

# Supporting Information

## **Metagenomics of Hydrocarbon Resource Environments Indicates Aerobic Taxa and Genes to be Unexpectedly Common**

Dongshan An, Sean M. Caffrey, Jung Soh, Akhil Agrawal, Damon Brown, Karen Budwill, Xiaoli Dong, Peter F. Dunfield, Julia Foght, Lisa M. Gieg, Steven J. Hallam, Niels W. Hanson, Zhiguo He, Thomas R. Jack, Jonathan Klassen, Kishori M. Konwar, Eugene Kuatsjah, Carmen Li, Steve R. Larter, Verlyn Leopatra, Camilla L. Nesbø, Thomas Oldenburg, Antoine P. Pagé, Esther Ramos-Padron, Fauziah F. Rochman, Alireeza Saidi-Mehrabad, Christoph W. Sensen, Payal Sipahimalani, Young C. Song, Sandra Wilson, Gregor Wolbring, Man-Ling Wong & Gerrit Voordouw.

correspondence to: [voordouw@ucalgary.ca](mailto:voordouw@ucalgary.ca)

### **This document file includes:**

Supplementary Text  
Figures. S1 to S4  
Tables S1 to S7  
20 pages

## Supplementary Text

### *Detailed description of public opinion survey*

**Public perceptions of hydrocarbon metagenomics.** A survey of public perceptions of energy production and metagenomics was conducted as part of the Hydrocarbon Metagenomics Project. The 215 survey respondents indicated that current hydrocarbon energy production technologies are viewed as environmentally damaging and that reducing this impact should receive high priority. Few participants responded that technologies currently used by oil and gas industries are acceptable (oil n=8, gas n=10). Over half of participants answered that reducing the environmental impact of oil (63.4%) and gas (65.7%) would positively impact their view of the fossil fuel industry. The groups that were the most negative towards greening efforts consisted of the 11.6% of the respondents, which felt that all oil sand exploration should cease, and the 8.4% of respondents, which felt all coalbed methane operations should cease, because these are non-renewable energy sources for which ecologically neutral extraction cannot be achieved. For this latter group, reduced environmental impact is not enough: only zero impact is acceptable. When given four choices of approaches to greening of energy production technologies (i.e., application of biotechnology), the majority of participants self-identified as having insufficient knowledge to identify particular technologies that could make the industry more environmentally friendly. Respondents who felt greening is needed and that it would lead them to think more positively about oil and gas operations could, therefore, still reject a given biotechnology approach after it has been shown to be feasible and is being implemented.

**Methodology.** An exploratory non-probability survey was designed, using the SurveyMonkey platform and transmitted online via Snowball using convenience sampling.<sup>42</sup> The survey consisted of 43 questions of the following types: simple yes or no, Likert scale, and opinion rating scale. The questions were developed to understand how groups and people think about (a) energy generation, (b) water security, (c) climate change, as well as (d) endeavors to make the extraction of oil and gas from oil sands and coal beds more environmental friendly by (i) minimizing the environmental impact of oil sands production, (ii) decreasing the use of water and emission of greenhouse gases and hydrogen sulphide, (iii) enhancing the production of clean burning natural gas from coal beds. The survey was open from Oct 2010-Feb 2011. The survey was distributed to listserves such as Eanth-I, a mailing list dedicated to the scholarly discussion of anything pertaining to the field of ecological/environmental anthropology, to listserves that cover NGOs working on the topics of the survey, to students and academics at the University of Calgary and to key people in industry and elsewhere who are seen as experts and who distributed the survey further through their networks. Following ethics approval of the survey, 215 participants replied to the survey with an answer database being automatically generated by Survey Monkey. Data were exported as CSV and PDF files for subsequent analysis. Frequency distribution analysis of answers, as well as cross tabulation analysis of the results were performed and related to demographics and to how certain questions were answered. As this was a non-probability sample no tests of significance were performed.<sup>43</sup>

**Results.** In order to investigate the potential influence of greening technologies on the perception of fossil fuel energy production/extraction, an exploratory non-probability survey was designed and transmitted online via snowball and convenience sampling. The majority of respondents were Canadians between the ages of 18-65, of which 31.4% were male and 68.6% female. Although a range of occupational demographics was represented, the majority of participants self-identified as being academically affiliated (students n=99, teachers/researchers n=44) with individuals also identifying from the following occupations: technical (n=24), non-governmental/civil society organization (NGO/CSO) (n=24), international organization (n=12), government (n=11), and industry (n=11). Participants identified other occupations in addition to those listed but these represented less than 5% of respondents. Two hundred and fifteen participants (n=215) responded to the survey; however, response rates to specific questions varied. Therefore, specific n values are reported for each result. Participants were asked a series of

questions related to their opinion on global and local energy issues. Choosing multiple responses or refusing to respond to questions was allowed.

In questions pertaining to possible changes in participant perceptions of energy extraction, over half of participants answered that reducing the environmental impact of oil (63.4%) and gas (65.7%) would positively impact their view of the fossil fuel industry (n=134). Of participants who identified reducing environmental impact as being necessary for oil extraction (n=60), 65.0% agreed that greening oil extraction would positively influence their perception of the industry (Table S1). Similarly, 80.8% of participants who identified the reduction of environmental impact as being necessary for natural gas extraction (n=52) agreed that making natural gas extraction and production more environmentally friendly would positively influence their perception of the gas industry (Table S1). Of the ones who stated that all oil sands recovery operations should cease (n=18) 64.7% would not change their views on the oil sand operation if this was more environmentally-friendly, whereas 23.5% would become more positive about this resource. For coal bed methane operations, 70.0% would not change their negative view while 23.5% would become more positive toward coal bed methane operations (Table S1). Of the ones who stated that coal bed methane operations should cease (n=18), 92.3% would not change their negative views on the coal bed methane operations and 15.4% would become more positive to oil sand extraction. Greening technologies were also identified, by participants who did not identify current energy production and extraction as environmentally damaging, as positively influencing their perspective of the oil and gas industry. Of participants who responded that the oil and gas industry was essential but had problems in general (oil n=37, gas n=19), 83.8% for oil and 73.7% for gas stated that they would view the production and extraction of these two fuel types more positively should greening of the industry take place (Table S1). Furthermore, participants who responded that oil and gas industries are acceptable as they are today (oil n=8, gas n=10), 75.0% (oil) and 60.0% (gas) agreed that greening technologies would also positively influence their view of these industries (table S1). Interestingly, when prompted to comment on the usefulness and applicability of four different green interventions to energy extraction, the majority of participants chose “Don’t know” as a response for every intervention; for stimulation via microbial communities 61.0% chose “Don’t know” with n=134, genetic alterations of metabolism of enzymes in bacteria 60% chose “Don’t know” with n=132, utilizing enzymes and other biological end products in target environments 53.8% chose “Don’t know” with n=130, and for creating unique biosynthetic organisms 51.5% chose “Don’t Know” with n=131. Furthermore, for the ones who stated that oil sand operations should cease, 46.7% rejected all four greening proposals whereas 40% had no opinion/did not know. There was no difference between the four approaches offered. For the ones who stated that coal bed methane operations should cease 60.0% rejected all four greening proposals. The highest rating was obtained for the first approach with a weighted mean of 2.11. The no opinion/did not know category was at 10%.

**Single-end Shotgun DNA Library Construction and 454 Sequencing.** Single-end shotgun DNA libraries were prepared from 0.5 to 1.0 µg of purified DNA and sequenced with the 454 Life Sciences GS-FLX genome sequencer using titanium chemistry and standard library construction procedures (Roche Applied Science, Laval, Quebec, Canada).

**Paired-end Shotgun DNA library Construction and Illumina Sequencing.** DNA libraries were prepared using the TruSeq DNA Sample Prep Kit v1 (Illumina, San Diego, CA) as per the manufacturer’s instructions starting with 2 µg of purified DNA. DNA fragmentation was performed using a Covaris E210 instrument (duty cycle, 5%; intensity, 5; burst per second, 200; duration, 100 s; temperature, 5.5-6°C). Library fragment size selection in the range of 500-600 bp was performed using a Pippin Prep (Sage Science, Beverly, MA) on a 1.5% gel cassette. The final libraries were verified on a Bioanalyzer DNA 1000 chip and quantified both by picogreen (Invitrogen) and qPCR (Kapa Biosystems). The libraries were loaded onto the flow cell, one per lane using a cBot (Illumina). The sequencing, paired end 150 bp reads, was performed on a HiSeq 2000 instrument (Illumina) as per the manufacturer’s instructions.

**Metagenome De Novo Assembly and Annotation.** The 454 reads of each sample were passed through a quality control pipeline, developed in-house to remove low-quality reads and minimize sequencing errors. Eliminated sequences included those that (i) had ambiguous bases, (ii) had an average

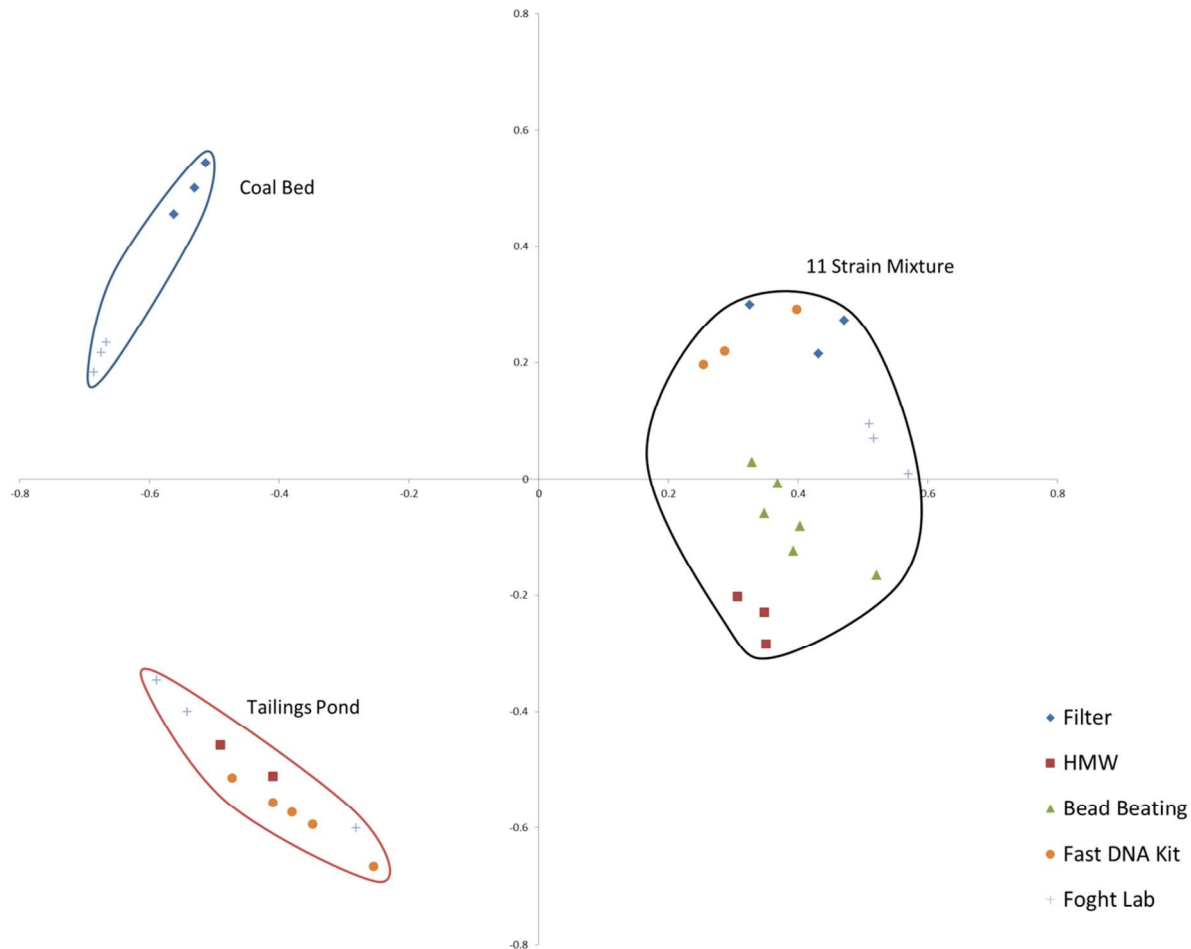
quality score below 25, (iii) contained homopolymer lengths greater than 6, (iv) were shorter than 100 bp, and (v) had artificial duplicates generated during sequencing as identified using UCLUST. Each 454 sequenced sample was assembled independently with Newbler (V2.6, Roche) using options “-mi 95 -ml 60 -a 200 -l 900” based on the quality controlled reads.

