Subsurface Hydrocarbon Degradation Strategies in Low- and High-Sulfate Coal Seam Communities Identified with Activity-Based Metagenomics

Authors: Hannah Schweitzer[1,2,†§**], Heidi Smith[1,2,§], Elliott P. Barnhart[1,3], Luke McKay[1,4], Robin Gerlach[1,5,6], Alfred B. Cunningham[1,5,7], Rex R. Malmstrom[8], Danielle Goudeau[8], and Matthew W. Fields[1,2,5]**

Affiliations:

[1]Center for Biofilm Engineering, Montana State University, Bozeman, MT 59717, USA

[2]Department of Microbiology & Cell Biology, Montana State University, Bozeman, MT 59717, USA

[3]US Geological Survey, Wyoming-Montana Water Science Center, Helena, MT 59601,USA

[4]Department of Land Resources and Environmental Sciences, Montana State University, Bozeman, MT 59717, USA

[5]Energy Research Institute, Montana State University, Bozeman, MT 59717, USA

[6]Department of Biological and Chemical Engineering, Montana State University, Bozeman, MT 59717, USA

[7]Department of Civil Engineering, Montana State University, Bozeman, MT 59717, USA

[8]DOE Joint Genome Institute, Berkeley, CA 94720, USA

[§]Indicates both authors contributed equally to this work

[†]Now at UiT - The Arctic University of Norway, 9019 Tromsø, Norway

[**]Corresponding authors

H.D. Schweitzer, Post Doctoral Researcher

UiT - The Arctic University of Norway

The Norweigian College of Fishery Science

Muninbakken 21

9019 Tromsø, Norway

hannah.schweitzer@uit.no

M.W. Fields, Professor

Montana State University

Center for Biofilm Engineering

366 EPS Building

Bozeman, MT 59717, USA

Matthew.fields@montana.edu

One Sentence Summary:

Identifying hydrocarbon degradation strategies across redox gradients via metagenomic analysis of environmental and translationally active (BONCAT-FACS) samples from subsurface coal beds.

Abstract

Environmentally relevant metagenomes and BONCAT-FACS derived translationally active metagenomes from Powder River Basin coal seams were investigated to elucidate potential genes and functional groups involved in hydrocarbon degradation to methane in coal seams with high- and low-sulfate levels. An advanced subsurface environmental sampler allowed the establishment of coal-associated microbial communities under *in situ* conditions for metagenomic analyses from environmental and translationally active populations. Metagenomic sequencing demonstrated that biosurfactants, aerobic dioxygenases, and anaerobic phenol degradation pathways were present in active populations across the sampled redox gradient. In particular, results suggested the importance of anaerobic degradation pathways under high-sulfate conditions with an emphasis on fumarate addition. Under low-sulfate conditions, a mixture of both aerobic and anaerobic pathways were observed but with a predominance of aerobic dioxygenases. The putative low-molecular weight biosurfactant, lichysein, appeared to play a more important role compared to rhamnolipids. The novel methods used in this study—subsurface environmental samplers in combination with metagenomic sequencing of both translationally active metagenomes and environmental genomes—offer a deeper and environmentally relevant perspective on community genetic potential from coal seams poised at different redox potentials broadening the understanding of degradation strategies for subsurface carbon.

Introduction

The terrestrial subsurface contains the majority of Earth's organic carbon (~90%)[1], and much of the carbon can be converted to methane under anaerobic conditions through biogasification (*i.e.*, biological decomposition of organic matter into methane and secondary gases). Biogasification can take place in coal, black shale, and petroleum reservoirs and is estimated to account for over 20% of the world's natural gas resources[2]. Factors influencing biogasification include coal rank, redox conditions (*e.g.*, presence or absence of oxygen and oxyanions), and the genetic potential and activity of the microbial community. Coal is a heterogeneous and highly complex hydrocarbon consisting of polycyclic aromatic hydrocarbons, alkylated benzenes, and long and short chain n-alkanes[3], and despite the recalcitrant nature of coal, degradation by microbial consortia has been shown in a variety of coal formations[4]. It is generally accepted that shallower coal beds that contain sulfate do not produce methane because sulfate-reducing bacteria (SRB) outcompete methanogens for substrates (*e.g.*, acetate, CO_2 and hydrogen)[5,6]. In methanogenic coal beds[6,7], hydrogenotrophic and acetoclastic methanogens are commonly identified, including different types of acetoclastic methanogens (*e.g.*, *Methanothrix*, *Methanosarcina*), which have distinct pathways for acetate utilization. It remains unknown what type of methanogenesis predominates *in situ* for different coal seams under different physicochemical conditions[8,9].

New coal degradation pathways are still being discovered and the involvement of different pathways in the turnover of refractory carbon under various redox conditions remains largely unresolved[10–12]. The majority of coal degradation research has focused on fumarate addition, while less is known about alternate coal degradation strategies such as phenol degradation by carboxylation and hydroxylation of alkanes,

benzene and ethylbenzene[13,14]. The fumarate addition pathway involves the activation of *n*-alkanes by the addition of fumarate via the double bonds at the terminal or sub terminal carbon[13,15–19]. Several fumarate addition genes (*e.g.*, *ass*-alkylsuccinate synthase for alkanes, *bss*-benzylsuccinate synthase for alkylbenzenes, and *nms*-naphylmethylsuccinate synthase) are often used as catabolic biomarkers for anaerobic hydrocarbon degradation[15–17]. These genes have been characterized from many subsurface hydrocarbon-containing environments,[13,16,20–25] but the importance under different redox conditions is still unclear. Carboxylation and hydroxylation strategies are less well documented mechanisms of anaerobic degradation, although, in recent years work has begun to suggest importance in anaerobic hydrocarbon degradation[14,18,26,27], yet how these strategies vary across redox transition zones *in situ* and detecting organisms responsible for degradation warrants further investigation.

While biosurfactants have not been identified *in situ* in coal seams and are not considered a necessary hydrocarbon degradation gene, previous laboratory-based research demonstrates a potentially important role of these compounds in decreasing the hydrophobicity of the solid coal surface, allowing for cellular and/or protein interactions at the coal surface[28]. Biosurfactant-producing microorganisms likely play direct and indirect roles in hydrocarbon degradation[28–31]. The accumulation of the esterase hydrolase enzyme has been correlated to biosurfactant production and is often used as a biomarker for biosurfactant production[32]. Biosurfactants are routinely observed in environments that consist of complex hydrocarbons and are therefore hypothesized to be an interdependent, complex, and coordinated means of increasing coal bioavailability. However, studies that have demonstrated active biosurfactant-producing microorganisms in coal environments are lacking.

Aerobic hydrocarbon degradation is the most documented form of hydrocarbon degradation via the aerobic activation of alkanes with dioxygenase enzymes that use oxygen as an electron acceptor and as a reactant in hydroxylation[33]. Aerobic hydrocarbon degradation in coal environments is often disputed due to the uncertainty of the presence of oxygen and more research into anaerobic hydrocarbon degradation strategies has been performed in the last few decades. However, recent metagenomic analyses have indicated the presence of aerobic hydrocarbon degradation genes in coal beds across Canada, but it is unknown whether these genes were present in active organisms *in situ*[23].

The rate-limiting step in coal biogasification has been attributed to initial biological breakdown of the refractory hydrocarbon matrix[34]. As discussed above, while several degradation strategies including the role of biosurfactants and aerobic hydrocarbon degradation have been observed, it remains challenging to link genetic potential to activity using traditional sequencing methodologies[35]. While traditional sequencing methodologies allow for an environmental snapshot they do not distinguish between dead, dormant, or active genomes in the environment. We were able to procure environmentally relevant samples from three separate coal seams located in the Powder River Basin (PRB) near Birney, Montana, with a Subsurface Environmental Sampler (SES; Patent # US10704993B2)[36]. The SES is a novel sampling tool which enabled harvesting of *in situ* coal-associated biofilms representative of local environmental conditions. The SES-retrieved samples were analyzed via total and activity-based metagenomes (using BONCAT- Bioorthogonal noncanonical amino acid tagging) to examine metabolic potential across hydrologically isolated coal seams that spanned vertical redox gradients. BONCAT methods have previously been used in combination of fluorescence-activated cell

sorting (BONCAT-FACS) and SSU rRNA gene sequencing[37,38], but metagenomic sequencing of active cells from BONCAT-FACS has yet to be performed. From these samples, we performed gene- and genome-centric approaches to assess the total hydrocarbon degradation/surfactant potential for each coal seam community as well as examine changes in microbial diversity for the different communities. Investigation of the active microbial groups and functional potential in CBM habitats provided new insight into the hydrocarbon degradation strategies and the associated capacities for carbon cycling that vary with depth and sulfate conditions through shallow coal seams.

Results and Discussion

Microbial communities and diversity under high and low-sulfate conditions

An SES[36] was used to collect *in situ* coal-associated microbial communities for shotgun metagenomic sequencing in three geochemically diverse coal seams. Samples from a high-sulfate (Nance, N-H) coal seam and two low-sulfate (Flowers Goodale, FG-L; Terret, T-L) coal seams were co-assembled and used to determine differences in hydrocarbon degradation genes and the overall communities based on the metagenome-assembled genomes (MAGs) (Fig. 1). After refinement, 86 MAGs had >80% estimated completeness for the co-assembled environmental shotgun metagenomes (Fig. 2, Table S1). There were 769,799 genes contained in the environmental metagenomes on 92,729 contigs.

The MAGs that were most closely related to organisms from the *Desulfobacteraceae* family were observed across all sampled coal seams regardless of sulfate level, although highest coverage was in the high-sulfate coal seam (Fig. 2). The *Desulfobacteraceae* family consists of many well studied sulfate reducing bacteria

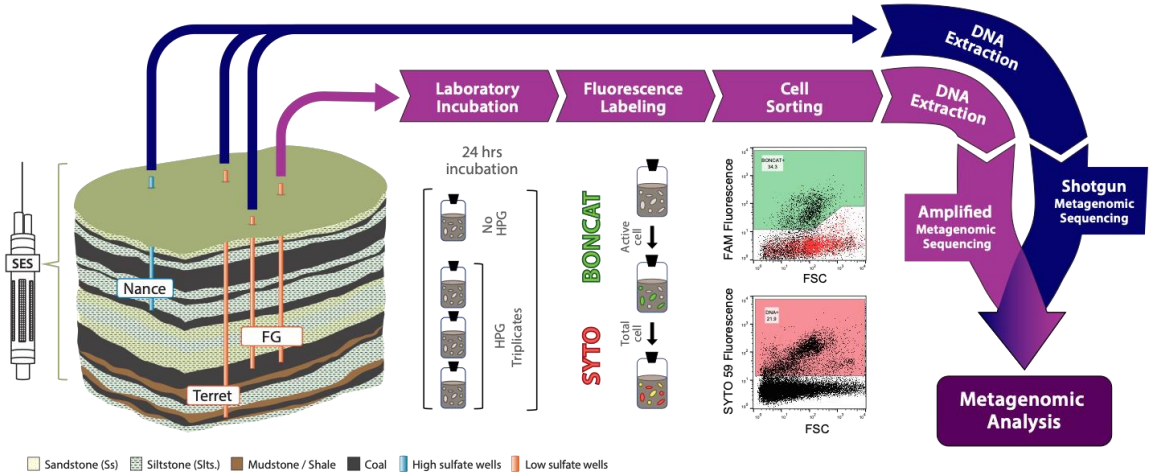


Figure 1. Schematic depicting experimental workflow beginning with *in situ* sampling of individual coal seams utilizing a Subsurface Environmental Sampler (SES) to retrieve CBM samples of the attached fraction under environmentally relevant conditions. Following retrieval, samples were either prepared for BONCAT incubation or shotgun metagenomics. For shotgun metagenomics, DNA was extracted and sequenced. BONCAT incubations lasted 24 hours followed by cell removal from coal, click chemistry, staining, and cell sorting of BONCAT+ and BONCAT-Total (SYTO stained) cell fractions, DNA extraction and amplified metagenomic sequencing.

