

UC Berkeley

UC Berkeley Electronic Theses and Dissertations

Title

Analyzing Microbial Physiology and Nutrient Transformation in a Model, Acidophilic Microbial Community using Integrated `Omics' Technologies

Permalink

<https://escholarship.org/uc/item/259113st>

Author

Justice, Nicholas Bruce

Publication Date

2013

Supplemental Material

<https://escholarship.org/uc/item/259113st#supplemental>

Peer reviewed|Thesis/dissertation

Analyzing Microbial Physiology and Nutrient Transformation in a Model, Acidophilic Microbial
Community using Integrated 'Omics' Technologies

By

Nicholas Bruce Justice

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Microbiology

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Jillian Banfield, Chair

Professor Mary Firestone

Professor Mary Power

Professor John Coates

Fall 2013

Abstract

Analyzing Microbial Physiology and Nutrient Transformation in a Model, Acidophilic Microbial Community using Integrated 'Omics' Technologies

by

Nicholas Bruce Justice

Doctor of Philosophy in Microbiology

University of California, Berkeley

Professor Jillian F. Banfield, Chair

Understanding how microorganisms contribute to nutrient transformations within their community is critical to prediction of overall ecosystem function, and thus is a major goal of microbial ecology. Communities of relatively tractable complexity provide a unique opportunity to study the distribution of metabolic characteristics amongst microorganisms and how those characteristics subscribe diverse ecological functions to co-occurring, and often closely related, species.

The microbial communities present in the low-pH, metal-rich environment of the acid mine drainage (AMD) system in Richmond Mine at Iron Mountain, CA constitute a model microbial community due to their relatively low diversity and extensive characterization over the preceding fifteen years. Here, chemoautotrophic biofilms form at the air-solution interface of the AMD solution, and carbon is fixed using energy derived from the oxidation of iron and sulfur species released from the dissolution of mineral sulfides.

The chemoautotrophic microbial communities that develop at the air-solution interface sink to the underlying sediment and degrade under microaerobic and anaerobic conditions. A transition from Bacteria- to Archaea-dominated communities coincides with this event. The Archaea identified in sunken biofilms are from the class Thermoplasmata, and in some cases, the highly divergent ARMAN nanoarchaeal lineage. Comparative community proteomic analyses showed a persistence of bacterial proteins in sunken biofilms, and evidence for amino acid modifications due to acid hydrolysis. Given the low representation of bacterial cells in sunken biofilms based on microscopy, hydrolyzed bacterial proteins were inferred to represent a population of lysed cells. These findings indicate dominance of acidophilic Archaea in degrading biofilms, and suggest that they play key roles in anaerobic nutrient cycling at low pH.

Biofilm submersion was recapitulated in microcosm experiments in which floating AMD microbial biofilms were submerged, amended with either $^{15}\text{NH}_4^+$ or deuterium oxide ($^2\text{H}_2\text{O}$), and proteomic stable isotope probing (protein-SIP) used to trace isotope incorporation into newly synthesized proteins of different community members. In ^{15}N -ammonia amended experiments, different $^{14}\text{N}/^{15}\text{N}$ atom% values reflect distinct modes of nitrogen acquisition, since ^{14}N is ultimately derived from extant organic biomass and ^{15}N is derived from inorganic ammonia

provided in the media. There were relatively few ^{15}N -enriched archaeal proteins and all showed low ^{15}N atom% enrichment in anaerobic iron-reducing, aerobic iron-reducing, and aerobic iron-oxidizing environments. These results are consistent with Archaea synthesizing protein using the ^{14}N derived from recycled biomolecules. This conclusion is further supported by results of parallel experiments using $^2\text{H}_2\text{O}$, in which extensive archaeal protein synthesis was detected. In contrast, the bacterial species showed little protein synthesis when incubated in $^2\text{H}_2\text{O}$. The nearly exclusive ability of Archaea to synthesize proteins using $^2\text{H}_2\text{O}$ may be due to archaeal heterotrophy (whereby Archaea offset deleterious effects of ^2H by accessing ^1H generated by respiration of organic compounds) or differences in how archaeal versus bacterial membranes (and their associated mechanisms of energy conservation) respond to $^2\text{H}_2\text{O}$. In biofilms incubated with ^{15}N -ammonium, bacteria synthesized proteins to different extents, with *Sulfobacillus* spp. synthesizing protein almost exclusively under iron-reducing conditions whereas *Leptospirillum* spp. synthesized protein in all conditions, with a clear emphasis on iron-oxidation metabolisms in the presence of Fe^{2+} and oxygen. These findings highlight distinct roles for *Sulfobacillus* vs. *Leptospirillum* in iron cycling. The greatest extent of ^{15}N atom incorporation was detected in proteins of *Leptospirillum*, whereas *Sulfobacillus* proteins had a low extent of ^{15}N incorporation, consistent with an autotrophic metabolism for *Leptospirillum* and heterotrophic metabolism for *Sulfobacillus*.

The role of *Sulfobacillus* organisms in biogeochemical cycling is poorly understood. The diversity of energy conservation and central carbon metabolism within this genus was analyzed using published *Sulfobacillus* genomes as well as five draft genomes of *Sulfobacillus* reconstructed by cultivation-independent sequencing of biofilms sampled from the Richmond Mine (AMDSBA1-5). Three of the newly sequenced species (AMDSBA1, AMDSBA2, and AMDSBA3) have no cultured representatives, and AMDSBA5 and AMDSBA4 represent strains of *S. thermosulfidooxidans* and *S. benefaciens*, respectively. Genomes were replete with pathways of sulfur oxidation, however the presence of enzymes involved with these pathways (and their copy numbers) varied considerably across the genus. Furthermore, several enzymes with putative sulfur and sulfur-compound reduction were identified, perhaps lending previously unknown anaerobic sulfur reduction capacity to *Sulfobacillus* species. Central carbon degradation pathways in *Sulfobacillus* lineages varied, with *S. thermosulfidooxidans* likely favoring the pentose phosphate pathway and lineages of *S. acidophilus*, AMDSBA1, AMDSBA2, AMDSBA3, and AMDSBA4 capable of using the semi-phosphorylative Entner-Doudoroff pathway. Proteins involved in dissimilatory nitrate reduction were limited to AMDSBA3, and amongst AMDSBA genomes, only AMDSBA5 encoded nickel-iron hydrogenase proteins. AMDSBA4 (*S. benefaciens*) is unusual in that its electron transport chain includes a *bc* complex, a unique cytochrome *c* oxidase, and an additional succinate dehydrogenase. It is also the only *Sulfobacillus* species with putative carboxysome proteins. Overall, the results demonstrate diverse ecological strategies for species of *Sulfobacillus* within the Richmond Mine.

Metabolomics methods lag behind other omics technologies due to a wide range of experimental complexities often associated with the environmental matrix. We identified key metabolites associated with acidophilic and metal-tolerant microorganisms using stable isotope labeling coupled with untargeted, high-resolution mass spectrometry. Initially, >3,500 metabolic features were observed in extracts of AMD biofilms, although the molecular identity of these features

remained unclear. Stable isotope labeling improved chemical formula prediction by >50% for larger metabolites (>250 atomic mass units), many of which were unrepresented in metabolic databases and may represent novel compounds. Taurine and hydroxyectoine were identified and likely provide protection from osmotic stress in the biofilms. Community genomic, transcriptomic and proteomic data were integrated to implicate fungi in taurine metabolism. *Leptospirillum* group II bacteria decrease production of ectoine and hydroxyectoine as biofilms mature, suggesting that biofilm structure provides some resistance to high metal and proton concentrations. The combination of taurine, ectoine, and hydroxyectoine may also constitute a sulfur, nitrogen, and carbon currency in the communities.

The genomic, proteomic, and metabolomic characterizations of the Richmond Mine microbial communities not only further our understanding of the physiology of acidophilic organisms but also help elucidate their functional roles within the ecosystem as a whole. Archaea dominate in anaerobic, submerged biofilms, where they synthesize protein using organic nitrogen derived from the degrading biofilm. *Sulfobacillus* are implicated in sulfur transformations, and encode diverse complements of proteins involved in sulfur, nitrate and hydrogen metabolisms, suggesting key niche differentiation within this genus. Metabolites that likely serve as organic nutrient sources for a variety of organisms were identified through use of stable isotope labeling techniques. The development and integration of novel 'omics' based technologies extends our knowledge of the Richmond Mine microbial communities and will ultimately help illuminate microbial contributions to ecosystem function in more complex environments.

Table of contents

Introduction	ii
Chapter 1 Heterotrophic Archaea Contribute to Carbon Cycling in Low-pH, Suboxic Biofilm Communities	
Chapter 2 ¹⁵ N- and ² H proteomic stable isotope probing resolves microbial responses to perturbation and links nitrogen flow to protein degradation by Archaea	28
Chapter 3 Comparison of Environmental and Isolate <i>Sulfobacillus</i> Genomes Reveals Diverse Carbon, Sulfur, Nitrogen, and Hydrogen Metabolisms	52
Chapter 4 Metabolites associated with adaptation of microorganisms to an acidophilic, metal-rich environment identified by stable isotope enabled metabolomics	84
References	102

Introduction

“Their life processes are played out in a very simple fashion; all their activities are driven by a purely inorganic chemical process” -- Sergei Winogradsky, describing chemolithotrophy, 1887

Around the same time that Winogradsky wrote these words (Winogradsky 1887), mining at the Richmond Mine in Iron Mountain, CA, began. Opening the mountain for its deposits of gold, silver, iron, copper, and zinc, man laid these minerals bare to oxygen and water, setting the stage for the proliferation of microorganisms. For these organisms, chemolithotrophy is the energetic engine that drives cellular metabolism, and more than one hundred and twenty-five years after Winogradsky, we are beginning to understand that the life processes of these microorganisms are played out in a fashion that is anything but “simple.”

The Richmond Mine, one of ten mines located at Iron Mountain, is an extraordinary place once described as a “toxic hellhole” (Fimrite 2010). Mining began in the 1860s and at one time, Iron Mountain was the largest copper producer in California, and the tenth largest copper producer in the world (United States Environmental Protection Agency 2006). Mining was discontinued in 1963, and twenty years later the Environmental Protection Agency designated Iron Mountain as a Superfund site due to the toxic effluent released by the mine. That effluent—acid mine drainage (AMD)—is formed from the dissolution of mineral sulfides, and is primarily comprised of sulfuric acid with high concentrations of iron, zinc, copper, and arsenic. In fact, at one point before cleanup efforts began, Iron Mountain was responsible for an estimated one-quarter of the national municipal and industrial discharge of copper and zinc into surface waters (United States Environmental Protection Agency 2006).

Today, within the now inoperative mines, microbial communities thrive. They form thick biofilms that float at the air-solution interface of the AMD flow and they exploit the energy-dense ore for their metabolic gain. Their activities vastly accelerate the dissolution of mineral sulfides and as such, they are a primary contributing factor in the formation of AMD. Although Richmond Mine may be a unique place, the organisms within are not. They are globally distributed in AMD sites, hydrothermal vents, solfatara fields, and industrial bioleaching reactors. As such, a more complete understanding of the physiological processes and interactions is of vast ecological and even biotechnological importance.

For almost two decades, dozens of researchers in the laboratory of Dr. Jill Banfield have investigated the microbiological processes and their inextricable link to the site’s geochemistry. Beginning in the mid-1990’s, Theresa Edwards made the first culture-independent analyses of the Richmond Mine microbial communities using 16S rRNA gene clone libraries. Shortly thereafter, Katrina Edwards used culture-based techniques to investigate the microbial colonization and dissolution of pyrite, ultimately isolating the extremely acidophilic Archeon *Ferroplasma acidarmanus* (Edwards, Bond, Gihring, et al. 2000b). Around the same time, Edwards, Shrenk, and Bond began using fluorescent in situ hybridization to demonstrate how environmental conditions like pH, temperature, and rainfall were intimately tied to the distribution of different organisms within the mine (Edwards et al. 1999; Schrenk et al. 1998; Edwards, Bond, Druschel, et al. 2000a; Bond, Druschel, et al. 2000a; Bond, Smriga, et al. 2000b). Greg Druschel later examined the chemistry of sulfur compound transformations, providing a geochemical framework within which to interpret the distribution of various microorganisms (Druschel et al. 2004; 2003). This early work established that the community

was of relatively tractable complexity, lending itself as a prime target for one of the first ‘metagenomic’ studies, carried out by Tyson et al. (Tyson et al. 2004). Work by Dick, Andersson, Simmons, Sun, and Deneff extended these metagenomic approaches, discovering genome-wide nucleotide signatures (Dick et al. 2009), patterns of phage and phage resistance (Tyson & Banfield 2008; Sun et al. 2013; Andersson & Banfield 2008), selection processes in natural populations (Simmons et al. 2008), and ultimately measuring evolutionary rates from environmentally-isolated genomic information (Deneff & Banfield 2012). Additional efforts by Yelton and Goltsman have further teased apart the metabolic potential of several important AMD organisms (Yelton et al. 2013; Aliaga Goltsman et al. 2013; Goltsman et al. 2009).

This bedrock of genomic information, in turn, created opportunities to use ‘metaproteomic’ approaches that examined functional activity of microorganisms in their environment (Ram et al. 2005). Proteomic data has been foundational in unraveling genomic recombination (Lo et al. 2007), ecological strategies (Mueller et al. 2010), stress responses (Belnap et al. 2011), and strain variation (Deneff et al. 2007) of microorganisms in the Richmond Mine. Additionally, early metabolomic work linked patterns of protein expression to specific metabolites within the community (Wilmes et al. 2010), and showed how stable isotopes can be used to identify novel lipids (Fischer et al. 2011). Finally, no mention of Richmond Mine researchers would be complete without the mention of Brett Baker, who among his numerous contributions identified a novel lineage of Archaea (Baker et al. 2006), and together with Luis Comolli showed that these organisms were amongst the smallest ever discovered (Comolli et al. 2009).

The nearly 20 years of effort by these individuals—and the many others not listed here—is humbling. Now, as the chapter on AMD research comes to a close in the Banfield group, this present work strives to unravel just a few of the many remaining mysteries of the Richmond Mine microbial communities.

The present work is primarily concerned with understanding the “life processes” of the organisms of the Richmond Mine and particularly how these underpin nutrient cycling within the community as a whole. A key component in these efforts is the integration of data from different ‘omics’ technologies, with a particular emphasis on stable-isotope mass spectrometry based approaches. The initial impetus for much of what follows began with the characterization of ‘sunken’ biofilms (Chapter 1). These biofilms are the product of the natural submersion of the Bacterial-dominated chemoautotrophic biofilms that float at the air-solution interface. Upon submersion, a transition to an Archaeal-dominated community was observed. This suggested that the archaeal community members were likely key players in anaerobic degradation of organic matter within the mine. Further experiments using stable-isotope proteomics showed that Archaea were actively synthesizing protein in submerged biofilm environments, and likely derive nitrogen from organic sources in the course of heterotrophic catabolism (Chapter 2). Although not dominant, bacteria of the genus *Sulfobacillus* were also observed in submerged biofilms. Metagenomic efforts allowed for the reconstruction of five near-complete *Sulfobacillus* genomes, and the metabolic potential of these organisms was analyzed with particular attention to iron and sulfur cycling capabilities (Chapter 3). Finally, stable-isotope metabolomic approaches were used to identify abundant metabolites associated with adaptation of microorganisms to the harsh acidic conditions of the Richmond Mine, and these results are integrated with proteomic and transcriptomic data to implicate specific species in the biosynthesis of certain metabolites (Chapter 4).

Understanding the metabolism of any given microorganism is far from trivial, and the task only becomes more challenging when trying to frame that metabolism within the context of a natural ecosystem. As a whole, the work performed at the Richmond Mine is a significant milestone towards developing a more comprehensive understanding of natural microbial populations and their attendant complexity. The results presented here are small part of that legacy. They serve to extend our knowledge of these acidophilic communities and further the development of ‘omics’ techniques that are being increasingly leveraged against more complex microbial communities. If there is anything that the Richmond Mine has taught us, it’s that the life processes of any microorganism—indeed any microbial community—are far from “simple,” and many mysteries yet lay waiting to be unearthed.

Chapter 1

Heterotrophic Archaea Contribute to Carbon Cycling in Low-pH, Suboxic Biofilm Communities

Abstract

Archaea are widely distributed, yet are most often not the most abundant members of microbial communities. Here, we document a transition from bacterial- to archaeal-dominated communities in microbial biofilms sampled from the Richmond Mine acid mine drainage (AMD) system (~ pH 1.0, ~ 38 °C) and in laboratory-cultivated biofilms. This transition occurs when chemoautotrophic microbial communities that develop at the air-solution interface sink to the sediment-solution interface and degrade under microaerobic and anaerobic conditions. The Archaea identified in these sunken biofilms are from the class *Thermoplasmata*, and in some cases, the highly divergent ARMAN nanoarchaeal lineage. In several of the sunken biofilms, nanoarchaea comprise 10-25% of the community, based on fluorescent *in situ* hybridization and metagenomic analyses. Comparative community proteomic analyses show a persistence of bacterial proteins in sunken biofilms, but there is clear evidence for amino acid modifications due to acid hydrolysis. Given the low representation of bacterial cells in sunken biofilms based on microscopy, we infer that hydrolysis reflects proteins derived from lysed cells. For Archaea, we detected ~ 2,400 distinct proteins, including a subset involved in proteolysis and peptide uptake. Laboratory cultivation experiments using complex carbon substrates demonstrated anaerobic enrichment of *Ferroplasma* and *Aplasma* coupled to reduction of ferric iron. These findings indicate dominance of acidophilic Archaea in degrading biofilms, and suggest that they play roles in anaerobic nutrient cycling at low pH.

Introduction

Our knowledge of archaeal contributions to the anaerobic carbon cycle continues to expand. Beyond their best-characterized role as methanogens, Archaea are increasingly implicated in nutrient cycling, for example via anaerobic methane oxidation (Hinrichs et al. 1999) and the dicarboxylate-hydroxybutyrate cycle (Huber et al. 2008). To date, Archaea have been shown to dominate in only a few environments, such as lake (Jiang et al. 2008) and ocean sediments (Orcutt et al. 2011), and in some ‘extreme’ environments, such as those characterized by high salinity (Oren 2002), or in acidic hot spring (Inskeep et al. 2010). Understanding the factors that enrich for Archaea and the physiological processes that sustain them *in situ* remain important questions.

In metal-rich, low-pH ecosystems associated with pyrite ores, several species of Archaea have been identified via clone library analysis (Bond, Druschel, et al. 2000a; Baker & Banfield 2003; Bruneel et al. 2008; Tan et al. 2008; Sanchez-Andrea et al. 2011), metagenomic sequencing (Tyson et al. 2004; Baker et al. 2006), and isolation (Edwards, Bond, Gihring, et al. 2000b; Golyshina et al. 2000). Archaea have also been identified as members of consortia employed in bioleaching systems for metal recovery (Rawlings 2002). Arguably, the best-studied low-pH, high-iron environment is the Richmond Mine AMD system at Iron Mountain, CA. In the Richmond Mine and in other acidic metal-rich systems, Archaea are typically found in relatively low abundance (Baker & Banfield 2003; Macalady et al. 2007; Kock & Schippers 2008). The Archaea in the system belong to two distinct lineages of the *Euryarcheota*: the class *Thermoplasmata*—which includes *Ferroplasma* species as well as the ‘alphabet plasmas’ referred to as A- through Iplasma (Baker & Banfield 2003)—and a deeply branching clade referred to as ARMAN (Baker et al. 2006; Comolli et al. 2009). Within the *Thermoplasmata*

belongs the order *Thermoplasmatales*, which includes the genera *Thermoplasma*, *Ferroplasma*, *Picrophilus*, and *Thermogymnomonas*. Iplasma is the most divergent of the AMD plasmas, and is almost certainly not in the order *Thermoplasmatales*, and may in fact represent a separate class (Yelton et al. 2013). Metagenomic sequencing has allowed for near-complete reconstruction of genomes for *Ferroplasma*, four *Thermoplasmata* and four ARMAN Archaea (Tyson et al. 2004; Baker et al. 2010; Yelton et al. 2011). Amongst the AMD plasmas, genomic analyses consistently indicate facultatively anaerobic and heterotrophic lifestyles (Tyson et al. 2004; Yelton et al. 2013). Furthermore, isolates from the *Thermoplasmatales* lineage, including *Thermoplasma volcanium*, *Thermoplasma acidophilum*, and *Ferroplasma* spp. have been characterized as facultatively anaerobic heterotrophs (Seeger et al. 1988; Dopson et al. 2006).

Generally, biofilms beginning to grow at the air-solution interface in the Richmond Mine are devoid of Archaea, but archaeal populations become more abundant with increasing biofilm age and thickness (Wilmes et al. 2009). Here we present evidence that, in contrast to floating biofilms, Archaea dominate the submerged suboxic biofilms and as such, may be key players in carbon and nutrient cycling in AMD environments.

Materials and Methods

Site description and sample collection

Biofilm samples were collected underground within the Richmond Mine at Iron Mountain, CA. Temperature was measured *in situ* and the pH determined shortly after sample collection. Samples for metagenomic analysis were frozen on site on dry ice whereas those for cultivation experiments were maintained on ice water during transport to the laboratory where they were used immediately for inoculation experiments. At the time of sampling, biofilm thickness was estimated visually to indicate growth (developmental) stage. Based on prior direct measurements, early growth stage biofilms are typically < 30 μm thick whereas late growth stage biofilms can be up to 200 μm thick (Wilmes et al. 2009).

Microscopy

Environmental samples were fixed and stained with lineage-specific fluorescent *in situ* hybridization (FISH) probes to determine archaeal and bacterial relative abundances, as described previously (Bond & Banfield 2001). Fixed cells were counted manually on a Leica DMRX epifluorescence microscope at 630 times magnification. At least five fields of view and >1000 cells were counted for each sample. FISH cell counts were used to calculate species percentages from total cell counts obtained using 4',6-diamidino-2-phenylindole. Probes used in this study are listed in Table 1.1.

16S rRNA sequencing and analysis

For all DNA extractions, ~1 g of frozen biofilm was washed twice with 0.5 mL cold PBS (pH 7.2) before being resuspended in 500 μl TE buffer (10 mM Tris-HCl, 1 mM EDTA [pH 8.0]) and 500 μl STEP buffer (0.5% SDS, 50 mM Tris-HCl [pH 7.5] 400 mM EDTA, 1 mg/mL Proteinase K, 0.5% Sarkosyl). The samples were cycled three times between liquid nitrogen and 50 $^{\circ}\text{C}$ water-bath to facilitate lysis. After the last cycle, an additional 500 μl of STEP buffer were added and the sample was left at 50 $^{\circ}\text{C}$ for 20 minutes. Samples were extracted 3 times with 1 volume each 25:24:1 phenol:chloroform:isoamyl alcohol mix (Fisher BioReagents, Pittsburgh, Pennsylvania) using 15 mL Phase Lock Gel tube (5 Prime, Gaithersburg, Maryland). Aqueous

layer was precipitated with 1 volume cold isopropanol and 0.1 volume cold sodium acetate (3 M pH 5.2). Pellets were washed with 1 mL cold ethanol and resuspended in TE buffer.

For clone library construction, 16S rRNA gene sequences were amplified by PCR in mixtures containing 1X PCR Master Mix (Promega, Madison Wisconsin), ~ 40 ng template DNA, 300 mM (each) forward and reverse primers, and 0.5 mg/mL BSA (New England Biolabs, Ipswich, Massachusetts). Primers for Bacteria-specific libraries were 27F (5' AGAGTTTGATCMTGGCTCAG) and 1492R (5' TACGGYTACCTTGTTACGACTT) (DeLong et al. 2006), and for Archaea-specific libraries were 522F (5' GGYAAGACSGGTGSCAGC) and 1354R (5' GCGRTTACTASGGAWTCC), designed to bind to both Archaea of the *Thermoplasmatales* and the ARMAN. Reactions were incubated in a DNA Engine Dyad thermocycler (MJ Research, Quebec, Canada) for 94 °C (2 min) followed by 30 cycles of 94 °C for 45 sec, 53 °C for 45 sec, 72 °C for 90 sec, and a final extension of 72°C for 10 min. PCR products were purified and cloned, as described previously (Bond, Druschel, et al. 2000a). Sequencing was carried out using the same amplification primers as used in PCR reactions on a 3730XL DNA Analyzer (Applied Biosystems, Carlsbad CA). Sequences were assembled with Phred/Phrap software and chimeric sequences were screened using Mallard (Ashelford et al. 2006). Sequences were clustered using UCLUST (Edgar 2010) at 97% percent identity. Sequences were aligned with MAFFT (Kato et al. 2002) and curated with MEGA (Tamura et al. 2011). Trees were constructed using the Maximum Likelihood algorithm in MEGA using the General Time Reversible model. Bootstrap values were calculated from 1000 iterations.

Metagenomic sequencing and analysis

For metagenomic sequencing, DNA was extracted from samples, as described above. Samples were treated with 10U/mL RNase (New England Biolabs, Ipswich, MA) for 15 minutes at room temperature to improve sequencing quality. DNA from both samples was sequenced with 454 FLX Titanium (454 Life Sciences, Branford, CT), as per manufacturer's protocols with the exception that fragment libraries were quantified with digital PCR rather than with titrations (White et al. 2009). Data was quality controlled by removing reads with at least one ambiguous base. Artificial duplicate reads formed as a result of the 454 FLX Titanium sequencing method were removed, as described previously (Gomez-Alvarez et al. 2009; Deneff & Banfield 2012). For analysis of community composition, reads over 100 nucleotides in length were compared by BLAST to genomes of organisms from the Richmond Mine recovered by prior metagenomic studies (Tyson et al. 2004; Baker et al. 2006). Only matches with >98% identity were reported and the number of hits for each species was normalized by genome size.

Protein analysis

Whole-cell fractions of proteins were extracted from frozen biofilms (~500 mg) via sonication and TCA-precipitation, as previously described (Deneff et al. 2009). Briefly, the cells were washed with H₂SO₄ (pH 1.1) to remove residual iron, then pellets resuspended in 6 mL 20 mM Tris-SO₄, pH 8.0 and sonicated on ice using a microtip sonicator. Five mL 0.4 M Na₂CO₃ pH 11 was added and non-lysed cells and polymers removed by centrifugation at 6,000 g for 20 minutes. Protein was precipitated with 1:10 TCA for at least 3 hours, centrifuged (14,000 g 10 minutes) and washed with cold methanol. The protein was resuspended in 6 M guanidine/10 mM DTT, and enzymatically digested with trypsin (Promega, Madison, WI). Peptides were separated via 24 h nano-2D-LC (strong cation exchange/reversed phase) and analyzed by tandem mass

spectrometry on a hybrid LTQ-Orbitrap mass spectrometer (Thermo Fisher Scientific, San Jose, CA), as described previously (Deneff et al. 2009; Mueller et al. 2010). At least two technical replicates were run for each sample. Unmodified peptides were identified by using the SEQUEST algorithm (Eng et al. 1994) to search tandem mass spectra against the predicted protein database `amdvl_reads_11212008_arman_Biofilm_AMD_CoreDB_04232008_B` constructed from previously published community genomic datasets (Tyson et al. 2004; Baker et al. 2006). The SEQUEST search results were filtered and sorted with the DTASelect algorithm (Tabb et al. 2002). False positive rates for this method of analysis have been determined to be between 1 and 5% (Ram et al. 2005; Deneff et al. 2009). The database is available: http://compbio.ornl.gov/amd_gtl_ms_results/databases/

Normalized spectral abundance factors (NSAF) were calculated as described previously (Florens et al. 2006; Mueller et al. 2010). For clustering analysis of Gplasma proteins, organism-specific NSAF values (`orgNSAF`) were calculated by normalizing NSAF values to the total detected spectral counts from that organism, after which they were arcsin transformed, mean centered, and scaled as described previously (Mueller et al. 2010). Only samples with >150 Gplasma proteins were used in order to prevent bias from zeros during normalization. Hierarchical clustering was carried out in Multiple Experiment Viewer (Saeed et al. 2003; 2006) using a Pearson Correlation coefficient and average linkage. Fisher's exact test was performed in R to compare enrichment of each COG functional category in the group of proteins enriched in sunken samples versus other proteins in the dataset.

Deamidation analysis

Tandem mass spectra were re-searched for deamidated peptides with the SEQUEST algorithm by considering the dynamic modification of deamidation for all glutamine (Q) and asparagine (N) residues. DTASelect output files were analyzed with custom Ruby scripts to calculate the spectral counts of deamidated Q and deamidated N residues as well as the total spectral counts of Q and N residues in identified peptides. These spectral counts were summed for each organism or Domain. The deamidation frequencies of Archaea and Bacteria were calculated as the percentages of the run-averaged spectral counts of deamidated Q or N over the total spectral counts of Q or N, respectively, in each domain. In order to ensure that deamidation percentages were calculated from a representative number of peptides, runs in which a single peptide contributed more than 50% of the total deamidation percentage for either archaeal or bacterial domains were not included in the analysis. By searching the reversed database (i.e., all peptide sequences reversed), false-discovery rates of deamidated peptides (~2%) and deamidated spectra (~3%) were identified. For analysis of differences of deamidation between growth stages and Domains, data were analyzed with a two-way ANOVA in R after verification that assumptions of normality and homoscedasticity were met.

Cultivation of aerobic biofilms

The floating biofilm was cultivated in laboratory bioreactors as previously described (Belnap et al. 2009; 2011). After growing to a thickness consistent with late-growth stage biofilms, the biofilm was dislodged from the reactor walls and submerged at the bottom of the reactor chamber under ~2 cm of slowly flowing AMD solution. After seven days, sunken biofilm samples were taken and either immediately fixed for FISH or frozen for later proteomic analyses.

Cultivation of anaerobic biofilms

To maximize diversity for anaerobic community culturing, inocula were derived from a variety of floating and sunken biofilm samples collected from several locations within the Richmond Mine (Redding, CA, USA). About five mL aliquots of the mixed inocula were added to replicate 120 mL serum vials filled with 50 mL of nitrogen-degassed ferric-9kBR medium containing 100 mM $\text{Fe}_2(\text{SO}_4)_3$, 1 mM $(\text{NH}_4)_2\text{SO}_4$, 0.5 mM KCl, 0.050 mM K_2HPO_4 , 5 mM $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$, 0.2 mM CaSO_4 , and trace minerals (Belnap et al. 2009), all adjusted to pH 1.1 with H_2SO_4 .

Triplicate cultures were supplemented with one of each of the following organic substrates: glycolate, formate, lactate, acetate, betaine (all 10 mM, Sigma Aldrich), casamino acids (0.02%, acid-hydrolyzed casein, BD, NJ, USA), or peptone (0.02%, enzymatic digest of protein, BD, NJ, USA). Additionally, some cultures were supplemented with naturally derived biofilm substrate obtained by lyophilization of a mature floating biofilm from the Richmond Mine, followed by re-dissolution into sterile deionized H_2O and sonication for 5 minutes with a microtip sonicator. This natural substrate was added to cultures at a concentration of 0.05% w/v. Sterility of this substrate was monitored via DNA extractions and microscopy in un-inoculated controls. Replicates of all cultures were cultivated with and without 0.01% yeast extract. A 10 % volume of each culture was transferred three times into 5 mL volumes fresh media every 3 to 4 weeks, and transferred into 40 mL in the fourth and final transfer. Ferrozine assays of dissolved iron concentration and DNA was extracted from cultures on the fourth transfer, 28 days after inoculation. Ferrozine assays were done as previously described (Stookey 1970), and DNA extractions were carried out as described above.

Results

Sunken biofilm sampling

Three sunken biofilms were collected from the 5-way (February, 2008, pH 0.98, 38 °C) and three from the UBA site (June 2009, pH 1.1, 38 °C) (see location map in Figure 1.1). At the 5-way site, we sampled a mature, thick (~0.1 mm) flexible biofilm floating on a ~ 5 cm deep flowing AMD solution (Floating Growth Stage 2, Figure 1.2A). In addition, we collected three fragile sunken biofilms of similar thickness that were stratified underneath the floating biofilm and suspended close to the stream base. We labeled these biofilms as 5way-Sunken1, reflecting its location immediately below the floating GS2 biofilm; 5way-Sunken2 the biofilm a few millimeters below 5way-Sunken1; and 5way-Sunken3, the biofilm resting on the stream channel (Figure 1.2A). Fragile (easily disintegrated) sunken biofilm samples collected from the UBA site (UBA-Sunken1, UBA-Sunken2, and UBA-Sunken3) were recovered from three different regions of a slowly draining AMD pool about 25 cm deep. All samples collected for this study, and those obtained in previous studies, are listed in Table 1.2.

Fluorescent In Situ Hybridization Microscopy of Natural Sunken Biofilms

Fluorescent In Situ Hybridization (FISH) was used to determine the relative proportion of Bacteria to Archaea in sunken biofilms (Figure 1.3, Table 1.3). In contrast to previously characterized early growth stage (GS1) and late growth stage (GS2) floating biofilms, sunken biofilms were dominated by Archaea (Figure 1.4). The relative community composition of the 5way-Sunken3 biofilm was qualitatively similar to the other 5way samples, but with clearly reduced cell density. High autofluorescence in non-DAPI filter channels, presumably caused by degraded biofilm and minerals, prohibited quantitative analysis of this sample. In the other five sunken samples analyzed quantitatively, non-ARMAN archaeal species dominated the

community. ARMAN and *Sulfobacillus* species were more abundant in 5way Sunken samples than UBA Sunken samples (Table 1.3).

Clone libraries and DNA sequencing

Clone libraries were constructed for 5way-Sunken1 to better determine the phylogenetic diversity of the Archaea and Bacteria present. Forty-five of the 89 sequences obtained from 5way-Sunken1 using Archaea-specific 16S rRNA primers belonged to the deeply branching clade known as ARMAN, a euryarchaeal lineage (Comolli et al. 2009; Baker et al. 2010) (Figure 1.5). Eleven clones were closely related (>99% identity) to *Ferroplasma acidarmanus*, a species previously isolated from the Richmond Mine (Edwards, Bond, Gihring, et al. 2000b). The remaining clones represented Aplasma, Bplasma, Gplasma, and Eplasma (Figure 1.5), a radiation of closely related *Thermoplasma* Archaea, referred to as the “alphabet plasmas” (Baker et al. 2003; Dick et al. 2009).

The Bacteria-specific 16S rRNA library contained 89 clones, 36 belonged to *Leptospirillum ferrodiazotrophum*, a member of *Leptospirillum* group III, 8 clones were related to, but distinct from *Leptospirillum* group III (3.0% divergent), 15 belonged to *Leptospirillum* group II (*Leptospirillum ferriphilum*), and the remainder to species of the genus *Sulfobacillus* (Figure 1.6). Within the *Sulfobacillus* lineage, 10 sequences were *Sulfobacillus thermosulfidooxidans* (100% identity), 14 were *Sulfobacillus benefaciens* (100% identity), and 4 derived from other *Sulfobacillus* species.

Proteomic comparisons of floating and sunken biofilms

Proteomic analyses of the six sunken biofilms and the 5way-Floating biofilm were used to analyze community structure in the context of previously sampled floating biofilms as well as to identify metabolic processes active in the sunken biofilms. The relative abundance of bacterial proteins was highest in early (93.0%) and late growth stage floating biofilms (84.9%), and was lowest in sunken biofilm samples (55.0%) (Figure 1.7). Archaeal proteins were most abundant in the sunken biofilm samples (24.5%), and less abundant in the early- (1.9%) and late-growth stage floating biofilms (7.1%).

The proteomes of the 5way-Floating biofilm and the three stratified sunken biofilms showed a striking increase in the relative abundance of archaeal proteins with increasing depth, 5way-Sunken1 contained relatively more archaeal protein (7.3%) than 5way-Floating (1.9%). Below 5way-Sunken1, a large increase in the relative abundance of the Archaea was found in 5way-Sunken2 (35.8% Archaea), with a similar relative abundance in 5way-Sunken3 (35.9% Archaea) (Figure 1.8).

Overall, 2,334 distinct archaeal proteins were identified across all samples analyzed (Table S1). Of those, 545 proteins were only identified in sunken samples, 761 in floating samples, and 1,128 proteins were identified in both. Proteins detected represented many major metabolic pathways, such as TCA cycle, glycolysis, fatty acid oxidation, and electron transport. Furthermore, many proteins associated with the acquisition and breakdown of organic carbon were detected, including peptide transporters, sugar transporters, peptidases, and extracellular glucoamylases (Table S2).

Gplasma was the only archaeal species with consistently high protein abundances across samples and growth stages, enabling proteome comparisons across growth stages by hierarchical clustering of orgNSAF values. The sunken samples formed a distinct cluster, as did a group of proteins showing relatively higher abundance in the sunken biofilms (Figure 1.9A). This group

of proteins had a distinct distribution of COG functional category counts (Fisher's Exact Test, $p = 0.0002$), with a lower abundance of proteins involved in protein biogenesis (ribosomal) and higher abundances of proteins involved in amino acid metabolism (Figure 1.9B). Elevated in abundance in the sunken biofilm protein cluster were two enzymes of the TCA cycle (malate dehydrogenase, aconitate hydratase), several enzymes associated with transformations of pyruvate (three subunits of the pyruvate dehydrogenase/2 oxoacid dehydrogenase complex, a cytochrome-associated pyruvate dehydrogenase, pyruvate:ferredoxin oxidoreductase, pyruvate phosphate dikinase, and acetolactate synthase), and multiple proteins that likely participate in electron transfer reactions, including an oxidoreductase, and aldehyde dehydrogenase, and a flavoprotein (Table S3).

454 sequencing

The relative percentage of bacterial proteins in sunken samples was higher than what was expected based on FISH analyses, so we analyzed metagenomic sequence from a subset of samples (5way-Sunken2 and UBA-Sunken2) as a third measure of community composition. These analyses used 50 Mb of 454 FLX Titanium sequence data obtained for each of the UBA-Sunken2 and 5way-Sunken2 biofilms (read lengths averaged ~ 290 bp for both samples).

BLAST comparisons of 454 sequencing reads to the genomic databases showed that the 5way-Sunken2 and UBA-Sunken2 biofilms were dominated by archaeal species (Figure 1.10). In the 5way-Sunken2 biofilm, 98.3% of reads mapping to the database matched archaeal sequences, whereas bacterial sequences represented only 1.3% of reads. In the UBA-Sunken2 sample, archaeal species comprised 69.24% of genomic reads whereas bacterial species made up 30.68%.

Deamidation:

Compared to FISH and metagenomic analyses, proteomic measurements showed a striking overrepresentation of Bacteria in sunken communities. Although this discrepancy could be partially explained by differences in cell size (and thus protein content) between the bacterial and archaeal community members, we hypothesized that bacterial proteins may derive from cell lysates rather than living cells. It is known that glutamyl and asparagyl residues can undergo an acid-dependent deamidation (Joshi & Kirsch 2002; Robinson 2002; Catak et al. 2006). Thus, if our hypothesis is true, bacterial proteins in sunken communities should have more extensive deamidation than in floating communities. In contrast, archaeal proteins should have the same degree of deamidation in floating and sunken biofilms if the cells were equally viable in both communities. We measured the degree of deamidation in proteins extracted from both floating and sunken biofilms by proteomics. On average, 28.3% of glutamines underwent deamidation in bacterial proteins from early growth-stage biofilms, 52.5% in late growth-stage biofilms, and 72.1% in sunken biofilms (Figure 1.11A). On the other hand, glutamines in archaeal proteins were deamidated at frequencies of 17.0% and 19.0%, in early and late-growth stage biofilms, respectively, and at 44.2% in sunken biofilms. Similarly, the frequencies of deamidation in bacterial asparagine residues were 22.2%, 36.4%, and 58.1% in early growth-stage, late growth-stage, and sunken biofilms, respectively, whereas archaeal asparagine deamidation remained relatively constant between 21.1% and 30.3% across all sample subsets (Figure 1.10B). Deamidation was not limited to any particular subset of proteins, and was evident for essential cytoplasmic proteins like ribosomal proteins, TCA cycle enzymes, and glycolysis enzymes (data not shown). A two-way ANOVA was used to test for differences among deamidation frequencies

between domains of life (Archaea and Bacteria) and growth stages (GS1, GS2, Sunken). Deamidation frequencies of asparagine residues differed significantly for domains [$F(1,43) = 14.11$, $p < 0.001$], as well as between growth stages, [$F(2,43) = 19.53$, $p < 0.001$]. Analysis of deamidation frequencies of glutamine residues showed similar results, with significant difference amongst domains [$F(1,44) = 52.88$, $p < 0.001$], and between growth stages [$F(2,44) = 36.97$, $p < 0.001$]. Importantly, there was an interaction between growth stage and domain for both asparagine [$F(2,43) = 6.74$, $p < 0.01$], and glutamine [$F(2,44) = 4.94$, $p < 0.05$] residues, indicating that the two domains respond differently to increasing growth stage and submersion. Post-hoc analysis using Tukey's honestly significant difference test was used to compare deamidation frequencies amongst growth stages and between domains for each residue (Table 1.4).

Bioreactor laboratory experiments

We submerged a laboratory-cultured biofilm in our laboratory bioreactors to test the hypothesis that a bacterial dominated floating biofilm would shift towards archaeal dominance if submerged. FISH results showed an increase in the relative abundance of Archaea after seven days of submersion (increasing from 23.5% of the community prior to sinking to 52.0% after submersion). Bacteria populations, on the other hand, showed a marked decrease from 76.5% of the community in the floating biofilm, to 47.9% of the sunken biofilm. Proteomic data also showed an increase in the relative abundance of Archaea, with archaeal proteins comprising 11.3% of the floating biofilm and 17.2% of the sunken biofilm, while bacterial proteins compromised 77.2% and 67.0% of the floating and sunken biofilms, respectively. Measures of protein deamidation, showed high frequencies of deamidation for bacterial asparagine (20.15% in floating; 34.8% in sunken) and glutamine residues (30.6% in floating; 52.6% in sunken), but not for archaeal residues (between 10.2-12.6% for glutamine and asparagine in either floating or sunken samples) (Figure 1.12).

Culturing:

In order to determine if Archaea are capable of carrying out anaerobic carbon oxidation in sunken biofilms, we enriched for anaerobic iron reducers using ferric sulfate media and a variety of different carbon sources. Iron was reduced to similar degrees in yeast-extract supplemented cultures of peptone, betaine, casamino acids, and AMD biofilm extract (Figure 1.13). Cultures without yeast extract were not viable even after the first transfer, except for native-biofilm cultures, which did not grow after the third transfer. No growth was apparent in cultures containing glycolate, formate, acetate, or lactate with or without yeast extract. Analysis of 16S rRNA sequences from cultures cultivated on peptone, betaine, casamino acids, and AMD biofilm were dominated by *Ferroplasma acidarmanus*, with *Aplasma* comprising 0-47% of the sequenced clones (Figure 1.13). We also detected no hydrogen consumption or production, nor methanogenesis (data not shown).

Discussion

Little is known about selection factors for Archaea or archaeal metabolism in AMD systems. Here we used a combination of metagenomic, proteomic and FISH analyses to compare community structures of thin floating biofilms, thicker, higher developmental stage floating biofilms, and biofilms submerged in suboxic/anoxic environments (Table 1.2). In contrast to floating biofilms that are Bacteria-dominated, we show here that field-collected communities in

submerged biofilms are dominated by Archaea and that a transition from bacterial to archaeal dominance can be induced by biofilm submersion in the laboratory. Based on the identification of proteins involved in peptide transport, sugar transport, fatty-acid oxidation, and extracellular protein and starch breakdown (Table S1), it is likely that the Richmond Mine plasmas may scavenge nutrients in the form of proteins and carbohydrates, in this case derived from the decaying biofilm.

For *Aplasma*, *Eplasma*, *Gplasma*, and *Ferroplasma*, many components of the electron transport chain were identified by proteomics, including subunits of NADH dehydrogenase, Rieske Fe-S proteins, electron transfer flavoprotein-quinone oxidoreductases, as well as other dehydrogenases and oxidoreductases that may play roles in electron transport (Table S1). Based on this, and the detection of almost all TCA-cycle proteins, it seems most likely that respiration is widely used for energy generation. Interestingly, the only protein known to be involved in a terminal electron accepting process identified was a *Gplasma* cytochrome *c* oxidase identified in one of the sunken samples. The detection of terminal oxidase activity is important, given that no methods exist for direct measurement of oxygen concentrations in AMD solutions. Oxygen may be introduced upstream in flowing solutions, but its concentration is calculated to be exceedingly low (Druschel et al. 2004). Based on the low-oxygen availability we contend that an alternate electron accepting process predominates, and ferric iron is an obvious candidate, especially given its high concentration in AMD solutions (tens of mM), the known ability of *Ferroplasma* to reduce ferric iron (Dopson et al. 2004), and our *Aplasma-Ferroplasma* enrichments showing iron reduction. To date, no terminal iron-reducing proteins have been annotated in these organisms, and indeed, knowledge of enzymes involved in this process is limited for the archaeal domain (Schroder et al. 2003).

Analysis of *Gplasma* proteins suggests a distinct metabolism in the sunken compared to floating biofilm environments. The observation that, when growing in the sunken biofilms, *Gplasma* appears to emphasize amino acid and carbohydrate metabolism relative to protein biogenesis and posttranslational modification, may indicate greater investment in substrate oxidation than in biosynthesis. This may reflect increased energy generation by substrate oxidation to compensate for lower energy yields associated with an anaerobic respiration. Notably, the relatively high abundance in sunken samples of proteins for transformations of pyruvate (Table S2) suggests that pyruvate may be an important node in the metabolic flux of *Gplasma* in this environment. The overabundance of superoxide dismutase in the sunken samples might indicate oxidative stress, although recent findings have shown high expression of this enzyme in *Geobacter* species carrying out iron reduction, regardless of oxygen exposure (Mouser et al. 2009).

ARMAN are typically rare members of AMD biofilm communities (Baker et al. 2010). Their apparent higher abundance by FISH and metagenomic measures than by proteomic analysis is likely due to the small cell volume and perhaps lower activity of these cells (Baker et al. 2010). While many of the ARMAN proteins identified were

associated with protein biosynthesis or were of unknown function, we also detected proteins involved in fatty acid oxidation, TCA cycle, and the Embden-Meyerhof glycolysis pathway.

As noted above, we surmise that *Ferroplasma* and/or *Aplasma* carry out chemoorganotrophic growth coupled to iron reduction. Interestingly, the native-biofilm extracts supported growth without yeast extract through three transfers, although the reasons for the failed growth in the fourth transfer are unclear (changes in the biofilm-extract due to prolonged

storage may have been important in this regard). The precise role of yeast extract for growth on other carbon sources is unclear, although it could function as an additional carbon source or provide micronutrients. For *T. acidophilum*, yeast extract has also been suggested protect against the high pH gradient, although the mechanism for this is unclear (Smith et al. 1975).

Leptospirillum group II and group III are important iron oxidizers in AMD systems (Tyson et al. 2004; Goltsman et al. 2009; Deneff, Kalnejais, et al. 2010a; Mueller et al. 2010) and *Leptospirillum* group III is able to fix nitrogen (Tyson et al. 2004). Compared to *Leptospirillum* spp., *Sulfobacillus* spp. tend to be in lower relative abundance (Baker & Banfield 2003). Studies have indicated that species of *Sulfobacillus* are facultative anaerobes capable of both autotrophic and heterotrophic modes of growth (Druschel et al. 2004; Johnson et al. 2008). Oxidation of Fe(II), S⁰, and sulfide minerals, anaerobic reduction of Fe(III), and utilization of yeast extract, glucose, mannose, and other carbon sources has been described for various *Sulfobacillus* species (Johnson et al. 2008). *Sulfobacillus* species were detected predominantly in 5way biofilms, but we have not clearly identified ecological roles for these organisms: their broad metabolic capabilities may be indicative of a generalist ecological strategy. Currently, incomplete genomic information for *Sulfobacillus* species prevents wide-spread detection of *Sulfobacillus* proteins, limiting insights into their functional contributions.