The assembled contigs and singletons longer than 200 bp of each 454 sequenced sample were submitted to the Integrated Microbial Genomes & Metagenomics (IMG/M) system<sup>23</sup> for annotation.

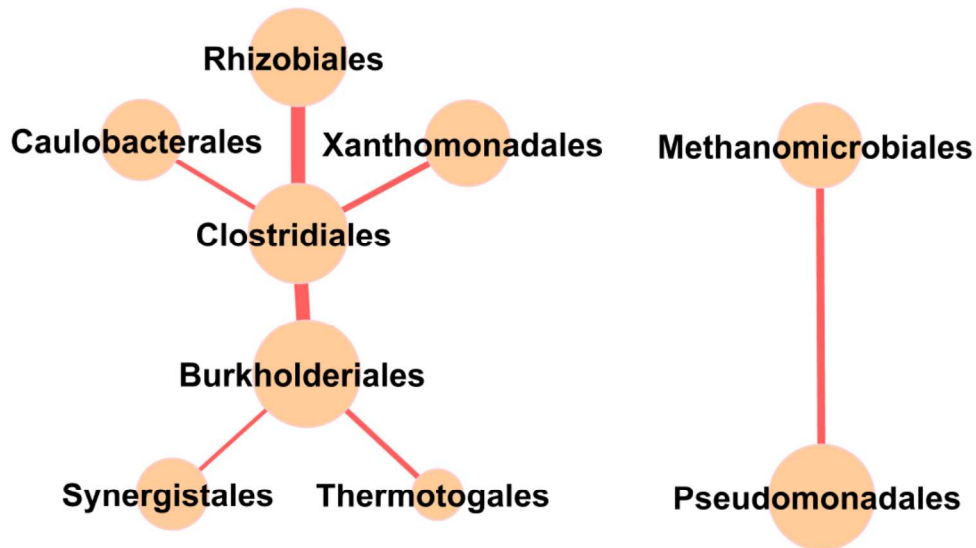
The raw Illumina reads of each sample were passed through an Illumina read quality control program developed in-house that filters out known Illumina sequencing and library preparation artifacts<sup>20</sup>, clipping off partial adaptors and low quality ends. Specifically, the following were removed: (i) reads mapping to spike-in PhiX sequence, (ii) reads shorter than 50 bases after clipping off partial primer-adaptor sequences and/or low quality bases at the 3'-end, and (iii) reads with low complexity. Duplicated read pairs produced during the PCR amplification step of library preparation were consolidated into a single consensus read pair as recommended by Hess *et al.*<sup>20</sup> Each sample's reads that passed the quality control stage were assembled independently using SOAPdenovo (V1.05, BGI, Shenzhen, China). As no optimal k-mer length exists for any de novo metagenome assembly, the multiple k-mer method was used to obtain the final assembly<sup>21</sup>. We performed eight assemblies for the library using different values of k-mer size (59, 63, 67, 71, 75, 79, 83, 87). The different assemblies were pair-wise merged using Newbler and AMOS minimus2<sup>22</sup> for the final assembly.

The assembled Illumina contigs longer than 200 bp from each sample were submitted to IMG/M for functional profiling. The latter focused on the presence of genes for O<sub>2</sub>-independent (anaerobic) and O<sub>2</sub>-dependent (aerobic) degradation of aromatic hydrocarbons<sup>44</sup>, as well as of genes for methanogenesis and methane oxidation (Table S6). The relative abundance of the functional genes (enzyme categories and Pfams) was obtained from the metagenome annotation and gene counts provided by the IMG/MER server and compared after normalization.

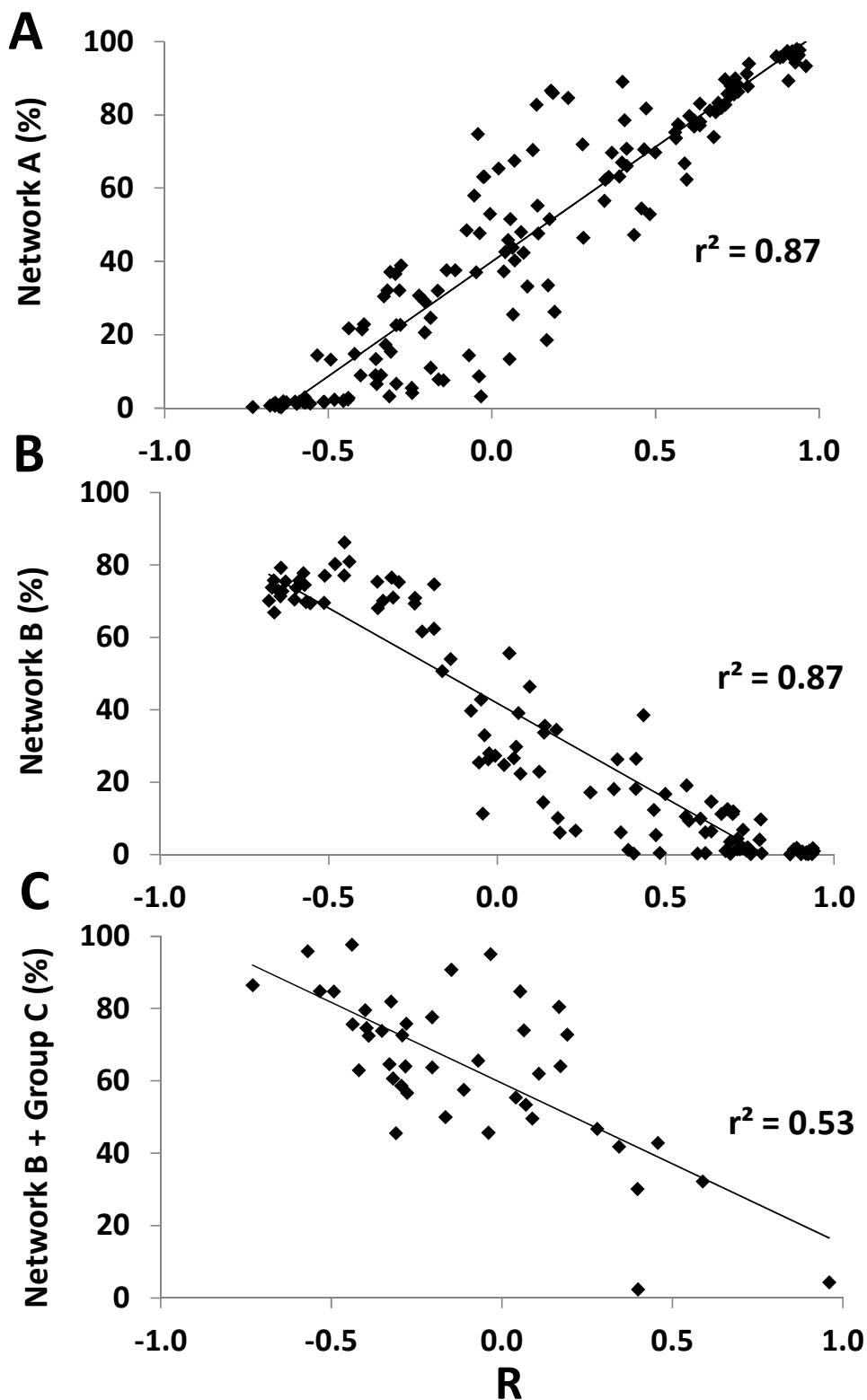
42. Yu, J.; Cooper, H. A. A quantitative review of research design effects on response rates to questionnaires. *J. Market. Res.* **1983**, *20*, 36-44.
43. Advisory Panel on Online Public Opinion Survey Quality Standards and Guidelines for Sampling. *Public Works and Government Services Canada*; <http://www.tpsgc-pwgsc.gc.ca/rop-por/rappports-reports/comiteenligne-panelonline/page-03-eng.html> (2011).
44. Fuchs, G., Boll, M., and Heider, J. Microbial degradation of aromatic compounds - from one strategy to four. *Nature Rev. Microbiol.* **2011**, *9*, 803-816.
45. Parks, D.H.; Beiko, R.G. Identifying biologically relevant differences between metagenomic communities. *Bioinformatics* **2010**, *26*, 715-721.
46. Schloss, P. Evaluating different approaches that test whether microbial communities have the same structure. *ISME Journal* **2008**, *2*, 265-275.