(SRB) and is known to be metabolically versatile and exhibits multiple electron transfer complexes suggesting alternative metabolic mechanisms in low-sulfate environments[39]. Previous research has demonstrated increased abundance of SRBs in high-sulfate coal seams compared to low-sulfate coals seams[6]. Although *dsr*A genes, a commonly used gene marker for SRB, have been detected in acetate-amended methane-producing coal seams, suggesting organisms with *dsr*A may play an important role in syntrophic biogasification[40]. The total *dsr*A gene coverage for the high-sulfate coal seam (N-H) was 2.2X and 2.4X fold greater than for the low-sulfate coal seams, T-L and FG-L, respectively (Table S2). In the high-sulfate coal seam (N-H), the *dsr*A gene was identified in MAGs taxonomically similar to *Desulfobacteraceae* (Bin 372) and *Desulfatibacillum alkenivorans* (Bin 369), whereas the FG-L and T-L low-sulfate coals were predominated by MAGs with < 80% completion as well as *Desulfococcus (*Bin 21*), Desulfuromonas* (Bin 275), and

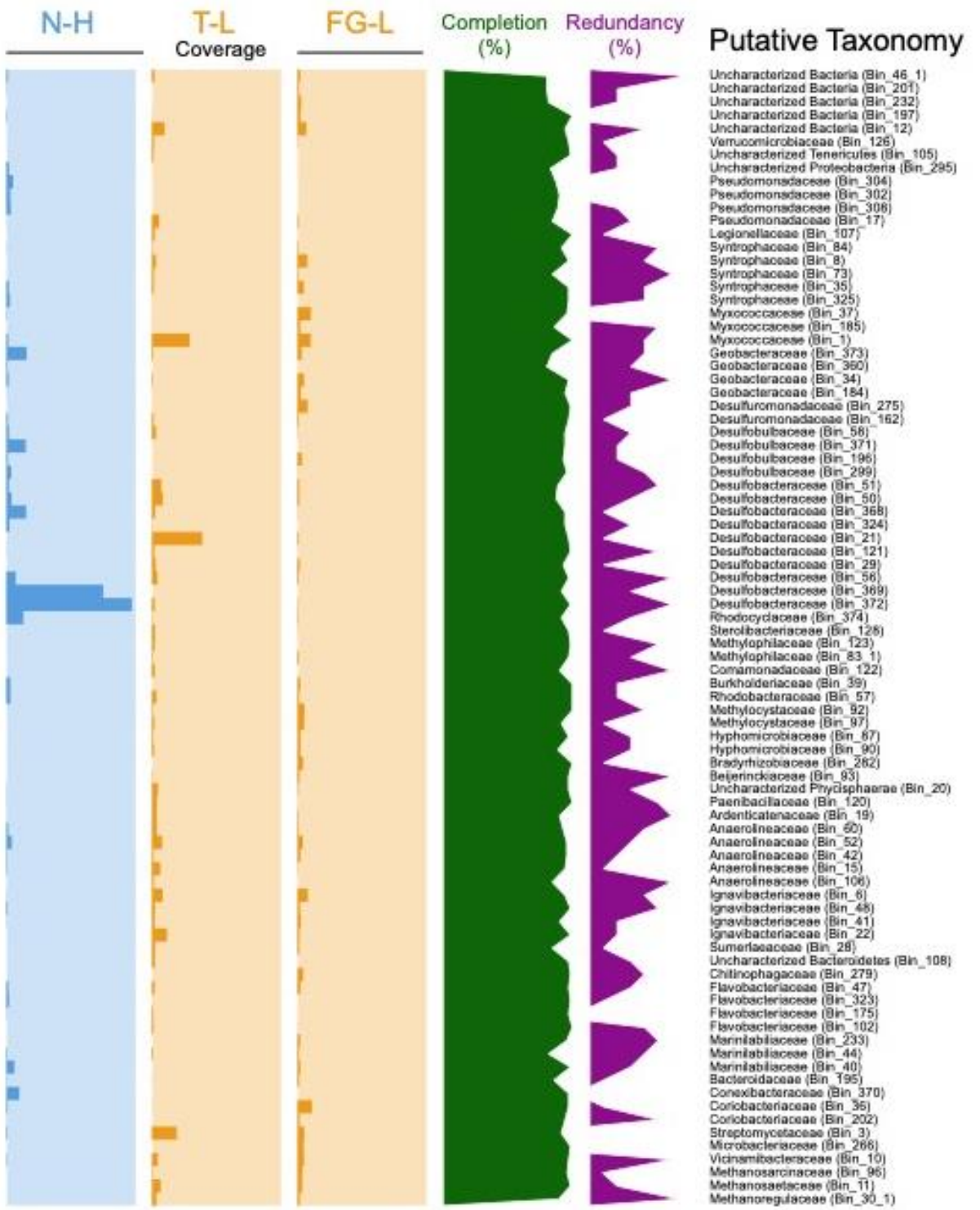


Figure 2. Comparison of coverages from the one high sulfate coal seam (N-H in blue) and two low sulfate coal seams (FG-L and T-L in orange) for each individual metagenome-assembled genome (MAG) with greater than 80% completion (green line graph). Coverage is calculated by finding the average depth of coverage across all contigs using Anvi'o compiled contigs. Coverage bar graphs axis is indicated with the highest possible sequencing depth being 550. All samples were below 10% redundancy (purple line graph).

178

Myxococcaceae (Bin 1 and Bin 37) (Fig. 3, Supplemental Data 1). Previous research on the *Desulfococcus* clade and the genus *Desulfatibacillum alkenivorans* reported the ability to anaerobically degrade alkanes[41,42]. *Desulfatibacillum alkenivorans* has also been described as being capable of using hexadecane as the sole carbon source in the presence of sulfate[42]. *Anaeromyxobacter dehalogenans* (from the *Myxococcaceae* family) is a facultative anaerobe that utilizes acetate with nitrate, oxygen, and fumarate[43].

The most predominant MAGs in the low-sulfate FG-L and T-L wells were most closely related to *Anaeromyxobacter* (Myxococcaceae), *Desulfococcus* (*Desulfobacteraceae*), and methylotrophic organisms from the *Methylocystaceae* family consisting of *Methylocystis* and *Methylosinus* (Fig. 2 and Table S1). Methane producing organisms such as acetoclastic methanogens, *Methanothrix* (*Methanosaetaceae*) and *Methanococcoides* (*Methanosarcinaceae*), and the hydrogentorphic methanogen, *Methanoregula formicica* (*Methanoregulaceae*), were also detected in the low-sulfate coal seams (Fig. 2 and Table S1). Hydrogenotrophic and acetoclastic methanogens are commonly identified in methanogenic coal beds[6,7], and different types of acetoclastic methanogens (*e.g.*, *Methanothrix*, *Methanosarcina*) have different pathways for acetate utilization. Active acetate utilization has been demonstrated in PRB coalbeds but the dominant microbial communities or metabolic pathways that generate or utilize acetate in these environments are generally unknown[35]. As stated above, *Anaeromyxobacter dehalogenans* typically thrive in environments with acetate further supporting the possibility of acetate in the low-sulfate communities[43]. Previous laboratory investigations have demonstrated that acetoclastic methanogens produce a large fraction of biogenic methane in coal enrichments and it is therefore hypothesized that high methane producing coal seams

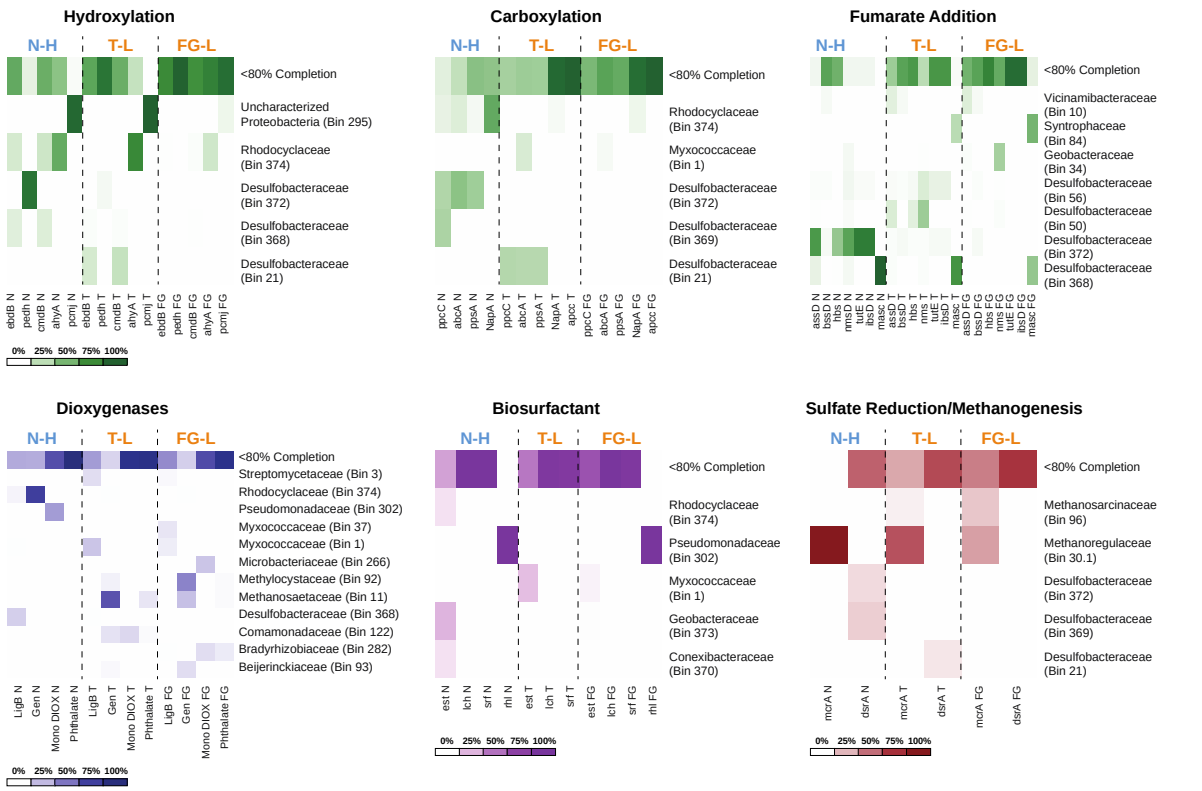


Figure 3. Heatmaps for abundant MAGs (>10% of the community) from shotgun metagenomic samples depicting the relative abundances for individual genes involved in hydrocarbon degradation for the high sulfate coal seam (N-H) and two low sulfate coal seams (T-L and FG-L). All MAGs that were below 80% completion were grouped together as <80% Completion. The six different heatmaps, displaying the genes of interest, were grouped based on method of anaerobic degradation (green heatmaps) via hydroxylation, carboxylation, and fumarate addition, aerobic degradation (blue heatmap) via dioxygenases, biosurfactant production (purple heat map) and sulfate reduction/ methanogenesis (red heatmap). All relative abundance of each gene found in the corresponding MAGs are represented by a color gradient (as seen in legend below heatmap) with the darkest color indicating the highest relative abundance for each sample and white indicating lack of the gene.

are abundant in acetoclastic methanogens[44,45]. In comparison to other methanogens, *Methanothrix spp*. have been reported to be better scavengers in low-acetate conditions[9]. However, recent isotopic analysis of PRB samples reveal that hydrogenotrophic methanogenesis can also be a significant methane producing pathway[44].

To further investigate the dominant methanogenic populations in the low-sulfate coal seams, the BONCAT positive (BONCAT+) metagenome from the FG-L coal seam was analyzed and resulted in 24 MAGs (Fig 4, Table S3). The BONCAT+

library had 888 contigs with 31,748 genes, while the total SYTO stained population (BONCAT-Total) from the same experiment yielded 71,455 genes on 1,748 contigs (Fig. S1, Table S4). Comparison of the BONCAT-Total and BONCAT+ samples demonstrated that 29.3 - 36.3% of the recovered genes belonged to populations that were BONCAT+ (Fig. S2). Six of the 24 BONCAT+ MAGs were taxonomically similar to known methanogens, *Methanothrix*, an acetoclastic methanogen and *Methanobacterium*, a hydrogenotrophic methanogen (Fig. 4, Table S3). These results suggest there are likely multiple (hydrogenotrophic and acetoclastic) active methanogenic pathways *in situ*[6,44], although the *Methanothrix* MAG exhibited significantly higher quality (Fig. 4).

Notably, three BONCAT+ MAGs had > 70% completeness and were related to the Bacteroidetes, Chlorobi and *Methanothrix* (Fig. 5, Table S3). The Chlorobi and *Methanothrix* MAGS overlapped between the BONCAT+ and the BONCAT-Total metagenome as determined by average nucleotide identity (Fig. 4, Fig. S1, Table S3, Table S4 and Supplemental Data 2). Bacteroidetes and Chlorobi have previously been observed to dominate acetate amended coal seams, and furthermore, organisms belonging to the Bacteroidetes phylum have been identified as a key lineage in hydrocarbon degradation[40,46]. These high quality BONCAT+ MAGs are examined in greater detail for potential involvement in coal degradation, acetate production, and subsequent acetoclastic methanogenesis in a separate study (McKay and Smith *et al*., in review).

Presence of Biosurfactant Genes

Presumptive biosurfactant genes for an esterase hydrolase enzyme (*est*), lichenysin synthetase (*lch*), and surfactin synthetase (*srf*) were detected with the

highest gene coverage in the FG-L coal seam (Fig. 5, Table S2). The *est* gene had the highest coverage (between 21.2-32.3 RPKM) compared to any of the investigated biosurfactant genes and was primarily in MAGs from the environmental metagenomes with < 80% estimated completeness (Fig. 3, Fig. 5, Table S2, and Supplemental Data 1). The *est* gene was also observed in three of the 24 BONCAT+ MAGs, two of which