Both FISH and DNA analyses showed a higher relative abundance of archaeal species than did proteomics analyses, suggesting a discrepancy between FISH, metagenomic, and proteomic measures of community composition. We attribute the discrepancy to the persistence of bacterial proteins in lysed and degrading cells, as indicated by extensive detection of acid-hydrolyzed peptides. Moreover, proteins with a high degree of solvent exposure (*i.e.*, unfolded) are more likely to undergo deamidation (Catak et al. 2006), and low pH can also greatly accelerate this reaction (Joshi & Kirsch 2002). Given the large difference in pH between the intracellular and extracellular environments of acidophilic organisms (Macalady et al. 2004), we suspect that acidophilic cells that have lost membrane integrity may show a greater degree of deamidation due to exposure of cytoplasmic proteins to low-pH solution. The difference between glutamine and asparagine deamidation frequencies is likely due to higher reactivity of glutamine residues in acidic conditions (Joshi & Kirsch 2002).

While some role for Bacteria (particularly species of *Sulfobacillus*) in nutrient cycling in the sunken communities cannot be ruled out, the dominance of Archaea by several measures of community composition, proteomic signature of heterotrophic growth, heterotrophic growth in anaerobic culture, and the high degree of amino acid deamidation in the Bacteria, indicate that Archaea drive nutrient cycling in suboxic and anoxic AMD environments. Complete submersion in AMD solution would appear to select against Bacteria, which may be less adapted to the low oxygen availability. The diversity of Archaea is particularly interesting, and may reflect a range of ecological niches in these high carbon, suboxic and anoxic environments.

Table 1.1: List of FISH probes.

Probe	<i>E. coli</i> position	Sequence (5' è 3')	Target	Formamide percent	Source
ARC915	915-934	GTGCTCCCCCGCCAATTCCT	Domain Archaea	35	(Amann, <i>et al.</i> , 1995)
EUB338 mix	338-355	GCTGCCTCCCGTAGGAGT	Domain Bacteria	35	(Amann, <i>et al.</i> , 1990, Daims, <i>et al.</i> , 1999)
	338-355	GCWGCCACCCGTAGGTGT	Domain Bacteria	35	(Amann, <i>et al.</i> , 1990, Daims, <i>et al.</i> , 1999)
SUL230	230-247	TAATGGGCCGCGRGCYCC	<i>Sulfobacillus</i>		Adapted from (Bond & Banfield, 2001)
FER656	656-674	CGTTTAACCTCACCCGATC	<i>Thermoplasmata</i>	25	(Edwards, <i>et al.</i> , 2000)
LF655	655-673	CGCTTCCCTCTCCCAGCCT	<i>Leptospirillum</i> groups I, II and III	35	(Bond & Banfield, 2001)
ARMAN mix	331-350	GTCTCAGTACCCTTCTGGGG	ARMAN	20	(Baker, <i>et al.</i> , 2010)
	329-348	CCGYAGTGCTAGGGTCCTTC	ARMAN	20	(Baker, <i>et al.</i> , 2010)
	424-442	GGCAAAAGTTCCCTTCCGG	ARMAN	20	(Baker, <i>et al.</i> , 2010)
	927-944	ACCCGTTTTTGTGCTCCC	ARMAN	20	(Baker, <i>et al.</i> , 2010)

Table 1.2: List of samples used in this study

Sample	Site	Date	pH	Temperature (°C)	Reference
GS1					
Sample 4	C drift + 8-10 m	Nov-06	1.01	39.5	(Mueller 2010)
Sample 11	AB Muck +20 m	Jun-06	0.72	36.0	(Mueller 2010)
Sample 14	AB Muck +20 m	Jun-06	0.72	36.0	(Mueller 2010)
Sample 17	AB Muck	Jun-06	0.72	36.0	(Mueller 2010)
Sample 28	C drift + 8-10 m	Nov-05	0.5	40.0	(Mueller 2010)
Sample 29	C drift + 30 m	Nov-04	ND	41.0	(Mueller 2010)
Sample 30	C drift +20 m	Nov-04	ND	41.0	(Mueller 2010)
Sample 31	C drift	Nov-04	ND	41.0	(Mueller 2010)
Sample 39	B drif + 1-5m	May-07	0.99	46.0	(Mueller 2010)
Sample 41	AB Muck	Aug-07	0.83	37.1	(Mueller 2010)
GS2 (Mueller 2010)					
Sample 7	AB Muck	Nov-06	1.01	39.5	(Mueller 2010)
Sample 8	AB Muck	Nov-06	1.18	42.7	(Mueller 2010)
Sample 9	C drift +8-10 m	Nov-06	1	43.0	(Mueller 2010)
Sample 10	C drift + 75 m	Nov-06	0.7	43.0	(Mueller 2010)
Sample 20	C drift + 75 m	Aug-06	0.72	36.0	(Mueller 2010)
Sample 25	C drift + 75 m	Jun-06	0.92	39.3	(Mueller 2010)
Sample 33	AB Muck	Jun-06	0.93	42.0	(Mueller 2010)
Sample 40	C drift +8-10 m	Aug-06	1	41.0	(Mueller 2010)
Sample 49	B drift	Mar-05	1.12	40.2	(Mueller 2010)
5way Floating GS2	5way	Feb-08	0.98	38.0	<i>this study</i>
Submerged					
5way-Sunken1	5way	Feb-08	0.98	38.0	<i>this study</i>
5way-Sunken2	5way	Feb-08	0.98	38.0	<i>this study</i>
5way-Sunken3	5way	Feb-08	0.98	38.0	<i>this study</i>
UBA-Sunken1	UBA	Jun-09	1.1	38.0	<i>this study</i>
UBA-Sunken2	UBA	Jun-09	1.1	38.0	<i>this study</i>
UBA-Sunken3	UBA	Jun-09	1.1	38.0	<i>this study</i>

Table 1.3: FISH analysis results for the sunken biofilm samples. Numbers indicate percentages (and SD) of DAPI-stained cells.

	5way-Sunken1	5way-Sunken2	UBA-Sunken1	UBA-Sunken2	UBA-Sunken3
EubMix	28.1 (8.9)	17.0 (8.75)	24.2 (8.5)	21.8 (10.3)	23.2 (9.9)
ARC915	50.1 (11.9)	65.4 (10.1)	73.1 (5.4)	78.2 (4.7)	76.9 (7.1)
LF655	10.8 (6.8)	<5	20.6 (7.9)	13.9 (10.1)	23.2 (5.9)
SUL230	15.6 (10.0)	13.1 (11.9)	<5	<5	<5
ARMAN 4/5	23.1 (7.4)	16.7 (6.5)	<1	<1	<1

*<1 and <5 indicates qualitative assessment only

Table 1.4: Results of pair-wise comparisons of deamidation frequencies between archaea and bacteria in early growth stage (GS1), late growth-stage (GS2) and sunken biofilms using Tukey's honestly significant difference test. NS = not significant; * = $p < 0.05$; ** = $p < 0.01$; *** $p < .001$.

	Archaea GS1	Archaea GS2	Archaea Sunken	Bacteria GS1	Bacteria GS2	Bacteria Sunken
Archaea GS1		NS	***	NS	***	***
Archaea GS2	NS		**	NS	***	***
Archaea Sunken	NS	NS		NS	NS	**
Bacteria GS1	NS	NS	NS		***	***
Bacteria GS2	*	NS	NS	*		*
Bacteria Sunken	***	***	***	***	**	

Figure 1.1 Map of Richmond Mine showing sampling locations

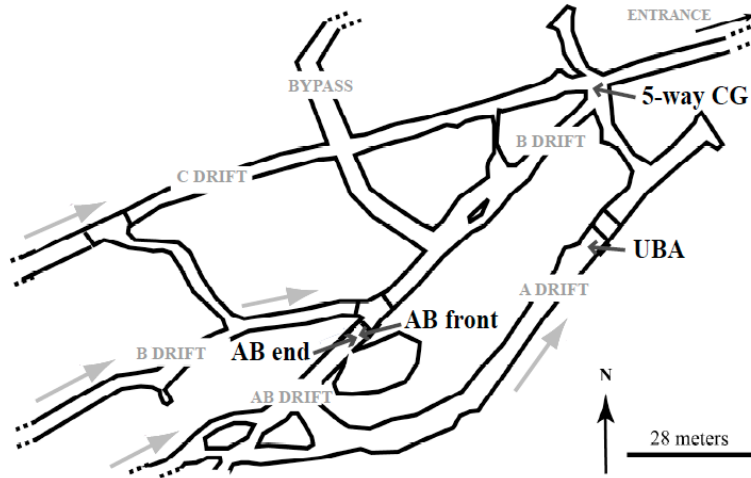


Figure 1.2 (A) Schematic showing the 5way sampling site and the stratification of the sunken samples. (B) Picture taken of the 5way sampling site showing the floating biofilm on top, with the sunken biofilm visible beneath.

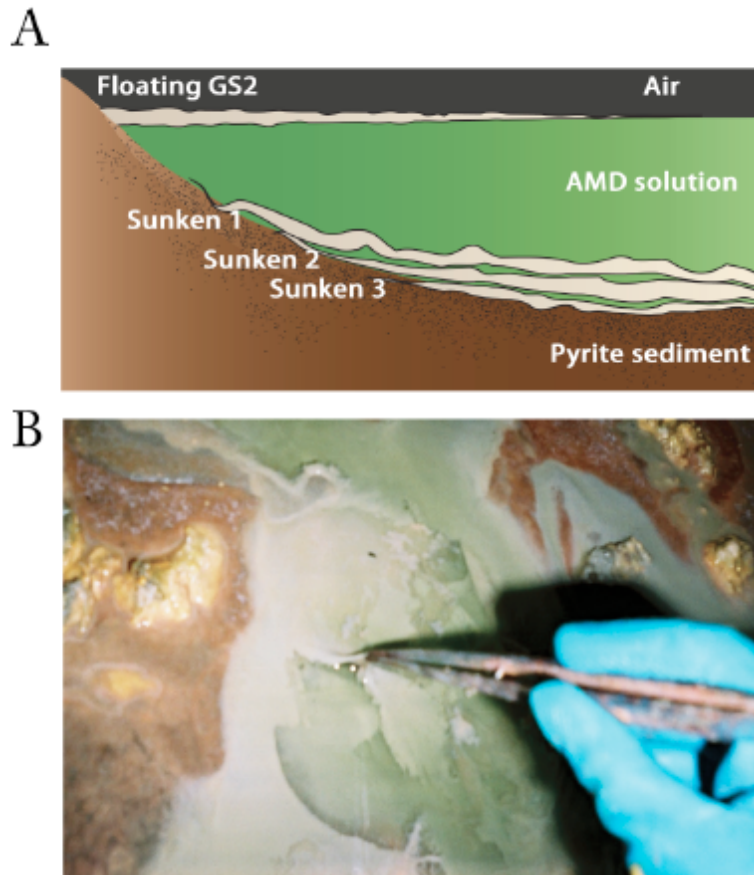


Figure 1.3 FISH images of 5way-Floating GS2 (left) and 5way-Sunken 1 biofilms for comparison of bacterial (eubmix probe, red) and archaeal (arc915 probe, green) populations

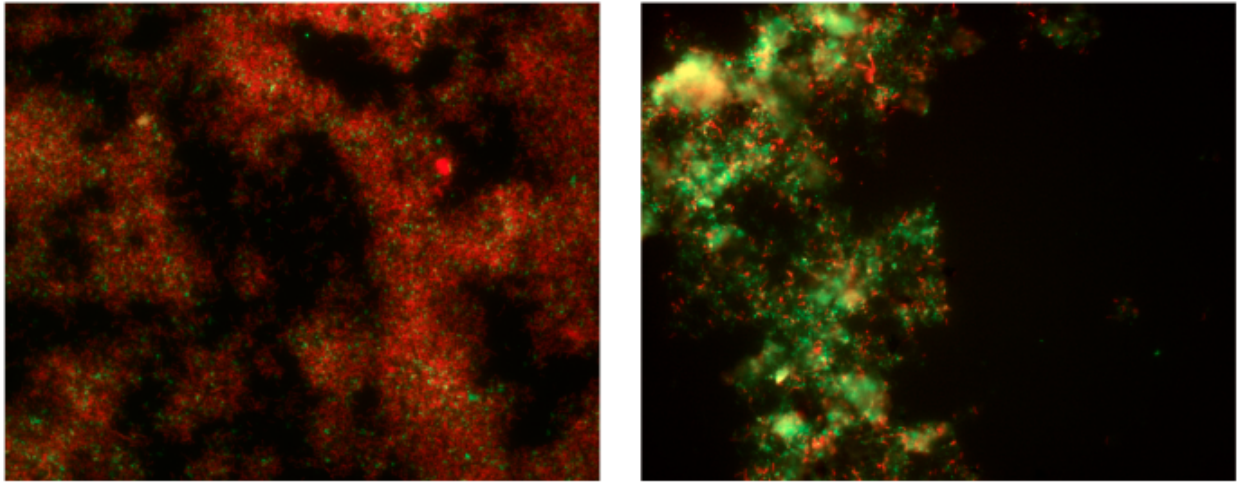


Figure 1.4. Percentages of DAPI-stained cells detected by FISH with either bacterial (EubMix) or archaeal-specific (Arc915) probes. Error bars represent +/- SD.

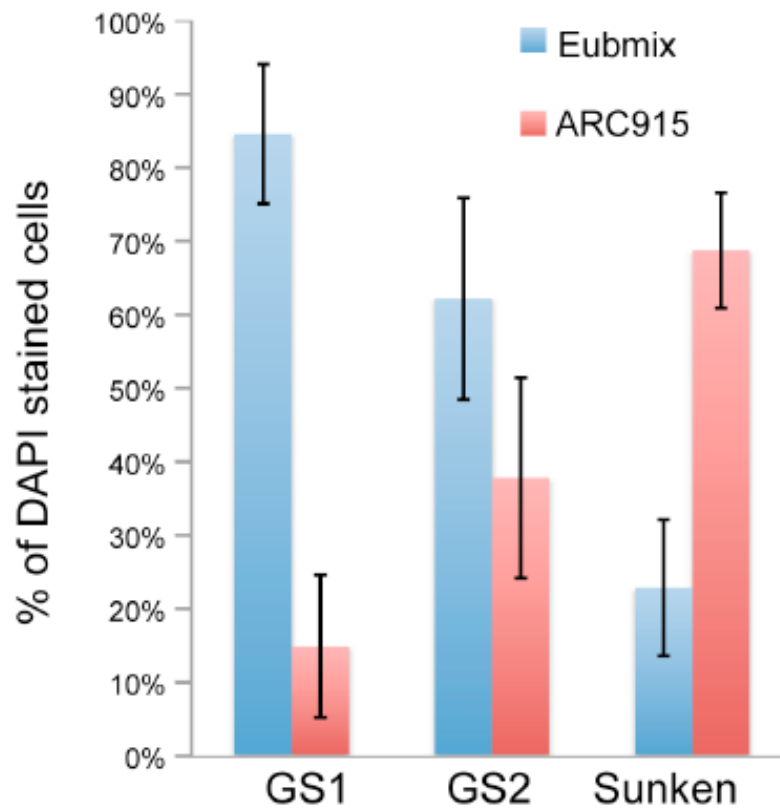


Figure 1.5. Phylogenetic tree of archaeal clones generated using the maximum-likelihood method. Asterisks next to clusters identify the number of clones belonging to that cluster. Bootstrap values are shown at the respective nodes. Scale bars are equal to 0.05 changes per site.

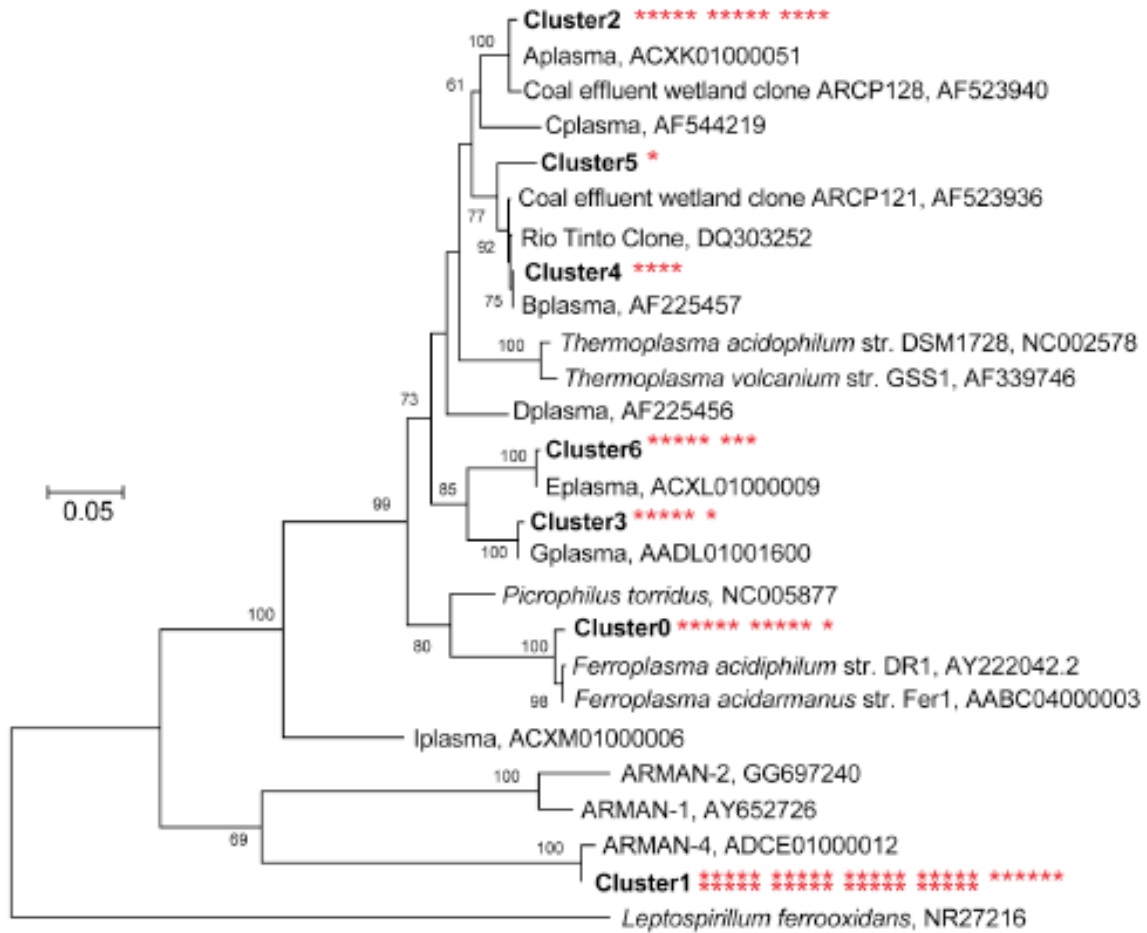


Figure 1.6. Phylogenetic tree of bacterial clones generated using the maximum-likelihood method. Asterisks next to clusters identify the number of clones belonging to that cluster. Bootstrap values are shown at the respective nodes. Scale bars are equal to 0.05 changes per site

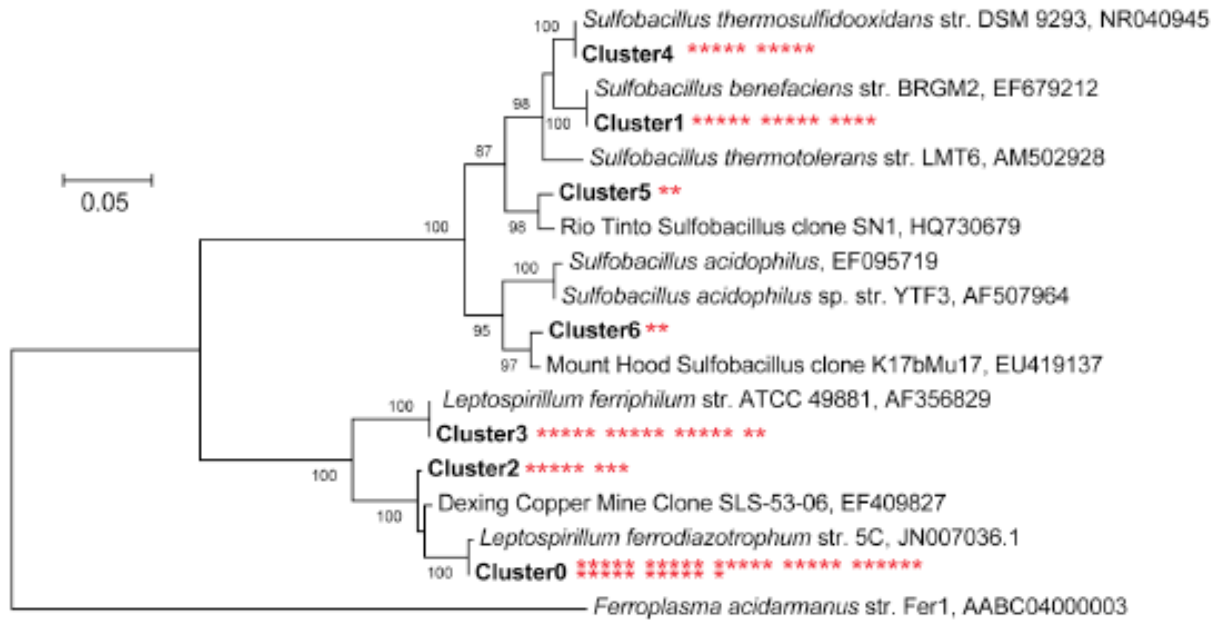


Figure 1.7. Bar graph showing average percent of proteome composition for floating early growth stage (GS1), floating late growth stage (GS2), and sunken biofilms. Percents are based on normalized spectral abundance factors. "Other" includes unassigned proteins, viral, and plasmid-associated proteins. Error bars represent +/- SD.

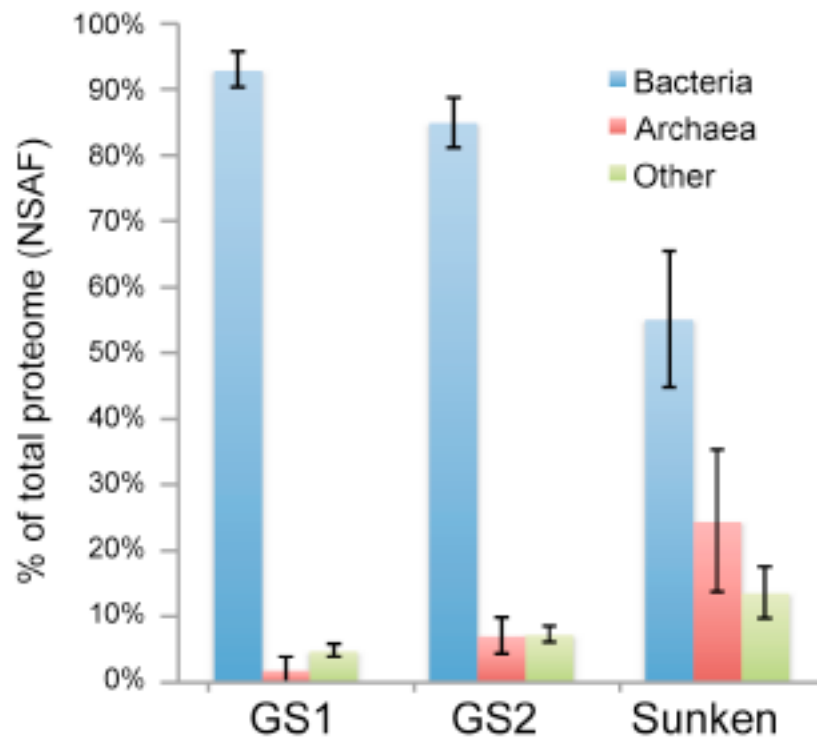


Figure 1.8. Percent of proteome composition for 5way samples. Percents are based on normalized spectral abundance factors.

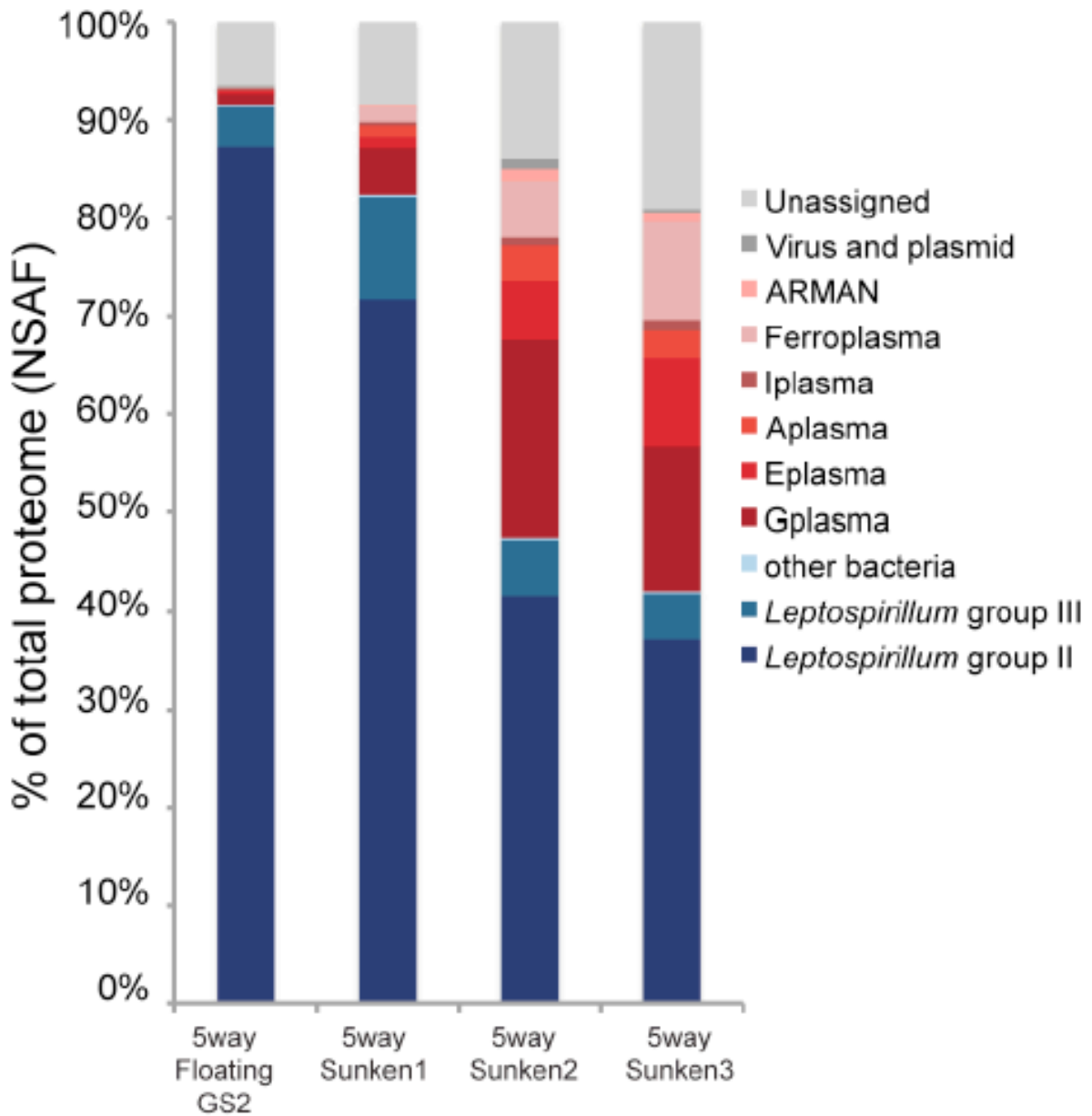


Figure 1.9. Hierarchical clustering of Gplasma proteins in >70% of samples using protein abundance data. Yellow is overrepresented and blue is underrepresented. The cluster of proteins uniformly overrepresented in sunken samples is indicated at the top. (B) The COG functional distribution of proteins in the sunken cluster versus the distribution of all the other proteins on the heatmap.

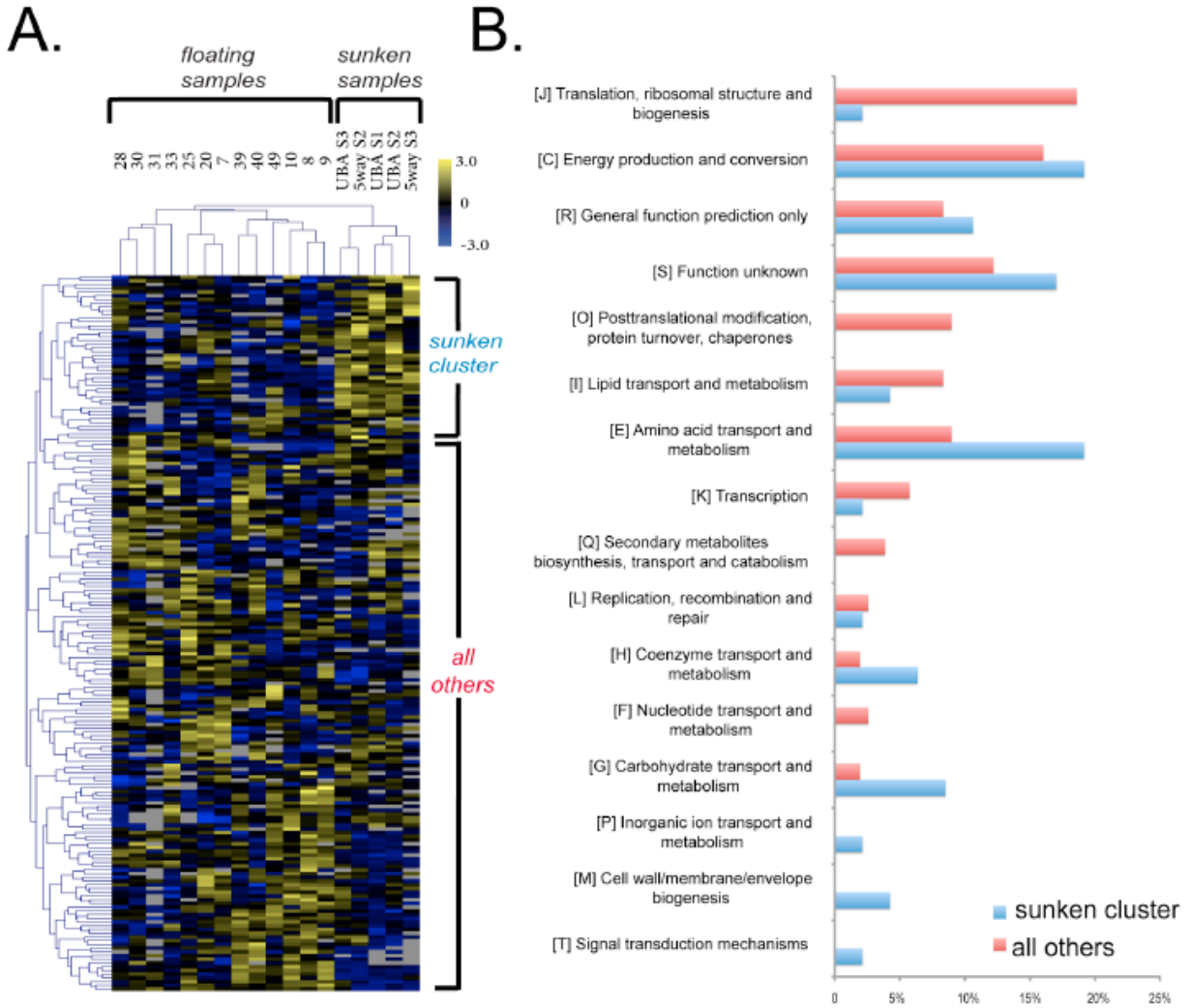


Figure 1.10. Composition of 5way-Sunken2 and UBA-Sunken2 communities based on percentages of 454 reads with BLAST matches against the genomic database (>98% identity), normalized by genome size.

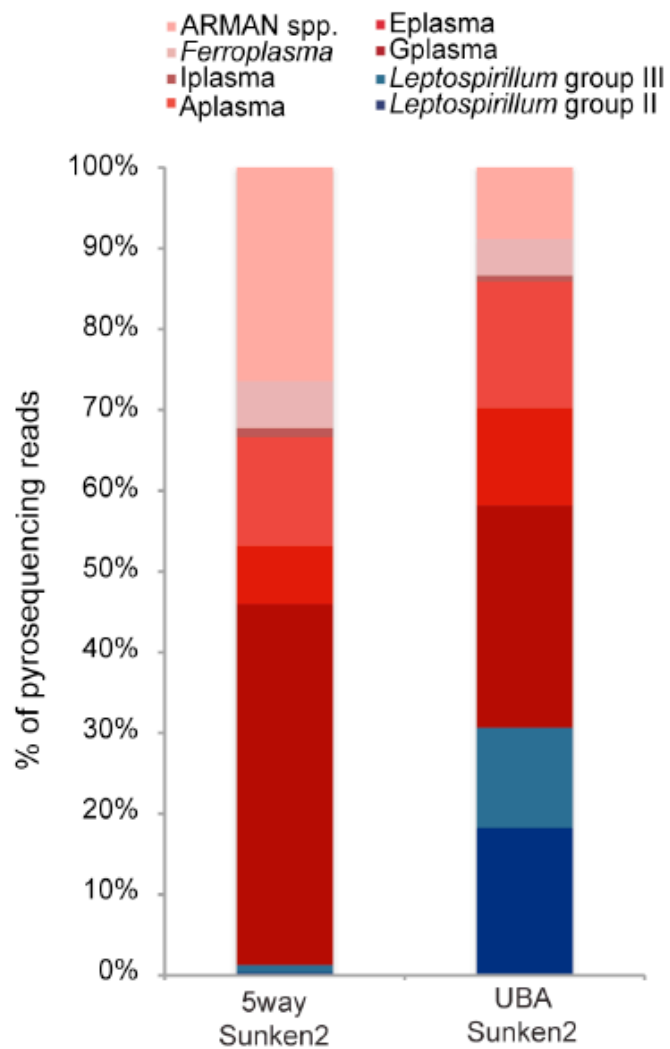


Figure 1.11. Deamidation of bacterial and archaeal asparagine (A) and glutamine residues (B) across growth stage 1 (GS1), growth stage 2 (GS2), and sunken samples. Error bars represent 95% confidence intervals, and italic letters correspond to groups that are not significantly different within the given bargraph by Tukey's honest significance test ($p < 0.05$).

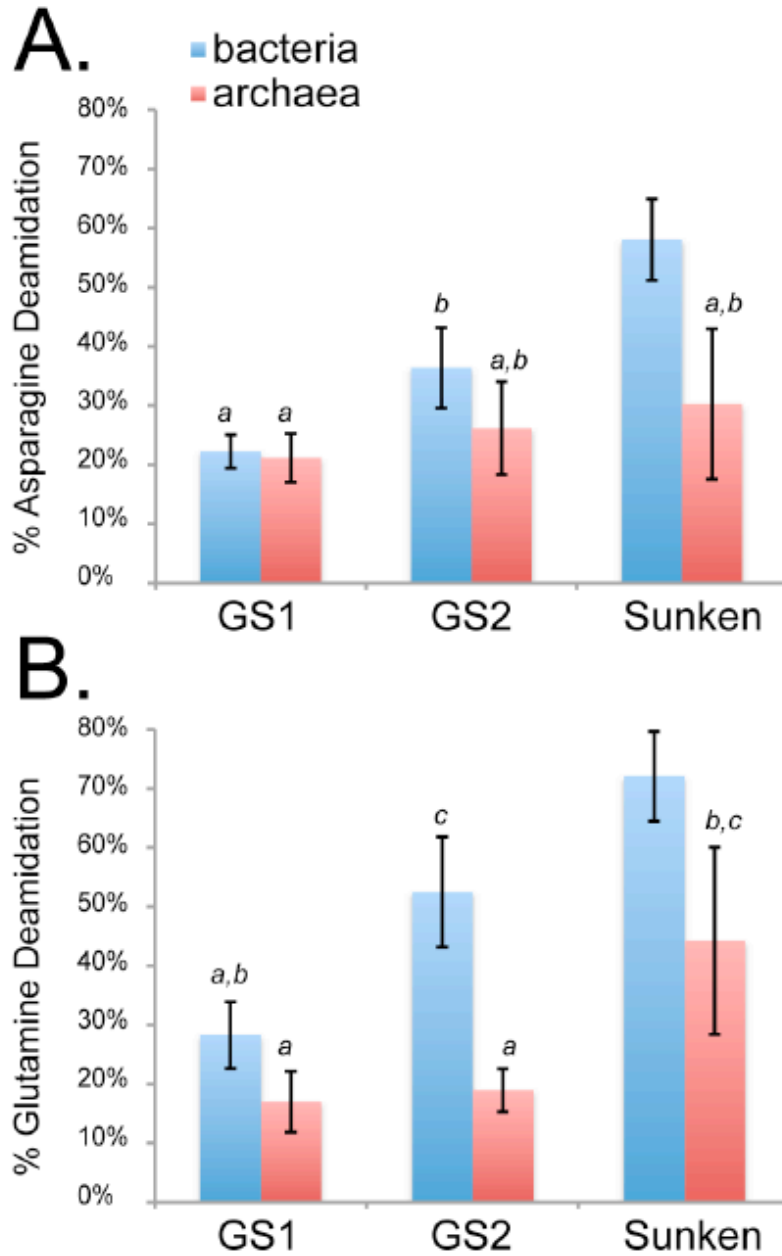


Figure 1.12. Analyses on bioreactor-cultivated floating and sunken biofilms. (A) FISH analysis of bioreactor floating and sunken biofilms. Error bars represent +/- SD. (B) Proteomic analysis of floating and sunken biofilms by NSAF. (C) Frequencies of glutamine deamidation and asparagine deamidation for Bacteria (blue) and Archaea (red).

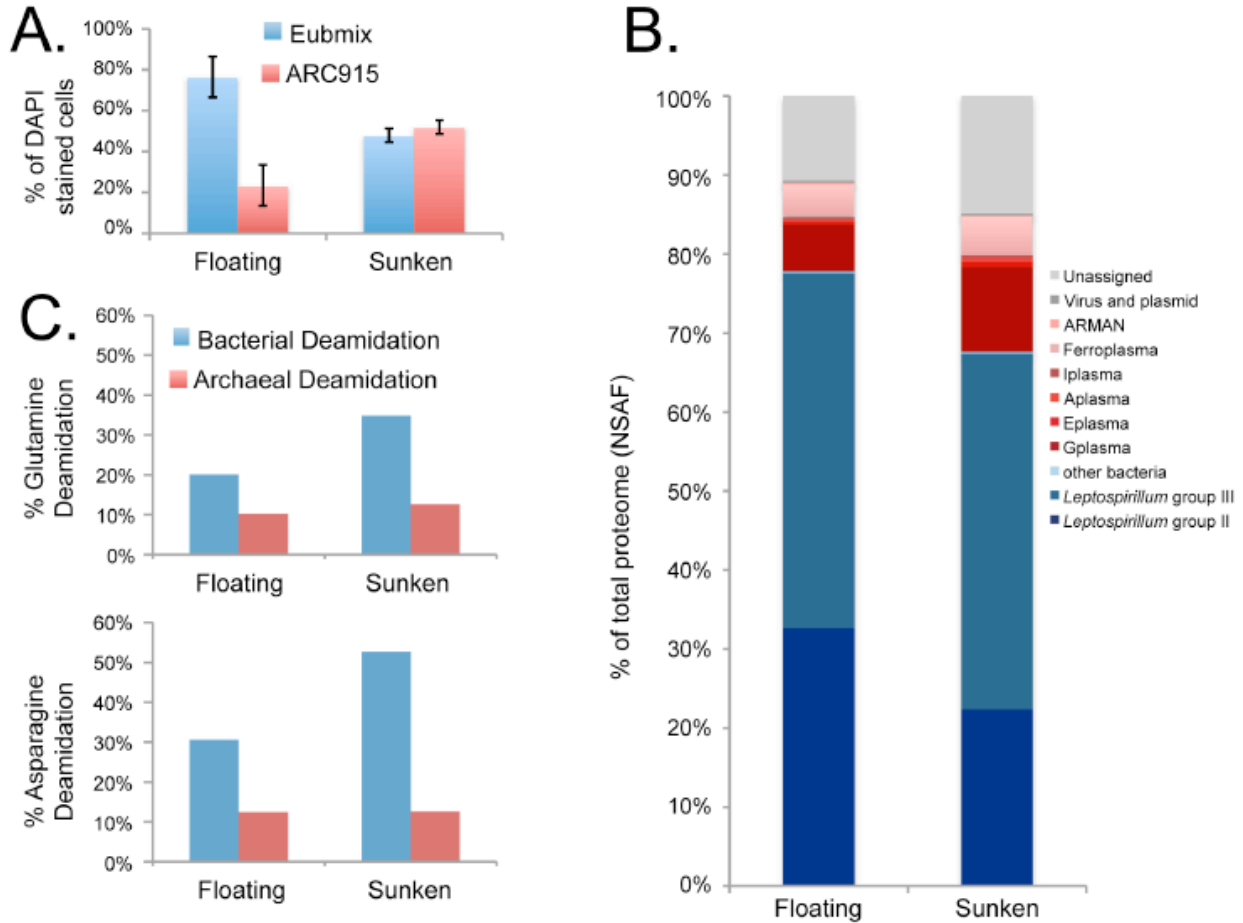
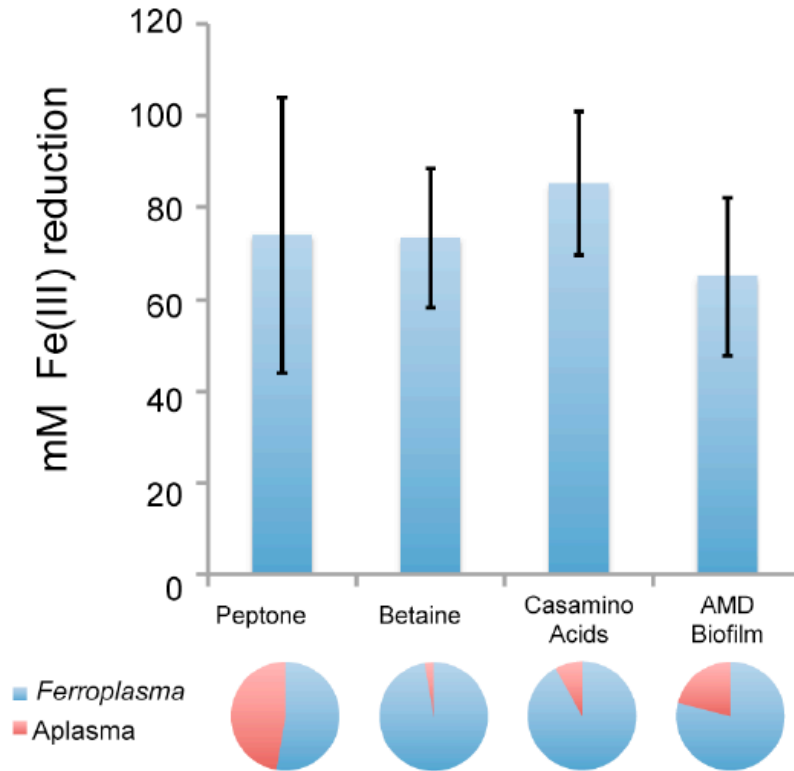


Figure 1.13 Bar graph (top) represents ferric iron reduction (mM) in anaerobic enrichments on different carbon substrates. Error bars are +/- SD. Pie charts (bottom) show representative community composition based on 16S clone libraries (at least 50 clones each)



Chapter 2

¹⁵N- and ²H proteomic stable isotope probing resolves microbial responses to perturbation and links nitrogen flow to protein degradation by Archaea

Abstract

Understanding how each species in a microbial community contributes to nutrient transformations in its environment is critical to prediction of overall ecosystem function, and thus is a major goal of microbial ecology. Proteomic stable isotope probing (SIP) provides a method of tracing nutrient flow by detection of stable isotope incorporation into thousands of proteins of coexisting members of a community. We conducted microcosm experiments in which floating acid-mine drainage (AMD) microbial biofilms were submerged—recapitulating the final stage in a natural biofilm lifecycle. Biofilms were amended with either ¹⁵NH₄⁺ or deuterium oxide (²H₂O) to examine to what extent these molecules were used in protein synthesis by different members of the community. Bacteria synthesized proteins to different extents in anaerobic iron-reducing, aerobic iron-reducing, and aerobic iron-oxidizing environments using the free ¹⁵N-ammonium in the media. Specifically, *Sulfobacillus* spp. synthesized protein almost exclusively under iron-reducing conditions whereas the *Leptospirillum* spp. synthesized protein in all conditions, with a clear emphasis on iron-oxidation metabolisms in the presence of Fe²⁺ and oxygen. These findings highlight distinct roles for *Sulfobacillus* vs. *Leptospirillum* in iron cycling. The greatest extent of ¹⁵N atom incorporation was detected in proteins of the *Leptospirillum*, whereas *Sulfobacillus* proteins had a low extent of ¹⁵N incorporation. Different atom% ¹⁴N to ¹⁵N values reflect distinct modes of nitrogen acquisition, since ¹⁴N is ultimately derived from extant organic biomass and ¹⁵N is derived from inorganic ammonia provided in the media. There were relatively few ¹⁵N-enriched archaeal proteins and all showed low atom% enrichment, consistent with Archaea synthesizing protein using the ¹⁴N derived from recycled biomolecules. Moreover, in parallel experiments using ²H₂O, extensive archaeal protein synthesis was detected in all conditions. In contrast, the bacterial species showed little protein synthesis when incubated in ²H₂O. The nearly exclusive ability of Archaea to synthesize proteins using ²H₂O may be due to archaeal heterotrophy (whereby Archaea offset deleterious effects of ²H by accessing ¹H generated by respiration of organic compounds) or differences in how archaeal versus bacterial membranes (and their associated mechanisms of energy conservation) respond to ²H₂O. Archaeal protein synthesis increased in the presence of oxygen and ferrous iron, and the newly synthesized proteins had similar functions between treatments, suggesting similar metabolisms under all conditions. This is one of the first demonstrations of the use of multiple isotopes in proteomic SIP experiments to resolve nutrient cycling in a microbial community.

Introduction

Microorganisms play pivotal roles in nutrient transformations and are major determinants of global elemental fluxes. Knowledge of how nutrients flow between members of a microbial community, as well as identification of organisms carrying out specific processes, are required for understanding how ecosystems function and respond to perturbation.

Stable-isotope probing (SIP) techniques are widely deployed to decipher the processes carried out by specific microorganisms (Dumont & Murrell 2005; Gutierrez-Zamora & Manfield 2010). Generally, SIP experiments involve incubation of an environmental sample with an isotopically-enriched substrate (e.g. ¹³CO₂), followed by the detection of isotopically-

enriched biomolecules like DNA, RNA, lipids, or protein. Lipid-based methods, such as ^{13}C -PLFA (phospholipid fatty acid) offer the advantage of being highly sensitive to low levels of ^{13}C incorporation, but often lack phylogenetic precision when identifying source organisms (Boschker et al. 1998). Nucleic acid based methods are commonly used because sequencing enables precise identification of the isotopically enriched taxa (Radajewski et al. 2000). However, due to the density-gradient based separation used in DNA or RNA-SIP, these methods require large degrees of isotope incorporation to differentiate labeled versus unlabeled molecules (Uhlík et al. 2009).

New protein-SIP methods allow quantification of isotope incorporation in proteins, and can thus provide both phylogenetic and functional information (Jehmlich, Schmidt, et al. 2010b; Pan et al. 2011). One well-described method uses a semi-automated approach to compare empirical peptide isotope distributions against theoretical distributions using a least-squares regression analysis (Taubert et al. 2011; 2013; 2012; Jehmlich, Schmidt, Hartwich, et al. 2008b; Jehmlich, Schmidt, Bergen, et al. 2008a). This approach was recently used to detect several hundred isotopically-enriched proteins in a mixed microbial consortia (Taubert et al. 2012). A second method uses the ‘half decimal place rule,’ which takes advantage of the characteristic shifts in the decimal place of measured peptide masses (Jehmlich, Fetzer, et al. 2010a; Fetzer et al. 2010). This approach offers the advantage of not requiring *a priori* knowledge of the protein sequence to determine isotope incorporation. A third approach, and the one employed in this study, determines the degree of ‘heavy’ isotope enrichment (%atom enrichment) across thousands of proteins, including closely-related strains and low abundance community members (Pan et al. 2011). This third approach was established in the study of acid mine drainage (AMD) biofilms (Pan et al. 2011), which have served as a model system for omics methods development (Denef, Mueller, et al. 2010b). The results showed that archaeal proteins tended to be less labeled by ^{15}N -ammonia than their bacterial counterparts in re-growing floating AMD biofilm.

