Potential for microbial anaerobic hydrocarbon degradation in naturally occurring petroleum-associated deep-sea sediments

Xiyang Dong [1,*], Chris Greening [2], Jayne E. Rattray [1], Anirban Chakraborty [1], Maria Chuvochina[2,3], Daisuke Mayumi [1,4], Carmen Li [1], James M. Brooks [5], Bernie B. Bernard [5], Ryan A. Groves[1], Ian A. Lewis [1], Casey R.J. Hubert [1,*]

[1] Department of Biological Sciences, University of Calgary, Calgary, T2N 1N4, Canada

[2] School of Biological Sciences, Monash University, Clayton, VIC 3800, Australia

[3] Australian Centre for Ecogenomics, School of Chemistry and Molecular Biosciences, The University of Queensland, QLD 4072, Australia

[4] Institute for Geo-Resources and Environment, Geological Survey of Japan, National Institute of Advanced Industrial Science and Technology (AIST), 1-1-1 Higashi, Tsukuba, 305-8567, Japan

[5] TDI Brooks International, College Station, Texas, TX 77845, USA

* Corresponding author. E-mail: xiyang.dong@ucalgary.ca.

* Corresponding author. E-mail: chubert@ucalgary.ca. Telephone: +1 (403) 220-7794.

The authors declare no conflict of interest.

Abstract

The lack of cultured isolates and microbial genomes from the deep seabed means that very little is known about the ecology of this vast habitat. Here, we investigated energy and carbon acquisition strategies of microbial communities from three deep seabed petroleum seeps (3 km water depth) in the Eastern Gulf of Mexico. Shotgun metagenomic analysis revealed that each sediment harbored diverse communities of chemoheterotrophs and chemolithotrophs. We recovered 82 metagenome-assembled genomes affiliated with 21 different archaeal and bacterial phyla. Multiple genomes encoded enzymes for acetogenic fermentation of aliphatic and aromatic compounds, specifically of candidate phyla *Aerophobetes*, *Aminicenantes*, TA06 and *Bathyarchaeota*. Microbial interactions in these communities are predicted to be driven by acetate and molecular hydrogen, as indicated by a high abundance of fermentation, acetogenesis, and hydrogen utilization pathways. These findings are supported by sediment geochemistry, metabolomics and thermodynamic modelling of hydrocarbon degradation. Overall, we infer that deep-sea sediments experiencing thermogenic hydrocarbon inputs harbor phylogenetically and functionally diverse communities potentially sustained through anaerobic hydrocarbon, acetate and hydrogen metabolism.

Deep-sea sediments, generally understood to be those occurring in water depths greater than ~500 meters, represent one of the largest habitats on Earth. In recent years, culture-independent 16S rRNA gene surveys and metagenomic studies have revealed these sediments host a vast abundance and diversity of bacteria and archaea [1-8]. Cell numbers decrease with sediment depth and age, from between 10^6 and 10^{10} cm^{-3} in the upper cm at the sediment-water interface to below 10^4 cm^{-3} several kilometers below the ocean floor [9, 10]. However, due to a lack of cultured representatives and genomes recovered from deep-sea sediments, it remains largely unresolved how microorganisms survive and function in these nutrient-limited ecosystems. Energy and carbon sources are essential requirements that allow the buried microorganisms to persist. With sunlight penetration not reaching the deep seabed, photosynthetic processes do not directly support these communities [11]. It has therefore been proposed that deep sea benthic and subsurface microbes are primarily sustained by complex detrital organic matter, including carbohydrates, proteinaceous compounds, and humic substances, derived from the overlying water column via sedimentation [11-13].

Another important potential carbon and energy source in deep-sea sediments are petroleum geofluids that migrate from subseafloor reservoirs up to the seafloor [14]. Petroleum compounds include smaller gaseous molecules, such as methane, propane and butane, and larger aliphatic and aromatic liquids. Numerous studies have investigated the role of methane oxidation in seabed sediments, which is mediated by anaerobic methanotrophic archaea (ANME), generally in syntrophy with bacteria respiring sulfate or other electron acceptors [4, 6, 8, 15, 16]. In contrast, little is known about the degradation of larger alkanes or aromatic compounds by deep seabed microorganisms. Vigneron et al. [2] performed a comparative gene-centric study of hydrocarbon and methane seeps of the Gulf of Mexico, and suggested that microorganisms in deep cold seeps

(water depth ~1 km) can potentially utilize a range of non-methane hydrocarbons. However, due to the absence of metagenome binning in that study, relevant metabolic functions were not assigned to specific pathways or taxa.

In addition to organic carbon compounds, microbial life in deep-sea sediments is also supported by inorganic electron donors. Some microorganisms have been isolated from deep sediments that are able to sustain themselves by oxidizing elemental sulfur, hydrogen sulfide, carbon monoxide, ammonia and molecular hydrogen (H_2) [6, 8, 11]. Of these, H_2 is a particularly important energy source given its production in large quantities by biological and geochemical processes. H_2 can be generated as a metabolic byproduct of fermentation, together with volatile fatty acids such as acetate, during organic matter degradation [9, 17]. H_2 can also be produced abiotically via serpentinization, radiolysis of water, or thermal alteration of sedimentary organic matter [18]. For example, the radiolysis of water by naturally occurring radionuclides (*e.g.* ^{40}K and ^{238}U) is estimated to produce 10^{11} mol H_2 per year [8, 19]. Depending on the availability of electron acceptors, H_2 oxidation can be coupled to sulfate, nitrate, and metal respiration, acetogenesis and methanogenesis [8, 11].

To understand the role of hydrocarbon substrates and hydrogen in supporting microbial life in deep-sea sediments, we performed metagenomic, geochemical and metabolomic analyses of three deep seabed sediments (water depth ~3km). The three sites exhibited different levels of migrated thermogenic hydrocarbons. Metagenomes generated from sediment samples obtained from each site were assembled and binned to obtain metagenome-assembled genomes (MAGs) and to reconstruct metabolic pathways for dominant members of the microbial communities. Complementing this genome-resolved metagenomics, a gene-centric analysis was performed by

directly examining unassembled metagenomic data. Through the combination of metagenomics with geochemistry and metabolomics, with supporting thermodynamic modeling, we provide evidence that (1) deep-sea sediments harbor phylogenetically diverse heterotrophic and lithotrophic microbial communities; (2) some members from the candidate phyla are engaged in degradation of aliphatic and aromatic thermogenic hydrocarbons; and (3) microbial community members are likely interconnected via acetate and hydrogen metabolism.

Results

Sediment geochemistry

This study tested three petroleum-associated near-surface sediments (referred to as Sites E26, E29 and E44) sampled from the Eastern Gulf of Mexico [20]. Petroleum content and other geochemical characteristics were analyzed for each of the three sites (Table 1). All sites had high concentrations of aromatic compounds and liquid alkanes; aromatic compounds were most abundant at Site E26, while liquid alkanes were at 2.5-fold higher concentration at Sites E26 and E29 than Site E44. Alkane gases were only abundant at Site E29 and were almost exclusively methane (CH_4). CH_4 sources can be inferred from stable isotopic compositions of CH_4 and molar ratios of CH_4 to higher hydrocarbons [15]. Ratios of $C_1/(C_2+C_3)$ were greater than 1,000 and $\delta^{13}C$ values of methane were more negative than -60‰, indicating that the CH_4 in these sediments is predominantly biogenic [15, 21]. GC-MS revealed an unresolved complex mixture (UCM) of saturated hydrocarbons in the C_{15+} range in all three sites. Such UCM signals correspond to degraded petroleum hydrocarbons and may indicate the occurrence of oil biodegradation at these sites [22]. Signature metabolites for anaerobic biodegradation of alkanes and aromatic compounds [23] were also detected, including benzoate, toluate and methyl- or trimethylsilyl esters (Table S1).

High concentrations of sulfate (>20 mM) were detected at each of the three sites (Table 1), consistent with sulfate being present in high concentrations in seawater and diffusing into the sediments. H_2 and acetate concentrations were both below limits of detection (0.005-0.02 ppm and 2.5 µM, respectively) consistent with H_2 and acetate being present at extremely low steady-state concentrations in deep-sea sediments due to tight coupling between producers and consumers [3, 15].

Deep-sea sediments harbor phylogenetically diverse bacterial and archaeal communities

Illumina NextSeq sequencing of genomic DNA from deep-sea sediment communities produced 85,825,930, 148,908,270, and 138,795,692 quality-filtered reads for Sites E26, E29, and E44, respectively (Table S2). The 16S rRNA gene amplicon sequencing results suggest the sediments harbor diverse bacterial and archaeal communities, with Chao1 richness estimates of 359, 1375 and 360 using bacterial primers as well as 195, 180 and 247 using archaeal primers, for Sites E26, E29 and E44, respectively (Table S3 and Figure S1). Taxonomic profiling of these metagenomes using small subunit ribosomal RNA (SSU rRNA) marker genes reveals that the most abundant phyla in the metagenomes were, in decreasing order, *Chloroflexi* (mostly classes *Dehalococcoidia* and *Anaerolineae*), *Candidatus* Atribacteria, *Proteobacteria* (mostly class *Deltaproteobacteria*), and *Candidatus* Bathyarchaeota (Figure 1a). While the three sites share a broadly similar community composition, notable differences were *Ca.* Bathyarchaeota and *Proteobacteria* being in higher relative abundance at the sites with more hydrocarbons (E29 and E26; Table 1), whereas the inverse is true for *Actinobacteria*, the Patescibacteria group, and *Ca.* Aerophobetes that are all present in higher relative abundance at Site E44 where hydrocarbon

levels are lower. Additional sampling is required to determine whether these differences are due to the presence of hydrocarbons or other factors.

Assembly and binning for the three metagenomes resulted in a total of 82 MAGs with >50% completeness and <10% contamination based on CheckM analysis [24]. Reconstructed MAGs comprise taxonomically diverse members from a total of 6 archaeal and 15 bacterial phyla (Figure 2 and Table S4). Within the domain Bacteria, members of the phylum *Chloroflexi* are highly represented in each sample, especially from the classes *Dehalococcoidia* and *Anaerolineae*. Within the domain Archaea, members of phylum *Bathyarchaeota* were recovered from all three sites. Most other MAGs belong to poorly understood candidate phyla that lack cultured representatives, including *Aminicenantes* (formerly OP8), *Aerophobetes* (formerly CD12), *Cloacimonas* (formerly WWE1), *Stahlbacteria* (formerly WOR-3), *Atribacteria* (formerly JS1 and OP9), TA06 and the Asgard superphylum including *Lokiarchaeota*, *Thorarchaeota*, and *Heimdallarchaeota*.

Candidate phylum TA06, also known as GN04 or AC1, was originally discovered in a hypersaline microbial mat [25]. First genomic representatives of this phylum were recovered from estuarine sediments [26] with a small number of other MAGs recently reported to belong to this lineage [27, 28]. Due to the paucity of available MAGs and misclassifications based on 16S rRNA gene sequences, members of TA06 are often 'confused' with members of the phylum WOR-3 (*Stahlbacteria*) [28]. In addition to the phylogenetic inference here based on 43 concatenated protein marker genes (Figure 2), the placement of two bins within the original TA06 phylum is further supported by genome classification based on concatenation of 120 ubiquitous, single-

copy marker genes [29] as well as classification of 16S rRNA genes using the SILVA database [30] (Tables S4 and S5).

In summary, while there are considerable community-level differences between the three sample locations, the recovered MAGs share common taxonomic affiliations at the phylum and class levels. Guided by sediment geochemistry (Table 1), we subsequently analyzed the metabolic potential of these MAGs to understand how dominant bacterial and archaeal community members generate energy and biomass in these natural petroleum-associated deep-sea environments. Hidden Markov models (HMMs) and homology-based models were used to search for the presence of different metabolic genes both in the recovered MAGs and unbinned metagenomes. Where appropriate, findings were further validated through metabolomic analyses, phylogenetic visualization, and analysis of gene context.