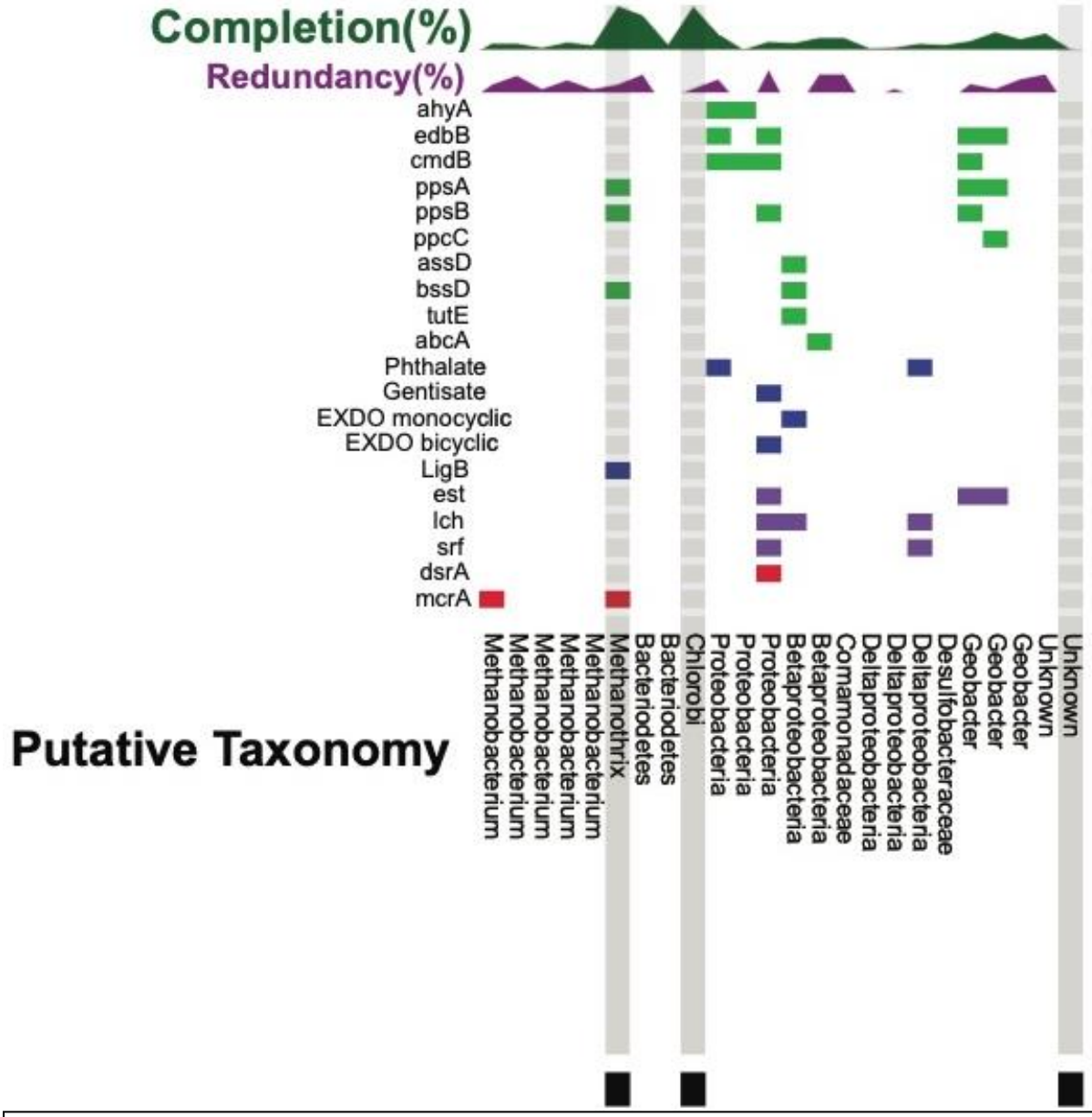


Figure 4. Evaluation of BONCAT positive metagenome assembled genomes (MAGs) and the presence of anaerobic hydrocarbon degrading (green), aerobic hydrocarbon degrading (blue), biosurfactant (purple) and sulfate reduction/methanogenesis (red) genes for each corresponding MAG. The percent completion (green line graph out of 100%) and redundancy (purple line graph out of 10%) were compared for each MAG. The three highlighted MAGs with the symbol (■) have a >99% average nucleotide identity similarity to BONCAT-Total MAGs indicated with the same symbol.

were most closely related to *Geobacter* and one related to Proteobacteria (Fig. 4, Table S3). Previous work has suggested the importance of esterases in the production of biosurfactants is due to the correlation between the accumulation of the esterase hydrolase enzyme and the production of lichenysin and surfactin[32]. Lichenysin, surfactin and the esterase enzyme genes had the highest coverage in the low-sulfate coal seams and were identified in the BONCAT+ metagenomes from FG-L (Fig. 4, Fig. 5, Table S2 and Table S3).

In the low-sulfate environmental metagenomes (FG-L and T-L), the *lch* and *srf* biosurfactant genes were most present in the MAGs that were < 80% estimated completeness (Fig. 3) and one highly complete (>80%) MAG that was taxonomically similar to *Legionella* (Supplemental Data 1). *Legionella* spp. have previously been shown to have the ability to produce biosurfactants that contain a lipid structure and had a similar environmental response as surfactin produced from *Bacillus subtilis*[47]. Lichenysin and surfactin are part of the surfactin operon family and contain four open reading frames that include *lch* and *srf*. Much like surfactin, lichenysin is a low molecular weight anionic cyclic lipopeptide biosurfactant previously shown to be produced by *Bacillus licheniformis* isolates[48]. Previous research has also demonstrated lichenysins are capable of enhancing oil recovery and degrading naphthalene and crude oil[48,49]. Biosurfactants have been suggested as having a potentially important role in decreasing the hydrophobicity of the solid coal surface, allowing for cellular and/or protein interactions at the coal surface[28]. Both lichenysin and surfactin are nonribosomal peptides and therefore are less energy intensive and could be ideal for oligotrophic conditions such as coal seams[48,50,51]. These results suggest biosurfactants are likely important in coal biogasification in low-sulfate environments and could facilitate the initial steps in hydrocarbon breakdown.

The *rhl* gene (rhamnolipids) had the lowest coverage of any of the investigated biosurfactant-related genes. The greatest *rhl* coverage was observed in the high-sulfate coal seam with little to no coverage in the low-sulfate coal seams (Fig. 5, Table S2). In addition, *rhl* was not detected in the BONCAT+ sample, although it was found in one MAG from the environmental shotgun metagenomes related to *Pseudomonadaceae* (Bin 302) (Fig. 3, Fig. 4, Table S2). Rhamnolipid biosurfactants produced by *Pseudomonas aeruginosa* and *Pseudomonas stutzeri* have previously shown the ability to enhance anaerobic degradation of the PAHs, phenanthrene and pyrene[52–54]. *Pseudomonas stutzeri*, a model biosurfactant producing organism, was detected in formation water originating from volatile bituminous coal in the Appalachian Basin coal bed and is suggested to play an important role in enhanced oil recovery[53–55]. The

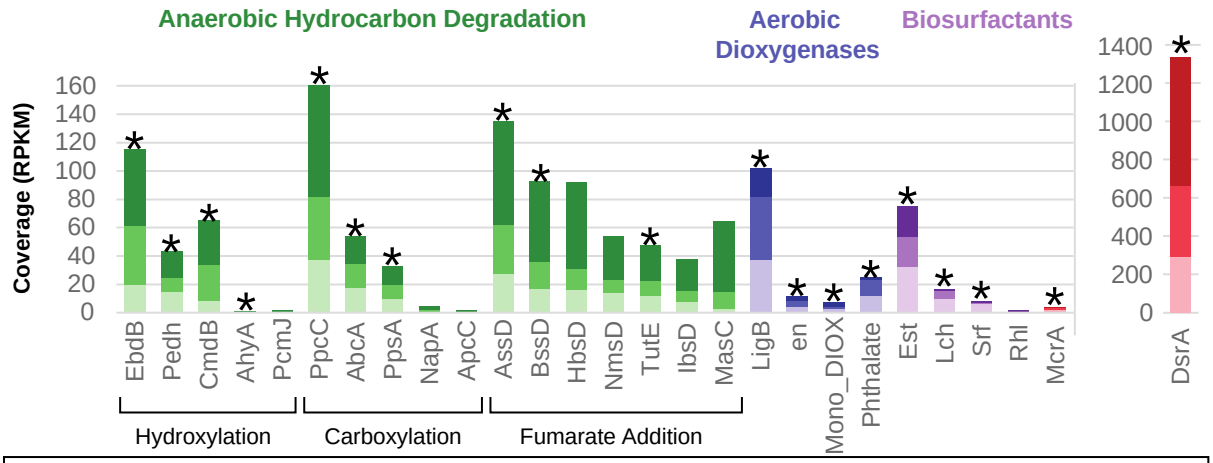


Figure 5. The coverage of each gene was calculated using the average depth of coverage across the contig the gene was identified in and all non-duplicate contigs containing the gene of interest above 40% identity were summed and reported here. Protein gene sequence coverage was normalized using the coverage of each gene per kilobase of transcript per million mapped reads (RPKM) for each gene of interest from each shotgun metagenomic sample. There are representative genes for each of the databases tested with genes of interest containing the highest coverage in one of the samples and/or being present in the BONCAT+ sample. The different genes are grouped based on anaerobic hydrocarbon degradation strategies (green), aerobic dioxygenases (blue), biosurfactant (purple) and sulfate reduction/ methanogenesis genes (red). All three environmental samples are represented in a color gradient with the high sulfate coal seam, N-H in the darkest shade (top), and two low sulfate coal seams, T-L (middle) with medium shade and FG-L (bottom) in the lightest shade. Stars indicate the presence of the individual gene in the BONCAT+ sample. The DsrA gene has its own axis because it had a greater overall coverage.

greater prevalence of *lch* and *srf* suggests that rhamnolipids may not be the dominant biosurfactants in all coal environments as previously hypothesized from benchtop coal enrichment experiments[52–55].

Hydrocarbon Degradation in High-Sulfate Environments

The high-sulfate coal seam (N-H) contained a higher average coverage for anaerobic hydrocarbon degradation genes compared to the low-sulfate coal seams (Fig. 5, Table S2). The phenylphosphate carboxylase gene (*ppc*C) had the highest coverage across all coal seam metagenomes regardless of sulfate concentration (Fig. 5, Table S2). Phenylphosphate carboxylase and phenylphosphate synthase can contribute to the anaerobic degradation of phenol, and phenylphosphate carboxylase is responsible for the conversion of phenylphosphate and CO_2 into 4-hydroxybenzoate and has previously been identified in *Geobacter metallireducens*[56]. Consistent with this, one of the most prominent populations represented by MAGs with > 80% estimated completeness in the high-sulfate coal seam (N-H) was taxonomically similar to *Geobacter* (Bin 360, Bin 373, Bin 34 and Bin 184) and contained the *ppc*C gene (Fig. 2, Table S1 and Supplemental Data 1).

The gene with the next highest coverage for the high-sulfate coal seam was the alkylsuccinate synthase subunit D (*ass*D) followed by hydroxybenzylsuccinate synthase (*hbs*D), and benzylsuccinate synthase subunit D (*bss*D) (Fig. 5, Table S2). These genes (*ass*D, *hbs*D and *bss*D) are all involved in hydrocarbon degradation via fumarate addition, which is currently the only well-described, oxygen-independent alkane activation[57]. Gene coverages for fumarate addition were higher for the high-sulfate coal seam (N-H) compared to the low-sulfate coal seams (Fig. 5, Table S2). For the high-sulfate coal seam, fumarate addition genes were detected in MAGs that

were most taxonomically similar to populations from the *Desulfobacteraceae* family (Bin 372) as well as others such as *Desulfatibacillum alkenivorans* (Bin 56) and the *Desulfococcus* clade (Bin 50, Bin 368) (Fig. 3). Previous research has shown these alkane-degrading organisms in the presence of sulfate[41,42]. Our results indicate fumarate addition could play a more crucial role in high-sulfate than low-sulfate coal seam environments for hydrocarbon degradation under high-sulfate conditions.

Low-sulfate hydrocarbon degradation: Aerobic and Anaerobic Strategies

Previously reported methane concentrations for low-sulfate FG-L and T-L (33-67 mg/L[7]) are consistent with our detection of higher *mcr*A gene coverages compared to the high-sulfate (< 0.15 mg/L methane) coal seam (Fig. 3, Fig. 5, Table S2). The rate-limiting step in coal biogasification has been attributed to initial biological breakdown of the refractory hydrocarbon matrix[34]. Analysis of the hydrocarbon degradation genes showed a mix of both aerobic and anaerobic strategies, yet there was a higher coverage of aerobic hydrocarbon degradation genes under low-sulfate compared to the high-sulfate coal seam. Homoprotocatechuate dioxygenase (LigB) was present in the BONCAT+ metagenome (Fig. 4 and Table S3), and in the environmental metagenomes LigB had the highest coverage of all investigated aerobic aromatic hydrocarbon degradation genes[58] regardless of sulfate level. Moreover, LigB had the highest coverage for any gene detected in the low-sulfate coal seams (Fig. 5, Table S2). The homoprotocatechuate dioxygenase is involved in the initial deoxygenation of 4-hydroxyphenylacetate to homoprotocatechuate (and eventually to fumarate). The monocyclic (Mono_DIOX) and bicyclic extradiol dioxygenases and gentisate dioxygenase (Gen) from the cupin superfamily were also in the BONCAT+ metagenome (Fig. 4, Table S3), although they were all at lower coverages across all

shotgun metagenomes compared to the LigB dioxygenase (Fig. 5, Table S2). The LigB dioxygenase was found in the putative methanogenic BONCAT+ MAG that was most closely related to *Methanothrix* (discussed above) indicating possible activity of aerobic hydrocarbon degradation (LigB) in addition to anaerobic hydrocarbon degradation (via *pps*AB and *bss*D). Possible oxygen-dependent or -tolerant properties of this putative methanogenic population are discussed in detail in a separate study (McKay and Smith *et al.*, in review). Further, the bicyclic dioxygenases and the gentisate dioxygenases (Gen) were in a BONCAT+ MAG taxonomically affiliated with the Proteobacteria, while the monocyclic dioxygenases were detected in a different MAG within the Betaproteobacteria (Fig. 4, Table S3) indicating further aerobic degradation capabilities.

The anaerobic hydrocarbon degradation gene with the greatest coverage in the low-sulfate environments was *ppc*C and was identified in BONCAT+ MAG most closely related to *Geobacter* (Fig. 4, Fig. 5, Table S2 and Table S3). The high coverage of *ppc*C in both high- and low-sulfate environments indicates that anaerobic phenol degradation is likely crucial in the degradation of PRB coal. Phenylphosphate carboxylase genes are often overlooked for the more well documented fumarate addition genes[13,14,18,26,27], but results presented herein suggest *ppc*C should be considered as a possible mechanisms for anaerobic hydrocarbon degradation. Phenol degradation via the activation of phenylphosphate followed by its carboxylation has previously been shown to only occur when in the presence of phenol and CO_2 and is believed to be a strictly anaerobic process[56]. However, previous research suggests *ppc* can exist in an inactive oxidized form until it is reduced and activated under anaerobic conditions[59]. Therefore, under fluctuating oxygen conditions, phenylphosphate carboxylase may play an important role.