Microbial organisms found in the Richmond Mine AMD system are adapted to growth at low pH (typically 0.5-1.2), elevated temperature (30-56 °C), and high concentrations of sulfate, iron, copper and zinc. *Leptospirillum* iron-oxidizing autotrophic bacteria create the initial biofilm by colonization of the air solution interface. Thicker, mature biofilms include other Bacteria (*Sulfobacillus* spp.), Archaea (of the *Thermoplasmatales* and ARMAN lineages) and microbial eukaryote (Baker et al. 2009; Denef, Mueller, et al. 2010b). Over time, the mature biofilms eventually sink into the AMD flow, a process that triggers a switch to an archaeal dominated community (Justice et al. 2012). Here, we replicated this process in laboratory bioreactors under three different growth conditions. ^{15}N SIP was used to determine the nitrogen source of Archaea and Bacteria members of the community (where ^{15}N is sourced from inorganic ammonia and ^{14}N from organic extant biomass). We expected that heterotrophic organisms consuming biomass with a natural abundance of ^{14}N would not show significant enrichment of ^{15}N . Furthermore, when submerged in $^2\text{H}_2\text{O}$, proteins synthesized by heterotrophic organisms can be confirmed through the detection of deuterium enrichment. We carried out the experiments under a range of environmentally relevant conditions to test for community structure and functional shifts.

Methods

Incubation and sampling: Late-growth stage biofilm was collected underground at the ‘AB Muck’ site of the Richmond Mine at Iron Mountain in Redding, CA (40°40’20”N, 122°31’40”W). The biofilm was stored at 4° C for approximately 8 hours while being transferred back to the laboratory, where it was cut into ~5 g pieces using a sterile scalpel. These pieces

were added to three different types of sterile media (50 mL): “Fe²⁺-aerobic”, “Fe³⁺-anaerobic”, and “Fe³⁺-aerobic.”

High concentrations of total iron (~150-350 mM) are measured in the Richmond Mine; typically most is present as ferrous iron (Fe²⁺) (Druschel et al. 2004). However, under some field conditions, ferric iron (Fe³⁺) can account for as much as 82 % of the dissolved iron (Druschel et al. 2004). High ferric iron solutions are common in bioreactor experiments, where re-reduction of ferric iron by interaction with pyrite is limited. 200 mM iron concentrations were chosen for the following experiments, given that iron-reducing enrichment cultures of archaeal organisms have been demonstrated to grow on 200 mM Fe³⁺ (Justice et al. 2012) and aerobic biofilm communities are routinely cultivated on 200 mM Fe²⁺ (Belnap et al. 2009; 2011; Pan et al. 2011).

The “Fe²⁺-aerobic” medium was the 9k-BR iron sulfate media published previously (Belnap et al. 2009; 2011). The medium consists of 200 mM FeSO₄ at pH 0.9, with trace mineral salts and 1 mM (NH₄)₂SO₄ 0.5 mM KCl, 0.050 mM K₂HPO₄, 5 mM MgSO₄·7H₂O, and 0.2 mM CaSO₄ (Belnap et al. 2009). The “Fe³⁺-anaerobic” and “Fe³⁺-aerobic” media are identical to the “Fe²⁺-aerobic” medium, except that ferric iron sulfate (100 mM Fe₂(SO₄)₃) was used in place of ferrous iron sulfate.

Fe³⁺-anaerobic treatments were vigorously degassed under a nitrogen stream for 20 minutes before the biofilm was added, and more gently for an additional 10 minutes following biofilm incubation. Both “Fe²⁺-aerobic” and “Fe³⁺-aerobic” treatments were incubated with exposure to atmospheric conditions. Iron measurements were carried out using the colorimetric ferrozine assay described previously (Stookey 1970).

For each of the three geochemical conditions, three replicates were treated with ¹⁵N, three with ²H₂O, and two without any heavy isotope. In treatments to be incubated with isotopically heavy nitrogen, the ammonium sulfate was replaced with ¹⁵N-ammonium sulfate (Cambridge Isotopes, Andover, MA). In treatments to be incubated with deuterated water, 97.96 atom% deuterium oxide was used and the pH lowered with several mL of >96 atom% ²H sulfuric acid (Cambridge Isotopes, Andover, MA). Iron measurements were made on all replicates. Protomic-SIP measurements were made on two of the three replicates of each geochemical condition-heavy isotope combination.

All samples were incubated, unstirred, for 14 days in a dark chamber at 40° C. Samples were disturbed only when media was extracted for iron measurements. On the 14th day, several grams of biomass were decanted into 50 mL centrifuge tubes, and separated from the media by spinning 6 minutes at 10,000 g, washed once with 20 mL H₂SO₄ (pH 1.0) and flash frozen in a dry ice ethanol bath and stored at -80° C until protein extraction.

Protein extraction and sample preparation

Proteins were extracted using an SDS protein extraction protocol as follows: between 500-750 mg of frozen biomass was resuspended in 1 mL SDS cell lysis buffer (5% SDS; 50 mM Tris-HCl, pH 8; 150 mM NaCl; 0.1 mM EDTA; 1 mM MgCl₂). 10 µl of fresh 5 M dithiothreitol (DTT) were added to reduce disulfide bonds and the samples heated in a 99° C water bath for 15 minutes. Lysed cells were then centrifuged 10 minutes at 10,000 g to remove cellular debris. The supernatant was transferred to a fresh tube, 300 µl 100% trichloroacetic acid were added, and the proteins precipitated overnight at 4° C. Precipitated proteins were centrifuged at 20,000 g for 20 minutes and the pellet washed three times with cold acetone. The pellet was resuspended in a guanidium chloride buffer (6 M guanidium chloride, 10 mM CaCl₂, 50 mM Tris pH 7.6) and

reduced with 10 mM DTT. The total protein concentrations were estimated with the BCA assay. 50 ug of protein from each sample was further processed with the Filter-aided Sample Preparation (FASP) method (Expedeon, CA), following the manufacture's protocol. Each sample was digested with sequencing-grade trypsin (Promega, WI) overnight in an enzyme:substrate ratio of 1:100(wt:wt) at room temperature with gentle shaking, followed by a second digestion for 4 hours. All digested peptide samples were stored at -80°C until analyses could be carried out.

Two-dimensional (2D) liquid chromatography (LC)-tandem MS (MS/MS) measurements

The multi-dimensional protein identification technology (MudPIT) was used in our analytical workflow. In each MudPIT run, 25 µg of peptides were loaded offline into a 150-µm-I.D. 2D back column (Polymicro Technologies, AZ) packed with 3 cm of C18 reverse phase (RP) resin (Luna, Phenomenex, CA) and 3 cm of strong cation exchange (SCX) resin (Luna, Phenomenex, CA). The back column was connected to a 100-µm-I.D. front column (New Objective, MA) packed in-house with 15 cm of C18 RP resin. The back column and front column were placed in-line with a U3000 quaternary HPLC (Dionex, CA). Prior to the measurement, the back column loaded with peptides was de-salted offline with 100% Solvent A (95% H₂O, 5% CH₃CN, and 0.1% formic acid), and washed with a 1 h gradient from 100% Solvent A to 100% Solvent B (30% H₂O, 70% CH₃CN, and 0.1% formic acid) to move peptides from RP resin to SCX resin. Each MudPIT run was configured with the 11 SCX-RP separations in 22 hours. 5%, 7%, 10%, 12%, 15%, 17%, 20%, 25%, 35%, 50%, and 100% of Solvent D (500 mM ammonium acetate dissolved in Solvent A) were used in the 11 SCX separations. Each SCX separation was followed by a 110 min RP separation with gradient from 100% Solvent A to 60% Solvent B. The MS/MS analysis was performed as described [Ref: Pan et al 2011] using an LTQ Orbitrap Elite mass spectrometer (Thermo Scientific). MS1 scans were acquired in the Orbitrap at the resolution of 30,000. The top 8 most abundant precursor peptide ions were selected for collisional-induced dissociation (CID). MS2 scans were measured in Orbitrap as well at the resolution of 15,000.

Peptide identification and isotope atom% estimation

Database searching was performed using SiproS 3.0 to identify peptides and proteins and estimate their stable isotope incorporation levels. The computation was carried out on a supercomputer, Titan. Single replicates of ¹⁵N-treated samples were initially searched against a database containing 79,633 proteins derived from ~80 Gb of genomic information obtained from previous metagenomic characterizations of the Richmond Mine AMD system (Tyson et al. 2004; Lo et al. 2007; Goltsman et al. 2009; Dick et al. 2009; Deneff & Banfield 2012) and unpublished). Based upon these preliminary results, additional searches were carried out with a refined database that excluded species not significantly detected in the inoculum (<1% abundance) and entirely absent in the isotopically-treated samples in order to reduce computational search space. Excluded species were the fungal organism *Acidomyces richmondensis* (Miller et al., in prep) and *Leptospirillum* Group IV (Aliaga Goltsman et al. 2013) final database included twenty-one complete or near-complete bacterial and archaeal genomes, several viral genomes, and an additional "unknown" category representing contigs that could not be assigned to specific organisms from assembly and binning procedures. These "unknown proteins" (labeled as AILUNK in the database) were assigned to broad phylogenetic lineages (e.g. *Archaea*, *Sulfobacillus*, *Leptospirillum*) based on BLAST similarity to the UniRef90

database. Peptide identification was carried out as described previously (Pan et al. 2011) using the following parameters: parent mass offsets of -3, -2, -1, 0, +1, +2, +3, parent mass tolerance of 0.05 Da, fragment mass tolerance of 0.02 Da, full tryptic peptides, up to 2 missed cleavages, isotopic enrichment levels from 0% to 100% at 1% interval. Because ^2H SIP can only measure the non-exchangeable hydrogen atoms in amino acid residues, exchangeable hydrogen atoms have a fixed natural enrichment level.

Peptide false discovery rates were calculated based on hits to reverse sequences as described previously (Junmin Peng et al. 2003). Peptide false discovery rates were found to be less than 0.1% in all samples.

Protein identification and isotope atom% estimation

Clustering of peptides from the ^{15}N labeling experiments was done as described previously (Pan et al. 2011). Briefly, identified peptides were assigned to proteins based on amino acid sequences predicted from the genome database. Because many proteins may exist in multiple isotopic variants (isotopologues, e.g., synthesized prior to vs. after isotope addition), peptides from the same protein were clustered based on their ^{15}N atom% using an agglomerative hierarchical clustering algorithm. In ^{15}N SIP, peptides were clustered into protein isotopologues until the average atom% difference between clusters was $>20\%$ and clusters containing a single peptide were discarded. Because of the lower false discovery rate of ^2H SIP peptide identification and the higher precision of ^2H atom% estimation, peptides were clustered to a minimum average atom% difference of 5% and clusters containing single peptides were retained.

Proteins were assigned to enriched categories if they had greater than 5% ^{15}N atom% enrichment in ^{15}N samples, and $>1\%$ ^2H atom% enrichment in $^2\text{H}_2\text{O}$ samples. These cutoffs were chosen because no enriched proteins were detected in the unlabeled inoculum above these values.

Proteome-based community compositional analysis

Spectral count information from the identified proteins in both labeled and unlabeled proteins was used for community composition analyses. Non-unique spectral counts (spectra that could not be uniquely assigned to a single peptide sequence) were split equally amongst all matching peptides.

Functional analysis

COGs were obtained based on the BLAST similarity results from eggNOG database version 2.0 (Muller et al. 2010). For counts of COG categories within each species, only proteins that could be uniquely assigned to that species were included.

Results

Iron reduction

Ferric and ferrous iron concentrations were determined during biofilm incubation to demonstrate active microbial respiration in the three geochemical environments. The cultures incubated in ^{15}N media that also contained Fe^{3+} but no oxygen (“ Fe^{3+} -anaerobic”) reduced approximately 76 (± 8) % of the Fe^{3+} after one week and approximately all of the iron (96 ± 9 %) by day 11 (Figure 2.1). In contrast, the cultures incubated in ^{15}N media that contained Fe^{3+} and oxygen (“ Fe^{3+} -aerobic”) reduced approximately 21 (± 3) % of the iron after one week and 43 (± 7) % after two weeks. Amongst acidophilic microorganisms, iron reduction does require strict

anaerobic conditions (Bridge & Johnson 1998), and compared to the Fe^{3+} -aerobic conditions, the lower rates of Fe^{3+} formation are expected due to the presence of oxygen as a more favorable terminal electron acceptor. Under the same conditions, the $^2\text{H}_2\text{O}$ incubations showed significantly less iron reduction, with only 57 (± 2) % iron reduced in the Fe^{3+} -anaerobic cultures and 9 (± 2)% in the Fe^{3+} -aerobic cultures. The difference in iron reduction between ^{15}N and $^2\text{H}_2\text{O}$ samples was expected due to the metabolic burden imposed by deuterium.

No net iron oxidation could be detected in experiments containing Fe^{2+} and oxygen (“ Fe^{2+} -aerobic”) in either $^2\text{H}_2\text{O}$ - or ^{15}N -treated samples. However, this does not preclude any iron oxidation activity, as the Fe^{3+} product may have been re-reduced by other community members.

Relative species composition of unlabeled proteins

Proteins without label incorporation are either existing proteins or proteins newly synthesized from recycled molecules. Greater than 96.4% of the proteome was unlabeled all samples, with most samples containing >99% unlabeled protein (Figure 2.2). We examined the species affiliation of unlabeled proteins, and compared results across the three experimental treatments. The unlabeled fractions of the $^2\text{H}_2\text{O}$ -treated samples had similar community structures across treatments and closely resembled the inoculum. In the $^2\text{H}_2\text{O}$ samples, *Leptospirillum* species contributed between 66 and 73 % of the community proteome (the inoculum was 74 % *Leptospirillum* spp.). Conversely, archaeal species contributed between 22 and 29 % of the community proteome (the inoculum was 23 % Archaea).

^{15}N -treated samples showed greater compositional differences across the three chemical environments than the deuterium-treated samples. Compared to the inoculum, the proteomes of biofilms from ^{15}N -treated ferric-iron experiments (Fe^{3+} -anaerobic and Fe^{3+} -aerobic) showed a greater percentage of archaeal protein (44 and 54 %, respectively), whereas the proteome from ^{15}N ferrous-iron conditions (Fe^{2+} -aerobic) samples more closely resembled the inoculum (29 % archaeal protein).

Relative species composition of labeled proteins

The incorporation of isotopically enriched substrate into protein is indicative of protein synthesis during the period of incubation. In ^{15}N -treated samples, proteins with >5% ^{15}N atom% enrichment were considered labeled, while in $^2\text{H}_2\text{O}$ -treated samples, proteins with >1% $^2\text{H}_2\text{O}$ atom% enrichment were considered labeled. These are empirically defined thresholds above which no labeled proteins are found in the inoculum controls. Compared to ^{15}N atom% estimation, the greater precision of $^2\text{H}_2\text{O}$ atom% estimation is believed to be an effect of the larger change in peptides’ isotopic distributions as measured by MS (caused by the greater number of non-exchangeable hydrogen atoms as compared to nitrogen atoms).

The numbers of $^2\text{H}_2\text{O}$ - or ^{15}N -labeled proteins belonging to different species are shown in Figure 2.3, and the spectral-count-weighted contribution to the total proteome shown in Figure 2.4. Bacteria showed the majority of isotope incorporation in ^{15}N -treated samples, indicating protein synthesis under these conditions. Conversely, Archaea showed the most isotope incorporation in $^2\text{H}_2\text{O}$ -treated samples. In combination, results for Archaea indicate protein synthesis, but not involving nitrogen derived from the labeled ammonium.

In the ^{15}N -treatments, *Leptospirillum* group II incorporated isotope under all conditions, and *Leptospirillum* group III showed isotope incorporation only in aerobic experiments. A

notable contrast between conditions was evident for *Sulfobacillus* species, where ^{15}N -labeled proteins were detected only in ferric-iron containing experiments (Fe^{3+} -anaerobic and Fe^{3+} -aerobic), with most found in biofilms cultivated in Fe^{3+} -aerobic conditions.

All archaeal species showed some evidence for the synthesis of ^2H -containing proteins, with the greatest synthesis of labeled proteins occurring in $^2\text{H}_2\text{O}$ -treated Fe^{2+} -aerobic conditions (Figure 2.3). The relative contribution of each archaeal species to the labeled archaeal proteome, however, was similar across all conditions (Figure 2.5). Amongst the Archaea, *Gplasma* accounted for the greatest amount of labeled-protein synthesis under all three conditions, an effect that remains evident when accounting for to the initial inoculum composition (Figure 2.5).

Metabolic function of proteins in ^{15}N -treated samples and varying degrees of isotopic enrichment
While the incorporation of isotopically enriched substrate into protein is indicative of amino acid biosynthesis during the period of incubation, the degree of isotope incorporation in labeled proteins reflects the $^{14}\text{N}/^{15}\text{N}$ or $^1\text{H}/^2\text{H}$ isotope ratios of the nitrogen and hydrogen pools as well as biological fractionation.

In ^{15}N -treated samples, isotopically enriched proteins were categorized in three distinct atom% groups (Figure 2.6). Category I—comprising the proteins with the lowest (5-12%) ^{15}N atom% enrichment—included proteins from *Leptospirillum*, *Sulfobacillus*, and archaeal species (Figure 2.7). Category II (between 22–63 ^{15}N atom% enrichment) is comprised exclusively of *Leptospirillum* proteins detected in biofilms incubated in the Fe^{3+} -aerobic environment. Finally, Category III proteins (>70 ^{15}N atom% enrichment), were derived predominantly from *Leptospirillum* spp. incubated in the Fe^{2+} -aerobic environment.

Category I is significant as it represent proteins synthesized during the course of the experiment wherein the majority of nitrogen was derived from sources other than the free ^{15}N ammonia in the media. All twelve ^{15}N -labeled archaeal proteins belonged to this category, and were only detected in biofilms incubated under aerobic conditions (both Fe^{3+} and Fe^{2+} ; Table 2.3). These included two proteins involved in amino acid metabolism (histidine ammonia-lyase and glutamine synthetase) as well a ribosomal protein, an flavin mononucleotide-reductase, aldo/keto reductase, and several proteins of unknown function. In addition, superoxide dismutases were identified in this group from both ARMAN4 and ARMAN5.

Nearly all isotopically-enriched *Sulfobacillus* proteins (33 of 34) belonged to the low ^{15}N atom% enrichment category (Category I; Table 2.2). Twenty-eight *Sulfobacillus* proteins were detected solely in the Fe^{3+} -aerobic environment, and included proteins involved in TCA cycle, transformations of pyruvate (and possible fermentation, e.g., lactate dehydrogenase), amino acid metabolism, and proteins involved in stress response (rubrerythrin and phage shock protein A; Table 2.2). Three labeled *Sulfobacillus* proteins identified in the Fe^{3+} -anaerobic environment were also identified in the Fe^{3+} -aerobic environment (succinate-CoA ligase, glucose 6-phosphate isomerase, and 2-methylcitrate synthase) and two were unique to the anaerobic environment (peroxiredoxin and ATP synthase). A glycerol kinase, also detected in the Fe^{3+} -aerobic environment, was the only *Sulfobacillus* protein detected in the high ^{15}N atom% enrichment category (Category III).

Leptospirillum proteins were found in each of the three enrichment categories. In Category I, nine *Leptospirillum* proteins were identified including all of the ^{15}N -labeled *Leptospirillum* proteins found in the Fe^{3+} -anaerobic sample (Table 2.1). In Category II (medium ^{15}N atom% enrichment), twelve *Leptospirillum* proteins were identified, and all were from biofilms incubated in Fe^{3+} -aerobic solution (Table 2.1). Among these proteins were two

cytochrome proteins (cytochrome 572 and a cytochrome 579) involved in iron-oxidation in *Leptospirillum* species (Singer et al. 2008; 2010). Proteins in Category III (high ^{15}N atom%, composed of proteins predominantly from Fe^{2+} -aerobic incubations) were similar in function to those found in Category II (middle ^{15}N atom%, composed of proteins predominantly from Fe^{3+} -aerobic incubations). Proteins shared between Categories II and III included both specific proteins (e.g., *Leptospirillum* group II 'UBA' cytochrome 572, *Leptospirillum* group III flagellin domain protein) as well as proteins with identical function although resolved to different *Leptospirillum* species (e.g., dnaK chaperone found in both *Leptospirillum* group III and *Leptospirillum* group II). These results indicate an emphasis on the synthesis of iron oxidation proteins by *Leptospirillum* spp. in both the Fe^{3+} -aerobic and Fe^{2+} -aerobic incubations.

Metabolic function of proteins in $^2\text{H}_2\text{O}$ -treated samples

In contrast to the discrete groups defined by the varying degree of ^{15}N isotope incorporation, the majority of the isotopologues in $^2\text{H}_2\text{O}$ -treated samples fell into a broad group between 5% and 45% ^2H atom%, centered around $\sim 20\%$ ^2H atom% (Figure 2.6). Only nine of the 1,043 ^2H -labeled proteins did not belong to archaeal species (Table 2.4). Notably, three of the non-archaeal proteins identified in the $^2\text{H}_2\text{O}$ samples were also detected in ^{15}N -enriched samples, including a cytochrome 572 and a chaperonin GroEL protein of *Leptospirillum*. *Leptospirillum* proteins varied between 12% ^2H atom% enrichment (Fe^{3+} -anaerobic environment) to 79% ^2H atom% enrichment (Fe^{2+} -aerobic environment). Three *Sulfobacillus* proteins were detected, varying from 10 to 30% ^2H atom% enrichment, including a chaperonin GroEL, and a peptidase. There were no systematic differences in the D atom% labeling between archaeal species, nor amongst COG functional categories (data not shown).

One thousand thirty-four different archaeal proteins were labeled in $^2\text{H}_2\text{O}$ -treated samples, of which 821 could be uniquely assigned to a single ORF. Biofilms incubated in the Fe^{2+} -aerobic environment showed the greatest degree of ^2H -labeled protein synthesis, with 709 labeled proteins detected of which 321 were detected in this condition only. In contrast, 55 proteins were found solely in biofilms cultivated in Fe^{3+} -anaerobic conditions and 34 were found in only biofilms cultivated in Fe^{3+} -aerobic conditions. Analysis of proteins unique to each condition did not reveal major metabolic differences between geochemical conditions (data not shown). The ^2H -labeled archaeal proteins unique to a geochemical condition were generally not found between replicate treatments, thus providing little indication that these proteins were part of a consistent response to the geochemical environment. Moreover, more than half (411) of all labeled archaeal proteins were detected in at least two conditions suggesting similar protein synthesis across conditions (Figure 2.8).

Differences were observed in functional category proportions when comparing the labeled versus unlabeled fractions of the $^2\text{H}_2\text{O}$ -treated samples. To investigate how *Gplasma*, the most abundant archaeal species, responded to submersion in $^2\text{H}_2\text{O}$, functional category comparison of labeled proteins (synthesized during the experiment) versus unlabeled proteins (presumably synthesized before the experiment) were carried out (Figure 2.9). The proportion of labeled proteins involved in carbohydrate metabolism, lipid metabolism, membrane biogenesis, and signal transduction was less than in the unlabeled fractions. In contrast, significantly greater fractions of labeled proteins were involved in energy production; inorganic ion transport; replication, recombination, and repair; and protein modification, chaperones, and protein turnover.

Discussion

There were two nitrogen sources for protein synthesis of the sunken community, $^{15}\text{NH}_4^+$ in the solution and ^{14}N -organic nitrogen in the decaying biomass—and the primary aim of this study was to identify which of these two sources each microbial community member used. This was accomplished both by the analysis of the degree of $^{14}\text{N}/^{15}\text{N}$ atom% enrichment as well as the use of $^2\text{H}_2\text{O}$ labeling which allowed for the detection of protein synthesis in the absence of ^{15}N incorporation. Archaeal protein synthesis from recycled biomass is supported by the incorporation of ^2H into archaeal protein in $^2\text{H}_2\text{O}$ -treated samples as well as the detection of few ^{15}N -labeled archaeal proteins, all at low atom% enrichment.

The low degree of bacterial protein synthesis in $^2\text{H}_2\text{O}$ -treated samples is not surprising, as the deleterious effect of deuterium on bacterial growth is well documented (Giovanni 1963; 1961; Thomson 1963). Although the $^2\text{H}_2\text{O}$ concentrations that inhibit bacterial growth have been shown to vary between species and strains, concentrations above 20% are often considered inhibitory (Thomson 1963). General responses to cultivation in deuterated water include long lag phases, slower rates of growth and respiration, and distorted cell shapes (Thomson 1963). Compared to a hydrogen atom, the heavier deuterium atom may disrupt molecular structures and processes through noncovalent interactions, and will particularly affect enzymatic reactions involving the formation or cleavage of hydrogen bonds via the primary kinetic isotope effect (Northrop 1975; Wade 1999).

Less is known about the effects of deuterium on Archaea, although similar principles should apply. Given the effects of deuterium on biological systems, it is interesting that Archaea were able to synthesize several orders of magnitude more protein than the Bacteria, suggesting fundamental differences in metabolic behaviors of these organisms. We posit that the Archaea, as heterotrophs, may have been better able to limit the relative concentration of ^2H in their cytoplasm via an increased import and degradation of various biomolecules (and thus increased catabolic release of ^1H as water or coupled to reducing equivalents like NAD(P)H). Additionally, the differences could be tied to Archaeal physiology, where perhaps differences in cell permeability lent the Archaeal organisms more resistance to cytoplasmic import of deuterium or, additionally, that the enzymatic reactions on which Archaea depend for energy are more resistant to the effects of deuterium (e.g., less dependence on hydride-dependent electron transfer mechanisms from NADH or NADPH).

While archaeal proteins showed clear incorporation of ^2H , the majority of ^2H atom% enrichments (8-40%, Figure 2.6), was much lower than might be expected based on the high ^2H concentration of the media (the media was ~96% $^2\text{H}_2\text{O}$ and allowing for a maximum 1/10 dilution from 5 g of biomass in 50 mL media leaves at least 86% ^2H). This discrepancy suggests some mechanism for enriching ^1H over ^2H in archaeal amino acids. In addition to the heterotrophic import and degradation of H-containing biomolecules described above, the kinetic isotope effect should contribute significantly to lower ^2H atom incorporation. Primary kinetic isotope effects are measures of the reaction rate differences of two isotopes of an atom when that atom is part of the bond being formed or broken. Primary kinetic isotope effects on reaction rates of 6 to 10-fold ($k_{\text{H}}/k_{\text{D}}$) have been characterized, reflecting a large enzymatic preference for the hydrogen atom over the deuterium atom (Northrop 1975). This effect is likely to account for the lower fractionation of ^2H into archaeal amino acids than would be expected based on ^2H concentration in the media alone.

In ^{15}N -treated biofilms, the kinetic isotope effects are negligible due to the small mass difference between ^{14}N and ^{15}N . In these treatments, labeled archaeal proteins were only slightly

enriched in ^{15}N , suggesting that that most of the nitrogen used in amino acid biosynthesis is derived from preexisting nitrogen-containing biomolecules. This mode of nitrogen acquisition is consistent with intracellular nutrient recycling or catabolism of externally derived peptides, nucleic acids, or other nitrogen containing molecules. Furthermore, compared to the inoculum, there was an increase in the relative abundance of unlabeled archaeal protein from *Thermoplasmatales* Archaea (in Fe^{3+} -containing environments) and ARMAN (in all treatments). This increase could be due to the synthesis of new archaeal protein with >95% organic ^{14}N , consequently below the 5% ^{15}N atom-enrichment threshold necessary for classification as ‘labeled.’

Heterotrophic metabolisms have been demonstrated for the Archaea, in enrichments of ARMAN and other members of the *Thermoplasmatales*, as well as with pure cultures of ‘*Ferroplasma acidarmanus*’ (Ziegler et al. 2013; Dopson et al. 2006). Previous work has shown that the *Thermoplasmatales* come to dominate the community in submerged biofilms (Justice et al. 2012), and the increased abundance of *Thermoplasmatales* protein in Fe^{3+} -containing environments indicates that Fe^{3+} is an important factor in this transition. It is possible that iron reduction with organic carbon as the electron donor is the dominant process here, as is seen in cultures of *Ferroplasma* spp. (Dopson et al. 2006). However, a loss of bacterial protein (i.e., through biotic or abiotic degradation) could also explain the increase in the relative abundance of archaeal protein in the unlabeled fraction of ^{15}N -treated, Fe^{3+} -containing cultures. We have shown previously that bacterial cell death via loss of membrane integrity and subsequent abiotic, acid-dependent deamidation of bacterial protein can occur under low oxygen/anaerobic conditions (Justice et al. 2012).

Compared to the inoculum, ARMAN were more abundant in all three conditions, suggesting Fe^{3+} was not a major factor in the increased abundance of protein from these Archaea. It must be noted that oxygen diffusion may have been limited even in the Fe^{2+} -aerobic and Fe^{3+} -aerobic conditions because the incubations were left unstirred and the biofilms were allowed to settle to the bottom of the culture bottles. This could potentially allow for anaerobic metabolisms to occur even in “aerobic” treatments. Evidence from ‘snottite’ biofilms in other AMD systems has shown that members of ARMAN are anaerobic, being sequestered in anaerobic niches in biofilms, and present in anaerobic enrichment cultures (Ziegler et al. 2013).

Like the Archaea, *Sulfobacillus* species showed low ^{15}N atom% incorporation, reflecting an organic- ^{14}N nitrogen assimilation strategy. However, unlike in the archaeal proteomes, protein synthesis in the absence of ^{15}N incorporation cannot be confirmed due to the lack of many ^2H -labeled proteins. *Sulfobacillus* species are facultative autotrophs, and can assimilate carbon using a wide range of organic molecules including monosaccharides, glutamic acid, citric acid, and glycerol (Johnson et al. 2008). During mixotrophic growth, however, as much as 20% of cellular carbon is still derived from the activity of the Calvin cycle (Wood & Kelly 1984). Although ribulose-1,5 biphosphate carboxylase (RuBisCo) proteins were not detected, putative carboxysome proteins (AMDSBA4_C47G00009) were, perhaps indicating autotrophic activity. Consistent with organic carbon assimilation, however, were the high frequency of labeled enzymes involved in the potential catabolism of protein, glycerol, and carbohydrates. The presence of labeled *Sulfobacillus* protein in Fe^{3+} -conditions is consistent with previous reports of *Sulfobacillus* conserving energy via ferric iron reduction (Tsaplina et al. 2010).

Leptospirillum spp. are aerobic iron-oxidizing autotrophs (Goltzman et al. 2009), and their proteins were detected at multiple ^{15}N atom% enrichments. Indeed, in anaerobic conditions, only a few *Leptospirillum* proteins were detected (all at low ^{15}N atom% enrichment). However,

these proteins were involved in biopolymer transport, redox reactions (flavin/NADH oxidase), and sugar degradation (phosphoglucomutase), suggesting that *Leptospirillum* may have a limited anaerobic capacity involving the consumption of biofilm carbon. In contrast, *Leptospirillum* proteins the middle ^{15}N atom% category (from the Fe^{3+} -aerobic treatment) and high ^{15}N atom% category (from the Fe^{2+} -aerobic environments) were similar, and clearly reflected an iron-oxidation metabolism (Table 2.1). The presence of iron-oxidation proteins in the Fe^{3+} -aerobic condition indicates iron cycling, whereby ferric iron—reduced by other community members—is made available as ferrous iron for *Leptospirillum* species. Moreover, *Leptospirillum* group III synthesized labeled flagellar proteins, which may be used to swim towards areas with higher oxygen concentrations.

Conclusion

Proteomic SIP offers a distinct advantage over other SIP methodologies in current use in environmental microbiology because both taxonomic and functional information are available. One of the most intriguing findings is the synthesis of archaeal protein in nearly 100% $^2\text{H}_2\text{O}$, in contrast to the minimal protein synthesis amongst the Bacteria. Further efforts towards understanding the cause of this discrepancy could provide a clearer picture of the physiological differences or lifestyles between these organisms. The use of both deuterium and ^{15}N isotopes in the proteomic SIP experiments enabled identification of nutrient cycling and identification of the roles of specific organisms. Both ^2H and ^{15}N isotopes provided evidence of archaeal derivation of nitrogen through the reclamation of nitrogen-containing biomolecules, further supported by the low degree of ^{15}N atom% incorporation into archaeal protein. The degree of ^{15}N atom incorporation also provides evidence of nutrient recycling in Bacteria, and the synthesized proteins reflect organism-specific responses to distinct geochemical conditions.

Co-author contributions:

Zhou Li, Yingfeng Wang, and Chongle Pan ran proteomics samples and carried out initial data analysis (peptide and protein atom% estimations). Susan E. Spaulding and Annika C. Mosier assisted with sample extractions. Chongle Pan, Annika C. Mosier, and Jillian F Banfield provided assistance editing text and interpreting results. Robert L. Hettich provided instrument support.

Figure 2.1. Diagram of experimental setup and iron reduction measurements. Top: Cartoon showing inoculum submerged in three different geochemical conditions in the presence of either ^{15}N or $^2\text{H}_2\text{O}$ isotopes. Bottom: percent iron reduction and oxidation for each of the conditions. Error bars represent +/- SD. Note that no detectable iron oxidation was observed in the ferrous iron aerobic conditions.

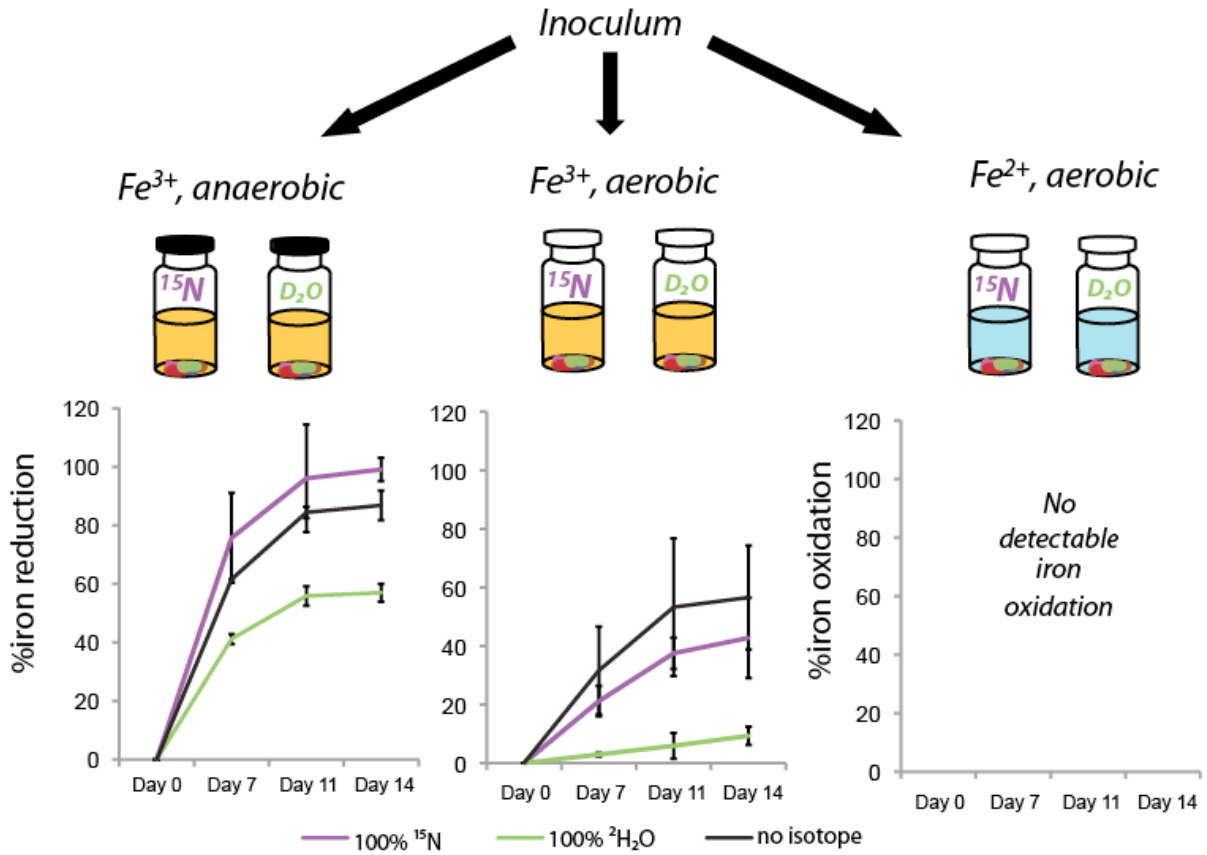


Figure 2.1. Species abundances calculated as a percent of total peptide-spectral matches (PSMs) averaged for each experimental treatment. “Nonunique *Leptospirillum*” includes PSMs that could not be confidently assigned to a single species or strain. “Other Archaea” includes archaeal PSMs that could not be matched to a single species as well as PSMs to *Dplasma* and *Iplasma* proteins. “Other” includes viral, plasmid, and unassigned proteins.

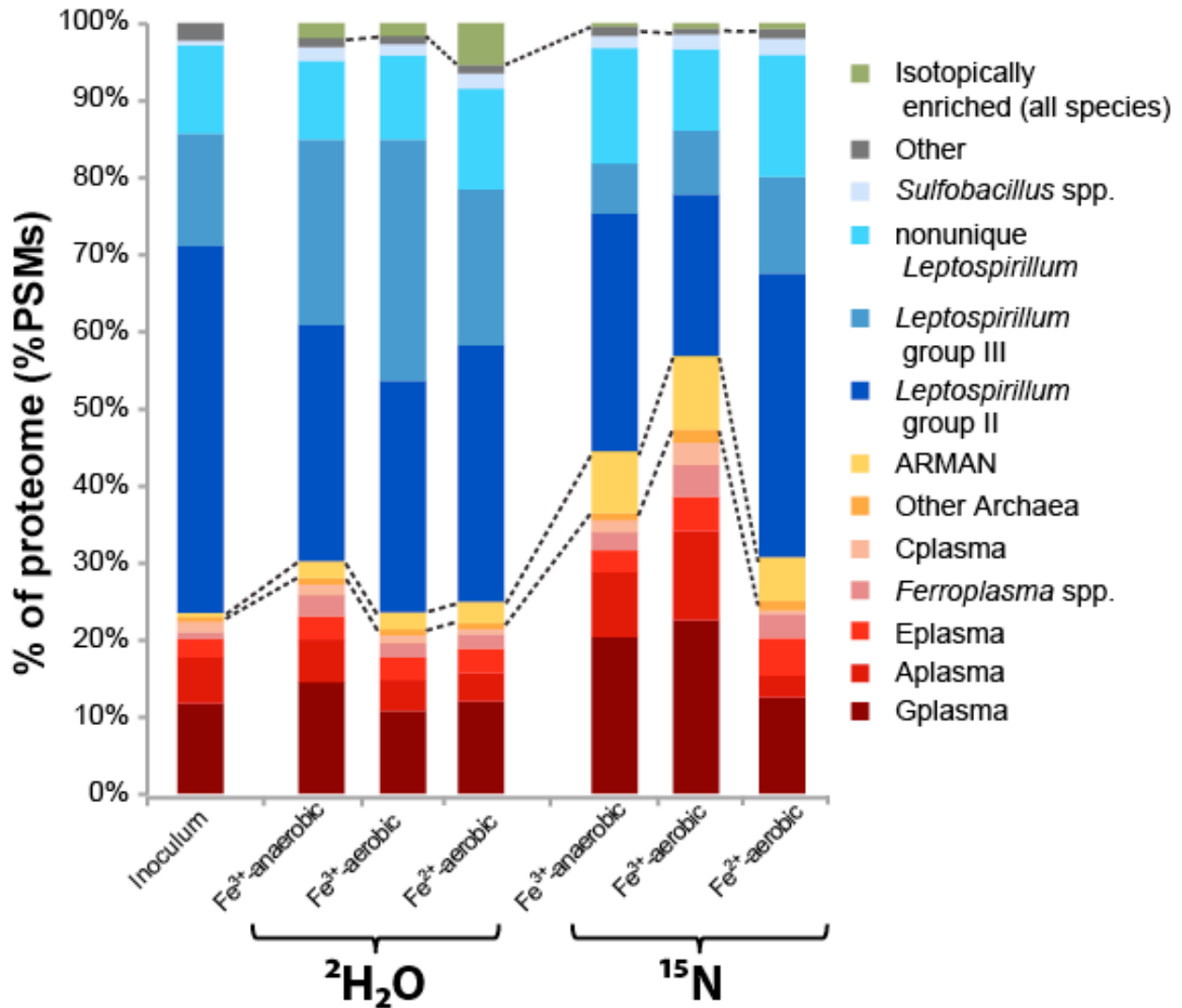


Figure 2.3. Counts of heavy-label enriched proteins detected for different species in $^2\text{H}_2\text{O}$ (top) and ^{15}N samples (bottom). Both replicates are shown within each category to indicate variation. “Nonunique *Leptospirillum*” includes proteins that could not be confidently assigned to a single species or strain *Leptospirillum*. “Other Archaea” includes archaeal PSMs that could not be matched to a single species of Archaea as well as proteins matched to *Diplasma* and *Iplasma* species.

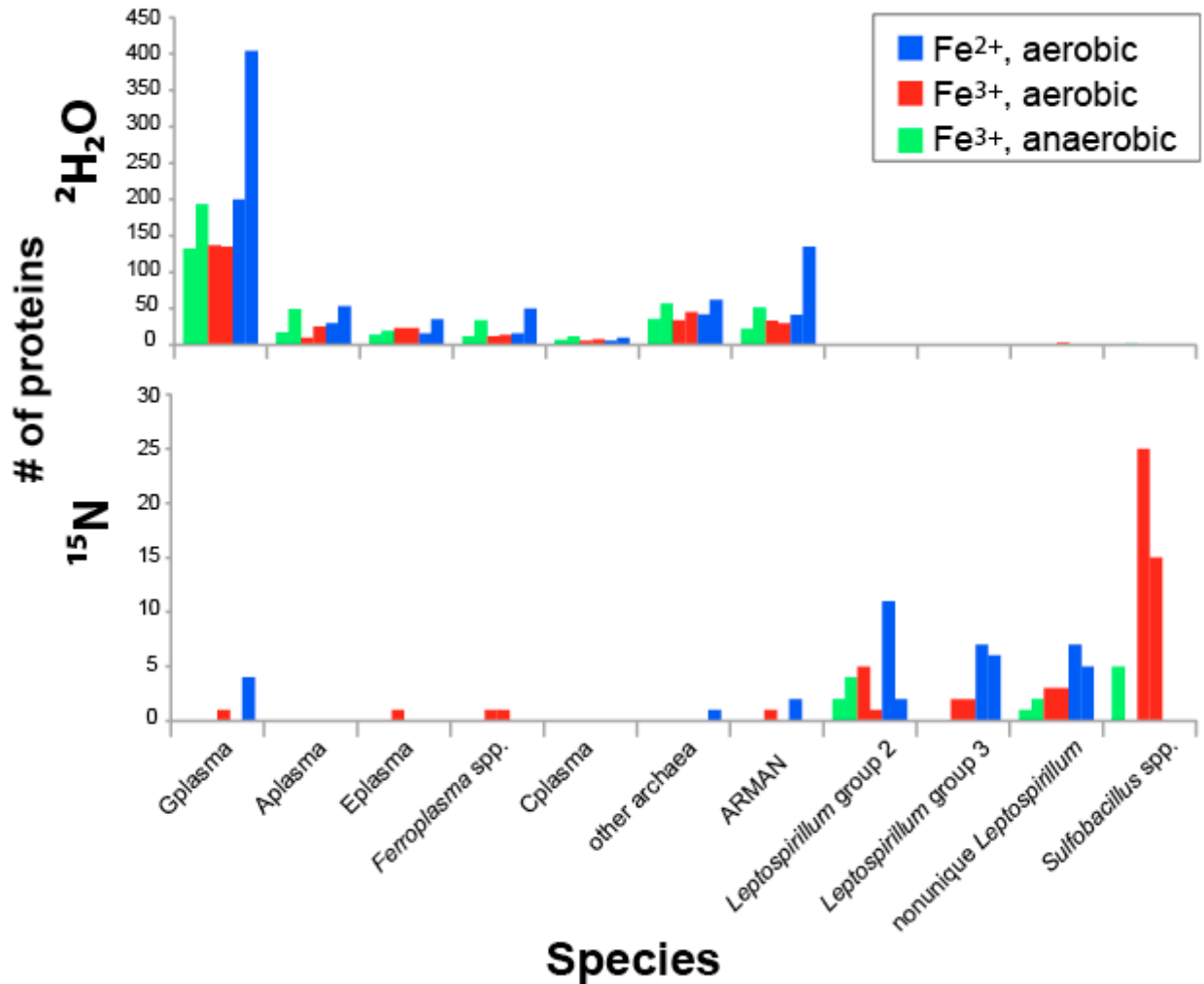


Figure 2.4. Relative abundance of heavy-label enriched protein for different species in $^2\text{H}_2\text{O}$ (top) and ^{15}N samples (bottom) calculated as a percent of all peptide-spectral matches (PSMs). Both replicates are shown within each category to indicate variation. “Nonunique *Leptospirillum*” includes PSMs that could not be confidently assigned to a single species or strain. “Other Archaea” includes archaeal PSMs that could not be matched to a single species as well as PSMs to Dplasma and Iplasma proteins.

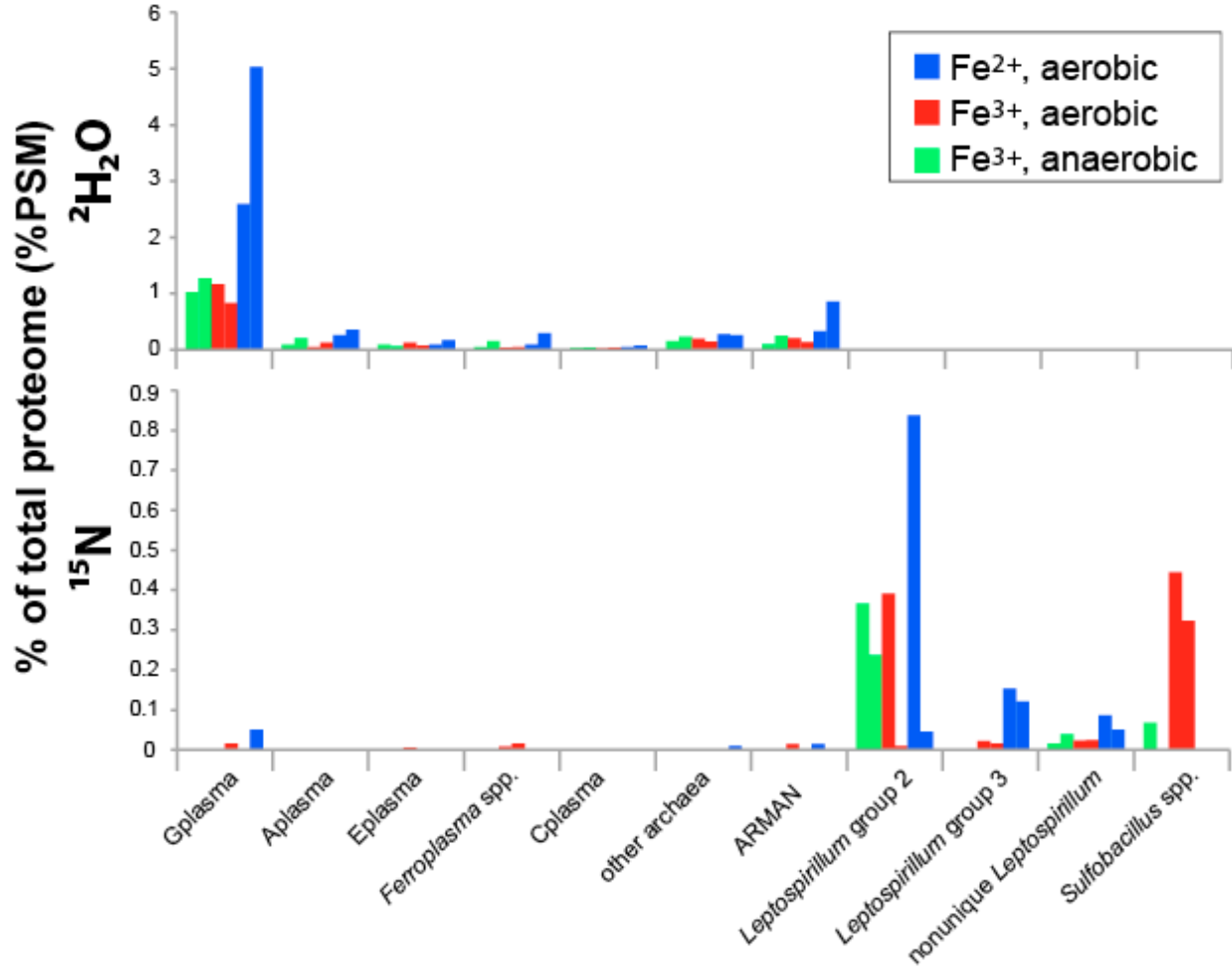


Figure 2.5. Species abundances calculated as a percent of averaged total peptide-spectral matches (PSMs) of the labeled fractions of experimental treatments. The archaeal fraction of the inoculum proteome is shown for comparison. “Other Archaea” includes archaeal PSMs that could not be matched to a single species as well as PSMs to Dplasma and Iplasma proteins.