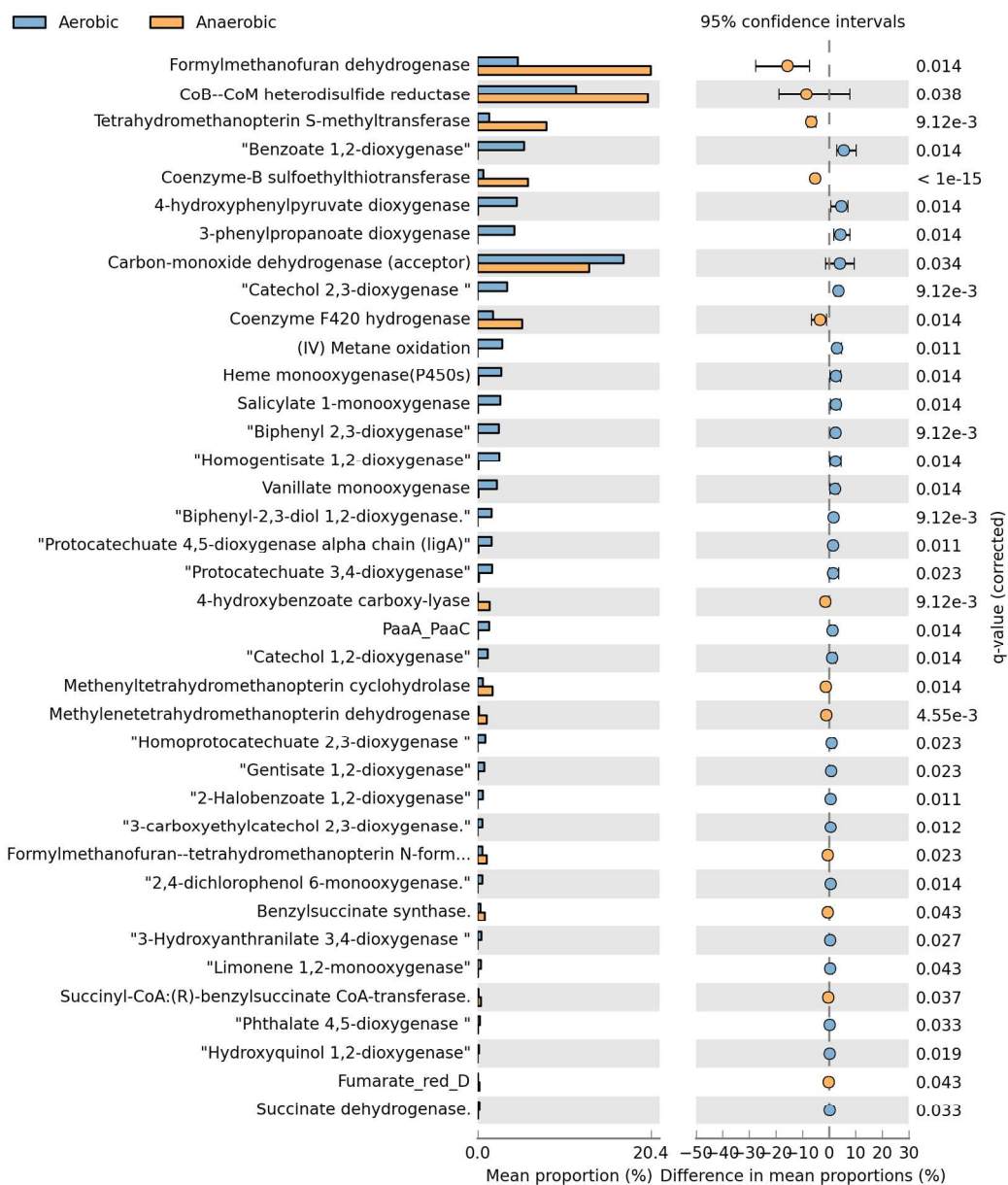
**Figure S1.** Non-metric multidimensional scaling (NMDS) analysis of pyrosequencing reads, illustrating the effect of different DNA extraction and PCR amplification protocols on analysis of replicate samples from CBM water, tailings ponds, and a constructed mixture of 11 genera. To determine reproducibility of the DNA extraction procedures used, three different labs processed an MLSB tailings pond sample, a CBM water sample and a mixture of 11 different prokaryotes, including *Escherichia coli* ( $1 \times 10^{10}$  cells/mL), *Bacillus subtilis* ( $1 \times 10^{10}$  cells/mL), *Streptomyces coelicolor* ( $7.5 \times 10^9$  cells/mL), *Porphyromonas gingivalis* ( $5 \times 10^9$  cells/mL), *Desulfovibrio vulgaris* ( $2.5 \times 10^9$  cells/mL), *Variovorax sp.* ( $1 \times 10^9$  cells/mL), *Campylobacter jejuni* ( $1 \times 10^9$  cells/mL), *Rhodococcus erythropolis* ( $7.5 \times 10^8$  cells/mL), *Thermotoga maritima* ( $5 \times 10^8$  cells/mL), *Paracoccus sp.* ( $2.5 \times 10^8$  cells/mL) and *Methanosarcina barkeri* ( $1 \times 10^8$  cells/mL). Once DNA was extracted from these samples in duplicate by each laboratory using their extraction procedures and amplified by PCR (representing five different protocols), the DNA was pyrosequenced. NMDS analysis of the pyrosequencing reads indicated that any lab-specific variations in DNA extraction and PCR amplification were insignificant compared to inter-sample differences. DNA extraction methods 1 (Filter), 2 (HMW), 3 (Bead Beating), 4 (Fast DNA Kit) and 5 (Foght Lab) were as described in Materials and Methods.



**Figure S2.** Negative co-occurrence analysis of 41 taxonomic orders representing  $\geq 0.1\%$  of total pyrosequencing reads. Negative correlation values among these orders were calculated using a threshold of -0.5 on the Spearman coefficient using the OTU association function of Mothur. The corresponding networks were visualized with Cytoscape.



**Figure S3.** Plots of R-score vs. fraction of sequences belonging to: (A) Network A for all 160 samples, (B) Network B for 116 samples from oil sands, tailings ponds and an oil field, (C) Network B plus Group C for 44 samples from CBM fields.



**Figure S4.** Statistical testing of significance of differences in gene counts for indicated genes or groups of genes for metagenomes PDSYNTPW5, CO182 and CO183 (PDCOCO, blue) as compared to metagenomes CG8, 10PW and CG13 (CGPWCG, orange), using White's non-parametric t-test and Storey false discovery rate for multiple test correction. Genes involved in aerobic hydrocarbon degradation and methane oxidation were significantly more abundant in metagenomes PDCOCO (harbouring predominantly aerobic communities), than in metagenomes CGPWCG (harbouring predominantly anaerobic communities), as indicated by P-values. Likewise genes for anaerobic hydrocarbon metabolism and methanogenesis were significantly more abundant in metagenomes CGPWCG than in metagenomes PDCOCO. The figure was created using STAMP.<sup>45</sup>



<b>Table S1. Shifts in Perspectives of Fossil Fuel Industries Based on Greening Technologies</b>				
	<b>Influence on Perspective of Oil &amp; Gas Production/Extraction</b>			
<b>Current Opinion of Production/Extraction</b>	View more positively		No effect	
	%	Response count	%	Response count
<b>Oil</b>				
Essential to do but has problems (n=37)	83.8	31	10.8	4
Environmental impact needs great improvement (n=60)	65.0	39	26.7	16
Should not be done (n=16)	6.3	1	75.0	12
Acceptable as extracted today (n=8)	75.0	6	25.0	2
<b>Natural Gas</b>				
Essential to do but has problems (n=19)	73.7	14	5.3	1
Environmental impact needs great improvement (n=52)	80.8	42	13.5	7
Should not be done (n=10)	20.0	2	80	8
Acceptable as extracted today (n=10)	60.0	6	30.0	3

**Table S2.** Additional information for the 160 samples analyzed in this study.

Group	ID	Sample name	Depth (m)	Sampling time (mm/year)	SRA Accession number (SRR)	Number of QC reads	Number of taxa	Network A (%)	Network B (%)	Group C (%)	R-score of sample
1_Oil_sands	001	2010Suncor Oil Sands Core Run 10 Subsample 1	295.00	05/2010	573806	6008	76	1	70	27	-0.68
	002	2010Suncor Oil Sands Core Run 10 Subsample 2	295.05	05/2010	573807	4849	91	2	76	21	-0.59
	003	2010Suncor Oil Sands Core Run 10 Subsample 3	295.10	05/2010	573808	4466	93	1	69	26	-0.55
	004	2010Suncor Oil Sands Core Run 10 Subsample 5	295.15	05/2010	573809	5450	99	2	70	26	-0.60
	005	2010Suncor Oil Sands Core Run 10 Subsample 6	295.20	05/2010	573810	3536	151	11	75	10	-0.19
	006	2010Suncor Oil Sands Core Run 10 Subsample 8	295.25	05/2010	573811	5022	94	2	80	16	-0.48
	007	2010Suncor Oil Sands Core Run 10 Subsample 11	295.30	05/2010	573812	3283	94	2	73	23	-0.64
	008	2010Suncor Oil Sands Core Run 10 Subsample 12	295.35	05/2010	573813	1806	80	2	75	21	-0.63
	009	2010Suncor Oil Sands Core Run 10 Subsample 13	295.40	05/2010	573814	4522	82	0	73	25	-0.65
	010	2010Suncor Oil Sands Core Run 11 Subsample 1	298.75	05/2010	573815	5109	94	1	79	18	-0.64
	011	2010Suncor Oil Sands Core Run 11 Subsample 2	298.80	05/2010	573816	3491	66	52	34	13	0.18
	012	2010Suncor Oil Sands Core Run 11 Subsample 3	298.85	05/2010	573817	3913	179	25	62	11	-0.19
	013	2010Suncor Oil Sands Core Run 11 Subsample 4	298.90	05/2010	573818	5221	91	44	39	16	0.06
	014	2010Suncor Oil Sands Core Run 11 Subsample 5	298.95	05/2010	573819	5002	77	37	43	18	-0.05
	015	2010Suncor Oil Sands Core Run 11 Subsample 6	299.00	05/2010	573820	5672	177	81	13	5	0.68
	016	2010Suncor Oil Sands Core Run 11 Subsample 7	299.05	05/2010	573821	3871	101	0	71	26	-0.64
	017	2010Suncor Oil Sands Core Run 11 Subsample 8	299.10	05/2010	573822	4616	137	82	12	4	0.70
	018	2010Suncor Oil Sands Core Run 11 Subsample 9	299.15	05/2010	573823	4995	141	2	70	23	-0.57
	019	2010Suncor Oil Sands Core Run 11 Subsample 10	299.20	05/2010	573824	2566	116	82	11	5	0.70
	020	2010Suncor Oil Sands Core Run 11 Subsample 11	299.25	05/2010	573825	3141	137	77	15	6	0.64
	021	2010Suncor Oil Sands Core Run 11 Subsample 13	299.30	05/2010	573826	4371	114	1	74	22	-0.60
	022	2010Suncor oil sands core Run 3 subsample 1	240.45	05/2010	617109	5876	102	1	76	21	-0.66
	023	2010Suncor oil sands core Run 3 subsample 2	240.55	05/2010	617110	3867	128	2	77	19	-0.51
	024	2010Suncor oil sands core Run 3 subsample 3	240.65	05/2010	617111	3851	112	13	68	17	-0.35
	025	2010Suncor oil sands core Run 3 subsample 4	240.75	05/2010	617113	3803	94	1	74	24	-0.67
	026	2010Suncor oil sands core Run 3 subsample 5	240.85	05/2010	617114	4983	135	2	77	12	-0.45
	027	2010Suncor oil sands core Run 3 subsample 6	240.95	05/2010	617115	5508	127	2	70	23	-0.51
	028	2010Suncor oil sands core Run 3 subsample 7	241.05	05/2010	617116	5383	122	2	78	18	-0.58
	029	2010Suncor oil sands core Run 3 subsample 8	241.15	05/2010	617117	4662	118	3	74	21	-0.57
	030	2010Suncor oil sands core Run 3 subsample 9	241.25	05/2010	617118	5315	101	1	67	29	-0.66
2_TPS	031	2009Suncor Tailings Pond 5 (6.5 ft)	1.98	07/2009	572718	4642	150	72	17	1	0.28
	032	2009Suncor Tailings Pond 5 (8.0 ft)	2.44	07/2009	573715	4173	118	86	6	1	0.19
	033	2009Suncor Tailings Pond 5 (15 ft)	4.57	07/2009	573716	4526	125	55	34	2	0.14
	034	2009Suncor Tailings Pond 5 (20 ft)	6.10	07/2009	573717	4032	143	71	18	4	0.43
	035	2009Suncor Tailings Pond 5 (25 ft)	7.62	07/2009	573718	4349	157	75	11	5	0.41
	036	2009Suncor Tailings Pond 5 (30 ft)	9.14	07/2009	573725	4418	143	49	40	2	-0.04
	037	2009Suncor Tailings Pond 5 (35 ft)	10.67	07/2009	573727	4454	86	42	46	2	-0.08
	038	2009Suncor Tailings Pond 5 (40 ft)	12.19	07/2009	573736	4195	130	63	26	2	0.10
	039	2009Suncor Tailings Pond 5 (45 ft)	13.72	07/2009	573737	4547	98	63	28	1	-0.03
	040	2009Suncor Tailings Pond 5 (50 ft)	15.24	07/2009	573738	5391	130	65	25	1	-0.02
	041	2009Suncor Tailings Pond 5 (65 ft)	19.81	07/2009	573739	3801	131	47	38	4	0.02
	042	2009Suncor Tailings Pond 5 (82 ft)	24.99	07/2009	573740	1471	96	58	25	6	-0.05
	043	2009Suncor Tailings Pond 5 (95 ft)	28.96	07/2009	573741	5217	157	52	30	8	0.06
3_TP6	044	2008Suncor Tailings Pond 6 (30 ft)-metagenome	9.14	10/2008	090658	3488	180	48	33	3	-0.04
	045	2010Suncor Tailings Pond 6 (pooled)-metagenome	3-22.00	07/2010	573768	15733	215	53	27	12	-0.01
	046	2010Suncor Tailings Pond 6 (3 m)	3.00	07/2010	573756	7512	173	91	4	1	0.78
	047	2010Suncor Tailings Pond 6 (6 m)	6.00	07/2010	573757	5335	149	78	15	4	0.64
	048	2010Suncor Tailings Pond 6 (9 m)	9.00	07/2010	573758	6219	156	74	19	4	0.56
	049	2010Suncor Tailings Pond 6 (12 m)	12.00	07/2010	573759	5969	150	63	26	7	0.36
	050	2010Suncor Tailings Pond 6 (15 m)	15.00	07/2010	573760	6811	155	79	6	11	0.62
	051	2010Suncor Tailings Pond 6 (18 m)	18.00	07/2010	573761	5801	165	70	17	9	0.50
	052	2010Suncor Tailings Pond 6 (21 m)	21.00	07/2010	573762	7147	171	75	11	11	0.56

