M., Liu, J.K., Al-Bassam, M.M., and Zengler, K. (2015) Networks of energetic and metabolic interactions define dynamics in microbial communities. *Proc Natl Acad Sci* **112**: 15450-15455.

Embree, M., Nagarajan, H., Movahedi, N., Chitsaz, H., and Zengler, K. (2014) Single-cell genome and metatranscriptome sequencing reveal metabolic interactions of an alkane-degrading methanogenic community. *ISME J* **8**: 757-767.

Finn, R.D., Clements, J., and Eddy, S.R. (2011) HMMER web server: Interactive sequence similarity searching. *Nucleic Acids Res* **39**: W29–W37.

Finn, R.D., Bateman, A., Clements, J., Coggill, P., Eberhardt, R.Y., Eddy, S.R. et al. (2014) Pfam: The protein families database. *Nucleic Acids Res* **42**: D222-D230.

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.

Gray, N.D., Sherry, A., Grant, R.J., Rowan, A.K., Hubert, C.R.J., Callbeck, C.M. et al. (2011) The quantitative significance of Syntrophaceae and syntrophic partnerships in methanogenic degradation of crude oil alkanes. *Environ Microbiol* **13**: 2957-2975.

Grundmann, O., Behrends, A., Rabus, R., Amann, J., Halder, T., Heider, J., and Widdel, F. (2008) Genes encoding the candidate enzyme for anaerobic activation of *n*-alkanes in the denitrifying bacterium, strain HxN1. *Environ Microbiol* **10**: 376-385.

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. *Nature Methods* **5**: 235-237.

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P. et al. (2010) Simple combinations of lineage-determining transcription factors prime *cis*-regulatory elements required for macrophage and B cell identities. *Mol Cell* **38**: 576-589.

Hungate, R.E. (1969) A roll tube method for cultivation of strict anaerobes. In *Methods in Microbiology*. Norris, J.R., and Ribbons, D.W. (eds). London: Academic Press Inc., pp. 117-132.

Hyatt, D., LoCascio, P.F., Hauser, L.J., and Uberbacher, E.C. (2012) Gene and translation initiation site prediction in metagenomic sequences. *Bioinformatics* **28**: 2223-2230.

Imachi, H., Sakai, S., Sekiguchi, Y., Hanada, S., Kamagata, Y., Ohashi, A., and Harada, H. (2008) *Methanolinea tarda* gen. nov., sp. nov., a methane-producing archaeon isolated from a methanogenic digester sludge. *Int J Syst Evol Bact* **58**: 294-301.

Jackson, B.E., Bhupathiraju, V.K., Tanner, R.S., Woese, C.R., and McInerney, M.J. (1999) *Syntrophus aciditrophicus* sp. nov., a new anaerobic bacterium that degrades fatty acids and benzoate in syntrophic association with hydrogen-using microorganisms. *Arch Microbiol* **171**: 107-114.

Jetten, M.S.M., Stams, A.J.M., and Zehnder, A.J.B. (1992) Methanogenesis from acetate: a comparison of the acetate metabolism in *Methanothrix soehngenii* and *Methanosarcina* spp. *FEMS Microbiol Rev* **88**: 181-197.

Johnson, J.M., Wawrik, B., Isom, C., Boling, W.B., and Callaghan, A.V. (2015) Interrogation of Chesapeake Bay sediment microbial communities for intrinsic alkane-utilizing potential under anaerobic conditions. *FEMS Microbiol Ecol* **91**: 1-14.

Jones, D.M., Head, I.M., Gray, N.D., Adams, J.J., Rowan, A.K., Aitken, C.M. et al. (2008) Crude-oil biodegradation via methanogenesis in subsurface petroleum reservoirs. *Nature* **451**: 176-180.

Kanehisa, M., and Goto, S. (2000) KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res* **28**: 27-30.

Kanehisa, M., Goto, S., Sato, Y., Furumichi, M., and Tanabe, M. (2012) KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Res* **40**: D109-D114.

Kim, M., Le, H., McInerney, M.J., and Buckel, W. (2013) Identification and characterization of *re*-citrate synthase in *Syntrophus aciditrophicus*. *J Bacteriol* **195**: 1689-1696.

Kuznetsov, S.I. (1950) A study of the possible current formation of methane in the gas- and oil-bearing facies of the Saratov and Buguruslan districts. *Mikrobiologiia* **19**: 193-202.

Lange, N.A. (1985) *Lange's Handbook of Chemistry, 13th Edition*. New York: McGraw-Hill Book Company.

Li, W., Wang, L.-Y., Duan, R.-Y., Liu, J.-F., Gu, J.-D., and Mu, B.-Z. (2012) Microbial community characteristics of petroleum reservoir production water amended with *n*-alkanes and incubated under nitrate-, sulfate-reducing and methanogenic conditions. *Int Biodeter Biodegr* **69**: 87-96.

Liang, B., Wang, L.-Y., Mbadinga, S.M., Liu, J.-F., Yang, S.-Z., Gu, J.-D., and Mu, B.-Z. (2015) *Anaerolineaceae* and *Methanosaeta* turned to be the dominant microorganisms in alkanes-dependent methanogenic culture after long-term of incubation. *AMB Express* **5**: Article 37.

Lindahl, P.A., and Chang, B. (2001) The evolution of acetyl-CoA synthase. *Origins Life Evol Bios* 31: 403-434.

Liu, Y., Balkwill, D.L., Aldrich, H.C., Drake, G.R., and Boone, D.R. (1999) Characterization of the anaerobic propionate-degrading syntrophs *Smithella propionica* gen. nov., sp. nov. and *Syntrophobacter wolinii*. *Int J Syst Evol Bact* **49**: 545-556.

Lyles, C.N., Le, H.M., Beasley, W.H., McInerney, M.J., and Suflita, J.M. (2014) Anaerobic hydrocarbon and fatty acid metabolism by syntrophic bacteria and their impact on carbon steel corrosion. *Front Microbiol* **5**: Article 114.

Maestrojuán, G.M., Boone, D.R., Xun, L., Mah, R.A., and Zhang, L. (1990) Transfer of *Methanogenium bourgense*, *Methanogenium marisnigri*, *Methanogenium olentangyi,* and *Methanogenium thermophilicum* to the Genus *Methanoculleus* gen. nov., Emendation of *Methanoculleus marisnigri* and *Methanogenium,* and Description of New Strains of *Methanoculleus bourgense* and *Methanoculleus marisnigri*. *Int J Syst Bacteriol* **40**: 117-122.

Martin, M. (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnetjournal* **17**: 10-12.

Mbadinga, S.M., Li, K.-P., Zhou, L., Wang, L.-Y., Yang, S.-Z., Liu, J.-F. et al. (2012) Analysis of alkane-dependent methanogenic community derived from production water of a high-temperature petroleum reservoir. *Appl Microbiol Biotech* **96**: 531-542.

McInerney, M.J., and Beaty, P.S. (1988) Anaerobic community structure from a nonequilibrium thermodynamic perspective. *Can J Microbiol* **34**: 487-493.

McInerney, M.J., Sieber, J.R., and Gunsalus, R.P. (2009) Syntrophy in anaerobic global carbon cycles. *Curr Opin Biotech* **20**: 623-632.

McInerney, M.J., Bryant, M.P., Hespell, R.B., and Costerton, J.W. (1981) *Syntrophomonas wolfei* gen. nov. sp. nov., an anaerobic, syntrophic, fatty acid-oxidizing bacterium. *Appl Environ Microbiol* **41**: 1029-1039.

McInerney, M.J., Rohlin, L., Mouttaki, H., Kim, U., Krupp, R.S., Rios-Hernandez, L. et al. (2007) The genome of *Syntrophus aciditrophicus:* Life at the thermodynamic limit of microbial growth. *Proc Natl Acad Sci USA* **104**: 7600-7605.

Michels, P.A.M., Michels, J.P.J., Boonstra, J., and Konings, W.N. (1979) Generation of an electrochemical proton gradient in bacteria by the excretion of metabolic end products. *FEMS Microbiol Lett* **5**: 357-364.

Nakatsu, C.H., and Marsh, T.L. (2007) Analysis of microbial communities with denaturing gradient gel electrophoresis and terminal restriction fragment length polymorphism. In *Methods for General and Molecular Microbiology*. C. A. Reddy, C.A., Beveridge, T.J., Breznak, J.A., Marzluf, G.A., Schmidt, T.M., and Snyder, L.R. (eds). Washington, DC: ASM Press.

Nobu, M.K., Narihiro, T., Hideyuki, T., Qiu, Y.-L., Sekiguchi, Y., Woyke, T. et al. (2015) The genome of *Syntrophorhabdus aromaticivorans* strain UI provides new insights for syntrophic aromatic compound metabolism and electron flow. *Environ Microbiol* **17**: 4861-4872..

Otto, R., Sonnenberg, A.S., Veldkamp, H., and Konings, W.N. (1980) Generation of an electrochemical proton gradient in *Streptococcus cremoris* by lactate efflux. *Proc Natl Acad Sci USA* **77**: 5502-5506.

Patel, G.B., and Sprott, G.D. (1990) *Methanosaeta concilii* gen. nov., sp. nov. ("*Methanothrix concilii*") and *Methanosaeta thermoacetophila* nom. rev., comb. nov. *Int J Syst Bacteriol* **40**: 79-82.

Penner, T.J., and Foght, J.M. (2010) Mature fine tailings from oil sands processing harbour diverse methanogenic communities. *Can J Microbiol* **56**: 459-470.

Petersen, T.N., Brunak, S., von Heijne, G., and Nielsen, H. (2011) SignalP 4.0: Discriminating signal peptides from transmembrane regions. *Nature Meth* **8**: 785-786.

Pfennig, N., and Biebl, H. (1976) *Desulfuromonas acetoxidans* gen. nov. and sp. nov., a new anaerobic, sulfur-reducing, acetate-oxidizing bacterium. *Arch Microbiol* **110**: 3-12.

Pruitt, K.D., Tatusova, T., and Maglott, D.R. (2007) NCBI reference sequences (RefSeq): A curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res* **35**: D61-D65.

Rainey, F.A., Ward-Rainey, N., Kroppenstedt, R.M., and Stackebrandt, E. (1996) The genus *Nocardiopsis* represents a phylogenetically coherent taxon and a distinct actinomycete lineage: Proposal of *Nocardiopsaceae* fam. nov. *Int J Syst Evol Bact* **46**: 1088-1092.

Romesser, J.A., Wolfe, R.S., Mayer, F., Spiess, E., and Walther-Mauruschat, A. (1979) *Methanogenium*, a new genus of marine methanogenic bacteria, and characterization of *Methanogenium cariaci* sp. nov. and *Methanogenium marisnigri* sp. nov. *Arch Microbiol* **121**: 147-153.

Sato, K., Nishina, Y., Setoyama, C., Miura, R., and Shiga, K. (1999) Unusually high standard redox potential of acrylyl-CoA/ propionyl-CoA couple among enoyl-CoA/acyl-CoA couples: A reason for the distinct metabolic pathway of propionyl-CoA from longer acyl-CoAs. *J Biochem* **126**: 668-675.

Schmidt, A., Müller, N., Schink, B., and Schleheck, D. (2013) A proteomic view at the biochemistry of syntrophic butyrate oxidation in *Syntrophomonas wolfei*. *PLoS ONE* **8**: e56905.

Siddique, T., Penner, T., Semple, K., and Foght, J.M. (2011) Anaerobic biodegradation of longer-chain *n*-alkanes coupled to methane production in oil sands tailings. *Environ Sci Technol* **45**: 5892-5899.

Siddique, T., Penner, T., Klassen, J., Nesbø, C., and Foght, J.M. (2012) Microbial communities involved in methane production from hydrocarbons in oil sands tailings. *Environ Sci Technol* **46**: 9802-9810.

Sieber, J.R., Crable, B.R., Sheik, C.S., Hurst, G.B., Rohlin, L., Gunsalus, R.P., and McInerney, M.J. (2015) Proteomic analysis reveals metabolic and regulatory systems involved in the syntrophic and axenic lifestyle of S*yntrophomonas wolfei*. *Front Microbiol* **6**: Article 115.

Sieber, J.R., Sims, D.R., Han, C., Kim, E., Lykidis, A., Lapidus, A.L. et al. (2010) The genome of *Syntrophomonas wolfei*: new insights into syntrophic metabolism and biohydrogen production. *Environ Microbiol* **12**: 2289-2301.

Smith, K.S., and Ingram-Smith, C. (2007) Methanosaeta, the forgotten methanogen? *Trends Microbiol* **15**: 150-155.

Song, L., Florea, L., and Langmead, B. (2014) Lighter: Fast and memory-efficient sequencing error correction without counting. *Genome Biol* **15**: 509.

Sousa, D.Z., Smidt, H., Alves, M.M., and Stams, A.J.M. (2007) *Syntrophomonas zehnderi* sp. nov., an anaerobe that degrades long-chain fatty acids in co-culture with *Methanobacterium formicicum*. *Int J Syst Evol Bact* **57**: 609-615.

Starai, V.J., and Escalante-Semerena, J.C. (2004) Acetyl-coenzyme A synthetase (AMP forming). *Cell Mol Life Sci* **61**: 2020-2030.

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

Tan, B., Nesbø, C., and Foght, J. (2014a) Re-analysis of omics data indicates *Smithella* may degrade alkanes by addition to fumarate under methanogenic conditions. *ISME J* **8**: 2353-2356.

Tan, B., Dong, X., Sensen, C.W., and Foght, J. (2013) Metagenomic analysis of an anaerobic alkane-degrading microbial culture: Potential hydrocarbon-activating pathways and inferred roles of community members. *Genome* **56**: 599-611.

Tan, B., de Araújo, E.S.R., Rozycki, T., Nesbø, C., and Foght, J. (2014b) Draft genome sequences of three *Smithella* spp. obtained from a methanogenic alkane-degrading culture and oil field produced water. *Genome Announc* **2**: e01085-01014.

Tan, B., Charchuk, R., Li, C., Nesbø, C., Abu Laban, N., and Foght, J. (2014c) Draft genome sequence of uncultivated *Firmicutes* (*Peptococcaceae* SCADC) single cells sorted from methanogenic alkane-degrading cultures. *Genome Announc* **2**: e00909-00914.

Tan, B., Fowler, S.J., Abu Laban, N., Dong, X., Sensen, C.W., Foght, J., and Gieg, L.M. (2015) Comparative analysis of metagenomes from three methanogenic hydrocarbon-degrading enrichment cultures with 41 environmental samples. *ISME J* **9**: 2028-2045.

Tatusov, R.L., Galperin, M.Y., Natale, D.A., and Koonin, E.V. (2000) The COG database: A tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Res* **28**: 33-36.

Ten Brink, B., and Konings, W.N. (1980) Generation of an electrochemical proton gradient by lactate efflux in membrane vesicles of *Escherichia coli*. *Europ J Biochem* **111**: 59-66.

Thauer, R.K., Jungermann, K., and Decker, K. (1977) Energy conservation in chemotropic anaerobic bacteria. *Bacteriol Rev* **41**: 100-180.

Townsend, G.T., Prince, R.C., and Suflita, J.M. (2003) Anaerobic oxidation of crude oil hydrocarbons by the resident microorganisms of a contaminated anoxic aquifer. *Environ Sci Technol* **37**: 5213-5218.

Wang, L.-Y., Gao, C.-X., Mbadinga, S.M., Zhou, L., Liu, J.-F., Gu, J.-D., and Mu, B.-Z. (2011) Characterization of an alkane-degrading methanogenic enrichment culture

from production water of an oil reservoir after 274 days of incubation. *Int Biodeter Biodegr* **65**: 444-450.

Wang, L.Y., Li, W., Mbadinga, S.M., Liu, J.F., Gu, J.D., and Mu, B.Z. (2012) Methanogenic microbial community composition of oily sludge and its enrichment amended with alkanes incubated for over 500 days. *Geomicrobiol J* **29**: 716-726.

Wawrik, B., Boling, W.B., Van Nostrand, J.D., Xie, J., Zhou, J., and Bronk, D.A. (2012a) Assimilatory nitrate utilization by bacteria on the West Florida Shelf as determined by stable isotope probing and functional microarray analysis. *FEMS Microbiol Ecol* **79**: 400-411.

Wawrik, B., Mendivelso, M., Parisi, V.A., Suflita, J.M., Davidova, I.A., Marks, C.R. et al. (2012b) Field and laboratory studies on the bioconversion of coal to methane in the San Juan Basin. *FEMS Microbiol Ecol* **81**: 26-42.

Webner, K. (2012) Die Gene der (1-Methylalkyl) succinat-Synthase im Anaeroben *n*-Alkanabbau des Betaproteobakteriums Stamm HxN1. In *Fachbereich Biologie/Chemie*. Bremen: Universität Bremen, p. 148.

Widdel, F., and Bak, F. (1992) Gram-negative mesophilic sulfate-reducing bacteria. In *The Prokaryotes*. Balows, A., Trüper, H.G., Dworkin, M., Harder, W., and Schleifer, K.H. (eds). New York: Springer-Verlag, pp. 3352-3378.

Wilke, A., Harrison, T., Wilkening, J., Field, D., Glass, E.M., Kyrpides, N. et al. (2012) The M5nr: A novel non-redundant database containing protein sequences and annotations from multiple sources and associated tools. *BMC Bioinformatics* **13**: 141.

Wu, D.Y., Jospin, G., and Eisen, J.A. (2013) Systematic identification of gene families for use as "markers" for phylogenetic and phylogeny-driven ecological studies of bacteria and archaea and their major subgroups. *PLoS ONE* **8**.

Wu, Y.-W., Tang, Y.-H., Tringe, S.G., Simmons, B.A., and Singer, S.W. (2014) MaxBin: An automated binning method to recover individual genomes from metagenomes using an expectation-maximization algorithm. *Microbiome* **2**: 26.

Xie, C., Mao, X., Huang, J., Ding, Y., Wu, J., Dong, S. et al. (2011) KOBAS 2.0: A web server for annotation and identification of enriched pathways and diseases. *Nucleic Acids Res* **39**: W316-W322.

Zengler, K., Richnow, H.H., Rosselló-Mora, R., Michaelis, W., and Widdel, F. (1999) Methane formation from long-chain alkanes by anaerobic microorganisms. *Nature* **401**: 266-269.

Figure 1. Phylogenetic analyses of the SDB consortium metagenomic data. (A) Phylogenetic affiliation of reads containing partial 16S rRNA genes as determined by USEARCH comparison (Edgar, 2010) to the Silva ribosomal RNA database release 111 (www.arb-silva.de). (B) Phylogenetic breakdown of all metagenomic reads as determined by blastN comparison to the M5NR database via MG-RAST (metagenomics.anl.gov).

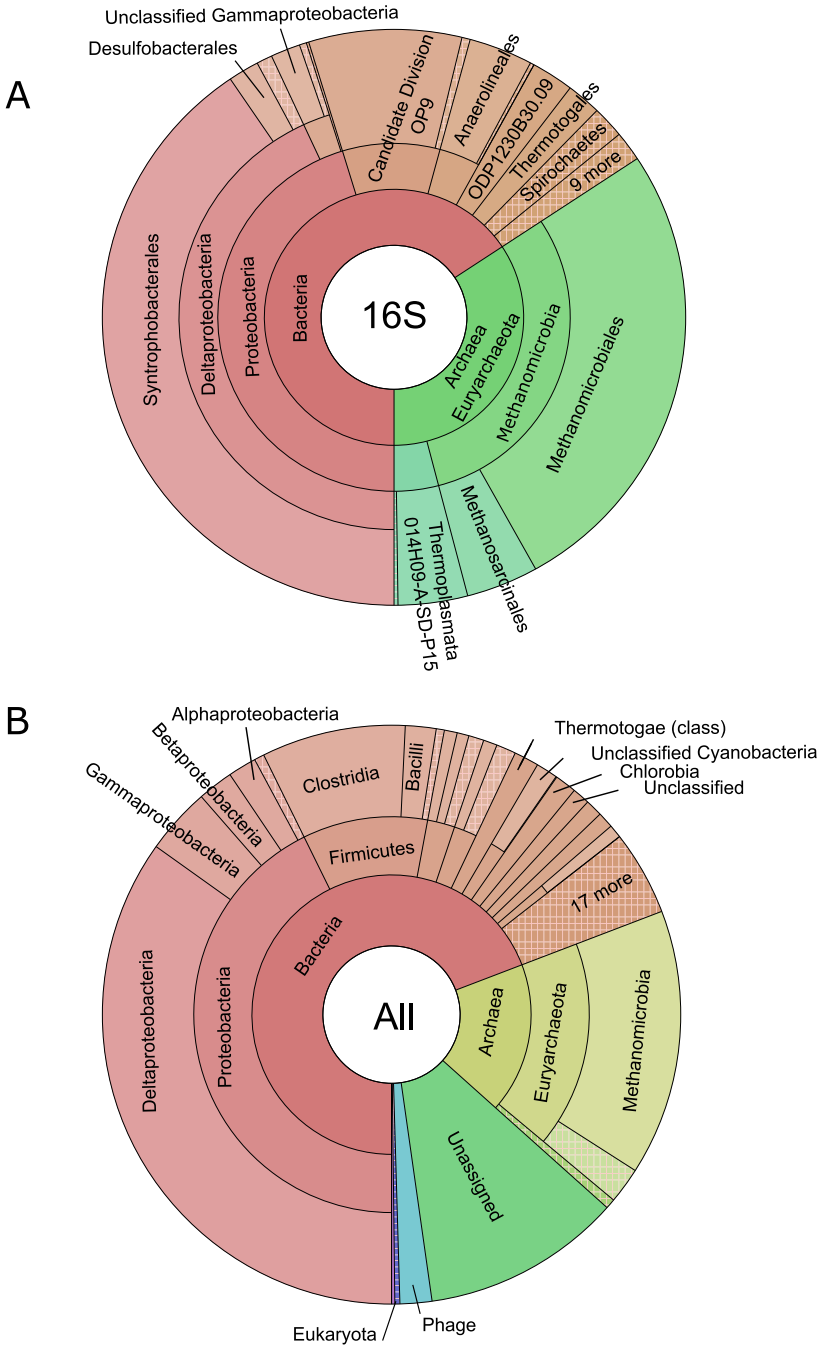


Figure 2. Phylogenetic analysis of the 16S rRNA gene sequence associated with the SDB genome bin that contained an *assA* gene. The neighbor-joining dendrogram (Tajima-Nei distance method; 5,000 bootstrap replicates; pairwise deletion) was generated from an alignment with the most closely associated sequences as determined by blastN analysis of the respective 16S rRNA gene. Sequences of related type strains within the order Syntrophobacterales are also shown. (Note: *"Smithella"* sp. ME-1 is not included because the draft genome does not contain a 16S rRNA gene sequence.)

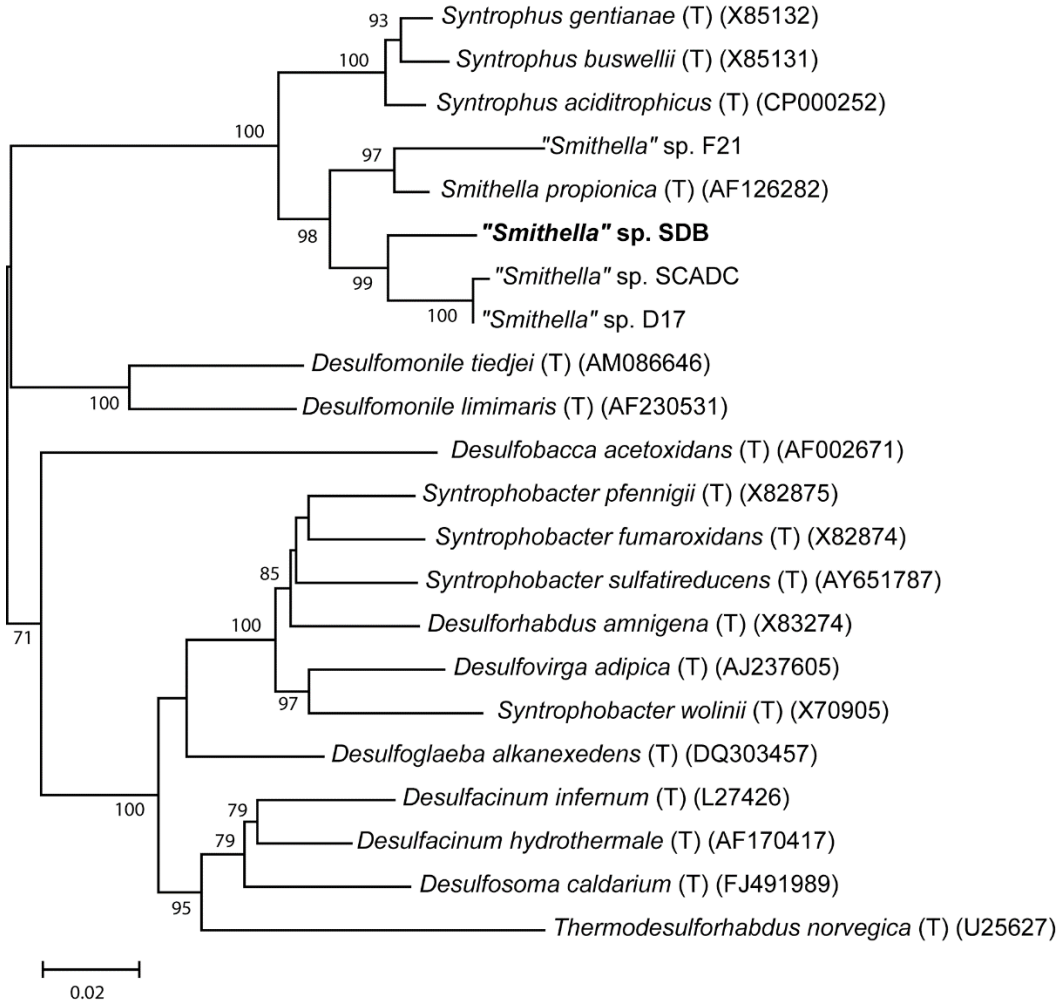


Figure 3. Phylogenetic analysis of *assA* genotypes observed in the SDB metagenome. Metagenomic reads were assembled using Ray (Boisvert et al., 2012), and the resulting contigs were screened for the presence of *assA*-like genes by blastN comparison with a custom database of *assA* and *bssA* sequences. Full length *assA* genes were extracted from contigs for phylogenetic analysis, and the neighbor-joining dendrogram (Poisson model distance method; 5,000 bootstrap replicates; pairwise deletion) was generated from an amino acid alignment with the most closely related known *assA* genotypes, as well as several representative sequences from the database, including the catalytic subunits of the related glycyl radical enzymes. Blue circles indicate the relative abundances of the five detected *assA* OTUs in the metagenome. RT-PCR transcription of the *assA* genotypes in cultures amended with octacosane with and without the HMN carrier is shown to the right of the blue circles. Boxes: Green – transcript detected and sequenced for confirmation; and Red – transcripts were not detected. Abbreviations: AssA – alkylsuccinate synthase alpha subunit; BssA – benzylsuccinate synthase alpha subunit; HbsA – (hydroxybenzyl)succinate synthase alpha subunit; HMN - 2,2,4,4,6,8,8-heptamethylnonane carrier; IbsA – (4-isopropylbenzyl)succinate synthase alpha subunit; MasD – methylalkylsuccinate synthase alpha subunit; and NmsA – 2-naphthylmethylsuccinate synthase alpha subunit.