Capacity for hydrocarbon and detrital biomass degradation in sediment microbial communities

In deep-sea marine sediments organic carbon is supplied either as detrital matter from the overlying water column or as aliphatic and aromatic petroleum compounds that migrate upwards from underlying petroleum-bearing sediments [11]. With respect to detrital matter, genes involved in carbon acquisition and breakdown were prevalent across both archaeal and bacterial MAGs. These include genes encoding intracellular and extracellular carbohydrate-active enzymes and peptidases, as well as relevant transporters and glycolysis enzymes (Figure 3 and Table S6). The importance of these carbon acquisition mechanisms is supported by the detection of corresponding intermediate metabolites, such as glucose and amino acids, in all three sediments (Table S1). The ability to break down fatty acids and other organic acids via the beta-oxidation

pathway was identified in 13 MAGs, including members of *Chloroflexi*, *Deltaproteobacteria*, *Aerophobetes* and *Lokiarchaeota* (Figure 3 and Table S6). These results align with many other studies suggesting that the majority of seabed microorganisms are involved in recycling of residual organic matter, including complex carbohydrates, proteins and lipids [13, 31, 32].

Unlike in other studies, the presence of petroleum hydrocarbons is a defining feature of the sediments investigated here; thus a key goal of this study was to identify the potential for microbial degradation of hydrocarbons as a source of energy and carbon. To this end, we focused on functional marker genes encoding enzymes that catalyze the activation of mechanistically sophisticated C-H bonds, to initiate hydrocarbon biodegradation [33]. For anaerobic hydrocarbon degradation, four oxygen-independent C-H activation reactions have been characterized: (1) addition of fumarate by glycyl-radical enzymes, *e.g.* for activation of alkylbenzenes and straight chain alkanes [34]; (2) hydroxylation with water by molybdenum cofactor-containing enzymes, *e.g.* for activation of ethylbenzene [33]; (3) carboxylation catalyzed by UbiD-like carboxylases, *e.g.* for activation of benzene and naphthalene [35]; and (4) reverse methanogenesis involving variants of methyl-coenzyme M reductase, *e.g.* for activation of methane and butane [36]. Most of the evidence for mechanisms 1-3 has come from studies of hydrocarbon contaminated aquifers, whereas mechanism 4 has been studied extensively in marine sediments [23, 37].

Evidence for glycyl radical enzymes that catalyze fumarate addition was found in 15 out of the 82 MAGs based on identifying genes encoding alkylsuccinate synthase (AssA) (Figures 3 and 4a). The *assA* sequences identified, while phylogenetically distant from canonical fumarate-adding enzymes and pyruvate formate lyases (Pfl), form a common clade with Pfl-like AssA from *Archaeoglobus fulgidus* VC-16 and *Abyssivirga alkaniphila* L81 (Figure 4a). Both of these

organisms have been shown experimentally to be capable of anaerobic alkane degradation [38, 39]. The putative *assA* genes identified here are present in all three samples regardless of hydrocarbon concentrations. They belong to MAGs affiliated with the bacterial phyla *Aerophobetes*, *Aminicenantes* and *Chloroflexi* as well as the archaeal phyla *Bathyarchaeota*, *Lokiarchaeota* and *Thorarchaeota*. The highest relative abundance of putative *assA* sequences was found in Site E29 as idnciated by quality-filtered reads, consistent with this sediment containing the highest concentration of aliphatic compounds (Tables 1 and S7). Additional searching for other genes encoding fumarate-adding enzymes in the quality-filtered reads (e.g. *bssA* and *nmsA*, and canonical *assA*) did not return significant counts, while some hits to canonical *assA* were indicated, suggesting it may also be possible existence in these habitats (Figure 4 and Table S7). Among the other three anaerobic hydrocarbon biodegradation mechanisms mentioned above, a MAG classified as *Dehalococcoidia* (*Chloroflexi* E29_bin2) contained genes encoding putative catalytic subunits of *p*-cymene dehydrogenase (Cmd) and alkane C_2-methylene hydroxylase (Ahy) (Figures 3 and S2), known to support *p*-cymene and alkane utilization [37]. Genes encoding enzymes catalyzing hydrocarbon carboxylation, reverse methanogenesis and aerobic hydrocarbon degradation (e.g. *alkB*, *nahC* and *nahG*) were not detected (Table S6). The latter result is expected due to the low concentrations of oxygen in the top 20 cm of organic rich seabed sediments.

Considering the degradation of aromatic hydrocarbons, genes responsible for reduction of benzoyl-CoA were detected in 12 MAGs (Figures 3 and 4b). Benzoyl-CoA is a universal biomarker for anaerobic degradation of monoaromatic compounds as it is a common intermediate to biochemical pathways catalyzing this process [40]. Benzoyl-CoA reduction to cyclohex-1,5-diene-1-carboxyl-CoA is performed by Class I ATP-dependent benzoyl-CoA

reductase (BCR; BcrABCD) in facultative anaerobes (*e.g. Thauera aromatica*) or Class II ATP-independent reductase (Bam; BamBCDEFGHI) in strict anaerobes like sulfate reducers [41]. The *bcr* genes detected are all Class I, and were found in bacterial MAGs (*i.e.*, *Dehalococcoidia*, *Anaerolineae*, *Deltaproteobacteria*, *Aminicenantes* and TA06) and archaeal MAGs (*i.e.*, *Thermoplasmata* and *Bathyarchaeota*) (Figures 3 and 4b). Genes for further transformation of dienoyl-CoA to 3-hydroxypimelyl-CoA were also identified (Figures 3 and 4b), *i.e.*, genes encoding for 6-oxo-cyclohex-1-ene-carbonyl-CoA hydrolase (Oah), cyclohex-1,5-diencarbonyl-CoA hydratase (Dch) and 6-hydroxycyclohex-1-ene-1-carbonyl-CoA dehydrogenases (Had) [42]. Together with detection of 23-162 nM benzoate in these sediments (Table 1) these results strongly suggest that the organisms represented by these MAGs are engaged in the typical downstream degradation of aromatic compounds employing the central benzoyl-CoA Bcr-Dch-Had-Oah pathway. However, the possible origins of benzoate are unclear based on current data.

Widespread capacity for fermentative production and respiratory consumption of acetate and hydrogen

Analysis of MAGs from these deep-sea hydrocarbon-associated sediments suggests that fermentation, rather than respiration, is the primary mode of organic carbon turnover in this environment. Most recovered MAGs with capacity for heterotrophic carbon degradation lacked respiratory primary dehydrogenases and terminal reductases, with exceptions being several *Proteobacteria* and one *Chloroflexi* (Table S6). In contrast, only 6 and 14 MAGs contained genes indicating the capability for fermentative production of ethanol and lactate, respectively, whereas 69 MAGs contained genes for fermentative acetate production (Figure 3 and Table S6).

These findings are consistent with other studies emphasizing the importance of fermentation, including acetate production, in deep-sea sediments [12, 43].

Acetate can also be produced by acetogenic CO_2 reduction through the Wood-Ljungdahl pathway using a range of substrates, including heterotrophic compounds [15]. Partial or complete sets of genes for the Wood-Ljungdahl pathway were found in 50 MAGs (Figures 3 and S3), including those affiliated with phyla previously inferred to mediate acetogenesis in deep-sea sediments through either the tetrahydrofolate-dependent bacterial pathway (*e.g. Chloroflexi* and *Aerophobetes*) [7, 44] or the tetrahydromethanopterin-dependent archaeal variant (*e.g. Bathyarchaeota* and Asgard group) [45, 46]. In addition, the signature diagnostic gene for the Wood-Ljungdahl pathway (*acsB*; acetyl-CoA synthase) is in high relative abundance in the quality-filtered metagenome reads at all three sites (Table S7). These observations are in agreement with mounting evidence that homoacetogens play a quantitatively important role in organic carbon cycling in the marine deep biosphere [45, 47, 48].

Evidence for H_2 metabolism was also found in MAGs from all three sites. We screened putative hydrogenase genes from various subgroups in MAGs as well as unbinned metagenomic sequences (Figures 1, 3 and Tables S6, S7). Surprisingly few H_2 evolving-only hydrogenases were detected, with only five Group A [FeFe]-hydrogenases and five Group 4 [NiFe]-hydrogenases detected across the bacterial and archaeal MAGs. Instead, the most abundant hydrogenases within MAGs as well as in quality-filtered unassembled reads were the Group 3b, 3c, and 3d [NiFe]-hydrogenases. Group 3b and 3d hydrogenases are physiologically reversible, but generally support fermentation in anoxic environments by coupling NAD(P)H reoxidation to fermentative H_2 evolution [49-51]. Group 3c hydrogenases mediate a central step in

hydrogenotrophic methanogenesis, bifurcating electrons from H_2 to heterodisulfides and ferredoxin [52]. The functional role of Group 3c hydrogenases in Bacteria and non-methanogenic Archaea remains unresolved [51] yet their corresponding genes co-occurred with heterodisulfide reductases across multiple archaeal and bacterial MAGs (Figure 3). Various Group 1 [NiFe]-hydrogenases were also detected, which supports hydrogenotrophic respiration in conjunction with a wide range of terminal reductases. This is consistent with previous studies in the Gulf of Mexico that experimentally measured the potential for hydrogen oxidation catalyzed by hydrogenase enzymes [53].

Given the genomic evidence for hydrogen and acetate production in these sediments, we investigated whether any of the MAGs encoded terminal reductases to respire using these compounds as electron donors. In agreement with the high sulfate concentrations (Table 1), the key genes for dissimilatory sulfate reduction (*dsrAB*) were widespread across the metagenome reads, particularly at Site E29 (Table S7). These genes were recovered from MAGs affiliated with *Deltaproteobacteria* and *Dehalococcoidia* (Table S6). We also identified 31 novel reductive dehalogenase (*rdhA*) genes across 22 MAGs, mainly from *Aminicenantes* and *Bathyarchaeota* (Figure 3 and Table S6), suggesting that organohalides – that can be produced through abiotic and biotic processes in marine ecosystems [54] – may be electron acceptors in these deep-sea sediments. All MAGs corresponding to putative sulfate reducers and dehalorespirers encoded the capacity to completely oxidize acetate and other organic acids to CO_2 using either the reverse Wood-Ljungdahl pathway or TCA cycle (Figure 3 and Table S6). Several of these MAGs also harbored the capacity for hydrogenotrophic dehalorespiration *via* Group 1a and 1b [NiFe]-hydrogenases (Figure 3). In addition to these dominant uptake pathways, one MAG belonging to the epsilonproteobacterial genus *Sulfurovum* (E29_bin29) included genes for the enzymes

needed to oxidize either H_2 (group 1b [NiFe]-hydrogenase), elemental sulfur (SoxABXYZ), and sulfide (Sqr), using nitrate as an electron acceptor (NapAGH); this MAG also has a complete set of genes for autotrophic CO_2 fixation *via* the reductive TCA cycle (Figure 3 and Table S6). The capacity for methanogenesis appears to be relatively low as revealed by the MAGs retrieved from the three sediments. None of the MAGs contained *mcrA* genes, but these genes for methanogenesis were detected in quality-filtered unassembled reads in all three sediments (Figures 1d and S4) and were mainly affiliated with acetoclastic methanogens at Site E29, and hydrogenotrophic methanogens at the other two sites (Figure 1d and S4). Overall, the collectively weak *mcrA* signal in the metagenomes suggests that the high levels of biogenic methane detected by geochemical analysis (Table 1) is due to methanogenesis in sediment layers deeper than the top 20 cm.

