

ORIGINAL ARTICLE

Metabolic interdependencies between phylogenetically novel fermenters and respiratory organisms in an unconfined aquifer

Kelly C Wrighton¹, Cindy J Castelle², Michael J Wilkins^{1,3}, Laura A Hug², Itai Sharon², Brian C Thomas², Kim M Handley⁴, Sean W Mullin², Carrie D Nicora⁵, Andrea Singh², Mary S Lipton⁵, Philip E Long⁶, Kenneth H Williams⁶ and Jillian F Banfield^{2,5}

¹Department of Microbiology, The Ohio State University, Columbus, OH, USA; ²Department of Earth and Planetary Science, University of California, Berkeley, Berkeley, CA, USA; ³School of Earth Sciences, The Ohio State University, Columbus, OH, USA; ⁴Department of Ecology and Evolution, University of Chicago, Chicago, IL, USA; ⁵Pacific Northwest National Laboratory, Department of Energy, Biological Sciences Department, Richland, WA, USA and ⁶Lawrence Berkeley National Laboratory, Department of Energy, Berkeley, CA, USA

Fermentation-based metabolism is an important ecosystem function often associated with environments rich in organic carbon, such as wetlands, sewage sludge and the mammalian gut. The diversity of microorganisms and pathways involved in carbon and hydrogen cycling in sediments and aquifers and the impacts of these processes on other biogeochemical cycles remain poorly understood. Here we used metagenomics and proteomics to characterize microbial communities sampled from an aquifer adjacent to the Colorado River at Rifle, CO, USA, and document interlinked microbial roles in geochemical cycling. The organic carbon content in the aquifer was elevated via acetate amendment of the groundwater occurring over 2 successive years. Samples were collected at three time points, with the objective of extensive genome recovery to enable metabolic reconstruction of the community. Fermentative community members include organisms from a new phylum, Melainabacteria, most closely related to Cyanobacteria, phylogenetically novel members of the Chloroflexi and Bacteroidales, as well as candidate phyla genomes (OD1, BD1-5, SR1, WWE3, ACD58, TM6, PER and OP11). These organisms have the capacity to produce hydrogen, acetate, formate, ethanol, butyrate and lactate, activities supported by proteomic data. The diversity and expression of hydrogenases suggests the importance of hydrogen metabolism in the subsurface. Our proteogenomic data further indicate the consumption of fermentation intermediates by Proteobacteria can be coupled to nitrate, sulfate and iron reduction. Thus, fermentation carried out by previously unknown members of sediment microbial communities may be an important driver of nitrogen, hydrogen, sulfur, carbon and iron cycling.

The ISME Journal (2014) 8, 1452–1463; doi:10.1038/ismej.2013.249; published online 13 March 2014

Subject Category: Integrated genomics and post-genomics approaches in microbial ecology

Keywords: metagenomics; proteomics; candidate phylum; hydrogenase; fermentation; sulfate reduction; microbial diversity

Introduction

Carbon turnover in reservoirs, such as the ocean, soil and subsurface sediments, occurs through a range of abiotic and biotic processes operating over highly divergent timescales, with short to very long-term impacts on atmospheric chemistry and global

climate. Information on microbial roles in acquisition, transformation and exchange of carbon and other resources is needed as inputs for global carbon-cycling models (Riley *et al.*, 2011; Grant *et al.*, 2012). At this time, few studies have focused on the microbial membership or carbon-induced biogeochemical cycling in anoxic sediments, which may contain a substantial fraction of Earth's biomass (Whitman *et al.*, 1998).

Stimulation of subsurface regions by organic carbon amendments can selectively increase the activity and abundance of selected microorganisms, thus more easily enabling 'omics'-based studies of microbial biogeochemical cycling in sediments.

Correspondence: J Banfield, Department of Earth and Planetary Sciences, Department of Environmental Science, Policy, and Management, University of California, 369 McCone Hall, Berkeley, CA 94720, USA.

E-mail: jbanfield@berkeley.edu

Received 10 July 2013; revised 7 November 2013; accepted 1 December 2013; published online 13 March 2014

Previously, we demonstrated that 2 years of successive acetate amendment to groundwater from a site of former heavy-metal contamination enabled recovery and physiological prediction for 49 members from previously genomically un-sampled and uncultivated bacterial phyla, designated as candidate phyla (CP). The results uncovered an obligatory fermentation-based lifestyle in CP organisms including members of SR1 (previously referred to as ACD80), OP11, OD1, PER, BD1-5 and WWE3 (three genomes previously assigned to OD1; Wrighton *et al.*, 2012). This finding was later supported by complete genomes for members of SR1, OD1 and WWE3 recovered from acetate-stimulated sediments from this same site (Kantor *et al.*, 2013). These phyla have been identified from non-carbon amended sediments from the same metal-contaminated aquifer (Castelle *et al.*, 2013) and pristine environments (Bri e *et al.*, 2007; Peura *et al.*, 2012).

Here, we describe the organisms responsible for biogeochemical processes ongoing in the aquifer via comprehensive analysis of the entire community metagenomic data set. We use proteogenomic-enabled metabolic analyses to identify roles for uncultivated and previously uncharacterized members of the subsurface bacterial community. We have partitioned biogeochemical traits into functional guilds and identified the metabolic interdependencies across the microbial community. The results uncover an integrated web of bacterial (and phage) interactions and linking fermentation and respiratory metabolisms to carbon, hydrogen, sulfur, nitrogen and iron cycling in the aquifer.

Materials and methods

Sampling, sequencing and assembly

Acetate, an abundant fermentation byproduct in anoxic systems, was added to groundwater in an alluvial aquifer adjacent to the Colorado River, CO, USA in a region previously stimulated the year before by acetate addition (Supplementary Figure S1A). Microbial community samples (denoted sample A, C and D) were collected 5, 7 and 10 days after the start of acetate addition to the aquifer at the Rifle Integrated Field Research Challenge site in Colorado, in August 2008. At the time of sampling during the second year stimulation (samples A, C, D), acetate concentrations ranged from ~0.6 to 1.2 mM (Supplementary Figure S1A). Additional details of the geochemistry and trace elements are provided (Supplementary Figure S1 and Williams *et al.*, 2011).

Previously, we reported details on the microbial community assembly and sampling (Wrighton *et al.*, 2012) and have summarized the methods here. Microbial cells from pumped groundwater that passed through a 1.2 µm pre-filter, but not a 0.2 µm filter, were frozen immediately upon collection for DNA and protein extractions. Illumina sequences

from DNA extracted from each of three samples were assembled individually, then co-assembled (denoted as assembly ACD). A total of 24 Gbp was used in the final iterative Velvet co-assembly (Sharon *et al.*, 2013).

From assembly to genomes: ESOM-based binning

Genome fragments were clustered using emergent self-organizing map (ESOM) analysis of their tetranucleotide sequence composition (Dick *et al.*, 2009). The primary map structure was established using 5-kb fragments (all fragments >10 kb were subdivided into 5-kb segments). In addition to tetranucleotide information, we projected the relative abundance of the genome fragments in early to late samples (ratio (A)/(D)) onto the ESOM to further differentiate clusters. To avoid potential mis-binnings, protein-coding genes associated with a genome bin were confirmed to have the same phylogenetic affiliation, guanine-cytosine content and coverage as the predominant core genes in the genome bin, and to match 16S ribosomal RNA (rRNA) genes where available. For additional details, see Supplementary Online Materials (SOM).

Genome annotation and proteomics

Details of the annotation and proteomic analysis were published previously (Wrighton *et al.*, 2012); information is also included in the SOM. Genes were predicted and annotated on assembled contigs in each genomic bin. These predicted proteins formed a database that was searched via SEQUEST with collected 2D-LC-MS/MS data from extracted biomass. SEQUEST peptide identifications were filtered using MSGF cutoffs (1e-10), with spectral count data for each identified protein subsequently normalized using NSAF calculations.

Genome bin completion estimates and phylogenetic assignment

Our primary method for assessing genome completeness was based on the presence or absence of orthologous groups representing a core gene set that typically occur only once per genome and are widely conserved among bacteria and archaea (Raes *et al.*, 2007). For ESOM bins with more than one genome, protein-coding genes were assigned to specific organisms within the bin by coverage, phylogenetic identity and guanine-cytosine content.

