

**University of Alberta**

Using "*omics*" approaches to study anaerobic hydrocarbon biodegradation by  
microbes indigenous to oil sands tailings ponds

By

Boon-Fei Tan

A thesis submitted to the Faculty of Graduate Studies and Research  
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy  
in  
Microbiology and Biotechnology

Department of Biological Sciences  
©Boon-Fei Tan  
Spring 2014  
Edmonton, Alberta

Permission is hereby granted to the University of Alberta Libraries to reproduce single copies of this thesis and to lend or sell such copies for private, scholarly or scientific research purposes only. Where the thesis is converted to, or otherwise made available in digital form, the University of Alberta will advise potential users of the thesis of these terms.

The author reserves all other publication and other rights in association with the copyright in the thesis and, except as herein before provided, neither the thesis nor any substantial portion thereof may be printed or otherwise reproduced in any material form whatsoever without the author's prior written permission.

## Abstract

In oil sands tailings ponds, methanogenesis is driven in part by the degradation of hydrocarbons in residual solvents used as a diluent during bitumen extraction, such as naphtha. Alkanes constitute a large proportion of these unrecovered hydrocarbons in mature fine tailings (MFT). Methanogenic degradation of alkanes has been poorly described in the literature. Fumarate addition, widely reported for activation of alkanes and monoaromatic compounds under sulfate- and nitrate-reducing conditions, has not been demonstrated conclusively for alkane degradation under methanogenic conditions because signature metabolites and key organisms have not been detected and/or isolated. In order to understand methanogenic alkane degradation by microorganisms indigenous to oil sands tailings ponds, a model alkane-degrading culture (SCADC) was established using MFT obtained from Mildred Lake Settling Basin (MLSB) tailings pond.

SCADC degraded many lower molecular weight alkanes, represented by  $n$ -C<sub>6</sub>-C<sub>10</sub>, 2-methylpentane and methylcyclopentane during year-long methanogenic incubation, but expected fumarate addition products were only detected for 2-methylpentane and methylcyclopentane. Nucleic acids isolated from SCADC were subjected to metagenomic and metatranscriptomic analysis using Illumina Hi-Seq. Metagenomic binning using multiple approaches recovered several partial genomes, including novel syntrophic *Desulfotomaculum* and *Smithella* spp. that are genetically capable of fumarate addition, which was previously unknown. Metatranscriptomic analysis further confirmed the high expression of genes encoding enzymes for alkane addition to fumarate by

*Desulfotomaculum* but not *Smithella* during active methanogenesis, indicating the importance of Firmicutes in fumarate activation of low molecular weight alkanes.

Data mining of metagenomes of MLSB and hydrocarbon-impacted environments recovered novel fumarate addition genes undescribed previously, indicating the overall ubiquitous nature of these genes in anoxic environments. Comparative metagenomic analysis of SCADC to two other metagenomes of methanogenic toluene- and naphtha-degrading cultures, in addition to physiological studies, suggests that fumarate addition may be the bottleneck reaction in these three cultures. The cultures have the genetic capability of degrading structurally diverse hydrocarbons and share highly conserved and streamlined functions for anaerobic respiration and methanogenesis, unlike *in situ* environments impacted by hydrocarbons, which are highly variable in their functional capabilities. This observation provides future prospects for development of commercial cultures for bioremediation and biomethanization.

## **Acknowledgements**

First and foremost, I would like to thank my family for understanding that I have to be away for so long to pursue a dream that seemed too far to reach. I quote a note my mom wrote me a long while ago when I was still a Master's student (translated from Chinese), "Son, time waits for no man; your youth will fade away quickly as time goes by if you do not treasure it; if you are sure that this is something that you must do, then you must begin and finish it". I also remember how I told her I wanted to go home so badly the first winter I was in Winnipeg (during my undergraduate years); She reprimanded me and determined that I had to stay and finish what I had promised to do. She is so loving and supportive - in her own special way.

Much gratitude to my supervisor Dr. Julia Foght for being a great mentor and who taught me a lot during the program. Her joyful character and love of work will always be a reminder that any work can be fun as long as you have a passion for it. Much thanks to friends and colleagues that I have spent years working alongside at the Foght Lab and in the Micro RIG. Some have helped directly in my research; others have provided feedback and moral support. Without them, the endeavour would for sure have been a lot less interesting. I would like to especially thank Kathy Semple, Rozyln Young and Anh Dao for their help in metabolite analyses; Carmen Li who has tirelessly prepared DNA samples for sequencing; Dr. Camilla Nesbø for advice in bioinformatics analysis (and reminders that names with an ø should always be spelt as such); Sara Ebert for teaching me anaerobic techniques; Drs. Abigail Adebusuyi and Nidal Abu-Laban for collaboration and advice in science and "life" in general. In particular, I must thank Albert Remus-Rosana who helped so much with optimizing RNA isolation (with many trials even on weekends and overnight), without whom the RNA-seq analysis presented in this thesis would not have been possible. Special thanks to Dr. Rebecca Case for advice and providing positive feedback on many questions, and Dr. Yan Boucher for being my candidacy examiner on short notice.

My gratitude also goes to members of the Hydrocarbon Metagenomics Project, especially the bioinformatics group and Jane Fowler (University of Calgary) who I collaborated with in comparative metagenomics analysis. Their technical expertise has proven to be very helpful in my data analysis.

Funding for this research was provided by Genome Canada and Genome Alberta. Syncrude Canada, Ltd. provided the original tailings pond samples for this project.

## Table of Contents

<b>1</b>	<b>Introduction .....</b>	<b>1</b>
1.1	Oil sands tailings ponds: a brief overview .....	1
1.1.1	Methanogenesis in oil sands tailings ponds .....	2
1.1.2	Microbiology of oil sands tailings ponds .....	3
1.1.3	Essential nutrients for microbial activities .....	4
1.2	Anaerobic biodegradation of petroleum hydrocarbons .....	6
1.2.1	Methanogenic alkane degradation: Community structures and primary alkane degraders .....	7
1.2.2	How are <i>n</i> -alkanes biodegraded under methanogenic conditions? .....	9
1.2.3	Proposed alkane-degradation pathways .....	12
1.2.4	Fumarate addition enzymes can exhibit relaxed specificity: How is this important?.....	13
1.2.5	Is addition to fumarate a ubiquitous mechanism for hydrocarbon degradation? .....	15
1.3	Thesis overview and research objectives .....	16
1.3.1	Research objectives .....	16
1.3.2	Thesis outline .....	17
1.4	Additional research outside the scope of the thesis .....	20
1.5	References .....	21
<b>2</b>	<b>Evidence for activation of aliphatic hydrocarbons by addition to fumarate under methanogenic conditions.....</b>	<b>28</b>
2.1	Abstract .....	28
2.2	Introduction .....	29
2.3	Experimental Procedures.....	31
2.3.1	Incubation of enrichment cultures .....	31
2.3.2	Analysis of methane, sulfide and volatile hydrocarbons in culture bottles .....	32
2.3.3	Detection of putative metabolites .....	33
2.3.4	Nucleic acid extraction, analyses of microbial community and RT-PCR of functional genes .....	34
2.3.5	Bioinformatics and phylogenetic analysis.....	35

2.4	Results .....	36
2.4.1	Alkane biodegradation under methanogenic and sulfidogenic conditions .....	36
2.4.2	Detection of putative succinylated metabolites of <i>n</i> -, iso- and <i>cyclo</i> -alkanes during methanogenesis and sulfidogenesis .....	39
2.4.3	Reverse transcription-PCR of expressed <i>assA</i> and <i>bssA</i> genes .....	45
2.4.4	Microbial community structure in methanogenic and sulfidogenic cultures .....	47
2.5	Discussion .....	53
2.5.1	Degradation of aliphatic alkanes by methanogenesis.....	53
2.5.2	Evidence that methanogenic degradation of alkanes proceeds by addition to fumarate.....	54
2.5.3	Inferred role of Peptococcaceae in methanogenic alkane degradation	56
2.6	Significance .....	58
2.7	References .....	59
<b>3</b>	<b>Metagenomic analysis of an anaerobic alkane-degrading microbial culture: Potential hydrocarbon-activating pathways and inferred roles of community members .....</b>	<b>64</b>
3.1	Abstract .....	64
3.2	Introduction .....	65
3.3	Materials and methods.....	66
3.3.1	Description of SCADC enrichment culture.....	66
3.3.2	Nucleic acid extraction and high throughput sequencing.....	67
3.3.3	Quality control and <i>de novo</i> assembly of metagenomic data .....	68
3.3.4	Taxonomic classification of metagenomics reads .....	69
3.3.5	Phylogenetic and bioinformatics analyses.....	69
3.4	Results and discussion.....	70
3.4.1	Anaerobic biodegradation of alkanes by SCADC.....	70
3.4.2	Features of the SCADC enrichment culture and metagenome.....	70
3.4.3	Taxonomic classification of PCR-amplified 16S rRNA genes, contigs and unassembled reads .....	71
3.4.4	Recruitment of SCADC reads to previously sequenced genomes .....	74
3.4.5	Presumptive hydrocarbon-activating mechanisms in SCADC.....	78

3.4.6	Evidence for genes encoding alternative hydrocarbon-activating pathways.....	85
3.4.7	Roles of community members: general overview .....	88
3.5	Significance .....	93
3.6	References .....	94
<b>4</b>	<b>Metagenomic and metatranscriptomic analyses of a model alkane-degrading enrichment culture implicate roles for <i>Desulfotomaculum</i> and <i>Smithella</i> in anaerobic alkane degradation .....</b>	<b>105</b>
4.1	Abstract .....	105
4.2	Introduction .....	106
4.3	Materials and Methods .....	107
4.3.1	Description of SCADC enrichment culture.....	107
4.3.2	Total RNA isolation and quality control of sequence reads.....	108
4.4	Bioinformatic analysis of SCADC metagenome and metatranscriptome .....	109
4.5	Results and discussion.....	115
4.5.1	Binning of SCADC metagenome .....	115
4.5.2	Metatranscriptomics reveals that <i>Desulfotomaculum</i> SCADC is the most metabolically active member.....	120
4.5.3	Metagenomic analysis implicates <i>Smithella</i> SCADC in alkane activation by addition to fumarate.....	129
4.5.4	Aceticlastic and hydrogenotrophic methanogenesis .....	131
4.5.5	Syntrophic interactions during alkane degradation .....	138
4.5.6	Expression of genes encoding transposases and prophage activation in SCADC.....	142
4.6	Conclusions .....	144
4.7	References .....	145
<b>5</b>	<b>Comparative metagenomic analysis of three methanogenic hydrocarbon-degrading enrichment cultures and relevant environmental metagenomes .....</b>	<b>153</b>
5.1	Abstract .....	153
5.2	Introduction .....	154

5.3	Materials and methods.....	156
5.3.1	Incubation conditions .....	156
5.3.2	Total DNA extraction for metagenomic sequencing.....	158
5.3.3	Quality control, <i>de novo</i> assembly and annotation of metagenomic data 158	
5.3.4	Comparative analysis of functional categories in metagenomic datasets 158	
5.3.5	Phylogenetic analysis of putative fumarate addition enzymes.....	159
5.4	Results and discussion.....	160
5.4.1	Descriptions of TOLDC, SCADC and NAPDC cultures.....	160
5.4.2	General features of NAPDC, SCADC and TOLDC metagenomes ...	160
5.4.3	TOLDC, SCADC, NAPDC share similar methanogenic communities but differ in dominant microbial communities putatively involved in hydrocarbon degradation.....	162
5.4.4	TOLDC, SCADC and NAPDC share many key metabolic and anaerobic hydrocarbon degradation pathways .....	169
5.4.5	Putative genes associated with hydrocarbon activation .....	169
5.4.6	TOLDC, SCADC and NAPDC communities have similar genetic capabilities but differ from natural environments previously exposed to hydrocarbons .....	173
5.5	Conclusions .....	181
5.6	References .....	183
<b>6</b>	<b>Assessing the functional diversity of fumarate addition genes in oil sands tailings ponds and other environments.....</b>	<b>190</b>
6.1	Abstract .....	190
6.2	Introduction .....	191
6.3	Materials and methods.....	192
6.3.1	Microcosm preparation and headspace gas chromatographic analyses.....	192
6.3.2	Nucleic acid extraction.....	193
6.3.3	Primer design and validation.....	194
6.3.4	PCR amplification, cloning and sequencing of partial <i>assA</i> and <i>bssA</i> .....	195

6.3.5	<i>assA</i> and <i>bssA</i> phylogeny and community analyses.....	196
6.3.6	qPCR quantification of <i>assA</i> and <i>bssA</i> phylogenies.....	197
6.4	Results and discussion.....	197
6.4.1	Bioconversion of naphtha components to methane.....	197
6.4.2	Diversity of fumarate addition genes in methanogenic hydrocarbon-degrading microcosms recovered through cloning and sequencing...	199
6.4.3	Increased abundance of <i>assA</i> and <i>bssA</i> during incubation suggests functional roles in hydrocarbon activation by fumarate addition.....	203
6.4.4	Diversity of fumarate addition genes in oil sands tailings ponds.....	207
6.4.5	MLSB tailings harbours novel clusters of fumarate addition genes ..	208
6.4.6	Higher abundance of <i>assA</i> and <i>nmsA</i> genes in MLSB and TP6 reflects the higher concentration of residual alkanes in tailings ponds.....	211
6.4.7	Abundance of fumarate addition genes in other environments.....	212
6.5	Significance .....	213
6.6	References .....	213
<b>7</b>	<b>Conclusion and synthesis .....</b>	<b>218</b>
7.1	Understanding alkane degradation using a model system.....	218
7.2	Opportunities in developing commercial cultures for bioremediation .....	220
7.3	Fumarate addition is a universal ubiquitous anaerobic hydrocarbon-activating mechanism.....	220
7.4	Petroleum microbiology in the genomic era .....	221
7.5	References .....	222
<b>8</b>	<b>Appendix A: Re-analysis of ‘omics data provides evidence for addition of <i>n</i>-hexadecane to fumarate under methanogenic conditions .....</b>	<b>223</b>
<b>9</b>	<b>Appendix B - Supporting information for Chapter 2 .....</b>	<b>229</b>
<b>10</b>	<b>Appendix C - Supporting information for Chapter 3 .....</b>	<b>233</b>
<b>11</b>	<b>Appendix D - Supporting information for Chapter 4 .....</b>	<b>246</b>
<b>12</b>	<b>Appendix E - Supporting information for Chapter 5 .....</b>	<b>253</b>
<b>13</b>	<b>Appendix F - Supporting information for Chapter 6 .....</b>	<b>255</b>

## List of Tables

<b>Table 2.1</b> Putative alkylsuccinate metabolites detected in silylated culture extracts from methanogenic and sulfidogenic alkane-degrading enrichment cultures.....	42
<b>Table 3.1</b> Taxonomic classification of the 15 most abundant OTUs in SCADC determined using 16S rRNA gene amplicon pyrosequencing followed by OTU clustering (at 5% distance level) and taxonomic assignment with the RDP classifier. ....	72
<b>Table 3.2</b> Summary of predicted protein and nucleotide recruitment of 454 and Illumina hybrid contigs to selected reference genomes acquired from NCBI. ....	77
<b>Table 4.1</b> Completeness of selected genomic bins obtained from SCADC metagenome sequence assembly .....	117
<b>Table 5.1</b> Enrichment of hydrocarbon-degrading methanogenic cultures and incubation conditions.....	157
<b>Table 5.2</b> Features of the metagenomes generated by 454 pyrosequencing of three hydrocarbon-degrading methanogenic enrichment cultures.	161
<b>Table 5.3</b> The five most abundant bacterial and archaeal operational taxonomic units (OTUs) in enrichment cultures determined by PCR amplification of 16S rRNA genes from metagenomic DNA.....	163
<b>Table 6.1</b> Primer sequence used in this study .....	195

## List of Figures

<b>Figure 1.1</b>	Anaerobic activation of substituted monoaromatic hydrocarbons (TEX: toluene, ethylbenzene and xylenes) and <i>n</i> -alkanes by fumarate addition.....	10
<b>Figure 1.2</b>	Proposed pathway for anaerobic degradation of <i>n</i> -alkanes, initiated by addition to fumarate and producing fatty acids for beta-oxidation.....	11
<b>Figure 1.3</b>	History of enrichment culture SCADC used in the experiments described in Chapters 2, 3, 4 and 5.....	20
<b>Figure 2.1</b>	(A) Methane and volatile hydrocarbons detected in the headspace of methanogenic cultures, and (B) soluble sulfide and volatile hydrocarbons detected in sulfidogenic cultures during incubation at 28°C.....	37
<b>Figure 2.2</b>	Methane detected in the headspace of methanogenic cultures amended with individual “hexanes”, <i>n</i> -C <sub>7</sub> and <i>n</i> -C <sub>8</sub> or <i>n</i> -C <sub>10</sub> incubated for 54 weeks.....	39
<b>Figure 2.3</b>	GC-MS profiles of silylated alkylsuccinates detected in the sulfate-reducing C <sub>6</sub> -C <sub>10</sub> -degrading culture incubated for 37 weeks.....	40
<b>Figure 2.4</b>	GC-MS profiles of silylated succinates detected in methanogenic enrichment culture with methycyclopentane as the only carbon source and incubated for 27 weeks (Top); methanogenic enrichment culture with 2-methylpentane as the only carbon source, incubated for 27 weeks (bottom). ....	43
<b>Figure 2.5</b>	Mass spectra of derivatized putative (1,3-dimethylpentyl)succinic acid and methylcyclopentylsuccinic acid arising from addition of 2-methylpentane and methylcyclopentane, respectively, to fumarate.....	44
<b>Figure 2.6</b>	Maximum Likelihood Tree of deduced amino acid sequences from cloned <i>assA</i> and <i>bssA</i> transcripts obtained from a methanogenic alkane-degrading culture. ....	46
<b>Figure 2.7</b>	Proportion of bacterial taxa in a methanogenic culture degrading a mixture of C <sub>6</sub> -C <sub>10</sub> alkanes after 1 day, 40 weeks and 81 weeks incubation.....	48
<b>Figure 2.8</b>	Bacterial community structure of methanogenic alkane-degrading enrichment cultures individually amended with ‘hexanes’ (containing <i>n</i> -C <sub>6</sub> , methylcyclopentane, 2-methylpentane and 3-methylpentane) or <i>n</i> -C <sub>7</sub> or <i>n</i> -C <sub>8</sub> and incubated for 54 weeks.....	49
<b>Figure 2.9</b>	Proportion of methanogen taxa in a methanogenic alkane-degrading culture after 1 day, 40 weeks and 81 weeks incubation.....	50

<b>Figure 2.10</b>	Proportion of bacterial taxa in an alkane-degrading culture 1 day, 22 weeks, 37 weeks and 45 weeks after transfer from methanogenic to sulfate-reducing conditions. ....51
<b>Figure 2.11</b>	Comparison of differentially abundant Operational Taxonomic Units (OTUs) in methanogenic and sulfidogenic enrichment cultures determined using Metastats. ....53
<b>Figure 3.1</b>	Relative abundance of taxa at the Domain level and Phylum level in unassembled 454 pyrosequencing reads, determined using SOrt-ITEMS.....74
<b>Figure 3.2</b>	Maximum likelihood tree of putative <i>assA</i> , <i>nmsA</i> and <i>bssA</i> genes recovered from SCADC metagenome in 454 and Illumina assemblies as contigs .....80
<b>Figure 3.3</b>	Comparison of the <i>ass</i> operon 1 in <i>D. alkenivorans</i> AK-01 with two SCADC contigs harbouring <i>ass</i> operon analogues.. .....84
<b>Figure 3.4</b>	Taxonomic distribution of selected COG categories representing broad metabolic functions pertaining to the SCADC enrichment culture .....91
<b>Figure 4.1</b>	Schematic representation of the "omics" workflow established for analysis of SCADC metagenome and metatranscriptome generated using Illumina Hi-seq. ....110
<b>Figure 4.2</b>	Example of clustering of metagenome contigs based on sequence coverage and GC content... .....111
<b>Figure 4.3</b>	Example of clustering of metagenome contigs based on sequence coverage and GC content. ....112
<b>Figure 4.4</b>	Example of Principal Component Analysis of the tetranucleotide frequencies composition of SCADC metagenome contigs binned using GC content versus coverage. ....113
<b>Figure 4.5</b>	Example of Emergent Self-Organizing Map (ESOM) analysis of the tetranucleotide composition of contigs binned in Figure 4.3.. .....114
<b>Figure 4.6</b>	Maximum likelihood tree of the 16S rRNA gene of “ <i>Desulfotomaculum</i> ” SCADC and “ <i>Smithella</i> ” SCADC obtained from 16S rRNA gene pyrotag sequencing compared to reference 16S rRNA sequences and sequences originating from hydrocarbon-impacted environments.....119
<b>Figure 4.7</b>	Metatranscriptomic analysis of the relative abundance of genes encoding $\alpha$ -subunit of alkylsuccinate ( <i>assA</i> ), benzylsuccinate ( <i>bssA</i> ) and naphthylsuccinate synthase ( <i>nmsA</i> ), and genes encoding the $\alpha$ -subunit of methyl-CoM reductase ( <i>mcrA</i> ).. .....120
<b>Figure 4.8</b>	Reconstruction of proposed alkane degradation pathway in " <i>Desulfotomaculum</i> " SCADC during active methanogenesis ..126

<b>Figure 4.9</b>	Reconstruction of methanogenesis pathway in <i>Methanoculleus</i> SCADC during active methanogenesis.....	133
<b>Figure 4.10</b>	Reconstruction of methanogenesis pathway by <i>Methanosaeta</i> SCADC during active methanogenesis.....	136
<b>Figure 4.11</b>	Possible interspecies metabolite and hydrogen/formate transfer among the SCADC microbial community.....	141
<b>Figure 5.1</b>	Proportion of microbial communities at the taxonomic class level in TOLDC, SCADC, NAPDC and eight other metagenomes based on assignment of 454 metagenomic reads..	166
<b>Figure 5.2</b>	Maximum likelihood tree of translated <i>assA/bssA/nmsA</i> homologs recovered from TOLDC, SCADC and NAPDC using tBLASTn.....	171
<b>Figure 5.3</b>	Principal component analysis of the functional categories in 41 published metagenomes from diverse environments plus the metagenomes of the three hydrocarbon-degrading cultures..	174
<b>Figure 5.4</b>	Comparison of shared functional categories among the three hydrocarbon-degrading cultures (NAPDC, SCADC and TOLDC).....	176
<b>Figure 5.5</b>	Ternary plot showing three-way comparisons of functional categories in SEED subsystems level 3 for three different groups of metagenomes (GM; three Gulf of Mexico deep marine sediments; TP; two oil sands tailings ponds; HC: the three hydrocarbon-degrading cultures).....	177
<b>Figure 5.6</b>	Ternary plot showing three-way comparison of functional categories in SEED subsystems level 3 for different groups of metagenomes. The plot shows the comparison of shared functions in NAPDC, SCADC, and TOLDC (HCD) with soil (six) and marine (eleven) metagenomes. ....	182
<b>Figure 6.1</b>	Headspace methane detection in heat-killed and substrate-free controls compared to microcosms amended with different hydrocarbon substrates.....	198
<b>Figure 6.2</b>	Diversity of translated <i>assA/bssA/nmsA</i> genes detected in hydrocarbon-degrading microcosms and a substrate-free control established from oil sands tailings mature fine tailings..	201
<b>Figure 6.3</b>	Relative abundance of <i>assA</i> and <i>bssA</i> subtypes in microcosms amended with different hydrocarbon substrates and incubated for 150 d.....	204
<b>Figure 6.4</b>	qPCR quantification of selected <i>assA</i> and <i>bssA</i> genotypes as shown in Figure 6.2 (G1, G2 and G3)..	206
<b>Figure 6.5</b>	Diversity of translated <i>assA/bssA/nmsA</i> genes detected in the metagenome of Syncrude MLSB.....	208

## **List of abbreviations**

<i>assA</i>	alpha-subunit of alkylsuccinate synthase
<i>bssA</i>	alpha-subunit of benzylsuccinate synthase
ESOM	Emergent Self Organizing Map
<i>nmsA</i>	alpha-subunit of naphthylsuccinate synthase
PCA	Principal Component Analysis
SCADC	Short Chain Alkane Degrading Culture

# 1 Introduction

## 1.1 Oil sands tailings ponds: a brief overview

The Athabasca oil sand deposits span an area of 100,000 km<sup>2</sup> of Northern Alberta and the extraction of oil sands bitumen yields 20% of oil production in Canada representing an estimated 1.7 - 2.5 trillion barrels of bitumen (Fedorak *et al.*, 2002; Holowenko *et al.*, 2000). For surface-mined oil sands, bitumen extraction from oil sands ore is achieved using caustic hot water extraction and froth flotation (Chalaturnyk *et al.*, 2002). In this process, alkaline water at 85°C is used in a conditioning process that promotes the disintegration of ore structure (containing bitumen, clays and fine particles) and allows the separation of bitumen from sand and clay particles (Chalaturnyk *et al.*, 2002). Light hydrocarbon diluents such as naphtha are used to recover additional bitumen and reduce its viscosity. The whole waste tailings slurry can contain 40 - 55 % solids (sands and clays) and 1 to 5% unrecovered bitumen and diluents. However, the tailings that flows into tailings ponds contain only ~ 8% solids with the sand dropping out to form dykes upon deposition. Much of the residual bitumen and diluent in tailings wastes is recovered before the tailings slurry is deposited in oil sands tailings ponds for long term storage (Chalaturnyk *et al.*, 2002). The tailings are allowed to dewater in conjunction with the slow settling (sedimentation) of fine tailings by gravitational force (Chalaturnyk *et al.*, 2002). About 80% of water released from the densification process is recycled and used in bitumen extraction. The settling of fine tailings can take a long time because the fines form a stable suspended network that densifies very slowly (Chalaturnyk *et al.*, 2002).

Current practice in the industry employs different techniques and strategies in accelerating settling of tailings fines. One approach is to use gypsum, whereby the Ca<sup>2+</sup> acts as the densification agent (Chalaturnyk *et al.*, 2002). Prior to being deposited into tailings ponds, tailings waste is mixed with gypsum to form composite tailings (CT) (Chalaturnyk *et al.*, 2002). The consolidation of composite tailings occurs in conjunction with water being released quickly through pores within the tailings layers (Fedorak *et al.*, 2003, Salloum *et al.*, 2002). The settled tailings, also termed mature fine tailings (MFT), constitute an

average 30% solids by weight, and contain other materials such as fine silts, unrecovered residual bitumen and diluents (e.g., naphtha), naphthenic acid, salts (NaCl) and ions such as  $\text{NO}_3^-$ ,  $\text{SO}_4^{2-}$ ,  $\text{Fe}^{3+}$  and  $\text{Mg}^{2+}$  (Fedorak *et al.*, 2002, Penner and Foght 2010). Over time, oil sands tailings beneath the water cap have become highly methanogenic in some ponds (Fedorak *et al.*, 2002, Holowenko *et al.*, 2000, Salloum *et al.*, 2002).

There are currently four surface-mining oil sands extraction companies in Fort McMurray: Suncor Energy Inc., Syncrude Canada Ltd., Shell Canada Ltd. and Canadian Natural Resources Limited (BGC Engineering Inc., 2010). Tailings from each company are stored in different ponds. The ponds referred to as Tailing pond 6 (TP6) and Mildred Lake Settling Basin (MLSB) in the later section of this Chapter are independently managed by Suncor and Syncrude, respectively. Because this thesis research has only been focusing on the materials obtained from MLSB, the remaining review will largely pertain to MLSB, unless stated otherwise.

### **1.1.1 Methanogenesis in oil sands tailings ponds**

An estimated 40 million L of methane (or 17 L/ m<sup>2</sup>) is released from the surface layer of Syncrude's MLSB on a daily basis (Holowenko *et al.*, 2000) High concentrations of  $\text{SO}_4^{2-}$  reported in tailings water [i.e., 0.1 - 36 mg/L depending on sample depth (Penner and Foght, 2010; Fedorak *et al.*, 2002)], due to gypsum addition for tailings consolidation, can inhibit methanogenesis (Salloum *et al.*, 2002). Once the sulfate concentration is reduced to  $\leq 20$  mg/L, as in MLSB, methanogenesis can begin, suggesting the long-term survival of methanogens within oil sands tailings ponds in the presence of competing SRB (Fedorak *et al.*, 2003, Holowenko *et al.*, 2000, Salloum *et al.*, 2002). Management of the sulfur cycle in oil sands tailings ponds therefore has been subject of interest since it is important for H<sub>2</sub>S and CH<sub>4</sub> controls (Chen *et al.*, 2013, Fedorak *et al.*, 2003, Fru *et al.*, 2013, Ramos-Padron *et al.*, 2011, Salloum *et al.*, 2002).

### 1.1.2 Microbiology of oil sands tailings ponds

The microbial communities of MLSB (Fedorak *et al.*, 2002, Fedorak *et al.*, 2003, Holowenko *et al.*, 2000, Salloum *et al.*, 2002, Siddique *et al.*, 2006, Siddique *et al.*, 2007) and TP6 (An *et al.*, 2013, Ramos-Padron *et al.*, 2011) have been described using various methods. The water cap of MLSB is dominated by aerobic microorganisms including methanotrophs capable of methane oxidation (Saidi-Mehrabad *et al.*, 2013). Microbial activities at the interface between the water cap and underlying tailings can rapidly deplete available O<sub>2</sub> and contribute to the formation of a sulfidogenic zone, where there is a high concentration of sulfide (Chen *et al.*, 2013, Fru *et al.*, 2013). Below the water cap, the tailings sediments can become highly stratified (Fedorak *et al.*, 2002, Fedorak *et al.*, 2003, Holowenko *et al.*, 2000, Ramos-Padron *et al.*, 2011, Salloum *et al.*, 2002) and dominated by a high diversity of anaerobes and facultative anaerobes (Holowenko *et al.*, 2000, Penner and Foght 2010).

In MLSB MFT, most probable number (MPN) analysis shows that there can be 10<sup>5</sup> to 10<sup>6</sup> methanogens/g MFT and 10<sup>4</sup> to 10<sup>5</sup> sulfate-reducing bacteria/g MFT (Holowenko *et al.*, 2000, Penner and Foght 2010). Denitrifiers and ferric iron (III) reducers also are typically present in high numbers (Bordenave *et al.*, 2010, Fedorak *et al.*, 2002, Salloum *et al.*, 2002). In the deep layers of Suncor TP6 MFT (e.g., 5 to 15 m below the water surface), there is an abundance of syntrophs, sulfate-reducing bacteria, secondary fermentors and methanogens (Ramos-Padron *et al.*, 2011). In the deepest layers of MFT (e.g., >15 to 30 m below water surface) in both MLSB and TP6, sulfate concentrations and methanogenesis rates are lower, corresponding to the decreased number of SRB and methanogens (Holowenko *et al.*, 2000, Penner and Foght 2010, Ramos-Padron *et al.*, 2011). The methanogen populations in both MLSB and TP6 are dominated by acetoclastic *Methanosaeta* and hydrogenotrophic methanogens such as *Methanoregula*, *Methanocalculus* and *Methanoculleus* (Penner and Foght 2010, Ramos-Padron *et al.*, 2011). The bacterial communities are more diverse than the Archaea, and populations vary across the length and depth of tailings ponds (Penner and Foght 2010, Ramos-Padron *et al.*, 2011).

At certain depths in MLSB and TP6, the bacterial communities can comprise primarily Proteobacteria, in particular Gamma- and Betaproteobacteria (An *et al.*, 2013, Penner and Foght 2010). Similar microbial communities with dominant Betaproteobacteria were also observed in biofilm structures formed by tailings microorganisms in laboratory conditions (from a water sample collected 0.45 m from the surface of the tailings pond) (Golby *et al.*, 2012). Members of the Betaproteobacteria can have a wide range of physiological capabilities. For example, nitrate-reducers such as *Thauera* and *Azoarcus* found in MLSB (Penner and Foght 2010) are capable of anaerobic hydrocarbon degradation coupled to nitrate reduction (reviewed by Foght 2008). However, the significance of this is unknown since nitrate is present at very low concentrations in MFT (Fru *et al.*, 2013, Penner and Foght 2010). Other community members including Pseudomonads and *Acidovorax* are capable of aerobic hydrocarbon degradation. The detection of high abundance of facultative anaerobes and aerobes in hydrocarbon resource environments has been highlighted by An *et al.*, (2013) who suggested the genetic potential for important *in situ* aerobic processes, although this remain to be demonstrated.

### **1.1.3 Essential nutrients for microbial activities**

Oil sands tailings ponds are generally limited in the source of labile organic carbon and inorganic essential nutrients, i.e., phosphate and nitrogen, essential for microbial life (Fedorak *et al.*, 2003, Fru *et al.*, 2013, Penner and Foght 2010). The methanogenic substrates and/or carbon source in MLSB has been attributed in part to the unrecovered residual naphtha biodegradable by microorganisms indigenous to the tailings ponds (Siddique *et al.*, 2006, Siddique *et al.*, 2007, Siddique *et al.*, 2011, Siddique *et al.*, 2012). In MFT, large proportions of the unrecovered residual naphtha consist of mainly aliphatic alkanes, i.e., *iso*-alkanes, *cyclo*-alkanes and *n*-alkanes (e.g., 6000 mg/kg C<sub>16</sub> to C<sub>21</sub>; 480 mg/kg C<sub>10</sub> to C<sub>12</sub>; 160 mg/kg C<sub>8</sub> to C<sub>10</sub>) and lower concentrations of aromatic compounds (e.g., <15 mg/kg of monoaromatics and other polyaromatic compounds such as phenanthrene and naphthalene) (Siddique *et al.*, 2006) However, not all components of naphtha, i.e., *cyclo*-, *iso*-alkanes and benzene,

could be readily and/or rapidly biodegraded under methanogenic conditions. Some naphtha components require a long acclimatization period before methane evolution (Siddique *et al.*, 2007, Siddique *et al.*, 2011). Naphthenic acids, which are present in high concentrations in tailings ponds and are toxic to fish and other organisms, are also likely not significant methanogenic substrates in tailings (Holowenko *et al.*, 2001). Conversely, microbial products such as acetate have been detected in MFT (Suncor TP6), indicating that these metabolites are being produced and used by the community for methane productions (Ramos-Padron *et al.*, 2011).

Essential minerals such as  $Mg^{2+}$ ,  $Zn^{2+}$  and other essential co-factors in enzyme and protein synthesis are either present in low or undetectable concentrations (Chen *et al.*, 2013, Penner and Foght 2010). The atmospheric temperature is highly variable and influenced by seasonal change, but in the deeper layers of the MFT temperature is more stable throughout the year ( $\sim 12$  °C) and can be favourable for growth and long term survival of psychrotolerant and mesophilic microbes (Penner and Foght 2010, Ramos-Padron *et al.*, 2011). Overall, oil sands tailings ponds are limited in labile carbon sources and essential nutrients for microbial growth and only microorganisms that are best suited for syntrophic growth with other community members under highly reduced conditions are able to thrive. These organisms therefore consist of (A) bacteria that can perform the initial activation and degradation of naphtha components to simple organic compounds; (B) methanogens that are able to use  $CO_2/H_2$  and/or acetate for methanogenesis; and (C) syntrophs that are essential in providing nutrients such as vitamins and amino acids to community members that are incapable of synthesising these compounds. Indeed, oil sands tailings ponds represent a challenging environment that may have enabled evolution of fascinating machineries for growth and survival under methanogenic environments.

## 1.2 Anaerobic biodegradation of petroleum hydrocarbons

Even though petroleum hydrocarbons in general are harmful to most life forms, some microbes can thrive by using these as electron donors and carbon sources. Aerobic hydrocarbon biodegradation is generally well studied and documented (Diaz *et al.*, 2013, Wang and Shao 2013) and microorganisms capable of initiating aerobic attack do so by using enzymes including mono- and di-oxygenases. In the degradation process, O<sub>2</sub> acts as both an electron acceptor and co-substrate (Diaz *et al.*, 2013). Contrary to the rapid biodegradation of alkanes and aromatics under aerobic conditions, the biochemical transformation of hydrocarbons under strictly anaerobic conditions and thermodynamically-constrained conditions is slower and under certain limiting conditions such as methanogenic conditions, a consortium of organisms is required for anaerobic hydrocarbon degradation to occur (reviewed by Agrawal and Gieg 2013, Callaghan 2013, Carmona *et al.*, 2009, Foght 2008, Heider 2007).

Naphtha, which is one of the diluents used for bitumen extraction, contains *n*-paraffins (C<sub>4</sub>-C<sub>12</sub>, 18% by weight), *iso*-paraffins (C<sub>5</sub>-C<sub>11</sub>, 31%), naphthenes (*cyclo*-C<sub>5</sub>-C<sub>10</sub>, 27%), monoaromatics BTEX compounds (benzene, toluene, ethylbenzene and xylenes, 15%) and others (Siddique *et al.*, 2007). Of these, only the *n*-alkanes, toluene, ethylbenzene, *m*- and *o*-xylenes have been reported to be biodegradable in methanogenic laboratory cultures (Siddique *et al.*, 2006, Siddique *et al.*, 2007, Siddique *et al.*, 2011). Other components in bitumen consist of longer chain alkanes and many refractory components, some of which such as tricyclic terpenes are recalcitrant to biodegradation under both aerobic and anaerobic conditions (Head *et al.*, 2003).

Under all anaerobic conditions such as nitrate-, sulfate- and iron-reducing and methanogenic conditions, monoaromatics (i.e., BTEX), polyaromatics (i.e., methylnaphthalene, phenanthrene) and *n*-alkanes can be readily biodegraded by pure isolates and enrichment cultures (Abu Laban *et al.*, 2010, Agrawal and Gieg 2013, Callaghan 2013, Carmona *et al.*, 2009, Foght 2008, Heider 2007, Zhang *et al.*, 2013). Substrate degradation coupled to methanogenesis is generally energetically less favourable than sulfate-, nitrate- and iron-reducing conditions

(Dolfing *et al.*, 2008). Even so, crude oil *n*-alkanes can be biodegraded by anaerobic microorganisms under methanogenic conditions (Gray *et al.*, 2011, Head *et al.*, 2003). Anaerobic degradation of organic compounds under methanogenic conditions can proceed to completion so long as end-products from fermentation are rapidly removed by methanogens or sulfate reducers (Dolfing *et al.*, 2008, Jackson and McInerney 2002).

Because alkanes constitute a large fraction of the hydrocarbons in oil sands tailings ponds and petroleum reservoirs, understanding methanogenic *n*-alkane degradation is crucial for optimizing bioremediation in oil sands tailings ponds for long-term management (Penner and Foght 2010, Siddique *et al.*, 2008), modelling formation of heavily biodegraded petroleum in oil reservoirs (Jones *et al.*, 2008) and potentially for converting unrecovered residual hydrocarbons in petroleum reservoirs to methane for recovery as a valuable fuel (Gieg *et al.*, 2008).

### **1.2.1 Methanogenic alkane degradation: Community structures and primary alkane degraders**

*n*-Alkane degradation under sulfate-, nitrate-reducing and methanogenic conditions has been frequently reported in the literature, but none has been reported under iron-reducing conditions (reviewed by Agrawal and Gieg 2013, Callaghan 2013, Mbadinga *et al.*, 2011). Alkane-degrading microorganisms have been isolated from sulfate- and nitrate-reducing enrichment cultures, but not from predominantly methanogenic enrichment cultures or environments. These *n*-alkane degraders exclusively consist of members of the Beta- and Deltaproteobacteria, with a single report in the literature that implicated a Firmicutes affiliated with *Desulfotomaculum* of being capable of using propane under sulfate-reducing conditions (Kniemeyer *et al.*, 2007). Following isolation, Betaproteobacterium *Azoarcus* spp. (Grundmann *et al.*, 2008, Wilkes *et al.*, 2003), Deltaproteobacterium *Desulfatibacillum alkenivorans* AK-01 (Callaghan *et al.*, 2008, So and Young 1999) and *Desulfococcus oleovorans* Hxd3 (So *et al.*, 2003) have been used as models to investigate degradation mechanisms under their respective reducing conditions. A limited number of sulfate- and nitrate-reducing

alkane-degraders have been isolated and many of these appear to have a substrate spectrum biased towards *n*-alkanes with a certain chain length (Mbadinga *et al.*, 2011). For example, the Firmicutes affiliated with *Desulfotomaculum* are only able to use propane and butane coupled to sulfate-reduction but not longer chain alkanes (Kniemeyer *et al.*, 2007).

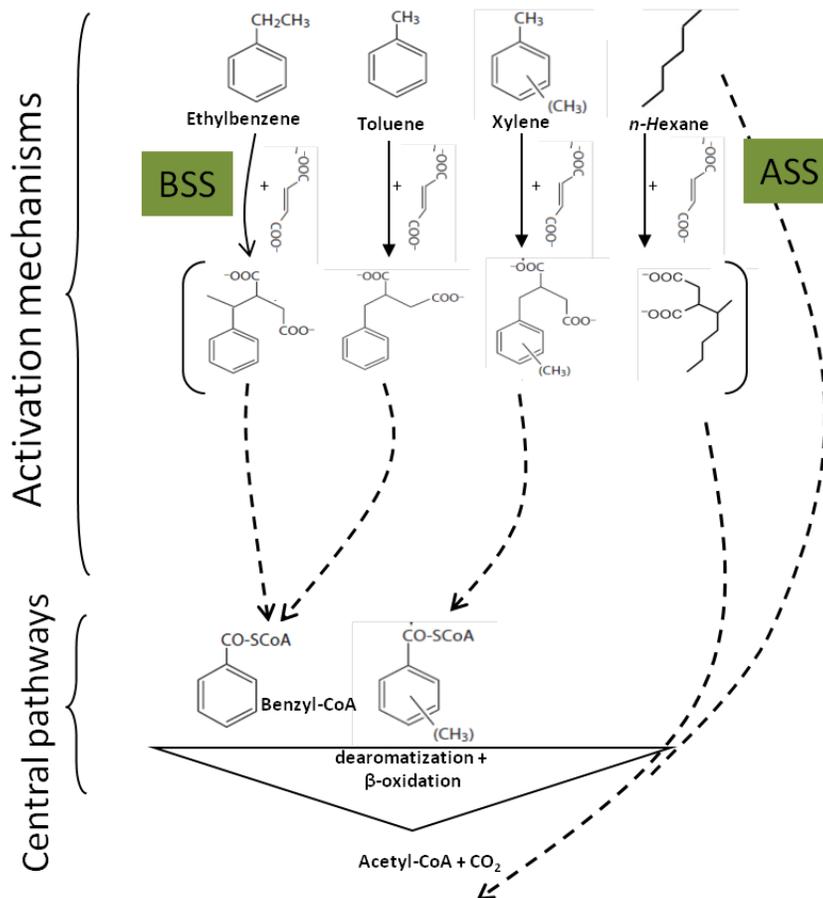
The biotransformation of *n*-hexadecane under methanogenic conditions was first shown to be carried out by a syntrophic community composed mainly of *Smithella* spp. and members of the hydrogenotrophic and acetoclastic methanogens (Zengler *et al.*, 1999). Microorganisms capable of alkane degradation under methanogenic conditions have rarely been isolated, with the exception of the sulfate-reducing bacterium *D. alkenivorans* AK-01 which is capable of growth on *n*-hexadecane with a methanogenic partner (Callaghan *et al.*, 2012). Many enrichment cultures shown to degrade *n*-alkanes typically comprise diverse bacterial and archaeal communities (Gray *et al.*, 2011, Li *et al.*, 2012, Siddique *et al.*, 2012, Zhou *et al.*, 2012), and the primary degraders in these systems are usually unknown. Numerous reports have previously indicated that *n*-alkanes can be readily biodegraded by a methanogenic community dominated by *Syntrophaceae* (Cheng *et al.*, 2013, Gray *et al.*, 2011, Siddique *et al.*, 2012). However, the only phylogenetically related isolate whose genome has been sequenced and is available for physiological study (*Syntrophus aciditrophicus* SB1) is not able to use *n*-alkanes as a growth substrate (McInerney *et al.*, 2007).

The inability or difficulty of isolating primary alkane degraders from methanogenic alkane degrading systems may be associated with the obligate or facultative syntrophic nature of the primary degrader with multiple secondary fermentors and/or methanogen partners (Kosaka *et al.*, 2008, Sieber *et al.*, 2012). The ability to study roles of primary degraders and potential mechanisms is significantly hindered because of the requirement to obtain an isolate for physiological study. To escape this requirement, high throughput sequencing of total DNA from a community has been used to obtain the total genetic information from a community, followed by computational inference of functions and taxonomy. Finally, physiological experimentation on the whole community

can be carried out in order to explain observations obtained from computational studies (Albertsen *et al.*, 2013, Haroon *et al.*, 2013).

### **1.2.2 How are *n*-alkanes biodegraded under methanogenic conditions?**

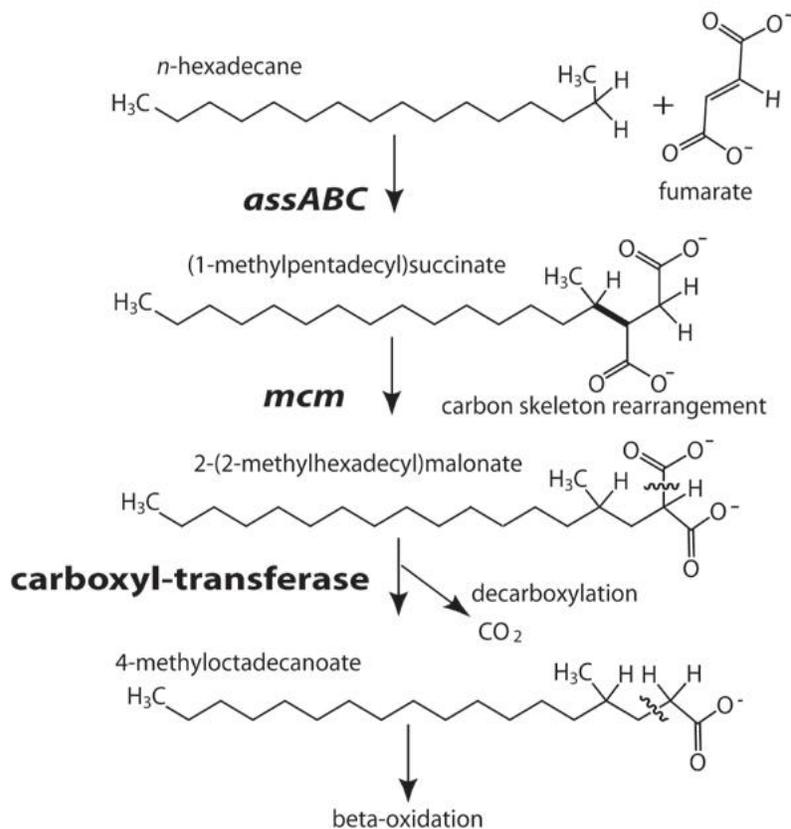
The only proven anaerobic alkane-activating mechanism is addition of the activated hydrocarbon across the double bond of fumarate (hereafter "fumarate addition"), catalyzed under sulfate- and nitrate-reducing conditions by a glycyl radical enzyme known as alkylsuccinate synthase (ASS) (Callaghan *et al.*, 2008, Grundmann *et al.*, 2008). A homolog of ASS is known as methyl-alkylsuccinate synthase (MAS) in the nitrate-reducing *Azoarcus* spp. HxN1 (Grundmann *et al.*, 2008). Other mechanisms have been proposed, but with limited evidence and requiring further confirmation (see below). Almost all knowledge about ASS is derived from studying benzylsuccinate synthase (BSS), a homolog of ASS that is involved in toluene addition to fumarate (Biegert *et al.*, 1996, Leuthner *et al.*, 1998, Rabus and Heider 1998). In brief, ASS, MAS and BSS all belong to the family of glycyl radical enzymes. These enzymes together with NMS (naphthyl-methylsuccinate synthase) are involved in the activation of structurally different hydrocarbons (reviewed by Foght 2008, Heider 2007). The primary substrates of ASS, BSS and NMS are *n*-alkanes, toluene and 2-methylnaphthalene, respectively; the enzymes add an activated substrate to the double bond of a fumarate molecule, forming a succinylated product (Figure 1.1) (Carmona *et al.*, 2009).



**Figure 1.1** Anaerobic activation of substituted monoaromatic hydrocarbons (TEX: toluene, ethylbenzene and xylenes) and n-alkanes by fumarate addition. Benzylsuccinate synthase (BSS) is involved in fumarate addition activation of TEX and has been reported to occur under methanogenic, sulfate-, nitrate- and iron-reducing conditions (reviewed by Foght, 2008). Alkylsuccinate synthase (ASS) is involved in fumarate addition activation of n-alkanes and had been reported to occur under mainly sulfate- and nitrate-reducing conditions (reviewed by Callaghan, 2013).

Alkane addition to fumarate was initially demonstrated in cultures of sulfate- and nitrate-reducing bacteria coupled with the detection of metabolites representing products of fumarate addition. Typically, the activating mechanism resulted in the formation of two diastereomers, produced from the addition to fumarate molecule of the subterminal carbon atom at two different positions (Callaghan *et al.*, 2006, Callaghan *et al.*, 2009, Kropp *et al.*, 2000, Wilkes *et al.*,

2002, Wilkes *et al.*, 2003). Further studies using the *n*-hexane-degrading *Azoarcus* sp. HxN1 showed that the fumarate addition product (1-methylpentyl)succinate is subsequently converted to a CoA-thioester, followed by carbon skeleton rearrangement to form 2-(2-methylhexyl)malonyl-CoA, which then undergoes decarboxylation to yield 4-methyloctanoyl-CoA (Figure 1.2). Subsequent conventional  $\beta$ -oxidation of 4-methyloctanoyl-CoA produces fatty acids that are two carbon atoms shorter than the substrates (Wilkes *et al.*, 2002). More recently the diastereomers produced from fumarate addition have been proposed to undergo an epimerization process to produce a single isomer prior to carbon skeleton rearrangement (Jarling *et al.*, 2012), however the gene and enzyme proposed for this function have not been confirmed experimentally.



**Figure 1.2** Proposed pathway for anaerobic degradation of *n*-alkanes, initiated by addition to fumarate and producing fatty acids for beta-oxidation. *assA*, aylsuccinate synthase; *mcm*, methylmalonyl-coA mutase. Figure adapted from Callaghan *et al.* (2012)

The genes encoding enzymes responsible for fumarate addition, carbon skeleton rearrangement and carboxylation have since been proposed based on proteomic study of *Azoarcus* spp. HxN1 and *D. alkenivorans* Ak-01 (Callaghan *et al.*, 2008, Grundmann *et al.*, 2008). Further genome annotation of *D. alkenivorans* Ak-01 showed that genes proposed for these processes are organized in a gene cluster, and therefore likely to be under similar genetic control (Callaghan *et al.*, 2012). Organisms capable of *n*-alkane addition to fumarate, therefore, can be expected to carry homologs of these genes. Currently only one genome of an *n*-alkane degrader (*D. alkenivorans* Ak-01) with known degrading mechanism (the other being *Desulfococcus oleovorans*, see below) is available in the public domain, precluding comparative genomic study to better understand the organization of gene clusters involved in fumarate addition.

The *assA* gene and its homolog *masD* (Grundmann *et al.*, 2008) encoding the  $\alpha$ -subunit of ASS is a homolog of *bssA* encoding the  $\alpha$ -subunit of benzylsuccinate synthase (BSS). Both genes are highly conserved in anaerobic hydrocarbon degraders and have been used as marker genes to indicate the functional potential of organisms and communities for hydrocarbon addition to fumarate (Acosta-González *et al.*, 2013, Callaghan *et al.*, 2010). Previously, the *assA* gene has been detected in methanogenic *n*-alkane-degrading cultures (Callaghan *et al.*, 2010, Gray *et al.*, 2011), but metabolites representing fumarate addition for alkanes have never been detected in methanogenic alkane-degrading cultures, invoking proposal of alternate mechanisms (Aitken *et al.*, 2013), discussed below.

### **1.2.3 Proposed alkane-degradation pathways**

Other anaerobic alkane activating mechanisms proposed in the literature include carboxylation (So *et al.*, 2003) and hydroxylation (Callaghan 2013). Alkane degradation by carboxylation is proposed to proceed with the addition of a carboxyl group (derived from bicarbonate) to the subterminal C3 atom of *n*-alkane, after which an ethyl group is removed, producing an odd-numbered fatty acid from an even numbered *n*-alkane, and vice versa. This putative mechanism was reported only for the Deltaproteobacterium *Desulfococcus oleovorans* Hxd3

(Callaghan *et al.*, 2006, Callaghan *et al.*, 2009, So *et al.*, 2003) whose genome did not contain an *assA* homolog and therefore must be using an alternative mechanism for alkane activation and degradation. However, the enzyme(s) and gene(s) proposed for the carboxylation mechanism have not been identified and confirmed. In addition, the proposed key metabolite 2-ethylpentadecanoic acid (following addition of a bicarbonate carbon at C3 of an *n*-alkane) was not detected (So *et al.*, 2003). Therefore, the proposed carboxylation mechanism of *n*-alkane activation is currently remains cryptic.

More recently an alternative mechanism has been proposed for the same organism. Instead of carboxylation, the activating mechanism is suggested to occur by means of hydroxylation using a homolog of ethylbenzene dehydrogenase, which is highly up-regulated in the presence of *n*-alkanes (Callaghan 2013). The proposed hydroxylation mechanism initiates attack on a C2 carbon and produces a secondary alcohol, which could be further oxidized to form a ketone. Following this, the ketone is then carboxylated at C3, followed by subsequent elimination of the C1 and C2 carbons to form a fatty acid that is one carbon shorter than the parent hydrocarbon (Callaghan 2013). Activation of hydrocarbons by means of hydroxylation is not unfounded because ethylbenzene has been shown to be degraded in a such way (Biegert *et al.*, 1996).

#### **1.2.4 Fumarate addition enzymes can exhibit relaxed specificity: How is this important?**

Most anaerobes capable of alkane addition to fumarate have a restricted substrate spectrum biased towards *n*-alkanes of certain chain length; therefore, the substrate spectrum is likely strain-specific. For example, nitrate-reducing Betaproteobacterium *Aromatoleum* OcN1 is capable of using *n*-C<sub>8</sub>-C<sub>12</sub>, whereas sulfate-reducing Deltaproteobacterium *Desulfoglaeba alkanexedens* ALDC is capable of using *n*-C<sub>6</sub>-C<sub>12</sub> (Davidova *et al.*, 2006). A marine sulfate reducer (BuS5) closely affiliated with *Desulfosarcina* or *Desulfococcus* can utilize *n*-propane by fumarate addition to C1 or C2 and it can also use *n*-butane by activating the molecule at the C2 carbon (Kniemeyer *et al.*, 2007). Despite this specificity, some enzymes catalyzing fumarate addition, in particular ASS, have

been shown to be highly promiscuous (Beller and Spormann 1999, Rabus *et al.*, 2011, Wilkes *et al.*, 2003) in that they can co-activate several structurally unrelated hydrocarbons, i.e. toluene and *cyclo*-alkanes present in a mixture, e.g., crude petroleum (Rabus *et al.*, 2011, Wilkes *et al.*, 2003).

Thus far, alkane-degrading isolates reported in the literature have never been reported to be capable of using individual *iso*- or *cyclo*-alkanes. This does not preclude their ability to do so because it is likely that these organisms have not been tested for their ability to use *iso*- and *cyclo*-alkanes, or that microorganisms capable of doing so have not yet been isolated. The only evidence of *cyclo*-alkane degradation came from Rios-Hernandez *et al.*, (2003) who showed that ethylcyclopentane could be degraded by means of fumarate addition by a sulfate-reducing microbial consortium. Supporting the notion of relaxed specificity of ASS, *Azoarcus* HxN1, which uses *n*-alkanes as a carbon source, is also capable of co-activating *cyclo*-alkanes in crude oil under nitrate-reducing conditions, even though the succinylated *cyclo*-alkane product has been suggested to be a dead-end metabolite (Wilkes *et al.*, 2003).

Several *n*-alkane degraders capable of alkane addition to fumarate can co-activate toluene in the presence of alkanes to produce benzylsuccinate. The benzylsuccinate is then completely degraded to benzoate (a dead-end metabolite for these cultures) by a process analogous to carbon skeleton rearrangement and decarboxylation essential for complete alkane degradation (Rabus *et al.*, 2011). But how is this important in the context of hydrocarbon degradation in a microbial community? Although the organisms with ASS do not necessarily benefit directly from co-activating toluene, benzoate produced in such manner can be excreted and readily used by syntrophs such as *Syntrophus aciditrophicus* SB1 which is capable of benzoate degradation (McInerney *et al.*, 2007). In turn, CO<sub>2</sub> produced from this syntroph can be used by an alkane-degrader like *D. alkenivorans* AK-01 which is capable of autotrophic carbon fixation (Callaghan *et al.*, 2012). Although such interactions have not been shown, it is not entirely impossible since hydrocarbon contaminants typically comprise both alkane and monoaromatic compounds.

BSS is more stereospecific and produces almost entirely (R)-benzylsuccinate, whereas ASS produce diastereomers in addition to being able to activate alkanes and toluene (Rabus *et al.*, 2011). On the other hand, NMS has been suggested to activate toluene and xylene in addition to 2-methylnaphthalene (Acosta-González *et al.*, 2013, von Netzer *et al.*, 2013). Organisms capable of toluene addition to fumarate contain suites of genes necessary for the downstream ring opening reaction, but these are absent in alkane degraders (Rabus *et al.*, 2011). Perhaps the ancestor of *bss* has been acquired by alkane degraders through horizontal gene transfer in an unknown evolutionary event, and then evolved to become *ass*? But the opposite can also be true; thus the evolutionary pattern of ASS, BSS and NMS remains a subject of interest.

### **1.2.5 Is addition to fumarate a ubiquitous mechanism for hydrocarbon degradation?**

Genes encoding fumarate addition (*assA/bssA/nmsA*) have been detected in many anoxic environments based on PCR amplification using degenerate primer pairs (Acosta-González *et al.*, 2013, Aitken *et al.*, 2013, Callaghan *et al.*, 2010, von Netzer *et al.*, 2013). In so doing, a repertoire of diverse *assA/bssA/nmsA* genes with unknown taxonomy have been discovered. *bssA* genes in organisms from the same taxonomic class tend to form phylogenetically congruent clades in a phylogenetic tree and assignment to their respective clades has been used to infer substrate specificity of organisms (Acosta-González *et al.*, 2013, von Netzer *et al.*, 2013). Organisms carrying the *bssA* gene almost exclusively belong to the Firmicutes, Delta- and Betaproteobacteria (von Netzer *et al.*, 2013). On the other hand, less is known about the *assA* and *nmsA* genes due to the lack of reference sequences with known taxonomic identity. However, several instances suggest that organisms capable of alkane degradation by addition to fumarate include Beta- and Deltaproteobacteria and Firmicutes (Tan *et al.*, 2013), whereas *nmsA* have only been found in a Deltaproteobacterium (Selesi *et al.*, 2010). How is this important from an ecological perspective? In most environments where hydrocarbons are presumably the primary organic carbon source, *assA/bssA/nmsA* genes have been detected. These environments include

deep marine sediments and oil seeps (Kimes *et al.*, 2013, von Netzer *et al.*, 2013), contaminated aquifers (Callaghan *et al.*, 2010, Winderl *et al.*, 2007), river sediments (Aitken *et al.*, 2013), hydrocarbon-contaminated sites (Acosta-González *et al.*, 2013), produced water from petroleum reservoirs (Zhou *et al.*, 2012) and others. Therefore, it is reasonable to conclude that genes encoding fumarate addition are ubiquitous (and perhaps represent a universal hydrocarbon-degrading mechanism) in anoxic environments and possibly are involved in the carbon cycle within the local community. Isolation of pure cultures and/or information linking gene phylogeny to host taxonomy could greatly help advance the field of petroleum microbiology by assigning functions to organisms.

### **1.3 Thesis overview and research objectives**

#### **1.3.1 Research objectives**

**Objective 1:** Investigation of *n*-alkane activation by fumarate addition under methanogenic conditions by microbes indigenous to oil sands tailings ponds using a model enrichment culture (SCADC; Short Chain Alkane Degrading Culture) (Chapter 2; see Figure 1.3 for culture history).

**Objective 2:** Examination of genetic capabilities in anaerobic hydrocarbon degradation and functional roles of SCADC microbial community using metagenomic approaches (Chapter 3)

**Objective 3:** Examination of gene expression by SCADC communities during *n*-alkane degradation under methanogenic conditions using metatranscriptomic approaches (Chapter 4).

**Objective 4:** Functional comparison of SCADC to two other methanogenic hydrocarbon-degrading cultures (Chapter 5).

**Objective 5:** Examination of the diversity of fumarate addition genes in oil sands tailings ponds and other environments (Chapter 6).

### 1.3.2 Thesis outline

#### **Chapter 1: Introduction**

#### **Chapter 2: Evidence for activation of aliphatic hydrocarbons by addition to fumarate under methanogenic conditions**

This Chapter describes the establishment of enrichment culture SCADC, a methanogenic short chain alkane-degrading culture originated from the mature fine tailings of Syncrude's Mildred Lake Settling Basin (see Figure 1.3 for culture descriptions). SCADC was used for intensive metabolite analysis during degradation of amended alkanes in search of products of fumarate addition. Metabolite detection identified fumarate addition products of 2-methylpentane and methylcyclopentane but not other amended *n*-alkanes, indicating that enzymes for fumarate addition are functional under methanogenic conditions. Here, community structure as well as potential microbial key players likely to be important during methanogenic alkane degradation were investigated over a time course using 16S rRNA gene pyrotag sequencing.

#### **Chapter 3: Metagenomic analysis of an anaerobic alkane-degrading microbial culture: Potential hydrocarbon-activating pathways and inferred roles of community members**

This chapter describes the community structure of SCADC enrichment culture based on total DNA sequencing using Illumina Hi-seq, as a complement to the time course community structure investigated in Chapter 2. The Illumina Hi-seq sequences were assembled and the community structure was investigated using sequence homology-based binning approach. The capability of SCADC community for hydrocarbon degradation was investigated by screening the sequence assembly for genes previously inferred to be involved in hydrocarbon degradation. Notably, a contig harbouring a novel *ass* gene cluster was putatively assigned to a member of the Firmicutes. The roles of community members not

involved directly in alkane degradation are described using selected COGs important for anaerobic respiration processes.

#### **Chapter 4: Metagenomic and metatranscriptomic analyses of a model alkane-degrading enrichment culture implicate *Desulfotomaculum* and *Smithella* in anaerobic alkane degradation**

This Chapter focuses on obtaining the partial genomes of microbial key players potentially involved in alkane degradation. This was achieved by taxonomic binning of the metagenome reported in Chapter 3 using a sequence composition-based method. RNA was isolated from SCADC subculture, followed by cDNA construction and Illumina Hi-seq sequencing. RNA-seq confirmed the observation in Chapter 2 that a gene encoding an enzyme subunit for alkane addition to fumarate was highly transcribed during active alkane degradation. Further investigation of active community based on pyrotag sequencing of the 16S rRNA using total RNA transcribed to cDNA, as well as RNA-seq of a novel bacterium genome affiliated with Peptococcaceae identified in this Chapter, confirmed our observation that a microbe belonging to Peptococcaceae is involved in fumarate addition activation of low molecular weight alkanes.

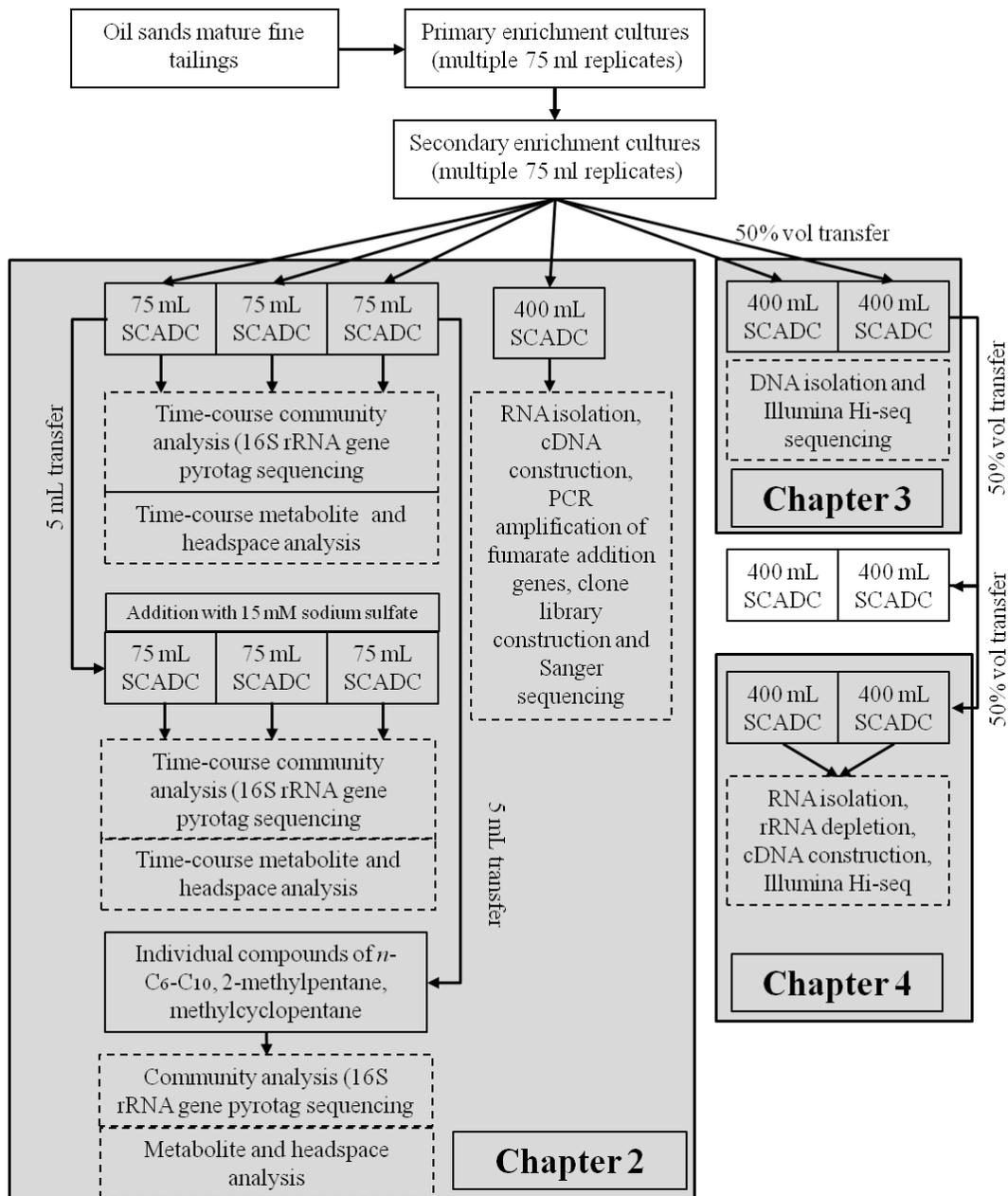
#### **Chapter 5: Comparative metagenomic analysis of three methanogenic hydrocarbon-degrading enrichment cultures and relevant environmental metagenomes**

This chapter examines the overall functional capability of SCADC compared to two other metagenomes obtained from a methanogenic naphtha-degrading culture enriched from mature fine tailings of oil sands tailings pond, and a methanogenic toluene-degrading culture enriched from gasoline-contaminated aquifer obtained from Fort Lupton. Comparison using the relative abundance of SEED categories suggests that three cultures had overall streamlined capabilities in anaerobic respiration, anaerobic hydrocarbon degradation and methanogenesis, and therefore these features are likely important in conversion of hydrocarbon to methane under methanogenic conditions.

## **Chapter 6: Assessing the functional diversity of fumarate addition genes in an oil sands tailings ponds and other environments**

This chapter describes the functional diversity of fumarate addition genes in oil sands tailings ponds by data mining the metagenome of an oil sands tailings pond. The functional potential of selected genes was investigated by preparing a new set of microcosms amended with different hydrocarbons (i.e., BTEX, C6-C10, C14-C16, naphtha) as the only organic substrates. Selected sequences representing different operational taxonomic units of fumarate addition genes were quantified using qPCR at two time points, showing the increase in abundance of fumarate addition genes in response to the addition of substrates, indicating that the microorganisms carrying these genes are important during degradation of amended substrates.

## **Chapter 7: Conclusion and synthesis**



**Figure 1.3** History of enrichment culture SCADC used in the experiments described in Chapters 2, 3, 4 and 5. Details of the microcosm preparations are explained in Chapter 2, Chapter 3 and Appendix C1.

#### 1.4 Additional research outside the scope of the thesis

In addition to my thesis project, I also spent a great amount of time trying to isolate organisms capable of alkane degradation under methanogenic conditions,

although none of that work has been fruitful. I also have been involved in the bioinformatic analyses (sequence assembly, taxonomic binning, sequence annotation, scripting, etc.) of other metagenomes generated for the Hydrocarbon Metagenomics Project (<http://www.hydrocarbonmetagenomics.com>), and taxonomic binning of the sequence assembly of a less complex community generated by other colleagues. This mainly involved identification of suitable bioinformatics tools (i.e., sequence assembler, etc.) that can be used in-house without heavy investment in computing tools; merging, adapting and modifying bioinformatics pipelines available in the literature; data mining of genes that may be essential in the ecology of oil sands tailings ponds; and comparative genomics using data generated by members of the lab.

## 1.5 References

- Abu Laban N, Selesi D, Rattei T, Tischler P, Meckenstock RU. (2010). Identification of enzymes involved in anaerobic benzene degradation by a strictly anaerobic iron-reducing enrichment culture. *Environ Microbiol* **12**: 2783-2796.
- Acosta-González A, Rossello-Mora R, Marques S. (2013). Diversity of benzylsuccinate synthase-like (*bssA*) genes in hydrocarbon-polluted marine sediments suggests substrate-dependent clustering. *Appl Environ Microbiol* **79**: 3667-3676.
- Agrawal A, Gieg LM. (2013). In situ detection of anaerobic alkane metabolites in subsurface environments. *Front Microbiol* **4**: 140.
- Aitken CM, Jones DM, Maguire MJ, Gray ND, Sherry A, Bowler BFJ *et al.*, (2013). Evidence that crude oil alkane activation proceeds by different mechanisms under sulfate-reducing and methanogenic conditions. *Geochim Cosmochim Acta* **109**: 162-174.
- Albertsen M, Hugenholtz P, Skarshewski A, Nielsen KL, Tyson GW, Nielsen PH. (2013). Genome sequences of rare, uncultured bacteria obtained by differential coverage binning of multiple metagenomes. *Nat Biotechnol* **31**: 533-538.
- An D, Brown D, Chatterjee I, Dong X, Ramos-Padron E, Wilson S, Bordenave S, Caffrey S, Gieg L, Sensen C, and Gerrit Voordouw, G. (2013) Microbial community and potential functional gene diversity involved in anaerobic hydrocarbon degradation and methanogenesis in an oil sands tailings pond. *Genome* 10.1139/gen-2013-0083.

- Beller HR, Spormann AM. (1999). Substrate range of benzylsuccinate synthase from *Azoarcus* sp strain T. *FEMS Microbiol Lett* **178**: 147-153.
- Biegert T, Fuchs G, Heider F. (1996). Evidence that anaerobic oxidation of toluene in the denitrifying bacterium *Thauera aromatica* is initiated by formation of benzylsuccinate from toluene and fumarate. *Eur J Biochem* **238**: 661-668.
- Bordenave S, Kostenko V, Dutkoski M, Grigoryan A, Martinuzzi RJ, Voordouw G. (2010). Relation between the activity of anaerobic microbial populations in oil sands tailings ponds and the sedimentation of tailings. *Chemosphere* **81**: 663-668.
- Callaghan AV, Gieg LM, Kropp KG, Suflita JM, Young LY. (2006). Comparison of mechanisms of alkane metabolism under sulfate-reducing conditions among two bacterial isolates and a bacterial consortium. *Appl Environ Microbiol* **72**: 4274-4282.
- Callaghan AV, Wawrik B, Chadhain SMN, Young LY, Zylstra GJ. (2008). Anaerobic alkane-degrading strain AK-01 contains two alkylsuccinate synthase genes. *Biochem Biophys Res Commun* **366**: 142-148.
- Callaghan AV, Tierney M, Phelps CD, Young LY. (2009). Anaerobic biodegradation of *n*-hexadecane by a nitrate-reducing consortium. *Appl Environ Microbiol* **75**: 1339-1344.
- Callaghan AV, Davidova IA, Savage-Ashlock K, Parisi VA, Gieg LM, Suflita JM *et al.*,. (2010). Diversity of benzyl- and alkylsuccinate synthase genes in hydrocarbon-impacted environments and enrichment cultures. *Environ Sci Technol* **44**: 7287-7294.
- Callaghan AV, Morris BEL, Pereira IAC, McInerney MJ, Austin RN, Groves JT *et al.*,. (2012). The genome sequence of *Desulfatibacillum alkenivorans* AK-01: a blueprint for anaerobic alkane oxidation. *Environ Microbiol* **14**: 101-113.
- Callaghan AV. (2013). Enzymes involved in the anaerobic oxidation of *n*-alkanes: from methane to long-chain paraffins. *Front Microbiol* **4**: 89.
- Carmona M, Zamarro MT, Blazquez B, Durante-Rodriguez G, Juarez JF, Valderrama JA *et al.*,. (2009). Anaerobic catabolism of aromatic compounds: a genetic and genomic view. *Microbiol Mol Biol Rev* **73**: 71.
- Chalaturnyk RJ, Scott JD, Ozum B. (2002). Management of oil sands tailings. *J Petrol Sci Eng* **20**: 1025-1046.
- Chen M, Walshe G, Fru EC, Ciborowski JJH, Weisener CG. (2013). Microcosm assessment of the biogeochemical development of sulfur and oxygen in oil sands fluid fine tailings. *Appl Geochem* **37**: 1-11.

Cheng L, Ding C, Li Q, He Q, Dai LR, Zhang H. (2013). DNA-SIP reveals that Syntrophaceae play an important role in methanogenic hexadecane degradation. *Plos One* **8**.

Davidova IA, Duncan KE, Choi OK, Suflita JM. (2006). *Desulfoglaeba alkanexedens* gen. nov., sp nov., an n-alkane-degrading, sulfate-reducing bacterium. *Int J Syst Evol Microbiol* **56**: 2737-2742.

Diaz E, Jimenez JI, Nogales J. (2013). Aerobic degradation of aromatic compounds. *Curr Opin Biotechnol* **24**: 431-442.

Dolfing J, Larter SR, Head IM. (2008). Thermodynamic constraints on methanogenic crude oil biodegradation. *Isme Journal* **2**: 442-452.

Fedorak PM, Coy DL, Salloum MJ, Dudas MJ. (2002). Methanogenic potential of tailings samples from oil sands extraction plants. *Canadian Journal of Microbiology* **48**: 21-33.

Fedorak PM, Coy DL, Dudas MJ, Simpson MJ, Renneberg AJ, MacKinnon MD. (2003). Microbially-mediated fugitive gas production from oil sands tailings and increased tailings densification rates. *J Environ Eng Sci* **2**: 199-211.

Foght J. (2008). Anaerobic biodegradation of aromatic hydrocarbons: Pathways and prospects. *J Mol Microbiol Biotechnol* **15**: 93-120.

Fru EC, Chen M, Walshe G, Penner T, Weisener C. (2013). Bioreactor studies predict whole microbial population dynamics in oil sands tailings ponds. *Appl Microbiol Biotechnol* **97**: 3215-3224.

Gieg LM, Duncan KE, Suflita JM. (2008). Bioenergy production via microbial conversion of residual oil to natural gas. *Appl Environ Microbiol* **74**: 3022-3029.

Golby S, Ceri H, Gieg LM, Chatterjee I, Marques LLR, Turner RJ. (2012). Evaluation of microbial biofilm communities from an Alberta oil sands tailings pond. *FEMS Microbiol Ecol* **79**: 240-250.

Gray ND, Sherry A, Grant RJ, Rowan AK, Hubert CRJ, Callbeck CM *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 *et al.*, (2008). Genes encoding the candidate enzyme for anaerobic activation of n-alkanes in the denitrifying bacterium, strain HxN1. *Environ Microbiol* **10**: 376-385.

Haroon MF, Hu SH, Shi Y, Imelfort M, Keller J, Hugenholtz P *et al.*, (2013). Anaerobic oxidation of methane coupled to nitrate reduction in a novel archaeal lineage. *Nature* **500**: 567.

Head IM, Jones DM, Larter SR. (2003). Biological activity in the deep subsurface and the origin of heavy oil. *Nature* **426**: 344-352.

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

Holowenko FM, MacKinnon MD, Fedorak PM. (2000). Methanogens and sulfate-reducing bacteria in oil sands fine tailings waste. *Canadian Journal of Microbiology* **46**: 927-937.

Holowenko FM, Mackinnon MD, Fedorak PM. (2001). Naphthenic acids and surrogate naphthenic acids in methanogenic microcosms. *Water Res* **35**: 2595-2606.

Jackson BE, McInerney MJ. (2002). Anaerobic microbial metabolism can proceed close to thermodynamic limits. *Nature* **415**: 454-456.

Jarling R, Sadeghi M, Drozdowska M, Lahme S, Buckel W, Rabus R *et al.*, (2012). Stereochemical investigations reveal the mechanism of the bacterial activation of *n*-alkanes without oxygen. *Angewandte Chemie-International Edition* **51**: 1334-1338.

Jones DM, Head IM, Gray ND, Adams JJ, Rowan AK, Aitken CM *et al.*, (2008). Crude-oil biodegradation via methanogenesis in subsurface petroleum reservoirs. *Nature* **451**: 176-U176.

Kimes NE, Callaghan AV, Aktas DF, Smith WL, Sunner J, Golding B *et al.*, (2013). Metagenomic analysis and metabolite profiling of deep-sea sediments from the Gulf of Mexico following the Deepwater Horizon oil spill. *Front Microbiol* **4**: 50.

Kniemeyer O, Musat F, Sievert SM, Knittel K, Wilkes H, Blumenberg M *et al.*, (2007). Anaerobic oxidation of short-chain hydrocarbons by marine sulfate-reducing bacteria. *Nature* **449**: 898-U810.

Kosaka T, Kato S, Shimoyama T, Ishii S, Abe T, Watanabe K. (2008). The genome of *Pelotomaculum thermopropionicum* reveals niche-associated evolution in anaerobic microbiota. *Genome Res* **18**: 442-448.

Kropp KG, Davidova IA, Suflita JM. (2000). Anaerobic oxidation of *n*-dodecane by an addition reaction in a sulfate-reducing bacterial enrichment culture. *Appl Environ Microbiol* **66**: 5393-5398.