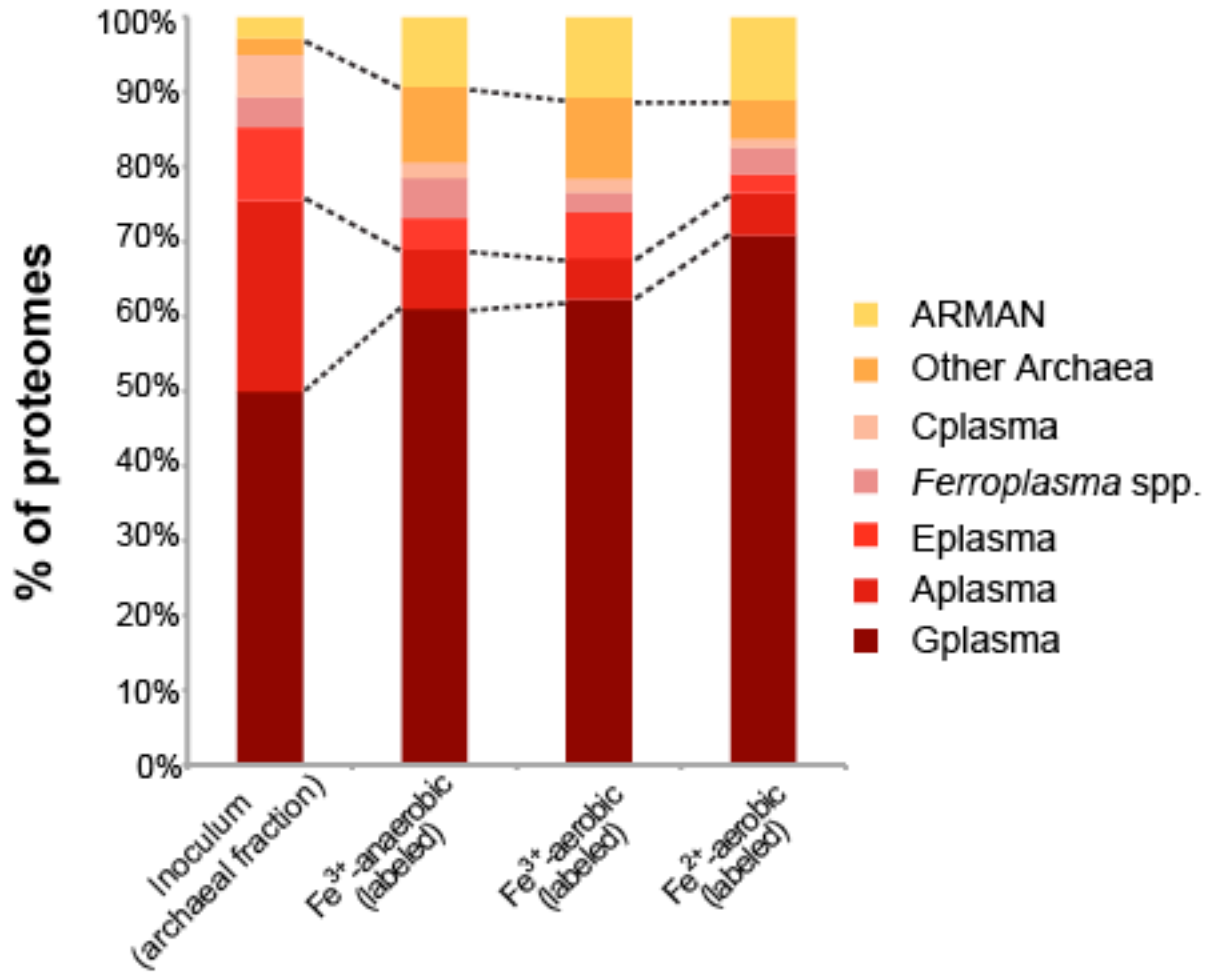


Figure 2.6. Lines trace 1% bins of protein counts (duplicate isotopologues are included). The top graph shows isotopologue distribution for the three treatments in the $^2\text{H}_2\text{O}$ -immersed samples, and the bottom graph shows the three treatments in the ^{15}N -enriched samples. Dashed lines in the bottom graph delineate the groups defined in the text (I, low ^{15}N atom% enrichment group; II, middle ^{15}N atom% enrichment group; III, high ^{15}N atom% enrichment group).

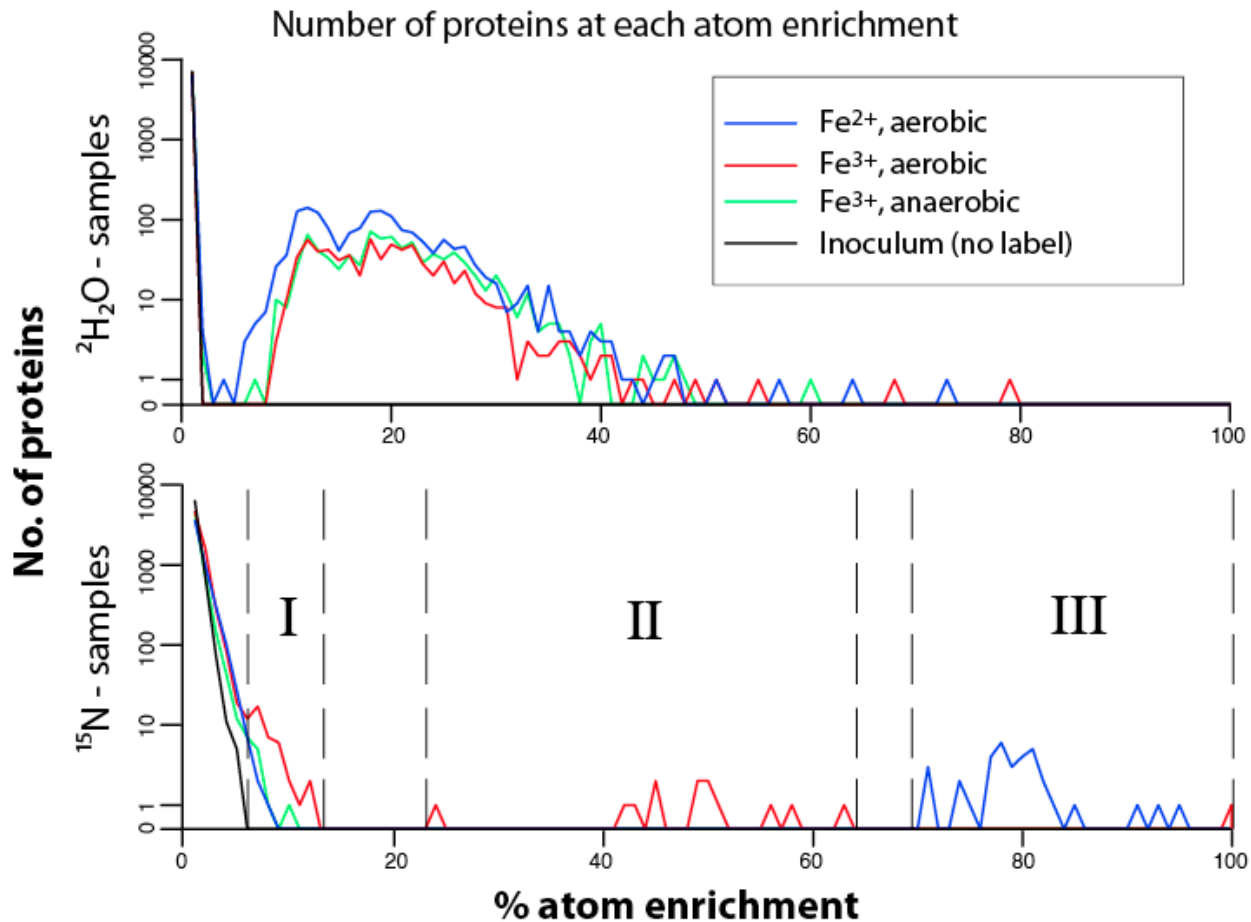


Figure 2.7. Counts of heavy-label enriched proteins detected for different species in different ^{15}N atom% enrichment categories. Both replicates are shown within each category to indicate variation. “Nonunique *Leptospirillum*” includes PSMs that could not be confidently assigned to a single species or strain. “Other Archaea” includes archaeal PSMs that could not be matched to a single species as well as PSMs to *Dplasma* and *Iplasma* proteins. “Other” includes viral, plasmid, and unassigned proteins.

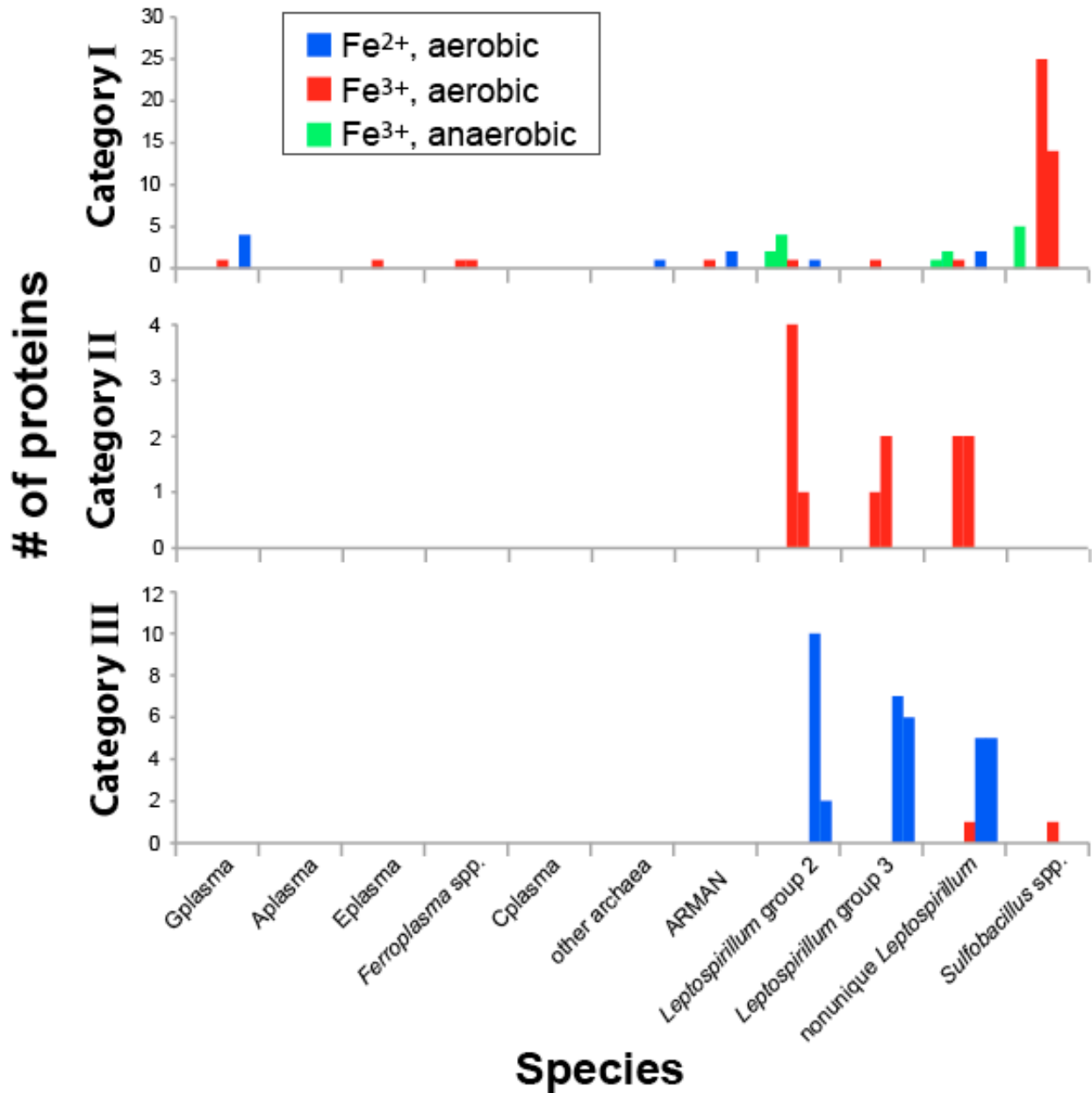


Figure 2.8. Venn diagram showing the number of archaeal proteins uniquely assigned to a single species in the three different treatments. Green, Fe³⁺-anaerobic; Red, Fe³⁺-aerobic; Blue, Fe²⁺-aerobic. Numbers in parentheses indicate the number of proteins found in 100% of replicates represented by the inscribed area of the diagram.

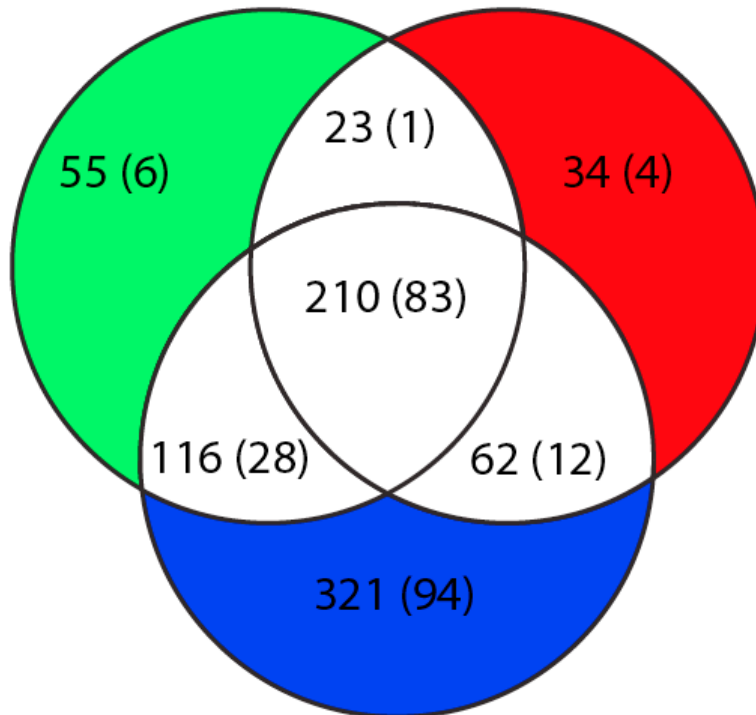


Figure 2.9. Difference in Gplasma COG category proportions between labeled and unlabeled fractions of proteomes $^2\text{H}_2\text{O}$ -treated samples. Statistically significant values are shown (Student's paired t test, $n=6$, $*p<0.05$, $**p<0.01$, $***p<0.001$). COG category labels are as follows: X, No Category Prediction; S, Function unknown; R, General function prediction only; Q, Secondary metabolites biosynthesis transport and catabolism; P, Inorganic ion transport and metabolism; I, Lipid transport and metabolism; H, Coenzyme transport and metabolism; F, Nucleotide transport and metabolism; E, Amino acid transport and metabolism; G, Carbohydrate transport and metabolism; C, Energy production and conversion; O, Posttranslational modification, protein turnover and chaperones; M, Cell wall/membrane/envelope biogenesis; T, Signal transduction mechanisms; L, Replication, recombination, and repair; K, Transcription; J, Translation, ribosomal structure and biogenesis

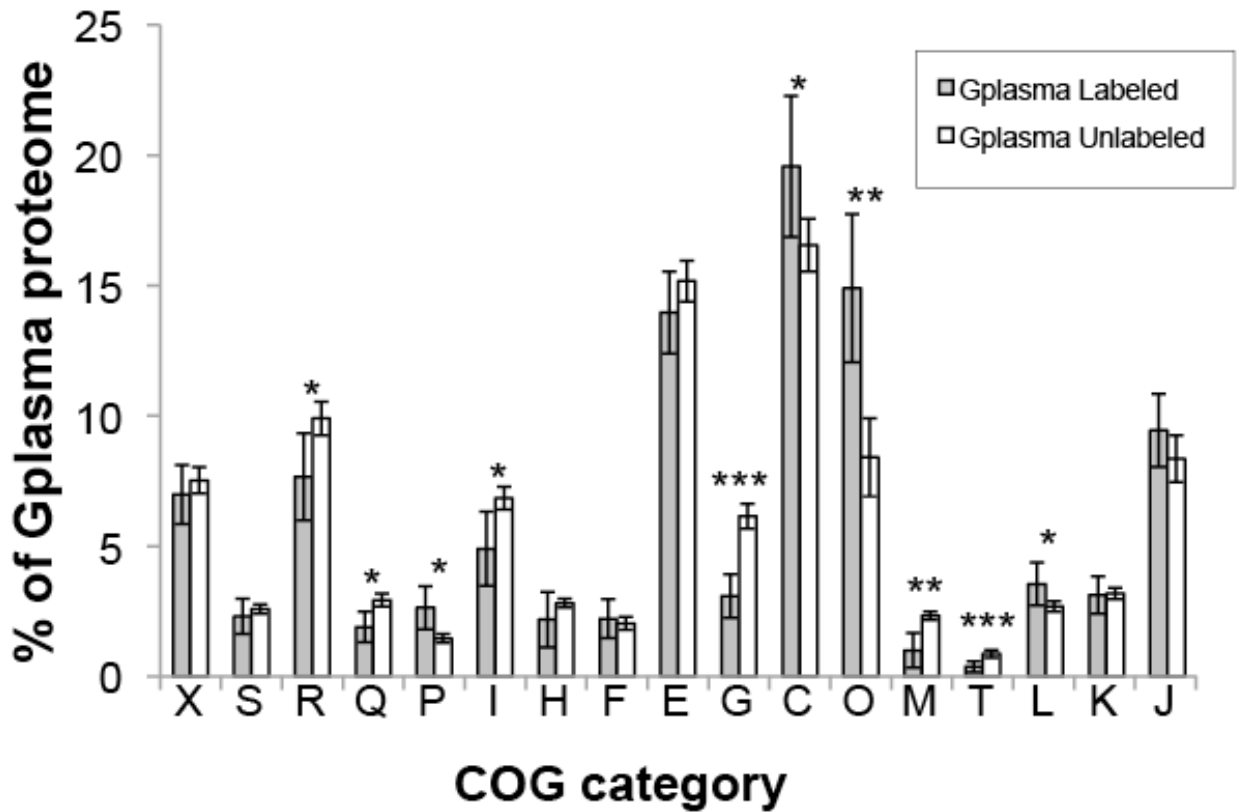


Table 2.1. ¹⁵N-enriched *Leptospirillum* proteins identified in different treatments and the corresponding level of ¹⁵N atom% enrichment (Category I: low; Category II; medium, Category III: high). Colored cells indicate identification of that protein in that Category and conditions, where Fe³⁺-anaerobic experiments are green, Fe³⁺-aerobic experiments are red, and Fe²⁺-aerobic experiments are blue.

organsim	locus	function	Category I			Category II			Category III		
			Fe3+	Fe3+ O ₂	Fe2+ O ₂	Fe3+	Fe3+ O ₂	Fe2+ O ₂	Fe3+	Fe3+ O ₂	Fe2+ O ₂
<i>Leptospirillum</i> group II strain 'C75'	{C75L2_68G0009}	Chaperonin GroEL									
nonunique <i>Leptospirillum</i>	{CGL2_11067G0009,CGL2_11212G0004,UBAL2_8692G0113,UBAL2_8692G0126}	elongation factor Tu (EC:3.6.5.3)]									
<i>Leptospirillum</i> group III	{UBAL3_8063G0025}	Flagellin domain protein									
<i>Leptospirillum</i> group III	{UBAL3_4481G0118}	groEL; chaperonin GroEL; K04077									
<i>Leptospirillum</i> group II strain 'UBA'	{UBAL2_8241G0149}	chaperonin GroEL									
nonunique <i>Leptospirillum</i>	{AILUNK_718986G0005}	Cytochrome 572									
nonunique <i>Leptospirillum</i>	{CGL2_11172G0057,UBAL2_8241G0533}	Putative uncharacterized protein									
<i>Leptospirillum</i> group II strain 'UBA'	{UBAL2_8524G0125}	Putative pyruvate:ferredoxin oxidoreductase gamma subunit									
<i>Leptospirillum</i> group III	{UBAL3_9627G0029}	Isocitrate dehydrogenase									
<i>Leptospirillum</i> group II strain 'UBA'	{UBAL2_8692G0142}	Cytochrome 572									
<i>Leptospirillum</i> group II strain 'UBA'	{UBAL2_8241G0598}	50S ribosomal protein L25									
<i>Leptospirillum</i> group II strain 'UBA'	{UBAL2_8241G0537}	Ribosomal protein S2									
<i>Leptospirillum</i> group II strain 'UBA'	{UBAL2_8135G0074}	Pyruvate:ferredoxin oxidoreductase alpha subunit									
<i>Leptospirillum</i> group II strain 'UBA'	{UBAL2_8135G0062}	Nucleoside diphosphate kinase serA; D-3-phosphoglycerate dehydrogenase (EC:1.1.1.95)									
<i>Leptospirillum</i> group II strain 'UBA'	{UBAL2_7931G0320}	alcohol dehydrogenase zinc-binding domain protein (EC:1.1.1.255)									
nonunique <i>Leptospirillum</i>	{UBAL3_9453G0086,CGL2_11181G0032,UBAL2_8241G0357}	ATP synthase subunit beta (EC:3.6.3.14)									
<i>Leptospirillum</i> group III	{UBAL2_8241G0357,UBAL3_9453G0086}	ATP synthase subunit beta (EC:3.6.3.14)									
nonunique <i>Leptospirillum</i>	{CGL2_11276G0154,UBAL2_8524G0126}	Isocitrate dehydrogenase									
nonunique <i>Leptospirillum</i>	{CGL2_11212G0034,UBAL2_8692G0085}	DNA-directed RNA polymerase subunit alpha (EC:2.7.7.6)									
nonunique <i>Leptospirillum</i>	{AILUNK_701418G0001,UBAL2_8241G0149}	Cytochrome 572									
nonunique <i>Leptospirillum</i>	{AILUNK_701418G0001,CGL2_11277G0153}	Cytochrome 572									
<i>Leptospirillum</i> group III	{UBAL3_9205G0168}	ribose-phosphate pyrophosphokinase (EC:2.7.6.1)									
<i>Leptospirillum</i> group III	{UBAL3_8063G0031}	Flagellin domain protein									
<i>Leptospirillum</i> group III	{UBAL3_8063G0028}	Flagellar hook-associated protein FlgK									
<i>Leptospirillum</i> group III	{UBAL3_8042G0058}	chaperone protein DnaK									
<i>Leptospirillum</i> group III	{UBAL3_7980G0004}	Putative uncharacterized protein									
<i>Leptospirillum</i> group III	{UBAL3_4866G0056}	Isocitrate dehydrogenase									
<i>Leptospirillum</i> group II strain 'Sway'	{CGL2_11077G0032}	Pyridine nucleotide-disulphide oxidoreductase									
<i>Leptospirillum</i> group II strain 'UBA'	{UBAL2_8062G0372}	Cytochrome 579									
<i>Leptospirillum</i> group III	{UBAL2_8524G0108,UBAL3_8042G0058}	Chaperone protein dnaK									
nonunique <i>Leptospirillum</i>	{CGL2_11233G0063,UBAL2_8135G0062}	D-3-phosphoglycerate dehydrogenase									
nonunique <i>Leptospirillum</i>	{CGL2_11111G0068,UBAL2_8241G0598}	Ribosomal protein S2									
nonunique <i>Leptospirillum</i>	{CGL2_11077G0003,CGL2_11212G0142,UBAL2_8049G0003,UBAL2_8135G0081}	Probable cytochrome C oxidase monoheme subunit									
<i>Leptospirillum</i> group II strain 'Sway'	{CGL2_11276G0172}	Chaperone protein DnaK									
<i>Leptospirillum</i> group II strain 'UBA'	{UBAL2_8241G0556}	Probable O-methyltransferase family protein									
<i>Leptospirillum</i> group II strain 'UBA'	{UBAL2_8062G0138}	Flavin oxidoreductase/NADH oxidase									
nonunique <i>Leptospirillum</i>	{CGL2_11238G0089,UBAL2_7931G0102}	Putative uncharacterized protein bifunctional									
<i>Leptospirillum</i> group II strain 'Sway'	{CGL2_11238G0069}	phosphoglucomutase/phosphomannomutase (EC:5.4.2.8 5.4.2.2)									
<i>Leptospirillum</i> group II strain 'Sway'	{CGL2_11233G0057}	Cold shock protein									
nonunique <i>Leptospirillum</i>	{CGL2_11276G0173,UBAL2_8524G0107}	Protein grpE									
<i>Leptospirillum</i> group II strain 'Sway'	{CGL2_11277G0176}	Biopolymer transport, ExbD									
nonunique <i>Leptospirillum</i>	{CGL2_11346G0059,UBAL2_8062G0235}	Undecaprenyl diphosphate synthase									
<i>Leptospirillum</i> group III	{UBAL3_9568G0092}	Elongation factor P									

Table 2.2. ¹⁵N-enriched *Sulfobacillus* proteins identified in different treatments and the corresponding level of ¹⁵N atom% enrichment (Category I: low; Category II; medium, Category III: high). Colored cells indicate identification of that protein in that Category and conditions, where Fe³⁺-anaerobic experiments are green, Fe³⁺-aerobic experiments are red, and Fe²⁺-aerobic experiments are blue.

organism	locus	function	Category I			Category II			Category III		
			Fe3+	Fe3+ O ₂	Fe2+ O ₂	Fe3+ O ₂	Fe3+ O ₂	Fe2+ O ₂	Fe3+ O ₂	Fe3+ O ₂	Fe2+ O ₂
<i>Sulfobacillus</i>	{AMDSBA4_C54G00005}	bifunctional transaldolase/phosphoglucose isomerase (EC:2.2.1.2, 5.3.1.9)	■	■							
<i>Sulfobacillus</i>	{AMDSBA4_C20G00053}	citrate synthase (EC:2.3.3.1)	■	■							
<i>Sulfobacillus</i>	{AMDSBA4_C5G000051}	succinate--CoA ligase (ADP-forming) (EC:6.2.1.5)	■	■							
<i>Sulfobacillus</i>	{AMDSBA4_C5G000049}	L-lactate dehydrogenase (EC:1.1.1.27)		■							
<i>Sulfobacillus</i>	{AMDSBA4_C47G00001}	allantoinase (EC:3.5.2.5)		■							
<i>Sulfobacillus</i>	{AMDSBA4_C21G00048}	hypothetical protein		■							
<i>Sulfobacillus</i>	{AMDSBA4_C19G00037}	diaminobutyrate-2-oxoglutarate transaminase (EC:2.6.1.76)		■							
<i>Sulfobacillus</i>	{AMDSBA4_C16G00078}	2-oxoglutarate ferredoxin oxidoreductase subunit alpha [EC:1.2.7.3]		■							
<i>Sulfobacillus</i>	{AILUNK_363018G0002}	Glyceraldehyde-3-phosphate dehydrogenase		■							
nonunique <i>Sulfobacillus</i>	{AILUNK_658428G0021,AMDSBA4_C11G00034}	Betaine-aldehyde dehydrogenase		■							
nonunique <i>Sulfobacillus</i>	{AILUNK_360667G0023,AMDSBA4_C50G00029}	30S ribosomal protein S5		■							
nonunique <i>Sulfobacillus</i>	{AILUNK_335206G0002,AMDSBA4_C29G00059}	Carboxymethylenebutenoldase		■							
<i>Sulfobacillus</i>	{AMDSBAU_1858_1}	propionyl-CoA carboxylase (EC:6.4.1.3)		■							
<i>Sulfobacillus</i>	{AMDSBA4_C7G00073}	ATP-dependent Clp protease, protease subunit (EC:3.4.21.92)		■							
<i>Sulfobacillus</i>	{AMDSBA4_C6G00095}	tryptophan synthase alpha chain (EC:4.2.1.20)		■							
<i>Sulfobacillus</i>	{AMDSBA4_C61G00013}	class II fumarate hydratase (EC:4.2.1.2)		■							
<i>Sulfobacillus</i>	{AMDSBA4_C5G00052}	succinyl-CoA synthetase subunit alpha (EC:6.2.1.5)		■							
<i>Sulfobacillus</i>	{AMDSBA4_C56G00009}	aconitate hydratase 1 (EC:4.2.1.3)		■							
<i>Sulfobacillus</i>	{AMDSBA4_C4G00007}	3-hydroxyisobutyryl-CoA hydrolase (EC:4.2.1.17)		■							
<i>Sulfobacillus</i>	{AMDSBA4_C47G00009}	carboxysome-like protein		■							
<i>Sulfobacillus</i>	{AMDSBA4_C3G00125}	Dihydrolipoamide dehydrogenase (EC:1.8.1.4)		■							
<i>Sulfobacillus</i>	{AMDSBA4_C3G00121}	pyruvate dehydrogenase E2 component (dihydrolipoamide acetyltransferase) (EC:2.3.1.12)		■							
<i>Sulfobacillus</i>	{AMDSBA4_C32G00108}	hypothetical protein		■							
<i>Sulfobacillus</i>	{AMDSBA4_C29G00007}	phage shock protein PspA		■							
<i>Sulfobacillus</i>	{AMDSBA4_C28G00082}	acetyl-CoA C-acyltransferase (EC:2.3.1.16)		■							
<i>Sulfobacillus</i>	{AMDSBA4_C28G00040}	rubrerythrin		■							
<i>Sulfobacillus</i>	{AMDSBA4_C1G00105}	glutamine synthetase (EC:6.3.1.2)		■							
<i>Sulfobacillus</i>	{AMDSBA4_C11G00035}	short-chain dehydrogenase/reductase SDR		■							
<i>Sulfobacillus</i>	{AMDSBA4_C11G00034}	betaine-aldehyde dehydrogenase (EC:1.2.1.8)		■							
<i>Sulfobacillus</i>	{AMDSBA4_C0G00028}	translation elongation factor Ts (EF-Ts)		■							
<i>Sulfobacillus</i>	{AMDSBA1_C17G00019}	malate dehydrogenase (EC:1.1.1.27)		■							
<i>Sulfobacillus</i>	{AILUNK_391805G0017}	Peroxioredoxin		■							
<i>Sulfobacillus</i>	{AILUNK_172907G0013}	ATP synthase subunit beta	■								
<i>Sulfobacillus</i>	{AMDSBA1_C15G00005}	glycerol kinase (EC:2.7.1.30)								■	

Table 2.3. ¹⁵N-enriched Archaeal proteins identified in different treatments and the corresponding level of ¹⁵N atom% enrichment (Category I: low; Category II; medium, Category III: high). Colored cells indicate identification of that protein in that Category and conditions, where Fe³⁺-anaerobic experiments are green, Fe³⁺-aerobic experiments are red, and Fe²⁺-aerobic experiments are blue.

organism	locus	function	Category I			Category II			Category III		
			Fe3+	Fe3+ O ₂	Fe2+ O ₂	Fe3+	Fe3+ O ₂	Fe2+ O ₂	Fe3+	Fe3+ O ₂	Fe2+ O ₂
Gplasma	{UNLGPL_13334G0168}	transmembrane_regions									
<i>Ferroplasma</i>	{CGFER2_24G0050}	glutamine synthetase (EC:6.3.1.2)									
<i>Ferroplasma</i>	{CGFER2_11G0006}	flavin mononucleotide-reductase									
Eplasma	{UNLEPL_15243G0439}	histidine ammonia-lyase (EC:4.3.1.3)									
ARMAN	{UNLAR4_1G0003}	Putative uncharacterized protein									
nonunique archaeal	{AILUNK_203221G0012}	Perosamine synthetase									
Gplasma	{UNLGPL_13477G0077}	aldo/keto reductase									
Gplasma	{UNLGPL_13477G0061}	prefoldin subunit beta; K04798									
Gplasma	{UNLGPL_13334G0103}	prefoldin beta subunit									
Gplasma	{UNLGPL_13290G0186}	30S ribosomal protein S19e									
ARMAN	{UNLAR5_1G0815}	hypothetical protein									
ARMAN	{UNLAR5_1G0815}	Superoxide dismutase									
ARMAN	{UNLAR4_1G0064}	Superoxide dismutase									

Table 2.4. Non-Archaeal ²H-enriched proteins identified in different treatments and the corresponding level of enrichment shown as black text in colored cells. Colored cells indicate identification of that protein in specific conditions, where Fe³⁺-anaerobic experiments are green, Fe³⁺-aerobic experiments are red, and Fe²⁺-aerobic experiments are blue.

organism	protein	function	Fe3+	Fe3+ O ₂	Fe2+ O ₂
<i>Sulfobacillus</i>	{AMDSBA1_C74G00002}	peptidase S53 propeptide	1		
<i>Leptospirillum</i> group II strain 'UBA'	{UBAL2_8049G0114}	Probable mechanosensitive ion channel	12		
<i>Sulfobacillus</i>	{AMDSBA1_C19G00019,AMDSBA3_C37G00005,AMDSBA4_C28G00006,AILUNK_456098G0037,AMDSBA5_C6G00037}	chaperonin GroEL	30		
nonunique <i>Leptospirillum</i>	{CGL2_11364G0015,UBAL2_8062G0357}	Putative histone-like DNA-binding protein		68	
nonunique <i>Leptospirillum</i>	{AILUNK_701418G0001,CGL2_11277G0153,UBAL2_8241G0149,UBAL3_9627G0029}	Cytochrome 572		79	
nonunique <i>Leptospirillum</i>	{C75L2_68G0009,UBAL3_4481G0118}	Chaperonin GroEL		25.0	
<i>Sulfobacillus</i>	{AMDSBA3_C64G00003}	ATP-dependent Clp protease ATP-binding subunit ClpB		55.0	10
Unassigned species	{AILUNK_379629G0001}	hypothetical protein			21
<i>Leptospirillum</i> group II strain '5way'	{CGL2_11277G0153}	Putative uncharacterized protein			73

Chapter 3

Comparison of Environmental and Isolate *Sulfobacillus* Genomes Reveals Diverse Carbon, Sulfur, Nitrogen, and Hydrogen Metabolisms

Abstract

Sulfobacillus organisms are found worldwide as members of microbial communities that accelerate sulfide mineral dissolution in acid mine drainage environments (AMD), acid-rock drainage environments (ARD), as well as in industrial bioleaching operations. Despite their frequent identification in these environments, their role in biogeochemical cycling is poorly understood. Here we report draft genomes of five species of the *Sulfobacillus* genus (AMDS1-5) reconstructed by cultivation-independent sequencing of biofilms sampled from the Richmond Mine, Iron Mountain Mine, CA. Three of these species (AMDSBA1, AMDSBA3, and AMDSBA2) have no cultured representatives. AMDSBA4 is a strain of *S. benefaciens*, and AMDSBA5 a strain of *S. thermosulfidoxidans*. We analyzed the diversity of energy conservation and central carbon metabolisms for these genomes and previously published *Sulfobacillus* genomes. Carbon monoxide dehydrogenase proteins, including several DMSO-family proteins of unknown function, were identified in all genomes. Enzymes with putative sulfur and sulfur-compound reduction were also identified in all of the genomes. However, pathways of sulfur oxidation vary considerably across the genus, including the number and type of subunits of putative heterodisulfide reductase complexes likely involved in sulfur oxidation. Central carbon degradation pathways in *Sulfobacillus* lineages vary, with *S. thermosulfidoxidans* likely favoring the pentose phosphate pathway. The semi-phosphorylative Entner Doudoroff pathway is present in lineages of *S. acidophilus*, AMDSBA1, AMDSAB3, AMDSAB4, and AMDSBA2. Only the AMDSBA3 genome encodes a dissimilatory nitrate reductase and only the AMDSBA5 and *S. thermosulfidoxidans* genomes encode assimilatory nitrate reductases. Among the AMDS genomes, only AMDSAB5 encodes nickel-iron hydrogenase proteins. Within the genus, AMDSBA4 (*S. benefaciens*) is unusual in that its electron transport chain includes a *bc* complex, a unique cytochrome *c* oxidase, and two distinct succinate dehydrogenase complexes. It is also the only *Sulfobacillus* species with putative carboxysome proteins. Overall, the results significantly expand our understanding of carbon, sulfur, nitrogen, and hydrogen metabolism within the *Sulfobacillus* genus.

Introduction

Species of the *Sulfobacillus* genus are Gram-positive spore forming bacteria belonging to the order Clostridiales, and are found globally in acidic environments such as thermal springs (Norris et al. 1996), hydrothermal vents (Bo Li et al. 2011), solfatara fields (Johnson et al. 2003), and acid mine drainage (AMD) environments (Baker & Banfield 2003). *Sulfobacillus* species are also frequently found in industrial reactors used in bioleaching operations (Watling et al. 2008). Thus, an understanding of their physiology is of both environmental and biotechnological importance.

In the AMD environment of the Richmond Mine (Iron Mountain, CA) *Sulfobacillus* are members of acidophilic, metal-tolerant microbial consortia that promote sulfide-mineral dissolution. In contrast to the dominant iron-oxidizing *Leptospirillum* bacteria, or certain

members of the heterotrophic *Thermoplasmatales* Archaea, *Sulfobacillus* generally appear in much lower relative abundances, having been identified only through 16S rRNA clone libraries, fluorescent in situ hybridization (Bond, Druschel, et al. 2000a; Justice et al. 2012; Druschel et al. 2004), and as unresolved fragments assembled from metagenomic sequence information (Dick et al. 2009). As their namesake implies, they are likely key players in sulfur cycling, yet relatively little is known about their metabolic pathways or how these vary across the *Sulfobacillus* genus.

Key features of *Sulfobacillus* metabolism have been described for several isolated *Sulfobacillus* species—*S. sibiricus* (Melamud et al. 2003), *S. thermotolerans* (Bogdanova et al. 2006), *S. acidophilus* (Norris et al. 1996), *S. thermosulfidoxidans* (Golovacheva & Karavaiko 1978), and *S. benefaciens* (Johnson et al. 2008). Each is a facultative anaerobe capable of assimilation of organic and inorganic forms of carbon, deriving energy from aerobic oxidation of iron and various sulfur species (e.g., tetrathionate and elemental sulfur), as well as from ferric iron respiration (Bridge & Johnson 1998; Tsaplina et al. 2010; Bogdanova et al. 2006; Melamud et al. 2003), and possibly fermentation (Tsaplina et al. 2010). Other phenotypic differences such as metal tolerance, temperature tolerance, and use of different carbon compounds set these organisms apart (Johnson et al. 2008; Bogdanova et al. 2006; Watling et al. 2008). Genomes for two of these species—*S. acidophilus* and *S. thermosulfidoxidans*—have been previously reported (Travisany et al. 2012; Anderson et al. 2012; Bo Li et al. 2011) and an additional *S. thermosulfidoxidans* genome is deposited in the publicly available IMG database. However, detailed metabolic predictions based on these genomes have not been reported. Here we present five new draft genomes of *Sulfobacillus* organisms assembled from metagenomic data. We analyze these genomes in the context of the previously published genomes of *Sulfobacillus* isolates, and describe key elements of carbon, sulfur, nitrogen, and hydrogen metabolism. In addition to illuminating the metabolic potential of a strain of *S. benefaciens*, and three other novel *Sulfobacillus* spp., this work provides a first reconstruction of Gram-positive, acidophilic sulfur oxidation pathways, and highlights both conserved and divergent metabolisms found in organisms of the *Sulfobacillus* genus.

Methods

DNA was collected from the 5way site at Iron Mountain Mine (Denef, Mueller, et al. 2010b) and extracted as previously described (Miller et al. 2011). This sample is referred to as ‘5way fungal streamer.’ Briefly, 4-5 ml of frozen biofilm was washed and thawed with cold acidified 0.9% NaCl (pH 1.0, H₂SO₄), homogenized with a 16 G needle in cold pH 7.0 phosphate buffered saline, and finally ground in a mortar and pestle in liquid nitrogen. The powdered biofilm pellets were then extracted with a PowerSoil DNA isolation kit (MoBio Laboratories, Carlsbad, CA, USA). Illumina Library preparation sequencing was carried out according to JGI protocols (Miller et al. 2011; JGI).

Three flow cells of Illumina HiSeq were used to obtain ~90 paired-end 76 bp reads (8.46 Gb). The raw reads were deposited in the NCBI Sequence Read Archive under accession number SRA:SRR191843. An additional overlapping paired-end library was sequenced to generate longer sequence reads. The initial paired-end reads were merged using SeqPrep (St John) into ~32 million long reads (6.17 Gb) with a median length of 194 bp. These reads were used solely in the assembly of the low abundance AMDSBA2 as described below. While the vast majority of sequencing information for the *Sulfobacillus* genomes came from the 5way fungal streamer

samples described above, reads from 17 additional previously published metagenomic datasets were used to assist binning and assembly (Denef & Banfield 2012).

Metagenomic sequence assembly

A total of 854 million paired-end (PE) Illumina reads (98 Gb) pooled from the 18 samples described above were assembled *de novo* using the Iterative De-Bruijn Assembler optimized for single-cell and metagenomic assemblies (idba-ud)(Y Peng et al. 2012). A kmer-range of 19-99 was used with the pre-correction option enabled. All PE-reads were quality-trimmed using Sickle (sickle pe options -q 15 and -l 40) prior to assembly. PE-reads were mapped back to the resulting scaffolds using bwa (H Li & Durbin 2009). All scaffolds larger than 1500 bp were binned using emergent self-organizing maps (ESOM) trained on tetranucleotide frequencies as previously described (Dick et al. 2009). Briefly, larger sequences were subdivided into 10 kb fragments, and trained for 100 epochs using the k-batch training method, along with any leftover fragments exceeding 3 kb in length. To minimize noise, sequences in the range 1500 - 3000 bp were not included during training, but instead projected onto the trained weight vectors generated with the larger fragments. In order to separate individual *Sulfobacillus* bins, a discrete bin containing putative members of the order Clostridiales, was isolated and further binned using log-normalized sample abundance patterns from 18 different samples, this time including fragments down to 1500 bp. Five distinct sub-bins (AMDSBA1, AMDSBA3, AMDSBA4, AMDSBA5 and AMDSBA2) were individually re-assembled *de novo* using every read-pair that mapped within these respective bins. Additionally, because of its high similarity to *S. thermosulfidooxidans* DSM9293, all PE-reads mapping to this organism were included in the AMDSBA5 sub-assembly as well. Due to its considerably lower abundance, AMDSBA2 was sub-assembled using PE-reads as described above, but with the inclusion of additional reads from the overlapping HiSeq PE library that mapped within AMDSBA2 bin (media read length 194 bp). Scaffolds belonging to the Clostridiales bin, but not to the 5 putative genome bins were not re-assembled.

Metabolic analysis

Open reading frames on assembled contigs for Iron Mountain *Sulfobacillus* and *S. thermosulfidooxidans* strain ‘Cutipay’ were predicted using Prodigal (Hyatt et al. 2010), and tRNAs predicted with tRNAscan (Lowe & Eddy 1997). Proteins were then compared with BLAST (Altschul et al. 1990) to UniRef90 (Suzek et al. 2007) and KEGG (Ogata et al. 1999) with matches greater than 60 bits being reported. Reverse BLASTs were also used to identify reciprocal best BLAST hits. Proteins were further analyzed with InterProScan (Quevillon et al. 2005) to identify conserved domains. Protein sequences and corresponding annotations for *S. thermosulfidooxidans* AT-1 (DSM9293), *S. acidophilus* TPY, and *S. acidophilus* NAL (DSM10332) were downloaded from the Integrated Microbial Genomes database available at <http://img.jgi.doe.gov/> (Markowitz et al. 2011). While *S. thermosulfidooxidans* strain ‘Cutipay’ and *S. acidophilus* TPY were included in most analyses here, results focus on the high quality genomes of the type strains *S. thermosulfidooxidans* AT-I (DSM9293) and *S. acidophilus* NAL (DSM10332). All references to *S. thermosulfidooxidans* and *S. acidophilus* in the text refer specifically to these type strains.

Orthologs shared between species were identified with amino acid sequence searches between all pair-wise combinations of genomes using USEARCH (Edgar 2010). Reciprocal best hits between each pair of genomes were considered orthologs if the alignments of their

sequences had an E-value less than 0.01, a Bit score greater than 40, and covered at least 65% of each amino acid sequence.

Protein phylogenetic analysis

Protein sequences were aligned using Muscle (Edgar 2004), and trimmed using Gblocks (Castresana 2000; Talavera & Castresana 2007), which stringently curates the alignment to phylogenetically informative sites. For each alignment, the optimum amino acid substitution model was estimated using ProtTest (Abascal et al. 2005). All trees were generated using RAxML (Stamatakis 2006) using the PROTCAT rate model and the ProtTest-determined amino acid substitution model. Support was evaluated using 500 bootstrap replications in RAxML. Trees were visualized in iTOL (Letunic & Bork 2007).

The concatenated ribosomal protein tree was made using 16 core ribosomal proteins (rpL2, 3, 4, 5, 6, 14, 15, 16, 18, 22, 24, rpS3, 8, 10, 17, 19), selected based on low frequencies of lateral gene transfer (Sorek et al. 2007; Wu & Eisen 2008). These proteins were aligned with Muscle (Edgar 2004), manually trimmed, then concatenated, resulting in an 2,052 residue alignment. Phylogenetic relationships were analyzed using RAxML using the LG + α + γ substitution model, and nodal support determined with 500 bootstrap replicates. Trees were visualized in FigTree (available: <http://tree.bio.ed.ac.uk/software/figtree/>)

16S rRNA gene sequence analysis

Near full-length 16S rRNA sequences for *Sulfobacillus* species were generated using EMIRGE—an iterative template-guided assembler that probabilistically generates 16S rRNA gene sequences using a 16S rRNA database (Miller et al. 2011). First, potential 16S rRNA gene regions were found by a BLAST of all assembled contigs against SILVA db v. 108 (Pruesse et al. 2007). Reads that mapped to these regions were extracted and trimmed with Sickle (available <https://github.com/najoshi/sickle>), allowing only paired-end reads with length >60 and quality scores >20. For the reference database, 186 sequences were downloaded from the SILVA SSU database representing the 174 sequences of ‘Family XVII Incertae Sedis’ (the family to which the *Sulfobacillus* genus belong) as well as twelve representative sequences of other organisms known to be represent the majority of the Richmond Mine microbial communities, including *Leptospirillum* species and Archaea of the ARMAN and *Thermoplasmatales* lineages. Potential chimeric sequences in this database were removed with DECIPHER (Wright et al. 2012) and UCHIME (Edgar et al. 2011) [searched against the 2011 Greengenes database (DeSantis et al. 2006)]. EMIRGE was run in 50 iterations with a “join_threshold” parameter of 0.99.