Open-access database for genome analyses

A summary of ESOM bin size, guanine-cytosine content and phylogenetic identity is provided at http://ggkbase.berkeley.edu/Rifle_ACD/organisms. All genomic data are publically available using this web service. ggKbase is designed around 'live data,' whereby projects are continuously updated and improved (updates may include bin content and

improvements to functional predictions; the project name, organism names and gene names remain consistent). We used the ‘lists’ and ‘genome summary’ functions to assess genome completeness and profile metabolic traits. Additional details on analysis and annotation are included in the SOM.

Results and Discussion

Generation of draft genome bins representing phylogenetically diverse lineages

The ESOM defined draft genome bins for many organisms (Supplementary Figure S2). We recovered 87 genome bins, of which four were identified as phage (ACD33, ACD84, ACD85 and ACD86) and two as potential mobile elements (ACD71 and ACD74). The remaining bins were bacterial; no genomic bins were affiliated with Archaea. Previously, we reported 49 CP genomes (Wrighton *et al.*, 2012). Here we focus on the remaining genomes and the proteomics-inferred metabolic networks that occur across the community. A summary table of the proteomic data are provided (Supplementary Table S1).

Bins were taxonomically assigned based on a coherent phylogenetic signal from single-copy gene markers, concatenated ribosomal protein trees and 16S rRNA gene trees, when recovered. We estimated genome bins to be near-complete if they contain >75% of 35 single marker genes (Supplementary Figure S3). The 20 near-complete genomes are visualized on a concatenated ribosomal protein tree, with the genomes first reported in this article highlighted in red, prior CP genomes reported (Wrighton *et al.*, 2011) are also denoted (Figure 1). For a detailed concatenated phylogenetic tree see Supplementary Figure 4.

Twenty-two of the genomes reported here can be assigned to phyla with cultivated representatives and prior genomic sampling including the Proteobacteria (19), Bacteroidetes (1), Chloroflexi (1) and Firmicutes (1) (Table 1). With the exception of three proteobacterial genomes, these ACD genomes were divergent from those previously sequenced (Figure 1). AAI comparisons between our ACD genomes to nearest neighbor genomes supported this conclusion. Only ACD10 (*Dechloromonas* spp., AAI 76 ± 17 , Supplementary Figure S5b), ACD54

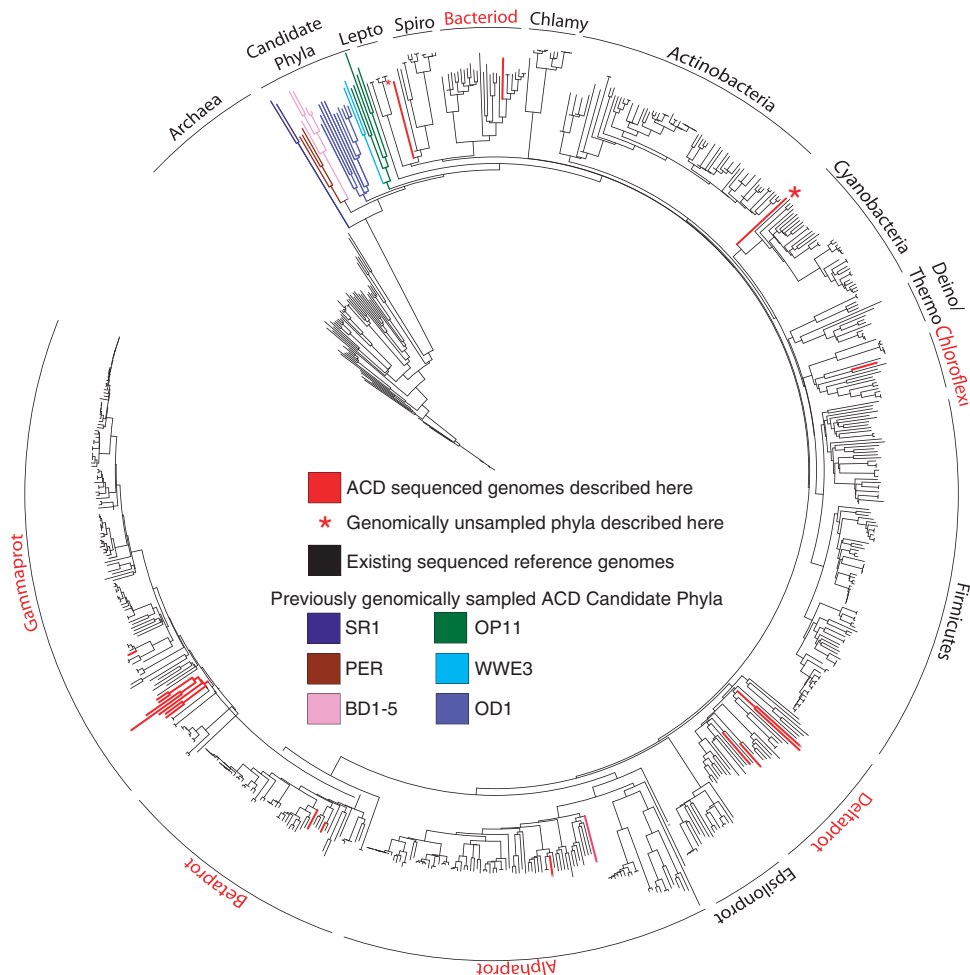


Figure 1 Maximum likelihood phylogenetic tree generated from concatenation of 16 ribosomal predicted proteins. The near-complete ACD sequences are shown in red with inferred taxonomic assignment summarized in Table 1.

Table 1 The phylogenetic affiliations of organisms discussed in this article

Near-complete single-copy genomes	Phylum	Class	Order	Family	Genus	Nearest neighbor
ACD77	Bacteroidetes	Bacteroidetes	Bacteroidia	Bacteroidales	Undefined	<i>Alistipes</i> and <i>Rikenella</i> spp.
ACD17 ^a	Chlamydiae	Chlamydii	Chlamydiales	Undefined	Undefined	<i>Simkania negevensis</i> strain Z
ACD79 ^a	Undefined	Undefined	Undefined	Undefined	Undefined	No clear phylogenetic signal (SI)
ACD34	Chloroflexi	Anaerolineae	Anaerolineales	Undefined	Undefined	<i>Anaerolinea thermophila</i> UNI-1
ACD54	Proteobacteria	Alpha	Rhodobacterales	Rhodobacteraceae	<i>Rhodobacter</i>	<i>Rhodobacter</i> sp. SW2
ACD16	Proteobacteria	Alpha	Rhodospirillales	Undefined	Undefined	Members of Rhodospirillales
ACD23	Proteobacteria	Beta	Burkholderiales	Comamonadaceae	Undefined	Members of the Comamonadaceae
ACD10	Proteobacteria	Beta	Rhodocyclales	Rhodocyclaceae	<i>Dechloromonas</i>	<i>Dechloromonas aromatica</i> RCB
ACD75_I	Proteobacteria	Delta	Desulfobacterales	Desulfobulbaceae	Undefined	<i>Desulfotalea</i> LVS4
ACD62	Proteobacteria	Delta	Undefined	Undefined	Undefined	ACD73 (no defined taxonomic affiliation)
ACD73	Proteobacteria	Delta	Undefined	Undefined	Undefined	ACD62 (no defined taxonomic affiliation)
ACD75_II	Proteobacteria	Delta	Syntrophobacterales	Undefined	Undefined	<i>Syntrophus aciditrophicus</i> SB
ACD21	Proteobacteria	Gamma	Legionellales	Coxellaceae	Undefined	(ACD69) and <i>Coxiella</i> spp.
ACD42	Proteobacteria	Gamma	Legionellales	Coxellaceae	Undefined	<i>Coxiella</i> spp.
ACD69	Proteobacteria	Gamma	Legionellales	Coxellaceae	Undefined	ACD21 and <i>Coxiella</i> spp.
ACD45	Proteobacteria	Gamma	Legionellales	Coxellaceae	Undefined	ACD60 and <i>Rickettsiella grylli</i>
ACD60	Proteobacteria	Gamma	Legionellales	Coxellaceae	Undefined	ACD45 and <i>Rickettsiella grylli</i>
ACD44	Proteobacteria	Gamma	Legionellales	Coxellaceae	Undefined	<i>Rickettsiella grylli</i>
ACD6_II	Proteobacteria	Gamma	Legionellales	Coxellaceae	Undefined	<i>Rickettsiella grylli</i>
ACD6_I	Proteobacteria	Gamma	Pseudomonadales	Moraxellaceae	<i>Acinetobacter</i>	<i>Acinetobacter lwoffii</i> SH145
ACD64 ^a	TM6	Undefined	Undefined	Undefined	Undefined	First genomic sampling
ACD20, Novel	Melainabacteria	Undefined	Undefined	Undefined	Undefined	Novel phylum near Cyanobacteria
ACD58, Novel	Berkelbacteria	Undefined	Undefined	Undefined	Undefined	Novel phylum
ACD47, Novel	Undefined	Undefined	Undefined	Undefined	Undefined	Novel phylum near Spirochaetes

Organisms are assigned to the deepest possible taxonomic level whereby the organism was placed in a defined group with >80% bootstrap support. Taxonomic levels for which no affiliation could be assigned based on existing reference genomes are marked as undefined. The 'Nearest neighbor' column refers to the taxonomy of publically available genomic sequences closest in similarity to ACD.