- Leuthner B, Leutwein C, Schulz H, Horth P, Haehnel W, Schiltz E *et al.*, (1998). Biochemical and genetic characterization of benzylsuccinate synthase from *Thauera aromatica*: a new glycyl radical enzyme catalysing the first step in anaerobic toluene metabolism. *Mol Microbiol* **28**: 615-628.
- Li W, Wang LY, Duan RY, Liu JF, Gu JD, Mu BZ. (2012). Microbial community characteristics of petroleum reservoir production water amended with *n*-alkanes and incubated under nitrate-, sulfate-reducing and methanogenic conditions. *Int Biodeterior Biodegrad* **69**: 87-96.
- Mbadinga SM, Wang LY, Zhou L, Liu JF, Gu JD, Mu BZ. (2011). Microbial communities involved in anaerobic degradation of alkanes. *Int Biodeterior Biodegrad* **65**: 1-13.
- McInerney MJ, Rohlin L, Mouttaki H, Kim U, Krupp RS, Rios-Hernandez L *et al.*, (2007). The genome of *Syntrophus aciditrophicus*: Life at the thermodynamic limit of microbial growth. *PNAS* **104**: 7600-7605.
- Penner TJ, Foght JM. (2010). Mature fine tailings from oil sands processing harbour diverse methanogenic communities. *Can J Microbiol* **56**: 459-470.
- Rabus R, Heider J. (1998). Initial reactions of anaerobic metabolism of alkylbenzenes in denitrifying and sulfate reducing bacteria. *Arch Microbiol* **170**: 377-384.
- Rabus R, Jarling R, Lahme S, Kuhner S, Heider J, Widdel F *et al.*, (2011). Co-metabolic conversion of toluene in anaerobic *n*-alkane-degrading bacteria. *Environ Microbiol* **13**: 2576-2585.
- Ramos-Padron E, Bordenave S, Lin SP, Bhaskar IM, Dong XL, Sensen CW *et al.*, (2011). Carbon and sulfur cycling by microbial communities in a gypsum-treated oil sands tailings pond. *Environ Sci Technol* **45**: 439-446.
- Rios-Hernandez LA, Gieg LM, Suflita JM. (2003). Biodegradation of an alicyclic hydrocarbon by a sulfate-reducing enrichment from a gas condensate-contaminated aquifer. *Appl Environ Microbiol* **69**: 434-443.
- Saidi-Mehrabad A, He ZG, Tamas I, Sharp CE, Brady AL, Rochman FF *et al.*, (2013). Methanotrophic bacteria in oilsands tailings ponds of northern Alberta. *Isme Journal* **7**: 908-921.
- Salloum MJ, Dudas MJ, Fedorak PM. (2002). Microbial reduction of amended sulfate in anaerobic mature fine tailings from oil sand. *Waste Manage Res* **20**: 162-171.

Selesi D, Jehmlich N, von Bergen M, Schmidt F, Rattei T, Tischler P *et al.*, (2010). Combined genomic and proteomic approaches identify gene clusters involved in anaerobic 2-methylnaphthalene degradation in the sulfate-reducing enrichment culture N47. *J Bacteriol* **192**: 295-306.

Siddique T, Fedorak PM, Foght JM. (2006). Biodegradation of short-chain *n*-alkanes in oil sands tailings under methanogenic conditions. *Environ Sci Technol* **40**: 5459-5464.

Siddique T, Fedorak PM, McKinnon MD, Foght JM. (2007). Metabolism of BTEX and naphtha compounds to methane in oil sands tailings. *Environ Sci Technol* **41**: 2350-2356.

Siddique T, Gupta R, Fedorak PM, MacKinnon MD, Foght JM. (2008). A first approximation kinetic model to predict methane generation from an oil sands tailings settling basin. *Chemosphere* **72**: 1573-1580.

Siddique T, Penner T, Semple K, Foght JM. (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, Nesbo C, Foght JM. (2012). Microbial communities involved in methane production from hydrocarbons in oil sands tailings. *Environ Sci Technol* **46**: 9802-9810.

Sieber JR, McInerney MJ, Gunsalus RP. (2012). Genomic insights into syntrophy: The paradigm for anaerobic metabolic cooperation. *Annu Rev Microbiol* **66**: 429-452.

So CM, Young LY. (1999). Initial reactions in anaerobic alkane degradation by a sulfate reducer, strain AK-01. *Appl Environ Microbiol* **65**: 5532-5540.

So CM, Phelps CD, Young LY. (2003). Anaerobic transformation of alkanes to fatty acids by a sulfate-reducing bacterium, strain Hxd3. *Appl Environ Microbiol* **69**: 3892-3900.

von Netzer F, Pilloni G, Kleindienst S, Kruger M, Knittel K, Grundger F *et al.*, (2013). Enhanced gene detection assays for fumarate-adding enzymes allow uncovering of anaerobic hydrocarbon degraders in terrestrial and marine systems. *Appl Environ Microbiol* **79**: 543-552.

Wang W, Shao Z. (2013). Enzymes and genes involved in aerobic alkane degradation. *Front Microbiol* **4**: 116.

Wilkes H, Rabus R, Fischer T, Armstroff A, Behrends A, Widdel F. (2002). Anaerobic degradation of *n*-hexane in a denitrifying bacterium: Further

degradation of the initial intermediate (1-methylpentyl)succinate via C-skeleton rearrangement. *Arch Microbiol* **177**: 235-243.

Wilkes H, Kuhner S, Bolm C, Fischer T, Classen A, Widdel F *et al.*, (2003). Formation of *n*-alkane- and cycloalkane-derived organic acids during anaerobic growth of a denitrifying bacterium with crude oil. *Org Geochem* **34**: 1313-1323.

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

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

Zhang T, Tremblay PL, Chaurasia AK, Smith JA, Bain TS, Lovley DR. (2013). Anaerobic benzene oxidation via phenol in *Geobacter metallireducens*. *Appl Environ Microbiol*.

Zhou L, Li K-P, Mbadinga SM, Yang S-Z, Gu J-D, Mu B-Z. (2012). Analyses of *n*-alkanes degrading community dynamics of a high-temperature methanogenic consortium enriched from production water of a petroleum reservoir by a combination of molecular techniques. *Ecotoxicol (London, England)* **21**: 1680-1691.

## 2 Evidence for activation of aliphatic hydrocarbons by addition to fumarate under methanogenic conditions<sup>1</sup>

### 2.1 Abstract

A methanogenic short-chain alkane-degrading enrichment culture (SCADC) was established from oil sands tailings and transferred several times in the laboratory with *n*-C<sub>6</sub> to *n*-C<sub>10</sub> alkanes as sole carbon source. The C<sub>6</sub> alkanes used in this study contained *n*-hexanes as well as 2-methylpentane, 3-methylpentane and methylcyclopentane as impurities. Cultures produced ~50% of the maximum theoretical methane during 18 months incubation at ~25°C. Headspace analyses at 18 months showed almost complete depletion of *n*-C<sub>6</sub> to *n*-C<sub>10</sub>, as well as 2-methylpentane and methylcyclopentane. The depletion of alkanes correlated with detection of signature metabolites characteristic of the fumarate addition activation of 2-methylpentane and methylcyclopentane, but not *n*-alkanes. During the active phase of methanogenesis, reverse-transcription PCR confirmed the expression of genes related to fumarate addition mechanisms (i.e., *assA* and *bssA*). Pyrotag sequencing of 16S rRNA genes obtained from DNA extracted during active degradation revealed the enrichment of bacteria related to Clostridia (i.e., Peptococcaceae) and methanogens (i.e., Methanosaetaceae and Methanomicrobiaceae). Transfer of methanogenic cultures into medium supplemented with sodium sulfate yielded sulfide and putative succinylated fumarate addition metabolites from *n*-C<sub>6</sub> to *n*-C<sub>10</sub>. Pyrotag sequencing of the 16S rRNA genes showed enrichment of Deltaproteobacteria relative to Clostridia. Concomitantly, headspace analyses at this point showed depletion of *n*-alkanes, whereas depletion of methylcyclopentane was minimal under sulfate-reducing conditions. This study demonstrates that *n*-alkanes and alicyclic hydrocarbons are degraded by an oil sands tailings enrichment community (possibly by Peptococcaceae) by fumarate addition under methanogenic conditions.

<sup>1</sup>A version of this chapter has been modified for publication to demonstrate that fumarate addition of *n*-alkanes under methanogenic conditions is possible. This was achieved by using "hexanes" containing 2-methylpentane and methylcyclopentane with the hypothesis that branched and *cyclo*-alkanes might accumulate during degradation and therefore be detected using GC-MS.

## 2.2 Introduction

Crude oil hydrocarbons such as alkanes and benzene, toluene, ethylbenzene, and xylenes (BTEX), polycyclic aromatic hydrocarbons and other refractory compounds are common contaminants in shallow and anoxic environments, e.g., contaminated aquifers (Parisi *et al.*, 2009), and can be released into the environment during events such as oil spills (Kimes *et al.*, 2013). In anoxic environments such as oil sands tailings ponds, anaerobic microorganisms are able to convert the unrecovered residual solvent (naphtha) used in bitumen extraction into methane (Siddique *et al.*, 2006, Siddique *et al.*, 2007, Siddique *et al.*, 2011, Siddique *et al.*, 2012), which is favourable because the process accelerates the densification of fine tailings for water recovery (Fedorak *et al.*, 2003). On the other hand, methane can also be harmful because it is a potent greenhouse gas. In petroleum reservoirs that did not experience paleopasteurization, methanogenic degradation of crude oil alkanes ( $n\text{-C}>12$ ) over geological time has been proposed to be responsible for the formation of heavy crude oil (Jones *et al.*, 2008). Degradation of crude oil alkanes occurs with stoichiometric conversion into methane and  $\text{CO}_2$  (Jones *et al.*, 2008, Zengler *et al.*, 1999), and the degradable fractions of hydrocarbon components in both crude oil and naphtha often contain high proportions of aliphatic hydrocarbons (Gray *et al.*, 2011, Siddique *et al.*, 2006). Therefore, understanding methanogenic alkane degradation is relevant to applied biotechnology such as microbially enhanced recovery of oil and/or production of natural gas from exhausted crude oil reservoirs (Gieg *et al.*, 2008, Jones *et al.*, 2008).

Enzymatic activation of hydrocarbons by addition to fumarate is the best described and most widely documented mechanism for initiating anaerobic hydrocarbon degradation under nitrate-, iron- and sulfate-reducing conditions (Agrawal and Gieg 2013, Callaghan 2013). The glycyl radical enzyme alkylsuccinate synthase (ASS: also called methylalkylsuccinate synthase, MAS) and its counterparts benzylsuccinate synthase (BSS) and naphthyl-2-methylsuccinate synthase (NMS) responsible for aliphatic and aromatic hydrocarbon activation, respectively, have been described and reviewed

(Mbadanga *et al.* 2011; Callaghan 2013; Agrawal and Gieg, 2013). The requisite genes (particularly those encoding the alpha-subunits, e.g. *assA*, *bssA*) have been detected in contaminated environments (Winderl *et al.*(Acosta-González *et al.*, 2013), 2007; Callaghan *et al.*, 2010), enrichment cultures (Callaghan *et al.*, 2010; Aitken *et al.*, 2013) and model alkane-degrading organisms such as nitrate-reducing *Azoarcus* sp. HXN1 (Grundmann *et al.*, 2008) and sulfate-reducing *Desulfatibacillum alkenivorans* AK-01 (Callaghan *et al.*, 2008). Detection of succinylated ‘signature metabolites’ from ASS activity under sulfate-reducing conditions is commonly used as evidence for *n*-alkane degradation (Duncan *et al.*, 2009; Gieg *et al.*, 2010). However, these metabolites are not commonly detected in methanogenic environments where alkane depletion is occurring (Gieg *et al.*, 2010, Parisi *et al.*, 2009), and rigorous examination of methanogenic alkane-degrading cultures in the laboratory similarly has failed to detect these characteristic metabolites (Callaghan *et al.*, 2010; Aitken *et al.* 2013). This has led to speculation that alternative mechanisms are used by methanogenic communities to activate alkanes and/or that the intermediates are neither excreted nor accumulate, but rather are metabolized intracellularly (Aitken *et al.*, 2013).

We sought evidence of alkane addition to fumarate indirectly by examining methanogenic *n*-alkane-degrading cultures that also contained *iso*- and *cyclo*-alkane analogues, hypothesizing that succinylated metabolites of the latter substrates might accumulate or be more persistent than their transient *n*-alkane homologues, and thus be detectable in culture supernatant. The assumption is based on previous observations that the ASS enzyme can be promiscuous in that it can co-activate structurally related and unrelated hydrocarbons in the presence of its usual *n*-alkanes substrates (Rabus *et al.*, 2011, Wilkes *et al.*, 2003). We generated total RNA from the SCADC community to determine (from cDNA) whether canonical *assA* and *bssA* genes were expressed during biodegradation of alkanes. We also incubated parallel cultures under methanogenic and sulfidogenic conditions to compare signature metabolite production and microbial community structure, based on 16S rRNA gene pyrosequencing. The cultures used in this study originated from oil sands mature fine tailings known to harbour microbes

capable of biodegrading short chain ( $n$ -C<sub>6</sub> to  $n$ -C<sub>10</sub>) and longer chain  $n$ -alkanes ( $n$ -C<sub>14</sub> to  $n$ -C<sub>18</sub>) as well as monoaromatics under strict methanogenic conditions (Siddique *et al.*, 2006, Siddique *et al.*, 2007, Siddique *et al.*, 2011). The results add to our understanding of methanogenic biodegradation of aliphatic hydrocarbons, suggesting that alkane biodegradation under methanogenic conditions can proceed by addition to fumarate.

## **2.3 Experimental Procedures**

### **2.3.1 Incubation of enrichment cultures**

Mature fine tailings from Mildred Lake Settling Basin were used to establish enrichment cultures capable of methanogenic degradation of a mixture of alkanes (Siddique *et al.*, 2007; Siddique *et al.*, 2011), an activity that was sustained upon transfer to fresh medium and substrate. Aliquots of these cultures were pooled for use as inoculum in the current study by transferring ~37-mL active methanogenic enrichment culture into serum bottles containing 37 mL of methanogenic medium (Widdel and Bak 1992) in triplicate 158-mL serum bottles. A filter-sterilized C6-C10 alkane stock comprising equal volumes of “hexanes”,  $n$ -C<sub>7</sub> (CAS 142-82-5, >97% purity, Fisher Scientific),  $n$ -C<sub>8</sub> (CAS 111-65-9, >98% purity, Sigma-Aldrich) and  $n$ -C<sub>10</sub> (CAS 124-18-5, >99% purity, Sigma-Aldrich) was added to a final concentration of 0.1 vol%. The “hexanes” (CAS 110-54-3, Fisher Scientific) comprised  $n$ -C<sub>6</sub> (62%) plus the C6 isomers 2-methylpentane (3%), 3-methylpentane (16%) and methylcyclopentane (19%), as determined by relative peak area (Tan *et al.*, 2013). Enrichment cultures of 75 mL amended individually with 75  $\mu$ L (or 0.1%vol) ‘hexanes’,  $n$ -C<sub>7</sub>,  $n$ -C<sub>8</sub>,  $n$ -C<sub>10</sub>, 2-methylpentane (EC number 2035234,  $\geq$ 95% purity, Fluka, USA) or methylcyclopentane (EC number 202-503-2, 97% purity, Sigma-Aldrich) were also prepared in methanogenic medium and inoculated with enrichment culture SCADC to measure methane and metabolite production. Duplicate inoculated control cultures that did not receive hydrocarbon substrate were prepared similarly. After incubation for approximately 18 months, 5 mL of culture fluids from the 75-mL methanogenic cultures containing “hexanes” and  $n$ -C<sub>7</sub> to  $n$ -C<sub>10</sub> described above were transferred into 45 mL of identical methanogenic medium

containing 15 mM sodium sulfate. A 400-mL methanogenic culture provided with 0.1 vol% alkane mixture was also established and incubated for 4 months for RNA isolation and metabolite detection (see below). All enrichment cultures had a headspace of 30% O<sub>2</sub>-free CO<sub>2</sub>-balance N<sub>2</sub> and were incubated in the dark at ~28 °C with gentle manual mixing by inversion once per week.

### **2.3.2 Analysis of methane, sulfide and volatile hydrocarbons in culture**

#### **bottles**

Methane was measured by sampling 50 µL of culture bottle headspace using a sterile needle and syringe, and immediately injected into a gas chromatograph with a flame-ionization detector as previously described (Siddique *et al.*, 2006). The maximum theoretical methane yields in methanogenic enrichment cultures were calculated using the Symons and Buswell equation (Roberts, 2002), accounting for the specific gravity of each hydrocarbon. Calculation of maximum theoretical methane yield is presented in Appendix Table B1. The soluble sulfide in sulfate-amended sulfidogenic enrichment cultures was determined using the methylene blue method described by Cline (1969).

For residual volatile hydrocarbon detection by gas-chromatography mass spectrometry (GC-MS), 50 µL of culture headspace were sampled using a sterile needle and syringe and injected into an Agilent 6890N gas chromatograph with a 5973 inert mass selective detector fitted with an Agilent HP-5MS capillary column (30 m × 0.25 µm film thickness; J + W Scientific). Helium was used as carrier gas at 0.8 mL min<sup>-1</sup> in splitless mode with the injector at 250 °C. The temperature program started at 35 °C for 7 min and ramped at 5 °C min<sup>-1</sup> to 100 °C. Headspace hydrocarbon depletion was calculated using method similar to that described by Prince and Suflita (2007), where a conserved internal marker was used for comparison. Briefly, the percentage of target compounds remaining in the headspace was calculated based on the following equation: % of headspace hydrocarbon =  $[(A_{\text{sample}}/C_{\text{sample}})/(A_{\text{uninoculated sterile medium}}/C_{\text{uninoculated sterile medium}})] \times 100$ , where A and C represent the headspace abundance of target analytes and an internal standard, respectively. We used 3-methylpentane as the conserved

internal standard in the short chain alkane-amended enrichment cultures because it was recalcitrant to degradation in our previously established enrichment cultures from oil sands tailings ponds incubated up to four years (unpublished data).

### 2.3.3 Detection of putative metabolites

Culture fluid (2 to 200 mL, depending on the culture) was periodically removed from culture bottles using a needle and syringe, and either immediately frozen at -80 °C or acidified to pH<2.5 using 10 M hydrochloric acid for extraction of metabolites (Gieg *et al.*, 2010). Before solvent extraction, 1 µg 4-fluoro-1-naphthoic acid (Sigma-Aldrich) was added as a surrogate standard (So *et al.*, 2003) into all samples and into extraction-derivatization control reactions containing only doubly-distilled water. The acidified samples were extracted three times with an equal volume of ethyl acetate (CAS 141-78-6, >99% purity, Fisher Scientific), dried overnight in a fume hood, concentrated to 100 µL and derivatized with an equal volume of *N*, *O*-bis-(trimethylsilyl) trifluoroacetamide (Thermo Scientific, Waltham, MA) at 70 °C for 30 min (Gieg *et al.*, 2010). The derivatives were analyzed by direct injection into the GC-MS described above. The oven was held at 65 °C for 5 min, then increased at 5 °C min<sup>-1</sup> to 270 °C, and held for 15 min. Identity of metabolites was determined by comparison of the mass spectral profiles and retention time of individual peaks to known standards and characteristic fragmentation patterns reported by Gieg and Suflita (2002) and Rios-Hernandez *et al.*, (2003). The standards were obtained from culture extracts from sulfate-reducing cultures of *Desulfoglaeba alkanexedens* ADLC grown individually on *n*-C<sub>6</sub>, *n*-C<sub>8</sub> or *n*-C<sub>10</sub>, and which were shown to contain (1-methylpentyl)succinic acid, (1-methylheptyl)succinic acid, and (1-methylnonyl)succinic acid corresponding to the signature fumarate addition metabolites of *n*-C<sub>6</sub>, *n*-C<sub>8</sub> and *n*-C<sub>10</sub> (donated by L. Gieg, University of Calgary). The detection limit of alkylsuccinate arising from fumarate addition of *n*-alkanes using GC-MS was not verified in this study, but other studies that report the detection limit for of alkylsuccinates was in the nanomolar range (Aitken *et al.*, 2013; Agrawal and Gieg 2013).

### **2.3.4 Nucleic acid extraction, analyses of microbial community and RT-PCR of functional genes**

DNA extraction was performed as previously described (Foght *et al.*, 2004) using sample fluids obtained from methanogenic enrichment cultures incubated for 0 days, 40 weeks and 81 weeks; sulfate-reducing enrichment cultures incubated for 0 days, 22, 37 and 45 weeks; and live control cultures that received neither hydrocarbons nor sulfate, incubated for 81 weeks. Amplification of archaeal and bacterial 16S rRNA genes for 454 pyrotag sequencing was performed using the primer set 454T-RA/454T-FB targeting the V6-V8 regions of the 16S rRNA gene universal for Bacteria and Archaea (Berdugo-Clavijo *et al.*, 2012, Fowler *et al.*, 2012). Further details about pyrotag sequencing, which was performed at McGill University Génome Québec Innovation Centre, Canada, have been published (Chapter 3).

For the isolation of total RNA, culture fluid from one single 400 mL methanogenic culture (Fig. 1.3) was treated with an equal volume of 'stop solution' containing 10% phenol and 90% ethanol. Subsequent total RNA isolation was performed using a modified phenol-chloroform method (Rosana *et al.*, 2012). Cells were harvested at 3000 x *g* for 20 min. The cell pellet was washed in TE buffer (50 mM Tris (ICN), 100 mM EDTA (BDH)) and resuspended in 400 µl RNA extraction buffer (0.5% Triton X-100 (Sigma-Aldrich), 0.4% N-lauroylsarcosine (Sigma-Aldrich) and 0.4% sodium dodecyl sulfate in TE buffer). Cells lysis was performed using 300 µl zirconium beads in the presence of 400 µl phenol (Fisher-Scientific) by alternate vortexing and cooling on ice for 1 min for 15 cycles. The clarified aqueous layer was harvested and RNA was sequentially extracted with an equal volume of phenol, 1:1 phenol-chloroform and 24:1 chloroform-isoamyl alcohol. Extracted RNA was precipitated with 4 M LiCl coupled with ethanol precipitation. RNA was resuspended in nuclease-free water (Ambion) and quantitated using a NanoDrop<sup>TM</sup> spectrophotometer (Thermo-Scientific, USA) before being subjected to two rounds of DNA removal using DNase I (Invitrogen, USA), following the manufacturers' protocol.

Two lines of evidence were used to demonstrate that the RNA was free of amplifiable DNA. First, the DNase-treated total RNA was analyzed using an Agilent 2100 Bioanalyzer (Agilent Technologies, Germany) with an RNA Nanochip (Agilent Technologies, Germany) to show that no genomic DNA remained. Second, a portion of the RNA eluate was subjected to PCR amplification of 16S rRNA genes; this reaction yielded no product, indicating that no DNA remained in the RNA preparation. Therefore, the DNA-free total RNA preparation was converted to complementary DNA (cDNA) by reverse transcription PCR using Superscript® III Reverse Transcriptase (Invitrogen, USA) with random hexamers (Invitrogen, USA) following the manufacturer's protocol. The resulting cDNA preparation was used for amplification of *assA* and *bssA* to generate a clone library of expressed functional genes as described below.

The *assA* and *bssA* genes present in the cDNA preparation were amplified using primers 1995F/2467R (Chapter 6) and 7772f/8546r (Winderl *et al.*, 2007), respectively. Primer 1995F had a sequence of 5'-CCNAARTGGGGHAAAYGACGA-3' and primer 246R was 5'-ANCCNGMNAYVCKNACRATVA-3'. Amplicons were cloned into pGEM-T (Promega, USA), transformed into Subcloning Efficiency™ DH5α™ Chemically Competent Cells (Invitrogen, USA) per the manufacturer's instructions, and the cloned inserts were sequenced as described by Siddique *et al.* (2012) at the Molecular Biology Services Unit at University of Alberta.

### **2.3.5 Bioinformatics and phylogenetic analysis**

Sequence reads obtained from pyrotag sequencing of 16S rRNA gene amplicons from metagenomic DNA were processed using Phoenix 2.0 (Soh *et al.*, 2013). Sequence reads were clustered into Operational Taxonomic Unit (OTUs) using average neighbour clustering with a 5% distance cutoff. Each OTU was assigned a taxonomic identity using the RDP training sets, and the proportions of taxonomic groups were visualized using Circos 2.0 (Krzywinski *et al.*, 2009). A distance cutoff of 5% was chosen for the clustering analysis of 16S rRNA pyrotag sequences based on reasoning discussed in section 5.4.3 in Chapter 5. Comparisons between groups of enrichment cultures incubated under

methanogenic (day 1, week 40 and 81) or sulfidogenic (week 22, 37 and 45) conditions were performed using Metastats analyses (White *et al.*, 2009). Selected representative sequences from OTUs of interest (>15% of differentially abundant taxa from each grouping, based on Metastats) were aligned with selected environmental sequences using Silva Aligner (Pruesse *et al.*, 2012) with the default settings. The alignment was inserted into the All-Species Living tree (Yarza *et al.*, 2010) in ARB software using the pos-var-ssuref:bacteria filter as recommended by the program (Ludwig *et al.*, 2004). Reference sequences from the All-Species Living tree and sequence reads from this study were subjected to phylogenetic analyses using the Neighbour-Joining method with Jukes-Cantor correction model (Guindon *et al.*, 2010) with 1,000 bootstrap replicates.

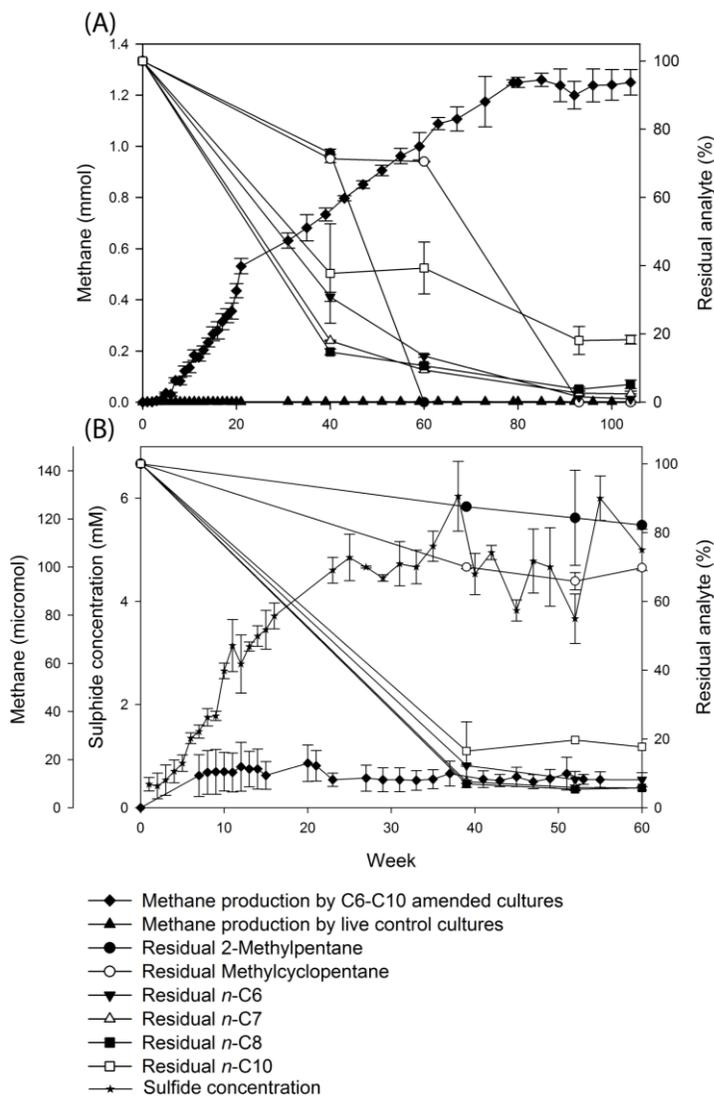
For phylogenetic analyses of cloned *assA* and *bssA* genes from cDNA, the deduced amino acid sequences of ~150 amino acids (~450 nt) were clustered at 2% distance cutoff using CD-Hit (Li and Godzik 2006). A representative sequence from each OTU was aligned with nearly complete or complete reference sequences from NCBI plus those reported by Tan *et al.* (2013) using MUSCLE v3.3 (Edgar 2004) followed by maximum likelihood analysis using PhyML v3.0 (Guindon *et al.*, 2010). The *assA/bssA* nucleotide sequences and their translated products obtained in this study had been deposited in the NCBI database with accession numbers KC934958-KC934965.

## **2.4 Results**

### **2.4.1 Alkane biodegradation under methanogenic and sulfidogenic conditions**

Multiple methanogenic *n*-alkane-degrading enrichment cultures established from oil sands mature fine tailings similar to those described by Siddique *et al.* (2006; 2011) were prepared and incubated for several years, after which these were pooled and used as inoculum for the current study (Figure 1.3). Replicate 75-mL cultures were incubated under methanogenic conditions with 0.1 vol% of an alkane mixture composed of equal volumes of “hexanes”, *n*-C<sub>7</sub>, *n*-C<sub>8</sub> and *n*-C<sub>10</sub>. The “hexanes” comprised *n*-C<sub>6</sub> plus the C<sub>6</sub> isomers 2-methylpentane, 3-methylpentane and methylcyclopentane as impurities (details are given in Experimental Procedures). Methane accumulated in the cultures with

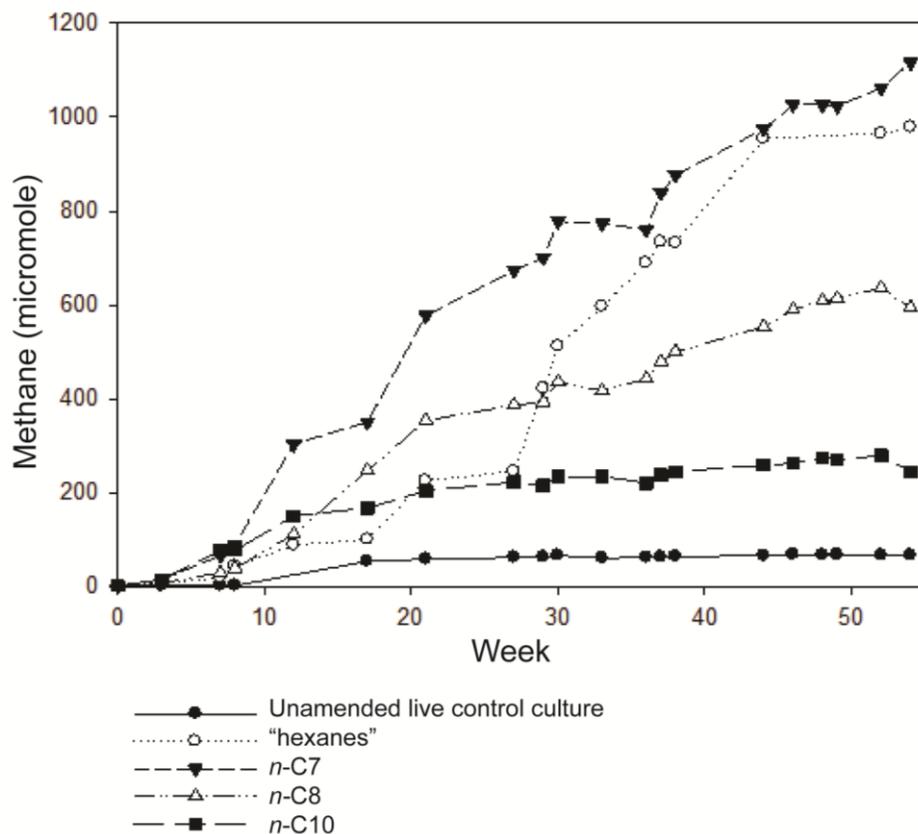
simultaneous depletion of *n*-C<sub>6</sub>, *n*-C<sub>7</sub>, *n*-C<sub>8</sub>, *n*-C<sub>10</sub>, 2-methylpentane and methylcyclopentane (Figure 2.1A). By 108 weeks (~ 2 years), ~1.2 mmol of methane had been produced, representing ~50% of the total theoretical methane yield based on the stoichiometric Symons and Buswell equation (Roberts, 2002).



**Figure 2.1** (A) Methane and volatile hydrocarbons detected in the headspace of methanogenic cultures, and (B) soluble sulfide and volatile hydrocarbons detected in sulfidogenic cultures during incubation at 28°C. Measurements were performed on independent triplicate cultures amended with 0.1 vol% of a C<sub>6</sub>-C<sub>10</sub> alkane mixture; live controls lacked the alkane amendment. Residual analyte was calculated from GC-MS peak areas as a percentage of the initial peak area after normalizing to the recalcitrant 3-methylpentane peak (see details in Experimental Procedures).

After incubation for approximately 18 months, 5 mL of methanogenic culture was transferred into 45 mL fresh medium containing 15 mM sodium sulfate and 0.1 vol% of alkane stock to establish replicate sulfidogenic cultures. In these cultures, sulfate amendment completely inhibited methanogenesis during 60 weeks of incubation but resulted in production of ~5 mM sulfide (Figure 2.1B). *n*-C<sub>6</sub>, *n*-C<sub>7</sub>, *n*-C<sub>8</sub> and *n*-C<sub>10</sub> were almost completely depleted by 60 weeks, whereas the depletion of 2-methylpentane and methylcyclopentane was minimal. The sulfide concentration remained unchanged by the time that almost all *n*-alkanes had been depleted, indicating that sulfate reduction was required for the oxidation of *n*-alkanes but not methylcyclopentane and 2-methylpentane. 3-methylpentane remained recalcitrant under both methanogenic and sulfate-reducing conditions throughout incubation, which was similarly observed in the initial inocula used in this study (and other unpublished data), which have been monitored for up to 4 years.

To confirm *n*-alkane degradation in the absence of 2-methylpentane and methylcyclopentane, the methanogenic enrichment cultures used in the current study were transferred into identical fresh methanogenic medium amended with only one of "hexanes", *n*-C<sub>7</sub>, *n*-C<sub>8</sub> or *n*-C<sub>10</sub>; A live culture lacking alkanes was included as a substrate-free baseline control. Accumulation of headspace methane in cultures amended with individual substrates compared to minimal methane production by the baseline control (<80 μmol) during 54 weeks incubation (Figure 2.2) confirmed that the methanogenic community was capable of degrading lower molecular weight *n*-alkanes in the absence of 2-methylpentane and methylcyclopentane; i.e., that co-metabolism was not required for methanogenic degradation of *n*-alkanes.



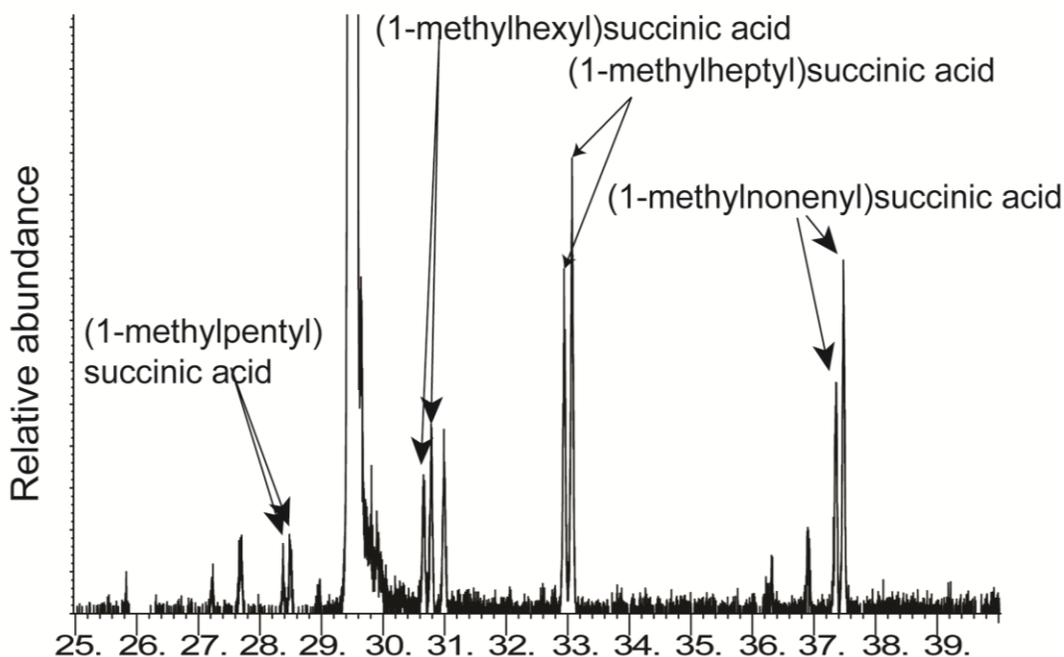
**Figure 2.2** Methane detected in the headspace of methanogenic cultures amended with individual “hexanes”, *n*-C<sub>7</sub> and *n*-C<sub>8</sub> or *n*-C<sub>10</sub> incubated for 54 weeks; unamended live control lacked the alkane amendment. The inoculum used in establishing these cultures was obtained from the enrichment cultures shown in Figure 2.1.

#### 2.4.2 Detection of putative succinylated metabolites of *n*-, iso- and cyclo-alkanes during methanogenesis and sulfidogenesis

Aliquots of culture fluids (2-200 mL) from methanogenic and sulfidogenic cultures were periodically collected, acid-extracted and derivatized for analysis by GC-MS and for comparison with known silylated alkylsuccinyl metabolites present in culture extracts from a sulfate-reducing enrichment culture of *Desulfoglaeba alkanexedens* ALDC grown with *n*-C<sub>6</sub>, *n*-C<sub>8</sub> and *n*-C<sub>10</sub>, provided by L. Gieg (University of Calgary). *D. alkanexedens* ADLC had been shown previously to use *n*-C<sub>6</sub> to *n*-C<sub>12</sub> and activate *n*-alkanes by addition to fumarate under sulfate-reducing condition (Davidova *et al.*, 2006).

#### 2.4.2.1.1 *n*-alkanes

In our methanogenic enrichment cultures, the alkylsuccinate metabolites expected from *n*-C<sub>6</sub>, *n*-C<sub>7</sub>, *n*-C<sub>8</sub>, and *n*-C<sub>10</sub> were not detected at any point during incubation (Table 2.1). However, when the same methanogenic cultures were transferred to sulfate-containing medium with the C<sub>6</sub>-C<sub>10</sub> alkane mixtures and incubated for 37 weeks, GC-MS detected putative metabolites having retention times and mass spectra identical to expected alkylsuccinates such as doubly-derivatized trimethylsilyl (di-TMS) esters of (1-methylpentyl)succinic acid, (1-methylhexyl)succinic acid, (1-methylheptyl)succinic acid, and (1-methylnonyl)succinic acids, corresponding to known metabolites arising from *n*-C<sub>6</sub>, *n*-C<sub>7</sub>, *n*-C<sub>8</sub>, and *n*-C<sub>10</sub> addition to fumarate (Table 2.1 and Figure 2.3).



**Figure 2.3** GC-MS profiles of silylated alkylsuccinates detected in the sulfate-reducing C<sub>6</sub>-C<sub>10</sub>-degrading culture incubated for 37 weeks

#### **2.4.2.1.2 *iso*-alkanes (2- methylpentane and 3-methylpentane)**

For silylated metabolites obtained from both methanogenic and sulfidogenic cultures, GC-MS also detected the di-TMS ester of the presumed (1,3-dimethylbutyl)succinic acid, putatively arising from addition of 2-methylpentane to fumarate. Supporting this attribution, the mass spectrum comprised ions with  $m/z$  331 ( $M+15=346$ ), plus other diagnostic ion fragments with  $m/z$  174, 262 and 217 (Figure 2.4 and 2.5). The methanogenic cultures grown with mixed substrates were further transferred into identical methanogenic medium, and incubated with 2-methylpentane as the only organic carbon source and incubated under identical conditions as the mixed substrates. GC-MS analysis of silylated extract from 2-methylpentane-amended cultures detected a metabolite with identical retention time and mass spectrum as the mixed substrate extract, confirming that this metabolite arose from 2-methylpentane. However, the presumed (1,3-dimethylbutyl)succinic acid appeared to be transient (Table 2.1), as it was detected in the methanogenic culture only at week 40, suggesting that this *iso*-alkane intermediate can be further metabolized and is not a dead-end product.

#### **2.4.2.1.3 *cyclo*-alkanes (methylcyclopentane and ethylcyclopentane)**

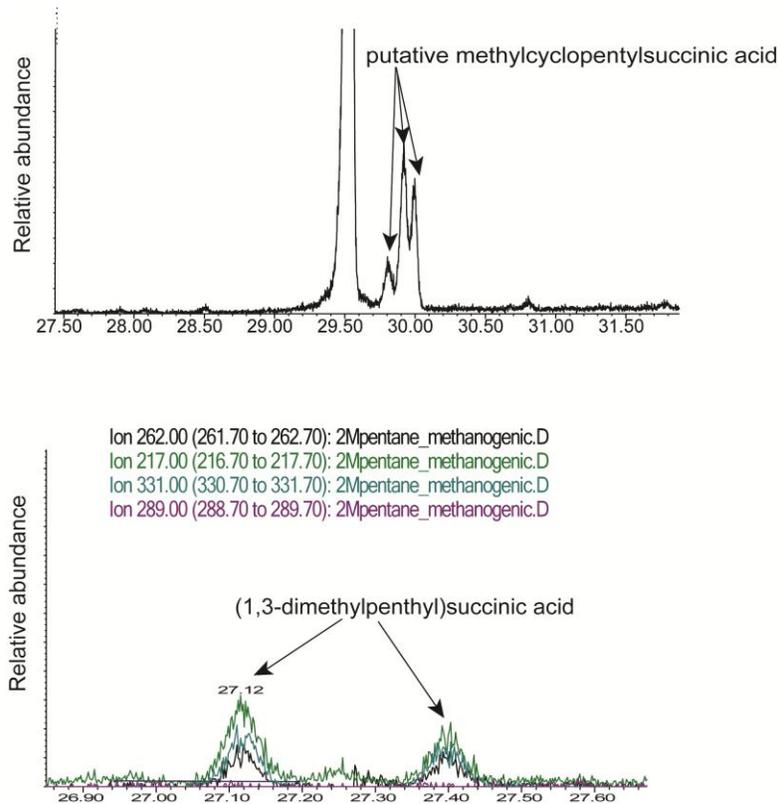
In methanogenic and sulfidogenic cultures with mixed substrates, GC-MS analysis of the silylated metabolites detected a cluster of three peaks at 29.9, 30.0 and 30.1 min that had identical mass spectra containing all the diagnostic ion fragments for fumarate addition and a fragment with  $m/z$  329 (Table 2.1, Figure 2.4 and Figure 2.5), likely representing the  $M-15$  ion of derivatized methylcyclopentylsuccinic acid (Rios-Hernandez *et al.*, 2003). The three peaks detected are putative isomers resulting from different positions of fumarate addition (Rios-Hernandez *et al.*, 2003). GC-MS analysis of culture fluids extracted from methanogenic cultures incubated for 27 and 37 weeks with methylcyclopentane as the only carbon source similarly detected three peaks with identical retention times and mass spectra (Table 2.1).

**Table 2.1** Putative alkylsuccinate metabolites detected in silylated culture extracts from methanogenic and sulfidogenic alkane-degrading enrichment cultures.

Parent compound	Retention times of metabolite peaks (min)	Selected ions (m/z)	Methanogenic conditions				Sulfidogenic conditions			Reference metabolites**
			C6-C10 mixture			Individual substrates		C6-C10 mixture		
			40 weeks	60 weeks	120 weeks	27 weeks	37 weeks	37 weeks	108 weeks	
<i>n</i> -C <sub>6</sub> *	28.4, 28.5	173, 217, 262, 331	-	-	-	-	-	+	-	+
<i>n</i> -C <sub>7</sub> *	30.7, 30.8	173, 217, 262, 345	-	-	-	-	-	+	-	<b>n/a</b>
<i>n</i> -C <sub>8</sub> *	32.9, 33.1	173, 217, 262, 359	-	-	-	-	-	+	-	+
<i>n</i> -C <sub>10</sub> *	37.4, 37.5	173, 217, 262, 387	-	-	-	-	-	+	-	+
2-methylpentane	27.1, 27.4	173, 217, 262, 331	+	-	-	+	+	+	-	<b>n/a</b>
3-methylpentane	None detected	None detected	-	-	-	<b>n/a</b>	-	-	-	<b>n/a</b>
methylcyclopentane	29.9, 30.0, 30.1	173, 217, 262, 329	+	+	+	+	+	+	+	<b>n/a</b>

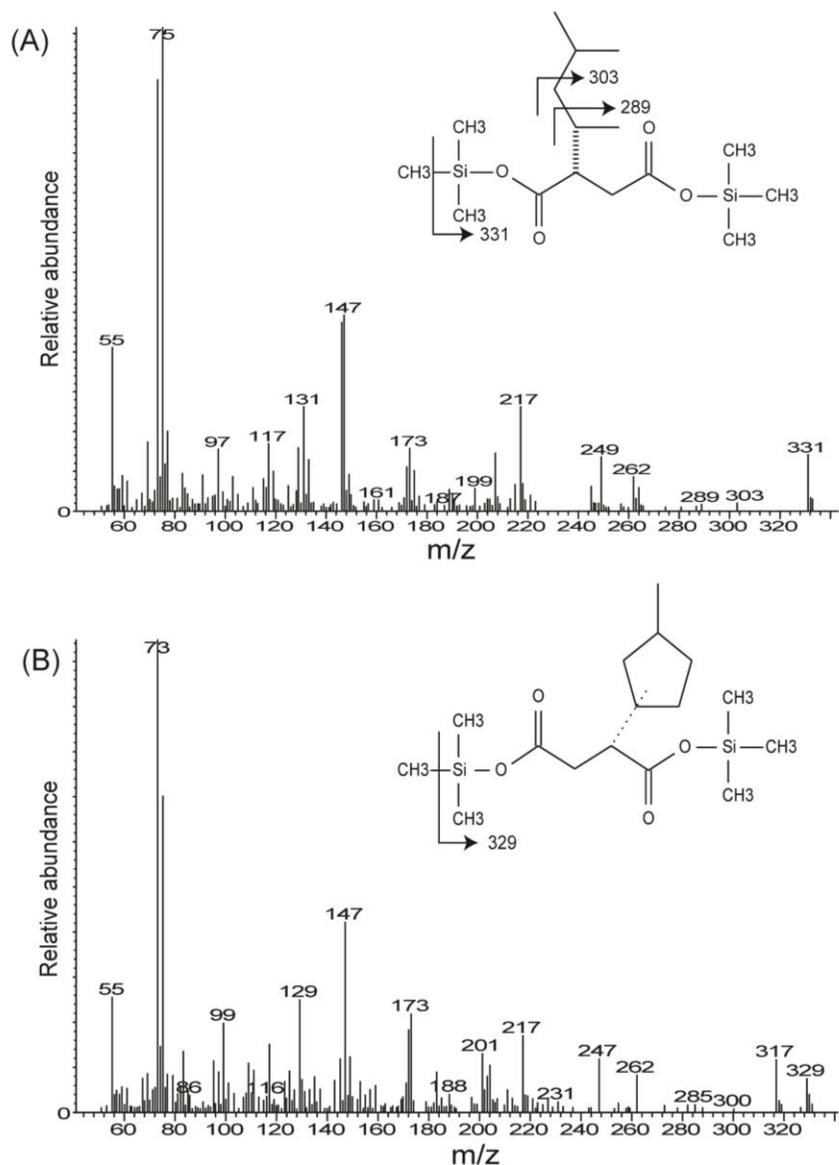
\*, inferred from metabolite spectrum; -, not detected; +, peak detected; n/a, not available

\*\* Reference metabolites obtained from culture extracts from *Desulfoglaeba alkanexedens* grown with *n*-C<sub>6</sub>, *n*-C<sub>7</sub> and *n*-C<sub>10</sub> under sulfate reducing-conditions, provided by L. Gieg, (University of Calgary)



**Figure 2.4** GC-MS profiles of silylated succinates detected in methanogenic enrichment culture with methylcyclopentane as the only carbon source and incubated for 27 weeks (Top); methanogenic enrichment culture with 2-methylpentane as the only carbon source, incubated for 27 weeks (bottom). Reconstructed ion chromatograms (RIC) were obtained using ions shown in Table 2.1.

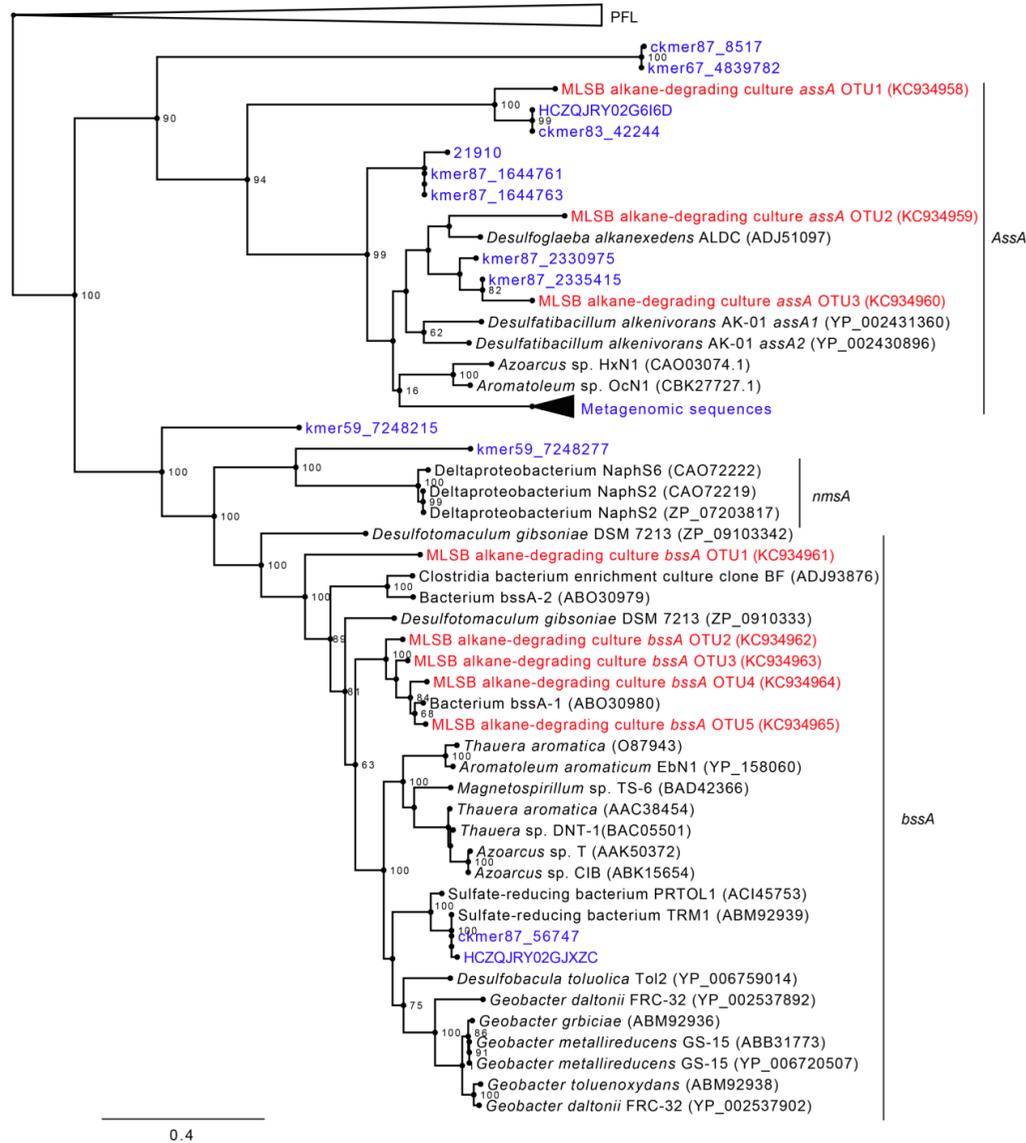
Similarly, transfer of the methanogenic cultures into identical methanogenic medium and addition of ethylcyclopentane as the only carbon source with incubation of ~ 1 year; followed by metabolite detection using GC-MS similarly detected three peaks with identical mass spectra representative of derivatized acids (results not shown). The long-term fate of methylcyclopentylsuccinic acid under both methanogenic and sulfidogenic conditions is currently being investigated because these compounds appear to be persistent, i.e., dead-end metabolites, even after incubation for >2 years when methanogenesis and sulfide production in respective cultures had reached plateaus (Table 2.1).



**Figure 2.5** Mass spectra of derivatized putative (1,3-dimethylpentyl)succinic acid (Top) and methylcyclopentylsuccinic acid (bottom) arising from addition of 2-methylpentane and methylcyclopentane, respectively, to fumarate. These were detected as silylated metabolites (Table 2.1) in both methanogenic and sulfidogenic cultures amended with either mixed substrates or single compounds. The deduced putative structures of TMS-derivatized (1,3-dimethylpentyl)succinic acid and methylcyclopentylsuccinic acid are shown in the insets with diagnostic m-15 ion fragments of m/z 331 and 329, respectively.

### 2.4.3 Reverse transcription-PCR of expressed *assA* and *bssA* genes

To investigate the expression of genes associated with fumarate activation of hydrocarbons, i.e., transcription of *assA* and *bssA* (encoding the  $\alpha$ -subunit of alkylsuccinate and benzylsuccinate synthase, respectively) under methanogenic conditions, total RNA was isolated from a 400-mL alkane-degrading culture (Fig. 1.3) during active methanogenesis and reverse transcribed. *assA* and *bssA* genes were amplified from this cDNA using primers 1995F/2467R (designed and tested as reported in Methods section of Chapter 6) and 7772f/8546r (Winderl *et al.*, 2007), followed by clone library construction and insert sequencing. Phylogenetic analysis of the translated clones plus reference and metagenomic sequences in Chapter 3 (Figure 3.3) confirmed that genes affiliated with *assA* and *bssA* were expressed by members of the enrichment culture during active methanogenesis and alkane biodegradation. Comparisons between the *assA* transcripts recovered in this study with those recovered from other environmental sources (Aitken *et al.*, 2013, Callaghan *et al.*, 2010, Zhou *et al.*, 2012) were not possible due to the non-overlapping positions of these sequences. Three *assA* genotypes (OTU1 to OTU3; n=18/20 clones) were detected, forming lineages distinct from those identified in pure cultures of the sulfate-reducing *Desulfatibacillum alkenivorans* AK-01 and nitrate-reducing *Azoarcus* HxN1 strains (Figure 2.6 and Appendix Table B1). The *assA* sequences representing OTU1 appear to be phylogenetically related to an *assA* gene presumably belonging to a Firmicutes member (ckmer83\_42244, Figure 2.6 and Figure 3.3 in Chapter 3). Five *bssA* genotypes were discerned, with a majority of these (OTU2 to OTU5; n=17/18 clones) forming a subclade with a known *bssA* transcript (accession number ABO30980.1) previously recovered from a methanogenic toluene-degrading culture dominated by an active bacterial species related to *Desulfotomaculum* spp. (Washer and Edwards 2007). The fifth *bssA* sequence (OTU1) appears to have no known close relatives.



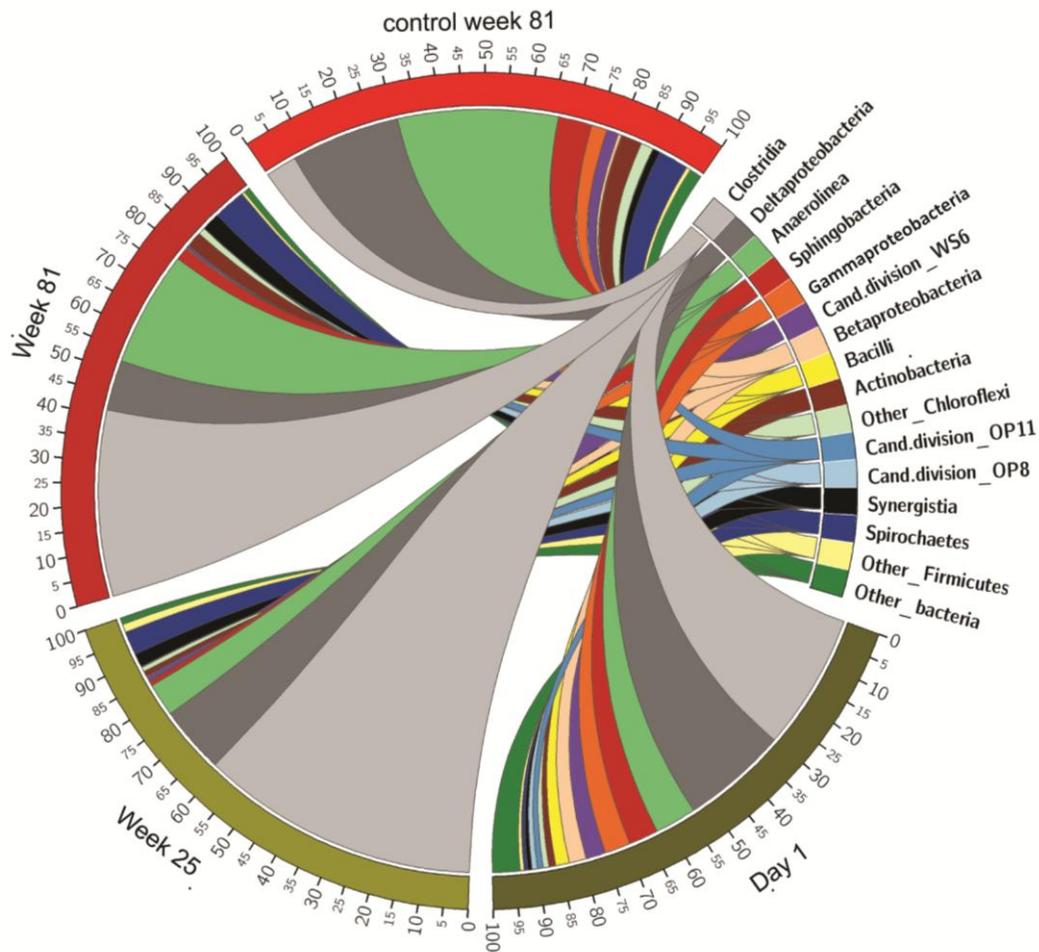
**Figure 2.6** Maximum Likelihood Tree of deduced amino acid sequences from cloned *assA* and *bssA* transcripts obtained from a methanogenic alkane-degrading culture (red text). All translated sequences obtained from the clone library were clustered at 2% distance cutoff, and a representative sequence from each operational taxonomic unit (OTU) was used in tree construction. Near-full length reference sequences are in black and sequences from metagenomes are shown in blue (Figure 3.3 in Chapter 3). Tree resampling was performed with 100 bootstrap replications; bootstrap values >60 are shown. GenBank or UniProt accession numbers are shown in parentheses. The tree was rooted with pyruvate formate lyase (PFL) in *Clostridium* sp. D5 (ZP\_08129105). The length of the trunk leading to the PFL outgroup has been shortened by 2/3. See Appendix Figure B1 for sequence alignment.

#### 2.4.4 Microbial community structure in methanogenic and sulfidogenic cultures

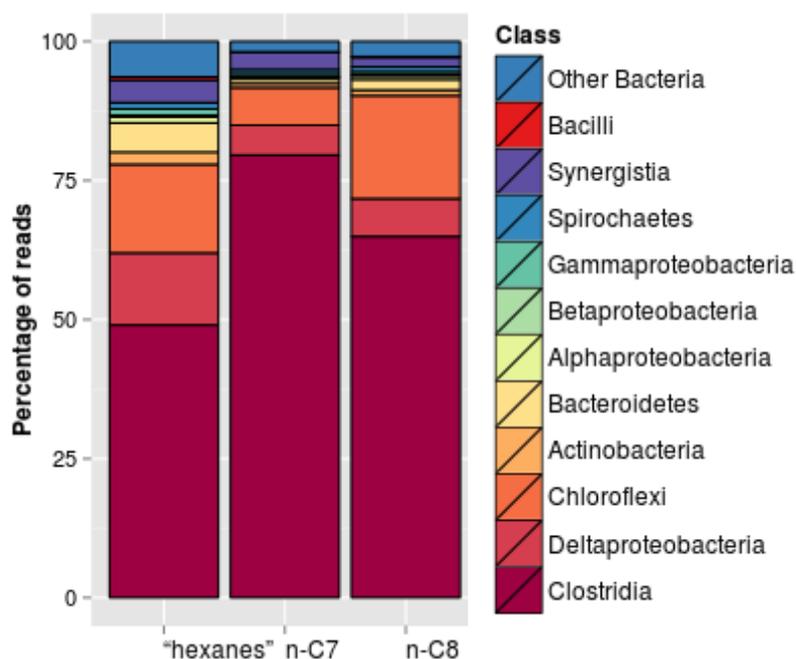
The microbial composition of methanogenic and sulfidogenic alkane-degrading enrichment cultures was investigated over a time course using pyrotag sequencing of the 16S rRNA genes. Aliquots from an established methanogenic enrichment culture sampled 1 d after transfer to fresh methanogenic medium with the C6-C10 alkane mixture showed a diverse bacterial community with Clostridia and Deltaproteobacteria as co-dominant classes (30% and 23% of total bacterial reads, respectively; day 1 in Figure 2.7). After incubation for 40 and 81 weeks while undergoing active alkane degradation and methanogenesis, pyrosequencing yielded bacterial reads dominated by Clostridia (~60 and 40% of total bacteria at weeks 40 and 81, respectively; Figure 2.7), Deltaproteobacteria (~15 and 11%), Anaerolinea (~7 and 26%) and Spirochaetes (~5 and 7%), with other bacterial classes comprising the remaining bacterial sequences. Notably, ~ 90% of the clostridial reads were taxonomically assigned to the family Peptococcaceae. In the live control culture incubated in methanogenic medium without alkanes, the bacterial community was dominated at 81 weeks by Anaerolinea (35%) and a host of other microbial taxa that collectively represented >70% of the total bacterial reads, whereas reads associated with Clostridiales (Peptococcaceae) had been reduced to less than 6% (Figure 2.7). Towards the end of the incubation with most alkanes being depleted (Figure 2.7; week 81), there was an increase of Anaerolinea.

The methanogenic cultures were further transferred into fresh methanogenic medium and incubated with individual alkanes, i.e., 'hexanes', *n*-C<sub>7</sub> or *n*-C<sub>8</sub> for 54 weeks. Pyrotag sequencing of the 16S rRNA genes from these cultures showed that the majority of the sequence reads were similarly dominated by the family Clostridia (~50% to 80%; Figure 2.8). The archaeal community was relatively stable during incubation under methanogenic conditions, with almost all archaeal reads being assigned to Euryarchaeota [approximately equal proportions of Methanosaetaceae and Methanomicrobiaceae persisting as co-dominant classes in amended and unamended live cultures (~80-90%, Figure

2.90)]. The remaining archaeal reads were mostly affiliated with *Methanolinea* and candidate genus *Methanoregula*.



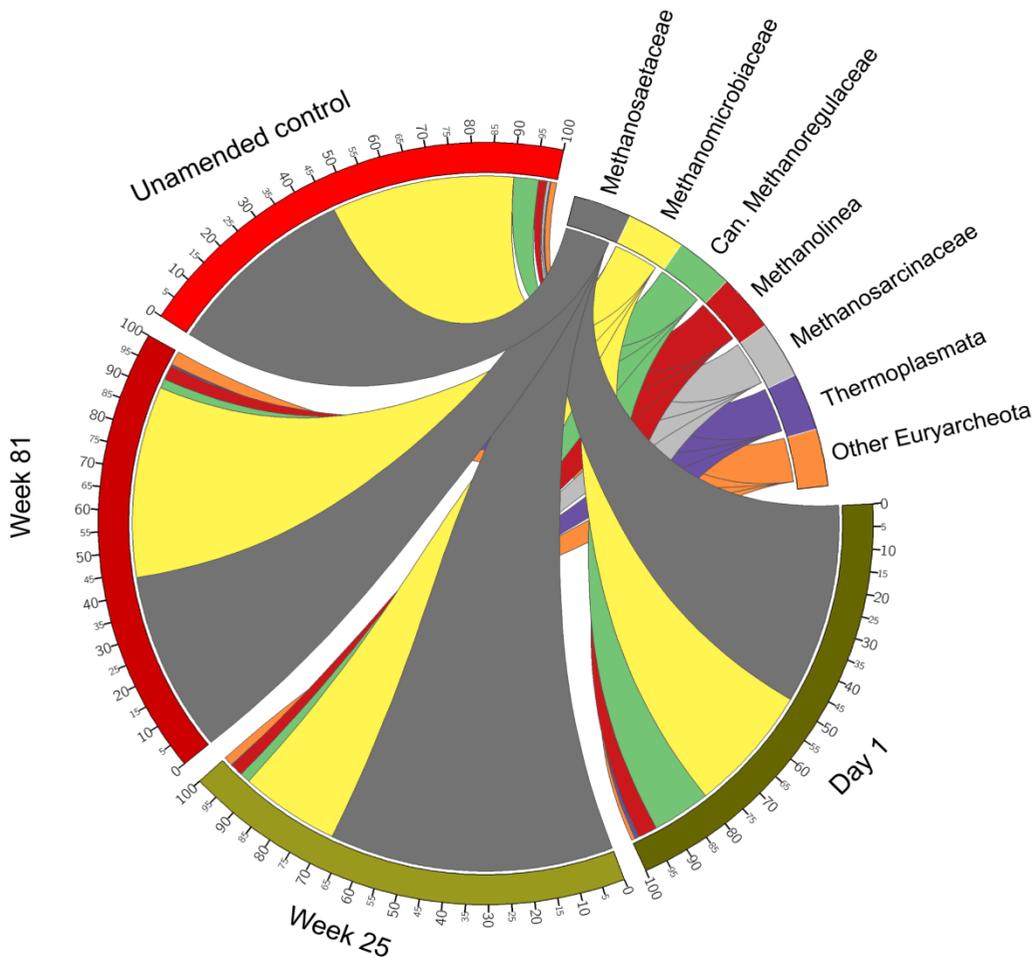
**Figure 2.7** Proportion of bacterial taxa in a methanogenic culture degrading a mixture of C<sub>6</sub>-C<sub>10</sub> alkanes after 1 day, 40 weeks and 81 weeks incubation, determined using pyrotag sequencing of 16S rRNA genes clustered at 5% distance cut-off and expressed as the percentage of total bacterial reads. A parallel live control culture incubated without alkanes for 81 weeks is included for comparison. Taxonomic affiliations were assigned using the RDP training sets.



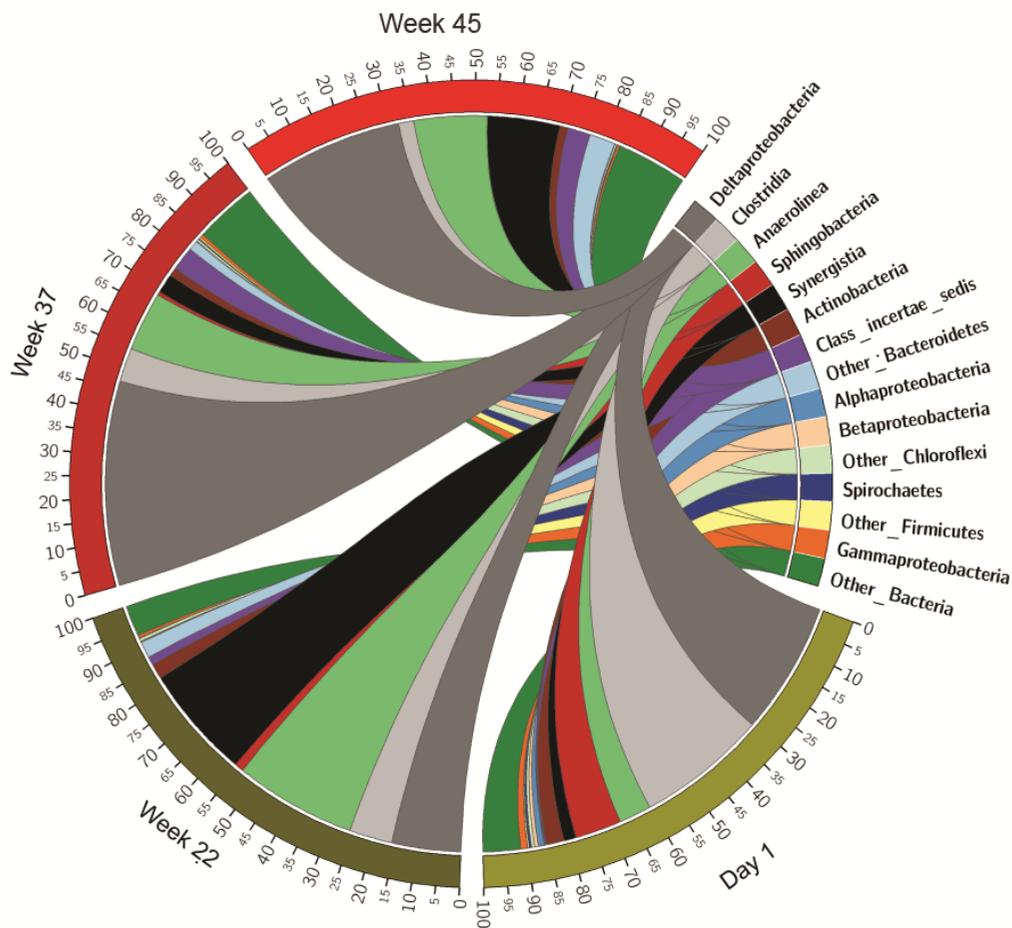
**Figure 2.8** Bacterial community structure of methanogenic alkane-degrading enrichment cultures individually amended with ‘hexanes’ (containing  $n$ -C<sub>6</sub>, methylcyclopentane, 2-methylpentane and 3-methylpentane) or  $n$ -C<sub>7</sub> or  $n$ -C<sub>8</sub> and incubated for 54 weeks, obtained from pyrotag sequencing of 16S rRNA genes clustered at 5% distance cut-off and expressed as a percentage of total bacterial reads. The columns are data from the methanogenic cultures shown in Figure 2.2. Taxa were identified from pyrotag sequences to the phylum level using the RDP training sets; “Other Bacteria” includes unclassified bacterial reads as well as singleton and doubleton reads.

To investigate the response of the microbial community to addition of sulfate, which yielded detectable succinylated alkane metabolites (Table 2.1), 5 mL of the methanogenic cultures were transferred into fresh methanogenic medium containing 15 mM sodium sulfate plus the C<sub>6</sub>-C<sub>10</sub> alkane mixture and incubated at room temperature in the dark. The microbial composition of the sulfidogenic enrichment culture at day 1 was similar to that in methanogenic enrichment cultures (Figure 2.7 and 2.10), as expected. After 22 weeks of incubation, which corresponded to the mid-phase of active sulfide production, sulfate-reduction and alkane-degradation (Figure 2.1), the bacterial 16S rRNA

gene were now dominated by Anaerolinea (28%), Synergistia (26%), Deltaproteobacteria (16%) and Clostridia (10%; Figure 2.10). By week 37, succinylated products of amended alkanes were detected (Table 2.1), and the bacterial reads were dominated by Deltaproteobacteria (~50%; Figure 2.10), whereas reads belonging to the Synergistia, Anaerolinea Clostridia had been reduced considerably (Figure 2.10).



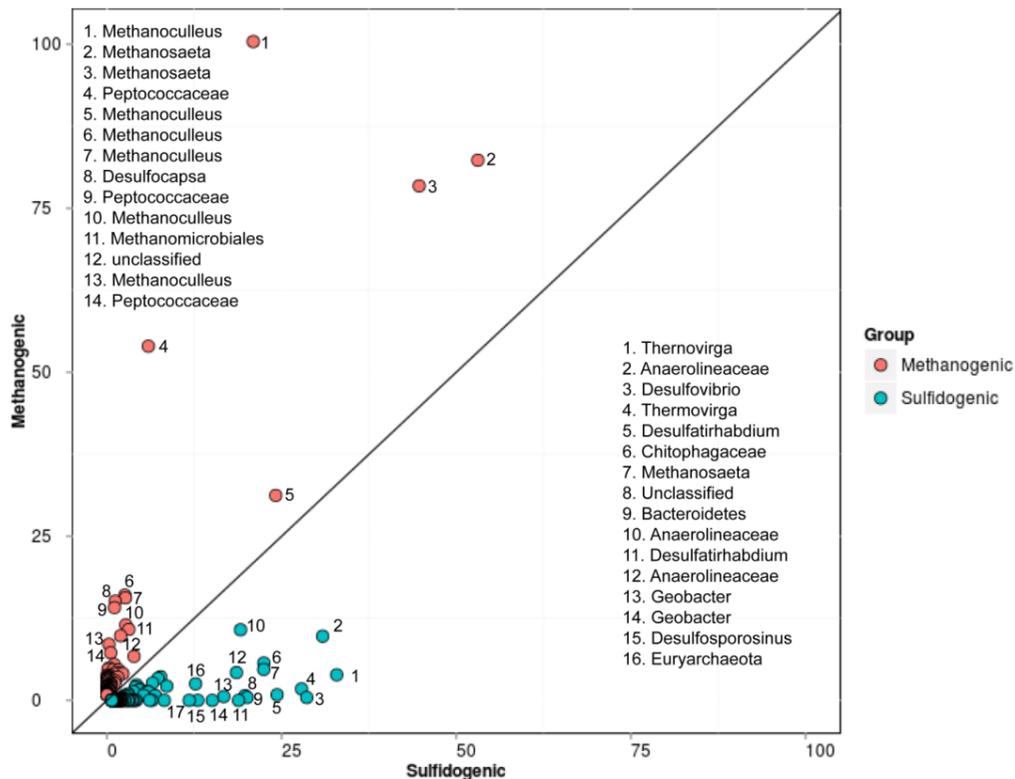
**Figure 2.9** Proportion of methanogen taxa in a methanogenic alkane-degrading culture after 1 day, 40 weeks and 81 weeks incubation, determined using pyrosequencing of 16S rRNA genes clustered at 5% distance cut-off and expressed as percentage of total archaeal reads. A parallel live control culture incubated without alkanes for 81 weeks is included for comparison. Taxonomic affiliations were assigned using the RDP training sets.



**Figure 2.10** Proportion of bacterial taxa in an alkane-degrading culture 1 day, 22 weeks, 37 weeks and 45 weeks after transfer from methanogenic to sulfate-reducing conditions, determined using pyrosequencing of 16S rRNA genes clustered at 5% distance cut-off and expressed as percentage of total bacterial reads. Taxonomic affiliations were assigned using the RDP training sets.

To further investigate community changes in response to sulfate amendment, Metastats (White *et al.*, 2009) was used to compare the differentially enriched 16S rRNA gene OTUs in cultures during active methanogenesis and sulfate reduction. Metastats employs false discovery rate and Fisher's exact test to detect differentially abundant features (OTU abundance in this case) in two treatment groups; in this case, the method is used to distinguish

changes in OTU proportions that are statistically significant (i.e. p-values <0.05) as a result of changing incubating conditions from methanogenic to sulfidogenic. Each point in Figure 2.11 represents a single 16S rRNA gene OTU clustered at 95% similarity with taxonomy assigned using RDP. Points falling below the diagonal are enriched under sulfate-reducing conditions whereas those falling above the diagonal are enriched under methanogenic conditions. The points falling close to the origin are significantly different but are present at low abundance. Overall, the analysis revealed significant enrichment of OTUs affiliated with Peptococcaceae (OTU 4 above the diagonal line in Figure 2.11), *Methanosaeta* and *Methanoculleus* in the methanogenic enrichment cultures (Figure 2.11), whereas there was significant enrichment of OTUs affiliated with Spirochaetales, Anaerolinea, Desulfobacterales, Desulfovibrionaceae and others in the sulfate-reducing enrichment cultures. Further phylogenetic analysis of the enriched Peptococcaceae-affiliated OTU showed that it is phylogenetically more closely related to *Desulfotomaculum* spp. and other environmental sequences recovered from hydrocarbon-contaminated sites (Appendix Figure B2).



**Figure 2.11** Comparison of differentially abundant Operational Taxonomic Units (OTUs) in methanogenic and sulfidogenic enrichment cultures determined using Metastats (see details in Experimental Procedures). All quality-controlled pyrotag sequences were clustered into OTUs at 5% distance cutoff, and taxonomic identity was assigned using the RDP training sets, with the lowest assigned taxon shown in the insets. Only OTUs that are significantly different ( $p$ -value  $< 0.05$ ) in abundance under either culturing condition are shown.

## 2.5 Discussion

### 2.5.1 Degradation of aliphatic alkanes by methanogenesis

Microbial consortia enriched from Mildred Lake Settling Basin (MLSB) have been shown previously to mineralize short chain  $n$ -alkanes ( $n$ -C<sub>6</sub> to  $n$ -C<sub>10</sub>), BTEX compounds (toluene,  $m$ - and  $p$ -xylenes) and long chain  $n$ -alkanes ( $n$ -C<sub>14</sub> to  $n$ -C<sub>18</sub>) under methanogenic conditions (Siddique *et al.*, 2006, Siddique *et al.*, 2007, Siddique *et al.*, 2011). In the present study, a methanogenic microbial consortium established from the same tailings pond simultaneously degraded low

molecular weight *n*-alkanes, methylcyclopentane and 2-methylpentane during 108 weeks incubation. The methane yield by week 108 was ~1.2 mmol, representing ~50% of the maximum theoretical yield according to the Simons and Buswell equation (Roberts, 2002). Previous reports of methane yields range from 64% of theoretical yield for pure hexadecane by an anaerobic mud enrichment culture (Zengler *et al.*, 1999), ~17% for a thermophilic C<sub>15</sub>-C<sub>20</sub> alkane-degrading consortium enriched from a petroleum reservoir (Zhou *et al.*, 2012) versus 77-79% for short chain *n*-alkanes (Siddique *et al.*, 2006, Siddique *et al.*, 2011) and 80-84% for longer chain *n*-alkanes (Siddique *et al.*, 2011) accomplished by oil sands tailings ponds enrichment consortia. The lower methane yields from the current tailings pond enrichments may be due in part to incomplete oxidation of substrates such as components of the 'hexanes' (Table 2.1): e.g., the persistent recalcitrance of 3-methylpentane, representing ~4% of the total C<sub>6</sub>-C<sub>10</sub> alkanes added to the culture, or accumulation of partially oxidized pathway intermediates of methylcyclopentane, representing ~4% of total alkanes. Other plausible explanations include greater assimilation of substrate into biomass or to differences in the microbial community structure reflected in metabolic pathway efficiencies, which may influence the observed methane yields. Questions about the mechanism(s) of alkane activation by bacterial members of the methanogenic community arise from these possibilities.

### **2.5.2 Evidence that methanogenic degradation of alkanes proceeds by addition to fumarate**

Anaerobic *n*-alkane activation by addition to fumarate has been widely reported to occur under sulfate- and nitrate-reducing conditions, with evidence provided by detection of characteristic succinylated metabolites from laboratory microbial consortia, pure isolates and contaminated environments (Agrawal and Gieg 2013, Callaghan 2013). In contrast, alkane activation by addition to fumarate has not yet been reported to occur under methanogenic or iron-reducing conditions (Mbadinga *et al.*, 2011). Although numerous studies have shown that alkanes could be biodegraded under methanogenic conditions in enrichment cultures or microcosms under laboratory conditions (Aitken *et al.*, 2013, Gray *et*

*al.*, 2011, Siddique *et al.*, 2006, Siddique *et al.*, 2011, Zengler *et al.*, 1999, Zhou *et al.*, 2012), alkylsuccinates have not been detected, invoking the possibility of an alternative activating mechanism (Aitken *et al.*, 2013). Recent single cell-sequencing and transcriptomic analysis of a methanogenic hexadecane-degrading culture (Embree *et al.*, 2013) concluded that *Smithella* spp., which had been implicated as the primary alkane-degrader under methanogenic conditions (Callaghan *et al.*, 2010, Gray *et al.*, 2011, Zengler *et al.*, 1999), may not be capable of addition to fumarate (See Chapter 4). However, in methanogenic cultures where *Syntrophus/Smithella* had been implicated as the primary degraders, the *assA* gene has been PCR-amplified and confirmed by cloning and sequencing (Aitken *et al.*, 2013, Callaghan 2013). The inability of these communities to catalyze fumarate addition under methanogenic condition would have been surprising because fumarate addition involving BSS had been demonstrated in other methanogenic toluene-degrading communities (Beller and Edwards 2000, Fowler *et al.*, 2012).