	053	2010Suncor Tailings Pond 6 (22 m)	22.00	07/2010	573763	7009	159	83	4	10		0.69	
	054	2011Suncor Tailings Pond 6 (3.5 m)	3.50	07/2011	573769	6588	214	83	6	2		0.64	
	055	2011Suncor Tailings Pond 6 (4 m)	4.00	07/2011	573770	7346	260	77	9	2		0.57	
	056	2011Suncor Tailings Pond 6 (7 m)	7.00	07/2011	573772	4923	172	80	10	4		0.60	
	057	2011Suncor Tailings Pond 6 (10 m)	10.00	07/2011	573773	3729	143	71	12	9		0.47	
	058	2011Suncor Tailings Pond 6 (13 m)	13.00	07/2011	573775	5149	181	62	18	10		0.35	
	059	2011Suncor Tailings Pond 6 (16 m)	16.00	07/2011	573776	5746	174	46	27	19		0.05	
	060	2011Suncor Tailings Pond 6 (18.5 m)	18.50	07/2011	573778	6744	279	48	36	6		0.14	
4_TP_MLSB	061	2011 Syncrude Tailings Pond MLSB-metagenome	NA	05/2011	629337	4441	160	67	22	4		0.07	
	062	2010Syncrude Tailings Pond MLSB M MFT- 11m	11.00	06/2010	573791	6810	91	98	1	0		0.93	
	063	2010Syncrude Tailings Pond MLSB M MFT - 1m	1.00	06/2010	573789	6759	192	88	7	1		0.73	
	064	2010Syncrude Tailings Pond MLSB M MFT - 6m	6.00	06/2010	573790	7064	102	97	1	0		0.92	
	065	2010Syncrude Tailings Pond MLSB N MFT - 16.1m	16.10	06/2010	573783	6103	221	66	26	1		0.41	
	066	2010Syncrude Tailings Pond MLSB N MFT - 1.1m	1.10	06/2010	573780	7036	206	81	11	1		0.67	
	067	2010Syncrude Tailings Pond MLSB N MFT -35.8m	35.80	06/2010	573784	6184	192	31	62	2		-0.22	
	068	2010Syncrude Tailings Pond MLSB N MFT - 6.1m	6.10	06/2010	573782	6961	217	37	56	1		0.04	
	069	2010Syncrude Tailings Pond MLSB S MFT - 12.3m	12.30	06/2010	573788	7596	101	88	10	0		0.78	
	070	2010Syncrude Tailings Pond MLSB S MFT - 1.1m	1.10	06/2010	573786	6429	117	96	2	0		0.89	
5_TP_surface surface	071	2010Syncrude Tailings Pond MLSB S MFT - 6.1m	6.10	06/2010	573787	5546	90	97	1	0		0.93	
	072	2011Center of pond WIP-metagenome	0-0.10	07/2011	516417	10991	173	9	75	8		-0.36	
	073	2010Suncor Tailings Pond 5 (0-10cm water)	0-0.10	05/2010	573797	3306	122	2	86	8		-0.45	
	074	2010Suncor Tailings Pond 6 (0-10cm water)	0-0.10	05/2010	573798	2753	119	3	81	9		-0.44	
	075	2009Suncor Tailings Pond 2/3 (0-10cm)	0-0.10	09/2009	573796	18842	152	87	10	1		0.18	
	076	2009Suncor Tailings Pond 1A, outflow of Pond 2/3 (0-10cm)	0-0.10	09/2009	573795	11975	116	83	14	1		0.14	
	077	2009Suncor Tailings Pond 1A (0-10cm)	0-0.10	09/2009	573792	9523	132	85	7	3		0.23	
	078	2009Suncor Tailings Pond 1A, vegetated marsh connected to pond (0-10cm)	0-0.10	09/2009	573793	8986	218	3	77	3		-0.31	
	079	2009Suncor Tailings Pond 1A (0-10cm)	0-0.10	09/2009	573794	12352	141	70	23	3		0.12	
	080	2010Syncrude Tailing Pond ASB (0-10cm)	0-0.10	07/2010	573799	3878	131	4	71	9		-0.24	
	081	2010Syncrude tailing pond MLSB (0-10cm)	0-0.10	07/2010	573800	4748	152	15	71	6		-0.31	
	082	2010Syncrude Tailing Pond WIP (0-10cm)	0-0.10	07/2010	573801	3690	158	5	69	12		-0.24	
	083	2010Syncrude Tailing Pond MLSB (0-10cm)	0-0.10	10/2010	573802	10352	218	9	70	15		-0.34	
	084	2010Syncrude Tailing Pond WIP (0-10cm)	0-0.10	10/2010	573803	6829	233	8	51	20		-0.16	
	085	2010Syncrude Tailing Pond MLSB (0-10cm)	0-0.10	11/2010	573804	7317	139	38	54	7		-0.14	
	086	2010Syncrude Tailing Pond WIP (0-10cm)	0-0.10	11/2010	573805	9919	162	7	75	9		-0.29	
	6_Oil_field	087	2012Baker Petrolite's MHGC PW10 (April)-metagenome	850.00	04/2012	631208	2058	97	90	2	6		0.74
		088	2011Baker Petrolite's MHGC PW18 (Jan)	850.00	01/2011	573828	9148	98	89	1	1		0.91
089		2011Baker Petrolite's MHGC PW3 (Jan)	850.00	01/2011	573829	12928	93	96	1	1		0.93	
090		2011Baker Petrolite's MHGC PW4 (Jan)	850.00	01/2011	573830	4038	61	94	0	0		0.93	
091		2011Baker Petrolite's MHGC PW3 (Jul)	850.00	07/2011	573832	12810	99	77	0	2		0.62	
092		2011Baker Petrolite's MHGC PW4 (Jul)	850.00	07/2011	573833	10278	105	53	0	0		0.48	
093		2011Baker Petrolite's MHGC PW10 (Jul)	850.00	07/2011	573835	13261	98	74	1	1		0.68	
094		2011Baker Petrolite's MHGC PW15 (Jul)	850.00	07/2011	573836	13848	151	86	1	0		0.74	
095		2011Baker Petrolite's MHGC PW3 (Aug)	850.00	08/2011	573839	10537	97	62	0	0		0.59	
096		2011Baker Petrolite's MHGC PW4 (Aug)	850.00	08/2011	573840	10515	145	70	6	9		0.37	
097		2011Baker Petrolite's MHGC PW13 (Aug)	850.00	08/2011	573843	10579	101	82	0	1		0.69	
098		2011Baker Petrolite's MHGC PW3 (Sep)	850.00	09/2011	573845	3094	55	88	0	6		0.75	
099		2011Baker Petrolite's MHGC PW12 (Sep)	850.00	09/2011	573846	8611	102	86	1	5		0.75	
100		2011Baker Petrolite's MHGC PW13 (Sep)	850.00	09/2011	573847	8490	66	97	0	2		0.92	
101		2011Baker Petrolite's MHGC PW15 (Sep)	850.00	09/2011	573848	6264	111	86	1	4		0.72	
102		2011Baker Petrolite's MHGC PW18 (Sep)	850.00	09/2011	573849	6494	115	63	1	32		0.39	
103		2011Baker Petrolite's MHGC PW19 (Sep)	850.00	09/2011	573851	3855	96	83	4	9		0.71	
104		2011Baker Petrolite's MHGC PW3 (Oct)	850.00	10/2011	617126	7366	91	79	0	2		0.41	
105		2011Baker Petrolite's MHGC PW19 (Oct)	850.00	10/2011	617130	7747	137	85	2	2		0.73	
106		2011Baker Petrolite's MHGC PW18 (Oct)	850.00	10/2011	617131	5540	193	82	5	5		0.47	
107	2012Baker Petrolite's MHGC PW18 (April)	850.00	04/2012	631211	5796	93	96	2	0		0.94		
108	2012Baker Petrolite's MHGC PW25 (April)	850.00	04/2012	631212	8721	113	90	1	1		0.71		
109	2012Baker Petrolite's MHGC PW3(March)	850.00	03/2012	631807	8558	79	97	0	0		0.90		
110	2012Baker Petrolite's MHGC PW5(March)	850.00	03/2012	631809	3854	83	96	0	1		0.92		
111	2012Baker Petrolite's MHGC PW9(March)	850.00	03/2012	631810	7711	45	97	0	0		0.92		
112	2012Baker Petrolite's MHGC PW16(March)	850.00	03/2012	631815	3712	86	96	1	1		0.88		
113	2012Baker Petrolite's MHGC PW18(March)	850.00	03/2012	631816	6015	92	98	1	0		0.94		
114	2012Baker Petrolite's MHGC PW25(March)	850.00	03/2012	631822	7377	59	96	0	0		0.93		
115	2012Baker Petrolite's MHGC PW9(May)	850.00	05/2012	631896	5860	58	96	0	0		0.87		
116	2012Baker Petrolite's MHGC PW30(April)	850.00	03/2012	631901	6129	79	94	0	0		0.78		
7_CBM_cores	117	2008 Core from well 8-30-28-24W4, 321.96 m	321.96	12/2008	573857	1552	114	14	31	35		-0.07	
	118	2008 Core from CBM well 8-30-28-24W4,	420.00	12/2008	573852	1899	30	0	3	84		-0.73	
	119	2008Core from well 5-29-28-24W4	314.60	12/2008	573861	2654	94	7	7	67		-0.35	
	120	2008Core from well 5-29-28-24W4	411.60	12/2008	573859	1156	61	9	14	66		-0.40	
	121	2008Coal core from Encana, well#5-29	585.00	12/2008	573862	2555	89	37	7	38		-0.31	
	122	2008Coal core from Encana, well#102/7	402.00	12/2008	573860	1682	81	9	22	23		-0.04	
	123	2008Core from CBM well 102/7-29-28-24W4	517.81	12/2008	573855	5066	81	23	3	73		-0.28	
	124	2008Core from well 102/7-29-28-24W4	576.26	12/2008	573858	2034	101	15	9	53		-0.42	
	125	Quicksilver CBM Cuttings, 686m-SHX26-CO183 metagenome	686.00	NA	619007	7134	96	22	18	58		-0.44	