^aDenotes a partial genome with phylogenetic identity based on single-copy gene or 16S rRNA analyses.

(*Rhodobacter* spp., AAI 69 ± 16 ; Supplementary Figure S5a) and the dominant genome in the ACD6 bin (*Acinetobacter* spp., Supplementary Figure S5c) had AAI values approaching 85%, which is the criterion used by Goris *et al.* (2007) to assign organisms to the same species.

Three additional genomes included here are members of phyla that lack cultivated representatives. A recovered 16S rRNA gene places ACD64, a partial genome bin, in the TM6 lineage and ACD58, previously reported as a divergent OD1 in Wrighton *et al.* (2012), is now recognized as a separate phylum level lineage, here named Berkelbacteria. For another near-complete genome, ACD20, we used a combination of core and 16S rRNA genes (<85% identity) assign the organism to a novel phylum called Melainabacteria, most closely related to the Cyanobacteria (Di Rienzi *et al.*, 2013; Figure 1). Another near-complete genome, ACD47, lacks a 16S rRNA gene sequence but, based on our concatenated protein phylogenetic analyses, this genome represents a previously genomically unsampled phylum (Figure 1, Supplementary Figure S5e). An additional seven genomic bins lacked sufficient markers for resolvable classification (for example, ACD79) and are reported as unknown. In this category, ACD17 had some conserved markers related to Chlamydiae (SOM). Overall, the results presented here underline

the phylogenetic novelty of these subsurface groundwater samples.

Community metabolic genomic potential and expression

Carbon degradation by phylogenetically novel bacteria. In light of the importance of fermentation previously identified in CP genomes (Wrighton *et al.*, 2012; Kantor *et al.*, 2013), we screened the entire community for genes encoding the hydrolysis of plant-derived materials and chitin, the two most abundant biopolymers (Berlemont and Martiny, 2013). We identified members of the OP11 (ACD38), OD1 (ACD8) and a genomic fragment with unknown affiliation (ACD79, 16x coverage) that could encode the capacity for the complete degradation of cellulose to monomeric carbon, which can be oxidized via glycolysis. A gammaproteobacterial genome (ACD69) contains an exocellulase with necessary residues for functionality, but this organism lacks a β -glucosidase, required to convert the cello-oligosaccharide product to glucose (Supplementary Table S2, Figure 2).

Additional potential exocellulosic activity may be associated, albeit by putative divergent cellulases, with members of the WWE3 CP (ACD24 and ACD25), OD1, OP11 and Proteobacteria.

Endocellulase genes were also identified in an incomplete Firmicutes genome (ACD 35) and a Bacteroidales genome (ACD77). ACD cellulose-degrading organisms may shape external carbon pools, as the ACD79 cellulase (from an unknown organism) is predicted to be localized to the outer cell membrane, and the OP11 and gammaproteobacterial enzymes are predicted to be localized extracellularly.

Many organisms in the community have β -glucosidases to degrade cello-oligosaccharides, but lack genes for the initial steps in cellulose breakdown mentioned above (Supplementary Table S2). The lower incidence of pathways for complete cellulose breakdown relative to those for utilization of cellobiose is consistent with a bioinformatic study that found genomes with only cello-oligosaccharide utilization outnumbered cellulase-containing genomes two to one (Berlemont and Martiny, 2013). Compared with compost and peat wetland soil metagenomic data sets (Allgaier *et al.*, 2010; Tveit *et al.*, 2013), the ratios of hemicellulose: cellulose genes were less, suggesting different carbon sources between the systems selectively enrich at the functional level (Figure 2). Genes for chitin and *N*-acetyl-glucosaminidase (almost in a 1:1 ratio), suggest the capacity to completely degrade this biopolymer to C-N residues is encoded in up to 30 different organisms within the community (Figure 2).

Carbon degradation can occur via respiration or fermentation, with many organisms capable of both processes. Of the aforementioned genomes that have carbon degradation machinery, only those of the Bacteroidales and Chloroflexi encode the machinery to reduce oxygen, but not other electron acceptors (SOM). Despite the broad genetic potential for aerobic respiration (for example, cytochrome and quinone oxidases) across the 87 genomes, we found no proteomic evidence indicating oxygen reductases are synthesized in this data set; a finding consistent with oxygen levels below 0.1 p.p.m. detection limit

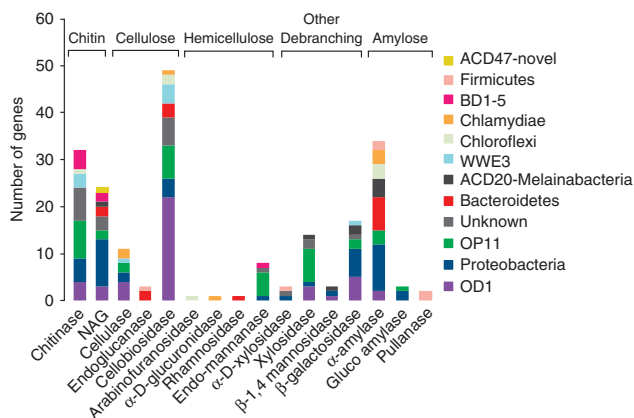


Figure 2 Inventory and phylogenetic assignment of putative glycoside hydrolases (GHs) identified in the ACD metagenome. See Supplementary Table S2 for details.

during the sampling period (Yabusaki *et al.*, 2011). The organisms capable of anaerobic respiration (discussed below) do not have extensive carbon degradation machinery. Taking these together, we suggest fermentation is the primary mode of carbon turnover in this system. Consistent with this hypothesis, a significant fraction of organisms, including Melainabacteria (ACD20) and members of the OP11 and OD1, appear to be obligately fermentative, and express genes for *in situ* carbon degradation. This inference is based on the lack of a complete tricarboxylic acid cycle, electron transport chain components and terminal reductase/oxidases (Wrighton *et al.*, 2012; Di Rienzi *et al.*, 2013).

Proteomics confirmed the synthesis of cellobiosidases (weakly supported), alpha-amylases and enzymes for monosaccharide degradation from Bacteroidales, OD1, Chloroflexi and ACD20 (Supplementary Table S1). Previously, we presented proteomic support for roles of OD1 and OP11 bacteria in the production of acetate, formate, lactate and ethanol as fermentation end products (Wrighton *et al.*, 2012). Here we expand the list of organisms capable of producing acetate (Bacteroidales, ACD20), butyrate (Bacteroidales, ACD20), ethanol (Bacteroidales, ACD20, Chloroflexi) and lactate (ACD20) (Figure 4). Overall, uncultivated and previously genomically unrepresented bacteria have key roles in carbon cycling in this aquifer; roles that could not have been assigned based on phylogeny alone or using unbinned metagenomic data.

We were initially surprised that a non-fermentable compound, like acetate, would stimulate a wide phylogenetic and metabolic diversity of fermentative bacteria. A similar phenomenon, where the addition of labile carbon-stimulated decomposition of more recalcitrant carbon, described as ‘priming’, has been well documented in soils and marine systems, but has yet to be defined in the terrestrial subsurface (Bianchi, 2011). Priming can be caused by direct or indirect mechanisms (Kuzyakov *et al.*, 2000); here we suggest the latter is more likely. We propose acetate amendment in the first year stimulated microbial blooms (Wilkins *et al.*, 2010) and this resulting biomass contained a diversity of carbon types that indirectly sustained a broader diversity of organisms after the stimulation event.