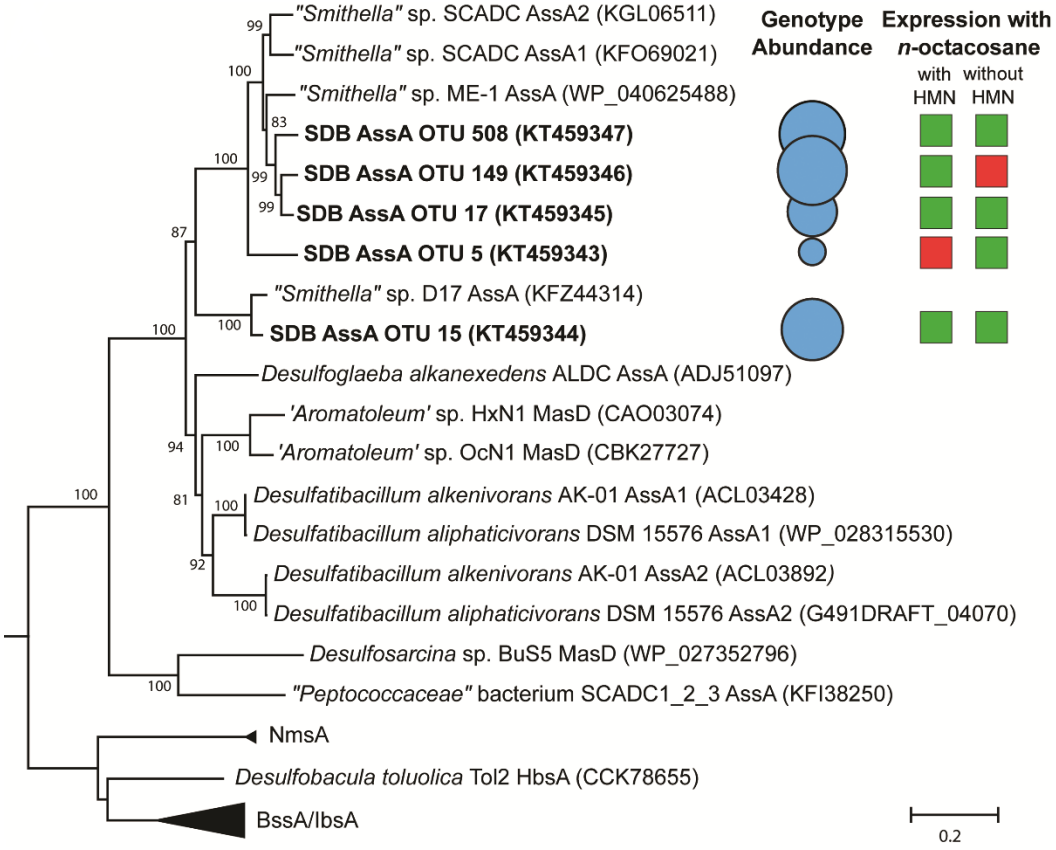


Figure 4. Genomic organization of genes involved in the anaerobic activation of hydrocarbons via addition to fumarate that were detected in the SDB metagenome (Contig/OTU - 345/149, 346/15, 347/5, 348/508, and 350/17) and the genomes of several cultivated and uncultivated hydrocarbon-degrading bacteria. The genomic organizations of alkylsuccinate synthase-associated genes in '*Aromatoleum*' sp. HxN1, '*Aromatoleum'* sp. OcN1, *Desulfothermus naphthae* TD3, and *Desulfatibacillum* sp. Pnd3 were adapted from Webner (Webner, 2012). The genomic organizations of alkylsuccinate synthase genes in *"Smithella"* strains D17, SCADC, and ME-1 were derived from their genomic sequences (Tan et al., 2014b). The gene organization shown for *"Peptococcaceae"* sp. SCADC was derived from single-cell genomics data (Tan et al., 2014c), whereas data for *Desulfatibacillum alkenivorans* AK-01 were derived from the completed genome (Callaghan et al., 2012).

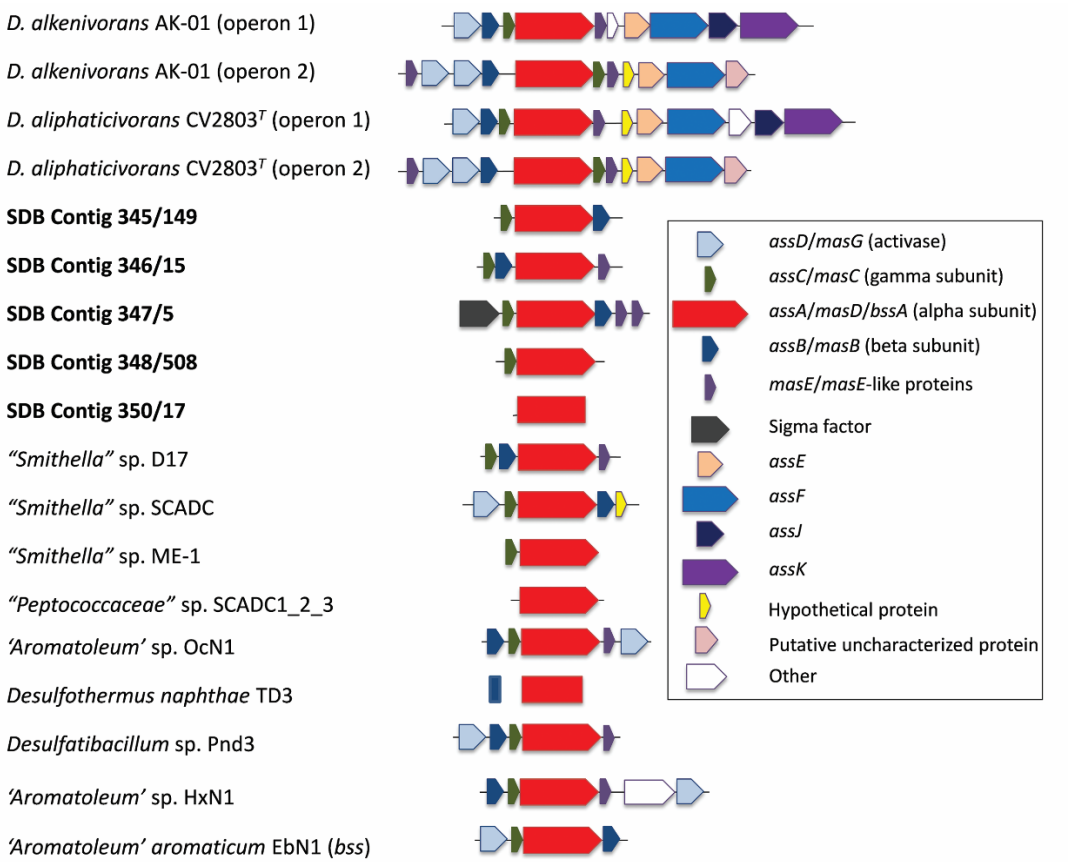
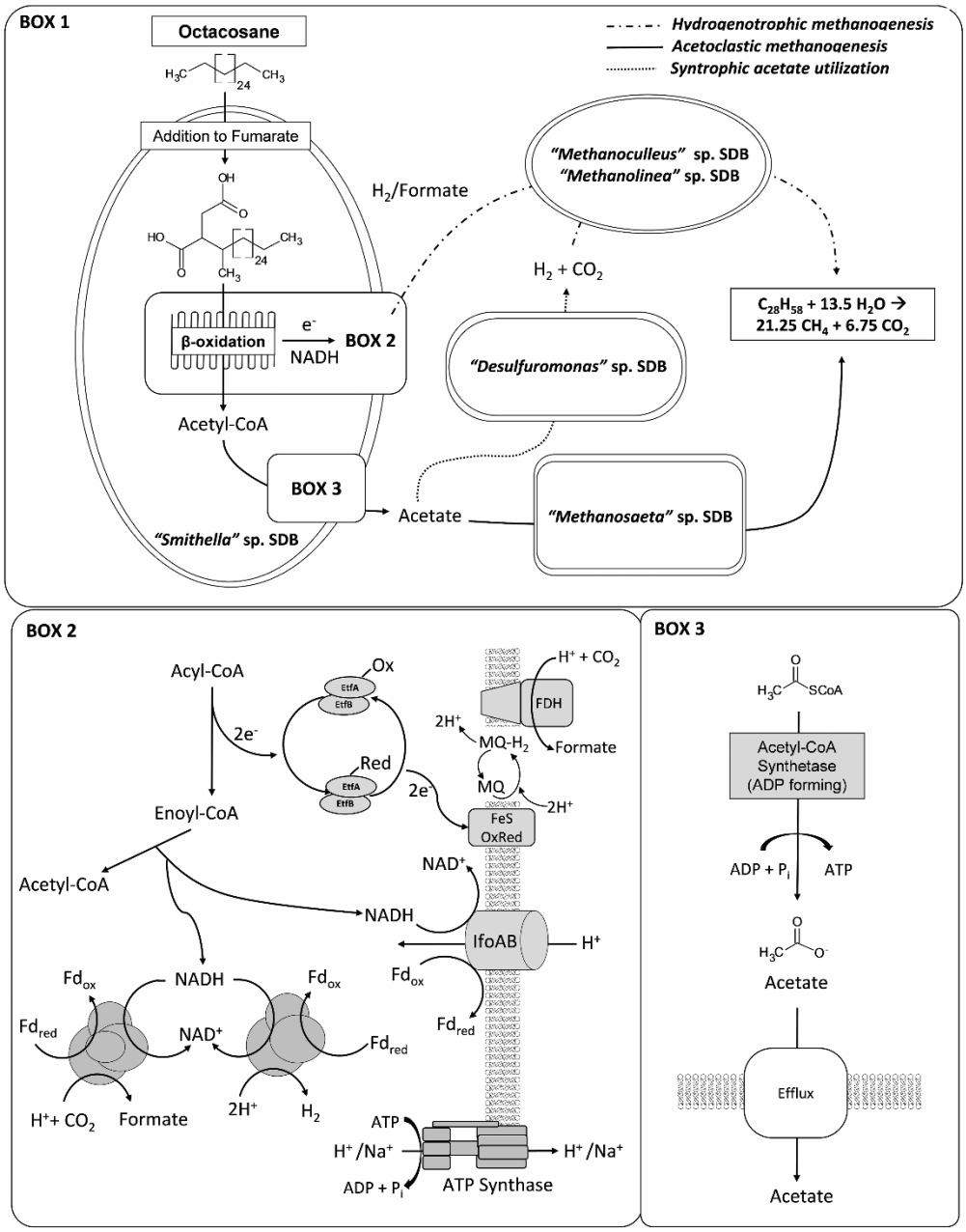


Figure 5. Schematic of the hypothetical carbon and energy flow in the SDB as it relates to interactions within the consortium (Box1), reverse electron transfer linked to beta oxidation and the generation of proton motive force (Box 2), and ATP synthesis (Box 3). Protein complexes identified in the binned genome data are shown in grey. Metagenomic data implicate *"Smithella"* in the activation of octacosane via addition to fumarate, generating H_2/formate, CO_2, and acetate, which are utilized by *"Desulfuromonas"* and the hydrogenotrophic and acetoclastic methanogens *"Methanoculleus/Methanolinea"* and *"Methanosaeta"*, respectively.



APPENDIX I: Genome Annotation and Analysis of Hydrocarbon Degradation Features of *Desulfoglaeba alkanexedens* ALDC.

Methods

Cultivation and DNA extraction

D. alkanexedens was cultivated for genomic DNA extraction on a reduced basal marine medium supplemented with yeast extract (100 mg L^{-1}) and butyrate (10 mM) as substrate at 31°C. Biomass was harvested from ~200 mL of culture during exponential phase by centrifugation at 5,500xg for 15 min in sealed polycarbonate vessels. The resulting cell pellet was resuspended in the Qiagen Genomic Tip Bacterial Lysis Buffer B1 modified to contain a 10-fold higher concentration of EDTA (46.5 g L^{-1} Na_2EDTA).

Genome sequencing and annotation

Genomic DNA was extracted using the Qiagen Genomic Tip DNA Extraction Kit (Qiagen Inc.; Valencia, CA) as the described by the manufacturer. Long genomic DNA fragments were prepared into libraries, sequenced on the PacBio RSII platform, and assembled with the HGAP2 algorithm by Washington State University Molecular Biology and Genomics Core Facility. The resulting single contig genome sequence was uploaded for automated annotation and public access in the JGI Integrated Microbial Genomes version 4 pipeline (Markowitz et al., 2014). All downstream phylogenetic analyses and refined sequence alignments were conducted using MEGA6 software package (Tamura et al., 2013). Manual annotation of open-reading frames

associated with alkane metabolism was performed using the DNASTAR software suite (DNASTAR; Madison, WI).

Results

General Genome Features

The genome of *D. alkanexedens* ALDC is a single circular chromosome 3,365,583 bp in length with 84.43% of bases coding and an average mol G+C content of 59.31%. No evidence of plasmids was found. There were 3020 protein-coding genes identified and 2270 of these (73.65%) were assigned a functional prediction by the IMG annotation pipeline. The genome contains 62 RNA genes, specifically: 6 rRNA genes (two copies of each 5S, 16S and 23S), 47 tRNA genes, and 9 other miscellaneous categorized RNA genes. Transfer RNA genes associated with at least a single codon for each of the 20 essential amino acids and selenocycsteine were present. The 16S rRNA genes are 99.9% similar to each other and are both located on the minus strand at locus tags 111218 and 113015. Based on the functional assignments of annotated putative protein-coding sequences, *D. alkanexedens* ALDC is predicted to be auxotrophic for 8 of the 20 essential amino acids (L-phenylalanine, L-tyrosine, L-histidine, L-cysteine, L-isoleucine, L-leucine, L-serine, and L-threonine).

Alkane Activation

Previous work demonstrated the activation of *n*-alkane substrates via the radical-mediated addition to fumarate (Kropp et al., 2000; Davidova et al., 2005). Two distinct gene clusters were identified encoding the requisite subunits for fully

functional alkylsuccinate synthases (Figs. 1 & 2). Gene cluster 1 is encoded on the minus strand and contains the requisite genes *assABCDE* and *masE,* as well as ancillary genes for large and small subunits of methylmalonyl-CoA mutase, an ATP-dependent LAO/AO transport system, and a MmgE/PrpD family protein. Gene cluster 2 on the plus strand contains functionally homologous genes as in cluster 1, but also contains *assF* and a gene coding for a putative carboxyl transferase.

The *assA* sequences identified within both gene clusters were highly similar to each other, maximum amino acid sequence dissimilarity of 3.1%, and form a distinct subclade within the polyphyletic group comprised of *Desulfatibacillum* spp., *Azoarcus* sp. HxN1, and '*Aromatoleum*' sp. OcN1 (Fig. 3). All of the AssA homologues were most closely affiliated with AssA1 (Dalk_1731) of *D. alkenivorans* AK-01 with 77% sequence identity and residue chemistry similarity of 88%. Alignment of each AssA protein sequence revealed the proposed catalytic glycine residue and single cysteine residue (distinct from the conserved tandem cysteine for pyruvate-formate lyases) to be conserved relative to other AssA homologues (Fig.4) (Callaghan et al., 2008).

A single copy of the alkylsuccinate synthase beta subunit gene (*assB1* and *assB2*) was identified within each of the gene clusters. These homologues share 92% identical amino acid sequence to each other and are most closely associated with AssB2 (Dalk_2200; 55% identical, 72% positives) and AssB1 (Dalk_1729; 56% identical, 74% positives) of *D. alkenivorans* AK-01 for *D. alkanexedens* AssB1 and AssB2, respectively. Full length alignment (125 amino acids) of the AssB homologues from *D. alkanexedens* to other reference proteins from *Desulfatibacillum* spp., "*Smithella*" spp., and *Aromatoleum* sp. HxN1 revealed the conserved cysteine residues at ^{41}C-X-X-

^{44}C-[X]$_{15}$-^{60}C-[X]$_{34}$-^{95}C, consistent with the distinctive iron sulfur cluster binding motif proposed previously (Callaghan et al., 2008).

Similar analyses were performed for all identified *assC* and *masE* genes. In each case, all three homologous protein sequences (one in the gene cluster 1 and two in gene cluster 2) for each allele conformed to the cysteine residue pattern previously proposed by Callaghan and colleagues (2008): AssC, ^{4}C-X-X - ^{7}C-[X]$_{14}$ - ^{22}C-[X]$_{25}$ -^{48}C; MasE, ^{3}C-X-X - ^{6}C-X $_{25}$ - ^{32}C-X-X - ^{35}C. The *assC* genes are highly similar to each other with identical amino acid sequence and a maximum nucleotide dissimilarity of 1.2%. Blast analysis of the protein sequences indicated the best match was to the AssC of the uncultivated "*Smithella*" sp. SCADC at 69% sequence identity and 81% residue chemistry similarity. The MasE homologues are also highly similar with a maximal amino acid divergence of 2.7%. All MasE sequences were closely affiliated with homologues from *D. alkenivorans* AK-01, with MasE1 and MasE2' both most similar to Dalk_2197 (64% identity, 75% positives), while MasE2 was to Dalk_1732 (60% identity, 76% positives).

Central Metabolism

Aside from C_6-C_{12} *n*-alkanes, *D. alkanexedens* ALDC was initially shown to metabolize a limited number of substrates including: pyruvate, butyrate, hexanoate, and 1-methyloctanoate (Davidova et al., 2006). In agreement with this initial description, all genes necessary for the oxidation of linear fatty acids to acetyl-CoA through the β-oxidation pathway were found. A pyruvate:ferredoxin oxidoreductase (*porABDG*) was encoded in the genome with a single copy of each subunit co-localized on the plus strand (locus tags 111576-79). No genes associated with a

benzoyl-CoA ligase (E.C. 6.1.2.25) or a benzoyl-CoA reductase (E.C. 1.3.7.8) required for the oxidation of aromatic substrates were found in the genome. The genome contains a nearly complete glycolysis pathway, but is missing the critical enzyme aldolase (E.C. 4.1.2.13) responsible for the cleavage of the 6 carbon skeleton. In addition, genes for several enzymes in the pentose phosphate pathway are present that should allow for the synthesis of key metabolic intermediates: erythrose-4P, Ribose-5P, and 5'-phosphoribosyl-diphosphate (PRPP).

Acetyl-CoA produced from the oxidation of various substrates has several metabolic fates encoded within the genome. A complete TCA cycle is present utilizing a *si*-citrate synthase that is 65% identical (amino acid alignment) to that of *Syntrophobacter fumaroxidans* MPOB (Sfum_2105). *D. alkanexedens* ALDC also possesses a complete Wood-Ljungdahl pathway and an AMP-dependent acetyl-CoA synthetase (ACS) for the conversion of acetyl-CoA to CO_2 or acetate under respiratory or fermentative conditions, respectively. The ACS of *D. alkanexedens* ALDC (locus tag 113073) shares 76% amino acid identity with the ACS from the order-level fatty acid-oxidizing relative, *S. fumaroxidans* MPOB (Sfum_0745). No genes for a phosphotransacetylase or acetate kinase were found.

Sulfur Metabolism

D. alkanexedens ALDC possesses all of the genes necessary for the activation of sulfate to adenylylsulfate by sulfate adenylyltransferase (*sat*) and subsequent reduction to sulfite and ultimately sulfide by adenosine-5'-phosphosulfate reductase (*aprAB*) and dissimilatory sulfite reductase (*dsrAB*), respectively. In addition to sulfate respiration, *D. alkanexedens* ALDC was shown to utilize thiosulfate as an

electron acceptor under heterotrophic growth conditions (Davidova et al., 2006). A gene encoding the catalytic subunit for a thiosulfate reductase (*phsA*) was identified, though the additional electron transfer and cytochrome b containing subunits (*phsBC*) were not observed. No genes encoding an anaerobic sulfite reductase (*asrAB*) were detected in the genome. In addition, no genes associated with the assimilatory reduction to sulfate via PAPS reductase and assimilatory sulfite reductase were identified. Nor was any evidence found of genes associated with the oxidation of sulfide by sulfur:quinone oxidoreductase or sulfur oxidation by the SOX enzyme system. Several membrane complexes are required to couple the reduction of terminal electron acceptors such as sulfate or thiosulfate to energy conservation. *D. alkanexedens* ALDC possesses all of the genes for the Quinone reductase complex (*qrcABCD*), the quinone-interacting membrane bound oxidoreductase complex (*qmoABC*) and the transmembrane electron transport complex (*dsrKMJOP*).

Nitrogen Metabolism

The initial description of *D. alkanexedens* ALDC does not detail the specific nitrogenous compounds used for either assimilatory or dissimilatory processes (Davidova et al., 2006). No genes associated with the dissimilatory reduction of nitrate to ammonia or denitrification were found in the genome. Similarly, there were no observed genes for an assimilatory nitrate reductase. Glutamine synthetase and two subunits of the glutamate synthase (*gltBD*) were found in the genome. The NADPH or NADH binding subunit of glutamate synthase (GLT1) was not detected. In the absence of an identifiable glutamate dehydrogenase, *D. alkanexedens* ALDC most likely assimilates ammonia through the ATP-dependent GS-GOGAT pathway tied to

the TCA cycle. Genes for all subunits of urease (*ureCBA*) were also detected, suggesting that urea could also serve as a nitrogen source via the assimilation of the resulting ammonia. Nitrogenase genes (*nifDHK*) were also identified and indicate the potential for diazotrophy, though neither the fixation of dinitrogen nor the utilization of urea have been demonstrated in cultivation by *D. alkanexedens* ALDC to date.

Discussion

Desulfoglaeba alkanexedens is one of the few currently known model organisms known to activate *n*-alkanes by addition to fumarate under sulfate-reducing conditions. As the sole member of the order *Syntrophobacterales* shown to utilize linear paraffins as substrates, a study to sequence and annotate the genome of this organism was undertaken to provide a framework for further physiological and ecological investigations. Previous studies utilizing [^{13}C]-*n*-dodecane revealed metabolic intermediates indicative of a glycyl-radical enzyme catalyzed activation of the subterminal carbon, and resulting addition to the unsaturation of fumarate (Davidova et al. 2005).

Genome analysis has confirmed *D. alkanexedens* ALDC to be a strictly anaerobic sulfate- (or thiosulfate) reducing organism with a limited range of substrates that should support growth, as initially described (Davidova et al., 2006). Two distinct homologous gene clusters encoding all of the necessary proteins for the alkylsuccinate synthase enzyme were identified. Though possessing homologous genes, these gene clusters are not co-localized on the chromosome and differ in strand orientation and gene arrangement. Multiple homologous ASS/MAS gene clusters have also been documented in the genomes of *Desulfatibacillum alkenivorans* AK-01 and *Azoarcus*

sp. HxN1 (A. V. Callaghan et al., 2008; Callaghan et al., 2012; Webner, 2012). While there is sequence divergence within the two ASS gene clusters of AK-01, the two MAS operons recently identified in HxN1 are completely identical at the nucleotide level. As seen with ASS cluster 2 in AK-01, the ASS gene clusters of *D. alkanexedens* ALDC also possess multiple copies of certain genes within each cluster (Figs. 1 & 2). Interestingly, ALDC is the first sequenced genome to show multiple copies of *assA* within a single gene cluster, having a total of three copies of *assA* within two gene clusters. The altered arrangement and multiple highly similar copies of several ASS genes within these two gene clusters, in addition the identified inverted repeat sequences may be evidence of previous intra-chromosomal homologous recombination events.

After the activation of the n-alkane substrate and subsequent rearrangement, the resulting fatty acid intermediate is further metabolized to acetyl-CoA via β-oxidation (Callaghan et al., 2012). The genome of ALDC encodes all of the requisite proteins associated with the repeated oxidative decarboxylation of linear fatty acids. Acetyl-CoA produced by the oxidation of substrates including pyruvate, n-alkanes, and fatty acids has several potential fates encoded within the genome. During respiration, acetyl-CoA can be completely oxidized to CO_2 via the Wood-Ljungdahl pathway providing further electrons for energy conservation. Under fermentative conditions such as during syntrophic paraffin degradation, acetyl-CoA is likely converted to acetate by substrate-level phosphorylation via the AMP-dependent ACS. Interestingly, the presence of a complete Wood Ljungdahl pathway suggests that chemolithoautotrophic growth coupled to sulfate or thiosulfate reduction may be

possible. The initial description of *D. alkanexedens* reported that H_2:CO_2 did not support growth (Davidova et al., 2006), but this metabolism has been reported in *D. alkenivorans* AK-01 which also possess similar pathways (So and Young, 1999; Callaghan et al., 2012). No genes were detected encoding any of the subunits of a benzoyl-CoA ligase or benzoyl-CoA reductase, supporting the previous assertion that aromatic compounds cannot be utilized as substrates for growth.

Acetyl-CoA may also play another critical role in the metabolism of *D. alkanexedens* ALDC by serving as the central link between the catabolic and anabolic processes. The genome contains a complete oxidative TCA cycle and significant portions of the glycolytic and pentose-phosphate pathways. These three pathways likely serve as the dominant routes for a number of anabolic intermediates within the cell.