While all the fumarate addition genes of interest were identified in the environmental shotgun metagenomes (Fig. 5) and belonged to predominant populations related to the *Geobacter* (Bin 34), *Syntrophus aciditrophicus* (Bin 84), *Anaeromyxobacter* (Bin 37), *Luteitalea pratensis* (Bin 10), and *Desulfococcus* (Bin 368 and Bin 50) (Fig. 3 and Supplemental Data 1), the coverage for these genes was much lower in the low-sulfate compared to the high-sulfate coal seam. Only two of the 24 BONCAT+ MAGs displayed the presence of fumarate genes (related to *Methanothrix* and Betaproteobacteria) and of the seven fumarate genes investigated only three (*assD*, *bssD*, and *tutE*) were present (Fig. 4, Fig. 5, Table S2, Table S3). It is possible that the low-sulfate and methanogenic conditions were not able to support fumarate addition, due to the lack of available electron acceptors, which are necessary for fumarate addition and therefore needed to rely on other methods of carbon degradation.

Together, our results suggest that fumarate is an important intermediate substrate in hydrocarbon degradation for both high- and low-sulfate environments, but likely plays a more important role in high-sulfate PRB environments. Previous laboratory investigations have demonstrated an increase in *ass*A gene abundance during carbon degradation in sulfate-reducing environments, with no such increase in corresponding methanogenic cultures[14,18,20]. In contrast, other oil field surveys and hydrocarbon degradation experiments demonstrate that fumarate addition is an important pathway under methanogenic conditions and organisms with the *ass*A, such as *Anaerolineaceae*, have shown the ability to degrade *n*-alkanes to produce acetate, potentially creating a syntrophic relationship with the acetoclastic methanogen, *Methanothrix*[6022,25]. Although, while fumarate addition likely plays a role in both high- and low- sulfate environments, our observations suggest that in low-sulfate coal

seams, prerequisite methanogenic substrates may result from both anaerobic and aerobic (homoprotocatechuate dioxygenase) hydrocarbon degradation as they are detected in higher coverages across all coal-dependent and active populations (Fig. 4, Fig. 5 and Table S2).

Conclusion

The microbial environment of the terrestrial subsurface is one of the most complex and poorly understood microbiome on Earth[61]. Due to diversity and high functional capacity, there have been many calls to better understand the diversity[62] and the activity of *in situ* microorganisms and the potential implications on the cycling of old and new carbon between the lithosphere and atmosphere[63]. There has been much debate on possible pathways actively involved in the subsurface carbon cycle, specifically, the degradation of complex hydrocarbons to methanogenic substrates under varying conditions (*e.g.* sulfate, oxygen). The subsurface environment constitutes a challenging environment to access and prior to this study, conclusions had been based primarily on laboratory benchtop enrichments, geochemical predictions, or shotgun metagenomic models[2,6,23]. The novel methods used in this work, accessing the subsurface and sampling of environmentally relevant *in situ* communities using the SES and the integration of metagenomes from active communities derived from BONCAT-FACS, expands both the gene and genome catalog of the subsurface environment as well as the activity of key carbon cycle genes indicative of hydrocarbon degradation. The active groups were putatively involved in both aerobic and anaerobic hydrocarbon degradation and biosurfactant production. The presented results demonstrated that active populations have the ability to produce diverse biosurfactants and could play an important role in coal-dependent

methanogenesis in low-sulfate coal seams. While this study proposes biosurfactants are important in low-sulfate coal seams, more work needs to be performed to determine the role and environmental triggers for biosurfactants from different coal ranks and sulfate concentrations.

Hydrocarbon degradation in coal seam environments is often assumed to be strictly anaerobic, although with advancements in metagenomics more research is beginning to suggest mixed aerobic and anaerobic hydrocarbon degradation[23,64]. Our results indicate that aerobic hydrocarbon degradation genes are associated with active populations in coal seam environments that span redox transition zones and are more prevalent in low-sulfate coal seams. Previously, enzymes belonging to the LigB superfamily had not been readily detected due to conventional screening techniques. In the advent of metagenomics, these enzymes have been broadly identified in the environment with an enrichment in hydrocarbon rich samples, suggesting enhanced importance in hydrocarbon degradation[65,66]. The BONCAT+ metagenomes were predominated by both aerobic and anaerobic degradation genes that included *ppc*C, *pps*A, *ebd*B, *cmd*B, *bss*D, *lig*B, monocyclic and bicyclic dioxygenases. While it is possible that active populations may not be using aerobic degradation genes, fluctuations in dissolved oxygen have been reported in different coal seams and coals have the ability to absorb free oxygen (Meslé and Fields, data not shown). Therefore, based on the active MAGs that contain aerobic genes, it is possible microaerophilic conditions could promote aerobic hydrocarbon degradation via dioxygenases such as the homoprotocatechuate dioxygenases from the LigB family. The source of oxygen in these environments is unknown, and likely depends on groundwater flow/recharge, surface water intrusion, water level changes, coal seam depth, and associated microbial activities. Further work should determine the extent of oxia, potential

oxygen sources, and potential segregation of hydrocarbon degradation and methanogenesis along heterogeneous critical zone transitions in the shallow subsurface.

Methods

Site Description and Sample Collection

The USGS Birney sampling site, located in the PRB of southeastern Montana (45.435606, -106.393309), has access to four subbituminous coal seams at different depths. Metagenomic samples were collected from a high-sulfate (23.78 – 26.04 mM) coal seam (N-H) at 65 meter depth and two low-sulfate (0.01 – 0.38 mM) coal seams (FG-L and T-L) located at 117 meter and 161 meter depths. The low-sulfate coal seams have measurable methane (33 – 67 mg/L) while the high-sulfate coal seams have low methane (<0.15 mg/L) as previously measured[7]. Coal associated microbial assemblages were collected with a DMS as previously described[7] or with a next-generation subsurface environmental sampler (SES; Patent# US10704993B2) following previous field protocols. The DMSs were incubated down-well for three months and retrieved on June 12th, 2017. Once retrieved, the slurry (coal and groundwater) from the DMSs were aseptically removed and stored on dry ice in sterile falcon tubes until brought back to the lab, where they were stored at -80°C until DNA extractions were performed.

DNA Extraction, Concentration and Sequencing for Shotgun Metagenomics

DNA was extracted from slurry in the DMS as previously described in Schweitzer et al.[6] using a FastDNA Spin Kit for Soil (MP Biomedical). DNA was purified using One Step PCR Clean Up (Zymo Research). The DNA was quantified

using Qubit dsDNA HS Assay Kit (Invitrogen) before being sent to the MBL Keck sequencing facility (The University of Chicago, IL). The genomic DNA was concentrated using a SpeedVac (Thermo Scientific) and quantified using Picogreen (Invitrogen). DNA concentrations ranged between 138.6ng (FG-L) and 1,450ng (T-L). Using a Covaris ultrasonicator, DNA was sheared to ~400 bp and libraries were constructed with NuGEN Ovation® Ultralow Library protocol (Tecan Genomics, CA). The amplified libraries were visualized on a Caliper HiSense Bioanalyzer and pooled to equimolar concentrations. The DNA was further size selected using a Sage PippinPrep 2% cassette. The pooled library was quantified using Kapa Biosystems qPCR library quantification before being sequenced in a 2x150 paired-end sequencing run using an Illumina NextSeq. The sequences were demultiplexed and the adapter sequences were trimmed using bcl2fastq conversion software. The sequences were archived with SRA (SRP292435) under the BioProject ID PRJNA678021.

<u>BONCAT Incubations, Fluorescent Labeling, Cell Sorting and Metagenomic Sequencing</u>

The following methods are also presented in our companion study, which examines high-quality BONCAT+ MAGs for their involvement in acetate production and acetoclastic methanogenesis (McKay and Smith *et al*., in review). Subsurface environmental samplers (SES, Patent # US10704993B2) that contained UV-sterile, crushed coal (size range 0.85-2.00mm) were deployed at the screened interval in a methane-producing coal seam (FG-L; September 2017) at the USGS Birney Field Site and, after approximately nine months of down-well incubation, were retrieved from CBM well FG11 at the USGS Birney Field Site[7]. In contrast to DMSs used for shotgun sequencing, the use of the SES allowed for samples to be kept under

anaerobic conditions and at formation pressures, an essential prerequisite for conducting representative BONCAT incubations. Upon retrieval, 10 mL of SES slurry was sampled via sealed ports (Swagelok) and were anaerobically transferred into gassed-out, sterile Balch tubes. Triplicate tubes were prepared for incubation with 250 μM L-homopropargylglycine (HPG, Click Chemistry Tools, Scottsdale, AZ, USA) prepared in sterile, degassed water (DEPC diethyl pyrocarbonate treated, pH 7). Control samples were prepared the same as other samples with the exception that HPG was not added (HPG negative control). Samples were incubated in the dark at 20 °C for 24 hrs. At the end of the incubation period cells were removed from coal by removing 1 mL of slurry and adding 5mls Tween® 20 dissolved in PBS (1X PBS) to a final concentration of 0.02% (Sigma-Aldrich). Samples were vortexed at maximum speed for 5 min followed by centrifugation at 500 × g for 5 min (Couradeau et al., 2019)[37]. The supernatant (containing the detached cells) was immediately cryopreserved at −20 °C in sterile 55% glycerol TE (11X) solution.

Translationally active cells were identified through a click reaction that added a fluorescent dye to HPG molecules that had been incorporated into newly synthesized protein. This BONCAT click reaction consisted of 5mM sodium ascorbate, 5mM aminoguanidine HCl, 500uM THPTA, 100uM $CuSO_4$, and 5uM FAM picolyl azide in 1X phosphate buffered saline. Incubation time was 30 minutes, followed by three washes in 20ml of 1X PBS for 5 minutes each. Cells were recovered from the filter by vortexing in 0.02% Tween for 5 minutes, and then stained using 0.5uM SYTO™59 (ThermoFisher Scientific, Invitrogen, Eugene OR, USA) DNA stain (Couradeau et al., 2019)[37].

Cells were sorted using a BD-InfluxTM (BD Biosciences, San Jose, CA, USA) configured to capture total cells labeled with SYTO™59 DNA stain and excited with a

640nm red laser and BONCAT-positive cells labeled with FAM picolyl azide dye and excited with a 488nm blue laser. BONCAT-positive cells were a subset to total SYTO™59 stained cells, and were differentiated from the BONCAT-negative based on their FAM fluorescence (530/40BP) to background fluorescence from identical cells inclubated without HPG, *i.e.*, HPG negative control, that also underwent the same click reaction to add a FAM label. The two populations were sorted. The first population contained all DNA+ cells (BONCAT-total), and the second population contained only those cells which were also BONCAT+ in comparison to the control. For each sample, we sorted 4 wells containing 5,000 cells each, and 20 wells containing 300 cells each into 384 wells plates. Plates were frozen at −80 °C until further processing.

Cells from samples containing 5,000 sorted cells were pelleted prior to sequencing via centrifugation at 6,000 x g for 1 hour at 10 °C. The supernatant was discarded by a brief inverted spin at <60 x g so as not to interfere with subsequent whole genome amplification reaction chemistry. Plates containing only 300 sorted cells were not pelleted. Sorted cells were lysed and amplified using 5 µl WGAX reactions following the optimized conditions described in Stepanauskas et al. (2017)[67]. Briefly, cells were lysed in 650 nl lysis buffer for 10 minutes at room temperature. The lysis buffer consisted of 300 nl TE + 350 nl of 400 mM KOH, 10 mM EDTA, 100 mM DTT. Lysis reactions were neutralized by the addition of 350 nl of 315 mM HCl in Tris-HCl. Amplification reactions were brought to 5 µL with final concentrations of 1X EquiPhi29 reaction buffer (Thermo), 0.2U/µl EquiPhi29 polymerase (Thermo), 0.4mM dNTPs, 50µM random heptamers, 10 mM DTT, and 0.5 µM SYTO13. Plates were incubated at 45 °C for 13 hours.

Libraries for BONCAT+ and BONCAT-total metagenomic sequencing were created using the Nextera XT v2 kit (Illumina) with 12 rounds of PCR amplification. All volumes and inputs to Nextera reactions were reduced 10-fold from the manufacturer's recommendations. Libraries were sequenced 2X150bp mode on Illumina's Nextseq platform. The sequences were archived under JGI GOLD Study ID Gs0141001.

Metagenomic Analysis with Anvi'o

NextSeq-generated fastq files were filtered at a minimum quality read length of 0.75 using Anvi'o for Illumina (http://merenlab.org). The filtered reads were co-assembled using MEGAHIT v1.1.2 (The University of Hong Kong and L3 Bioinformatics Limited). All shotgun, BONCAT+, and BONCAT-Total replicates were co-assembled and kept separate for subsequent comparisons. Alignments were accomplished using bowtie2 v2.3.6 and samtools v1.9. All bam files were merged into individual profile databases for each sample prior to binning using Anvi'o. Bins were assembled using tetranucleotide frequency into 376 genomic clusters for the shotgun metagenomics samples and 20 bins for the BONCAT+ amplified metagenome and 38 bins for the BONCAT-Total amplified metagenome. To determine completeness, scanning of ribosomal single copy genes for bacteria (Campbell et al. 2013[68]) and archaea (Rinke et al. 2013[69]) in each cluster was performed in Anvi'o. To identify open reading frames in contigs, Prodigal was used within Anvi'o. Functions were assigned in Anvi'o using NCBI's Clusters of Orthologous Genes (COGs) and a Hidden Markov Model hit database was generated with Anvi'o. Coverage estimates, GC content, and N50 were calculated. The phylogenetic tree was created using MUSCLE (drive5.com/muscle) for multiple sequence alignment. Anvi'o Interactive

Server (anvi-server.org) was used to create phylogenomic evaluation of genomic parameters for each MAG. All shotgun metagenome high quality (>80% completion), BONCAT+ and BONCAT-Total MAGs are archived under JGI GOLD (available upon publication).