Phage may also shape carbon pools in the subsurface (Engelhardt *et al.*, 2011). Phage abundances and proteomic identifications vary over time during the stimulation event studied here (Supplementary Table S1). We recovered four partial phage genomes, all of which have genomic similarity to sequenced phage that target Gammaproteobacteria. Phage lysis of gammaproteobacterial cells (ACD44, ACD45, ACD46 and ACD60) or other taxa responding to acetate amendment may explain the decreased abundance of these taxa over time. This could have resulted in release of fermentable compounds, possibly accounting for the major increase in

obligatory fermentative bacteria at later time points (for example, BD1-5 such as ACD3; OD1, such as ACD1 and ACD5). The organismal changes in relative abundance across the samples are discussed in the SOM (Supplementary Figure S6). Surprisingly, no spacer sequences extracted from the CRISPR loci matched (even imperfectly) to any phage in the data set, raising the possibility that sample collection by filtration separated most phage from their bacterial hosts (SOM).

Hydrogen economy may link fermentation and respiratory processes. Molecular hydrogen is an important metabolic intermediary in wetlands, sewage sludge, serpentinized rocks and the intestinal track of insects and animals (Schmidt *et al.*, 2010; Ballor and Leadbetter, 2011; Brazelton *et al.*, 2012). It is possible a similar hydrogen exchange may link fermentation and respiratory metabolisms in aquifer communities, given the extensive evidence for fermentation-based metabolism identified in these genomes.

To this end, we identified all three genes in confurcating FeFe hydrogenase complexes, which are found in fermenters known to produce high molar ratios of H₂ (Schut and Adams, 2009; Sieber *et al.*, 2012). Two copies exist in the Bacteroidales (ACD77) genome, while the novel phylum Melainabacteria (ACD20) contains three copies (SOM). Phylogenetic analyses confirmed that the hydrogenase sequences were most closely related to those from fermentative organisms (Supplementary Figure S7), as such we predict both organisms ferment carbon to produce H₂ in these samples.

Unlike FeFe hydrogenases, which are generally considered to catalyze hydrogen production in fermentative bacteria, the four phylogenetically distinct groups of NiFe hydrogenases can be involved in either hydrogen production or consumption. With NiFe hydrogenases, phylogenetic affiliation may provide insight into the co-factors and potential physiological function. For instance, group 1 NiFe hydrogenases are membrane-associated enzymes commonly found in organisms that use H₂ as a donor for respiratory metabolism (Vignais and Billoud, 2007). We recovered three sequences in our data set, from genomes affiliated with *Geobacter*, the Desulfobulbaceae (ACD75) and a partial sequence on a genome fragment from a plasmid likely associated with *Geobacter* spp. (ACD74) (Supplementary Figure S8). We posit that these link hydrogen uptake to a respiratory metabolism.

The vast majority of hydrogenase catalytic subunit genes in our data set belong to group 3 (types b, c, and d) hydrogenases, which are physiologically reversible with either H₂ production or consumption. In addition to the group 3b hydrogenases previously reported from OD1 and OP11 genomes (Wrighton *et al.*, 2012), we recovered group 3b sequences from two gammaproteobacterial genomes (ACD46 and

ACD21) and a partial sequence from the ACD79 genomic bin of unknown taxonomic assignment (Supplementary Figure S8). Unlike the Archaeal sulf-hydrogenase homologs in the obligately fermentative Thermococcales, the physiological role for NADP group 3b hydrogenases in the Proteobacteria (for example, *A. vinelandii* and *T. denitrificans*) is not yet known (Beller *et al.*, 2006). The ACD79 genome bin has a second hydrogenase, and along with sequences from the Chloroflexi (ACD34) and Deltaproteobacteria (ACD 62) genomes, is most closely related to group 3c sequences (Supplementary Figure S8). A physiological role could not be predicted for the group 3b hydrogenases in ACD79, ACD46 and ACD21 or for the group 3c hydrogenase in ACD62 and ACD79.

We also identified a group 3d hydrogenase in the Chloroflexi genome (ACD34), as well as a partial group 3d sequence from a genome related to *Dechloromonas* (ACD10). We predict that the Chloroflexi hydrogenases support a fermentative, not respiratory, metabolism *in situ*, yielding hydrogen as a byproduct. This finding is supported by cultivation studies using isolates closely related to the Chloroflexi studied here (Yamada *et al.*, 2006), and a 75% shared amino-acid similarity between the Chloroflexi and ACD34 hydrogenases. Alternatively, for the ACD10 group 3d hydrogenase we infer a role in hydrogen uptake, not production, based on the high shared amino-acid identity (67%) to *Dechloromonas* spp. that have been shown to use hydrogen as an electron donor (Shrout *et al.*, 2005).

Proteomics confirmed the *in situ* synthesis of the FeFe hydrogenase from ACD20, and the NiFe Chlamydiae group 3c, OD1 group 3b and *Geobacter* group 1 hydrogenases (Supplementary Table S1). The expression of *Geobacter*-affiliated uptake hydrogenase gene detected here is consistent with proteomic investigations from other acetate stimulations at the site, where the authors concluded hydrogen may be a donor for metal reduction (Wilkins *et al.*, 2013). From the ACD data, we detected greater proteomic support for H₂ production relative to other fermentation end products (Figure 4). Together our data suggest that H₂ produced by phylogenetically novel fermentative organisms is an important ecosystem currency, potentially fueling a diversity of respiratory metabolisms in the subsurface.

Diversity of multi-heme c-type cytochromes. Multi-heme c-type cytochromes (MHCs) are metalloproteins that can have various biochemical roles, including substrate catalysis and electron transfer in many respiratory metabolisms, including anaerobic ammonia oxidation as well as nitrite, oxygen and iron reduction. Where iron-reducing bacteria require direct contact with a mineral, MHCs localized in both the periplasm and outer membrane transfer electrons across the cell envelope. In iron-reducing bacteria, the physiological importance for MHCs are

clear from genomic data, with an abundance of MHCs (average >37 per genome), each containing multiple predicted heme-binding motifs (average >6 per protein) (Wrighton *et al.*, 2011).

In addition to MHCs from *Geobacter* spp. that have been extensively researched at Rifle (Aklujkar *et al.*, 2010), we recovered 43 MHCs (only 8 belonged to *Geobacter*) and 39 mono-heme *c*-type cytochromes. Of the MHCs, 7 were predicted to be localized in the periplasm, 10 others had a signal peptide but localization could not be predicted and 3 were predicted to be extracellular (Supplementary Table S3). One of the extracellular cytochromes was associated with a phage, and another periplasmic MHC was associated with a plasmid. These results suggest that mobile elements and phage may transfer physiological capabilities amongst microorganisms in the subsurface.

Most MHCs were recovered from genomes affiliated with members of the Proteobacteria. In addition to *Geobacter* spp., we recovered MHCs from organisms most closely related to *Desulfotalea psychrophila* LVS4 (ACD75) and *Dechloromonas* spp. (ACD10), which can both grow by soluble ferric iron reduction (Knoblauch *et al.*, 1999; Weber *et al.*, 2006). Although there are multiple organisms with MHCs in the ACD75 bin, we could confidently assign four *c*-type cytochromes (with 6, 4, 4 and 1 heme) to scaffolds from organisms most closely related to *Desulfotalea*. Interestingly, the genome from the most closely related isolate (*D. psychrophila* LVS4) lacks annotated MHCs and reduces iron via a yet unknown mechanism (Rabus *et al.*, 2004).

We also recovered MHCs from phylogenetically novel organisms. Three of these MHCs contain 23, 22 or 16 hemes, and are encoded by the ACD73 genome, which represents a novel order within the Deltaproteobacteria (Supplementary Figure S5). One *c*-type cytochrome with 10 hemes is encoded by a very partial genome (ACD39) that lacks a confirmed phylogenetic affiliation. Genomes that encode multiple copies of MHCs with high heme content (>16) are rare, and are typically associated with organisms capable of reducing insoluble iron minerals (for example, *Shewanella*, *Anaeromyxobacter* and *Geobacter*; Sharma *et al.*, 2010). Our findings provide detailed and direct genomic evidence indicating that the capacity for metal reduction within the aquifer extends beyond the *Geobacteraceae*.