References

Callaghan, A. V, Morris, B.E.L., Pereira, I.A.C., McInerney, M.J., Austin, R.N., Groves, J.T., et al. (2012) The genome sequence of Desulfatibacillum alkenivorans AK-01: a blueprint for anaerobic alkane oxidation. *Environ. Microbiol.* **14**: 101–113.

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

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

Davidova, I.A., Duncan, K.E., Choi, O.K., and Suflita, J.M. (2006) Desulfoglaeba alkanexedens gen. nov., sp. nov., an n-alkane-degrading, sulfate-reducing bacterium. *Int J Syst Evol Microbiol* **56**: 2737–2742.

Davidova, I.A., Gieg, L.M., Nanny, M., Kropp, K.G., and Suflita, J.M. (2005) Stable isotopic studies of n-alkane metabolism by a sulfate-reducing bacterial enrichment culture. *Appl. Environ. Microbiol.* **71**: 8174–8182.

Kropp, K.G., Davidova, I.A., and Suflita, J.M. (2000) Anaerobic oxidation of n-dodecane

by an addition reaction in a sulfate-reducing bacterial enrichment culture. *Appl Env. Microbiol* **66**: 5393–5398.

Markowitz, V.M., Chen, I.-M.A., Palaniappan, K., Chu, K., Szeto, E., Pillay, M., et al. (2014) IMG 4 version of the integrated microbial genomes comparative analysis system. *Nucleic Acids Res.* **42**: D560–7.

So, C.M. and Young, L.Y. (1999) Isolation and characterization of a sulfate-reducing bacterium that anaerobically degrades alkanes. *Appl. Environ. Microbiol.* **65**: 2969–76.

Tamura, K., Stecher, G., Peterson, D., Filipski, A., and Kumar, S. (2013) MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol. Biol. Evol.* **30**: 2725–9.

Webner, K. (2012) Die Gene der (1-Methylalkyl)succinat-Synthase im anaeroben n-Alkanabbau des Betaproteobakteriums Stamm HxN1.

Figure 1. Alkylsuccinate synthase gene cluster 1. Blue arrows denote open-reading frame orientation and frame. Spacing of all figure elements is proportional to genome arrangement. Red blocks indicate identified inverted repeat sequences.

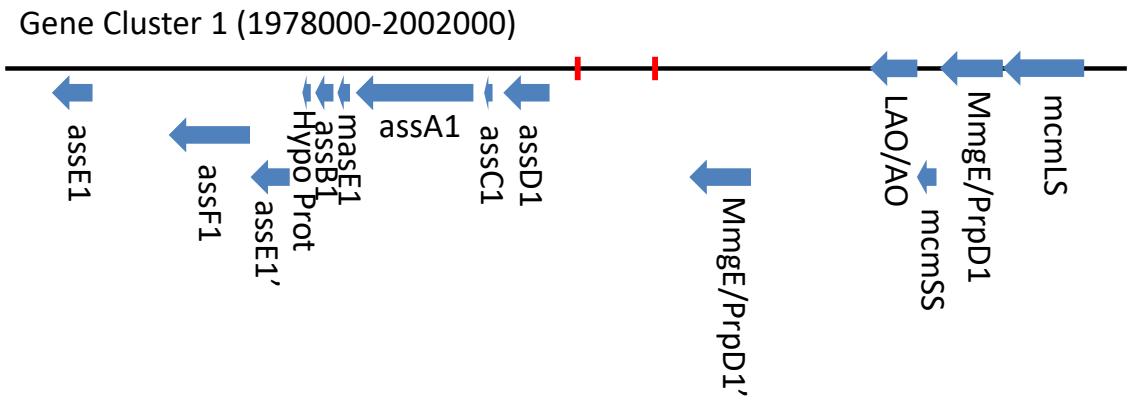


Figure 2. Alkylsuccinate synthase gene cluster 2. Blue arrows denote open-reading frame orientation and frame. Spacing of all figure elements is proportional to genome arrangement. Red blocks indicate identified inverted repeat sequences.

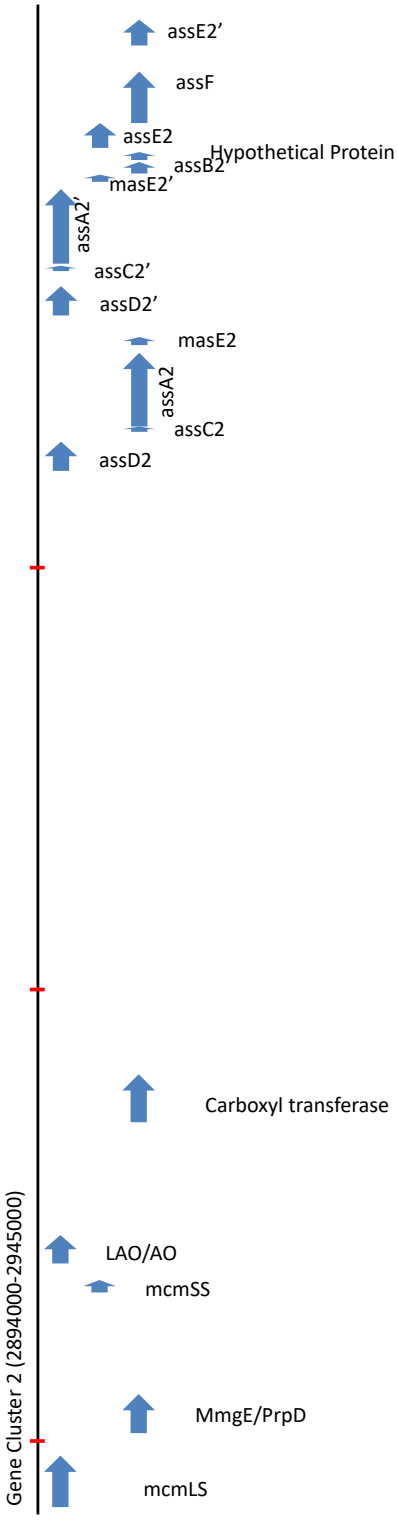


Figure 3. Phylogenetic analysis of protein-coding sequences of fumarate-addition glycyl-radical enzyme alpha subunits. Neighbor-joining dendrogram was calculated based on a poisson method with complete deletion. Statistical evaluation was performed by bootstrap method with 500 replications.

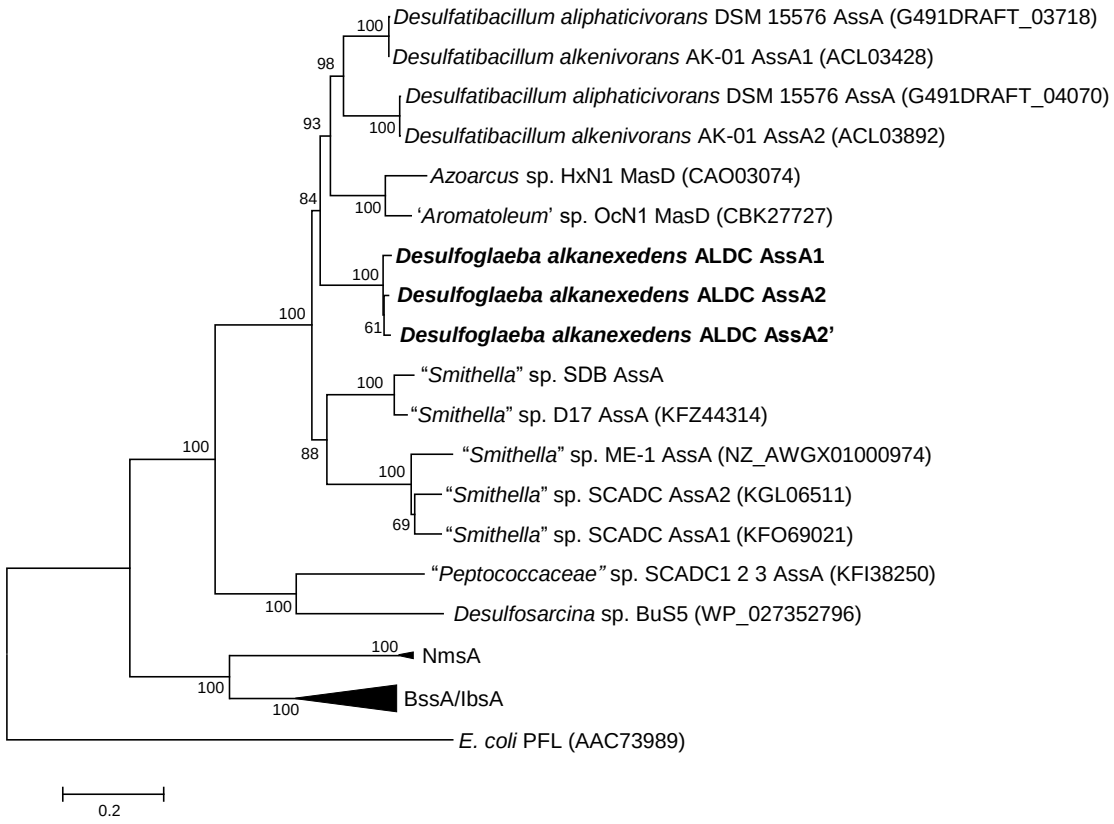
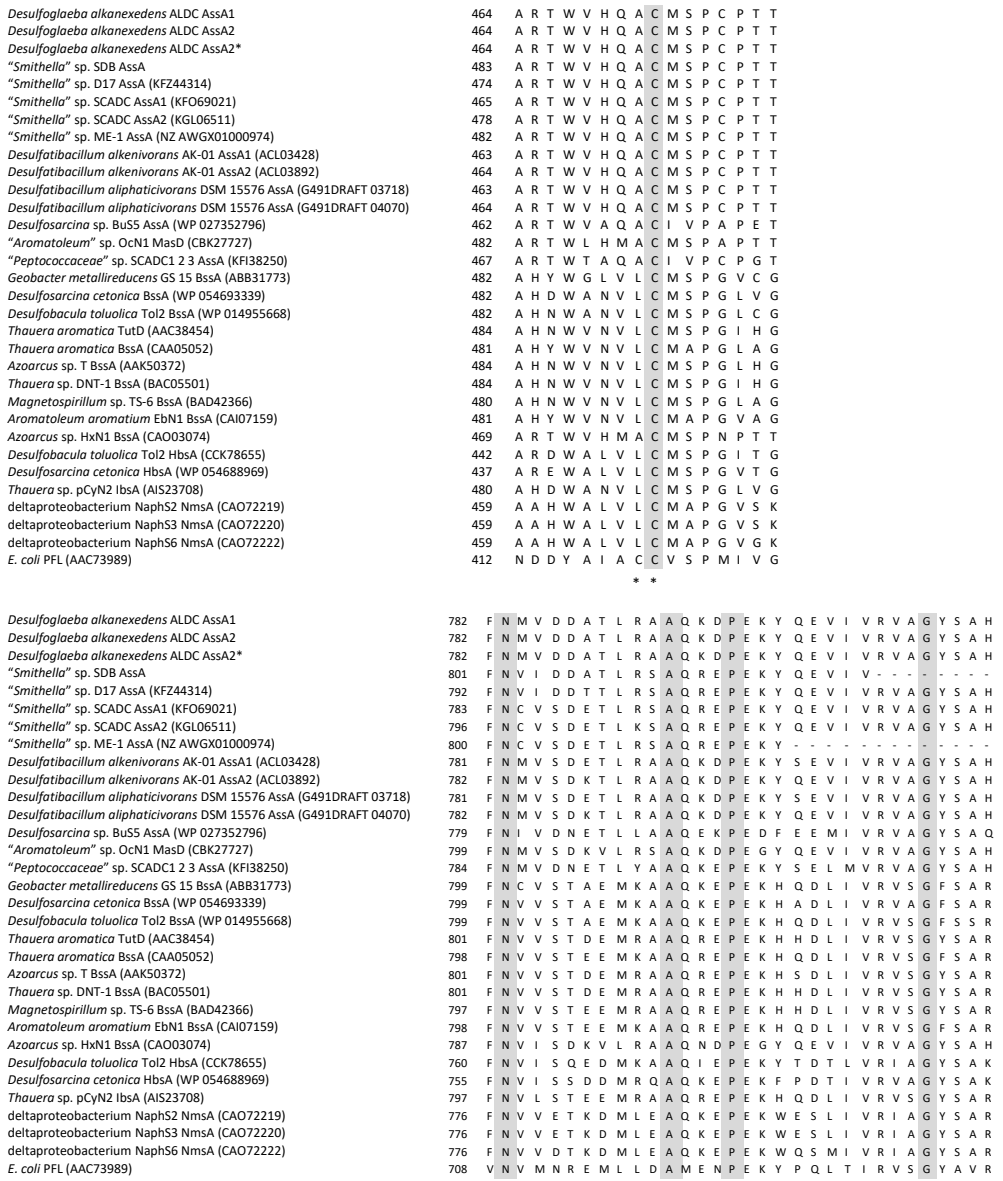


Figure 4. Translated amino acid alignments of fumarate-addition glycyl-radical alpha subunits showing distinct conserved catalytic glycine (top) and cysteine (bottom) residues.



Appendix II: Chapter 3 Supplemental Materials.

References

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. *Nature Meth* **5**: 235-237.

Johnson, J.M., Wawrik, B., Isom, C., Boling, W.B., and Callaghan, A.V. (2015) Interrogation of Chesapeake Bay sediment microbial communities for intrinsic alkane-utilizing potential under anaerobic conditions. *FEMS Microbiol Ecol* **91**: 1-14.

Figure S1. Methane production measured in the sediment-free SDB cultures amended with (A) octacosane and (B) other medium- and long- chain *n*-alkanes. Data points represent the averages of triplicate measurements. Error bars correspond to one standard deviation. For some measurements, the standard deviations were very small, and the error bars are masked by the corresponding data points.

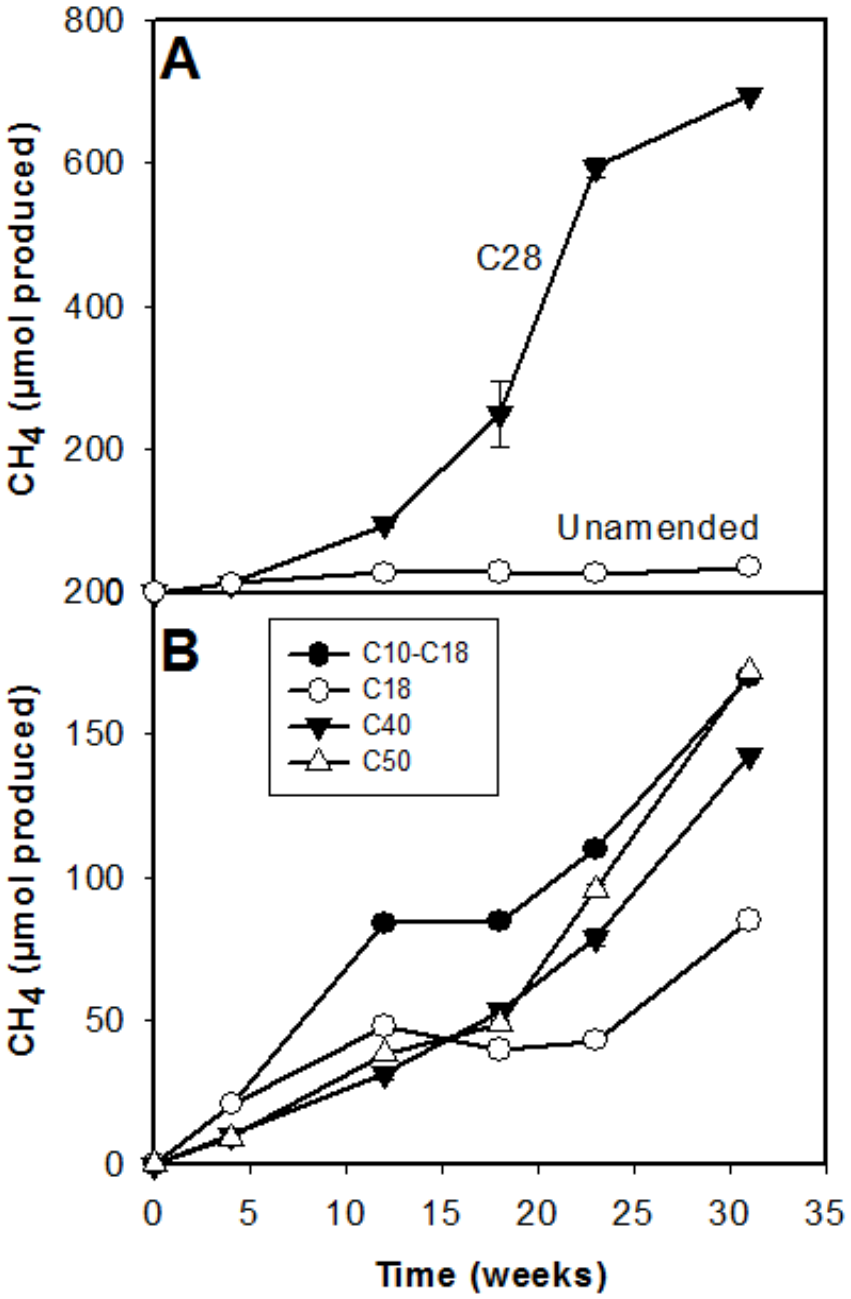


Figure S2. Phylogenetic profile of 16S rRNA genes obtained by 454 pyrosequencing of (A) bacterial and (B) archaeal 16S rRNA gene PCR products. PCR products were generated, barcoded, and sequenced as previously described (Johnson et al., 2015). Briefly, bacterial 16S rRNA gene PCR products were generated with primers 27F and 338R, which contained 5' Titanium Fusion adapter sequences (forward primer A-tag: CCATCTCATCCCTGCGTGTCTCCGACTCAG; reverse primer B-tag: CCTATCCCCTGTGTGCCTTGGCAGTCTCAG) and a unique 8-nucleotide barcode tag (Hamady et al., 2008). PCR reactions contained 0.2 μM of the 'tagged' forward primer, 0.25 μM of the reverse primer, 0.25 μL of DreamTaq (5 units μL^{-1}) (Thermo Fisher Scientific, Waltham, MA), PCR Supermix (Life Technologies, Carlsbad, CA) and 2 μL of template DNA (1:15 dilution). PCR cycles were as follows: 95°C for 7 min. and 30 cycles of 95°C for 20 sec., 55°C for 20 sec., and 72°C for 40 sec. Archaeal 16S rRNA genes were amplified using primers A8F (5'-TCCGGTTGATCCTGCC-3') and A344R (5'-TCGCGCCTGCTGCICCCCGT-3') with Titanium adaptors described above and by applying an extension step of 55°C for 60 seconds. PCR products were sequenced by loading equimolar quantities on a 454 sequencer using GS FLX Titanium chemistry.

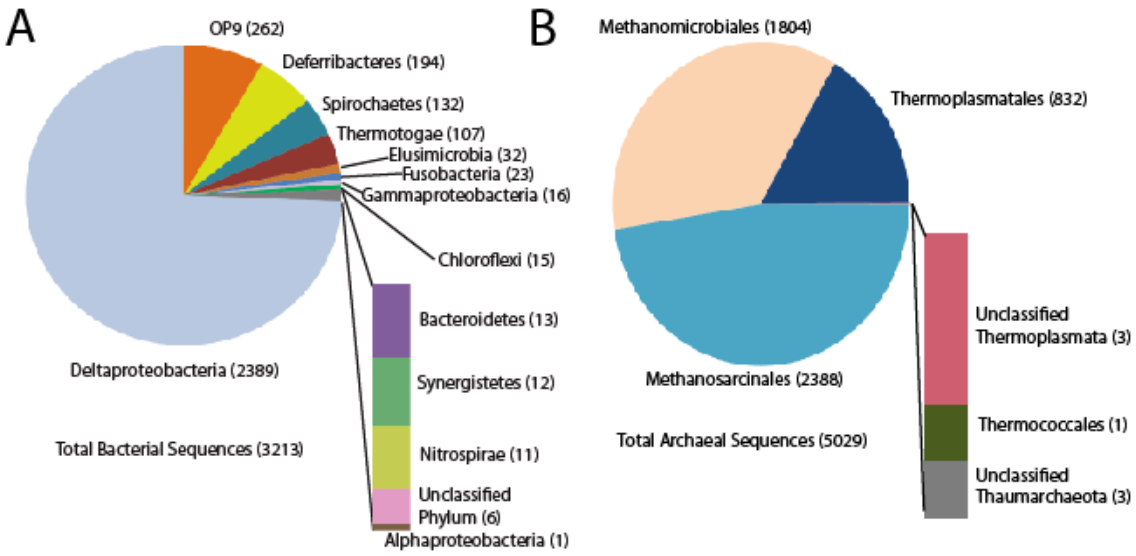


Figure S3. Phylogenetic analysis of the 16S rRNA gene sequences associated with archaeal genome bins that affiliated with 16S rRNA genes. The neighbor-joining dendrogram (Tajima-Nei distance method; 5,000 bootstrap replicates; pairwise deletion) was generated from an alignment with most closely associated sequences as determined by blastN analysis of the respective 16S rRNA gene.

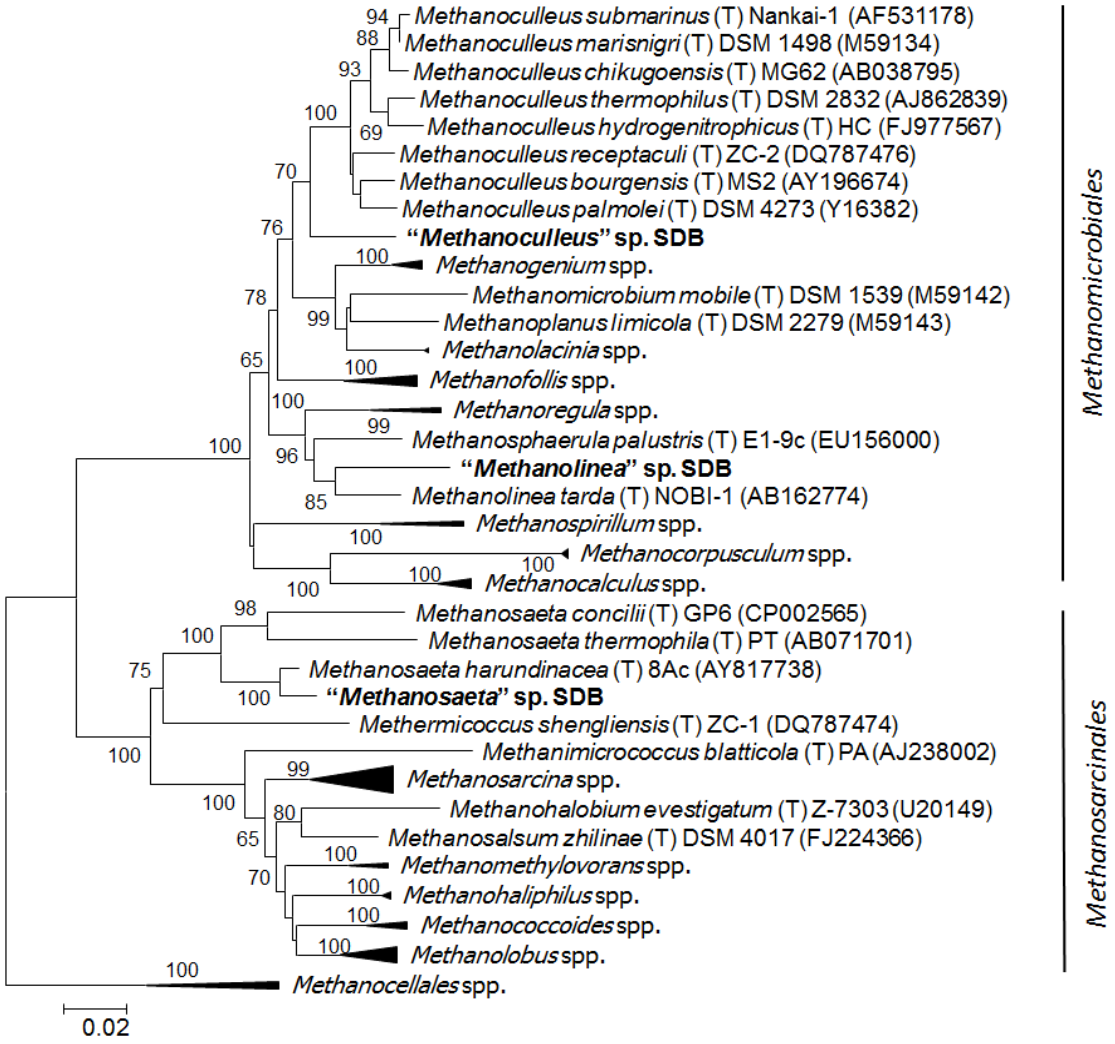


Table S1. The stoichiometry of octacosane mineralization to methane in SDB cultures after 16 weeks of incubation.

	C_{28}-alkane Remaining (mg)	C_{28}-alkane Consumed* (mg)	C_{28}-alkane Consumed (mmol)	CH_4 Produced (mmol)	CH_4 Expected (mmol)	CH_4 Produced/Expected (%)
Replicate 1	1.28	0.40	1.00×10^{-3}	0.023	0.020	115
Replicate 2	1.44	0.24	6.10×10^{-4}	0.010	0.013	77
Replicate 3	1.10	0.58	1.47×10^{-3}	0.027	0.031	87
Sterile control	1.68	n.a.	n.a.	not detected	n.a.	n.a.
Uninoculated control	2.20	n.a.	n.a.	not detected	n.a.	n.a.

n.a. = not applicable

**Consumption was calculated as the difference between live and sterile controls.*

Table S2. Characteristics of binned genomes in the SDB culture obtained via MaxBin analysis of Illumina MiSeq data.