In the methanogenic cultures described here, putative succinylated metabolites of *iso*- and *cyclo*-alkanes (2-methylpentane and methylcyclopentane) were detected with concurrent detection of several genotypically distinct *assA* and *bssA* transcripts (i.e., cloned cDNA reverse transcribed from total RNA) during active alkane degradation, suggesting the presence of active ASS and BSS enzymes. It is interesting that expression of both *assA* and *bssA* was observed, implying that both alkyl- and benzylsuccinate synthases may be involved in the activation of *n*-alkanes, 2-methylpentane and methylcyclopentane. In fact, it is possible that the activity of the two succinyl synthases is responsible for the multiple peaks representing putative metabolite isomers (Table 2.1), although the substrate spectrum of the presumed ASS and BSS enzymes was not ascertained in this study. Because ASS appears to be promiscuous, and BSS tends to be substrate-specific (Rabus *et al.*, 2011), the depletion of *n*-alkanes is likely due to the activity of ASS instead of BSS. The ability of ASS-like enzymes to co-activate *n*- and *cyclo*-alkanes has been implicated previously. During growth in crude oil under nitrate-reducing conditions, the model organism *Azoarcus* spp.

HxN1, which activates *n*-alkanes by addition to fumarate, was found to be capable of co-activating *n*-alkanes, cyclopentane and methycyclopentane; although the downstream metabolites from *cyclo*-alkanes are likely to be dead-end products that do not support growth (Wilkes *et al.*, 2003). Furthermore, Callaghan *et al.* (2012) previously showed that *D. alkenivorans* AK-01, which activates *n*-alkanes under sulfate-reducing condition by addition to fumarate, is capable of producing metabolites from hexadecane to support the growth of *Methanospirillum hungatei* JF-1 in a co-culture under methanogenic conditions. These reports lend support to the notion that alkane activation by addition to fumarate can proceed under methanogenic conditions.

The reason(s) for lack of evidence for alkylsuccinate production in methanogenic conditions is currently unclear and may reflect limitations in detection and analytical method, and/or tight coupling of alkane biodegradation with high turnover of metabolites under methanogenic conditions may prevent accumulation of pathway intermediates in the culture medium. This assumption is plausible because in some studies, syntrophic consortia have been shown to accumulate very low amounts of metabolic by-products, and in some cases the expected metabolite such as formate was not detected (Walker *et al.*, 2012). For example, a previous study of a methanogenic toluene-degrading culture showed that benzylsuccinate (from toluene activation by addition to fumarate) appeared only transiently during incubation and only accumulated to maximum concentrations of <0.01 mol% of total toluene consumed (Beller and Edwards 2000). Substrate degradation and accumulation of metabolic by-products by syntrophic microbial consortia can also be dependent on multiple factors such as enzyme kinetics and affinity, and the thermodynamic requirements of the community (Dolfing *et al.*, 2008, Dwyer *et al.*, 1988, Hopkins *et al.*, 1995, Warikoo *et al.*, 1996).

### **2.5.3 Inferred role of Peptococcaceae in methanogenic alkane degradation**

The abundance of Peptococcaceae possibly related to *Desulfotomaculum* spp. detected in the present study in methanogenic alkane-degrading cultures suggests that these Firmicutes are likely responsible for primary alkane

degradation. This assumption is supported by the detection of abundant *assA* transcripts (*assA* OTU1) that are phylogenetically related to the Firmicutes-affiliated *assA* gene detected in the metagenome of a parallel alkane-degrading culture (Chapter 3 and 4). Species related to *Desulfotomaculum* spp. were previously implicated in fumarate activation of propane (Kniemeyer *et al.*, 2007) and ethylcyclopentane (Rios-Hernandez *et al.*, 2003) under sulfate-reducing conditions, but no isolate or enrichment culture is known to be capable of this process under methanogenic conditions (reviewed by Agrawal and Gieg 2013, Callaghan 2013). More often, members of *Desulfotomaculum* spp. are implicated in the degradation of monaromatic compounds such as toluene and xylene (Morasch *et al.*, 2004, Washer and Edwards 2007). This is in contrast to the findings of others (Gray *et al.*, 2011, Zengler *et al.*, 1999) that implicated species affiliated with *Syntrophus/Smithella* as the primary degraders of longer chain *n*-alkanes (e.g., *n*-hexadecane). The apparent differences between the microbial key players in methanogenic degradation of shorter chain *n*-alkanes (*n*-C<sub>6</sub>-C<sub>10</sub>) and longer chain *n*-alkanes (i.e., *n*-C<sub>16</sub>) is unknown but are possibly due to differences in the activating enzyme or the ability of the key microbial species to attach to oil droplets within culture fluids, since short chain *n*-alkanes such as decane have greater water solubility (9.5 mg/L) than hexadecane (0.0009 mg/L).

Alkane activation by addition to fumarate under sulfate-reducing conditions has been demonstrated in many studies, and bacterial species that are capable of doing so are mostly Proteobacteria, in particular sulfate-reducing Deltaproteobacteria (Agrawal and Gieg 2013, Mbadinga *et al.*, 2011). In the current study, amending an active methanogenic community with sulfate followed by prolonged incubation enriched for members of Deltaproteobacteria and not Peptococcaceae. Members of Deltaproteobacteria belonging to the order Desulfobacterales (e.g., *Desulfatirhabdium* and Desulfobulbaceae) detected in the sulfate-reducing enrichment culture are likely responsible for fumarate activation of *n*-alkanes under sulfate-reducing conditions. In contrast, the Peptococcaceae detected in abundance in the methanogenic culture likely are involved in fermentative alkane oxidation and do not couple this process to sulfate reduction.

Members of *Desulfotomaculum* couple substrate oxidation to sulfate reduction, but exceptions do exist for *Pelotomaculum* spp. that are incapable of sulfate reduction (Plugge *et al.*, 2011). The results presented here, therefore, suggest that degradation of aliphatic alkanes by oil sands microorganisms involved several primary alkane degraders that have very different physiological capabilities depending upon the terminal electron acceptor.

## 2.6 Significance

The detection of both hydrogenotrophic *Methanoculleus* spp. and acetoclastic *Methanosaeta* spp. in the current methanogenic alkane-degrading enrichment cultures signifies that methane evolution may be mediated by both groups of methanogens. Other microbes including members of Sphingobacteria, Synergistia and Anaerolinea have unknown roles in activating hydrocarbon substrates, and may be involved in autotrophic carbon fixation and hydrogen scavenging (Chapter 3). In other methanogenic alkane-degrading cultures, syntrophic alkane oxidizers such as Clostridiales have been proposed to be involved in acetate oxidation to yield CO<sub>2</sub> and H<sub>2</sub> for hydrogenotrophic methanogens (Gray *et al.*, 2011). Therefore, methanogenic alkane degradation may involve a large consortium of community members having different thermodynamic requirements (Dwyer *et al.*, 1988, Hopkins *et al.*, 1995, Warikoo *et al.*, 1996). The key reaction in all of these cultures relies on key players that are capable of initiating the primary attack on the hydrocarbon substrate, representing the limiting factor in methanogenesis. The detection of *assA* and *bssA* transcripts, along with putative succinylated products of 2-methylpentane and methylcyclopentane confirmed that fumarate activation is an important process in the initial attack on aliphatic hydrocarbons by oil sands microorganisms under methanogenic conditions. Under methanogenic conditions, syntrophs related to Peptococcaceae are likely to be involved in the primary attack on low molecular weight *n*-alkanes. Follow-up experimentation is expected to give new insights into the potential of this organism for alkane activation under methanogenic condition.

## 2.7 References

- Acosta-González A, Rossello-Mora R, Marques S. (2013). Diversity of benzylsuccinate synthase-like (*bssA*) genes in hydrocarbon-polluted marine sediments suggests substrate-dependent clustering. *Appl Environ Microbiol* **79**: 3667-3676.
- Agrawal A, Gieg LM. (2013). In situ detection of anaerobic alkane metabolites in subsurface environments. *Front Microbiol* **4**: 140.
- Aitken CM, Jones DM, Maguire MJ, Gray ND, Sherry A, Bowler BFJ *et al.*, (2013). Evidence that crude oil alkane activation proceeds by different mechanisms under sulfate-reducing and methanogenic conditions. *Geochim Cosmochim Acta* **109**: 162-174.
- Beller HR, Edwards EA. (2000). Anaerobic toluene activation by benzylsuccinate synthase in a highly enriched methanogenic culture. *Appl Environ Microbiol* **66**: 5503-5505.
- Berdugo-Clavijo C, Dong XL, Soh J, Sensen CW, Gieg LM. (2012). Methanogenic biodegradation of two-ringed polycyclic aromatic hydrocarbons. *FEMS Microbiol Ecol* **81**: 124-133.
- Callaghan AV, Wawrik B, Chadhain SMN, Young LY, Zylstra GJ. (2008). Anaerobic alkane-degrading strain AK-01 contains two alkylsuccinate synthase genes. *Biochem Biophys Res Commun* **366**: 142-148.
- Callaghan AV, Davidova IA, Savage-Ashlock K, Parisi VA, Gieg LM, Suflita JM *et al.*, (2010). Diversity of benzyl- and alkylsuccinate synthase genes in hydrocarbon-impacted environments and enrichment cultures. *Environ Sci Technol* **44**: 7287-7294.
- Callaghan AV. (2013). Enzymes involved in the anaerobic oxidation of *n*-alkanes: from methane to long-chain paraffins. *Front Microbiol* **4**: 89.
- Cline JD. (1969). Spectrophotometric determination of hydrogen sulfide in nature waters. *Limnol Oceanogr* **14**: 454-458.
- Davidova IA, Duncan KE, Choi OK, Suflita JM. (2006). *Desulfoglaeba alkanexedens* gen. nov., sp nov., an *n*-alkane-degrading, sulfate-reducing bacterium. *Int J Syst Evol Microbiol* **56**: 2737-2742.
- Dolfing J, Larter SR, Head IM. (2008). Thermodynamic constraints on methanogenic crude oil biodegradation. *Isme Journal* **2**: 442-452.
- Dwyer DF, Weegaerssens E, Shelton DR, Tiedje JM. (1988). Bioenergetic conditions of butyrate metabolism by a syntrophic, anaerobic bacterium in

coculture with hydrogen-oxidizing methanogenic and sulfidogenic bacteria. *Appl Environ Microbiol* **54**: 1354-1359.

Edgar RC. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32**: 1792-1797.

Fedorak PM, Coy DL, Dudas MJ, Simpson MJ, Renneberg AJ, MacKinnon MD. (2003). Microbially-mediated fugitive gas production from oil sands tailings and increased tailings densification rates. *J Environ Eng Sci* **2**: 199-211.

Foght J, Aislabie J, Turner S, Brown CE, Ryburn J, Saul DJ *et al.*,. (2004). Culturable bacteria in subglacial sediments and ice from two Southern Hemisphere glaciers. *Microb Ecol* **47**: 329-340.

Fowler SJ, Dong XL, Sensen CW, Suflita JM, Gieg LM. (2012). Methanogenic toluene metabolism: community structure and intermediates. *Environ Microbiol* **14**: 754-764.

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

Gieg LM, Duncan KE, Suflita JM. (2008). Bioenergy production via microbial conversion of residual oil to natural gas. *Appl Environ Microbiol* **74**: 3022-3029.

Gieg LM, Davidova IA, Duncan KE, Suflita JM. (2010). Methanogenesis, sulfate reduction and crude oil biodegradation in hot Alaskan oilfields. *Environ Microbiol* **12**: 3074-3086.

Gray ND, Sherry A, Grant RJ, Rowan AK, Hubert CRJ, Callbeck CM *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 *et al.*,. (2008). Genes encoding the candidate enzyme for anaerobic activation of *n*-alkanes in the denitrifying bacterium, strain HxN1. *Environ Microbiol* **10**: 376-385.

Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. *Syst Biol* **59**: 307-321.

Hopkins BT, Mcinerney MJ, Warikoo V. (1995). Evidence for an anaerobic syntrophic benzoate degradation threshold and isolation of the syntrophic benzoate degrader. *Appl Environ Microbiol* **61**: 526-530.

Jones DM, Head IM, Gray ND, Adams JJ, Rowan AK, Aitken CM *et al.*, (2008). Crude-oil biodegradation via methanogenesis in subsurface petroleum reservoirs. *Nature* **451**: 176-U176.

Kimes NE, Callaghan AV, Aktas DF, Smith WL, Sunner J, Golding B *et al.*, (2013). Metagenomic analysis and metabolite profiling of deep-sea sediments from the Gulf of Mexico following the Deepwater Horizon oil spill. *Front Microbiol* **4**: 50.

Kniemeyer O, Musat F, Sievert SM, Knittel K, Wilkes H, Blumenberg M *et al.*, (2007). Anaerobic oxidation of short-chain hydrocarbons by marine sulphate-reducing bacteria. *Nature* **449**: 898-U810.

Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D *et al.*, (2009). Circos: An information aesthetic for comparative genomics. *Genome Res* **19**: 1639-1645.

Li W, Godzik A. (2006). Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **22**: 1658-1659.

Ludwig W, Strunk O, Westram R, Richter L, Meier H, Yadhukumar *et al.*, (2004). ARB: a software environment for sequence data. *Nucleic Acids Res* **32**: 1363-1371.

Mbadinga SM, Wang LY, Zhou L, Liu JF, Gu JD, Mu BZ. (2011). Microbial communities involved in anaerobic degradation of alkanes. *Int Biodeterior Biodegrad* **65**: 1-13.

Morasch B, Schink B, Tebbe CC, Meckenstock RU. (2004). Degradation of *o*-xylene and *m*-xylene by a novel sulfate-reducer belonging to the genus *Desulfotomaculum*. *Arch Microbiol* **181**: 407-417.

Parisi VA, Brubaker GR, Zenker MJ, Prince RC, Gieg LM, da Silva MLB *et al.*, (2009). Field metabolomics and laboratory assessments of anaerobic intrinsic bioremediation of hydrocarbons at a petroleum-contaminated site. *Microbial Biotechnology* **2**: 202-212.

Plugge CM, Zhang W, Scholten JC, Stams AJ. (2011). Metabolic flexibility of sulfate-reducing bacteria. *Front Microbiol* **2**: 81.

Prince RC, Suflita JM. (2007). Anaerobic biodegradation of natural gas condensate can be stimulated by the addition of gasoline. *Biodegradation* **18**: 515-523.

Pruesse E, Peplies J, Glockner FO. (2012). SINA: accurate high-throughput multiple sequence alignment of ribosomal RNA genes. *Bioinformatics* **28**: 1823-1829.

Rabus R, Jarling R, Lahme S, Kuhner S, Heider J, Widdel F *et al.*, (2011). Co-metabolic conversion of toluene in anaerobic *n*-alkane-degrading bacteria. *Environ Microbiol* **13**: 2576-2585.

Rios-Hernandez LA, Gieg LM, Suflita JM. (2003). Biodegradation of an alicyclic hydrocarbon by a sulfate-reducing enrichment from a gas condensate-contaminated aquifer. *Appl Environ Microbiol* **69**: 434-443.

Roberts D J. Methods for assessing anaerobic biodegradation potential. In *Manual of Environmental Microbiology*, 2nd ed.; Hurst, C. J., Crawford, R. L., Knudson, G. R., McInerney, M. J., Stetzenbach, L. D., Eds.; ASM Press: Washington, DC, 2002; pp 1008–1017.

Rosana ARR, Chamot D, Owttrim GW. (2012). Autoregulation of RNA helicase expression in response to temperature stress in *Synechocystis* sp PCC 6803. *Plos One* **7**.

Siddique T, Fedorak PM, Foght JM. (2006). Biodegradation of short-chain *n*-alkanes in oil sands tailings under methanogenic conditions. *Environ Sci Technol* **40**: 5459-5464.

Siddique T, Fedorak PM, McKinnon MD, Foght JM. (2007). Metabolism of BTEX and naphtha compounds to methane in oil sands tailings. *Environ Sci Technol* **41**: 2350-2356.

Siddique T, Penner T, Semple K, Foght JM. (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, Nesbo C, Foght JM. (2012). Microbial communities involved in methane production from hydrocarbons in oil sands tailings. *Environ Sci Technol* **46**: 9802-9810.

So CM, Phelps CD, Young LY. (2003). Anaerobic transformation of alkanes to fatty acids by a sulfate-reducing bacterium, strain Hxd3. *Appl Environ Microbiol* **69**: 3892-3900.

Soh J, Dong X, Caffrey SM, Voordouw G, Sensen CW. (2013). Phoenix 2: A locally installable large-scale 16S rRNA gene sequence analysis pipeline with Web interface. *J Biotechnol* **167**: 393-403.

Tan B, Dong XL, Sensen CW, Foght JM. (2013). Metagenomic analysis of an anaerobic alkane-degrading microbial culture: potential hydrocarbon-activating pathways and inferred roles of community members. *Genome* **56**: 599-611.

- Walker CB, Redding-Johanson AM, Baidoo EE, Rajeev L, He Z, Hendrickson EL *et al.*, (2012). Functional responses of methanogenic archaea to syntrophic growth. *ISME J* **6**: 2045-2055.
- Warikoo V, McInerney MJ, Robinson JA, Suflita JM. (1996). Interspecies acetate transfer influences the extent of anaerobic benzoate degradation by syntrophic consortia. *Appl Environ Microbiol* **62**: 26-32.
- Washer CE, Edwards EA. (2007). Identification and expression of benzylsuccinate synthase genes in a toluene-degrading methanogenic consortium. *Appl Environ Microbiol* **73**: 1367-1369.
- White JR, Nagarajan N, Pop M. (2009). Statistical methods for detecting differentially abundant features in clinical metagenomic samples. *PLoS Comp Biol* **5**: e1000352.
- Wilkes H, Kuhner S, Bolm C, Fischer T, Classen A, Widdel F *et al.*, (2003). Formation of *n*-alkane- and *cyclo*-alkane-derived organic acids during anaerobic growth of a denitrifying bacterium with crude oil. *Org Geochem* **34**: 1313-1323.
- Winderl C, Schaefer S, Lueders T. (2007). Detection of anaerobic toluene and hydrocarbon degraders in contaminated aquifers using benzylsuccinate synthase (*bssA*) genes as a functional marker. *Environ Microbiol* **9**: 1035-1046.
- Yarza P, Ludwig W, Euzéby J, Amann R, Schleifer K-H, Gloeckner FO *et al.*, (2010). Update of the All-Species Living Tree Project based on 16S and 23S rRNA sequence analyses. *Syst Appl Microbiol* **33**: 291-299.
- Zengler K, Richnow HH, Rossello-Mora R, Michaelis W, Widdel F. (1999). Methane formation from long-chain alkanes by anaerobic microorganisms. *Nature* **401**: 266-269.
- Zhou L, Li K-P, Mbadinga SM, Yang S-Z, Gu J-D, Mu B-Z. (2012). Analyses of *n*-alkanes degrading community dynamics of a high-temperature methanogenic consortium enriched from production water of a petroleum reservoir by a combination of molecular techniques. *Ecotoxicology (London, England)* **21**: 1680-1691.

### 3 Metagenomic analysis of an anaerobic alkane-degrading microbial culture: Potential hydrocarbon-activating pathways and inferred roles of community members<sup>1</sup>

#### 3.1 Abstract

A microbial community (short-chain alkane-degrading culture, SCADC) enriched from an oil sands tailings pond was shown to degrade C<sub>6</sub>–C<sub>10</sub> alkanes under methanogenic conditions. Total genomic DNA from SCADC was subjected to 454 pyrosequencing, Illumina paired-end sequencing, and 16S rRNA gene amplicon pyrotag sequencing; the latter revealed 320 operational taxonomic units (OTUs) at 5% distance. Metagenomic sequences were subjected to in-house quality control and co-assembly, yielding 984 086 contigs, followed by annotation using MG-RAST and IMG. Substantial nucleotide and protein recruitment to *Methanosaeta concilii*, *Syntrophus aciditrophicus* and *Desulfobulbus propionicus* reference genomes suggested the presence of closely related strains in SCADC; other genomes were not well mapped, reflecting the paucity of suitable reference sequences for such communities. Nonetheless, we detected numerous homologues of putative hydrocarbon succinate synthase genes (e.g., *assA*, *bssA*, and *nmsA*) implicated in anaerobic hydrocarbon degradation, suggesting the ability of the SCADC microbial community to initiate methanogenic alkane degradation by addition to fumarate. Annotation of a large contig revealed analogues of the *ass* operon 1 in the alkane-degrading sulfate-reducing bacterium *Desulfatibacillum alkenivorans* AK-01. Despite being enriched under methanogenic–fermentative conditions, additional metabolic functions inferred by COG profiling indicated multiple CO<sub>2</sub> fixation pathways, organic acid utilization, hydrogenase activity, and sulfate reduction.

<sup>1</sup>A version of this chapter has been published. Tan BF, Dong XL, Sensen CW 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.

### 3.2 Introduction

Anaerobic biodegradation of petroleum hydrocarbons can be an important process in anoxic environments (Berdugo-Clavijo *et al.*, 2012, Gray *et al.*, 2010, Head *et al.*, 2003). For example, biodegradation occurring in petroleum reservoirs over millennia can result in crude oils with altered hydrocarbon profiles and reduced economic value (Head *et al.*, 2003, Jones *et al.*, 2008). Field studies at sites contaminated with spilled petroleum or refined products often show that hydrocarbon contaminants undergo anaerobic bioremediation through the activities of the indigenous microflora (Parisi *et al.*, 2009).

Anaerobic biodegradation of hydrocarbons requires activation of the substrate, with the most widely reported mechanism being enzymatic addition of the hydrocarbon across the double bond of fumarate (Beller 2000, Grundmann *et al.*, 2008, Selesi *et al.*, 2010). Currently, three phylogenetically related enzymes catalyzing this addition have been identified: benzylsuccinate synthase (BSS) involved in toluene and xylene activation (Beller 2000); alkylsuccinate synthase (ASS; also known as 1-methylalkyl-succinate synthase [MAS]) involved in *n*-alkane activation (Callaghan *et al.*, 2008, Grundmann *et al.*, 2008); and 2-methylnaphthylsuccinate synthase (NMS) involved in 2-methylnaphthalene activation (Selesi *et al.*, 2010). ASS-mediated activation is believed to play a major role *in situ* in *n*-alkane-contaminated environments under nitrate- and sulfate-reducing conditions, documented by detection of homologous genes (Callaghan *et al.*, 2010) and succinylated ‘signature metabolites’ (Parisi *et al.*, 2009).

However, the possibility of *n*-alkane activation by addition to fumarate under strictly methanogenic conditions has been recently questioned (Aitken *et al.* 2013) and alternate mechanisms have been proposed (Callaghan 2013). Regardless of the initial reaction, hydrocarbon biodegradation coupled to methanogenesis requires the participation of a diverse microbial community including Bacteria capable of an initial anaerobic attack on hydrocarbons, subsequent conversion of degradation products by bacterial syntrophs and

ultimately by archaeal methanogens to yield methane and carbon dioxide (Berdugo-Clavijo *et al.*, 2012, Fowler *et al.*, 2012, Gray *et al.*, 2011).

In northern Alberta, Canada, surface mining of oil sands ores and extraction of bitumen using hot water and hydrocarbon solvents produces tailings wastes that are retained in large, anaerobic ‘tailings ponds’ (Chalaturnyk *et al.*, 2002). Several of these tailings ponds have become methanogenic (Holowenko *et al.*, 2000; Fedorak *et al.*, 2002; Fedorak *et al.*, 2003) due to biodegradation of residual bitumen-extraction solvents by diverse microbial communities indigenous to the ponds (Siddique *et al.* 2006; Siddique *et al.* 2007; Penner and Foght 2010). Increased understanding of the mechanisms and microbes involved in conversion of hydrocarbons to methane in oil sands tailings ponds should contribute to the long term management and monitoring of the ponds, and also the design of remediation techniques for their future reclamation.

In the current study, we analyze the metagenome of an enrichment culture established from indigenous microbes in oil sands tailings collected from Mildred Lake Settling Basin. This short-chain alkane-degrading culture (SCADC) is capable of degrading alkanes with the simultaneous production of methane during incubation in the dark under strictly methanogenic conditions. Total genomic DNA extracted from this culture was subjected to high-throughput sequencing using the 454 GS FLX Titanium system and paired-end HiSeq2000 Illumina sequencing, followed by bioinformatic analysis of the sequence assembly, as well as pyrosequencing of PCR-amplified 16S rRNA genes, allowing us to study the potential alkane biodegradation pathways and to infer the metabolic roles of community members within the SCADC enrichment culture.

### **3.3 Materials and methods**

#### **3.3.1 Description of SCADC enrichment culture**

A description of the establishment of the SCADC from oil sands mature fine tailings is presented in Figure B1. SCADC was maintained on aliphatic hydrocarbons by amending mineral methanogenic medium (Widdel and Bak 1992) with 0.1 vol% of an alkane mixture comprising equal volumes of ‘hexanes’ (CAS 110-54-3, Fisher Scientific), *n*-heptane (*n*-C<sub>7</sub>; CAS 142-82-5, >97% purity,

Fisher Scientific), *n*-octane (*n*-C<sub>8</sub>; CAS 111-65-9, >98% purity, Sigma-Aldrich) and *n*-decane (*n*-C<sub>10</sub>; CAS 124-18-5, >99% purity, Sigma-Aldrich). The 'hexanes' comprised predominantly *n*-C<sub>6</sub> (62% of 'hexanes') but also contained the hexane isomers 2-methylpentane (3%), 3-methylpentane (16%) and methylcyclopentane (19%; Figure S2). Incubation was carried out at 28°C under strictly anaerobic conditions using culture medium described in Figure S1 with a headspace of O<sub>2</sub>-free 30% CO<sub>2</sub>-balance N<sub>2</sub>. Duplicate large-volume cultures (400 mL each) were used for extraction of genomic DNA and parallel triplicate small-volume cultures (75 mL each) were used to measure methanogenesis and volatile hydrocarbon depletion, as described below. Duplicate 75-mL unamended enrichment cultures were incubated in parallel as baseline controls for methanogenesis.

Methane production was measured by sampling 50 mL of culture headspace using a sterile needle and syringe followed by gas chromatography with a flame ionization detector (GC-FID) as previously described (Siddique *et al.*, 2006, Siddique *et al.*, 2007b). To analyze residual volatile hydrocarbons during incubation, 50 mL of culture headspace were sampled and injected into an Agilent 6890N gas chromatograph with an inert mass selective detector (GC-MS; Agilent model 5973) fitted with an Agilent HP-5MS capillary column (30 m × 0.25 mm ID, 0.25 µm film thickness; J + W Scientific). Headspace hydrocarbon depletion was assessed using the GC-MS method described by (Prince and Suflita 2007) with 3-methylpentane (a recalcitrant component of 'hexanes') as the internal standard.

### **3.3.2 Nucleic acid extraction and high throughput sequencing**

After incubation for approximately 4 months, 2 x 400 ml of SCADC were subjected to DNA extraction at the University of British Columbia using the method of Wright *et al.* (2009). Briefly, culture fluid was filtered through a 0.22 µm Sterivex filter (Millipore, US), followed by total DNA extraction using phenol chloroform and further purification using cesium chloride density gradient ultracentrifugation (Wright *et al.* 2009). Total genomic DNA was sequenced using 454 GS FLX and Illumina HiSeq2000 instruments.

Amplification of 16S rRNA genes from the genomic DNA was carried out in a Mastercycler® PRO thermocycler (Eppendorf) using the primer pair described by Berdugo-Clavijo *et al.* (2012) with the following touchdown thermocycling protocol: 95°C for 5 min, followed by 10 cycles of 95°C for 30 s, 60°C for 30 s (decreasing by 0.5°C/cycle) and 72°C for 30 s, followed by 30 cycles of 95°C for 30 s, 55°C for 30 s and 72°C for 30 s, followed by final extension at 72°C for 10 min. The PCR reaction contained 2.5 µM of each primer, 5x KAPA2G reaction buffer A (Kapa Biosystems, Woburn MA), 5 µL of KAPA enhancer 1 solution, 1.25 µL DMSO, 0.1 µL DNA polymerase (KAPA2G), 0.5 µL dNTP mix (10mM, KAPA2G), 2 µL MgCl<sub>2</sub> (25mM), 0.5 µL of template DNA and 5.15 µL of sterile nuclease free water. The 16S rRNA amplicons were sequenced using a GS FLX Titanium Series Kit XLR70 (Roche). All sequencing was performed at McGill University Génome Québec Innovation Centre, Canada.

### **3.3.3 Quality control and *de novo* assembly of metagenomic data**

All raw 454 reads were passed through an in-house quality control program that filters out the low quality reads, specifically reads that: (i) contained ambiguous bases, (ii) had an average quality score below 25, (iii) contained homopolymers of lengths >6, or (iv) were shorter than 100 bp. The artificial duplicates generated during the 454 sequencing were identified using UCLUST with “--id 0.90 --idprefix 5” options and removed. All reads that passed quality control were assembled with Newbler V2.6 (Roche) using options “-mi 95 -ml 60 -a 200 -l 900”. Illumina raw reads were subjected to quality control and assembly process in a manner similar to that described by Saidi-Mehrabad *et al.* (2012). 454 contigs and singletons longer than 200 bp were merged with Illumina contigs using minimus2 from the AMOS package (Sommer *et al.*, 2007). All 454 reads passing quality control and the contigs from the 454 and Illumina assemblies were submitted to IMG/M (Markowitz *et al.*, 2012) and MG-RAST v3.3.3.1 (Meyer *et al.*, 2008) for automated annotation. All sequence data have been submitted to the Short Read Archive under SRA number SAMN01828453.

### 3.3.4 Taxonomic classification of metagenomics reads

Taxonomic classification of metagenomic reads was conducted using two approaches. Contigs longer than 1000 bp were classified using PhylopythiaS with the generic model (Patil *et al.*, 2012), whereas unassembled 454 reads were subjected to taxonomic classification using Sort-ITEMS (Haque *et al.*, 2009) and RITA (MacDonald *et al.*, 2012). RITA was executed with the following pipeline parameters: NB\_DCMEGABLAST, NB\_BLASTN, NB\_BLASTX, DCMEGABLAST\_RATIO, BLASTN\_RATIO, BLASTX\_RATIO, NB\_ML, NULL\_LABELLER --dblastne 10 --blastne 10 --blastxe 10 --dblastne 10 blastne 10 --blastxratio 10. Nucleotide recruitment of the 454 and Illumina hybrid assembly to reference genomes was conducted using Geneious Pro (Kearse *et al.*, 2012) with the following parameters: maximum gaps allowed per read, 10%; maximum gap size, 10; minimum overlap identity, 95%; maximum mismatches per read, 10%; and maximum ambiguity allowed, 4. The sequence reads obtained from pyrosequencing of 16S rRNA amplicons were analyzed using the Phoenix 2 pipeline developed in-house and previously described (Soh *et al.*, 2013).

### 3.3.5 Phylogenetic and bioinformatics analyses

Many genes involved in anaerobic hydrocarbon degradation, including *assA/bssA/nmsA* homologues and *ebdA* (Figure 1.2 and Table Appendix C1), share high sequence homology to other phylogenetically-related genes. For example, in the case of *assA/bssA/nmsA*, these genes are often mis-annotated as pyruvate-formate lyases when using the annotation pipeline in IMG and MG-RAST. Therefore, in addition to using the annotation output from MG-RAST and IMG, reference sequences of genes of interest (Appendix Table C1) were used in tBLASTn (Gertz *et al.*, 2006) searches against the SCADC sequence assembly. Translated amino acid sequences were first aligned using MUSCLE V3.3 (Edgar 2004) along with sequences obtained from the NCBI database. The sequence alignment was manually edited by removing stop codons that had been introduced as a result of sequencing or assembly errors. The alignment was subjected to maximum likelihood tree construction using the PhyML v 3.0 web server (Guindon and Gascuel 2003) with the WAG model and four substitution

rate categories; tree topology and branch length were optimized using the nearest neighbour interchange, and branch support was obtained with 100 bootstrap replications. The resulting Newick file was visualized and manipulated in ETE2 (Huerta-Cepas *et al.*, 2010).

### **3.4 Results and discussion**

#### **3.4.1 Anaerobic biodegradation of alkanes by SCADC**

In the present study, we observed methanogenic alkane degradation by oil sands microbes in enrichment cultures established using mature fine tailings from the Mildred Lake Settling Basin (Chapter 2). The enrichment cultures had been transferred twice in the laboratory by growth on alkanes in the dark under strictly methanogenic conditions. During incubation of the 75-mL triplicate cultures, detection of volatile hydrocarbons and methane using GC-MS and GC-FID, respectively, showed cumulative methane production with simultaneous depletion of *n*-C<sub>6</sub>–*n*-C<sub>10</sub>, methylcyclopentane and 2-methylpentane, the latter two compounds being C6 isomers present in ‘hexanes’ (Appendix Figure C2). The third isomer, 3-methylpentane, did not degrade and was used as an internal standard. At 108 weeks of incubation, ~1.2 mmol of methane had been produced concomitant with depletion of ~80%–100% of the alkanes with the exception of 3-methylpentane that was undegraded (Chapter 2). In duplicate unamended baseline control cultures, a negligible amount of methane (~1.8 ± 0.1 μmol) was detected after 108 weeks of incubation. Culture fluids were withdrawn periodically, solvent-extracted, derivatized and analyzed using a method similar to that described by Gieg *et al.*, (2010). GC-MS analysis detected putative succinylated products indicative of activation of 2-methylpentane and methylcyclopentane by addition to fumarate, whereas putative succinylated products of the *n*-alkanes were not detected (Chapter 2).

#### **3.4.2 Features of the SCADC enrichment culture and metagenome**

Duplicate 400-mL SCADC enrichments were incubated for 4 months before being sacrificed for genomic DNA extraction and high-throughput sequencing using the 454-FLX platform and paired-end sequencing using the Illumina platform. Pair-wise assembly of quality-controlled reads generated by

the 454 pyrosequencing and Illumina platforms generated 984,086 contigs, with a maximum contig length of 513,327 bp and a minimum contig length of 200 bp. The mean contig assembly length was 657 bp, with a standard deviation of 2,620 bp and N50 of 1,002 bp. The general features of assembled SCADC metagenome and 454 pyrosequencing data are provided in Appendix Table C3 and Figure B3. Gene prediction of the combined Illumina and 454 assembly conducted using the IMG pipeline (Markowitz *et al.*, 2012) produced 1,501,746 predicted protein coding regions, with 774,691 (52%) being assigned an annotation (with product name).

### **3.4.3 Taxonomic classification of PCR-amplified 16S rRNA genes, contigs and unassembled reads**

To investigate the diversity of the microbial community in SCADC, a subsample of the total DNA used for metagenomic sequencing was subjected to 16S rRNA amplicon pyrosequencing using the GS-FLX platform. The resulting sequence reads were subjected to quality control, followed by Operational Taxonomic Unit (OTU) clustering at 5% distance level and taxonomic assignment with the RDP classifier using the Phoenix 2 pipeline, as described by Soh *et al.* (2013). The SCADC community comprised 320 OTUs, of which 167 were singletons. The 15 most abundant OTUs, representing ~80% of the total reads, were related to Methanomicrobiales, Methanosarcinales, Spirochaetales, Syntrophobacterales, Desulfovibrionales, Desulfobacterales, Anaerolineales and Clostridiales (Table 3.1).

In contrast, taxonomic binning of contigs >1000 bp from the SCADC 454 and Illumina hybrid assembly, conducted using PhylopythiaS (Patil *et al.*, 2012), represented only 91,374 or 9.3% of the total hybrid contigs. The difficulty experienced in obtaining a greater percentage of contigs >1000 bp for PhylopythiaS analysis is likely due to the high diversity (or presence of clonal populations) in SCADC and poor assembly reported in Chapter 4. This assumption is based on the detection of a large number of OTUs (153 OTUs, excluding 167 singletons) as 16S rRNA pyrotags, and multiple genotypes of genes associated with hydrocarbon degradation and methanogenesis in the

metagenome, including *assA/bssA*, *bcrA/bamB* and *mcrA*. Further efforts to link the 20 longest contigs in the SCADC metagenome using the “*in silico* gap closure method” (Tang *et al.*, 2012) also were unsuccessful. However, most of the contigs classified by PhylopythiaS (i.e., taxon bins with reads >1%) were assigned to relatives of Clostridiales, Syntrophobacterales, Desulfobacterales, Spirochaetales, Methanosarcinales, and Methanomicrobiales (Appendix Table C4), corresponding to the abundant OTUs listed in Table 3.1.

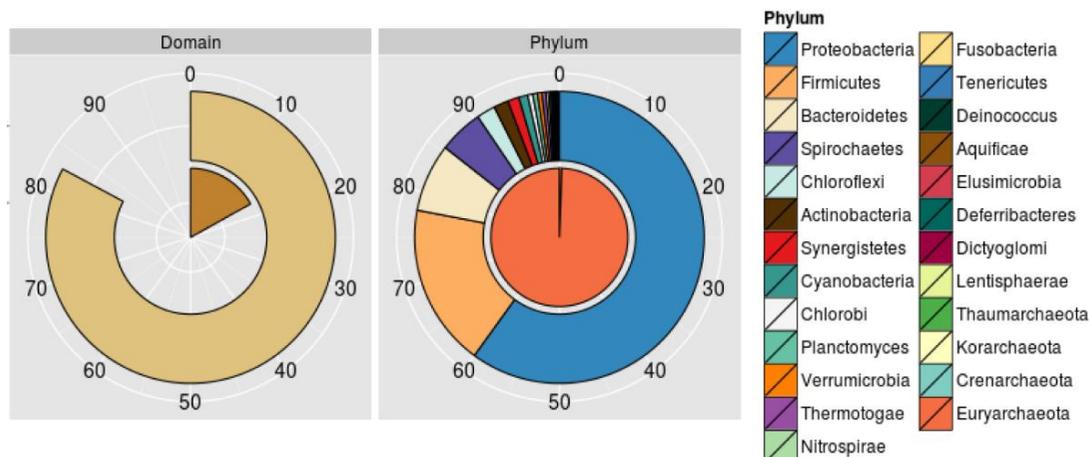
**Table 3.1** Taxonomic classification of the 15 most abundant OTUs in SCADC determined using 16S rRNA gene amplicon pyrosequencing followed by OTU clustering (at 5% distance level) and taxonomic assignment with the RDP classifier.

OTU number	Taxonomic affiliation	Percentage of reads
1	Methanosarcinales	36.3
2	Methanomicrobiales	13.8
3	Spirochaetales	5.0
4	Syntrophobacterales	4.1
5	Methanosarcinales	3.5
6	Methanomicrobiales	3.1
7	Syntrophobacterales	2.4
8	Spirochaetales	2.3
9	Methanomicrobiales	1.6
10	Methanomicrobiales	1.5
11	Desulfovibrionales	1.5
12	Desulfobacterales	1.3
13	Anaerolineales	1.1
14	Clostridiales	1.0
15	Syntrophobacterales	1.0

**Note:** In total, 320 operational taxonomic units (OTUs) were detected, with 167 of these being singletons. The sum of OTUs presented in this table represents ~80% of the total sequence reads.

On a broader scale, the classification of unassembled 454 sequence reads using Sort-ITEMS (Haque *et al.*, 2009) showed that most reads (~85%) were assigned to the domain Bacteria, with a minor proportion (15%) assigned to the Archaea (Figure 3.2). The bacterial taxa appeared to be dominated by Proteobacteria, Firmicutes, Bacteroidetes, Spirochaetes, Chloroflexi, with a host of other microbial phyla each comprising <1% of total reads. Reads assigned to the archaeal domain were almost exclusively related to the Euryarchaeota, with most being assigned to *Methanosarcinales* (30% of archaeal reads) and *Methanomicrobiales* (41% of Archaea).

Community profiling of SCADC revealed using three different methods (16S pyrosequencing, Sort-ITEMS and PhylopythiaS) is largely in agreement regarding the microbial taxa present in SCADC after 4 months of incubation, i.e., during active methanogenesis supported by alkane biodegradation. Discrepancies mainly arise regarding the relative abundance of the microbial taxa. More specifically, greater numbers of 16S rRNA pyrotags were assigned to *Methanosarcinales* and *Methanomicrobiales* (Table 3.1, combined archaeal reads > 60%) versus the greater abundance of bacterial reads determined using Sort-ITEMS analysis of metagenomic sequences (Figure 3.1, combined bacterial reads >70%). Several reasons can explain this observation including the presence of multiple 16S rRNA gene copies in some taxa, the polyploid nature of certain archaeal genomes (Hildenbrand *et al.*, 2011), and/or biases introduced by PCR amplification.



**Figure 3.1** Relative abundance of taxa at the Domain level (Left panel) and Phylum level (Right panel) in unassembled 454 pyrosequencing reads, determined using Sort-ITEMS. Proportions of bacterial taxa are shown on the outer ring and archaeal taxa on the inner ring in both panels. In the Domain-level panel, 82% of the reads are bacterial (tan colour) and 18% are archaeal reads (brown). The legend describes the Phylum-level panel, proceeding clockwise from the top and beginning with the Bacteria; the final four legend items correspond to the inner archaeal circle.

### 3.4.4 Recruitment of SCADC reads to previously sequenced genomes

To further elucidate the relatedness of SCADC community to previously sequenced genomes, nucleotide recruitment to selected reference genomes in the NCBI and IMG databases was conducted using the 454 and Illumina hybrid contig sequences. To retain high stringency and selectivity, the unassembled Illumina and 454 reads were not used in this process. The reference genomes were selected on the basis of the results obtained in Table 3.1 and Appendix Table C4, as well as their relatedness to potential primary degraders (species known to activate hydrocarbons by addition to fumarate); potential syntrophs and homoacetogens; and methanogenic Archaea. In general, the metagenome yielded high nucleotide coverage of the reference genome of *Methanosaeta concilii* GP6 (reference ID: NC\_015416, 86% coverage, >98% nucleotide identity; Table 2), *Syntrophus aciditrophicus* SB1 (NC\_00759, 55% coverage, 98% nucleotide identity) and *Desulfobulbus propionicus* DSM2032 (NC\_014972, 73% coverage,

>99% nucleotide identity), indicating the presence of highly related species within SCADC (Table 3.2). Previous reports have implicated *Syntrophus* spp. as key alkane degraders in complex methanogenic communities (Gray *et al.*, 2011, Zengler *et al.*, 1999) even though the genome sequence of the type strain *S. aciditrophicus* SB1 implies limited fermentative and respiratory functions (McInerney *et al.*, 2007). Although a relatively high proportion of reads were assigned to Syntrophobacterales, of which *Syntrophus* spp. is a member (Tables 3.1, .32 and Appendix Table C4), there is only circumstantial evidence that *Syntrophus* spp. is an important alkane-degrader in the SCADC enrichment culture.

Comparison of predicted ORFs in SCADC to protein-coding genes encoded by relatives of microbial species that mediate hydrocarbon addition to fumarate (i.e., *Desulfotomaculum kuznetsovii*, NC\_015573; *Desulfosporosinus* sp. OT ASM22451v1, AGAF00000000 and *Desulfatibacillum alkenivorans* AK-01, NC\_0117685; Table 3.2), showed relatively low coverage of sequence homologues (<30% coverage). Therefore, the microbial species in SCADC that harbour hydrocarbon activation genes (e.g., *assA*, *bssA* and *nmsA*) are unlikely to be closely related to these species. This assumption is based on the observation that the *assA/bssA/nmsA*-affiliated sequences detected in SCADC were not closely related phylogenetically to any of the primary degraders that have been fully sequenced thus far (Figure 3.2). Undoubtedly, this disconnect is a result of the current paucity of isolates and therefore reference genome sequences from anaerobic hydrocarbon-impacted environments such as tailings ponds.

Many taxa that were detected in SCADC (Table 3.1 and Appendix Table C4) can potentially participate in syntrophic relationships with primary hydrocarbon-degraders and methanogens. Comparisons of predicted ORFs in SCADC to protein-coding genes in *Desulfomicrobium baculatum* DSM4028 (NC\_013173) and *Syntrophorhabdus aromaticivorans* UI (NC\_909663) show moderate to high coverage of homologous genes (>60%), suggesting the presence of these functions in SCADC. Although the SCADC sequence assembly did not map well to the reference genome of *D. baculatum* DSM4028, the presence of

related species in SCADC is supported by our cultivation from SCADC of several isolates closely related to *Desulfomicrobium* spp. at the level of 16S rRNA gene sequence (unpublished results). However, homology searches of the 20 longest contigs in SCADC against the NCBI genome database using Megablast, as performed according to (Brisson *et al.*, 2012), did not return a high proportion of matches (<1% of contigs). Therefore, it is likely that many of the potential syntrophs in SCADC do not have closely related strains that have been sequenced yet.

For the taxonomic assignment of methanogenic Archaea, the *mcrA* gene (Appendix Table C1) has been proposed as an alternative marker to the more widely-used 16S rRNA genes (Luton *et al.*, 2002). The *mcrA* gene encodes the  $\alpha$ -subunit of methyl-CoenzymeM reductase, which is involved in the last step of methanogenesis in all known methanogenic Archaea (Luton *et al.*, 2002). Phylogenetic analysis of the *mcrA* genes recovered from the SCADC metagenome using tBLASTn revealed the presence of diverse phylotypes of methanogens related to *Methanomicrobiales*, *Methanosarcinales*, and *Methanobacteriales* (Appendix Figure C4).

Further comparison of predicted ORFs in the SCADC metagenome to the protein-coding genes in selected methanogens (Table 3.2) shows that the SCADC metagenome harbours a high percentage of homologous genes (>60%) found in *M. concilii* GP6 (NC\_015416), *Methanoculleus marisnigri* JR1 (NC\_009051), *Methanolinea tarda* NOBI-1 (NZ\_AGIIY000000000) and *Methanoregula formicum* SMSP (NC\_019943). The presence of multiple *mcrA* genes is consistent with the detection of multiple OTUs affiliated with Methanosarcinales and Methanomicrobiales. The high percentage of sequence homologues in these methanogens, which utilize different methanogenesis pathways, implies the potential for more than one route of methane production in SCADC (see below).

**Table 3.2** Summary of predicted protein and nucleotide recruitment of 454 and Illumina hybrid contigs to selected reference genomes acquired from NCBI.

Reference genomes	Nucleotide		Protein	
	% geno- me mapp	% ident ical sites	# of homo- logs	% cove- rage <sup>b</sup>
<b>Euryarchaeota</b>				
<i>Methanosaeta concilii</i> GP6 (NC_015416; 3)	85.7	98.3	2728	96
<i>Methanoculleus marisnigri</i> JR1 (NC_009051; 2.4)	1.1	95.9	2095	84
<i>Methanolinea tarda</i> NOBI-1 (NZ_AG1Y000000000;	NR <sup>b</sup>	0	1508	73
<i>Methanoregula boonei</i> 6A8 (NC_009712; 2.5)	NR	0	1230	50
<i>Methanoregula formicum</i> SMSP (NC_019943; 2.8)	NR	0	1817	63
<b>Proteobacteria</b>				
<i>Syntrophus aciditrophicus</i> SB1 (NC_00759; 3.2)	55	98	1781	56
<i>Desulfobulbus propionicus</i> DSM2032 (NC_014972;	73.2	99.6	3066	91
<i>Desulfovibrio aespoensis</i> DSM10631 (NC_014844;	39.3	97.3	1101	33
<i>Desulfatibacillum alkenivorans</i> AK-01 (NC_011768;	NR	0	921	17
<i>Desulfococcus oleovorans</i> Hxd3 (NC_009943; 3.9)	NR	0	823	25
<i>Desulfomicrobium baculatum</i> DSM4028	NR	0	2871	81
<i>Desulfosarcina</i> sp.BuS5 (PRJNA165293; 3.6)	ND <sup>c</sup>	ND	1063	30
<i>Syntrophorhabdus aromaticivorans</i> UI (NC_909663;	ND	ND	2198	61
<b>Firmicutes</b>				
<i>Clostridium</i> sp. 7_3_54FAA (PRJNA40017; 5.4)	NR	0	1168	23
<i>Desulfosporosinus</i> sp. OT	ND	ND	1352	22
<i>Desulfotomaculum kuznetsovii</i> (NC_015573; 3.6)	NR	0	1017	28
<b>Chloroflexi</b>				
<i>Anaerolinea thermophila</i> UN-1 (NC_014960; 3.5)	NR	0	511	16
<b>Spirochaetes</b>				
<i>Sphaerochaeta globus</i> Buddy (NC_015152; 3.3)	0.7	96.1	825	27

<sup>a</sup> defined by Kearsse *et al.* (2012).

<sup>b</sup> defined as the percentage of predicted ORFs in the SCADC metagenome that had BLASTp hits >60% to protein-coding genes in selected reference genomes. The number of protein-coding genes in selected reference genomes was obtained from IMG (Markowitz *et al.*, 2012).

<sup>b</sup>NR, no sequence assembly could be mapped to reference genome

<sup>c</sup>ND, not determined.

### 3.4.5 Presumptive hydrocarbon-activating mechanisms in SCADC

The canonical proposed pathway for anaerobic alkane degradation comprises three key steps: (1) alkane is first activated by the glycyl radical enzyme alkylsuccinate synthase, followed by addition to fumarate; (2) carbon skeleton rearrangement; and (3) decarboxylation yielding a fatty acid for further degradation via beta-oxidation (Figure 1.2; Callaghan *et al.*, 2012). Based on genomic analysis of *D. alkenivorans* AK-01, it has been proposed that carbon-skeleton rearrangement and decarboxylation of the resulting malonate could potentially be catalyzed by methylmalonyl-CoA mutase and carboxyl transferase, respectively (Callaghan *et al.* 2012). However, this has not been confirmed.

Much of the knowledge of the biochemistry and operon structure of alkylsuccinate synthase (ASS) is based on research done with benzylsuccinate synthase (BSS), an homolog that activates toluene, ethylbenzene and xylenes (Callaghan *et al.*, 2008, Callaghan *et al.*, 2012, Grundmann *et al.*, 2008). BSS is a heterotrimer, encoded by *bssABC* (Leuthner *et al.*, 1998). The *bssABC* genes are homologues to the alkylsuccinate synthase-encoding genes *masDEF* and *assABC* found in *Azoarcus* HxN1(Grundmann *et al.*, 2008) and *D. alkenivorans* AK-01, respectively (Callaghan *et al.*, 2012). Activation of 2-methylnaphthalene via addition to fumarate is similarly catalyzed by the related enzyme naphthyl-2-methylsuccinate synthase (NMS) encoded by *nmsABC*, which are homolog of *bssABC* and *assABC* (Slesi *et al.*, 2010). A putative *ass* operon encoding enzymes for *n*-alkane addition to fumarate, carbon-skeleton rearrangement and carboxyl transferase has been reported in the Deltaproteobacterium *D. alkenivorans* AK-01, where the functions of genes related to carbon skeleton rearrangement and carboxyl transferase were inferred based on the annotation of whole genome sequence of the organism (Callaghan *et al.*, 2012). In addition to the *assA* genes that are present in *D. alkenivorans* AK-01 and *Azoarcus* HxN1, full length putative *assA* sequences have also been amplified from the Deltaproteobacterium *Desulfoglaeba alkanexedens* ALDC (ADJ51097) and Betaproteobacterium *Aromatoleum* sp. OcN1 (CBK27727)(Callaghan *et al.*, 2010).

Contigs harbouring homologues of *ass* and *bss* in the SCADC metagenome were identified based on homology searches with tBLASTn using reference amino acid sequences of *bssA* and *assA* (Appendix Table C1). Most sequence homologues recovered were phylogenetically more closely related to the glycol radical enzyme pyruvate formate lyase (PFL) than to *assA/bssA/nmsA*. Sequences that clustered away from *assA/bssA/nmsA* phylogenetic clades were further subjected to BLASTp searches against the NCBI non-redundant database to confirm their identities. All sequences had the lowest E-values and highest bitscore hits (with values varying from sequence to sequence) and therefore the strongest relatedness and similarity to pyruvate-formate lyase, formate C-acetyltransferase and glycerol dehydratase; none was similar to *assA/ bssA/nmsA* gene sequences, confirming closer relatedness to pyruvate-formate lyase gene sequences than alkyl- or benzylsuccinate synthase sequences. Furthermore, these PFL-like sequences were located in contigs that either contained few genes or lacked the full complement genes in the *ass* and *bss* operons (i.e., *assB*, *assC*, *bssB*, *bssC*); therefore these sequences are not likely to be functionally related to authentic genes responsible for alkane addition to fumarate. Notably, all SCADC sequence homologues recovered that are phylogenetically related to *assA*, *bssA* and *nmsA* cluster in groups distinct from the full-length sequences present in cultivated isolates, with a few exceptions (Figure 3.2).

More specifically, the putative *bssA* homologues recovered are closely related to those in the sulfate-reducing bacteria PRTOL (ACI45753) and TRM1 (ABM92939). The putative *bssA*-containing contig (ckmer87\_56747) also contains putative subunits of *bssBC* and *bssD*, suggesting that the organism carrying this contig is capable of making a functionally intact enzyme for the activation of toluene or its analogs through addition to fumarate. In contrast, contig kmer59\_7248277 harbouring putative *nmsA* homologues contains only putative *nmsB* and *nmsD* due to its short length. If the organism that carries this contig is present at low abundance, its genome may have been incompletely sequenced and a complete *nms* might be recovered through more in-depth targeted sequencing, e.g., through single cell sequencing.

Most of the putative *assA* homologues recovered from the SCADC metagenome appear to be located within the same clade as putative *assA* sequences recovered from other environmental sources (Figure 3.2). Notably, the putative *assA* genes located on the two longest contigs carrying *ass* genes (ckmer83\_42244 and contig 21910; Figures 3.2 and 3.3) seem to be distinct from all other *assA* genes, suggesting the presence of different putative primary alkane-degraders in the SCADC enrichment than in other environments. Many recovered putative *assA/bssA/nmsA* genes were located in contigs and singlets shorter than 1000 bp and therefore could not be classified with PhylopythiaS (Appendix Table C5).

**Figure 3.2** (Next page) Maximum likelihood tree of putative *assA*, *nmsA* and *bssA* genes recovered from SCADC metagenome in 454 and Illumina assemblies as contigs. Tree resampling was performed with 100 bootstrap replications. Branches with bootstrap values >60 are represented with black dots. The tree was rooted with pyruvate formate lyase (PFL) in *Clostridium* sp. D5 (ZP\_08129105). Sequences retrieved from the SCAD metagenome are shown in red, reference sequences retrieved from NCBI database are shown in black, and sequences recovered from clone libraries of other alkane-degrading enrichment cultures (BF Tan, unpublished) are shown in blue text on the tree. The approximate length of each contig is shown below the graphic. The long contig in the centre contains reference sequences from *D. alkenivorans* AK-01 with approximate genome location indicated below the graphic. Genes on the reference contig are named and colour-coded for reference to the SCADC contigs. Numbered boxes in light blue are genes not belonging to known *ass*, *nms* or *bss* operons; the numbers are identified in the inset text. Two contigs, ckmer83\_42244 and contig 21910, are shown in greater detail in Figure 3.3. The branch labelled “SCAD metagenome” represents six leaves collapsed into a single branch.



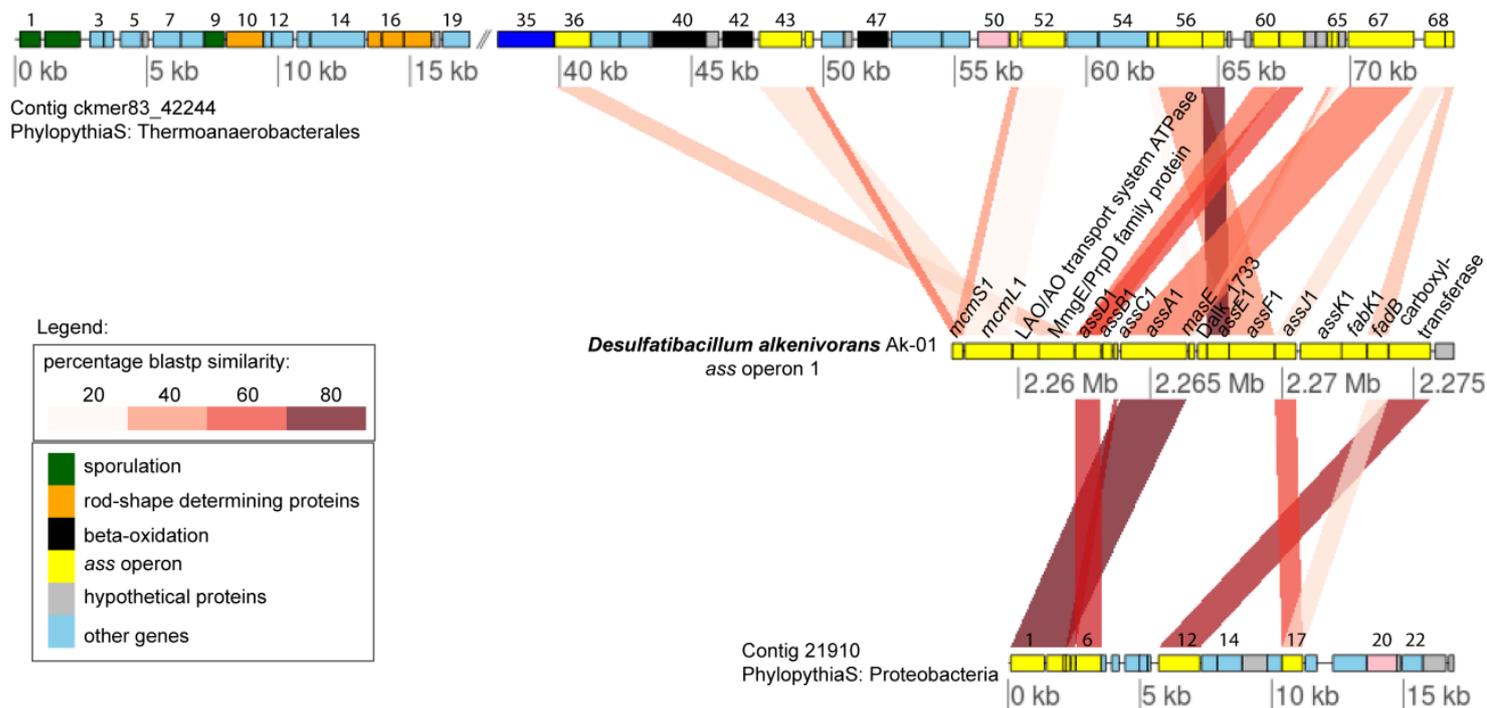
*al.*, 2004, Cravo-Laureau *et al.*, 2005, Davidova *et al.*, 2006, Kniemeyer *et al.*, 2007) and nitrate-reducing bacteria (Ehrenreich *et al.*, 2000). Furthermore, almost all of the pure isolates reported in the literature thus far are related to either Delta- or Betaproteobacteria (Callaghan 2013, Mbadinga *et al.*, 2011). However, the longest SCADC contig harbouring putative *assA* homolog (ckmer83\_42244; 74,073 bp; Figure 3.3) likely is a member of the Firmicutes, based on its classification in the Thermoanaerobacterales using PhylopythiaS (Patil *et al.*, 2012). This conclusion is supported by BLASTp searches of all predicted ORFs on this contig against the IMG database (minimum cut off 60%), where the best matches for 17 ORFs were to clostridial sequences, with only one ORF affiliated with Bacteroidetes and two ORFs with Deltaproteobacteria. The remaining ORFs had no strong affiliation with any taxonomic group. Furthermore, this contig also contains putative genes related to sporulation (stage IV sporulation protein B and FA, and stage 0 sporulation two component response regulator) and rod-shape determining genes (*rodA* and *mreBCD*, Figure 3.3, Appendix Table C6), which are morphological characteristics of many Firmicutes. Multiple genes putatively encoding methylmalonyl-coA mutase (*mcmL* and *mcmS*) and alkylsuccinate synthase glycyl activating enzyme (*assD*) were also present within this contig, although neither the *assBC* homologue nor genes related to carboxyl transfer were detected (Figure 3.3, Appendix Tables C6 and C7). Consistent with the current finding implicating members of the Firmicutes as important alkane-degraders in SCADC, we have previously observed an increase in abundance of microorganisms related to Firmicutes (i.e., Clostridiales) during incubation of an alkane-degrading enrichment culture (Chapter 2). Furthermore, contig ckmer83\_42244 also appears to be better assembled than other *assA*-bearing contigs because this contig is relatively longer than all other *assA*-bearing contigs, indicating that the organism bearing contig ckmer83\_42244 is relatively well represented in the SCADC enrichment culture compared to other organisms that bear *ass* homologues. Nonetheless, the complexity of the SCADC metagenome and lack of reference genomes complicates *in silico* efforts to obtain

the complete genome sequence of the putative alkane-degrader that bears contig ckmer83\_42244.

The second-longest contig harbouring a putative *ass* gene (contig 21910) forms a distinct branch on the phylogenetic tree shown in Figure 3.2. Of the 24 predicted ORFs located on this contig, eight showed best BLASTp hits to deltaproteobacterial sequences, whereas the remaining predicted ORFs affiliated with unassigned taxonomic groups. Further classification using PhylopythiaS related this contig to proteobacteria. Besides the putative *assA* genes, this contig also contains putative *assB*, *assD* and a gene encoding a carboxyl transferase.

Interestingly, a gene homologous to that encoding  $\alpha$ -methylacyl-CoA racemase appears to be located near the *ass* operon genes in both contigs ckmer83\_42244 and contig 21910. The enzyme  $\alpha$ -methylacyl-CoA racemase has been identified as a crucial component in the degradation of *iso*-alkanes by *Mycobacterium* sp. under aerobic conditions (Sakai *et al.*, 2004). More recently Jarling *et al.* (2012) proposed that, prior to carbon-skeleton rearrangement (Callaghan *et al.* 2012), the products of *n*-alkane addition to fumarate, i.e. 1-methylalkylsuccinate isomers, are first epimerized to generate a diastereomer, only after which they undergo carbon-skeleton rearrangement and subsequent decarboxylation. Interestingly, the initial epimerization reaction of 1-methylalkylsuccinate has been postulated to be catalyzed by the enzyme  $\alpha$ -methylacyl-CoA racemase (Jarling *et al.* 2012). Although the genes encoding for  $\alpha$ -methylacyl-CoA racemase appear to be located within the *ass* operon (Figure3.3), any role of this enzyme in the anaerobic degradation of *iso*-alkanes or epimerization reaction of 1-methylalkylsuccinate is currently circumstantial.

Overall, the detection in SCADC of putative genes encoding BssA, NmsA and other subunits of the benzylsuccinate- and 2-methylnaphthylsuccinate synthase enzymes suggest that, despite enrichment on alkanes, the SCADC microbial community is capable of activating toluene and 2-methylnaphthalene by addition to fumarate. Supporting this inference, we have observed that a SCADC inoculum was capable of methane production after transfer into fresh medium containing toluene (0.1% v/v) as sole organic carbon source and



**Figure 3.3** Comparison of the *ass* operon 1 in *D. alkenivorans* AK-01 with two SCADC contigs harbouring *ass* operon analogues. The central graphic represents *ass* operon 1 in *D. alkenivorans* AK-01, named according to Callaghan *et al.* (2012) with the approximate genome location shown below. Each ORF is represented by a coloured box, with the broad functional category (where known) represented in the legend inset. Numbers above boxes in contigs ckmer83\_42244 and 21910 refer to annotations that can be found in Appendix Table C6. Numbers below coloured boxes refer to the length or position within each contig. Comparisons between translated protein sequences were conducted using BLASTp (Appendix Table C7); similarity is indicated by colour intensity (see legend inset).

incubation under methanogenic conditions for 1 year (not shown). The ability of SCADC to degrade 2-methylnaphthalene via methanogenesis has not yet been investigated, so it remains uncertain whether the detected *nms* gene homologues are involved in the activation and degradation of 2-methylnaphthalene.

### **3.4.6 Evidence for genes encoding alternative hydrocarbon-activating pathways**

Other than activation by addition to fumarate, several additional mechanisms have been proposed for the activation of *n*-alkanes and monoaromatic compounds (BTEX; benzene, toluene, ethylbenzene and xylenes) under different electron-accepting conditions. The putative genes involved in these novel mechanisms (Appendix Table C1) include: anaerobic benzene carboxylase (encoded by *abcAD*) proposed for the carboxylation of benzene to yield benzoate (Abu Laban *et al.*, 2010); ethylbenzene dehydrogenase (*ebdABC* genes) that has been experimentally proven to be involved in hydroxylation of ethylbenzene (Kniemeyer and Heider 2001); and various types of monooxygenases (Appendix Table C1) proposed to be involved in the oxidation of alkane/methane coupled to the reduction of nitrogen species (Ettwig *et al.*, 2010, Zedelius *et al.*, 2011).

Expression of a putative anaerobic benzene carboxylase (ABC) was up-regulated in an iron-reducing benzene-degrading enrichment culture dominated by Peptococcaceae (Abu Laban *et al.*, 2010). The enzyme ABC is hypothesized to be encoded by *abcAD* genes, where *abcA* shares high sequence homology with phenylphosphate carboxylase that is involved in phenol metabolism in many well-known hydrocarbon-degraders including *Aromatoleum aromaticum* EbN1, *Thauera aromatica* and *Azoarcus* spp. (Carmona *et al.*, 2009). tBLASTn homology searches of the SCADC 454 and Illumina contigs against reference sequence of *abcA* and *abcD* (Abu Laban *et al.*, 2010) detected multiple *abcA* sequence homologues, whereas no *abcD* homologue was detected. Notably, the *abcA* homologues in SCADC were exclusively annotated by the IMG annotation pipeline as 3-polyprenyl-4-hydroxybenzoate decarboxylase and other related decarboxylases. The enzyme 3-polyprenyl-4-hydroxybenzoate decarboxylase and

its homologs are involved in chorismate metabolism or ubiquinol-8 biosynthesis (Caspi *et al.*, 2012). No other hydrocarbon-degrading genes (e.g., benzoyl-coA reductase or benzoyl-coA ligase) could be identified within these contigs; therefore the *abcA* homologues detected in SCADC are likely functionally distinct from the putative *abcA* gene described for enrichment culture BF (Abu Laban *et al.*, 2010).

Ethylbenzene dehydrogenase (EBD) is a heterotrimer involved in hydroxylation of ethylbenzene (Carmona *et al.*, 2009). The enzyme is encoded by *ebdABC* genes and was previously found in the denitrifier *A. aromaticum* EbN1 (Carmona *et al.*, 2009, Kniemeyer and Heider 2001). Because the enzyme shares high sequence homology to other dehydrogenases such as DMSO and nitrate reductases (Carmona *et al.*, 2009), sequences in the SCADC sequence assembly that were annotated as putative ethylbenzene dehydrogenase genes were further examined for downstream and upstream genes. Based on the annotation in IMG and further examination through BlastX searches for genes located near the putative *ebdA* genes, there is no evidence to suggest that SCADC encodes any of the other subunits of ethylbenzene dehydrogenase (i.e., *ebdB* and *ebdC*) and therefore it is unlikely that SCADC community members are capable of producing functional ethylbenzene dehydrogenase. Further experimental work is required to test whether SCADC has the ability to utilize ethylbenzene as a carbon source.

Anaerobic alkane degradation by a Gammaproteobacterium has been recently reported under denitrifying conditions (Zedelius *et al.*, 2011). The degradation pathway was proposed to be initiated by at least two different types of monooxygenase (Appendix Table C1) using oxygen species derived from partial reduction and dismutation of nitrogen species (Ettwig *et al.*, 2010). Homology searches using the reference sequence of monooxygenase in the Gammaproteobacterium HdN1 (Appendix Table C1; Zedelius *et al.*, 2011), which has been suggested to mediate such mechanism, did not detect any sequence homologues in the SCADC metagenome.

Finally, alkane activation has been proposed to proceed via carboxylation in *Desulfococcus oleovorans* Hxd3 (So *et al.*, 2003): the requisite genes and enzymes have not yet been reported, precluding searches of the SCADC metagenome. More recently, the carboxylation mechanism hypothesized to be catalyzed by *D. oleovorans* Hxd3 has now been proposed as a degradation mechanism for a ketone generated by the activation of alkanes via hydroxylation (Callaghan 2013). The whole genome sequence of *D. oleovorans* Hxd3 shows that this organism contains an ethylbenzene dehydrogenase-like complex, postulated to be involved in the anaerobic hydroxylation of alkanes, based on preliminary proteomic experiments (Callaghan 2013). tBLASTn searches of the SCADC metagenome using translated sequence of the putative ethylbenzene dehydrogenase-like complex in *D. oleovorans* Hxd3 (Dole\_0194) detected multiple copies of homologous genes in the SCADC metagenome. The sequence with the highest pairwise similarity (45%) had only 13% query coverage and an E-value of 1.5E-27, whereas the sequence with the highest E-value (3.6E-95) had 28% pairwise similarity and nearly complete sequence coverage compared to the translated sequence of Dole\_0914. A BLASTp search of this sequence against the NCBI non-redundant database shows that it has an E-value of 0 and 38 – 75% pairwise similarity to multiple sequences of DMSO dehydrogenase and nitrate reductase found in other bacterial species (e.g., *Sulfuricurvum kujiense* DSM16994 and *Thiovulum* sp. ES). The function of the Dole\_0194 sequence homologues in SCADC, nevertheless, remains elusive and requires further investigation such as cell isolation and physiological testing to confirm their activities.

Taken together, there is no concrete evidence to show that the SCADC enrichment culture harbours any of the known genes encoding alternative hydrocarbon-activating mechanisms, and evidence only for *ass*-like genes. By default, then, we are able to propose only that the alkane degradation mechanism in SCADC is through addition to fumarate. However, further experimental research and additional reference sequences are needed to further explore the full extent of alkane activation mechanisms under methanogenic conditions.

### 3.4.7 Roles of community members: general overview

Because SCADC is a “closed” system without gas exchange and with finite carbon source (0.1% v/v alkanes) and nutrients, the efficient coordination of substrate utilization among different competing and cooperating microbial groups becomes important to sustain methane production. In other model systems, metagenomic analyses have revealed the importance of metabolic interactions between community members for mutual growth and survival (Brisson *et al.*, 2012, Hug *et al.*, 2012, Lykidis *et al.*, 2011). Such interactions undoubtedly occur among the >300 OTUs detected in SCADC and potential examples of resource sharing are evident from metagenomic analysis, as described below.

#### 3.4.7.1 Carbon utilization pathways in SCADC

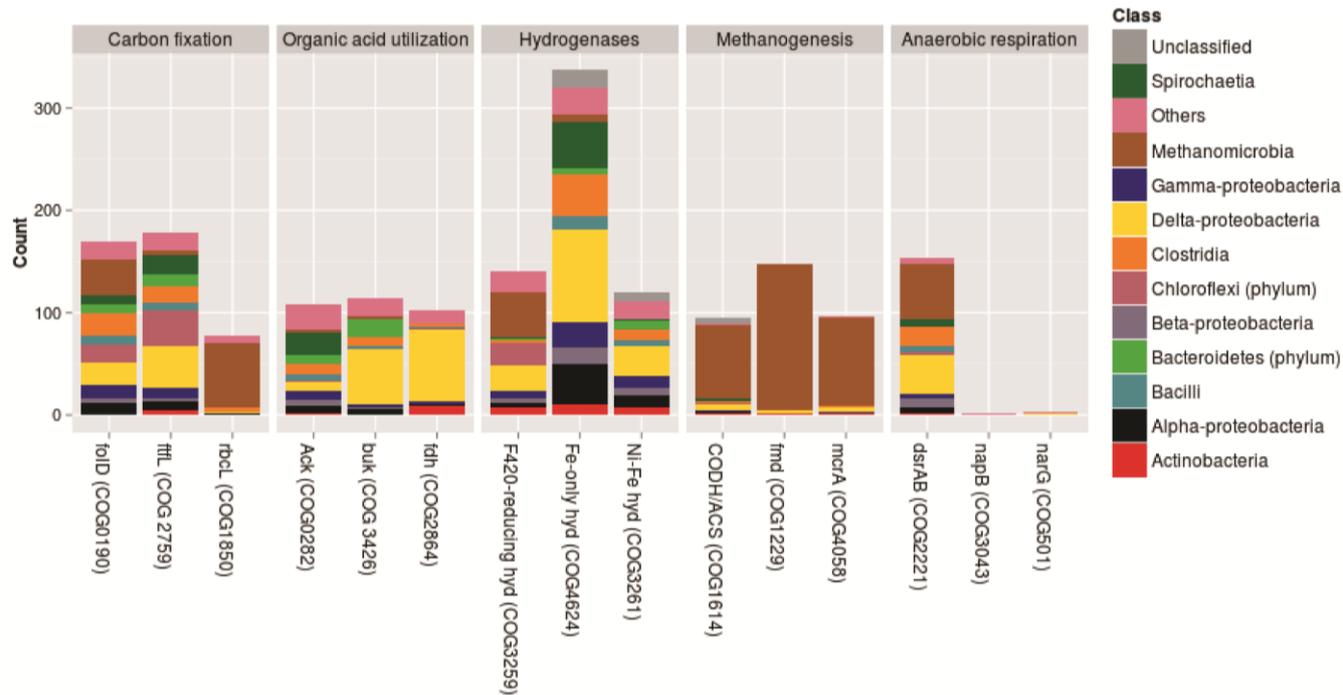
Bacterial species that utilize hydrocarbons under fermentative conditions produce volatile fatty acids such as acetate and formate and CO<sub>2</sub> and H<sub>2</sub> as end products (Callaghan *et al.*, 2012, Edwards and Grbić-Galić 1994, Mbadanga *et al.*, 2012). The CO<sub>2</sub> and H<sub>2</sub> contained in the culture bottle headspace could be utilized by hydrogenotrophic methanogens to yield methane and/or by homoacetogens to synthesize biomass, for example. To explore potential alternate strategies for autotrophic carbon fixation (Berg *et al.*, 2010), pathway reconstruction was conducted in the KEGG module mapper using KEGG orthologues (KOs; Appendix Figure C7; (Kanehisa *et al.*, 2012)). In general, the SCADC metagenome appears to contain partial to complete pathways for several carbon-fixation pathways, including the Wood-Ljungdahl (reductive acetyl-coA) pathway and Calvin (RubisCO), dicarboxylate-hydroxybutyrate, 3-hydroxypropionate, hydroxypropionate-hydroxybutyrate and Arnon-Buchanan (reductive citric acid) cycles (Appendix Figure C7).

Genome-enabled metabolic pathway re-construction revealed the presence of Wood-Ljungdahl pathway genes in many anaerobes including *D. alkenivorans* AK-01, *D. oleovorans* Hxd3, *Desulfobacula toluolica* Tol2 and others that are known to degrade hydrocarbons under different anaerobic conditions, as revealed by pathways searches in Biocyc v 17.0 (Caspi *et al.*, 2012). Selected COGs

annotated with functions of interest were further subjected to taxonomic classification using RITA (MacDonald *et al.*, 2012). For example, homologues of *folD* and *ftfL* components in the Wood-Ljungdahl pathway detected in SCADC were assigned to a repertoire of taxonomic groups (Figure 3.4), many of which were detected in high abundance in the SCADC metagenome (Tables 3.1, 3.2 and C4). Members of these taxa include *M. marisnigri* JR1, *S. aciditrophicus* SB1, *Anaerolinea thermophila* UNI-1, *Desulfovibrio aespoeensis* Aspo-2, *D. baculatum* DSM 4028 and *D. kuznetsovii* DSM 6115 are capable of CO<sub>2</sub> fixation through the Wood-Ljungdahl pathway, as revealed by pathways searches in Biocyc v 17.0 (Caspi *et al.*, 2012). Other microorganisms such as *Spirochaeta thermophila* DSM 6192 and *D. propionicus* DSM2032, and possibly the closely related taxa in SCADC, can fix CO<sub>2</sub> through the reductive TCA cycle (Caspi *et al.*, 2012). Interestingly, putative RubisCO homologues of the Calvin-Benson-Bassham cycle detected in the SCADC metagenome appear to be related most frequently to *Methanomicrobia* (Figure 3.4). RubisCO in methanogenic Archaea has been previously suggested to be involved in carbon fixation (Finn and Tabita 2003) and AMP metabolism (Sato *et al.*, 2007). The ecological role of this enzyme in SCADC and oil sands tailing ponds community is currently unexplored.

Finally, organic acids such as acetate produced as metabolic by-products of incomplete beta-oxidation of hydrocarbon substrates, or produced by homoacetogens from H<sub>2</sub> and CO<sub>2</sub> are other potential carbon sources to sustain growth and survival of syntrophic community members. Organic acids such as acetate and formate have been detected previously in methanogenic hydrocarbon-degrading enrichment cultures (Edwards and Grbić-Galić 1994, Mbadinga *et al.*, 2012). *D. propionicus* and *D. baculatum*, which are closely related to sequences detected in SCADC, can grow on H<sub>2</sub> in the presence of CO<sub>2</sub> with acetate as the carbon source (Widdel and Pfennig 1982). A common way of utilizing volatile fatty acids like acetate and butyrate requires kinases that convert, for example, acetate to acetyl-coA, which then enters various central metabolic pathways. A putative *ack* gene encoding acetate kinase was detected in SCADC and appears to

be assigned to diverse microbial species (Figure 3.4). Although there is no evidence to suggest that butyrate is a key metabolite in SCADC, organisms such



**Figure 3.4** Taxonomic distribution of selected COG categories representing broad metabolic functions pertaining to the SCADC enrichment culture. Annotation was conducted using the COG profilers in IMG. Metagenomic reads were classified using RITA with the parameters listed in Materials and Methods. Abbreviations: dsrAB, dissimilatory sulfite reductase  $\alpha$ - and  $\beta$ -subunits; napB, periplasmic nitrate reductase; narG, nitrate reductase; fold, methenyltetrahydrofolate cyclohydrolase; ftfL, formyltetrahydrofolate synthetase; rbcL, ribulose biphosphate carboxylase; hyd, hydrogenase; CODH/ACS, carbon monoxide dehydrogenase/acetyl-coA synthase; fmd, formylmethanofuran dehydrogenase; mcrA, methyl coenzyme M reductase  $\alpha$ -subunit; ack, acetate kinase; buk, butyrate kinase; fdh, formate dehydrogenase.

as Clostridiales bacterium 1\_7\_47FAA and *D. aespoensis* Aspo-2 can utilize butyrate as carbon source in addition to acetate (Caspi *et al.*, 2012). For many syntrophs such as *S. aciditrophicus*, the ability to efficiently compete for the utilization of limited amounts of organic carbon in a system like SCADC becomes important to sustain its growth and persistence.

### **3.4.7.2 Hydrogenases**

Thermodynamically, increased H<sub>2</sub> partial pressure has been identified as a key factor inhibiting methanogenesis and substrate degradation (Ahring *et al.*, 1991, Schmidt and Ahring 1993). Thus, it is crucial for SCADC community members to maintain low H<sub>2</sub> partial pressure to sustain carbon flux towards methane production.

Syntrophic microbial species can scavenge hydrogen for production and conservation of energy (Vignais and Billoud 2007). Hydrogenases involved in the reversible oxidation of H<sub>2</sub> appear to be quite common in many anaerobic systems (Brisson *et al.*, 2012, Hug *et al.*, 2012, Lykidis *et al.*, 2011) including the SCADC enrichment culture (Figure 3.4). Most of the 454 sequences classified as Fe- and Fe-Ni hydrogenases in SCADC were assigned to Deltaproteobacteria, Clostridia, Spirochaetes and others (Figure 3.4). This is consistent with members of these phyla being present in high abundance in SCADC (Tables 3.1 and 3.2) and the observed diversity of hydrogenases present in anaerobes across different phyla (Vignais and Billoud 2007). In contrast, co-enzyme F420-dependent hydrogenase sequences were mostly affiliated with *Methanomicrobia*, consistent with this enzyme being required for hydrogenotrophic methanogenesis (Deppenmeier *et al.*, 1992).