Metagenomic Analysis with Curated Database

Taxonomic identity was determined using NCBI BLASTn of all contigs for each bin. All blasted contigs with the best e-value and percent identity were then compared to select the most common ancestor to represent the most likely taxonomic identification for all the contigs for MAGs > 20X coverage. Genes were analyzed using curated databases via direct protein sequence comparison. Curated databases for hydrocarbon degradation genes included aerobic (AromaDeg)[58] and anaerobic (AnHyDeg)[19] protein databases. Other functional gene databases (*e.g.*, biosurfactant genes, *dsr*A, and *mcr*A) were generated from protein sequences by extracting HMM files from the Pfam database (EMBL-EBI) and from McKay et al. (2017).[12] These protein databases were BLAST-searched against each protein sequence with an e-value cutoff of 1E-10 and above 40% identity. All duplicate gene hits were removed with the highest e-value hit being kept. All protein gene sequence coverages were normalized by using Reads/Coverage per kilobase of transcript per million mapped reads (RPKM).

$$RPKM = \frac{Gene\ Coverage}{\frac{Gene\ Length}{1000} * \frac{Total\ \#\ of\ Reads}{1{,}000{,}000}}$$

Genes were selected from the complete list of all genes (Table S1) and classified as genes of interest for further analyses. The genes of interest were selected first by ensuring there were representatives from each of the four databases tested (AnHyDeg,

AromaDeg, Biosurfactant, and Redox) and that one subunit representative from the same operon family was further analyzed. To determine which of the gene representatives were used, at least one of the following criteria had to be met: identified in the BONCAT+ sample or contained the highest gene coverage in one of the samples (FG-L, T-L, or N-H).

Acknowledgements

Metagenome sequencing was made possible by the Deep Carbon Observatory's Census of Deep Life and was performed at the Marine Biological Laboratory (Woods Hole, MA, United States) and Joint Genome Institute CSP 503725. We are grateful for the assistance of Mitch Sogin, Joseph Vineis, Amelia Brumbaugh, and Hilary Morrison at MBL. The work conducted by the U.S. Department of Energy Joint Genome Institute, a DOE Office of Science User Facility, is supported under contract No. DE-AC02-05CH11231. Any use of trade, firm, or product name is for descriptive purposes only and does not imply endorsement by the U.S. Government. The US Geological Survey (E.P.B) supported field work and sampling. The work was further supported by NSF 1736255 (L.J.M., R.G., and M.W.F.). The authors would also like to acknowledge Dr. Katie Davis and George Platt for their help in sample collection in the Powder River Basin.

References (Limited to 70 references)

1. Javoy, M. The major volatile elements of the Earth: Their origin, behavior, fate. *Geophys. Res. Lett.* **24**, 177–180 (1997).
2. Ritter, D. *et al.* Enhanced microbial coalbed methane generation: A review of research, commercial activity, and remaining challenges. *Int. J. Coal Geol.* **146**, 28–41 (2015).
3. Fakoussa, R. M. & Hofrichter, M. Biotechnology and microbiology of coal degradation. *Appl. Microbiol. Biotechnol.* **52**, 25–40 (1999).
4. Elder, D. J. E. & Kelly, D. J. The bacterial degradation of benzoic acid and benzenoid compounds under anaerobic conditions: Unifying trends and new perspectives. *FEMS Microbiol. Rev.* **13**, 441–468 (1994).
5. Muyzer, G. & Stams, A. J. M. The ecology and biotechnology of sulphate-reducing

bacteria. *Nat. Rev. Microbiol.* **6**, 441–454 (2008).
6. Schweitzer, H. *et al.* Changes in microbial communities and associated water and gas geochemistry across a sulfate gradient in coal beds: Powder River Basin, USA. *Geochim. Cosmochim. Acta* **245**, (2019).
7. Barnhart, E. P. *et al.* Hydrogeochemistry and coal-associated bacterial populations from a methanogenic coal bed. *Int. J. Coal Geol.* **162**, 14–26 (2016).
8. Barnhart, E. P. *et al.* Potential Role of Acetyl-CoA Synthetase (acs) and Malate Dehydrogenase (mae) in the Evolution of the Acetate Switch in Bacteria and Archaea. *Sci. Rep.* **5**, 12498 (2015).
9. Min, H. & Zinder, S. H. Kinetics of Acetate Utilization by Two Thermophilic Acetotrophic Methanogens: Methanosarcina sp. Strain CALS-1 and Methanothrix sp. Strain CALS-1. *Appl. Environ. Microbiol.* **55**, 488–91 (1989).
10. Borrel, G. *et al.* Comparative genomics highlights the unique biology of Methanomassiliicoccales, a Thermoplasmatales-related seventh order of methanogenic archaea that encodes pyrrolysine. *BMC Genomics* **15**, 679 (2014).
11. Haroon, M. F. *et al.* Anaerobic oxidation of methane coupled to nitrate reduction in a novel archaeal lineage. *Nature* **500**, 567–570 (2013).
12. McKay, L. J., Hatzenpichler, R., Inskeep, W. P. & Fields, M. W. Occurrence and expression of novel methyl-coenzyme M reductase gene (mcrA) variants in hot spring sediments. *Sci. Rep.* **7**, (2017).
13. Callaghan, A. V. *et al.* Diversity of benzyl- and alkylsuccinate synthase genes in hydrocarbon-impacted environments and enrichment cultures. *Environ. Sci. Technol.* **44**, 7287–7294 (2010).
14. Callaghan, A. V., Tierney, M., Phelps, C. D. & Young, L. Y. Anaerobic biodegradation of n-hexadecane by a nitrate-reducing consortium. *Appl. Environ. Microbiol.* **75**, 1339–1344 (2009).
15. Beller, H. R. & Edwards, E. A. Anaerobic toluene activation by benzylsuccinate synthase in a highly enriched methanogenic culture. *Appl. Environ. Microbiol.* **66**, 5503–5505 (2000).
16. von Netzer, F. *et al.* Enhanced gene detection assays for fumarate-adding enzymes allow uncovering of anaerobic hydrocarbon degraders in terrestrial and marine systems. *Appl. Environ. Microbiol.* **73**, 543–552 (2013).
17. Acosta-González, A., Rosselló-Móra, R. & Marqués, S. Diversity of benzylsuccinate synthase-like (bssA) genes in hydrocarbon-polluted marine sediments suggests substrate-dependent clustering. *Appl. Environ. Microbiol.* **79**, 3667–3676 (2013).
18. Callaghan, A. V., Gieg, L. M., Kropp, K. G., Suflita, J. M. & Young, L. Y. Comparison of mechanisms of alkane metabolism under sulfate-reducing conditions among two bacterial isolates and a bacterial consortium. *Appl. Environ. Microbiol.* **72**, 4274–4282 (2006).
19. Callaghan, A. V. & Wawrik, B. *AnHyDeg: A Curated Database of Anaerobic Hydrocarbon Degradation Genes.* (2016).
20. Aitken, C. M. *et al.* Evidence that crude oil alkane activation proceeds by different mechanisms under sulfate-reducing and methanogenic conditions. *Geochim. Cosmochim. Acta* **109**, 162–174 (2013).
21. Tan, B., Dong, X., Sensen, C. W. & Foght, J. Metagenomic analysis of an anaerobic alkane-degrading microbial culture: Potential hydrocarbon-activating pathways and inferred roles of community members. *Genome* **56**, 599–611 (2013).
22. Berdugo-Clavijo, C. & Gieg, L. M. Conversion of crude oil to methane by a microbial consortium enriched from oil reservoir production waters. *Front. Microbiol.* **5**, 197 (2014).

23. An, D. *et al.* Metagenomics of hydrocarbon resource environments indicates aerobic taxa and genes to be unexpectedly common. *Environ. Sci. Technol.* **47**, 10708–10717 (2013).
24. Johnson, J. M., Wawrik, B., Isom, C., Boling, W. B. & Callaghan, A. V. Interrogation of Chesapeake Bay sediment microbial communities for intrinsic alkane-utilizing potential under anaerobic conditions. *FEMS Microbiol. Ecol.* **91**, 1–14 (2015).
25. Gieg, L. M. & Toth, C. R. A. Signature Metabolite Analysis to Determine In Situ Anaerobic Hydrocarbon Biodegradation. in *Anaerobic Utilization of Hydrocarbons, Oils, and Lipids* (2020). doi:10.1007/978-3-319-50391-2_19.
26. Zhou, J. *et al.* Synthesis and Characterization of Anaerobic Degradation Biomarkers of *n* -Alkanes via Hydroxylation/Carboxylation Pathways. *Eur. J. Mass Spectrom.* **22**, 31–37 (2016).
27. Meckenstock, R. U. & Mouttaki, H. Anaerobic degradation of non-substituted aromatic hydrocarbons. *Curr. Opin. Biotechnol.* **22**, 406–414 (2011).
28. Yin, S., Tao, X. & Shi, K. The role of surfactants in coal bio-solubilisation. *Fuel Process. Technol.* **92**, 1554–1559 (2011).
29. Makkar, R. S. & Rockne, K. J. Comparison of synthetic surfactants and biosurfactants in enhancing biodegradation of polycyclic aromatic hydrocarbons. *Environ. Toxicol. Chem.* **22**, 2280–2292 (2003).
30. Barin, R., Talebi, M., Biria, D. & Beheshti, M. Fast bioremediation of petroleum-contaminated soils by a consortium of biosurfactant/bioemulsifier producing bacteria. *Int. J. Environ. Sci. Technol.* **11**, 1701–1710 (2014).
31. Gojgic-Cvijovic, G. D. *et al.* Biodegradation of petroleum sludge and petroleum polluted soil by a bacterial consortium: A laboratory study. *Biodegradation* **23**, 1–14 (2012).
32. Sekhon, K. K., Khanna, S. & Cameotra, S. S. Enhanced biosurfactant production through cloning of three genes and role of esterase in biosurfactant release. *Microb. Cell Fact.* **10**, 49 (2011).
33. Wentzel, A., Ellingsen, T. E., Kotlar, H. K., Zotchev, S. B. & Throne-Holst, M. Bacterial metabolism of long-chain n-alkanes. *Appl. Microbiol. Biotechnol.* **76**, 1209–1221 (2007).
34. Wawrik, B. *et al.* Field and laboratory studies on the bioconversion of coal to methane in the San Juan Basin. *FEMS Microbiol. Ecol.* **81**, 26–42 (2012).
35. Ulrich, G. & Bower, S. Active methanogenesis and acetate utilization in Powder River Basin coals, United States. *Int. J. Coal Geol.* **76**, 25–33 (2008).
36. Barnhart, E., Hyatt, R., Fields, M. W. & Cunningham, A. B. Subsurface environment sampler with actuator movable collection chamber. (2020).
37. Couradeau, E. *et al.* Probing the active fraction of soil microbiomes using BONCAT-FACS. *Nat. Commun.* **10**, (2019).
38. Hatzenpichler, R. *et al.* Visualizing in situ translational activity for identifying and sorting slow-growing archaeal−bacterial consortia. *Proc. Natl. Acad. Sci.* **113**, E4069–E4078 (2016).
39. Plugge, C. M., Zhang, W., Scholten, J. C. M. & Stams, A. J. M. Metabolic flexibility of sulfate-reducing bacteria. *Front. Microbiol.* **2**, (2011).
40. Beckmann, S. *et al.* Long-term succession in a coal seam microbiome during in situ biostimulation of coalbed-methane generation. *ISME J.* **13**, 632–650 (2019).
41. Kleindienst, S. *et al.* Diverse sulfate-reducing bacteria of the Desulfosarcina/Desulfococcus clade are the key alkane degraders at marine seeps. *ISME J.* **8**, 2029–2044 (2014).
42. Shin, B. *et al.* Anaerobic degradation of hexadecane and phenanthrene coupled to

sulfate reduction by enriched consortia from northern Gulf of Mexico seafloor sediment. *Sci. Rep.* **9**, (2019).