Using the QIIME software suite (Caporaso, Kuczynski, et al. 2010b), the EMIRGE-generated sequences and representative sequences of *Sulfobacillus* isolates were aligned using PyNAST with default parameters (Caporaso, Bittinger, et al. 2010a), and filtered with *filter_alignment.py* in QIIME. Edges were trimmed leaving 1,154 unambiguously aligned positions. A maximum likelihood tree generated with RAxML using the GTRCAT model. Nodal support was evaluated with 500 bootstrap replicates. Trees were visualized in FigTree.

Microscopy

Environmental samples were fixed and stained with lineage-specific fluorescent *in situ* hybridization (FISH) probes to examine *Sulfobacillus* populations, as described previously (Bond & Banfield 2001). *Sulfobacillus* (SUL230; 5'- TAATGGGCCCGCGRGCYCC) and

archaeal-specific probes (ARC915; 5'-GTGCTCCCCCGCCAATTCCT) have been previously reported (Justice et al. 2012; Amann et al. 1995)

Results

Overall Genome Statistics and Phylogeny

We reconstructed the near complete genomes of five *Sulfobacillus* species from the Richmond Mine, represented as five discrete bins on abundance-pattern ESOMs (Figure 3.1). Each genome had between 129 and 409 scaffolds with coverage ranging from 10 to 258X (Table 3.1). Genomes were between 3.07 and 4.56 Mbp, and encoded between 3,312 and 4,629 protein-coding genes. Figure 3.2 summarizes the completeness of each genome estimated using the presence of a set of conserved, single copy genes with low frequency of horizontal gene transfer (Wu & Eisen 2008; Sorek et al. 2007).

Based on phylogenetic trees of concatenated protein alignments and EMIRGE-generated 16S rRNA genes, AMDSBA5 was classified as a strain of *S. thermosulfidooxidans* (99.7% 16S rRNA gene similarity), AMDSBA1 was classified as a strain of *S. benefaciens* (100% 16S rRNA gene similarity), and AMDSBA3 was shown to related to be most closely related to *S. acidophilus* strains (97.2% 16S rRNA gene similarity, Figure 3.3A and B, Table 3.2). AMDSBA2 may represent another strain of *S. thermosulfidooxidans* with 99.2% 16S similarity, while AMDSBA4 is at least 97% divergent from any *Sulfobacillus* isolates.

Sulfobacillus Abundance and Distribution

Previous data have indicated the presence of *Sulfobacillus* species in low abundance in late growth stage and submerged biofilms (Justice et al. 2012), as well as in warmer (~43 °C) areas of the Richmond Mine (Bond, Druschel, et al. 2000a) 6. They are sometimes found to form compact structures of rope-like assemblages (Figure 3.4), and similar filamentous-like growth as been reported elsewhere (Norris et al. 1996; Tsaplina et al. 2010). Metagenomic sequencing data indicated that species AMDSBA1 and AMDSBA4 most abundant late-growth stage biofilms (Figure 3.5B). The higher abundances in the 5way fungal streamer (Figure 3.5A) may reflect the unique DNA-extraction procedure (grinding in liquid nitrogen) used on this sample alone, potentially improving lysis of Gram-positives and their vegetative spores.

Energy Metabolism: electron transport chain

Prior studies have shown that *Sulfobacillus* are capable of aerobic growth. Genomes indicate that oxygen reduction can occur via several different terminal oxidases. Each of the *Sulfobacillus* genomes analyzed in this study has genes encoding the four-subunit proton pumping quinol oxidases, although only a fragment of subunit II of the quinol oxidases from AMDSBA4 and AMDSBA2 were obtained. AMDSBA3 and AMDSBA1 contained an additional quinol oxidase (Figure 3.6), and this second copy may have a different affinity for oxygen. Except for AMDSBA3, all of the *Sulfobacillus* genomes also contain cytochrome *c* oxidase encoding-genes. The number of cytochrome *c* oxidase-encoding genes varies between species with *S. acidophilus* containing 2 copies, with *S. thermosulfidooxidans* strains containing at least 4 (Figure 3.6). Furthermore, a unique terminal oxidase from AMDSBA4 (AMDSBA4_44_9-12) was orthologous to the quinol oxidases of the other *Sulfobacillus* (and contained four subunits), however it contained a Cu_A-binding site in subunit II, indicating it accepts electrons from cytochrome *c* and not quinones, (Saraste 1994; Beinert 1997). Additionally, cytochrome *c*-

binding sites (CXXCH) were identified in the C-terminal end of both subunits II and IV which are not present in any other terminal oxidases here, suggesting the electron transport chain of AMDSBA4 is distinct from other *Sulfobacillus* species.

AMDSBA4 is also unique in that it contains a cytochrome *bc* complex. Other *Sulfobacillus* have a di-heme cytochrome *b* homologous to the *petB* gene found in other cytochrome *bc* complexes, however the absence of cytochrome *c* and the iron sulfur protein in these operons suggests that these are not canonical *bc* complexes and may perform some other function.

c-type cytochromes are annotated in all genomes except AMDSBA1, although the presence of cytochrome *c* oxidases and cytochrome *c* assembly factors would indicate that *c*-type cytochromes are likely present.

All of the *Sulfobacillus* genomes contain multiple genes encoding the two-subunit *bd*-type quinol oxidases. The cytochrome *bd* respiratory oxygen reductases contribute to the proton motive force by transmembrane charge separation without a proton pump (Borisov et al. 2011).

Each genome contains genes encoding a classic 14-subunit NADH:quinone oxidoreductase (Complex 1). In addition, an 11-subunit of Complex 1 is conserved in AMDSBA1, AMDSBA5, AMDSBA2, and *S. acidophilus*. This 11 subunit gene cluster lacks genes encoding the “N-module” subunits (*nuoE*, *nuoF*, and *nuoG*) that are involved in NADH binding, and has been suggested to be the evolutionary ancestor of the 14 subunit complex (Moparthy & Hägerhäll 2011; Kerscher et al. 2008). The lack of the N-module suggests it can receive electrons from donors other than NADH (Moparthy & Hägerhäll 2011).

Genes encoding succinate:quinone oxidoreductases (SQORs) are found in all genomes. These SQORs are characterized by having a single hydrophobic subunit anchor with two heme groups, and thus can be classified as ‘Type B’ (Lancaster 2002). AMDSBA4, however, has an additional SQOR adjacent to the first (AMDSBA4_13_54-57), distinct in that it contains two hydrophobic subunit anchors, each predicted to bind a heme group. As such, it most similar to ‘Type A’ SQOR enzymes most commonly associated with Archaea (Lancaster 2002; Schafer et al. 2002). The functional significance of this second SQOR in AMDSBA4 is unclear.

Sulfur metabolism

Oxidation of sulfur, tetrathionate, and sulfide minerals is a well-documented characteristic of *Sulfobacillus* spp. (Johnson et al. 2008; Melamud et al. 2003; Bogdanova et al. 2006; Watling et al. 2008), and the genomes analyzed here are replete with sulfur oxidation-reduction machinery. Elemental sulfur and various polythionates such as tetrathionate and pentathionate are ultimately formed by the reaction of acid-insoluble sulfide minerals (e.g., pyrite) with ferric iron (Schippers et al. 1996; Dopson & Johnson 2012; Luther 1987).

Aerobically, sulfur may be oxidized in a disproportionation reaction involving sulfur oxygenase reductase (SOR), which disproportionates linear polysulfide in the presence of oxygen to sulfite and hydrogen sulfide (Urich et al. 2006). SOR is found in *S. acidophilus*, AMDSBA2, and in two copies in *S. thermosulfidooxidans* and AMDSBA5. Best studied in the archaea *Acidianus ambivalens* (Urich et al. 2006), the SOR protein is cytoplasmic, and the lack of cofactors means it is not coupled to electron transport, supported by work done in *Acidithiobacillus caldus* (Lin Xu Chen et al. 2012). All *Sulfobacillus* SOR proteins possess the identified conserved residues for a functional enzyme (Urich et al. 2006; Zhang et al. 2013).

Another possible route of sulfur compound oxidation may be through heterodisulfide reductase-like (Hdr) proteins encoded in all *Sulfobacillus* genomes. In methanogens,

heterodisulfide reductases reduce heterodisulfide bonds in the final step of methanogenesis (Hedderich et al. 2005). However, proteins related to heterodisulfide reductases are widely distributed and are suspected to be involved more generally in various forms of sulfur metabolism (Grein et al. 2013; Hedderich et al. 1999). Clusters of Hdr-like proteins are conserved in other acidophilic sulfur oxidizing Bacteria and Archaea and are upregulated during aerobic growth on sulfur in *Acidithiobacillus ferrooxidans*, where they are predicted to oxidize disulfide intermediates to sulfite (Quatrini et al. 2009). The proposed disulfide substrates are sulfane sulfur species (RSS_nH), formed by the reaction of cellular thiols (e.g., cysteine residues) with extracellular elemental sulfur (Rohwerder 2003; Rohwerder & Sand 2007).

Two different types of *Sulfobacillus* *hdr* gene clusters are found with varying gene content (Figure 3.7). One group, *hdr* gene cluster I, was identified only in *S. thermosulfidooxidans*, AMDSBA1 and AMDSBA5. This cluster includes an HdrA-like protein with an FAD-binding site and a four conserved-cysteine motif ($\text{CxGCRD}_{x_6-8}\text{CSx}_2\text{CC}$) typical in binding a Fe-S cluster. Two HdrB proteins are present which contain the cysteine-rich domain CCG domain ($\text{Cx}_n\text{CCG}_{x_m}\text{C}$) predicted to bind [4Fe-4S] clusters (Hamann et al. 2007), and another [4Fe-4S] binding center is encoded by the HdrC protein. Furthermore, a DsrE-family protein encoded in this locus is predicted to be involved in intracellular sulfur trafficking (Pott & Dahl 1998), as is the protein TusA (Ikeuchi et al. 2006). It is also interesting to note that a homolog to the glycine cleavage protein H (a lipoic acid containing protein) is found in this locus as is a lipoate-protein ligase that catalyzes the attachment of lipoic acid to target proteins (Morris et al. 1994). The lipoic acid moiety of glycine cleavage protein H acts as a “swinging arm” that serves to transfer reaction intermediates between the various catalytic sites of the other proteins in the glycine cleavage system complex (Perham 2000). It is tempting to speculate that lipoic acid, with its disulfide bond, may play a key role in sulfur species transfers amongst this heterodisulfide complex.

The second group (*hdr* gene cluster II) is found in all *Sulfobacillus* genomes and contains many of the same subunits as *hdr* gene cluster I (Figure 3.7). However, there are additional genes encoding electron transfer flavoproteins as well as a second, longer *hdrC* gene containing 6 transmembrane helices and 2 cysteine-rich CCG domains not found in the HdrC of *hdr* gene cluster I. These three genes are syntenous, and similar arrangements have been observed in the sulfate reducing bacterium *Desulfobacterium autotrophicum* (Strittmatter et al. 2009).

When SQR catalyzes the disproportionation reaction of linear polysulfide to H_2S and sulfite, the sulfide may be oxidized back to sulfur by the membrane-bound sulfide:quinone oxidoreductases (SQR), with the concomitant reduction of quinones. Thus, sulfur disproportionation is linked to the electron transport chain (Marcia et al. 2009; WAKAI et al. 2007). Across the *Sulfobacillus* genomes, 74 proteins were annotated as SQRs or more generally as “FAD-dependent pyridine nucleotide disulfide oxidoreductases,” the family to which SQR proteins belong. There are three active-site cysteine residues shown to be important for SQR activity, with the second and third being essential (Cherney et al. 2010; Marcia et al. 2009). Alignment of these 74 *Sulfobacillus* putative SQRs and their related oxidoreductases with the characterized SQR of *Acidithiobacillus ferrooxidans* showed 9 as having all three active-site cysteine residues, and 27 as having the second and third essential active-site cysteine residues. Protein trees constructed with those 36 *Sulfobacillus* proteins containing at least the second and third essential active-site cysteine residues indicated that all proteins belonged to either Group III or Group V SQRs, based on the protein phylogeny laid out by Marcia et al 2009 (Figure 3.8). Both groups have representatives with demonstrated sulfide:oxidase activities (Marcia et al.

2009; Lencina et al. 2012). Based on this analysis, SQR is likely present in the published genomes of all *Sulfobacillus* species analyzed here.

Sulfite oxidation may be carried out by putative sulfite oxidoreductases homologous to the single molybdopterin-cofactor containing sulfite oxidase family of proteins (Kappler & Dahl 2001). All *Sulfobacillus* except for AMDSAB1 contained putative sulfite oxidoreductase proteins. Unlike the well-characterized sulfite oxidoreductase of *Starkeya Novella* (formerly *Thiobacillus novellus*) which consists of a two subunit complex encoded by syntenous genes (*SorAB*), the sulfite oxidoreductases of *Sulfobacillus* are encoded by a single gene, without the cytochrome-*c* containing SorB subunit. In contrast, the *Sulfobacillus* sequences contain a longer N-terminus region reminiscent of the heme-*b* binding domain found in eukaryotic sulfite reductases (Kappler 2000). It is unclear whether these enzymes transfer electrons directly to oxygen (i.e., they are sulfite reductases like those found in eukaryotes) or if electrons are transferred to an alternative electron acceptor like cytochrome *c* (i.e., they are sulfite dehydrogenases like that found in *S. novella*).

Tetrathionate can be disproportionated by tetrathionate hydrolase (TTH) into sulfate, thiosulfate, and other sulfur compounds (De Jong et al. 1997; Kanao et al. 2007; Bugaytsova & Lindström 2004). TTH activity has been detected in a strain of *S. thermosulfidooxidans* (Egorova et al. 2004), and tetrathionate oxidation is widespread amongst *Sulfobacillus* isolates (Shiers et al. 2010). TTH homologs were found in all but AMDSBA3 of the *Sulfobacillus* genomes. Based on TTH-family phylogeny laid out by Protze et al (2011), only *S. thermosulfidooxidans*, *S. acidophilus*, and AMDSBA1 have proteins that fall within characterized TTH-family clades (Figure 3.9). The remaining enzymes belong to yet uncharacterized families of TTH homologs. Furthermore, TTH proteins have been shown to be periplasmic in Gram-negative organisms [(Bugaytsova & Lindström 2004) and refs therein], and thought to be extracellular in *A. ambivalens* (Protze et al. 2011), and thus could also be extracellular in *Sulfobacillus* species.

Membrane-associated TQO enzymes were found in all *Sulfobacillus* genomes analyzed, except for AMDSBA3. Thiosulfate:quinone oxidoreductase (TQO) may oxidize thiosulfate formed by TTH back to tetrathionate and concomitantly passing electrons into the electron transport chain via the quinone pool (Müller et al. 2004).

Another mechanism of tetrathionate metabolism could occur via tetrathionate reductases (TTR), an enzyme belonging to the dimethylsulfoxide oxidoreductase (DMSOR) family of enzymes with a molybdo-bis (pyranopterin guanine dinucleotide) cofactor (Rothery et al. 2008). Several *Sulfobacillus* DMSOR catalytic subunits appear closely related to tetrathionate reductases, including the characterized tetrathionate reductase from *Salmonella enterica* (Figure 3.10). In *S. enterica*, TTR is a three-subunit enzyme that can reduce trithionate and tetrathionate to thiosulfate and sulfite (Hinsley & Berks 2002). TTR proteins were found in *S. acidophilus*, AMDSBA1, and a partial sequence is available from AMDSBA2. Except for a second putative TTR in AMDSBA1 (AMDSBA1_3_25, which also lacks the second and third subunits), the putative TTR first and second subunits each contain a Tat-signal sequence, indicating an extracellular location. No obvious heme or quinone-binding subunits were observed in the membrane-anchor subunit, leaving questions as to how this enzyme interacts with an electron transport chain of the cell.

Notably, phylogenetic analysis of other catalytic subunits of *Sulfobacillus* DMSOR-family proteins showed that several *Sulfobacillus* proteins are closely related to the sulfur reductase of *A. aeolicus* (Figure 3.10). The three-subunit sulfur reductase complex in *Aquifex aeolicus* is involved in the reduction of tetrathionate, sulfur, and polysulfide coupled to H₂

oxidation via a quinone pool (Guiral et al. 2005). Similar three-subunit gene operons are found in all *Sulfobacillus* species analyzed here. These proteins lack any signal peptides, and thus, like the sulfur reductase from *A. aeolicus*, may be cytoplasmically located and anchored to the membrane (Guiral et al. 2005). Whether or not these proteins are involved in sulfur reduction is unknown, and no sulfur reduction metabolism has yet been detected for any *Sulfobacillus* species.

Sulfate assimilation is presumed to occur in all *Sulfobacillus* species by the combined action of an ATP sulfurylase, an adenosine 5'-phosphosulfate (APS) reductase, a ferredoxin-dependent sulfite reductase, and an O-acetylserine sulfhydrylase. The APS-reductase is homologous to 3'-phosphoadenosine 5'-phosphosulfate (PAPS) reductases often found in sulfate assimilation pathways, but the presence of an iron-sulfur binding subunit suggests the preferred substrate is APS (Kopriva et al. 2002; Bick et al. 2000). AMDSBA5 lacked the ATP sulfurylase and APS reductase, thus inorganic sulfur assimilation may proceed from sulfite. Furthermore, AMDSBA2 and AMDSBA4 did not contain the sulfite reductase, which may reflect incomplete genomic coverage or point to inorganic sulfur assimilation directly from H₂S.

Iron oxidation and reduction

Although sulfur oxidation is more energetically favorable than iron oxidation (more electronegative potential and greater number of electrons per mole of substrate) many *Sulfobacillus* isolates are demonstrated iron oxidizers (Melamud et al. 2003; Bogdanova et al. 2006; Norris et al. 1996; Johnson et al. 2008). In other acidophilic microorganisms, outer-membrane cytochromes have been implicated in the oxidation of iron, including in the Gram-negative bacteria *Leptospirillum* spp. and *Acidithiobacillus ferrooxidans* [(Bonney & Holmes 2011) and refs therein]. Electrons from iron are ultimately transferred to either a terminal oxidase complex to reduce oxygen and consume and pump protons (“downhill”) or to an NADH dehydrogenase (“uphill”) (Bonney & Holmes 2011). There are membrane-associated *c*-type cytochromes that could be involved in iron oxidation in all *Sulfobacillus* species examined here except AMDSBA1. Electrons could be transferred from these membrane *c*-type cytochromes to the terminal cytochrome *c* oxidases in a “downhill” oxidation of iron. The “uphill” reactions require harnessing the proton-motive force in order to push the electrons in a thermodynamically unfavorable direction, and this is accomplished by the *bc* complex in *Leptospirillum* spp. and *A. ferrooxidans* (Bonney & Holmes 2011). Only AMDSBA4 is predicted to have this complex, and it is unclear whether it functions in reverse electron transport as in the Gram-negative iron oxidizers.

Ferric iron reduction has been shown for all isolate species of *Sulfobacillus* (Bridge & Johnson 1998; Johnson et al. 2008; Bogdanova et al. 2006; Melamud et al. 2003). Although Fe³⁺ reduction is widespread in acidophilic microorganisms (Coupland & Johnson 2008; Johnson et al. 2012), little is known about the precise enzymes and pathways involved. In *A. ferrooxidans*, a *c*-type cytochrome has been implicated in ferric iron respiration (Ohmura et al. 2002) as has a rusticyanin (Kucera et al. 2011). Rusticyanin proteins are found annotated only in *S. thermosulfidooxidans* and AMDSB5, and membrane-bound *c*-type cytochrome proteins are found in all *Sulfobacillus* genomes except AMDSBA1. Osorio et al (2013) also proposed a model for ferric iron reduction in *A. ferrooxidans* when elemental sulfur was an electron donor, including one in which Fe³⁺ was indirectly reduced by H₂S generated from anaerobic sulfur disproportionation. This H₂S ‘scavenging’ allows normally endergonic sulfur disproportionation to become thermodynamically favorable (Thamdrup et al. 1993; Hardisty et al. 2013). While

generation of hydrogen sulfide may play a role in ferric iron reduction, *Sulfobacillus* have shown to reduce iron in cultures without any sulfur species present (Karavaiko et al. 2001; Johnson et al. 2008), implying that some mechanism for the direct reduction of iron is present. However, there is scant evidence present at this time as to what that mechanism might be.

Hydrogen metabolism

The catalytic subunits for twenty-seven different nickel-iron (NiFe) hydrogenase proteins were identified across all *Sulfobacillus* species, which cluster into 5 distinct subfamilies (Figure 3.11). One conserved 16-subunit gene cluster was identified within *S. acidophilus*, *S. thermosulfidooxidans*, and AMDSBA5, and the catalytic subunit was found to cluster with Group 1 respiratory-uptake [NiFe]-hydrogenases. Like other Group I hydrogenases, these hydrogenases possess a Tat-motif at the N-terminus of the small subunit and are thus predicted to be located extracellularly. There, they are predicted to oxidize H₂, transferring protons to the quinone pool via a cytochrome *b*-containing subunit (Vignais & Billoud 2007). The second 16-subunit hydrogenase gene cluster was identified in *S. acidophilus* and *S. thermosulfidooxidans* species, although the operon structure and catalytic subunits differ. In *S. acidophilus*, the catalytic subunits share the greatest homology with subgroup A of the Group 2 hydrogenases, which are cytoplasmic hydrogen-utilizing complexes most commonly associated with nitrogen fixing cyanobacteria [Figure 3.11, (Vignais & Billoud 2007)]. The hydrogenases of *S. thermosulfidooxidans*, however, are most similar to a relatively recently identified fifth group of hydrogenases [Figure 3.11, (Constant et al. 2010)]. This fifth group is composed of putative high-affinity hydrogenases and has been detected predominantly in Actinobacteria, Acidobacteria, and Chloroflexi (Constant et al. 2011). Neither the hydrogenases of Group 2 nor Group 5 contain a signal peptide, and are thus predicted to remain in the cytoplasm. Moreover, they do not contain the *b*-type cytochrome subunit thought to aid in electron transfer to quinones, and the electron acceptor is unclear. No Group 1, 2 or 5 hydrogenases were found in AMDSBA1, AMDSBA2, AMDSBA3, or AMDSBA4.

The catalytic subunits of the remaining hydrogenases indicate that they belong to the Group 4 hydrogenase family, and at least one representative is found in each *Sulfobacillus* species (Figure 3.11). Six syntenic genes are found near each catalytic subunit, and bear homology to Complex I enzymes (*nuoH*, *nuoL*) and formate hydrogen-lyase enzymes (*HyfE* and *HyfB*). The catalytic large subunits of these hydrogenases lack the binding site motifs normally found in [NiFe] hydrogenases [N-terminal: RxCGxCxxxH; C-terminal DPCxxCxxH/R; (Thauer et al. 2010)]. As such, these enzymes are similar to energy-converting hydrogenase-related complexes (Ehr) and their function remains unknown (Coppi 2005; Vignais & Billoud 2007; Marreiros et al. 2013).

Carbon monoxide dehydrogenases

Aerobic-type carbon monoxide dehydrogenases (CODH) of both Form 1 and Form 2 were identified in *Sulfobacillus* genomes. Form 1 CODHs are known to be involved in carbon monoxide oxidation to CO₂ with release of reducing equivalents introduced into the electron transport chain usually via cytochromes (King & Weber 2007). Form I enzymes were found in all organisms except AMDSBA2 and AMDSBA3. The function of Form II enzymes—differentiated by the active site motif AYRGAGR in place of the Form I motif AYXCSFR—are not fully understood, and may also play roles in CO oxidation as well as oxidation of other substrates (King & Weber 2007). Form II CODH enzymes are found in all *Sulfobacillus* species

Nitrogen metabolism

AMDSBA3 is the only *Sulfobacillus* species analyzed here that contains genes for dissimilatory nitrate reduction, observed as a single *NarGHJI* operon (Figure 3.10, Figure 3.6). No other components for denitrification or nitrite reduction were found. AMDSBA3 is thus the only *Sulfobacillus* species analyzed here capable of using nitrate as an alternative electron acceptor in anaerobic environments.

AMDSBA1 and *S. thermosulfidooxidans* contained genes encoding assimilatory nitrate reductases (Figure 3.6). The proteins encoded by the nitrate reductase genes (*NasC*) are members of the DMSOR family of enzymes, and phylogenetically cluster near the assimilatory nitrate reductase of *Bacillus subtilis* (Figure 3.10). However, unlike the nitrate reductase of *B. subtilis*, the nitrate reductases in AMDSBA1 and *S. thermosulfidooxidans* do not seem to possess an electron transfer subunit (*NasB*) required for the transfer of electrons from NADH to nitrate (Lin & Stewart 1998), making the electron donor unclear (e.g., NADH, flavodoxin, or ferredoxin). Nearby genes encoding nitrite reductases (*NasDE*) in both species contain the necessary siroheme and NADH binding sites required for the six-electron reduction of nitrite to ammonia using NADH as the electron donor (Lin & Stewart 1998; Richardson et al. 2001). Copper-containing nitric oxide-forming nitrite reductases were found in *S. acidophilus* strains, as were nitric oxide dioxygenase enzymes, which may oxidize nitric oxide to nitrate as a detoxification mechanism (Gardner 2005). It has also been suggested that these enzymes may reduce nitric oxide to nitrous oxide anaerobically, which if true, would allow denitrification activity (nitrite to nitrous oxide) in these organisms (Poole & Hughes 2000).

Ammonium transporters were identified in all genomes, and ammonia can be assimilated into central metabolic pathways via glutamate dehydrogenase, glutamine synthetase and glutamate synthase.

Autotrophy

Sulfobacillus spp. are known autotrophs, and published genomes of *S. thermosulfidooxidans* DSM 9293, and *S. acidophilus* indicate the presence of carbon fixation via Calvin-Benson-Bassham (CBB) cycle (Caldwell et al. 2007; Anderson et al. 2012; Bo Li et al. 2011; Travisany et al. 2012). The *Sulfobacilli* also have complete CBB cycle genes, including a type I ribulose-1,5-bisphosphate carboxylase-oxygenase (RuBisCO) (Figure 3.12). In addition to the type I RuBisCO, AMDSBA4 and *S. acidophilus* strains have additional Type 4 RuBisCO-like proteins (Figure 3.12) that likely participate in the methionine salvage pathway (Tabita et al. 2007). Interestingly, AMDSBA4 is the only genome with an annotated carboxysome (AMDSBA4_48_6-9).

Central carbon pathways

In addition to autotrophic growth, *Sulfobacillus* isolates have been reported to grow mixotrophically and heterotrophically on a variety of carbon substrates, including glucose, fructose, glycerol and other various organic carbon compounds (Bogdanova et al. 2006; Bridge & Johnson 1998; Johnson et al. 2008). All organisms except *S. thermosulfidooxidans*, AMDSBA5, and AMDSBA1 have a complete Embden-Meyerhoff pathway for glycolysis. *S. thermosulfidooxidans*, AMDBA5, and AMDSAB1, lack a 6-phosphofructokinase. While genes encoding 6-phosphofructokinase could not be identified in *S. thermosulfidooxidans*, cell-free enzymatic assays suggest these cells can carry out this transformation (Karavaiko et al. 2001).

Glucose-6-phosphate isomerase activity in AMDSBA4 and both *S. acidophilus* can be accounted for by the inclusion of unique bifunctional transaldolases/phosphoglucose isomerase (Sugiyama et al. 2003). The oxidative portion of the pentose phosphate pathway (glucose dehydrogenase, 6-phosphogluconolactonase, and 6-phosphogluconate dehydrogenase) was present in all organisms.

Complete semi-phosphorylative Entner-Doudoroff carbon degradation pathway (2-keto-3-deoxygluconate 6-phosphate aldolase, 2-dehydro-3-deoxygluconokinase, and gluconate dehydratase) was present in *S. acidophilus* strains, AMDSAB1, AMDSBA3, and AMDSBA4. While no gluconate dehydratases were annotated, deoxy-acid dehydratases likely carry out this function as they do in other bacteria (Kim & Lee 2008). Phosphogluconate dehydratases, indicative of the phosphorylative Entner-Doudoroff pathway, were not found in any genomes. Enzyme assays from cell-free extracts of *S. thermosulfidooxidans* and *S. sibiricus* strains have shown this phosphogluconate dehydratase activity, however, as well as keto-3-deoxy-6-phosphogluconate aldolase activity *S. thermosulfidooxidans*, which would indicate a phosphorylative Entner-Douroroff pathway (Karavaiko et al. 2001; Krasil'nikova et al. 2010).

Pyruvate can be oxidatively decarboxylated to acetyl-CoA with pyruvate dehydrogenase (found in all *Sulfobacillus*) or in AMDSBA1 with a putative pyruvate ferredoxin oxidoreductase (AMDSBA1_57_11-13). Excepting a few gaps from the low-coverage genome of AMDSBA2, the tricarboxylic acid cycle was complete in all genomes. Furthermore, a putative oxoglutarate synthase that may allow oxidations of 2-oxoglutarate to be coupled to the reduction of ferredoxin was found in all *Sulfobacillus* genomes. Finally, a complete glyoxylate bypass system (malate synthase and isocitrate lyase) was found in only AMDSBA1 and *S. acidophilus*, possibly permitting carbon assimilation of C₂ compounds like acetate. Malate synthase was found in all but AMDSBA3 and AMDSAB4, which together with glycolate oxidase, may provide a route of glycolic acid assimilation. Glycolic acid is produced as an exudate by key primary producing acidophiles like *Leptospirillum*, and the degradation of glycolic acid sets *Sulfobacillus* apart amongst other acidophiles. In fact, it is thought to contribute to an organic-carbon degrading niche in stirred-tank bioleaching systems (Nancuqueo & Johnson 2010). Lactate utilization proteins similar to those found in *Bacillus subtilis* were identified in all organisms, and are thought to allow growth on lactate (Chai et al. 2009).

Formate dehydrogenase

Putative cytoplasmic selenocysteine-containing formate dehydrogenases are found in all *Sulfobacillus* genomes (Figure 3.10). These genes possess the binding site and three catalytic residues necessary for formate oxidation. Each catalytic subunit is syntenous to cluster of genes with homology to the NADH-binding NuoEFG subunits, and thus may represent coupling of formate oxidation to NADH reduction.

Fermentative metabolism

Strains of *S. thermosulfidooxidans*, *S. sibiricus*, and *S. thermotolerans* have been shown to produce propionate and acetate in the growth media during growth in hypoxic conditions (Tsaplina et al. 2010). Complete fermentative pathways that result in propionate formation, however, could not be identified. In all *Sulfobacillus* species, neither the enzymes lactoyl-CoA dehydratase [of the acrylate pathway (Tholozan et al. 1992)] nor propionyl-CoA:succinate-CoA transferase and methylmalonyl-CoA carboxyltransferase [of the succinate-propionate fermentative pathway (Schink et al. 1987)] could be identified. Acetate production could be

produced with substrate level phosphorylation by a phosphate acetyltransferase (found only in AMDSBA2) and an acetate kinase (identified in AMDSBA1 and AMDSBA2). Traditional fermentative pathways thus cannot account for propionate and acetate produced in isolate strains.

Lipid and lipopolysaccharide biogenesis

All *Sulfobacillus* species have the genes required to synthesize fatty acids. Species of *Sulfobacillus* predominantly contain branched-chain anteiso and ω -alicyclic fatty acids (Tsaplina et al. 1994). In general, branched chain fatty acids can be synthesized from branched short-chain carboxylic-acid CoA esters or from α -keto acids (Kaneda 1991). In the *Sulfobacillus*, the former pathway is probably used, as branched chain α -keto acid decarboxylases were not identified. The chorismate pathway and genes encoding biosynthesis of ω -alicyclic fatty acids were identified in all *Sulfobacillus* species. The chorismate pathway produces shikimic acid as an intermediate (OSHIMA & ARIGA 1975), which is ultimately used to form cyclic carboxylic acid-coA esters used in the synthesis of ω -alicyclic fatty acids. Peptidoglycan biosynthesis was accounted for in all *Sulfobacillus*. An S-layer has been reported to be present on some cells of *S. thermosulfidooxidans* (Duda et al. 2001), however only a single annotated S-layer protein was found in AMDSBA1 (AMDSBA1_76_63). Notably, a gene cluster present only in AMDSBA3 encodes for a putative squalene biosynthetic pathway (AMDSBA3_32_15-22), including a squalene-cyclase that could be involved in the production of hopanoids (Kannenberg & Poralla 1999; Spanova & Daum 2011). Hopanoids have been previously identified in acidophilic organisms (Jones et al. 2011), and are thought to alter membrane fluidity and permeability to protons (Welander et al. 2012).

Vitamin and coenzyme biosynthesis

Menaquinone biosynthetic pathways were identified in all *Sulfobacillus* isolates, in agreement with detection of a menaquinone with 7 isoprenoid units in *Sulfobacillus thermotolerans* (Bogdanova et al. 2006). Complete biosynthetic pathways for lipoic acid, coenzyme A, nicotinamide adenine dinucleotide, flavin adenine dinucleotide, and biotin were identified. Vitamin B12 biosynthetic proteins were found in all organisms except AMDSBA3 and AMDSBA4. All *Sulfobacillus* genomes have B12-dependent methylmalonyl-CoA genes, and thus all likely require the B12 cofactor. The corrinoid-transporting enzyme encoded by the *btuF* gene was found in all species, thus AMDSBA3 and AMDSBA4 probably scavenge vitamin B12 from the environment while the other *Sulfobacillus* can either scavenge or synthesize B12 *de novo*.

Environmental stress

Genes encoding alkyl hydroperoxide reductases/peroxiredoxins and superoxide dismutases were found in all species. Oxidative stress responses are important in all aerobic Bacteria and Archaea, and defense against reactive oxygen species in metal-rich acidic environments is especially important due to redox-active metals creating reactive-oxygen species via the Fenton reaction (Cárdenas et al. 2012). Thioredoxin reductases were also identified and may play an important role in ROS defense as identified in acidophilic *Leptospirillum* bacteria (Norambuena et al. 2012).

Genes encoding for the biosynthetic pathways of ectoine and hydroxyectoine were found solely in AMDSBA1 (AMDSBA1_21_55-58). In all *Sulfobacillus*, trehalose could be synthesized by the combined actions of trehalose-6-phosphate synthase and trehalose-6-

phosphate phosphatase. However, a second trehalose biosynthetic pathway, involving trehalose synthase (Empadinhas & da Costa 2006), was found in *S. thermosulfidooxidans* as well as AMDSBA1, AMDSBA2, and AMDSBA5. Compatible solutes play important roles in adaptation to the osmotic stress induced by the high-ionic strength environments like those found in metal-rich, acidic environments.

All *Sulfobacillus* isolates are known spore formers (Melamud et al. 2003; Bogdanova et al. 2006; Johnson et al. 2008), and all genomes analyzed here are consistent with sporulation capability.

Transport

Transporters for citrate, potassium, phosphate were identified in all species as were chromate, copper, and arsenite efflux pumps. Numerous amino acid permease-like proteins, oligopeptide and dipeptide transporters, and sugar ABC transport proteins were identified in all species, consistent with a heterotrophic mode of carbon assimilation. Polar amino acid ABC transporters were found only in AMDSBA1, AMDSBA4, and both *S. acidophilus* species and branched amino-acid transport systems were found in all but AMDSBA3. A putative taurine transporter was identified in all but AMDSAB3 and AMDSBA4. Taurine has been identified as a high-abundance compatible solute putatively synthesized by Eukaryotes in the Richmond Mine AMD system, and may provide an important source of nitrogen, carbon, and sulfur for *Sulfobacillus* organisms (Mosier et al. 2013).

Motility

All *Sulfobacillus* contained operons for the synthesis of flagella and chemotaxis signal transduction systems. Notably, both *S. acidophilus* strains contain nine separate methyl-accepting chemotaxis proteins involved in the transduction of environmental signals to motility response (Bren & Eisenbach 2000). The other genomes contained far fewer, with AMDSBA2 containing six; AMDSBA3 containing four; and *S. thermosulfidooxidans*, AMDSBA5, AMDSBA4 and AMDSBA1 all containing two or less.

Discussion

Analysis of the *Sulfobacillus* genomes illuminates metabolic pathways used in various transformations of carbon, nitrogen, and sulfur metabolism (summarized Figure 3.13 and Supplemental Table 3.1). Moreover, potential new metabolisms not yet observed in isolate cultures are potentially encoded within these genomes, including hydrogen oxidation, and putative anaerobic sulfur reduction capabilities.

Carbon fixation via the Calvin cycle is a well-described characteristic of *Sulfobacillus* species (Z W Chen et al. 2007), and type I RuBisCO genes were identified in all *Sulfobacillus* genomes. It has been noted that *Sulfobacillus* growth rates increase when they are cultivated in the presence of yeast extract and simple sugars (Ghuri & Johnson 1991; Karavaiko et al. 2001). Optimum growth under mixotrophic conditions using inorganic electron donors (usually ferrous iron) and organic carbon (usually glucose or yeast extract) is widely observed (Wood & Kelly 1983; 1984; Karavaiko et al. 2001; Tsaplina et al. 2008). To our knowledge, only one isolated strain (*S. acidophilus* NAL) is capable of purely heterotrophic growth, although growth rates are greatly decreased (Norris et al. 1996). In another strain of *S. acidophilus* (ALV), it was estimated

that about 20% of cellular carbon was derived from CO₂ fixation during growth on glucose (Wood & Kelly 1984).

Prior work demonstrated that glucose consumption proceeded via the oxidative pentose phosphate pathway (Wood & Kelly 1984), and that enzymes of the Entner-Doudoroff pathway (6-phosphogluconate dehydratase and 2-keto-3-deoxy-6-phosphogluconate aldolase) were not present. While our results are consistent with the absence of 6-phosphogluconate dehydratase in all genomes studied here, the 2-keto-3-deoxy-6-phosphogluconate aldolase was found in several AMDS genomes as well as in the genomes of *S. acidophilus* strains TPY and NAL—allowing for a complete semi-phosphorylative Entner-Doudoroff pathway. The presence of this enzyme in *S. acidophilus* NAL may be a key component contributing to its ability to grow heterotrophically.

Across all genomes analyzed here, only two of the enzymes involved in sulfur oxidation were conserved: sulfide:quinone oxidoreductases (SQR) and the heterodisulfide reductase cluster II (Hdr II). While the Hdr II proteins were only present in a single copy amongst all genomes, the SQR proteins were found in between three and six copies for all organisms except AMDSBA1 (Figure 3.6). Although the precise reason for the multiple copies of the SQR is unclear, it suggests the importance of sulfide oxidation for many of these organisms.

Sulfide, along with sulfite, can be formed aerobically through a disproportionation reaction of elemental sulfur with sulfur oxygenase (SOR). The subsequent oxidation of sulfide by sulfide:quinone reductase allows for the disproportionation reaction of SOR to be coupled to the electron transport chain. While all organisms contain SQR proteins, AMDSBA1, AMDSBA3 and AMDSBA4 lack the SOR proteins necessary to disproportionate sulfur. In these cases, the source of hydrogen sulfide is unclear. Sulfide may be formed by the reduction of sulfite using sulfite reductase (AMDSBA1, 5, 3), or perhaps hydrogen sulfide is scavenged from the surrounding environment. Another source of sulfide could be derived from anaerobic sulfur disproportionation. In this pathway, electrons from sulfane oxidation at the heterodisulfide reductase complex could be used to reduce sulfur, as was proposed for *A. ferroxidans* (Osorio et al. 2013). This would require the production of 3 moles of H₂S for every one mole of sulfate to balance electrons, however, and the reaction is thus predicted to be thermodynamically favorable only if H₂S is ‘scavenged’ by reactions with ferric iron in the environment (Thamdrup et al. 1993; Hardisty et al. 2013).

Regardless of whether anaerobic sulfur disproportionation reactions occur, the presence of enzymes putatively involved in sulfur-compound reduction (sulfur reductase and tetrathionate reductase) is interesting, as no sulfur reducing capacity has ever been noted for a *Sulfobacillus* species. To our knowledge, hydrogen utilization has also never been described, although the presence of several different uptake hydrogenase complexes suggests that hydrogen may be an important source of low-potential electrons for these organisms. In the AMDS genomes, only AMDSBA5 contained hydrogenases and these complexes may be a key component of niche specialization for AMDSBA5 in the Richmond Mine. The source of hydrogen, however, is unclear, although flammable, odorless bubbles of gas have been observed to be trapped within the biofilms found in the mine (unpublished observations).

Thiosulfate has also been detected in sediments as well as in solutions within the Richmond Mine (Ma & Banfield 2011). Thiosulfate is predicted to be the first aqueous reaction product of pyrite oxidation (Druschel et al. 2004; Luther 1987). However, because thiosulfate is readily oxidized by ferric iron in acidic media, it is predicted to be found only on the surface of pyritic sediments in the presence of ferric iron. Indeed, thiosulfate was found in bulk solution

only in highly reduced conditions essentially devoid of Fe^{3+} (Ma & Banfield 2011). The availability of thiosulfate may explain why *Sulfobacillus* spp. are found so infrequently in the floating biofilms of the Richmond Mine, and are more sometimes associated with the sediment or as members of submerged biofilms (Justice et al. 2012). Tetrathionate, formed by the oxidation of thiosulfate by ferric iron, is far more stable at acidic pH (Druschel et al. 2003), and thus it is surprising that it was not detected in AMD solutions in the Richmond Mine (Druschel et al. 2004). The absence of tetrathionate, however, can be explained if it is rapidly oxidized by *Sulfobacillus* organisms.

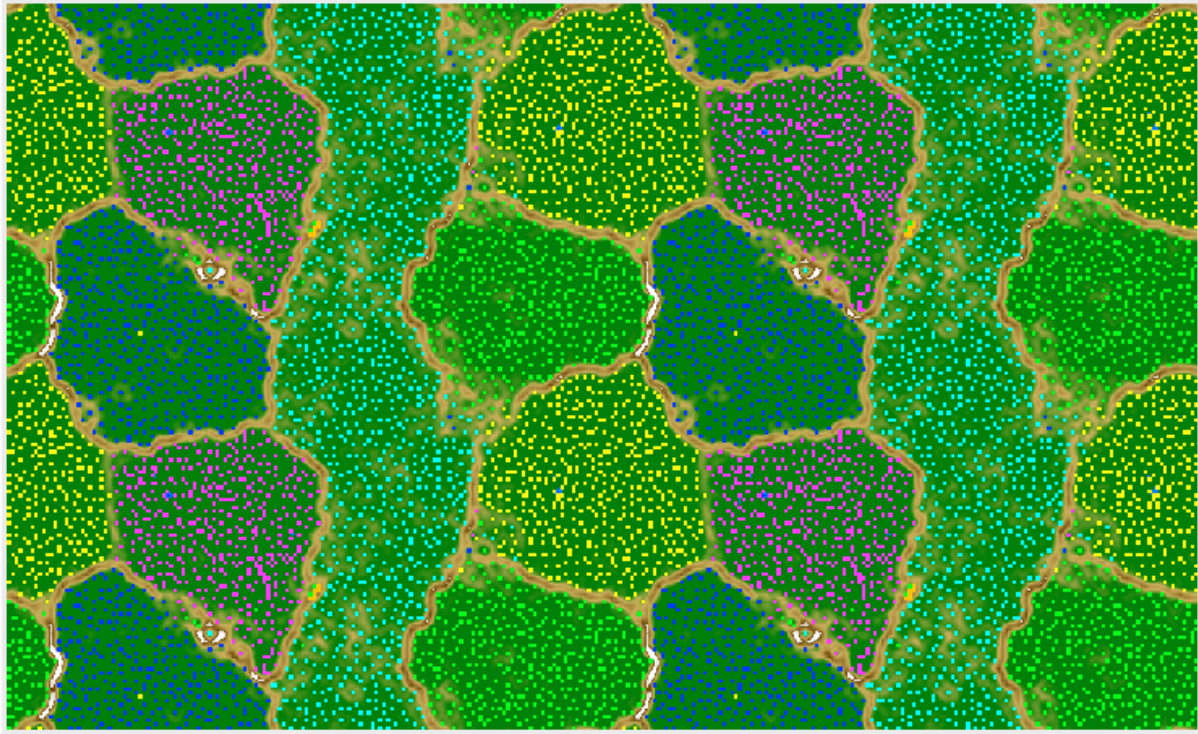
Amongst all of the genomes analyzed here, AMDSBA4 is unique in many respects. In addition to the putative carboxysome proteins, AMDSBA4 possesses the only *bc* complex amongst all of the *Sulfobacillus*. This is significant as the *bc* complex has been implicated in reverse electron transport during aerobic iron oxidation in *Acidithiobacillus* and *Leptospirillum* species (Bonney & Holmes 2011). Furthermore, a second succinate dehydrogenase and a unique cytochrome *c*-oxidase suggest that there may be many distinct aspects of electron transport in AMDSBA4 that warrant further investigation.

The genomic comparisons presented here advance our understanding of the metabolic potential within the genus *Sulfobacillus* and demonstrate diverse ecological strategies for these organisms. Pathways involved in inorganic sulfur, hydrogen, and nitrogen metabolisms are distributed unevenly amongst species, suggesting key differences in energy metabolism that may be foundational to niche differentiation. Common to all *Sulfobacillus* genomes, however, is a high degree of metabolic versatility, enabling these species to survive in a wide range of acidic environments and adapt to changing biogeochemical conditions. Information gained from this study will help unravel the ecological function of these organisms and their interactions with other organisms in both natural and industrial environments.

Co-author contributions:

Anders Norman carried out all sequence assembly and binning procedures. Christopher T. Brown assisted and wrote scripts for finding orthologs amongst the genomes. Andrea Singh and Brian C. Thomas provided bioinformatic support. Jillian F. Banfield helped interpret results and edit text.

Figure 3.1 Emergent self-organizing map (ESOM) of Clostridiales sequence fragments binned by time-series abundance patterns (>1500 bp). Note that the map is continuous (top and bottom edges, left and right edges). Each point represents a sequence fragment with colors indicative of the genome bin it belongs too (see legend below).