### **3.4.7.3 Methanogenesis and other anaerobic respiration processes**

The SCADC enrichment culture harbours two dominant orders of methanogens: Methanosarcinales and Methanomicrobiales (Tables 3.1, 3.2 and C4), members of which are commonly detected in methanogenic hydrocarbon-degrading cultures (Callaghan *et al.*, 2010, Siddique *et al.*, 2011, Siddique *et al.*, 2012) and oil sands tailings (Penner and Foght 2010, Ramos-Padron *et al.*, 2011). Thus, both acetoclastic and hydrogenotrophic methanogenesis pathways were

expected in SCADC and, indeed, complete pathways for both were detected (Appendix Figure C7). SCADC also had high coverage (96% and 84%, respectively; Table 3.2) of homologous genes in *M. concilii* GP6 (an acetoclastic methanogen) and *M. marisnigri* JR1 (a hydrogenotrophic methanogen).

Although metabolism in SCADC is predominantly methanogenic and/or fermentative because of the lack of inorganic electron acceptors other than CO<sub>2</sub>, several highly abundant taxa capable of dissimilatory sulfate reduction were detected in SCADC, including members of the Deltaproteobacteria and Clostridia (Tables 3.1, 3.2 and C4). Supporting this evidence, sequences of *dsrAB* genes encoding dissimilatory sulphite reductase  $\alpha$ - and  $\beta$ -subunits were detected in the SCADC metagenome, associated with Deltaproteobacteria and Clostridia (Figure 3.4). This suggests that SCADC has the potential to shift to sulfate or sulfite reduction, although some *dsrAB*-related sequences in SCADC were affiliated with Methanomicrobia. The latter finding is consistent with recent evidence for Dsr-like proteins in a wide variety of methanogens (Susanti and Mukhopadhyay 2012).

### **3.5 Significance**

The establishment of the SCADC culture enriched from oil sands tailings has enabled study of putative hydrocarbon activation and degradation processes. More importantly, metagenomic analysis of SCADC facilitated the prediction of the metabolic potential of this microbial community. Knowing the genetic capabilities and operon structure, particularly for genes with putative functions in SCADC metagenomes, provides a basis for future strain isolation, metatranscriptomics and proteomics work to verify the functions of genes of interest. Revealing the diversity of microbes and estimating the repertoire of their metabolic capabilities provides information relevant for prediction of greenhouse gas emissions from tailings ponds and for managing tailings reclamation. On a fundamental basis it highlights important areas for further research into methanogenic hydrocarbon degradation processes that are widespread in contaminated environments, and for biotechnological processes such as methanization of exhausted petroleum reservoirs.

### 3.6 References

- Abu Laban N, Selesi D, Rattei T, Tischler P, Meckenstock RU. (2010). Identification of enzymes involved in anaerobic benzene degradation by a strictly anaerobic iron-reducing enrichment culture. *Environ Microbiol* **12**: 2783-2796.
- Agrawal A, Gieg LM. (2013). In situ detection of anaerobic alkane metabolites in subsurface environments. *Front Microbiol* **4**: 140.
- Ahring BK, Westermann P, Mah RA. (1991). Hydrogen inhibition of acetate metabolism and kinetics of hydrogen consumption by *Methanosarcina thermophila* TM-1. *Arch Microbiol* **157**: 38-42.
- Aitken CM, Jones DM, Maguire MJ, Gray ND, Sherry A, Bowler BFJ *et al.*,. (2013). Evidence that crude oil alkane activation proceeds by different mechanisms under sulfate-reducing and methanogenic conditions. *Geochim Cosmochim Acta* **109**: 162-174.
- Beller HR, Spormann AM. (1997). Anaerobic activation of toluene and *o*-xylene by addition to fumarate in denitrifying strain T. *J Bacteriol* **179**: 670-676.
- Beller HR, Spormann AM. (1998). Analysis of the novel benzylsuccinate synthase reaction for anaerobic toluene activation based on structural studies of the product. *J Bacteriol* **180**: 5454-5457.
- Beller HR, Spormann AM. (1999). Substrate range of benzylsuccinate synthase from *Azoarcus* sp strain T. *FEMS Microbiol Lett* **178**: 147-153.
- Beller HR. (2000). Metabolic indicators for detecting in situ anaerobic alkylbenzene degradation. *Biodegradation* **11**: 125-139.
- Beller HR, Edwards EA. (2000). Anaerobic toluene activation by benzylsuccinate synthase in a highly enriched methanogenic culture. *Appl Environ Microbiol* **66**: 5503-5505.
- Beller HR, Kane SR, Legler TC, McKelvie JR, Lollar BS, Pearson F *et al.*,. (2008). Comparative assessments of benzene, toluene, and xylene natural attenuation by quantitative polymerase chain reaction analysis of a catabolic gene, signature metabolites, and compound-specific isotope analysis. *Environ Sci Technol* **42**: 6065-6072.
- Berdugo-Clavijo C, Dong XL, Soh J, Sensen CW, Gieg LM. (2012). Methanogenic biodegradation of two-ringed polycyclic aromatic hydrocarbons. *FEMS Microbiol Ecol* **81**: 124-133.

Berg IA, Kockelkorn D, Ramos-Vera WH, Say RF, Zarzycki J, Hugler M *et al.*, (2010). Autotrophic carbon fixation in archaea. *Nature Reviews Microbiology* **8**: 447-460.

Biegert T, Fuchs G, Heider F. (1996). Evidence that anaerobic oxidation of toluene in the denitrifying bacterium *Thauera aromatica* is initiated by formation of benzylsuccinate from toluene and fumarate. *Eur J Biochem* **238**: 661-668.

Bombach P, Chatzinotas A, Neu TR, Kastner M, Lueders T, Vogt C. (2010). Enrichment and characterization of a sulfate-reducing toluene-degrading microbial consortium by combining in situ microcosms and stable isotope probing techniques. *FEMS Microbiol Ecol* **71**: 237-246.

Brisson VL, West KA, Lee PKH, Tringe SG, Brodie EL, Alvarez-Cohen L. (2012). Metagenomic analysis of a stable trichloroethene-degrading microbial community. *Isme Journal* **6**: 1702-1714.

Callaghan AV, Wawrik B, Chadhain SMN, Young LY, Zylstra GJ. (2008). Anaerobic alkane-degrading strain AK-01 contains two alkylsuccinate synthase genes. *Biochem Biophys Res Commun* **366**: 142-148.

Callaghan AV, Davidova IA, Savage-Ashlock K, Parisi VA, Gieg LM, Suflita JM *et al.*, (2010). Diversity of benzyl- and alkylsuccinate synthase genes in hydrocarbon-impacted environments and enrichment cultures. *Environ Sci Technol* **44**: 7287-7294.

Callaghan AV, Morris BEL, Pereira IAC, McInerney MJ, Austin RN, Groves JT *et al.*, (2012). The genome sequence of *Desulfatibacillum alkenivorans* AK-01: a blueprint for anaerobic alkane oxidation. *Environ Microbiol* **14**: 101-113.

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

Carmona M, Zamarro MT, Blazquez B, Durante-Rodriguez G, Juarez JF, Valderrama JA *et al.*, (2009). Anaerobic catabolism of aromatic compounds: a genetic and genomic view. *Microbiol Mol Biol Rev* **73**: 71-+.

Caspi R, Altman T, Dreher K, Fulcher CA, Subhraveti P, Keseler IM *et al.*, (2012). The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. *Nucleic Acids Res* **40**: D742-D753.

Chalaturnyk RJ, Scott JD, Ozum B. (2002). Management of oil sands tailings. *Petroleum Science and Technology* **20**: 1025-1046.

Cline JD. (1969). Spectrophotometric determination of hydrogen sulfide in natural waters. *Limnol Oceanogr* **14**: 454-&.

Cravo-Laureau C, Matheron R, Joulian C, Cayol JL, Hirschler-Rea A. (2004). *Desulfatibacillum alkenivorans* sp nov., a novel *n*-alkene-degrading, sulfate-reducing bacterium, and emended description of the genus *Desulfatibacillum*. *Int J Syst Evol Microbiol* **54**: 1639-1642.

Cravo-Laureau C, Grossi V, Raphael D, Matheron R, Hirschler-Rea A. (2005). Anaerobic *n*-alkane metabolism by a sulfate-reducing bacterium, *Desulfatibacillum aliphaticivorans* strain CV2803. *Appl Environ Microbiol* **71**: 3458-3467.

Davidova IA, Duncan KE, Choi OK, Suflita JM. (2006). *Desulfoglaeba alkanexedens* gen. nov., sp nov., an *n*-alkane-degrading, sulfate-reducing bacterium. *Int J Syst Evol Microbiol* **56**: 2737-2742.

Deppenmeier U, Blaut M, Schmidt B, Gottschalk G. (1992). Purification and properties of a F420-non reactive, membrane-bound hydrogenase from *Methanosarcina* Strain G01. *Arch Microbiol* **157**: 505-511.

Duncan KE, Gieg LM, Parisi VA, Tanner RS, Tringe SG, Bristow J *et al.*, (2009). Biocorrosive thermophilic microbial communities in Alaskan North Slope oil facilities. *Environ Sci Technol* **43**: 7977-7984.

Dwyer DF, Weegaerssens E, Shelton DR, Tiedje JM. (1988). Bioenergetic conditions of butyrate metabolism by a syntrophic, anaerobic bacterium in coculture with hydrogen-oxidizing methanogenic and sulfidogenic bacteria. *Appl Environ Microbiol* **54**: 1354-1359.

Edgar RC. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32**: 1792-1797.

Edwards EA, Grbić-Galić D. (1994). Anaerobic degradation of toluene and *o*-xylene by a methanogenic consortium. *Appl Environ Microbiol* **60**: 313-322.

Ehrenreich P, Behrends A, Harder J, Widdel F. (2000). Anaerobic oxidation of alkanes by newly isolated denitrifying bacteria. *Arch Microbiol* **173**: 58-64.

Ettwig KF, Butler MK, Le Paslier D, Pelletier E, Mangenot S, Kuypers MMM *et al.*, (2010). Nitrite-driven anaerobic methane oxidation by oxygenic bacteria. *Nature* **464**: 543.

Fedorak PM, Hrudehy SE. (1984). The effects of phenol and some alkyl phenolics on batch anaerobic methanogenesis. *Water Res* **18**: 361-367.

Fedorak PM, Coy DL, Salloum MJ, Dudas MJ. (2002). Methanogenic potential of tailings samples from oil sands extraction plants. *Can J Microbiol* **48**: 21-33.

Fedorak PM, Coy DL, Dudas MJ, Simpson MJ, Renneberg AJ, MacKinnon MD. (2003). Microbially-mediated fugitive gas production from oil sands tailings and increased tailings densification rates. *J Environ Eng Sci* **2**: 199-211.

Feisthauer S, Siegert M, Seidel M, Richnow HH, Zengler K, Grundger F *et al.*,. (2010). Isotopic fingerprinting of methane and CO<sub>2</sub> formation from aliphatic and aromatic hydrocarbons. *Org Geochem* **41**: 482-490.

Finn MW, Tabita FR. (2003). Synthesis of catalytically active form III ribulose 1,5-bisphosphate carboxylase/oxygenase in archaea. *J Bacteriol* **185**: 3049-3059.

Foght J, Aislabie J, Turner S, Brown CE, Ryburn J, Saul DJ *et al.*,. (2004). Culturable bacteria in subglacial sediments and ice from two Southern Hemisphere glaciers. *Microb Ecol* **47**: 329-340.

Foght J. (2008). Anaerobic biodegradation of aromatic hydrocarbons: Pathways and prospects. *J Mol Microbiol Biotechnol* **15**: 93-120.

Fowler SJ, Dong XL, Sensen CW, Suflita JM, Gieg LM. (2012). Methanogenic toluene metabolism: community structure and intermediates. *Environ Microbiol* **14**: 754-764.

Gertz EM, Yu YK, Agarwala R, Schaffer AA, Altschul SF. (2006). Composition-based statistics and translated nucleotide searches: Improving the TBLASTN module of BLAST. *BMC Biol* **4**.

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

Gieg LM, Duncan KE, Suflita JM. (2008). Bioenergy production via microbial conversion of residual oil to natural gas. *Appl Environ Microbiol* **74**: 3022-3029.

Gieg LM, Davidova IA, Duncan KE, Suflita JM. (2010). Methanogenesis, sulfate reduction and crude oil biodegradation in hot Alaskan oilfields. *Environ Microbiol* **12**: 3074-3086.

Gray ND, Sherry A, Hubert C, Dolfing J, Headt IM. (2010). Methanogenic degradation of petroleum hydrocarbons in subsurface environments: remediation, heavy oil formation, and energy recovery. *Adv Appl Microbiol* **72**: 137-161.

Gray ND, Sherry A, Grant RJ, Rowan AK, Hubert CRJ, Callbeck CM *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 *et al.*, (2008). Genes encoding the candidate enzyme for anaerobic activation of *n*-alkanes in the denitrifying bacterium, strain HxN1. *Environ Microbiol* **10**: 376-385.
- Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. *Syst Biol* **59**: 307-321.
- Haque MM, Ghosh TS, Komanduri D, Mande SS. (2009). SOrt-ITEMS: Sequence orthology based approach for improved taxonomic estimation of metagenomic sequences. *Bioinformatics* **25**: 1722-1730.
- Head IM, Jones DM, Larter SR. (2003). Biological activity in the deep subsurface and the origin of heavy oil. *Nature* **426**: 344-352.
- Hildenbrand C, Stock T, Lange C, Rother M, Soppa J. (2011). Genome copy numbers and gene conversion in methanogenic archaea. *J Bacteriol* **193**: 734-743.
- Holowenko FM, MacKinnon MD, Fedorak PM. (2000). Methanogens and sulfate-reducing bacteria in oil sands fine tailings waste. *Can J Microbiol* **46**: 927-937.
- Hopkins BT, Mcinerney MJ, Warikoo V. (1995). Evidence for an anaerobic syntrophic benzoate degradation threshold and isolation of the syntrophic benzoate degrader. *Appl Environ Microbiol* **61**: 526-530.
- Huerta-Cepas J, Dopazo J, Gabaldon T. (2010). ETE: a python Environment for Tree Exploration. *BMC Bioinformatics* **11**.
- Hug LA, Beiko RG, Rowe AR, Richardson RE, Edwards EA. (2012). Comparative metagenomics of three Dehalococcoides-containing enrichment cultures: the role of the non-dechlorinating community. *BMC Genomics* **13**.
- Jackson BE, Bhupathiraju VK, Tanner RS, Woese CR, McInerney MJ. (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.
- Jehmlich N, Kleinstüber S, Vogt C, Benndorf D, Harms H, Schmidt F *et al.*, (2010). Phylogenetic and proteomic analysis of an anaerobic toluene-degrading community. *J Appl Microbiol* **109**: 1937-1945.
- Jones DM, Head IM, Gray ND, Adams JJ, Rowan AK, Aitken CM *et al.*, (2008). Crude-oil biodegradation via methanogenesis in subsurface petroleum reservoirs. *Nature* **451**: 176-U176.

- Kanehisa M, Goto S, Sato Y, Furumichi M, Tanabe M. (2012). KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Res* **40**: D109-D114.
- Kimes NE, Callaghan AV, Aktas DF, Smith WL, Sunner J, Golding B *et al.*, (2013). Metagenomic analysis and metabolite profiling of deep-sea sediments from the Gulf of Mexico following the Deepwater Horizon oil spill. *Front Microbiol* **4**: 50.
- Kniemeyer O, Heider J. (2001). Ethylbenzene dehydrogenase, a novel hydrocarbon-oxidizing molybdenum/iron-sulfur/heme enzyme. *J Biol Chem* **276**: 21381-21386.
- Kniemeyer O, Musat F, Sievert SM, Knittel K, Wilkes H, Blumenberg M *et al.*, (2007). Anaerobic oxidation of short-chain hydrocarbons by marine sulphate-reducing bacteria. *Nature* **449**: 898-U810.
- Krieger CJ, Beller HR, Reinhard M, Spormann AM. (1999). Initial reactions in anaerobic oxidation of *m*-xylene by the denitrifying bacterium *Azoarcus* sp strain T. *J Bacteriol* **181**: 6403-6410.
- Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D *et al.*, (2009). Circos: An information aesthetic for comparative genomics. *Genome Res* **19**: 1639-1645.
- Kunapuli U, Griebler C, Beller HR, Meckenstock RU. (2008). Identification of intermediates formed during anaerobic benzene degradation by an iron-reducing enrichment culture. *Environ Microbiol* **10**: 1703-1712.
- Kunapuli U, Jahn MK, Lueders T, Geyer R, Heipieper HJ, Meckenstock RU. (2010). *Desulfitobacterium aromaticivorans* sp nov and *Geobacter toluenoxydans* sp nov., iron-reducing bacteria capable of anaerobic degradation of monoaromatic hydrocarbons. *Int J Syst Evol Microbiol* **60**: 686-695.
- Leuthner B, Leutwein C, Schulz H, Horth P, Haehnel W, Schiltz E *et al.*, (1998). Biochemical and genetic characterization of benzylsuccinate synthase from *Thauera aromatica*: a new glycyl radical enzyme catalysing the first step in anaerobic toluene metabolism. *Mol Microbiol* **28**: 615-628.
- Li W, Godzik A. (2006). Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **22**: 1658-1659.
- Liu A, Garcia-Dominguez E, Rhine ED, Young LY. (2004). A novel arsenate respiring isolate that can utilize aromatic substrates. *FEMS Microbiol Ecol* **48**: 323-332.

- Ludwig W, Strunk O, Westram R, Richter L, Meier H, Yadhukumar *et al.*, (2004). ARB: a software environment for sequence data. *Nucleic Acids Res* **32**: 1363-1371.
- Luton PE, Wayne JM, Sharp RJ, Riley PW. (2002). The *mcrA* gene as an alternative to 16S rRNA in the phylogenetic analysis of methanogen populations in landfill. *Microbiology-Sgm* **148**: 3521-3530.
- Lykidis A, Chen CL, Tringe SG, McHardy AC, Copeland A, Kyrpides NC *et al.*, (2011). Multiple syntrophic interactions in a terephthalate-degrading methanogenic consortium. *Isme Journal* **5**: 122-130.
- MacDonald NJ, Parks DH, Beiko RG. (2012). Rapid identification of high-confidence taxonomic assignments for metagenomic data. *Nucleic Acids Res* **40**.
- Markowitz VM, Chen IMA, Chu K, Szeto E, Paliappan K, Grechkin Y *et al.*, (2012). IMG/M: the integrated metagenome data management and comparative analysis system. *Nucleic Acids Res* **40**: D123-D129.
- Matsuki T, Watanabe K, Fujimoto J, Miyamoto Y, Takada T, Matsumoto K *et al.*, (2002). Development of 16S rRNA-gene-targeted group-specific primers for the detection and identification of predominant bacteria in human feces. *Appl Environ Microbiol* **68**: 5445-5451.
- Mbadinga SM, Wang LY, Zhou L, Liu JF, Gu JD, Mu BZ. (2011). Microbial communities involved in anaerobic degradation of alkanes. *Int Biodeter Biodegr* **65**: 1-13.
- Mbadinga SM, Li KP, Zhou L, Wang LY, Yang SZ, Liu JF *et al.*, (2012). Analysis of alkane-dependent methanogenic community derived from production water of a high-temperature petroleum reservoir. *Appl Microbiol Biotechnol* **96**: 531-542.
- McInerney MJ, Rohlin L, Mouttaki H, Kim U, Krupp RS, Rios-Hernandez L *et al.*, (2007). The genome of *Syntrophus aciditrophicus*: Life at the thermodynamic limit of microbial growth. *PNAS* **104**: 7600-7605.
- Meyer F, Paarmann D, D'Souza M, Olson R, Glass EM, Kubal M *et al.*, (2008). The metagenomics RAST server - a public resource for the automatic phylogenetic and functional analysis of metagenomes. *Bmc Bioinformatics* **9**.
- Morasch B, Schink B, Tebbe CC, Meckenstock RU. (2004). Degradation of *o*-xylene and *m*-xylene by a novel sulfate-reducer belonging to the genus *Desulfotomaculum*. *Arch Microbiol* **181**: 407-417.

- Muller JA, Galushko AS, Kappler A, Schink B. (2001). Initiation of anaerobic degradation of *p*-cresol by formation of 4-hydroxybenzylsuccinate in *Desulfobacterium cetonicum*. *J Bacteriol* **183**: 752-757.
- Parisi VA, Brubaker GR, Zenker MJ, Prince RC, Gieg LM, da Silva MLB *et al.*, (2009). Field metabolomics and laboratory assessments of anaerobic intrinsic bioremediation of hydrocarbons at a petroleum-contaminated site. *Microbial Biotechnology* **2**: 202-212.
- Patil KR, Rouné L, McHardy AC. (2012). The PhyloPythiaS Web Server for Taxonomic Assignment of Metagenome Sequences. *Plos One* **7**: 9.
- Penner TJ, Foght JM. (2010). Mature fine tailings from oil sands processing harbour diverse methanogenic communities. *Can J Microbiol* **56**: 459-470.
- Pilloni G, von Netzer F, Engel M, Lueders T. (2011). Electron acceptor-dependent identification of key anaerobic toluene degraders at a tar-oil-contaminated aquifer by Pyro-SIP. *FEMS Microbiol Ecol* **78**: 165-175.
- Plugge CM, Zhang W, Scholten JC, Stams AJ. (2011). Metabolic flexibility of sulfate-reducing bacteria. *Front Microbiol* **2**: 81.
- Prince RC, Suflita JM. (2007). Anaerobic biodegradation of natural gas condensate can be stimulated by the addition of gasoline. *Biodegradation* **18**: 515-523.
- Pruesse E, Peplies J, Glockner FO. (2012). SINA: accurate high-throughput multiple sequence alignment of ribosomal RNA genes. *Bioinformatics* **28**: 1823-1829.
- Rabus R, Wilkes H, Behrends A, Armstroff A, Fischer T, Pierik AJ *et al.*, (2001). Anaerobic initial reaction of *n*-alkanes in a denitrifying bacterium: Evidence for (1-methylpentyl)succinate as initial product and for involvement of an organic radical in *n*-hexane metabolism. *J Bacteriol* **183**: 1707-1715.
- Rabus R, Jarling R, Lahme S, Kuhner S, Heider J, Widdel F *et al.*, (2011). Co-metabolic conversion of toluene in anaerobic *n*-alkane-degrading bacteria. *Environ Microbiol* **13**: 2576-2585.
- Ramos-Padron E, Bordenave S, Lin SP, Bhaskar IM, Dong XL, Sensen CW *et al.*, (2011). Carbon and sulfur cycling by microbial communities in a gypsum-treated oil sands tailings pond. *Environ Sci Technol* **45**: 439-446.
- Rios-Hernandez LA, Gieg LM, Suflita JM. (2003). Biodegradation of an alicyclic hydrocarbon by a sulfate-reducing enrichment from a gas condensate-contaminated aquifer. *Appl Environ Microbiol* **69**: 434-443.

Rosana ARR, Chamot D, Owttrim GW. (2012). Autoregulation of RNA helicase Expression in Response to Temperature Stress in *Synechocystis* sp PCC 6803. *Plos One* **7**.

Salloum MJ, Dudas MJ, Fedorak PM. (2002). Microbial reduction of amended sulfate in anaerobic mature fine tailings from oil sand. *Waste Manage Res* **20**: 162-171.

Sato T, Atomi H, Imanaka T. (2007). Archaeal type III RuBisCOs function in a pathway for AMP metabolism. *Science* **315**: 1003-1006.

Savage KN, Krumholz LR, Gieg LM, Parisi VA, Suflita JM, Allen J *et al.*,. (2010). Biodegradation of low-molecular-weight alkanes under mesophilic, sulfate-reducing conditions: metabolic intermediates and community patterns. *FEMS Microbiol Ecol* **72**: 485-495.

Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB *et al.*,. (2009). Introducing mothur: Open-Source, Platform-Independent, Community-Supported Software for Describing and Comparing Microbial Communities. *Appl Environ Microbiol* **75**: 7537-7541.

Schmidt JE, Ahring BK. (1993). Effects of hydrogen and formate on the degradation of propionate and butyrate in thermophilic granules from an upflow anaerobic sludge blanket reactor. *Appl Environ Microbiol* **59**: 2546-2551.

Selesi D, Jehmlich N, von Bergen M, Schmidt F, Rattei T, Tischler P *et al.*,. (2010). Combined genomic and proteomic approaches identify gene clusters involved in anaerobic 2-methylnaphthalene degradation in the sulfate-reducing enrichment culture N47. *J Bacteriol* **192**: 295-306.

Siddique T, Fedorak PM, Foght JM. (2006). Biodegradation of short-chain *n*-alkanes in oil sands tailings under methanogenic conditions. *Environ Sci Technol* **40**: 5459-5464.

Siddique T, Fedorak PM, McKinnon MD, Foght JM. (2007). Metabolism of BTEX and naphtha compounds to methane in oil sands tailings. *Environ Sci Technol* **41**: 2350-2356.

Siddique T, Gupta R, Fedorak PM, MacKinnon MD, Foght JM. (2008). A first approximation kinetic model to predict methane generation from an oil sands tailings settling basin. *Chemosphere* **72**: 1573-1580.

Siddique T, Penner T, Semple K, Foght JM. (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, Nesbo C, Foght JM. (2012). Microbial communities involved in methane production from hydrocarbons in oil sands tailings. *Environ Sci Technol* **46**: 9802-9810.
- So CM, Phelps CD, Young LY. (2003). Anaerobic transformation of alkanes to fatty acids by a sulfate-reducing bacterium, strain Hxd3. *Appl Environ Microbiol* **69**: 3892-3900.
- Soh J, Dong X, Caffrey SM, Voordouw G, Sensen CW. (2013). Phoenix 2: A locally installable large-scale 16S rRNA gene sequence analysis pipeline with Web interface. *J Biotechnol* **167**: 393-403.
- Tamura K, Dudley J, Nei M, Kumar S. (2007). MEGA4: Molecular evolutionary genetics analysis (MEGA) software version 4.0. *Mol Biol Evol* **24**: 1596-1599.
- Tang SQ, Gong YC, Edwards EA. (2012). Semi-automatic *in silico* gap closure enabled *de novo* assembly of two dehalobacter genomes from metagenomic data. *Plos One* **7**. doi. 10.1371/journal.pone.0052038
- Vignais PM, Billoud B. (2007). Occurrence, classification, and biological function of hydrogenases: An overview. *Chemical Reviews* **107**: 4206-4272.
- Warikoo V, McInerney MJ, Robinson JA, Suflita JM. (1996). Interspecies acetate transfer influences the extent of anaerobic benzoate degradation by syntrophic consortia. *Appl Environ Microbiol* **62**: 26-32.
- Washer CE, Edwards EA. (2007). Identification and expression of benzylsuccinate synthase genes in a toluene-degrading methanogenic consortium. *Appl Environ Microbiol* **73**: 1367-1369.
- Weelink SAB, van Doesburg W, Saia FT, Rijpstra WIC, Roling WFM, Smidt H *et al.*, (2009). A strictly anaerobic betaproteobacterium *Georgfuchsia toluolica* gen. nov., sp nov degrades aromatic compounds with Fe(III), Mn(IV) or nitrate as an electron acceptor. *FEMS Microbiol Ecol* **70**: 575-585.
- White JR, Nagarajan N, Pop M. (2009). Statistical methods for detecting differentially abundant features in clinical metagenomic samples. *PLoS Comp Biol* **5**: e1000352.
- Widdel F, Pfennig N. (1982). Studies on dissimilatory sulfate-reducing bacteria that decompose fatty acids. *Arch Microbiol* **131**: 360-365.
- Widdel F, Bak F (1992). Gram-negative mesophilic sulfate-reducing bacteria. In: Balows A, Trüper, H. G., Dworkin, M., Harder, W. and Schleifer, K.H. (ed). *The Prokaryotes*. Springer-Verlag: New York, USA. pp 3352-3378.

Wilkes H, Kuhner S, Bolm C, Fischer T, Classen A, Widdel F *et al.*, (2003). Formation of *n*-alkane- and cycloalkane-derived organic acids during anaerobic growth of a denitrifying bacterium with crude oil. *Org Geochem* **34**: 1313-1323.

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

Winderl C, Penning H, von Netzer F, Meckenstock RU, Lueders T. (2010). DNA-SIP identifies sulfate-reducing *Clostridia* as important toluene degraders in tar-oil-contaminated aquifer sediment. *Isme Journal* **4**: 1314-1325.

Yarza P, Ludwig W, Euzéby J, Amann R, Schleifer K-H, Gloeckner FO *et al.*, (2010). Update of the All-Species Living Tree Project based on 16S and 23S rRNA sequence analyses. *Syst Appl Microbiol* **33**: 291-299.

Zedelius J, Rabus R, Grundmann O, Werner I, Brodkorb D, Schreiber F *et al.*, (2011). Alkane degradation under anoxic conditions by a nitrate-reducing bacterium with possible involvement of the electron acceptor in substrate activation. *Environmental Microbiology Reports* **3**: 125-135.

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

Zhou L, Li K-P, Mbadinga SM, Yang S-Z, Gu J-D, Mu B-Z. (2012). Analyses of *n*-alkanes degrading community dynamics of a high-temperature methanogenic consortium enriched from production water of a petroleum reservoir by a combination of molecular techniques. *Ecotoxicology (London, England)* **21**: 1680-1691.

## **4 Metagenomic and metatranscriptomic analyses of a model alkane-degrading enrichment culture implicate roles for *Desulfotomaculum* and *Smithella* in anaerobic alkane degradation<sup>1</sup>**

### **4.1 Abstract**

Methanogenic degradation of alkanes, i.e., 2-methylpentane, methylcyclopentane and *n*-C<sub>6</sub> to *n*-C<sub>10</sub> was investigated using an enrichment culture (SCADC) established from mature fine tailings of an oil sands tailing pond (Chapter 2). The SCADC enrichment culture was transferred twice, and the rRNA-depleted total RNA was isolated during active methanogenesis, converted to cDNA and subjected to RNA-seq using the Illumina platform. *De novo* assembly of the metagenome reported in Chapter 3, followed by metagenomic binning using multiple approaches based on sequence composition and homology resulted in several partial to almost complete genomic bins, in particular a bin of an unculturable Peptococcaceae affiliated with *Desulfotomaculum* spp. and of a Syntrophaceae affiliated with *Smithella* spp. that both have the genetic capability for alkane degradation by fumarate addition. Mapping of transcripts to these genomes revealed high expression of genes involved in alkane activation to fumarate, fumarate regeneration and beta-oxidation by "*Desulfotomaculum*". The fumarate addition gene in *Smithella* has low raw read mapping value, indicating that it is not responsible for low molecular weight alkane activation, and likely uses longer-chain alkanes as growth substrates. Multiple genes encoding syntrophic processes such as interspecies formate and hydrogen transfer, and pili/flagella for aggregate formation were highly expressed. Overall, metatranscriptomics results show that SCADC is highly dynamic and provide the first evidence for novel clades of "*Desulfotomaculum*" and "*Smithella*" being able to catalyze fumarate activation of alkanes.

<sup>1</sup>A version of this chapter has been modified for publication to demonstrate gene expression of SCADC community during active methanogenesis and alkane biodegradation. Particularly, this chapter demonstrates the expression of genes related to fumarate addition activation of *n*-alkanes by a novel microbe related to *Desulfotomaculum* within the family Peptococcaceae.

## 4.2 Introduction

In surface-mined oil sands, hot water and solvents are used to separate bitumen from the oil sand ores (Chalaturnyk *et al.*, 2002). The resulting slurry of water, sand, clays and hydrocarbons is retained in tailing ponds, which have been the subject of scrutiny due to environmental concerns about water usage, greenhouse gas emissions and long-term reclamation (Chalaturnyk *et al.*, 2002). Components of the residual solvent in tailings, in particular the alkanes, can be anaerobically biodegraded by indigenous tailings pond microorganisms to produce methane and carbon dioxide (Siddique *et al.*, 2006, Siddique *et al.*, 2007b, Siddique *et al.*, 2011, Siddique *et al.*, 2012). Biodegradation of alkanes of various chain lengths has been reported to occur under primarily sulfate- and nitrate-reducing condition (reviewed by Agrawal and Gieg 2013, Callaghan 2013). Recently, a number of alkane-degrading sulfate- and nitrate-reducers have been isolated, and their degradation mechanism elucidated (Callaghan *et al.*, 2008, Grundmann *et al.*, 2008). Under sulfate- and nitrate-reducing conditions, alkane activation has been attributed to the "hydrocarbon addition to fumarate" mechanism catalyzed by alkylsuccinate synthase (Callaghan *et al.*, 2008, Grundmann *et al.*, 2008). The susceptibility of different alkanes to methanogenic degradation has also been reported (Aitken *et al.*, 2013, Gray *et al.*, 2011, Li *et al.*, 2012, Mbadanga *et al.*, 2012, Siddique *et al.*, 2011, Zengler *et al.*, 1999, Zhou *et al.*, 2012), although the initial activating mechanism has not been conclusively shown, due to the lack of metabolite evidence under methanogenic conditions (Aitken *et al.*, 2013, Callaghan 2013). Also, although numerous studies of methanogenic degradation of longer-chain alkanes (e.g. *n*-hexadecane) have implicated *Smithella/Syntrophus* spp. as the primary alkane-degrader (Cheng *et al.*, 2013, Gray *et al.*, 2011, Zengler *et al.*, 1999), the activating mechanism is currently unknown and there is no evidence that the closest cultivated relative *Syntrophus aciditrophicus* SB1 is capable of initiating alkane degradation (McInerney *et al.*, 2007). Similarly, the degradation of low molecular weight alkanes (i.e.,  $n < 10$ ) has mostly been reported under sulfidogenic and nitrate-reducing conditions (Davidova *et al.*, 2005, Ehrenreich *et al.*, 2000, Kniemeyer *et*

*al.*, 2007). Microbes in these communities are mostly related to the Proteobacteria (Agrawal and Gieg 2013), although limited study has also implicated Firmicutes related to *Desulfotomaculum* spp. as being capable of degrading low molecular weight alkanes (e.g., propane) under sulfate-reducing conditions (Kniemeyer *et al.*, 2007).

In Chapter 3, we reported the metagenomic analysis of SCADC (short chain alkane degrading culture) used as a model for intensive study of methanogenic alkane degradation by oil sands microbes. SCADC, which comprised approximately 300 Operational Taxonomic Units (OTUs) at the time of analysis, completely degraded *n*-C<sub>6</sub>, *n*-C<sub>7</sub>, *n*-C<sub>8</sub> and *n*-C<sub>10</sub>, 2-methylpentane and methylcyclopentane within 18 months incubation in the laboratory. In this Chapter, metatranscriptomic analysis of the SCADC enrichment culture was conducted by first isolating RNA from duplicate cultures of 400 mL during active methanogenesis. The RNA was subsequently depleted of rRNA, reverse-transcribed to cDNA and sequenced using Illumina Hi-seq. Mapping of transcripts to annotated Open Reading Frames (ORFs) of novel genomic bins, followed by pathway analysis revealed alkane degradation and carbon assimilation pathways and putative syntrophic relationships by novel organisms previously unknown to be capable of anaerobic *n*-alkane degradation. Highly expressed phage-like sequences detected in the metatranscriptome are considered as a mechanism of gene dissemination and reorganization in SCADC community during methanogenic growth.

### **4.3 Materials and Methods**

#### **4.3.1 Description of SCADC enrichment culture**

Replicate bottles of SCADC enrichment culture were previously enriched using mature fine tailings (MFT) from Syncrude Mildred Lake Settling Basin by using alkanes as the only organic carbon source (Chapter 2; Figure 1.3). Cultures had been transferred twice before metagenomic sequencing using Illumina Hi-Seq and 454 Pyrotag sequencing platforms (Chapter 3). The SCADC microbial community degrades *n*-C<sub>6</sub> to *n*-C<sub>10</sub>, 2-methylpentane and methylcyclopentane, with concomitant production of methane during incubation (Chapter 2 and 3;

Figure 1.3). To prepare a metatranscriptome, a 400 mL SCADC culture was transferred twice (incubating 4 months at 28°C for each transfer), before total RNA isolation (Figure 1.3). Further description of the SCADC enrichment culture and assembly of SCADC metagenome is provided in Chapter 3 and Figure 1.3.

#### **4.3.2 Total RNA isolation and quality control of sequence reads**

Total RNA was extracted from duplicate 400-mL enrichment cultures of SCADC (Figure 1.3) during active methanogenesis using a modified phenol-bead lysis method (Rosana *et al.*, 2012). Upon opening the anaerobic culture bottle, a volume of ice-cold stop solution (10% phenol in ethanol) equal to the culture volume was added to inactivate cells and preserve RNA integrity. The subsequent steps to isolate total RNA have been described in Chapter 2. Thereafter, total isolated RNA was treated twice with DNase I (Ambion, US) to remove DNA, and RNA <200 bp was removed using the RNA Clean and Concentrator™ kit (ZymoResearch, US). To assess the presence of contaminating DNA, the RNA was subjected to PCR amplification of the 16S rRNA gene using primer pair 27F (5'-GAGTTTGMTTCCTGGCTCAG-3') and 1492R (5'-ACG GYTACCTTGTTACGACTT-3'); no amplification was observed after gel electrophoresis, indicating the absence of contaminating DNA (data not shown).

The total RNA was subjected to PCR amplification of the 16S rRNA gene, pyrotag sequencing and bioinformatic analysis as previously described by Tan *et al.* (2013) and in Chapter 3. A portion of the total RNA was subsequently depleted of rRNA using a terminator 5'-phosphate-dependent exonuclease kit (Epicentre, US), followed by the Ribo-Zero rRNA removal kit (Epicentre, US). The quality of total RNA and rRNA-depleted total RNA was assessed using a Bioanalyzer instrument (Agilent Technologies, US). Results indicated the absence of high molecular weight nucleic acid (i.e., absence of DNA) and the removal of 16S and 23S rRNA. This rRNA-depleted total RNA was subsequently used for double stranded cDNA construction using random hexamers and SuperscriptIII reverse transcriptase (Invitrogen, US) following the manufacturer's protocol. Double stranded cDNA was subjected to Illumina

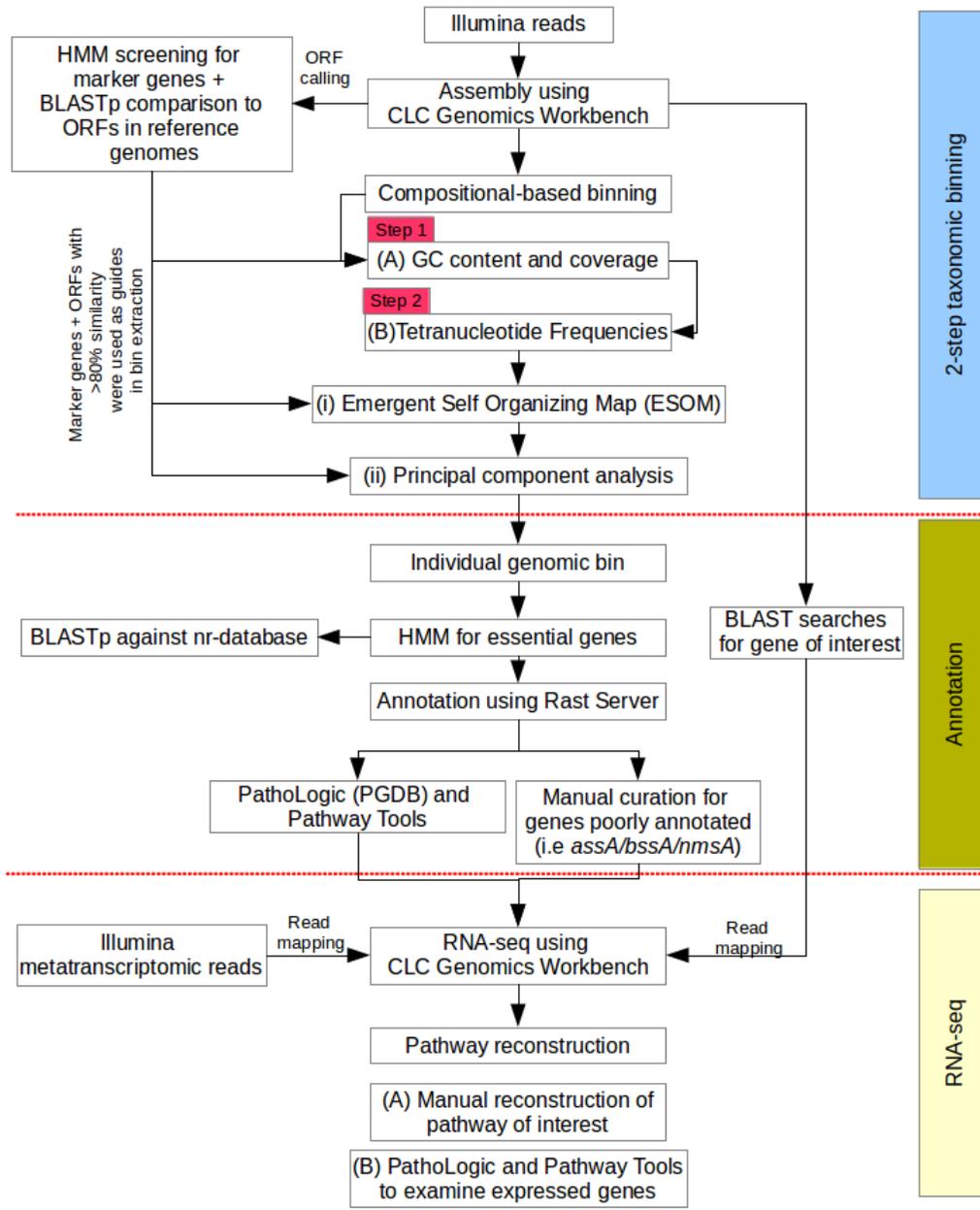
Hi-seq sequencing at McGill University Génome Québec Innovation Centre, Canada.

#### **4.4 Bioinformatic analysis of SCADC metagenome and metatranscriptome**

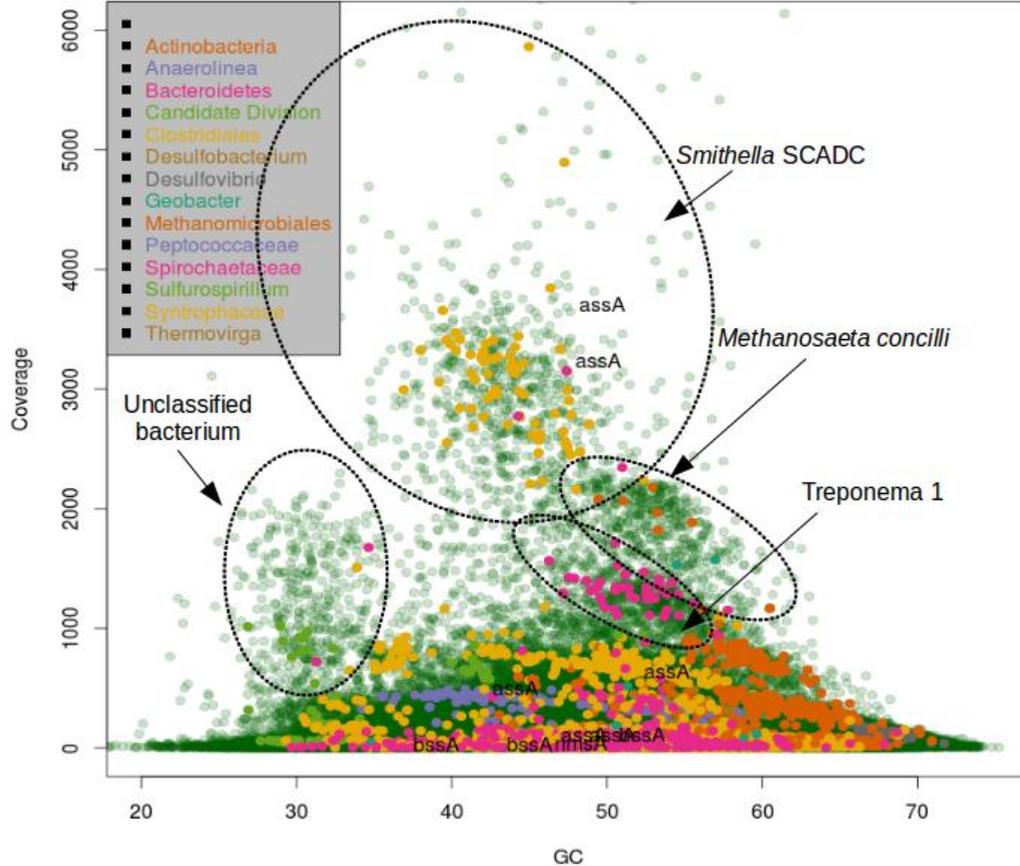
Metagenomic reads generated using the Illumina platforms reported in Chapter 3 were re-assembled using CLC Genomic Workbench with the default settings (CLC Bio). A combination of strategies previously reported (Haroon *et al.*, 2013, Ishii *et al.*, 2013) were adapted and used in the current study for contig binning (Figure 4.1; Appendix D Methods). Briefly, contigs were first binned based on GC content and sequence coverage (Figure 4.2 and 4.3). Each bin was further refined using tetranucleotide frequencies, followed by principal component analysis in R (<http://www.r-project.org/>; Figure 4.4) or Emergent Self-Organizing Map (ESOM) analysis (Figure 4.5) (Dick *et al.*, 2009). The completeness of selected genomic bins was assessed by comparing the number of essential marker genes to those present in the closest affiliated reference genomes, followed by BLASTp searches against the nr-database to confirm their identity (Appendix Tables D2 and D3). Selected taxonomic bins were submitted for automated annotation using Rapid Annotation Using Subsystem Technology (RAST) server (Aziz *et al.*, 2008), and the annotation was used in reconstruction of metabolic pathways using Pathway Tools (Karp *et al.*, 2010). Genes of interest including *assA/bssA/nmsA* and *mcrA* were recovered from SCADC sequence assembly and each genomic bin using BLASTn searches with a list of reference sequences (Chapter 3).

Metatranscriptomic reads generated using Illumina paired-end sequencing were subjected to the same quality control procedures as the metagenome described in Chapter 3. The quality-filtered SCADC metatranscriptomic reads consisted of 700 million paired-end Illumina sequences with read length ranging from 50 bp to 138 bp and mean G+C content of 44%. Metatranscripts were mapped to annotated reads in selected genomic bins and sequence reads using the RNA-seq platform in CLC Genomics Workbench (CLC Bio) with the following parameters: 96% similarity over 95% of transcript length. Additional RNA-seq

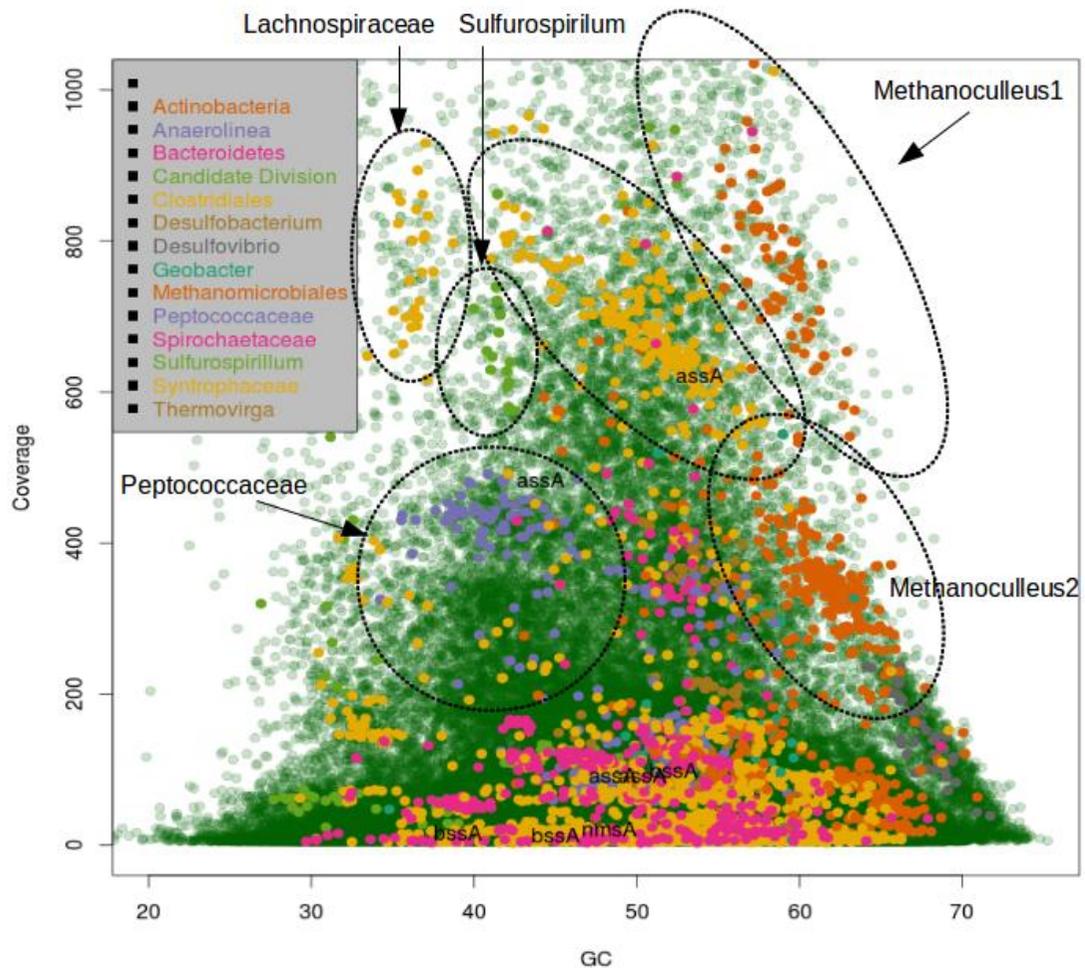
was also conducted using the reference genome of *Methanosaeta concilii* GP6 (NC\_015416) because previous bioinformatic analysis had indicated the presence of a closely related strain (Chapter 3).



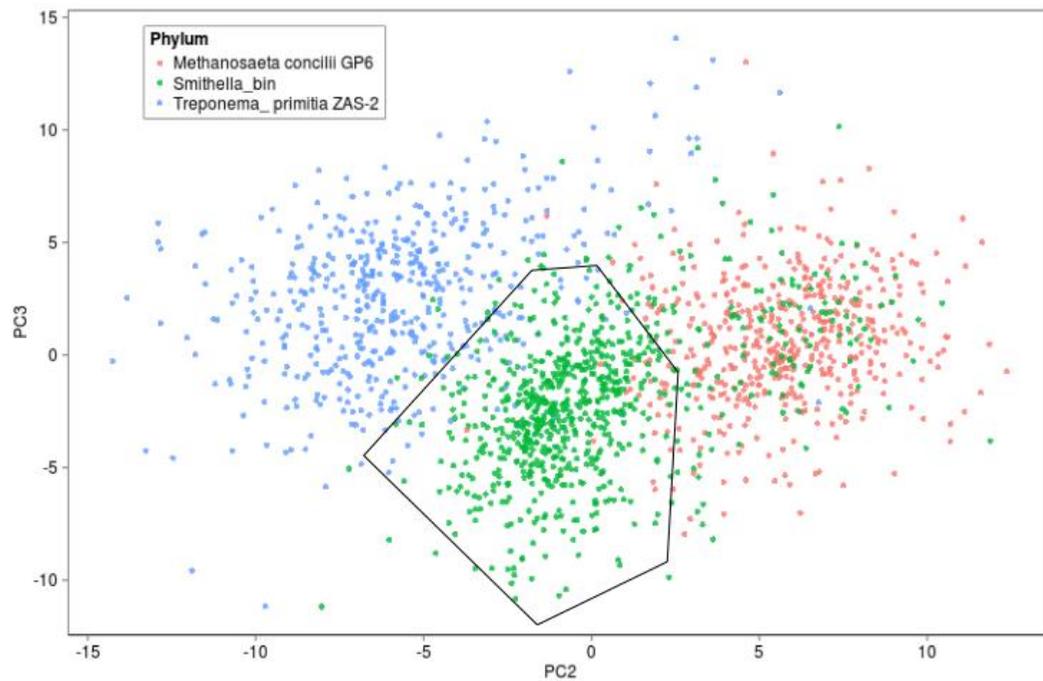
**Figure 4.1** Schematic representation of the "omics" workflow established for analysis of SCADC metagenome and metatranscriptome generated using Illumina Hi-seq. A detailed description of the workflow is presented in text in Appendix Method D.



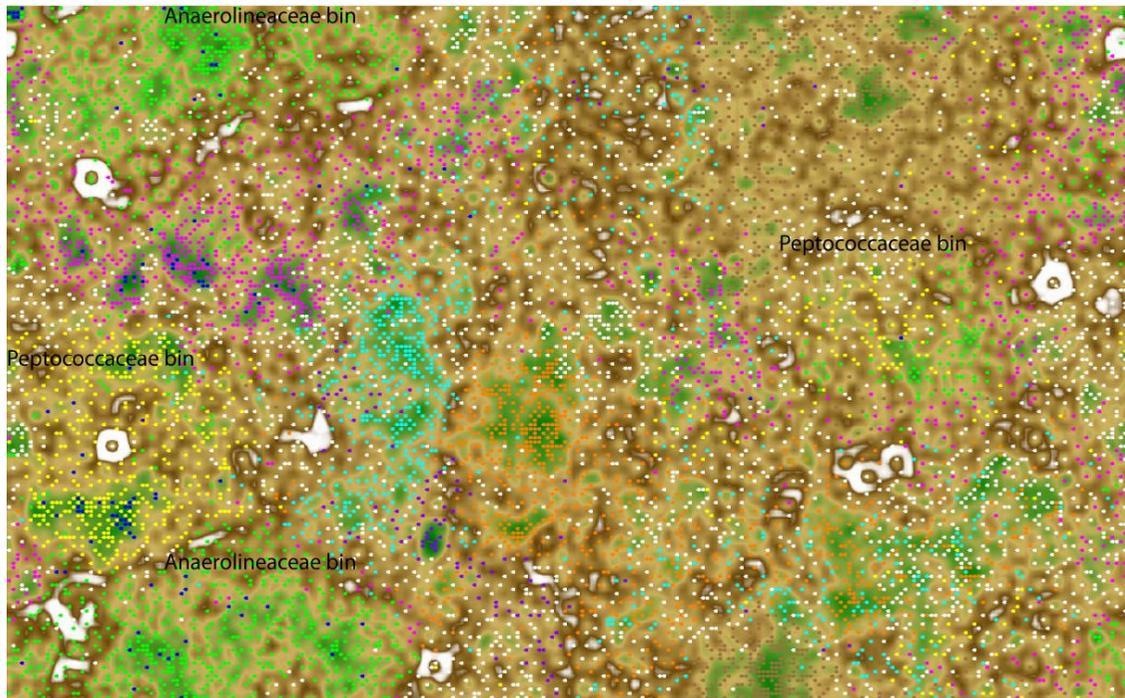
**Figure 4.2** Example of clustering of metagenome contigs based on sequence coverage and GC content. Each dot represents a contig of specific sequence length. Colours were overlaid on each contig based on the presence of (i) essential genes classified based on best BLAST hits against the NCBI nr-database, and (ii) homologous genes (>80% similarity) present in related species. Genes assigned to *assA*/*bssA*/*nmsA* were recovered from the SCADC metagenome sequence assembly by BLASTn searches and overlaid on the plot to guide binning. This plot shows the clear separation of subsets of contigs, in particular for *Smithella* SCADC (see main text), *Methanosaeta* and *Treponema*, and an unclassified genome. Contigs bearing *assA*, *bssA* or *nmsA* genes are labelled.



**Figure 4.3** Example of clustering of metagenome contigs based on sequence coverage and GC content. This plot is similar to Figure 4.2 but shows contigs that have lower sequence coverage. See Figure 4.2 legend for plot description. Contigs bearing *assA*, *bssA* or *nmsA* genes are labelled. This plot shows the separation of subsets of contigs for *Desulfotomaculum* SCADC (see main text), *Methanoculleus* 1 and 2, Syntrophaceae 2, *Sulfurospirillum* and Lachnospiraceae.



**Figure 4.4** Example of Principal Component Analysis of the tetranucleotide frequencies composition of SCADC metagenome contigs binned using GC content versus coverage (Figure 4.2). The genome of *Methanosaeta concilii* GP6 (NC\_015416) and *Treponema primitia* ZAS-2 (NC\_015578) were seeded in the analysis so that contigs sharing similar tetranucleotide frequencies with these two reference genomes would be distinguished from the *Smithella* SCADC bin (see main text).



**Figure 4.5** Example of Emergent Self-Organizing Map (ESOM) analysis of the tetranucleotide composition of contigs binned in Figure 4.3. Figure shows clear separation of contigs associated with *Desulfotomaculum* SCADC (yellow dots) from other metagenomic bins. Dark blue dots represent single copy essential genes used in guidance of bin extraction.

## **4.5 Results and discussion**

### **4.5.1 Binning of SCADC metagenome**

#### **4.5.1.1 Description of SCADC enrichment culture**

In order to investigate the roles of microbial community members in oil sands tailings capable of degrading alkanes under methanogenic conditions, a model enrichment culture (SCADC) was established using oil sands mature fine tailings (MFT) as the inoculum (Chapters 2 and 3; Figure 1.3). SCADC was maintained in basal mineral medium with N<sub>2</sub>/CO<sub>2</sub> in the headspace and was found to degrade *n*-C<sub>6</sub> to *n*-C<sub>10</sub>, and the minor components 2-methylpentane and methylcyclopentane in the "hexanes" supplied as the only carbon source (Chapter 2 and 3). The culture had been transferred several times in the laboratory before total DNA was isolated for metagenomic sequencing (Chapter 3; Figure 1.3). The metatranscriptome was obtained from duplicate 400 mL SCADC cultures (Figure 1.3) during the active phase of methanogenesis when the SCADC community was actively degrading alkanes.

#### **4.5.1.2 Metagenomic binning of SCADC metagenome**

Previously, binning of Illumina and 454 sequence assemblies obtained from the SCADC metagenome using a number of sequence-homology based methods (i.e., PhylopythiaS, T2Binning and SOrt-ITEMS) had proven to be unsuccessful, particularly for identifying genomes that might be directly involved in the activation of alkanes (Chapter 3). This failure was likely due to the complexity of the culture, presence of several highly related strains (i.e., Clostridiales, Syntrophaceae and Anaerolineaceae) and lack of closely related reference genomes that could be used in dataset training.

In the current study, SCADC Illumina reads (Chapter 3) were re-assembled using CLC Genomics Workbench and binned with a combination of strategies (Figure 4.1 and Appendix D Methods) shown to be successful in separating metagenomic reads from complex microbial communities (Haroon *et al.*, 2013, Ishii *et al.*, 2013). Sequence coverage for each contig was obtained using CLC Genomic Workbench during the assembly process whereby Illumina reads were mapped back to contigs that shared 95% similarity over 95% of

sequence read length. The differences in sequence coverage and GC content, as well as tetranucleotide frequency of different microbial populations allowed the binning of several partial to nearly complete genomes (Table 4.1; Appendix Table D1). The completeness of selected genomic bins was determined by quantifying the number of essential genes compared to that in the closest affiliated reference genomes (Table 4.1; Appendix Table D2 and D3).

#### **4.5.1.3 Identification of genomes capable of hydrocarbon addition to fumarate**

We first focus on identifying genomes that are capable of hydrocarbon addition to fumarate by BLASTn and BLASTx searches of individual genomic bins and SCADC sequence assemblies for putative *assA/bssA/nmsA* homologs encoding the  $\alpha$ -subunit of alkylsuccinate (ASS), benzylsuccinate (BSS) and 2-methylnaphthalenesuccinate synthase (NMS). All recovered sequences were subjected to phylogenetic analysis to assign them to their respective phylogenetic clades performed as described in Chapter 3. Full length *assA* genes were identified in genomes affiliated with “*Desulfotomaculum*” SCADC, “*Smithella*” SCADC, Syntrophaceae 2, Desulfobacteraceae and Clostridiales (Table 4.1). The genus names are placed in quotation marks to signify that these OTUs form clades phylogenetically distinct from published genera, but that biochemical characterization of these uncultivated OTUs not been conducted to permit the proposal of a new, named genus.

The taxonomic affiliation of the *ass* operon in “*Desulfotomaculum*” SCADC was supported by identification on the contig of valyl-tRNA synthetase (best BLASTX hit *Pelotomaculum thermopropionicum*, 71% similar), *arsR* transcriptional regulator (*Desulfotomaculum gibsoniae* DSM 7213, 54%) and other functional genes, as reported in Chapter 3. BLASTn comparison of the *Desulfotomaculum* SCADC-affiliated 16S rRNA gene obtained from 454 pyrotag sequencing indicates that the closest related sequences were recovered mainly from methanogenic environments with historical exposure to hydrocarbons (Figure 4.6).

**Table 4.1** Completeness of selected genomic bins obtained from SCADC metagenome sequence assembly

Taxonomic affiliation of genomic bin	Genome Size (Mbp)	% GC	Completeness (%) <sup>a</sup>	Closest affiliation (based on RAST)	Activating genes detected
<b>Potential primary degrader</b>					
<i>Desulfotomaculum</i> SCADC	3.1	43	96	<i>Pelotomaculum thermopropionicum</i> SI	<i>assA</i> <sup>c</sup>
<i>Smithella</i> SCADC	3.6	44	93	<i>Syntrophus aciditrophicus</i> SB 1	<i>assA</i> <sup>c,d</sup>
Syntrophaceae 2	2.0	50	89	<i>Syntrophus aciditrophicus</i> SB 1	<i>assA</i>
Desulfobacteraceae	5.2	50	94	<i>Desulfatibacillum alkenivorans</i> AK-01	<i>assA</i>
Clostridiales	3.7	50	100	<i>Heliobacterium modesticaldum</i> Ice1	<i>assA</i>
Deltaproteobacteria	3.5	53	>100 <sup>b</sup>	<i>Desulfuromonas acetoxidans</i>	<i>assA</i>
<b>Methanogens</b>					
<i>Methanoculleus</i> 1	2.7	60	85	<i>Methanoculleus marisnigri</i> JR1	-
<i>Methanoculleus</i> 2	2.4	63	85	<i>Methanoculleus marisnigri</i> JR1	-
Methanoregulaceae 1	1.7	60	80	<i>Methanoregula boonei</i> 6A8	-
Methanoregulaceae 2	2.9	56	100	<i>Methanoregula boonei</i> 6A8	-
Methanoregulaceae 3	2.2	48	80	<i>Methanoregula boonei</i> 6A8	-

<sup>a</sup>, Single copy essential genes were determined using marker genes reported by Albertsen *et al.* (2013). Completeness of archaeal genomic bin was based on number of marker genes present in genome of *Methanoculleus marisnigri* (NC\_009051.1) and *Methanoregula boonei* 6A8 (NC\_009712.1), determined using the same Hidden Markov Model (HMM) for bacterial species. For detailed description, see Appendix Table D2 and 3

<sup>b</sup>, indicates potential of containing mixed genomes.

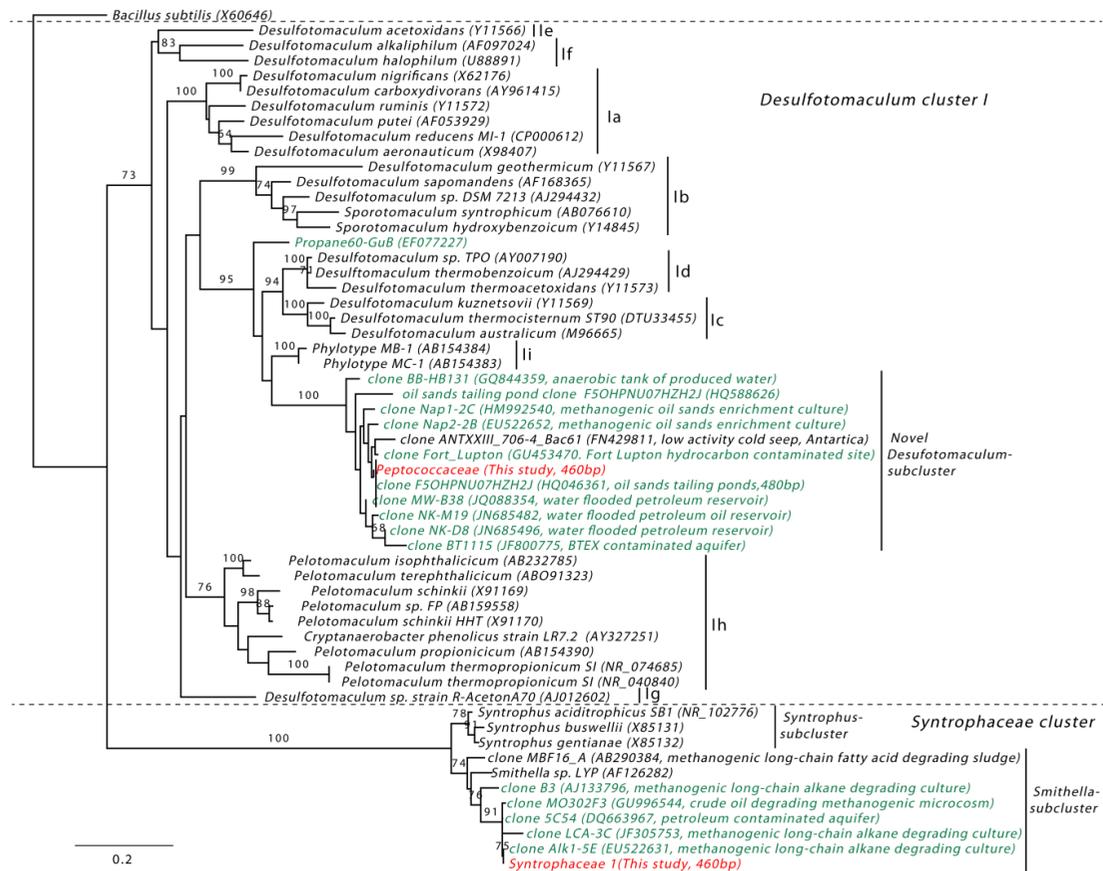
<sup>c</sup>, taxonomic genes were identified in contigs containing these genes (see main text for details).

<sup>d</sup>, also contains a gene related to pyruvate formate lyases (Appendix Figure D1).

Further phylogenetic analysis indicates that these novel sequences from putative *Desulfotomaculum* were also distantly related to *Desulfotomaculum* subcluster Id, Ic and Ii and the only other Firmicutes implicated in degrading low molecular weight hydrocarbon (propane) by coupling to sulfate-reduction (*Desulfotomaculum*-related Firmicutes in enrichment culture Propane-GutB; Kniemeyer *et al.*, 2007) (Figure 4.6). Thus, the “*Desulfotomaculum*” SCADC likely represents a novel *Desulfotomaculum* clade, of which the closest relatives were recovered mainly from methanogenic hydrocarbon-impacted environments.

The *ass* contig in “*Smithella*” SCADC bears no taxonomic signature. Bioinformatic analysis of a fosmid sequence of 53 kbp originating from a parallel oil sands tailings enrichment culture (NAPDC reported in Chapter 5) indicated that it belongs to “*Smithella* SCADC”. The *ass* contig of “*Smithella*” SCADC assembly is 20,483 bp and shares 99.8% nucleotide similarity (overlapping region of ~4000 bp) with the fosmid sequence. The fosmid sequence contains a gene encoding glutamyl tRNA synthetase (best BLASTx hit to *Syntrophus aciditrophicus* SB1, 70% similar) and 50S ribosomal protein L32 (*S. aciditrophicus* SB1, 67.2%), thus supporting the proposal that this *ass* operon is associated with the Syntrophaceae, which include both *Smithella* and *Syntrophus*.

Phylogenetic analysis of the “*Smithella*” SCADC 16S rRNA gene showed that it is highly similar (>97%; Figure 4.6) to the *Smithella* 16S rRNA gene sequences that had been implicated in methanogenic degradation of crude oil alkanes (clone MO302F3 in Gray *et al.*, 2011), but more dissimilar to the *Smithella* spp. reported by Embree *et al.*, (2013). Other *ass* contigs were too short to yield any taxonomic signatures (Table 4.1), however, the putative *assA* in “*Smithella*” SCADC and Syntrophaceae 2 appear to belong to the same subcluster, along with the putative *assA* gene recovered from the *Smithella* genome reported by Embree *et al.*, (2013) and partial sequence of *assA* obtained from other methanogenic alkane-degrading enrichment cultures where members of Syntrophaceae have been implicated as the primary alkane-degraders (Aitken *et al.*, 2013)(Appendix Figure D1 and Appendix A).



**Figure 4.6** Maximum likelihood tree of the 16S rRNA gene of “*Desulfotomaculum*” SCADC and “*Smithella*” SCADC obtained from 16S rRNA gene pyrotag sequencing (labelled in red) compared to reference 16S rRNA sequences and sequences originating from hydrocarbon-impacted environments (labelled in green). All sequences were at least 1000 bp (unless indicated otherwise in brackets following the GenBank accession number). All sequences were aligned using the SINA aligner (Pruesse *et al.*, 2012), followed by phylogenetic construction in Geneious (Kearse *et al.*, 2012) using the TM9 model with 1000 bootstrap replicates. The scale bar denotes 20% sequence divergence and bootstrap values >60 are shown on the nodes.

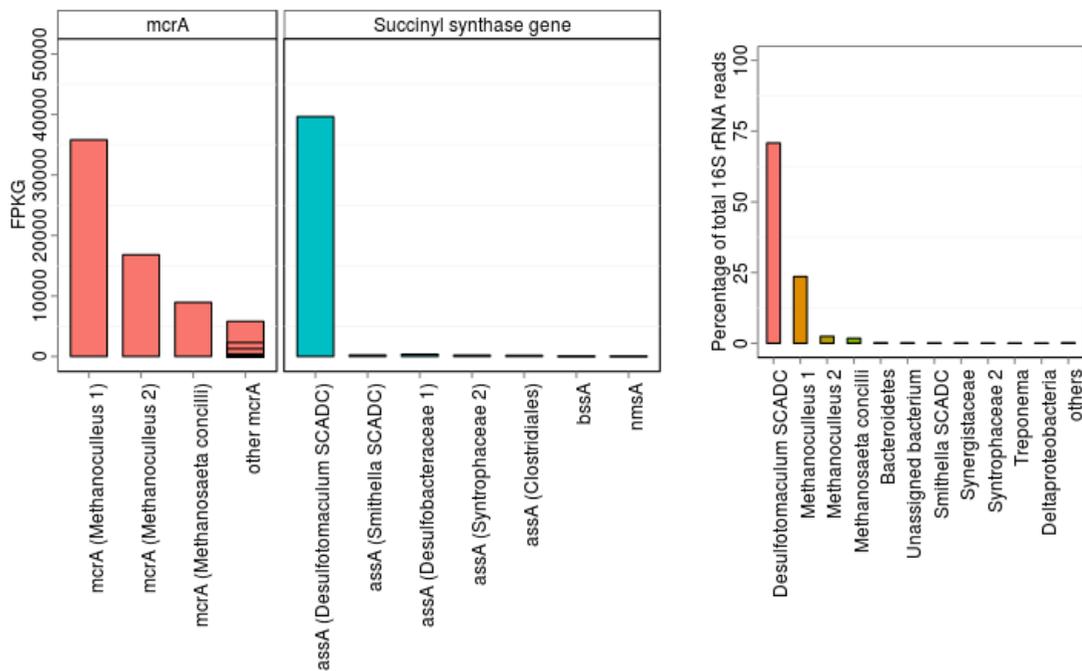
Genomes containing *bssA/nmsA* remained un-binned, perhaps due to the presence of multiple genomes with similar tetranucleotide frequency and/or low contig coverage. However, a single contig harbouring the putative *nmsAB* gene also bore multiple taxonomic genes affiliated with *Desulfotomaculum* spp.

(Appendix Table D4), suggesting that this member of the Clostridiales is capable of activating 2-methylnaphthalene by addition to fumarate. Taxonomic binning also recovered partial genome and mixed genomes (containing multiple strains) of the members of *Methanoculleus*, Bacteroidetes, Anaerolineaceae and Clostridiales, which are better known to be secondary by product users and may be involved in syntrophy with other community members (Appendix Table D1).

#### **4.5.2 Metatranscriptomics reveals that *Desulfotomaculum* SCADC is the most metabolically active member**

The active microbial population was investigated by PCR co-amplification of the bacterial and archaeal 16S rRNA genes using reverse-transcribed total RNA, followed by Pyrotag sequencing, and operational taxonomic unit (OTU) analysis using the Phoenix pipeline (Soh *et al.*, 2013). Results indicate that the most abundant (and therefore presumably most active) 16S rRNA OTUs were represented by that of “*Desulfotomaculum*” SCADC (>70% of the total reads), followed by two OTUs affiliated with *Methanoculleus* and one with *Methanosaeta*. Sequences affiliated with Bacteroidetes and Syntrophaceae collectively made up <2% of the total sequences (Figure 4.7). Thus, based on the pyrotag analysis of the active microbial community, I focused metatranscriptomic analysis on the taxonomic bins affiliated with “*Desulfotomaculum*” SCADC, *Methanoculleus* and *Methanosaeta*.

**Figure 4.7** (Next page) Metatranscriptomic analysis of the relative abundance of (left) genes encoding  $\alpha$ -subunit of alkylsuccinate (*assA*), benzylsuccinate (*bssA*) and naphthylsuccinate synthase (*nmsA*), and genes encoding the  $\alpha$ -subunit of methyl-CoM reductase (*mcrA*). All functional genes were recovered from SCADC Illumina assembly using BLASTn and BLASTx and transcript mapping was conducted in CLC Genomics Workbench and expressed as “Fragments mapped per kilobase of gene length (FPKG)”(right). Metatranscriptomic analysis of the relative abundance of active microbial populations was determined using 16S rRNA gene pyrotag sequencing of cDNA converted from total RNA extracted during active methanogenesis.



#### 4.5.2.1 Metatranscriptomics suggests a primary role for *Desulfotomaculum* SCADC in alkane activation

Activation of alkanes by addition to fumarate has been the most widely reported mechanism for anaerobic alkane degradation in environments with predominantly sulfate-and nitrate-reducing conditions, with evidence provided by metabolite and marker gene detection in fields and enrichment cultures (Aitken *et al.*, 2013, Callaghan *et al.*, 2010, Parisi *et al.*, 2009). Whether the same mechanism prevails under methanogenic condition is, however, currently questionable because the characteristic succinylated products were not detected in alkane-degrading degrading enrichment cultures, although *assA* genes responsible for alkane addition to fumarate have been cloned and sequenced from cultures (Aitken *et al.*, 2013, Callaghan *et al.*, 2010). An alternative alkane activation mechanism recently has been proposed and reported based on proteomic analysis of the Deltaproteobacterium *Desulfococcus oleovorans* Hxd3: hexadecane is first hydroxylated by enzyme an homologous to ethylbenzene dehydrogenase to form a ketone, which is then followed by carboxylation to yield a substrate for  $\beta$ -oxidation (Callaghan 2013). The organism was initially suggested to activate

alkanes by carboxylation coupled to sulfate reduction (Callaghan *et al.*, 2006), although this capability has not been reported under methanogenic conditions.

In order to investigate gene expression during methanogenic growth with alkanes in SCADC, genes encoding enzymes for addition to fumarate were first recovered from SCADC sequence assembly and individual genomics bins. Mapping of metatranscripts to all reported *assA/bssA/nmsA* genes using RNA-Seq in CLC Genomic Workbench (CLC Bio, USA), followed by transcript normalization using "fragments mapped per kilobase of gene length (FPKG)" (Haroon *et al.*, 2013), indicates that the *assA* genes affiliated with “*Desulfotomaculum*” SCADC had the highest expression values (>39 000 FPKG; Figure 4.1) of all recovered *assA/bssA/nsmA* sequences (<100 FPKG; Figure 4.1). This observation corroborates the results of 16S rRNA pyrotag sequencing of the metatranscriptome showing that “*Desulfotomaculum*” SCADC is the most metabolically active OTU in SCADC and suggesting that it activates alkanes by addition to fumarate. The active expression of the *assA* gene is supported by metabolite detection of putative succinylated products of 2-methylpentane and methylcyclopentane activation in SCADC, although metabolite extraction and analysis at different times failed to detect any succinylated product of *n*-alkanes (Chapter 2). Methylcyclopentane is likely not the primary substrate of the ASS enzyme since its putative metabolite continued to accumulate in SCADC culture fluids even after methane production reached plateau (Chapter 2). The inability to detect the characteristic metabolite of *n*-alkane addition to fumarate has been previously reported for other methanogenic alkane-degrading cultures, even though the microbial community carried copies of *assA* genes (Aitken *et al.*, 2013).

In the current study, high expression of *assA* by “*Desulfotomaculum*” SCADC strongly hinted at alkane activation by addition to fumarate under methanogenic conditions, although the substrate range was not fully determined (Chapter 2). It is hypothesized here that, since the ASS enzyme is known to co-activate several hydrocarbon classes, i.e., *n*-alkanes, toluene and *cyclo*-alkanes (Rabus *et al.*, 2011, Wilkes *et al.*, 2003), it is also possible that the ASS

enzyme in "*Desulfotomaculum* SCADC" is capable of co-activating *n*-alkanes and 2-methylpentane. Failure to detect the fumarate addition metabolites under methanogenic conditions may be due to the typical short half-life of metabolites reported for syntrophic microbial consortia that degrade organic acids (e.g., formate involved in formate transfer; Walker *et al.*, 2012). A proposed alkane activation mechanism through hydroxylation was suggested, based on proteomic analysis, to be encoded by genes homologous to ethylbenzene dehydrogenase (Dole\_0914) in *Desulfococcus oleovorans* Hxd3 (Callaghan 2013). Sequence homologs of Dole\_0914 with unknown taxonomic affiliation in SCADC were recovered and reported earlier (Chapter 3). Mapping of metatranscripts to these sequences, however, showed very low FPKG values (0 to <50; not shown). Also, homologous genes were not found in the genomic bins of "*Desulfotomaculum*" SCADC and "*Smithella*" SCADC, indicating that hydroxylation is likely not the alkane activating mechanism in SCADC.