Scaffold	Contigs	Length (bp)	N50	% G+C	Coverage	Complete	% of reads	ORFs	Best Recruitment
1	398	2971898	11111	41	646.82	95%	34.24	2900	* *Syntrophus aciditrophicus* SB
2	109	2085722	39717	57	274.23	98%	10.19	2259	¥ *Methanospaerula palustris* E1-9c
3	433	3139690	12300	38	213.67	98%	11.95	2960	*Desulfuromonas acetoxidans*
4	550	2936005	8388	37	90.62	95%	4.74	2492	*Cloacamonas acidaminovorans*
5	511	2169164	6455	54	80.26	95%	3.10	2365	λ *Methanospaerula palustris* E1-9c
6	829	3071038	4924	43	52.46	78%	2.87	3056	*Syntrophus aciditrophicus SB*
7	388	1681062	6441	49	50.09	90%	1.50	1494	*Desulfuromonas acetoxidans*
8	329	2526504	12272	46	44.7	95%	2.01	2379	*Anaerolinea thermophila* UNI-1
9	631	1584395	2939	34	40.9	90%	1.15	1444	*Thermincola* sp. JR
10	897	2754247	3666	49	30.09	83%	1.48	3288	*Desulfatibacillum alkenivorans* AK-01
11	665	2580768	6176	45	17.45	88%	0.80	2504	*Kosmotoga olearia* TBF 19.5.1
12	1099	2207150	2178	34	16.41	75%	0.65	1982	*Thermincola* sp. JR
13	452	3408119	15125	50	15.9	93%	0.97	3618	*Pyrococcus furiosus* DSM 3638
14	1559	3596862	2714	42	15.25	98%	0.98	3616	*Syntrophus aciditrophicus* SB
15	1158	2304839	2165	58	12.91	55%	0.53	2269	┼ *Methanosaeta thermophila* PT
16	1085	2463402	2566	52	10.91	60%	0.48	2416	*Spirochaeta smaragdinae* DSM 11293
17	1100	1912728	1764	67	9.05	75%	0.31	1534	*Thioalkalivibrio* sp. HL-EbGR7
18	2739	4118536	1456	47	8.57	53%	0.63	4178	*Syntrophus aciditrophicus* SB
19	1113	1738014	1456	55	8.25	63%	0.26	1627	*Methanoculleus marisnigri* JR1

Genomic bins affiliated with 16S rRNA genes: * *"Smithella"* sp. SDB in Figure 3; ¥ *"Methanoculleus"* sp. SDB; λ *"Methanolinea"* sp. SDB; ┼ *"Methanosaeta"* sp. SDB

Table S3. Number of genes with and without homology between the binned *"Smithella"* sp. SDB genome and previously reported *"Smithella"*-like genome bins from methanogenic alkane-degrading cultures and *Syntrophus aciditrophicus* SB. All protein-coding genes were compared via pairwise blastp comparison. The number of protein features that have blastp scores of e <1E-10 in pairwise comparison is shown.

Organism	# of genes with homology in *"Smithella"* sp. SDB	# of genes without homology in *"Smithella"* sp. SDB
"Smithella" sp. ME-1	1843	996
"Smithella" sp. SCADC	1993	1614
"Smithella" sp. F21	900	873
"Smithella" sp. D17	1059	687
Syntrophus aciditrophicus SB	1384	1782

Table S4. *assA* genotypes detected in PCR-based clone libraries, 454- based metagenomes, and Illumina-based metagenomes.

454 Metagenome Contig #	# of Sequences	Illumina Metagenome Contig #	# of Sequences	PCR Library OTU	# of Clones	mRNA Detected Via RT-PCR with HMN	mRNA Detected Via RT-PCR without HMN
345	83	149	9423	-	-	Yes	No
346	48	15	7447	-	-	Yes	Yes
347	32	5	1416	OTU 1	12	No	Yes
348	50	508	8511	-	-	Yes	Yes
350	35	17	4822	OTU 4	6	Yes	Yes
-	-	-	-	OTU 2	14	*n.a.*	*n.a.*
-	-	-	-	OTU 3	1	*n.a.*	*n.a.*
-	-	-	-	OUT 5	1	*n.a.*	*n.a.*

'-' = not detected; '*n.a.*' = not assayed; HMN = heptamethylnonane carrier phase for octacosane substrate

Table S5. Number of mismatches between primers used for *assA* RT-PCR and *assA* OTUs detected in metagenome.

RELATIVE POSITION*	PRIMER	*assA* Genes Detected in SDB contig ID: 454 (Illumina) metagenome 345 (149)	346 (15)	347 (5)	348 (508)	350 (17)	PRIMER SEQUENCE
1410	C345_F	0	5	4	2	4	AGA GGC CTT GGA TCT GGC
2250	C345_R	0	13	6	0	4	ATC GGT GAT AGC CTT CT
1505	C345/47/48/50_F	0	2	0	0	0	AAG GMA CCT GGG TCT GGA
2320	C345/47/48/50_R	0	5	0	0	0	TCG TCA TCA TTG CCC CAC
1885	C346_F	10	0	8	7	9	GCC TTA TTT AAC GGC TAC
2308	C346_R	5	0	3	5	5	TCC CCA TTT AGG CGC GTT
395	C350_F	6	7	12	9	0	AAA ATG CTA GTA TTC AAG AT
712	C350_R	5	6	9	5	0	GTT GGA TCC CAG CCA ATT

* Relative position to *assA* gene in *Desulfatibacillum alkenivorans* strain AK-01

Table S6. Comparison of genomic content of available *"Smithella"* draft genomes originating from methanogenic alkane-degrading consortia. (NOTE: Search parameters are shown at the bottom of the table. Text in parentheses refers to EC numbers or locus tags within searched genomes.) Search parameters were as follows:

n.d. = not detected
Λ = as determined by examining RAST annotation
¥ = as determine by blastp (E<1e-10; Identities >25%; hit across full length of query)
£ = as determined by blastp against genes from *Desulfatibacillum alkenivorans* AK-01 (E<1e-10; Identities >25%; hit across full length of query)
§ = as determined by blastp (Diamond) search against genes from *Syntrophorhabdus aromaticivorans* (E<1e-10)
┼ = as determined by blastp (Diamond) search against genes from *Syntrophobacter fumaroxidans* (E<1e-10)
= as determined by blastp (Diamond) search against the KOBAS database and subsequent KEGG Map reconstruction (E<1e-10)

"Smithella" strain	SDB	ME-1	SCADC	F21	D17
General Features					
# of Contigs	398	1037	244	251	335
Draft Genome Size	2.97E+06	2.91E+06	3.19E+06	1.63E+05	1.63E+06
# of Coding Sequences	2900	2839	3457	1737	1746
% Completeness Based on Presence of 111 Bacterial Single Copy Marker Genes	94.6	81.1	99.1	77.5	82.0
Features Related To Alkane Degradation ¥					
Methylmalonyl-CoA mutase, small subunit (Dalk_1724)	+	+	+	+	n.d.
Methylmalonyl-CoA mutase, large subunit (Dalk_1725)	+	+	+	+	n.d.
Putative LAO/AO transport system ATPase (Dalk_1726)	+	+	+	+	n.d.
Putative MmgE/PrpD family protein (Dalk_1727)	+	+	+	+	n.d.
Alkylsuccinate synthase (I), activating enzyme (AssD1) (Dalk_1728) (1.97.1.4)	+	+	+	n.d.	n.d.
Alkylsuccinate synthase (I), beta subunit (AssB1) (Dalk_1729)	+	+	+	n.d.	+
Alkylsuccinate synthase (I), gamma subunit (AssC1) (Dalk_1730)	+	n.d.	+	n.d.	n.d.
Alkylsuccinate synthase (I),subunit (AssA1) (Dalk_1731)	+	+	+	n.d.	+
Putative uncharacterized protein similar to MasE (Dalk_1732)	+	+	+	n.d.	+
Hypothetical protein (Dalk_1733)	+	+	+	n.d.	n.d.

Chaperone (AssE1) (Dalk_1734)	+	+	+	n.d.	+
Putative uncharacterized protein (AssF1) (Dalk_1735)	+	+	+	n.d.	+
Carboxyl transferase (Dalk_1740)	+	n.d.	+	n.d.	n.d.
Beta Oxidation ^					
Long-chain-fatty-acid--CoA ligase (6.2.1.3)	+	+	+	+	+
Carnitine O-palmitoyltransferase (2.3.1.21)	n.d.	n.d.	n.d.	n.d.	n.d.
Acyl-CoA oxidase (1.3.3.6)	n.d.	n.d.	n.d.	n.d.	n.d.
Acyl-CoA dehydrogenase (1.3.99.3/1.3.8.7)	+	+	n.d.	n.d.	n.d.
Long-chain-acyl-CoA dehydrogenase (1.3.99.13/1.3.8.8)	n.d.	n.d.	n.d.	n.d.	n.d.
Enoyl-CoA hydratase (4.2.1.17)	+	+	+	+	+
Long-chain-enoyl-CoA hydratase (4.2.1.74)	n.d.	n.d.	n.d.	n.d.	n.d.
3-Hydroxyacyl-CoA dehydrogenase (1.1.1.35)	+	+	+	+	+
Long-chain-3-hydroxyacyl-CoA Dehydrogenase (1.1.1.211)	n.d.	n.d.	n.d.	n.d.	n.d.
3-Ketoacyl-CoA thiolase (2.3.1.16)	+	+	+	+	+
Short-chain acyl-CoA dehydrogenase (1.3.99.2/1.3.8.1)	+	+	+	+	+
Acetyl-CoA C-acetyltransferase (2.3.1.9)	+	+	+	+	+
Long-chain-fatty-acid--[acyl-carrier-protein] Ligase(6.2.1.20)	n.d.	n.d.	n.d.	n.d.	n.d.
Dodecenoyl-CoA Isomerase (5.3.3.8)	n.d.	+	n.d.	+	+
3-Hydroxybutyryl-CoA epimerase (5.1.2.3)	n.d.	+	n.d.	+	+
Putative sirtuin-type regulator for CoA-ligase (3.5.1.98)	n.d.	n.d.	n.d.	n.d.	n.d.
Electron transfer flavoprotein, alpha subunit	+	+	+	+	+
Electron transfer flavoprotein, beta subunit	+	+	+	+	+
Acetyl-CoA hydrolase (3.1.2.1)	+	+	+	+	n.d.
Glutaconate CoA-transferase subunit A (2.8.3.12)	+	n.d.	n.d.	n.d.	+
Glutaconate CoA-transferase subunit B (2.8.3.12)	+	+	n.d.	+	+
Alpha-methylacyl-CoA racemase (5.1.99.5)	+	n.d.	+	n.d.	n.d.
Isovaleryl-CoA dehydrogenase (1.3.99.10/1.3.8.4)	n.d.	n.d.	+	+	n.d.
Succinyl-CoA synthase (6.2.1.5)	n.d.	n.d.	n.d.	+	n.d.
Phenylacetate-CoA ligase (6.2.1.30)	+	+	+	+	+
Acetate-CoA synthetase (6.2.1.1)	+	+	+	+	+

Trans-feruloyl-CoA synthase (6.2.1.34)	n.d.	n.d.	n.d.	n.d.	n.d.
Alcohol and Aldehyde Metabolism ^					
Long-chain-alcohol dehydrogenase (1.1.1.192)	n.d.	n.d.	n.d.	n.d.	n.d.
Long-chain-aldehyde dehydrogenase (1.2.1.48)	n.d.	n.d.	n.d.	n.d.	n.d.
Alcohol dehydrogenase (1.1.1.1)	+	+	+	+	+
Aldehyde dehydrogenase (1.2.1.3)	+	+	+	n.d.	n.d.
Polyhydroxybutyrate Metabolism ^					
3-Hydroxybutyryl-CoA epimerase (5.1.2.3)	n.d.	+	n.d.	+	n.d.
3-Hydroxybutyryl-CoA dehydratase (4.2.1.55)	+	+	+	n.d.	+
Polyhydroxyalkanoic acid synthase	+	+	+	n.d.	+
Acetyl-CoA acetyltransferase (2.3.1.9)	+	+	+	+	+
Acetoacetyl-CoA synthetase (6)	+	+	n.d.	n.d.	n.d.
Enoyl-CoA hydratase (4.2.1.17)	+	+	+	+	+
3-Ketoacyl-CoA thiolase (2.3.1.16)	+	+	+	+	+
3-Hydroxyacyl-CoA dehydrogenase (1.1.1.35)	+	+	+	+	+
3-Hydroxybutyryl-CoA dehydrogenase (1.1.1.157)	+	+	+	+	+
Synthesis of Oxaloacetate and Fumarate ^					
Pyruvate-ferredoxin oxidoreductase (1.2.7.1)	+	n.d.	+	+	n.d.
Pyruvate phosphate dikinase (2.7.9.1)	+	+	+	+	+
Pyruvate carboxylase, biotin carboxylase subunit (6.4.1.1)	+	+	+	n.d.	n.d.
Pyruvate carboxylase, carboxyl transferase subunit (6.4.1.1)	n.d.	n.d.	n.d.	n.d.	n.d.
Phosphoenolpyruvate carboxylase (4.1.1.31)	n.d.	n.d.	n.d.	n.d.	n.d.
ATP citrate lyase (2.3.3.8)	n.d.	n.d.	n.d.	n.d.	n.d.
Malate synthase (2.3.3.9)	n.d.	n.d.	n.d.	n.d.	n.d.
Malate dehydrogenase, NAD-dependent (1.1.1.37)	+	n.d.	+	+	+
Class I fumarate hydratase, FumA (4.2.1.2)	+	+	n.d.	+	n.d.
Class II fumarate hydratase, FumC (4.2.1.2)	n.d.	n.d.	n.d.	n.d.	n.d.
Methylmalonyl-CoA:pyruvate transcarboxylase	n.d.	n.d.	n.d.	n.d.	n.d.

TCA Cycle (Oxidative) ^					
Citrate synthase (2.3.3.1)	n.d.	n.d.	n.d.	+	n.d.
Re-Citrate synthase (NCBI Accession #: CAI83711) ¥	+	+	+	n.d.	+
Aconitate hydratase (4.2.1.3)	+	+	+	n.d.	+
Isocitrate dehydrogenase (1.1.1.42)	+	+	+	n.d.	+
Isocitrate dehydrogenase (1.1.1.41)	n.d.	n.d.	n.d.	n.d.	n.d.
2-Oxoglutarate dehydrogenase (1.2.4.2 / 2.3.1.61)	n.d.	n.d.	n.d.	n.d.	n.d.
2-Oxoglutarate ferredoxin oxidoreductase (1.2.7.3)	+	+	+	+	n.d.
Dihydrolipoamide dehydrogenase (1.8.1.4)	+	+	+	+	n.d.
Succinyl-CoA synthetase alpha subunit (6.2.1.4)	n.d.	n.d.	n.d.	n.d.	n.d.
Succinyl-CoA synthetase alpha subunit (6.2.1.5)	n.d.	n.d.	n.d.	+	n.d.
Succinyl-CoA:acetate CoA-transferase (2.8.3.18)	n.d.	n.d.	n.d.	n.d.	n.d.
Fumarate reductase flavoprotein (1.3.5.4)	n.d.	n.d.	n.d.	n.d.	n.d.
Succinate dehydrogenase flavoprotein (1.3.5.1)	n.d.	+	n.d.	n.d.	n.d.
Fumarate hydratase (4.2.1.2)	+	+	n.d.	+	n.d.
Malate dehydrogenase (1.1.1.37)	+	n.d.	+	+	+
Pyruvate carboxylase (6.4.1.1)	+	+	+	+	+
TCA Cycle (Reductive) ^					
PEP-carboxylase (4.1.1.32/4.1.1.49)	+	n.d.	n.d.	+	n.d.
Malate dehydrogenase (NAD-dependent) (1.1.1.37)	+	n.d.	+	+	+
Class I fumarate hydratase, FumA (4.2.1.2)	+	+	n.d.	+	n.d.
Class II fumarate hydratase, FumC (4.2.1.2)	n.d.	n.d.	n.d.	n.d.	n.d.
Fumarate reductase (1.3.1.6)	n.d.	n.d.	n.d.	n.d.	n.d.
2-Oxoglutarate synthase (alpha subunit) (1.2.7.3)	+	+	+	+	n.d.
2-Oxoglutarate synthase (beta subunit) (1.2.7.3)	+	+	+	+	n.d.
2-Oxoglutarate synthase (gamma subunit) (1.2.7.3)	+	+	+	+	n.d.
2-Oxoglutarate synthase (delta subunit) (1.2.7.3)	+	+	+	+	n.d.
Isocitrate dehydrogenase, NADP-dependent (1.1.1.42)	+	+	+	n.d.	+

Aconitate hydratase (4.2.1.3)	+	n.d.	+	n.d.	+
ATP-citrate lyase (2.3.3.8)	n.d.	n.d.	n.d.	n.d.	n.d.
2-Oxoglutarate synthase (1.2.7.3)	+	+	+	+	n.d.
Pyruvate-ferredoxin oxidoreductase (1.2.7.1)	+	n.d.	+	+	n.d.
Pyruvate phosphate dikinase (2.7.9.1)	+	+	+	+	+
Pyruvate and Acetate Metabolism ^					
Pyruvate synthase (1.2.7.1)	+	n.d.	+	+	n.d.
Pyruvate dehydrogenase E1 component (1.2.4.1)	+	+	n.d.	+	n.d.
Pyruvate dehydrogenase E2 component (2.3.1.12)	n.d.	n.d.	n.d.	n.d.	n.d.
Dihydrolipoamide Dehydrogenase (1.8.1.4)	+	+	+	+	n.d.
Pyruvate carboxylase (6.4.1.1)	+	+	+	+	+
Acetyl-CoA hydrolase (3.1.2.1)	n.d.	n.d.	n.d.	n.d.	n.d.
Acetyl-CoA ligase (AMP forming) (6.2.1.1)	+	+	+	+	n.d.
Acetate-CoA synthetase (ADP-forming) (6.2.1.13)	+	+	+	+	+
Propionate-CoA transferase (2.8.3.1)	n.d.	n.d.	n.d.	n.d.	n.d.
Acetate-CoA ligase (ADP forming) (6.2.1.13)	+	n.d.	+	+	n.d.
Phosphate acetyltransferase (2.3.1.8)	n.d.	n.d.	+	n.d.	n.d.
Acetate kinase (2.7.2.1)	+	+	+	n.d.	+
Acylphosphatase (3.6.1.7)	+	+	+	n.d.	+
Acetaldehyde dehydrogenase (1.2.1.10)	n.d.	n.d.	n.d.	n.d.	n.d.
Aldehyde dehydrogenase (NAD+) (1.2.1.3)	+	+	+	n.d.	n.d.
Methylmalonyl-CoA mutase (5.4.99.2)	+	+	+	+	n.d.
Methylmalonyl-CoA epimerase (5.1.99.1)	n.d.	n.d.	n.d.	+	n.d.
Wood-Ljungdahl Pathway ^					
Carbon monoxide dehydrogenase (acceptor) (1.2.99.2)	n.d.	n.d.	n.d.	n.d.	n.d.
Carbon monoxide dehydrogenase (ferredoxin) (1.2.7.4)	n.d.	n.d.	n.d.	n.d.	n.d.
Acetyl-CoA synthase (2.3.1.169)	n.d.	n.d.	n.d.	n.d.	n.d.
5-Methyltetrahydrofolate methyltransferase (2.1.1.258)	n.d.	n.d.	n.d.	n.d.	n.d.

Methylene-tetrahydrofolate reductase (1.5.1.20)	+	+	+	+	+
Methylene-tetrahydrofolate dehydrogenase (NADPH+) (1.5.1.5)	+	+	+	n.d.	+
Methenyltetrahydrofolate cyclohydrolase (3.5.4.9)	+	+	+	n.d.	+
Formate-tetrahydrofolate ligase (6.3.4.3)	n.d.	n.d.	n.d.	n.d.	n.d.
Formate Dehydrogenases ^					
Formate dehydrogenase-O FdoG, alpha subunit (1.2.1.2)	+	+	+	+	n.d.
Formate dehydrogenase-O FdoH, beta subunit (1.2.1.2)	+	+	+	+	+
Formate dehydrogenase-O FdoI, gamma subunit (1.2.1.2)	+	+	+	+	+
NAD-dependent formate dehydrogenase, alpha subunit (1.2.1.2)	+	+	+	+	n.d.
Formate dehydrogenase chain D (EC 1.2.1.2)	+	+	+	+	n.d.
Formate dehydrogenase formation protein FdhE	+	+	+	+	n.d.
Hydrogenases ^					
Hydrogenase (NADH) subunit HoxE (1.12.1.2)	+	+	+	+	n.d.
Hydrogenase (NADH) subunit HoxF (1.12.1.2)	+	+	+	+	n.d.
Hydrogenase (NADH) subunit B (1.12.1.2)	+	+	+	+	n.d.
Hydrogenase (Fe), HymA subunit	+	+	+	+	n.d.
[NiFe] Hydrogenase nickel incorporation protein HypA	+	+	+	+	+
[NiFe] Hydrogenase nickel incorporation protein HypB	+	+	+	+	+
Periplasmic [Fe] hydrogenase (ferredoxin) large subunit (1.12.7.2)	+	n.d.	+	+	n.d.
Periplasmic [Fe] hydrogenase (ferredoxin) small subunit (1.12.7.2)	n.d.	n.d.	+	+	n.d.
CoB--CoM Reducing hydrogenase (ferredoxin) delta subunit	+	+	+	+	n.d.
Coenzyme F_{420}-reducing hydrogenase (ferredoxin) beta subunit	+	+	+	+	n.d.
Terminal Electron Acceptors ^					
Nitrate reductase (1.7.99.4)	n.d.	n.d.	n.d.	n.d.	n.d.
Nitrite reductase (1.7.2.1)	n.d.	n.d.	n.d.	n.d.	n.d.
Nitric oxide reductase (1.7.99.7)	n.d.	n.d.	n.d.	n.d.	n.d.

Glutamine synthetase (6.3.1.2)	+	+	+	+	n.d.
Nitrogenase (1.18.6.1)	n.d.	n.d.	+	n.d.	n.d.
ATP sulfurylase (2.7.7.4)	+	n.d.	n.d.	n.d.	n.d.
Adenylyl-sulfate kinase (2.7.1.25)	n.d.	n.d.	n.d.	n.d.	n.d.
PAPS reductase (1.8.4.8)	n.d.	n.d.	n.d.	n.d.	n.d.
Adenylyl-sulfate reductase (1.8.99.2)	n.d.	n.d.	n.d.	n.d.	n.d.
Thiosulfate reductase (2.8.1.5)	n.d.	n.d.	n.d.	n.d.	n.d.
Sulfite reductase (1.8.99.1)	n.d.	n.d.	n.d.	n.d.	n.d.
Sulfite reductase (NADH) (1.8.1.2)	n.d.	n.d.	n.d.	n.d.	n.d.
Sulfite reductase (ferredoxin) (1.8.7.1)	n.d.	n.d.	n.d.	n.d.	n.d.
Cysteine synthase	+	+	+	+	n.d.
Sulfate permease / transporter	n.d.	n.d.	n.d.	n.d.	+
Membrane Complexes [£ §]					
DsrMKJOP [£]					
Sulfite reduction-associated complex , DsrM	n.d.	n.d.	n.d.	n.d.	n.d.
Sulfite reduction-associated complex , DsrK	n.d.	n.d.	n.d.	n.d.	n.d.
Sulfite reduction-associated complex, DsrJ	n.d.	n.d.	n.d.	n.d.	n.d.
Sulfite reduction-associated complex , DsrO	n.d.	n.d.	n.d.	n.d.	n.d.
Sulfite reduction-associated complex , DsrP	n.d.	n.d.	n.d.	n.d.	n.d.
QmoABC [£]					
Quinone-interacting oxidoreductase, QmoA (1.8.98.1)	+	+	+	+	n.d.
Quinone-interacting oxidoreductase, QmoB (1.8.98.1)	+	+	+	+	n.d.
Quinone-interacting oxidoreductase, QmoC (1.8.98.1)	n.d.	n.d.	n.d.	n.d.	n.d.
QrcABCD/MopABCD [£]					
Quinone reductase complex, QrcA(1.2.7.- -)	n.d.	n.d.	n.d.	n.d.	n.d.
Quinone reductase complex, QrcB (1.2.7.- -)	n.d.	n.d.	n.d.	n.d.	n.d.
Quinone reductase complex, QrcC (1.2.7.- -)	n.d.	n.d.	n.d.	n.d.	n.d.
Quinone reductase complex, QrcD (1.2.7.- -)	n.d.	n.d.	n.d.	n.d.	n.d.
Ohc Complex [£]					

FAD-dependent pyridine nucleotide-disulphide oxidoreductase	n.d.	n.d.	n.d.	+	+
Iron-sulfur binding domain protein with TAT signal OhcB1	+	n.d.	+	n.d.	+
Protein with transmembrane helices OhcB2	n.d.	n.d.	n.d.	n.d.	n.d.
Octaheme, *c*-type cytochrome OhcA	n.d.	n.d.	n.d.	n.d.	n.d.
Membrane-associated cytochrome OhcC	n.d.	n.d.	n.d.	n.d.	n.d.
Putative equivalent of Hmc or 9Hc complex [£]					
Putative uncharacterized periplasmic protein (Dalk_1817)	n.d.	n.d.	n.d.	n.d.	+
Periplasmic 4Fe-4S ferredoxin transmembrane protein (Dalk_1818)	+	n.d.	+	n.d.	n.d.
NrfD family membrane protein (Dalk_1819)	n.d.	n.d.	n.d.	n.d.	n.d.
Cytochrome *c* family protein (Dalk_1820)	n.d.	n.d.	n.d.	n.d.	n.d.
Putative uncharacterized protein (Dalk_1821)	n.d.	n.d.	n.d.	n.d.	n.d.
Rnf Complex [£]					
Cytochrome *c*, class III family protein, Rnf-associated (dhcA)	n.d.	n.d.	n.d.	n.d.	n.d.
Electron transport complex protein, RnfB	+	+	+	+	n.d.
Electron transport complex protein, RnfC	n.d.	n.d.	n.d.	n.d.	n.d.
Electron transport complex protein, RnfD	n.d.	n.d.	n.d.	n.d.	n.d.
Electron transport complex protein, RnfG	n.d.	n.d.	n.d.	n.d.	n.d.
Electron transport complex protein, RnfE	n.d.	n.d.	n.d.	n.d.	n.d.
Electron transport complex protein, RnfA	n.d.	n.d.	n.d.	n.d.	n.d.
Oxidoreductase FAD/NAD(P)-binding domain protein (EC:1.18.1.2)	n.d.	n.d.	n.d.	n.d.	n.d.
NADPH-dependent glutamate synthase GltA (1.4.1.13/1.4.1.14)	+	+	+	+	n.d.
Cytochrome bd quinol oxidase [£]	n.d.	n.d.	n.d.	n.d.	n.d.
Cytochrome *bd* quinol oxidase subunit I (cydA (EC:1.10.3.- -)	+	+	+	n.d.	n.d.
Cytochrome *bd* quinol oxidase subunit II (cydB (EC:1.10.3.- -)	+	+	+	+	n.d.
Electron Transport Flavoprotein [£]					
Electron transfer flavoprotein alpha subunit (Dalk_3215)	+	+	+	n.d.	+
Electron transfer flavoprotein alpha/beta subunit (Dalk_3216)	+	+	+	n.d.	+
Putative Fe-S Oxidoreductase (Dalk_3217)	+	+	+	n.d.	+
ETF-driven Reduced Ferredoxin Generation (Fix System) [§]					
Electron transfer protein beta subunit (SynarDRAFT_2457)	n.d.	+	+	n.d.	+
Electron transfer protein alpha subunit (SynarDRAFT_2458)	+	+	+	n.d.	+