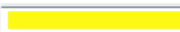



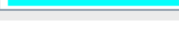
SULFO_1	
SULFO_3	
SULFO_4	
SULFO_5	
SULFO_6	

Figure 3.2 Genome completeness as estimated by the number of conserved single copy genes identified in each of the *Sulfolobus* genomes

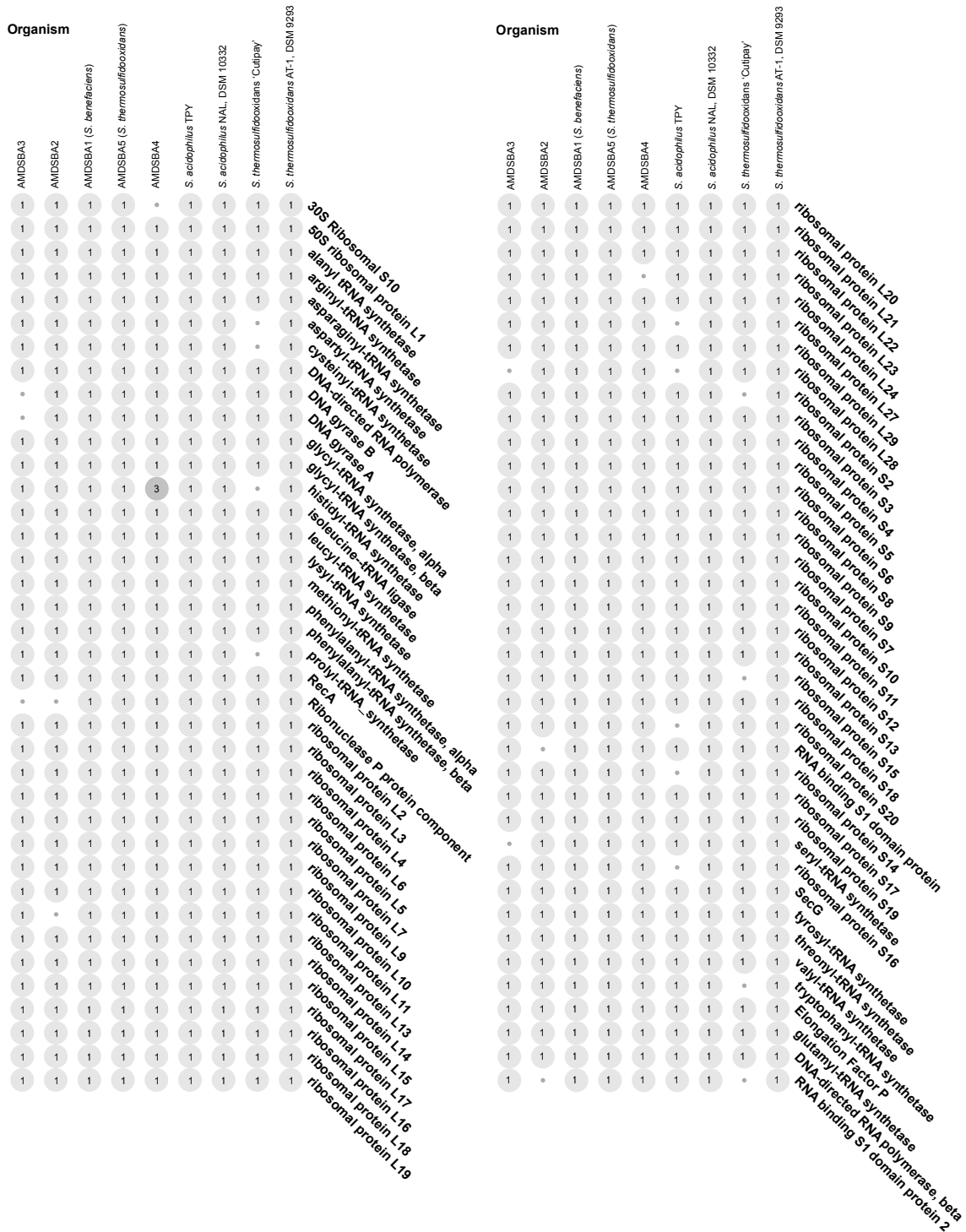


Figure 3.3 Maximum likelihood phylogenetic trees showing relationship amongst Iron Mountain and published *Sulfobacillus* species using A) concatenated alignment of sixteen ribosomal proteins and B) EMIRGE-generated 16S rRNA genes. Black diamonds indicate nodes with greater than 90% bootstrap support.

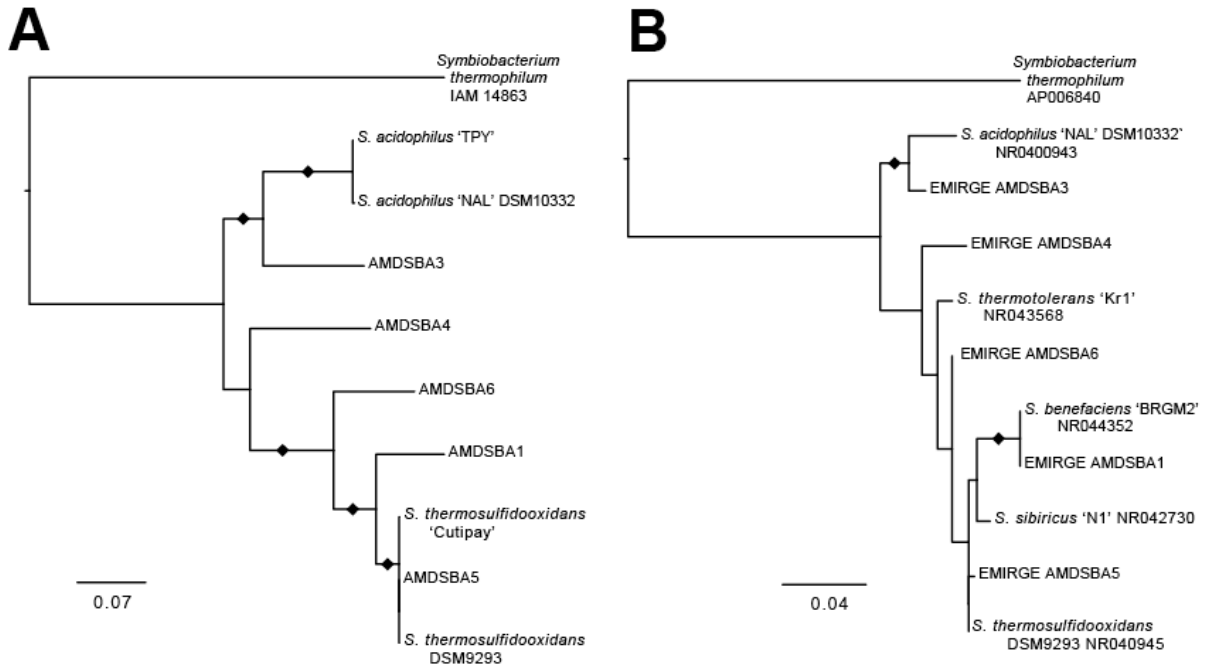


Figure 3.4 FISH images of AB Muck submerged biofilm showing *Sulfobacillus* (SUL230 probe, red) and archaeal (ARC15 probe, green) populations.

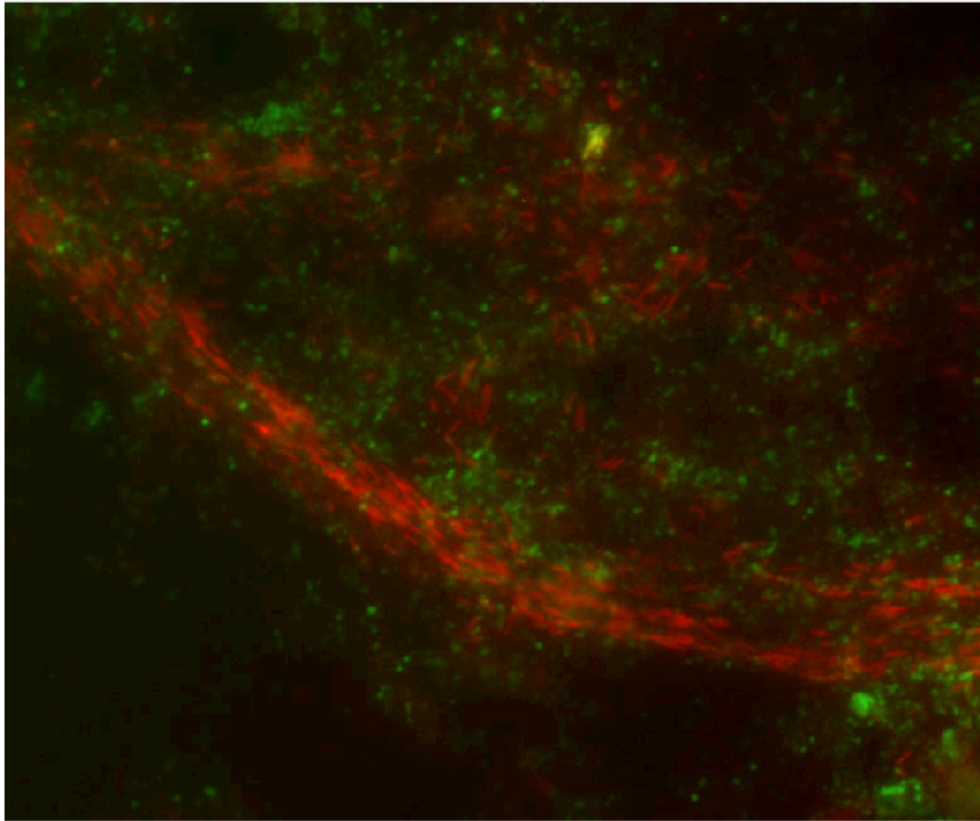


Figure 3.5 *Sulfobacillus* abundance estimated as a percentage of basepairs mapped to each organism over total sequenced DNA in each sample. 5way fungal streamer data is removed in (B) to better visualize low-abundance organisms. Sample and growth stage is depicted on the X-axis, with GS0-1 indicating low growth-stage biofilms, and GS1.5-2 indicating more mature, thicker growth stage biofilms.

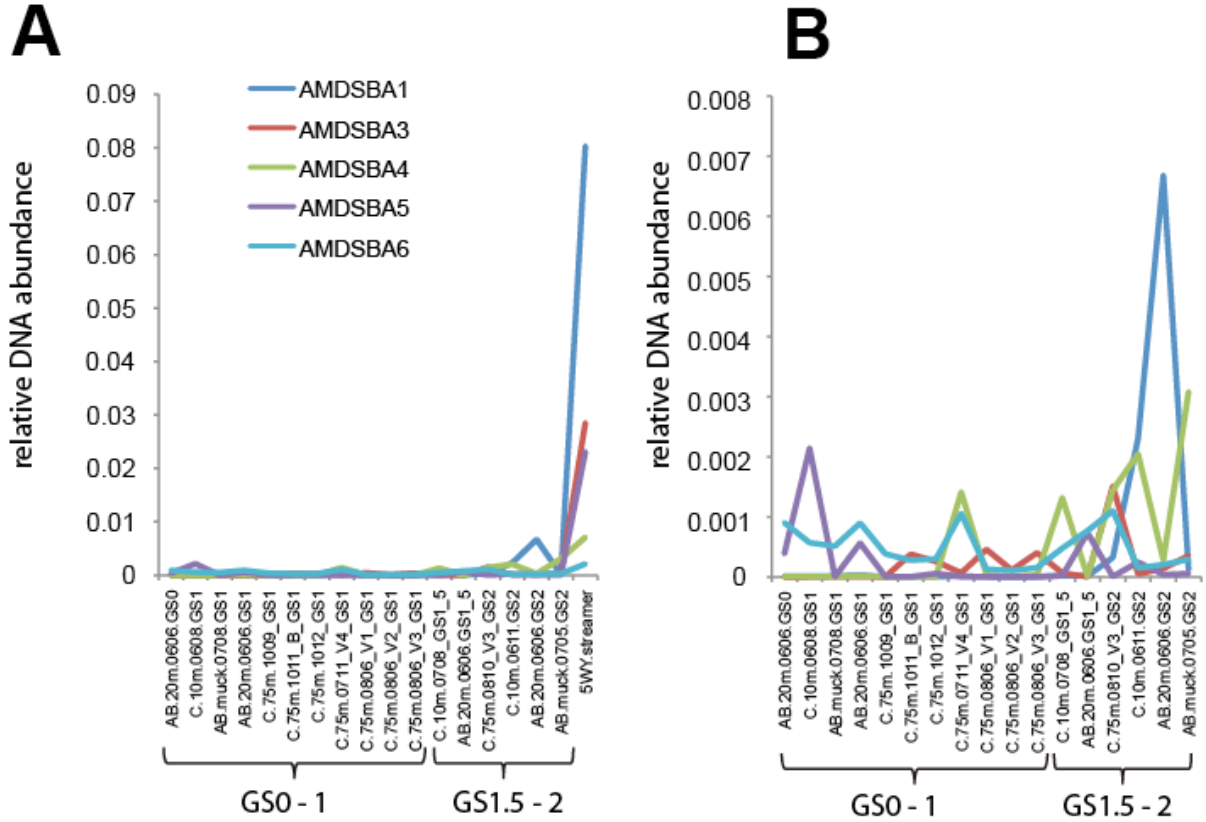


Figure 3.6 Gene copy numbers for electron transport chain, sulfur, hydrogen, nitrogen metabolisms in *Sulfobacillus* genomes. A filled blue square indicates presence, with a white number indicating copy numbers >1. Filled grey squares are presumed present due to presence in closely related organisms or partial gene identification. Colored circles and triangles at the top of the graph indicate species identifier in Figure 3.13.

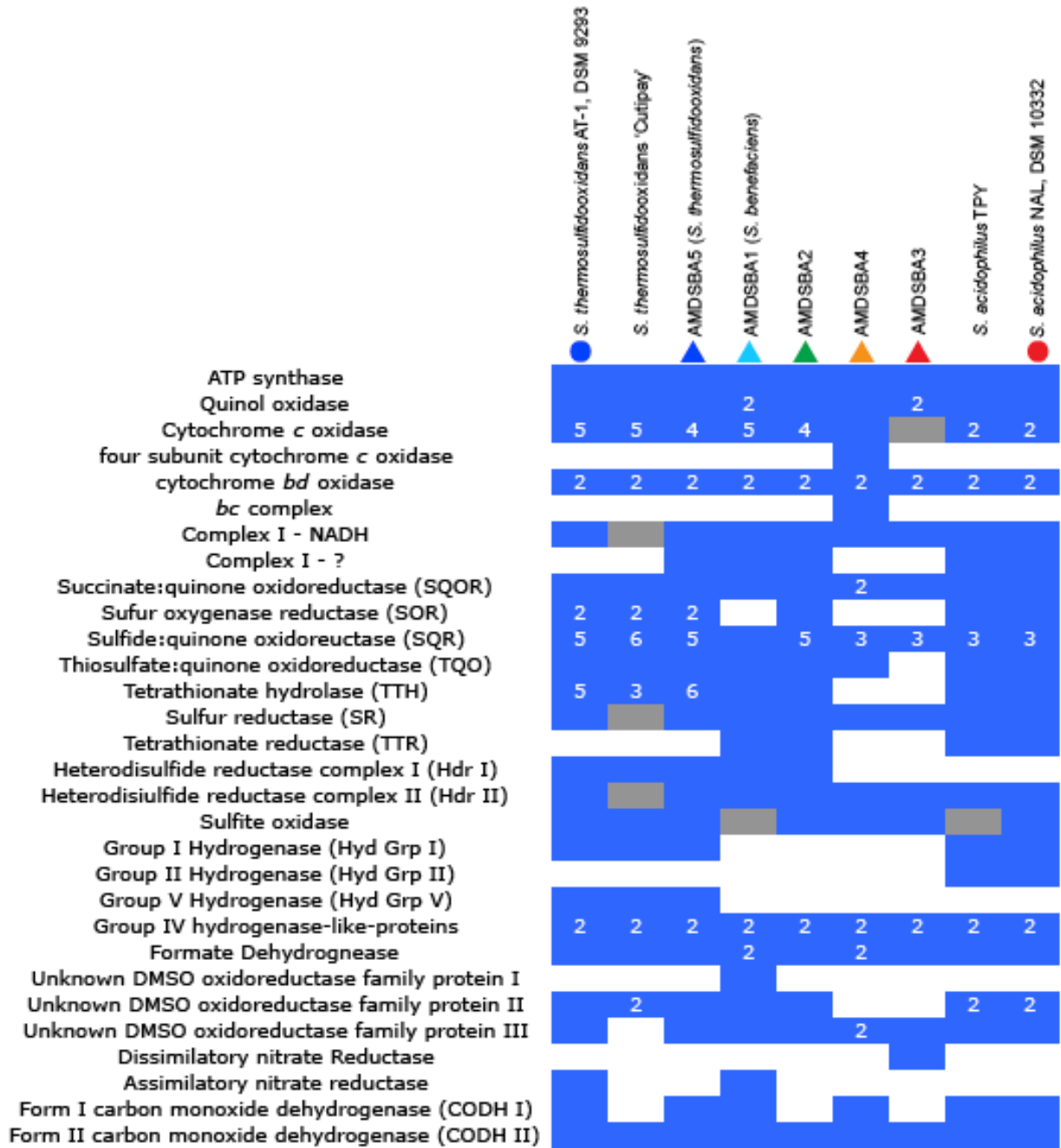


Figure 3.7 Operon structure of putative heterodisulfide reductase complexes. Hdr cluster I is found in *S. thermosulfidooxidans*, AMDSBA1, AMDSBA2, and AMDSBA5. Hdr cluster IIa is found in *S. thermosulfidooxidans* AMDSBA1, AMDSBA2, AMDSBA4, and AMDSBA5. Hdr cluster IIb is found in *S. acidophilus* and AMDSBA3. Gene abbreviations are as follows: *thdx*, thioredoxin; *gcsH*, glycine cleavage system protein H; *hdrABC*, heterodisulfide reductases; *hypo.*, hypothetical; *dsrE*, dsrE-like sulfur relay protein; *tusA*, sulfur transfer protein; *lpl*, lipote protein ligase; *etfAB*, electron transfer flavoprotein. Gene colors correspond to protein orthology as defined in methods.

Heterodisulfide reductase gene cluster I



Heterodisulfide reductase gene cluster IIa



Heterodisulfide reductase gene cluster IIb



Figure 3.8 Phylogenetic analysis of sulfide:quinone oxidoreductases proteins (SQR). Sequences from *Sulfobacillus* genomes are listed in red. Diamonds indicate nodes with >90% bootstrap support. Bootstrap values greater than 55% are shown as text. Asterisks indicate proteins containing all three conserved active site cysteine residues, all other *Sulfobacillus* sequences contain only the second and third residues. Protein tree adapted from SQR phylogeny laid out by Marcia et al (2009). The tree is rooted midway to the sulfur oxygenase reductase from *Thioalkalivibrio nitratireducens* (YP 007217840), which was used as an outgroup.



Figure 3.9 Phylogenetic analysis of tetrathionate hydrolase proteins (TTH). Sequences from *Sulfobacillus* genomes are listed in red. Diamonds indicate nodes with >90% bootstrap support. Bootstrap values greater than 55% are shown as text. Protein tree adapted from Protze et al (2011).

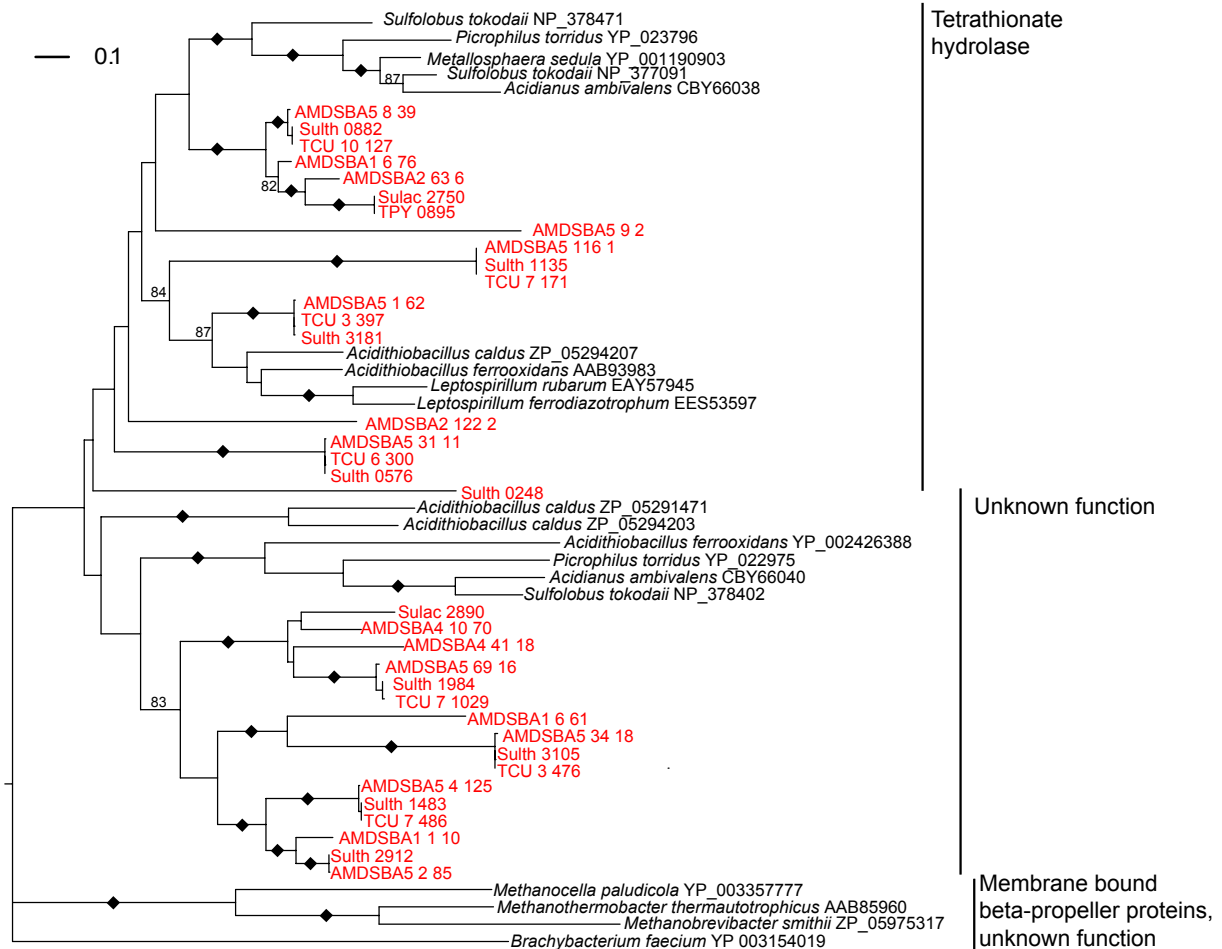


Figure 3.12 Phylogenetic analysis of RuBisCO proteins. Sequences from *Sulfobacillus* genomes are listed in red. Diamonds indicate nodes with >90% bootstrap support. Bootstrap values greater than 55% are shown as text.

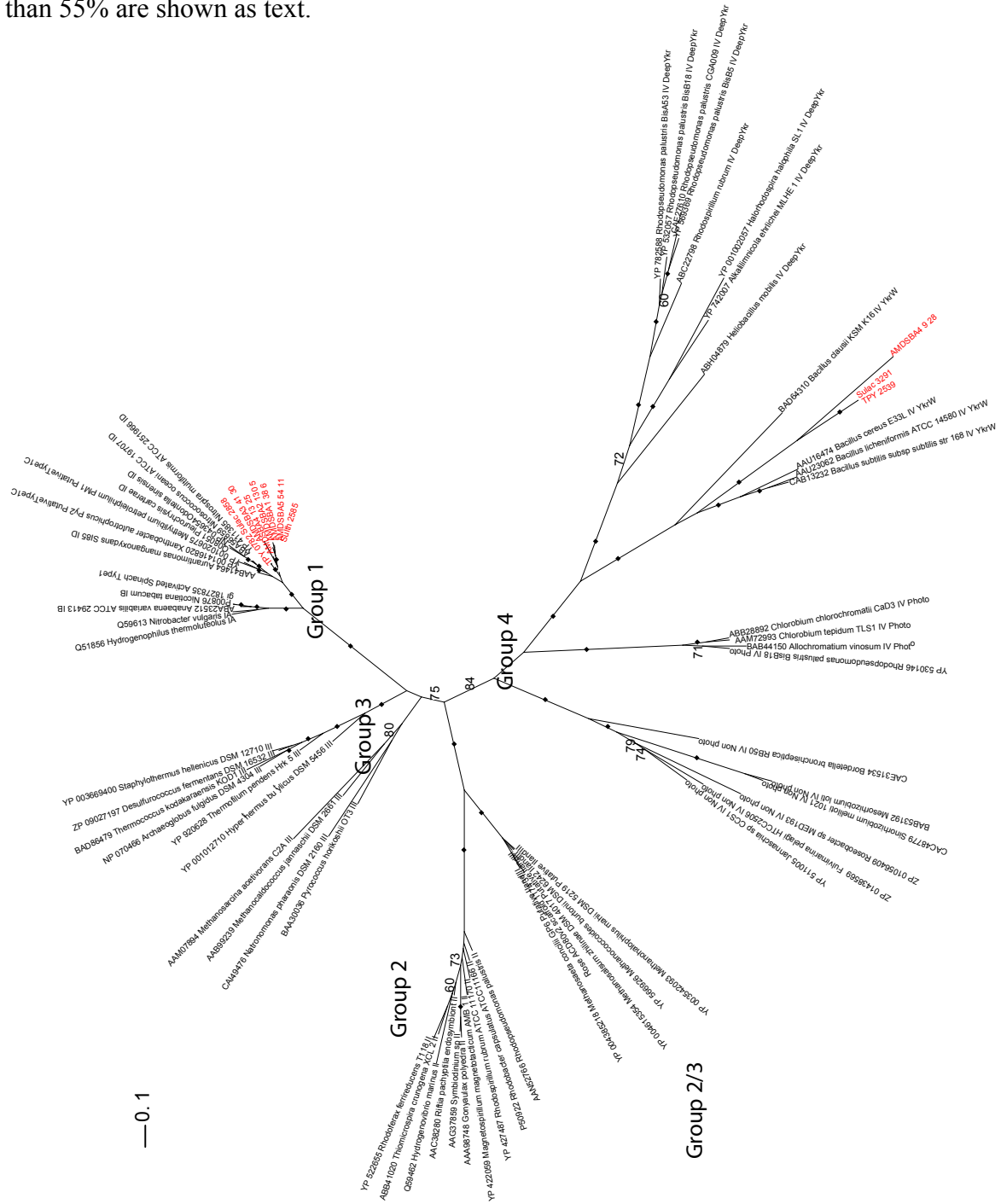


Figure 3.13 (next page) Reconstruction of central metabolism of *Sulfobacillus* species. Dark bordered boxes represent enzymes or enzyme complexes found in all species, while boxes bordered by dashed lines represent those present in only a subset. Full gene information can be found in Supplementary Table 3.1. Colored circles (previously published genomes) and triangles (AMDS genomes) represent which organisms those enzymes are found in, as indicated in the figure legend. CODH, carbon monoxide dehydrogenase complexes; DHA-P, dihydroxyacetone phosphate; Fdh, formate dehydrogenase; G3P, glyceraldehyde-3-phosphate; Hdr, heterodisulfide reductase complexes; KDG, 2-keto-3-deoxygluconate; KDPG, 2-keto-3-deoxygluconate 6-phosphate; SQR, sulfide:quinone oxidoreductase; SQOR, succinate:quinone oxidoreductase; SR, sulfur reductase; TTH, tetrathionate hydrolase; TQO, thiosulfate:quinone oxidoreductase; TTR, tetrathionate reductase;

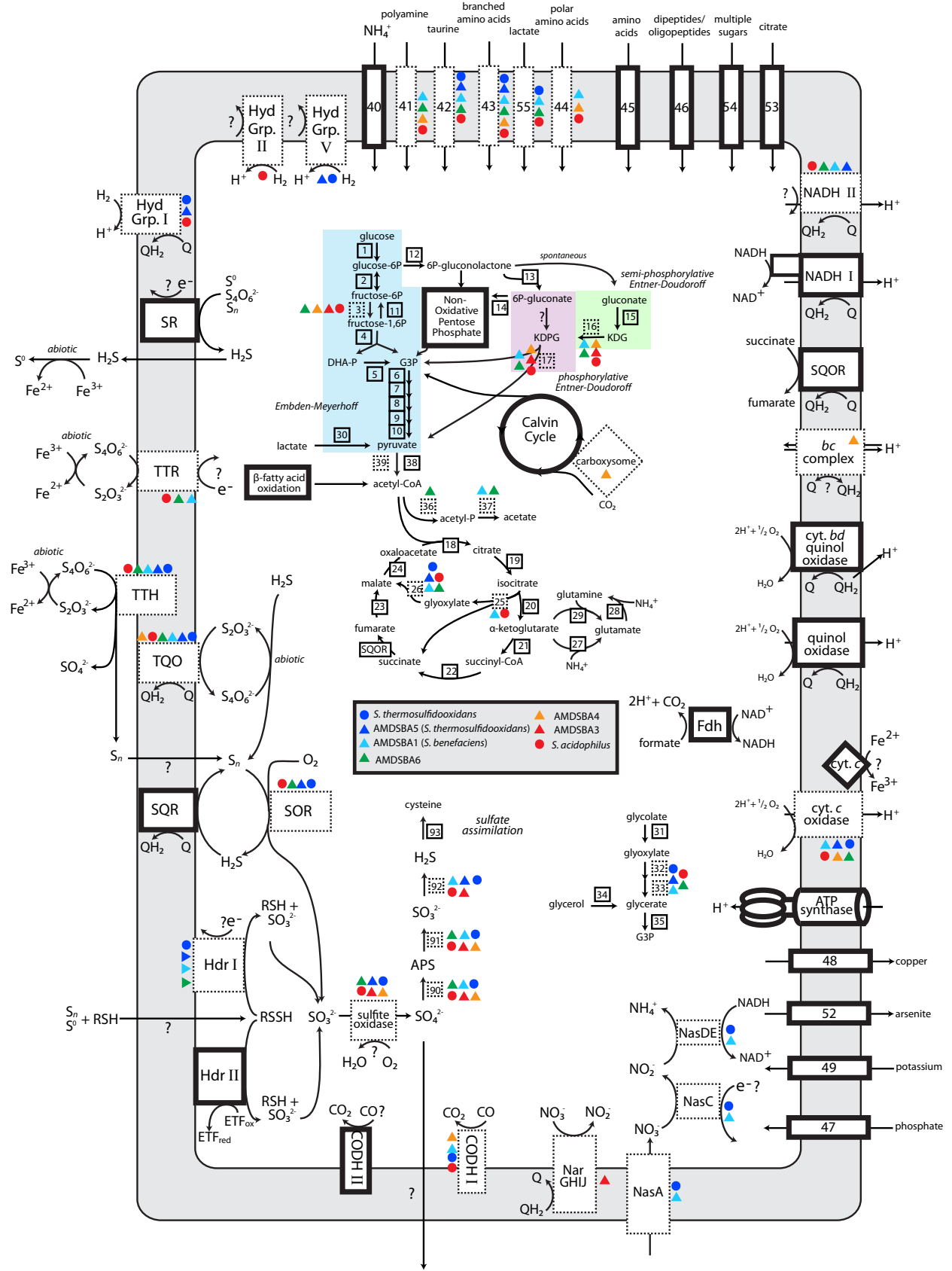


Table 3.1 General statistics of *Sulfobacillus* genomes

Name	Genome Size (Mbp)	GC Content (%)	Coverage (X)	Scaffolds	Protein Coding Genes	Reference
AMDSBA1	4.56	52.19	258	155	4629	this study
AMDSBA2	3.07	52.37	10	409	3312	this study
AMDSBA3	3.68	55.54	113	129	3785	this study
AMDSBA4	4.11	51.67	25	179	4254	this study
AMDSBA5	3.65	49.15	92	161	3870	this study
<i>S. thermosulfidooxidans</i> AT-1 (DSM 9293)	3.86	49.64	N/D	2	3875	IMG JGI
<i>S. thermosulfidooxidans</i> 'Cutipay'	3.86	49.30	116	35	3600	Travisany 2012
<i>S. acidophilus</i> NAL ^T (DSM 10332)	3.56	56.75	168	1	3626	Anderson 2012
<i>S. acidophilus</i> 'TPY'	3.55	56.70	26	1	3770	Li 2011

Table 3.2 16S rRNA gene similarity (across 1,124 aligned positions) for published and reconstructed *Sulfobacillus* species

	<i>Sulfobacillus thermotolerans</i> Kr1 NR043568	AMDSBA2	AMDSBA4	AMDSBA3	AMDSBA5
<i>Sulfobacillus thermotolerans</i> AT-1 (DSM 9293)	-	-	-	-	-
<i>S. thermosulfidoxidans</i> N1 NR042730	93.8	97.6	-	-	-
<i>Sulfobacillus sibiricus</i> N1 NR044352	94.2	97.8	99	-	-
<i>S. thermosulfidoxidans</i> AT-1 (DSM 9293)	94.6	96.8	97.9	98	-
<i>Sulfobacillus thermotolerans</i> Kr1 NR043568	94.6	97.1	98.5	99.2	98.6
AMDSBA2	94.3	96.4	96.5	97	96.8
AMDSBA4	97.2	94.9	94.8	95.1	95.6
AMDSBA3	94	97.9	98.8	99.7	97.7
AMDSBA5	93.7	100	97.6	97.8	96.8
AMDSBA1	-	-	-	-	-

Chapter 4

Metabolites associated with adaptation of microorganisms to an acidophilic, metal-rich environment identified by stable isotope enabled metabolomics

Abstract

Microorganisms grow in a remarkable range of extreme conditions. Environmental transcriptomic and proteomic studies have highlighted metabolic pathways active in extremophilic communities. However, metabolites directly linked to their physiology are less well-defined because metabolomics methods lag behind other ‘omics’ technologies due to a wide range of experimental complexities often associated with the environmental matrix. We identified key metabolites associated with acidophilic and metal-tolerant microorganisms using stable isotope labeling coupled with untargeted, high resolution mass spectrometry. We observed >3,500 metabolic features in biofilms growing in pH ~0.9 acid mine drainage solutions containing millimolar concentrations of iron, sulfate, zinc, copper, and arsenic. Stable isotope labeling improved chemical formula prediction by >50% for larger metabolites (>250 atomic mass units), many of which were unrepresented in metabolic databases and may represent novel compounds. Taurine and hydroxyectoine were identified and likely provide protection from osmotic stress in the biofilms. Community genomic, transcriptomic and proteomic data implicate fungi in taurine metabolism. *Leptospirillum* group II bacteria decrease production of ectoine and hydroxyectoine as biofilms mature, suggesting that biofilm structure provides some resistance to high metal and proton concentrations. The combination of taurine, ectoine, and hydroxyectoine may also constitute a sulfur, nitrogen, and carbon currency in the communities.

Introduction

Over the past decade, metagenomic approaches have illuminated the metabolic potential of communities without the need for cultivation and isolation (Tyson et al. 2004; Breitbart et al. 2002). Leveraging the genomic context uncovered by this method, community proteomics and transcriptomics can provide insight into the potential function of coexisting microorganisms *in situ*. However, these analyses are blind to the flux of small molecule metabolites that are foundational to the physiological or phenotypic state of an organism. Metabolomic measurements can bring to light the key intra- and extra-cellular metabolites involved in cellular processes such as ion homeostasis, redox status, nutrient cycling, energetics, and cell-cell signaling [e.g., (Wasser et al. 1996; Kristal et al. 1998)]. By capturing relative sizes of the metabolite pools, metabolomics is a reflection of the net expression of many genes, pathways, and processes.

Metabolomics studies may prove particularly useful in studying adaptation to extreme environments since metabolites essential to survival in these environments may be relatively abundant (e.g. for ion homeostasis). Yet, studies of metabolites have typically targeted specific molecules of interest from isolated organisms. Few studies have leveraged untargeted, high-throughput metabolomics techniques to study microbial adaptation [e.g., adaptation of *Pyrococcus furiosus* to temperature stress (Trauger et al. 2008) and *Streptomyces coelicolor* to salt stress (Trauger et al. 2008)].

Traditional physiological studies of microbial isolates have defined many adaptation mechanisms to extreme conditions, including improved membrane selectivity and stability, detoxification, enhancement of cellular repair capabilities, and alteration of macromolecular

structures. However, adaptation of organisms to their environments in part relates to behavior within a community context—a facet that is not captured in typical laboratory based studies of isolates but is captured by community metabolomics studies. The abundance of some compounds (e.g., sugars and amino acids) results from uptake, consumption, and excretion by many different organisms, thus the overall concentration reflects the net metabolic state of the community. Untargeted metabolomics has tremendous potential for hypothesis generation in microbial communities since it provides a direct biochemical observation of the community metabolism. However, only a few studies [e.g., (Halter et al. 2012)] have used untargeted metabolomics to study adaptation to environmental challenges in a community context. This is in part because metabolomics methods are still challenging (relative to the more standardized *omics* approaches such as transcriptomics) due to a wide range of experimental complexities often associated with the environmental matrix (e.g., abundant salt can result in extensive experimental artifacts).

To investigate community-level adaptations to the simultaneous challenges of high proton and metal concentrations, we examined the metabolome of microbial biofilm communities in an acid mine drainage (AMD) environment (Richmond Mine, Iron Mountain, California). Previous research in the system demonstrated that proteins and metabolite features exhibited correlative patterns reflective of functional differentiation of bacterial species (Wilmes et al. 2010), suggesting that combining *omics* approaches may prove useful in defining adaptation strategies specific to particular groups of organisms. Microbial biofilms found at Richmond Mine grow at low pH (typically 0.5–1.2) and elevated temperature (30–56°C) in solutions containing millimolar concentrations of sulfate, iron, zinc, copper, and arsenic (Druschel et al. 2004)—conditions that together make untargeted metabolomics extremely challenging. Here, we used a combination of stable isotope probing and untargeted metabolomics to facilitate the identification of metabolites from *in situ* and laboratory-cultivated AMD microbial biofilms. We integrated metabolite identifications with previously acquired genomic and proteomic data to elucidate adaptation to acidophilic and metal-rich conditions on the metabolic level.

Methods

Sample Collection

AMD biofilms from Iron Mountain Mine (near Redding, California) were flash-frozen on-site in a dry ice/ethanol bath, and then transferred to –80°C upon return to the laboratory. Laboratory grown biofilms were cultured in bioreactors as previously described (Belnap et al. 2011). In addition to the traditional media, AMD biofilms were also grown with ¹⁵N-ammonium sulfate. The flow rate of the bioreactors was approximately 225 µl per minute. Biofilms were harvested after ~4-7 weeks of growth and frozen at –80°C for use in metabolite analysis.

Fluorescence in situ hybridization

Fluorescence *in situ* hybridization (FISH) was carried out on fixed (4% paraformaldehyde) AMD biofilm samples as described previously (Amann et al. 1995; Bond & Banfield 2001). For estimation of abundance, cells were counted in three replicate fields of view (average of 888 total cells counted per probe) and converted to a percentage of the total cell count found using the general nucleic acid stain 4',6-diamidino-2-phenylindole (DAPI). Oligonucleotide probes used in this study for identification of individual species and groups were as follows: EUBMIX (all Bacteria); ARC915 (all Archaea); EUKMIX (all Eukaryotes); LF655 (all *Leptospirillum* bacteria); LF1252 (*Leptospirillum* group III bacteria); L2UBA288 (*Leptospirillum* group II

UBA-genotype); L2CG288 (*Leptospirillum* group II 5-way genotype); and SUL230 [*Sulfobacillus* spp.; (Miller et al. 2011)].

Metabolite Extraction

For metabolomics, three replicates with approximately 75 mg of biomass (wet weight) were used from each sample (natural biofilm, cultured biofilm, and ¹⁵N-labeled-cultured biofilm) and were extracted in 700 µl of methanol:isopropanol:water (3:3:2 v/v/v). Metabolites were also extracted from pelleted cells of a pure culture of *Acidomyces richmondensis* (purity confirmed by microscopy and negative PCR amplification of archaeal and bacterial 16S rRNA genes). The frozen samples were lyophilized for 12-24 hours, and then bead beat for 4 seconds (with a steel ball). 700 µl of extraction buffer (3:3:2 methanol:isopropanol:water) was added to each sample. Samples were then bead beat again for 4 seconds, vortexed for 20 seconds, and sonicated for 10 minutes. After sonication, samples were centrifuged for 9 min at 15,000 rpm. The supernatant was transferred to a clean tube, dried using a SpeedVac (typically 4-8 hours), resuspended again in the 3:3:2 extraction buffer and then dried. Dried, pelleted samples were resuspended in either 75% acetonitrile:25% methanol or 3:3:2 methanol:isopropanol:water for HILIC and C18 chromatography, respectively, and then filtered (Millipore, Ultrafree-MC, Durapore PVDF 0.22 µm, catalogue number UFC30GVNB). Extraction blank controls followed this complete procedure precisely, beginning with the addition of extraction solvent to empty tubes.

Liquid Chromatography-Mass Spectrometry Analysis

Samples were analyzed using an Agilent 1200 capillary LC system with an Agilent 6520 dual-ESI-Q-TOF mass spectrometer (± 2 ppm MS mass accuracy and mass resolution of 20,000). An Acquity BEH HILIC column (1.7 µm, 1.0 x 150 mm; catalogue no. 186003459) was used in both positive and negative modes for LC separation with the following liquid chromatography conditions: 2 µL injection volume; 40 µL min⁻¹ flow rate; solvent A, 5 mM ammonium acetate; solvent B, 90% acetonitrile with 5 mM ammonium acetate; timetable: 0 min 100% B, 5 min 100% B, 20.25 min 72% B, 35 min 72% B; 40 min 100% B, 60 min 100% B. MS/MS analysis was performed by specifying a preferred list of precursor ions based on the analysis of profiling data (collision energy, 10 eV, 20 eV, 40 eV).

Reverse phase chromatography was performed using an Agilent Zorbax SB-C18 column (5 µm, 0.5 x 150 mm; catalogue no. 5064-8256) in positive mode. The LC conditions were as follows: 2 µL injection volume; 20 µL min⁻¹ flow rate; solvent A was 0.1% formic acid in water, and solvent B was 0.1% formic acid in acetonitrile; timetable: 0 min 3% B; 3 min 3% B; 5 min 50% B; 30 min 99% B; 40 min 99% B. Columns were washed with 1:1 acetone:isopropanol for 10 min, then re-equilibrated in 3% B for 10 min between analytical runs. MS/MS analysis was performed by specifying a preferred list of precursor ions based on the analysis of profiling data (collision energy, 10 eV, 20 eV, 40 eV).

Data analysis

The total number of features in each dataset was determined using MZmine 2.3 (Pluskal et al. 2010) with the following parameters: 2-20 minute retention time (RT) filter for HILIC dataset; centroid mass detector (noise level 100); chromatogram builder (0.2 minimum time span, 500 minimum height, 500 ppm m/z tolerance); chromatogram deconvolution (noise amplitude, 500 minimum peak height, 0.2 minimum peak duration, amplitude of 10); gap filling

(same m/z and RT range gap filler, 100 ppm m/z tolerance); duplicate peak filtering (500 ppm m/z tolerance, 2 min. absolute RT tolerance).

Raw datasets were preprocessed using the MathDAMP package (Baran et al. 2006) to generate three-way comparison visualizations (Baran et al. 2007) of the ^{15}N , ^{14}N , and extraction blank controls. From this visualization, we discriminated compounds of biological origin and identified their approximate molecular mass, retention time, and the number of nitrogen atoms (atomic mass unit shifts between ^{14}N - and ^{15}N -labeled samples). Precise m/z values and retention times were determined using Agilent MassHunter Workstation Software Qualitative Analysis (version B.03.01). Additional low abundance features were found by overlaying MS spectra from ^{14}N and ^{15}N -labeled samples at similar elution times in MassHunter software. Features were grouped according to retention time, and redundant features corresponding to common adducts (positive mode: $[\text{M} + \text{H}]^+$, $[\text{M} + \text{Na}]^+$, $[\text{M} + \text{NH}_4]^+$, $[\text{M} + \text{Fe}]^{++}$, $[\text{M} + \text{Ca}]^{++}$, $[\text{M} + \text{ACN} + \text{H}]^+$; negative mode: $[\text{M} - \text{H}]^-$ $[\text{M} + \text{Na} - 2\text{H}]^-$) and fragments (e.g., $[\text{M} - n\text{H}_2\text{O} + \text{H}]^+$) were identified. $[\text{M} + \text{H}]^+$ ions extracted from this curated list, and these ions were inspected in MassHunter to predict chemical formulae. Chemical formulae were assigned if they had a single high (>85) molecular formula score consistent with the number of measured nitrogen atoms (MassHunter molecular formula scores are out of 100, and use a proprietary algorithm that take into account monoisotopic mass, isotope spacing, and isotope abundances). Peak heights of MS spectra were used to estimate the abundance of taurine in each sample. No other metabolites were quantified due to an extraction artifact (esterification of carboxylic acids resulting from the low pH of our samples and alcohol extraction) that may have affected the peak heights of metabolites containing carboxylic acids.

Community genomic analysis

Bacterial, archaeal, and eukaryal genes involved in taurine and ectoine metabolism were identified from a well curated metagenomic and genomic dataset containing nearly 80,000 gene sequences and capturing essentially all organisms representing more than a few percent of the community. MycoCosm (Grigoriev et al. 2012) was used to evaluate gene annotations for the *Acidomyces richmondensis* fungal genome. Reciprocal blast searches against the KEGG database were conducted using the KAAS server (Moriya et al. 2007). Transcript and protein abundances of *Acidomyces richmondensis* were determined by Miller et al., (in prep). Abundances of proteins involved in ectoine and hydroxyectoine synthesis were determined by Mueller et al., (Mueller et al. 2011).

Results and Discussion

Community composition of AMD biofilms with different growth strategies

For metabolomic analyses presented here, biofilm samples were collected from the air-AMD solution interface of the AB-muck site within the Richmond mine on July 15, 2011 (hereafter referred to as the Mine biofilm sample). The Mine biofilm was categorized as late developmental stage, based primarily on biofilm thickness. AMD biofilms were also grown in laboratory bioreactors using the Mine biofilm as inoculum for three different lengths of cultivation time: 26, 35, and 51 days (hereafter identified as BR-26days, BR-35days, and BR-51days indicating bioreactor growth and length of cultivation). In all cases, the cultivated biofilms were thick and well developed at the time of sampling. BR-26days and BR-35days were grown with ^{15}N -labeled ammonium sulfate (resulting in ^{15}N -labeled nitrogen atoms in the

biofilm metabolites) in order to identify the number of nitrogen atoms in each metabolite and also to confirm biological origin.

The community composition of these biofilm samples was determined using fluorescence *in situ* hybridization (FISH) targeting identification of broad phylogenetic groups, as well as individual species and strains (Figure 4.1). Insufficient biomass precluded FISH analyses on BR-35days. Prior studies showed that biofilms in the Richmond mine are dominated by *Leptospirillum* group II, a chemoautotrophic iron-oxidizing bacteria (Tyson et al. 2004; Bond, Druschel, et al. 2000a). Genomic reconstructions revealed two distinct strains of *Leptospirillum* group II referred to as the 5-way and UBA genotypes (Lo et al. 2007; Simmons et al. 2008). In this study, the *Leptospirillum* group II 5-way genotype was more abundant than the UBA genotype in both bioreactors. Conversely, nearly all of the Bacteria in the Mine biofilm belonged to the *Leptospirillum* group II UBA genotype, which is consistent with prior studies showing that the UBA genotype typically predominates over the 5-way genotype in the mine (Denef, Kalnejais, et al. 2010a).

Previous studies have shown that as the biofilms mature and thicken, they also diversify with increasing proportions of *Leptospirillum* group III bacteria and Archaea, as well as other low abundance taxa from the Eukarya, Firmicute and Actinobacteria lineages (Tyson et al. 2004; Dick et al. 2009). The bioreactor biofilms had a much higher percentage of *Leptospirillum* group III bacteria than seen in the Mine biofilm (22-32% compared to only 2% in the mine). Geochemical and FISH data collected previously in the mine suggest a positive correlation between the abundance of *Leptospirillum* group III and ammonium concentrations ($r = 0.96$, $n = 6$), which may suggest that ammonium concentrations in the bioreactors (2 mM compared to an average concentration of 177 μM in the mine) favors *Leptospirillum* group III. Recent studies also indicate that high ferric iron concentrations in the bioreactors relative to mine solutions may also select for *Leptospirillum* group III (Ma et al., in review).

The abundance of Archaea varied between the mine and bioreactor biofilms. Archaea made up 29-35% of the communities in the BR-26days and Mine biofilms. Conversely, BR-51days had much higher percentages of Archaea (61%), which may be the result of longer cultivation time compared to BR-26days. The low-abundance community member *Sulfobacillus* was found in all samples (0.2-1.2%). Eukaryotes were not identified by FISH but are not necessarily absent from the biofilms; the inherent heterogeneity of the biofilms and patchy eukaryal distribution may prohibit visualization in microscopy with a random sampling design.

Detection and chemical formula prediction of metabolites found in AMD enabled by stable isotope labeling

Metabolites were extracted from natural and cultivated AMD biofilms (Mine, BR-26days, BR-35days, and BR-51days) and analyzed using liquid-chromatography coupled to high-resolution mass spectrometry (LC-MS). Both hydrophilic interaction liquid chromatography (HILIC) and C18 reversed phase (RP) columns were used to separate polar and non-polar organics, respectively. LC-ESI-MS (LC-MS with electrospray ionization) is one of the most widely used metabolomic platforms given its versatility in separation techniques coupled with the wide range of compounds that can be desorbed/ionized (Garcia et al. 2008). Generally, LC-MS approaches fall into one of two categories: (1) targeted analyses, which aim to quantify changes within a defined set of known metabolites diagnostic of a given phenotype of interest; and (2) untargeted analyses, which seek to discern both novel and previously characterized metabolites (Oldiges et al. 2007; Fiehn 2001). An untargeted LC-MS approach was used in the

current study, rather than the commonly used GC-MS approach, to maximize the diversity of metabolites detected.