Sulfate and iron reduction co-occur during secondary stimulation. Sulfate reduction in the Rifle sediments and groundwater has been previously attributed to *Desulfobacter* spp. (Miletto *et al.*, 2011; Handley *et al.*, 2012, 2013), while metagenome reconstruction suggested sulfide re-oxidation was attributed to *Sulfurovum*- and *Sulfurimonas*-like Epsilonproteobacteria (Handley *et al.*, 2013). None of these genomically characterized sulfur-cycling organisms were identified in these samples, indicating the value of sampling different material

(planktonic vs sediment-attached) under varying geochemical conditions (iron reduction vs sulfate reduction) to capture the vast physiological diversity in subsurface communities.

Here we recovered key genes for the transport, activation and reduction of sulfate only from contigs with best hits to the Desulfobulbaceae (ACD75). The dominant genome within the bin (coverage >60X) is from an organism most closely related to *Desulfotalea psychrophila* LVS4. Genomic fragments with this coverage encode key genes for the sulfate transport and the activation and reduction of sulfate, including ATP sulfurylase (*sat*), APS reductase (*aprAB*) and dissimilatory sulfite reductase (*dsrABCD*). We also identified *sat* and *dsrAB* encoded on medium coverage (~40X) scaffolds and *aprAB* encoded on low coverage (5X) scaffolds from this Desulfobulbaceae-like bin (Supplementary Table S4, Supplementary Figure S9). Proteomics confirmed the *sat*, *apr* and *dsr* genes are synthesized from the three strains across all time points (Supplementary Figure S10).

In addition, the presence of a sulfite:cytochrome *c* oxidoreductase in conjunction with the DSR pathway in the medium coverage genome (~40X), may indicate the capacity for sulfur disproportionation, as occurs in other members of the Desulfobulbaceae (Finster *et al.*, 1998). Along these lines, cultures in a defined medium with elemental sulfur (S⁰) and amorphous ferric hydroxide (FeOOH) formed sulfate via disproportionation (Thamdrup *et al.*, 1993). This process is consistent with the geochemical conditions of the Rifle aquifer and requires the coexistence of sulfate, sulfide, reactive metals (FeOOH) and a high-turnover pool of elemental sulfur (Supplementary Figure S1).

Sulfate reduction by ACD75 can be coupled to the oxidation of lactate, ethanol, hydrogen and formate (Figure 3). We were not able to reconstruct a complete CODH/ACS pathway from any of our genomes. Members of the Desulfobulbaceae (for example, *D. psychrophila* LVS4) closely related to our sulfate-reducing bacteria (SRB) are incomplete oxidizers, generally not known to grow solely with acetate (Friedrich *et al.*, 2001; Knoblauch *et al.*, 1999). This is despite a complete tricarboxylic acid cycle and genes for the conversion of acetate to acetyl-CoA (Rabus *et al.*, 2004). Supporting this, *dsrB* gene transcripts from another Rifle experiment found members of the Desulfobulbaceae were constant before and during acetate stimulation (Miletto *et al.*, 2011). Our proteomics results support the previously proposed decoupling of sulfate reduction from acetate amendment for the Desulfobulbaceae, with the synthesis of sulfate reduction genes coinciding with the usage of fermentation end-products (for example, ethanol) rather than acetate (Figure 3).

Our proteogenomic findings indicating concurrent iron reduction and sulfate reduction are consistent with the studies of Druhan *et al.* (2012), who used sulfur isotopes to identify early onset of sulfate

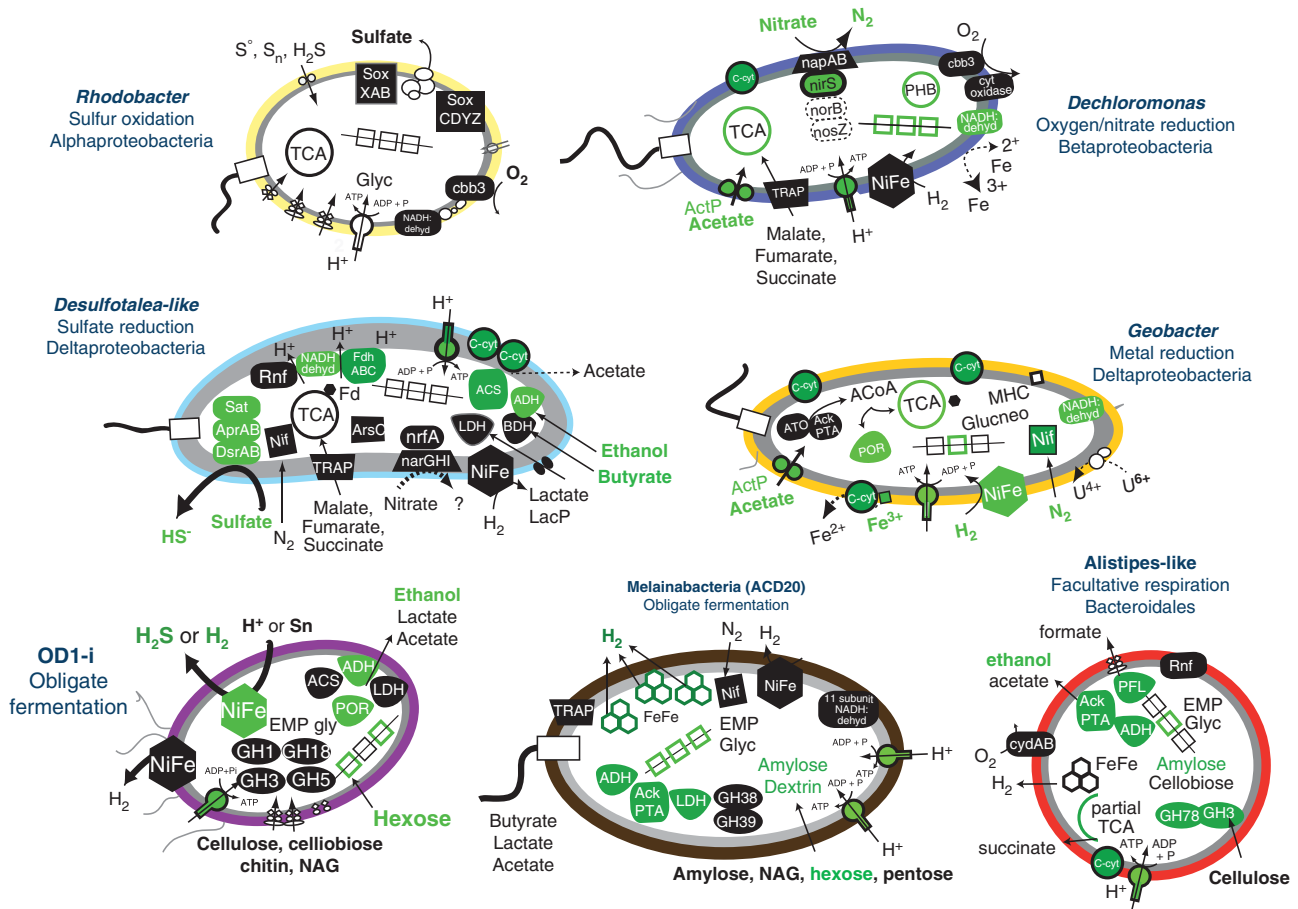


Figure 3 Genome-enabled metabolic potential of key functional members. Confirmation of protein synthesis by proteomics was denoted in green. Members of the OD1-i (purple, for example, ACD5, 7, 15), Melainabacteria (brown) and Bacteroidales (red) are inferred to ferment various carbon sources to produce hydrogen, ethanol, acetate and lactate. Fermentative end-products (for example, hydrogen, ethanol) can be consumed by respiratory members of the community including a sulfate-reducing member of the Desulfobulbaceae most closely related to *Desulfotalea* spp. (blue), *Geobacter* spp. that have been documented to reduce ferric iron and uranium at the site and *Dechloromonas* spp. that have the capacity for nitrate and oxygen reduction (dark blue). Reduced sulfide produced from the Desulfobulbaceae, and possibly OD1, can serve as an electron donor for sulfur-oxidizing *Rhodobacter* spp. (pale yellow).

reduction (during acetate stimulation) before detection of sulfide or statistically significant decreases in groundwater sulfate from this same aquifer during acetate stimulation. In addition, we offer a mechanism for co-existence of iron-reducing and SRB by niche partitioning, with *Geobacter* utilizing acetate and certain SRB consuming end-products of fermentation, thus explaining the large relative abundance (> 1:1) of Desulfobulbaceae to Geobacteraceae by 16S rRNA gene copy number (Wrighton et al., 2012) and genomic coverage in our data set. Together our findings may have important ramifications for reactive transport models (Yabusaki et al., 2011) that may currently underestimate the biomass and activity of SRB during early acetate stimulation and do not yet incorporate alternative carbon sources and hydrogen into biogeochemical predictions.