43. Sanford, R. A., Cole, J. R. & Tiedje, J. M. Characterization and description of Anaeromyxobacter dehalogenans gen. nov., sp. nov., an Aryl-halorespiring facultative anaerobic myxobacterium. *Appl. Environ. Microbiol.* **68**, 893–900 (2002).
44. Vinson, D. S., Blair, N. E., Ritter, D. J., Martini, A. M. & McIntosh, J. C. Carbon mass balance, isotopic tracers of biogenic methane, and therole of acetate in coal beds: Powder River Basin (USA). *Chem. Geol.* **530**, 119329 (2019).
45. Jones, E. J. P., Voytek, M. A., Corum, M. D. & Orem, W. H. Stimulation of methane generation from nonproductive coal by addition of nutrients or a microbial consortium. *Appl. Environ. Microbiol.* **76**, 7013–7022 (2010).
46. Strapoć, D. *et al.* Biogeochemistry of microbial coal-bed methane. *Annu. Rev. Earth Planet. Sci.* **39**, (2011).
47. Stewart, C. R., Burnside, D. M. & Cianciotto, N. P. The surfactant of legionella pneumophila is secreted in a TolC-dependent manner and is antagonistic toward other legionella species. *J. Bacteriol.* **193**, 5971–5984 (2011).
48. Suthar, H. & Nerurkar, A. Characterization of Biosurfactant Produced by Bacillus licheniformis TT42 Having Potential for Enhanced Oil Recovery. *Appl. Biochem. Biotechnol.* **180**, 248–260 (2016).
49. Kumar, A. P. *et al.* Evaluation of orange peel for biosurfactant production by Bacillus licheniformis and their ability to degrade naphthalene and crude oil. *3 Biotech* **6**, 43 (2016).
50. McIntosh, J. A., Donia, M. S. & Schmidt, E. W. Ribosomal peptide natural products: Bridging the ribosomal and nonribosomal worlds. *Nat. Prod. Rep.* **26**, 537–559 (2009).
51. Strieker, M., Tanović, A. & Marahiel, M. A. Nonribosomal peptide synthetases: Structures and dynamics. *Curr. Opin. Struct. Biol.* **20**, 234–240 (2010).
52. Yu, H., Huang, G. H., Xiao, H., Wang, L. & Chen, W. Combined effects of DOM and biosurfactant enhanced biodegradation of polycylic armotic hydrocarbons (PAHs) in soil-water systems. *Environ. Sci. Pollut. Res.* **21**, 10536–10549 (2014).
53. Zhao, F. *et al.* Heterologous production of Pseudomonas aeruginosa rhamnolipid under anaerobic conditions for microbial enhanced oil recovery. *J. Appl. Microbiol.* **118**, 379–389 (2015).
54. Singh, D. N. & Tripathi, A. K. Coal induced production of a rhamnolipid biosurfactant by Pseudomonas stutzeri, isolated from the formation water of Jharia coalbed. *Bioresour. Technol.* **128**, 215–221 (2013).
55. Ross, D. E. & Gulliver, D. Reconstruction of a Nearly Complete *Pseudomonas* Draft Genome Sequence from a Coalbed Methane-Produced Water Metagenome. *Genome Announc.* **4**, e01024-16 (2016).
56. Schleinitz, K. M. *et al.* Phenol degradation in the strictly anaerobic iron-reducing bacterium Geobacter metallireducens GS-15. *Appl. Environ. Microbiol.* **75**, 3912–3919 (2009).
57. Wilkes, H., Buckel, W., Golding, B. T. & Rabus, R. Metabolism of Hydrocarbons in n-Alkane-Utilizing Anaerobic Bacteria. *J. Mol. Microbiol. Biotechnol.* **26**, 138–151 (2016).
58. Duarte, M., Jauregui, R., Vilchez-Vargas, R., Junca, H. & Pieper, D. H. AromaDeg, a novel database for phylogenomics of aerobic bacterial degradation of aromatics. *Database (Oxford)* doi:10.1093/database/bau118 (2014) doi:10.1093/database/bau118.
59. Schmeling, S. & Fuchs, G. Anaerobic metabolism of phenol in proteobacteria and further studies of phenylphosphate carboxylase. *Arch. Microbiol.* **191**, 869–878 (2009).
60. Liang, B. *et al.* Anaerolineaceae and Methanosaeta turned to be the dominant

microorganisms in alkanes-dependent methanogenic culture after long-term of incubation. *AMB Express* **5**, (2015).

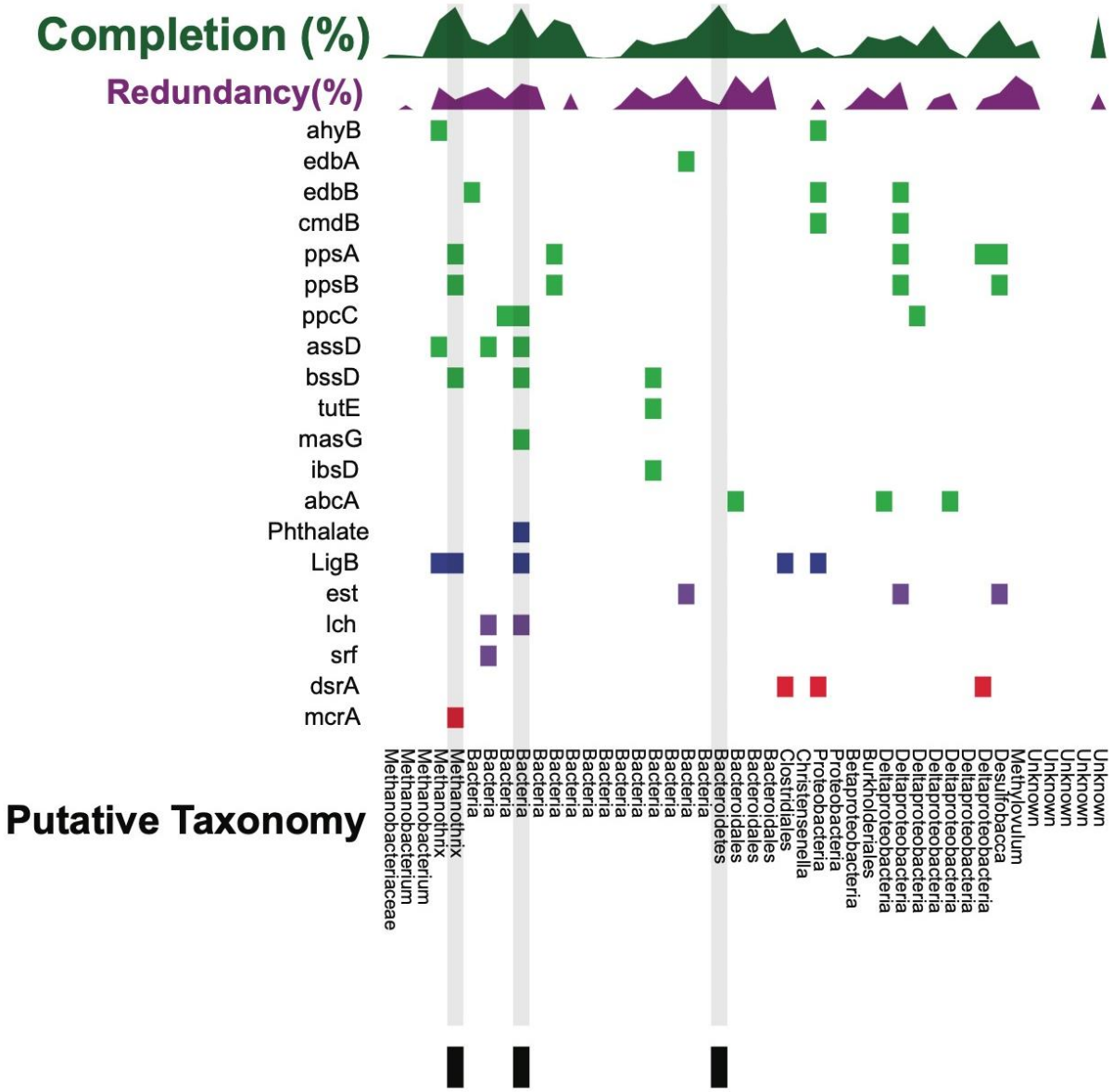
61. Momper, L., Jungbluth, S. P., Lee, M. D. & Amend, J. P. Energy and carbon metabolisms in a deep terrestrial subsurface fluid microbial community. *ISME J.* **11**, 2319–2333 (2017).
62. Lemos, L. N., Mendes, L. W., Baldrian, P. & Pylro, V. S. Genome-Resolved Metagenomics Is Essential for Unlocking the Microbial Black Box of the Soil. *Trends in Microbiology* vol. 29 279–282 (2021).
63. Schweitzer, H. *et al.* Innovating carbon-capture biotechnologies through ecosystem-inspired solutions. *One Earth* **4**, 49–59 (2021).
64. da Cruz, G. F. *et al.* Could petroleum biodegradation be a joint achievement of aerobic and anaerobic microrganisms in deep sea reservoirs? *AMB Express* **1**, (2011).
65. Suenaga, H., Ohnuki, T. & Miyazaki, K. Functional screening of a metagenomic library for genes involved in microbial degradation of aromatic compounds. *Environ. Microbiol.* **9**, 2289–2297 (2007).
66. Terrón-González, L., Martín-Cabello, G., Ferrer, M. & Santero, E. Functional metagenomics of a biostimulated petroleum-contaminated soil reveals an extraordinary diversity of extradiol dioxygenases. *Appl. Environ. Microbiol.* **82**, 2467–2478 (2016).
67. Stepanauskas, R. *et al.* Improved genome recovery and integrated cell-size analyses of individual uncultured microbial cells and viral particles. *Nat. Commun.* **8**, (2017).
68. Campbell, J. H. *et al.* UGA is an additional glycine codon in uncultured SR1 bacteria from the human microbiota. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 5540–5545 (2013).
69. Rinke, C. *et al.* Insights into the phylogeny and coding potential of microbial dark matter. *Nature* **499**, 431–437 (2013).

Supplemental Information Contents:

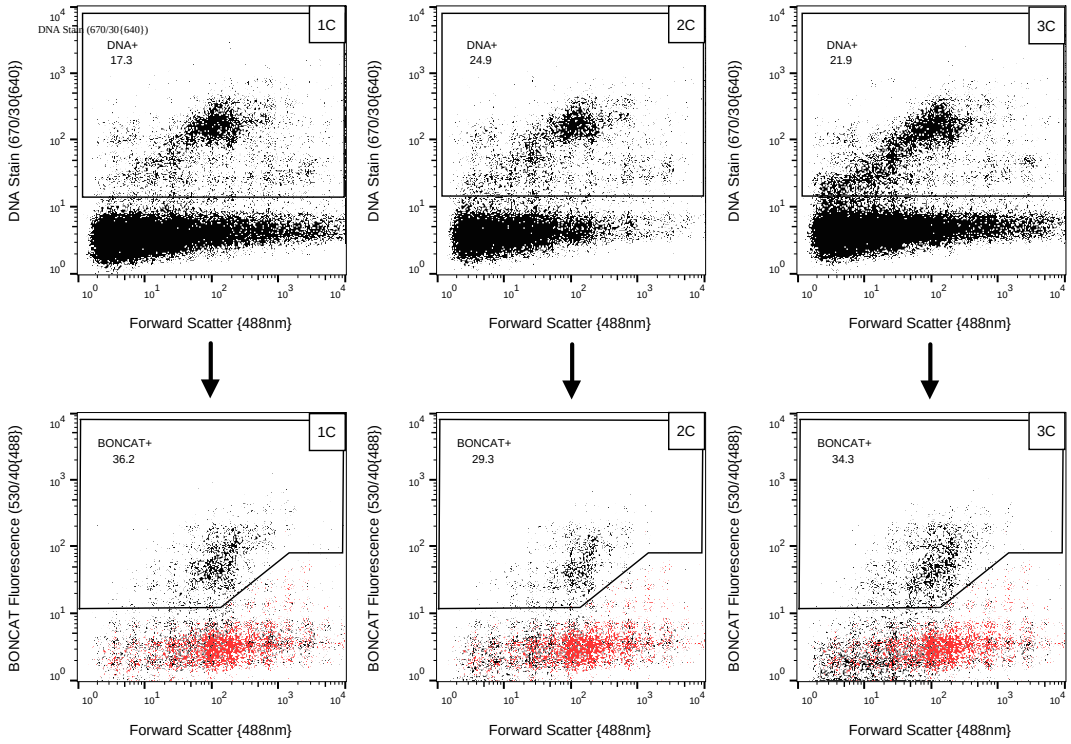
Supplemental Figures S1-S2

Supplemental Tables 1-4 and Supplemental Data 1-2

Supplemental Figures:



Supplemental Figure S1.) Evaluation of SYTO-Total metagenome assembled genomes (MAGs) and the presence or absence of anaerobic (green) and aerobic hydrocarbon degrading (blue), biosurfactant (purple) and sulfate reduction/methanogenesis (red) genes for each corresponding MAG. The AnHyDeg and AromaDeg databases were used to determine the different hydrocarbon degradation genes used for comparison. The percent completion (green line graph out of 100%) and redundancy (purple line graph out of 10%) were compared for each MAG. MAGs with putative taxonomy Methanothrix, Bacteria and Bacteriodetes were all greater than 92% complete and contain an average nucleotide identity that is >99% similar to their matching BONCAT MAGs which is represented by the symbol (■).



Supplemental Figure S2.) Gate drawing accomplished in two steps for the biological replicates of FG-11 BONCAT positive samples. The top panel 1A, 2A, and 3A is gated to separate total cells from background particles and gates indicated in black are based on a DNA stain (Syto59, ex: 640nm em: 655-685nm). Following the initial gating, SYTO+ cells were analyzed further for BONCAT fluorescence with the FAM (Picolyl dye (Ex: 488nm/Em: 530nm) (bottom panel). Gate for 1B, 2B, and 3B. Gating of the BONCAT+ populations was achieved by comparing each replicate to both HPG negative and water controls (indicated in red overlaid in each graph). The number in the top left of each of the bottom panel graphs indicates the percentage of BONCAT+ cells in comparison to total cells (SYTO+) for each replicate.

Supplemental Tables/Data:

***Uploaded as a data file with table explanations provided below**

Supplemental Table S1.) List of environmental metagenome assembled genomes (MAGs) from high sulfate and low sulfate coal seams from the Powder River Basin and the corresponding sequencing and analysis parameters.