Dehydrogenases (flavoprotein) (SynarDRAFT_2459)	n.d.	n.d.	n.d.	+	n.d.
Ferredoxin-like protein (SynarDRAFT_2460)	n.d.	n.d.	n.d.	+	n.d.
Electron-Confurcating Hydrogenases [§]					
NADH:ubiquinone oxidoreductase, 24 kD subunit (SynarDRAFT_0747)	n.d.	n.d.	n.d.	n.d.	n.d.
NADH:ubiquinone oxidoreductase, NADH-binding (SynarDRAFT_0748)	+	+	+	+	n.d.
Hydrogenase, Fe-only (SynarDRAFT_0749)	+	n.d.	+	+	n.d.
Ion-Translocating Fd:NADH oxidoreductase [§]					
Flavodoxin oxidoreductases (SynarDRAFT_0709)	+	+	+	+	+
Hypothetical protein (SynarDRAFT_0710)	+	+	+	+	n.d.
Coenzyme F_{420} hydrogenase/dehydrogenase, beta subunit (SynarDRAFT_0711)	+	+	+	+	n.d.
Coenzyme F_{420}-reducing hydrogenase, delta subunit (SynarDRAFT_0712)	+	+	+	+	n.d.
Heterodisulfide reductase, subunit A and related polyferredoxins (SynarDRAFT_0713)	+	+	+	+	n.d.
Heterodisulfide reductase, subunit B (SynarDRAFT_0714)	+	n.d.	n.d.	n.d.	n.d.
Hypothetical protein (SynarDRAFT_0715)	n.d.	n.d.	n.d.	n.d.	n.d.
Benzoyl-CoA Reductase Cassette [§]					
Coenzyme F_{420}-reducing hydrogenase, delta subunit (SynarDRAFT_0933)	+	+	+	+	n.d.
Pyridine nucleotide-disulphide oxidoreductase (SynarDRAFT_0934)	+	+	+	+	n.d.
Thi4 family (SynarDRAFT_0935)	+	+	+	+	n.d.
Fe-S oxidoreductase (SynarDRAFT_0936)	+	+	+	n.d.	n.d.
Fe-S-cluster-containing hydrogenase components 1 (SynarDRAFT_0937)	n.d.	n.d.	n.d.	n.d.	n.d.
Aldehyde:ferredoxin oxidoreductase (SynarDRAFT_0938)	n.d.	n.d.	n.d.	+	n.d.
NADH-quinone oxidoreductase, E subunit (SynarDRAFT_3490)	+	n.d.	n.d.	n.d.	n.d.
NADH:ubiquinone oxidoreductase, NADH-binding (SynarDRAFT_3491)	+	+	+	+	n.d.
NADH Dehydrogenase [†]					
NADH-ubiquinone/plastoquinone oxidoreductase (Sfum_0199)	n.d.	n.d.	n.d.	n.d.	n.d.
NADH-quinone oxidoreductase, B subunit (Sfum_0200)	n.d.	n.d.	n.d.	n.d.	n.d.
NADH dehydrogenase (ubiquinone), 30 kDa subunit (Sfum_0201)	n.d.	n.d.	n.d.	n.d.	n.d.
NADH dehydrogenase (ubiquinone) (Sfum_0202)	n.d.	n.d.	n.d.	n.d.	n.d.
NADH dehydrogenase I, D subunit (Sfum_1942)	n.d.	n.d.	n.d.	n.d.	n.d.

NADH-ubiquinone/plastoquinone oxidoreductase, chain 3 (Sfum_1943)	n.d.	n.d.	n.d.	n.d.	n.d.
Respiratory-chain NADH dehydrogenase, subunit 1 (Sfum_0203/1941)	n.d.	n.d.	n.d.	n.d.	n.d.
4Fe-4S ferredoxin, iron-sulfur binding domain protein (Sfum_0204)	n.d.	n.d.	n.d.	n.d.	n.d.
NADH-quinone oxidoreductase, chain I (Sfum_1940)	n.d.	+	n.d.	n.d.	n.d.
NADH-ubiquinone/plastoquinone oxidoreductase, chain 6 (Sfum_0205/1939)	n.d.	n.d.	n.d.	n.d.	n.d.
NADH-ubiquinone oxidoreductase, chain 4L (Sfum_0206/1938)	n.d.	+	n.d.	n.d.	n.d.
Proton-translocating NADH-quinone oxidoreductase, chain L (Sfum_0207/1937)	n.d.	n.d.	n.d.	n.d.	n.d.
Proton-translocating NADH-quinone oxidoreductase, chain M (Sfum_0208/1936)	n.d.	n.d.	n.d.	n.d.	n.d.
Proton-translocating NADH-quinone oxidoreductase, chain N (Sfum_0209/1935)	n.d.	n.d.	n.d.	n.d.	n.d.
V-type H(+)-translocating pyrophosphatase (Sfum_2995/3037)	n.d.	+	n.d.	n.d.	n.d.
F_0F_1 ATP Synthases ^					
ATP Synthase F0 Subunit A	+	+	+	n.d.	n.d.
ATP Synthase F0 Subunit B	+	+	+	n.d.	n.d.
ATP Synthase F0 Subunit C	+	+	+	n.d.	n.d.
ATP Synthase F1 alpha Chain	+	+	+	n.d.	n.d.
ATP Synthase F1 alpha Chain	+	+	+	n.d.	n.d.
ATP Synthase F1 alpha Chain	+	+	+	n.d.	n.d.
ATP Synthase F1 alpha Chain	+	+	+	n.d.	n.d.
ATP Synthase F1 alpha Chain	+	+	+	n.d.	n.d.
ATP Synthase F1 Protein I	n.d.	n.d.	n.d.	n.d.	n.d.
ATP Synthase F1 Protein I2	+	+	+	n.d.	n.d.
Protection from Oxygen Species ^					
Manganese superoxide dismutase SodA (1.15.1.1)	n.d.	n.d.	n.d.	n.d.	n.d.
Superoxide dismutase [Fe]sodB (1.15.1.1)	n.d.	n.d.	+	+	+
Superoxide dismutase [Cu-Zn] Precursor sodC (1.15.1.1)	n.d.	n.d.	n.d.	n.d.	n.d.
Catalase HPII (1.11.1.6)	+	+	+	n.d.	n.d.
Peroxidase HPI (1.11.1.7)	+	+	+	n.d.	n.d.
Cytochrome C551 Peroxidase (1.11.1.5)	n.d.	n.d.	n.d.	n.d.	n.d.

Miscellaneous Genome Features ^					
Genes for motility and chemotaxis	n.d.	n.d.	n.d.	n.d.	n.d.
Genes for production of flagellum	n.d.	n.d.	n.d.	n.d.	n.d.
Genes for social mobility and non-flagellar motility	n.d.	n.d.	n.d.	n.d.	n.d.
Quorum sensing	n.d.	n.d.	n.d.	n.d.	n.d.

ABC Transporters (only detected shown) #					
Tungstate TupA	+	+	+	+	+
Tungstate TupB	+	+	+	+	+
Tungstate TupC	+	+	+	+	+
Molybdate/Tungstate WtpA	+	n.d.	n.d.	n.d.	n.d.
Molybdate/Tungstate WtpB	+	n.d.	n.d.	n.d.	n.d.
Molybdate/Tungstate WtpC	+	n.d.	n.d.	n.d.	n.d.
Molybdate ModA	+	n.d.	+	+	n.d.
Molybdate ModB	+	n.d.	+	+	n.d.
Molybdate ModC	+	n.d.	+	+	n.d.
Molybdate ModF	n.d.	n.d.	n.d.	n.d.	n.d.
Phospholipid MlaC	+	+	+	n.d.	n.d.
Phospholipid MlaD	+	+	+	+	+
Phospholipid MlaE	+	+	+	+	+
Phospholipid MlaB	n.d.	n.d.	n.d.	n.d.	n.d.
Phospholipid MlaF	+	+	+	n.d.	+
Phosphate PstS	+	n.d.	+	+	+
Phosphate PstC	+	n.d.	+	+	+
Phosphate PstA	+	+	+	+	+
Phosphate PstB	+	+	+	+	+
Branched chain amino acid LivK	+	+	+	+	+
Branched chain amino acid LivH	+	+	+	+	+
Branched chain amino acid LivM	+	+	+	+	+

Branched chain amino acid LivG	+	+	+	+	+
Branched chain amino acid LivF	+	+	+	+	+
Iron complex FhuD	+	n.d.	+	+	n.d.
Iron complex FhuB	+	+	+	n.d.	+
Iron complex FhuC	n.d.	+	+	n.d.	+
Zinc ZnuA	+	+	+	+	+
Zinc ZnuB	+	+	+	+	+
Zinc ZnuC	+	n.d.	+	+	+
Cobalt CbiN	n.d.	n.d.	n.d.	n.d.	n.d.
Cobalt CbiM	n.d.	+	+	n.d.	+
Cobalt CbiQ	n.d.	+	+	n.d.	+
Cobalt CbiO	n.d.	+	+	n.d.	+
Nickel CbiK	n.d.	n.d.	n.d.	n.d.	n.d.
Nickel CbiN/L	n.d.	n.d.	n.d.	n.d.	n.d.
Nickel CbiM	n.d.	+	+	n.d.	+
Nickel CbiQ	n.d.	+	+	n.d.	+
Nickel CbiO	n.d.	+	+	n.d.	+
Biotin BioY	n.d.	n.d.	n.d.	n.d.	n.d.
Biotin BioN	n.d.	n.d.	n.d.	n.d.	n.d.
Biotin BioM	n.d.	n.d.	n.d.	n.d.	n.d.
Biotin EcfT	+	n.d.	+	n.d.	+
Biotin EcA1	+	n.d.	+	n.d.	+
Biotin EcA2	+	n.d.	+	n.d.	+
Lipo-polysaccharide RfbA	n.d.	+	+	n.d.	+
Lipo-polysaccharide RfbB	+	+	+	n.d.	+
Lipo-oligosaccharide NodJ	+	n.d.	+	n.d.	n.d.
Lipo-oligosaccharide NodI	+	n.d.	+	n.d.	n.d.
Lipoprotein LolC	+	+	+	n.d.	n.d.
Lipoprotein LolE	+	+	+	n.d.	n.d.
Lipoprotein LolD	+	+	+	n.d.	n.d.
Heme CcmD	n.d.	n.d.	n.d.	n.d.	n.d.

Heme CcmC	n.d.	n.d.	+	n.d.	n.d.
Heme CcmB	n.d.	n.d.	+	n.d.	n.d.
Heme CcmA	n.d.	n.d.	+	n.d.	n.d.
Lipo-polysaccharide LptF	+	+	+	+	+
Lipo-polysaccharide LptG	+	+	+	n.d.	+
Lipo-polysaccharide LptB	+	+	+	+	+

Table S7. Protein-coding genes predicted from the *"Smithella"* sp. SDB binned genome for which homologs are also found in the binned genomes of *"Smithella"* spp. SCADC, ME-1, F21, and D17. A total of 295 genes shared homology among all five partial genomes. Of these, 246 returned positive blastp matches when searched against NCBI's swissprot/uniprot database. Best blastp matches are shown. Proteins without blastp matches are omitted from the table.

Unitprot Sequence Identifier	Annotation
Q9I4V0|2NPD_PSEAE	Nitronate monooxygenase OS=Pseudomonas aeruginosa (strain ATCC 15692 / PAO1 / 1C / PRS 101 / LMG 12228) GN=PA1024
O07587|AAT3_BACSU	Putative aspartate aminotransferase YhdR OS=Bacillus subtilis (strain 168) GN=yhdR
P79274|ACADL_PIG	Long-chain specific acyl-CoA dehydrogenase, mitochondrial OS=Sus scrofa GN=ACADL
O29057|ACD1_ARCFU	Acetate--CoA ligase [ADP-forming] I OS=Archaeoglobus fulgidus (strain ATCC 49558 / VC-16 / DSM 4304 / JCM 9628 / NBRC 100126) GN=AF_1211
P45857|ACDB_BACSU	Acyl-CoA dehydrogenase OS=Bacillus subtilis (strain 168) GN=mmgC
O34421|ACDC_BACSU	Probable acyl-CoA dehydrogenase YngJ OS=Bacillus subtilis (strain 168) GN=yngJ
Q2LTJ7|ACPS_SYNAS	Holo-[acyl-carrier-protein] synthase OS=Syntrophus aciditrophicus (strain SB) GN=acpS
P33224|AIDB_ECOLI	Putative acyl-CoA dehydrogenase AidB OS=Escherichia coli (strain K12) GN=aidB
C3JXY0|AK_PSEFS	Aspartate kinase OS=Pseudomonas fluorescens (strain SBW25) GN=PFLU_4747
Q9K0V3|AMIC_NEIMB	N-acetylmuramoyl-L-alanine amidase AmiC OS=Neisseria meningitidis serogroup B (strain MC58) GN=amiC
D4GSY9|ANTRB_HALVD	Putative ABC transporter anion-binding protein HVO_1888 OS=Haloferax volcanii (strain ATCC 29605 / DSM 3757 / JCM 8879 / NBRC 14742 / NCIMB 2012 / VKM B-1768 / DS2) GN=HVO_1888
D4GSY8|ANTRP_HALVD	Probable anion ABC transporter permease protein HVO_1887 OS=Haloferax volcanii (strain ATCC 29605 / DSM 3757 / JCM 8879 / NBRC 14742 / NCIMB 2012 / VKM B-1768 / DS2) GN=HVO_1887
C6DAW5|ARNA_PECCP	Bifunctional polymyxin resistance protein ArnA OS=Pectobacterium carotovorum subsp. carotovorum (strain PC1) GN=arnA
A1JPN5|ARNA_YERE8	Bifunctional polymyxin resistance protein ArnA OS=Yersinia enterocolitica serotype O:8 / biotype 1B (strain NCTC 13174 / 8081) GN=arnA
Q57M57|ARNB_SALCH	UDP-4-amino-4-deoxy-L-arabinose--oxoglutarate aminotransferase OS=Salmonella choleraesuis (strain SC-B67) GN=arnB
Q93PD9|ARNC_YERPS	Undecaprenyl-phosphate 4-deoxy-4-formamido-L-arabinose transferase OS=Yersinia pseudotuberculosis serotype I (strain IP32953) GN=arnC
B1JJ31|ARND_YERPY	Probable 4-deoxy-4-formamido-L-arabinose-phosphoundecaprenol deformylase ArnD OS=Yersinia pseudotuberculosis serotype O:3 (strain YPIII) GN=arnD
Q48HY9|ARNT_PSE14	Undecaprenyl phosphate-alpha-4-amino-4-deoxy-L-arabinose arabinosyl transferase OS=Pseudomonas syringae pv. phaseolicola (strain 1448A / Race 6) GN=arnT

B2IYE4|AROA_NOSP7	3-phosphoshikimate 1-carboxyvinyltransferase OS=Nostoc punctiforme (strain ATCC 29133 / PCC 73102) GN=aroA
Q2LUC8|AROC_SYNAS	Chorismate synthase OS=Syntrophus aciditrophicus (strain SB) GN=aroC
Q2LUD6|AROK_SYNAS	Shikimate kinase OS=Syntrophus aciditrophicus (strain SB) GN=aroK
P45946|ARSB_BACSU	Arsenite resistance protein ArsB OS=Bacillus subtilis (strain 168) GN=arsB
Q06065|ATOC_ECOLI	Acetoacetate metabolism regulatory protein AtoC OS=Escherichia coli (strain K12) GN=atoC
Q06067|ATOS_ECOLI	Signal transduction histidine-protein kinase AtoS OS=Escherichia coli (strain K12) GN=atoS
Q2LQZ7|ATPA1_SYNAS	ATP synthase subunit alpha 1 OS=Syntrophus aciditrophicus (strain SB) GN=atpA1
Q72B03|BAMD_DESVH	Outer membrane protein assembly factor BamD OS=Desulfovibrio vulgaris (strain Hildenborough / ATCC 29579 / NCIMB 8303) GN=bamD
P21175|BRAC_PSEAE	Leucine-, isoleucine-, valine-, threonine-, and alanine-binding protein OS=Pseudomonas aeruginosa (strain ATCC 15692 / PAO1 / 1C / PRS 101 / LMG 12228) GN=braC
P21629|BRAF_PSEAE	High-affinity branched-chain amino acid transport ATP-binding protein BraF OS=Pseudomonas aeruginosa (strain ATCC 15692 / PAO1 / 1C / PRS 101 / LMG 12228) GN=braF
Q9XCD6|BTB7_MYCS2	Biotinylated protein TB7.3 homolog OS=Mycobacterium smegmatis (strain ATCC 700084 / mc(2)155) GN=MSMEG_1917
P38942|CAT2_CLOK5	4-hydroxybutyrate coenzyme A transferase OS=Clostridium kluyveri (strain ATCC 8527 / DSM 555 / NCIMB 10680) GN=cat2
Q8VPF2|CATJ_PSESB	3-oxoadipate CoA-transferase subunit B OS=Pseudomonas sp. (strain B13) GN=catJ
Q74FF1|CLPB_GEOSL	Chaperone protein ClpB OS=Geobacter sulfurreducens (strain ATCC 51573 / DSM 12127 / PCA) GN=clpB
Q39054|CNX1_ARATH	Molybdopterin biosynthesis protein CNX1 OS=Arabidopsis thaliana GN=CNX1
Q9KVD1|COABC_VIBCH	Coenzyme A biosynthesis bifunctional protein CoaBC OS=Vibrio cholerae serotype O1 (strain ATCC 39315 / El Tor Inaba N16961) GN=coaBC
P39695|COMEC_BACSU	ComE operon protein 3 OS=Bacillus subtilis (strain 168) GN=comEC
P52046|CRT_CLOAB	Short-chain-enoyl-CoA hydratase OS=Clostridium acetobutylicum (strain ATCC 824 / DSM 792 / JCM 1419 / LMG 5710 / VKM B-1787) GN=crt
P96349|CSP2_LACPL	Cold shock protein 2 OS=Lactobacillus plantarum (strain ATCC BAA-793 / NCIMB 8826 / WCFS1) GN=cspL
P05384|DBHB_PSEAE	DNA-binding protein HU-beta OS=Pseudomonas aeruginosa (strain ATCC 15692 / PAO1 / 1C / PRS 101 / LMG 12228) GN=hupB
P15263|DEGT_GEOSE	Pleiotropic regulatory protein OS=Geobacillus stearothermophilus GN=degT
Q39T85|DER_GEOMG	GTPase Der OS=Geobacter metallireducens (strain GS-15 / ATCC 53774 / DSM 7210) GN=der
P73248|DHPS_SYNY3	Dihydropteroate synthase OS=Synechocystis sp. (strain PCC 6803 / Kazusa) GN=folP
Q9SQT8|DHQSD_ARATH	Bifunctional 3-dehydroquinate dehydratase/shikimate dehydrogenase, chloroplastic OS=Arabidopsis thaliana GN=EMB3004
P59650|DHSL_COXBU	Deoxyhypusine synthase-like protein OS=Coxiella burnetii (strain RSA 493 / Nine Mile phase I) GN=CBU_0721
P37469|DNAC_BACSU	Replicative DNA helicase OS=Bacillus subtilis (strain 168) GN=dnaC

P33655|DNAG_CLOAB	DNA primase OS=Clostridium acetobutylicum (strain ATCC 824 / DSM 792 / JCM 1419 / LMG 5710 / VKM B-1787) GN=dnaG
Q2LUH6|DNAK_SYNAS	Chaperone protein DnaK OS=Syntrophus aciditrophicus (strain SB) GN=dnaK
Q9VIU7|DPM1_DROME	Probable dolichol-phosphate mannosyltransferase OS=Drosophila melanogaster GN=CG10166
O60762|DPM1_HUMAN	Dolichol-phosphate mannosyltransferase subunit 1 OS=Homo sapiens GN=DPM1
Q9XDH5|DPO3A_THEAQ	DNA polymerase III subunit alpha OS=Thermus aquaticus GN=dnaE
B8FBE8|DPO4_DESAA	DNA polymerase IV OS=Desulfatibacillum alkenivorans (strain AK-01) GN=dinB
Q2LTS7|DTD_SYNAS	D-aminoacyl-tRNA deacylase OS=Syntrophus aciditrophicus (strain SB) GN=dtd
Q185R8|DXR_PEPD6	1-deoxy-D-xylulose 5-phosphate reductoisomerase OS=Peptoclostridium difficile (strain 630) GN=dxr
Q7M9K2|ECTB_WOLSU	Diaminobutyrate--2-oxoglutarate transaminase OS=Wolinella succinogenes (strain ATCC 29543 / DSM 1740 / LMG 7466 / NCTC 11488 / FDC 602W) GN=ectB
Q2LUL6|EFG2_SYNAS	Elongation factor G 2 OS=Syntrophus aciditrophicus (strain SB) GN=fusA2
B8D0C1|EFG_HALOH	Elongation factor G OS=Halothermothrix orenii (strain H 168 / OCM 544 / DSM 9562) GN=fusA
Q2LTQ6|EFTS_SYNAS	Elongation factor Ts OS=Syntrophus aciditrophicus (strain SB) GN=tsf
Q2LQA3|EFTU_SYNAS	Elongation factor Tu OS=Syntrophus aciditrophicus (strain SB) GN=tuf
P0AB02|ELYC_ECO57	Envelope biogenesis factor ElyC OS=Escherichia coli O157:H7 GN=elyC
P71062|EPSL_BACSU	Uncharacterized sugar transferase EpsL OS=Bacillus subtilis (strain 168) GN=epsL
Q2LVR8|ERA_SYNAS	GTPase Era OS=Syntrophus aciditrophicus (strain SB) GN=era
P94551|ETFA_BACSU	Electron transfer flavoprotein subunit alpha OS=Bacillus subtilis (strain 168) GN=etfA
P51831|FABG_BACSU	3-Oxoacyl-[acyl-carrier-protein] reductase FabG OS=Bacillus subtilis (strain 168) GN=fabG
P9WQ36|FAC13_MYCTO	Long-chain-fatty-acid--CoA ligase FadD13 OS=Mycobacterium tuberculosis (strain CDC 1551 / Oshkosh) GN=fadD13
A6VVM8|FADA_MARMS	3-Ketoacyl-CoA thiolase OS=Marinomonas sp. (strain MWYL1) GN=fadA
O67150|FDHE_AQUAE	Protein FdhE homolog OS=Aquifex aeolicus (strain VF5) GN=fdhE
Q6T1W6|FDTB_ANETH	dTDP-3-amino-3,6-dideoxy-alpha-D-galactopyranose transaminase OS=Aneurinibacillus thermoaerophilus GN=fdtB
C1F8X6|FTSH_ACIC5	ATP-dependent zinc metalloprotease FtsH OS=Acidobacterium capsulatum (strain ATCC 51196 / DSM 11244 / JCM 7670) GN=ftsH
Q9K9T7|FTSZ_BACHD	Cell division protein FtsZ OS=Bacillus halodurans (strain ATCC BAA-125 / DSM 18197 / FERM 7344 / JCM 9153 / C-125) GN=ftsZ
Q4MQ58|G3P1_BACCE	Glyceraldehyde-3-phosphate dehydrogenase 1 OS=Bacillus cereus GN=gap1
Q2LXM9|GATA_SYNAS	Glutamyl-tRNA(Gln) amidotransferase subunit A OS=Syntrophus aciditrophicus (strain SB) GN=gatA
Q2LXN0|GATB2_SYNAS	Aspartyl/glutamyl-tRNA(Asn/Gln) amidotransferase subunit B 2 OS=Syntrophus aciditrophicus (strain SB) GN=gatB2
Q2LXN4|GATC_SYNAS	Aspartyl/glutamyl-tRNA(Asn/Gln) amidotransferase subunit C OS=Syntrophus aciditrophicus (strain SB) GN=gatC