#### **4.5.2.2 *Desulfotomaculum* SCADC is not a sulfate-reducer**

Genes necessary for complete sulfate reduction to H<sub>2</sub>S such as those encoding transmembrane electron transport complexes (*dsrMKJOP*), dissimilatory sulphite reductase  $\alpha$ - and  $\beta$ -subunits (*dsrAB*) and ATP sulfurylase were not detected in "*Desulfotomaculum*" SCADC, with the exception of the *dsrC* gene. Phylogenetic analysis of the *dsrC* gene indicates that it is distinct from the *dsrC* gene of other *Desulfotomaculum* spp., but more similar to the *dsrC* genes in *Pelobacter propionicus* DSM2379 (NC\_008609.1) and the *dsrAB* (may be in fact *dsrC* gene; see Visser *et al.*, 2013) in *Pelotomaculum thermopropionicum* SI (NC\_009454.1) (Appendix Figure D2). Both *P. propionicus* (which also lacks *dsrAB*) and *P. thermopropionicum* also lack *dsrMKJOP* and are incapable of sulfate, thiosulfate and sulfite reduction (Imachi *et al.*, 2002, Imachi *et al.*, 2006, Schink 1984). Because the *dsrC* gene product is not a subunit of DsrAB and does not necessarily always occur in a gene cluster (Pereira *et al.*, 2011), tBLASTx was conducted to search for related *dsrAB* genes from SCADC sequence assembly. Phylogenetic analysis of *dsrAB* genes recovered from SCADC indicates that most of the sequences recovered were

related to the *dsrAB* genes in methanogenic Archaea (Susanti and Mukhopadhyay 2012), similarly reported in Chapter 3; only two SCADC contigs were related to the *Desulfotomaculum* clade, and likely do not belong to the *Desulfotomaculum* SCADC bin since they both had very low sequence coverage (not shown). The function of the *dsrC* gene in “*Desulfotomaculum*” SCADC is unknown but it appears that it was expressed (FPKG>1000) under methanogenic conditions (Appendix Table D5; see therein for discussion on transcriptomic analysis of *Desulfotomaculum* SCADC). Previous studies have also reported the expression of *dsrAB*-like genes under both sulfate-reducing and methanogenic conditions by *Pelotomaculum* strain MGP, which is incapable of reducing sulfite, sulfate, or organosulfonates (Imachi *et al.*, 2006). Because the *dsrAB* gene in *P. thermopropionicum* SI (NC\_009454.1) recently has been suggested to be related to *dsrC* based on sequence comparison (Visser *et al.*, 2013), the *dsrAB* gene in strain MGP may in fact also be related to *dsrC* instead of *dsrAB*, therefore rendering strain MGP incapable of sulfite reduction. In other organisms that do not perform dissimilatory sulfate reduction such as *Escherichia coli*, enzymes encoded by genes related to *dsrC* are found to be involved in sulfur transfer for tRNA 2-thiouridine synthesis (Ikeuchi *et al.*, 2006).

In the last three years, efforts in our laboratory to isolate *Desulfotomaculum* SCADC using numerous anaerobic techniques (e.g., pasteurization followed by plating on basal mineral media with alkanes as the only carbon source) under both fermentative and sulfate-reducing condition have not been successful. This is perhaps due to the requirement for a syntrophic partner to provide compounds for mutual growth. For example, syntrophic microorganisms such as *Syntrophus aciditrophicus* SB1 (McInerney *et al.*, 2007) and *P. thermopropionicum* (Kosaka *et al.*, 2008) have reduced gene content and functionality for degrading high molecular weight compounds, and therefore require a partner in syntrophic growth for substrate degradation (Kato and Watanabe 2010). Also, we only observed enrichment of *Desulfotomaculum* in SCADC under methanogenic conditions and not when the culture was amended with sulfate (Chapter 2). Therefore, based on the existing evidence, it is tempting

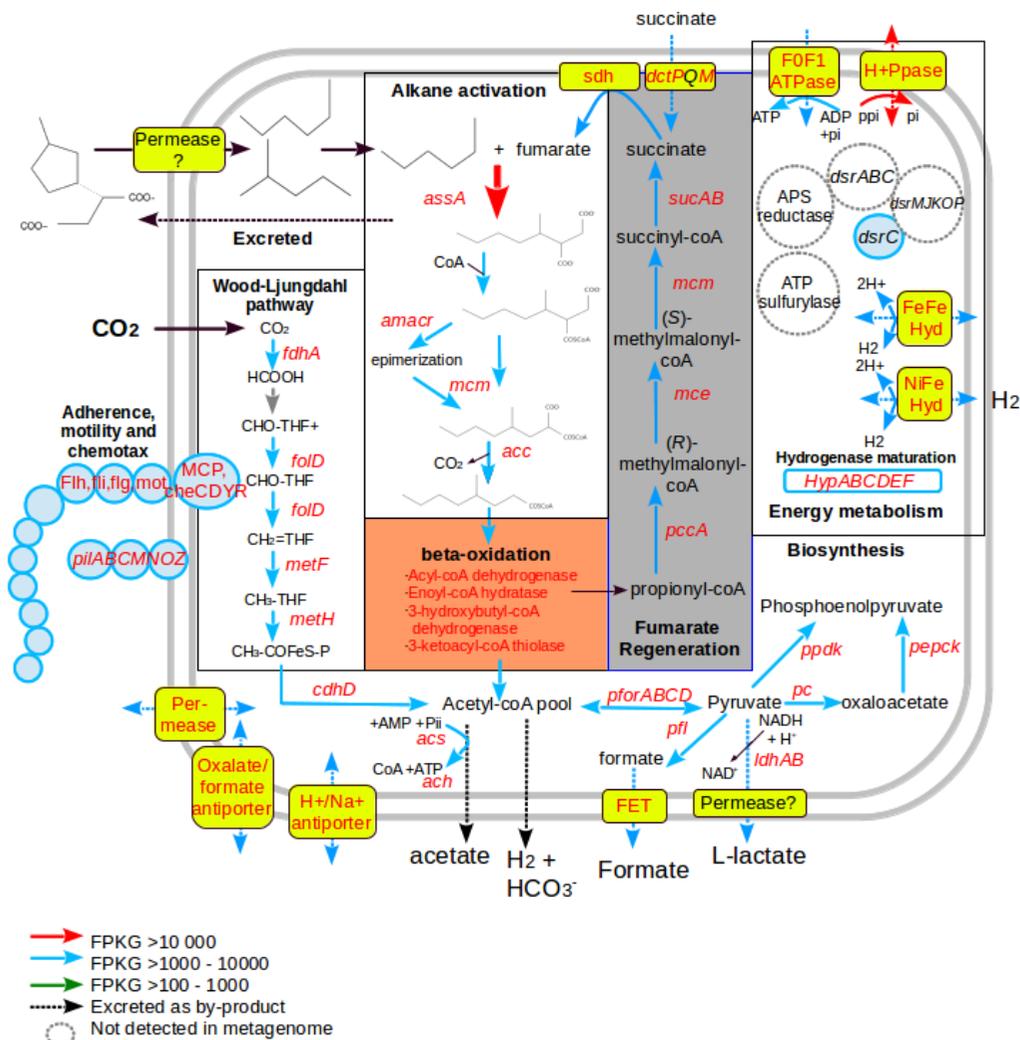
to hypothesize that the “*Desulfotomaculum*” SCADC represents a novel clade of microbe related to *Desulfotomaculum* that lacks the capability for sulfate reduction, similar to *Pelotomaculum* spp., and their physiological role in alkane degradation is through fermentative/syntrophic growth in methanogenic consortia. Nonetheless, we cannot rule out the possibility that erroneous sequence assembly, annotation and/or binning could have resulted in the exclusion of contigs harbouring the *dsrAB* and *dsrMKJOP* genes.

#### **4.5.2.3 Elucidation of anaerobic alkane degradation by “*Desulfotomaculum*” SCADC**

To investigate gene expression by “*Desulfotomaculum*” SCADC during methanogenic growth on low molecular weight alkanes, metatranscripts were mapped to all annotated open reading frames (ORFs). Genes hypothesized to be involved in the initial activation of alkanes and further degradation involving putative carbon skeleton rearrangement and decarboxylations, as well as those involved in the methylmalonyl-coA pathway proposed for fumarate regeneration, were inferred based on genome annotation in *Desulfatibacillum alkenivorans* (Callaghan *et al.*, 2012). Exceptions were *assA* and its homolog that were found to be translated during growth of *D. alkenivorans* on *n*-alkanes of under sulfate-reducing conditions (Callaghan *et al.*, 2012) and *Azoarcus* sp. HxN1 under nitrate-reducing conditions (Grundmann *et al.*, 2011). Most genes hypothesized to be involved in alkane degradation including those encoding enzymes for processes such as glycyl radical formation (*assD*), fumarate addition (*assA*), carbon skeleton rearrangement (*mcm*) and decarboxylation (*acc*) and many genes involved in beta-oxidation annotated for “*Desulfotomaculum*” SCADC were located in a single contig (contig ckmer83\_42244 in Chapter 3). A metabolic pathway with the focus on alkane addition to fumarate and regeneration of fumarate based on that proposed by Callaghan *et al.* (2012) was reconstructed for “*Desulfotomaculum*” SCADC based on transcriptomic analysis. Overall, there was high expression of putative genes encoding the proposed pathway of alkane activation by addition to fumarate, fumarate regeneration,  $\beta$ -oxidation, and genes for formation of central metabolic products such as acetyl-coA and pyruvate for

biomass synthesis were highly expressed (Figure 4.8), confirming that these processes are essential during alkane degradation.

**Figure 4.8** (Next page) Reconstruction of proposed alkane degradation pathway in *Desulfotomaculum* SCADC during active methanogenesis, based on Callaghan *et al.* (2012) and Jarling *et al.* (2012) with reference to MetaCyc (Caspi *et al.*, 2012). The pathways were reconstructed in Pathway Tools (Karp *et al.*, 2011) using annotations obtained from the RAST server (Aziz *et al.*, 2008). Metatranscripts were mapped to annotated gene fragments using CLC Genomic Workbench (CLC-Bio) and values were normalized and expressed as "fragments mapped per kilobase of gene length (FPKG)". All genes encoding enzymes involved in the initial fumarate addition mechanism and continuing to formation of acyl-coA for  $\beta$ -oxidation and metabolic by-products of acetate, formate, L-lactate and hydrogen utilization were highly expressed and are highlighted in red. Differently coloured lines denote relative gene expression as FPKG. Other abbreviations: pcc, propionyl-coA carboxylase; suc, succinyl-coA ligase; pepc, phosphoenolpyruvate carboxylase; pfor, pyruvate: ferredoxin oxidoreductase; ppdk, pyruvate phosphate dikinase/pyruvate synthase; pepck, phosphoenolpyruvate carboxykinase; pc, pyruvate carboxylase; dctPQM, C4-dicarboxylate TRAP transporter; MFT, major facilitator transporter; DMT, drug and metabolite transporter.



According to the scheme shown in Figure 4.8, drawn according to the pathways annotated and proposed for *D. alkenivorans* AK-01 (Callaghan *et al.*, 2012), low molecular weight *n*-alkanes represented by *n*-hexane are hypothesized to be first (i) activated by addition to fumarate to form (1-methylpentyl)succinic acid, catalyzed by alkylsuccinate synthase genes (*assABC*). This is followed by (ii) CoA transfer to form (1-methylpentyl)succinate-CoA, a process catalyzed by CoA ligases. The downstream degradation mechanism involves (iii) carbon skeleton rearrangement by methylmalonyl-CoA mutase (*mcm*) to form (2-methylhexyl)malonyl-CoA; followed by (iv) decarboxylation by acetyl-CoA decarboxylase (*acc*) to form 4-methyloctanyl-CoA, which can then be further degraded through  $\beta$ -oxidation. Prior to step (iii) carbon-skeleton rearrangement,

(1-methylpentyl)succinic acid diastereomers have been proposed (Jarling *et al.*, 2012) to undergo (v) epimerization, catalyzed by  $\alpha$ -methylacyl-CoA racemase (*amacr*) to form the diastereomer required in step (iii). The candidate enzyme  $\alpha$ -methylacyl-CoA racemase proposed for this epimerization process has been demonstrated to be crucial for the degradation of methyl-branched alkanes by aerobic *Mycobacterium* sp. (Sakai *et al.*, 2004), and intriguingly was located on the same contig as the *ass* operon and genes encoding  $\beta$ -oxidation in “*Desulfotomaculum*” SCADC (Chapter 3). tBLASTx searches using the *amacr* gene annotated in the current study in *Desulfotomaculum* SCADC against all annotated ORFs in *Desulfatibacillum alkenivorans* AK-01 recovered only a single homologous gene (locus tag Dalk\_0909, annotated as L-carnitine dehydratase/bile-acid inducible protein F; amino acid sequence similarity, 41%) that is located 1.5 Mbp upstream of the annotated Ak-01 *ass* operon (*ass* locus 1 and 2). That is, the *amacr* gene in AK-01 is not located within the annotated *ass* operon, and therefore is not likely directly under the same genetic control. The high expression of this gene in *Desulfotomaculum* SCADC suggests that it may be in fact essential for methanogenic growth with alkanes, although the actual function can only be ascertained with more stringent biochemical and genetic analysis.

#### **4.5.2.4 Other carbon assimilation/excretion pathways in *Desulfotomaculum* SCADC**

In addition to carbon assimilation through the mineralization of alkanes, “*Desulfotomaculum*” SCADC also likely functions as a homoacetogen in using the Wood-Ljungdahl pathway for autotrophic carbon fixation during methanogenic growth (Figure 4.8). Hydrogen uptake and electron bifurcating mechanisms likely provide in part the reducing power necessary for CO<sub>2</sub> reduction, evidently with the expression of genes encoding [Fe-Fe] hydrogenases and [NiFe] hydrogenases. The Wood-Ljungdahl pathway and genes encoding multiple hydrogenases were also detected in *D. alkenivorans* AK-01, which couples alkane oxidation to sulfate reduction and is capable of methanogenic growth with an archaeal partner (Callaghan *et al.*, 2012). The role of carbon

fixation during methanogenic alkane degradation by “*Desulfotomaculum*” SCADC is unknown; whether or not methanogenic alkane oxidation by “*Desulfotomaculum*” SCADC and *D. alkenivorans* AK-01 could be coupled to the reduction of CO<sub>2</sub> via the Wood-Ljungdahl pathway remains untested. Nonetheless, the ability to fix CO<sub>2</sub> likely becomes essential for long-term survival when a hydrocarbon source becomes limiting in the environment. Energy conservation appears to be largely reliant on proton pumping and hydrolysis of pyrophosphate by membrane-bound pyrophosphatases (H<sup>+</sup>-ppases; FPKG>10 000), and also through ATP synthesis by a proton motive force generated through the action of F1-F0 ATP synthase (Figure 4.8). The observed expression of *acs* (acetyl-coA synthetase), *ach* (acetyl-coA hydrolase), *pfl* (pyruvate formate lyase), FET (formate efflux transporter), and *ldhAB* (lactate dehydrogenase large and small subunits) indicate that the by-products of alkane degradation can include acetate, formate and possibly L-lactate (these enzymes are bi-functional and may be involved in the consumption of these metabolites under other circumstances).

#### **4.5.3 Metagenomic analysis implicates *Smithella* SCADC in alkane activation by addition to fumarate**

Microorganisms related to *Syntrophus* and the phylogenetically related species of largely uncultivated *Smithella* spp. have been routinely implicated in methanogenic degradation of crude oil and long-chain alkanes (Cheng *et al.*, 2013, Gray *et al.*, 2011, Zengler *et al.*, 1999). However, there is little non-circumstantial evidence that members of *Syntrophus/Smithella* spp. are capable of fumarate addition. In the present study, metagenomic binning recovered a partial to almost complete genome of *Syntrophus/Smithella* from the SCADC metagenome that is putatively capable of alkane addition to fumarate, evident by the detection of a single copy of *assA* gene (Figure 4.6). Mapping of the “*Smithella*” SCADC assembled reads to the genome of *S. aciditrophicus* SB1, the only cultivated representative of *Syntrophus*, indicates that these organisms do not share high similarity: <10% of the genome of *S. aciditrophicus* SB1 could be mapped with *Smithella* SCADC contigs (not shown). Transcript mapping with a overall low FPKG value (<100) for the “*Smithella*” SCADC *assA* sequence

suggests that the organism is not responsible for the activation of low molecular weight alkanes in the SCADC culture. Instead, the organism may be involved in activation of longer-chain alkanes, consistent with their frequent detection in crude oil and hexadecane-degrading enrichment cultures (Cheng *et al.*, 2013, Gray *et al.*, 2011, Zengler *et al.*, 1999). Annotation of the SCADC metagenome shows that “*Smithella*” SCADC contains multiple copies of genes encoding hydrogenase maturation (*hypABCDEF*), formate dehydrogenase (*fdh*), periplasmic Fe-only dehydrogenase and FAD-dependent dehydrogenase (*HOXEF*), but lacks genes encoding sulfate (*dsrABC*, ATP sulfurylase) and nitrate reduction (e.g. *narG*, *nirK*), as was reported for *S. aciditrophicus* SB1 (McInerney *et al.*, 2007) and *Smithella* spp. reported by Embree *et al.* (2013). The presence of multiple hydrogenases and formate dehydrogenases suggest that *Smithella* SCADC is capable of interspecies electron/hydrogen transfer and may be a facultative or obligate syntroph like its closest relative *S. aciditrophicus* SB1 (McInerney *et al.*, 2007).

Very recently, Embree *et al.* (2013) published the single cell draft genome of a *Smithella* spp. (DDBJ/EMBL/GenBank accession number AWGX00000000) from a methanogenic hexadecane-degrading culture (Zengler *et al.*, 1999). Contrary to our findings reported here, the authors reported that *Smithella* spp. is not capable of fumarate activation of *n*-hexadecane because the *assA* gene was not detected. In order to investigate this claim, tBLASTn was conducted to manually screen the *Smithella* spp. genome reported by Embree *et al.* (2013) using several annotated *assA* genes as probes; This revealed the presence of seven genes encoding putative glycyl radical enzymes including alkylsuccinate synthase (Appendix A). Phylogenetic analysis of the recovered genes indicated that the *Smithella* spp. draft genome reported by Embree *et al.* (2013) does indeed contain a single copy of full-length *assA* (AWGX1000974) which was not annotated in the genome deposited in GenBank. Further evidence by re-analysis of the metatranscriptome reported by Embree *et al.* (2013), discussed in Appendix A, confirm that *Smithella* spp. is indeed capable of alkane activation of long-chain alkanes by addition to fumarate under methanogenic conditions.

#### 4.5.4 Aceticlastic and hydrogenotrophic methanogenesis

Hydrocarbon degradation with methanogenesis often yields substantial amounts of acetate and formate in addition to CO<sub>2</sub> and H<sub>2</sub> (Edwards and Grbić-Galić 1994, Mbadanga *et al.*, 2012). Formation of acetate from acetyl-CoA has been a favourable strategy by fermentative bacteria to obtain adenosine ATP for energy conservation. Thus, methanogenic hydrocarbon-degrading enrichment cultures can be dominated by both aceticlastic and hydrogenotrophic methanogens (Chapter 5). The diversity and proportion of both classes of methanogens likely differ depending on the source of inoculum used in establishing these cultures and thermodynamic requirements of community members. For example, in methanogenic crude oil-degrading cultures, the increase in hydrogenotrophic methanogens relative to aceticlastic methanogens has been attributed to the action of syntrophic acetate oxidizers (SAO) oxidizing acetate to yield CO<sub>2</sub> and H<sub>2</sub> for hydrogenotrophic methanogens (Gray *et al.*, 2011). Other studies indicate that non-CO<sub>2</sub>/H<sub>2</sub>-utilizing methanogens are equally important in methanogenic alkane-degrading cultures (Callaghan *et al.*, 2010, Zhou *et al.*, 2012). Acetate oxidation by syntrophic acetate oxidizers (SAO) such as Clostridiales is likely to have less effect on the overall CO<sub>2</sub>/H<sub>2</sub> contribution since transcriptomic analysis did not detect such organisms in abundance in SCADC (Figure 4.7). Other organisms, e.g., “*Smithella*” SCADC (see below) are likely to be homoacetogens and to compete with the hydrogenotrophic methanogens by converting CO<sub>2</sub> into formate using formate dehydrogenase. However, this mechanism is not known and can only be ascertained using more conclusive methods such as <sup>13</sup>C-acetate stable isotope probing.

##### 4.5.4.1 Roles of *Methanoculleus* SCADC

*Methanoculleus* are hydrogenotrophic methanogens and are capable of using H<sub>2</sub>/CO<sub>2</sub> or formate as the primary energy source in the presence of acetate as carbon source for growth (Dianou *et al.*, 2001, Mikucki *et al.*, 2003, Zellner *et al.*, 1998), with exceptions such as *Methanoculleus hydrogenitrophicus* that does not use acetate or formate (Tian *et al.*, 2010). In the current study, pyrotag sequencing of 16S rRNA gene amplified from translated cDNA and mapping of

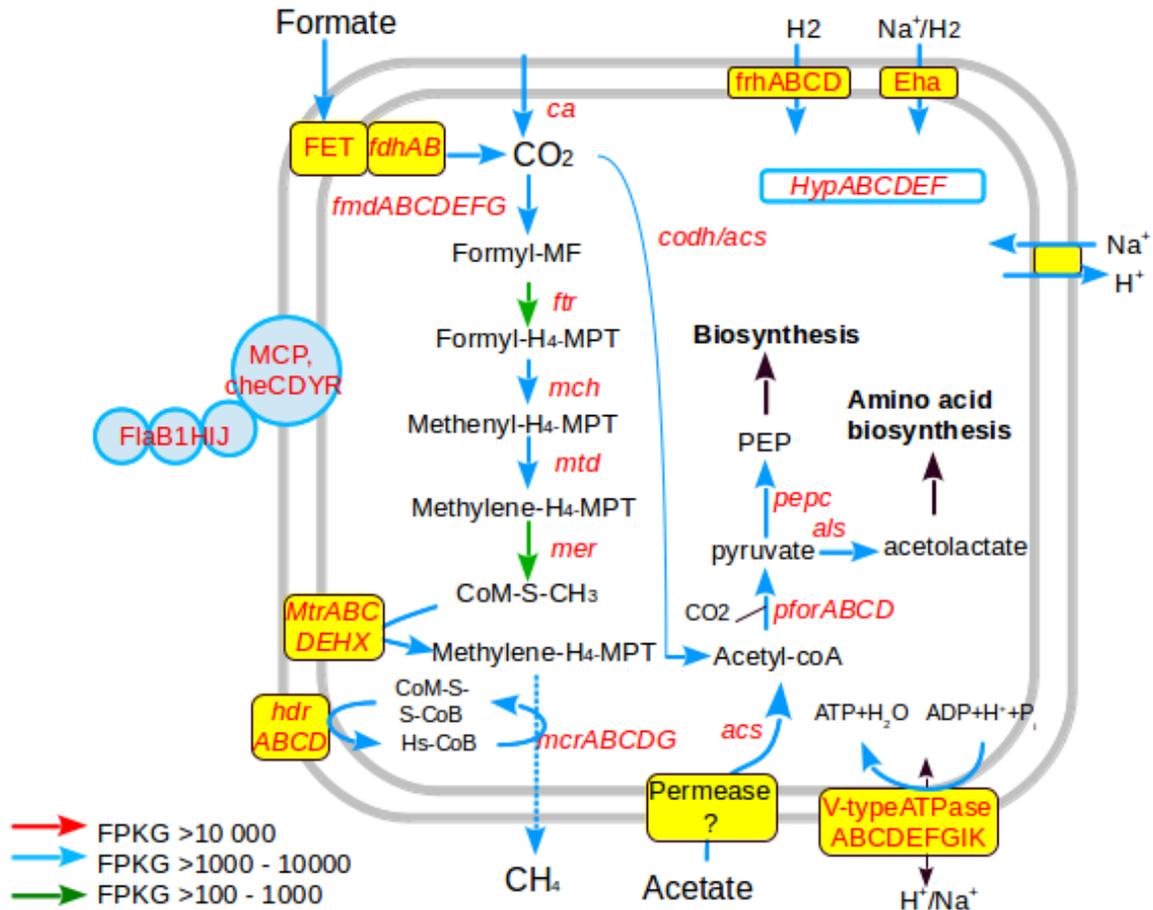
metatranscripts to *mcrA* genes indicate that hydrogenotrophic methanogens affiliated with *Methanoculleus* are the most metabolically active, followed by acetoclastic *Methanosaeta* (Figure 4.7). Thus, the main route of methanogenesis is likely through CO<sub>2</sub>/H<sub>2</sub>-utilizing hydrogenotrophic methanogens. Acetoclastic methanogenesis is likely secondary in SCADC since the *mcrA* gene of *Methanosaeta* is less dominant in the SCADC metatranscriptome (most *Methanosaeta* spp. is obligate acetoclastic methanogens). Although the two *Methanoculleus*-related species in SCADC could be separated using sequence coverage (Figure 4.3), there appears to be mixing of some contigs between *Methanoculleus* 1 and 2 (Table 4.1), which were inseparable using tetranucleotide frequency binning (not shown). Thus, these two metagenomic bins were merged into single genome and discussed below (Figure 4.9; Appendix Table D6 and D7).

Based on RNA-seq analysis of SCADC *Methanoculleus*, the main route of methanogenesis is likely through CO<sub>2</sub> uptake and/or conversion into bicarbonate by carbonic anhydrase (encode by *ca*; Figure 4.9) and also through oxidation of formate to CO<sub>2</sub> by formate dehydrogenase (*fdhAB*). The expression of multiple copies of acetyl-coA synthetase (*acs*) indicates that acetate likely is being converted to acetyl-coA and used as a carbon source for growth but not for energy generation and methane production (Dianou *et al.*, 2001, Mikucki *et al.*, 2003, Zellner *et al.*, 1998). CO<sub>2</sub> produced from formate oxidation could also be converted to acetyl-coA by carbon-monoxide dehydrogenase/acetyl-coA synthase (*codh*). All genes involved in methane formation from CO<sub>2</sub> (*ptr*, *mch*, *mtd*, *mer*) and various steps of hydrogen uptake (*hdr*, *frh* and *eha*), and the last step of methanogenesis involving the reduction of the coenzyme M-bound methyl group to methane by methyl-coenzyme reductase (*mcrABCDG*) were also highly expressed.

Surprisingly, genes encoding enzymes for depolymerisation of complex carbohydrates (i.e.,  $\alpha$ -amylase and 4- $\alpha$ -glucanotransferase) were also highly expressed (FPKG>1000). Genes annotated as amylases were also found in the genomes of some methanogens such as *Methanoculleus bourgensis* MS2

(YP\_006545908.1) and *Methanoculleus marisnigri* (YP\_001046301.1), although their functions may not be related to hydrolysis of carbohydrates since these polymers were absent in the SCADC enrichment culture and methanogens are not known to utilize or hydrolyze polymers.

**Figure 4.9** (Next page) Reconstruction of methanogenesis pathway in *Methanoculleus* SCADC during active methanogenesis. Pathways were reconstructed and predicted using Pathway Tools (Karp *et al.*, 2011) and with reference to (Hendrickson *et al.* (2007), Walker *et al.* (2012) and MetaCyc (Caspi *et al.*, 2012). Metatranscripts were mapped to annotated gene fragments using CLC Genomic Workbench (CLC-Bio) and values were normalized and expressed as "fragments mapped per kilobase of gene length (FPKG)". All genes related to hydrogenotrophic methanogenesis were highly expressed and are highlighted in red. Differently coloured lines denote relative gene expression as FPKG. **Abbreviations:** ca, carbonic anhydride; fdh, formate dehydrogenase; mtd, F<sub>420</sub>-dependent methylene-tetrahydromethanopterin dehydrogenase; ftr, formylmethanofuran-tetrahydromethanopterin-N-formyltransferase; mch, N(5),N(10)-methenyl-tetrahydromethanopterin cyclohydrolase; mer, N(5),N(10)-methylene-tetrahydromethanopterin reductase; hdr, heterodisulfide reductase; als, acetolactate synthase; frh, F<sub>420</sub>-dependent hydrogenase; fhc, formylmethanofuran-tetrahydromethanopterin formyltransferase; ppaC, manganese-dependent inorganic pyrophosphatase



Expression of genes encoding putative permeases annotated for transport of Vitamin B12, ferrous iron, molybdenum, zinc and tungsten (FPKG>1000), indicate that these compounds may be essential for growth, and are probably used as co-factors in protein synthesis (e.g., biosynthesis of tungsten-containing formylmethanofuran dehydrogenase). *Methanoculleus* SCADC also appears to be obtaining nitrogen from headspace N<sub>2</sub> gas with the expression of nitrogenase ( $\alpha$ - and  $\beta$ -chain, *nifEHN*, FPKG>1000).

#### 4.5.4.2 Roles of *Methanosaeta* SCADC

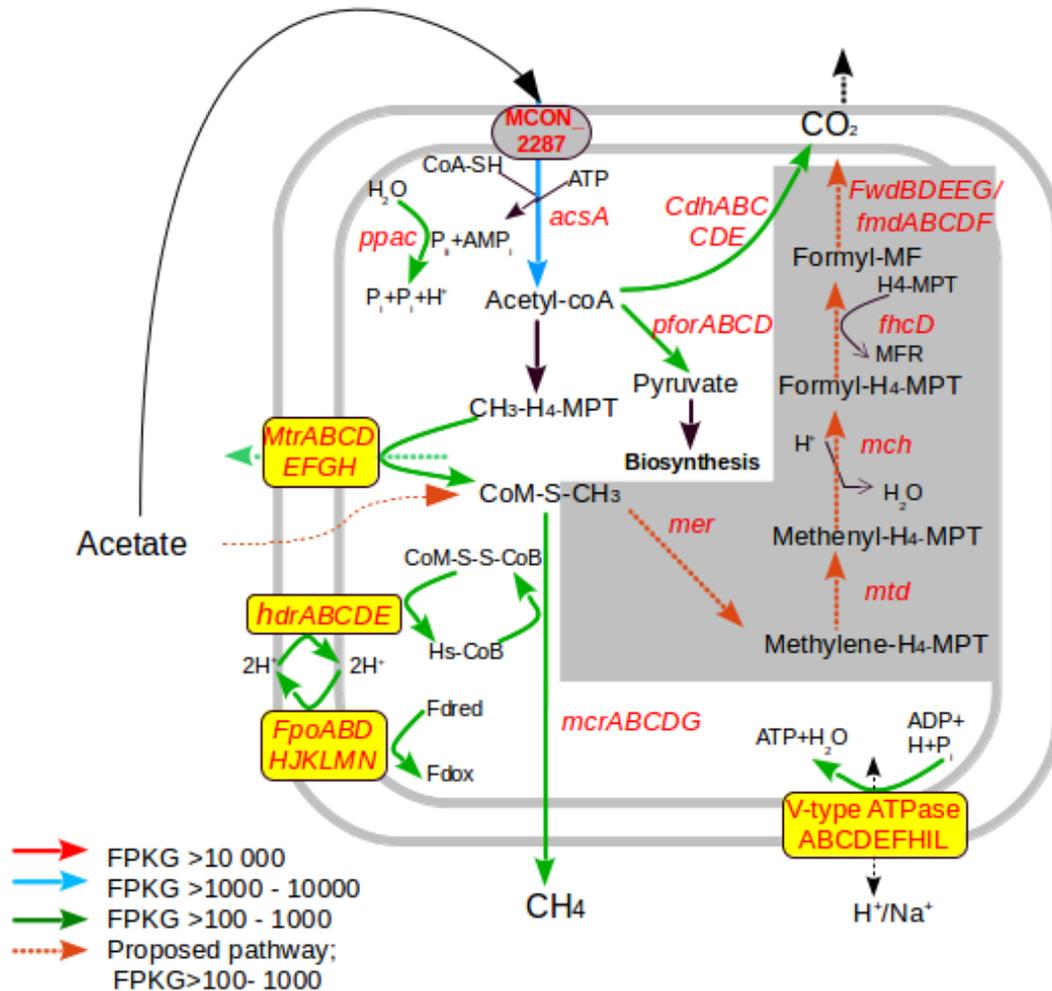
The active *Methanosaeta* in SCADC is closely related to *Methanosaeta concilii* GP6 (Chapter 3), whose genome (NC\_015416) has been sequenced. Mapping of the SCADC sequence assembly to *M. concilii* GP6 showed >80% sequence coverage with >98% similarity (Chapter 3), which is supported by

>98% similarity in 16S rRNA and *mcrABG* genes (not shown). Almost the entire genomic region of *M. concilii* GP6 containing ORFs could be mapped with SCADC assembled contigs, with the exception of regions encoding transposases and tandem repeats, suggesting that *Methanosaeta* SCADC is highly related to strain GP6 but may have undergone an active genetic reorganization (see below). *Methanosaeta* spp. are considered to be obligate acetoclastic methanogens (Ma *et al.*, 2006, Patel and Sprott 1990). However, the genomes of *M. concilii*, *M. thermophila* and *M. harundinacea* 6Ac contain genes encoding both acetoclastic (*acsA*, *cdhABC*, *mtr*) and hydrogenotrophic/methylotrophic (*fwd*, *fmd*, *mch*, *mtd*, *mer*) methanogenesis pathways (Zhu *et al.*, 2012). Phylogenetically, genes encoding the hydrogenotrophic pathway appear to be related to those found in methylotrophic methanogens rather than hydrogenotrophic methanogens (Zhu *et al.*, 2012) even though these three *Methanosaeta* are not known to use methylotrophic substrates, e.g., methylamine, as carbon or energy sources (Ma *et al.*, 2006). In the current study, RNA-seq using the annotated genome of *M. concilii* GP6 as a reference showed that SCADC genes involved in both acetoclastic methanogenesis (Figure 4.10; solid green lines) and methyl group oxidation (Figure 4.10; dotted orange line; FPKG>100) were highly expressed, especially the *acs* gene encoding acetyl-coA synthetase for conversion of acetate to acetyl-coA (FPKG >1000), indicating that acetate is being used for methanogenesis. Genes encoding carbon monoxide dehydrogenase (*cdh*) were also highly expressed, consistent with physiological evidence of *M. concilii* being able to convert acetate into nearly equimolar amounts of CO<sub>2</sub> and methane (Patel and Sprott 1990). The CO<sub>2</sub> produced from acetoclastic methanogenesis in turn may be utilized by *Methanoculleus* SCADC for methane generation (Figure 4.6).

When grown with acetate, *M. harundinaceae* 6Ac has been shown to express genes necessary for both acetoclastic and a proposed methylotrophic pathway but it has been suggested, based on <sup>13</sup>C-acetate experimentation, that the methylotrophic pathway is used as carbon shunt to generate essential reduced ferredoxin or F<sub>420</sub> for biomass synthesis, instead of CO<sub>2</sub> conversion to methane (Zhu *et al.*, 2012). This, however, accounts for <1% of total CO<sub>2</sub> generated from

<sup>13</sup>C-acetate (Zhu *et al.*, 2012). The expression of genes encoding both acetoclastic and hydrogenotrophic pathways by *M. concilii* was similarly reported by Embree *et al.* (2013) in methanogenic cultures grown with hexadecane. Nevertheless, the functions of genes encoding the methyl-oxidation pathway remain to be elucidated in the literature.

**Figure 4.10** (Next page) Reconstruction of methanogenesis pathway by *Methanosaeta* SCADC during active methanogenesis. The genome of *Methanosaeta concilli* GP6 (NC\_05416) was used in RNA-seq (see main text) to map metatranscripts to annotated gene fragments using CLC Genomic Workbench (CLC-Bio); values were normalized and expressed as "fragments mapped per kilobase of gene length (FPKG)". All genes related to methanogenesis were highly expressed (most genes encoding metabolic functions had FPKG<1000, with the exception of *acs*) and are highlighted in red. Differently coloured lines denote relative gene expression as FPKG. The metabolic shunt for acetate conversion to CO<sub>2</sub> (<1% total conversion) proposed by Zhu *et al.*, (2012) is shaded in grey and the pathway is shown in dotted orange arrows (all genes have FPKG between 100 and 1000).



#### 4.5.4.3 Other methanogenesis pathways

Methylotrophic methanogenesis is likely to have negligible contribution to methane production in SCADC because of the lack of representative methylotrophic methanogens, e.g., Methanosarcinales, in SCADC (Chapter 3). Genes encoding the methylotrophic pathways for methylamine (*mtbABC*) and methanol (*mttABC*) conversion to methyl-S-CoM were recovered from the SCADC metagenome using KEGG Orthologues (KO) but not the genes involved in conversion of di- and tri-methylamine (*mtmBC* and *mtbBC*, not shown). BLASTX searches of the putative *mtbABC* and *mttABC* sequences against the NCBI nr-database revealed that these sequences were more related to non-*mtbABC* sequences; for example, *mtbA* to uroporphyrinogen decarboxylase.

Overall, bioinformatic analysis of RNA-seq for both *Methanoculleus* SCADC and *Methanosaeta* SCADC (Figure 4.9 and 4.10) shows the expression of genes encoding various membrane-bound hydrogenases, e.g., *hdrABCDE* and *frhABCD*, consistent with the requirement for regenerating reducing power during methanogenic growth. However, these different membrane bound hydrogenases are likely to be involved in very different mechanisms. For example, it has been suggested that Eha in *Methanococcus* functions primarily in methanogenesis, whereas Ehb is involved in the generation of a low potential electron carrier for anabolic processes (i.e., formation of acetyl-coA for biomass synthesis). Furthermore, the functions of some of the hydrogenases are currently unclear. For example, in *Methanosaeta*, CoM-CoB produced by the reduction of methyl Coenzyme M, shown in Figure 4.10, has been suggested to be reduced by the heterodisulfide reductase (encoded by *hdrABCD*) for regeneration of redox potential. The electron donor for regenerating the heterodisulfide reductase complex is currently unknown but it has been proposed that the F<sub>420</sub> H<sub>2</sub>-hydrogenases (encoded by *fpoABCDHIJKLMN*) might be involved in this (Welte and Deppenmeier 2011). Although we observed the expression of all Fpo subunits (FPKG>100, with the exception of *fpoO*, FPKG=0), the physiological role of the Fpo complex can only be determined using a pure culture.

The transcriptome of many methanogens can appear to be very different depending upon the growth conditions and thermodynamic requirements, and therefore expression can vary depending on whether they are grown in pure culture or defined co-culture (Hendrickson *et al.*, 2007, Walker *et al.*, 2012). The transcriptome presented in this Chapter can be expected to help in further experimentation to understand the complex roles of methanogens in hydrocarbon-degrading environments.

## **4.5.5 Syntrophic interactions during alkane degradation**

### **4.5.5.1 Hydrogen/electron and formate transfers in SCADC**

Syntrophic degradation of high molecular weight compounds under methanogenic conditions often involves interspecies hydrogen/electron and

formate transfers (Kato and Watanabe 2010, Stams and Plugge 2009). Evidently, genes encoding these processes such as hydrogen/electron transfers (i.e.,  $F_{420}$ -dependent cytoplasmic Ni-Fe and Fe-Fe hydrogenases) and formate transfer (i.e., formate dehydrogenase) were highly expressed by “*Desulfotomaculum*” SCADC (Figure 4.8). The expression of these hydrogenases suggests that *Desulfotomaculum* SCADC likely either produces  $H_2$  to dispose of excess reductant from alkane oxidation ( $H_2$  sink) or  $H_2$  is consumed and used to produce NADPH for anabolic metabolism or as a reductant in  $CO_2$  fixation in the Wood-Ljungdahl pathway.

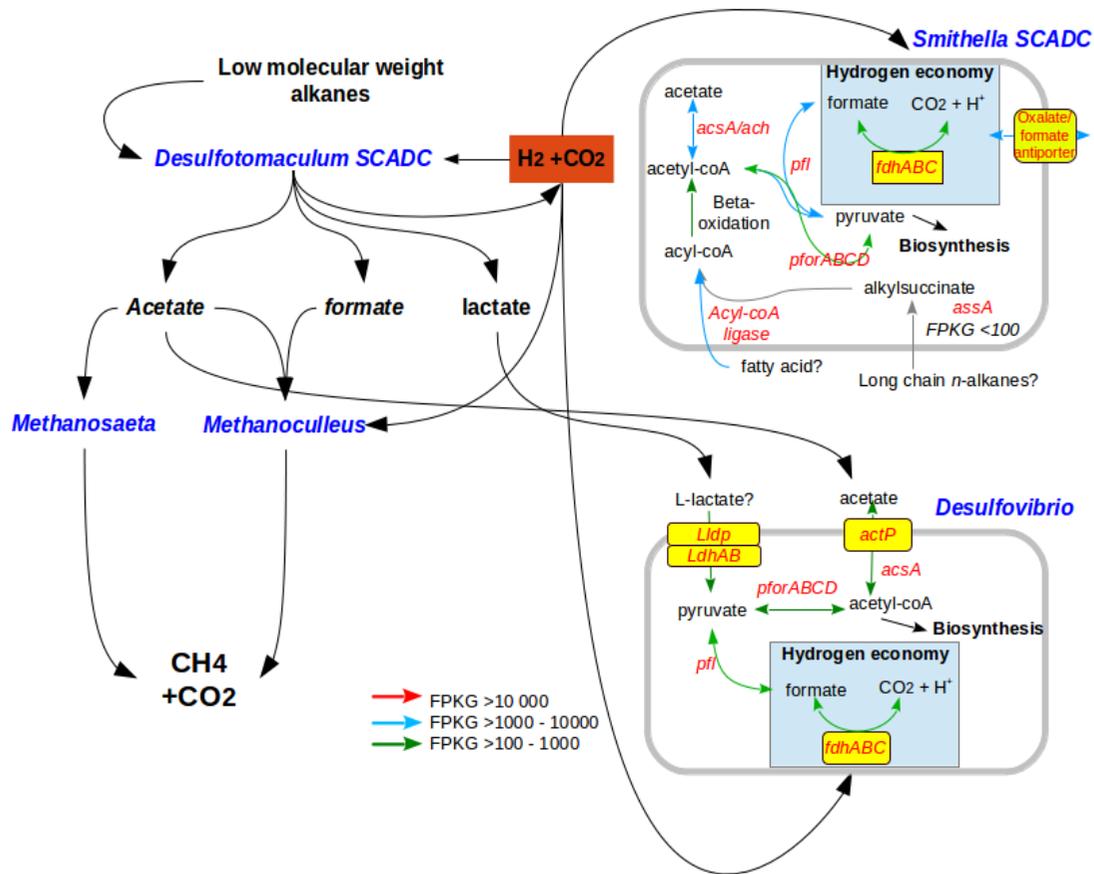
In fact, organisms capable of syntrophic growth such as *D. alkenivorans* AK-01 (Callaghan *et al.*, 2012) and *P. thermopropionicum* (Imachi *et al.*, 2002) often carry multiple genes encoding hydrogenases and formate dehydrogenases. Some anaerobes can also carry genes encoding terminal electron-accepting process such as sulfate and sulfite reduction genes. This allows them to be metabolically flexible when grown under different conditions. For example, in the absence of sulfate, the sulfate-reducing bacterium *Desulfovibrio vulgaris* Hildenborough reduces  $CO_2$  with  $H_2$  using a periplasmic formate dehydrogenase to produce formate. The formate is then proposed to either act as a hydrogen storage compound or to be converted to pyruvate by pyruvate formate lyases acting in reverse (da Silva *et al.*, 2013) (Figure 4.11).

The co-expression of both hydrogenases and formate dehydrogenases by syntrophs can be important to sustain methanogenic growth of all community members. For example, when grown in pure culture or in co-culture with *Methanospirillum hungatei* SB1 on either benzoate or cyclohexane-1-carboxylate, *Syntrophomonas wolfei* and *S. aciditrophicus* SB1 co-expressed genes encoding multiple copies of hydrogenases and formate dehydrogenases; notably, these genes were only highly expressed in co-culture (Sieber *et al.*, 2013). Similar observations have also been reported for *Syntrophobacter fumaroxidans* grown in the presence of *M. hungatei*, whereby the former co-expressed hydrogenases and formate dehydrogenase during syntrophic growth (Worm *et al.*, 2011). In SCADC, the hydrogen producers (e.g.,

“*Desulfotomaculum*” SCADC) and consumers (hydrogenotrophic methanogens and homoacetogens) can have different thermodynamic requirements and hydrogen thresholds (Kotsyurbenko *et al.*, 2001). Therefore, the syntrophic relationships developed in SCADC likely depend on efficiency in H<sub>2</sub>/CO<sub>2</sub> sequestration that satisfies the energy requirements of both hydrogenotrophic methanogens and homoacetogens.

#### **4.5.5.2 Metabolite transfer in SCADC at a glance**

According to the scheme in Figure 4.11, CO<sub>2</sub> + H<sub>2</sub> produced from alkane oxidation is hypothesized to be the main carbon and energy source for syntrophs (homoacetogens) in SCADC. *Desulfovibrio*, for example, is hypothesized to use lactate produced by “*Desulfotomaculum*” SCADC, evident by the expression of lactate permease and lactate dehydrogenase (FPKG>100, not shown). The organism may also be capable of converting CO<sub>2</sub> into formate with the enzyme formate dehydrogenase (*fdhABC*), as proposed by da Silva *et al.*, (2013). “*Smithella*” SCADC, which is hypothesized to be capable of activating only longer-chain alkanes, acts as a syntroph in SCADC by either assimilating acetate (with acetyl-coA synthetase/acetyl-coA hydrolase) to form acetyl-coA for biomass synthesis, or reduces CO<sub>2</sub> using cytoplasmic formate dehydrogenase (*fdhABC*) to form formate, which is then converted to pyruvate using pyruvate formate lyases or through the Wood-Ljungdahl pathway (FPKG>100, not shown). Overall, metatranscriptomics of SCADC provides evidence of interspecies metabolite/H<sub>2</sub>/electron transfer involving the primary degrader (*Desulfotomaculum* SCADC), homoacetogens, and hydrogenotrophic and acetoclastic methanogens.



**Figure 4.11** Possible interspecies metabolite and hydrogen/formate transfer among the SCADC microbial community. Metatranscripts were mapped to the genomic bins of *Smithella* SCADC and *Desulfovibrio* SCADC (not shown) using CLC Genomic Workbench (CLC-Bio); values were normalized and expressed as "fragments mapped per kilobase of gene length (FPKG)"

#### 4.5.5.3 Other syntrophic interactions in the SCADC community

Exchange of metabolites during syntrophic growth is not limited to interspecies hydrogen/electron or formate transfer. For example, a defined syntrophic co-culture of *Desulfovibrio vulgaris* and *Methanococcus marispludis* has been suggested to involve alanine exchange, whereby the methanogen benefits by using the alanine produced by *D. vulgaris* as a carbon and nitrogen source (Walker *et al.*, 2012). Similarly, *Dehalococcoides*, which requires cobalamin for dechlorinating processes, benefits when grown syntrophically with

bacteria that synthesize cobalamin *de novo* (He *et al.*, 2007). The high expression in SCADC of genes encoding various permeases implies the uptake/export of compounds, although the nature of metabolite transfers, e.g., valine and cobalamin, can only be verified by a differential transcriptomic analysis or experimentation with defined co-cultures.

Finally, direct electron transfer (DET) could also be involved in syntrophic degradation of substrates (Sieber *et al.*, 2012). This is typically achieved by formation of extracellular structures that connect members of different species in a network, e.g., for metabolite exchange. Genes encoding multiple subunits of pili and flagella, implicated in formation of nanowires and cell aggregates for electron transfer (Leang *et al.*, 2010) and syntrophic metabolite exchange (Shimoyama *et al.*, 2009), respectively, were highly expressed by “*Desulfotomaculum*” SCADC (Figure 4.8). However, genes encoding the outer membrane *c*-type cytochrome system necessary for electron transfer from nanowires (Leang *et al.*, 2010) were not detected in “*Desulfotomaculum*” SCADC (with the exception of an ORF annotated as cytochrome *c*-type biogenesis protein *dsbD*, which is not known to be involved in electron transfer). The formation of flagella is associated with motility and may not be involved in aggregate formation. Due to the nature of the transcriptomics study presented in the current work, it is not possible to definitely conclude whether syntrophy in SCADC involves direct electron transfer or formation of cell aggregates for metabolite exchange. However, further physiological studies using SCADC as a model system can be expected to provide insight into syntrophic degradation of hydrocarbons.

#### **4.5.6 Expression of genes encoding transposases and prophage activation in SCADC**

In the current study, mapping of metatranscripts to the annotated genomic bin of “*Desulfotomaculum*” SCADC showed that most genes on a single contig encoding prophage activation had FPKG values >10 000 (Appendix Table D5 and text, therein), indicating very high expression. Other than genes associated with prophage, many non-phage ORFs such as one annotated as putative

ferredoxin, were also highly expressed. Concatenated sequences in the “*Desulfotomaculum*” SCADC bin were submitted to PFAST (Zhou *et al.*, 2011) for prophage prediction. PFAST predicted that more than half of contig\_7984 (28,975 of 45,971 bp) encodes an intact prophage and contains genes encoding phage plate protein, tail tape measure protein, phage tail sheath protein, major head protein, phage portal protein, terminase and other putative prophage proteins (Appendix Figure D3). Searches of the entire sequence of the predicted prophage against the NCBI nr-database and Viral genome database using discontinuous MEGABLAST did not return any hits, indicating that the genome of this phage may not have been sequenced yet.

Due to the abundance of “*Desulfotomaculum*” SCADC prophage-activating transcripts in the SCADC metatranscriptome, we further explored the possibility of transfer of essential genes through prophage activation by performing BLASTx searches of all ORFs along this contig against the nr-database. Based on annotations obtained from RAST and BLAST search results, all genes on this contig outside the predicted boundary for the prophage genome were assigned as hypothetical proteins but had best BLASTp hits to members of the Peptococcaceae (i.e., *Desulfotomaculum*). Previous metagenomic and metatranscriptomic study of a vinyl chloride-degrading culture (KB-1) suggested that starvation due to limited electron acceptor (vinyl chloride) could induce prophage activation in the primary vinyl chloride degrader (*Dehalococcoides*) (Waller *et al.*, 2012). In the same study, however, the authors found no evidence that the *tce* gene encoding dechlorination was packaged in viral particles even though the *tce* gene was located in close proximity to the prophage gene, indicating that the essential gene may not be transferred within the community through prophage activation. The reason for high expression of prophage genes in “*Desulfotomaculum*” SCADC is unknown, and may reflect a stress response since an abundance of transcripts (FPKG generally >2000) also mapped to genes annotated as stage II, III, IV, V sporulation proteins and phosphate starvation-inducible protein PhoH.

Expression of some transposase-encoding genes was also observed, with transcripts mapping to the ORF of *Methanosaeta concilii* GP6 (Appendix Table D8). The genome of *M. concilii* GP6 harbours at least 110 copies of putative transposases, with many occurring as tandem repeats. The active transcription of genes encoding putative mobile genetic elements by *M. concilii* SCADC implies active genetic rearrangements during methanogenic growth. Events such as prophage activation and transcription of integrase genes by “*Desulfotomaculum*” SCADC could allow acquisition of essential genes for processes such as anaerobic degradation of pollutants, of which the importance of integron in transferring hydrocarbon degrading genes in Sydney Tar Pond was inferred based on cloning and sequencing of integron gene cassette (Koenig *et al.*, 2009). The expression of genes encoding mobile genetic elements has been detected under *in situ* conditions and therefore could facilitate increased genetic variation and capability, such as demonstrated by metatranscriptomic analysis of microbial communities in marine (Hewson *et al.*, 2009) and hydrothermal vent environments (Sanders *et al.*, 2013).

#### **4.6 Conclusions**

Microorganisms capable of alkane degradation under methanogenic conditions have proven difficult to isolate for characterization since they are likely obligate syntrophs that require a methanogenic partner. In addition, signature metabolites from anaerobic *n*-alkane degradation have never been detected under methanogenic conditions. Therefore, the mechanism(s) for activating alkanes under methanogenic condition remains cryptic. In Chapter 2, we showed that low molecular weight alkanes (*n*-alkanes and an *iso*-alkane) were degraded through addition to fumarate. Based on this, we proposed that the SCADC community activates alkanes through addition to fumarate. In this chapter, we were able to obtain the partial genome of a novel “*Desulfotomaculum*” SCADC and “*Smithella*” SCADC. Metatranscriptomic analysis suggested that the “*Desulfotomaculum*” SCADC is highly active and “*Smithella*” SCADC is not. High expression of genes encoding alkane addition to fumarate, hydrogen/electron and formate transfer, and prophage activation by

“*Desulfotomaculum*” SCADC, indicates that “*Desulfotomaculum*” SCADC is highly active and the population is possibly involved in active genetic exchange or genome reorganization. The expression of genes encoding acetoclastic and hydrogenotrophic pathways indicates that the route of methanogenesis is primarily through CO<sub>2</sub>/H<sub>2</sub> reduction and secondary through acetate conversion to CO<sub>2</sub> and methane.

Furthermore, by reanalyzing the draft genome of *Smithella* spp. and metatranscriptomes obtained from cultures grown under different growth conditions reported by Embree *et al.* (2013), we further confirmed our conclusion in Chapter 2, 3 and 4 that alkane activation by addition to fumarate can occur under methanogenic conditions. Finally, understanding the methanogenic alkane-degradation process involving multiple species can be expected to improve strategies for bioremediation of anoxic sites and management of oil sands tailings ponds.

#### 4.7 References

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

Aitken CM, Jones DM, Maguire MJ, Gray ND, Sherry A, Bowler BFJ *et al.*, (2013). Evidence that crude oil alkane activation proceeds by different mechanisms under sulfate-reducing and methanogenic conditions. *Geochim Cosmochim Acta* **109**: 162-174.

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

Callaghan AV, Gieg LM, Kropp KG, Suflita JM, Young LY. (2006). Comparison of mechanisms of alkane metabolism under sulfate-reducing conditions among two bacterial isolates and a bacterial consortium. *Appl Environ Microbiol* **72**: 4274-4282.

Callaghan AV, Wawrik B, Chadhain SMN, Young LY, Zylstra GJ. (2008). Anaerobic alkane-degrading strain AK-01 contains two alkylsuccinate synthase genes. *Biochem Biophys Res Commun* **366**: 142-148.

Callaghan AV, Davidova IA, Savage-Ashlock K, Parisi VA, Gieg LM, Suflita JM *et al.*, (2010). Diversity of benzyl- and alkylsuccinate synthase genes in

hydrocarbon-impacted environments and enrichment cultures. *Environ Sci Technol* **44**: 7287-7294.

Callaghan AV, Morris BEL, Pereira IAC, McInerney MJ, Austin RN, Groves JT *et al.*, (2012). The genome sequence of *Desulfatibacillum alkenivorans* AK-01: a blueprint for anaerobic alkane oxidation. *Environ Microbiol* **14**: 101-113.

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

Caspi R, Altman T, Dreher K, Fulcher CA, Subhraveti P, Keseler IM *et al.*, (2012). The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. *Nucleic Acids Res* **40**: D742-D753.

Chalaturnyk RJ, Scott JD, Ozum B. (2002). Management of oil sands tailings. *Petroleum Science and Technology* **20**: 1025-1046.

Cheng L, Ding C, Li Q, He Q, Dai LR, Zhang H. (2013). DNA-SIP reveals that Syntrophaceae play an important role in methanogenic hexadecane degradation. *Plos One* **8**.

da Silva SM, Voordouw J, Leitao C, Martins M, Voordouw G, Pereira IA. (2013). Function of formate dehydrogenases in *Desulfovibrio vulgaris* Hildenborough energy metabolism. *Microbiology* **159**: 1760-1769.

Davidova IA, Gieg LM, Nanny M, Kropp KG, Suflita JM. (2005). Stable isotopic studies of *n*-alkane metabolism by a sulfate-reducing bacterial enrichment culture. *Appl Environ Microbiol* **71**: 8174-8182.

Dianou D, Miyaki T, Asakawa S, Morii H, Nagaoka K, Oyaizu H *et al.*, (2001). *Methanoculleus chikugoensis* sp nov., a novel methanogenic archaeon isolated from paddy field soil in Japan, and DNA-DNA hybridization among *Methanoculleus* species. *Int J Syst Evol Microbiol* **51**: 1663-1669.

Dick GJ, Andersson AF, Baker BJ, Simmons SL, Thomas BC, Yelton AP *et al.*, (2009). Community-wide analysis of microbial genome sequence signatures. *Genome Biol* **10**: R85.

Edwards EA, Grbić-Galić D. (1994). Anaerobic degradation of toluene and *o*-xylene by a methanogenic consortium. *Appl Environ Microbiol* **60**: 313-322.

Ehrenreich P, Behrends A, Harder J, Widdel F. (2000). Anaerobic oxidation of alkanes by newly isolated denitrifying bacteria. *Arch Microbiol* **173**: 58-64.

Embree M, Nagarajan H, Movahedi N, Chitsaz H, Zengler K. (2013). Single-cell genome and metatranscriptome sequencing reveal metabolic interactions of an

alkane-degrading methanogenic community. *ISME Journal*. Advance online publication.

Gray ND, Sherry A, Grant RJ, Rowan AK, Hubert CRJ, Callbeck CM *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 *et al.*, (2008). Genes encoding the candidate enzyme for anaerobic activation of *n*-alkanes in the denitrifying bacterium, strain HxN1. *Environ Microbiol* **10**: 376-385.

Haroon MF, Hu SH, Shi Y, Imelfort M, Keller J, Hugenholtz P *et al.*, (2013). Anaerobic oxidation of methane coupled to nitrate reduction in a novel archaeal lineage. *Nature* **500**: 567.

He J, Holmes VF, Lee PK, Alvarez-Cohen L. (2007). Influence of vitamin B12 and cocultures on the growth of *Dehalococcoides* isolates in defined medium. *Appl Environ Microbiol* **73**: 2847-2853.

Hendrickson EL, Haydock AK, Moore BC, Whitman WB, Leigh JA. (2007). Functionally distinct genes regulated by hydrogen limitation and growth rate in methanogenic Archaea. *PNAS* **104**: 8930-8934.

Hewson I, Poretsky RS, Dyhrman ST, Zielinski B, White AE, Tripp HJ *et al.*, (2009). Microbial community gene expression within colonies of the diazotroph, *Trichodesmium*, from the Southwest Pacific Ocean. *ISME J* **3**: 1286-1300.

Ikeuchi Y, Shigi N, Kato J, Nishimura A, Suzuki T. (2006). Mechanistic insights into sulfur relay by multiple sulfur mediators involved in thiouridine biosynthesis at tRNA wobble positions. *Mol Cell* **21**: 97-108.

Imachi H, Sekiguchi Y, Kamagata Y, Hanada S, Ohashi A, Harada H. (2002). *Pelotomaculum thermopropionicum* gen. nov., sp. nov., an anaerobic, thermophilic, syntrophic propionate-oxidizing bacterium. *Int J Syst Evol Microbiol* **52**: 1729-1735.

Imachi H, Sekiguchi Y, Kamagata Y, Loy A, Qiu YL, Hugenholtz P *et al.*, (2006). Non-sulfate-reducing, syntrophic bacteria affiliated with *Desulfotomaculum* cluster I are widely distributed in methanogenic environments. *Appl Environ Microbiol* **72**: 2080-2091.

Ishii S, Suzuki S, Norden-Krichmar TM, Tenney A, Chain PSG, Scholz MB *et al.*, (2013). A novel metatranscriptomic approach to identify gene expression dynamics during extracellular electron transfer. *Nature Communications* **4**.

- Jarling R, Sadeghi M, Drozdowska M, Lahme S, Buckel W, Rabus R *et al.*, (2012). Stereochemical investigations reveal the mechanism of the bacterial activation of *n*-Alkanes without oxygen. *Angew Chem Int Ed Engl* **51**: 1334-1338.
- Karp PD, Paley SM, Krummenacker M, Latendresse M, Dale JM, Lee TJ *et al.*, (2010). Pathway Tools version 13.0: integrated software for pathway/genome informatics and systems biology. *Brief. Bioinform* **11**: 40-79.
- Karp PD, Latendresse M, Caspi R. (2011). The Pathway Tools pathway prediction algorithm. *Stand Genomic Sci* **5**: 424-429.
- Kato S, Watanabe K. (2010). Ecological and evolutionary interactions in syntrophic methanogenic consortia. *Microbes Environ* **25**: 145-151.
- Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S *et al.*, (2012). Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28**: 1647-1649.
- Kniemeyer O, Musat F, Sievert SM, Knittel K, Wilkes H, Blumenberg M *et al.*, (2007). Anaerobic oxidation of short-chain hydrocarbons by marine sulphate-reducing bacteria. *Nature* **449**: 898-U810.
- Koenig JE, Sharp C, Dlutek M, Curtis B, Joss M, Boucher Y *et al.*, (2009). Integron gene cassettes and degradation of compounds associated with industrial waste: The case of the Sydney Tar Ponds. *Plos One* **4**.
- Kosaka T, Kato S, Shimoyama T, Ishii S, Abe T, Watanabe K. (2008). The genome of *Pelotomaculum thermopropionicum* reveals niche-associated evolution in anaerobic microbiota. *Genome Res* **18**: 442-448.
- Kotsyurbenko OR, Glagolev MV, Nozhevnikova AN, Conrad R. (2001). Competition between homoacetogenic bacteria and methanogenic archaea for hydrogen at low temperature. *FEMS Microbiol Ecol* **38**: 153-159.
- Leang C, Qian XL, Mester T, Lovley DR. (2010). Alignment of the *c*-Type cytochrome *OmcS* along pili of *Geobacter sulfurreducens*. *Appl Environ Microbiol* **76**: 4080-4084.
- Li W, Wang LY, Duan RY, Liu JF, Gu JD, Mu BZ. (2012). Microbial community characteristics of petroleum reservoir production water amended with *n*-alkanes and incubated under nitrate-, sulfate-reducing and methanogenic conditions. *Int Biodeterior Biodegrad* **69**: 87-96.
- Ma K, Liu XL, Dong XZ. (2006). *Methanosaeta harundinaceae* sp nov., a novel acetate-scavenging methanogen isolated from a UASB reactor. *Int J Syst Evol Microbiol* **56**: 127-131.

- Mbadinga SM, Li KP, Zhou L, Wang LY, Yang SZ, Liu JF *et al.*, (2012). Analysis of alkane-dependent methanogenic community derived from production water of a high-temperature petroleum reservoir. *Appl Microbiol Biotechnol* **96**: 531-542.
- McInerney MJ, Rohlin L, Mouttaki H, Kim U, Krupp RS, Rios-Hernandez L *et al.*, (2007). The genome of *Syntrophus aciditrophicus*: Life at the thermodynamic limit of microbial growth. *PNAS* **104**: 7600-7605.
- Mikucki JA, Liu YT, Delwiche M, Colwell FS, Boone DR. (2003). Isolation of a methanogen from deep marine sediments that contain methane hydrates, and description of *Methanoculleus submarinus* sp nov. *Appl Environ Microbiol* **69**: 3311-3316.
- Parisi VA, Brubaker GR, Zenker MJ, Prince RC, Gieg LM, da Silva MLB *et al.*, (2009). Field metabolomics and laboratory assessments of anaerobic intrinsic bioremediation of hydrocarbons at a petroleum-contaminated site. *Microbial Biotechnology* **2**: 202-212.
- Patel GB, Sprott GD. (1990). *Methanosaeta concilii* Gen-Nov, Sp-Nov (*Methanothrix concilii*) and *Methanosaeta thermoacetophila* Nom-Rev, Comb-Nov. *Int J Syst Bacteriol* **40**: 79-82.
- Pereira IA, Ramos AR, Grein F, Marques MC, da Silva SM, Venceslau SS. (2011). A comparative genomic analysis of energy metabolism in sulfate reducing bacteria and archaea. *Front Microbiol* **2**: 69.
- Pruesse E, Peplies J, Glockner FO. (2012). SINA: accurate high-throughput multiple sequence alignment of ribosomal RNA genes. *Bioinformatics* **28**: 1823-1829.
- Rabus R, Jarling R, Lahme S, Kuhner S, Heider J, Widdel F *et al.*, (2011). Co-metabolic conversion of toluene in anaerobic *n*-alkane-degrading bacteria. *Environ Microbiol* **13**: 2576-2585.
- Rosana ARR, Chamot D, Owtrim GW. (2012). Autoregulation of RNA helicase expression in response to temperature stress in *Synechocystis* sp PCC 6803. *Plos One* **7**.
- Sakai Y, Takahashi H, Wakasa Y, Kotani T, Yurimoto H, Miyachi N *et al.*, (2004). Role of alpha-methylacyl coenzyme A racemase in the degradation of methyl-branched alkanes by *Mycobacterium* sp strain P101. *J Bacteriol* **186**: 7214-7220.

- Sanders JG, Beinart RA, Stewart FJ, Delong EF, Girguis PR. (2013). Metatranscriptomics reveal differences in *in situ* energy and nitrogen metabolism among hydrothermal vent snail symbionts. *ISME J* **7**: 1556-1567.
- Schink B. (1984). Fermentation of 2,3-butanediol by *Pelobacter carbinolicus* sp. nov and *Pelobacter propionicus* sp. nov, and evidence for propionate formation from C-2 compounds. *Arch Microbiol* **137**: 33-41.
- Shimoyama T, Kato S, Ishii S, Watanabe K. (2009). Flagellum mediates symbiosis. *Science* **323**: 1574-1574.
- Siddique T, Fedorak PM, Foght JM. (2006). Biodegradation of short-chain *n*-alkanes in oil sands tailings under methanogenic conditions. *Environ Sci Technol* **40**: 5459-5464.
- Siddique T, Fedorak PM, McKinnon MD, Foght JM. (2007). Metabolism of BTEX and naphtha compounds to methane in oil sands tailings. *Environ Sci Technol* **41**: 2350-2356.
- Siddique T, Penner T, Semple K, Foght JM. (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, Nesbo C, Foght JM. (2012). Microbial communities involved in methane production from hydrocarbons in oil sands tailings. *Environ Sci Technol* **46**: 9802-9810.
- Sieber JR, McInerney MJ, Gunsalus RP. (2012). Genomic insights into syntrophy: The paradigm for anaerobic metabolic cooperation. *Annu Rev Microbiol* **66**: 429-452.
- Soh J, Dong X, Caffrey SM, Voordouw G, Sensen CW. (2013). Phoenix 2: A locally installable large-scale 16S rRNA gene sequence analysis pipeline with Web interface. *J Biotechnol* **167**: 393-403.
- Stams AJM, Plugge CM. (2009). Electron transfer in syntrophic communities of anaerobic bacteria and archaea. *Nature Rev Microbiol* **7**: 568-577.
- Susanti D, Mukhopadhyay B. (2012). An intertwined evolutionary history of methanogenic archaea and sulfate reduction. *Plos One* **7**: e45313.
- Tian JQ, Wang YF, Dong XZ. (2010). *Methanoculleus hydrogenitrophicus* sp. nov., a methanogenic archaeon isolated from wetland soil. *Int J Syst Evol Microbiol* **60**: 2165-2169.
- Visser M, Worm P, Muyzer G, Pereira IA, Schaap PJ, Plugge CM *et al.*, (2013). Genome analysis of *Desulfotomaculum kuznetsovii* strain 17(T) reveals a

physiological similarity with *Pelotomaculum thermopropionicum* strain SI(T). *Stand Genomic Sci* **8**: 69-87.

Walker CB, Redding-Johanson AM, Baidoo EE, Rajeev L, He Z, Hendrickson EL *et al.*, (2012). Functional responses of methanogenic archaea to syntrophic growth. *ISME J* **6**: 2045-2055.

Waller AS, Hug LA, Mo K, Radford DR, Maxwell KL, Edwards EA. (2012). Transcriptional analysis of a *Dehalococcoides*-containing microbial consortium reveals prophage activation. *Appl Environ Microbiol* **78**: 1178-1186.

Welte C, Deppenmeier U. (2011). Membrane-bound electron transport in *Methanosaeta thermophila*. *J Bacteriol* **193**: 2868-2870.

Wilkes H, Kuhner S, Bolm C, Fischer T, Classen A, Widdel F *et al.*, (2003). Formation of *n*-alkane and *cyclo*-alkane-derived organic acids during anaerobic growth of a denitrifying bacterium with crude oil. *Org Geochem* **34**: 1313-1323.

Worm P, Stams AJM, Cheng X, Plugge CM. (2011). Growth- and substrate-dependent transcription of formate dehydrogenase and hydrogenase coding genes in *Syntrophobacter fumaroxidans* and *Methanospirillum hungatei*. *Microbiology-Sgm* **157**: 280-289.

Zellner G, Messner P, Winter J, Stackebrandt E. (1998). *Methanoculleus palmolei* sp. nov., an irregularly coccoid methanogen from an anaerobic digester treating wastewater of a palm oil plant in North-Sumatra, Indonesia. *Int J Syst Bacteriol* **48**: 1111-1117.

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

Zhou L, Li K-P, Mbadinga SM, Yang S-Z, Gu J-D, Mu B-Z. (2012). Analyses of *n*-alkanes degrading community dynamics of a high-temperature methanogenic consortium enriched from production water of a petroleum reservoir by a combination of molecular techniques. *Ecotoxicology (London, England)* **21**: 1680-1691.

Zhou Y, Liang YJ, Lynch KH, Dennis JJ, Wishart DS. (2011). PHAST: A Fast Phage Search Tool. *Nucleic Acids Res* **39**: W347-W352.

Zhu JX, Zheng HJ, Ai GM, Zhang GS, Liu D, Liu XL *et al.*, (2012). The genome characteristics and predicted function of methyl-group oxidation pathway in the obligate acetoclastic methanogens, *Methanosaeta* spp. *Plos One* 7.

## 5 Comparative metagenomic analysis of three methanogenic hydrocarbon-degrading enrichment cultures and relevant environmental metagenomes<sup>1</sup>

### 5.1 Abstract

Three model methanogenic hydrocarbon-degrading cultures were independently established from oil sands tailings ponds and gasoline-condensate contaminated aquifer sediments for intensive investigation of hydrocarbon degradation under methanogenic conditions. These cultures were enriched for growth on toluene (TOLDC), short chain alkanes (*n*-C<sub>6</sub> to *n*-C<sub>10</sub>; SCADC) or naphtha (NAPDC) and were subjected to 454 pyrosequencing and 16S rRNA gene pyrotag sequencing. Comparative metagenomic analysis<sup>1</sup> of the three metagenomes showed that these cultures each harbour unique dominant bacterial populations but share highly similar methanogen communities. These cultures are highly conserved in their overall functional capabilities for anaerobic respiration and methanogenesis and are capable of degrading of several structurally unrelated hydrocarbons by fumarate addition. However, their functional capabilities are distinct from environments with historical exposure to hydrocarbon pollutants including deep sea sediments from the Gulf of Mexico and oil sand tailings ponds. This indicates the high enrichment of functions relevant to methanogenic processes and provides the basis for future investigations into the development of consortia that potentially can be used in bioremediation and biomethanization.

<sup>1</sup>This work was done collaboratively with members of the Hydrocarbon Metagenomics Project. BoonFei Tan, Jane Fowler and Nidal Abu Laban conceived and discussed experimental designs. BFT, JF and NAL maintained SCADC, TOLDC, and NAPDC cultures, respectively. BFT, JF and NAL generated Table 5.1 and 5.2. BFT generated other Figures presented in this Chapter. He conceived and performed all associated bioinformatics analysis. JF and NAL generated figures and results and performed all associated bioinformatics analyses that are not included this Chapter. The written material presented in this Chapter is modified from a draft manuscript contributed to equally by BFT and JF.

## 5.2 Introduction

Methanogenesis and syntrophic biodegradation of organic compounds are important processes in polluted and engineered environments such as contaminated groundwater, organic chemical waste disposal sites, former gas plants, landfills, anaerobic digestors and oil sands tailings ponds (Lykidis *et al.*, 2011, Qiu *et al.*, 2004, Siddique *et al.*, 2011). When electron acceptors such as oxygen, nitrate, iron(III) and sulfate are depleted, methanogenesis is the dominant process by which organic matter is degraded in the subsurface. Furthermore, methanogenic biodegradation of hydrocarbons is an important biogeochemical process which, over geological time, has contributed to the formation of natural gas deposits in shales and the conversion of certain light oil deposits to heavy oil and bitumen deposits (Jones *et al.*, 2008). An improved understanding of the bioconversion of hydrocarbons to methane may contribute to the development of technologies for improved fossil fuel energy extraction, as well as to the development of improved bioremediation techniques for subsurface hydrocarbon-contaminated and oil sands tailings environments.

The mineralization of hydrocarbons by methanogenesis involves the coordinated metabolism of bacteria that catalyze the initial attack on hydrocarbon substrates with subsequent conversion to methanogenic substrates, and methanogens that produce CH<sub>4</sub>, CO<sub>2</sub>, and H<sub>2</sub>O as end products. These reactions have relatively low energy yields compared to aerobic or anaerobic metabolism in the presence of electron acceptors, and this energy must be shared amongst all microbes acting on the substrate (Schink 1997). The adaptations that allow sufficient energy conservation for growth and survival at these low energy yields are not well understood.

Previous studies showed that methanogenic hydrocarbon-degrading communities are highly diverse (Chang *et al.*, 2005, Penner and Foght 2010). They typically consist of hydrogenotrophic and/or acetotrophic methanogens (e.g., *Methanoculleus*, *Methanolinea*, *Methanoregula*, *Methanospirillum* and *Methanosaeta* respectively), as well as a variety of bacteria, with members of the Deltaproteobacteria (e.g., *Syntrophus/Smithella*, *Desulfovibrio*, *Geobacter*) and

Firmicutes (e.g., *Desulfotomaculum*, *Desulfosporosinus*, *Pelotomaculum*) being the most common and abundant (Gray *et al.*, 2011, Kleinsteuber *et al.*, 2012). In addition, members of the Spirochaetes, Bacteroidetes, Chloroflexi and Betaproteobacteria are frequently found, but generally in lower abundance (Kleinsteuber *et al.*, 2012). Methanogenic hydrocarbon-degrading co-cultures, consisting of a single bacterium and archaeon, are notoriously difficult to obtain, and therefore the specific roles of this wide diversity of organisms in methanogenic communities remain cryptic. The specific community structure may depend on the range of hydrocarbon substrates present, as well as the available nutrients and specific biogeochemical characteristics. Despite the recent report of a methanogenic hydrocarbon-degrading co-culture (Callaghan *et al.*, 2012), the key enzymes and genes that mediate the syntrophic biodegradation of hydrocarbons are not yet well described. Genes encoding anaerobic hydrocarbon-degrading enzymes have been shown to be present and active in hydrocarbon-degrading methanogenic cultures (Fowler *et al.*, 2012, Washer and Edwards 2007, Wawrik *et al.*, 2012). Other genes of importance in syntrophic degradation include those involved in methanogenesis as well as electron transfer processes, amino acid biosynthesis, metabolism, and transport (Kato and Watanabe 2010, Walker *et al.*, 2012). To gain insight into the metabolic processes and the microbial communities involved in methanogenic hydrocarbon metabolism, we carried out metagenomic sequencing of three methanogenic hydrocarbon-degrading cultures enriched on different substrates: toluene, a mixture of low molecular weight alkanes, or naphtha. To identify functions of importance in methanogenic hydrocarbon degradation, we compared these metagenomic datasets to publicly available metagenomes from a range of environments. The metagenomes were screened for the presence of genes and pathways known to be involved in the anaerobic degradation of hydrocarbon substrates, as well as for a number of central metabolic and key anaerobic pathways.

## 5.3 Materials and methods

### 5.3.1 Incubation conditions

Three methanogenic hydrocarbon-degrading enrichment cultures were established and maintained on different substrates: toluene; *n*-C<sub>6</sub> to *n*-C<sub>10</sub> alkanes (plus minor proportions of 2-methylpentane and methylcyclopentane present as impurities in commercial ‘hexanes’; Chapter 2 and 3); or naphtha (a refinery cut consisting of a mixture of monoaromatic hydrocarbons, *n*-, *iso*- and *cyclo*-alkanes) (Table 5.1). The toluene-degrading methanogenic enrichment culture (TOLDC) was enriched more than 10 years from gas condensate-contaminated aquifer sediments (Fowler *et al.*, 2012). The short-chain alkane-degrading culture (SCADC) completely biodegrades *n*-C<sub>6</sub> to C<sub>10</sub> alkanes (Chapters 2 and 3; Figure 1.3). The naphtha-degrading culture (NAPDC) was maintained with naphtha (CAS no. 64742-49-0) as the sole organic carbon substrate (Abu Laban, unpublished). The SCADC and NAPDC cultures were enriched from methanogenic mature fine tailings collected from Mildred Lake Settling Basin (MLSB), an oil sands tailings pond in northeastern Alberta, Canada (Siddique *et al.*, 2006). All cultures were incubated stationary in the dark under mesothermic conditions (20-28°C) and transferred to fresh medium at intervals (Table 5.1).

#### 5.1.1. DNA extraction and 16S rRNA gene pyrotag sequencing

The microbial communities present in TOLDC and SCADC have been previously described using 16S rRNA gene pyrotag sequencing (Fowler *et al.*, 2012, Tan *et al.*, 2013). Total DNA from NAPDC used for 16S rRNA gene pyrosequencing was extracted (Foght *et al.*, 2004), followed by PCR amplification with primers 926F and 1392R, targeting the V6-V8 regions of the 16S rRNA gene (Tan *et al.* 2013). All 16S rRNA gene pyrotag sequencing reads were generated using the same primer set (Fowler *et al.* 2012) and were submitted simultaneously to Phoenix 2 (Soh *et al.*, 2013) for quality control, cluster analysis (0.05 distance cut-off) and taxonomic classification with RDP classifier.

**Table 5.1** Enrichment of hydrocarbon-degrading methanogenic cultures and incubation conditions

	NAPDC	SCADC	TOLDC
<b>In situ conditions of inoculum source</b>			
Sample environment	Mature fine tailings from oil sands tailings	Mature fine tailings from oil sands tailings	Gas condensate-contaminated aquifer sediments <sup>2</sup>
Country	non <sup>1</sup> Alberta, Canada	non <sup>1</sup> Alberta, Canada	Colorado, USA
Depth (m below surface)	31 m	35 m	1.5 m
Bulk pH	7-8	7-8	7-8
Temperature	12-20°C	12-20°C	15-20°C
Redox condition	Anaerobic	Anaerobic	Anaerobic
<b>Incubation conditions for enrichment cultures</b>			
Inoculum %	50% (v/v)	50% (v/v)	30% (v/v)
Number of transfers	2	4	>10
Inoculum proportion	10% (v/v)	50% (v/v)	30% (v/v)
Substrate	0.2% (v/v) Naphtha	0.1% (v/v) C6-C10 Alkane mix	0.01% (v/v) Toluene
Culture Medium	anaerobic mineral medium <sup>3</sup>	anaerobic mineral medium <sup>3</sup>	anaerobic freshwater
Initial medium pH	7-8	7-8	7-8
Initial headspace gas	20/80 % (v/v) CO <sub>2</sub> /N <sub>2</sub>	20/80 % (v/v) CO <sub>2</sub> /N <sub>2</sub>	20/80 % (v/v) CO <sub>2</sub> /N <sub>2</sub>
Incubation temperature	28°C	28°C	20-25°C
Incubation time	120 days	120 days	150 days
Light exposure	Dark	Dark	Dark

<sup>1</sup> Mildred Lake Settling Basin, Alberta, Canada (Siddique *et al.*, 2007a).

<sup>2</sup> Ft. Lupton, Colorado, USA (Fowler *et al.*, 2012).

<sup>3</sup> Reported by Widdel and Bak (1992).

### **5.3.2 Total DNA extraction for metagenomic sequencing**

For high throughput sequencing of total DNA, approximately 500 mL of filtered culture fluids from each of TOLDC and NAPDC were used for phenol-chloroform DNA extraction, followed by cesium chloride purification (Wright *et al.* 2009) performed at University of British-Columbia. Total DNA for each culture was sequenced on a half plate using the GS FLX Titanium Sequencing Kit XLR70 (Roche) at McGill University Génome Québec Innovation Centre, Canada.

### **5.3.3 Quality control, *de novo* assembly and annotation of metagenomic data**

All metagenomic 454 reads were subjected to quality control (QC) and sequence assembly using an in-house pipeline based on NewblerV.2.3 using options “mi 95 -ml 60 -a 200 -l 900” (Tan *et al.* 2013). The QC reads and assembled contigs were submitted to MG-RAST v 3.3.7.3 (Meyer *et al.*, 2008) for automated gene calling and annotation. The QC and unassembled metagenomic reads obtained from 454 pyrosequencing of total DNA were subjected to taxonomic classification using (i) the taxonomic classifier in MG-RAST with the following parameters: M5NR annotation, maximum e-value  $1 \times 10^{-10}$ , minimum alignment length 50 bp, minimum identity 60%, and (ii) homology-based binning using SOrt-ITEMS (Haque *et al.*, 2009). Additionally, unassembled reads were used in genome mapping in Geneious R7 (Biomatters Ltd., Auckland, New Zealand) with default settings. Metagenomic sequences from SCADC are available on the Short Read Archive under the SRA number SAMN01828453; sequences from TOLDC and NAPDC will be made public upon manuscript publication.