Thermodynamic modelling of hydrocarbon degradation

Both the geochemistry data and biomarker gene survey suggest that hydrocarbon degradation is occurring in the three deep-sea sediments sampled (Tables 1 and S1). Recreating environmental conditions for cultivating the organisms represented by the retrieved MAGs is a challenging and time consuming process, preventing further direct evidence of hydrocarbon degradation capabilities (and other metabolisms) among the majority of the lineages represented by the MAGs retrieved here [48]. Instead, we provide theoretical evidence that hydrocarbon degradation is feasible in this environment by modelling whether these processes are thermodynamically favorable in the conditions typical of deep sea sediments, namely high pressure and low temperature.

As concluded from the genome analysis and supported by metabolomics (Table 1), it is likely that most hydrocarbon oxidation occurs through fermentation rather than respiration. Taking hydrogen production and the Wood-Ljungdahl pathway into consideration (Figures 3 and 4), we compared the thermodynamic constraints on hydrocarbon biodegradation for two plausible scenarios: (1) fermentation with production of hydrogen and acetate, and (2) fermentation with production of acetate alone. Hexadecane and benzoate are used as representative aliphatic and aromatic compounds, respectively, based on the geochemistry results (e.g. C_{2+} alkane detection) and genomic analysis (e.g. *bcr* genes) [47, 55]. The calculated results show that the threshold concentrations of acetate that result in favorable energetics ($\Delta G' < 0$) for fermentative co-generation of acetate and hydrogen require acetate to be extremely low in a hexadecane degradation scenario ($< 10^{-12}$ mM acetate), and acetate to be at moderate levels in a benzoate degradation scenario (< 3.8 mM acetate) (Figure 5). By contrast, for fermentation leading to production of only acetate, its concentration can be as high as 470 mM in a benzoate degradation scenario and as high as 300 mM in a hexadecane degradation scenario (Figure 5). Fermentative degradation of hexadecane to hydrogen and acetate in the deep seabed could therefore be less favorable than acetate production alone via the Wood-Ljungdahl pathway. If microbial communities are consuming hexadecane or more complex hydrocarbons as carbon and energy sources, it might be likely that they employ the Wood-Ljungdahl pathway to produce acetate. These lines of evidence emphasize the potential importance of acetogenic fermentation of hydrocarbons in the deep seabed. However, other reactions such as fermentation to H_2 still cannot be excluded, *e.g.*, for less complex hydrocarbons such as benzoate and related compounds.

Discussion

In this study, metagenomics revealed that the majority of the Bacteria and Archaea comprising deep-sea sediment microbial communities belong to candidate phyla that lack cultured representatives and sequenced genomes (Figures 1 and 2). As a consequence, it is challenging to link phylogenetic patterns with the microbial functional traits underpinning the biogeochemistry of deep seabed habitats. Here, we were able to address this by combining *de novo* assembly and binning of metagenomic data with geochemical and metabolomic analyses, and complementing our observations with thermodynamic modeling. Pathway reconstruction from 82 MAGs recovered from the three deep-sea surface sediments revealed that many community members were capable of anaerobic hydrocarbon degradation as well as acquiring and hydrolyzing residual organic matter (Figure 3), whether supplied as detritus from the overlying water column or as autochthonously produced necromass (Figure 6). Heterotrophic fermenters and acetogens were in considerably higher relative abundance than heterotrophic respirers, despite the abundance of sulfate in the sediments (Table 1). For example, while genomic coverage of putative sulfate reducers is relatively low (< 1% of the communities), the most abundant MAG at each site were all putative acetogenic heterotrophs, *i.e. Dehalococcoidia* E26_bin16, *Actinobacteria* E44_bin5, and *Aminicenantes* E29_bin47 for Sites E26, E44 and E29 respectively (~3.3-4.5% relative abundance, Table S4). Therefore, in contrast with coastal sediments [56], microbial communities in the deep seabed are likely influenced by the capacity to utilize available electron donors more so than by the availability of oxidants.

In this context, multiple lines of evidence indicate degradation of aliphatic or aromatic petroleum compounds as carbon and energy sources for anaerobic populations in these deep-sea

hydrocarbon seep environments (Table 1, Figures 3 - 5). Whereas capacity for detrital organic matter degradation is a common feature in the genomes retrieved in this study, and from many other environments [26], anaerobic hydrocarbon degradation is a more exclusive feature that was detected in 23 out of 82 MAGs. Evidence of anaerobic alkane oxidation via fumarate addition and hydroxylation pathways, as well as anaerobic aromatic compound degradation by the Class I benzoyl-CoA reductase pathway, was found in all three sediments. The ability to utilize hydrocarbons may explain the ecological dominance (high relative abundance) of certain lineages of Bacteria and Archaea in these microbial communities (Figure 1a) as many of those phyla have previously been found to be associated with hydrocarbons in various settings. For example, *Aerophobetes* have been detected in other cold seep environments [7], *Aminicenantes* are often found associated with fossil fuels [27], and *Chloroflexi* harboring genes for anaerobic hydrocarbon degradation have been found in hydrothermal vent sediments [4]. While Archaea have been reported to mediate oxidation of methane and other short-chain alkanes in sediments [5, 36], few have been reported to anaerobically degrade larger hydrocarbons [37]. The finding of *Bathyarchaeota* and other archaeal phyla potentially capable of anaerobic hydrocarbon degradation extends the potential hydrocarbon substrate spectrum for Archaea. More broadly, these findings extend the breadth of lineages that include putative hydrocarbon-degrading bacteria and archaea in deep-sea sediments. Current knowledge of anaerobic hydrocarbon degradation remains limited, with the majority of studies focused on environments subject to anthropogenic hydrocarbon contamination, most notably groundwater aquifers [37]. It is possible that microorganisms inhabiting deep-sea sediments harbor novel mechanisms for anaerobic hydrocarbon degradation that may be relevant for biotechnology and bioremediation in a variety

of other settings, *e.g.*, other cold habitats. Future studies of genome-enabled hydrocarbon degradation using samples such as the sediments studied here may elucidate this further.

Genomic analyses of 12 MAGs harboring genes for central benzoyl-CoA pathway reveal that they are likely a mixture of obligate fermenters and sulfate reducers. The finding that these organisms use the ATP-consuming class I, not the reversible class II, benzoyl-CoA reductase is surprising. It is generally thought that strict anaerobes must use class II BCRs because the amount of energy available from benzoate oxidation during sulfate reduction or fermentation is not sufficient to support the substantial energetic requirement of the ATP-dependent class I BCR reaction [42]. However, acetogenic fermentation of hydrocarbons may explain how the Class I reaction could be thermodynamically favorable, as shown in Figure 5. In agreement with this, there are reported exceptions to the general Class I vs Class II observations, such as the hyperthermophilic archaeon *Ferroglobus placidus* that couples benzoate degradation via the Class I system with iron reduction [42], and fermentative deep-sea *Chloroflexi* strains DscP3 and Dsc4 that contain genes for class I benzoyl-CoA reductases [44]. Indeed, acetogens can utilize many different substrates and have relatively high ATP yields as well as thermodynamic efficiencies toward heterotrophic substrates, which is consistent with the proposed importance of acetogens in energy-limited seafloor ecosystems [45, 47].

Based on the evidence presented here we propose that acetate and hydrogen are the central intermediates underpinning community interactions and biogeochemical cycling in these deep-sea sediments (Figure 6). Maintaining low acetate and hydrogen concentrations in the environment is important for promoting continuous fermentation of organic substrates, consistent with thermodynamic constraints (Figure 5). Acetate and hydrogen in sediment porewater were

below detection limits, consistent with the high turnover rates of both compounds. This may correspond with the genomic potential within these microbial communities for the coupling of acetate consumption to sulfate reduction, organohalide respiration and acetoclastic methanogenesis, as suggested in other studies [55, 57]. Some community members also appear to be capable of H_2 consumption, including *via* putative heterodisulfide reductase-coupled hydrogenases. In turn, hydrogen oxidation can support autotrophic carbon fixation and therefore may provide a feedback loop for regeneration of organic carbon. Acetate- and hydrogen-oxidizing community members are likely to promote upstream fermentative degradation of necromass and hydrocarbons.

Overall, this metagenome dataset extended the knowledge of metabolic potential of microbial communities inhabited in deep-sea sediments that receive an input of thermogenic hydrocarbon. They are mostly likely sustained through fermentation, acetogenesis and hydrogen metabolisms. More importantly, supported by geochemical and metabolomics analysis of hydrocarbons and other metabolites as well as by thermodynamic modelling of different hydrocarbon degradation scenarios, our findings expanded the diversity of microbial lineages with the potential for anaerobic hydrocarbon degradation through *e.g.* the activity of glycyl radical enzymes. Together with the recent discovery of anaerobic butane degradation in gas-rich hydrothermally-heated sediments [36], it can be inferred that anaerobic degradation of hydrocarbons heavier than methane might be more widespread than previously expected. They may play as an important chemical energy in dark deep-sea sediments just like hydrogen, sulfide and methane [58].

Methods

Sampling and geochemical measurements

The three marine sediment samples used in this study were collected from the near-surface (top 20 cm) of the seafloor in the Eastern Gulf of Mexico as part of a piston coring survey, as described previously [20]. Samples for hydrocarbon characterization were sectioned on board the research vessel immediately following piston core retrieval, flushed with N_2 and sealed in hydrocarbon-free gas tight metal canisters then frozen until analysis. Interstitial gas analysis was later performed on the headspace in the canisters using GC with Flame Ionization Detector (GC-FID). Sediment samples for gas/liquid chromatography and stable isotope analysis were frozen, freeze-dried and homogenized then extracted using accelerated solvent extraction (ACE 200). Extracts were subsequently analyzed using GC/FID, a Perkin-Elmer Model LS 50B fluorometer, GC/MS and Finnigan MAT 252 isotope mass spectrometry as detailed elsewhere [59].

Sulfate and chloride concentrations were measured in a Dionex ICS-5000 reagent-free ion chromatography system (Thermo Scientific, CA, USA) equipped with an anion-exchange column (Dionex IonPac AS22; 4 × 250 mm; Thermo Scientific), an EGC-500 K_2CO_3 eluent generator cartridge and a conductivity detector. Organic acids were analysed in the 0.2 µm filtered sediment porewater using a Thermo RS3000 HPLC fitted with an Ultimate 3000 UV detector. Separation was achieved over an Aminex HPX-87H organic acid column (Biorad, USA) under isocratic conditions (0.05 mM H_2SO_4) at 60°C with a run time of 20 minutes. Organic acids were compared to the retention time of known standards and the limit of detection for acetate was determined to be 2.5 µM.

For the analysis of metabolites, sediment was spun down, the supernatant collected, diluted 1:1 in pure methanol, and filtered through 0.2 µm Teflon syringe filters. Extracts were separated using Ultra High-Performance Liquid Chromatography (UHPLC) equipped with a hydrophilic interaction liquid chromatography column (Syncronis HILIC, Thermo Fisher). A Thermo Fisher Scientific Q-Exactive HF mass spectrometer in negative-mode electrospray ionization was used to collect high-resolution full-scan MS data. In addition, benzoate ion (*m*/*z* 121.02943) was subjected to fragmentation using collision induced dissociation with a collision energy of 10% and activation time 20 ms (*m*/*z* 121.02943 > 77.03948). Metabolites were further identified using accurate mass and retention times of standards using Thermo Xcalibur software and MAVEN freeware. Larger compound lists were assigned identification using a combination of MAVEN and the KEGG database.