Q9K786|GCSH_BACHD	Glycine cleavage system H protein OS=Bacillus halodurans (strain ATCC BAA-125 / DSM 18197 / FERM 7344 / JCM 9153 / C-125) GN=gcvH
Q2LRC1|GLMM_SYNAS	Phosphoglucosamine mutase OS=Syntrophus aciditrophicus (strain SB) GN=glmM
O26760|GLNB2_METTH	Nitrogen regulatory protein P-II 2 OS=Methanothermobacter thermautotrophicus (strain ATCC 29096 / DSM 1053 / JCM 10044 / NBRC 100330 / Delta H) GN=MTH_664
B5YJT7|GMHA_THEYD	Phosphoheptose isomerase OS=Thermodesulfovibrio yellowstonii (strain ATCC 51303 / DSM 11347 / YP87) GN=gmhA
P77851|HBD_THETC	3-hydroxybutyryl-CoA dehydrogenase OS=Thermoanaerobacterium thermosaccharolyticum (strain ATCC 7956 / DSM 571 / NCIB 9385 / NCA 3814) GN=hbd
Q3J7H2|HIS82_NITOC	Histidinol-phosphate aminotransferase 2 OS=Nitrosococcus oceani (strain ATCC 19707 / NCIMB 11848) GN=hisC2
B8FB71|HLDE_DESAA	Bifunctional protein HldE OS=Desulfatibacillum alkenivorans (strain AK-01) GN=hldE
A4YI89|HPCD_METS5	3-hydroxypropionyl-coenzyme A dehydratase OS=Metallosphaera sedula (strain ATCC 51363 / DSM 5348) GN=Msed_2001
Q8KDT8|HPPA_CHLTE	K(+)-insensitive pyrophosphate-energized proton pump OS=Chlorobium tepidum (strain ATCC 49652 / DSM 12025 / TLS) GN=hppA
O66821|HPRT_AQUAE	Hypoxanthine-guanine phosphoribosyltransferase OS=Aquifex aeolicus (strain VF5) GN=hpt
Q4UJB1|HSPC4_RICFE	Small heat shock protein C4 OS=Rickettsia felis (strain ATCC VR-1525 / URRWXCal2) GN=hspc4-1
Q8U2E4|HYD1G_PYRFU	Sulfhydrogenase 1 subunit gamma OS=Pyrococcus furiosus (strain ATCC 43587 / DSM 3638 / JCM 8422 / Vc1) GN=hydG
Q3Z6L1|HYPA_DEHM1	Probable hydrogenase nickel incorporation protein HypA OS=Dehalococcoides mccartyi (strain ATCC BAA-2266 / KCTC 15142 / 195) GN=hypA
P0AAN4|HYPB_ECOL6	Hydrogenase isoenzymes nickel incorporation protein HypB OS=Escherichia coli O6:H1 (strain CFT073 / ATCC 700928 / UPEC) GN=hypB
Q2LQB5|IF1_SYNAS	Translation initiation factor IF-1 OS=Syntrophus aciditrophicus (strain SB) GN=infA
A7HBJ3|IHFA_ANADF	Integration host factor subunit alpha OS=Anaeromyxobacter sp. (strain Fw109-5) GN=ihfA
Q11C35|IHFB_CHESB	Integration host factor subunit beta OS=Chelativorans sp. (strain BNC1) GN=ihfB
P38674|ILV5_NEUCR	Ketol-acid reductoisomerase, mitochondrial OS=Neurospora crassa (strain ATCC 24698 / 74-OR23-1A / CBS 708.71 / DSM 1257 / FGSC 987) GN=ilv-2
O67820|IMDH_AQUAE	Inosine-5'-monophosphate dehydrogenase OS=Aquifex aeolicus (strain VF5) GN=guaB
Q2LUJ9|ISPE_SYNAS	4-diphosphocytidyl-2-C-methyl-D-erythritol kinase OS=Syntrophus aciditrophicus (strain SB) GN=ispE
Q9RCA6|JAG_BACHD	Protein jag OS=Bacillus halodurans (strain ATCC BAA-125 / DSM 18197 / FERM 7344 / JCM 9153 / C-125) GN=jag
Q2LQ62|KGUA_SYNAS	Guanylate kinase OS=Syntrophus aciditrophicus (strain SB) GN=gmk
O31661|KINE_BACSU	Sporulation kinase E OS=Bacillus subtilis (strain 168) GN=kinE
Q6AJL7|KPRS_DESPS	Ribose-phosphate pyrophosphokinase OS=Desulfotalea psychrophila (strain LSv54 / DSM 12343) GN=prs
P94547|LCFA_BACSU	Long-chain-fatty-acid--CoA ligase OS=Bacillus subtilis (strain 168) GN=lcfA

P44446|LCFH_HAEIN	Putative long-chain-fatty-acid--CoA ligase OS=Haemophilus influenzae (strain ATCC 51907 / DSM 11121 / KW20 / Rd) GN=HI_0002
Q2LTN3|LEPA_SYNAS	Elongation factor 4 OS=Syntrophus aciditrophicus (strain SB) GN=lepA
Q8FWP6|LIVB7_BRUSU	Leu/Ile/Val-binding protein homolog 7 OS=Brucella suis biovar 1 (strain 1330) GN=BRA0400
Q39T32|LNT_GEOMG	Apolipoprotein N-acyltransferase OS=Geobacter metallireducens (strain GS-15 / ATCC 53774 / DSM 7210) GN=lnt
A0LEE9|LON1_SYNFM	Lon protease 1 OS=Syntrophobacter fumaroxidans (strain DSM 10017 / MPOB) GN=lon1
P45073|LPTB_HAEIN	Lipopolysaccharide export system ATP-binding protein LptB OS=Haemophilus influenzae (strain ATCC 51907 / DSM 11121 / KW20 / Rd) GN=lptB
Q2LTI8|MIAA1_SYNAS	tRNA dimethylallyltransferase 1 OS=Syntrophus aciditrophicus (strain SB) GN=miaA1
P57531|MLTA_BUCAI	Membrane-bound lytic murein transglycosylase A homolog OS=Buchnera aphidicola subsp. Acyrthosiphon pisum (strain APS) GN=mltA
P0AEZ8|MLTD_ECOL6	Membrane-bound lytic murein transglycosylase D OS=Escherichia coli O6:H1 (strain CFT073 / ATCC 700928 / UPEC) GN=mltD
Q2LSF6|MNME_SYNAS	tRNA modification GTPase MnmE OS=Syntrophus aciditrophicus (strain SB) GN=mnmE
Q92AG0|MNTC_LISIN	Manganese transport system membrane protein MntC OS=Listeria innocua serovar 6a (strain CLIP 11262) GN=mntC
Q2LTB7|MTAD1_SYNAS	5-methylthioadenosine/S-adenosylhomocysteine deaminase 1 OS=Syntrophus aciditrophicus (strain SB) GN=mtaD1
Q2LXN1|MTNA_SYNAS	Methylthioribose-1-phosphate isomerase OS=Syntrophus aciditrophicus (strain SB) GN=mtnA
P0A5L7|NADE_MYCBO	Glutamine-dependent NAD(+) synthetase OS=Mycobacterium bovis (strain ATCC BAA-935 / AF2122/97) GN=nadE
B9M5P5|NADK_GEODF	NAD kinase OS=Geobacter daltonii (strain DSM 22248 / JCM 15807 / FRC-32) GN=nadK
P32382|NADO_THEBR	NADH oxidase OS=Thermoanaerobacter brockii
O30124|NPD2_ARCFU	NAD-dependent protein deacylase 2 OS=Archaeoglobus fulgidus (strain ATCC 49558 / VC-16 / DSM 4304 / JCM 9628 / NBRC 100126) GN=cobB2
Q04849|NTRX_AZOC5	Nitrogen assimilation regulatory protein NtrX OS=Azorhizobium caulinodans (strain ATCC 43989 / DSM 5975 / ORS 571) GN=ntrX
Q04850|NTRY_AZOC5	Nitrogen regulation protein NtrY OS=Azorhizobium caulinodans (strain ATCC 43989 / DSM 5975 / ORS 571) GN=ntrY
Q9HWC4|NUSG_PSEAE	Transcription termination/antitermination protein NusG OS=Pseudomonas aeruginosa (strain ATCC 15692 / PAO1 / 1C / PRS 101 / LMG 12228) GN=nusG
Q2LR77|OBG_SYNAS	GTPase Obg OS=Syntrophus aciditrophicus (strain SB) GN=obg
O26715|OGT_METTH	Methylated-DNA--protein-cysteine methyltransferase OS=Methanothermobacter thermautotrophicus (strain ATCC 29096 / DSM 1053 / JCM 10044 / NBRC 100330 / Delta H) GN=ogt
P77181|PAAY_ECOLI	Phenylacetic acid degradation protein PaaY OS=Escherichia coli (strain K12) GN=paaY
Q8NVT1|PCRA_STAAW	ATP-dependent DNA helicase PcrA OS=Staphylococcus aureus (strain MW2) GN=pcrA
P0A5S1|PHOL_MYCBO	PhoH-like protein OS=Mycobacterium bovis (strain ATCC BAA-935 / AF2122/97) GN=Mb2389c

P23545|PHOR_BACSU	Alkaline phosphatase synthesis sensor protein PhoR OS=Bacillus subtilis (strain 168) GN=phoR
P22608|PILB_PSEAE	Type 4 fimbrial assembly protein PilB OS=Pseudomonas aeruginosa (strain ATCC 15692 / PAO1 / 1C / PRS 101 / LMG 12228) GN=pilB
P22609|PILC_PSEAE	Type 4 fimbrial assembly protein PilC OS=Pseudomonas aeruginosa (strain ATCC 15692 / PAO1 / 1C / PRS 101 / LMG 12228) GN=pilC
O28486|PNPH_ARCFU	Probable 6-oxopurine nucleoside phosphorylase OS=Archaeoglobus fulgidus (strain ATCC 49558 / VC-16 / DSM 4304 / JCM 9628 / NBRC 100126) GN=AF_1788
P13794|PORF_PSEAE	Outer membrane porin F OS=Pseudomonas aeruginosa (strain ATCC 15692 / PAO1 / 1C / PRS 101 / LMG 12228) GN=oprF
Q42910|PPDK_MESCR	Pyruvate, phosphate dikinase, chloroplastic OS=Mesembryanthemum crystallinum GN=PPD
O66105|PPIB_TREPA	Probable peptidyl-prolyl cis-trans isomerase OS=Treponema pallidum (strain Nichols) GN=ppiB
C3KW94|PRSA_CLOB6	Foldase protein PrsA OS=Clostridium botulinum (strain 657 / Type Ba4) GN=prsA
Q2LTG0|PSTB_SYNAS	Phosphate import ATP-binding protein PstB OS=Syntrophus aciditrophicus (strain SB) GN=pstB
O83018|PT1_BACSI	Phosphoenolpyruvate-protein phosphotransferase OS=Bacillus sp. (strain S) GN=ptsI
Q8Q0J2|PUR2_METMA	Phosphoribosylamine--glycine ligase OS=Methanosarcina mazei (strain ATCC BAA-159 / DSM 3647 / Goe1 / Go1 / JCM 11833 / OCM 88) GN=purD
Q55135|PURU_SYNY3	Formyltetrahydrofolate deformylase OS=Synechocystis sp. (strain PCC 6803 / Kazusa) GN=purU
Q2LQ82|PYRF_SYNAS	Orotidine 5'-phosphate decarboxylase OS=Syntrophus aciditrophicus (strain SB) GN=pyrF
Q2LRK2|PYRG_SYNAS	CTP synthase OS=Syntrophus aciditrophicus (strain SB) GN=pyrG
O29809|QUED_ARCFU	Putative 6-carboxy-5,6,7,8-tetrahydropterin synthase OS=Archaeoglobus fulgidus (strain ATCC 49558 / VC-16 / DSM 4304 / JCM 9628 / NBRC 100126) GN=queD
Q5XAQ7|RAFY_STRP6	Ribosome-associated factor Y OS=Streptococcus pyogenes serotype M6 (strain ATCC BAA-946 / MGAS10394) GN=M6_Spy1371
Q9K974|RECN_BACHD	DNA repair protein RecN OS=Bacillus halodurans (strain ATCC BAA-125 / DSM 18197 / FERM 7344 / JCM 9153 / C-125) GN=recN
A5IM04|RF2_THEP1	Peptide chain release factor 2 OS=Thermotoga petrophila (strain RKU-1 / ATCC BAA-488 / DSM 13995) GN=prfB
P27833|RFFA_ECOLI	dTDP-4-amino-4,6-dideoxygalactose transaminase OS=Escherichia coli (strain K12) GN=rffA
Q2LQ89|RL10_SYNAS	50S ribosomal protein L10 OS=Syntrophus aciditrophicus (strain SB) GN=rplJ
Q2LQ91|RL11_SYNAS	50S ribosomal protein L11 OS=Syntrophus aciditrophicus (strain SB) GN=rplK
Q2LQB0|RL14_SYNAS	50S ribosomal protein L14 OS=Syntrophus aciditrophicus (strain SB) GN=rplN
Q2LQB3|RL16_SYNAS	50S ribosomal protein L16 OS=Syntrophus aciditrophicus (strain SB) GN=rplP
Q2LQC1|RL18_SYNAS	50S ribosomal protein L18 OS=Syntrophus aciditrophicus (strain SB) GN=rplR
Q2LQ90|RL1_SYNAS	50S ribosomal protein L1 OS=Syntrophus aciditrophicus (strain SB) GN=rplA
Q2LR75|RL21_SYNAS	50S ribosomal protein L21 OS=Syntrophus aciditrophicus (strain SB) GN=rplU

Q2LQ96|RL22_SYNAS	50S ribosomal protein L22 OS=Syntrophus aciditrophicus (strain SB) GN=rplV
Q2LQA9|RL24_SYNAS	50S ribosomal protein L24 OS=Syntrophus aciditrophicus (strain SB) GN=rplX
Q2LUK6|RL25_SYNAS	50S ribosomal protein L25 OS=Syntrophus aciditrophicus (strain SB) GN=rplY
Q2LR76|RL27_SYNAS	50S ribosomal protein L27 OS=Syntrophus aciditrophicus (strain SB) GN=rpmA
Q2LQB2|RL29_SYNAS	50S ribosomal protein L29 OS=Syntrophus aciditrophicus (strain SB) GN=rpmC
Q2LQA0|RL4_SYNAS	50S ribosomal protein L4 OS=Syntrophus aciditrophicus (strain SB) GN=rplD
Q2LQA8|RL5_SYNAS	50S ribosomal protein L5 OS=Syntrophus aciditrophicus (strain SB) GN=rplE
Q2LQ88|RL7_SYNAS	50S ribosomal protein L7/L12 OS=Syntrophus aciditrophicus (strain SB) GN=rplL
Q2LUJ5|RL9_SYNAS	50S ribosomal protein L9 OS=Syntrophus aciditrophicus (strain SB) GN=rplI
P55294|RMLB_NEIMB	dTDP-glucose 4,6-dehydratase OS=Neisseria meningitidis serogroup B (strain MC58) GN=rfbB1
A0LGM1|RNC_SYNFM	Ribonuclease 3 OS=Syntrophobacter fumaroxidans (strain DSM 10017 / MPOB) GN=rnc
Q2LRA0|RNY_SYNAS	Ribonuclease Y OS=Syntrophus aciditrophicus (strain SB) GN=rny
Q9F0J6|ROO_DESGI	Rubredoxin-oxygen oxidoreductase OS=Desulfovibrio gigas GN=roo
P0A171|RP54_PSEPK	RNA polymerase sigma-54 factor OS=Pseudomonas putida (strain KT2440) GN=rpoN
Q2LQ87|RPOB_SYNAS	DNA-directed RNA polymerase subunit beta OS=Syntrophus aciditrophicus (strain SB) GN=rpoB
Q2LQ86|RPOC_SYNAS	DNA-directed RNA polymerase subunit beta' OS=Syntrophus aciditrophicus (strain SB) GN=rpoC
Q2LQ83|RPOZ_SYNAS	DNA-directed RNA polymerase subunit omega OS=Syntrophus aciditrophicus (strain SB) GN=rpoZ
Q01624|RPSB_STIAD	RNA polymerase sigma-B factor OS=Stigmatella aurantiaca (strain DW4/3-1) GN=sigB
Q2LTQ4|RRF_SYNAS	Ribosome-recycling factor OS=Syntrophus aciditrophicus (strain SB) GN=frr
Q2LQA2|RS10_SYNAS	30S ribosomal protein S10 OS=Syntrophus aciditrophicus (strain SB) GN=rpsJ
Q2LQ85|RS12_SYNAS	30S ribosomal protein S12 OS=Syntrophus aciditrophicus (strain SB) GN=rpsL
Q2LUJ7|RS18_SYNAS	30S ribosomal protein S18 OS=Syntrophus aciditrophicus (strain SB) GN=rpsR
Q2LQ97|RS19_SYNAS	30S ribosomal protein S19 OS=Syntrophus aciditrophicus (strain SB) GN=rpsS
Q2LQK7|RS20_SYNAS	30S ribosomal protein S20 OS=Syntrophus aciditrophicus (strain SB) GN=rpsT
Q2LTQ7|RS2_SYNAS	30S ribosomal protein S2 OS=Syntrophus aciditrophicus (strain SB) GN=rpsB
Q2LQ95|RS3_SYNAS	30S ribosomal protein S3 OS=Syntrophus aciditrophicus (strain SB) GN=rpsC
A8ZRQ3|RS6_DESOH	30S ribosomal protein S6 OS=Desulfococcus oleovorans (strain DSM 6200 / Hxd3) GN=rpsF
Q2LQA4|RS7_SYNAS	30S ribosomal protein S7 OS=Syntrophus aciditrophicus (strain SB) GN=rpsG
Q2LQA6|RS8_SYNAS	30S ribosomal protein S8 OS=Syntrophus aciditrophicus (strain SB) GN=rpsH

Q97FZ9|RUBY1_CLOAB	Rubrerythrin-1 OS=Clostridium acetobutylicum (strain ATCC 824 / DSM 792 / JCM 1419 / LMG 5710 / VKM B-1787) GN=rbr1
A7Z671|SCPA_BACA2	Segregation and condensation protein A OS=Bacillus amyloliquefaciens subsp. plantarum (strain DSM 23117 / BGSC 10A6 / FZB42) GN=scpA
Q2LTP4|SECA_SYNAS	Protein translocase subunit SecA OS=Syntrophus aciditrophicus (strain SB) GN=secA
A1APP9|SELD_PELPD	Selenide, water dikinase OS=Pelobacter propionicus (strain DSM 2379) GN=selD
P39663|SPHR_SYNE7	Alkaline phosphatase synthesis transcriptional regulatory protein SphR OS=Synechococcus elongatus (strain PCC 7942) GN=sphR
O59179|STOPP_PYRHO	Membrane-bound protease PH1510 OS=Pyrococcus horikoshii (strain ATCC 700860 / DSM 12428 / JCM 9974 / NBRC 100139 / OT-3) GN=PH1510
Q9UYB2|SUA5_PYRAB	Threonylcarbamoyl-AMP synthase OS=Pyrococcus abyssi (strain GE5 / Orsay) GN=sua5
Q2LUH7|SURE_SYNAS	5'-nucleotidase SurE OS=Syntrophus aciditrophicus (strain SB) GN=surE
Q2LPL7|SYA_SYNAS	Alanine--tRNA ligase OS=Syntrophus aciditrophicus (strain SB) GN=alaS
Q8PVQ1|SYC_METMA	Cysteine--tRNA ligase OS=Methanosarcina mazei (strain ATCC BAA-159 / DSM 3647 / Goe1 / Go1 / JCM 11833 / OCM 88) GN=cysS
Q2LTE0|SYDN2_SYNAS	Aspartate--tRNA(Asp/Asn) ligase 2 OS=Syntrophus aciditrophicus (strain SB) GN=aspS2
Q2LR26|SYFB_SYNAS	Phenylalanine--tRNA ligase beta subunit OS=Syntrophus aciditrophicus (strain SB) GN=pheT
Q2LVI0|SYGA_SYNAS	Glycine--tRNA ligase alpha subunit OS=Syntrophus aciditrophicus (strain SB) GN=glyQ
Q2LVI9|SYGB_SYNAS	Glycine--tRNA ligase beta subunit OS=Syntrophus aciditrophicus (strain SB) GN=glyS
A8ZUR6|SYH_DESOH	Histidine--tRNA ligase OS=Desulfococcus oleovorans (strain DSM 6200 / Hxd3) GN=hisS
Q9WYW2|SYW_THEMA	Tryptophan--tRNA ligase OS=Thermotoga maritima (strain ATCC 43589 / MSB8 / DSM 3109 / JCM 10099) GN=trpS
Q2LR98|SYY_SYNAS	Tyrosine--tRNA ligase OS=Syntrophus aciditrophicus (strain SB) GN=tyrS
P12243|THIO1_SYNE7	Thioredoxin-1 OS=Synechococcus elongatus (strain PCC 7942) GN=trxA
O26981|THIRX_METTH	Thioredoxin OS=Methanothermobacter thermautotrophicus (strain ATCC 29096 / DSM 1053 / JCM 10044 / NBRC 100330 / Delta H) GN=MTH_895
P45359|THLA_CLOAB	Acetyl-CoA acetyltransferase OS=Clostridium acetobutylicum (strain ATCC 824 / DSM 792 / JCM 1419 / LMG 5710 / VKM B-1787) GN=thlA
Q29RZ5|TNG2_BOVIN	Transport and Golgi organization protein 2 homolog OS=Bos taurus GN=TANGO2
O83845|TSAE_TREPA	tRNA threonylcarbamoyladenosine biosynthesis protein TsaE OS=Treponema pallidum (strain Nichols) GN=tsaE
Q2LR71|UVRC_SYNAS	UvrABC system protein C OS=Syntrophus aciditrophicus (strain SB) GN=uvrC
Q9XC90|WAAE_KLEPN	Lipopolysaccharide core biosynthesis glycosyltransferase WaaE OS=Klebsiella pneumoniae GN=waaE
P9WMY2|WBBL_MYCTO	N-acetylglucosaminyl-diphospho-decaprenol L-rhamnosyltransferase OS=Mycobacterium tuberculosis (strain CDC 1551 / Oshkosh) GN=wbbL

Q8U4K3|WTPC_PYRFU	Molybdate/tungstate import ATP-binding protein WtpC OS=Pyrococcus furiosus (strain ATCC 43587 / DSM 3638 / JCM 8422 / Vc1) GN=wtpC
Q58408|Y1002_METJA	Uncharacterized protein MJ1002 OS=Methanocaldococcus jannaschii (strain ATCC 43067 / DSM 2661 / JAL-1 / JCM 10045 / NBRC 100440) GN=MJ1002
Q58466|Y1066_METJA	Uncharacterized protein MJ1066 OS=Methanocaldococcus jannaschii (strain ATCC 43067 / DSM 2661 / JAL-1 / JCM 10045 / NBRC 100440) GN=MJ1066
O27262|Y1194_METTH	Uncharacterized protein MTH_1194 OS=Methanothermobacter thermautotrophicus (strain ATCC 29096 / DSM 1053 / JCM 10044 / NBRC 100330 / Delta H) GN=MTH_1194
Q2LSN7|Y1217_SYNAS	UPF0042 nucleotide-binding protein SYNAS_12170 OS=Syntrophus aciditrophicus (strain SB) GN=SYNAS_12170
P73475|Y1230_SYNY3	Universal stress protein Slr1230 OS=Synechocystis sp. (strain PCC 6803 / Kazusa) GN=slr1230
O30107|Y130_ARCFU	Uncharacterized protein AF_0130 OS=Archaeoglobus fulgidus (strain ATCC 49558 / VC-16 / DSM 4304 / JCM 9628 / NBRC 100126) GN=AF_0130
O28938|Y1331_ARCFU	Uncharacterized protein AF_1331 OS=Archaeoglobus fulgidus (strain ATCC 49558 / VC-16 / DSM 4304 / JCM 9628 / NBRC 100126) GN=AF_1331
Q6LXF3|Y1398_METMP	Uncharacterized metallohydrolase MMP1398 OS=Methanococcus maripaludis (strain S2 / LL) GN=MMP1398
O28852|Y1420_ARCFU	Uncharacterized protein AF_1420 OS=Archaeoglobus fulgidus (strain ATCC 49558 / VC-16 / DSM 4304 / JCM 9628 / NBRC 100126) GN=AF_1420
A5UN65|Y1438_METS3	MEMO1 family protein Msm_1438 OS=Methanobrevibacter smithii (strain PS / ATCC 35061 / DSM 861) GN=Msm_1438
O67444|Y1464_AQUAE	Uncharacterized RNA pseudouridine synthase aq_1464 OS=Aquifex aeolicus (strain VF5) GN=aq_1464
Q58939|Y1544_METJA	Uncharacterized protein MJ1544 OS=Methanocaldococcus jannaschii (strain ATCC 43067 / DSM 2661 / JAL-1 / JCM 10045 / NBRC 100440) GN=MJ1544
Q59027|Y1633_METJA	Uncharacterized protein MJ1633 OS=Methanocaldococcus jannaschii (strain ATCC 43067 / DSM 2661 / JAL-1 / JCM 10045 / NBRC 100440) GN=MJ1633
P64924|Y2031_MYCBO	Uncharacterized protein Mb2031c OS=Mycobacterium bovis (strain ATCC BAA-935 / AF2122/97) GN=Mb2031c
P59599|Y2212_CHLTE	Uncharacterized N-acetyltransferase CT2212 OS=Chlorobium tepidum (strain ATCC 49652 / DSM 12025 / TLS) GN=CT2212
Q9LK65|Y3307_ARATH	Putative DUF21 domain-containing protein At3g13070, chloroplastic OS=Arabidopsis thaliana GN=CBSDUFCH1
Q9ZD73|Y471_RICPR	Uncharacterized protein RP471 OS=Rickettsia prowazekii (strain Madrid E) GN=RP471
P55680|Y4WB_RHISN	Uncharacterized zinc protease-like protein y4wB OS=Rhizobium sp. (strain NGR234) GN=NGR_a01030
Q58029|Y612_METJA	Probable arogenate/prephenate dehydrogenase OS=Methanocaldococcus jannaschii (strain ATCC 43067 / DSM 2661 / JAL-1 / JCM 10045 / NBRC 100440) GN=MJ0612
O26884|Y793_METTH	Uncharacterized protein MTH_793 OS=Methanothermobacter thermautotrophicus (strain ATCC 29096 / DSM 1053 / JCM 10044 / NBRC 100330 / Delta H) GN=MTH_793
O26945|Y857_METTH	Protein MTH_857 OS=Methanothermobacter thermautotrophicus (strain ATCC 29096 / DSM 1053 / JCM 10044 / NBRC 100330 / Delta H) GN=MTH_857
Q8D4U6|Y8J3_VIBVU	Uncharacterized response regulatory protein VV2_1193 OS=Vibrio vulnificus (strain CMCP6) GN=VV2_1193

Q58376|Y966_METJA	Uncharacterized protein MJ0966 OS=Methanocaldococcus jannaschii (strain ATCC 43067 / DSM 2661 / JAL-1 / JCM 10045 / NBRC 100440) GN=MJ0966
Q06753|YACO_BACSU	Putative TrmH family tRNA/rRNA methyltransferase YacO OS=Bacillus subtilis (strain 168) GN=yacO
Q2LTL4|YBEY_SYNAS	Endoribonuclease YbeY OS=Syntrophus aciditrophicus (strain SB) GN=ybeY
P37518|YCHF_BACSU	Ribosome-binding ATPase YchF OS=Bacillus subtilis (strain 168) GN=ychF
O07585|YHDP_BACSU	UPF0053 protein YhdP OS=Bacillus subtilis (strain 168) GN=yhdP
O07592|YHDW_BACSU	Putative glycerophosphoryl diester phosphodiesterase YhdW OS=Bacillus subtilis (strain 168) GN=yhdW
P37665|YIAD_ECOLI	Probable lipoprotein YiaD OS=Escherichia coli (strain K12) GN=yiaD
Q2LSF9|YIDC_SYNAS	Membrane protein insertase YidC OS=Syntrophus aciditrophicus (strain SB) GN=yidC
O06745|YITJ_BACSU	Bifunctional homocysteine S-methyltransferase/5,10-methylenetetrahydrofolate reductase OS=Bacillus subtilis (strain 168) GN=yitJ
O31608|YJBJ_BACSU	Putative murein lytic transglycosylase YjbJ OS=Bacillus subtilis (strain 168) GN=yjbJ
O34441|YLOC_BACSU	UPF0701 protein YloC OS=Bacillus subtilis (strain 168) GN=yloC
O31775|YMDB_BACSU	Uncharacterized protein ymdB OS=Bacillus subtilis (strain 168) GN=ymdB
P0AA93|YPDA_ECOLI	Sensor histidine kinase YpdA OS=Escherichia coli (strain K12) GN=ypdA
P54464|YQEY_BACSU	Uncharacterized protein YqeY OS=Bacillus subtilis (strain 168) GN=yqeY
P46344|YQFF_BACSU	Uncharacterized protein YqfF OS=Bacillus subtilis (strain 168) GN=yqfF
P46340|YQGI_BACSU	Probable ABC transporter permease protein YqgI OS=Bacillus subtilis (strain 168) GN=yqgI
P54515|YQHQ_BACSU	Uncharacterized protein YqhQ OS=Bacillus subtilis (strain 168) GN=yqhQ
O34966|ZNUA_BACSU	High-affinity zinc uptake system binding-protein ZnuA OS=Bacillus subtilis (strain 168) GN=znuA
Q9APD9|ZRAR_KLEOX	Transcriptional regulatory protein ZraR OS=Klebsiella oxytoca GN=zraR
Q8Z332|ZRAS_SALTI	Sensor protein ZraS OS=Salmonella typhi GN=zraS
Q926D8|ZURA_LISIN	Zinc uptake system ATP-binding protein ZurA OS=Listeria innocua serovar 6a (strain CLIP 11262) GN=zurA

Table S8. Protein-coding genes predicted from the *"Smithella"* sp. SDB binned genome for which homologs are also found in the binned genomes of *"Smithella"* spp. SCADC, ME-1, and D17, in addition to those found in Table S7. The table herein does not include *"Smithella"* sp. F21 because the draft genome does not appear to contain genes for known alkane activation pathways, and it is unclear whether the requisite organism has alkane-utilizing capacity. A total of 358 genes shared homology among the four partial genomes (in addition to those in Table S7). Of these, 253 retuned positive blastp matches when searched against NCBI's swissprot/uniprot database. Best blastp matches are shown. Proteins without blastp matches are omitted from the table.