### **5.3.4 Comparative analysis of functional categories in metagenomic datasets**

NAPDC, SCADC, and TOLDC unassembled metagenomic sequences were annotated in MG-RAST v 3.3.7.3 (Meyer *et al.*, 2008) using SEED with the following parameters: maximum e-value,  $1 \times 10^{-10}$ ; minimum percent identity, 60%; minimum alignment length, 50 bp (Meyer *et al.* 2008). Forty-one additional

metagenomes representing a variety of environments including potentially closely-related hydrocarbon-impacted sites and dechlorinating cultures (Appendix Table E1) were compared based on their functional capabilities. Metagenomes representing hydrocarbon-impacted sites that were included in this analysis include three metagenomes of the Gulf of Mexico deep marine sediments (Kimes *et al.*, 2013) and one metagenome of an oil sands tailing pond (An *et al.*, 2013). At the time of the analysis, no other similar metagenomes (i.e., from hydrocarbon-degrading cultures) were available in MG-RAST for comparison. All 44 metagenomes were compared using principal component analysis (PCA) by comparing relative abundances of SEED functional categories using R software with the *ade4* package (R Development Core Team, 2008). Datasets from oil sands tailings and Gulf of Mexico deep marine sediments were further compared in groups with pooled NAPDC, SCADC, and TOLDC metagenomes using STAMP (Table D1) (Parks and Beiko 2010) as described by (Hug *et al.*, 2012). For subsystem categories with rare counts (e.g., 0.1% of total counts), a fixed number of pseudocounts proportional to the total gene count was added to all categories across all groups used in comparisons to prevent selection of rare categories (Hug *et al.*, 2012). For all comparisons, unassembled metagenomic reads were used unless stated otherwise.

### **5.3.5 Phylogenetic analysis of putative fumarate addition enzymes**

Amino acid sequences of genes encoding fumarate addition (i.e. *assA*, *bssA* and *nmsA*) in TOLDC and NAPDC metagenomes were recovered using tBLASTn (Gertz *et al.*, 2006). For this part of the analysis, assembled 454 reads were used. Illumina reads reported by Tan *et al.*, (2013) were re-assembled using CLC Genomics Workbench (CLC Bio, USA), and the *assA/bssA/nmsA* genes were similarly recovered from the sequence assembly using tBLASTn. All genes, along with reference sequences obtained from the GenBank database, were aligned using MUSCLE V3.3 (Edgar 2004). Poorly aligned regions were manually edited and indels were removed. Phylogenetic analysis was performed using PhyML (Guindon *et al.*, 2010) with WAG model and 100 bootstrap replicates. Trees were visualized and edited in FigTree V14.0 (Rambaut 2012).

## 5.4 Results and discussion

### 5.4.1 Descriptions of TOLDC, SCADC and NAPDC cultures

TOLDC was originally enriched from gasoline condensate-contaminated aquifer sediments in Fort Lupton (Table 5.1). *assA* and *bssA* genes encoding enzymes for fumarate addition had been detected previously in samples collected from the site, suggesting that hydrocarbon contaminants in the polluted area could be degraded via fumarate addition (Callaghan *et al.*, 2010). This was further confirmed by metabolite analysis of TOLDC that detected putative products of benzylsuccinate, indicating that toluene was activated by fumarate addition (Fowler *et al.*, 2012). NAPDC and SCADC were enriched from the mature fine tailings of Mildred Lake Settling Basins (MLSB). Previous studies had indicated that microorganisms indigenous to MLSB are capable of methanogenic degradation of monoaromatics and aliphatic alkanes (Siddique *et al.*, 2006, Siddique *et al.*, 2007b, Siddique *et al.*, 2011, Siddique *et al.*, 2012). SCADC degrades short chain *n*-alkanes (*n*-C-6 to C-10), 2-methylpentane and methylcyclopentane during 108 days of incubation (Tan *et al.*, 2013; Chapter 2). Metabolite analysis detected succinylated products for the activation of 2-methylpentane and methylcyclopentane by fumarate addition, but not alkylsuccinates from *n*-alkane addition to fumarate (Chapter 2). NAPDC was degrading toluene and *o*-xylene at the time when DNA was extracted for metagenomic sequencing. Metabolite analysis using the method reported in Chapter 2 did not detect products of fumarate addition (Abu Laban, unpublished). However, benzylsuccinate and products of *o*-xylene addition to fumarate were routinely detected in a parallel methanogenic naphtha-degrading culture enriched from oil sands tailings ponds (unpublished).

### 5.4.2 General features of NAPDC, SCADC and TOLDC metagenomes

Metagenomic 454 pyrosequencing of total DNA extracted from NAPDC, SCADC, and TOLDC generated ~370,000, 670,000, and 550,000 quality controlled reads, respectively (Table 5.2). Rarefaction analysis based on total DNA metagenomic analysis generated in MG-Rast and the size of the raw datasets revealed that SCADC was more deeply sequenced than either NAPDC or

TOLDC, but that none of the cultures was sequenced to saturation (Appendix Figure E1).

**Table 5.2** Features of the metagenomes generated by 454 pyrosequencing of three hydrocarbon-degrading methanogenic enrichment cultures.

	<b>NAPDC</b>	<b>SCADC</b>	<b>TOLDC</b>
Number of raw reads post-quality control (QC)	368,209	667,134	550,247
Unassembled reads post-Mean read length post-QC (bp)	130,215,413 353.65	230,716,341 345.83	215,982,194 392.52
<b>Newbler assembly</b>			
Length of contigs (bp)	9,473,370	17,382,962	20,280,655
Number of contigs	8,471	15,274	10,888
Range of contig lengths (bp)	200-30,000	200-30,000	200-30,000
Mean contig size (bp)	1,118	1,138	1,863
Largest contig (bp)	25,813	26,073	28,253
Number of singletons	161,851	326,382	179,067
N50	1,330	1,343	2,813
Number of predicted proteins	133,107	261,378	170,842
Number of rRNA genes <sup>1</sup>	23/54/155	46/135/267	29/104/189
Number of tRNA genes	816	1,322	891
<b>MG-RAST data (assembled)</b>			
MG-RAST ID	4492772.3	4492619.3	4492778.3
Metagenome size (bp)	64,676,632	127,644,733	89,513,841
Average sequence length (bp)	379 ± 352	373 ± 310	471 ± 647
Number of sequences	170,322	341,656	189,955
GC content (%)	49 ± 8%	50 ± 9%	53 ± 9%
Number of predicted ORFs	162,642	325,794	187,741
ORFs with predicted function	94.562	184.014	105.547
Alpha diversity <sup>2</sup>	662	771	907

<sup>1</sup> 5S, 16S, 23S, respectively.

<sup>2</sup> Number of distinct species in a given metagenome sample as calculated by MG-RAST.

The metagenomes assembled into approximately 8,500 (NAPDC), 15,000 (SCADC), and 11,000 (TOLDC) contigs, with ~162,000, 326,000 and 179,000 singletons, respectively, and had GC contents of 49-53% (Table 5.2). The

average assembled sequence lengths were quite short (1118–1863 bp) but contigs varied greatly in length (200–30,000 bp). This likely reflects the relative abundance of different community members, with more abundant members being better assembled due to a higher proportion of reads. The longest contigs ranged from 25.8 to 28.2 Kb, which is relatively short compared to other metagenomes (Hug *et al.* 2012), suggesting that these cultures are particularly biodiverse.

#### **5.4.3 TOLDC, SCADC, NAPDC share similar methanogenic communities but differ in dominant microbial communities putatively involved in hydrocarbon degradation**

The microbial community compositions of TOLDC, SCADC and NAPDC (Table 5.3) were determined by 16S rRNA gene pyrotag sequencing using previously reported primer sets (Mason *et al.*, 2012; Allers *et al.*, 2013 and An *et al.*, 2013), and also by taxonomic profiling of unassembled metagenomic reads in MG-RAST using best hit classification against the M5NR database with a minimum e-value of  $10^{-10}$ , minimum alignment length of 60 amino acids and minimum identity (% similarity) cut off 60%.

Currently, there is a lack of consensus on the best similarity threshold values used in OTU clustering analysis (i.e. 97% or 95%) (Schloss and Westcott 2011), but it has been suggested that OTU clustering with similarity of less than 97% may be able to accommodate potential biases arising from homopolymers and sequence artifacts introduced during 454 pyrotag sequencing. Therefore a threshold of less than 97% may be better able to ensure accuracy and sensitivity for phylogenetic profiling (Kunin *et al.*, 2010). Based on that premise, 95% similarity was chosen as the threshold for OTU clustering in the present study.

Overall, the total number of OTUs were 153 (NAPDC), 320 (SCADC) and 147 (TOLDC), including 65, 167, and 65 singletons, respectively (Table 5.3). Clustering analysis indicated that all cultures in general shared dominant methanogen OTUs (i.e., *Methanosaeta*, *Methanoculleus*), but the dominant bacterial OTUs in each culture may be unique (i.e., Lachnospiraceae and Sedimentibacter in TOLDC, Peptococcaceae *Desulfotomaculum* in NAPDC, Peptococcaceae, *Smithella* and *Syntrophus* in SCADC). SCADC and NAPDC

shared more OTUs with each other than with TOLDC (Table 5.3), consistent with their similar origin and growth substrates (i.e., both contain alkanes as substrates).

**Table 5.3** The five most abundant bacterial and archaeal operational taxonomic units (OTUs) in enrichment cultures determined by PCR amplification of 16S rRNA genes from metagenomic DNA.

Phylum	Taxonomic affiliation <sup>1</sup>	Percentage of reads in each		
		NAPDC	SCADC	TOLDC
<b><u>Bacterial OTUs</u></b>				
Firmicutes	<i>Anaerobacter</i>	9.5	3.0	30.2
Firmicutes	Lachnospiraceae	0.0	0.0	23.8
Firmicutes	Peptococcaceae	19.5	0.2	0.0
Firmicutes	<i>Desulfotomaculum</i>	15.3	0.0	0.0
Firmicutes	<i>Sedimentibacter</i>	0.0	0.0	11.1
Spirochaetes	Spirochaetaceae	2.7	14.3	0.2
Deltaproteobacteria	<i>Smithella</i>	0.6	11.6	0.0
Deltaproteobacteria	<i>Syntrophus</i>	0.1	6.8	0.0
Deltaproteobacteria	<i>Geobacter</i>	2.8	2.6	0.1
Deltaproteobacteria	<i>Desulfovibrio</i>	0.6	4.2	0.0
Deltaproteobacteria	<i>Desulfobacterium</i>	0.1	3.6	0.0
Synergistetes	<i>Thermanaerovibrio</i>	4.2	2.5	0.0
Chloroflexi	Anaerolineaceae	0.0	0.0	5.4
Candidate division	Candidate division OP11	2.4	0.0	0.0
Candidate division	Candidate division OP8	0.0	0.0	3.3
<b><u>Archaeal OTUs</u></b>				
Euryarchaeota	<i>Methanosaeta</i> 1	51.5	55.9	50.6
Euryarchaeota	<i>Methanosaeta</i> 2	5.4	5.4	2.2
Euryarchaeota	<i>Methanoculleus</i> 1	32.1	21.3	15.1
Euryarchaeota	<i>Methanoculleus</i> 2	3.6	2.4	0.4
Euryarchaeota	<i>Methanolinea</i>	0.4	1.3	27.3
Euryarchaeota	<i>Candidatus Methanoregula</i>	0.6	4.8	0.6
Euryarchaeota	<i>Methanomicrobia</i>	1.4	1.4	0.0
Euryarchaeota	WCHA1	0.2	0.3	3.0

<sup>1</sup> Each OTU was assigned a taxonomic affiliation using RDP classifier, and only the lowest identified taxon is shown.

The proportion of OTUs associated with the methanogens was 8% for TOLDC, 65% for SCADC and 89% for NAPDC as determined by pyrotag sequencing. In contrast, taxonomic assignment of metagenomic reads using MG-RAST detected higher abundance of bacterial versus archaeal reads in SCADC and NAPDC (>70% bacterial reads in both cultures; Figure 5.1). One explanation for the discrepancy is potential primer bias that favours the methanogen populations in NAPDC and SCADC. However, the primer sets used in this study (J. Klassen, unpublished) had been tested by subjecting a mixed culture of 11 bacterial and archaeal strains to PCR co-amplification, followed by pyrotag sequencing of the amplicons: results were as expected and biases towards certain taxonomic groups were not significant. Other likely causes of discrepancies are variations in the copy number of 16S rRNA genes or polyploidy of the genomes of different methanogen OTUs (Hildenbrand *et al.*, 2011) observed in the three cultures. Notably, the methanogen OTUs in TOLDC are represented by the high abundance of OTUs affiliated with *Methanolinea* (~27%), but the same OTU was <2% in SCADC and NAPDC. Nonetheless, microbial profiling was largely consistent at the intra-domain level as assessed using OTU analysis and classification of metagenomic reads using SOrt-ITEMSs (Haque *et al.*, 2009) and MG-RAST (Appendix Figure E2).

#### **5.4.3.1 Community members putatively involved in bottleneck reactions of anaerobic hydrocarbon degradation**

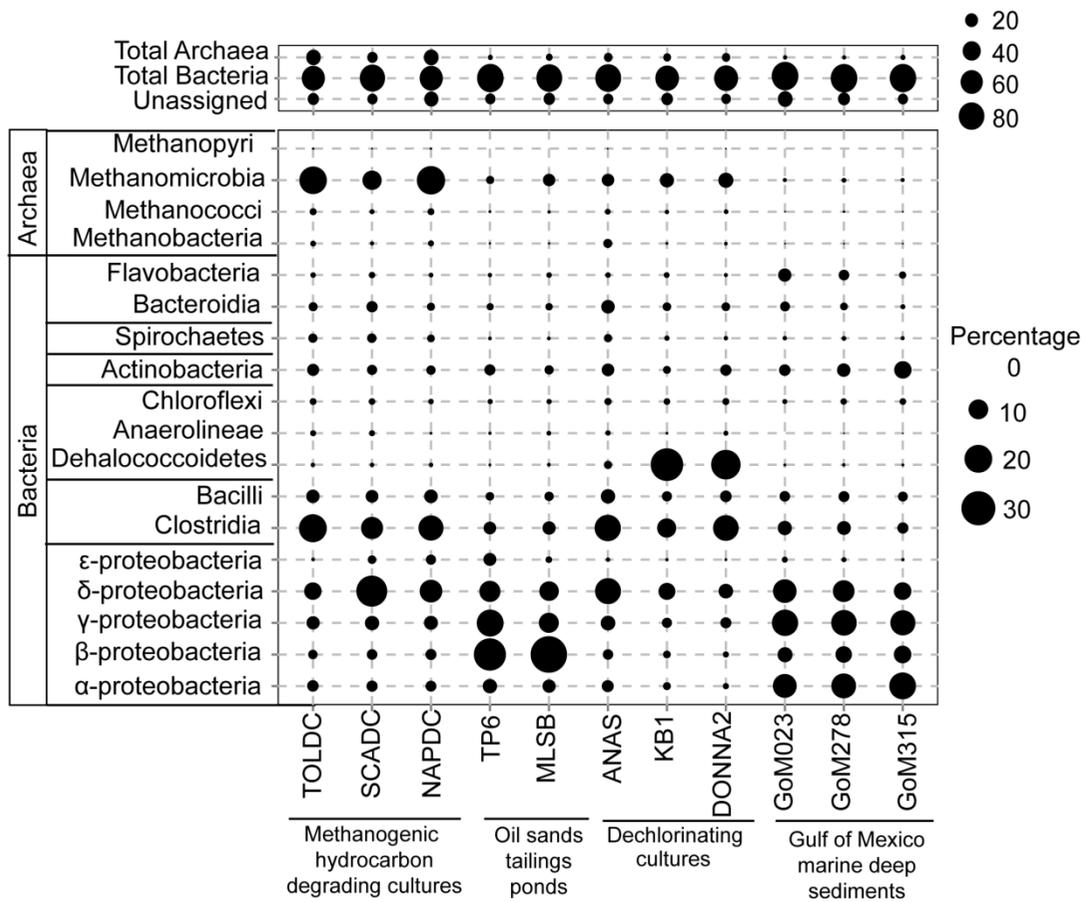
In methanogenic hydrocarbon-impacted environments, microorganisms that are capable of initiating degradation of substrates are essential in the methanogenesis process. Therefore, the initial activation and degradation of substrates represent the bottleneck reaction of methanogenesis. In environments and enrichment cultures studied under laboratory conditions, dominant microorganisms detected using techniques such as cloning and sequencing of 16S rRNA genes have been inferred to have crucial roles in initial substrate degradation (Gray *et al.*, 2011, Rios-Hernandez *et al.*, 2003, Zengler *et al.*, 1999).

Based on that premise, the dominant microorganisms detected in TOLDC, SCADC and NAPDC are likely crucial in the bottleneck reactions of hydrocarbon

degradation in their respective cultures. In general, the combination of pyrotag and MG-RAST taxonomic profiling methods used here revealed that the NAPDC and TOLDC bacterial communities are dominated by members of the phylum Firmicutes, followed by Proteobacteria, Bacteroidetes, Spirochaetes, and other phyla in smaller proportions (< 1% of total reads). Despite the similarities in the overall bacterial community structure at the phylum level (Appendix Figure D1), there are noteworthy taxonomic variations in the dominant OTU affiliated with Firmicutes. The dominant Firmicutes-related OTUs in NAPDC were related to Peptococcaceae (i.e., *Desulfotomaculum*), whereas Clostridiales (i.e., *Anaerobacter*, Lachnospiraceae and *Sedimentibacter*) were more dominant in TOLDC (Table 5.3). The bacterial community in SCADC, as determined using metagenomic reads, was dominated by members of the Deltaproteobacteria, Spirochaetes, Bacteroidetes, with numerous other microorganisms each representing <1% of reads (Figure 5.1 and Table 5.3). Community analysis using 16S rRNA gene pyrotag sequencing resulted in greater apparent abundance of Spirochaetes and Syntrophaceae (e.g., *Syntrophus* and *Smithella*) and smaller proportions of Firmicutes and Bacteroidetes in SCADC compared to TOLDC and NAPDC (Table 5.3).

The high abundance of Firmicutes in TOLDC and NAPDC aligns with recent descriptions of strictly anaerobic monoaromatic hydrocarbon-degrading cultures that implicate Firmicutes (e.g., Peptococcaceae) as primary hydrocarbon degraders (Abu Laban *et al.*, 2010, Winderl *et al.*, 2010) or playing key roles alongside Deltaproteobacteria (Ficker *et al.*, 1999, Sakai *et al.*, 2009, Ulrich and Edwards 2003). This is in contrast to the microbial communities capable of anaerobic degradation of *n*-alkanes, in which members of the Deltaproteobacteria such as *Syntrophus*, and phylogenetically related species of *Smithella* sp. (Gray *et al.*, 2011, Jones *et al.*, 2008, Zengler *et al.*, 1999) have been routinely implicated in methanogenic alkane degradation. Furthermore, all of the known strictly anaerobic (sulfate-reducing/syntrophic) alkane degraders, including *Desulfococcus oleovorans* HxD3, strain Pnd3, *Desulfatibacillum aliphaticivorans* CV2803, *Desulfothermus naphthae* TD3, *Desulfatibacillum alkenivorans* AK-01,

*Desulfoglaeba alkanexedens*, and *Desulfosarcina* sp., are Deltaproteobacteria (Agrawal and Gieg 2013, Callaghan 2013). There may also be a role for Firmicutes in alkane degradation, as suggested by the recent implication of a *Desulfotomaculum* sp. in the degradation of low-molecular weight *n*-alkanes, i.e., propane, under sulfate-reducing conditions (Kniemeyer *et al.*, 2007). Likewise, we reported in Chapter 4 that *Desulfotomaculum* spp. is hypothesized to be involved in the syntrophic degradation of low molecular weight alkanes. Additional discussion of this point appears in Sections 2.4.3 and 4.3.2.3.



**Figure 5.1** Proportion of microbial communities at the taxonomic class level in TOLDC, SCADC, NAPDC and eight other metagenomes based on assignment of 454 metagenomic reads using best hit classification against the M5NR database in MG-RAST. All metagenomes are available on the MG-RAST website and the associated identifications are reported in Appendix Table E1.

In all three cultures, the dominant microbial communities putatively involved in hydrocarbon degradation may include novel primary degraders where genomes have not yet been sequenced, since metagenomic sequence reads from these cultures did not map closely to the genomes of well-known hydrocarbon-degraders or their close relatives (Appendix Table E2). Based on the dominant OTUs in these cultures (Table 5.3) and on the phylogenetic distribution of genes involved in activation of hydrocarbons by fumarate addition (see below), the primary hydrocarbon degraders in NAPDC, SCADC, and TOLDC consist of different combinations of a few bacterial taxa. In methanogenic hydrocarbon-impacted environments, microorganisms that are involved in bottleneck reactions such as initial activation form the "specialist" guild and are crucial for substrate degradation, whereas diverse phylogenetically distinct microorganisms may be capable of common anaerobic processes such as autotrophic carbon fixation and hydrogen uptake, and thus can be assumed to be the "generalists" in these systems (discussed in section 3.4.7 in Chapter 3).

#### **5.4.3.2 Presumptive syntrophs and secondary fermenters**

The high diversity and persistence of microorganisms in TOLDC, SCADC and NAPDC even after multiple transfers is likely due to the presence of secondary fermenters and syntrophs (e.g. Spirochaetes, Clostridiales, Synergistetes, Candidate division OP11) that are essential to methanogenic growth of the whole community. Methanogenic degradation of hydrocarbons typically yields by-products such as CO<sub>2</sub>, H<sub>2</sub>, acetate and formate (Edwards and Grbić-Galić 1994, Mbadanga *et al.*, 2011). In general, these organisms are likely to be involved in the transformation of partially degraded hydrocarbons to methanogenic substrates (Beller and Edwards 2000), maintaining a low redox potential and hydrogen partial pressure through H<sub>2</sub> sequestration and/or production (Ahring *et al.*, 1991), and/or scavenging or recycling of waste products and dead biomass within the enrichments. More specifically, homoacetogens such as Spirochaetes can be capable of autotrophic CO<sub>2</sub> fixation. Syntrophic acetate oxidizers (e.g., Clostridiales) can further convert acetate produced from hydrocarbon degradation into CO<sub>2</sub> and H<sub>2</sub> that may be utilized by

hydrogenotrophic methanogens for methanogenesis (Gray *et al.*, 2011). Other members of the community such as Anaerolineaceae, of which *Dehalococcoides* is a member, and *Desulfovibrio* spp. may be involved in providing essential nutrients such as vitamin B12 and amino acids (e.g., valine) for other community members (Brisson *et al.*, 2012, Walker *et al.*, 2012). The stability and persistence of diverse community members in methanogenic hydrocarbon degrading environments such as TOLDC even after incubation for >10 years indicates that they must be essential for syntrophic growth of the community.

### 5.4.3.3 Methanogens

Almost all metagenomics reads associated with the archaeal communities in the three cultures were exclusively related to Euryarcheota (Appendix Figure E2). The archaeal populations in all three cultures were less diverse and were dominated primarily by Methanosarcinales (2 OTUs affiliated with *Methanosaeta*) and Methanomicrobiales (2 *Methanoculleus* OTUs and 1 *Methanolinea* OTU; Table 5.3). Most members of Methanosarcinales (i.e., Methanosarcinaceae) are generally metabolically versatile and can use acetate, CO<sub>2</sub>/H<sub>2</sub> and methylotrophic substrates such as methanol as growth substrates (Maestrojuan and Boone 1991). However, the genus *Methanosaeta* is considered to comprise obligate acetate-utilizing methanogens (Patel and Sprott 1990). Recent genomic sequencing and annotation suggest that *Methanosaeta* is genetically capable of methylotrophic and/or hydrogenotrophic methanogenesis (Zhu *et al.*, 2012), although this remains to be confirmed. In contrast, most *Methanoculleus* spp. use formate and CO<sub>2</sub>/H<sub>2</sub> as energy sources in the presence of acetate as carbon source (Dianou *et al.*, 2001, Zellner *et al.*, 1998). Mapping of metagenomic reads to selected reference genomes of the methanogens shows that TOLDC, SCADC and NAPDC all harbour a strain of *Methanosaeta* that is closely related to *Methanosaeta concilii* GP6, but the *Methanoculleus* in these cultures is not similar to *Methanoculleus marisnigri* or *Methanoculleus bourgenis*, whose genomes have been sequenced (Appendix Table E2). Based on this, methanogenesis in TOLDC, SCADC and NAPDC likely proceeds

primarily through CO<sub>2</sub>/H<sub>2</sub> (hydrogenotrophic) and acetate (acetotrophic methanogenesis) conversion to CH<sub>4</sub>.

#### **5.4.4 TOLDC, SCADC and NAPDC share many key metabolic and anaerobic hydrocarbon degradation pathways**

Hydrocarbon addition to fumarate was first discovered for the degradation of toluene by *Thauera aromatica* under denitrifying conditions (Biegert *et al.*, 1996). This mechanism has since been demonstrated in activation of *o*-, *m*- and *p*-xylene, ethylbenzene and 2-methylnaphthalene (2-MN) under sulfate-, iron-reducing and methanogenic conditions (reviewed by Foght 2008, Heider 2007). Fumarate addition to *n*-alkanes has only been demonstrated under sulfate- and nitrate-reducing conditions thus far but is questioned under methanogenic conditions (Aitken *et al.*, 2013). Re-analysis of the draft genome of *Smithella* spp. and the metatranscriptome of an associated culture grown under methanogenic conditions with hexadecane (Embree *et al.*, 2013), discussed in Chapter 4 and Appendix A, proved that the Deltaproteobacterium *Smithella* is genetically capable of hexadecane addition to fumarate under methanogenic conditions.

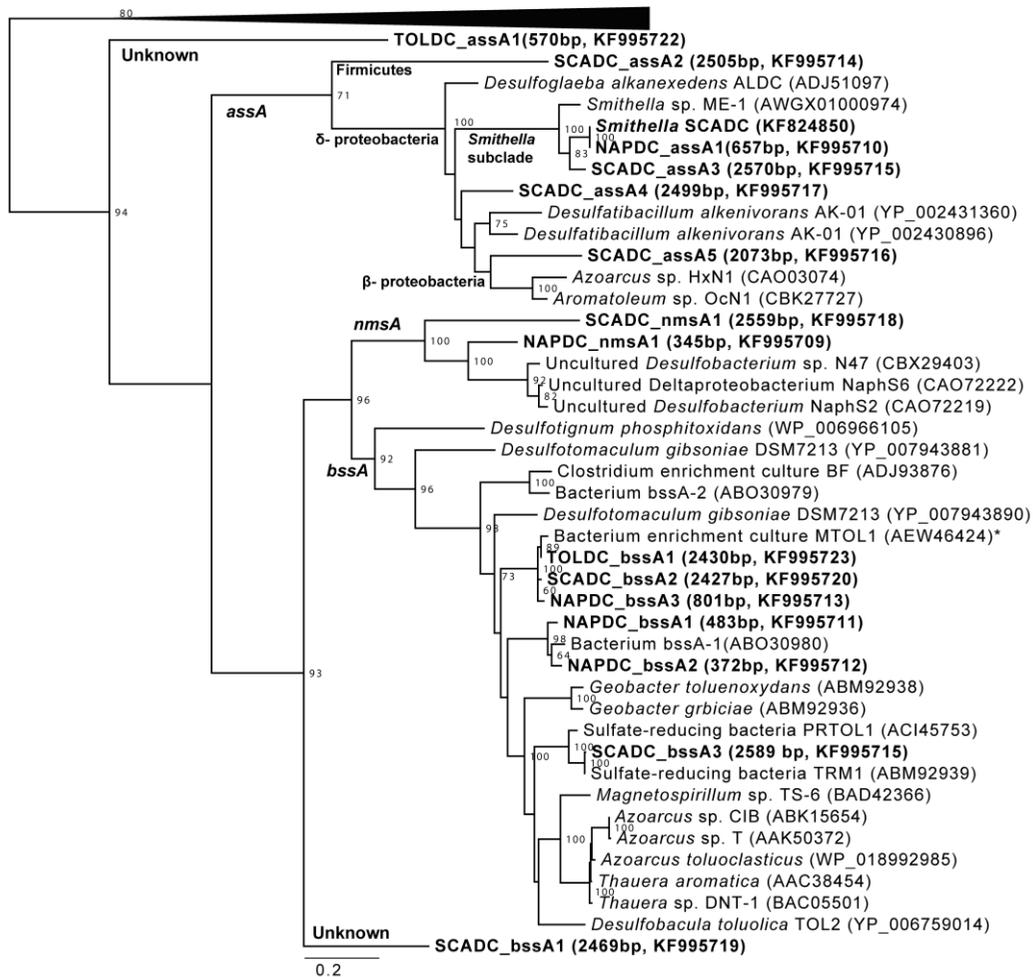
Several anaerobic hydrocarbon activating mechanisms besides fumarate addition have been proposed for the carboxylation (So *et al.*, 2003) and hydroxylation (Callaghan 2013) of *n*-alkanes; and carboxylation, hydroxylation or methylation of benzene (Abu Laban *et al.*, 2010, Ulrich and Edwards 2003), although only the hydroxylation of ethylbenzene has been confirmed (Kniemeyer and Heider 2001). Hydrocarbon addition to fumarate continues to be the most widely recognized and reported hydrocarbon activating mechanism based on detection of metabolites and genes encoding enzymes responsible for this process (reviewed by Callaghan 2013; Agrawal and Gieg, 2013).

#### **5.4.5 Putative genes associated with hydrocarbon activation**

The *assA*, *bssA*, and *nmsA* genes encoding the  $\alpha$ -subunits of alkylsuccinate (ASS), benzylsuccinate (BSS) and naphthylsuccinate synthase (NMS), respectively, have been used as markers to indicate the functional potential of microbial communities and bacterial isolates to activate hydrocarbons with fumarate (Callaghan, 2013; Agrawal and Gieg, 2013). Metagenomic

analysis of SCADC in Chapters 3 and 4 indicated the presence of these genes, even though the SCADC culture has been enriched only on alkanes (Tan *et al.*, 2013). Furthermore, *assA* and *bssA* had also been detected in the contaminated aquifer in Fort Lupton, the source of TOLDC (Callaghan *et al.*, 2010) and in an oil sands tailings pond (Suncor TP 6) (Chapter 6).

In this study, we screened for evidence of genes encoding other fumarate addition mechanisms in TOLDC and NAPDC using tBLASTn (Gertz *et al.*, 2006) in order to infer the substrate spectrum of all cultures. Several full length sequences (2.4 kbp) of putative genes encoding fumarate addition were recovered from the SCADC metagenome. However, sequences recovered from assembly of the NAPDC metagenome were not full length and most were <1000 bp. Phylogenetic analysis of the fumarate addition genes identified in the three hydrocarbon-degrading metagenomes revealed distinct groups of *assA*, *bssA* and *nmsA* genes (Figure 5.2). The *assA* genes associated with the *Smithella* SCADC genome assembly reported in Chapter 4 were also found in the NAPDC metagenome but not in TOLDC, corroborating the paucity of 16S rRNA gene OTU representing Syntrophaceae in TOLDC (Table 5.3). Several full length *assA* sequences putatively related to  $\delta$ -proteobacteria and Firmicutes detected in SCADC (reported in Chapter 3) were not detected in NAPDC and TOLDC. A single partial *assA* sequence detected in TOLDC (TOLDC\_assA1) did not appear to have close affiliation with any known reference sequences, and appeared to form a deep branch in the phylogenetic tree (Figure 5.2). The amino acid sequence of this putative *assA* gene had a best BLAST hit to the translated *assA* gene in *Desulfoglaeba alkanexedens* ALDC (ADJ51097, similarity 44.5%) and to other *assA* sequences but was less similar to pyruvate formate lyase (PFL).



**Figure 5.2** Maximum likelihood tree of translated *assA*/*bssA*/*nmsA* homologs recovered from TOLDC, SCADC and NAPDC using tBLASTn (Gertz *et al.*, 2006). Closely related sequences were recovered from the NCBI nr-database through BLASTX searches. All sequences were aligned using Muscle 3.3 (Edgar 2004), followed by manual adjustment of poorly aligned regions and deletion of indels. Maximum likelihood tree was constructed using PhyML (Guindon *et al.*, 2010) with WAG model and 100 bootstrap replicates in Geneious R7. Bootstrap support  $\geq 60\%$  is indicated. Full length translated pyruvate formate lyase (PFL) genes were used as an outgroup. Sequence length of genes recovered from TOLDC, SCADC and NAPDC are indicated in brackets. A tree with the same overall topology was obtained when including only full-length sequences with gaps.

Full-length *bssA* genes with high sequence homology were detected in TOLDC, SCADC and NAPDC (SCADC\_ bssA2, TOLDC\_ bssA1 and NAPDC\_ bssA3; Figure 5.2), implying similar substrate spectra and/or phylogeny. Only a single copy of *bssA* was detected in TOLDC, as reported

previously (Fowler *et al.*, 2012), indicating that TOLDC is highly enriched in a dominant toluene degrader (likely to be the dominant Firmicutes-related OTU, Table 5.3) unlike SCADC which continues to harbour diverse *assA* genes. The SCADC metagenome appears to contain a phylogenetically distinct *bssA* gene (SCADC\_bssA1) with the best BLASTp hit to the *bssA* of *Desulfobacula toluolica* (YP\_006759359, similarity 55%); this sequence was not detected in the NAPDC metagenome. Whereas all three hydrocarbon-degrading metagenomes contained *assA* and *bssA*, only NAPDC and SCADC contained *nmsA*. NmsA-like enzymes are not necessarily unique to degradation of polyaromatic hydrocarbon, i.e., 2-methylnaphthalene, such as that demonstrated in enrichment culture N47 and NaphS2/NaphS6 because *Desulfotomaculum* sp. Ox39 carrying a *nmsA* homolog (EF123665) is incapable of growth on 2-methylnaphthalene. NmsA homologs have instead been implicated in the degradation of xylenes and toluene (Acosta-González *et al.*, 2013, von Netzer *et al.*, 2013). This possibility is unknown in SCADC and NAPDC because these cultures have not been tested for their capability to degrade 2-methylnaphthalene. A number of related genes encoding glyceryl radical enzymes that did not group with the fumarate addition clades were identified (not shown in Figure 5.2), but their function is currently unknown and they appear to be more closely related to pyruvate formate lyase than to known fumarate addition genes. Although NAPDC had been enriched on naphtha, which contains both alkanes and monoaromatic hydrocarbons, the fumarate addition genes did not appear to be more diverse than genes encoding similar functions in TOLDC and SCADC. Our results demonstrate that all three cultures have the genetic potential to degrade alkanes, and that, additionally, NAPDC and SCADC might degrade PAHs like 2-MN as well as monoaromatic hydrocarbons.

Previous studies have shown that fumarate addition enzymes can be promiscuous and can co-activate several structurally unrelated substrates (Beller and Spormann 1997, Rabus *et al.*, 2011, Wilkes *et al.*, 2003). This is likely due to co-metabolism and gene induction in the presence of target compounds. For example, *Azoarcus* spp. HxN1 which activates *n*-alkanes with ASS can also co-

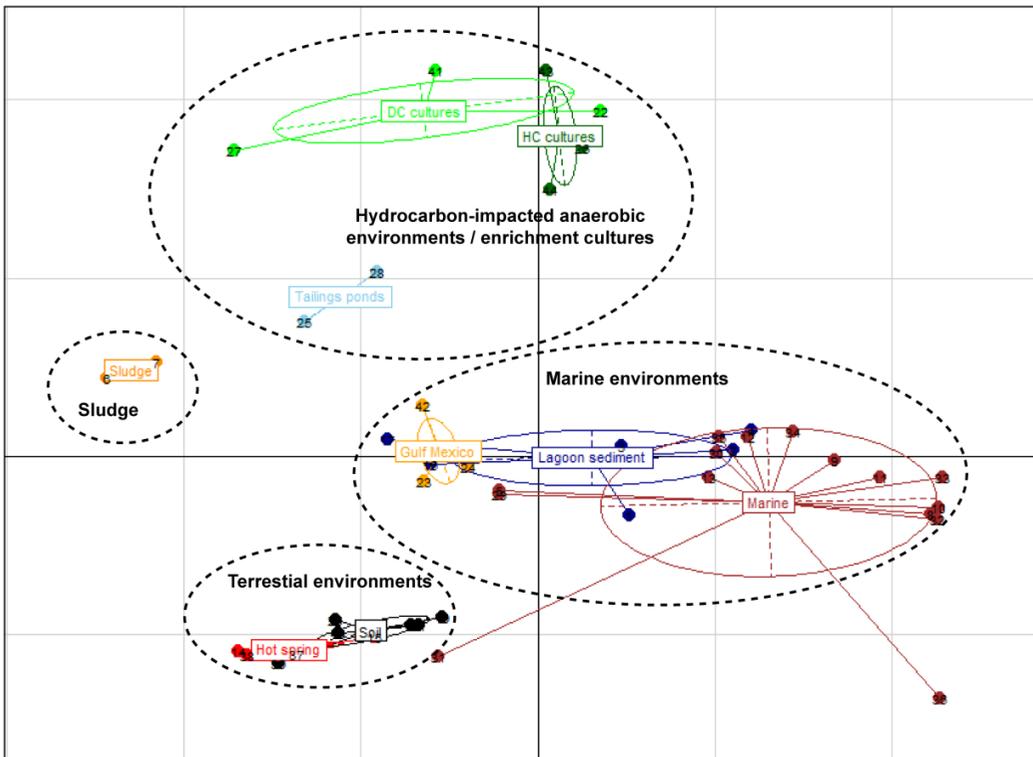
activate *iso*- and *cyclo*-alkanes in the presence of *n*-alkanes, but unlikely to activate *iso*- and *cyclo*-alkanes in the absence of *n*-alkane; inferring that the target compound is needed for gene induction; Similarly, ASS in sulfate-reducing bacteria is capable of activating toluene in the presence of *n*-alkanes (Rabus *et al.*, 2011). Therefore, cultures with broad substrate range like TOLDC, SCADC and NAPDC may have biotechnological applications such as hydrocarbon bioremediation, since petroleum pollutants often comprise structurally diverse compounds. Nonetheless, further laboratory studies are needed to define the complete range of substrates that these cultures can activate by fumarate addition.

#### **5.4.6 TOLDC, SCADC and NAPDC communities have similar genetic capabilities but differ from natural environments previously exposed to hydrocarbons**

In order to explore the functional similarity of TOLDC, SCADC and NAPDC, we performed a principal component analysis (PCA) based on the abundance of SEED functional categories of three metagenomes relative to each other, and to 41 metagenomes from a range of environments including soil, sediment, sludge, marine terrestrial environments, oil sands tailings ponds, and Gulf of Mexico deep marine sediments that had recently experienced oil spills (Appendix Table E1). At the time of analysis, no other metagenomes from methanogenic hydrocarbon environments were available in the MG-RAST database, precluding comparisons to environments that might exhibit highly similar geochemical processes.

Overall, the 44 selected environments and enrichment cultures were more functionally related to each other than to metagenomes derived from other environments (Figure 5.3). This is likely due to selection pressures imposed by environmental conditions. TOLDC, SCADC and NAPDC are more similar to each other than to all other environments including Gulf of Mexico deep marine sediments recently impacted by oil spills (GoM) and oil sands tailings ponds (MLSB and TP6; discussed further in subsequent section). Three-way analysis as described by Hug *et al.*, (2012) was further performed using SEED subsystems category level three to compare TOLDC, SCADC and NAPDC. Overall, all

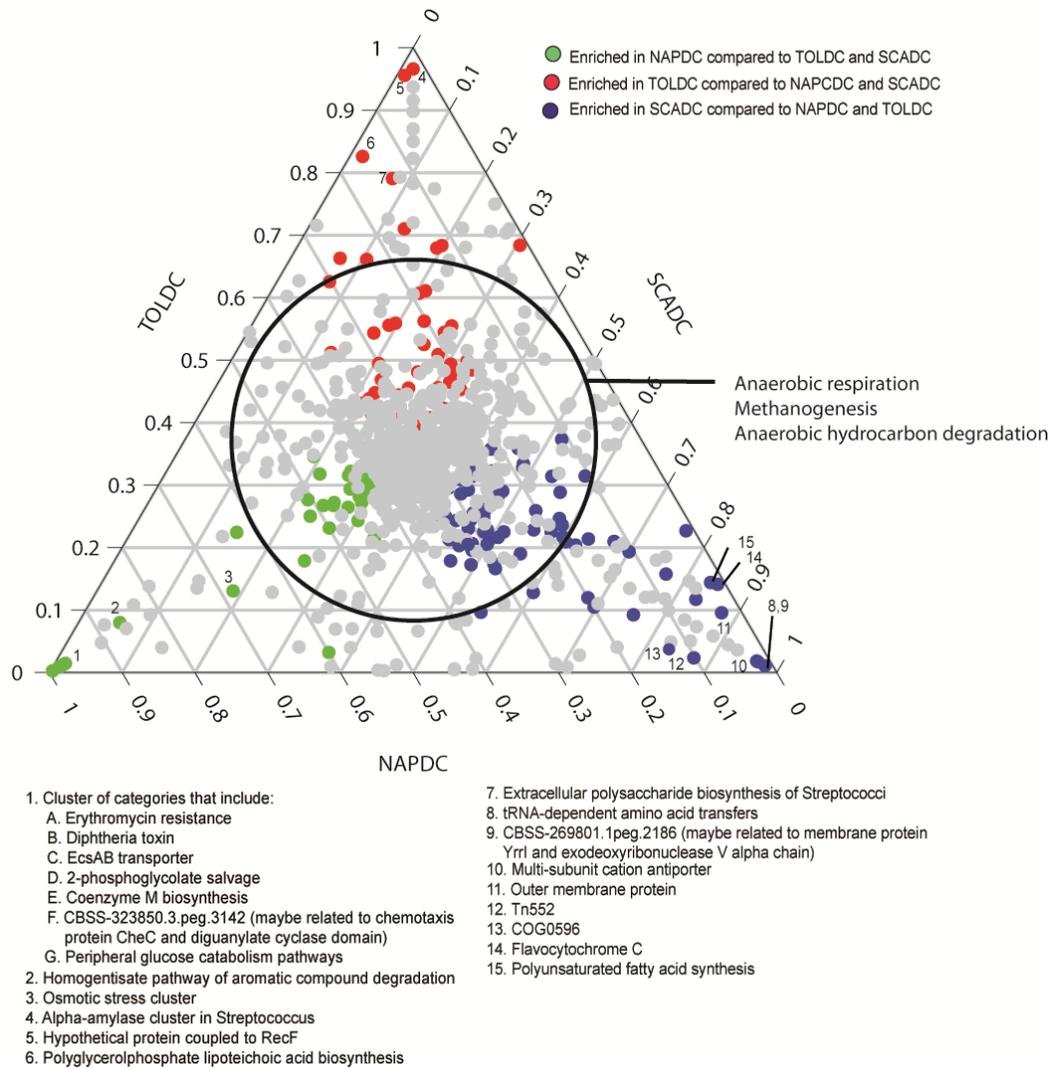
functions crucial for methanogenesis and anaerobic processes were highly conserved in the three metagenomes, and only a small proportion of functions not known to be crucial in anaerobic processes were conserved in each metagenome. These functions are likely associated with the dominant microorganisms detected in each metagenome and likely do not have important roles in hydrocarbon degradation (Figure 5.4 and Appendix Table E3).



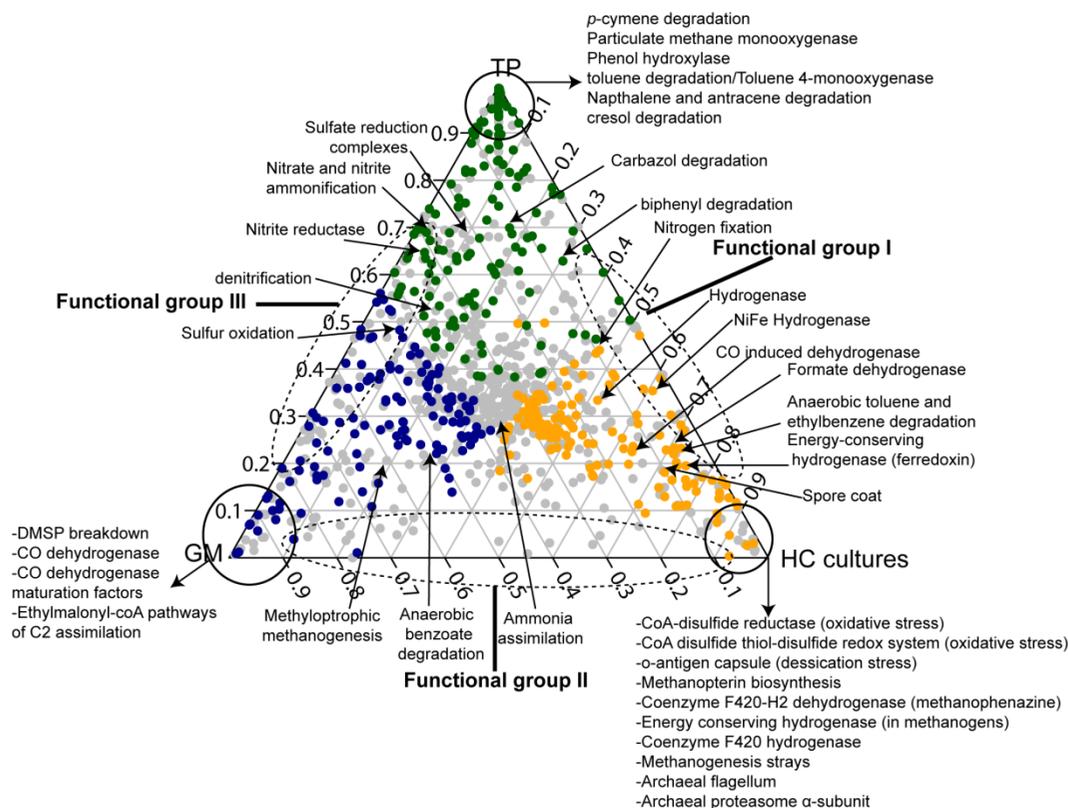
**Figure 5.3** Principal component analysis of the functional categories in 41 published metagenomes from diverse environments (Appendix Table D1) plus the metagenomes of the three hydrocarbon-degrading (HC) cultures. DC, dechlorinating cultures. All counts were normalized against total annotated sequences of each metagenome. Symbols are colored according to their environment of origin and metagenomes that share similar environmental origin are in broken circles.

In environments that had previous exposure to hydrocarbons (in addition to the dechlorinating cultures), reads associated with Methanomicrobia in different proportions could be detected (i.e., <5% in Gulf of Mexico to ~ 10% in oil sands tailings ponds) (Figure 5.1), suggesting overall anoxic and/or

methanogenic conditions. Three-way analysis as described by Hug *et al.*, (2012) was further performed using SEED subsystems category level three to compare three groups of environments: methanogenic hydrocarbon degrading enrichment cultures (TOLDC, SCADC and NAPDC), oil sands tailings ponds (MLSB and TP6), and Gulf of Mexico marine deep sediments (GoM23, GoM278 and GoM315). Even though these environments have in general very different physiochemical properties, genes encoding enzymes responsible for anaerobic hydrocarbon addition to fumarate had been detected previously (An *et al.*, 2013; Kimes *et al.*,2013). In addition, metabolite analysis of samples collected from Gulf of Mexico marine deep sediments detected putative benzylsuccinate (Kimes *et al.*, 2013). Together, this suggests that anaerobic hydrocarbon addition to fumarate may be a common mechanism of hydrocarbon degradation in these environments.



**Figure 5.4** Comparison of shared functional categories among the three hydrocarbon-degrading cultures (NAPDC, SCADC and TOLDC). Each point on the ternary plot represents a subsystem category of the three metagenomes, with the proportion of each SEED being normalized to a value of 1.0. Data points are coloured according to each metagenome. Points located near the vertices are enriched within the metagenome associated with that vertex, whereas points located near the centre have similar proportions in all three metagenomes (i.e., show no specific enrichment), as explained by Hug *et al.* (2012). Grey dots represents functional categories present at lower abundance (not statistically significant).



**Figure 5.5** Ternary plot showing three-way comparisons of functional categories in SEED subsystems level 3 for three different groups of metagenomes (GM; three Gulf of Mexico deep marine sediments; TP; two oil sands tailings ponds; HC: the three hydrocarbon-degrading cultures). Each point on the ternary plot represents a subsystem category of the three groups of pooled metagenomes, with the proportion of each being normalized to a value of 1.0. Data points are coloured according to each pooled group: blue, Gulf of Mexico sediments; green, oil sands tailings ponds; yellow, three enrichment cultures. Points located near the vertices are enriched within the metagenomes associated with that vertex, whereas points located near the centre have similar proportions in all eight metagenomes (e.g., show no specific enrichment; Hug *et al.* 2012). Specific functional categories are indicated with arrows and identified. Three groups of functional categories are indicated by broken circles (see main text).

#### 5.4.6.1 Oil sands tailings ponds

Syncrude's Mildred Lake Settling Basin (MLSB) from which SCADC and NAPDC were originally enriched from is highly methanogenic (Fedorak *et al.*, 2002). Tailings Pond 6 is managed by Suncor, a different oil sands extraction company, and both companies use naphtha as a diluent in the bitumen extraction

process (Chalaturnyk *et al.*, 2002). Previous studies have shown that certain fractions of naphtha can serve as methanogenic substrates (Siddique *et al.*, 2006, Siddique *et al.*, 2007b, Siddique *et al.*, 2011, Siddique *et al.*, 2012). The metagenomes obtained from these two different oil sands tailings ponds [Tailings pond 6 reported by An *et al.*, (2013) and MLSB, unpublished] had very similar overall genetic potentials based on the PCA analysis in Figure 5.3 but were not closely related to SCADC and NAPDC that were originally enriched from the mature fine tailings of MLSB (Figure 5.3). Suncor Tailings Pond 6 harbours microorganisms capable of anaerobic degradation of alkanes and aromatic compounds by addition to fumarate (An *et al.*, 2013) and similar genetic potentials have been detected in Syncrude MLSB (Chapter 6). Both tailings pond metagenomes analyzed in the current study harbour a high abundance of Betaproteobacteria, most of which are facultative anaerobes, and less abundant of methanogens compared to NAPDC and SCADC (Figure 5.1). In deeper layer of TP6, microbial communities were represented by an abundance of syntrophs (i.e., *Syntrophus*, *Smithella* and *Pelotomaculum*), sulfate-reducing bacteria (SRB) and methanogens (Ramos-Padron *et al.*, 2011).

The metagenomes of the two oil sands tailings ponds are highly enriched in functions associated with aerobic degradation of pollutants such as toluene, naphthalene, anthracene, cresol and methane (Figure 5.5), which are not detected in TOLDC, SCADC and NAPDC. These results reflect the high abundance of reads assigned to Beta- and Gammaproteobacteria, including *Burkholderia*, *Pseudomonas*, and *Rhodococcus* that are well known to be capable of catalyzing aerobic degradation of various pollutants using mono- and dioxygenases (Das and Chandran 2011). Supporting this, studies using microcosms (Semple and Foght, unpublished) have confirmed that naphtha hydrocarbons are rapidly degraded aerobically by tailings microbes. The importance of aerobic processes in methanogenic oil sands tailings ponds under *in situ* conditions has not been confirmed and warrants further investigation. Although genes for denitrification are present it is unlikely to be an important process *in situ* because of the lack of nitrate as the terminal electron acceptors (Penner and Foght 2010). Stimulation of

nitrate-reducing bacteria using various means may be favourable for tailings management as this increases the densification rate of fine tailings, and therefore allows rapid water recycling (Bordenave *et al.*, 2010). Detection of sulfate reduction and sulfur oxidation genes (Figure 5.5) in addition to physiochemical experiments reported earlier (Holowenko *et al.*, 2000, Salloum *et al.*, 2002) implicate the importance of sulfur cycling *in situ*.

Overall, oil sands tailings ponds appear to share some functions common to anaerobic respiration (i.e., NiFe hydrogenases, hydrogenases; denoted in Functional Group I in Figure 5.5) with the three hydrocarbon-degrading cultures but overall had fewer functions associated with methanogenesis, which are enriched only in the three cultures (Figures 5.1 and 5.5). The low abundance of Betaproteobacteria in SCADC and NAPDC is most likely due to the culturing process that is favourable for anaerobic microbial communities, in particular methanogens that are capable of syntrophic growth with primary degraders capable of anaerobic hydrocarbon degradation. It is not surprising that TOLDC, SCADC and NAPDC share similar genetic potentials even though TOLDC was enriched from a very different environment (Fort Lupton, Table 5.1) as this illustrates that methanogenic cultures consist of very specialized microbial communities and therefore the establishment of these cultures can allow detailed studies of syntrophic processes of complex communities.

#### **5.4.6.2 Environments impacted by hydrocarbons**

Three metagenomes from Gulf of Mexico deep marine sediments (GoM23, GoM278 and GoM 315) that recently experienced the Deepwater Horizon (DWH) oil spill appear to have overall similar functional potentials (Kime *et al.*, 2013) and also with microbial communities in lagoon sediments and marine water (Figure 5.3), consistent with their marine origin. The Gulf of Mexico deep marine sediments do not share much functional potential with the three enrichment cultures (Functional Group II in Figure 5.5) but had more in common with oil sands tailings ponds regarding functional potentials common in anoxic environments such as sulfur oxidation and denitrification processes (Functional Group III in Figure 5.5). Even though these samples were collected

near the DWH oil spill, the microbial communities are not similar to the microbial communities of oil sands tailings ponds and the methanogenic enrichment cultures, which are respectively dominated by Betaproteobacteria and methanogens (Figure 5.1). Instead, the microbial communities in the Gulf of Mexico marine sediments are dominated by Alphaproteobacteria (26-37% of total metagenomic reads with assigned taxa) and Gammaproteobacteria (33%), especially at sites further away from the blowout (GoM278 and GoM023)(Kimes *et al.*, 2013). The metagenomes from Gulf of Mexico had a relatively high abundance of reads assigned to CO dehydrogenase, dimethylsulfoniopropionate (DMSP) breakdown and a number of clustering-based subsystems with unknown functions (Figure 5.5). DMSP is produced by marine phytoplankton as an osmoprotectant, and can be released into the environments during bloom and died-off (Yoch 2002). Many microorganisms involved in DMSP breakdown are anaerobes and are capable of CO<sub>2</sub> fixation using CO dehydrogenase (vanderMaarel *et al.*, 1996, Yoch 2002).

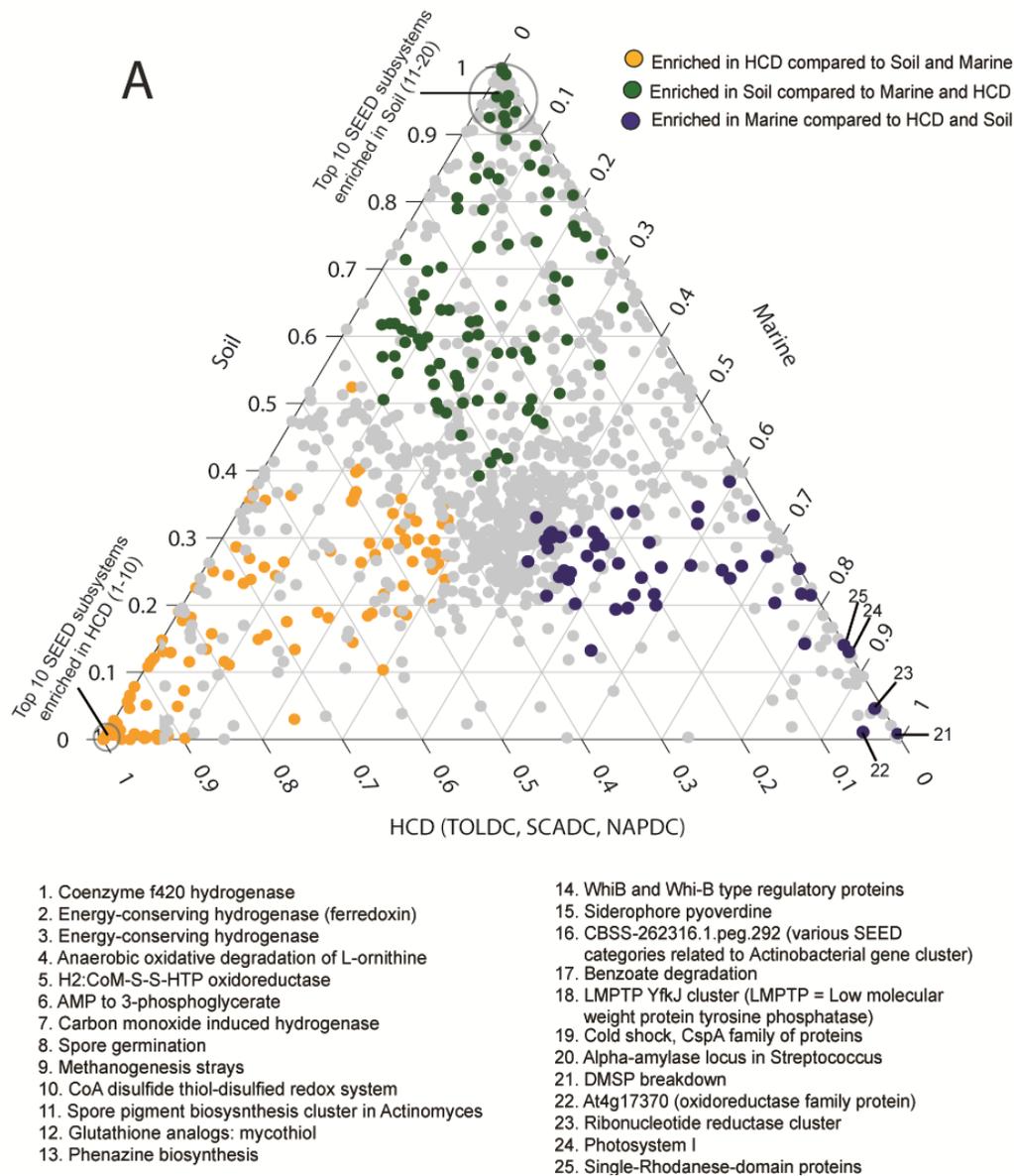
#### **5.4.6.3 Functional comparisons to other environments**

To further explore factors that distinguish methanogenic hydrocarbon-degrading enrichments from other environmental metagenomes, we performed a three-way analysis by pooling groups of metagenomic reads from NAPDC, SCADC and TOLDC with the metagenomes from soil and marine environments, then compared functional categories at SEED subsystems level 3 as described above (Figure 5.6). Not surprisingly, functional categories that were most highly enriched in the methanogenic hydrocarbon-degrading enrichment cultures included those related to methanogenesis (Coenzyme F<sub>420</sub> hydrogenase, hydrogen CoM-S-S-HTP oxidoreductase, methanogenesis strays (related to unidentified component of methyl-CoA reductase), energy conservation and electron and/or hydrogen transfer (energy-conserving hydrogenases), oxidative stress (CoA disulfide thiol-disulfide redox system) and Archaea-specific pathways (AMP to 3-phosphoglycerate). Functions enriched in soil and marine metagenomes are consistent with expected metabolic processes in these environments. For example, soil metagenomes are dominated by functions related to Actinomycetes,

Mycobacteria and Pseudomonads, in particular degradation pathways for benzoate and other toxic compounds, as well as functions related to interspecies competition. Functions enriched in marine communities were primarily related to sulfur cycling and photosynthesis.

## **5.5 Conclusions**

Overall, the current study integrated and compared three metagenomes of methanogenic hydrocarbon degrading cultures enriched with toluene (TOLDC), alkanes (SCADC) or naphtha (NAPDC). Metabolite profiling and phylogenetic assessment of key genes indicate that hydrocarbon addition to fumarate may be an essential bottleneck reaction in these methanogenic cultures. Repeated transfers of TOLDC, SCADC and NAPDC preserve functions that are essential in activation of structurally different hydrocarbons (i.e., alkanes and monoaromatics). More importantly, these three cultures are capable of degrading their expected substrates even after multiple transfers. The knowledge based on inferred substrate spectrum obtained from metagenomic analysis in addition to physiological characterization of these cultures is important for the development of cultures with biotechnological applications such as bioremediation. Comparison of these three enrichment cultures to hydrocarbon-impacted environments suggests that common functional potentials are preserved but environments select for functions important for *in situ* geochemical processes. Detection of genes encoding fumarate addition in these cultures and other methanogenic environments such as Gulf of Mexico deep marine sediments and oil sands tailings ponds suggest that this mechanism is perhaps wide-spread in many other methanogenic environments. Overall, the findings provide a solid foundation for comparison with other anaerobic hydrocarbon-impacted ecosystems.



**Figure 5.6** Ternary plot showing three-way comparison of functional categories in SEED subsystems level 3 for different groups of metagenomes. The plot shows the comparison of shared functions in NAPDC, SCADC, and TOLDC (HCD) with soil (six) and marine (eleven) metagenomes. Each point on the ternary plot represents a subsystem category of the three groups of pooled metagenomes, with the proportion of each being normalized to a value of 1.0. Points located near the vertices are enriched within the metagenome associated with that vertex, whereas points located near the centre have similar proportions in all twenty metagenomes (e.g., show no specific enrichment; Hug *et al.* (2012)).

## 5.6 References

- Abu Laban N, Selesi D, Rattei T, Tischler P, Meckenstock RU. (2010). Identification of enzymes involved in anaerobic benzene degradation by a strictly anaerobic iron-reducing enrichment culture. *Environ Microbiol* **12**: 2783-2796.
- Acosta-González A, Rossello-Mora R, Marques S. (2013). Diversity of benzylsuccinate synthase-like (*bssA*) genes in hydrocarbon-polluted marine sediments suggests substrate-dependent clustering. *Appl Environ Microbiol* **79**: 3667-3676.
- Agrawal A, Gieg LM. (2013). In situ detection of anaerobic alkane metabolites in subsurface environments. *Front Microbiol* **4**: 140.
- Ahring BK, Westermann P, Mah RA. (1991). Hydrogen inhibition of acetate metabolism and kinetics of hydrogen consumption by *Methanosarcina Thermophila* TM-1. *Arch Microbiol* **157**: 38-42.
- Aitken CM, Jones DM, Maguire MJ, Gray ND, Sherry A, Bowler BFJ *et al.*, (2013). Evidence that crude oil alkane activation proceeds by different mechanisms under sulfate-reducing and methanogenic conditions. *Geochim Cosmochim Acta* **109**: 162-174.
- An D, Brown D, Chatterjee I, Dong X, Ramos-Padron E, Wilson S, Bordenave S, Caffrey S, Gieg L, Sensen C, and Gerrit Voordouw, G. (2013) Microbial community and potential functional gene diversity involved in anaerobic hydrocarbon degradation and methanogenesis in an oil sands tailings pond. *Genome*, 10.1139/gen-2013-0083.
- Beller HR, Spormann AM. (1997). Anaerobic activation of toluene and o-xylene by addition to fumarate in denitrifying strain T. *J Bacteriol* **179**: 670-676.
- Beller HR, Edwards EA. (2000). Anaerobic toluene activation by benzylsuccinate synthase in a highly enriched methanogenic culture. *Appl Environ Microbiol* **66**: 5503-5505.
- Biegert T, Fuchs G, Heider F. (1996). Evidence that anaerobic oxidation of toluene in the denitrifying bacterium *Thauera aromatica* is initiated by formation of benzylsuccinate from toluene and fumarate. *Eur J Biochem* **238**: 661-668.
- Bordenave S, Kostenko V, Dutkoski M, Grigoryan A, Martinuzzi RJ, Voordouw G. (2010). Relation between the activity of anaerobic microbial populations in oil sands tailings ponds and the sedimentation of tailings. *Chemosphere* **81**: 663-668.

Brisson VL, West KA, Lee PKH, Tringe SG, Brodie EL, Alvarez-Cohen L. (2012). Metagenomic analysis of a stable trichloroethene-degrading microbial community. *ISME* **6**: 1702-1714.

Callaghan AV, Davidova IA, Savage-Ashlock K, Parisi VA, Gieg LM, Suflita JM *et al.*,. (2010). Diversity of benzyl- and alkylsuccinate synthase genes in hydrocarbon-impacted environments and enrichment cultures. *Environ Sci Technol* **44**: 7287-7294.

Callaghan AV, Morris BEL, Pereira IAC, McInerney MJ, Austin RN, Groves JT *et al.*,. (2012). The genome sequence of *Desulfatibacillum alkenivorans* AK-01: a blueprint for anaerobic alkane oxidation. *Environ Microbiol* **14**: 101-113.

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

Chalaturnyk RJ, Scott JD, Ozum B. (2002). Management of oil sands tailings. *Petrol Sci Technol* **20**: 1025-1046.

Chang W, Um Y, Holoman TRP. (2005). Molecular characterization of anaerobic microbial communities from benzene-degrading sediments under methanogenic conditions. *Biotechnol Prog* **21**: 1789-1794.

Das N, Chandran P. (2011). Microbial degradation of petroleum hydrocarbon contaminants: an overview. *Biotechnol Res Int* **2011**: 941810.

Dianou D, Miyaki T, Asakawa S, Morii H, Nagaoka K, Oyaizu H *et al.*,. (2001). *Methanoculleus chikugoensis* sp nov., a novel methanogenic archaeon isolated from paddy field soil in Japan, and DNA-DNA hybridization among *Methanoculleus* species. *Int J Syst Evol Microbiol* **51**: 1663-1669.

Edgar RC. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32**: 1792-1797.

Edwards EA, Grbić-Galić D. (1994). Anaerobic degradation of toluene and o-xylene by a methanogenic consortium. *Appl Environ Microbiol* **60**: 313-322.

Embree M, Nagarajan H, Movahedi N, Chitsaz H, Zengler K. (2013). Single-cell genome and metatranscriptome sequencing reveal metabolic interactions of an alkane-degrading methanogenic community. *ISMEI*. Advance online publication.

Fedorak PM, Coy DL, Salloum MJ, Dudas MJ. (2002). Methanogenic potential of tailings samples from oil sands extraction plants. *Canadian Journal of Microbiology* **48**: 21-33.

- Ficker M, Krastel K, Orlicky S, Edwards E. (1999). Molecular characterization of a toluene-degrading methanogenic consortium. *Appl Environ Microbiol* **65**: 5576-5585.
- Foght J, Aislabie J, Turner S, Brown CE, Ryburn J, Saul DJ *et al.*,. (2004). Culturable bacteria in subglacial sediments and ice from two Southern Hemisphere glaciers. *Microb Ecol* **47**: 329-340.
- Foght J. (2008). Anaerobic biodegradation of aromatic hydrocarbons: Pathways and prospects. *J Mol Microbiol Biotechnol* **15**: 93-120.
- Fowler SJ, Dong XL, Sensen CW, Suflita JM, Gieg LM. (2012). Methanogenic toluene metabolism: community structure and intermediates. *Environ Microbiol* **14**: 754-764.
- Gertz EM, Yu YK, Agarwala R, Schaffer AA, Altschul SF. (2006). Composition-based statistics and translated nucleotide searches: Improving the TBLASTN module of BLAST. *BMC Biol* **4**.
- Gray ND, Sherry A, Grant RJ, Rowan AK, Hubert CRJ, Callbeck CM *et al.*,. (2011). The quantitative significance of Syntrophaceae and syntrophic partnerships in methanogenic degradation of crude oil alkanes. *Environ Microbiol* **13**: 2957-2975.
- Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. *Syst Biol* **59**: 307-321.
- Haque MM, Ghosh TS, Komanduri D, Mande SS. (2009). SOrt-ITEMS: Sequence orthology based approach for improved taxonomic estimation of metagenomic sequences. *Bioinformatics* **25**: 1722-1730.
- Heider J. (2007). Adding handles to unhandy substrates: anaerobic hydrocarbon activation mechanisms. *Curr Opin Chem Biol* **11**: 188-194.
- Hildenbrand C, Stock T, Lange C, Rother M, Soppa J. (2011). Genome copy numbers and gene conversion in methanogenic archaea. *J Bacteriol* **193**: 734-743.
- Holowenko FM, MacKinnon MD, Fedorak PM. (2000). Methanogens and sulfate-reducing bacteria in oil sands fine tailings waste. *Can J Microbiol* **46**: 927-937.
- Hug LA, Beiko RG, Rowe AR, Richardson RE, Edwards EA. (2012). Comparative metagenomics of three *Dehalococcoides*-containing enrichment cultures: the role of the non-dechlorinating community. *BMC Genomics* **13**.

Jones DM, Head IM, Gray ND, Adams JJ, Rowan AK, Aitken CM *et al.*, (2008). Crude-oil biodegradation via methanogenesis in subsurface petroleum reservoirs. *Nature* **451**: 176-U176.

Kato S, Watanabe K. (2010). Ecological and evolutionary interactions in syntrophic methanogenic consortia. *Microbes Environ* **25**: 145-151.

Kimes NE, Callaghan AV, Aktas DF, Smith WL, Sunner J, Golding B *et al.*, (2013). Metagenomic analysis and metabolite profiling of deep-sea sediments from the Gulf of Mexico following the Deepwater Horizon oil spill. *Front Microbiol* **4**: 50.

Kleinsteuber S, Schleinitz KM, Vogt C. (2012). Key players and team play: anaerobic microbial communities in hydrocarbon-contaminated aquifers. *Appl Microbiol Biotechnol* **94**: 851-873.

Kniemeyer O, Heider J. (2001). Ethylbenzene dehydrogenase, a novel hydrocarbon-oxidizing molybdenum/iron-sulfur/heme enzyme. *J Biol Chem* **276**: 21381-21386.

Kniemeyer O, Musat F, Sievert SM, Knittel K, Wilkes H, Blumenberg M *et al.*, (2007). Anaerobic oxidation of short-chain hydrocarbons by marine sulphate-reducing bacteria. *Nature* **449**: 898-U810.

Kunin V, Engelbrektson A, Ochman H, Hugenholtz P. (2010). Wrinkles in the rare biosphere: pyrosequencing errors can lead to artificial inflation of diversity estimates. *Environ Microbiol* **12**: 118-123.

Lykidis A, Chen CL, Tringe SG, McHardy AC, Copeland A, Kyrpides NC *et al.*, (2011). Multiple syntrophic interactions in a terephthalate-degrading methanogenic consortium. *ISME Journal* **5**: 122-130.

Maestrojuan GM, Boone DR. (1991). Characterization of *Methanosarcina barkeri* Mst and 227, *Methanosarcina mazei* S-6t, and *Methanosarcina vacuolata* Z-761t. *Int J Syst Bacteriol* **41**: 267-274.

Mbadinga SM, Wang LY, Zhou L, Liu JF, Gu JD, Mu BZ. (2011). Microbial communities involved in anaerobic degradation of alkanes. *Int Biodeterior Biodegrad* **65**: 1-13.

Meyer F, Paarmann D, D'Souza M, Olson R, Glass EM, Kubal M *et al.*, (2008). The metagenomics RAST server - a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC Bioinformatics* **9**.

Parks DH, Beiko RG. (2010). Identifying biologically relevant differences between metagenomic communities. *Bioinformatics* **26**: 715-721.

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

Penner TJ, Foght JM. (2010). Mature fine tailings from oil sands processing harbour diverse methanogenic communities. *Canadian Journal of Microbiology* **56**: 459-470.

Qiu YL, Sekiguchi Y, Imachi H, Kamagata Y, Tseng IC, Cheng SS *et al.*, (2004). Identification and isolation of anaerobic, syntrophic phthalate isomer-degrading microbes from methanogenic sludges treating wastewater from terephthalate manufacturing. *Appl Environ Microbiol* **70**: 1617-1626.

Rabus R, Jarling R, Lahme S, Kuhner S, Heider J, Widdel F *et al.*, (2011). Co-metabolic conversion of toluene in anaerobic n-alkane-degrading bacteria. *Environ Microbiol* **13**: 2576-2585.

Ramos-Padron E, Bordenave S, Lin SP, Bhaskar IM, Dong XL, Sensen CW *et al.*, (2011). Carbon and sulfur cycling by microbial communities in a gypsum-treated oil sands tailings pond. *Environ Sci Technol* **45**: 439-446.

Rios-Hernandez LA, Gieg LM, Suflita JM. (2003). Biodegradation of an alicyclic hydrocarbon by a sulfate-reducing enrichment from a gas condensate-contaminated aquifer. *Appl Environ Microbiol* **69**: 434-443.

Sakai N, Kurisu F, Yagi O, Nakajima F, Yamamoto K. (2009). Identification of putative benzene-degrading bacteria in methanogenic enrichment cultures. *J Biosci Bioeng* **108**: 501-507.

Salloum MJ, Dudas MJ, Fedorak PM. (2002). Microbial reduction of amended sulfate in anaerobic mature fine tailings from oil sand. *Waste Manage Res* **20**: 162-171.

Schink B. (1997). Energetics of syntrophic cooperation in methanogenic degradation. *Microbiol Mol Biol Rev* **61**: 262-&.

Schloss PD, Westcott SL. (2011). Assessing and improving methods used in operational taxonomic unit-based approaches for 16S rRNA gene sequence analysis. *Appl Environ Microbiol* **77**: 3219-3226.

Siddique T, Fedorak PM, Foght JM. (2006). Biodegradation of short-chain n-alkanes in oil sands tailings under methanogenic conditions. *Environ Sci Technol* **40**: 5459-5464.

Siddique T, Fedorak PM, McKinnon MD, Foght JM. (2007). Metabolism of BTEX and naphtha compounds to methane in oil sands tailings. *Environ Sci Technol* **41**: 2350-2356.

- Siddique T, Penner T, Semple K, Foght JM. (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, Nesbo C, Foght JM. (2012). Microbial Communities Involved in Methane Production from Hydrocarbons in Oil Sands Tailings. *Environ Sci Technol* **46**: 9802-9810.
- So CM, Phelps CD, Young LY. (2003). Anaerobic transformation of alkanes to fatty acids by a sulfate-reducing bacterium, strain Hxd3. *Appl Environ Microbiol* **69**: 3892-3900.
- Soh J, Dong X, Caffrey SM, Voordouw G, Sensen CW. (2013). Phoenix 2: A locally installable large-scale 16S rRNA gene sequence analysis pipeline with Web interface. *J Biotechnol* **167**: 393-403.
- Tan B, Dong XL, Sensen CW, Foght JM. (2013). Metagenomic analysis of an anaerobic alkane-degrading microbial culture: potential hydrocarbon-activating pathways and inferred roles of community members. *Genome* **56**: 599-611.
- Ulrich AC, Edwards EA. (2003). Physiological and molecular characterization of anaerobic benzene-degrading mixed cultures. *Environ Microbiol* **5**: 92-102.
- vanderMaarel MJEC, Jansen M, Haanstra R, Meijer WG, Hansen TA. (1996). Demethylation of dimethylsulfoniopropionate to 3-S-methylmercaptopropionate by marine sulfate-reducing bacteria. *Appl Environ Microbiol* **62**: 3978-3984.
- von Netzer F, Pilloni G, Kleindienst S, Kruger M, Knittel K, Grundger F *et al.*, (2013). Enhanced gene detection assays for fumarate-adding enzymes allow uncovering of anaerobic hydrocarbon degraders in terrestrial and marine systems. *Appl Environ Microbiol* **79**: 543-552.
- Walker CB, Redding-Johanson AM, Baidoo EE, Rajeev L, He Z, Hendrickson EL *et al.*, (2012). Functional responses of methanogenic archaea to syntrophic growth. *ISME J* **6**: 2045-2055.
- Washer CE, Edwards EA. (2007). Identification and expression of benzylsuccinate synthase genes in a toluene-degrading methanogenic consortium. *Appl Environ Microbiol* **73**: 1367-1369.
- Wawrik B, Mendivelso M, Parisi VA, Suflita JM, Davidova IA, Marks CR *et al.*, (2012). Field and laboratory studies on the bioconversion of coal to methane in the San Juan Basin. *FEMS Microbiol Ecol* **81**: 26-42.

- Widdel F, Bak F (1992). Gram-negative mesophilic sulfate-reducing bacteria. In: Balows A, Trüper, H. G., Dworkin, M., Harder, W. and Schleifer, K.H. (ed). *The Prokaryotes*. Springer-Verlag: New York, USA. pp 3352-3378.
- Wilkes H, Kuhner S, Bolm C, Fischer T, Classen A, Widdel F *et al.*, (2003). Formation of n-alkane- and cycloalkane-derived organic acids during anaerobic growth of a denitrifying bacterium with crude oil. *Org Geochem* **34**: 1313-1323.
- Winderl C, Penning H, von Netzer F, Meckenstock RU, Lueders T. (2010). DNA-SIP identifies sulfate-reducing Clostridia as important toluene degraders in tar-oil-contaminated aquifer sediment. *Isme Journal* **4**: 1314-1325.
- Yoch DC. (2002). Dimethylsulfoniopropionate: Its sources, role in the marine food web, and biological degradation to dimethylsulfide. *Appl Environ Microbiol* **68**: 5804-5815.
- Zellner G, Messner P, Winter J, Stackebrandt E. (1998). *Methanoculleus palmolei* sp. nov., an irregularly coccoid methanogen from an anaerobic digester treating wastewater of a palm oil plant in North-Sumatra, Indonesia. *Int J Syst Bacteriol* **48**: 1111-1117.
- Zengler K, Richnow HH, Rossello-Mora R, Michaelis W, Widdel F. (1999). Methane formation from long-chain alkanes by anaerobic microorganisms. *Nature* **401**: 266-269.
- Zhu JX, Zheng HJ, Ai GM, Zhang GS, Liu D, Liu XL *et al.*, (2012). The genome characteristics and predicted function of methyl-group oxidation pathway in the obligate acetoclastic methanogens, *Methanosaeta* spp. *Plos One* **7**.

## 6 Assessing the functional diversity of fumarate addition genes in oil sands tailings ponds and other environments<sup>1</sup>

### 6.1 Abstract

Hydrocarbons in residual naphtha that are deposited along with tailings slurry into oil sands tailing ponds are substrates for methanogenesis used by tailings microorganisms. Fumarate addition through enzymes homologous to benzylsuccinate- (BSS), alkylsuccinate- (ASS) and naphthylsuccinate synthase (NMS) are well established as the primary enzymes involved in anaerobic degradation of toluene, alkanes and 2-methylnaphthalene, respectively. In order to understand methanogenic hydrocarbon degradation by tailings ponds organisms, microcosms using tailings pond mature fine tailings were established using naphtha, C<sub>6</sub>-C<sub>10</sub>, C<sub>14</sub>-C<sub>16</sub> or BTEX (benzene, toluene, ethylbenzene, xylenes) as the only carbon source plus substrate-free and heat-killed controls. Incubation of microcosms over ~490 days detected greater methane production in amended microcosms compared to heat-killed and substrate-free controls, indicating methane was produced from the degradation of hydrocarbons. PCR amplification targeting alpha-subunits of genes encoding homologs of BSS (*bssA*) and ASS (*assA*) from microcosms incubated for up to 150 d detected *bssA* only in microcosms amended with BTEX and naphtha; *assA* affiliated with *Smithella* and *Desulfotomaculum* had increased abundance in C<sub>14</sub>-C<sub>16</sub> and C<sub>6</sub>-C<sub>10</sub>-cultures, respectively, but not in BTEX-amended cultures. Results obtained from clone libraries could be corroborated with those obtained from qPCR amplification of total DNA isolated from microcosms incubated for ~150 and ~450 days by targeting selected *assA* and *bssA* genotypes. This indicates that a variety of tailings microorganisms are responsible for the anaerobic degradation of different hydrocarbons in naphtha. Further data mining for fumarate addition genes in oil sands tailing ponds and other related environments detected a diverse repertoire including novel clusters of fumarate addition genes, implicating their ubiquitous nature in these environments.