	126	Quicksilver CBM Cuttings, 730m-SHX25-CO182 metagenome	730.00	NA	619006	6724	89	13	21	64	-0.49
	127	2010Coal cuttings, ARC Resources	140.00	10/2010	573863	1969	46	32	33	31	-0.28
	128	2010Coal cuttings from ARC Resources	327.00	10/2010	573864	2303	51	37	21	38	-0.30
	129	2010Coal cuttings, ARC Resources.,	327.00	10/2010	573865	2062	46	39	20	36	-0.28
	130	2010Coal cuttings from ARC Resources	145.00	10/2010	573866	2178	57	23	43	30	-0.39
	131	2010Coal cuttings, ARC Resources.,	145.00	10/2010	573867	1986	52	21	45	30	-0.40
	132	2010Coal cuttings from ARC Resources	261.00	10/2010	573868	2700	65	30	28	37	-0.33
	133	2010Coal cuttings, ARC Resources.	261.00	10/2010	573869	2503	62	32	30	31	-0.32
	134	2009Cuttings, Quicksilver, Ardely formation	324.00	07/2009	573870	1234	89	38	27	30	-0.11
	135	2009Cuttings, Quicksilver, Rockyford formation,	793.00	07/2009	573871	2341	116	29	19	44	-0.20
9_CBM_dcuts	136	2010T1 (1045.5 m cuttings)	1045.5	02/2010	573880	1056	42	40	3	50	0.07
	137	2010TRIDENT EXPLORATION CAN 1 (cuttings)	1045.5	02/2010	573882	1459	39	1	2	94	-0.57
	138	2009Trident coal cuttings	1047.1	06/2009	573877	10053	140	57	2	40	0.34
	139	2009Trident coal cuttings	1044.9	06/2009	573878	8595	137	8	4	87	-0.15
	140	2009Trident coal cuttings	1042.3	06/2009	573879	9702	143	67	2	28	0.40
	141	2009Trident_well_1560m- Trident_1560D metagenome	1560.0	06/2009	573886	4519	116	33	4	58	0.11
	142	2009Trident_well	1610.0	06/2009	573887	5313	128	46	5	42	0.28
	143	2009Nexen coal cuttings	1063.0	04/2009	573874	9421	67	54	0	43	0.46
	144	2009Nexen coal cuttings	1065.7	04/2009	573875	12890	53	19	0	80	0.17
	145	2009Nexen coal cuttings	1064.4	04/2009	573876	12908	56	26	0	73	0.19
10_CBM_water	146	2009Nexen water	1063.0	04/2009	573890	10099	60	33	0	64	0.17
	147	2009Nexen water	1064.4	04/2009	573891	13001	45	26	0	74	0.06
	148	2009Trident water from well#100/03-28-060-05W5	NA	05/2009	573888	10557	129	32	2	48	-0.17
	149	2009Trident water from well#100/03-28-060-05W5	NA	05/2009	573889	12232	103	23	0	72	-0.29
	150	2009Conoco-Phillips produced water - CG13 metagenome	NA	08/2009	573895	8984	72	67	0	32	0.59
	151	2009Conoco-Phillips produced water - CG19 metagenome	NA	08/2009	573898	4424	101	89	2	1	0.40
	152	2009Conoco-Phillips produced water - CG22	NA	08/2009	573899	9139	61	13	7	78	0.05
	153	2009Conoco-Phillips produced water - CG23	NA	08/2009	573900	3145	63	43	3	52	0.04
	154	2009Conoco-Phillips produced water - CG1	NA	08/2009	573892	6589	66	14	69	16	-0.53
	155	2009Conoco-Phillips produced water - CG4	NA	08/2009	573901	3497	39	3	0	95	-0.03
	156	2009Conoco-Phillips produced water - CG7 metagenome	NA	08/2009	573893	8284	81	17	0	82	-0.32
157	2009Conoco-Phillips produced water - CG8 metagenome	NA	08/2009	573894	3498	39	93	0	4	0.96	
158	2009Conoco-Phillips produced water - CG16	NA	08/2009	573896	9351	31	2	65	33	-0.44	
159	2009Conoco-Phillips produced water - CG18	NA	08/2009	573897	7528	76	21	1	77	-0.20	
160	2009Conoco-Phillips produced water - CG10	NA	08/2009	573902	6619	97	48	9	40	0.09	

**Table S3A.** Pairwise comparison of community structures using Mother's implementation of AMOVA.<sup>46</sup> Table shows F<sub>s</sub> values. The p-values calculated from 1000 randomizations, all were less than 0.001. AMOVA test of all samples had an F<sub>s</sub> value of 13.5 and a p-value less than 0.001.

#	1_Oil_sands	2_TP5	3_TP6	4_TP_MLSB	5_TP_surface	6_Oil_field	7_CBM_cores	8_CBM_cuts	9_CBM_dcuts	10_CBM_water
1	-									
2	39.6584	-								
3	30.2411	16.4107	-							
4	29.5108	12.7484	7.85257	-						
5	23.3002	13.7259	17.5163	10.1439	-					
6	59.4742	32.6847	34.1501	18.7096	25.0218	-				
7	17.1099	18.4198	16.6876	14.0087	9.16582	20.4943	-			
8	31.5622	23.5099	26.3363	20.5667	11.8283	30.7785	7.28112	-		
9	23.8181	19.0433	19.8587	14.1503	9.67842	22.7126	6.98942	11.2481	-	
10	18.3739	11.7393	12.5002	8.7104	7.34758	15.7863	4.17919	6.94137	4.04419	-

**Table S3B.** Pairwise comparison of community structures using weighted UniFrac. Table shows WScores. All wSig were less than 0.001, except for CBM\_cuts-CBM\_dcuts (0.007), CBM\_cores-CBM\_water (0.002), CBM\_dcuts-CBM\_water (0.005) and CBM\_cores-TP\_surface (0.011).

#	1_Oil_sands	2_TP5	3_TP6	4_TP_MLSB	5_TP_surface	6_Oil_field	7_CBM_cores	8_CBM_cuts	9_CBM_dcuts	10_CBM_water
1	-									
2	0.994991	-								
3	0.986975	0.975148	-							
4	0.918538	0.902331	0.712606	-						
5	1	1	1	1	-					
6	1	1	1	0.9508	1	-				
7	1	1	1	0.975847	1	1	-			
8	1	1	1	0.975847	1	1	0.799603	-		
9	0.989433	0.989433	0.989433	0.989433	0.995196	0.989433	0.869404	0.950534	-	
10	0.960804	0.960804	0.960804	0.960804	0.981932	0.960804	0.895782	0.870115	0.873427	-