Unitprot Sequence Identifier	Annotation
P53001|AAT1_BACSU	Aspartate aminotransferase OS=Bacillus subtilis (strain 168) GN=aspB
O29057|ACD1_ARCFU	Acetate--CoA ligase [ADP-forming] I OS=Archaeoglobus fulgidus (strain ATCC 49558 / VC-16 / DSM 4304 / JCM 9628 / NBRC 100126) GN=AF_1211
P39533|ACON2_YEAST	Homocitrate dehydratase, mitochondrial OS=Saccharomyces cerevisiae (strain ATCC 204508 / S288c) GN=ACO2
Q2LTM9|ACYP_SYNAS	Acylphosphatase OS=Syntrophus aciditrophicus (strain SB) GN=acyP
O28608|ALADH_ARCFU	Alanine dehydrogenase OS=Archaeoglobus fulgidus (strain ATCC 49558 / VC-16 / DSM 4304 / JCM 9628 / NBRC 100126) GN=ala
Q46629|AMSH_ERWAM	Amylovoran export outer membrane protein AmsH OS=Erwinia amylovora GN=amsH
P9WMK8|AP4A_MYCTO	AP-4-A phosphorylase OS=Mycobacterium tuberculosis (strain CDC 1551 / Oshkosh) GN=MT2688
Q2LT95|ARGB_SYNAS	Acetylglutamate kinase OS=Syntrophus aciditrophicus (strain SB) GN=argB
Q2LT96|ARLY_SYNAS	Argininosuccinate lyase OS=Syntrophus aciditrophicus (strain SB) GN=argH
A8ZNS1|ATP62_ACAM1	ATP synthase subunit a 2 OS=Acaryochloris marina (strain MBIC 11017) GN=atpB2
Q2LY34|ATPA2_SYNAS	ATP synthase subunit alpha 2 OS=Syntrophus aciditrophicus (strain SB) GN=atpA2
A6VWP9|ATPB1_MARMS	ATP synthase subunit beta 1 OS=Marinomonas sp. (strain MWYL1) GN=atpD1
Q2LR05|ATPB_SYNAS	ATP synthase subunit beta OS=Syntrophus aciditrophicus (strain SB) GN=atpD
Q2LQZ8|ATPD_SYNAS	ATP synthase subunit delta OS=Syntrophus aciditrophicus (strain SB) GN=atpH
Q2LQZ9|ATPF1_SYNAS	ATP synthase subunit b 1 OS=Syntrophus aciditrophicus (strain SB) GN=atpF1
Q2LY33|ATPF2_SYNAS	ATP synthase subunit b 2 OS=Syntrophus aciditrophicus (strain SB) GN=atpF2
A4SUT3|ATPG_POLSQ	ATP synthase gamma chain OS=Polynucleobacter necessarius subsp. asymbioticus (strain DSM 18221 / CIP 109841 / QLW-P1DMWA-1) GN=atpG

Q2LQZ6|ATPG_SYNAS	ATP synthase gamma chain OS=Syntrophus aciditrophicus (strain SB) GN=atpG
A3PS63|ATPL2_RHOS1	ATP synthase subunit c 2 OS=Rhodobacter sphaeroides (strain ATCC 17029 / ATH 2.4.9) GN=atpE2
O87943|BSSA_THAAR	Benzylsuccinate synthase alpha subunit OS=Thauera aromatica GN=bssA ***(assA – see main text)**
O87941|BSSD_THAAR	Benzylsuccinate synthase activating enzyme OS=Thauera aromatica GN=bssD ***(assD - see main text)**
A8ALR6|CAIC_CITK8	Probable crotonobetaine/carnitine-CoA ligase OS=Citrobacter koseri (strain ATCC BAA-895 / CDC 4225-83 / SGSC4696) GN=caiC
Q2LQP0|CHCOA_SYNAS	Cyclohexane-1-carbonyl-CoA dehydrogenase OS=Syntrophus aciditrophicus (strain SB) GN=SYN_02586
Q2LTS1|COAD_SYNAS	Phosphopantetheine adenylyltransferase OS=Syntrophus aciditrophicus (strain SB) GN=coaD
P31773|COMF_HAEIN	Competence protein F OS=Haemophilus influenzae (strain ATCC 51907 / DSM 11121 / KW20 / Rd) GN=comF
P45049|COMM_HAEIN	Competence protein ComM OS=Haemophilus influenzae (strain ATCC 51907 / DSM 11121 / KW20 / Rd) GN=comM
Q8ZD92|CPDP_YERPE	Probable 3',5'-cyclic-nucleotide phosphodiesterase OS=Yersinia pestis GN=cpdP
C5AZ74|CROR_METEA	3-hydroxybutyryl-CoA dehydratase OS=Methylobacterium extorquens (strain ATCC 14718 / DSM 1338 / AM1) GN=croR
P0A6J8|DDLA_ECOLI	D-alanine--D-alanine ligase A OS=Escherichia coli (strain K12) GN=ddlA
Q8XA87|DEAD_ECO57	ATP-dependent RNA helicase DeaD OS=Escherichia coli O157:H7 GN=deaD
P94598|DHE3_BACTN	Glutamate dehydrogenase OS=Bacteroides thetaiotaomicron (strain ATCC 29148 / DSM 2079 / NCTC 10582 / E50 / VPI-5482) GN=gdhA
Q54Q69|DHKG_DICDI	Hybrid signal transduction histidine kinase G OS=Dictyostelium discoideum GN=dhkG
Q57749|DHPSL_METJA	7,8-dihydropterin-6-methyl-4-(beta-D-ribofuranosyl)-aminobenzene-5'-phosphate synthase OS=Methanocaldococcus jannaschii (strain ATCC 43067 / DSM 2661 / JAL-1 / JCM 10045 / NBRC 100440) GN=MJ0301
Q06004|DHSO_BACSU	Sorbitol dehydrogenase OS=Bacillus subtilis (strain 168) GN=gutB
P0ABS3|DKSA_ECO57	RNA polymerase-binding transcription factor DksA OS=Escherichia coli O157:H7 GN=dksA
A0LJ41|DNAJ_SYNFM	Chaperone protein DnaJ OS=Syntrophobacter fumaroxidans (strain DSM 10017 / MPOB) GN=dnaJ
Q2LTN8|DNLJ_SYNAS	DNA ligase OS=Syntrophus aciditrophicus (strain SB) GN=ligA
P52026|DPO1_GEOSE	DNA polymerase I OS=Geobacillus stearothermophilus GN=polA
P45672|DUS_AZOBR	Probable tRNA-dihydrouridine synthase OS=Azospirillum brasilense GN=dus
Q5XIC0|ECI2_RAT	Enoyl-CoA delta isomerase 2, mitochondrial OS=Rattus norvegicus GN=Eci2
P97089|ETFB_THETC	Electron transfer flavoprotein subunit beta OS=Thermoanaerobacterium thermosaccharolyticum (strain ATCC 7956 / DSM 571 / NCIB 9385 / NCA 3814) GN=etfB
P45127|ETTA_HAEIN	Energy-dependent translational throttle protein EttA OS=Haemophilus influenzae (strain ATCC 51907 / DSM 11121 / KW20 / Rd) GN=ettA
P0A1A9|EX3_SALTY	Exodeoxyribonuclease III OS=Salmonella typhimurium (strain LT2 / SGSC1412 / ATCC 700720) GN=xthA

P16118|F261_HUMAN	6-phosphofructo-2-kinase/fructose-2,6-bisphosphatase 1 OS=Homo sapiens GN=PFKFB1
O34340|FABF_BACSU	3-oxoacyl-[acyl-carrier-protein] synthase 2 OS=Bacillus subtilis (strain 168) GN=fabF
P51831|FABG_BACSU	3-oxoacyl-[acyl-carrier-protein] reductase FabG OS=Bacillus subtilis (strain 168) GN=fabG
Q899P3|FABH_CLOTE	3-oxoacyl-[acyl-carrier-protein] synthase 3 OS=Clostridium tetani (strain Massachusetts / E88) GN=fabH
B6EGU1|FADA_ALISL	3-ketoacyl-CoA thiolase OS=Aliivibrio salmonicida (strain LFI1238) GN=fadA
P38135|FADK_ECOLI	Short-chain-fatty-acid--CoA ligase OS=Escherichia coli (strain K12) GN=fadK
Q6T1W8|FDTA_ANETH	TDP-4-oxo-6-deoxy-alpha-D-glucose-3,4-oxoisomerase OS=Aneurinibacillus thermoaerophilus GN=fdtA
Q57563|FER2_METJA	Uncharacterized ferredoxin MJ0099 OS=Methanocaldococcus jannaschii (strain ATCC 43067 / DSM 2661 / JAL-1 / JCM 10045 / NBRC 100440) GN=MJ0099
O67866|FLAV_AQUAE	Flavodoxin OS=Aquifex aeolicus (strain VF5) GN=fldA
Q93AL9|FLDB_CLOSG	R-phenyllactate dehydratase subunit alpha OS=Clostridium sporogenes GN=fldB
Q2LRX4|FMT_SYNAS	Methionyl-tRNA formyltransferase OS=Syntrophus aciditrophicus (strain SB) GN=fmt
Q9HS44|FOLCP_HALSA	Probable bifunctional folylpolyglutamate synthase/dihydropteroate synthase OS=Halobacterium salinarum (strain ATCC 700922 / JCM 11081 / NRC-1) GN=folP
B9M769|FOLD_GEODF	Bifunctional protein FolD OS=Geobacter daltonii (strain DSM 22248 / JCM 15807 / FRC-32) GN=folD
E1WS50|FTN_BACF6	Bacterial non-heme ferritin OS=Bacteroides fragilis (strain 638R) GN=ftnA
Q8P993|FTSK_XANCP	DNA translocase FtsK OS=Xanthomonas campestris pv. campestris (strain ATCC 33913 / NCPPB 528 / LMG 568) GN=ftsK
Q99XR4|GLFT_STRP1	Glutamate formimidoyltransferase OS=Streptococcus pyogenes serotype M1 GN=M5005_Spy1772
Q8TZ14|GLMS_METKA	Glutamine--fructose-6-phosphate aminotransferase [isomerizing] OS=Methanopyrus kandleri (strain AV19 / DSM 6324 / JCM 9639 / NBRC 100938) GN=glmS
P43799|GLPA_HAEIN	Anaerobic glycerol-3-phosphate dehydrogenase subunit A OS=Haemophilus influenzae (strain ATCC 51907 / DSM 11121 / KW20 / Rd) GN=glpA
A4WCP1|GLPB_ENT38	Anaerobic glycerol-3-phosphate dehydrogenase subunit B OS=Enterobacter sp. (strain 638) GN=glpB
P0A997|GLPC_ECO57	Anaerobic glycerol-3-phosphate dehydrogenase subunit C OS=Escherichia coli O157:H7 GN=glpC
B0K643|GLPK_THEPX	Glycerol kinase OS=Thermoanaerobacter sp. (strain X514) GN=glpK
P24943|GLTT_GEOSE	Proton/sodium-glutamate symport protein OS=Geobacillus stearothermophilus GN=gltT
Q72VB8|GPMI_LEPIC	2,3-bisphosphoglycerate-independent phosphoglycerate mutase OS=Leptospira interrogans serogroup Icterohaemorrhagiae serovar copenhageni (strain Fiocruz L1-130) GN=gpmI
Q2LR03|GREA_SYNAS	Transcription elongation factor GreA OS=Syntrophus aciditrophicus (strain SB) GN=greA
P35818|GSPD_PSEAE	Type II secretion system protein D OS=Pseudomonas aeruginosa (strain ATCC 15692 / PAO1 / 1C / PRS 101 / LMG 12228) GN=xcpQ
P31742|GSPE_XANCP	Type II secretion system protein E OS=Xanthomonas campestris pv. campestris (strain ATCC 33913 / NCPPB 528 / LMG

	568) GN=xpsE
O18404|HCD2_DROME	3-hydroxyacyl-CoA dehydrogenase type-2 OS=Drosophila melanogaster GN=scu
P9WMR0|HELY_MYCTO	Probable helicase HelY OS=Mycobacterium tuberculosis (strain CDC 1551 / Oshkosh) GN=helY
A1AV12|HLDE_PELPD	Bifunctional protein HldE OS=Pelobacter propionicus (strain DSM 2379) GN=hldE
P43777|HPPK_HAEIN	2-amino-4-hydroxy-6-hydroxymethyldihydropteridine pyrophosphokinase OS=Haemophilus influenzae (strain ATCC 51907 / DSM 11121 / KW20 / Rd) GN=folK
C0QJ36|HSLU_DESAH	ATP-dependent protease ATPase subunit HslU OS=Desulfobacterium autotrophicum (strain ATCC 43914 / DSM 3382 / HRM2) GN=hslU
B8GTA0|HSLV_THISH	ATP-dependent protease subunit HslV OS=Thioalkalivibrio sulfidiphilus (strain HL-EbGR7) GN=hslV
Q834W6|HYEP_ENTFA	Hydrophobic dipeptide epimerase OS=Enterococcus faecalis (strain ATCC 700802 / V583) GN=EF_1511
O13294|IDH2_CANTR	Isocitrate dehydrogenase [NADP] peroxisomal OS=Candida tropicalis GN=IDP2
Q74D03|IF3_GEOSL	Translation initiation factor IF-3 OS=Geobacter sulfurreducens (strain ATCC 51573 / DSM 12127 / PCA) GN=infC
P37251|ILVB_BACSU	Acetolactate synthase large subunit OS=Bacillus subtilis (strain 168) GN=ilvB
Q2LXP6|ILVD_SYNAS	Dihydroxy-acid dehydratase OS=Syntrophus aciditrophicus (strain SB) GN=ilvD
P39576|ILVE2_BACSU	Branched-chain-amino-acid aminotransferase 2 OS=Bacillus subtilis (strain 168) GN=ilvK
O67703|ILVH_AQUAE	Acetolactate synthase small subunit OS=Aquifex aeolicus (strain VF5) GN=ilvH
P62590|INT2_ECOLX	Integrase/recombinase OS=Escherichia coli GN=int
Q9KD89|IOJAP_BACHD	Ribosomal silencing factor RsfS OS=Bacillus halodurans (strain ATCC BAA-125 / DSM 18197 / FERM 7344 / JCM 9153 / C-125) GN=rsfS
Q1MSE5|ISPG_LAWIP	4-hydroxy-3-methylbut-2-en-1-yl diphosphate synthase OS=Lawsonia intracellularis (strain PHE/MN1-00) GN=ispG
A5G5T2|KDSA_GEOUR	2-dehydro-3-deoxyphosphooctonate aldolase OS=Geobacter uraniireducens (strain Rf4) GN=kdsA
Q2LY80|KDSB_SYNAS	3-deoxy-manno-octulosonate cytidylyltransferase OS=Syntrophus aciditrophicus (strain SB) GN=kdsB
Q6GV12|KDSR_MOUSE	3-ketodihydrosphingosine reductase OS=Mus musculus GN=Kdsr
P0AC77|KDTA_ECO57	3-deoxy-D-manno-octulosonic acid transferase OS=Escherichia coli O157:H7 GN=waaA
A8ALQ4|KEFC_CITK8	Glutathione-regulated potassium-efflux system protein KefC OS=Citrobacter koseri (strain ATCC BAA-895 / CDC 4225-83 / SGSC4696) GN=kefC
A4WFE5|KEFG_ENT38	Glutathione-regulated potassium-efflux system ancillary protein KefG OS=Enterobacter sp. (strain 638) GN=kefG
A8X2R1|L2HDH_CAEBR	L-2-hydroxyglutarate dehydrogenase, mitochondrial OS=Caenorhabditis briggsae GN=CBG06643
Q5SKN9|LCFCS_THET8	Long-chain-fatty-acid--CoA ligase OS=Thermus thermophilus (strain HB8 / ATCC 27634 / DSM 579) GN=TTHA0604
A2T195|LEP4_AERS4	Type 4 prepilin-like proteins leader peptide-processing enzyme OS=Aeromonas salmonicida (strain A449) GN=tapD
Q8RCF9|LEU12_CALS4	2-isopropylmalate synthase 2 OS=Caldanaerobacter subterraneus subsp. tengcongensis (strain DSM 15242 / JCM 11007 /

	NBRC 100824 / MB4) GN=leuA2
Q2LWJ3|LEU1_SYNAS	2-isopropylmalate synthase OS=Syntrophus aciditrophicus (strain SB) GN=leuA
Q607H7|LOLA_METCA	Outer-membrane lipoprotein carrier protein OS=Methylococcus capsulatus (strain ATCC 33009 / NCIMB 11132 / Bath) GN=lolA
Q2LVS9|LON_SYNAS	Lon protease OS=Syntrophus aciditrophicus (strain SB) GN=lon
Q48IV1|LOVHK_PSE14	Blue-light-activated protein OS=Pseudomonas syringae pv. phaseolicola (strain 1448A / Race 6) GN=PSPPH_2483
Q9LLY4|LPAT1_BRANA	1-acyl-sn-glycerol-3-phosphate acyltransferase 1, chloroplastic OS=Brassica napus GN=LPAT1
Q7N8V4|LPTD_PHOLL	LPS-assembly protein LptD OS=Photorhabdus luminescens subsp. laumondii (strain TT01) GN=lptD
Q50414|MAUE_METFK	Methylamine utilization protein MauE OS=Methylobacillus flagellatus (strain KT / ATCC 51484 / DSM 6875) GN=mauE
Q5F336|MBLC2_CHICK	Metallo-beta-lactamase domain-containing protein 2 OS=Gallus gallus GN=MBLAC2
P39582|MENA_BACSU	1,4-dihydroxy-2-naphthoate octaprenyltransferase OS=Bacillus subtilis (strain 168) GN=menA
B0RCZ0|MENG_CLAMS	Demethylmenaquinone methyltransferase OS=Clavibacter michiganensis subsp. sepedonicus (strain ATCC 33113 / DSM 20744 / JCM 9667 / LMG 2889 / C-1) GN=menG
P0AFV6|MEPS_ECO57	Murein DD-endopeptidase MepS/Murein LD-carboxypeptidase OS=Escherichia coli O157:H7 GN=mepS
Q3A388|METK_PELCD	S-adenosylmethionine synthase OS=Pelobacter carbinolicus (strain DSM 2380 / Gra Bd 1) GN=metK
P9WJ78|MFTC_MYCTO	Putative mycofactocin radical SAM maturase MftC OS=Mycobacterium tuberculosis (strain CDC 1551 / Oshkosh) GN=mftC
P82597|MGLP_BAC25	Thermostable monoacylglycerol lipase OS=Bacillus sp. (strain H-257)
Q2LT94|MIAB_SYNAS	tRNA-2-methylthio-N(6)-dimethylallyladenosine synthase OS=Syntrophus aciditrophicus (strain SB) GN=miaB
P30769|MKL_MYCLE	Probable ribonucleotide transport ATP-binding protein mkl OS=Mycobacterium leprae (strain TN) GN=mkl
Q2RGL2|MOAA_MOOTA	Cyclic pyranopterin monophosphate synthase OS=Moorella thermoacetica (strain ATCC 39073) GN=moaA
B0TI15|MOAC_HELMI	Cyclic pyranopterin monophosphate synthase accessory protein OS=Heliobacterium modesticaldum (strain ATCC 51547 / Ice1) GN=moaC
Q44243|MOEA_NOSS1	Molybdopterin molybdenumtransferase OS=Nostoc sp. (strain PCC 7120 / UTEX 2576) GN=moeA
P0A9X4|MREB_ECOLI	Rod shape-determining protein MreB OS=Escherichia coli (strain K12) GN=mreB
B8H610|MREC_CAUCN	Cell shape-determining protein MreC OS=Caulobacter crescentus (strain NA1000 / CB15N) GN=mreC
Q2YBM4|MTGA_NITMU	Monofunctional biosynthetic peptidoglycan transglycosylase OS=Nitrosospira multiformis (strain ATCC 25196 / NCIMB 11849) GN=mtgA
P37169|MURJ_SALTY	Putative lipid II flippase MurJ OS=Salmonella typhimurium (strain LT2 / SGSC1412 / ATCC 700720) GN=murJ
Q51363|NADB_PSEAE	L-aspartate oxidase OS=Pseudomonas aeruginosa (strain ATCC 15692 / PAO1 / 1C / PRS 101 / LMG 12228) GN=nadB
Q2LUY2|NIKR_SYNAS	Putative nickel-responsive regulator OS=Syntrophus aciditrophicus (strain SB) GN=SYNAS_20170
Q52524|NIRG_PSEST	Protein NirG OS=Pseudomonas stutzeri GN=nirG

P28903|NRDD_ECOLI	Anaerobic ribonucleoside-triphosphate reductase OS=Escherichia coli (strain K12) GN=nrdD
Q8PXY1|NRPR1_METMA	Global nitrogen regulator NrpRI OS=Methanosarcina mazei (strain ATCC BAA-159 / DSM 3647 / Goe1 / Go1 / JCM 11833 / OCM 88) GN=nrpRI
P09431|NTRB_RHOCB	Nitrogen regulation protein NtrB OS=Rhodobacter capsulatus (strain ATCC BAA-309 / NBRC 16581 / SB1003) GN=ntrB
P56558|OGT1_RAT	UDP-N-acetylglucosamine--peptide N-acetylglucosaminyltransferase 110 kDa subunit OS=Rattus norvegicus GN=Ogt
Q2LT98|OTC_SYNAS	Ornithine carbamoyltransferase OS=Syntrophus aciditrophicus (strain SB) GN=argF
Q9K0U7|PA1_NEIMB	Putative phospholipase A1 OS=Neisseria meningitidis serogroup B (strain MC58) GN=NMB0464
Q72K16|PAAK_THET2	Phenylacetate-coenzyme A ligase OS=Thermus thermophilus (strain HB27 / ATCC BAA-163 / DSM 7039) GN=TT_C0602
P0A138|PAL_PSEPK	Peptidoglycan-associated lipoprotein OS=Pseudomonas putida (strain KT2440) GN=pal
Q2LTJ5|PANB_SYNAS	3-methyl-2-oxobutanoate hydroxymethyltransferase OS=Syntrophus aciditrophicus (strain SB) GN=panB
Q2LSQ5|PANC_SYNAS	Pantothenate synthetase OS=Syntrophus aciditrophicus (strain SB) GN=panC
Q2LSQ4|PAND_SYNAS	Aspartate 1-decarboxylase OS=Syntrophus aciditrophicus (strain SB) GN=panD
P44469|PBP2_HAEIN	Penicillin-binding protein 2 OS=Haemophilus influenzae (strain ATCC 51907 / DSM 11121 / KW20 / Rd) GN=mrdA
P70997|PBPG_BACSU	Penicillin-binding protein 2D OS=Bacillus subtilis (strain 168) GN=pbpG
P46322|PGSA_BACSU	CDP-diacylglycerol--glycerol-3-phosphate 3-phosphatidyltransferase OS=Bacillus subtilis (strain 168) GN=pgsA
I3R9Z4|PHAC_HALMT	Poly(3-hydroxyalkanoate) polymerase subunit PhaC OS=Haloferax mediterranei (strain ATCC 33500 / DSM 1411 / JCM 8866 / NBRC 14739 / NCIMB 2177 / R-4) GN=phaC
O67085|PHEA_AQUAE	P-protein OS=Aquifex aeolicus (strain VF5) GN=pheA
P22263|PORF_PSESY	Outer membrane porin F OS=Pseudomonas syringae pv. syringae GN=oprF
Q9LJL3|PREP1_ARATH	Presequence protease 1, chloroplastic/mitochondrial OS=Arabidopsis thaliana GN=PREP1
P94461|PRIA_BACSU	Primosomal protein N' OS=Bacillus subtilis (strain 168) GN=priA
Q97E99|PRSA_CLOAB	Foldase protein PrsA OS=Clostridium acetobutylicum (strain ATCC 824 / DSM 792 / JCM 1419 / LMG 5710 / VKM B-1787) GN=prsA
Q9L3Q5|PRXU_EUBAC	Selenocysteine-containing peroxiredoxin PrxU OS=Eubacterium acidaminophilum GN=prxU
Q2LTR6|PSD_SYNAS	Phosphatidylserine decarboxylase proenzyme OS=Syntrophus aciditrophicus (strain SB) GN=psd
Q48269|PSS_HELPY	CDP-diacylglycerol--serine O-phosphatidyltransferase OS=Helicobacter pylori (strain ATCC 700392 / 26695) GN=pssA
P26379|PTFA_BACSU	Fructose-specific phosphotransferase enzyme IIA component OS=Bacillus subtilis (strain 168) GN=levD
P26380|PTFB_BACSU	Fructose-specific phosphotransferase enzyme IIB component OS=Bacillus subtilis (strain 168) GN=levE
P00497|PUR1_BACSU	Amidophosphoribosyltransferase OS=Bacillus subtilis (strain 168) GN=purF
A8ZVI6|PUR7_DESOH	Phosphoribosylaminoimidazole-succinocarboxamide synthase OS=Desulfococcus oleovorans (strain DSM 6200 / Hxd3) GN=purC