<sup>1</sup>A version of this chapter has been modified for publication to demonstrate the diversity of fumarate addition genes in oil sands tailings ponds and other environments

## 6.2 Introduction

Oil sands tailings ponds are artificial reservoirs created for the storage of oil sands tailings waste from bitumen extraction of oil sands ore (Chalaturnyk *et al.*, 2002). After less than two decades of operation, the Mildred Lake Settling Basin (MLSB), the largest oil sands tailings ponds operated by Syncrude, became highly methanogenic (Fedorak *et al.*, 2003). The sources of methanogenesis in MLSB have been unambiguously linked to the anaerobic biodegradation of hydrocarbon components in naphtha, which is used as a diluent in bitumen extraction (Siddique *et al.*, 2006, Siddique *et al.*, 2007). Naphtha is composed primarily of short-chain *n*-alkanes (C<sub>6</sub>, C<sub>7</sub>, C<sub>8</sub>, C<sub>9</sub>, C<sub>10</sub>), monoaromatic hydrocarbons (BTEX: benzene, toluene, ethylbenzene, *m*-, *o*-, and *p*-xylenes), *iso*- and *cyclo*-alkanes, at varying concentrations (Siddique *et al.*, 2007). Incubation of MLSB mature fine tailings (MFT), a consolidated form of settled tailings slurry containing naphtha hydrocarbons indicated that short-chain *n*-alkanes, toluene, ethylbenzene, *o*- and *m*-xylenes were preferentially utilized by indigenous microorganisms to generate methane (Siddique *et al.*, 2006, Siddique *et al.*, 2007). Indigenous microorganisms enriched from MLSB were shown to degrade longer-chain *n*-alkanes under methanogenic conditions, although these hydrocarbons are not part of naphtha diluent (Siddique *et al.*, 2011).

Studies conducted previously determined that the bacterial populations native to MLSB belonged to different physiological groups of sulfate-, nitrate-, and iron-reducing bacteria (Fedorak *et al.*, 2003, Holowenko *et al.*, 2000, Salloum *et al.*, 2002). Later, 16S rRNA gene surveys of samples collected from MLSB and Syncrude's West-In-Pit using PCR and cloning techniques revealed the presence of a diverse array of bacterial species and an archaeal population dominated by acetivlastic methanogens (Penner and Foght 2010). Although it has now become very clear that anaerobic biodegradation of naphtha components is one source of methanogenesis in MLSB, the potential hydrocarbon activation process, which is thought to be the bottleneck reaction in methanogenesis, remains undescribed.

Although anaerobic hydrocarbon degradation can proceed through several pathways (Callaghan 2013), the fumarate addition mechanism appear to be a more ubiquitous universal mechanism for anaerobic hydrocarbon degradation (Agrawal and Gieg 2013). Because of this, we investigated the importance of this process in oil sands tailings ponds using methanogenic microcosms established from Syncrude MLSB. Metagenomic approaches using cloning and sequencing of the *assA* (encoding the  $\alpha$ -subunit of alkylsuccinate synthase) and *bssA* genes ( $\alpha$ -subunit of benzylsuccinate synthase), followed by qPCR quantification of selected genotypes suggested that amended hydrocarbons in microcosms were degraded by fumarate addition. Furthermore, data mining of a MLSB metagenome generated using Illumina Hi-seq recovered a diverse array of *assA* and *nmsA* genes but lower diversity of *bssA* genes. The results presented here are important for the general understanding of hydrocarbon degradation in oil sands tailings ponds and provide consideration for the development of molecular tools in monitoring hydrocarbon degradation in oil sands tailings ponds and other anoxic hydrocarbon-impacted environments.

### **6.3 Materials and methods**

#### **6.3.1 Microcosm preparation and headspace gas chromatographic analyses**

A different set of methanogenic microcosms (not related to SCADC in Figure 1.3) was aseptically prepared by adding 25 mL of mature fine tailings from Mildred Lake Settling Basin to 50 ml of sterile carbonate-buffered methanogenic medium in 158 ml serum bottles under a headspace of O<sub>2</sub>-free 30% CO<sub>2</sub>-balance N<sub>2</sub> gas. The methanogenic medium was prepared according to Widdel and Bak (1992) and gassed with O<sub>2</sub>-free 30% CO<sub>2</sub>, balance N<sub>2</sub> for at least 30 min. A redox indicator (resazurin) and a reducing agent (sulfide) were added as described by Fedorak and Hrudey (1984). All microcosms were individually supplemented with one of the following as the only added organic carbon source: 0.2% v/v naphtha; or 0.1% v/v *n*-alkane mixture (equimolar C<sub>6</sub>, C<sub>7</sub>, C<sub>8</sub>, C<sub>10</sub>); or 0.1% v/v long-chain *n*-alkanes (equimolar C<sub>14</sub>, C<sub>16</sub>, C<sub>18</sub>); or 0.05% v/v BTEX (equimolar benzene, toluene, ethylbenzene, *m*-, *p*-, *o*-xylenes); or no addition (endogenous carbon as "baseline" control). All sources of pure hydrocarbon substrates used in

this study were obtained commercially from Sigma-Aldrich (US). The naphtha source used in this study, provided by Syncrude Canada Ltd., contained short-chain *n*-alkanes (C<sub>6</sub> to C<sub>10</sub>), BTEX compounds, *iso*- and *cyclo*-alkanes as previously reported (Siddique *et al.*, 2007). Substrate-free baseline and unamended heat-killed controls were incubated parallel to the amended microcosms to account for background methane production. All microcosms were incubated stationary in the dark at room temperature, and methane production during incubation was periodically measured by sampling 50 µL of culture headspace using a sterile needle and syringe followed by gas chromatography with a flame ionization detector (GC-FID) as previously described (Siddique *et al.*, 2006, Siddique *et al.*, 2007). Depletion of hydrocarbon substrates was detected by sampling 50 µL of culture headspace and injecting into an Agilent 6890N gas chromatograph with an inert mass selective detector (GC-MS; Agilent model 5973) fitted with an Agilent HP-5MS capillary column (30 m × 0.25 mm ID, 0.25 µm film thickness; J + W Scientific). The depletion of added hydrocarbons was qualitatively assessed by comparing the analyte peak area to that of methylcyclohexane and 1,1,3-trimethylcyclohexane present as recalcitrant background hydrocarbons in oil sands mature fine tailings (i.e., internal standards). For the enrichment culture amended with long-chain *n*-alkanes, samples were first extracted with hexane and acetone with pristane as an internal standard, followed by GC-analyses as described by Siddique *et al.*, (2011). The headspace of a heat-killed unamended control was analyzed to account for residual hydrocarbon carry-over from fresh MFT.

### **6.3.2 Nucleic acid extraction**

Nucleic acid extraction was performed as described earlier (Foght *et al.*, 2004). Briefly, 1 mL of sample was divided equally into triplicate tubes for DNA extraction by bead-beating. After isopropanol precipitation, DNA from triplicate extractions was eluted into 50 µL of sterile MilliQ water and stored at -20°C until analysis. A negative control containing sterile MilliQ water was used in each extraction trial, followed by PCR amplification (see below) to ensure reagents were devoid of contaminating DNA.

### 6.3.3 Primer design and validation

The detection of *assA* genes from our enrichment cultures was first performed using primer set 9 reported by Callaghan *et al.*, (2010). Amplicons of correct size were excised, cloned and sequenced as detailed below. Other primers targeting the *assA* genes have been recently reported (Aitken *et al.*, 2013, von Netzer *et al.*, 2013) but not tested in the current study. The selective amplification of *assA* genes using different primer sets was noted in Callaghan *et al.* (2010), and to increase chances of detecting *assA* genes that might be highly divergent, highly degenerate primers targeting the conserved regions of *bssA* and *assA* genes as well as the glycyl radical conserved motif were designed. Nucleotide sequences of *bssA* genes in *Aromatoleum aromaticum* EbN1 (NC\_006513.1), *Geobacter metallireducens* GS-15 (AF441130.1), *Thauera aromatica* K172 (AJ001848.3), *Azoarcus* sp. T (AY032676.1); *masD* gene in *Azoarcus* sp. HxN1 (AM748709); and *assA1* and *2* in *Desulfatibacillum alkenivorans* AK-01 (DQ826035, DQ826036, respectively ) were aligned with ClustalX, and primers were designed manually around the nucleotide regions that harbor the characteristic conserved motif of a glycyl radical enzyme (Table 6.1). The specificity of primers 1995F and 2467R was determined using DNA from *Azoarcus* sp. HXN1 (kindly donated by Max Planck Institute, Germany), *Thauera aromatica* K172 (DSMZ 6984), *Azoarcus* sp. T (DSMZ 9506), *Desulfobacula toluolica* (DSMZ 7467), *Desulfobacterium toluolica* (DSMZ 7467), *Desulfobacterium cetonicum* (DSMZ 7267) and *Desulfatibacillum alkenivorans* AK-01 (DSMZ 16219). Template DNA from *Desulfotomaculum acetoxidans* (DSMZ771) that does not contain fumarate addition gene was used as a negative control. All PCR reactions with pure cultures harbouring *bssA/assA* tested positive, while negative controls did not produce any amplicons.

**Table 6.1** Primer sequence used in this study

Primer	Target	Sequences	Source
1995F <sup>a</sup>	<i>assA/bssA</i>	CCNAARTGGGGHAAYGACGA	This study
2467R	<i>assA/bssA</i>	ANCCNGMNAYVCKNACRATVA	This study
assA-G1F <sup>b</sup>	Group 1	CGATGTCTGCGGGAAGTGGCC	This study
assA-G1R	Group 1	GCCCACAAGGCTGGAATAATG	This study
assA-G2F <sup>c</sup>	Group 2	ACTCGACAAGAAGGGACCAACCG	This study
assA-G2R	Group 2	AAAGCCTCGTCCGACACACAGT	This study
bssA-G3F <sup>e</sup>	Group 3	CATCATCTCCGGAGAGATGCG	This study
bssA-G3R	Group 3	CAGCCTTCTGGGTGGCAGCG	This study
454T_FB <sup>f</sup>	16S rRNA	<u>(F)</u> AAACTYAAAKGAATTGRCGG	See legend
454T_RA	16S rRNA	<u>(F)</u> (N) <sub>10</sub> ACGGGCGGTGTGTRC	See legend

<sup>a</sup>, Primers 1995F and 2467R correspond to the nucleotide position 1936 to 1955 bp and 2411 to 2431bp, respectively, in *D. alkenivorans* AK-01 *assA* (DQ826035); position 1948 to 1967 bp and 2423 to 2443 bp, respectively in *Azoarcus* sp. HxN1 *masD* (AM748709); and position 1995 to 2014 bp and 2467 to 2487 bp, respectively, in *T. aromatica* *bssA* (AJ001848.3)

<sup>b</sup>, Primers assA-G1F/assA-G1R were used in qPCR quantification of group 1 *assA* (see Fig. 6.2 in text)

<sup>c</sup>, Primers assA-G2F/assA-G2R were used in qPCR quantification of group 2 *assA* (see Fig. 6.2 in text)

<sup>e</sup>, Primers bssA-G3F/bssA-G3R were used in qPCR quantification of group 3 *bssA* (see Fig. 6.2 in text)

<sup>f</sup>, Primers 454T\_FB/454T\_RA were obtained from primers 926F and 1392R (Matsuki *et al.*, 2002). The underlined letter F represents the 454 sequencing adaptors, and (N)<sub>10</sub> is a sample-specific 10 nucleotide barcode sequence which allows sample multiplexing during pyrosequencing.

### 6.3.4 PCR amplification, cloning and sequencing of partial *assA* and *bssA*

All PCR reactions were performed in 25 µL mixtures containing 12.5 µL Promega PCR Master Mix (Promega, Madison, WI), 1 µL extracted DNA (approximately 50 ng/µL), 0.04 µM of each primer, and 0.5 µg/µL BSA. The touchdown PCR conditions were as follows: 95°C for 5 min, followed by 10 cycles of 95°C for 45 s, 55°C for 45 s (decreasing by 1°C/cycle) and 72°C for 60 s. The subsequent 30 cycles were 95°C for 45 s, 50°C for 45 s, 72°C for 60 s, and

a final extension of 72°C for 10 min. The amplicons from all PCR assays were resolved by agarose gel electrophoresis, followed by staining in ethidium bromide (Bio-rad), and visualization with a UV detector. All samples amplified with primers 1995F and 2467R produced amplicons of the expected size (bp) and these were cloned into pGEM-T vectors (Promega, Madison, WI) and transformed into Subcloning Efficiency DH5 $\alpha$  competent cells (Invitrogen, Carlsbad, CA) according to manufacturer's instructions. All clones were first checked for inserts with primers T7 and SP6 flanking the insert, using standard PCR conditions. Sequencing reactions were performed using T7 primers and BigDye Terminator v3.1 cycle sequencing kit (Applied Biosystem, Foster City, CA). Sequences were resolved with an ABI3130 sequencer (Applied Biosystem) at the University of Alberta Molecular Biology Services Unit.

### **6.3.5 *assA* and *bssA* phylogeny and community analyses**

All sequences recovered from *assA/bssA* clone libraries were trimmed using SeqMan in Lasergene package (Madison, US), exported to MEGA5.0 (Tamura *et al.*, 2007), aligned with MUSCLE V3.3 program (Edgar 2004) and manually edited. Sequences that did not align properly and contained no primer regions were removed from further analysis. Nucleotide sequences were translated and clustered into operational taxonomic units (OTUs) at 0.02 distance using MOTHUR with the furthest clustering method (Schloss *et al.*, 2009). The 0.02 distance level was chosen based on the following observations: (i) The translated *assA* gene region in *Desulfoglaeba alkanexedens* (ADJ51097.1) and *D. alkenivorans* AK-01 (ABH11460.1) targeted by primers 1995F and 2467R share 92% similarity when compared using BLASTP with BLOSUM62 matrix, and (ii) the same region in most *bssA* genes, for example in *T. aromatica* (CAA05052.1) and *A. aromaticum* EbN1 (CAA05052.1), share 98% similarity. Thus, a distance level of 0.02 appeared to be necessary to group all putative *assA* and *bssA* amino acid sequences into distinct OTUs. Phylogenetic trees were constructed using representative sequences from OTUs obtained from each clone library using the RAxML package in Geneious R7 (Biomatters Ltd., New Zealand) with the

following parameters: maximum-likelihood and WAP model of amino acid substitution, with 1000 bootstrap replicate analysis.

### **6.3.6 qPCR quantification of *assA* and *bssA* phylogenies**

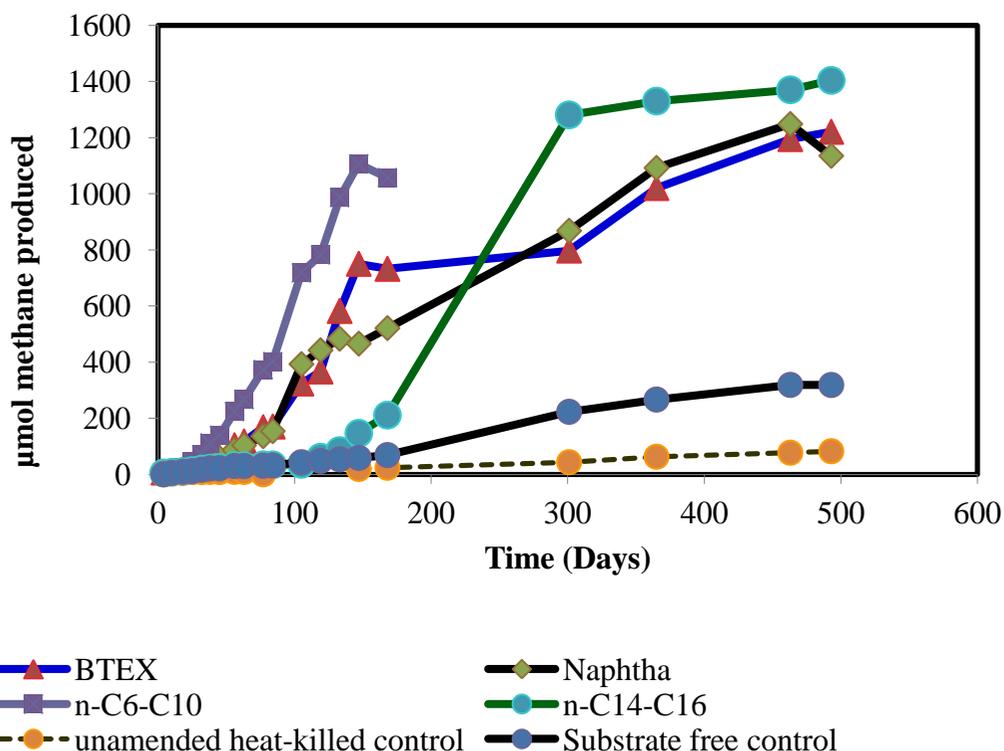
qPCR primers targeting different genotypes of *assA* and *bssA* were designed manually by targeting the variable region in the sequence of interest. The specificity of these primers was confirmed by PCR using different clones recovered in Chapter section 6.3.5. The abundance of fumarate addition genotypes 150 and 450 d during incubation was quantified using qPCR with the primers listed in Table 6.1. All triplicate 10- $\mu$ L qPCR reactions contained 5  $\mu$ L of 5X proprietary reaction buffer (MBSU, University of Alberta), 0.2  $\mu$ M of each primer, 200ng/ $\mu$ L bovine serum albumin (Invitrogen), 2  $\mu$ L template DNA and 1  $\mu$ L MilliQ water. All qPCR reactions were performed using a 7500 real time PCR system (Applied Biosystems) with the following conditions: 95°C for 2 min, followed by 40 cycles of 95°C for 15 s, 60°C for 30 s. Fluorescence signals were collected after each annealing cycle at 60°C, and dissociation curve analysis was included in every run to determine the specificity of the reaction. Selected plasmid clones were recovered using a MiniElute Kit (Qiagen), quantified using a spectrometer (Nanodrop), and serially diluted to gene copy numbers ranging from  $10^2$  to  $10^8$ / $\mu$ L, and included in every qPCR run. The standard curves for all qPCR reactions had R values of at  $\geq 0.99$  (Appendix Figure F1).

## **6.4 Results and discussion**

### **6.4.1 Bioconversion of naphtha components to methane**

In order to investigate the roles of fumarate addition mechanisms by oil sands microorganisms, microcosms were established using oil sands mature fine tailings and incubated under methanogenic conditions with different hydrocarbon substrates representing components of naphtha. These microcosms were amended with whole naphtha, BTEX compounds, short-chain *n*-alkanes (C<sub>6</sub>, C<sub>7</sub>, C<sub>8</sub>, C<sub>10</sub>) and longer-chain *n*-alkanes (C<sub>14</sub>, C<sub>16</sub>, C<sub>18</sub>) as the only added organic carbon source. During incubation for ~490 d at room temperature, headspace methane analysis using GC-FID detected accumulation of >1000  $\mu$ mol methane in microcosms amended with naphtha, BTEX, and C<sub>14</sub>-C<sub>16</sub>. In the heat-killed and

substrate-free controls, less than 300  $\mu\text{mol}$  of methane was detected throughout incubation. Concomitant with methane evolution, we detected the loss of hydrocarbon substrates compared to the heat-killed control (Appendix Table F1). This indicates that methane evolution is associated with methanogenic degradation of amended hydrocarbons. The  $\text{C}_6\text{-C}_{10}$  culture was lost after incubation for  $\sim 150$  days, but parallel methanogenic  $\text{C}_6\text{-C}_{10}$ -degrading cultures established from oil sands mature fine tailings could similarly degrade  $\text{C}_6\text{-C}_{10}$  alkanes coupled to methanogenesis (not shown). The results presented here are consistent with previous findings that organisms enriched from oil sands tailings ponds are capable of methanogenic degradation of toluene, ethylbenzene, *o*- and *m*-xylenes,  $\text{C}_6\text{-C}_{10}$  and  $\text{C}_{14}\text{-C}_{16}$  (Siddique *et al.*, 2006, Siddique *et al.*, 2007, Siddique *et al.*, 2011).



**Figure 6.1** Headspace methane detection in heat-killed and substrate-free controls compared to microcosms amended with different hydrocarbon substrates. The culture amended with  $\text{C}_6$  to  $\text{C}_{10}$  was only monitor up to  $\sim 160$  days due to culture loss. Microcosms set up from non-parallel cultures were used in the subsequent qPCR analysis (see below).

#### **6.4.2 Diversity of fumarate addition genes in methanogenic hydrocarbon-degrading microcosms recovered through cloning and sequencing**

Although several hydrocarbon-activating mechanisms such as carboxylation of benzene (Abu Laban *et al.*, 2010) and hydroxylation of alkanes (Callaghan 2013) have been previously elucidated or proposed, fumarate addition continues to be the most intensively studied mechanism based on functional gene and metabolite detection in cultures and environments (reviewed by Agrawal and Gieg 2013, Callaghan 2013). In order to investigate the diversity of *assA/nmsA/bssA* genes as an indication of hydrocarbon activation by fumarate addition in methanogenic hydrocarbon-degrading cultures established in the current study, PCR amplification was initially performed using primers reported by Winderl *et al.* (2007) and primer set 9 reported by Callaghan *et al.* (2010) to screen a BTEX-degrading culture. Sixty nucleotide sequences of ~794 bp were obtained from clone libraries constructed using primer set 7772f/8546r (Winderl *et al.*, 2007). Primer set 9 (Callaghan *et al.*, 2010) recovered non-specific PCR products (59/79 are not related to pyruvate formate lyase, PFL) and were not used in subsequent studies. Phylogenetic analysis of the post quality-controlled sequences indicates that several *bssA* sequences (recovered using 7772f/8546r) in the BTEX-amended microcosm were related to uncultivated sulfate-reducing bacteria with unknown taxonomy (Appendix Figure F2). The *assA* sequences recovered using Primer set 9 were phylogenetically related to the *assA* genes recovered by Callaghan *et al.*, (2010) from a hydrocarbon-impacted environment, i.e., Fort Lupton, although the taxonomy could not be ascertained.

The specificity of primers used in amplification of *assA/bssA/nmsA* in cultures and environmental samples can result in under-detection of fumarate addition genes (Callaghan *et al.*, 2010, Kimes *et al.*, 2013, von Netzer *et al.*, 2013). In the present study, a highly degenerate primer pair targeting a different conserved region than that previously used (Aitken *et al.*, 2013, Callaghan *et al.*, 2010, von Netzer *et al.*, 2013) was designed in order to detect *assA/bssA/nmsA* that may be highly divergent from the canonical *assA* gene reported by Callaghan *et al.*, (2010). More recently, von Netzer *et al.*, (2013) discussed in detail the

efficiency of different primers sets available in the literature for amplifying different groups of *assA/bssA/nmsA*. The authors subsequently introduced several newly designed primer sequences claiming to recover broad taxonomic groups that were not amplifiable with the existing primer sequences available in the literature. The primer sequences designed by Aitken *et al.*, (2013) and von Netzer *et al.*, (2013) were not available at the time of this analysis, and therefore were not tested in our study.

The primers designed for this study co-amplified the *assA* and *bssA* genes in model toluene- and alkane-degrading cultures (See Methods). Using primers 1995F/2467R (Table 6.1), a total of 268 nucleotide sequences of approximately 490 bp were obtained from clone libraries constructed from each of the four hydrocarbon-amended microcosms and a substrate-free microcosm. Translated sequences were clustered into Operational Taxonomic Units (OTUs) at 98% similarity, followed by maximum likelihood phylogenetic analysis. All sequences were subsequently inserted into a reference *assA/nmsA/bssA* tree using MLTreeMap (Stark *et al.*, 2010) (Figure 6.2). Following this, all recovered *assA/bssA/nmsA* sequences were assigned to different clusters according to Acosta-González *et al.*, (2013) in order to infer their substrate spectrum.

Sequences in Cluster IV (a) are described as being associated with methanogenic environments (Acosta-González *et al.*, 2013). The *assA* sequences that fall within this cluster have previously been associated with unknown taxa (Acosta-González *et al.*, 2013), but here, we assigned Cluster IV (a) to *Smithella* spp. because two *assA* gene sequences within this cluster were recently detected in single cell genome sequencing of *Smithella* ME-1 (Appendix A) and the genomic bin of a *Smithella* spp. obtained from the metagenome of a short-chain alkane degrading culture (Chapter 4). Supporting this taxonomic assignment, most other sequences within this cluster were recovered through PCR screening of methanogenic longer-chain (C>14) *n*-alkane degrading cultures where *Smithella* spp. or its closely related genus *Syntrophus* were detected as 16S rRNA genes (Aitken *et al.*, 2013, Callaghan *et al.*, 2010, Zhou *et al.*, 2012).

Cluster IV (b) has been described as being associated with non-methanogenic divergent *assA* (Acosta-González *et al.*, 2013) and contains canonical *assA* genes reported for Deltaproteobacterium *D. alkenivorans* Ak-01 and Betaproteobacterium *Azoarcus* sp. HxN1. Most sequences recovered from hydrocarbon-impacted environments including terrestrial and marine habitats belong to this cluster (Aitken *et al.*, 2013, Callaghan *et al.*, 2010, Kimes *et al.*, 2013, von Netzer *et al.*, 2013). A large number of clones detected in this study were phylogenetically related to this cluster but did not appear to be correlated with particular hydrocarbon amendments (Figure 6.2 and 6.3). Cluster IV (c) is newly described in this Chapter and a single *assA* within this cluster has been putatively assigned to the taxon of *Desulfotomaculum* (Chapter 4).

Cluster III is associated with *nmsA*, and appears to contain three subgroups based on sequences detected in oil sands tailings ponds (Figure 6.5; see later section) in contrast to two subgroups described by Acosta-González *et al.*, (2013). A single *nmsA* gene recovered from SCADC (SCADC\_*nmsA*1 in Figure 6.2) has been putatively assigned to Clostridiales (Chapter 4).

**Figure 6.2** (Next page) Diversity of translated *assA/bssA/nmsA* genes detected in hydrocarbon-degrading microcosms and a substrate-free control established from oil sands tailings mature fine tailings. Using MLTreeMap (Stark *et al.*, 2010), post quality-controlled reads obtained from clone libraries of different microcosms were inserted into a reference tree containing reference sequences having known taxonomy and metagenomic reads (See Figure 6.6). The reference tree was constructed using maximum likelihood analysis with RAxML and 1000 bootstrap replicates (not shown) and rooted using pyruvate formate lyases (omitted from this Figure, see subsequent section). Pie charts embedded within tree branches represent placement of clone library sequences that are phylogenetically related to the reference sequence at the tip of each branch. Chart size corresponds to the frequency of sequence recovery within a sample/clone library. Sequences recovered from different clone libraries are represented with unique colours. "G1", "G2" and G3" and taxa shown in red represent targets for qPCR quantification (Table 6.1 and Figure 6.3). Sequences (i.e., Mallorca and Figueiras recovered from hydrocarbon-impacted environments) reported by Acosta-González *et al.* (2013) were used as references.



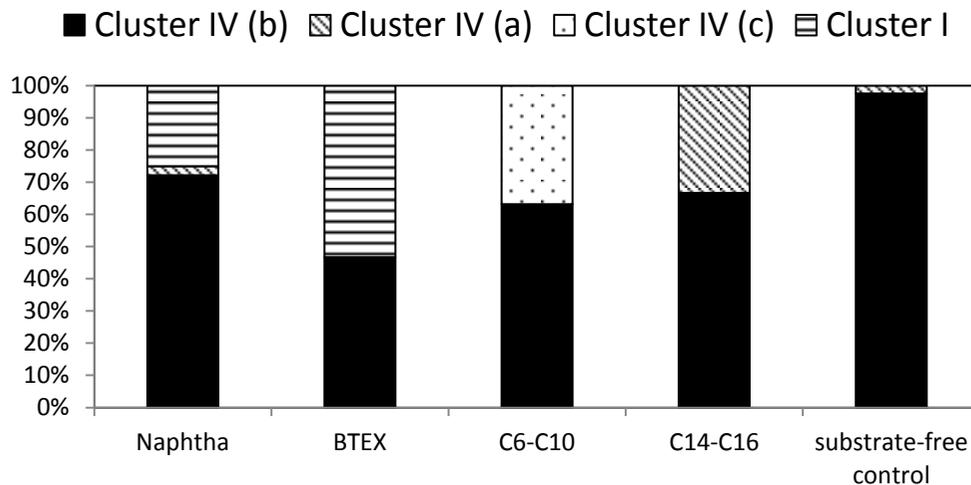
However, no sequences affiliated with *nmsA* were detected in any hydrocarbon-degrading cultures using primer 1995F/2467R, possibly due to specificity of the primers or that organisms carrying the *nmsA* gene are not enriched in the microcosms since polyaromatic substrates, e.g., 2-methylnaphthalene, were not present in amendments. Despite that, *Desulfotomaculum* sp. Ox39 carrying a *nmsA*-like gene is incapable of growth using 2-methylnaphthalene (Acosta-González *et al.*, 2013), and the gene has since been suggested to be involved in the activation of monoaromatic compounds such as *m*-xylene and toluene (Acosta-González *et al.*, 2013, von Netzer *et al.*, 2013).

Clusters I and II are both associated with *bssA* genes (Acosta-González *et al.*, 2013), and sequences related to Cluster I *bssA* were detected in only BTEX and naphtha-degrading microcosms in the present study (Figure 6.2). Cluster II *bssA* genes are likely not present in the methanogenic degrading cultures used in this study, as metagenomic sequencing of MLSB and a short-chain alkane-degrading culture (Chapter 2) did not recover genes belonging to Cluster II. Cluster V likely represents a novel undescribed fumarate addition gene cluster for which the functional role has not been investigated (discussed further below). The selectivity of the primers used in this study has not been investigated, but overall, PCR amplification recovered genes within Cluster IV (a, b and c), and Cluster I but not Clusters III and V. The degenerate primer designed and used in the current study does not produce amplicons that overlap with sequences amplified using primers reported previously (Aitken *et al.*, 2013, Callaghan *et al.*, 2010, von Netzer *et al.*, 2013). Although this has made direct comparison difficult, phylogenetic comparisons with sequences generated through high throughput sequencing is possible (see below).

#### **6.4.3 Increased abundance of *assA* and *bssA* during incubation suggests functional roles in hydrocarbon activation by fumarate addition**

All hydrocarbon-amended microcosms and substrate-free controls shared an abundance of Cluster IV (b) *assA* genes that remained high in abundance even in the substrate-free control, but decreased in abundance in microcosms that received naphtha, BTEX, C<sub>6</sub>-C<sub>10</sub> and C<sub>14</sub>-C<sub>16</sub> after incubation for ~150 d (Figure

6.3). Within this cluster, a single dominant *assA* and several minor *assA* monophyletic clades were detected (Figure 6.2). Cluster IV (b) *assA* is likely involved in the degradation of endogenous hydrocarbons (also present in the substrate-free control) which include mostly low molecular weight alkanes (not shown) by coupling this process to the reduction of background sulfate. During early onset of methane production in the C<sub>6</sub>-C<sub>10</sub> culture, pyrotag sequencing of the 16S rRNA gene detected higher abundance of sulfate-reducing bacteria and low abundance of sequences affiliated with Peptococcaceae (i.e., *Desulfotomaculum*). However, the abundance of Peptococcaceae increased over time during a year-long incubation and transfer with C<sub>6</sub>-C<sub>10</sub> alkanes under strictly methanogenic conditions (not shown). The delayed onset of methanogenesis perhaps is due to the presence of sulfate in tailings, which was not measured in this study but has been reported (Salloum *et al.*, 2002).



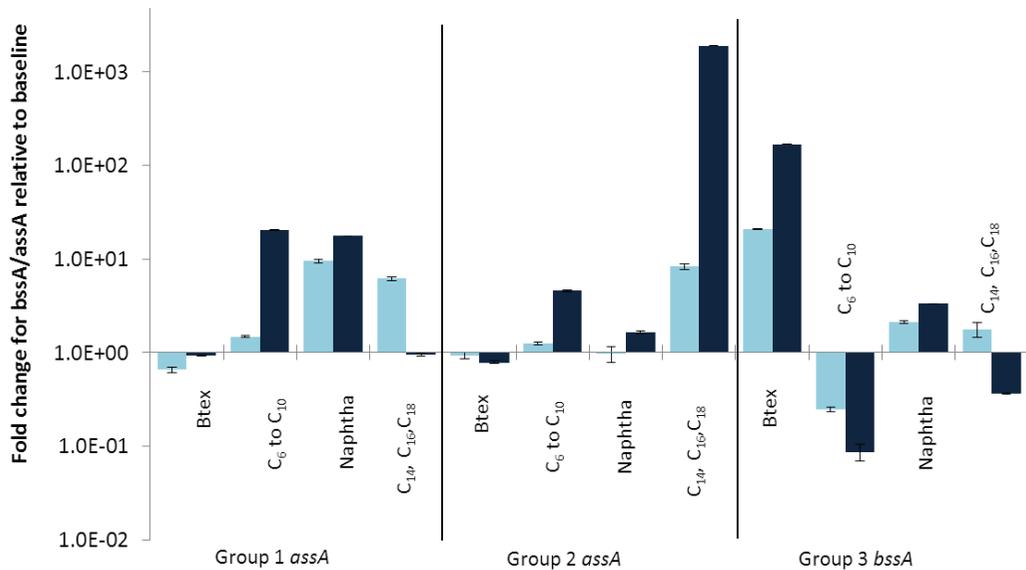
**Figure 6.3** Relative abundance of *assA* and *bssA* subtypes (see Figure 6.2) in microcosms amended with different hydrocarbon substrates and incubated for 150 d

In microcosms amended with additional hydrocarbons, there was an apparent enrichment of a few different *assA* and *bssA* genotypes relative to background Cluster IV (b) *assA* after incubation of ~150 d. For example, amendment with BTEX and naphtha resulted in the enrichment of a large

proportion of clones affiliated with Cluster I *bssA*, which was not detected in alkane-amended and substrate-free microcosms, indicating that the *bssA* is involved in toluene addition to fumarate. Here, the result is consistent with detection of benzylsuccinate metabolites in a parallel methanogenic toluene-degrading culture enriched from MLSB (Semple and Young, unpublished). Together, the results presented here suggest that Cluster I *bssA* in tailings organisms is functional in toluene degradation by fumarate addition. In the microcosm amended with C<sub>14</sub>-C<sub>16</sub>, the enrichment of Cluster IV (a) *assA* putatively affiliated with *Smithella* spp. is consistent with reports in Chapter 4 and Appendix A that implicate *Smithella* in activating longer-chain alkanes instead of short-chain *n*-alkanes. The expression of *assA* genes by *Smithella* sp. TM-1 during methanogenic degradation of hexadecane further confirms that Cluster IV (a) *Smithella*-affiliated *assA* is capable of alkane activation by fumarate addition (Embree *et al.*, 2013 and Appendix A). In the C<sub>6</sub>-C<sub>10</sub> amended microcosm, there was a greater abundance of Cluster IV (c) *assA* affiliated with *Desulfotomaculum* spp., which has been shown to highly transcribe its *assA* gene during active degradation of low molecular weight alkanes (Chapter 4).

To confirm our observations here, the relative abundance of selected genotypes shown in Figure 6.2 were quantified by qPCR using DNA isolated at two time points during active methanogenesis. qPCR primers were designed to specifically target three groups of interest: *Smithella*-affiliated *assA* in Cluster IV (a) (G2 in Figure 6.2), the dominant Cluster IV (b) *assA* present in abundance in all microcosms (G1 in Figure 6.2), and a *bssA* phylogenetically related to the sulfate-reducing bacterium TRM1 (ABM92939) (G3 in Figure 6.2). Compared to the substrate-free control, there was a large increase in *Smithella*-affiliated *assA* only in the microcosm amended with long-chain *n*-alkanes (Figure 6.4). In contrast, there was a large increase in abundance of *assA* genes in short-chain alkane and naphtha-amended microcosms, but not in the microcosm amended with BTEX. In naphtha and BTEX-amended microcosms, only *bssA* (Group 3) increased in abundance (Figure 6.4). Naphtha consists of both alkanes and BTEX compounds; the increase of both *assA* and *bssA* suggests that these compounds

are likely being degraded by fumarate addition simultaneously during incubation. Overall, the results obtained from qPCR analysis corroborate those obtained from clone libraries; the increase in abundance of selected genotypes together with the detection of expressed genes in response to growth on their expected substrates confirm that microorganisms in oil sands tailings ponds are functionally capable of hydrocarbon degradation by fumarate addition.



**Figure 6.4** qPCR quantification of selected *assA* and *bssA* genotypes as shown in Figure 6.2 (G1, G2 and G3). DNA was isolated from microcosms that received amendments of different hydrocarbon substrates followed by incubation (light blue= 150 d, dark blue=450 d). The absolute copy numbers of genes was quantified using a standard curve shown in Appendix Figure F1 and expressed as relative abundance compared to that detected in the substrate-free control. Quantification of genes was conducted for a different C<sub>6</sub>-C<sub>10</sub> degrading microcosm due to loss of the parallel short chain culture. Values represent the mean of technical qPCR replicates (n=3) from a single DNA extract; error bars represent standard deviation.

#### 6.4.4 Diversity of fumarate addition genes in oil sands tailings ponds

The Illumina Hi-seq sequences generated for the MLSB tailings pond metagenome were first assembled into contigs using CLC Genomics Workbench (CLC-Bio, USA) with the default settings. Fumarate addition genes are frequently miss-annotated using automated annotation pipelines (Aziz *et al.*, 2008, Markowitz *et al.*, 2012, Meyer *et al.*, 2008) possibly due to the lack of reference sequences and presence of a huge number of phylogenetically related genes (i.e., glycerol dehydratase and pyruvate formate lyases; PFL) in the database. To overcome this problem, the *assA/bssA/nmsA* genes in the assembled contigs were recovered using tBLASTn with maximum E-values of  $10^{-1}$ . The fumarate addition genes were then assigned to their respective phylogenetic clades using MLTreeMap (Stark *et al.*, 2010). To do this, all sequences recovered from MLSB metagenomic contigs in addition to sequences reported by Acosta-González *et al.* (2013) and Kimes *et al.* (2013) and selected reference sequences in NCBI were aligned using Muscle 3.3 (Edgar 2004). The alignment was manually edited and sequences that were poorly aligned (usually PFL) were removed. This was followed by maximum likelihood phylogenetic analysis using RAxML (Stamatakis *et al.*, 2005).

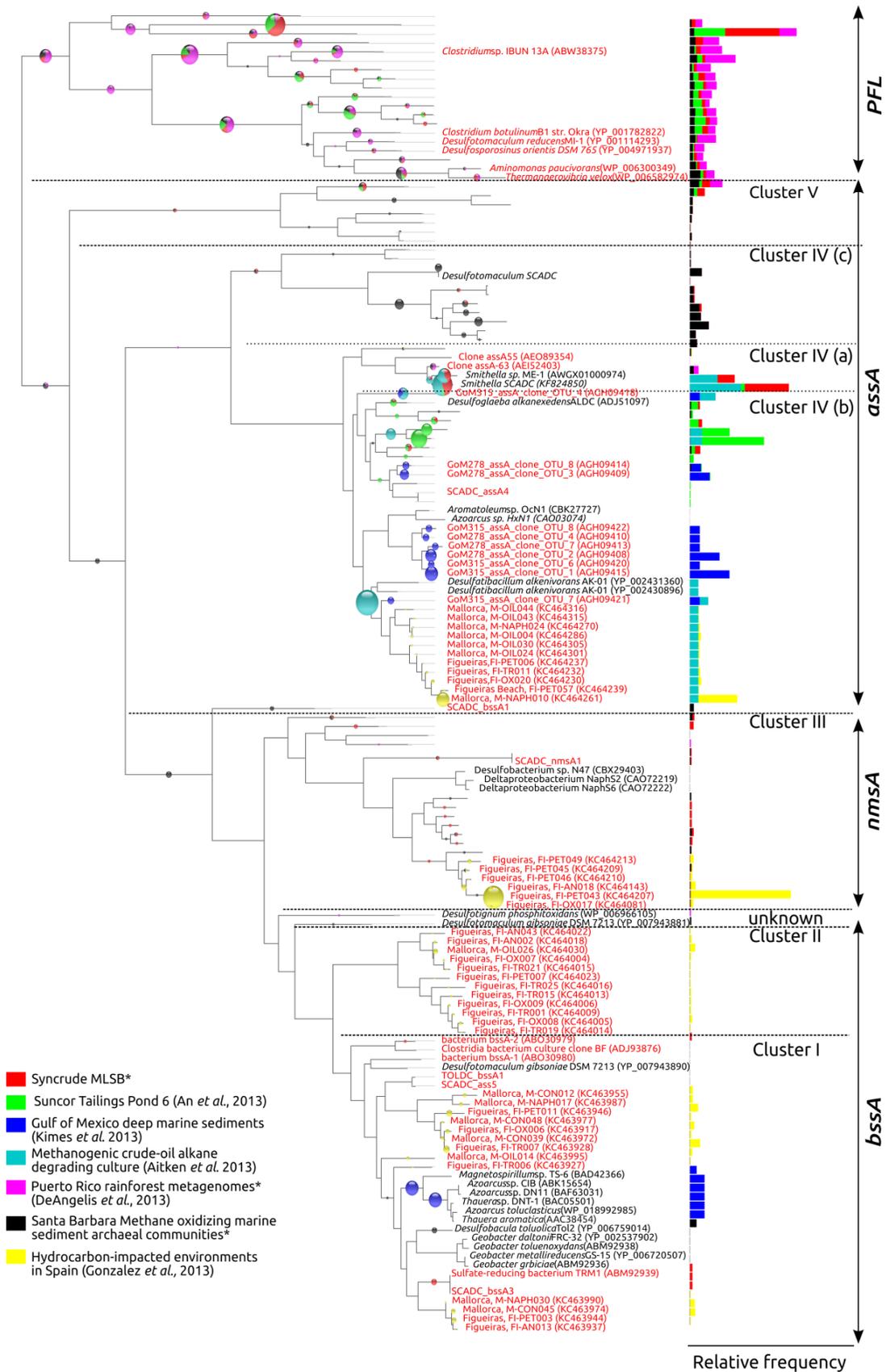
The alignment and resulting tree were subsequently used as a reference template in MLTreeMap analysis (Stark *et al.*, 2010) to assign unassembled 454 pyrotag short reads from MLSB, Suncor Tailing Pond 6 (TP6; (An *et al.*, 2013) and Gulf of Mexico deep marine sediments (GoM; Kimes *et al.*, 2013) into respective phylogenetic clusters based on the naming scheme described earlier (this Chapter) and in Acosta-González *et al.* (2013). In so doing, the contribution of each cluster to the overall detection of *assA/bssA/nmsA* could be semi-quantified. In addition, contigs from five metagenomes of soil communities enriched from a Puerto Rico rainforest capable of decomposing switch grass and/or wood (Project ID: Gm00353) and six metagenomes of Santa Barbara methane-oxidizing marine sediment archaeal communities (Project ID Gm0026) were downloaded from the Joint Genome Institute IMG website (Markowitz *et al.*, 2012) and included in this analysis. For comparison purposes, all sequences

recovered from clone libraries (Acosta-González *et al.*, 2013, Aitken *et al.*, 2013, Kimes *et al.*, 2013) were included in the MLTreeMap analysis.

#### 6.4.5 MLSB tailings harbours novel clusters of fumarate addition genes

Phylogenetic analysis of the fumarate addition genes recovered from the MLSB metagenome detected a high diversity of *assA* and *nmsA* but few *bssA* sequences (Figure 6.5). This is in contrast to studies done in other environments where *bssA* is more prevalent (Winderl *et al.*, 2007, Winderl *et al.*, 2010). Notably, several full length *assA/bssA/nmsA* sequences (~2400 bp) recovered in MLSB were also detected in the SCADC metagenome (Chapter 4) with >98% similarity, indicating that sequence assembly, at least for these genes, is less likely to be chimeric since these two metagenomes were assembled independently of each other. The differences in overall diversity and phylogenetic relationship of fumarate addition genes between MLSB and other environments (i.e., GoM and those reported by Acosta-González *et al.*, 2013) is apparent and likely due to substrate specificity of the detected genes.

**Figure 6.5** (Next page) Diversity of translated *assA/bssA/nmsA* genes detected in the metagenome of Syncrude MLSB. All sequences are shown as unlabelled branches and had >450 amino acids (with >80% of these recovered sequences almost full length, i.e., ~800 amino acids). The *assA/bssA/nmsA* amino acid sequences were assigned into clusters based on (Acosta-González *et al.*, 2013). The reference tree was constructed using maximum likelihood analysis using RAxML with reference sequences having known taxonomy (in black) or those from environmental sources (in red) and rooted using pyruvate formate lyases (PFL). The pyrotag short reads and assembled contigs from several metagenomes (see main text) were assigned into their respective clusters using MLTreeMap (Stark *et al.*, 2010). Pie charts embedded within tree branches represent placement of reads that are phylogenetically related to the reference sequence at the tip of each branch. Chart size corresponds to the frequency of recovery within a metagenome. The bar chart on the right shows the relative abundance of different phylogenies within a single metagenome (relevant only within samples, not between samples). The relative proportions of reads at a lower taxonomic placement (not at the tips) were equally distributed among all extended tips and are shown in the bar chart on the right. \* indicates assembled contigs were used.



The *assA* gene sequences recovered from MLSB are associated with three clusters discussed earlier: Cluster IV (a) and (b) according to Acosta-González *et al.* (2013) and Cluster IV (c) described in this Chapter. BLASTX/BLASTP searches were conducted using all sequences of Cluster IV (c) against the NCBI database, but all closest relatives recovered (best BLAST hits) were represented by sequences belonging to Cluster IV (b), indicating that Cluster IV (c) comprise novel *assA* that (i) may have escaped detection in other environments due to selectivity of primers (but was recovered using primers in Table 6.1), or (ii) that organisms carrying these gene sequences are only found in oil sands tailings ponds. The earlier assumption is more plausible since selectivity of primers has been reported (von Netzer *et al.*, 2013). Interestingly, Cluster IV (c) was also not detected in the metagenome of Suncor Tailings pond 6 (more discussion in the following section), possibly due to the substrates that are present *in situ* or differing pond conditions. Multiple genes that are phylogenetically related to novel Cluster IV (c) and a minor number of *nmsA*-like genes were detected in the metagenomes of Santa Barbara methane-oxidizing marine sediment archaeal communities (Project ID Gm0026) but not in other metagenomes screened in the current study. *nmsA*-like genes have also been detected in Guaymas Basin and Gulf of Mexico hydrocarbon seeps, indicating the potentially ubiquitous nature of *nmsA* in marine hydrocarbon seeps. Coal Oil Point Seeps in Santa Barbara release petroleum hydrocarbons including methane and *n*-alkanes into the surrounding sea water (Farwell *et al.*, 2009). The presence of fumarate addition genes, in particular Cluster IV (c) *assA*, which was shown to be functional based on metatranscriptomic analysis (Chapter 4), indicates that related genes potentially may be involved in *in situ* anaerobic degradation of some of the released hydrocarbons. The presence of Cluster IV (c) in other related environments, i.e., Guaymas Basin hydrocarbon seeps in the Gulf of California and the Gulf of Mexico hydrocarbon seeps is not known, since related genes were not detected using PCR amplification (von Netzer *et al.*, 2013).

Further, a single Subcluster (Cluster V) of PFL-like genes with unknown taxonomic identity was recovered from MLSB (Figure 6.5). Phylogenetically

related sequences were also recovered from TP6, Puerto Rico rainforest soil communities (Project ID Gm00353), and Santa Barbara (Project ID Gm0026). BLASTX searches of MLSB Cluster V sequences against the NCBI database indicate that they had best BLASTP matches to *assA* (35 to 38% similarity in amino acids) in *D. alkanexedens* ALDC and *D. alkenivorans* AK-01, and had no hits to PFL. All glycyl-radical forming enzymes, including canonical ASS, BSS and NMS contain the catalytic conserved amino acids Cys492 and Gly828 [position corresponds to the BSSA residues in *Thauera* sp. DNT1 (BAC05501)] (Leuthner *et al.*, 1998, Li *et al.*, 2009), which were similarly detected in Cluster IV (c) and V at the position expected for glycyl-forming enzymes (not shown); this confirms that these sequences are not likely to be chimeric and are indeed related to PFL. The functional role of Cluster V is currently unknown but the differences in substrate spectrum for ASS, BSS and NSM has been in part attributed to changes in the amino acid residues within substrate binding pockets (Acosta-González *et al.*, 2013, Bharadwaj *et al.*, 2013).

The newly described cluster of fumarate addition-like genes may in fact be involved in the activation of novel substrates. Supporting this inference, fumarate activation of hydrocarbons structurally different from that catalyzed by ASS, BSS and NMS has been previously reported. For example, a Firmicutes related to *Desulfotomaculum* spp. has been implicated in the complete degradation of ethylcyclopentane under sulfate-reducing conditions (Rios-Hernandez *et al.*, 2003). However, the genes involved (i.e., *assA*, *bssA*, *nmsA*, etc.) have not been reported, precluding further phylogenetic and docking analysis (Bharadwaj *et al.*, 2013).

#### **6.4.6 Higher abundance of *assA* and *nmsA* genes in MLSB and TP6 reflects the higher concentration of residual alkanes in tailings ponds**

In general, the *assA*- and *nmsA*-like genes detected in MLSB and TP6 are more diverse and higher in abundance than the *bssA* genes (Figure 6.5). This is likely due to the larger proportion of alkanes present in residual naphtha in the tailings ponds compared to monoaromatic compounds, i.e., toluene and xylenes (Siddique *et al.*, 2006, Siddique *et al.*, 2007); this may have selected for

organisms that are capable of anaerobic alkane degradation. Phylogenetic assignment of short reads using MLTreeMap indicates that there is a higher abundance of Cluster IV (c) *Smithella*-affiliated *assA* in MLSB in contrast to the higher abundance of Cluster IV (b) *assA* in TP6 (Figure 6.5), which is suggested to be associated with communities that are frequently detected in non-methanogenic environments (Acosta-González *et al.*, 2013). This likely reflects differences in management practices of two tailings ponds (i.e., greater use of gypsum in tailings consolidation in the Suncor TP6) that cannot be verified in this study.

The lack of detection *bssA* sequences in unassembled 454 data obtained from TP6 was further verified by first assembling the 454 data into contigs, followed by tBLASTn searches using reference *bssA* sequences. Recovered sequences were aligned and subjected to phylogenetic analysis described in Methods (data not shown). This approach recovered essentially only *assA* and not *bssA* sequences, confirming that TP6 has a high abundance of *assA* genes but may have a smaller representation of *bssA* sequences. Most sequences detected in MLSB and TP6, with the exception of a few sequences, have unknown taxonomy. The taxonomic analysis conducted by An *et al.*, (2013) assigned the fumarate addition genes detected in TP6 to Actinobacteria and Euryarchaeota, amongst others: those recovered genes are likely pyruvate formate lyases or other non-fumarate addition genes since members of these phyla are not known to be capable of fumarate addition let alone to carry genes capable of hydrocarbon addition to fumarate.

#### **6.4.7 Abundance of fumarate addition genes in other environments**

Screening the three metagenomes from GoM (Kimes *et al.*, 2013) using the Hidden Markov Model constructed in the present study for MLTreeMap analysis did not recover any fumarate addition genes. Therefore, the three GoM metagenomes were separately assembled into contigs and used in tBLASTn screening for fumarate addition genes: this recovered only three sequences related to fumarate addition genes, in contrast to the higher diversity of *assA* and *bssA* reported by (Kimes *et al.*, 2013). This result is possibly due to the low abundance

of organisms capable of fumarate addition at the GoM sites sampled for metagenomic analysis by Kimes *et al.*, (2013). Because of this, the sequences obtained from clone libraries reported by Kimes *et al.*, (2013) were used in the phylogenetic comparison in Figure 6.5. The lack of *assA/bssA/nmsA* genes in TP6 and GoM does not necessarily preclude the presence of alternate fumarate addition genes, as PCR amplification or Illumina sequencing (i.e., MLSB) may be able to recover sequences that are lower in abundance.

## 6.5 Significance

By mining metagenomes available in the public domain, we have shown that MLSB and other environments harbour phylogenetically distinct putative fumarate addition genes which may not be readily recoverable by PCR using primers that were designed based on known sequences. Furthermore, qPCR quantification of recovered fumarate addition genes coupled to transcriptomics analysis using enrichment cultures (Chapter 4 and Appendix A) confirmed that these genes are functionally involved in hydrocarbon degradation by fumarate addition. The divergence of Cluster V PFL-like genes from canonical *assA* and *bssA* and Cluster IV (c) from IV (a) and (b) implicate potential differences in their functional roles (Acosta-González *et al.*, 2013). The elucidation of the substrate spectrum based on sequence similarity and docking analysis is of particular interest because this could potentially be useful in the development of molecular models to infer substrate degradability *in situ*. Further data mining of other metagenomes can be expected to provide additional insights into the ubiquitous nature of fumarate addition genes in anoxic environments and to further expand the known diversity of the fumarate addition genes.

## 6.6 References

Abu Laban N, Selesi D, Rattei T, Tischler P, Meckenstock RU. (2010). Identification of enzymes involved in anaerobic benzene degradation by a strictly anaerobic iron-reducing enrichment culture. *Environ Microbiol* **12**: 2783-2796.

Acosta-González A, Rossello-Mora R, Marques S. (2013). Diversity of benzylsuccinate synthase-like (*bssA*) genes in hydrocarbon-polluted marine

sediments suggests substrate-dependent clustering. *Appl Environ Microbiol* **79**: 3667-3676.

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

Aitken CM, Jones DM, Maguire MJ, Gray ND, Sherry A, Bowler BFJ *et al.*, (2013). Evidence that crude oil alkane activation proceeds by different mechanisms under sulfate-reducing and methanogenic conditions. *Geochim Cosmochim Acta* **109**: 162-174.

An D, Brown D, Chatterjee I, Dong X, Ramos-Padron E, Wilson S, Bordenave S, Caffrey S, Gieg L, Sensen C, and Gerrit Voordouw, G. (2013) Microbial community and potential functional gene diversity involved in anaerobic hydrocarbon degradation and methanogenesis in an oil sands tailings pond. *Genome* 10.1139/gen-2013-0083.

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

Bharadwaj VS, Dean AM, Maupin CM. (2013). Insights into the glycy radical enzyme active site of benzylsuccinate synthase: a computational study. *J Am Chem Soc* **135**: 12279-12288.

Callaghan AV, Davidova IA, Savage-Ashlock K, Parisi VA, Gieg LM, Suflita JM *et al.*, (2010). Diversity of benzyl- and alkylsuccinate synthase genes in hydrocarbon-impacted environments and enrichment cultures. *Environ Sci Technol* **44**: 7287-7294.

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

Chalaturnyk RJ, Scott JD, Ozum B. (2002). Management of oil sands tailings. *Petrol Sci Technol* **20**: 1025-1046.

Edgar RC. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32**: 1792-1797.

Farwell C, Reddy CM, Peacock E, Nelson RK, Washburn L, Valentine DL. (2009). Weathering and the fallout plume of heavy oil from strong petroleum seeps near Coal Oil Point, CA. *Environ Sci Technol* **43**: 3542-3548.

Fedorak PM, Hrudey SE. (1984). The effects of phenol and some alkyl phenolics on batch anaerobic methanogenesis. *Water Res* **18**: 361-367.

- Fedorak PM, Coy DL, Dudas MJ, Simpson MJ, Renneberg AJ, MacKinnon MD. (2003). Microbially-mediated fugitive gas production from oil sands tailings and increased tailings densification rates. *J Environ Eng Sci* **2**: 199-211.
- Foght J, Aislabie J, Turner S, Brown CE, Ryburn J, Saul DJ *et al.*,. (2004). Culturable bacteria in subglacial sediments and ice from two Southern Hemisphere glaciers. *Microb Ecol* **47**: 329-340.
- Holowenko FM, MacKinnon MD, Fedorak PM. (2000). Methanogens and sulfate-reducing bacteria in oil sands fine tailings waste. *Canadian Journal of Microbiology* **46**: 927-937.
- Kimes NE, Callaghan AV, Aktas DF, Smith WL, Sunner J, Golding B *et al.*,. (2013). Metagenomic analysis and metabolite profiling of deep-sea sediments from the Gulf of Mexico following the Deepwater Horizon oil spill. *Front Microbiol* **4**: 50.
- Leuthner B, Leutwein C, Schulz H, Horth P, Haehnel W, Schiltz E *et al.*,. (1998). Biochemical and genetic characterization of benzylsuccinate synthase from *Thauera aromatica*: a new glycyl radical enzyme catalysing the first step in anaerobic toluene metabolism. *Mol Microbiol* **28**: 615-628.
- Li L, Patterson DP, Fox CC, Lin B, Coschigano PW, Marsh ENG. (2009). Subunit structure of benzylsuccinate synthase. *Biochemistry* **48**: 1284-1292.
- Markowitz VM, Chen IMA, Chu K, Szeto E, Palaniappan K, Grechkin Y *et al.*,. (2012). IMG/M: the integrated metagenome data management and comparative analysis system. *Nucleic Acids Res* **40**: D123-D129.
- Matsuki T, Watanabe K, Fujimoto J, Miyamoto Y, Takada T, Matsumoto K *et al.*,. (2002). Development of 16S rRNA-gene-targeted group-specific primers for the detection and identification of predominant bacteria in human feces. *Appl Environ Microbiol* **68**: 5445-5451.
- Meyer F, Paarmann D, D'Souza M, Olson R, Glass EM, Kubal M *et al.*,. (2008). The metagenomics RAST server - a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC Bioinformatics* **9**.
- Penner TJ, Foght JM. (2010). Mature fine tailings from oil sands processing harbour diverse methanogenic communities. *Can J Microbiol* **56**: 459-470.
- Rios-Hernandez LA, Gieg LM, Suflita JM. (2003). Biodegradation of an alicyclic hydrocarbon by a sulfate-reducing enrichment from a gas condensate-contaminated aquifer. *Appl Environ Microbiol* **69**: 434-443.

- Salloum MJ, Dudas MJ, Fedorak PM. (2002). Microbial reduction of amended sulfate in anaerobic mature fine tailings from oil sand. *Waste Manage Res* **20**: 162-171.
- Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB *et al.*, (2009). Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl Environ Microbiol* **75**: 7537-7541.
- Siddique T, Fedorak PM, Foght JM. (2006). Biodegradation of short-chain *n*-alkanes in oil sands tailings under methanogenic conditions. *Environ Sci Technol* **40**: 5459-5464.
- Siddique T, Fedorak PM, McKinnon MD, Foght JM. (2007). Metabolism of BTEX and naphtha compounds to methane in oil sands tailings. *Environ Sci Technol* **41**: 2350-2356.
- Siddique T, Penner T, Semple K, Foght JM. (2011). Anaerobic biodegradation of longer-chain *n*-alkanes coupled to methane production in oil sands tailings. *Environ Sci Technol* **45**: 5892-5899.
- Stamatakis A, Ludwig T, Meier H. (2005). RAxML-II: a program for sequential, parallel and distributed inference of large phylogenetic. *Concurrency and Concurrency Prac Ex* **17**: 1705-1723.
- Stark M, Berger SA, Stamatakis A, von Mering C. (2010). MLTreeMap - accurate Maximum Likelihood placement of environmental DNA sequences into taxonomic and functional reference phylogenies. *BMC Genomics* **11**.
- Tamura K, Dudley J, Nei M, Kumar S. (2007). MEGA4: Molecular evolutionary genetics analysis (MEGA) software version 4.0. *Mol Biol Evol* **24**: 1596-1599.
- von Netzer F, Pilloni G, Kleindienst S, Kruger M, Knittel K, Grundger F *et al.*, (2013). Enhanced gene detection assays for fumarate-adding enzymes allow uncovering of anaerobic hydrocarbon degraders in terrestrial and marine systems. *Appl Environ Microbiol* **79**: 543-552.
- Widdel F, Bak F (1992). Gram-negative mesophilic sulfate-reducing bacteria. In: Balows A, Trüper, H. G., Dworkin, M., Harder, W. and Schleifer, K.H. (ed). *The Prokaryotes*. Springer-Verlag: New York, USA. pp 3352-3378.
- Winderl C, Schaefer S, Lueders T. (2007). Detection of anaerobic toluene and hydrocarbon degraders in contaminated aquifers using benzylsuccinate synthase (*bssA*) genes as a functional marker. *Environ Microbiol* **9**: 1035-1046.

Winderl C, Penning H, von Netzer F, Meckenstock RU, Lueders T. (2010). DNA-SIP identifies sulfate-reducing *Clostridia* as important toluene degraders in tar-oil-contaminated aquifer sediment. *Isme Journal* **4**: 1314-1325.

Zhou L, Li K-P, Mbadinga SM, Yang S-Z, Gu J-D, Mu B-Z. (2012). Analyses of *n*-alkanes degrading community dynamics of a high-temperature methanogenic consortium enriched from production water of a petroleum reservoir by a combination of molecular techniques. *Ecotoxicology (London, England)* **21**: 1680-1691.

## 7 Conclusion and synthesis

### 7.1 Understanding alkane degradation using a model system

Arguably, oil sands tailings ponds have been criticized as being a highly toxic environment that does not support life. Contrary to that belief, oil sands tailings ponds are teeming with microbial life. More importantly, the growth and survival of this microbial life depends on the degradation of hydrocarbons in the form of residual solvents used as bitumen diluents. Because naphtha diluent contains a large fraction of alkanes, it is deemed important to understand how the substrates can be used as a carbon source by tailings organisms. In order to better understanding methanogenic alkane degradation that has been poorly described in the literature, SCADC was established from oil sands tailings mature fine tailings for intensive metagenomics, metatranscriptomics and metabolomics studies. Hydrocarbon degradation in SCADC and therefore tailings ponds is a dynamic process, and can be catalyzed by both syntrophic microorganisms under methanogenic conditions or by sulfate-reducing bacteria under sulfidogenic conditions (Chapter 1). Under both conditions, the tailings ponds organisms degrade alkanes by fumarate addition and possibly other pathways as well.

Using "omics" approaches many insights into methanogenic alkane degradation were obtained using the SCADC culture. For example, bioinformatics analysis has provided a snapshot of the potential role of a large consortium of bacterial communities in SCADC. These microbial communities are likely present as secondary fermentors by removing by-products such as acetate and hydrogen, thus making the initial process of hydrocarbon degradation feasible (Chapter 4). The focus of this research, however, was understanding bottleneck reactions by identifying potential key alkane degraders and the mechanisms involved. The SCADC community is capable of fumarate addition activation of monoaromatics, alkanes and 2-methylnaphthalene (Chapter 3). The organism capable of degrading low molecular weight alkanes in SCADC under methanogenic conditions has been identified to be a novel *Desulfotomaculum* (Chapter 4 and 5). This organism activates low molecular weight alkanes including *n*-alkanes and *iso*-alkanes by fumarate addition and completely

degrades their intermediates during syntrophic growth in SCADC in the absence of sulfate (Chapter 1 and Chapter 5). In the presence of sulfate, this organism can be outcompeted by sulfate-reducing bacteria that are similarly capable of alkane activation by fumarate addition coupled to sulfate reduction (Chapter 2). The *assA* gene encoding the alpha-subunit of the fumarate addition enzyme ASS in *Desulfotomaculum* spp. is located in a gene cluster that has novel organization compared to the canonical *ass* gene cluster in *Desulfatibacillum alkenivorans* Ak-01 (Chapter 4). Furthermore, the *assA* in *Desulfotomaculum* SCADC belongs to a phylogenetic Cluster IV (c) *assA*, which is distinctive compared to all other *assA* genes reported in the literature, with the exception of multiple putative *assA* genes detected in marine hydrocarbon seeps (Chapter 6). Based on this, I hypothesised that *assA* Cluster IV (c) is likely involved in the degradation of novel hydrocarbon substrates. This hypothesis will be difficult to prove because previous efforts to isolate *Desulfotomaculum* SCADC have been difficult, likely due to it being an obligate syntroph. However, SCADC culture, which has now been transferred several times in the laboratory, is highly enriched in *Desulfotomaculum* SCADC (Chapter 4) and it can be expected to be an extremely useful culture for future physiological analysis and intensive isolation efforts.

Another key organism (*Smithella* spp.) which has previously been suggested to be involved in alkane degradation in cultures reported by others (Chapter 4) is also present in SCADC as a closely related species. In SCADC, *Smithella* spp. is not able to degrade low molecular weight alkanes (Chapter 4) and instead is hypothesized to be involved in fumarate addition activation of longer-chain alkanes (i.e., *n*-hexadecane) under methanogenic conditions (Appendix A). Because SCADC contains at least two strains of *Smithella* spp., the true substrate spectrum of these organisms cannot be ascertained in this study. With the technological advances of single cell sorting and genome sequencing, isolation in the traditional sense almost seems unnecessary. Furthermore, physiological aspects of a single bacterial species in a community can be probed using community-wide physiological tests such as RNA-seq, and metabolomics (Haroon *et al.*, 2013). The ability to obtain a single genome through progressive

bioinformatic binning methods from the metagenome of a mixed culture will have major implications for biotechnological purposes because information obtained from these methods can provide the basis for creation of artificial genomes (Gibson *et al.*, 2008) that can serve specific purposes such as being used in a microbial agent for bioremediation of contaminated sites.

## **7.2 Opportunities in developing commercial cultures for bioremediation**

Comparing the functional capability of SCADC to the metagenomes of two other methanogenic hydrocarbon-degrading cultures indicates that methanogenic hydrocarbon-degrading cultures are highly conserved in their overall functions, being able to degrade multiple hydrocarbon substrates, having streamlined methanogenesis pathways and anaerobic respiration. In contrast, in other anoxic (methanogenic) environment, the microbial communities are capable of a wide variety of functions that are absent in model methanogenic hydrocarbon-degrading cultures (Chapter 5). This does not necessarily suggest inflexible metabolic functions in these methanogenic hydrocarbon-degrading cultures. Rather, streamlined methanogenic hydrocarbon degrading functions in these cultures suggests that the communities are stable overall, and are enriched in key functions relevant for hydrocarbon conversion to methane. Further investigation and physiological probing of these cultures to investigate their substrate spectrum and nutrient requirements for optimum growth can be expected to yield insights into ways to manipulate community-wide functions for desirable properties such as increased methane production and acceleration of hydrocarbon degradation. This could have potential future commercial applications in the biomethanization of petroleum reservoirs for tertiary hydrocarbon recovery in the form of methane (Gieg *et al.*, 2008).

## **7.3 Fumarate addition is a universal ubiquitous anaerobic hydrocarbon-activating mechanism**

Genes encoding alpha-subunits of ASSA, BSSA and NMSA have been used as indicators of the ability of microbial communities to use fumarate addition to activate hydrocarbons for degradation. PCR amplification using DNA

isolated from different environments has detected a high diversity of *assA/bssA/nmsA* genes in many anoxic environments (Chapter 6). Further data mining of several metagenomes including two metagenomes obtained from oil sands tailings ponds, one each from marine hydrocarbon seeps and forest compost in Puerto Rico, further identified phylogenetically distinct *assA* genes (Cluster IV (c) and Cluster V in Chapter 6). Thus, it's reasonable to conclude that fumarate addition genes are ubiquitous and possibly a universal hydrocarbon activating mechanism for substituted aromatic compounds such as toluene and saturated hydrocarbons such as *n*-alkanes under anoxic conditions. Even though phylogenetic analysis of fumarate addition genes may not necessarily be accurate in assigning substrate spectrum, the assignment of sequences into novel clusters can certainly be used as a basis for high throughput screening of novel substrates that can be degraded by novel clusters of genes (not limited to fumarate addition genes). Further data mining of functionally important genes will be important in providing the basis for future docking analysis and protein bio-engineering (i.e., engineering of proteins that have broad substrate spectrum to target different groups of hydrocarbons for biomethanization).

#### **7.4 Petroleum microbiology in the genomic era**

Above and all, metagenomics analysis in the present study has provided insights into hydrocarbon degradation in oil sands tailings ponds based on studies with enrichment cultures. The results presented in this thesis and by others (An *et al.*, 2013) in the Hydrocarbon Metagenomic Project (<http://www.hydrocarbonmetagenomics.com/>) are expected to provide a sound basis for accurate modelling of methane emissions from oil sands tailings ponds based on degradability of different classes of unrecovered residual naphtha and bitumen and genetic capabilities of microbes indigenous to oil sands tailings ponds (Siddique *et al.*, 2008). More importantly, the detection of novel genes and organisms in oil sands tailings ponds and cultures derived from these systems can be expected to provide a basis for the development of molecular tools that can be used as indicators for *in situ* anaerobic hydrocarbon degradation in contaminated anoxic environments such as aquifers or marine sediments. The milestones and

legacy of this research project is long term in part because the data are now available for future data comparison and mining of sequences (i.e., hypothetical genes) that may provide future insights into processes unknown currently.

## 7.5 References

An D, Caffrey SM, Soh J, Agrawal A, Brown D, Budwill K *et al.*, (2013). Metagenomics of hydrocarbon resource environments indicates aerobic taxa and genes to be unexpectedly common. *Environ Sci Technol* **47**: 10708-10717.

Gibson DG, Benders GA, Andrews-Pfannkoch C, Denisova EA, Baden-Tillson H, Zaveri J *et al.*, (2008). Complete chemical synthesis, assembly, and cloning of a *Mycoplasma genitalium* genome. *Science* **319**: 1215-1220.

Gieg LM, Duncan KE, Suflita JM. (2008). Bioenergy production via microbial conversion of residual oil to natural gas. *Appl Environ Microbiol* **74**: 3022-3029.

Haroon MF, Hu SH, Shi Y, Imelfort M, Keller J, Hugenholtz P *et al.*, (2013). Anaerobic oxidation of methane coupled to nitrate reduction in a novel archaeal lineage. *Nature* **500**: 567.

Siddique T, Gupta R, Fedorak PM, MacKinnon MD, Foght JM. (2008). A first approximation kinetic model to predict methane generation from an oil sands tailings settling basin. *Chemosphere* **72**: 1573-1580.

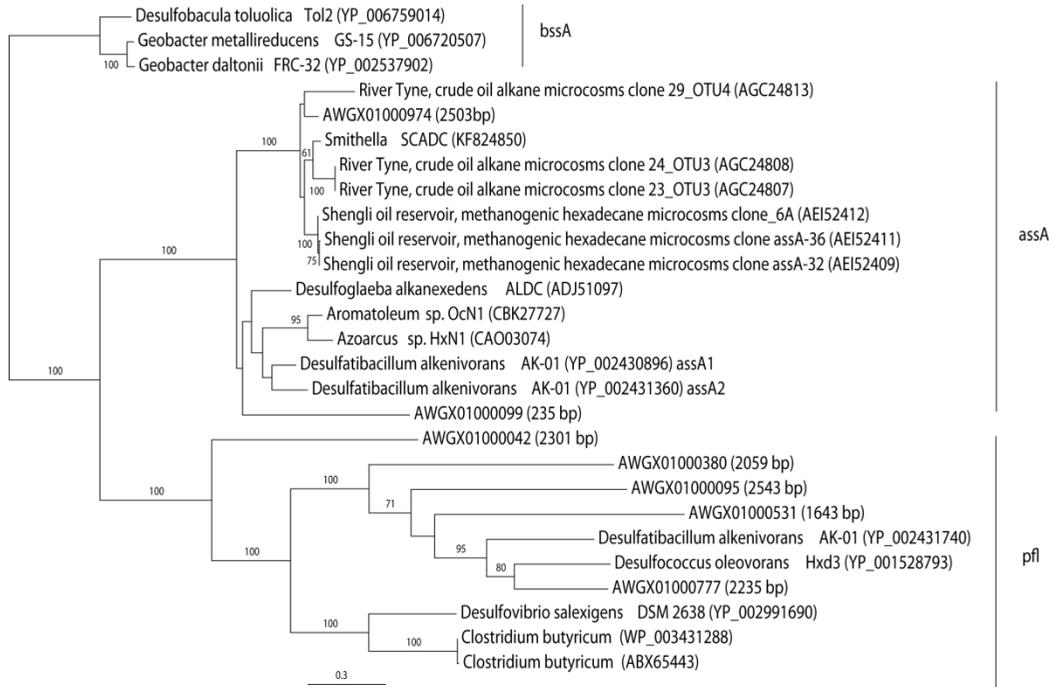
## 8 Appendix A: Re-analysis of 'omics data provides evidence for addition of *n*-hexadecane to fumarate under methanogenic conditions

Biodegradation of *n*-alkanes is rapid and well understood under aerobic conditions. In contrast, under methanogenic conditions alkane degradation is slower and requires a consortium of microbes, most of which have not been isolated; hence the metabolic pathways are largely unknown (Aitken *et al.*, 2013, Callaghan 2013). Understanding methanogenic *n*-alkane degradation is crucial for optimizing bioremediation in anaerobic environments, modelling formation of heavily biodegraded petroleum in oil reservoirs and potentially for converting unrecovered residual hydrocarbons in petroleum reservoirs to methane for recovery as a valuable fuel. The most widely reported mechanism for anaerobic *n*-alkane degradation is addition to fumarate. The initial hydrocarbon-activating reaction under nitrate- and sulfate-reducing conditions is catalyzed by the glycyl radical enzyme alkylsuccinate synthase (ASS), encoded by the *assABC* genes (Callaghan 2013). Due to sequence homology and conservation of the *assA* gene encoding the  $\alpha$ -subunit of ASS, its presence has been used as a diagnostic marker to indicate the genetic capability for alkane activation by addition to fumarate (Callaghan 2013). Alternative pathways, i.e., carboxylation and hydroxylation, have been proposed for the sulfate reducer *Desulfococcus oleovorans* Hxd3, but the corresponding genes and enzymes have yet to be confirmed (Callaghan 2013). Under methanogenic conditions several studies have implicated *Syntrophus* spp. and/or the phylogenetically similar but largely uncultivated genus *Smithella* as primary *n*-alkane degraders (Cheng *et al.*, 2013, Gray *et al.*, 2011, Zengler *et al.*, 1999). However, the activating mechanism(s) remain cryptic because signature alkylsuccinate metabolites have not been detected *in situ* or in cultures, and *assA* genes have not been assigned to either genus (Aitken *et al.*, 2013, Callaghan 2013). Thus, alternative but unknown alkane activation mechanisms have been proposed (Aitken *et al.*, 2013).

Embree *et al.* (2013) recently published the draft genome of a *Smithella* spp. (DDBJ/EMBL/GenBank accession number AWGX00000000) from an *n*-hexadecane-degrading methanogenic enrichment culture. The authors used

fluorescence activated cell sorting to separate six bacterial cells related to *Smithella* from the community, amplified DNA using whole-genome multiple displacement amplification, sequenced using Illumina Hi-seq, assembled using a *de novo* co-assembler, and annotated the draft genome using the RAST server (Embree *et al.*, 2013). Furthermore, they obtained metatranscriptomes from enrichment cultures grown on hexadecane, butyric acid or caprylic acid as the only organic carbon source. RAST failed to annotate *assABC* genes in the draft *Smithella* genome and therefore mapping of metatranscripts from the enrichment cultures to the genome did not indicate expression of these genes in any of the three cultures. Embree *et al.* (2013) therefore concluded that *Smithella* is incapable of *n*-alkane activation by addition to fumarate, even though they observed transcription of genes homologous to those encoding activating enzymes in hexadecane-degraders and expression of fatty acid utilization genes required for  $\beta$ -oxidation of *n*-alkanes. We recently analyzed the metagenome of a methanogenic short-chain alkane-degrading enrichment culture (SCADC; Tan *et al.*, 2013) and recovered a partial *Smithella* genome in which we detected a single copy of *assA* (KF824850; unpublished). Hence, we used this sequence plus several annotated *assA* genes (e.g., Callaghan *et al.*, 2012) as probes in BLASTX and tblastn screening of the *Smithella* draft genome reported by Embree *et al.* (2013). We found seven copies of genes encoding glyceryl radical enzymes including homologs of ASS and pyruvate formate lyase (PFL) (Fig. A1). Phylogenetic analysis of the recovered genes indicated that the draft *Smithella* genome does indeed contain a single copy of nearly full-length *assA* (accession number AWGX01000974; contig\_5325; contig length, 2584 bp) that was not previously annotated in the draft genome (Embree *et al.*, 2013). The translated sequence of this gene and a gene fragment (AWGX01000099; contig\_9960; contig length, 235 bp) are related to the *assA* annotated in *Smithella* SCADC (Tan *et al.*, 2013) and to other *assA* genes recovered from enrichment cultures where *Syntrophus* or *Smithella* have been implicated as primary alkane degraders (Cheng *et al.*, 2013, Gray *et al.*, 2011) (Fig. A1). The likely reason for the inadvertent failure of Embree *et al.* (2013) to detect *assA* is that it was not

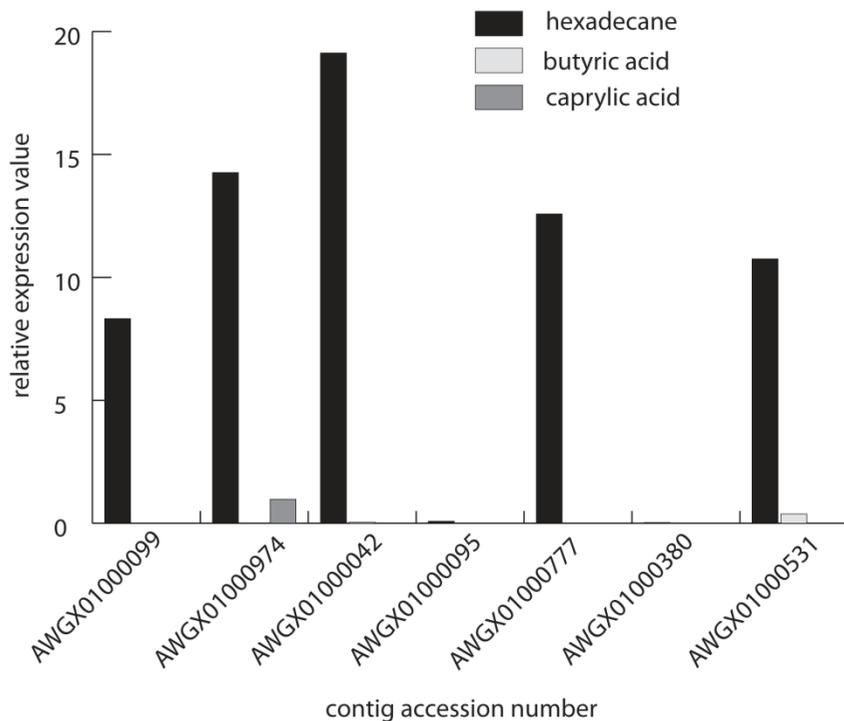
annotated by the automated RAST server, possibly due to the short sequence length (<3.0 kb); however, it was readily detected using manual sequence similarity searches.



**Figure A1** (Next page). Maximum likelihood tree of translated *assA* homologs in draft *Smithella* genome (Embree *et al.*, 2013) recovered through BLASTX and tblastn 155 searches (shown in boldface, with the SCADC *assA* sequence). Closely related sequences were recovered from the NCBI nr-database through BLASTX searches. All sequences were aligned using Muscle 3.3, followed by phylogenetic tree construction using PhyML with WAG model and 100 bootstrap replicates in Geneious R7. Bootstrap support  $\geq 60\%$  is indicated. Full length translated benzylsuccinate synthase  $\alpha$ -subunit (*bssA*) sequences were used as the outgroup. The length of nucleotide sequences recovered from the draft *Smithella* spp. genome (Embree *et al.*, 2013) is shown. The *assA* sequences from clones (indicated as such on the tree) were not full-length and ranged from 414-662 bp. All other sequences used in the tree were full length (>2400 bp). A tree with the same overall topology was obtained when including only full-length sequences and removing gaps.