DNA extraction and sequencing

For the three sediment samples, DNA was extracted from 10 g of sediment using the PowerMax Soil DNA Isolation Kit (12988-10, QIAGEN) according to the manufacturer's protocol with minor modifications for the step of homogenization and cell lysis *i.e.*, cells were lysed in PowerMax Bead Solution tubes for 45□s at 5.5□m□s^{-1} using a Bead Ruptor 24 (OMNI International). DNA concentrations were assessed using a Qubit 2.0 fluorometer (Thermo Fisher Scientific, Canada). Metagenomic library preparation and DNA sequencing was conducted at the Center for Health Genomics and Informatics in the Cumming School of Medicine, University of Calgary. DNA fragment libraries were prepared by shearing genomic DNA using a Covaris sonicator and the NEBNext Ultra II DNA library preparation kit (New England BioLabs). DNA

was sequenced on a ~40 Gb (*i.e.* 130 M reads) mid-output NextSeq 500 System (Illumina Inc.) 300 cycle (2 × 150 bp) sequencing run.

To provide a high-resolution microbial community profile, the three samples were also subjected to 16S rRNA gene amplicon sequencing on a MiSeq benchtop sequencer (Illumina Inc.). DNA was extracted from separate aliquots of the same sediment samples using the DNeasy PowerLyzer PowerSoil kit (MO BIO Laboratories, a Qiagen Company, Carlsbad, CA, USA) and used as the template for different PCR reactions. The v3-4 region of the bacterial 16S rRNA gene and the v4-8 region of the archaeal 16S rRNA gene were amplified using the primer pairs SD-Bact-0341-bS17/SD-Bact-0785-aA21 and SD-Arch-0519-aS15/SD-Arch-0911-aA20, respectively [60] as described previously [20] on a ~15 Gb 600-cyce (2 x 300 bp) sequencing run (for results see Figure S1).

Metagenomic assembly and binning

Raw reads were quality-controlled by (1) clipping off primers and adapters and (2) filtering out artifacts and low-quality reads as described previously [61]. Filtered reads were assembled using metaSPAdes version 3.11.0 [62] and short contigs (<500 bp) were removed. Sequence coverage was determined by mapping filtered reads onto assembled contigs using BBmap version 36 (https://sourceforge.net/projects/bbmap/). Binning of metagenome contigs was performed using MetaBAT version 2.12.1 (--minContig 1500) [63]. Contaminated contigs in produced bins were further removed based on genomic properties (GC, tetranucleotide signatures, and coverage) and taxonomic assignments using RefineM version 0.0.22 [64]. Resulting bins were further examined for contamination and completeness using CheckM version 1.0.8 with the lineage-specific workflow [24].

Annotation

For MAGs, genes were called by Prodigal (-p meta) [65]. Metabolic pathways were predicted against the KEGG GENES database using the GhostKOALA tool [66] and against the Pfam, TIGRfam and custom HMM databases (https://github.com/banfieldlab/metabolic-hmms) using MetaErg (https://sourceforge.net/projects/metaerg/). The dbCAN web server was used for carbohydrate-active gene identification (cutoffs: coverage fraction: 0.40; e-value: 1e-18) [67]. Genes encoding proteases and peptidases were identified using BLASTP against the MEROPS database release 12.0 (cutoffs: e-value, 1e-20; sequence identity, 30%) [68]. Genes involved in anaerobic hydrocarbon degradation were identified using BLASTP against a custom database (Table S8) (cutoffs: e-value, 1e-20; sequence identity, 30%). Hydrogenases were identified and classified using a web-based search using the hydrogenase classifier HydDB with option of Check sequences using CDD [69].

Full-length 16S rRNA genes were reconstructed from metagenomic reads using phyloFlash version 3.1 (https://hrgv.github.io/phyloFlash/) together with the SILVA SSU 132 rRNA database [30]. Diversity calculations were based on phyloFlash counts as well as the separate 16S rRNA gene amplicon library results. Functional and taxonomic McrA gpkgs were used to assess the diversity of methanogens against the metagenomic reads using GraftM with default parameters [70]. Genes encoding the catalytic subunits of hydrogenases, *dsrA*, *acsB*, *assA*, *nmsA* and *bssA* were retrieved from metagenomic reads through diamond BLASTx [71] queries against comprehensive custom databases [46, 69, 72] (cutoffs: e-value, 1e-10; sequence identity, 70%).

Phylogenetic analyses

For taxonomic classification of each MAG, two methods were used to produce genome trees that were then used to validate each other. In the first method the tree was constructed using concatenated proteins of up to 16 syntenic ribosomal protein genes following procedures reported elsewhere [73]; the second tree was constructed using concatenated amino acid sequences of up to 43 conserved single-copy genes following procedures described previously [74]. Both trees were calculated using FastTree version 2.1.9 (-lg -gamma) [75] and resulting phylogenies were congruent. Reference genomes for relatives were accessed from NCBI GenBank including genomes selected from several recent studies representing the majority of candidate bacterial and archaeal phylogenetic groups [4, 64, 76-79]. The tree in Figure 2 was inferred based on concatenation of 43 conserved single-copy genes (Database S1). To be accurate, it was built using RAxML version 8 [80] implemented by the CIPRES Science Gateway [81]. It was called as follows: raxmlHPC-HYBRID -f a -n result -s input -c 25 -N 100 -p 12345 -m PROTCATLG -x 12345. The phylogeny resulting from RAxML is consistent with the taxonomic classification of MAGs that resulted from FastTree. Interactive tree of life (iTOL) version 3 [82] was used for tree visualization and modification.

For phylogenetic placements of functional genes, sequences were aligned using the MUSCLE algorithm [83] included in MEGA7 [84]. All positions with less than 95% site coverage were eliminated. Maximum likelihood phylogenetic trees were constructed in MEGA7 using a general time reversible substitution model and uniform rates among sites. These trees were bootstrapped with 100 replicates.

Taxonomic classification of MAGs inferred to belong to candidate phylum TA06 after phylogenetic analyses were additionally confirmed by performing classify workflow using GTDB-Tk version 0.0.6+ (https://github.com/Ecogenomics/GtdbTk).

Thermodynamic calculations

The values of Gibbs free energy of formation for substances were taken from Madigan et al. [85] and Dolfing et al. [55]. The pH used in all calculations was 8.0 as reported in a previous thermodynamic study of deep marine sediments [47], partial pressure was 300 atm based on water depths at the three sites (http://docs.bluerobotics.com/calc/pressure-depth/), and temperature was set as 4°C to represent deep sea conditions [86]. Calculations followed accepted protocols for determining reaction kinetics and thermodynamics [87].

Acknowledgements

The work was supported by a Genome Canada Genomics Applied Partnership Program (GAPP) grant to CRJH. We thank Xiaoli Dong and Marc Strous for bioinformatics support, the Centre for Health Genomics and Informatics at University of Calgary for NextSeq sequencing, Alexander Probst for providing the database of 16 syntenic ribosomal proteins, and Nina Dombrowski and Brett Baker for providing a custom blast database for hydrocarbon degradation genes.

Data availability

DNA sequences (amplicon sequences, genomes and raw sequence reads) have been deposited in the NCBI BioProject database with accession number PRJNA415828 and PRJNASUB3936075

(https://www.ncbi.nlm.nih.gov/bioproject/). The authors declare that all other data supporting the findings of this study are available within the article and its supplementary information files, or from the corresponding authors upon request.

References

1. Petro, C., Starnawski, P., Schramm, A. & Kjeldsen, K.U. Microbial community assembly in marine sediments. *Aquat Microb Ecol* **79**, 177-195 (2017).
2. Vigneron, A. et al. Comparative metagenomics of hydrocarbon and methane seeps of the Gulf of Mexico. *Sci Rep* **7**, 16015 (2017).
3. Marshall, I.P.G., Karst, S.M., Nielsen, P.H. & Jørgensen, B.B. Metagenomes from deep Baltic Sea sediments reveal how past and present environmental conditions determine microbial community composition. *Mar Genomics* **37**, 58-68 (2018).
4. Dombrowski, N., Seitz, K.W., Teske, A.P. & Baker, B.J. Genomic insights into potential interdependencies in microbial hydrocarbon and nutrient cycling in hydrothermal sediments. *Microbiome* **5**, 106 (2017).
5. Lloyd, K.G. et al. Predominant archaea in marine sediments degrade detrital proteins. *Nature* **496**, 215-218 (2013).
6. Orsi, W.D. Ecology and evolution of seafloor and subseafloor microbial communities. *Nat Rev Microbiol* (2018).
7. Wang, Y. et al. Draft genome of an *Aerophobetes* bacterium reveals a facultative lifestyle in deep-sea anaerobic sediments. *Sci Bull* **61**, 1176-1186 (2016).
8. Orcutt, B.N., Sylvan, J.B., Knab, N.J. & Edwards, K.J. Microbial ecology of the dark ocean above, at, and below the seafloor. *Microbiol Mol Biol Rev* **75**, 361-422 (2011).
9. Inagaki, F. et al. Exploring deep microbial life in coal-bearing sediment down to ~2.5 km below the ocean floor. *Science* **349**, 420-424 (2015).
10. Jørgensen, B.B. & Marshall, I.P. Slow Microbial Life in the Seabed. *Ann Rev Mar Sci* **8**, 311-332 (2016).
11. Jørgensen, B.B. & Boetius, A. Feast and famine--microbial life in the deep-sea bed. *Nat Rev Microbiol* **5**, 770-781 (2007).
12. Arndt, S. et al. Quantifying the degradation of organic matter in marine sediments: A review and synthesis. *Earth-Sci Rev* **123**, 53-86 (2013).
13. Orsi, W.D., Richards, T.A. & Francis, W.R. Predicted microbial secretomes and their target substrates in marine sediment. *Nat Microbiol* **3**, 32-37 (2018).

14. Brooks, J.M. et al. Deep-sea hydrocarbon seep communities: evidence for energy and nutritional carbon sources. *Science* **238**, 1138-1142 (1987).

15. Ijiri, A. et al. Deep-biosphere methane production stimulated by geofluids in the Nankai accretionary complex. *Sci Adv* **4**, eaao4631 (2018).

16. Ruff, S.E. et al. Global dispersion and local diversification of the methane seep microbiome. *Proc Natl Acad Sci U S A* **112**, 4015-4020 (2015).

17. Dong, X. et al. Fermentative Spirochaetes mediate necromass recycling in anoxic hydrocarbon-contaminated habitats. *ISME J* **12**, 2039-2050 (2018).

18. Lin, Y.-S. et al. Towards constraining H_2 concentration in subseafloor sediment: A proposal for combined analysis by two distinct approaches. *Geochim Cosmochim Acta* **77**, 186-201 (2012).

19. Sauvage, J., Flinders, A., Spivack, A. & D'Hondt, S. Global distribution of radiolytic H_2 production in marine sediment and implications for subsurface life. *AGU Fall Meeting Abstracts* (2017).

20. Chakraborty, A. et al. Thermophilic endospores associated with migrated thermogenic hydrocarbons in deep Gulf of Mexico marine sediments. *ISME J* **12**, 1895-1906 (2018).

21. Whiticar, M.J. Carbon and hydrogen isotope systematics of bacterial formation and oxidation of methane. *Chem Geol* **161**, 291-314 (1999).

22. Killops, S.D. & Aljuboori, M.A.H.A. Characterization of the Unresolved Complex Mixture (UCM) in the gas chromatograms of biodegraded petroleums. *Org Geochem* **15**, 147-160 (1990).

23. Gieg, L.M. & Toth, C.R.A. in Anaerobic Utilization of Hydrocarbons, Oils, and Lipids. (ed. M. Boll) 1-30 (Springer International Publishing, Cham; 2017).

24. Parks, D.H., Imelfort, M., Skennerton, C.T., Hugenholtz, P. & Tyson, G.W. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res* **25**, 1043-1055 (2015).

25. Ley, R.E. et al. Unexpected diversity and complexity of the Guerrero Negro hypersaline microbial mat. *Appl Environ Microbiol* **72**, 3685-3695 (2006).