Q2LSI8|PURA_SYNAS	Adenylosuccinate synthetase OS=Syntrophus aciditrophicus (strain SB) GN=purA
B3E4T4|PYRDB_GEOLS	Dihydroorotate dehydrogenase B (NAD(+)), catalytic subunit OS=Geobacter lovleyi (strain ATCC BAA-1151 / DSM 17278 / SZ) GN=pyrD
P58885|PYRK_CALS4	Dihydroorotate dehydrogenase B (NAD(+)), electron transfer subunit OS=Caldanaerobacter subterraneus subsp. tengcongensis (strain DSM 15242 / JCM 11007 / NBRC 100824 / MB4) GN=pyrK
Q81I72|QUEG_BACCR	Epoxyqueuosine reductase OS=Bacillus cereus (strain ATCC 14579 / DSM 31) GN=queG
O52177|RELA_MYXXA	GTP pyrophosphokinase OS=Myxococcus xanthus GN=relA
Q73B22|RESA_BACC1	Thiol-disulfide oxidoreductase ResA OS=Bacillus cereus (strain ATCC 10987) GN=resA
Q56903|RFBE_YEREN	O-antigen export system ATP-binding protein RfbE OS=Yersinia enterocolitica GN=rfbE
P25888|RHLE_ECOLI	ATP-dependent RNA helicase RhlE OS=Escherichia coli (strain K12) GN=rhlE
O05517|RIMI_BACSU	Putative ribosomal-protein-alanine acetyltransferase OS=Bacillus subtilis (strain 168) GN=rimI
Q2LVU6|RIMM_SYNAS	Ribosome maturation factor RimM OS=Syntrophus aciditrophicus (strain SB) GN=rimM
H8EUF2|RIP3_MYCTE	Putative zinc metalloprotease Rip3 OS=Mycobacterium tuberculosis (strain ATCC 35801 / TMC 107 / Erdman) GN=rip3
Q2LVU4|RL19_SYNAS	50S ribosomal protein L19 OS=Syntrophus aciditrophicus (strain SB) GN=rplS
Q2LR24|RL20_SYNAS	50S ribosomal protein L20 OS=Syntrophus aciditrophicus (strain SB) GN=rplT
Q2LR20|RL35_SYNAS	50S ribosomal protein L35 OS=Syntrophus aciditrophicus (strain SB) GN=rpmI
Q2LUM5|RLMN_SYNAS	Dual-specificity RNA methyltransferase RlmN OS=Syntrophus aciditrophicus (strain SB) GN=rlmN
P44443|RNE_HAEIN	Ribonuclease E OS=Haemophilus influenzae (strain ATCC 51907 / DSM 11121 / KW20 / Rd) GN=rne
Q2LVU5|RNH2_SYNAS	Ribonuclease HII OS=Syntrophus aciditrophicus (strain SB) GN=rnhB
P44468|RODA_HAEIN	Rod shape-determining protein RodA OS=Haemophilus influenzae (strain ATCC 51907 / DSM 11121 / KW20 / Rd) GN=mrdB
O34557|RPE_BACSU	Ribulose-phosphate 3-epimerase OS=Bacillus subtilis (strain 168) GN=rpe
Q4UU85|RPFG_XANC8	Cyclic di-GMP phosphodiesterase response regulator RpfG OS=Xanthomonas campestris pv. campestris (strain 8004) GN=rpfG
Q2LVU9|RS16_SYNAS	30S ribosomal protein S16 OS=Syntrophus aciditrophicus (strain SB) GN=rpsP
Q2LSQ6|RSMA_SYNAS	Ribosomal RNA small subunit methyltransferase A OS=Syntrophus aciditrophicus (strain SB) GN=rsmA
P94464|RSMB_BACSU	Ribosomal RNA small subunit methyltransferase B OS=Bacillus subtilis (strain 168) GN=rsmB
Q2LR39|RSMH_SYNAS	Ribosomal RNA small subunit methyltransferase H OS=Syntrophus aciditrophicus (strain SB) GN=rsmH
O27657|RUBPS_METTH	Putative ribose 1,5-bisphosphate isomerase OS=Methanothermobacter thermautotrophicus (strain ATCC 29096 / DSM 1053 / JCM 10044 / NBRC 100330 / Delta H) GN=MTH_1620
Q72EH1|SAHH_DESVH	Adenosylhomocysteinase OS=Desulfovibrio vulgaris (strain Hildenborough / ATCC 29579 / NCIMB 8303) GN=ahcY

Q9M8Y0|SEC_ARATH	Probable UDP-N-acetylglucosamine--peptide N-acetylglucosaminyltransferase SEC OS=Arabidopsis thaliana GN=SEC
P43885|SERA_HAEIN	D-3-phosphoglycerate dehydrogenase OS=Haemophilus influenzae (strain ATCC 51907 / DSM 11121 / KW20 / Rd) GN=serA
P52878|SERC_METBF	Phosphoserine aminotransferase OS=Methanosarcina barkeri (strain Fusaro / DSM 804) GN=serC
P37105|SRP54_BACSU	Signal recognition particle protein OS=Bacillus subtilis (strain 168) GN=ffh
P47154|STE24_YEAST	CAAX prenyl protease 1 OS=Saccharomyces cerevisiae (strain ATCC 204508 / S288c) GN=STE24
Q2LR28|SYFA_SYNAS	Phenylalanine--tRNA ligase alpha subunit OS=Syntrophus aciditrophicus (strain SB) GN=pheS
Q747X9|SYI_GEOSL	Isoleucine--tRNA ligase OS=Geobacter sulfurreducens (strain ATCC 51573 / DSM 12127 / PCA) GN=ileS
Q2LQK5|SYL_SYNAS	Leucine--tRNA ligase OS=Syntrophus aciditrophicus (strain SB) GN=leuS
Q2LTR7|SYP_SYNAS	Proline--tRNA ligase OS=Syntrophus aciditrophicus (strain SB) GN=proS
B5XZG7|SYQ_KLEP3	Glutamine--tRNA ligase OS=Klebsiella pneumoniae (strain 342) GN=glnS
Q2LVV5|SYR_SYNAS	Arginine--tRNA ligase OS=Syntrophus aciditrophicus (strain SB) GN=argS
B5E8F6|SYS_GEOBB	Serine--tRNA ligase OS=Geobacter bemidjiensis (strain Bem / ATCC BAA-1014 / DSM 16622) GN=serS
A1ARE5|SYT_PELPD	Threonine--tRNA ligase OS=Pelobacter propionicus (strain DSM 2379) GN=thrS
A8ZW56|TAL_DESOH	Probable transaldolase OS=Desulfococcus oleovorans (strain DSM 6200 / Hxd3) GN=tal
C0QD59|TATC_DESAH	Sec-independent protein translocase protein TatC OS=Desulfobacterium autotrophicum (strain ATCC 43914 / DSM 3382 / HRM2) GN=tatC
P07464|THGA_ECOLI	Galactoside O-acetyltransferase OS=Escherichia coli (strain K12) GN=lacA
A7N7S3|THIE_VIBCB	Thiamine-phosphate synthase OS=Vibrio campbellii (strain ATCC BAA-1116 / BB120) GN=thiE
Q2LVU2|TRMD_SYNAS	tRNA (guanine-N(1)-)-methyltransferase OS=Syntrophus aciditrophicus (strain SB) GN=trmD
Q2LSP8|TSAD_SYNAS	tRNA N6-adenosine threonylcarbamoyltransferase OS=Syntrophus aciditrophicus (strain SB) GN=tsaD
Q9FZE1|UGDH1_ARATH	UDP-glucose 6-dehydrogenase 1 OS=Arabidopsis thaliana GN=UGD1
Q8XNI5|UVRA_CLOPE	UvrABC system protein A OS=Clostridium perfringens (strain 13 / Type A) GN=uvrA
Q8VZC0|UXS1_ARATH	UDP-glucuronic acid decarboxylase 1 OS=Arabidopsis thaliana GN=UXS1
Q46558|VAPB_DICNO	Virulence-associated protein B OS=Dichelobacter nodosus GN=vapB
O06662|VAPC_SHIFL	tRNA(fMet)-specific endonuclease VapC OS=Shigella flexneri GN=vapC
G4NYJ6|WAPA_BACPT	tRNA3(Ser)-specific nuclease WapA OS=Bacillus subtilis subsp. spizizenii (strain TU-B-10) GN=wapA
Q2LT92|XERC_SYNAS	Tyrosine recombinase XerC OS=Syntrophus aciditrophicus (strain SB) GN=xerC
Q9KCP0|XERD_BACHD	Tyrosine recombinase XerD OS=Bacillus halodurans (strain ATCC BAA-125 / DSM 18197 / FERM 7344 / JCM 9153 / C-125) GN=xerD
P45373|Y064_ALLVD	Uncharacterized protein Alvin_0064 OS=Allochromatium vinosum (strain ATCC 17899 / DSM 180 / NBRC 103801 / D)

	GN=Alvin_0064
Q2LSD6|Y1120_SYNAS	Maf-like protein SYNAS_11200 OS=Syntrophus aciditrophicus (strain SB) GN=SYNAS_11200
Q58536|Y1136_METJA	Uncharacterized protein MJ1136 OS=Methanocaldococcus jannaschii (strain ATCC 43067 / DSM 2661 / JAL-1 / JCM 10045 / NBRC 100440) GN=MJ1136
P74178|Y1178_SYNY3	Uncharacterized protein sll1178 OS=Synechocystis sp. (strain PCC 6803 / Kazusa) GN=sll1178
P45103|Y1198_HAEIN	Uncharacterized protein HI_1198 OS=Haemophilus influenzae (strain ATCC 51907 / DSM 11121 / KW20 / Rd) GN=HI_1198
Q58624|Y1227_METJA	Uncharacterized protein MJ1227 OS=Methanocaldococcus jannaschii (strain ATCC 43067 / DSM 2661 / JAL-1 / JCM 10045 / NBRC 100440) GN=MJ1227
O27329|Y1261_METTH	UPF0104 membrane protein MTH_1261 OS=Methanothermobacter thermautotrophicus (strain ATCC 29096 / DSM 1053 / JCM 10044 / NBRC 100330 / Delta H) GN=MTH_1261
Q9CLF7|Y1278_PASMU	Uncharacterized aldolase PM1278 OS=Pasteurella multocida (strain Pm70) GN=PM1278
Q1RGU3|Y1340_RICBR	Probable ABC transporter permease protein RBE_1340 OS=Rickettsia bellii (strain RML369-C) GN=RBE_1340
P45245|Y1544_HAEIN	Uncharacterized NAD(P)H oxidoreductase HI_1544 OS=Haemophilus influenzae (strain ATCC 51907 / DSM 11121 / KW20 / Rd) GN=HI_1544
P73695|Y1697_SYNY3	Thylakoid-associated protein sll1697 OS=Synechocystis sp. (strain PCC 6803 / Kazusa) GN=sll1697
Q9I310|Y1727_PSEAE	Uncharacterized signaling protein PA1727 OS=Pseudomonas aeruginosa (strain ATCC 15692 / PAO1 / 1C / PRS 101 / LMG 12228) GN=PA1727
Q97I96|Y1756_CLOAB	UPF0109 protein CA_C1756 OS=Clostridium acetobutylicum (strain ATCC 824 / DSM 792 / JCM 1419 / LMG 5710 / VKM B-1787) GN=CA_C1756
P9WLR4|Y1829_MYCTO	Uncharacterized protein MT1877 OS=Mycobacterium tuberculosis (strain CDC 1551 / Oshkosh) GN=MT1877
P64924|Y2031_MYCBO	Uncharacterized protein Mb2031c OS=Mycobacterium bovis (strain ATCC BAA-935 / AF2122/97) GN=Mb2031c
O67893|Y2135_AQUAE	Uncharacterized protein aq_2135 OS=Aquifex aeolicus (strain VF5) GN=aq_2135
Q2LVT7|Y2322_SYNAS	UPF0102 protein SYNAS_23220 OS=Syntrophus aciditrophicus (strain SB) GN=SYNAS_23220
Q8PUE7|Y2387_METMA	Putative ABC transporter ATP-binding protein MM_2387 OS=Methanosarcina mazei (strain ATCC BAA-159 / DSM 3647 / Goe1 / Go1 / JCM 11833 / OCM 88) GN=MM_2387
Q6MEM7|Y248_PARUW	Uncharacterized RNA methyltransferase pc0248 OS=Protochlamydia amoebophila (strain UWE25) GN=pc0248
O26373|Y273_METTH	UPF0098 protein MTH_273 OS=Methanothermobacter thermautotrophicus (strain ATCC 29096 / DSM 1053 / JCM 10044 / NBRC 100330 / Delta H) GN=MTH_273
Q55914|Y309_SYNY3	Uncharacterized methyltransferase slr0309 OS=Synechocystis sp. (strain PCC 6803 / Kazusa) GN=slr0309
Q57951|Y531_METJA	Universal stress protein MJ0531 OS=Methanocaldococcus jannaschii (strain ATCC 43067 / DSM 2661 / JAL-1 / JCM 10045 / NBRC 100440) GN=MJ0531
A0R3D6|Y5435_MYCS2	Putative ligase MSMEG_5435/MSMEI_5285 OS=Mycobacterium smegmatis (strain ATCC 700084 / mc(2)155) GN=MSMEG_5435
Q58000|Y580_METJA	Uncharacterized protein MJ0580 OS=Methanocaldococcus jannaschii (strain ATCC 43067 / DSM 2661 / JAL-1 / JCM 10045

	/ NBRC 100440) GN=MJ0580
Q9V0X6|Y663_PYRAB	Probable S-adenosyl-L-methionine-binding protein PYRAB06630 OS=Pyrococcus abyssi (strain GE5 / Orsay) GN=PYRAB06630
Q04737|Y751_SYNY3	TPR repeat-containing protein slr0751 OS=Synechocystis sp. (strain PCC 6803 / Kazusa) GN=slr0751
Q58218|Y808_METJA	Uncharacterized protein MJ0808 OS=Methanocaldococcus jannaschii (strain ATCC 43067 / DSM 2661 / JAL-1 / JCM 10045 / NBRC 100440) GN=MJ0808
O58549|Y819_PYRHO	Uncharacterized methyltransferase PH0819 OS=Pyrococcus horikoshii (strain ATCC 700860 / DSM 12428 / JCM 9974 / NBRC 100139 / OT-3) GN=PH0819
Q58242|Y832_METJA	Uncharacterized protein MJ0832 OS=Methanocaldococcus jannaschii (strain ATCC 43067 / DSM 2661 / JAL-1 / JCM 10045 / NBRC 100440) GN=MJ0832
Q58298|Y888_METJA	Probable metallo-hydrolase MJ0888 OS=Methanocaldococcus jannaschii (strain ATCC 43067 / DSM 2661 / JAL-1 / JCM 10045 / NBRC 100440) GN=MJ0888
Q58382|Y972_METJA	UPF0056 membrane protein MJ0972 OS=Methanocaldococcus jannaschii (strain ATCC 43067 / DSM 2661 / JAL-1 / JCM 10045 / NBRC 100440) GN=MJ0972
P37543|YABB_BACSU	Uncharacterized protein YabB OS=Bacillus subtilis (strain 168) GN=yabB
P39830|YBAL_ECOLI	Inner membrane protein YbaL OS=Escherichia coli (strain K12) GN=ybaL
O05519|YDIF_BACSU	Uncharacterized ABC transporter ATP-binding protein YdiF OS=Bacillus subtilis (strain 168) GN=ydiF
O34789|YDJI_BACSU	Uncharacterized protein YdjI OS=Bacillus subtilis (strain 168) GN=ydjI
P76221|YDJZ_ECOLI	TVP38/TMEM64 family inner membrane protein YdjZ OS=Escherichia coli (strain K12) GN=ydjZ
Q797A7|YFNA_BACSU	Uncharacterized amino acid permease YfnA OS=Bacillus subtilis (strain 168) GN=yfnA
P54592|YHCH_BACSU	Uncharacterized ABC transporter ATP-binding protein YhcH OS=Bacillus subtilis (strain 168) GN=yhcH
Q45500|YKTD_BACSU	Putative S-adenosyl-L-methionine-dependent methyltransferase YktD OS=Bacillus subtilis (strain 168) GN=yktD
Q5UQE9|YL477_MIMIV	Uncharacterized peptidase C1-like protein L477 OS=Acanthamoeba polyphaga mimivirus GN=MIMI_L477
O34331|YLBH_BACSU	Putative rRNA methyltransferase YlbH OS=Bacillus subtilis (strain 168) GN=ylbH
O34731|YLBK_BACSU	Uncharacterized NTE family protein YlbK OS=Bacillus subtilis (strain 168) GN=ylbK
P54304|YQER_BACSU	Oxygen-independent coproporphyrinogen-III oxidase-like protein YqeR OS=Bacillus subtilis (strain 168) GN=hemN
P54463|YQEW_BACSU	Uncharacterized protein YqeW OS=Bacillus subtilis (strain 168) GN=yqeW
P54525|YQII_BACSU	Uncharacterized protein YqiI OS=Bacillus subtilis (strain 168) GN=yqiI
P54553|YQJP_BACSU	Probable metallo-hydrolase YqjP OS=Bacillus subtilis (strain 168) GN=yqjP
O05243|YUGU_BACSU	UPF0047 protein YugU OS=Bacillus subtilis (strain 168) GN=yugU
O34451|YVRB_BACSU	Uncharacterized ABC transporter permease protein YvrB OS=Bacillus subtilis (strain 168) GN=yvrB
P39156|YWLF_BACSU	Putative sugar phosphate isomerase YwlF OS=Bacillus subtilis (strain 168) GN=ywlF

P42306|YXIO_BACSU	Uncharacterized MFS-type transporter YxiO OS=Bacillus subtilis (strain 168) GN=yxiO
P25852|ZRAR_SALTY	Transcriptional regulatory protein ZraR OS=Salmonella typhimurium (strain LT2 / SGSC1412 / ATCC 700720) GN=zraR
Q8Z332|ZRAS_SALTI	Sensor protein ZraS OS=Salmonella typhi GN=zraS

JOHN WILEY AND SONS LICENSE TERMS AND CONDITIONS

Aug 11, 2016

This Agreement between Christopher R Marks ("You") and John Wiley and Sons ("John Wiley and Sons") consists of your license details and the terms and conditions provided by John Wiley and Sons and Copyright Clearance Center.

License Number	3921040279819
License date	Aug 02, 2016
Licensed Content Publisher	John Wiley and Sons
Licensed Content Publication	Environmental Microbiology
Licensed Content Title	Methanogenic paraffin degradation proceeds via alkane addition to fumarate by 'Smithella' spp. mediated by a syntrophic coupling with hydrogenotrophic methanogens
Licensed Content Author	Boris Wawrik,Christopher R. Marks,Irene A. Davidova,Michael J. McInerney,Shane Pruitt,Kathleen E. Duncan,Joseph M. Suflita,Amy V. Callaghan
Licensed Content Date	Jun 27, 2016
Licensed Content Pages	1
Type of use	Dissertation/Thesis
Requestor type	Author of this Wiley article
Format	Electronic
Portion	Full article
Will you be translating?	No
Title of your thesis / dissertation	The Ecology of Anaerobic Hydrocarbon-Degrading Microorganisms Associated with the Built-Environment.
Expected completion date	Sep 2016
Expected size (number of pages)	200
Requestor Location	Christopher R Marks 770 Van Vleet Oval NORMAN, OK 73019 United States Attn: Christopher R Marks
Publisher Tax ID	EU826007151
Billing Type	Invoice
Billing Address	Christopher R Marks 770 Van Vleet Oval NORMAN, OK 73019 United States Attn: Christopher R Marks

basis, or any of the rights granted to you hereunder to any other person.

- The Wiley Materials and all of the intellectual property rights therein shall at all times remain the exclusive property of John Wiley & Sons Inc, the Wiley Companies, or their respective licensors, and your interest therein is only that of having possession of and the right to reproduce the Wiley Materials pursuant to Section 2 herein during the continuance of this Agreement. You agree that you own no right, title or interest in or to the Wiley Materials or any of the intellectual property rights therein. You shall have no rights hereunder other than the license as provided for above in Section 2. No right, license or interest to any trademark, trade name, service mark or other branding ("Marks") of WILEY or its licensors is granted hereunder, and you agree that you shall not assert any such right, license or interest with respect thereto

- NEITHER WILEY NOR ITS LICENSORS MAKES ANY WARRANTY OR REPRESENTATION OF ANY KIND TO YOU OR ANY THIRD PARTY, EXPRESS, IMPLIED OR STATUTORY, WITH RESPECT TO THE MATERIALS OR THE ACCURACY OF ANY INFORMATION CONTAINED IN THE MATERIALS, INCLUDING, WITHOUT LIMITATION, ANY IMPLIED WARRANTY OF MERCHANTABILITY, ACCURACY, SATISFACTORY QUALITY, FITNESS FOR A PARTICULAR PURPOSE, USABILITY, INTEGRATION OR NON-INFRINGEMENT AND ALL SUCH WARRANTIES ARE HEREBY EXCLUDED BY WILEY AND ITS LICENSORS AND WAIVED BY YOU.

- WILEY shall have the right to terminate this Agreement immediately upon breach of this Agreement by you.

- You shall indemnify, defend and hold harmless WILEY, its Licensors and their respective directors, officers, agents and employees, from and against any actual or threatened claims, demands, causes of action or proceedings arising from any breach of this Agreement by you.

- IN NO EVENT SHALL WILEY OR ITS LICENSORS BE LIABLE TO YOU OR ANY OTHER PARTY OR ANY OTHER PERSON OR ENTITY FOR ANY SPECIAL, CONSEQUENTIAL, INCIDENTAL, INDIRECT, EXEMPLARY OR PUNITIVE DAMAGES, HOWEVER CAUSED, ARISING OUT OF OR IN CONNECTION WITH THE DOWNLOADING, PROVISIONING, VIEWING OR USE OF THE MATERIALS REGARDLESS OF THE FORM OF ACTION, WHETHER FOR BREACH OF CONTRACT, BREACH OF WARRANTY, TORT, NEGLIGENCE, INFRINGEMENT OR OTHERWISE (INCLUDING, WITHOUT LIMITATION, DAMAGES BASED ON LOSS OF PROFITS, DATA, FILES, USE, BUSINESS OPPORTUNITY OR CLAIMS OF THIRD PARTIES), AND WHETHER OR NOT THE PARTY HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES. THIS LIMITATION SHALL APPLY NOTWITHSTANDING ANY FAILURE OF ESSENTIAL PURPOSE OF ANY LIMITED REMEDY PROVIDED HEREIN.

- Should any provision of this Agreement be held by a court of competent jurisdiction to be illegal, invalid, or unenforceable, that provision shall be deemed amended to achieve as nearly as possible the same economic effect as the original provision, and the legality, validity and enforceability of the remaining provisions of this Agreement shall not be affected or impaired thereby.

- The failure of either party to enforce any term or condition of this Agreement shall not constitute a waiver of either party's right to enforce each and every term and condition of this Agreement. No breach under this agreement shall be deemed waived or excused by either party unless such waiver or consent is in writing signed by the party granting such waiver or consent. The waiver by or consent of a party to a breach of any provision of this Agreement shall not operate or be construed as a waiver of or consent to any other or subsequent breach by such other party.

- This Agreement may not be assigned (including by operation of law or otherwise) by you without WILEY's prior written consent.

- Any fee required for this permission shall be non-refundable after thirty (30) days from receipt by the CCC.

- These terms and conditions together with CCC's Billing and Payment terms and conditions (which are incorporated herein) form the entire agreement between you and WILEY concerning this licensing transaction and (in the absence of fraud) supersedes all prior agreements and representations of the parties, oral or written. This Agreement may not be amended except in writing signed by both parties. This Agreement shall be binding upon and inure to the benefit of the parties' successors, legal representatives, and authorized assigns.

- In the event of any conflict between your obligations established by these terms and conditions and those established by CCC's Billing and Payment terms and conditions, these terms and conditions shall prevail.

- WILEY expressly reserves all rights not specifically granted in the combination of (i) the license details provided by you and accepted in the course of this licensing transaction, (ii) these terms and conditions and (iii) CCC's Billing and Payment terms and conditions.

- This Agreement will be void if the Type of Use, Format, Circulation, or Requestor Type was misrepresented during the licensing process.

- This Agreement shall be governed by and construed in accordance with the laws of the State of New York, USA, without regards to such state's conflict of law rules. Any legal action, suit or proceeding arising out of or relating to these Terms and Conditions or the breach thereof shall be instituted in a court of competent jurisdiction in New York County in the State of New York in the United States of America and each party hereby consents and submits to the personal jurisdiction of such court, waives any objection to venue in such court and consents to service of process by registered or

certified mail, return receipt requested, at the last known address of such party.