We also re-analyzed transcription of the newly annotated *assA* gene under the three growth substrates (*n*-hexadecane, butyric acid or caprylic acid) used by

Embree *et al.* (2013). We mapped the metatranscripts (GSE49830) to the seven *Smithella assA* homologs (Fig. 1A) using CLC Genomics Workbench with two mismatches allowed per read alignment. In so doing, we discovered that *assA* homologs on contigs AWGX01000099 and AWGX01000974 were actually highly transcribed during growth on hexadecane but not on caprylic or butyric acid (Fig. A2). Expression of putative PFL genes on contigs AWGX01000531 and AWX01000777 (Fig. A2) is likely involved in converting pyruvate to acetyl-coA and formate (Lu *et al.*, 2012), common in methanogenic substrate degradation.



**Figure A2** (Next page) Re-analysis of relative transcription of *assA* homologs in the metatranscriptome of enrichment cultures grown under methanogenic conditions with *n*-hexadecane, butyric acid or caprylic acid (Embree *et al.*, 2013). Relative transcription was calculated as:  $(\text{Total transcripts mapped to a gene fragment} \times 10^9) / (\text{fragment length} \times \text{total transcript abundance in that library})$

Embree *et al.* (2013) reported detection of genes in the *Smithella* draft genome encoding  $\alpha$ -methylacyl-coA racemase and methyl-malonyl-coA that are proposed, respectively, to epimerize and carboxylate metabolic intermediates in the fumarate activation pathway used under nitrate- and sulfate-reducing conditions (Jarling *et al.*, 2012; Callaghan *et al.*, 2012). Genes for  $\beta$ -oxidation of fatty acids were also highly transcribed (Embree *et al.*, 2013), as would be expected if *Smithella* was utilizing *n*-hexadecane via a fumarate activation pathway. Based upon our re-analysis of published sequences, we reach a conclusion opposite to that of Embree *et al.* (2013), confirming instead that *Smithella* spp. is indeed capable of activating and utilizing long-chain alkanes like *n*-hexadecane under methanogenic conditions by addition to fumarate.

## References

- Aitken CM, Jones DM, Maguire MJ, Gray ND, Sherry A, Bowler BFJ *et al.*, (2013). Evidence that crude oil alkane activation proceeds by different mechanisms under sulfate-reducing and methanogenic conditions. *Geochim Cosmochim Acta* **109**: 162-174.
- Callaghan AV, Morris BEL, Pereira IAC, McInerney MJ, Austin RN, Groves JT *et al.*, (2012). The genome sequence of *Desulfatibacillum alkenivorans* AK-01: a blueprint for anaerobic alkane oxidation. *Environ Microbiol* **14**: 101-113.
- Callaghan AV. (2013). Enzymes involved in the anaerobic oxidation of *n*-alkanes: from methane to long-chain paraffins. *Front Microbiol* **4**: 89.
- Cheng L, Ding C, Li Q, He Q, Dai LR, Zhang H. (2013). DNA-SIP reveals that Syntrophaceae play an important role in methanogenic hexadecane degradation. *PLoS ONE* 8(7): e66784. doi:10.1371/journal.pone.0066784
- Embree M, Nagarajan H, Movahedi N, Chitsaz H, Zengler K. (2013). Single-cell genome and metatranscriptome sequencing reveal metabolic interactions of an alkane-degrading methanogenic community. *ISME J* advance on-line, Oct 24, 2013. doi:10.1038/ismej.2013.187
- Gray ND, Sherry A, Grant RJ, Rowan AK, Hubert CRJ, Callbeck CM *et al.*, (2011). The quantitative significance of Syntrophaceae and syntrophic partnerships in methanogenic degradation of crude oil alkanes. *Environ Microbiol* **13**: 2957-2975.

Jarling R, Sadeghi M, Drozdowska M, Lahme S, Buckel W, Rabus R *et al.*, (2012). Stereochemical investigations reveal the mechanism of the bacterial activation of *n*-alkanes without oxygen. *Angewandte Chemie-Int Ed* **51**: 1334-1338.

Lu W, Du J, Schwarzer NJ, Gerbig-Smentek E, Einsle O, Andrade SLA. (2012). The formate channel FocA exports the products of mixed-acid fermentation. *Proc Nat Acad Sci USA* **109**: 13254-13259.

Tan B, Dong XL, Sensen CW, Foght JM. (2013). Metagenomic analysis of an anaerobic alkanedegrading microbial culture: potential hydrocarbon-activating pathways and inferred roles of community members. *Genome* 56:599-611

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

## 9 Appendix B - Supporting information for Chapter 2

Table B1. Calculation of the maximum theoretical methane yield from *n*-alkanes added to active 75-mL SCADC enrichment cultures.

**Table B2.** Pairwise BLASTp similarity between the *assA* genes recovered in the current study compared to the almost-complete *assA* genes detected in pure isolates.

**Figure B1.** Amino acid sequence alignment of translated *assA* OTU1, OTU2 and OTU3 (KC934958-KC934960) reverse transcribed from methanogenic alkane-degrading culture with *assA* reference sequences obtained from GenBank database. Alignment was conducted using MUSCLE v3.3 (Edgar 2004). Numbers on top of the alignment denote position of amino acid. All sequences contain the conserved motif of IVRVX for glyceryl radical enzymes at position 152 to 156.

**Figure B2.** Neighbour-joining tree constructed with representative 16S rRNA gene sequences from selected OTUs (in bold) shown in Figure 2.11. Numbers on branches are bootstrap values based on 1000 replicates.

**Table B1.** Calculation of the maximum theoretical methane yield from *n*-alkanes added to active 75-mL SCADC enrichment cultures.

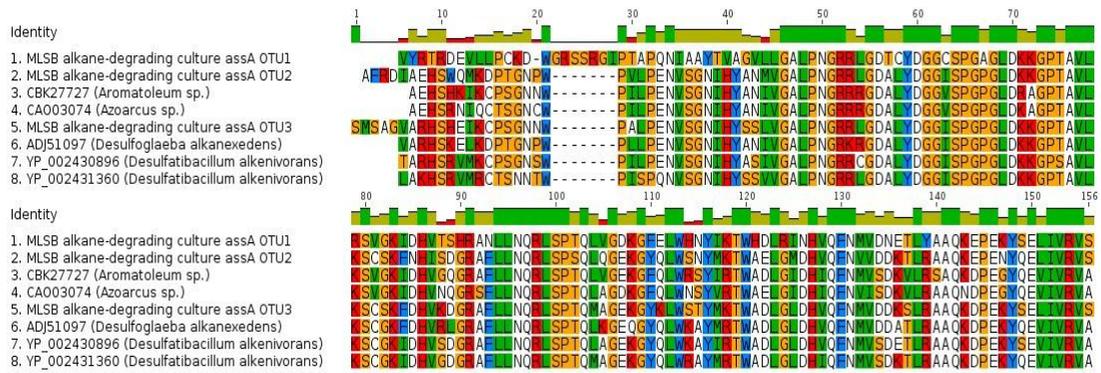
substrates	Density (mg/ $\mu$ L)	Weight (mg)	Molar mass (mg/mmol)	mmol	Mol methane expected ( $\mu$ mol) <sup>a</sup>
C6	0.6548	12.278	86.18	0.1425	677
C7	0.6840	12.825	100.21	0.1280	704
C8	0.7080	13.275	114.23	0.1162	726
C10	0.7300	13.688	142.29	0.0962	745
Maximum theoretical methane yield from C <sub>6</sub> -C <sub>10</sub> mixture					2852

Note: Substrate stock was prepared by mixing equal volumes of C<sub>6</sub>-C<sub>10</sub>. Seventy-fix microlitres of the substrate stock was added to 75 mL of culture, representing 0.1% vol. Each added alkane therefore represents 18.75  $\mu$ L of the total added substrate stock.

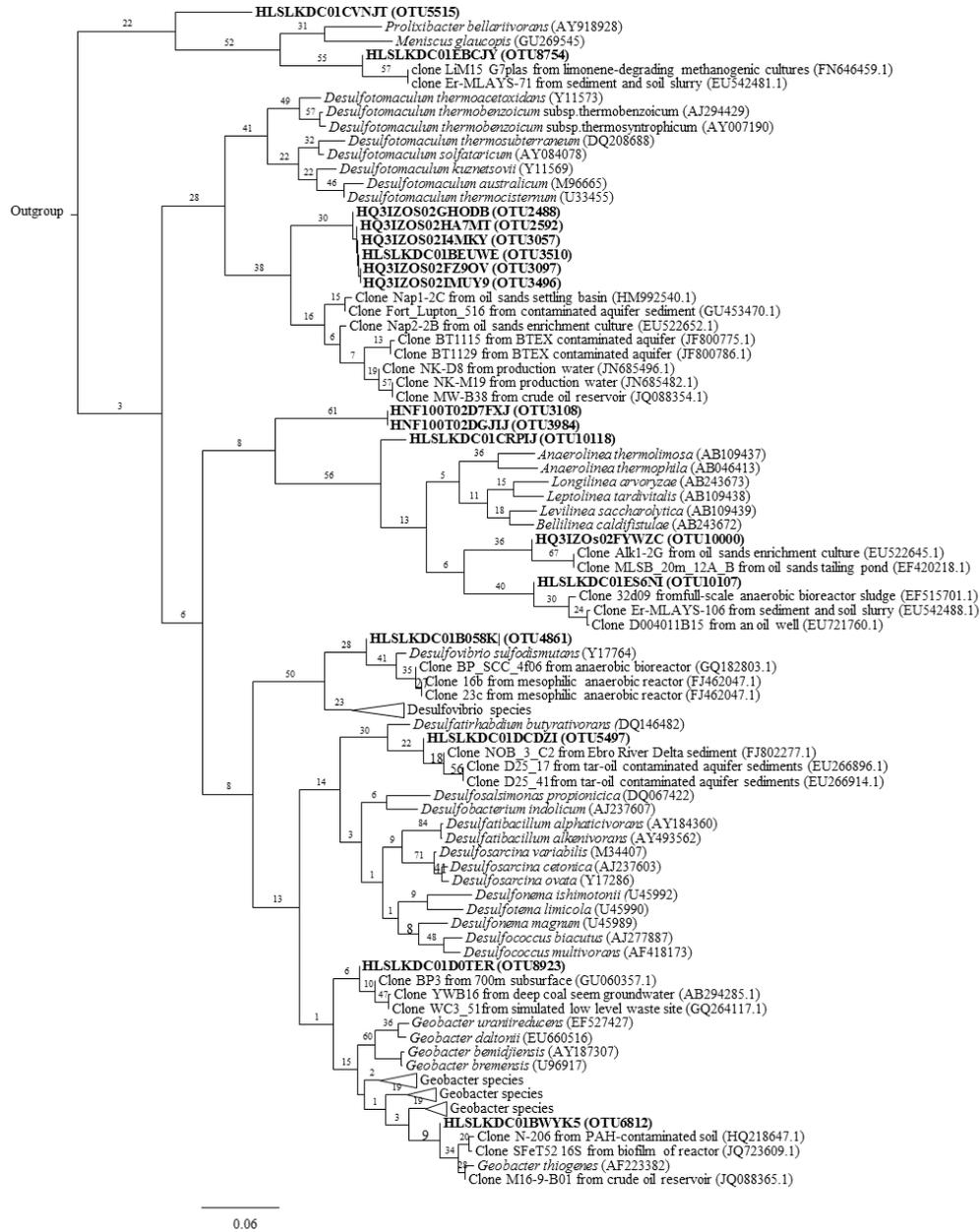
<sup>a</sup> Mol methane expected from added substrate was calculated according to Symons and Buswell equation (Roberts, 2002).

**Table B2.** Pairwise BLASTp similarity between the *assA* genes recovered in the current study compared to the almost-complete *assA* genes detected in pure isolates.

	<i>D. alkenivorans</i> AK-01 assA1 (YP_002430896)	<i>D. alkenivorans</i> AK-01 assA2 (YP_002431360)	<i>Aromatoleum</i> sp. OcN1 (CBK27727)	<i>Desulfoglaeba</i> <i>alkanexedens</i> ALDC (ADJ51097)	<i>Azoarcus</i> sp. HxN1 (CAO03074)
<i>assA</i> OTU1	59.1	63.2	62.4	63.3	60.9
<i>assA</i> OTU2	76.4	73.6	74.1	78.4	72.7
<i>assA</i> OTU3	83.3	81.9	77.6	81.9	72.7



**Figure B1.** Amino acid sequence alignment of translated *assA* OTU1, OTU2 and OTU3 (KC934958-KC934960) reverse transcribed from methanogenic alkane-degrading culture with *assA* reference sequences obtained from GenBank. Alignment was conducted using MUSCLE v3.3 (Edgar 2004). Numbers at the top of the alignment denote amino acid position. All sequences contain the conserved motif of IVRVX for glycyl radical enzymes at positions 152 to 156.



**Figure B2.** Neighbour-joining tree constructed with representative 16S rRNA gene sequences from selected OTUs (in bold) shown in Figure 2.11. Numbers on branches are bootstrap values based on 1000 replicates.

## References

Edgar RC. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32**: 1792-1797.

Roberts D J. Methods for assessing anaerobic biodegradation potential. In *Manual of Environmental Microbiology*, 2nd ed.; Hurst, C. J., Crawford, R. L., Knudson, G. R., McInerney, M. J., Stetzenbach, L. D., Eds.; ASM Press: Washington, DC, 2002; pp 1008–1017.

## 10 Appendix C - Supporting information for Chapter 3

**Table C1.** Reference genes used in tBLASTn searches of the SCADC metagenome.

**Table C2.** Headspace methane and volatile hydrocarbons measured in triplicate 75 mL SCADC enrichment cultures after incubation in the dark for 108 weeks under strictly methanogenic conditions with a mixture of C6-C10 alkanes, as described in Methods section.

**Table C3:** General features of SCADC metagenome.

**Table C4.** Taxonomic binning of contigs >1000 bp using PhylopythiaS, using the generic model.

**Table C5 (Excel).** Putative *assA/bssA/nmsA* homologs recovered from SCADC metagenome.

**Table C6 (Excel).** Annotation of ORFs for contigs ckmer83\_42244 and 21910.

**Table C7 (Excel).** BLASTP comparison of ORFs detected in SCADC contigs ckmer83\_42244 and 21910 against ORFs in the *assA1* operon in the reference genome of *D. alkenivorans* AK-01.

**Figure C1.** Schematic flow diagram showing the steps used in establishing SCADC enrichment cultures used for the metagenomic analysis in this study.

**Figure C2.** Total ion GC-MS chromatogram of headspace volatile hydrocarbons in uninoculated control medium, showing the relative proportions of *n*-alkanes and 'hexanes' isomers.

**Figure C3.** Top panel: Length distribution of SCADC Illumina and 454 assembled contigs; Bottom panel: Length distribution of sequence reads obtained from 454 pyrosequencing.

**Figure C4:** Maximum likelihood tree of putative *mcrA* genes (Table S1) retrieved from SCADC 454 and Illumina assembly. Sequences retrieved from SCADC metagenome are shown in red, and partial to near-full length reference sequences retrieved from NCBI database are shown in black. Numbers on

branches are bootstrap values based on 100 replications. The tree was rooted with *Methanopyrus kandleri* AV19 (NC\_003551.1). Numbers next to leaves that have been collapsed are number of contigs recovered from the SCADC metagenome.

**Figure C5:** Maximum likelihood tree of putative *bcrA/bzdQ* genes retrieved from SCADC 454 and illumina assembly. Sequences retrieved from the SCADC metagenome are shown in red, and partial to near full-length reference sequences retrieved from NCBI database are shown in black. Numbers on branches are bootstrap values based on 100 replications. The tree is rooted at the midpoint.

**Figure C6:** Maximum likelihood tree of putative *bamB* genes (Table S1) recovered from SCADC 454 and Illumina assembly. Sequences retrieved from the SCADC metagenome are shown in red, and partial to near full-length reference sequences retrieved from the NCBI database are shown in black. The tree is rooted at the midpoint. Numbers on branches are bootstrap values based on 100 replications. Numbers next to leaves that have been collapsed are the number of contigs recovered from SCADC metagenome.

**Figure C7:** Completeness of pathways related to methanogenesis and carbon fixation. Pathway completeness was determined as follows: All KEGG orthologs (KOs) annotated in IMG were imported into the KEGG module mapper (Kanehisa *et al.*, 2012) for pathway prediction. Completeness was scored as the percentage of the number of detected KO components divided by the total expected number of components within a module. However, the incompleteness of a pathway cannot be construed as lack of a pathway, due to possible mis-annotation of genes and possible substitution of function by related genes.

**Table C1.** Reference genes used in tBLASTn searches of the SCADC metagenome.

Gene name	Organism	NCBI accession number	Substrate
alkylsuccinate synthase, $\alpha$ -subunit ( <i>assA</i> )	<i>Desulfoglaeba alkanexedens</i> ALDC	ADJ51097	alkanes
benzylsuccinate synthase, $\alpha$ -subunit ( <i>bssa</i> )	<i>Azoarcus</i> sp. T	AAK50372	toluene
naphthylmethylsuccinate synthase, $\alpha$ -subunit ( <i>nmsA</i> )	NaphS6	CAO72222	2-methyl-naphthalene
putative benzene carboxylase, $\alpha$ -subunit ( <i>abcA</i> )	Bacteria enrichment culture BF	ADJ94001	benzene
putative benzene carboxylase, $\square$ -subunit ( <i>abcD</i> )	Bacteria enrichment culture BF	ADJ94002	benzene
particulate methane monooxygenase ( <i>pmoA</i> )	<i>Methylomicrobium japonense</i>	BAE86885	methane
alkane-1-monooxygenase ( <i>alkM</i> )	Gammaaproteobacterium HdN1	YP_003809668.1	alkanes
monooxygenase flavin-binding protein ( <i>FMO</i> )	Gammaaproteobacterium HdN1	YP_003810690.1	alkanes
ethylbenzene dehydrogenase, $\alpha$ -subunit ( <i>ebdA</i> )	<i>Aromatoleum aromaticum</i> EbN1	Q5NZV2_AZOSE	ethylbenzene
ethylbenzene dehydrogenase, $\beta$ -subunit ( <i>ebdB</i> )	<i>Aromatoleum aromaticum</i> EbN1	YP_158332	ethylbenzene
ethylbenzene dehydrogenase, $\gamma$ -subunit ( <i>ebdC</i> )	<i>Aromatoleum aromaticum</i> EbN1	YP_158331	ethylbenzene
benzoyl-CoA reductase, $\alpha$ -subunit ( <i>bcrA</i> )	<i>Azoarcus</i> sp. CIB	AAQ08809	benzoyl-CoA
benzoyl-CoA reductase, $\alpha$ -subunit ( <i>bamB</i> )	<i>Geobacter metallireducens</i> GS-15	YP_006720765	benzoyl-CoA
methyl-coenzyme M reductase ( <i>mcrA</i> )	<i>Methanosaeta concilii</i> GP6	YP_004383383	-
methyl-coenzyme M reductase ( <i>mcrA</i> )	<i>Methanolinea tarda</i> NOB1-1	ZP_09042590	-
Putative ethylbenzene dehydrogenase (Dole_0914)	<i>Desulfococcus oleovorans</i> Hxd3	YP_001528081.1	-

**Table C2.** Headspace methane and volatile hydrocarbons measured in triplicate 75 mL SCADC enrichment cultures after incubation in the dark for 108 weeks under strictly methanogenic conditions with a mixture of C6-C10 alkanes, as described in Methods section.

Headspace analyte	Residual analyte detected in headspace <sup>a, b</sup>
<i>n</i> -C6	1.0 ± 1.4 %
<i>n</i> -C7	2.5 ± 0.7 %
<i>n</i> -C8	5.2 ± 1.2 %
<i>n</i> -C10	18.3 ± 1.3 %
2-methylpentane	0 %
methylcyclopentane	0 %
3-methylpentane (used as internal standard)	100%

<sup>a</sup>Values are the mean of triplicate cultures ± 1 standard deviation.

<sup>b</sup>GC-MS analysis of C6-C10 hydrocarbons remaining in the culture bottles after 108 weeks incubation, expressed as a percentage of the initial volatile hydrocarbon after normalization to the recalcitrant compound 3-methylpentane (used as an internal standard), and determined using the method described by Prince and Suflita (2007). Briefly, the percentage of target compounds remaining in the headspace was calculated based on the following equation: % headspace analyte =  $[(A_{\text{sample}}/C_{\text{sample}})/(A_{\text{sterile}}/C_{\text{sterile}})] \times 100$ , where A and C represent the headspace abundance of target analyte and 3-methylpentane, respectively. The uninoculated sterile control contains sterile medium with identical amount of hydrocarbon substrate as described in Materials and Methods.

**Table C3:** General features of SCADC metagenome.

	Illumina +454 assembly	454 sequencing data
<b>General features<sup>a</sup></b>		
Number of reads	984,086	667,134
Maximum contig read length	513,327	641
Minimum contig read length	200	100
Mean contig read length	657	355
N50	1002	391
<b>Annotation features from IMG</b>		
<b>RNA genes</b>	11,899 (0.8) <sup>b</sup>	4,955 (0.7)
rRNA genes	2,277 (0.2)	2,007 (2)
5 S rRNA	399 (0.03)	151 (0.02)
16 S rRNA	690 (0.1)	637 (0.1)
18 S rRNA	6 (<0.01)	3 (<0.01)
23 S rRNA	1,166 (0.1)	1,214 (0.2)
28 S rRNA	16 (<0.01)	2 (<0.01)
tRNA genes	9,622 (0.6)	2,948 (0.4)
<b>Protein coding genes</b>	1,501,746 (99)	751,978 (99)
With product name	774,691 (51)	318,243 (42)
With COG	786,927 (52)	337,993 (45)
With Pfam	944,525 (62)	397,292 (52)
With KO	575,143 (38)	273,719 (36)
With Enzyme	325,109 (21)	157,838 (21)
With MetaCyc	222,187 (15)	108,961 (14)
With KEGG	339,712 (22)	165,856 (22)

<sup>a</sup> General features including number of reads, maximum contig length, minimum contig length, mean contig length, and N50 were obtained for reads that passed quality control as described in Materials and Methods, generated using the in-house pipeline.

<sup>b</sup>Numbers in parentheses denote percentage of reads obtained from analysis using the DOE Joint Genome Institute IMG web server.

**Table C4.** Taxonomic binning of contigs >1000 bp using PhylopythiaS, using the generic model.

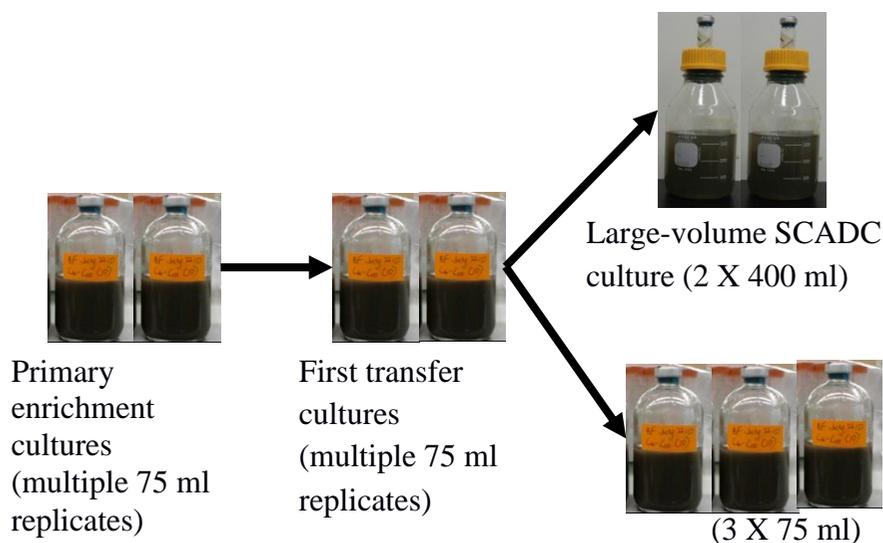
Phylum	Class/Order	Contig count	Percentage
<b><u>Bacteria</u></b>		<b>60,065</b>	<b>91.4</b>
Firmicutes	<b>Clostridia</b>	3,536	5.4
	Clostridiales	2,850	4.3
Firmicutes	<b>Negativicutes</b>	687	1.1
	Selenomonadales	687	1.1
Chlorobi	<b>Chlorobia</b>	797	1.2
	Chlorobiales	797	1.2
Bacteroidetes	-	1,740	2.7
Proteobacteria	<b>Alphaproteobacteria</b>	1,995	3.0
Proteobacteria	<b>Betaproteobacteria</b>	11,172	1.8
	Nitrosomonadales	792	1.2
Proteobacteria	<b>Gammaproteobacteria</b>	2,412	3.7
Proteobacteria	<b>Deltaproteobacteria</b>	7,968	12.1
	Syntrophobacterales	2,392	3.6
	Desulfuromonadales	765	1.2
	Desulfobacterales	1,956	3.0
Spirochaetes	<b>Spirochaetales</b>	2,677	4.1
<b><u>Archaea</u></b>		<b>5,654</b>	<b>8.6</b>
Euryarchaeota	<b>Methanomicrobia</b>	3,464	5.3
	Methanosarcinales	1,392	2.1
	Methanomicrobiales	1,712	2.6

Note: Only taxa at the class level representing >1% of the total reads are reported.

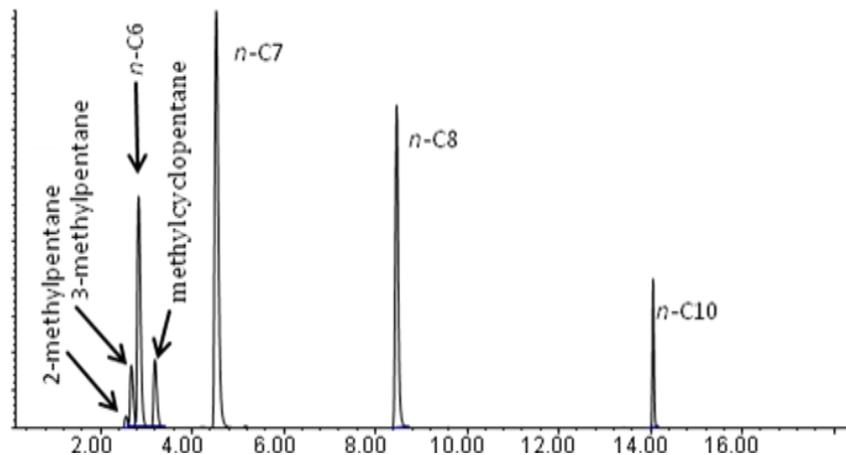
**Table C5 (Excel).** Putative *assA/bssA/nmsA* homologs recovered from SCADC metagenome.

**Table C6 (Excel).** Annotation of ORFs for contigs ckmer83\_42244 and 21910.

**Table C7 (Excel).** BLASTP comparison of ORFs detected in SCADC contigs ckmer83\_42244 and 21910 against ORFs in the *assA1* operon in the reference genome of *D. alkenivorans* AK-01.



**Figure C1.** Schematic flow diagram showing the steps used in establishing SCADC enrichment cultures used for the metagenomic analysis in this study. Multiple “primary enrichment” cultures were established by adding 25 ml of mature fine tailings from Mildred Lake Settling Basin to 50 ml of methanogenic medium in 158 ml serum bottles under a headspace of O<sub>2</sub>-free 30% CO<sub>2</sub>-balance N<sub>2</sub> gas. After incubating in the dark at room temperature for ~1 year, the primary cultures were pooled and used to inoculate multiple replicate “first transfer” cultures by adding ~37 ml of culture to 37 ml fresh methanogenic medium (50 vol% transfer). After incubation for at least 6 months, these first transfer cultures were pooled as an inoculum for the SCADC enrichment cultures (replicate 75 ml and duplicate 400 ml volumes; 50 vol% transfer). Duplicate 75 ml unamended enrichment cultures were similarly prepared. The 400 ml SCADC enrichment cultures were incubated for 4 months before being sacrificed for total DNA extraction. The 75 ml enrichment cultures were used for headspace and metabolite analysis during an additional 2 years of incubation. All enrichment cultures were amended with a 0.1 vol% of the alkane mixture described in the Methods section of the main text. The methanogenic medium used in establishing SCADC was prepared according to Widdel and Bak (1992), gassed with O<sub>2</sub>-free 30% CO<sub>2</sub> balance N<sub>2</sub> for at least 30 min, and immediately capped with a butyl-rubber seal and aluminum crimp before autoclaving. A redox indicator (resazurin) and a reducing agent (sulfide) were added as described by Fedorak and Hruddy (1984). All enrichment cultures were incubated at 28 °C in the dark with occasional manual agitation (once per week), unless stated otherwise. To demonstrate continued viability of the cultures, methane measurement in headspace gas was conducted routinely and microscopic observation was performed periodically.



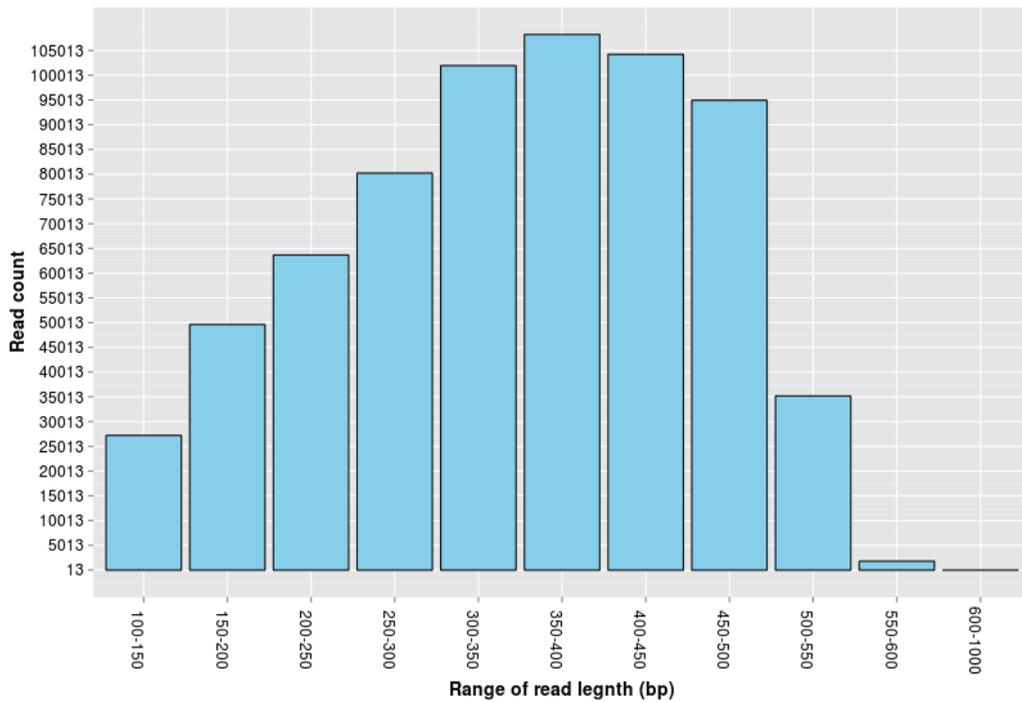
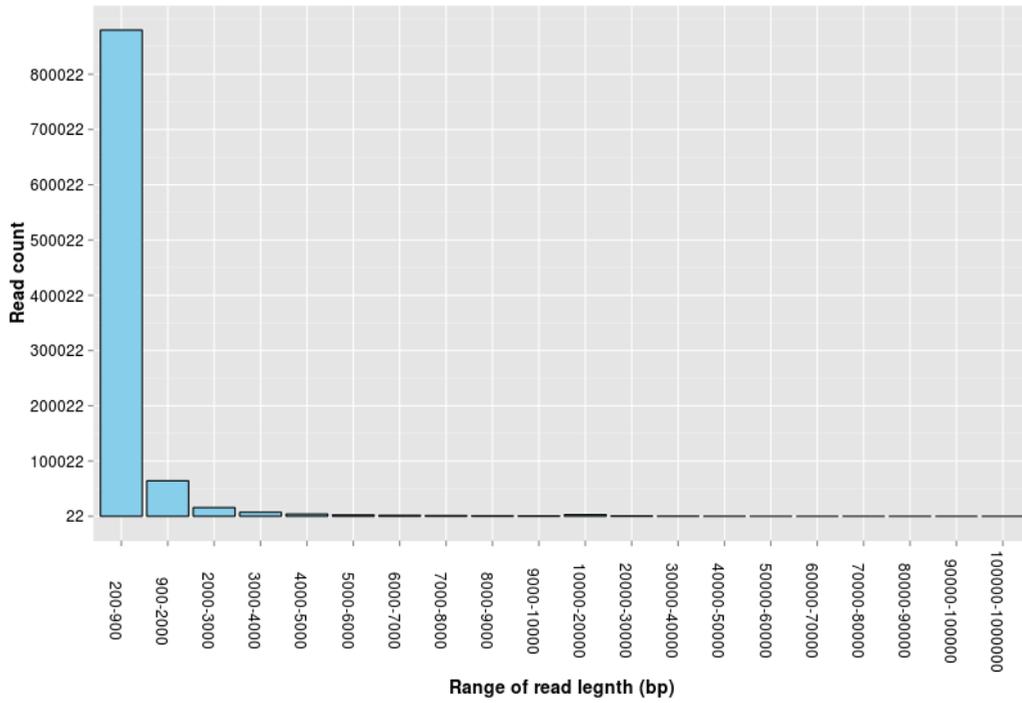
**Figure C2.** Total ion GC-MS chromatogram of headspace volatile hydrocarbons in uninoculated control medium, showing the relative proportions of *n*-alkanes and ‘hexanes’ isomers.

Percentage of 2-methylpentane in total ‘hexanes’ = 3%

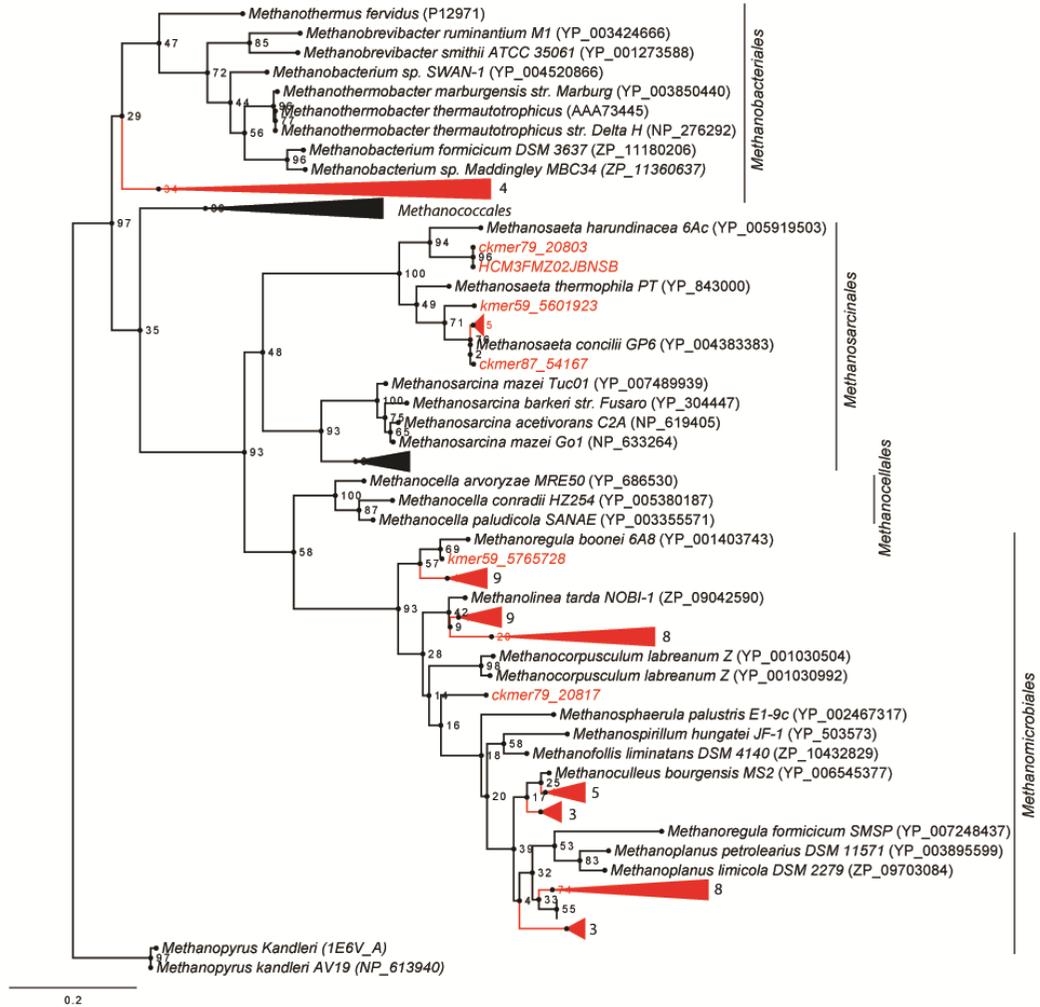
Percentage of 3-methylpentane in total ‘hexanes’ = 16%

Percentage of *n*-C<sub>6</sub> in total ‘hexanes’ = 62%

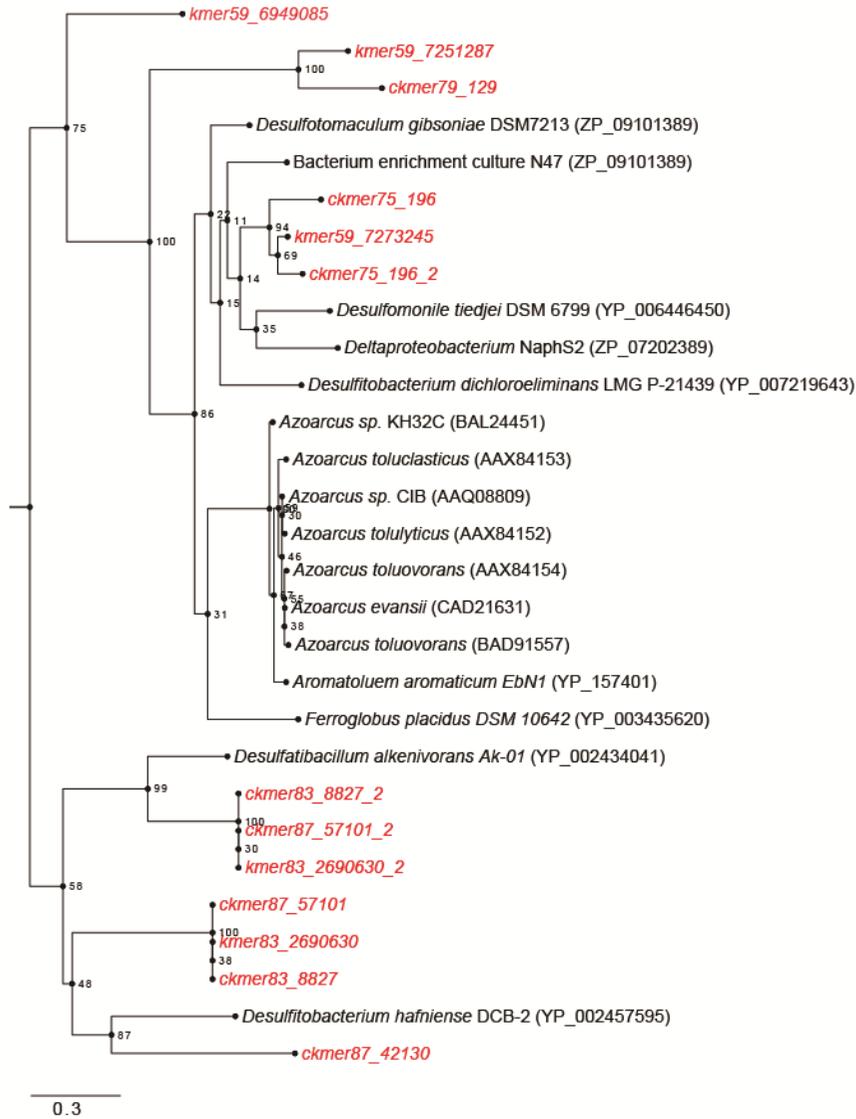
Percentage of methylcyclopentane in total ‘hexanes’ = 19%



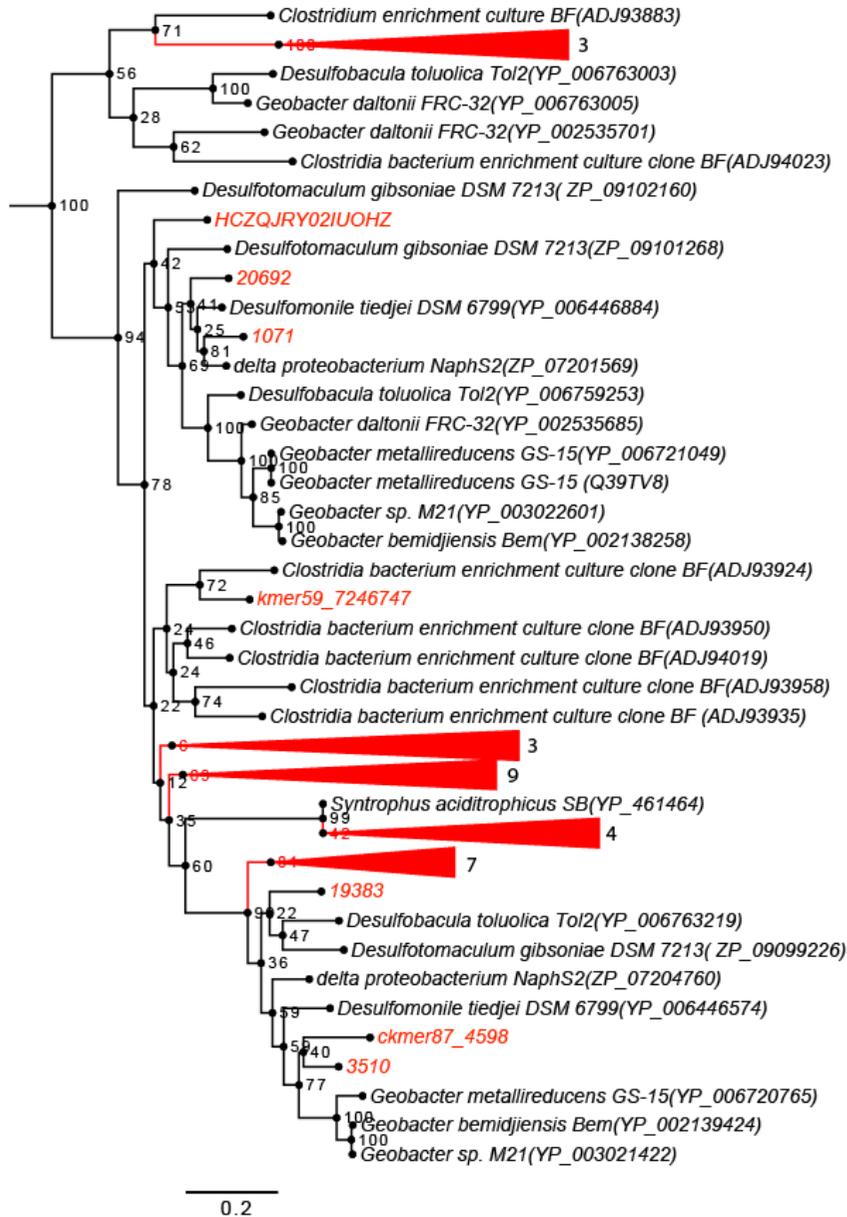
**Figure C3.** Top panel: Length distribution of SCADC Illumina and 454 assembled contigs; Bottom panel: Length distribution of sequence reads obtained from 454 pyrosequencing.



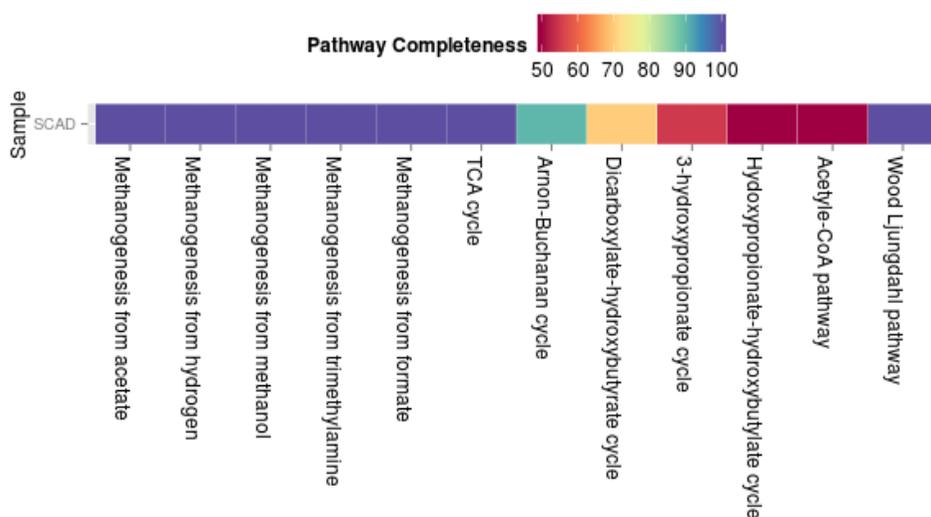
**Figure C4:** Maximum likelihood tree of putative *mcrA* genes retrieved from SCADC 454 and Illumina assembly. Sequences retrieved from SCADC metagenome are shown in red, and partial to near-full length reference sequences retrieved from NCBI database are shown in black. Numbers on branches are bootstrap values based on 100 replications. The tree was rooted with *Methanopyrus kandleri* AV19 (NC\_003551.1). Numbers next to leaves that have been collapsed are the number of contigs recovered from the SCADC metagenome.



**Figure C5:** Maximum likelihood tree of putative *bcrA/bzdQ* genes retrieved from SCADC 454 and illumina assembly. Sequences retrieved from the SCADC metagenome are shown in red, and partial to near full-length reference sequences retrieved from NCBI database are shown in black. Numbers on branches are bootstrap values based on 100 replications. The tree is rooted at the midpoint.



**Figure C6:** Maximum likelihood tree of putative *bamB* genes (Table S1) recovered from SCADC 454 and Illumina assembly. Sequences retrieved from the SCADC metagenome are shown in red, and partial to near full-length reference sequences retrieved from the NCBI database are shown in black. The tree is rooted at the midpoint. Numbers on branches are bootstrap values based on 100 replications. Numbers next to leaves that have been collapsed are the number of contigs recovered from SCADC metagenome.



**Figure C7:** Completeness of pathways related to methanogenesis and carbon fixation. Pathway completeness was determined as follows: All KEGG orthologs (KOs) annotated in IMG were imported into the KEGG module mapper (Kanehisa *et al.*, 2012) for pathway prediction. Completeness was scored as the percentage of the number of detected KO components divided by the total expected number of components within a module. However, the incompleteness of a pathway cannot be construed as lack of a pathway, due to possible mis-annotation of genes and possible substitution of function by related genes.

## Appendix C -References

- Fedorak, P.M., and Hrudey, S.E. 1984. The effects of Phenol and some alkyl phenolics on batch anaerobic methanogenesis. *Water Res* **18**(3): 361-367. doi:10.1016/0043-1354(84)90113-1.
- 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**(D1): D109-D114. doi:10.1093/nar/gkr988.
- Prince, R.C., and Suflita, J.M. 2007. Anaerobic biodegradation of natural gas condensate can be stimulated by the addition of gasoline. *Biodegradation* **18**(4): 515-523. doi:10.1007/s10532-006-9084-4.
- Widdel, F., and Bak, F. 1992. Gram-negative mesophilic sulfate-reducing bacteria. In *The Prokaryotes*. Edited by A. Balows, Trüper, H. G., Dworkin, M., Harder, W. and Schleifer, K.H. Springer-Verlag, New York, USA. pp. 3352-3378.

## 11 Appendix D - Supporting information for Chapter 4

### Methods

**Figure D1.** Maximum likelihood tree of the translated amino acid sequences of *assA* genes recovered from *Desulfotomaculum* SCADC, *Smithella* SCADC and Syntrophaceae 2 (red). All sequences were aligned using MUSCLE (Edgar 2004), followed by phylogenetic analysis using PhyML v3.3 (Guindon *et al.*, 2010) with WAG model and 100 bootstrap replicates. The tree is rooted with the glycerol dehydratase gene of *Clostridium butyricum* (ABW38357) and bootstrap values >60 are shown on branches. The naming of *assA* clusters (i.e., cluster IV a and b) was done according to (Acosta-González *et al.*, 2013).

**Figure D2.** Maximum likelihood tree of the translated amino acid sequences of *dsrC* genes recovered from SCADC (red) and SCADC metagenome (green). Translated *dsrC* and *dsrAB* sequence from the sulfate non-reducers *Pelobacter propionicus* DSM2379 and *Pelotomaculum thermopropionicum* SI, respectively, are shown in purple. All sequences were aligned using MUSCLE v3.3 (Edgar 2004), followed by phylogenetic analysis using PhyML (Guindon *et al.*, 2010) with WAG model and 100 bootstrap replicates. The tree is rooted in the midpoint and bootstrap values >60 are shown on branches.

**Figure D3.** Prophage prediction for concatenated *Desulfotomaculum* SCADC metagenomic bin using PHAST (Zhou *et al.*, 2011), showing the presence of genes presumptive encoding an intact phage. All genes associated with this contig can be found in Appendix Table C5.

**Table D1.** Completeness of SCADC metagenomic bins not-associated with *assA* and methanogens.

**Table D2.** Number of essential marker genes present in selected SCADC metagenomic bins for bacterial species.

**Table D3.** Number of essential marker genes present in selected SCADC metagenomic bins for archaeal species.

**Table D4.** Annotation of contig containing *nmsAB* genes and genes affiliated with members of Clostridiales.

**Table D5.** FPKG of annotated "*Desulfotomaculum*" SCADC.

**Table D6.** FPKG of annotated SCADC *Methanoculleus* 1 bin.

**Table D7.** FPKG of annotated SCADC *Methanoculleus* 2 bin.

**Table D8.** FPKG of *Methanosaeta concilii* GP6 using SCADC metatranscriptomic reads.

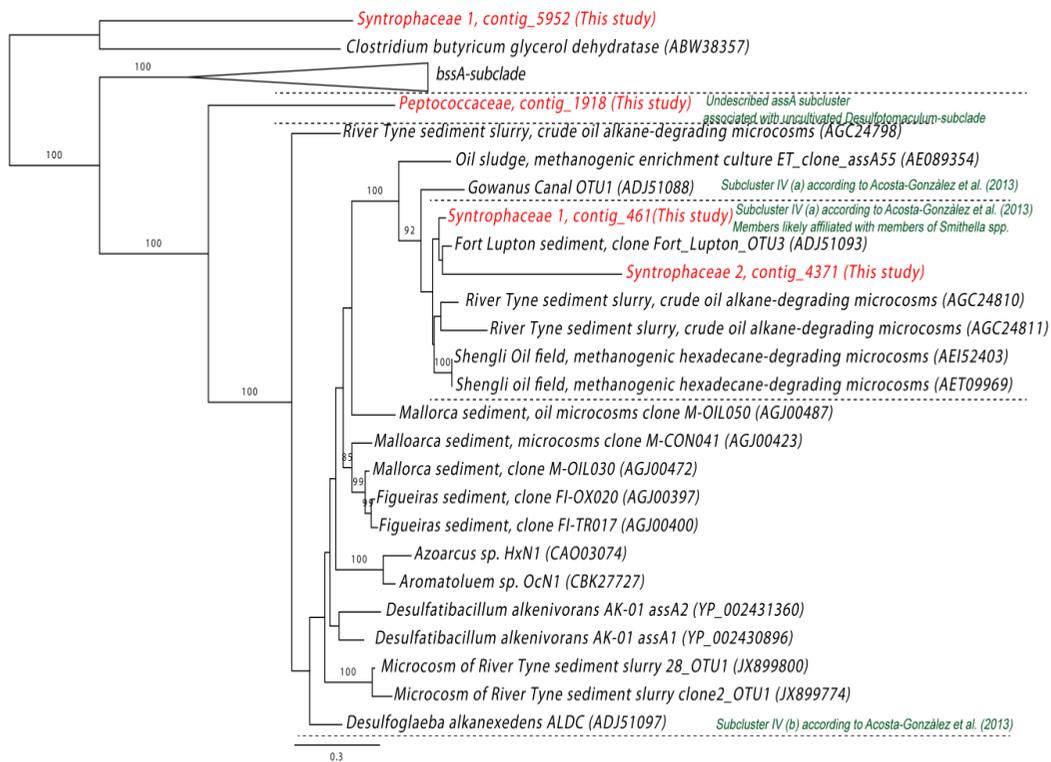
## Appendix D Methods

Illumina sequences generated from the SCADC enrichment culture, reported in Tan *et al.*, (2013) were assembled using CLC Genomics Workbench (CLC Bio, USA) using the default settings. The workflow used in the current study was adapted from Haroon *et al.* (2013) and Ishii *et al.* (2013). A comprehensive workflow of the binning method is presented in Figure 4.1 (main text).

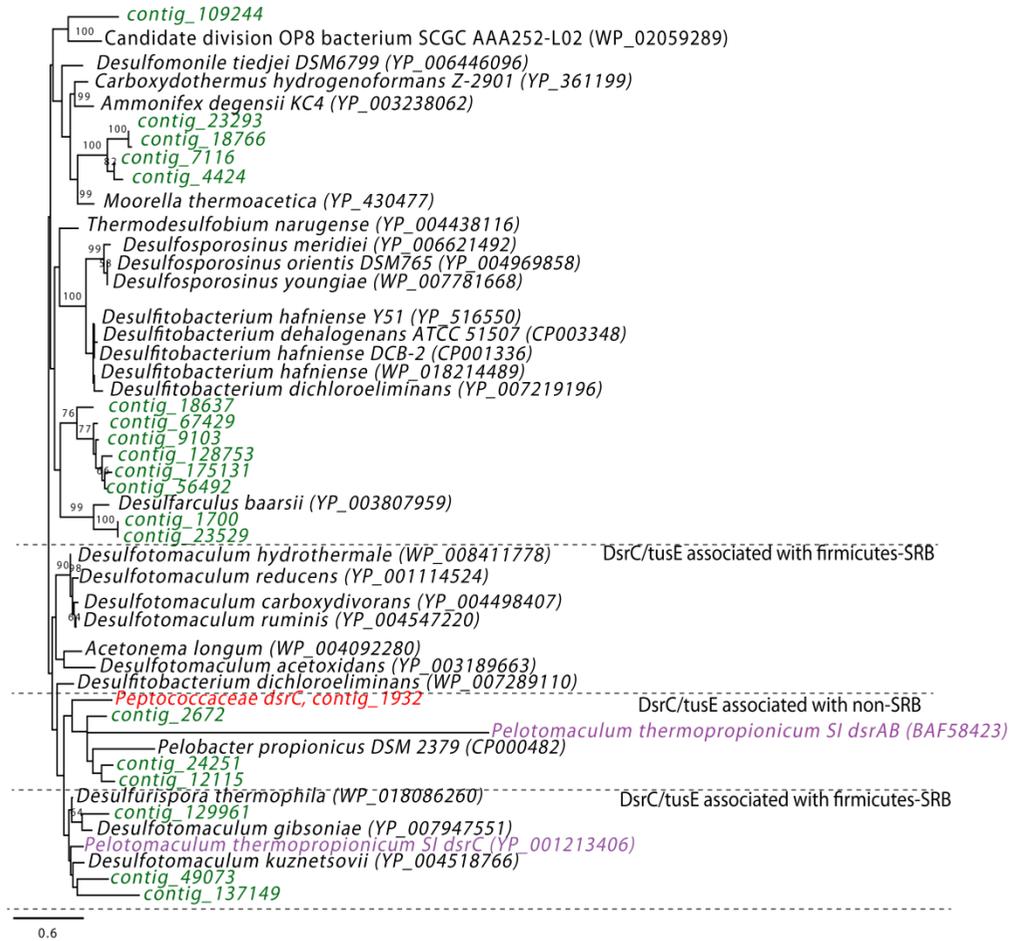
The frequencies or sequence coverage of each contig were obtained by mapping Illumina reads to each contig that shared 96% similarity over 95% of sequence length. Contigs were first binned based on GC content and sequence coverage (Figures 4.2 and 4.3). Open reading frames (ORFs) were predicted using PRODIGAL (Hyatt *et al.*, 2010), followed by best BLAST hits against the NCBI nr-database for taxonomic classification. Contigs were highlighted with identical colours when containing taxonomic signatures, i.e., essential marker genes determined using the gene list compiled by Albertsen *et al.*, (2013), and BLASTP of open reading frames (ORFs) that shared >80% similarity to selected reference genomes (i.e., *Methanosaeta concilii* GP6, *Syntrophus aciditrophicus* SB1, etc.).

Contigs within a selected region were extracted and subjected to tetranucleotide frequency analysis, followed by clustering and binning using principal component analysis (Figure 4.4), and Emergent Self-Organizing Map (ESOM) (Dick *et al.*, 2009)(Figure 4.5) based on the tetranucleotide composition of each contig. Essential marker genes were overlaid on the ESOM map to guide genomic bin extraction. Completeness of selected bins was determined by comparing the number of essential genes to that in the closest affiliated organisms using the gene list compiled by Albertsen *et al.*, (2013). Extracted bins were subjected to automated annotation using the RAST server (Aziz *et al.*, 2008). Pathway reconstruction was in part performed using PathoLogic and Pathway tools (Karp *et al.*, 2011) based on the taxonomic class of each bin, and also by manual reconstruction based on the pathway inferred for model organisms, such as the alkane-degrader *Desulfatibacillum alkenivorans* AK-01 (Callaghan *et al.*,

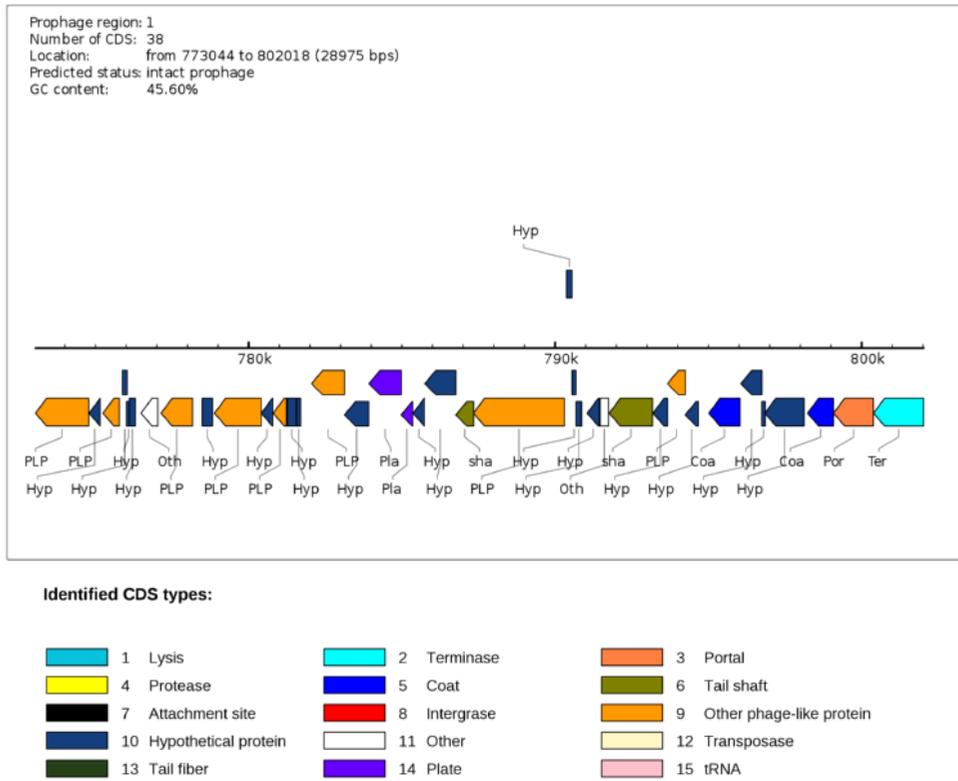
2012), or by referring to MetaCyc (Caspi *et al.*, 2012) for the closest related species. Genes of interest (i.e., *assA*/*bssA*/*nmsA*) were manually curated by selecting best BLASTX searches against the NCBI nr-database. All ORFs on contigs of interest were searched against the NCBI database to confirm the taxonomic class of that contig. Annotated sequences of selected bins were subjected to RNA-seq using the CLC Genomics Workbench and transcript abundance was normalized against fragment length and expressed as "fragments mapped per kilobase of gene length" (Haroon *et al.*, 2013).



**Figure D1.** Maximum likelihood tree of the translated amino acid sequences of *assA* genes recovered from *Desulfotomaculum* SCADC, *Smithella* SCADC and Syntrophaceae 2 (red). All sequences were aligned using MUSCLE (Edgar 2004), followed by phylogenetic analysis using PhyML v3.3 (Guindon *et al.*, 2010) with WAG model and 100 bootstrap replicates. The tree is rooted with the glycerol dehydratase gene of *Clostridium butyricum* (ABW38357) and bootstrap values >60 are shown on branches. The naming of *assA* clusters (i.e., cluster IV a and b) was done according to Acosta-González *et al.* (2013).



**Figure D2.** Maximum likelihood tree of the translated amino acid sequences of *dsrC* genes recovered from *Desulfotomaculum* SCADC (red) and SCADC metagenome (green). Translated *dsrC* and *dsrAB* sequence from the sulfate non-reducers *Pelobacter propionicus* DSM2379 and *Pelotomaculum thermopropionicum* SI, respectively, are shown in purple. All sequences were aligned using MUSCLE v3.3 (Edgar 2004), followed by phylogenetic analysis using PhyML (Guindon *et al.*, 2010) with WAG model and 100 bootstrap replicates. The tree is rooted in the midpoint and bootstrap values >60 are shown on branches.



**Figure D3.** Prophage prediction for concatenated *Desulfotomaculum* SCADC metagenomic bin using PHAST (Zhou *et al.*, 2011), showing the presence of genes presumptively encoding an intact phage. Names of genes associated with this contig can be found in Appendix Table D5.

**Table D1. Completeness of SCADC metagenomic bins not associated with *assA* and methanogens**

Metagenomic bins <sup>a</sup>	Size (Mbp)	% GC	# essential genes	Closest affiliation (based on RAST)
Unclassified bacterium	1.2	31	118 (>100)	<i>Fervidobacterium nodosum</i> Rt17-B1
Treponema 1	2.3	53	88 (82)	<i>Treponema vincentii</i> ATCC 35580
Treponema 2	2.2	54	100 (93)	<i>Treponema vincentii</i> ATCC 35580
Lachnospiraceae	4.2	36	108 (100)	<i>Clostridium phytofermentans</i> ISDg
Desulfovibrio	3.1	67	90 (84)	<i>Desulfovibrio aespoeensis</i> Aspo-2
Syntrophaceae 3	3.1	52	106 (99)	<i>Syntrophus aciditrophicus</i> SB
Sphaerochaetaceae	2.7	52	92(84)	<i>Treponema denticola</i> ATCC
Sulfurospirillum	2.7	42	106 (99)	<i>Sulfurospirillum deleyianum</i> DSM
Thermovirga	2.1	54	101(94)	<i>Thermanaerovibrio</i>
Anaerolinea 1	2.9	49	104 (97)	<i>Anaerolinea thermophila</i> UNI-1
Anaerolinea 2	3.2	49	123	<i>Anaerolinea thermophila</i> UNI-1
Anaerolinea 3	2.8	54	111	<i>Herpetosiphon aurantiacus</i> ATCC
Anaerolinea 4	4.0	42	92 (86)	<i>Anaerolinea thermophila</i> UNI-1
Bacteroidetes 1	10.2	50	269	<i>Chitinophaga pinensis</i> DSM 2588
Bacteroidetes 2	4.8	42	207	<i>Parabacteroides distasonis</i> ATCC
Coriobacteraceae 1	3.0	65	97 (90)	<i>Eggerthella lenta</i> DSM 2243
Clostridium 1	5.3	55	208	<i>Alkaliphilus metalliredigens</i>
Clostridium 2	4.8	47	103 (96)	<i>Clostridium</i> sp. 7_2_43FAA
Firmicutes	3.7	50	107 (100)	<i>Thermincola</i> sp. JR
Elusimicrobium	1.8	61	82 (77)	<i>Elusimicrobium minutum</i> Pei191

<sup>a</sup>, Naming of metagenomic bin was on based taxonomic class associated with the closest related species determined using RAST automated annotation server (Aziz *et al.*, 2008).

<sup>b</sup>, The focus of this study was on genomic bins harbouring *assA*/*bsaA*; therefore the number of single copy genes in the closest reference genome presented in this table was not determined. Instead, completeness was calculated based on 108 single copy genes and may vary depending on species (completeness shown in bracket). Multiple genomes of closely related strains were present in some genomic bins, in particular those that contain >108 copies.

**Table D2.** Number of essential marker genes present in selected SCADC metagenomic bins for bacterial species

**Table D3.** Number of essential marker genes present in selected SCADC metagenomic bins for archaeal species

**Table D4.** Annotation of contig containing *nmsAB* genes and genes affiliated with members of Clostridiales

**Table D5.** FPKG of annotated *Desulfotomaculum* SCADC

**Table D6.** FPKG of annotated SCADC *Methanoculleus* 1

**Table D7.** FPKG of annotated SCADC *Methanoculleus* 2

**Table D8.** FPKG of *Methanosaeta concilii* GP6 using SCADC metatranscriptomic reads

Note: Table D2 to D8 can be downloaded from:

<https://dl.dropboxusercontent.com/u/67859694/Supplement-TableD2-D8.xls>

## 12 Appendix E - Supporting information for Chapter 5

**Figure E1.** Rarefaction curve of annotated species richness for NAPDC, SCADC and TOLDC. This curve is a plot of the total number of distinct species annotations as a function of the number of metagenome reads

**Figure E2.** Composition of microbial communities in NAPDC, SCADC and TOLDC using pyrosequencing of 16S rRNA gene amplicons and unassembled 454 metagenomic reads accessed using MG-RAST and SOrt-ITEMS. Relative abundance is expressed as the proportion of total archaeal reads (top panel) or bacterial reads (bottom panel).

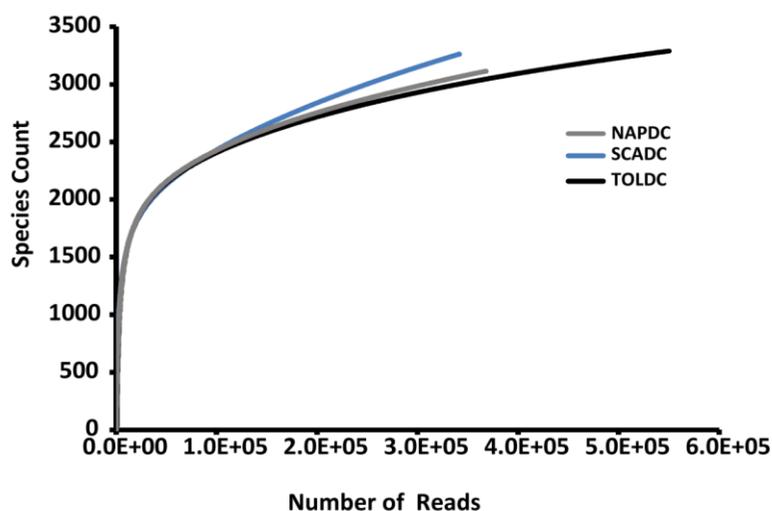
**Table E1.** List of metagenomes acquired from MG-RAST representing a variety of different environments and hydrocarbon-degrading microcosms

**Table E2.** Summary of recruitment of quality filtered 454 raw datasets to selected reference genomes.

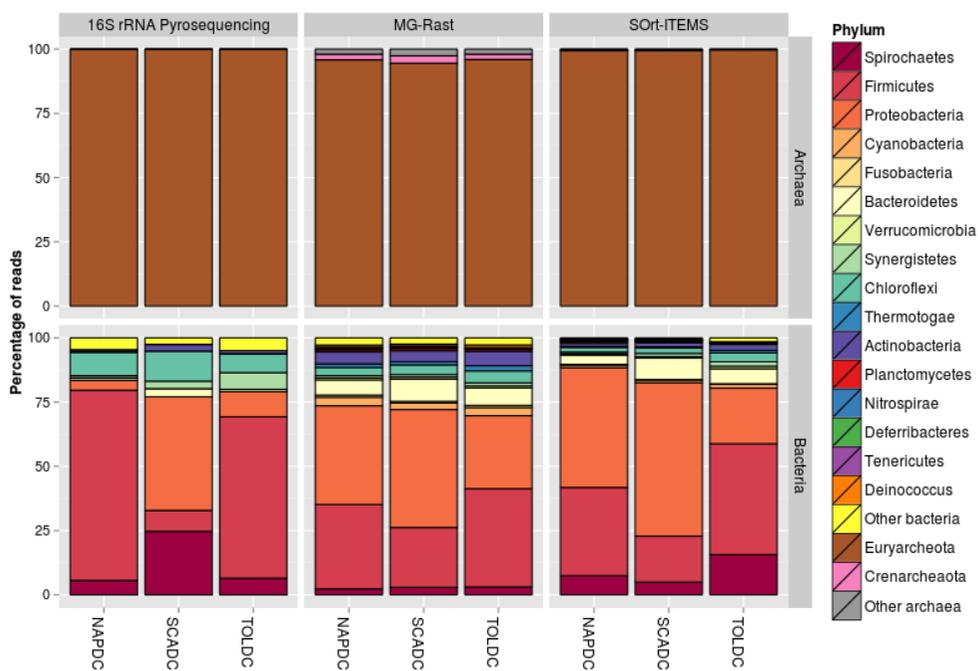
**Table E3.** Proportion of the functional categories the third-level SEED subsystem counts of TOLDC, SCAD, and NAPDC.

Note: Supplementary Table E1-E3 can be downloaded at:

<https://dl.dropboxusercontent.com/u/67859694/Supplement-TableE1-E3.xlsx>



**Figure E1.** Rarefaction curve of annotated species richness for NAPDC, SCADC and TOLDC. This curve is a plot of the total number of distinct species annotations as a function of the number of metagenome reads



**Figure E2.** Composition of microbial communities in NAPDC, SCADC and TOLDC using pyrosequencing of 16S rRNA gene amplicons and unassembled 454 metagenomic reads accessed using MG-Rast and Sort-ITEMS. Relative abundance is expressed as the proportion of total archaeal reads (top panel) or bacterial reads (bottom panel).

### 13 Appendix F - Supporting information for Chapter 6

**Table F1.** Qualitative measurement of volatile hydrocarbons remaining after 490 d incubation using GC-MS

**Figure F1.** Standard curves for the quantification of G1, G2, G3 shown in Figure 6.4 in main text

**Figure F2** Phylogenetic tree of translated *bssA* sequences amplified using the primer set of Winderl *et al.* (2007) and translated *assA* amplified using primer set 9 (Callaghan *et al.*, 2010). Putative amino acid sequences were grouped at 0.02 difference using MOTHUR, and a representative OTU from each clone library was used in tree construction using maximum likelihood with 100 bootstrap replicates. Bootstrap values >50 are shown at each branch. Pyruvate formate lyase from *Desulfotomaculum reducens* M1-1(PFLYP001114293) was used as the outgroup.

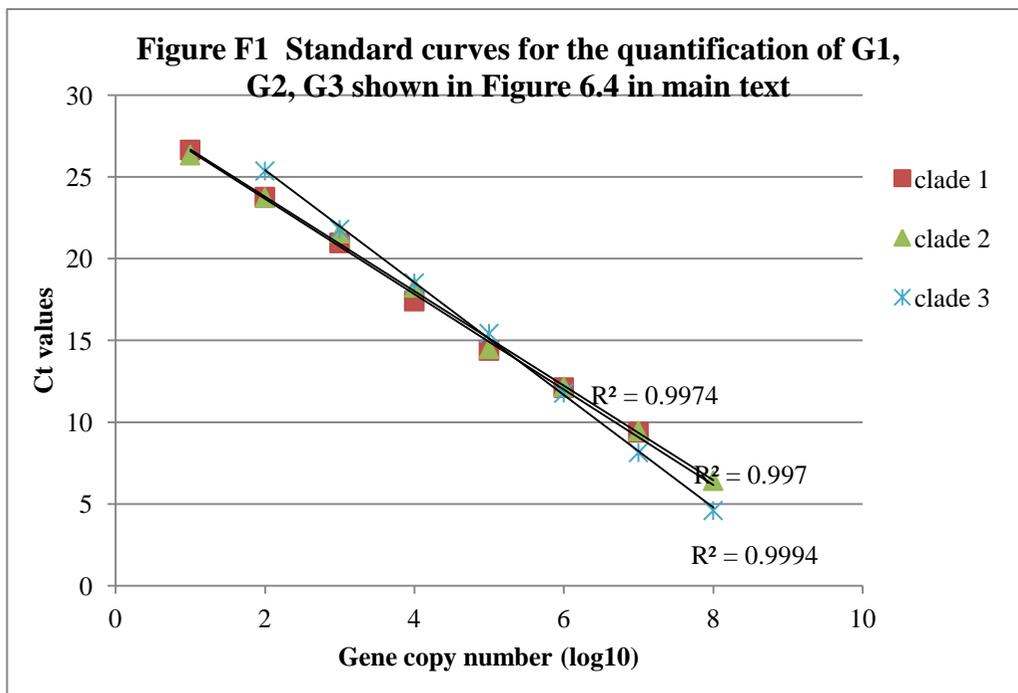
**Table F1.** Qualitative measurement of volatile hydrocarbons remaining after 490 d incubation using GC-MS

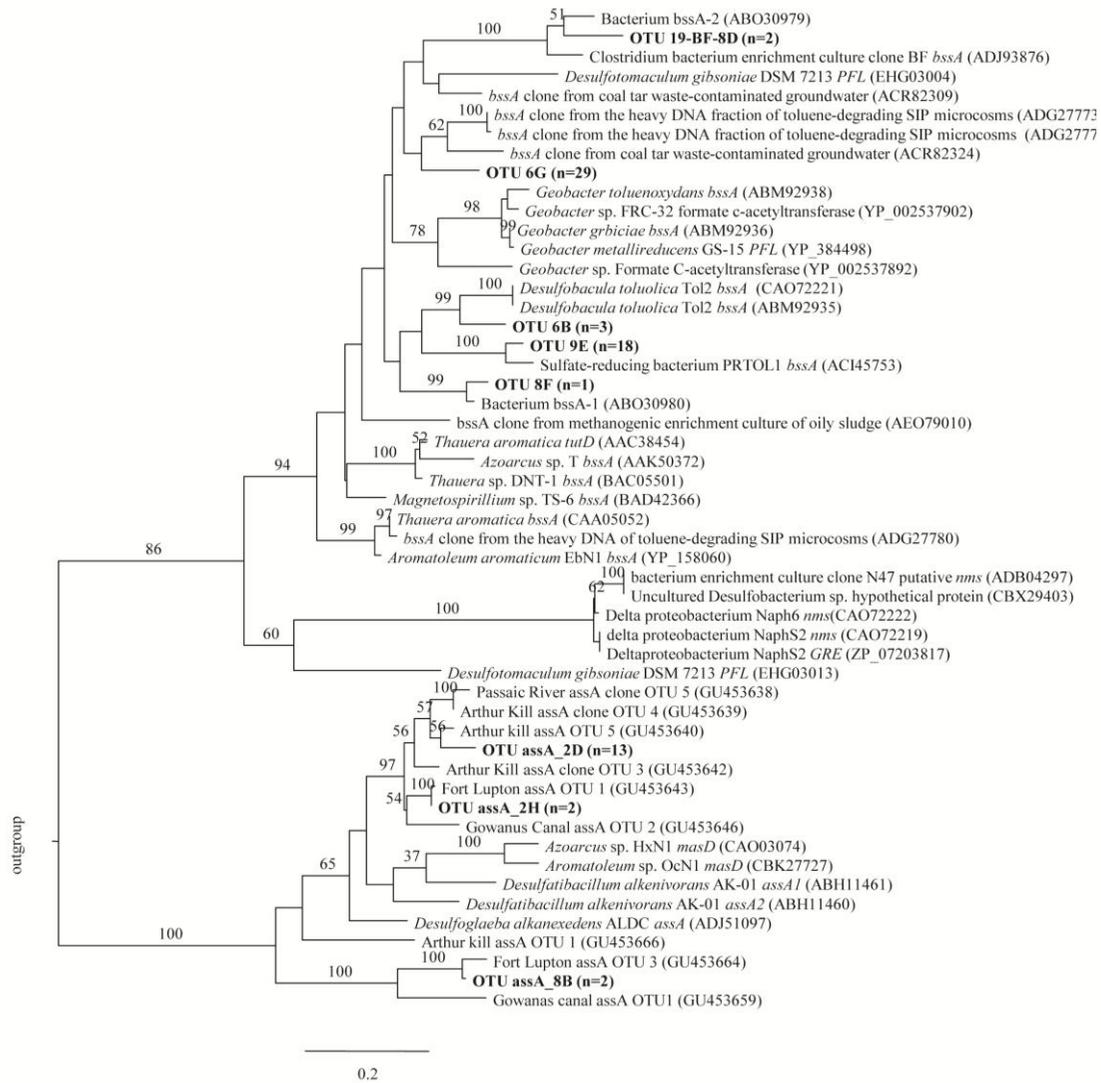
Treatments	Toluene	Benzene	Ethylbenzene	<i>m-</i> or <i>p</i> -xylene	<i>o</i> -xylene	<i>n</i> -C10	<i>n</i> -C8	<i>n</i> -C7	<i>n</i> -C6
Unamended heat- killed control <sup>a</sup>	- <sup>b,c</sup>	-	+	+	-	-	-	+	+
Unamended baseline control	-	-	-	-	-	-	-	+	+
Naphtha (200 µl)	-	±	-	±	-	-	±	±	±
BTEX (50 µl)	-	+	-	±	-	-	-	+	-
<i>n</i> -C14, C16, C18 (100 µl)	-	-	-	-	-	-	-	+	+
<i>n</i> -C6, C7, C8, C10 (100 µl)	-	-	-	-	-	±	±	±	±

<sup>a</sup>Hydrocarbons detected in unamended heat-killed control represent endogenous hydrocarbons

<sup>b</sup> -, undetected; +,detected; +/-,reduced over time but still present

<sup>c</sup> , qualitative measurements were based on the concentration of internal standards :methylcyclohexane and 1,1,3-trimethylcyclohexane, which remained unchanged during incubation period





**Figure F2** Phylogenetic tree of translated *bssA* sequences amplified using the primer set of Winderl *et al.*, (2007) and translated *assA* amplified using primer set 9 (Callaghan *et al.*, 2010). Putative amino acid sequences were grouped at 0.02 difference using MOTHUR, and a representative OTU from each clone library was used in tree construction using maximum likelihood with 100 bootstrap replicates. Bootstrap values >50 are shown at each branch. Pyruvate formate lyase from *Desulfotomaculum reducens* M1-1(PFLYP001114293) was used as the outgroup.

## Reference

Callaghan AV, Davidova IA, Savage-Ashlock K, Parisi VA, Gieg LM, Suflita JM *et al.*,. (2010). Diversity of benzyl- and alkylsuccinate synthase genes in hydrocarbon-impacted environments and enrichment cultures. *Environ Sci Technol* **44**: 7287-7294.

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