26. Baker, B.J., Lazar, C.S., Teske, A.P. & Dick, G.J. Genomic resolution of linkages in carbon, nitrogen, and sulfur cycling among widespread estuary sediment bacteria. *Microbiome* **3**, 14 (2015).

27. Hu, P. et al. Genome-resolved metagenomic analysis reveals roles for candidate phyla and other microbial community members in biogeochemical transformations in oil reservoirs. *MBio* **7**, e01669-01615 (2016).

28. Sieber, C.M.K. et al. Recovery of genomes from metagenomes via a dereplication, aggregation and scoring strategy. *Nat Microbiol* **3**, 836-843 (2018).

29. Parks, D.H. et al. A proposal for a standardized bacterial taxonomy based on genome phylogeny. *bioRxiv* (2018).

30. Quast, C. et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res* **41**, D590-596 (2013).

31. Bradley, J.A., Amend, J.P. & LaRowe, D.E. Necromass as a limited source of energy for microorganisms in marine sediments. *J Geophys Res-Biogeo* **123**, 577-590 (2018).

32. Lipp, J.S., Morono, Y., Inagaki, F. & Hinrichs, K.U. Significant contribution of Archaea to extant biomass in marine subsurface sediments. *Nature* **454**, 991-994 (2008).

33. Boll, M. & Heider, J. in Handbook of Hydrocarbon and Lipid Microbiology. (ed. K.N. Timmis) 1011-1024 (Springer Berlin Heidelberg, Berlin, Heidelberg; 2010).

34. Tan, B. et al. Comparative analysis of metagenomes from three methanogenic hydrocarbon-degrading enrichment cultures with 41 environmental samples. *ISME J* **9**, 2028-2045 (2015).

35. Meckenstock, R.U. et al. Anaerobic degradation of benzene and polycyclic aromatic hydrocarbons. *J Mol Microbiol Biotechnol* **26**, 92-118 (2016).

36. Laso-Perez, R. et al. Thermophilic archaea activate butane via alkyl-coenzyme M formation. *Nature* **539**, 396-401 (2016).

37. Rabus, R. et al. Anaerobic Microbial Degradation of Hydrocarbons: From Enzymatic Reactions to the Environment. *J Mol Microbiol Biotechnol* **26**, 5-28 (2016).

38. Schouw, A. et al. *Abyssivirga alkaniphila* gen. nov., sp. nov., an alkane-degrading, anaerobic bacterium from a deep-sea hydrothermal vent system, and emended descriptions of *Natranaerovirga pectinivora* and *Natranaerovirga hydrolytica*. *Int J Syst Evol Microbiol* **66**, 1724-1734 (2016).

39. Khelifi, N. et al. Anaerobic oxidation of long-chain n-alkanes by the hyperthermophilic sulfate-reducing archaeon, *Archaeoglobus fulgidus*. *ISME J* **8**, 2153-2166 (2014).

40. Porter, A.W. & Young, L.Y. in Advances in Applied Microbiology, Vol. 88. (eds. S. Sariaslani & G.M. Gadd) 167-203 (Academic Press, 2014).
41. Dong, X. et al. Reconstructing metabolic pathways of a member of the genus *Pelotomaculum* suggesting its potential to oxidize benzene to carbon dioxide with direct reduction of sulfate. *FEMS Microbiol Ecol* **93**, fiw254-fiw254 (2017).
42. Holmes, D.E., Risso, C., Smith, J.A. & Lovley, D.R. Genome-scale analysis of anaerobic benzoate and phenol metabolism in the hyperthermophilic archaeon *Ferroglobus placidus*. *ISME J* **6**, 146-157 (2012).
43. Beulig, F., Røy, H., Glombitza, C. & Jørgensen, B.B. Control on rate and pathway of anaerobic organic carbon degradation in the seabed. *Proc Natl Acad Sci U S A* **115**, 367-372 (2018).
44. Sewell, H.L., Kaster, A.K. & Spormann, A.M. Homoacetogenesis in Deep-Sea *Chloroflexi*, as inferred by single-cell genomics, provides a link to reductive dehalogenation in terrestrial *Dehalococcoidetes*. *MBio* **8** (2017).
45. He, Y. et al. Genomic and enzymatic evidence for acetogenesis among multiple lineages of the archaeal phylum *Bathyarchaeota* widespread in marine sediments. *Nat Microbiol* **1**, 16035 (2016).
46. Adam, P.S., Borrel, G. & Gribaldo, S. Evolutionary history of carbon monoxide dehydrogenase/acetyl-CoA synthase, one of the oldest enzymatic complexes. *Proc Natl Acad Sci U S A* **115**, E1166-E1173 (2018).
47. Lever, M.A. Acetogenesis in the energy-starved deep biosphere - a paradox? *Front Microbiol* **2**, 284 (2011).
48. Yu, T. et al. Growth of sedimentary *Bathyarchaeota* on lignin as an energy source. *Proc Natl Acad Sci U S A* **115**, 6022-6027 (2018).
49. Berney, M., Greening, C., Conrad, R., Jacobs, W.R., Jr. & Cook, G.M. An obligately aerobic soil bacterium activates fermentative hydrogen production to survive reductive stress during hypoxia. *Proc Natl Acad Sci U S A* **111**, 11479-11484 (2014).
50. Ma, K., Hao, Z. & Adams, M.W.W. Hydrogen production from pyruvate by enzymes purified from the hyperthermophilic archaeon, *Pyrococcus furiosus*: A key role for NADPH. *FEMS Microbiol Lett* **122**, 245-250 (1994).

51. Greening, C. et al. Genomic and metagenomic surveys of hydrogenase distribution indicate H_2 is a widely utilised energy source for microbial growth and survival. *ISME J* **10**, 761 (2015).

52. Wagner, T., Koch, J., Ermler, U. & Shima, S. Methanogenic heterodisulfide reductase (HdrABC-MvhAGD) uses two noncubane [4Fe-4S] clusters for reduction. *Science* **357**, 699-703 (2017).

53. Adhikari, R.R. et al. Hydrogen utilization potential in subsurface sediments. *Front Microbiol* **7**, 8 (2016).

54. Kawai, M. et al. High frequency of phylogenetically diverse reductive dehalogenase-homologous genes in deep subseafloor sedimentary metagenomes. *Front Microbiol* **5**, 80 (2014).

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

56. Canfield, D.E., Thamdrup, B. & Hansen, J.W. The anaerobic degradation of organic matter in Danish coastal sediments: iron reduction, manganese reduction, and sulfate reduction. *Geochim Cosmochim Acta* **57**, 3867-3883 (1993).

57. Liu, Y.F. et al. Metabolic capability and in situ activity of microorganisms in an oil reservoir. *Microbiome* **6**, 5 (2018).

58. Colman, D.R., Poudel, S., Stamps, B.W., Boyd, E.S. & Spear, J.R. The deep, hot biosphere: Twenty-five years of retrospection. *Proceedings of the National Academy of Sciences* **114**, 6895-6903 (2017).

59. Bernard, B.B. et al. Surface geochemical exploration and heat flow surveys in fifteen (15) frontier Indonesian basins. *Thirty-Second Annual Convention & Exhibition May 2008* (2008).

60. Klindworth, A. et al. Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. *Nucleic Acids Research* **41**, e1-e1 (2013).

61. Kleiner, M. et al. Assessing species biomass contributions in microbial communities via metaproteomics. *Nat Commun* **8**, 1558 (2017).

62. Nurk, S., Meleshko, D., Korobeynikov, A. & Pevzner, P.A. metaSPAdes: a new versatile metagenomic assembler. *Genome Res* **27**, 824-834 (2017).

63. Kang, D.D., Froula, J., Egan, R. & Wang, Z. MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities. *PeerJ* **3**, e1165 (2015).

64. Parks, D.H. et al. Recovery of nearly 8,000 metagenome-assembled genomes substantially expands the tree of life. *Nat Microbiol* **2**, 1533-1542 (2017).

65. Hyatt, D. et al. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* **11**, 119 (2010).

66. Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P.M. & Henrissat, B. The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res* **42**, D490-495 (2014).

67. Yin, Y. et al. dbCAN: a web resource for automated carbohydrate-active enzyme annotation. *Nucleic Acids Res* **40**, W445-451 (2012).

68. Rawlings, N.D. et al. The MEROPS database of proteolytic enzymes, their substrates and inhibitors in 2017 and a comparison with peptidases in the PANTHER database. *Nucleic Acids Res* **46**, D624-D632 (2018).

69. Søndergaard, D., Pedersen, C.N. & Greening, C. HydDB: A web tool for hydrogenase classification and analysis. *Sci Rep* **6**, 34212 (2016).

70. Boyd, J.A., Woodcroft, B.J. & Tyson, G.W. GraftM: a tool for scalable, phylogenetically informed classification of genes within metagenomes. *Nucleic Acids Res* **46**, e59 (2018).

71. Buchfink, B., Xie, C. & Huson, D.H. Fast and sensitive protein alignment using DIAMOND. *Nature Methods* **12**, 59 (2014).

72. Anantharaman, K. et al. Expanded diversity of microbial groups that shape the dissimilatory sulfur cycle. *ISME J* **12**, 1715-1728 (2018).

73. Tully, B.J., Sachdeva, R., Graham, E.D. & Heidelberg, J.F. 290 metagenome-assembled genomes from the Mediterranean Sea: a resource for marine microbiology. *PeerJ* **5**, e3558 (2017).

74. Kato, S. et al. Genome-enabled metabolic reconstruction of dominant chemosynthetic colonizers in deep-sea massive sulfide deposits. *Environ Microbiol* **20**, 862-877 (2018).

75. Price, M.N., Dehal, P.S. & Arkin, A.P. FastTree 2--approximately maximum-likelihood trees for large alignments. *PLoS One* **5**, e9490 (2010).

76. Anantharaman, K. et al. Thousands of microbial genomes shed light on interconnected biogeochemical processes in an aquifer system. *Nat Commun* **7**, 13219 (2016).
77. Zaremba-Niedzwiedzka, K. et al. Asgard archaea illuminate the origin of eukaryotic cellular complexity. *Nature* **541**, 353-358 (2017).
78. Probst, A.J. et al. Differential depth distribution of microbial function and putative symbionts through sediment-hosted aquifers in the deep terrestrial subsurface. *Nat Microbiol* **3**, 328-336 (2018).
79. Hug, L.A. et al. A new view of the tree of life. *Nat Microbiol* **1**, 16048 (2016).
80. Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312-1313 (2014).
81. Miller, M.A., Pfeiffer, W. & Schwartz, T. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. *2010 Gateway Computing Environments Workshop (GCE)*, 1-8 (2010).
82. Letunic, I. & Bork, P. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res* **44**, W242-245 (2016).
83. Edgar, R.C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32**, 1792-1797 (2004).
84. Kumar, S., Stecher, G. & Tamura, K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol Biol Evol* **33**, 1870-1874 (2016).
85. Madigan, M.T., Martinko, J.M. & Parker, J. Brock biology of microorganisms, Vol. 13. (Pearson, 2017).
86. Xu, W.Y., Lowell, R.P. & Peltzer, E.T. Effect of seafloor temperature and pressure variations on methane flux from a gas hydrate layer: Comparison between current and late Paleocene climate conditions. *J Geophys Res-Sol Ea* **106**, 26413-26423 (2001).
87. Dolfing, J. in Hydrocarbon and Lipid Microbiology Protocols. (eds. T.J. McGenity, K.N. Timmis & B. Nogales Fernández) 155-163 (Springer Berlin Heidelberg, Berlin, Heidelberg; 2015).