Methylocaldum	Methylococcales		0.07	0.00	0.00	0.00	0.01	<b>0.71</b>	0.00	0.00	0.00	0.00	0.00
Kosmotoga	Thermotogales	A	0.07	0.00	0.00	0.00	0.01	0.00	<b>0.69</b>	0.00	0.00	0.00	0.01
Desulfotomaculum	Clostridiales	A	0.07	0.00	0.02	0.04	0.07	0.01	<b>0.40</b>	0.00	0.00	0.01	0.18
Limnobacter	Burkholderiales	B	0.07	0.00	0.09	0.01	0.01	<b>0.52</b>	0.01	0.00	0.00	0.00	0.00
Coprothermobacter	Thermoanaerobacterales	C	0.07	0.02	0.01	0.04	0.01	0.00	0.00	0.00	0.00	0.00	<b>0.57</b>
Desulfoglaeba	Syntrophobacterales	A	0.06	0.00	0.00	<b>0.61</b>	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Legionella	Legionellales		0.06	0.03	<b>0.40</b>	0.02	0.00	0.10	0.00	0.00	0.00	0.00	0.00
Derxia	Burkholderiales	B	0.05	0.01	<b>0.26</b>	0.02	0.01	0.21	0.00	0.00	0.03	0.00	0.00
Prevotella	Bacteroidales	C	0.05	0.00	0.00	0.00	0.00	0.00	0.00	<b>0.53</b>	0.00	0.01	0.00
Halomonas	Oceanospirillales	C	0.05	0.02	0.00	0.04	0.16	0.00	0.01	0.01	0.01	<b>0.23</b>	0.07
Geoalkalibacter	Desulfuromonadales	A	0.05	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	<b>0.50</b>	0.00
Anaerococcus	Clostridiales	A	0.05	0.00	0.00	0.00	0.00	0.00	0.00	<b>0.49</b>	0.00	0.00	0.00
Vibrio	Vibrionales		0.05	0.01	0.00	0.00	0.01	0.00	0.00	0.00	0.00	<b>0.34</b>	0.10
Dietzia	Actinomycetales	B	0.04	0.01	0.01	0.02	0.03	0.00	0.00	<b>0.16</b>	0.01	0.09	0.11
Desulfurivibrio	Desulfobacterales	A	0.04	0.00	0.01	0.03	<b>0.36</b>	0.00	0.00	0.00	0.00	0.00	0.00
Peptoniphilus	Clostridiales	A	0.04	0.00	0.00	0.00	0.00	0.00	0.00	<b>0.38</b>	0.00	0.00	0.00
Anaerovorax	Clostridiales	A	0.04	0.00	0.00	0.02	<b>0.21</b>	0.11	0.01	0.00	0.00	0.01	0.00
Porphyromonas	Bacteroidales	C	0.03	0.00	0.00	0.00	0.00	0.00	0.00	<b>0.32</b>	0.00	0.00	0.00
Thermus	Thermales		0.03	0.00	0.00	0.10	<b>0.19</b>	0.00	0.00	0.00	0.00	0.00	0.00
Anaerobaculum	Synergistales	A	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	<b>0.24</b>
Alkalibacter	Clostridiales	A	0.02	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	<b>0.17</b>	0.03
Dialister	Clostridiales	A	0.02	0.00	0.00	0.00	0.00	0.00	0.00	<b>0.20</b>	0.00	0.00	0.00
Desulfuromusa	Desulfuromonadales	A	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	<b>0.17</b>	0.02
Methylophaga	Thiotrichales		0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	<b>0.12</b>	0.01

**Table S4B.** Bottom up evaluation of the ability of 109 genera (Table 4A) to grow using aerobic respiration. All species of each genus listed at <http://www.bacterio.cict.fr> were classified as anaerobic, facultative, aerobic or unassigned through inspection of the literature. The R-score was calculated for each genus as:  $R = (N_{an}-N_{ar})/(N_{an}+N_{fac}+N_{ar})$ , where  $N_{an}$ ,  $N_{fac}$  and  $N_{ar}$  are the number of anaerobic, facultative and aerobic species, listed in the table. Individual entries for species are not listed. The genera are grouped by order. Affiliation of orders with Network A or Network B or with group C is indicated. Of these, eight orders are missing from the table, because although present in excess of 0.1% overall, these did not have a genus that was among the most abundant 22 in any of the HREs (Table S6A). Eight genera at the bottom of the table belonged to orders not assigned to Network A, Network B or Group C.

Order	Network/Group	Genus	Number of species assigned in each category					R-score
			Total	$N_{an}$	$N_{fac}$	$N_{ar}$	Not assigned	
Anaerolineales	A	Leptolinea	1	1	0	0	0	1.00
Campylobacterales	A	Arcobacter	15	0	8	6	1	-0.43
Chromatiales	A	Alishewanella	5	0	3	2	0	-0.40
Chromatiales	A	Alkalimonas	3	0	0	3	0	-1.00
Chromatiales	A	Thiocapsa	6	3	3	0	0	0.50
Clostridiales	A	Acetobacterium	8	8	0	0	0	1.00
Clostridiales	A	Alkalibacter	1	1	0	0	0	1.00
Clostridiales	A	Anaerococcus	7	7	0	0	0	1.00
Clostridiales	A	Anaerovirgula	1	1	0	0	0	1.00
Clostridiales	A	Anaerovorax	1	1	0	0	0	1.00
Clostridiales	A	Desulfosporosinus	7	7	0	0	0	1.00
Clostridiales	A	Desulfotomaculum	30	30	0	0	0	1.00
Clostridiales	A	Dialister	5	5	0	0	0	1.00
Clostridiales	A	Finegoldia	1	1	0	0	0	1.00
Clostridiales	A	Fusibacter	2	2	0	0	0	1.00
Clostridiales	A	Peptoniphilus	10	10	0	0	0	1.00
Clostridiales	A	Sporomusa	9	9	0	0	0	1.00
Desulfobacterales	A	Desulfobacterium	9	9	0	0	0	1.00
Desulfobacterales	A	Desulfobulbus	6	6	0	0	0	1.00
Desulfobacterales	A	Desulfocapsa	2	2	0	0	0	1.00
Desulfobacterales	A	Desulfurivibrio	2	2	0	0	0	1.00
Desulfovibrionales	A	Desulfomicrobium	7	7	0	0	0	1.00
Desulfovibrionales	A	Desulfovibrio	69	69	0	0	0	1.00
Desulfuromonadales	A	Desulfuromonas	7	7	0	0	0	1.00
Desulfuromonadales	A	Desulfuromusa	4	4	0	0	0	1.00
Desulfuromonadales	A	Geoalkalibacter	2	2	0	0	0	1.00
Desulfuromonadales	A	Geobacter	16	16	0	0	0	1.00
Erysipelotrichales	A	Erysipelothrix	3	0	3	0	0	0.00
Hydrogenophilales	A	Thiobacillus	22	0	7	15	0	-0.68
Methanobacteriales	A	Methanobacterium	31	31	0	0	0	1.00
Methanobacteriales	A	Methanothermobacter	7	7	0	0	0	1.00
Methanomicrobiales	A	Methanocalculus	4	4	0	0	0	1.00
Methanomicrobiales	A	Methanoculleus	8	8	0	0	0	1.00
Methanomicrobiales	A	Methanofollis	5	5	0	0	0	1.00
Methanomicrobiales	A	Methanolinea	2	2	0	0	0	1.00
Methanomicrobiales	A	Methanoregula	2	2	0	0	0	1.00
Methanomicrobiales	A	Methanospirillum	2	2	0	0	0	1.00
Methanosarcinales	A	Methanolobus	7	7	0	0	0	1.00
Methanosarcinales	A	Methanomethylovorans	2	2	0	0	0	1.00
Methanosarcinales	A	Methanosaeta	4	4	0	0	0	1.00
Methanosarcinales	A	Methanosarcina	11	11	0	0	0	1.00
Rhodocyclales	A	Azospira	2	0	1	1	0	-0.50
Rhodocyclales	A	Dechloromonas	3	0	3	0	0	0.00
Rhodocyclales	A	Methylovorsatilis	1	0	0	1	0	-1.00
Rhodocyclales	A	Thauera	9	0	9	0	0	0.00
Synergistales	A	Anaerobaculum	3	3	0	0	0	1.00
Syntrophobacterales	A	Desulfoglaeba	1	1	0	0	0	1.00
Syntrophobacterales	A	Smithella	1	1	0	0	0	1.00
Syntrophobacterales	A	Syntrophus	3	3	0	0	0	1.00



Thermotogales	A	Kosmotoga	2	2	0	0	0	1.00
Actinomycetales	B	Corynebacterium	111	0	92	19	0	-0.17
Actinomycetales	B	Dietzia	14	0	6	6	2	-0.50
Actinomycetales	B	Microbacterium	78	0	3	50	25	-0.94
Actinomycetales	B	Propionibacterium	14	8	6	0	0	0.57
Bacillales	B	Staphylococcus	47	0	45	1	1	-0.02
Burkholderiales	B	Bordetella	8	0	3	5	0	-0.63
Burkholderiales	B	Derxia	1	0	1	0	0	0.00
Burkholderiales	B	Cupriavidus	13	0	2	11	0	-0.85
Burkholderiales	B	Limnobacter	2	0	0	2	0	-1.00
Burkholderiales	B	Acidovorax	15	0	5	10	0	-0.67
Burkholderiales	B	Aquabacterium	5	0	5	0	0	0.00
Burkholderiales	B	Brachymonas	2	0	1	1	0	-0.50
Burkholderiales	B	Delftia	4	0	0	4	0	-1.00
Burkholderiales	B	Diaphorobacter	2	0	2	0	0	0.00
Burkholderiales	B	Hydrogenophaga	9	0	8	1	0	-0.11
Burkholderiales	B	Pelomonas	3	0	2	0	1	0.00
Burkholderiales	B	Rhodoferax	3	0	3	0	0	0.00
Burkholderiales	B	Simplicispira	3	0	1	2	0	-0.67
Burkholderiales	B	Duganella	2	0	0	2	0	-1.00
Caulobacterales	B	Brevundimonas	23	0	8	15	0	-0.65
Caulobacterales	B	Phenylobacterium	8	0	1	7	0	-0.88
Enterobacteriales	B	Citrobacter	12	1	5	6	0	-0.42
Enterobacteriales	B	Escherichia	7	0	7	0	0	0.00
Flavobacteriales	B	Flavobacterium	113	0	22	91	0	-0.81
Flavobacteriales	B	Lutibacter	3	0	0	3	0	-1.00
Lactobacillales	B	Streptococcus	116	4	112	0	0	0.03
Rhizobiales	B	Ochrobactrum	17	0	1	16	0	-0.94
Rhizobiales	B	Xanthobacter	5	0	0	5	0	-1.00
Rhizobiales	B	Methylobacterium	42	0	0	42	0	-1.00
Rhizobiales	B	Rhizobium	61	0	9	52	0	-0.85
Sphingobacteriales	B	Sediminibacterium	1	0	0	1	0	-1.00
Sphingobacteriales	B	Algoriphagus	23	0	8	15	0	-0.65
Sphingobacteriales	B	Pedobacter	36	6	1	29	0	-0.64
Sphingomonadales	B	Porphyrobacter	6	0	0	6	0	-1.00
Sphingomonadales	B	Novosphingobium	20	0	2	18	0	-0.90
Sphingomonadales	B	Sphingomonas	73	0	0	73	0	-1.00
Xanthomonadales	B	Arenimonas	6	0	1	5	0	-0.83
Aeromonadales	C	Aeromonas	29	0	29	0	0	0.00
Alteromonadales	C	Marinobacter	33	0	12	21	0	-0.64
Alteromonadales	C	Shewanella	60	0	60	0	0	0.00
Bacteroidales	C	Porphyromonas	16	16	0	0	0	1.00
Bacteroidales	C	Prevotella	48	48	0	0	0	1.00
Fusobacteriales	C	Ilyobacter	4	4	0	0	0	1.00
Oceanospirillales	C	Halomonas	82	0	25	57	0	-0.70
Oceanospirillales	C	Marinobacterium	11	0	1	8	2	-0.89
Pseudomonadales	C	Acinetobacter	27	0	1	26	0	-0.96
Pseudomonadales	C	Pseudomonas	150	0	43	64	43	-0.60
Rhodobacterales	C	Paracoccus	35	0	5	30	0	-0.86
Rhodobacterales	C	Rhodobacter	16	0	16	0	0	0.00
Rhodospirillales	C	Magnetospirillum	4	0	1	3	0	-0.75
Thermoanaerobacteriales	C	Coprothermobacter	2	2	0	0	0	1.00
Acholeplasmatales		Acholeplasma	18	0	18	0	0	0.00
Legionellales		Legionella	59	0	0	59	0	-1.00
Methylococcales		Methylocaldum	3	0	0	3	0	-1.00
Opitutales		Opitutus	1	1	0	0	0	1.00
Thermales		Thermus	14	0	3	11	0	-0.79
Thermococcales		Thermococcus	27	27	0	0	0	1.00
Thiotrichales		Methylophaga	8	0	5	3	0	-0.38
Vibrionales		Vibrio	29	0	29	0	0	0.00
Sum			1955					