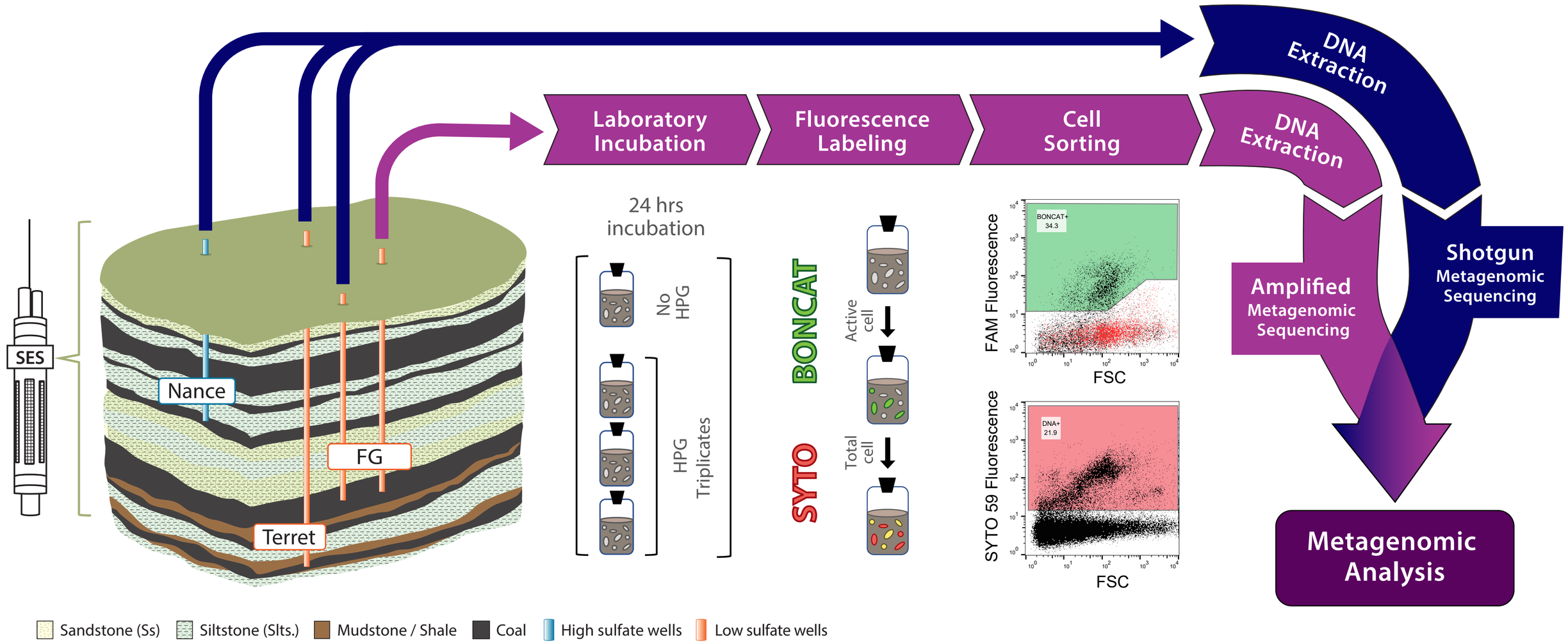
Supplemental Table 2.) Complete list of the four databases and all the genes compared in this study. The function of each gene and the coverage of each gene in each sample is listed and the presence of gene in the BONCAT+ sample is indicated. The genes further analyzed are marked as genes of interest.

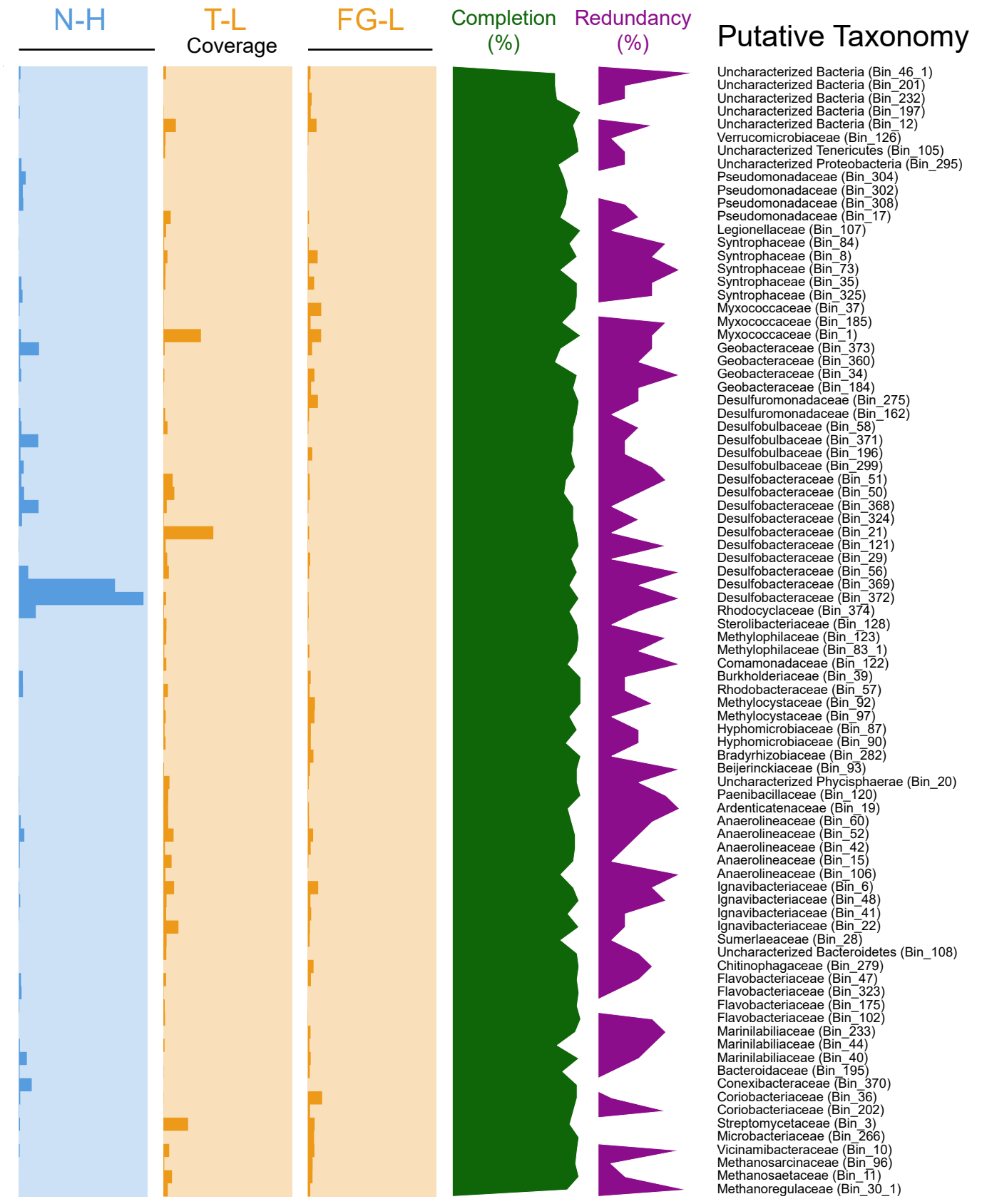
Supplemental Table 3.) List of BONCAT+ metagenome assembled genomes (MAGs) and the corresponding sequencing and analysis parameters. The corresponding genes present in each MAG are listed under their corresponding database.

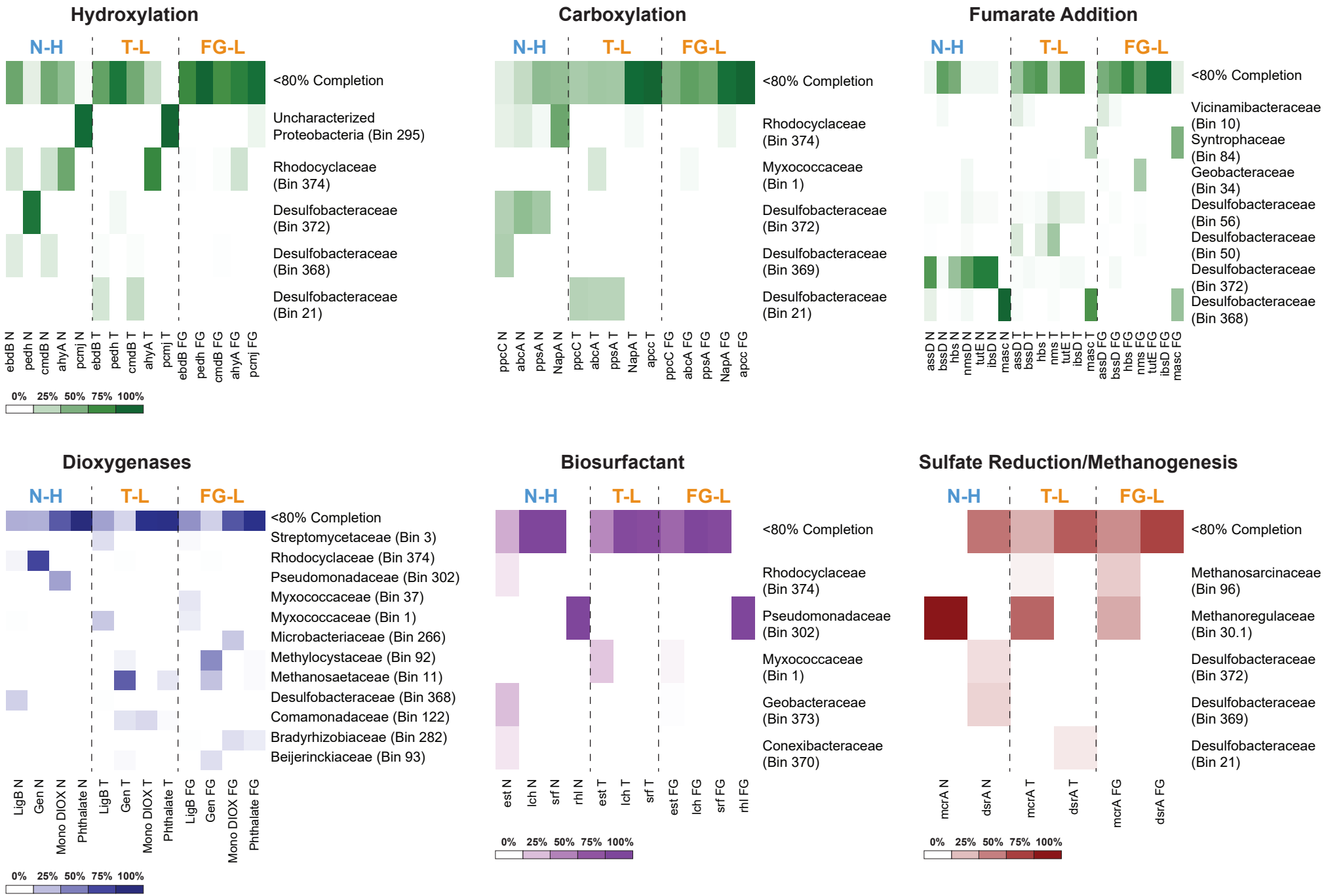
Supplemental Table 4.) List of SYTO-Total metagenome assembled genomes (MAGs) and there corresponding sequencing and analysis parameters. The corresponding genes that are present in each MAG are listed under their corresponding database.

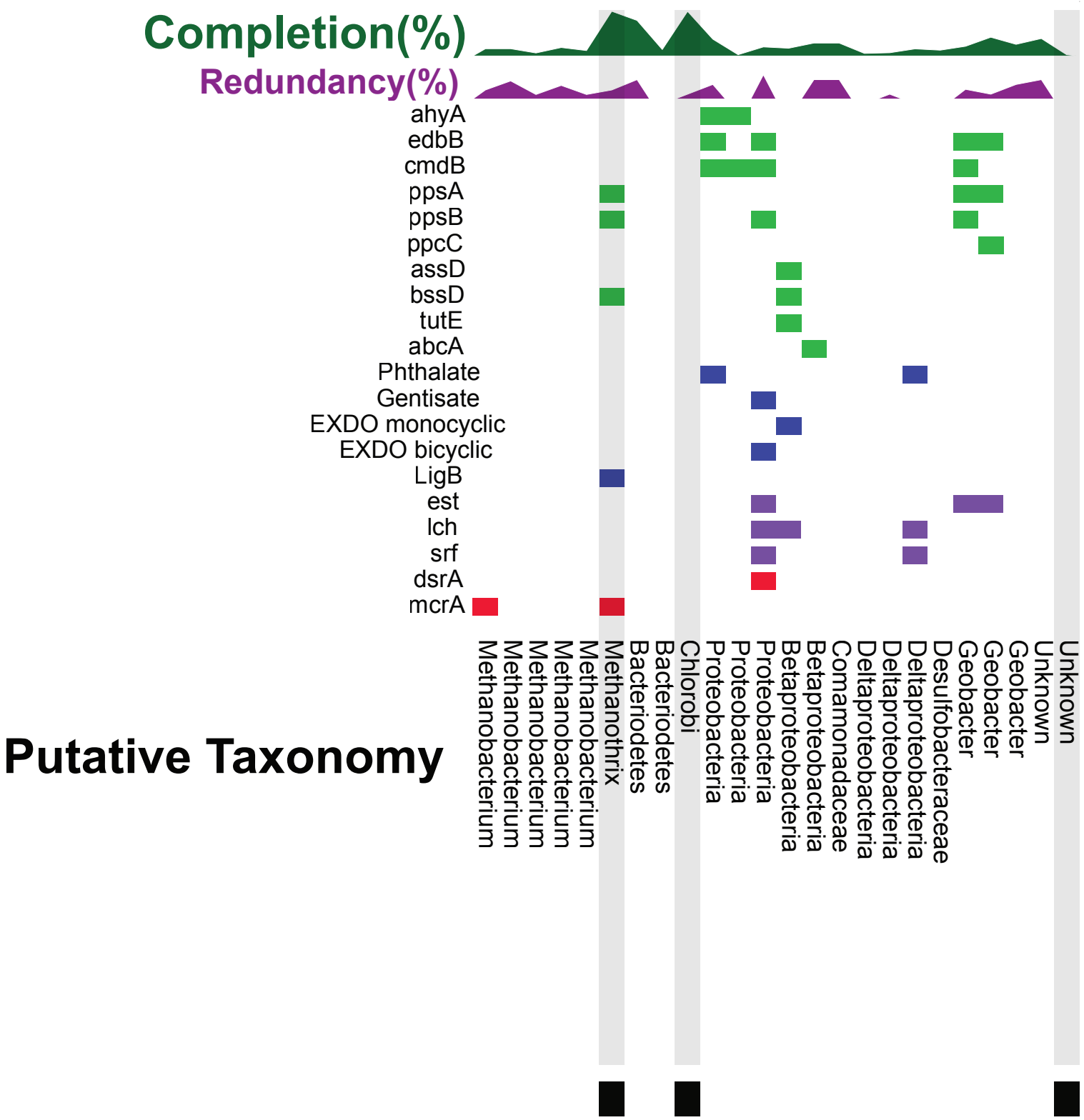
Supplemental Data 1.) Relative abundance of individual genes of interest involved in hydrocarbon degradation for the shotgun environmental MAGs from the high-sulfate coal seam (N-H) and two low-sulfate coal seams (T-L and FG-L). All MAGs that were below 80% completion were grouped together as <80% Completion.