Tables and Figures

Table 1 Geochemical description of sediment samples from Sites E26, E29 and E44. TSF Max: total scanning fluorescence maximum intensity. UCM: uncharacterized complex mixture. Σn-Alk: sum of C_{15}-C_{34} n-alkanes. ΣAlk Gas: total alkane gases. C_{2+} Alk: sum of alkane gases larger than methane. T/D: thermogenic/diagenetic n-alkane ratio. BDL: below detection limit. NA: not analyzed.

Core ID	***Site E26***	***Site E29***	***Site E44***
Latitude (N)	26.59	27.43	26.28
Longitude (W)	87.51	86.01	86.81
Water depth (km)	2.8	3.2	3.0
Sulfate (mM)	20.01	33.73	31.72
Benzoate (nM)	93.6	22.6	161.7
Succinate (nM)	11.7	5.0	16.6
Acetate (μM)	BDL	BDL	BDL
Chloride (g L^{-1})	21.04	20.15	21.05
Total Scanning Fluorescence MAX	57326.7	26738.3	13502.3
Unresolved Complex Mixture (μg g^{-1})	32	13	7.3
Σn-Alkanes (ng g^{-1})	2845.3	2527	1045
Thermogenic/Diagenetic Ratio	1.0	2.6	0.8
ΣAlkane Gas (ppm)	9	36012	9.9
C_{2+} Alkanes (ppm)	0.3	17.5	0.5
$C_1/(C_2+C_3)$	NA	3974.2	NA
$\delta^{13}CH_4$ (‰, vs. PDB)	NA	-85.1	NA
H_2 (ppm)	BDL	BDL	BDL

Figure 1 Relative frequency of metagenomic sequence reads for different marker genes at Sites E26, E29 and E44. (a) Community compositions based on reconstruction of full-length 16S rRNA genes from the metagenomes. Eukaryotes and unassigned reads are not shown. (b) Relative occurrences of hydrogenases with different metal co-factors. (c) Relative occurrences of different subtypes of NiFe hydrogenases. (d) Relative occurrences of *mcrA* genes indicative of different types of methanogenesis.

Figure 2 Phylogenetic placement of 82 reconstructed metagenome-assembled genomes. A maximum-likelihood phylogenomic tree was built based on concatenated amino acid sequences of 43 conserved single copy genes using RAxML with the PROTGAMMALG model. Sequences of *Altiarchaeales* ex4484_43 were used as an outgroup. The scale bar represents 1 amino acid substitution per sequence position. Bootstrap values > 70% are indicated. Blue for Site E26 (E26_binX), red for Site E29 (E29_binY), and green for Site E44 (E44_binZ).

Figure 3 Identification of functional genes or pathways present in MAGs. The presence of genes or pathways are indicated by orange shaded boxes. Aor, aldehyde:ferredoxin oxidoreductase; Kor, 2-oxoglutarate/2-oxoacid ferredoxin oxidoreductase; Por, pyruvate:ferredoxin oxidoreductase; Ior, indolepyruvate ferredoxin oxidoreductase. GHs, glycoside hydrolases. Presence of the Embden-Meyerhof-Parnas pathway is indicated if at least five genes were detected. Presence of the Beta-oxidation pathway is indicated if at least three genes were detected. AssA, catalytic subunit of alkylsuccinate synthase. CmdA, catalytic subunit of p-cymene dehydrogenase; AhyA, catalytic subunit of alkane C2-methylene hydroxylase. Additional detail for the central benzoyl-CoA degradation pathway can be found in Figure 4. Presence of the Wood-Ljungdahl pathway is indicated if at least four genes were detected.

Presence of the TCA cycle is indicated if at least six key genes were detected. Lactate and ethanol fermentation are indicated if genes encoding respective dehydrogenases were detected. H_2ase, hydrogenase. DsrAB, dissimilatory sulfite reductase. More details about these functional genes and pathways can be found in the text and in Table S6.

Figure 4 Evidence for anaerobic hydrocarbon degradation in MAGs. (a) Phylogenetic relationship of identified genes in MAGs with currently known alkyl-arylalkylsuccinate synthases based on the respective catalytic alpha-subunits. Ass/Mas, n-alkanes (1-methylalkyl) succinate synthase; Nms, 2-naphthylmethyl succinate synthase; Bss, benzyl succinate synthase; Ibs, 4-isopropylbenzyl succinate synthase; Hbs, 4-hydroxybenzyl succinate synthase. Sequences of pyruvate formate lyase (Pfl) from *E. coli* were used as an outgroup. The scale bar represents 0.1 amino acid substitutions per sequence position. Bootstrap values > 70% are indicated. The full sequences can be found in Text S1. (b) Summary of identified enzymes involved in central benzoyl-CoA processing in anaerobic aromatic hydrocarbon biodegradation. The MAGs were shown only if it was at least partially complete (presence of at least three subunits within one cluster for BcrABCD). Presence of genes or pathways are indicated by green boxes. Bcr, benzoyl-CoA reductase; Oah, 6-oxo-cyclohex-1-ene-carbonyl-CoA hydrolase; Dch, cyclohex-1,5-diencarbonyl-CoA hydratase; Had, 6-hydroxycyclohex-1-ene-1-carbonyl-CoA dehydrogenases.

Figure 5 Thermodynamic constraints on anaerobic benzoate and hexadecane degradation in deep sea sediments. Two possible scenarios are illustrated depending on the end products based on metabolic predictions in Figure 3: (1) fermentation with production of hydrogen and acetate, and (2) fermentation with production of acetate alone. Thermodynamics for each

reaction are indicated by a line in its corresponding color. If $\Delta G' < 0$, the reaction is energetically favorable (yellow-shaded area), and if $\Delta G' > 0$ the reaction is assumed not to occur. The graph shows that $\Delta G'$ for hexadecane fermentation to acetate alone (green reaction) will not reach negative values unless the concentration for acetate is extremely low (far lower than the detection limit for acetate of 2.5 μM in this study, see dash line) such that the other three reactions are more realistic scenarios for anaerobic hydrocarbon degradation in the marine sediments studied here.

Figure 6 Common organotrophic and hydrogenotrophic potentials of three hydrocarbon-impacted microbial communities inferred from metagenomics and metabolomics.

Figure 1

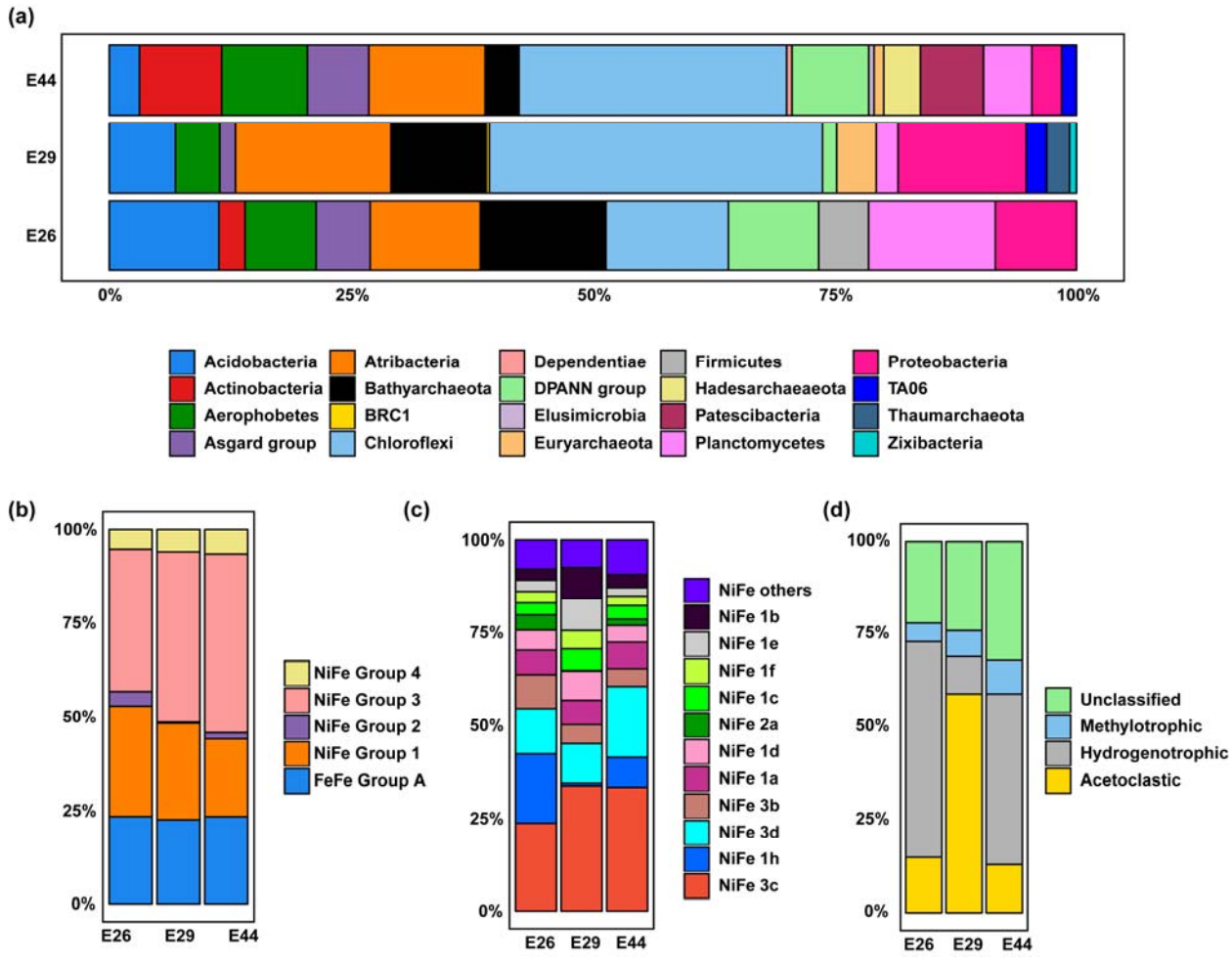

(a)
E44
E29
E26
0%
25%
50%
75%
100%
Acidobacteria
Actinobacteria
Aerophobetes
Asgard group
Atribacteria
Bathyarchaeota
BRC1
Chloroflexi
Dependentiae
DPANN group
Elusimicrobia
Euryarchaeota
Firmicutes
Hadesarchaeaeota
Patescibacteria
Planctomycetes
Proteobacteria
TA06
Thaumarchaeota
Zixibacteria
(b)
NiFe Group 4
NiFe Group 3
NiFe Group 2
NiFe Group 1
FeFe Group A
(c)
NiFe others
NiFe 1b
NiFe 1e
NiFe 1f
NiFe 1c
NiFe 2a
NiFe 1d
NiFe 1a
NiFe 3b
NiFe 3d
NiFe 1h
NiFe 3c
(d)
Unclassified
Methylotrophic
Hydrogenotrophic
Acetoclastic