More than 3500 raw features (ions with unique retention time and mass-to-charge ratio combinations) were identified in each of the RP and HILIC datasets. These uncurated data, however, included features associated with background chromatography noise, features also found in extraction blanks, as well as multiple adducts and fragment ions of the same compounds. In order to obtain the highest quality data possible, we used a strict manual curation strategy to identify compounds of interest in our samples and prevent errors associated with automated peak detection, deconvolution, and alignment. Three-way visualization plots (Baran et al. 2007) of the ^{14}N -biofilm, ^{15}N -biofilm, and extraction blank datasets (available: <http://geomicrobiology.berkeley.edu/pages/metabolites.html>) were used to narrow down the large list of raw features to “pure spectra,” that is the spectra that would likely result from a pure compound within the biological matrix. From this visualization, as well as manual observation of very abundant peaks, we generated a list of 241 likely parent ions features (the highest intensity feature from a pure spectrum) from the RP and HILIC analyses (Table 4.1).

Eighty parent compounds (56 in RP and 24 in HILIC; Table 4.1) were confirmed after grouping co-eluting features and identifying common adducts, fragment ions, and neutral losses including those associated with esterification of carboxylic acids (presumably as a result of extraction buffer and residual acidic AMD). The number of metabolites found here is consistent with other untargeted LC-MS metabolomic studies from complex, natural biological matrices (Baran et al. 2010; 2011; van der Werf et al. 2007; Brauer et al. 2006). 48% of the reverse phase (RP) metabolites and 79% of the HILIC metabolites contained at least one nitrogen atom (determined by stable isotope labeling), providing a strong level of certainty that the compounds are of biological origin. Mass spectra of all compounds were manually visualized to ensure absence in extraction blanks.

The use of stable isotopes greatly aids in determining the chemical formulae of unknown metabolites by constraining the possible elemental composition (Baran et al. 2011; Hegeman et al. 2007; Rodgers et al. 2000). Here, using the number of nitrogen atoms informed by stable isotope labeling, chemical formulae were determined for 38 of the RP features and 18 of the HILIC features (Table 4.2). The nitrogen labeling method also informs the formulae of compounds without nitrogen, as their chemical formulae are constrained by the “zero” nitrogen count. Generally, high-mass accuracy (<5 ppm) mass spectrometers can confidently assign unique chemical formula for features under ~200-250 atomic mass units (amu) (Hegeman et al. 2007; Bowen et al. 2011). Indeed, we were able to confidently assign chemical formulae for most features <250 amu without the assistance of nitrogen labeling. For those thirty-seven features above 250 amu in Table 4.2, nitrogen labeling allowed us to confidently assign chemical formulae to twenty of them, twice as many formulae as were possible with spectral information alone. Features for which chemical formulae could not be determined were either of high m/z , low spectral quality, or had an unidentified adduct.

Metabolite annotation and identification

We obtained MS/MS spectra at collision energies of 10, 20 and 40 eV on all features with sufficient peak heights (available: <http://geomicrobiology.berkeley.edu/pages/metabolites.html>). The chemical formulae and MS/MS data were matched with metabolites in online databases (MetaCyc, KEGG, MassBank, and METLIN) (Table 4.2). MS/MS data from more than 90% of the features had no match to metabolites in MS/MS databases (Massbank and METLIN), and as

per protocols established by the Chemical Analysis Working Group of the Metabolomics Standards Initiative, these metabolites are classified as “unknown” (Dunn et al. 2012; Sumner et al. 2007). While these data may speak to the novelty of some AMD metabolites, it must also be noted that these databases rely on commercially available standards, which are estimated to represent only half of all biological metabolites (Garcia et al. 2008).

From this analysis we found three metabolites that were particularly interesting and warranted further investigation within the context of the biofilm community: 1) phosphatidylethanolamine lipids, 2) taurine, and 3) hydroxyectoine. Unusual *lyso* phosphatidylethanolamine (PE) lipids and methylated derivatives previously identified in the Richmond mine (Fischer et al. 2011) were also found in the mine and bioreactor biofilms presented here. Fischer et al., (Fischer et al. 2011) suggested a link between these lipids and the *Leptospirillum* group II UBA genotype based on correlations of lipid and proteome abundance patterns. Interestingly, we found features consistent with some of these same lipids (same *m/z* and retention time for 454.294 and 480.309) in pure cultures of *Acidomyces richmondensis*, a fungus known to be abundant in the mine and often the dominant eukaryal species (Baker et al. 2009; 2004). The *A. richmondensis* genome contained 8 genes predicted to be involved in the synthesis, methylation, and binding of PEs, some of which were expressed in community transcriptomic and proteomic data (Miller et al., in prep). Together, these results show that fungi and bacteria may both be involved in the metabolism of these lipids. Fischer et al., (Fischer et al. 2011) suggested that these lipids may prevent uptake of toxic levels of iron cations in AMD biofilms.

Taurine: Metabolic interactions and abundance within the AMD community

Among the metabolites present in the AMD biofilms, we identified taurine (2-aminoethanesulfonic acid) by comparing MS/MS spectra with a taurine chemical standard (Figure 4.2). Taurine is a phylogenetically ancient compound (Huxtable 1992) and is involved in numerous physiological functions across disparate forms of life, including membrane stabilization, stimulation of glycolysis and glycogenesis, regulation of phosphorylation, and antioxidation [(Huxtable 1992; Levis & Park 2003) and references therein]. Some microbes can use taurine as an exclusive source of carbon, nitrogen and sulfur [(Stapley & Starkey 1970; Denger et al. 2006) and references therein]. Taurine is a particularly effective osmoregulator and is used as a compatible solute by a variety of microorganisms [e.g., (Huxtable 1992; Cayley et al. 1991) and references therein]. Compatible solutes (generally very soluble, low molecular weight organic molecules) can be accumulated in the cytoplasm as a mechanism for coping with hyperosmotic stress. Compatible solutes can also protect proteins, nucleic acids and membranes from the harmful effects of heat, freezing, drying, and oxygen radicals [(da Costa et al. 1998; Lentzen & Schwarz 2006; Pastor et al. 2010) and references therein].

Although many of the potential roles for taurine are relevant, its properties as a compatible solute may be particularly useful in microbial adaptation to the high ionic strength waters within the Richmond Mine (Druschel et al. 2004). We explored genomic sequences of ~20 AMD biofilm community members (including bacteria, archaea, and fungi) to determine which organisms may produce taurine (Figure 4.3). This dataset contains nearly 80,000 gene sequences and captures essentially all organisms representing more than a few percent of the community. Reciprocal BLAST searches against the KEGG database indicated that archaea and bacteria in the AMD communities are unable to generate taurine (no archaea or bacteria are known to synthesize taurine). The only archaeal or bacterial enzyme potentially involved in

taurine biosynthesis was a glutamate decarboxylase (E.C. 4.1.1.15), which has broad functionality in several different metabolic pathways.

We evaluated the likelihood for eukaryotic taurine biosynthesis using the genome of the dominant fungus, *A. richmondensis* (Figure 4.3). The *A. richmondensis* genome encodes cysteine dioxygenase (EC:1.13.11.20) and glutamate decarboxylase (EC: 4.1.1.15) genes involved in two routes of taurine metabolism. Enzymes mediating the oxidation of hypotaurine to taurine and 3-sulfino-L-alanine to L-cysteate were not evident, however, it has been shown that these reactions may occur non-enzymatically [(Fontana et al. 2006; Coloso et al. 2006) and references therein]. Ferric iron, found in high concentrations in the mine, may act as a chemical oxidant of these compounds, thereby completing the pathways of taurine production in *A. richmondensis*. Efforts to identify taurine in a pure culture of *A. richmondensis* failed, although culture conditions may not have favored taurine production.

We also assessed the potential for taurine degradation using community genomic sequence data (Figure 4.3). While bacteria and archaea in the mine encode genes for some enzymes involved in taurine degradation pathways, they do not appear to have a definitive or complete mechanism for the breakdown of taurine. The most likely candidate route of degradation is via gamma-glutamyltranspeptidase (EC 2.3.2.2) found in *Sulfobacillus* and three archaeal species (Ferroplasma, Cplasma, and Gplasma); however, this enzyme has broad activity and is also involved in cyanoamino acid, glutathione, and arachidonic acid metabolism. *Sulfobacillus* also has genes encoding two taurine transport proteins (TauA and TauC), suggesting taurine may indeed have a biological role in this organism. Taurine diffuses slowly through cell membranes and taurine biotransformation enzymes are usually soluble and intracellular, so transport of taurine into the bacterial cell is required for utilization of the compound (Huxtable 1992; Cook & Denger 2006). Cells responding to hyperosmolar conditions can increase intracellular taurine content via active transport of taurine into the cell.

Genomic evidence suggests that the fungal species *A. richmondensis* is capable of degrading taurine via taurine catabolism dioxygenase TauD/TfdA enzymes. TauD is a dioxygenase that converts taurine to sulfite and aminoacetaldehyde, with reaction requirements of oxygen, Fe^{2+} , and α -ketoglutarate (Eichhorn 1997). In *Escherichia coli*, the *tauD* gene is expressed only under conditions of sulfate starvation (Eichhorn 1997; van der Ploeg et al. 1996). There are 12 copies of the taurine catabolism dioxygenase *tauD/tfdA* genes in the *A. richmondensis* genome. Interestingly, transcripts of all 12 *tauD/tfdA* genes were detected in a fungal streamer biofilm community from the mine (Miller et al., in prep). Two of these transcripts were relatively abundant in the transcriptome (ranked in the top 1500 transcripts out of 10,305 total genes) and their proteins were also detected in a community metaproteome (Miller et al., in prep). Other Dothidiomycetes (the fungal class including *A. richmondensis*) genomes contain between one and 10 copies of taurine catabolism dioxygenase genes per genome (based on BLAST searches).

Given the genomic potential for taurine biosynthesis and likelihood for degradation in the AMD biofilms, we evaluated the abundance of taurine across the different growth conditions (natural mine biofilms and biofilms grown in bioreactors for 26, 35, and 51 days) based on peak heights in the MS spectra (Figure 4.4). Taurine concentrations were an order of magnitude higher in all three bioreactor biofilms than in the mine biofilm. This discrepancy is likely explained by different biogeochemical conditions in the mine and bioreactor biofilms. It is possible that the bioreactor communities generate more taurine relative to the mine communities or, equally possible, that more taurine is consumed in the mine. Metagenomic evidence

implicates *A. richmondensis* as the dominant organism involved in taurine biosynthesis and degradation. The low-abundance community member *Sulfobacillus* may also have the potential for taurine consumption; while FISH shows higher numbers of *Sulfobacillus* in the mine biofilms, the percentage of the total community is only on the order of 1%.

Hydroxyectoine: Metabolic interactions within the AMD community

We identified hydroxyectoine (confirmed with chemical standards and MS/MS spectra; Figure 4.5) and possibly ectoine (correct mass and chemical formula but incomplete MS/MS data) in natural and cultivated AMD biofilms and were interested in their role as compatible solutes used in adaptation to hyperosmotic stress. Ectoine biosynthesis occurs in three enzymatic steps: (1) L-diaminobutyric acid transaminase (EctB or ThpB) converts L-aspartate-beta-semialdehyde into L-diaminobutyric acid; (2) acetylation to N- γ -acetyldiaminobutyric acid occurs via L-diaminobutyric acid acetyl transferase (EctA or ThpA); and (3) cyclic condensation then leads to the formation of ectoine through ectoine synthase (EctC or ThpC) (Caspi et al. 2010; Louis & Erwin A Galinski 1997; Canovas et al. 1997). Hydroxyectoine is primarily generated through the hydroxylation of ectoine by ectoine hydroxylase (EctD or ThpD) (Garcia-Estepa et al. 2006; Prabhu et al. 2004). Compared to ectoine, hydroxyectoine can confer additional protective properties against heat stress (Garcia-Estepa et al. 2006) and freeze-drying (Lippert & Erwin A Galinski 1992).

Complete ectoine and hydroxyectoine biosynthesis pathways have been identified previously in one archaeal genome, *Nitrosopumilus maritimus* (Walker et al. 2010), and in over 50 bacterial genomes with particular representation among the α - and γ -Proteobacteria and Actinobacteria [(Pastor et al. 2010) and references therein]. In the AMD biofilms, both the 5-way and UBA genotypes of *Leptospirillum* group II have all of the genes necessary for ectoine and hydroxyectoine biosynthesis (ectABCD) and their protein products have been identified by community proteomics (Goltsman et al. 2009). Previous studies showed high numbers of ectoine synthase proteins in *Leptospirillum* group II grown under non-optimized culture conditions; however, ectoine synthases were found in similar levels to the natural biofilm upon optimization of the culturing media (Belnap et al. 2009). *Leptospirillum* group II EctB proteins were significantly more abundant in low pH cultures (pH 0.85 versus pH 1.45), suggesting greater osmotic stress during growth in more acidic solutions (Belnap et al. 2011). Interestingly, *Leptospirillum* group II was the first acidophilic bacterium described with a complete pathway for biosynthesis of ectoine and hydroxyectoine (Aliaga Goltsman et al. 2013). Other AMD community members also have genes in the ectoine biosynthesis pathway. Genes encoding *ectB* were found in some archaea (*Cplasma*, *Eplasma*, and *Ferroplasma*) and *Sulfobacillus* has both *ectA* and *ectB* genes; however, ectoine biosynthesis by these organisms cannot be confirmed since the complete Ect operon was not found.

We evaluated the abundance of ectoine and hydroxyectoine biosynthesis proteins in *Leptospirillum* group II bacteria across different growth stages of AMD biofilms (early, mid, and late growth stages; Figure 4.6), using previously acquired quantitative proteomics data (Mueller et al. 2011). EctA proteins were not found in any of the samples, which may suggest low abundance. Protein abundance of EctB, EctC, and EctD generally decreased with biofilm development suggesting that more ectoine and hydroxyectoine are produced in the early growth stages. In early biofilm development, the organisms may have greater exposure to the AMD solution because the biofilm is still thin and friable and contains less extracellular polymeric substances (EPS) (Jiao et al. 2010). EPS has been reported to provide protection from a variety

of environmental stresses, including osmotic shock [e.g., (Davey & O'toole 2000) and references therein]. Thus, with less protection in early biofilm development, *Leptospirillum* group II bacteria may produce more ectoine and hydroxyectoine in order to cope with greater exposure to the high ionic strength AMD solution.

Some organisms are capable of concurrently using ectoine as an osmoprotectant and as an energy and carbon substrate (Jebbar et al. 2005; Schwibbert et al. 2010; Vargas et al. 2006). Genes involved in ectoine metabolism have been identified in *Sinorhizobium meliloti* (eutABCD) (Jebbar et al. 2005) and *Halomonas elongata* (doeABCD) (Schwibbert et al. 2010). In the AMD biofilms, some of these genes involved in ectoine utilization have been identified in *Sulfobacillus* and Firmicute genomes, but not the complete operons.

Conclusion

We used an untargeted approach based on stable isotope labeling coupled with high resolution mass spectrometry to uncover metabolites within natural and cultivated AMD microbial communities. We confirmed the identification of taurine and hydroxyectoine and used community proteogenomic data to determine which organisms are capable of producing or consuming these osmolytes. From this we suggest that to mitigate less protection from EPS in early biofilm development, *Leptospirillum* group II bacteria may produce more ectoine and hydroxyectoine in order to cope with greater exposure to the high ionic strength AMD solution. We determined the specific chemical formulae of many other metabolites that were not identified because they are not present in current metabolic databases and suitable pure compound standards for a defined set of candidate molecules were not available. The set of abundant but not identified compounds could include novel metabolites, with potentially high biological and chemical relevance.

Table 4.1. Comparison of the number of features and metabolites resulting from varying levels of curation.

Analytical column	No. of features after three-way visualization analysis	No. of metabolites after manual curation	No. of metabolites containing nitrogen	No. of metabolites with chemical formula
Agilent Zorbax SB-C18 Column	191	56	27	38
AQUITY UPLC BEH HILIC Column	50	24	19	18

Table 4.2. Acid mine drainage metabolites with assigned chemical formula. All features listed are identified in positive mode.

Number	RT for ¹⁴ N biofilm (min)	RT for ¹⁵ N biofilm (min)	No. of Nitrogens (using SIP)	Measured Mass	Difference from Theoretical Mass (ppm)	Formula	Formula Matches in KEGG	Formula Matches in MassBank	Formula Matches in MetaCyc	Formula Matches in METLIN	MS/MS match in databases
<i>Agilent Zorbax SB-C18 Column</i>											
1	8.39	8.38	1	131.0944	-1.6	C ₆ H ₁₃ NO ₂	6	0	0	6	no
2	8.82	8.84	1	145.0527	-0.3	C ₉ H ₇ NO	7	4	7	8	no
3	8.51	8.50	1	147.0529	-1.6	C ₅ H ₉ NO ₄	10	6	4	11	n/d
4	1.58	1.55	2	158.0691	-0.1	C ₆ H ₁₀ N ₂ O ₃	2	0	1	2	yes*
5	10.36	10.38	1	161.0476	-0.3	C ₉ H ₇ NO ₂	7	3	5	8	n/d
6	8.65	8.64	1	161.0688	0.1	C ₆ H ₁₁ NO ₄	9	5	3	12	no
7	1.29	1.29	2	172.0848	0.2	C ₇ H ₁₂ N ₂ O ₃	1	1	1	5	no
8	8.12	8.09	1	217.1314	0.1	C ₁₀ H ₁₉ NO ₄	1	0	0	2	no
9	28.70	28.37	0	254.2249	1.4	C ₁₆ H ₃₀ O ₂	1	0	0	53	n/d
10	13.91	13.88	0	254.2240	-2.2	C ₁₆ H ₃₀ O ₂	1	0	0	53	yes
11	16.20	16.22	0	254.2240	-2.2	C ₁₆ H ₃₀ O ₂	1	0	0	53	yes
12	25.17	25.05	0	254.2240	-2.2	C ₁₆ H ₃₀ O ₂	1	0	0	53	yes
13	32.51	32.24	0	254.2240	-2.2	C ₁₆ H ₃₀ O ₂	1	0	0	53	n/d
14	8.58	8.60	2	258.1924	-7.4	C ₁₃ H ₂₆ N ₂ O ₃	0	0	0	0	no
15	11.91	11.88	0	268.2404	0.7	C ₁₇ H ₃₂ O ₂	2	0	0	47	no
16	21.19	21.21	0	268.2404	0.7	C ₁₇ H ₃₂ O ₂	2	0	0	47	no
17	8.60	8.62	2	274.1899	2.4	C ₁₃ H ₂₆ N ₂ O ₄	0	0	0	0	n/d
18	14.95	14.92	0	278.1531	4.7	C ₁₆ H ₂₂ O ₄	3	0	1	7	no
19	14.25	14.19	0	280.2402	0.0	C ₁₈ H ₃₂ O ₂	5	0	0	178	n/d
20	9.23	9.22	2	286.2256	-0.1	C ₁₅ H ₃₀ N ₂ O ₃	0	0	0	0	no
21	23.44	23.41	0	320.1992	1.5	C ₁₉ H ₂₈ O ₄	2	0	2	4	no
22	9.59	9.56	0	326.2464	2.2	C ₁₉ H ₃₄ O ₄	0	0	0	6	no
23	10.36	10.38	0	330.2412	1.8	C ₁₈ H ₃₄ O ₅	3	0	1	16	no
24	10.92	10.94	0	330.2412	1.8	C ₁₈ H ₃₄ O ₅	3	0	1	16	no
25	11.47	11.47	0	330.2412	1.8	C ₁₈ H ₃₄ O ₅	3	0	1	16	no
26	18.85	18.90	0	330.2776	1.9	C ₁₉ H ₃₈ O ₄	0	0	2	5	no
27	10.10	10.12	0	358.2356	0.2	C ₁₉ H ₃₄ O ₆	0	0	0	0	no
28	11.16	11.18	0	358.2356	0.2	C ₁₉ H ₃₄ O ₆	0	0	0	0	no
29	30.05	29.75	0	376.2622	2.3	C ₂₃ H ₃₆ O ₄	2	0	0	5	no
30	22.04	22.06	0	406.2717	-0.5	C ₂₄ H ₃₈ O ₅	3	0	0	82	n/d
31	25.95	25.80	0	434.2676	1.8	C ₂₅ H ₃₈ O ₆	0	0	0	4	n/d
32	35.04	34.72	0	446.3407	2.5	C ₂₆ H ₄₆ O ₄	4	0	1	28	no
33	13.07	13.01	1	451.2709	2.3	C ₂₁ H ₄₄ NO ₇ P	0	0	1	5	yes**
34	14.63	14.60	1	453.2864	2.0	C ₂₁ H ₄₄ NO ₇ P	0	0	2	5	yes**
35	13.76	13.73	1	465.2864	1.9	C ₂₂ H ₄₄ NO ₇ P	0	0	0	4	yes**
36	15.50	15.50	1	467.3014	0.5	C ₂₂ H ₄₆ NO ₇ P	0	0	0	7	yes**
37	14.13	14.10	1	479.3020	1.7	C ₂₃ H ₄₆ NO ₇ P	0	0	1	7	yes**
38	15.98	15.98	1	481.3192	5.0	C ₂₃ H ₄₈ NO ₇ P	0	0	2	10	yes**
<i>AQUITY BEH HILIC Column</i>											
39	6.13	6.26	1	115.0637	3.4	C ₅ H ₆ NO ₂	4	3	5	5	yes
40	8.27	8.43	1	125.0147	0.5	C ₂ H ₇ NO ₃ S	1	1	1	1	yes
41	5.85	5.98	1	129.0793	2.7	C ₆ H ₁₁ NO ₂	5	5	7	8	n/d
42	6.61	6.77	1	129.0797	5.8	C ₆ H ₁₁ NO ₂	5	5	7	8	yes
43	6.18	6.18	5	135.0545	0.2	C ₅ H ₅ N ₅	1	2	2	1	yes
44	6.05	6.18	5	151.0495	0.8	C ₅ H ₅ N ₅ O	1	1	2	3	yes
45	17.60	17.63	2	158.0698	4.3	C ₆ H ₁₀ N ₂ O ₃	2	0	1	2	yes*
46	3.04	3.12	1	161.0679	-5.5	C ₆ H ₁₁ NO ₄	9	5	3	12	no
47	14.33	14.43	2	172.0846	-1.0	C ₇ H ₁₂ N ₂ O ₃	0	1	1	5	no
48	19.28	19.34	1	229.0718	1.3	C ₆ H ₁₆ NO ₆ P	0	0	0	0	n/d
49	2.73	2.81	1	231.1478	3.3	C ₁₁ H ₂₁ NO ₄	2	0	0	4	n/d
50	3.85	3.91	1	263.1374	2.0	C ₁₁ H ₂₁ NO ₆	0	0	0	0	no
51	4.85	5.00	0	330.2412	1.8	C ₁₈ H ₃₄ O ₅	3	0	1	0	no
52	11.72	12.41	0	342.1162	0.0	C ₁₂ H ₂₂ O ₁₁	35	5	33	28	yes
53	5.03	5.82	0	342.1162	0.0	C ₁₂ H ₂₂ O ₁₁	35	5	33	28	yes
54	11.52	11.55	2	400.3308	1.8	C ₂₂ H ₄₄ N ₂ O ₄	0	0	0	0	no
55	10.04	9.89	2	428.3617	0.7	C ₂₄ H ₄₆ N ₂ O ₄	0	0	0	0	no
56	7.66	7.46	1	689.4996	0.1	C ₃₇ H ₇₂ NO ₈ P	0	0	0	0	no

*MS/MS data from Kol et al. (2010)

**MS/MS data from Fischer et al. (2011)

n/d indicates incomplete MS/MS

Figure 4.1. Community composition of natural and cultivated AMD communities based on fluorescence *in situ* hybridization.

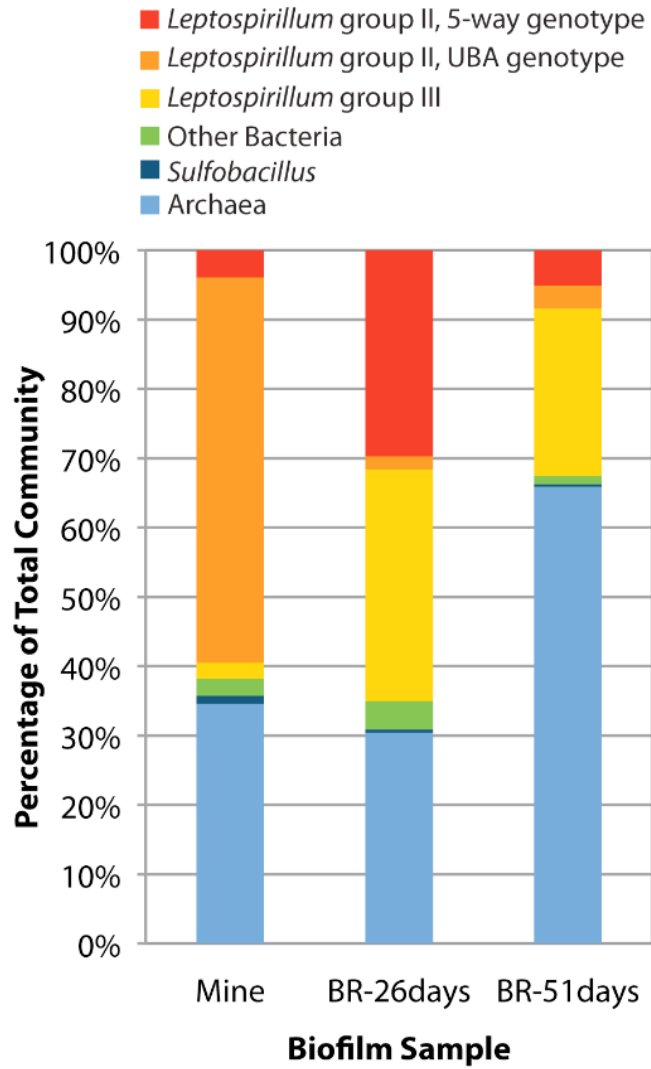


Figure 4.2. Identification of taurine the AMD biofilm using MS/MS comparisons with a chemical reference standard. Taurine MS/MS spectra at three collision energies (10, 20, and 40 eV) shown in the top panels for the AMD biofilm and the bottom panels for the reference standard.

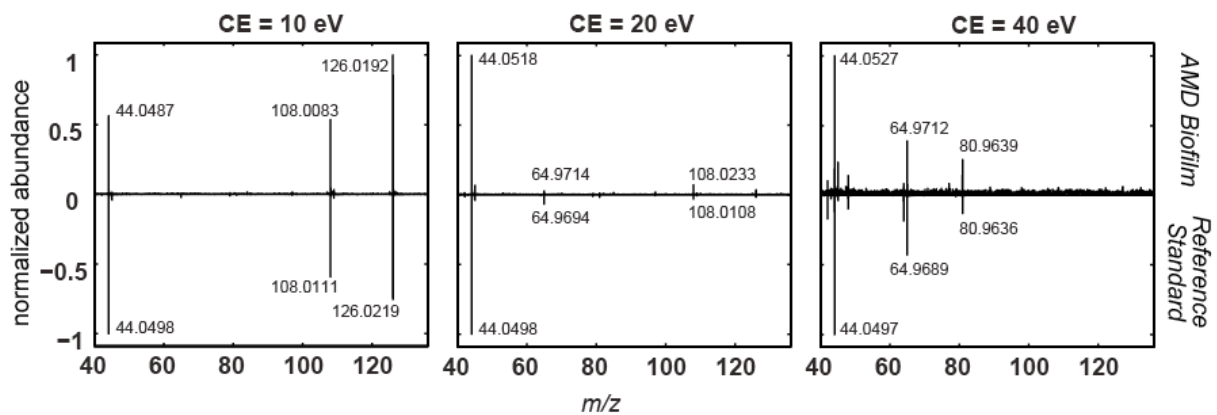


Figure 4.3. Taurine metabolism by prokaryotes and eukaryotes in the AMD biofilms.

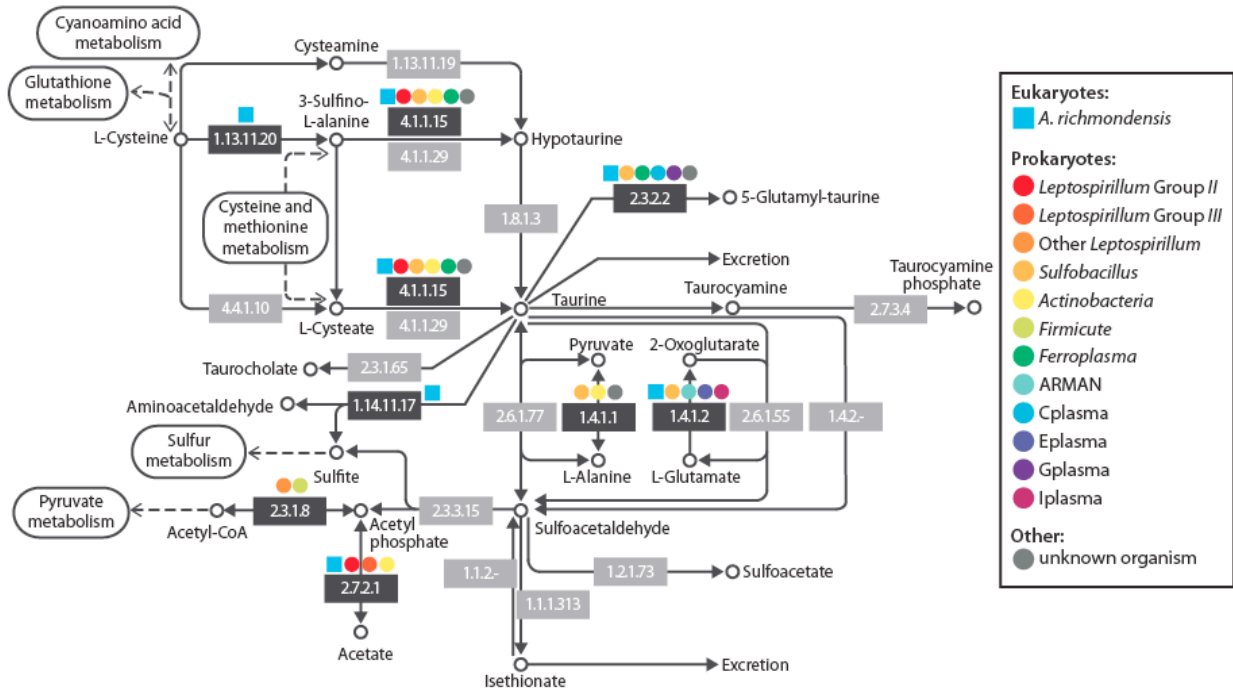


Figure 4.4. Abundance of taurine under different growth conditions of AMD biofilms.

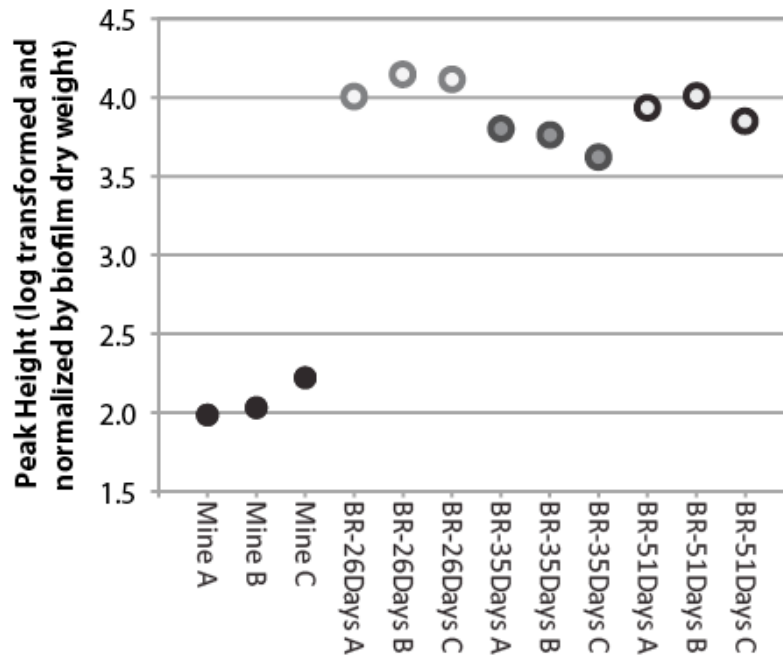


Figure 4.5. Identification of hydroxyectoine in the AMD biofilm using MS/MS comparisons with a chemical reference standard. Hydroxyectoine MS/MS spectra at three collision energies (10, 20, and 40 eV) shown in the top panels for the AMD biofilm and the bottom panels for the reference standard.

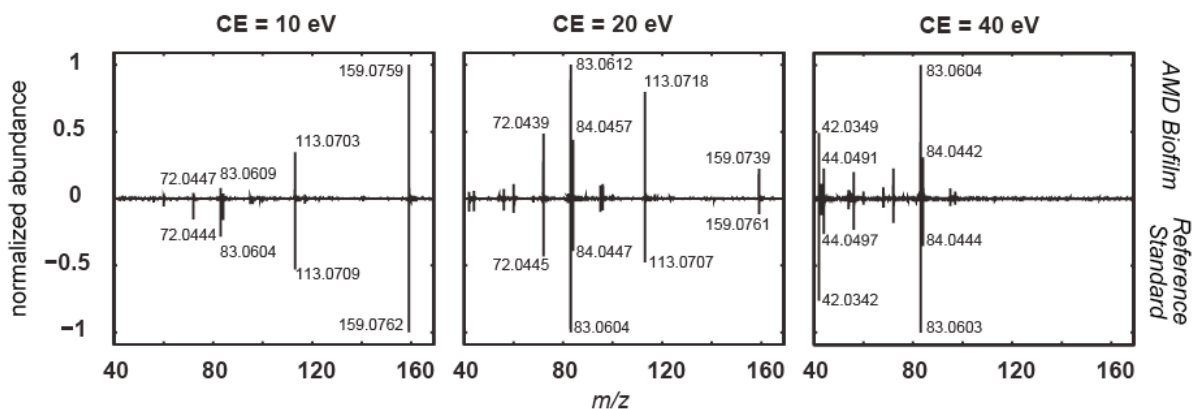
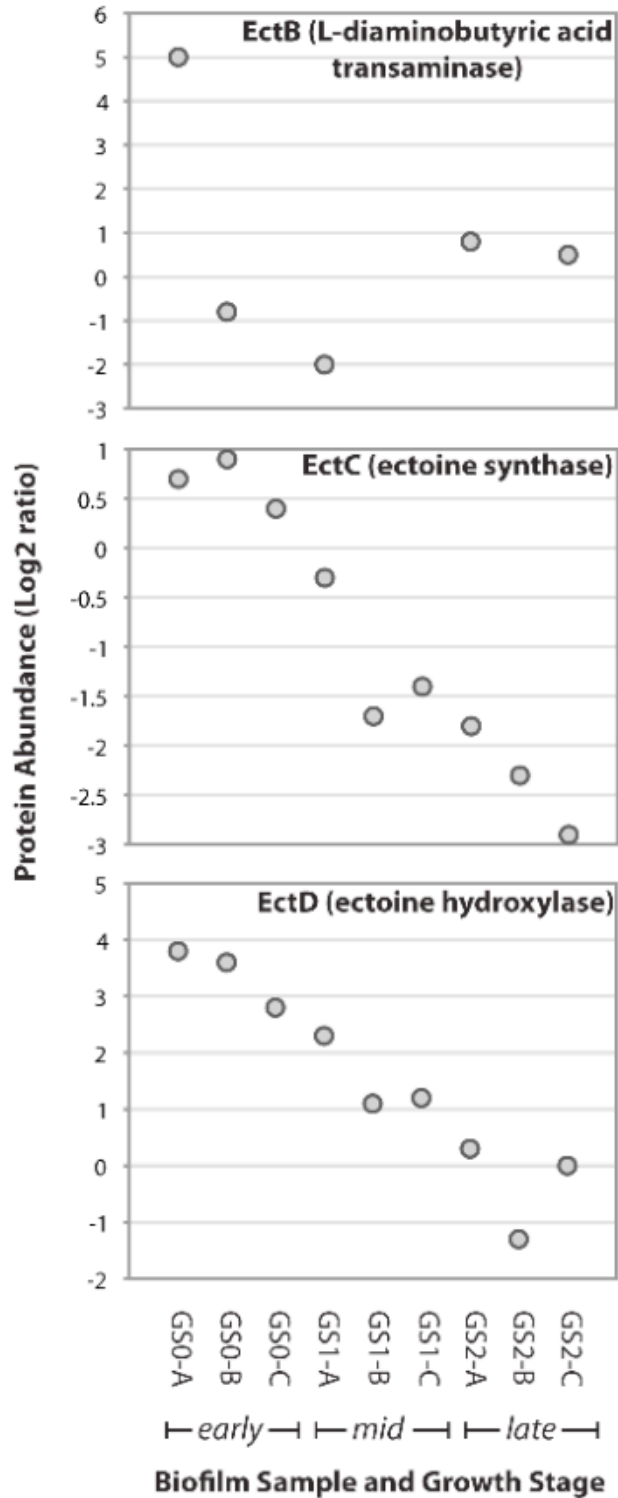


Figure 4.6. Abundance of ectoine and hydroxyectoine biosynthesis proteins across different growth stages of AMD biofilms. Data based on previously acquired quantitative proteomics data (Mueller et al., 2011).



References

- Abascal F, Zardoya R, Posada D. (2005). ProtTest: selection of best-fit models of protein evolution. *Bioinformatics* 21:2104–2105.
- Aliaga Goltsman DS, Dasari M, Thomas BC, Shah MB, VerBerkmoes NC, Hettich RL, et al. (2013). A new group in the *Leptospirillum* clade: cultivation-independent community genomics, proteomics and transcriptomics of the new species *Leptospirillum* group IV UBA BS. *Applied and Environmental Microbiology*.
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. (1990). Basic Local Alignment Search Tool. *Journal of Molecular Biology* 215:403–410.
- Amann R, Ludwig W, Schleifer KH. (1995). Phylogenetic Identification and in-Situ Detection of Individual Microbial-Cells Without Cultivation. *Microbiol. Rev.* 59:143–169.
- Anderson I, Chertkov O, Chen A, Saunders E, Lapidus A, Nolan M, et al. (2012). Complete genome sequence of the moderately thermophilic mineral-sulfide-oxidizing firmicute *Sulfobacillus acidophilus* type strain (NALT). *Standards in Genomic Sciences*.
- Andersson AF, Banfield JF. (2008). Virus population dynamics and acquired virus resistance in natural microbial communities. *Science* 320:1047–1050.
- Ashelford KE, Chuzhanova NA, Fry JC, Jones AJ, Weightman AJ. (2006). New screening software shows that most recent large 16S rRNA gene clone libraries contain chimeras. *Applied and Environmental Microbiology* 72:5734–5741.
- Baker BJ, Banfield JF. (2003). Microbial communities in acid mine drainage. *FEMS Microbiol. Ecol.* 44:139–152.
- Baker BJ, Comolli LR, Dick GJ, Hauser L, Hyatt D, Dill BD, et al. (2010). Enigmatic, ultrasmall, uncultivated Archaea. *Proc. Natl. Acad. Sci. U. S. A.* 107.
- Baker BJ, Hugenholtz P, Dawson SC, Banfield JF. (2003). Extremely acidophilic protists from acid mine drainage host Rickettsiales-lineage endosymbionts that have intervening sequences in their 16S rRNA genes. *Applied and Environmental Microbiology* 69:5512–5518.
- Baker BJ, Lutz MA, Dawson SC, Bond PL, Banfield JF. (2004). Metabolically active eukaryotic communities in extremely acidic mine drainage. *Applied and Environmental Microbiology* 70:6264–6271.
- Baker BJ, Tyson GW, Lindsey G, Banfield J. (2009). Insights into the Diversity of Eukaryotes in Acid Mine Drainage Biofilm Communities. *Applied and Environmental Microbiology* 2192–2199.
- Baker BJ, Tyson GW, Webb RI, Flanagan J, Hugenholtz P, Allen EE, et al. (2006). Lineages of acidophilic archaea revealed by community genomic analysis. *Science* 314:1933–1935.
- Baran R, Bowen BP, Bouskill NJ, Brodie EL, Yannone SM, Northen TR. (2010). Metabolite

Identification in *Synechococcus* sp. PCC 7002 Using Untargeted Stable Isotope Assisted Metabolite Profiling.

Baran R, Bowen BP, Northen TR. (2011). Untargeted metabolic footprinting reveals a surprising breadth of metabolite uptake and release by *Synechococcus* sp. PCC 7002.

Baran R, Kochi H, Saito N, Suematsu M, Soga T, Nishioka T, et al. (2006). MathDAMP: a package for differential analysis of metabolite profiles. *BMC Bioinformatics* 7.

Baran R, Robert M, Suematsu M, Soga T, Tomita M. (2007). Visualization of three-way comparisons of omics data. *BMC Bioinformatics* 8.

Beinert H. (1997). Copper A of Cytochrome c Oxidase, A Novel, Long-Embattled, Biological Electron-Transfer Site. *European journal of biochemistry / FEBS* 245:521–532.

Belnap CP, Pan C, Denev VJ, Samatova NF, Hettich RL, Banfield J. (2011). Quantitative Proteomic Analyses of the response of acidophilic microbial communities to different pH conditions. *ISME J*.

Belnap CP, Pan C, VerBerkmoes NC, Power ME, Samatova NF, Carver RL, et al. (2009). Cultivation and quantitative proteomic analyses of acidophilic microbial communities. *ISME J* 4:520–530.

Bick JA, Dennis JJ, Zylstra GJ, Nowack J, Leustek T. (2000). Identification of a New Class of 5'-Adenylylsulfate (APS) Reductases from Sulfate-Assimilating Bacteria. *Journal of Bacteriology* 182:135–142.

Bogdanova TI, Tsaphna IA, Kondrat'eva TF, Duda VI, Suzina NE, Melamud VS, et al. (2006). *Sulfobacillus thermotolerans* sp. nov., a thermotolerant, chemolithotrophic bacterium. *International Journal of Systematic and Evolutionary Microbiology* 56:1039–1042.

Bond PL, Banfield JF. (2001). Design and performance of rRNA targeted oligonucleotide probes for in situ detection and phylogenetic identification of microorganisms inhabiting acid mine drainage environments. *Microb. Ecol.* 41:149–161.

Bond PL, Druschel GK, Banfield JF. (2000a). Comparison of acid mine drainage microbial communities in physically and geochemically distinct ecosystems. *Applied and Environmental Microbiology* 66:4962–4971.

Bond PL, Smriga SP, Banfield JF. (2000b). Phylogeny of microorganisms populating a thick, subaerial, predominantly lithotrophic biofilm at an extreme acid mine drainage site. *Applied and Environmental Microbiology* 66:3842–3849.

Bonnefoy V, Holmes DS. (2011). Genomic insights into microbial iron oxidation and iron uptake strategies in extremely acidic environments. *Environmental Microbiology* 14:1597–1611.

Borisov VB, Gennis RB, Hemp J, Verkhovsky MI. (2011). The cytochrome bd respiratory oxygen reductases. *BBA - Bioenergetics* 1807:1398–1413.

- Boschker HTS, Nold SC, Wellsbury P, Bos D, de Graaf W, Pel R, et al. (1998). Direct linking of microbial populations to specific biogeochemical processes by ¹³C-labelling of biomarkers. *Nature* 392:801–805.
- Bowen BP, Fischer CR, Baran R, Banfield JF, Northen T. (2011). Improved genome annotation through untargeted detection of pathway-specific metabolites. *BMC Genomics* 12:S6.
- Brauer MJ, Yuan J, Bennett BD, Lu W, Kimball E, Botstein D, et al. (2006). Conservation of the metabolomic response to starvation across two divergent microbes. *103:19302–19307*.
- Breitbart M, Salamon P, Andresen B, Mahaffy JM, Segall AM, Mead D, et al. (2002). Genomic analysis of uncultured marine viral communities. *Proc. Natl. Acad. Sci. U. S. A.* 99:14250–14255.
- Bren A, Eisenbach M. (2000). How Signals Are Heard during Bacterial Chemotaxis: Protein-Protein Interactions in Sensory Signal Propagation. *Journal of Bacteriology* 182:6865–6873.
- Bridge TAM, Johnson DB. (1998). Reduction of Soluble Iron and Reductive Dissolution of Ferric Iron-Containing Minerals by Moderately Thermophilic Iron-Oxidizing Bacteria. 1–7.
- Bruneel O, Pascault N, Egal M, Bancon-Montigny C, Goni-Urriza MS, Elbaz-Poulichet F, et al. (2008). Archaeal diversity in a Fe-As rich acid mine drainage at Carnoules (France). *Extremophiles* 12:563–571.
- Bugaytsova Z, Lindström EB. (2004). Localization, purification and properties of a tetrathionate hydrolase from *Acidithiobacillus caldus*. *European journal of biochemistry / FEBS* 271:272–280.
- Caldwell PE, MacLean MR, Norris PR. (2007). Ribulose biphosphate carboxylase activity and a Calvin cycle gene cluster in *Sulfobacillus* species. *Microbiology (Reading, Engl.)* 153:2231–2240. 0.
- Canovas D, Vargas C, IglesiasGuerra F, Csonka LN, Rhodes D, Ventosa A, et al. (1997). Isolation and characterization of salt-sensitive mutants of the moderate halophile *Halomonas elongata* and cloning of the ectoine synthesis genes. *272:25794–25801*.
- Caporaso JG, Bittinger K, Bushman FD, DeSantis TZ, Andersen GL, Knight R. (2010a). PyNAST: a flexible tool for aligning sequences to a template alignment. *Bioinformatics* 26:266–267.
- Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, et al. (2010b). QIIME allows analysis of high-throughput community sequencing data. *Nature Methods* 7:335–336.
- Caspi R, Altman T, Dale JM, Dreher K, Fulcher CA, Gilham F, et al. (2010). The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. *38:D473–9*.
- Castresana J. (2000). Selection of conserved blocks from multiple alignments for their use in

phylogenetic analysis. *Molecular Biology and Evolution* 17:540–552.

Catak S, Monard G, Aviyente V, Ruiz-Lopez MF. (2006). Reaction mechanism of deamidation of asparaginyl residues in peptides: Effect of solvent molecules. *Journal of Physical Chemistry A* 110:8354–8365.

Cayley S, Lewis BA, Guttman HJ, Record MT. (1991). Characterization of the cytoplasm of *Escherichia coli* K-12 as a function of external osmolarity. Implications for protein-DNA interactions in vivo. 222:281–300.

Cárdenas JP, Moya F, Covarrubias P, Shmaryahu A, Levicán G, Holmes DS, et al. (2012). Comparative genomics of the oxidative stress response in bioleaching microorganisms. *Hydrometallurgy* 127-128:162–167.

Chai Y, Kolter R, Losick R. (2009). A Widely Conserved Gene Cluster Required for Lactate Utilization in *Bacillus subtilis* and Its Involvement in Biofilm Formation. *Journal of Bacteriology* 191:2423–2430.

Chen L, Ren Y, Lin J, Liu X, Pang X, Lin J. (2012). *Acidithiobacillus caldus* Sulfur Oxidation Model Based on Transcriptome Analysis between the Wild Type and Sulfur Oxygenase Reductase Defective Mutant Lebedev, N, ed. *PLoS ONE* 7:e39470.

Chen ZW, Liu YY, Wu JF, She Q, Jiang CY, Liu SJ. (2007). Novel bacterial sulfur oxygenase reductases from bioreactors treating gold-bearing concentrates. *Appl Microbiol Biotechnol* 74:688–698.

Cherney MM, Zhang Y, Solomonson M, Weiner JH, James MNG. (2010). Crystal Structure of Sulfide: Quinone Oxidoreductase from *Acidithiobacillus ferrooxidans*: Insights into Sulfidotrophic Respiration and Detoxification. *Journal of Molecular Biology* 398:292–305.