Given the proteomic support for sulfide production by SRB and also potentially from OD1 3b sulf-hydrogenases (Wrighton et al., 2012), we evaluated the potential for sulfur oxidation in the

ACD community. A genome assigned to *Rhodobacter* spp. (ACD54) contained a complete 16-gene sulfur oxidation pathway (Sox). Like other purple non-sulfur photosynthetic bacteria (for example, *Rhodovulum sulfidophilum*; Friedrich et al., 2001), ACD54 may be capable of chemolithotrophic growth in the absence of light by coupling sulfur oxidation to oxygen reduction via either the identified *aa*₃-type cytochrome *c* oxidase (80.8% AAI, *Oceanicola granulosus*) or the high oxygen affinity *cbb*₃-type oxidase (73.68% AAI, *Rhodobacter sphaeroides*). This organism may also degrade carbon compounds (Supplementary Table S2), by either an aerobic respiratory or fermentative metabolism. Genes for the reduction of alternative terminal electron acceptors or for phototrophy were not identified, but ACD54 is a partial genome and some physiology may not be sampled. Proteomics suggest that sulfur cycling by ACD54 is not physiologically active during these sampled time points or that the synthesized protein was below detection (Figure 3).

Proteobacteria are responsible for denitrification. Nitrate is the most prevalent groundwater contaminant and impacts drinking water resources on a global scale (Rivett *et al.*, 2008). Despite this, the diversity and activity of nitrate-reducing bacteria remains surprisingly understudied in the subsurface relative to other ecosystems, such as soils or wastewater treatment systems (Green *et al.*, 2010). In our samples, only members of the Proteobacteria encoded the capacity for dissimilatory nitrate reduction.

The ACD betaproteobacterial genomes are closely related to known subsurface denitrifying bacteria from *Dechloromonas* (ACD10) and members of the Comamonadaceae, most closely related to *Acidovorax* spp. (ACD23) (Coates *et al.*, 2001; Byrne-Bailey *et al.*, 2010). Organisms closely related to those studied here have been demonstrated to have a role in nitrate, as well as selenium reduction at this and other subsurface sites (Byrne-Bailey *et al.*, 2010). In ACD10, we recovered genes for *napAB* (periplasmic nitrate reductase), *nirS* (nitrite reductase), while ACD23 encodes *narG* (membrane-bound nitrate reductase), *nirS* and *nosZ* (nitrous oxide reductase) (Supplementary Table S5). Given that these are partial genomes, it might be possible that both organisms, like their nearest neighbors, completely denitrify using acetate, hydrogen and organic acids as donors. We also recovered genes potentially involved in nitrate reduction (*narGHJ*) by sulfate-reducing ACD75 Desulfobulbaceae, as well as the capacity for other nitrogen transformations by *Geobacter* spp., Bacteroidales and novel members of the Deltaproteobacteria (SOM). Proteomics confirmed the ACD10 and ACD23 *nirS* proteins were synthesized *in situ* (Supplementary Table S1). However, given proteomic evidence for nitrite reductase only, it is possible these proteins may have other functions, including the detoxification or reduction of oxygen or sulfite (Averill, 1996; Pereira *et al.*, 2000).

The ability to fix nitrogen was used to explain the increasing dominance of *Geobacter* relative to other iron-reducing bacteria during the course of biostimulation (Zhuang *et al.*, 2010). We therefore examined the capacity for nitrogen fixation in our genomes, based on the presence of the catalytic subunit of the nitrogenase (*nifH*). In addition to *Geobacter* (*nifH* confirmed by proteomics), we identified cluster III *nifH* genes from obligately fermentative ACD20 and sulfate-reducing Desulfobulbaceae ACD75 (Zehr *et al.*, 2003). Ultimately, the use of proteogenomic data to infer nitrogen cycle processes highlights the complementary value of this approach when compared with geochemical measurements that may fail to account for pathways where substrates are in low abundance or consumed in close concert with their production.

Interconnected metabolic networks are driven by phylogenetically novel organisms. There is a

limited amount of research about the encoded and manifested physiology of the vast majority of microorganisms in subsurface sediments. Here, we used community proteogenomics to predict overlapping resource utilization (that is, two species consume shared resources) and cooperative interactions (for example, where the metabolites produced by one organism are consumed by another; Figure 4). Ultimately, our research connects the functional traits of carbon, hydrogen, metal, sulfur and nitrogen cycling to a phylogenomic framework, assigning phylogenetic identity to many processes that were previously unknown or unassigned.

Our approach sheds light on the physiology of phylogenetically novel fermentative bacteria from at least six previously genomically unrepresented CP and two novel phyla. We expanded the fermentative capacity of CP (for example, OD1, WWE3 and OP11) to include possible roles in cellulose and chitin degradation, as well as assigning roles in carbon degradation and hydrogen production to a member of a new phylum, sibling to Cyanobacteria (ACD20, Melainabacteria), Bacteroidales (ACD77) and Anaerolineae (ACD34) (Figure 4). Draft genomes for members of the Anaerolineae have been documented in unamended sediments from this same aquifer (Hug *et al.*, 2013) and complete genomes from the Melainabacteria lineage were recovered from the human intestinal tract (Di Rienzi *et al.*, 2013). Conclusions from these studies are consistent with our proposed roles in carbon degradation, suggesting a conserved functionality across ecosystems may exist for some of the organisms identified here.

Our results show that the production of organic acids and hydrogen by phylogenetically novel fermenters fuels respiratory processes driven by members of the Proteobacteria. Of the fermentative metabolic end-products, the production of hydrogen and ethanol was most strongly supported by our proteomic data (Figure 4). The reasons for this may include a selective advantage gleaned by an increase in ATP production for organisms capable of hydrogen over organic acid production (Herrmann *et al.*, 2008).

A notable finding from our research was proteomic evidence for the co-existence of nitrite, sulfate and iron reduction during early acetate stimulation. Our proteogenomic results, in conjunction with isotopic data (Druhan *et al.*, 2012), is consistent with findings where carbon excess resulted in overlapping redox zones with co-occurring reduction of multiple terminal electron accepting processes (Roychoudhury and Merret, 2006). Our genomic analyses suggest that the sulfate reduction processes that occur during iron reduction phase of the aquifer may be decoupled from acetate amendment and instead rely on other fermentation end-products (for example, hydrogen and ethanol; Figure 4) and that SRB (for example, ACD75; Desulfobulbaceae), fermentative sulfur reduction via sulf-hydrogenases (for example, ACD1, 5, 15;