Figure 2

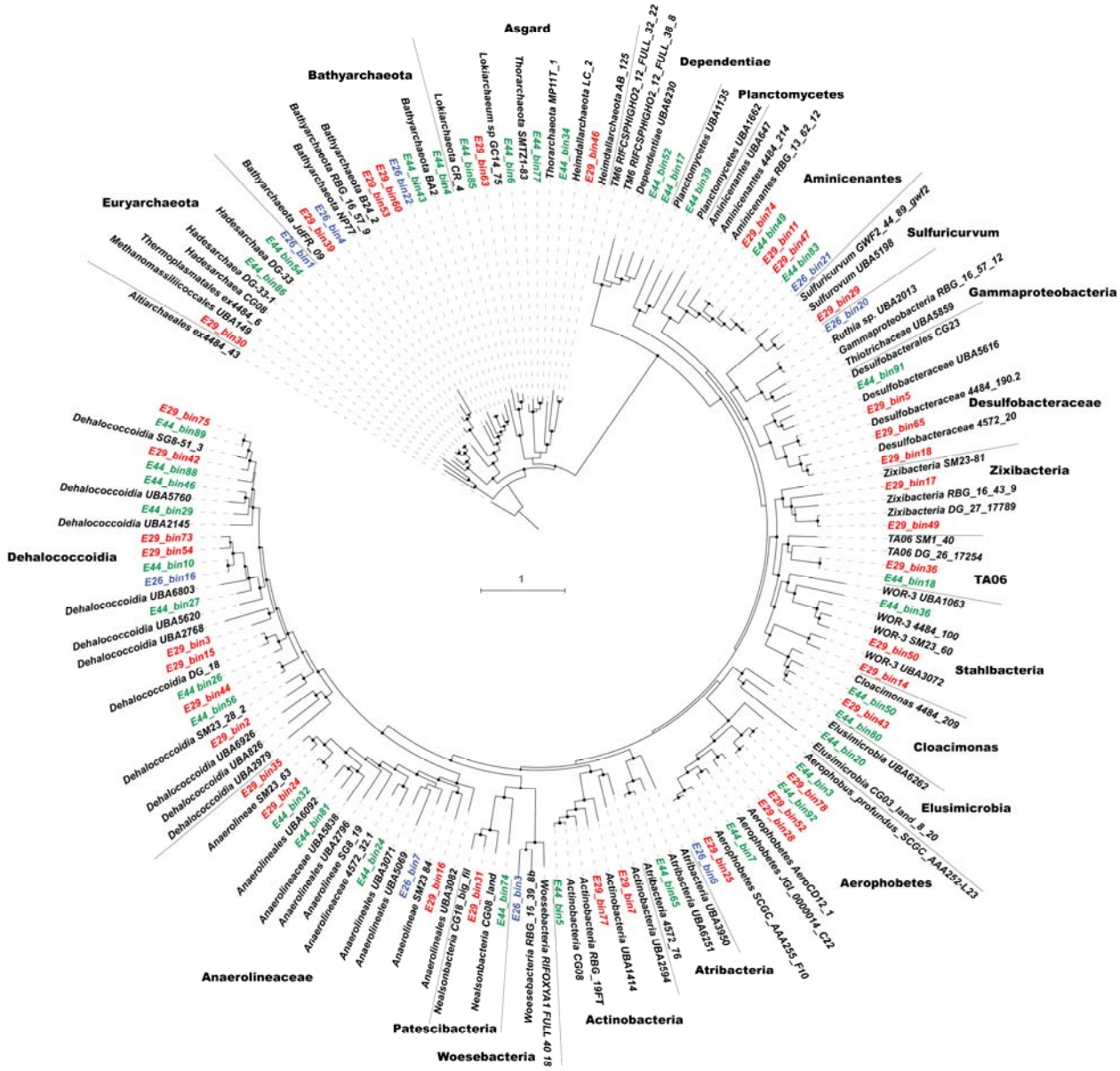

Asgard
Dependentiae
Planctomycetes
Aminicenantes
Sulfuricurvum
Gammaproteobacteria
Desulfobacteraceae
Zixibacteria
TA06
Stahlbacteria
Cloacimonas
Elusimicrobia
Aerophobetes
Atribacteria
Actinobacteria
Woesebacteria
Patescibacteria
Anaerolineaceae
Dehalococcoidia
Euryarchaeota
Bathyarchaeota
1

Figure 3

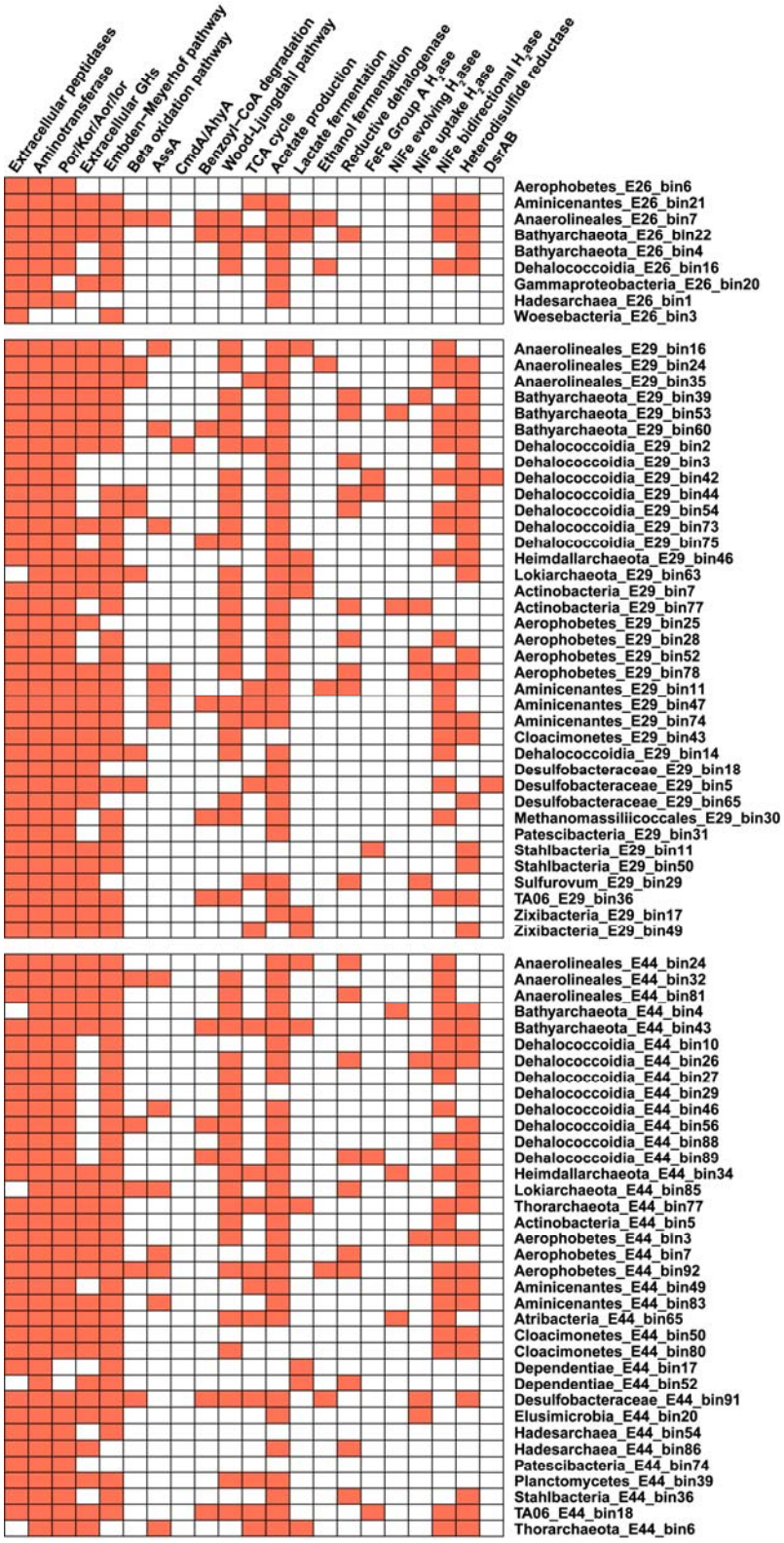

Extracellular peptidases
Aminotransferase
Por/Kor/Aor/Ior
Extracellular GHs
Embden-Meyerhof pathway
Beta oxidation pathway
AssA
CmdA/AhyA
Benzoyl-CoA degradation
Wood-Ljungdahl pathway
TCA cycle
Acetate production
Lactate fermentation
Ethanol fermentation
Reductive dehalogenase
FeFe Group A H2ase
NiFe evolving H2ase
NiFe uptake H2ase
NiFe bidirectional H2ase
Heterodisulfide reductase
DsrAB
Aerophobetes_E26_bin6
Aminicenantes_E26_bin21
Anaerolineales_E26_bin7
Bathyarchaeota_E26_bin22
Bathyarchaeota_E26_bin4
Dehalococcoidia_E26_bin16
Gammaproteobacteria_E26_bin20
Hadesarchaea_E26_bin1
Woesebacteria_E26_bin3
Anaerolineales_E29_bin16
Anaerolineales_E29_bin24
Anaerolineales_E29_bin35
Bathyarchaeota_E29_bin39
Bathyarchaeota_E29_bin53
Bathyarchaeota_E29_bin60
Dehalococcoidia_E29_bin2
Dehalococcoidia_E29_bin3
Dehalococcoidia_E29_bin42
Dehalococcoidia_E29_bin44
Dehalococcoidia_E29_bin54
Dehalococcoidia_E29_bin73
Dehalococcoidia_E29_bin75
Heimdallarchaeota_E29_bin46
Lokiarchaeota_E29_bin63
Actinobacteria_E29_bin7
Actinobacteria_E29_bin77
Aerophobetes_E29_bin25
Aerophobetes_E29_bin28
Aerophobetes_E29_bin52
Aerophobetes_E29_bin78
Aminicenantes_E29_bin11
Aminicenantes_E29_bin47
Aminicenantes_E29_bin74
Cloacimonetes_E29_bin43
Dehalococcoidia_E29_bin14
Desulfobacteraceae_E29_bin18
Desulfobacteraceae_E29_bin5
Desulfobacteraceae_E29_bin65
Methanomassiliicoccales_E29_bin30
Patescibacteria_E29_bin31
Stahlbacteria_E29_bin11
Stahlbacteria_E29_bin50
Sulfurovum_E29_bin29
TA06_E29_bin36
Zixibacteria_E29_bin17
Zixibacteria_E29_bin49
Anaerolineales_E44_bin24
Anaerolineales_E44_bin32
Anaerolineales_E44_bin81
Bathyarchaeota_E44_bin4
Bathyarchaeota_E44_bin43
Dehalococcoidia_E44_bin10
Dehalococcoidia_E44_bin26
Dehalococcoidia_E44_bin27
Dehalococcoidia_E44_bin29
Dehalococcoidia_E44_bin46
Dehalococcoidia_E44_bin56
Dehalococcoidia_E44_bin88
Dehalococcoidia_E44_bin89
Heimdallarchaeota_E44_bin34
Lokiarchaeota_E44_bin85
Thorarchaeota_E44_bin77
Actinobacteria_E44_bin5
Aerophobetes_E44_bin3
Aerophobetes_E44_bin7
Aerophobetes_E44_bin92
Aminicenantes_E44_bin49
Aminicenantes_E44_bin83
Atribacteria_E44_bin65
Cloacimonetes_E44_bin50
Cloacimonetes_E44_bin80
Dependentiae_E44_bin17
Dependentiae_E44_bin52
Desulfobacteraceae_E44_bin91
Elusimicrobia_E44_bin20
Hadesarchaea_E44_bin54
Hadesarchaea_E44_bin86
Patescibacteria_E44_bin74
Planctomycetes_E44_bin39
Stahlbacteria_E44_bin36
TA06_E44_bin18
Thorarchaeota_E44_bin6