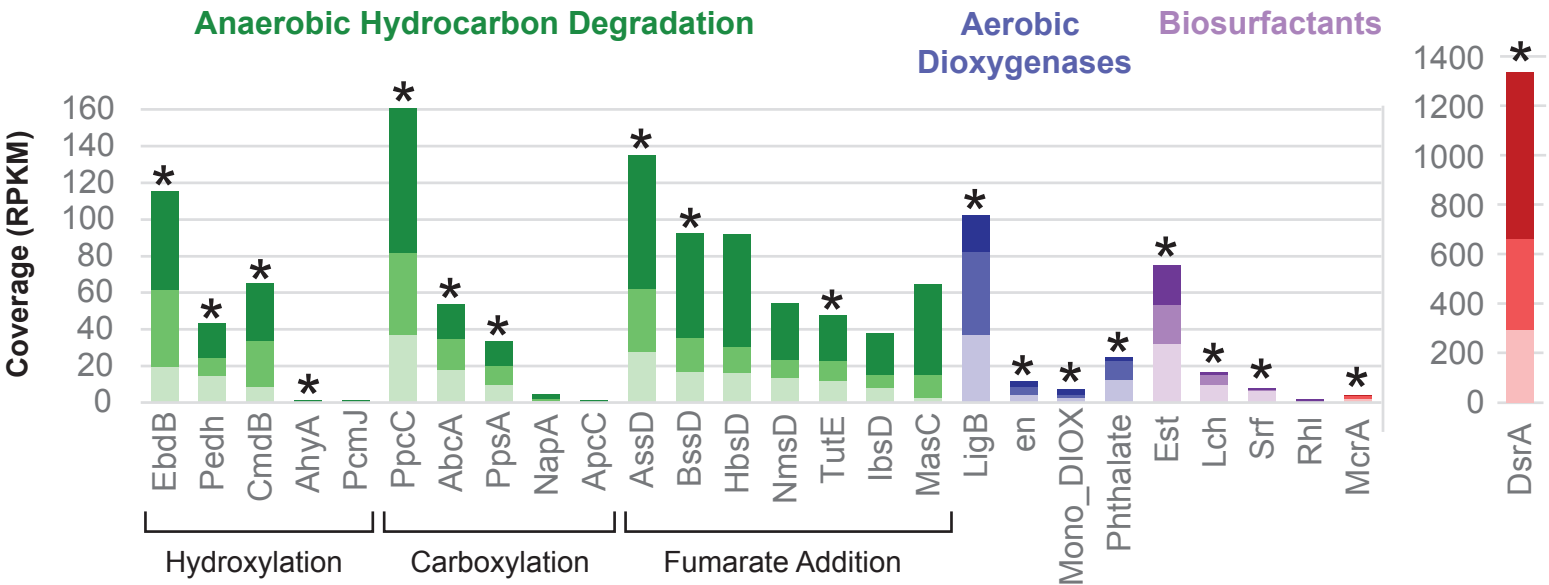
Supplemental Data 2.) Average Nucleotide Identity comparison of all shotgun, BONCAT+ and SYTO-Total metagenomes.


SES
Nance
FG
Terret
Laboratory Incubation
Fluorescence Labeling
Cell Sorting
DNA Extraction
DNA Extraction
24 hrs incubation
No HPG
HPG Triplicates
BONCAT
SYTO
Active cell
Total cell
FAM Fluorescence
FSC
BONCAT+ 34.3
SYTO 59 Fluorescence
FSC
DNA+ 21.9
Amplified Metagenomic Sequencing
Shotgun Metagenomic Sequencing
Metagenomic Analysis
Sandstone (Ss)
Siltstone (Slts.)
Mudstone / Shale
Coal
High sulfate wells
Low sulfate wells


N-H
T-L
Coverage
FG-L
Completion (%)
Redundancy (%)
Putative Taxonomy
Uncharacterized Bacteria (Bin_46_1)
Uncharacterized Bacteria (Bin_201)
Uncharacterized Bacteria (Bin_232)
Uncharacterized Bacteria (Bin_197)
Uncharacterized Bacteria (Bin_12)
Verrucomicrobiaceae (Bin_126)
Uncharacterized Tenericutes (Bin_105)
Uncharacterized Proteobacteria (Bin_295)
Pseudomonadaceae (Bin_304)
Pseudomonadaceae (Bin_302)
Pseudomonadaceae (Bin_308)
Pseudomonadaceae (Bin_17)
Legionellaceae (Bin_107)
Syntrophaceae (Bin_84)
Syntrophaceae (Bin_8)
Syntrophaceae (Bin_73)
Syntrophaceae (Bin_35)
Syntrophaceae (Bin_325)
Myxococcaceae (Bin_37)
Myxococcaceae (Bin_185)
Myxococcaceae (Bin_1)
Geobacteraceae (Bin_373)
Geobacteraceae (Bin_360)
Geobacteraceae (Bin_34)
Geobacteraceae (Bin_184)
Desulfuromonadaceae (Bin_275)
Desulfuromonadaceae (Bin_162)
Desulfobulbaceae (Bin_58)
Desulfobulbaceae (Bin_371)
Desulfobulbaceae (Bin_196)
Desulfobulbaceae (Bin_299)
Desulfobacteraceae (Bin_51)
Desulfobacteraceae (Bin_50)
Desulfobacteraceae (Bin_368)
Desulfobacteraceae (Bin_324)
Desulfobacteraceae (Bin_21)
Desulfobacteraceae (Bin_121)
Desulfobacteraceae (Bin_29)
Desulfobacteraceae (Bin_56)
Desulfobacteraceae (Bin_369)
Desulfobacteraceae (Bin_372)
Rhodocyclaceae (Bin_374)
Sterolibacteriaceae (Bin_128)
Methylophilaceae (Bin_123)
Methylophilaceae (Bin_83_1)
Comamonadaceae (Bin_122)
Burkholderiaceae (Bin_39)
Rhodobacteraceae (Bin_57)
Methylocystaceae (Bin_92)
Methylocystaceae (Bin_97)
Hyphomicrobiaceae (Bin_87)
Hyphomicrobiaceae (Bin_90)
Bradyrhizobiaceae (Bin_282)
Beijerinckiaceae (Bin_93)
Uncharacterized Phycisphaerae (Bin_20)
Paenibacillaceae (Bin_120)
Ardenticatenaceae (Bin_19)
Anaerolineaceae (Bin_60)
Anaerolineaceae (Bin_52)
Anaerolineaceae (Bin_42)
Anaerolineaceae (Bin_15)
Anaerolineaceae (Bin_106)
Ignavibacteriaceae (Bin_6)
Ignavibacteriaceae (Bin_48)
Ignavibacteriaceae (Bin_41)
Ignavibacteriaceae (Bin_22)
Sumerlaeaceae (Bin_28)
Uncharacterized Bacteroidetes (Bin_108)
Chitinophagaceae (Bin_279)
Flavobacteriaceae (Bin_47)
Flavobacteriaceae (Bin_323)
Flavobacteriaceae (Bin_175)
Flavobacteriaceae (Bin_102)
Marinilabiliaceae (Bin_233)
Marinilabiliaceae (Bin_44)
Marinilabiliaceae (Bin_40)
Bacteroidaceae (Bin_195)
Conexibacteraceae (Bin_370)
Coriobacteriaceae (Bin_36)
Coriobacteriaceae (Bin_202)
Streptomycetaceae (Bin_3)
Microbacteriaceae (Bin_266)
Vicinamibacteraceae (Bin_10)
Methanosarcinaceae (Bin_96)
Methanosaetaceae (Bin_11)
Methanoregulaceae (Bin_30_1)


Hydroxylation
N-H
T-L
FG-L
<80% Completion
Uncharacterized Proteobacteria (Bin 295)
Rhodocyclaceae (Bin 374)
Desulfobacteraceae (Bin 372)
Desulfobacteraceae (Bin 368)
Desulfobacteraceae (Bin 21)
ebdB N
pedh N
cmdB N
ahyA N
pcmj N
ebdB T
pedh T
cmdB T
ahyA T
pcmj T
ebdB FG
pedh FG
cmdB FG
ahyA FG
pcmj FG
0% 25% 50% 75% 100%
Carboxylation
N-H
T-L
FG-L
<80% Completion
Rhodocyclaceae (Bin 374)
Myxococcaceae (Bin 1)
Desulfobacteraceae (Bin 372)
Desulfobacteraceae (Bin 369)
Desulfobacteraceae (Bin 21)
ppcC N
abcA N
ppsA N
NapA N
ppcC T
abcA T
ppsA T
NapA T
apcc T
ppcC FG
abcA FG
ppsA FG
NapA FG
apcc FG
Fumarate Addition
N-H
T-L
FG-L
<80% Completion
Vicinamibacteraceae (Bin 10)
Syntrophaceae (Bin 84)
Geobacteraceae (Bin 34)
Desulfobacteraceae (Bin 56)
Desulfobacteraceae (Bin 50)
Desulfobacteraceae (Bin 372)
Desulfobacteraceae (Bin 368)
assD N
bssD N
hbs N
nmsD N
tutE N
ibsD N
masc N
assD T
bssD T
hbs T
nms T
tutE T
ibsD T
masc T
assD FG
bssD FG
hbs FG
nms FG
tutE FG
ibsD FG
masc FG
Dioxygenases
N-H
T-L
FG-L
<80% Completion
Streptomycetaceae (Bin 3)
Rhodocyclaceae (Bin 374)
Pseudomonadaceae (Bin 302)
Myxococcaceae (Bin 37)
Myxococcaceae (Bin 1)
Microbacteriaceae (Bin 266)
Methylocystaceae (Bin 92)
Methanosaetaceae (Bin 11)
Desulfobacteraceae (Bin 368)
Comamonadaceae (Bin 122)
Bradyrhizobiaceae (Bin 282)
Beijerinckiaceae (Bin 93)
LigB N
Gen N
Mono DIOX N
Phthalate N
LigB T
Gen T
Mono DIOX T
Phthalate T
LigB FG
Gen FG
Mono DIOX FG
Phthalate FG
0% 25% 50% 75% 100%
Biosurfactant
N-H
T-L
FG-L
<80% Completion
Rhodocyclaceae (Bin 374)
Pseudomonadaceae (Bin 302)
Myxococcaceae (Bin 1)
Geobacteraceae (Bin 373)
Conexibacteraceae (Bin 370)
est N
lch N
srf N
rhl N
est T
lch T
srf T
est FG
lch FG
srf FG
rhl FG
0% 25% 50% 75% 100%
Sulfate Reduction/Methanogenesis
N-H
T-L
FG-L
<80% Completion
Methanosarcinaceae (Bin 96)
Methanoregulaceae (Bin 30.1)
Desulfobacteraceae (Bin 372)
Desulfobacteraceae (Bin 369)
Desulfobacteraceae (Bin 21)
mcrA N
dsrA N
mcrA T
dsrA T
mcrA FG
dsrA FG
0% 25% 50% 75% 100%


Completion(%)
Redundancy(%)
ahyA
edbB
cmdB
ppsA
ppsB
ppcC
assD
bssD
tutE
abcA
Phthalate
Gentisate
EXDO monocyclic
EXDO bicyclic
LigB
est
lch
srf
dsrA
mcrA
Putative Taxonomy
Methanobacterium
Methanobacterium
Methanobacterium
Methanobacterium
Methanobacterium
Methanobacterium
Methanothrix
Bacteriodetes
Bacteriodetes
Chlorobi
Proteobacteria
Proteobacteria
Proteobacteria
Betaproteobacteria
Betaproteobacteria
Comamonadaceae
Deltaproteobacteria
Deltaproteobacteria
Deltaproteobacteria
Desulfobacteraceae
Geobacter
Geobacter
Geobacter
Geobacter
Unknown
Unknown


Anaerobic Hydrocarbon Degradation
Aerobic Dioxygenases
Biosurfactants
Coverage (RPKM)
160
140
120
100
80
60
40
20
0
1400
1200
1000
800
600
400
200
0
EbdB
Pedh
CmdB
AhyA
PcmJ
PpcC
AbcA
PpsA
NapA
ApcC
AssD
BssD
HbsD
NmsD
TutE
IbsD
MasC
LigB
en
Mono_DIOX
Phthalate
Est
Lch
Srf
Rhl
McrA
DsrA
Hydroxylation
Carboxylation
Fumarate Addition