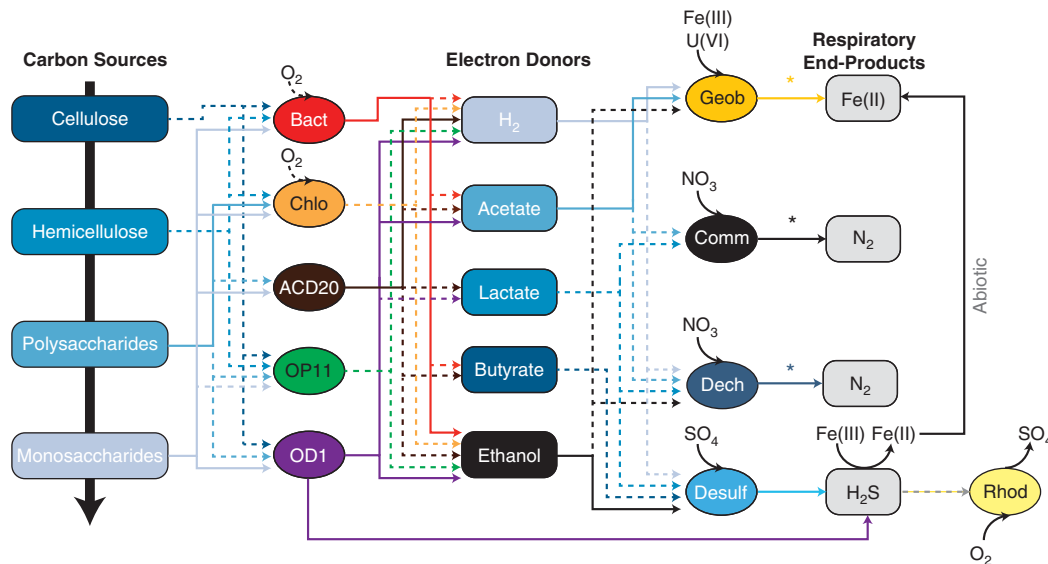


Figure 4 Predicted metabolic and geochemical interactions supported by genomic (dashed lines) and proteomic (solid lines) analyses. Carbon degradation by phylogenetically diverse fermentative bacteria including members of the Bacteroidales (Bact, red), Chloroflexi (Chlo, orange), Melainabacteria (ACD20, brown) and select members of OD1-i (purple) and OP11 (green). Fermentative end-products are linked to respiratory iron reduction by *Geobacter* (Geob, gold), nitrite reduction by *Dechloromonas* and *Comamonadaceae* (Dech, dark blue; Comm, black) sulfate reduction by *Desulfobulbaceae* (Desulf, pale blue) reduction. Sulfide produced by *Desulfobulbaceae* and OD1 has the potential to be oxidized by *Rhodobacter* (Rhod, pale yellow) or abiotically reduce ferric iron. Asterisks (*) indicate partial pathways (nitrite reductase) or the presence of genes indicative of function when no clearly defined reductases are known (for example, MHCs).

OD1) and sulfur disproportionation (for example, ACD75; *Desulfobulbaceae*) activity may produce sulfide. Biogenic sulfide during early acetate amendment could contribute to the abiotic reduction of iron geochemically detected in the aquifer and previously almost exclusively attributed to bioreduction by *Geobacteraceae* (Figure 4). Our findings suggest that other phylogenetically unclassified Deltaproteobacteria (other than *Geobacter* spp.) may contribute to metal reduction (Supplementary Table S3).

Future research is needed to discern the broader obligate or facultative organism metabolic interdependencies that exist for members of the Rifle subsurface community. For instance, in addition to the metabolic interdependencies outlined in Figure 4, a recent study documented that some members of CP discussed here (for example, SR1, ACD80) lack identifiable biosynthetic pathways and may be dependent on members of the community for key metabolites (Kantor *et al.*, 2013). Our research demonstrates how a proteogenomic approach can assign microbial identity and metabolic roles for bacteria that previously lacked characterized physiologies. Ultimately this research can untangle the metabolic interdependencies that shape the structure, function and stability of complex microbial communities.

Acknowledgements

This material is based upon work supported through the Integrated Field Research Challenge Site (IFRC) at Rifle, Colorado, the US Department of Energy (DOE), Office of

Science, Office of Biological and Environmental Research funded the work under contract DE-AC02-05CH11231 (Lawrence Berkeley National Laboratory; operated by the University of California).

References

- Aklujkar M, Young ND, Holmes D, Chavan M, Risso C, Kiss HE *et al.* (2010). The genome of *Geobacter bemidjiensis*, exemplar for the subsurface clade of *Geobacter* species that predominate in Fe(III)-reducing subsurface environments. *BMC Genomics* **11**: 490.
- Allgaier M, Reddy A, Park JI, Ivanova N, D'haeseleer P, Lowry S *et al.* (2010). Targeted discovery of glycoside hydrolases from a switchgrass-adapted compost community. *PLoS One* **5**: e8812.
- Averill BA. (1996). Dissimilatory nitrite and nitric oxide reductases. *Chem Rev* **96**: 2951–2964.
- Ballor NR, Leadbetter JR. (2011). Analysis of extensive [FeFe] hydrogenase gene diversity within the gut microbiota of insects representing five families of Dictyoptera. *Microbial Ecol* **63**: 586–595.
- Beller HR, Chain PS, Letain TE, Chakicherla A, Larimer FW, Richardson PM *et al.* (2006). The genome sequence of the obligately chemolithoautotrophic, facultatively anaerobic bacterium *Thiobacillus denitrificans*. *J Bacteriol* **188**: 1473–1488.
- Berlemont R, Martiny AC. (2013). Phylogenetic distribution of potential cellulases in bacteria. *Appl Environ Microbiol* **79**: 1545–1554.
- Bianchi TS. (2011). The role of terrestrially derived organic carbon in the coastal ocean: a changing

- paradigm and the priming effect. *Proc Natl Acad Sci* **108**: 19473–19481.
- Brazelton WJ, Nelson B, Schrenk MO. (2012). Metagenomic evidence for H₂ oxidation and H₂ production by serpentinite-hosted subsurface microbial communities. *Front Microbiol* **2**: 268.
- Bri e C, Moreira D, L pez-Garc a P. (2007). Archaeal and bacterial community composition of sediment and plankton from a suboxic freshwater pond. *Res Microbiol* **158**: 213–227.
- Byrne-Bailey KG, Weber KA, Chair AH, Bose S, Knox T, Spanbauer TL *et al.* (2010). Completed genome sequence of the anaerobic iron-oxidizing bacterium *Acidovorax ebreus* strain TPSY. *J Bacteriol* **192**: 1475–1476.
- Castelle CJ, Hug LA, Wrighton KC, Thomas BC, Williams KH, Wu D *et al.* (2013). Extraordinary phylogenetic diversity and metabolic versatility in aquifer sediment. *Nat Commun* **4**: 2120.
- Coates JD, Chakraborty R, Lack JG, O'Connor SM, Cole KA, Bender KS *et al.* (2001). Anaerobic benzene oxidation coupled to nitrate reduction in pure culture by two strains of *Dechloromonas*. *Nature* **411**: 1039–1043.
- Dick GJ, Andersson AF, Baker BJ, Simmons SL, Thomas BC, Yelton AP *et al.* (2009). Community-wide analysis of microbial genome sequence signatures. *Genome Biol* **10**: R85.
- Di Rienzi SC, Sharon I, Wrighton KC, Koren O, Hug LA, Thomas BC *et al.* The human gut and groundwater harbor non-photosynthetic bacteria belonging to a new candidate phylum sibling to Cyanobacteria. *eLife* (2013). **2**: e01102.
- Druhan JL, Steefel CI, Molins S, Williams KH, Conrad ME, DePaolo DJ. (2012). Timing the onset of sulfate reduction over multiple subsurface acetate amendments by measurement and modeling of sulfur isotope fractionation. *Environ Sci Technol* **46**: 8895–8902.
- Engelhardt T, Sahlberg M, Cypionka H, Engelen B. (2011). Induction of prophages from deep-subseafloor bacteria. *Environ Microbiol* **3**: 459–465.
- Finster K, Liesack W, Thamdrup B. (1998). Elemental sulfur and thiosulfate disproportionation by *Desulfocapsa sulfoexigens* sp. nov., a new anaerobic bacterium isolated from marine surface sediment. *Appl Environ Microbiol* **64**: 119–125.
- Friedrich CG, Rother D, Bardischewksy F, Quentmeier A, Fischer J. (2001). Oxidation of reduced inorganic sulfur compounds by bacteria: emergence of a common mechanism? *Appl Environ Microbiol* **67**: 2873–2882.
- Goris J, Konstantinidis KT, Klappenbach JA, Coenye T, Vandamme P, Tiedje JM. (2007). DNA-DNA hybridization values and their relationship to whole-genome sequence similarities. *Int J Systemic Evolutionary Microbiol* **57**: 81–91.
- Grant RF, Desai AR, Sulman BN. (2012). Modelling contrasting responses of wetland productivity to changes in water table depth. *Biogeosciences Discussions* **9**: 5579–5623.
- Green SJ, Prakash O, Gihring TM, Akob DM, Jasrotia P, Jardine PM *et al.* (2010). Denitrifying Bacteria isolated from terrestrial subsurface sediments exposed to mixed-waste contamination. *Appl Environ Microbiol* **76**: 3244–3254.
- Handley KM, VerBerkmoes NC, Steefel CI, Williams KH, Sharon I, Miller CS *et al.* (2013). Biostimulation induces syntrophic interactions that impact C, S and N cycling in a sediment microbial community. *ISME J* **7**: 800–816.
- Handley KM, Wrighton KC, Piceno YM, Andersen GL, DeSantis TZ, Williams KH *et al.* (2012). High-density PhyloChip profiling of stimulated aquifer microbial communities reveals a complex response to acetate amendment. *FEMS Microbiol Ecol* **81**: 188–204.
- Herrmann G, Jayamani E, Mai G, Buckel W. (2008). Energy conservation via electron-transferring flavoprotein in anaerobic bacteria. *J Bacteriol* **190**: 784–791.
- Hug LA, Castelle CJ, Wrighton KC, Thomas BC, Sharon I, Frischkorn KR *et al.* (2013). Community genomic analyses constrain the distribution of metabolic traits across the Chloroflexi phylum and indicate roles in sediment carbon cycling. *Microbiome* **1**: 22.
- Kantor RS, Wrighton KC, Handley KM, Sharon I, Hug LA, Castelle CJ *et al.* (2013). Small genomes and sparse metabolisms of sediment-associated bacteria from four candidate phyla. *mBio* **4**: e00708–e00713.
- Knoblauch C, Sahn K, J rgensen BB. (1999). Psychrophilic sulfate-reducing bacteria isolated from permanently cold arctic marine sediments: description of *Desulfofrigus oceanense* gen. nov., sp. nov., *Desulfofrigus fragile* sp. nov., *Desulfotalea gelida* gen. nov., sp. nov., *Desulfotalea psychrophila* gen. nov., sp. nov. and *Desulfotalea arctica* sp. nov. *Int J Syst Bacteriol* **49**: 1631–1643.
- Kuzyakov Y, Friedel JK, Stahr K. (2000). Review of mechanisms and quantification of priming effects. *Soil Biol Biochem* **32**: 1485–1498.
- Miletto M, Williams KH, N'Guessan AL, Lovley DR. (2011). Molecular analysis of the metabolic rates of discrete subsurface populations of sulfate reducers. *Appl Environ Microbiol* **77**: 6502–6509.
- Pereira IA, LeGall J, Xavier AV, Teixeira M. (2000). Characterization of a heme c nitrite reductase from a non-ammonifying microorganism, *Desulfovibrio vulgaris* Hildenborough. *Biochimica et Biophysica Acta* **1481**: 119–130.
- Peura S, Eiler A, Bertilsson S, Tirola M, Jones RI. (2012). Distinct and diverse anaerobic bacterial communities in boreal lakes dominated by candidate division OD1. *ISME J* **1–13**.
- Raes J, Korbelt JO, Lercher MJ, Mering VC, Bork P. (2007). Prediction of effective genome size in metagenomic samples. *Genome Biol* **8**: 1.
- Rabus R, Ruepp A, Frickey T, Rattei T, Fartmann B, Stark M *et al.* (2004). The genome of *Desulfotalea psychrophila*, a sulfate-reducing bacterium from permanently cold Arctic sediments. *Environ Microbiol* **6**: 887–902.
- Riley WJ, Subin ZM, Lawrence DM, Swenson SC, Torn MS, Meng L *et al.* (2011). Barriers to predicting changes in global terrestrial methane fluxes: analyses using CLM4Me, a methane biogeochemistry model integrated in CESM. *Biogeosciences* **8**: 1925–1953.
- Rivett MO, Buss SR, Morgan P, Smith JWN, Bemment CD. (2008). Nitrate attenuation in groundwater: a review of biogeochemical controlling processes. *Water Res* **42**: 4215–4232.
- Roychoudhury AN, Merrett GL. (2006). Redox pathways in a petroleum contaminated shallow sandy aquifer:

- Iron and sulfate reductions. *Sci Total Environ* **366**: 262–274.
- Schmidt O, Drake HL, Horn MA. (2010). Hitherto unknown [Fe-Fe]-hydrogenase gene diversity in anaerobes and anoxic enrichments from a moderately acidic fen. *Appl Environ Microbiol* **76**: 2027–2031.
- Schut GJ, Adams MWW. (2009). The iron-hydrogenase of *Thermotoga maritima* utilizes ferredoxin and NADH synergistically: a new perspective on anaerobic hydrogen production. *J Bacteriol* **191**: 4451–4457.
- Sharma S, Cavallaro G, Rosato A. (2010). A systematic investigation of multiheme c-type cytochromes in prokaryotes. *J Biol Inorganic Chem* **15**: 559–571.
- Sharon I, Morowitz MJ, Thomas BC, Costello EK, Relman DA, Banfield JF. (2013). Time series community genomics analysis reveals rapid shifts in bacterial species, strains, and phage during infant gut colonization. *Genome Res* **23**: 111–120.
- Shrout JD, Scheetz TE, Casavant TL, Parkin GF. (2005). Isolation and characterization of autotrophic, hydrogen-utilizing, perchlorate-reducing bacteria. *Appl Microbiol Biotechnol* **67**: 261–268.
- Sieber JR, McNerney MJ, Gunsalus RP. (2012). Genomic insights into syntrophy: the paradigm for anaerobic metabolic cooperation. *Annu Rev Microbiol* **66**: 429–452.
- Thamdrup B, Finster K, Wurgler Hansen J, Bak F. (1993). Bacterial disproportionation of elemental sulfur coupled to chemical reduction of iron or manganese. *Appl Environ Microbiol* **59**: 101–108.
- Tveit A, Schwacke R, Svenning MM, Urich T. (2013). Organic carbon transformation in high-Arctic peat soils: key functions and microorganisms. *ISME J* **7**: 299–311.
- Vignais PM, Billoud B. (2007). Occurrence, classification, and biological function of hydrogenases: an overview. *Chem Rev* **107**: 4206–4272.
- Weber KA, Achenbach LA, Coates JD. (2006). Microorganisms pumping iron: anaerobic microbial iron oxidation and reduction. *Nat Rev Microbiol* **4**: 752–764.
- Whitman WB, Coleman DC, Wiebe WJ. (1998). Prokaryotes: the unseen majority. *Proc Natl Acad Sci USA* **95**: 6578–6583.
- Wilkins MJ, Callister SJ, Miletto M, Williams KH, Nicora CD, Lovley DR *et al.* (2010). Development of a biomarker for *Geobacter* activity and strain composition; Proteogenomic analysis of the citrate synthase protein during bioremediation of U(VI). *Microbiol Biotechnol* **4**: 55–63.
- Wilkins MJ, Wrighton KC, Nicora CD, Williams KH, McCue LA, Handley KM *et al.* (2013). Fluctuations in species-level protein expression occur during element and nutrient cycling in the subsurface. *PLoS One* **8**: e57819.
- Williams KH, Long PE, Davis JA, Wilkins MJ, N'Guessan AL, Steefel CI *et al.* (2011). Acetate availability and its influence on sustainable bioremediation of uranium-contaminated groundwater. *Geomicrobiol J* **28**: 519–539.
- Wrighton KC, Thomas BC, Sharon I, Miller CS, Castelle CJ, VerBerkmoes NC *et al.* (2012). Fermentation, hydrogen, and sulfur metabolism in multiple uncultivated bacterial phyla. *Science* **337**: 1661–1665.
- Wrighton KC, Thrash JC, Melnyk RA, Bigi JP, Byrne-Bailey KG, Remis JP *et al.* (2011). Evidence for direct electron transfer by a Gram-positive bacterium isolated from a microbial fuel cell. *Appl Environ Microbiol* **77**: 7633–7639.
- Yabusaki SB, Fang Y, Williams KH, Murray CJ, Ward AL, Dayvault RD *et al.* (2011). Variably saturated flow and multicomponent biogeochemical reactive transport modeling of a uranium bioremediation field experiment. *J Contaminant Hydrol* **126**: 271–290.
- Yamada T, Sekiguchi Y, Hanada S, Imachi H, Ohashi A, Harada H *et al.* (2006). *Anaerolinea thermolimosa* sp. nov., *Levilinea saccharolytica* gen. nov., sp. nov. and *Leptolinea tardivitalis* gen. nov., sp. nov., novel filamentous anaerobes, and description of the new classes *Anaerolineae* classis nov. and *Caldilineae* classis nov. in the bacterial phylum Chloroflexi. *Int J Syst Evolutionary Microbiol* **56**: 1331–1340.
- Zehr JP, Jenkins BD, Short SM, Steward GF. (2003). Nitrogenase gene diversity and microbial community structure: a cross-system comparison. *Environ Microbiol* **5**: 539–554.
- Zhuang K, Izallalen M, Mouser P, Richter H, Risso C, Mahadevan R *et al.* (2010). Genome-scale dynamic modeling of the competition between *Rhodospirillum rubrum* and *Geobacter* in anoxic subsurface environments. *ISME J* **5**: 305–316.

Supplementary Information accompanies this paper on The ISME Journal website (<http://www.nature.com/ismej>)