Figure 4

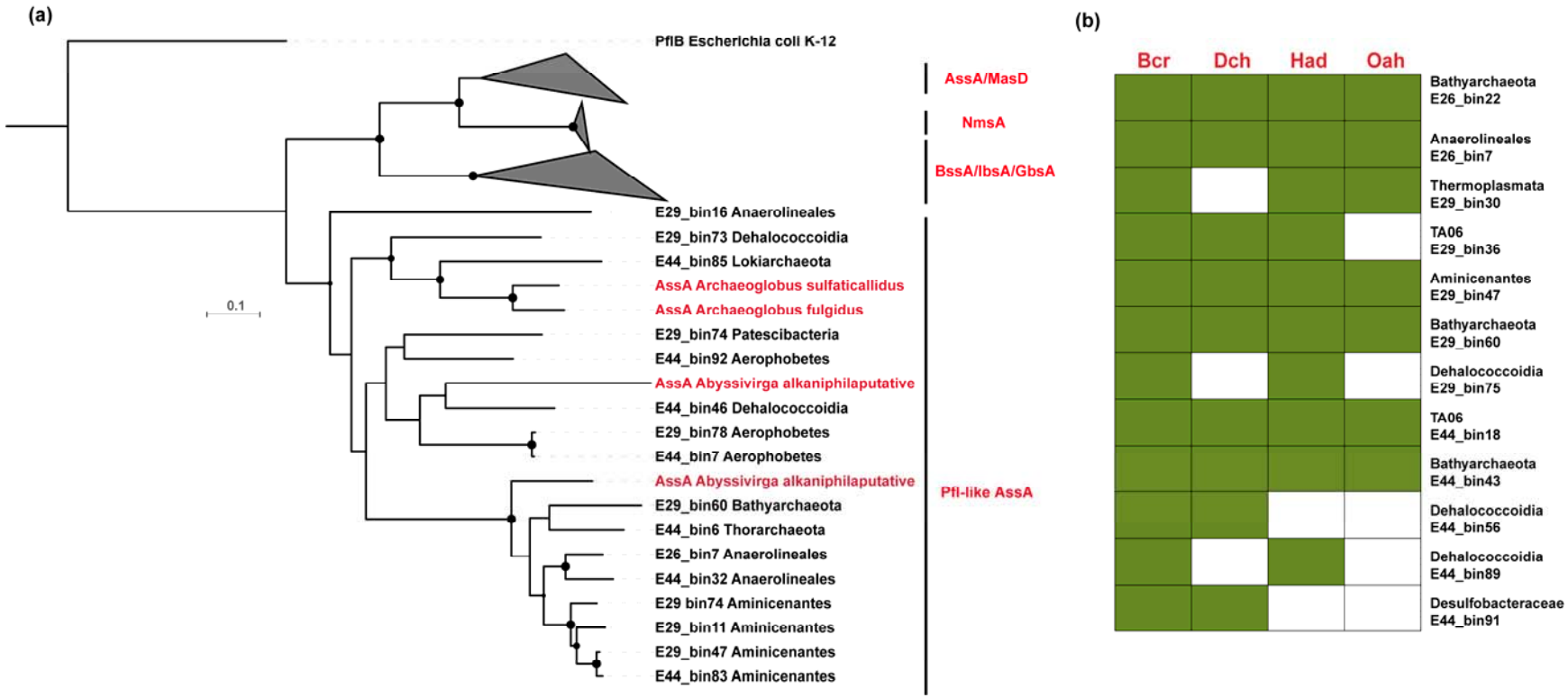

(a)
PflB Escherichia coli K-12
AssA/MasD
NmsA
BssA/IbsA/GbsA
E29_bin16 Anaerolineales
E29_bin73 Dehalococcoidia
E44_bin85 Lokiarchaeota
AssA Archaeoglobus sulfaticallidus
AssA Archaeoglobus fulgidus
E29_bin74 Patescibacteria
E44_bin92 Aerophobetes
AssA Abyssivirga alkaniphilaputative
E44_bin46 Dehalococcoidia
E29_bin78 Aerophobetes
E44_bin7 Aerophobetes
AssA Abyssivirga alkaniphilaputative
Pfl-like AssA
E29_bin60 Bathyarchaeota
E44_bin6 Thorarchaeota
E26_bin7 Anaerolineales
E44_bin32 Anaerolineales
E29 bin74 Aminicenantes
E29_bin11 Aminicenantes
E29_bin47 Aminicenantes
E44_bin83 Aminicenantes
0.1
(b)
Bcr
Dch
Had
Oah
Bathyarchaeota E26_bin22
Anaerolineales E26_bin7
Thermoplasmata E29_bin30
TA06 E29_bin36
Aminicenantes E29_bin47
Bathyarchaeota E29_bin60
Dehalococcoidia E29_bin75
TA06 E44_bin18
Bathyarchaeota E44_bin43
Dehalococcoidia E44_bin56
Dehalococcoidia E44_bin89
Desulfobacteraceae E44_bin91

Figure 5

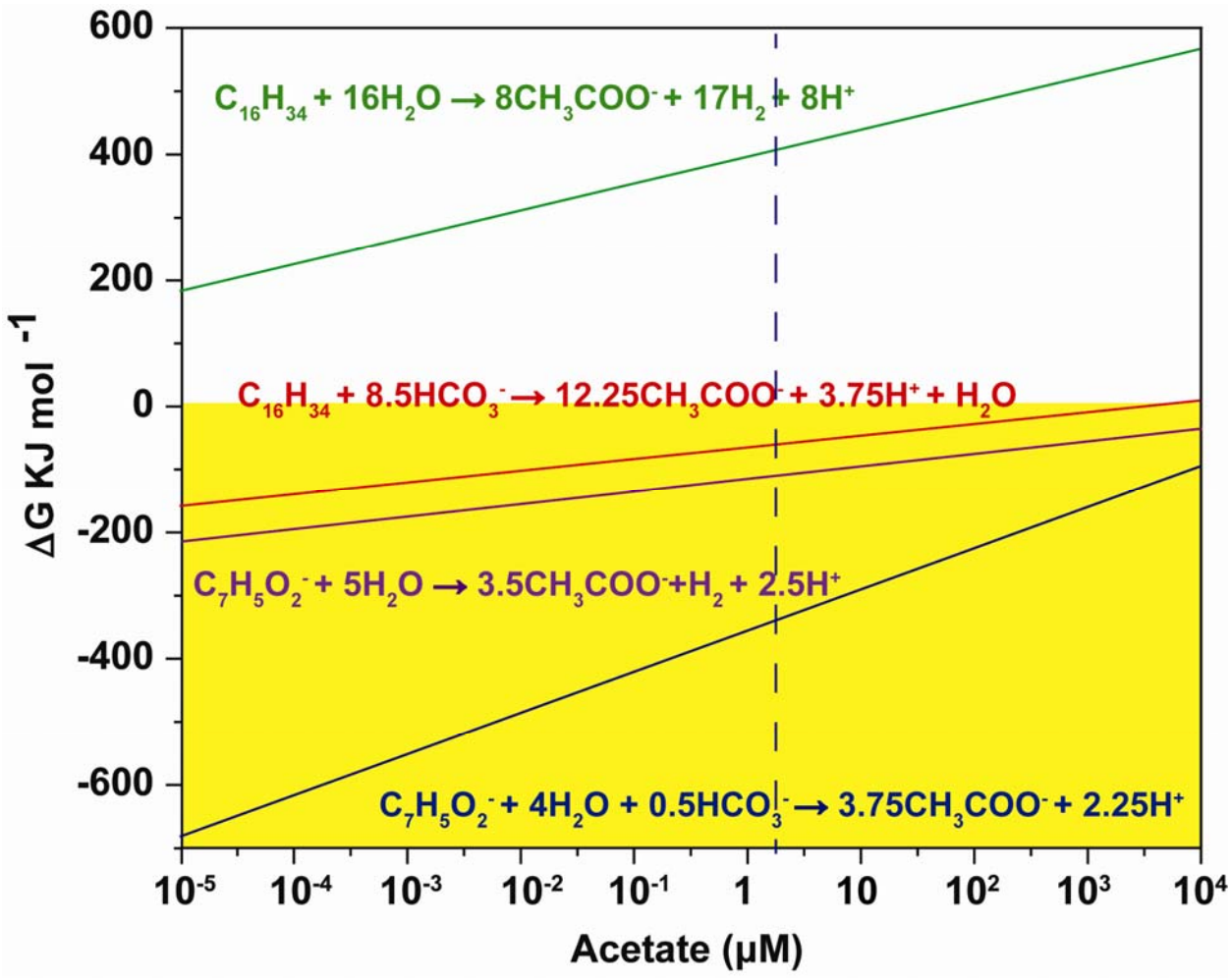
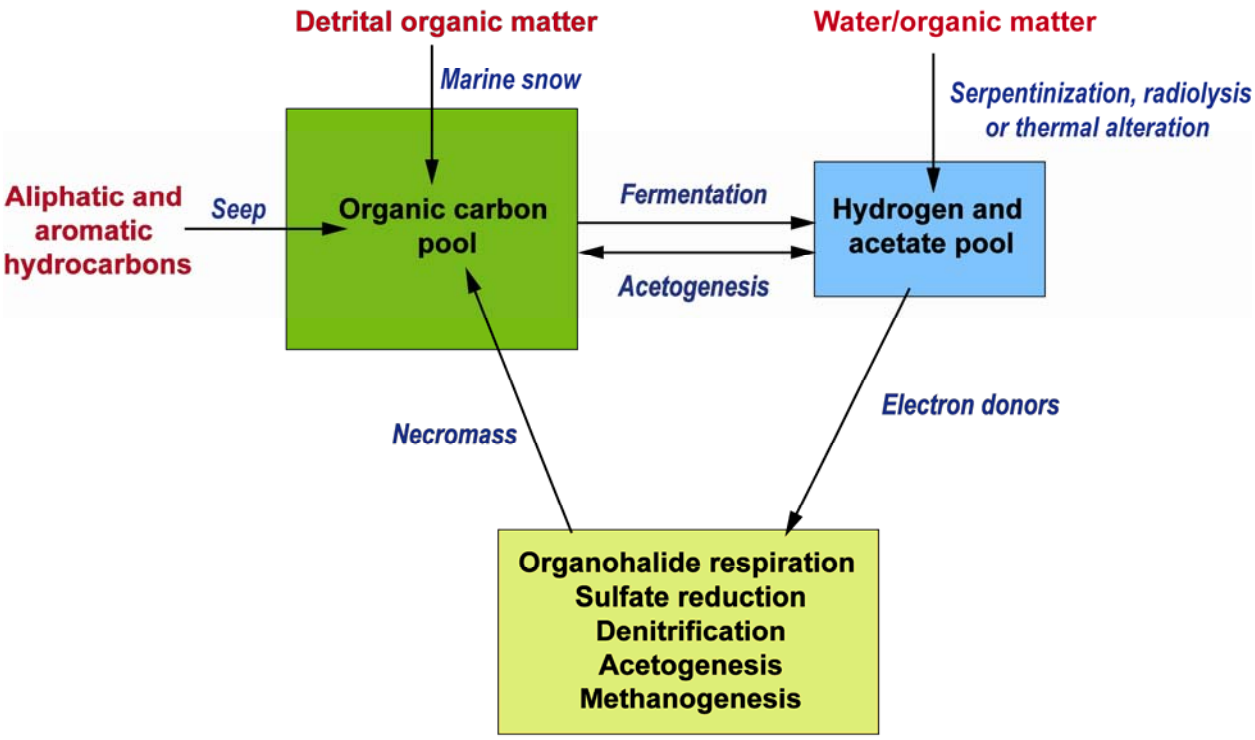

600
400
200
0
-200
-400
-600
ΔG KJ mol -1
C16H34 + 16H2O → 8CH3COO- + 17H2 + 8H+
C16H34 + 8.5HCO3- → 12.25CH3COO- + 3.75H+ + H2O
C7H5O2- + 5H2O → 3.5CH3COO-+H2 + 2.5H+
C7H5O2- + 4H2O + 0.5HCO3- → 3.75CH3COO- + 2.25H+
10-5
10-4
10-3
10-2
10-1
1
10
102
103
104
Acetate (μM)

Figure 6


Detrital organic matter
Marine snow
Water/organic matter
Serpentinization, radiolysis or thermal alteration
Aliphatic and aromatic hydrocarbons
Seep
Organic carbon pool
Fermentation
Acetogenesis
Hydrogen and acetate pool
Electron donors
Necromass
Organohalide respiration
Sulfate reduction
Denitrification
Acetogenesis
Methanogenesis