**Table S5.** Distribution of community compositions derived from phylogenetically identified 16S rRNA gene pyrosequencing reads after PCR (16S P), from 16S sequence reads identified in the metagenomes (16S M), and from all phylogenetically identified genes in the metagenome (EGTs M). The proportions (%) of the most prevalent orders identified by 16S pyrotag PCR were determined (not shown) and compared with proportions of the other two categories. Sums for these three categories for all 12 samples for which metagenomic sequencing was done are indicated. The Network or Group assignment for each order is shown. Eight genera at the bottom of the table belonged to orders not assigned to Network A Network B or Group C.

Phylum_Class	Order	Network or group	Sum (16S P)	Sum (16S M)	Sum (EGTs)
Actinobacteria_Coriobacteridae	Coriobacteriales	A	2	1	0
Thermotogae_Thermotogae	Thermotogales	A	1	0	0
Euryarchaeota_Methanomicrobia	Methanomicrobiales	A	96	12	19
Euryarchaeota_Thermoplasmata	Thermoplasmatales	A	1	1	0
Synergistetes_Synergistia	Synergistales	A	3	0	0
Proteobacteria_Deltaproteobacteria	Syntrophobacterales	A	9	9	7
Chloroflexi_Anaerolineae	Anaerolineales	A	11	14	9
Euryarchaeota_Methanobacteria	Methanobacteriales	A	35	5	5
Spirochaetes_Spirochaetes	Spirochaetales	A	3	4	4
Euryarchaeota_Methanomicrobia	Methanosarcinales	A	88	12	22
Proteobacteria_Deltaproteobacteria	Desulfobacterales	A	19	9	6
Proteobacteria_Betaproteobacteria	Hydrogenophilales	A	34	25	13
Firmicutes_Clostridia	Clostridiales	A	154	91	38
Proteobacteria_Deltaproteobacteria	Desulfuromonadales	A	29	20	32
Firmicutes_Erysipelotrichi	Erysipelotrichales	A	1	1	0
Proteobacteria_Betaproteobacteria	Rhodocyclales	A	83	39	28
Proteobacteria_Gammaproteobacteria	Chromatiales	A	6	15	9
Proteobacteria_Deltaproteobacteria	Desulfovibrionales	A	20	21	34
Proteobacteria_Epsilonproteobacteria	Campylobacterales	A	7	203	97
Firmicutes_Bacilli	Lactobacillales	B	1	0	0
Firmicutes_Bacilli	Bacillales	B	1	6	12
Actinobacteria_Actinobacteridae	Actinomycetales	B	4	20	11
Bacteroidetes_Sphingobacteria	Sphingobacteriales	B	6	16	1
Proteobacteria_Alphaproteobacteria	Sphingomonadales	B	6	5	6
Proteobacteria_Betaproteobacteria	Flavobacteriales	B	24	27	7
Proteobacteria_Alphaproteobacteria	Rhizobiales	B	5	12	12
Proteobacteria_Gammaproteobacteria	Xanthomonadales	B	2	6	2
Proteobacteria_Gammaproteobacteria	Enterobacteriales	B	0	1	0
Proteobacteria_Betaproteobacteria	Burkholderiales	B	155	118	102
Proteobacteria_Alphaproteobacteria	Caulobacterales	B	1	8	2
Proteobacteria_Gammaproteobacteria	Pseudomonadales	C	218	117	146
Proteobacteria_Alphaproteobacteria	Rhodobacterales	C	33	45	51
Proteobacteria_Gammaproteobacteria	Alteromonadales	C	37	14	11
Proteobacteria_Gammaproteobacteria	Oceanospirillales	C	19	14	4
Fusobacteria_Fusobacteria	Fusobacteriales	C	0	0	0
Proteobacteria_Alphaproteobacteria	Rhodospirillales	C	3	7	3
Bacteroidetes_Bacteroidia	Bacteroidales	C	6	6	7
Proteobacteria_Gammaproteobacteria	Aeromonadales	C	3	5	1
Planctomycetes_Planctomycetacia	Planctomycetales	C	1	0	0
Proteobacteria_Betaproteobacteria	Methylophilales	C	2	2	0
Firmicutes_Clostridia	Thermoanaerobacterale	C	8	27	20
Proteobacteria_Betaproteobacteria	Nitrosomonadales	C	0	5	1
Proteobacteria_Gammaproteobacteria	Methylococcales	C	1	7	3
Proteobacteria_Betaproteobacteria	Neisseriales	C	1	3	2
Deferribacteres_Deferribacteres	Deferribacterales	C	0	3	7
Acidobacteria_Acidobacteria	Acidobacteriales	C	0	7	0
Euryarchaeota_Methanomicrobia	Methanocellales	C	1	15	8
Proteobacteria_Deltaproteobacteria	Bdellovibrionales	C	0	7	2
Verrucomicrobia_Opitutae	Puniceicoccales	C	3	4	3
	Others		7	209	454
	Sum		1150	1198	1200

**Table S6.** Characterization of 12 metagenomes from samples identified in Figure. 1. The R-score for the sequenced sample and the distribution of PCR amplified 16S rRNA reads over Network A and Network B and Group C are indicated. The sequencing strategy 454 single end (454S) or Illumina paired end (IlluminaP) is indicated, together with SRA and IMG/G accession numbers, the total length of sequenced bases (Mbp), the median contig size (bp) and the total length of assembled bases (Mbp). The numbers of reads matching with the indicated gene categories functioning in O<sub>2</sub>-dependent or O<sub>2</sub>-independent hydrocarbon degradation, methanogenesis or methane oxidation in normalized metagenomic datasets (100,000 sequences) are indicated. CG7, CG8, CG13 and CG19 are Conoco-Phillips produced waters; 10PW is produced water from the MHGC oil field; T\_1560D are Trident coal cuttings from 1560 mbs; Syncrude is a TP\_MLSB sample collected in 2011; TP6\_2010 is a Suncor TP6 sample, pooled from different depths and collected in 2010; TP6\_2008 is a Suncor TP6 sample, collected from 10 mbs in 2008; CO182 and CO183 are Quicksilver CBM cuttings; PDSYNTPWS is Syncrude tailing pond surface water. Normalized gene counts were obtained from the IMG/MER annotation.<sup>23</sup>

	HRE	CBM_water	Oil_field	CBM_water	CBM_water	CBM_deuts	TP_MLSB	TP6	TP6	TP6	CBM_water	TP_surface	CBM_cuts	CBM_cuts
Metagenome Name	CG8	10PW	CG13	CG19	T_1560D	Syncrude	TP6_2010	TP6_2010	TP6_2008	CG7	PDSYNTPWS	CBM182	CBM183	
Clade (Figure 1)	42	17	23	25	43	7	8	4	18_c	33	18_a	18_b	18_b	
SRA accession number	SRX210870	SRX210984	SRX210867	SRX210868	SRX210875	SRX210980	SRX210872	SRX210871	SRX210869	SRX210986	SRX211003	SRX211004	SRX211004	
IMG Genome ID	3300000589	3300000501	3300000584	3300000594	2228664011	3300000185	2228664013	2228664008	3300000586	3300000513	330000052	330000052	330000053	
Strategy	454S	454S	454S	454S	454S	454S	454S	454S	454S	454S	IlluminaP	IlluminaP	IlluminaP	
Total sequenced bases (Mbp)	60	240	57	68	242	309	188	294	70	320	72.283	36.706	577	
N50, median contig size of the assembly (bp)	1145	2817	1755	1104	1937	1476	1124	2357	1872	1490	590	577	577	
Total assembled bases (Mbp)	4.73	16.84	2.16	2.12	10.41	14.46	8	14.72	6.64	11.18	283.51	201.43	40.49	
R-score of sample	0.96	0.74	0.59	0.4	0.11	0.07	-0.01	-0.04	-0.32	-0.36	-0.44	-0.49	-0.49	
Network A	93.314	89.997	66.771	89.064	33.17	67.472	52.988	47.738	17.262	8.986	13.192	21.768	21.768	
Network B	0.229	1.996	0.156	1.719	4.333	22.344	27.292	32.974	0.338	75.364	20.796	17.968	17.968	
Group C	4.06	6.026	32.035	0.634	57.623	3.696	11.688	2.754	81.541	7.723	63.892	57.667	57.667	
<b>(I) Activation (Peripheral pathways), O<sub>2</sub> dependent</b>														
<b>1. Dioxxygenases</b>														
Benzoate 1,2-dioxygenase	EC 1.14.12.10	0	0	0	87	49	70	47	29	167	109	35	45	
2-Halobenzoate 1,2-dioxygenase	EC 1.14.12.13	0	0	0	0	14	22	3	7	34	10	5	7	
Biphenyl 2,3-dioxygenase	EC 1.14.12.18	0	0	0	32	17	19	6	63	77	29	17	49	
4-hydroxyphenylpyruvate dioxygenase	EC 1.13.11.27	0	5	0	64	71	79	24	2	13	66	71	11	
Phthalate 4,5-dioxygenase	EC 1.14.12.7	0	5	0	0	4	3	0	6	2	6	2	0	
3-phenylpropanoate dioxygenase	EC 1.14.12.19	0	0	0	0	26	2	0	59	184	85	17	54	
Paa_A_PaaC	PF05138	0	5	0	0	16	12	4	5	4	14	24	6	
<b>2. Monoxygenases</b>														
Limonene 1,2-monoxygenase	EC 1.14.12.107	0	5	0	0	0	0	0	0	0	13	0	0	
2,4-dichlorophenol 6-monoxygenase	EC 1.14.13.20	0	0	0	0	12	7	0	1	1	10	7	0	
Salicylate 1-monoxygenase	EC 1.14.13.1	0	5	0	0	35	29	3	15	8	44	35	7	
Vanillate monoxygenase	EC 1.14.13.82	0	9	0	0	30	22	0	6	7	34	35	5	
Heme monoxygenase(P450s)	PF00067	0	8	0	3	33	19	5	4	14	47	39	3	
<b>(I) Activation (Peripheral pathways), O<sub>2</sub> independent</b>														
<b>3. Carboxylation</b>														
4-hydroxybenzoate carboxyl-lyase	EC 4.1.1.61	45	33	50	0	2	2	3	8	2	2	0	0	
Phenylphosphate carboxylase gamma subunit	PF09662	0	2	0	0	1	1	0	0	0	1	0	0	
<b>4. Fumarate addition</b>														
Benzylsuccinate synthase	EC 4.1.99.11	22	54	0	0	11	4	32	51	8	0	0	16	
Succinyl-CoA:(R)-benzylsuccinate CoA-transferase	EC 2.8.3.15	0	23	14	0	4	5	15	10	14	4	0	0	
(R)-benzylsuccinyl-CoA dehydrogenase	EC 1.3.99.21	0	27	7	0	1	1	3	3	0	6	0	0	
Succinate dehydrogenase	EC 1.3.99.1	0	5	0	0	13	8	0	1	6	5	0	3	
Fumarate_red_D	PF02313	15	0	0	5	1	2	1	2	1	1	0	0	
<b>(II) Dearomatization (Central pathways), O<sub>2</sub> dependent</b>														
<b>1. Oxygen-dependent intradiol ring-cleavage</b>														
Protocatechuate 3,4-dioxygenase	EC 1.13.11.3	0	18	0	32	38	42	12	1	4	24	64	11	
Catechol 1,2-dioxygenase	EC 1.13.11.1	0	0	0	16	9	11	3	1	2	12	38	4	
Hydroxyquinol 1,2-dioxygenase	EC 1.13.11.37	0	0	0	0	2	3	0	0	1	4	2	0	
<b>2. Oxygen-dependent extradiol ring-cleavage</b>														
Catechol 2,3-dioxygenase	EC 1.13.11.2	0	0	0	16	24	26	41	25	43	40	45	36	
Homogentisate 1,2-dioxygenase	EC 1.13.11.5	0	9	0	16	26	30	24	2	7	49	31	3	
Protocatechuate 4,5-dioxygenase alpha chain (ligA)	EC 1.13.11.8	0	9	0	0	20	18	6	5	21	23	21	12	
Homoprotocatechuate 2,3-dioxygenase	EC 1.13.11.15	0	0	0	0	9	8	0	0	1	4	21	4	
Biphenyl-2,3-diol 1,2-dioxygenase	EC 1.13.11.39	0	0	0	16	15	16	9	11	25	22	19	17	
Gentisate 1,2-dioxygenase	EC 1.13.11.4	0	0	7	0	14	13	6	1	2	18	5	3	
3-carboxyethylcatechol 2,3-dioxygenase	EC 1.13.11.16	0	0	0	0	4	4	0	3	8	10	5	4	
3-Hydroxyanthranilate 3,4-dioxygenase	EC 1.13.11.6	0	0	0	0	3	3	0	0	0	13	2	0	
<b>(II) Dearomatization (Central pathways), O<sub>2</sub> independent</b>														
<b>3. ATP-dependent ring cleavage</b>														
2-hydroxyglutaryl-CoA dehydratase, D-component (HGD-D)	PF06050	74	75	44	48	7	5	32	57	58	5	21	45	
<b>4. ATP-independent ring cleavage</b>														
Aldehyde:ferredoxin oxidoreductases (AOR_BamB)	EC 1.2.7.5	245	447	895	541	20	20	194	260	256	29	175	270	
<b>(III) Methanogenesis</b>														
<b>1. Hydrogenotrophic</b>														
Formylmethanofuran dehydrogenase	EC 1.2.99.5	758	510	462	119	9	4	379	113	138	33	28	131	
Formylmethanofuran-tetrahydromethanopterin N-formyltransferase	EC 2.3.1.101	22	50	28	8	3	1	18	7	14	9	0	13	
Methylenetetrahydromethanopterin cyclohydrolase	EC 3.5.4.27	45	72	43	8	2	1	21	10	10	13	0	11	
Methylenetetrahydromethanopterin dehydrogenase	EC 1.5.99.9	22	36	43	0	0	0	41	7	3	2	0	4	
Coenzyme F420 hydrogenase	EC 1.12.98.1	89	275	121	40	11	13	88	42	45	16	12	45	
<b>2. Acetotrophic</b>														
Carbon-monoxide dehydrogenase (acceptor)	EC 1.2.99.2	245	334	661	183	62	43	150	219	235	223	184	201	
<b>3. Hydrogenotrophic &amp; Acetotrophic</b>														
Tetrahydromethanopterin S-methyltransferase	EC 2.1.1.86	201	266	263	16	0	0	97	44	37	1	7	52	
Coenzyme-B sulfoethylthioesterase	EC 2.8.4.1	134	185	227	40	0	0	82	31	19	0	5	25	
CoB-CoM heterodisulfide reductase	EC 1.8.98.1	423	740	675	302	0	3	291	395	442	16	57	432	
<b>(IV) Methane oxidation</b>														
Methane monoxygenase	EC 1.14.13.25	0	0	0	16	27	30	15	42	44	51	21	30	

**Table S7.** Distribution of the 10 most prevalent bacterial taxa, highlighted in yellow at each depth (as in Figure 3) of oil sands core FB11. The values in parentheses are the number of quality controlled reads.

phylum_class_order_family_genus	Depth (cm)											Aver	
	5 (5106)	15 (3510)	25 (3921)	35 (5252)	45 (5026)	55 (5593)	65 (3901)	75 (4641)	85 (5073)	95 (2608)	105 (3158)		115 (4404)
Proteobacteria_Gammaproteobacteria_Pseudomonadales_Moraxellaceae_Acinetobacter	17.20	12.05	5.94	14.62	17.17	1.82	23.64	1.57	20.66	2.76	2.38	20.71	11.71
Proteobacteria_Alphaproteobacteria_Rhizobiales_Rhizobiaceae_Rhizobium	14.18	4.99	1.94	5.24	5.47	0.86	9.90	0.65	7.22	1.34	1.30	8.88	5.16
Proteobacteria_Betaproteobacteria_Burkholderiales_Burkholderiaceae_Cupriavidus	10.13	4.36	2.65	5.79	5.39	0.72	9.69	0.58	8.48	0.35	1.05	8.92	4.84
Proteobacteria_Alphaproteobacteria_Caulobacteriales_Caulobacteraceae_Brevundimonas	7.74	4.62	2.42	4.59	4.82	0.82	6.82	0.75	5.05	0.61	0.48	4.61	3.61
Proteobacteria_Betaproteobacteria_Burkholderiales_Comamonadaceae_Delftia	6.84	3.16	1.71	3.62	3.76	0.73	6.56	0.78	5.24	0.42	1.01	5.84	3.31
Proteobacteria_Alphaproteobacteria_Rhizobiales_Methylobacteriaceae_Methylobacterium	6.76	2.99	1.61	3.43	3.86	0.77	4.64	0.67	5.09	0.77	0.48	7.36	3.20
Proteobacteria_Betaproteobacteria_Burkholderiales_Burkholderiaceae_Ralstonia	2.15	0.68	26.93	0.65	0.12	0.13	0.33	0	0.71	0.04	0.03	1.20	2.75
Proteobacteria_Betaproteobacteria_Burkholderiales_Comamonadaceae_Pelomonas	4.11	1.48	1.02	1.79	2.89	0.22	3.49	0.37	3.33	0.38	0.35	2.91	1.86
Proteobacteria_Betaproteobacteria_Burkholderiales_Comamonadaceae_Aquabacterium	1.61	1.20	1.45	0.86	1.33	0.25	2.82	0.30	2.25	0.15	0.38	4.22	1.40
Proteobacteria_Betaproteobacteria_Burkholderiales_Comamonadaceae	1.63	0.40	1.51	1.30	0.96	0.39	2.54	0.28	2.43	0.31	0.41	3.36	1.29
Actinobacteria_Actinobacteridae_Actinomycetales_Microbacteriaceae_Microbacterium	3.47	1.71	0.71	1.41	0.94	0.16	2.74	0.15	1.96	0.19	0.29	1.52	1.24
Proteobacteria_Gammaproteobacteria_Enterobacteriales_Enterobacteriaceae_Enterococcus	1.88	0.86	0.56	1.09	2.15	0.16	2.54	0.15	2.09	0.23	0.32	1.82	1.15
Proteobacteria_Alphaproteobacteria_Rhizobiales	0.63	0.29	0.79	0.36	0.36	2.32	1.00	2.82	0.63	1.38	1.58	1.23	1.12
Proteobacteria_Alphaproteobacteria_Caulobacteriales_Caulobacteraceae_uncultured	2.62	0.83	0.77	1.28	1.33	0.18	1.18	0.19	1.83	0.19	0.76	2.02	1.10
Proteobacteria_Alphaproteobacteria_Sphingomonadales_Sphingomonadaceae_Sphingomonas	0.37	0.40	0.49	0.34	0.74	0.32	0.56	0.30	3.00	0.46	0.54	0.75	0.69
Proteobacteria_Gammaproteobacteria_Pseudomonadales_Pseudomonadaceae_Pseudomonas	0.33	0.20	1.38	0.36	0.70	0.73	0.62	0.69	0.63	0.46	0.63	0.45	0.60
Proteobacteria_Betaproteobacteria_Burkholderiales_Comamonadaceae_Acidovorax	0.22	0.14	0.33	0.10	0.44	0.45	1.10	0.22	0.75	0.19	0.29	2.91	0.59
Bacteroidetes_Sphingobacteria_Sphingobacteriales_Chitinophagaceae_Sediminibacterium	2.23	1.31	0.38	0.23	0.36	0.07	0.21	0.04	0.47	0.12	0.13	0.77	0.53
Proteobacteria_Betaproteobacteria_Burkholderiales_Burkholderiaceae_Burkholderia	0.74	0	2.55	0.17	0	0.05	0.18	0.04	0.39	0	0	0.09	0.35
Firmicutes_Bacilli_Lactobacillales_Carnobacteriaceae_Trichococcus	0	0	1.35	0	0	0.61	0.36	0.37	0.20	0.54	0.51	0	0.33
Proteobacteria_Gammaproteobacteria_Alteromonadales_Alteromonadaceae_Marinobacter	0	0	0	0	0	0.95	0	0.88	0	0.46	1.20	0	0.29
Proteobacteria_Gammaproteobacteria_Enterobacteriales_Enterobacteriaceae_Enterococcus	0.02	0	0.03	0.13	0	0.39	0.08	0.52	0.14	0.77	0.54	0.05	0.22
Sum	84.84	41.65	56.52	47.33	52.77	13.11	80.98	12.33	72.13	12.12	14.63	79.61	47.33