Coloso RM, Hirschberger LL, Dominy JE, Lee J-I, Stipanuk MH. (2006). Cysteamine dioxygenase: evidence for the physiological conversion of cysteamine to hypotaurine in rat and mouse tissues. 583:25–36.

Comolli LR, Baker BJ, Downing KH, Siegerist CE, Banfield JF. (2009). Three-dimensional analysis of the structure and ecology of a novel, ultra-small archaeon. *Isme J.* 3:159–167.

Constant P, Chowdhury SP, Hesse L, Pratscher J, Conrad R. (2011). Genome Data Mining and Soil Survey for the Novel Group 5 NiFe -Hydrogenase To Explore the Diversity and Ecological Importance of Presumptive High-Affinity H₂-Oxidizing Bacteria. *Applied and Environmental Microbiology* 77:6027–6035.

Constant P, Chowdhury SP, Pratscher J, Conrad R. (2010). Streptomycetes contributing to atmospheric molecular hydrogen soil uptake are widespread and encode a putative high-affinity [NiFe]-hydrogenase. *Environmental Microbiology* 12:821–829.

Cook AM, Denger K. (2006). Metabolism of taurine in microorganisms - A primer in molecular biodiversity? In:Vol. 583, pp. 3–13.

- Coppi MV. (2005). The hydrogenases of *Geobacter sulfurreducens*: a comparative genomic perspective. *Microbiology (Reading, Engl.)* 151:1239–1254.
- Coupland K, Johnson DB. (2008). Evidence that the potential for dissimilatory ferric iron reduction is widespread among acidophilic heterotrophic bacteria. *FEMS Microbiol. Lett.* 279:30–35.
- da Costa MS, Santos H, Galinski EA. (1998). An overview of the role and diversity of compatible solutes in Bacteria and Archaea. 61:117–153.
- Davey ME, O'toole GA. (2000). Microbial biofilms: from ecology to molecular genetics. *Microbiology and Molecular Biology Review* 64:847–867.
- De Jong GA, Hazeu W, Bos P, Kuenen JG. (1997). Isolation of the tetrathionate hydrolase from *Thiobacillus acidophilus*. *European journal of biochemistry / FEBS* 243:678–683.
- DeLong EF, Preston CM, Mincer T, Rich V, Hallam SJ, Frigaard NU, et al. (2006). Community genomics among stratified microbial assemblages in the ocean's interior. *Science* 311:496–503.
- Denef VJ, Banfield JF. (2012). In Situ Evolutionary Rate Measurements Show Ecological Success of Recently Emerged Bacterial Hybrids. *Science* 336:462–466.
- Denef VJ, Kalnejais LH, Mueller RS, Wilmes P, Baker BJ, Thomas BC, et al. (2010a). Proteogenomic basis for ecological divergence of closely related bacteria in natural acidophilic microbial communities. *Proc. Natl. Acad. Sci. U. S. A.* 107:2383–2390.
- Denef VJ, Mueller RS, Banfield JF. (2010b). AMD Biofilms: Using model communities to study microbial evolution and ecological complexity in nature. *Isme J.* 4:599–610.
- Denef VJ, Shah MB, VerBerkmoes NC, Hettich RL, Banfield JF. (2007). Implications of strain- and species-level sequence divergence for community and isolate shotgun proteomic analysis. *J. Proteome Res.* 6:3152–3161.
- Denef VJ, VerBerkmoes NC, Shah MB, Abraham P, Lefsrud M, Hettich RL, et al. (2009). Proteomics-inferred genome typing (PIGT) demonstrates inter-population recombination as a strategy for environmental adaptation. *Environmental Microbiology* 11:313–325.
- Denger K, Smits THM, Cook AM. (2006). Genome-enabled analysis of the utilization of taurine as sole source of carbon or of nitrogen by *Rhodobacter sphaeroides* 2.4.1. *Microbiology* 152:3197–3206.
- DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL, Keller K, et al. (2006). Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Applied and Environmental Microbiology* 72:5069–5072.
- 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.

- Dopson M, Baker-Austin C, Bond P. (2006). Towards determining details of anaerobic growth coupled to ferric iron reduction by the acidophilic archaeon 'Ferroplasma acidarmanus' Fer1. *Extremophiles* 11:159–168.
- Dopson M, Baker-Austin C, Hind A, Bowman JP, Bond PL. (2004). Characterization of *Ferroplasma* isolates and *Ferroplasma acidarmanus* sp nov., extreme acidophiles from acid mine drainage and industrial bioleaching environments. *Applied and Environmental Microbiology* 70:2079–2088.
- Dopson M, Johnson DB. (2012). Biodiversity, metabolism and applications of acidophilic sulfur-metabolizing microorganisms. *Environmental Microbiology* 14:2620–2631.
- Druschel GK, Baker BJ, Gihring TM, Banfield JF. (2004). Acid mine drainage biogeochemistry at Iron Mountain, California. *Geochemical Transactions* 5:13–32.
- Druschel GK, Hamers RJ, Banfield JF. (2003). Kinetics and mechanism of polythionate oxidation to sulfate at low pH by O₂ and Fe³⁺. *Geochimica et Cosmochimica Acta* 67:4457–4469.
- Duda VI, Suzina NE, Severina LO, Dmitriev VV, Karavaiko GI. (2001). Formation of Flat Lamellar Intramembrane Lipid Structures in Microorganisms. *Journal of Membrane Biology* 180:33–48.
- Dumont MG, Murrell JC. (2005). Stable isotope probing - linking microbial identity to function. *Nature Reviews Microbiology* 3:499–504.
- Dunn WB, Erban A, Weber RJM, Creek DJ, Brown M, Breitling R, et al. (2012). Mass appeal: metabolite identification in mass spectrometry-focused untargeted metabolomics. *Metabolomics* 1–23.
- Edgar RC. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32:1792–1797.
- Edgar RC. (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 26:2460–2461.
- Edgar RC, Haas BJ, Clemente JC, Quince C, Knight R. (2011). UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* 27:2194–2200.
- Edwards KJ, Bond PL, Druschel GK, McGuire MM, Hamers RJ, Banfield JF. (2000a). Geochemical and biological aspects of sulfide mineral dissolution: lessons from Iron Mountain, California. *Chemical Geology* 169:383–397.
- Edwards KJ, Bond PL, Gihring TM, Banfield JF. (2000b). An archaeal iron-oxidizing extreme acidophile important in acid mine drainage. *Science* 287:1796–1799.
- Edwards KJ, Gihring TM, Banfield JF. (1999). Seasonal variations in microbial populations and environmental conditions in an extreme acid mine drainage environment. *Applied and*

Environmental Microbiology 65:3627–3632.

Egorova MA, Tsaplina IA, Zakharchuk LM, Bogdanova TI, Krasil'nikova EN. (2004). Effect of cultivation conditions on the growth and activities of sulfur metabolism enzymes and carboxylases of *Sulfobacillus thermosulfidooxidans* subsp *asporogenes* strain 41. *Applied Biochemistry and Microbiology* 40:381–387.

Eichhorn E. (1997). Characterization of alpha -Ketoglutarate-dependent Taurine Dioxygenase from *Escherichia coli*. 272:23031–23036.

Empadinhas N, da Costa MS. (2006). Diversity and biosynthesis of compatible solutes in hyper/thermophiles. *International Microbiology* 9:199–206.

Eng JK, McCormack AL, Yates JR. (1994). An Approach to Correlate Tandem Mass-Spectral Data of Peptides with Amino-acid dequences in a Protein Database. *J. Am. Soc. Mass Spectrom.* 5:976–989.

Fetzer I, Jehmlich N, Vogt C, Richnow H-H, Seifert J, Harms H, et al. (2010). Calculation of partial isotope incorporation into peptides measured by mass spectrometry. *BMC Res Notes* 3:178.

Fiehn O. (2001). Combining Genomics, Metabolome Analysis, and Biochemical Modelling to Understand Metabolic Networks. *Comparative and Functional Genomics* 2:155–168.

Fimrite P. (2010). Inside a toxic hellhole, Iron Mountain Mine. *San Francisco Chronicle*, August 29.

Fischer CR, Wilmes P, Bowen BP, Northen TR, Banfield JF. (2011). Deuterium-exchange metabolomics identifies N-methyl lyso phosphatidylethanolamines as abundant lipids in acidophilic mixed microbial communities. *Metabolomics*.

Florens L, Carozza MJ, Swanson SK, Fournier M, Coleman MK, Workman JL, et al. (2006). Analyzing chromatin remodeling complexes using shotgun proteomics and normalized spectral abundance factors. *Methods* 40:303–311.

Fontana M, Duprè S, Pecci L. (2006). The reactivity of hypotaurine and cysteine sulfinic acid with peroxyxynitrite. *Adv. Exp. Med. Biol* 583:15–24.

Garcia DE, Baidoo EE, Benke PI, Pingitore F, Tang YJ, Villa S, et al. (2008). Separation and mass spectrometry in microbial metabolomics. *Current Opinion in Microbiology* 11:233–239.

Garcia-Esteva R, Argandona M, Reina-Bueno M, Capote N, Iglesias-Guerra F, Nieto JJ, et al. (2006). The *ectD* Gene, Which Is Involved in the Synthesis of the Compatible Solute Hydroxyectoine, Is Essential for Thermoprotection of the Halophilic Bacterium *Chromohalobacter salexigens*. *Journal of Bacteriology* 188:3774–3784.

Gardner PR. (2005). Nitric oxide dioxygenase function and mechanism of flavohemoglobin, hemoglobin, myoglobin and their associated reductases. *Journal of Inorganic Biochemistry*

99:247–266.

Ghauri MA, Johnson DB. (1991). Physiological Diversity Amongst Some Moderately Thermophilic Iron-Oxidizing Bacteria. *FEMS Microbiol. Ecol.* 85:327–334.

Giovanni R. (1963). Studies on Incorporation of Deuterium into Bacteria. *Biochemical Journal* 87:79–82.

Giovanni R. (1961). The Effects of Deuterium Oxide on Bacteria. *Zeitschrift für Vererbungslehre* 92:389–402.

Golovacheva RS, Karavaiko GI. (1978). *Sulfobacillus*, a new genus of thermophilic sporulating bacteria. *Mikrobiologiya* 47:815–822.

Goltsman DSA, Deneff VJ, Singer SW, VerBerkmoes NC, Lefsrud M, Mueller RS, et al. (2009). Community Genomic and Proteomic Analyses of Chemoautotrophic Iron-Oxidizing "*Leptospirillum rubrum*" (Group II) and '*Leptospirillum ferrodiazotrophum*' (Group III) Bacteria in Acid Mine Drainage Biofilms. *Applied and Environmental Microbiology* 75:4599–4615.

Golyshina OV, Pivovarova TA, Karavaiko GI, Kondrat'eva TF, Moore ERB, Abraham WR, et al. (2000). *Ferroplasma acidiphilum* gen. nov., sp nov., an acidophilic, autotrophic, ferrous-iron-oxidizing, cell-wall-lacking, mesophilic member of the *Ferroplasmaceae* fam. nov., comprising a distinct lineage of the Archaea. *International Journal of Systematic and Evolutionary Microbiology* 50:997–1006.

Gomez-Alvarez V, Teal TK, Schmidt TM. (2009). Systematic artifacts in metagenomes from complex microbial communities. *ISME J.* 3:1314–1317.

Grein F, Ramos AR, Venceslau SS, Pereira IAC. (2013). Unifying concepts in anaerobic respiration: Insights from dissimilatory sulfur metabolism. *biochimica et biophysica acta* 1827:145–160.

Grigoriev IV, Nordberg H, Shabalov I, Aerts A, Cantor M, Goodstein D, et al. (2012). The genome portal of the Department of Energy Joint Genome Institute. 40:D26–32. <http://eutils.ncbi.nlm.nih.gov/entrez/eutils/elink.fcgi?dbfrom=pubmed&id=22110030&retmode=ref&cmd=prlinks>.

Guiral M, Tron P, Aubert C, Gloter A, Iobbi-Nivol C, Giudici-Ortoni M-T. (2005). A membrane-bound multienzyme, hydrogen-oxidizing, and sulfur-reducing complex from the hyperthermophilic bacterium *Aquifex aeolicus*. *Journal of Biological Chemistry* 280:42004–42015.

Gutierrez-Zamora M-L, Manfield M. (2010). An appraisal of methods for linking environmental processes to specific microbial taxa. *Rev Environ Sci Biotechnol* 9:153–185.

Halter D, Goulhen-Chollet F, Gallien SEB, Casiot C, Hamelin JEROM, Gilard FCO, et al. (2012). In situ proteo-metabolomics reveals metabolite secretion by the acid mine drainage bio-

indicator, *Euglena mutabilis*. *ISME J* 6:1391–1402.
<http://www.nature.com/doi/10.1038/ismej.2011.198>.

Hamann N, Mander GJ, Shokes JE, Scott RA, Bennati M, Hedderich R. (2007). A cysteine-rich CCG domain contains a novel [4Fe-4S] cluster binding motif as deduced from studies with subunit B of heterodisulfide reductase from *Methanothermobacter marburgensis*. *Biochemistry* 46:12875–12885.

Hardisty DS, Olyphant GA, Bell JB, Johnson AP, Pratt LM. (2013). Acidophilic sulfur disproportionation. *Geochimica et Cosmochimica Acta* 113:136–151.

Hedderich R, Hamann N, Bennati M. (2005). Heterodisulfide reductase from methanogenic archaea: a new catalytic role for an iron-sulfur cluster. *Biol. Chem.* 386:961–970.

Hedderich R, Klimmek O, Kröger A, Dirmeier R, Keller M, Stetter KO. (1999). Anaerobic respiration with elemental sulfur and with disulfides. *FEMS Microbiol. Rev.* 22:353–381.

Hegeman AD, Schulte CF, Cui Q, Lewis IA, Huttlin EL, Eghbalnia H, et al. (2007). Stable Isotope Assisted Assignment of Elemental Compositions for Metabolomics. *Anal. Chem.* 79:6912–6921.

Hinrichs KU, Hayes JM, Sylva SP, Brewer PG, DeLong EF. (1999). Methane-consuming archaeobacteria in marine sediments. *Nature* 398:802–805.

Hinsley AP, Berks BC. (2002). Specificity of respiratory pathways involved in the reduction of sulfur compounds by *Salmonella enterica*. *Microbiology (Reading, Engl.)* 148:3631–3638.

Huber H, Gallenberger M, Jahn U, Eylert E, Berg IA, Kockelkorn D, et al. (2008). A dicarboxylate/4-hydroxybutyrate autotrophic carbon assimilation cycle in the hyperthermophilic *Archaeum Ignicoccus hospitalis*. *Proc. Natl. Acad. Sci. U. S. A.* 105:7851–7856.

Huxtable RJ. (1992). Physiological actions of taurine. *Physiological Reviews* 72:101–163.

Hyatt D, Chen G-L, LoCasio PF, Land ML, Larimer FW, Hauser LJ. (2010). Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 11.

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

Inskeep WP, Rusch DB, Jay ZJ, Herrgard MJ, Kozubal MA, Richardson TH, et al. (2010). Metagenomes from High-Temperature Chemotrophic Systems Reveal Geochemical Controls on Microbial Community Structure and Function. *PLoS ONE* 5.

Jebbar M, Sohn-Bosser L, Bremer E, Bernard T, Blanco C. (2005). Ectoine-Induced Proteins in *Sinorhizobium meliloti* Include an Ectoine ABC-Type Transporter Involved in Osmoprotection and Ectoine Catabolism. *Journal of Bacteriology* 187:1293–1304.

Jehmlich N, Fetzer I, Seifert J, Mattow J, Vogt C, Harms H, et al. (2010a). Decimal place slope, a fast and precise method for quantifying ¹³C incorporation levels for detecting the metabolic activity of microbial species. *Mol. Cell Proteomics* 9:1221–1227.

Jehmlich N, Schmidt F, Bergen von M, Richnow H-H, Vogt C. (2008a). Protein-based stable isotope probing (Protein-SIP) reveals active species within anoxic mixed cultures. *ISME J* 2:1122–1133.

Jehmlich N, Schmidt F, Hartwich M, Bergen von M, Richnow H-H, Vogt C. (2008b). Incorporation of carbon and nitrogen atoms into proteins measured by protein-based stable isotope probing (Protein-SIP). *Rapid Communications in Mass Spectrometry* 22:2889–2897.

Jehmlich N, Schmidt F, Taubert M, Seifert J, Bastida F, Bergen von M, et al. (2010b). Protein-based stable isotope probing. *Nature Protocols* 5:1957–1966.

JGI. Protocols in Production Sequencing.

http://www.jgi.doe.gov/sequencing/protocols/protos_production.html.

Jiang H, Dong H, Yu B, Ye Q, Shen J, Rowe H, et al. (2008). Dominance of putative marine benthic Archaea in Qinghai Lake, north-western China. *Environmental Microbiology* 10:2355–2367.

Jiao Y, Cody GD, Harding AK, Wilmes P, Schrenk MO, Wheeler KE, et al. (2010). Characterization of Extracellular Polymeric Substances from Acidophilic Microbial Biofilms. *Applied and Environmental Microbiology* 76:2916–2922.

Johnson DB, Joulain C, d'Hugues P, Hallberg KB. (2008). *Sulfobacillus benefaciens* sp. nov., an acidophilic facultative anaerobic Firmicute isolated from mineral bioleaching operations. *Extremophiles* 12:789–798. <http://www.springerlink.com/index/10.1007/s00792-008-0184-4>.

Johnson DB, Kanao T, Hedrich S. (2012). Redox transformations of iron at extremely low pH: fundamental and applied aspects. *Frontiers in microbiology* 3:1–13.

Johnson DB, Okibe N, Roberto FF. (2003). Novel thermo-acidophilic bacteria isolated from geothermal sites in Yellowstone National Park: physiological and phylogenetic characteristics. *Arch. Microbiol.* 180:60–68.

Jones DS, Albrecht HL, Dawson KS, Schaperdorth I, Freeman KH, Pi Y, et al. (2011). Community genomic analysis of an extremely acidophilic sulfur-oxidizing biofilm. *ISME J* 6:158–170.

Joshi AB, Kirsch LE. (2002). The relative rates of glutamine and asparagine deamidation in glucagon fragment 22-29 under acidic conditions. *Journal of Pharmaceutical Sciences* 91:2332–2345.

Justice NB, Pan C, Mueller R, Spaulding SE, Shah V, Sun CL, et al. (2012). Heterotrophic Archaea Contribute to Carbon Cycling in Low-pH, Suboxic Biofilm Communities. *Applied and Environmental Microbiology* 78:8321–8330.

- Kanao T, Kamimura K, Sugio T. (2007). Identification of a gene encoding a tetrathionate hydrolase in *Acidithiobacillus ferrooxidans*. *J. Biotechnol.*
- Kaneda T. (1991). Iso-and anteiso-fatty acids in bacteria: biosynthesis, function, and taxonomic significance. *Microbiol. Rev.* 55:288–302.
- Kannenbergh EL, Poralla K. (1999). Hopanoid biosynthesis and function in bacteria. *Naturwissenschaften* 86:168–176. <http://link.springer.com/article/10.1007/s001140050592>.
- Kappler U. (2000). Sulfite: Cytochrome c Oxidoreductase from *Thiobacillus novellus*. Purification, Characterization, and Molecular Biology of Heterodimeric Member of the Sulfite Oxidase Family. *Journal of Biological Chemistry* 275:13202–13212.
- Kappler U, Dahl C. (2001). Enzymology and molecular biology of prokaryotic sulfite oxidation. *FEMS Microbiol. Lett.* 203:1–9.
- Karavaiko GI, Krasil'nikova EN, Tsaplina IA, Bogdanova TI, Zakharchuk LM. (2001). Growth and carbohydrate metabolism of sulfobacilli. *Microbiology* 70:245–250.
- Katoh K, Misawa K, Kuma K, Miyata T. (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* 30:3059–3066.
- Kerscher S, Dröse S, Zickermann V, Brandt U. (2008). The three families of respiratory NADH dehydrogenases. *Results Probl Cell Differ* 45:185–222.
- Kim S, Lee SB. (2008). Identification and characterization of the bacterial d-gluconate dehydratase in *Achromobacter xylosoxidans*. *Biotechnol Bioproc E* 13:436–444.
- King GM, Weber CF. (2007). Distribution, diversity and ecology of aerobic CO-oxidizing bacteria. *Nature Reviews Microbiology* 5:107–118.
- Kock D, Schippers A. (2008). Quantitative microbial community analysis of three different sulfidic mine tailing dumps generating acid mine drainage. *Applied and Environmental Microbiology* 74:5211–5219.
- Kopriva S, Buchert T, Fritz G, Suter M, Benda RD, Schunemann V, et al. (2002). The presence of an iron-sulfur cluster in adenosine 5'-phosphosulfate reductase separates organisms utilizing adenosine 5'-phosphosulfate and phosphoadenosine 5'-phosphosulfate for sulfate assimilation. *Journal of Biological Chemistry* 277:21786–21791.
- Krasil'nikova EN, Zakharchuk LM, Egorova MA, Bogdanova TI, Zhuravleva AE, Tsaplina IA. (2010). Regulation of metabolic pathways in sulfobacilli under different aeration regimes. *Microbiology* 79:147–152.
- Kristal BS, Vigneau-Callahan KE, Matson WR. (1998). Simultaneous analysis of the majority of low-molecular-weight, redox-active compounds from mitochondria. *Anal. Biochem.* 263:18–25.
- Kucera J, Bouchal P, Cerna H, Potesil D, Janiczek O, Zdrahal Z, et al. (2011). Kinetics of

anaerobic elemental sulfur oxidation by ferric iron in *Acidithiobacillus ferrooxidans* and protein identification by comparative 2-DE-MS/MS. *Antonie van Leeuwenhoek* 101:561–573.

Lancaster CRD. (2002). Succinate:quinone oxidoreductases: an overview. *biochimica et biophysica acta* 1–6.

Lencina AM, Ding Z, Schurig-Briccio LA. (2012). Characterization of the Type III sulfide:quinone oxidoreductase from *Caldivirga maquilingsensis* and its membrane binding. *Biochimica et Biophysica*

Lentzen G, Schwarz T. (2006). Extremolytes: natural compounds from extremophiles for versatile applications. *Appl Microbiol Biotechnol* 72:623–634.

Letunic I, Bork P. (2007). Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. *Bioinformatics* 23:127–128.

Levis GS, Park E. (2003). Taurine: new implications for an old amino acid. *FEMS Microbiol. Lett.* 226:195–202.

Li B, Chen Y, Liu Q, Hu S, Chen X. (2011). Complete genome analysis of *Sulfobacillus acidophilus* strain TPY, isolated from a hydrothermal vent in the Pacific Ocean. *Journal of Bacteriology* 193:5555–5556.

Li H, Durbin R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25:1754–1760.

Lin JT, Stewart V. (1998). Nitrate assimilation by bacteria. *Adv. Microb. Physiol.* 39:1–30– 379.

Lippert K, Galinski E. (1992). Enzyme stabilization by ectoine-type compatible solutes: protection against heating, freezing and drying. *Appl Microbiol Biotechnol* 37.

Lo I, Denev VJ, VerBerkmoes NC, Shah MB, Goltsman D, DiBartolo G, et al. (2007). Strain-resolved community proteomics reveals recombining genomes of acidophilic bacteria. *Nature* 446:537–541.

Louis P, Galinski EA. (1997). Characterization of genes for the biosynthesis of the compatible solute ectoine from *Marinococcus halophilus* and osmoregulated expression in *Escherichia coli*. *Microbiology (Reading, Engl.)* 143:1141–1149.

Lowe TM, Eddy SR. (1997). tRNAscan-SE: A program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* 25:955–964.

Luther GW. (1987). Pyrite oxidation and reduction: molecular orbital theory considerations. *Geochimica et Cosmochimica Acta* 51:3193–3199.

Ma S, Banfield JF. (2011). Micron-scale Fe(2+)/Fe(3+), intermediate sulfur species and O(2) gradients across the biofilm-solution-sediment interface control biofilm organization. *Geochimica et Cosmochimica Acta* 75:3568–3580.

- Macalady JL, Jones DS, Lyon EH. (2007). Extremely acidic, pendulous cave wall biofilms from the Frasassi cave system, Italy. *Environmental Microbiology* 9:1402–1414.
- Macalady JL, Vestling MM, Baumler D, Boekelheide N, Kaspar CW, Banfield JF. (2004). Tetraether-linked membrane monolayers in *Ferroplasma* spp: a key to survival in acid. *Extremophiles* 8:411–419.
- Marcia M, Ermler U, Peng G, Michel H. (2009). A new structure-based classification of sulfide:quinone oxidoreductases. *Proteins* 78:1073–1083.
- Markowitz VM, Chen IMA, Palaniappan K, Chu K, Szeto E, Grechkin Y, et al. (2011). IMG: the integrated microbial genomes database and comparative analysis system. *Nucleic Acids Res.* 40:D115–D122.
- Marreiros BC, Batista AP, Duarte AMS, Pereira MM. (2013). A missing link between complex I and group 4 membrane-bound [NiFe] hydrogenases. *BBA - Bioenergetics* 1827:198–209.
- Melamud VS, Pivovarova TA, Tourova TP, Kolganova TV, Osipov GA, Lysenko AM, et al. (2003). *Sulfobacillus sibiricus* sp. nov., a New Moderately Thermophilic Bacterium - Springer. *Microbiology* 72:605–612.
- Miller CS, Baker BJ, Thomas BC, Singer SW, Banfield JF. (2011). EMIRGE: reconstruction of full-length ribosomal genes from microbial community short read sequencing data. *Genome Biol.* 12:R44.
- Moparthi VK, Hägerhäll C. (2011). The Evolution of Respiratory Chain Complex I from a Smaller Last Common Ancestor Consisting of 11 Protein Subunits. *J Mol Evol* 72:484–497.
- Moriya Y, Itoh M, Okuda S, Yoshizawa AC, Kanehisa M. (2007). KAAS: an automatic genome annotation and pathway reconstruction server. *Nucleic Acids Res.* 35:W182–W185.
- Morris TW, Reed KE, Cronan JE. (1994). Identification of the Gene Encoding Lipoate-Protein Ligase-a of *Escherichia-Coli* - Molecular-Cloning and Characterization of the Lpla Gene and Gene-Product. *Journal of Biological Chemistry* 269:16091–16100.
- Mosier AC, (null), Bowen BP, Baran R, Thomas BC, Northen TR, et al. (2013). Metabolites Associated with Adaptation of Microorganisms to an Acidophilic, Metal-Rich Environment Identified by Stable-Isotope-Enabled Metabolomics. *mBio* 4:e00484–12–e00484–12.
- Mouser PJ, Holmes DE, Perpetua LA, DiDonato R, Postier B, Liu A, et al. (2009). Quantifying expression of *Geobacter* spp. oxidative stress genes in pure culture and during in situ uranium bioremediation. *Isme J.* 3:454–465.
- Mueller RS, Denev VJ, Kalnejais LH, Suttle BK, Thomas BC, Wilmes P, et al. (2010). Ecological distribution and population physiology defined by proteomics in a natural microbial community. *Molecular Systems Biology* 6.
- Mueller RS, Dill BD, Pan CL, Belnap CP, Thomas BC, VerBerkmoes NC, et al. (2011).

- Proteome changes in the initial bacterial colonist during ecological succession in an acid mine drainage biofilm community. *Environmental Microbiology* 13:2279–2292.
- Muller J, Szklarczyk D, Julien P, Letunic I, Roth A, Kuhn M, et al. (2010). eggNOG v2.0: extending the evolutionary genealogy of genes with enhanced non-supervised orthologous groups, species and functional annotations. *Nucleic Acids Res.* 38:D190–D195.
- Müller FH, Bandejas TM, Urich T, Teixeira M, Gomes CM, Kletzin A. (2004). Coupling of the pathway of sulphur oxidation to dioxygen reduction: characterization of a novel membrane-bound thiosulphate:quinone oxidoreductase. *Mol. Microbiol.* 53:1147–1160.
- Nancucheo I, Johnson DB. (2010). Production of glycolic acid by chemolithotrophic iron- and sulfur-oxidizing bacteria and its role in delineating and sustaining acidophilic sulfide mineral-oxidizing consortia. *Applied and Environmental Microbiology* 76:461–467.
- Norambuena J, Flores R, Cárdenas JP, Quatrini R, Chávez R, Levicán G. (2012). Thiol/Disulfide System Plays a Crucial Role in Redox Protection in the Acidophilic Iron-Oxidizing Bacterium *Leptospirillum ferriphilum* Meijler, MM, ed. *PLoS ONE* 7:e44576.
- Norris PR, Clark DA, Owen JP, Waterhouse S. (1996). Characteristics of *Sulfobacillus acidophilus* sp. nov. and other moderately thermophilic mineral-sulphide-oxidizing bacteria. *Microbiology (Reading, Engl.)* 142 (Pt 4):775–783.
- Northrop DB. (1975). Steady-state analysis of kinetic isotope effects in enzymic reactions. *Biochemistry* 14:2644–2651.
- Ogata H, Goto S, Sato K, Fujibuchi W, Bono H, Kanehisa M. (1999). KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* 27:29–34.
- Ohmura N, Sasaki K, Matsumoto N, Saiki H. (2002). Anaerobic Respiration Using Fe³⁺, S₀, and H₂ in the Chemolithoautotrophic Bacterium *Acidithiobacillus ferrooxidans*. *Journal of Bacteriology* 184:2081–2087.
- Oldiges M, Lütz S, Pflug S, Schroer K, Stein N, Wiendahl C. (2007). Metabolomics: current state and evolving methodologies and tools. *Appl Microbiol Biotechnol* 76:495–511..
- Orcutt BN, Sylvan JB, Knab NJ, Edwards KJ. (2011). Microbial Ecology of the Dark Ocean above, at, and below the Seafloor. *Microbiol. Mol. Biol. Rev.* 75:361—.
- Oren A. (2002). Molecular ecology of extremely halophilic Archaea and Bacteria. *FEMS Microbiol. Ecol.* 39:1–7.
- OSHIMA M, ARIGA T. (1975). Omega-Cyclohexyl Fatty-Acids in Acidophilic Thermophilic Bacteria - Studies on Their Presence, Structure, and Biosynthesis Using Precursors Labeled with Stable Isotopes and Radioisotopes. *Journal of Biological Chemistry* 250:6963–6968.
- Osorio H, Mangold S, Denis Y, Nancucheo I, Esparza M, Johnson DB, et al. (2013). Anaerobic Sulfur Metabolism Coupled to Dissimilatory Iron Reduction in the Extremophile

- Acidithiobacillus ferrooxidans*. *Applied and Environmental Microbiology* 79:2172–2181.
- Pan CL, Fischer CR, Hyatt D, Bowen BP, Hettich RL, Banfield JF. (2011). Quantitative Tracking of Isotope Flows in Proteomes of Microbial Communities. *Molecular & Cellular Proteomics* 10:M110.006049–M110.006049.
- Pastor JM, Salvador M, Argandoña M, Bernal V, Reina-Bueno M, Csonka LN, et al. (2010). Ectoines in cell stress protection: Uses and biotechnological production. *Biotechnology Advances* 28:782–801. <http://linkinghub.elsevier.com/retrieve/pii/S0734975010000911>.
- Peng J, Elias JE, Thoreen CC, Licklider LJ, Gygi SP. (2003). Evaluation of Multidimensional Chromatography Coupled with Tandem Mass Spectrometry (LC/LC–MS/MS) for Large-Scale Protein Analysis: The Yeast Proteome. *J. Proteome Res.* 2:43–50.
- Peng Y, Leung HCM, Yiu SM, Chin FYL. (2012). IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics* 28:1420–1428.
- Perham RN. (2000). Swinging Arms and Swinging Domains in Multifunctional Enzymes: Catalytic Machines for Multistep Reactions. *Annu. Rev. Biochem.* 69:961–1004.
- Pluskal T, Castillo S, Villar-Briones A, Oresic M. (2010). MZmine 2: Modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. *BMC Bioinformatics* 11:395.
- Poole RK, Hughes MN. (2000). New functions for the ancient globin family: bacterial responses to nitric oxide and nitrosative stress. *Mol. Microbiol.* 36:775–783.
- Pott AS, Dahl C. (1998). Sirohaem sulfite reductase and other proteins encoded by genes at the *dsr* locus of *Chromatium vinosum* are involved in the oxidation of intracellular sulfur. *Microbiology (Reading, Engl.)* 144:1881–1894.
- Prabhu J, Schauwecker F, Grammel N, Keller U, Bernhard M. (2004). Functional Expression of the Ectoine Hydroxylase Gene (*thpD*) from *Streptomyces chrysomallus* in *Halomonas elongata*. *Applied and Environmental Microbiology* 70:3130–3132.
- Protze J, Müller F, Lauber K, Naß B, Mentele R, Lottspeich F, et al. (2011). An Extracellular Tetrathionate Hydrolase from the Thermoacidophilic Archaeon *Acidianus Ambivalens* with an Activity Optimum at pH 1. *Frontiers in microbiology* 2:68.
- Pruesse E, Quast C, Knittel K, Fuchs BM, Ludwig W, Peplies J, et al. (2007). SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Res.* 35:7188–7196.
- Quatrini R, Appia-Ayme C, Denis Y, Jedlicki E, Holmes DS, Bonnefoy V. (2009). Extending the models for iron and sulfur oxidation in the extreme Acidophile *Acidithiobacillus ferrooxidans*. *BMC Genomics* 10:394.
- Quevillon E, Silventoinen V, Pillai S, Harte N, Mulder N, Apweiler R, et al. (2005).

- InterProScan: protein domains identifier. *Nucleic Acids Res.* 33:W116–W120.
- Radajewski S, Ineson P, Parekh NR, Murrell JC. (2000). Stable-isotope probing as a tool in microbial ecology. *Nature* 403:646–649.
- Ram RJ, VerBerkmoes NC, Thelen MP, Tyson GW, Baker BJ, Blake RC, et al. (2005). Community proteomics of a natural microbial biofilm. *Science* 308:1915–1920.
- Rawlings DE. (2002). Heavy metal mining using microbes. *Annu. Rev. Microbiol.* 56:65–91.
- Richardson DJ, Berks BC, Russell DA, Spiro S, Taylor CJ. (2001). Functional, biochemical and genetic diversity of prokaryotic nitrate reductases. *CMLS, Cell. Mol. Life Sci.* 58:165–178.
- Robinson NE. (2002). Protein deamidation. *Proc. Natl. Acad. Sci. U. S. A.* 99:5283–5288.
- Rodgers RP, Blumer EN, Hendrickson CL, Marshall AG. (2000). Stable isotope incorporation triples the upper mass limit for determination of elemental composition by accurate mass measurement. *J. Am. Soc. Mass Spectrom.* 11:835–840.
- Rohwerder T. (2003). The sulfane sulfur of persulfides is the actual substrate of the sulfur-oxidizing enzymes from *Acidithiobacillus* and *Acidiphilium* spp. *Microbiology* 149:1699–1710.
- Rohwerder T, Sand W. (2007). Oxidation of Inorganic Sulfur Compounds in Acidophilic Prokaryotes. *Eng. Life Sci.* 7:301–309.
- Rothery RA, Workun GJ, Weiner JH. (2008). The prokaryotic complex iron-sulfur molybdoenzyme family. *biochimica et biophysica acta* 1778:1897–1929.
- Saeed AI, Hagabati NK, Braisted JC, Liang W, Sharov V, Howe EA, et al. (2006). TM4 microarray software suite. In: *DNA Microarrays, Part B: Databases and Statistics*, Kimmel, AOB, ed (ed). Vol. 411, pp. 134–193.
- Saeed AI, Sharov V, White J, Li J, Liang W, Bhagabati N, et al. (2003). TM4: A free, open-source system for microarray data management and analysis. *Biotechniques* 34:374—.
- Sanchez-Andrea I, Rodriguez N, Amils R, Luis Sanz J. (2011). Microbial Diversity in Anaerobic Sediments at Rio Tinto, a Naturally Acidic Environment with a High Heavy Metal Content. *Applied and Environmental Microbiology* 77:6085–6093.
- Saraste M. (1994). Structure and evolution of cytochrome oxidase. 1–3.
- Schafer G, Anemuller S, Moll R. (2002). Archaeal complex II: 'classical' and 'non-classical' succinate : quinone reductases with unusual features. *BBA - Bioenergetics* 1553:57–73.
- Schink B, Kremer DR, Hansen TA. (1987). Pathway of propionate formation from ethanol in *Pelobacter propionicus*. *Arch. Microbiol.* 147:321–327.
- Schippers A, Jozsa P, Sand W. (1996). Sulfur chemistry in bacterial leaching of pyrite. *Applied and Environmental Microbiology* 62:3424–3431.

- Schrenk MO, Edwards KJ, Goodman RM, Hamers RJ, Banfield JF. (1998). Distribution of *Thiobacillus ferrooxidans* and *Leptospirillum ferrooxidans*: Implications for generation of acid mine drainage. *Science* 279:1519–1522.
- Schroder I, Johnson E, de Vries S. (2003). Microbial ferric iron reductases. *FEMS Microbiol. Rev.* 27:427–447.
- Schwibbert K, Marin-Sanguino A, Bagyan I, Heidrich G, Lentzen G, Seitz H, et al. (2010). A blueprint of ectoine metabolism from the genome of the industrial producer *Halomonas elongata* DSM 2581T. *Environmental Microbiology* 13:1973–1994.
- Seegerer A, Langworthy TA, Stetter KO. (1988). *Thermoplasma acidophilum* and *Thermoplasma volcanium* sp. nov. from Solfatara Fields. *Systematic and Applied Microbiology* 10:161–171.
- Shiers DW, Ralph DE, Watling HR. (2010). A comparative study of substrate utilisation by *Sulfobacillus* species in mixed ferrous ion and tetrathionate growth medium. *Hydrometallurgy* 104:363–369.
- Simmons SL, DiBartolo G, Deneff VJ, Goltsman DSA, Thelen MP, Banfield JF. (2008). Population genomic analysis of strain variation in *Leptospirillum* group II bacteria involved in acid mine drainage formation. *Plos Biology* 6:1427–1442.
- Singer SW, Chan CS, Zemla A, VerBerkmoes NC, Hwang M, Hettich RL, et al. (2008). Characterization of cytochrome 579, an unusual cytochrome isolated from an iron-oxidizing microbial community. *Applied and Environmental Microbiology* 74:4454–4462.
- Singer SW, Erickson BK, VerBerkmoes NC, Hwang M, Shah MB, Hettich RL, et al. (2010). Posttranslational modification and sequence variation of redox-active proteins correlate with biofilm life cycle in natural microbial communities. *Isme J.* 4:1398–1409.
- Smith PF, Langworthy TA, Smith MR. (1975). Polypeptide Nature of Growth Requirement in Yeast Extract for *Thermoplasma Acidophilum*. *Journal of Bacteriology* 124:884–892.
- Sorek R, Zhu Y, Creevey CJ, Francino MP, Bork P, Rubin EM. (2007). Genome-wide experimental determination of barriers to horizontal gene transfer. *Science* 318:1449–1452.
- Spanova M, Daum G. (2011). Squalene - biochemistry, molecular biology, process biotechnology, and applications. *Eur. J. Lipid Sci. Technol.* 113:1299–1320.
- St John J. SeqPrep. <https://github.com/jstjohn/SeqPrep>.
- Stamatakis A. (2006). RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22:2688–2690.
- Stapley EO, Starkey RL. (1970). Decomposition of Cysteic Acid and Taurine by Soil Microorganisms. *Journal of General Microbiology* 64:77–84.
- Stookey LL. (1970). Ferrozine - a New Spectrophotometric Reagent for Iron. *Anal. Chem.*

42:779–&.

Strittmatter AW, Liesegang H, Rabus R, Decker I, Amann J, Andres S, et al. (2009). Genome sequence of *Desulfobacterium autotrophicum* HRM2, a marine sulfate reducer oxidizing organic carbon completely to carbon dioxide. *Environmental Microbiology* 11:1038–1055.

Sugiyama M, Suzuki S-I, Tonouchi N, Yokozeki K. (2003). Transaldolase/glucose-6-phosphate isomerase bifunctional enzyme and ribulokinase as factors to increase xylitol production from D-arabitol in *Gluconobacter oxydans*. *Biosci Biotechnol Biochem* 67:2524–2532.

Sumner LW, Amberg A, Barrett D, Beale MH, Beger R, Daykin CA, et al. (2007). Proposed minimum reporting standards for chemical analysis. *Metabolomics* 3:211–221.

Sun CL, Barrangou R, Thomas BC, Horvath P, Fremaux C, Banfield JF. (2013). Phage mutations in response to CRISPR diversification in a bacterial population. *Environmental Microbiology* 15:463–470.

Suzek BE, Huang H, McGarvey P, Mazumder R, Wu CH. (2007). UniRef: comprehensive and non-redundant UniProt reference clusters. *Bioinformatics* 23:1282–1288.

Tabb DL, McDonald WH, Yates JR. (2002). DTASelect and contrast: Tools for assembling and comparing protein identifications from shotgun proteomics. *J. Proteome Res.* 1:21–26.

Tabita FR, Hanson TE, Li HY, Satagopan S, Singh J, Chan S. (2007). Function, structure, and evolution of the RubisCO-like proteins and their RubisCO homologs. *Microbiol. Mol. Biol. Rev.*

Talavera G, Castresana J. (2007). Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Syst. Biol.* 56:564–577.

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. (2011). MEGA5: Molecular Evolutionary Genetics Analysis Using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. *Molecular Biology and Evolution* 28:2731–2739.

Tan GL, Shu WS, Hallberg KB, Li F, Lan CY, Zhou WH, et al. (2008). Culturable and molecular phylogenetic diversity of microorganisms in an open-dumped, extremely acidic Pb/Zn mine tailings. *Extremophiles* 12:657–664.

Taubert M, Baumann S, Bergen von M, Seifert J. (2011). Exploring the limits of robust detection of incorporation of ¹³C by mass spectrometry in protein-based stable isotope probing (protein-SIP). *Anal Bioanal Chem* 401:1975–1982.

Taubert M, Bergen M, Seifert J. (2013). Limitations in detection of ¹⁵N incorporation by mass spectrometry in protein-based stable isotope probing (protein-SIP). *Anal Bioanal Chem* 405:3989–3996.

Taubert M, Vogt C, Wubet T, Kleinstüber S, Tarkka MT, Harms H, et al. (2012). Protein-SIP enables time-resolved analysis of the carbon flux in a sulfate-reducing, benzene-degrading microbial consortium. *Isme J.* 6:2291–2301.

- Thamdrup B, Finster K, Hansen JW, Bak F. (1993). Bacterial Disproportionation of Elemental Sulfur Coupled to Chemical Reduction of Iron or Manganese. *Applied and Environmental Microbiology* 59:101–108.
- Thauer RK, Kaster A-K, Goenrich M, Schick M, Hiromoto T, Shima S. (2010). Hydrogenases from Methanogenic Archaea, Nickel, a Novel Cofactor, and H₂ Storage. *Annu. Rev. Biochem.* 79:507–536.
- Tholozan JL, Touzel JP, Samain E, Grivet JP, Prensier G, Albagnac G. (1992). *Clostridium neopropionicum* sp. nov., a strict anaerobic bacterium fermenting ethanol to propionate through acrylate pathway. *Arch. Microbiol.* 157:249–257.
- Thomson JF. (1963). *Biological Effects of Deuterium*. Pergamon Press.
- Trauger SA, Kalisak E, Kalisiak J, Morita H, Weinberg MV, Menon AL, et al. (2008). Correlating the transcriptome, proteome, and metabolome in the environmental adaptation of a hyperthermophile. *7*:1027–1035.
- Travisany D, Di Genova A, Sepulveda A, Bobadilla-Fazzini RA, Parada P, Maass A. (2012). Draft genome sequence of the *Sulfobacillus thermosulfidooxidans* Cutipay strain, an indigenous bacterium isolated from a naturally extreme mining environment in Northern Chile. *Journal of Bacteriology* 194:6327–6328.
- Tsaplina IA, Krasil'nikova EN, Zhuravleva AE, Egorova MA, Zakharchuk LM, Suzina NE, et al. (2008). The dependence of intracellular ATP level on the nutrition mode of the acidophilic bacteria *Sulfobacillus thermotolerans* and *Alicyclobacillus tolerans*. *Microbiology* 77:654–664.
- Tsaplina IA, Osipov GA, Bogdanova TI, NEDOREZOVA TP, Karavaiko GI. (1994). Fatty-Acid Composition of Lipids in Thermoacidophilic Bacteria of the Genus *Sulfobacillus*. *Microbiology* 63:459–464.
- Tsaplina IA, Zhuravleva AE, Egorova MA, Bogdanova TI, Krasil'nikova EN, Zakharchuk LM, et al. (2010). Response to oxygen limitation in bacteria of the genus *sulfobacillus*. *Microbiology* 79:13–22.
- Tyson GW, Banfield JF. (2008). Rapidly evolving CRISPRs implicated in acquired resistance of microorganisms to viruses. *Environmental Microbiology* 10:200–207.
- Tyson GW, Chapman J, Hugenholtz P, Allen EE, Ram RJ, Richardson PM, et al. (2004). Community structure and metabolism through reconstruction of microbial genomes from the environment. *Nature* 428:37–43.
- Uhlík O, Jecná K, Leigh MB, Macková M, Macek T. (2009). DNA-based stable isotope probing: a link between community structure and function. *Sci. Total Environ.* 407:3611–3619.
- United States Environmental Protection Agency. (2006). Abandoned Mine Lands Case Study. United States Environmental Protection Agency <http://www.epa.gov/aml/tech/imm.pdf>.

- Urich T, Gomes CM, Kletzin A, Frazão C. (2006). X-ray Structure of a self-compartmentalizing sulfur cycle metalloenzyme. *Science* 311:996–1000.
- van der Ploeg JR, Weiss MA, Saller E, Nashimoto H, Saito N, Kertesz MA, et al. (1996). Identification of sulfate starvation-regulated genes in *Escherichia coli*: A gene cluster involved in the utilization of taurine as a sulfur source. *J Biol Chem* 271:5438–5446.
- van der Werf MJ, Overkamp KM, Muilwijk B, Coulier L, Hankemeier T. (2007). Microbial metabolomics: Toward a platform with full metabolome coverage. *Anal Chem* 79:17–25.
- Vargas C, Jebbar M, Carrasco R, Blanco C, Calderon MI, Iglesias-Guerra F, et al. (2006). Ectoine as compatible solutes and carbon and energy sources for the halophilic bacterium *Chromohalobacter salexigens*. *J Appl Microbiol* 100:98–107.
<http://eutils.ncbi.nlm.nih.gov/entrez/eutils/elink.fcgi?dbfrom=pubmed&id=16405689&retmode=ref&cmd=prlinks>.
- Vignais PM, Billoud B. (2007). Occurrence, Classification, and Biological Function of Hydrogenases: An Overview. *Chemical Reviews* 107:4206–4272.
- Wade D. (1999). Deuterium isotope effects on noncovalent interactions between molecules. *Chemico-Biological Interactions* 117:191–217.
- WAKAI S, TSUJITA M, KIKUMOTO M, MANCHUR MA, Kanao T, Kamimura K. (2007). Purification and Characterization of Sulfide:Quinone Oxidoreductase from an Acidophilic Iron-Oxidizing Bacterium, *Acidithiobacillus ferrooxidans*. *Biosci Biotechnol Biochem* 71:2735–2742.
- Walker CB, la Torre de JR, Klotz MG, Urakawa H, Pinel N, Arp DJ, et al. (2010). *Nitrosopumilus maritimus* genome reveals unique mechanisms for nitrification and autotrophy in globally distributed marine crenarchaea. *Proc Natl Acad Sci U S A* 107:8818–8823.
- Wasser JS, Lawler RG, Jackson DC. (1996). Nuclear magnetic resonance spectroscopy and its applications in comparative physiology. *Physiological Zoology* 69:1–34.
- Watling HR, Perrot FA, Shiers DW. (2008). Comparison of selected characteristics of *Sulfobacillus* species and review of their occurrence in acidic and bioleaching environments. *Hydrometallurgy* 93:57–65.
- Welander PV, Doughty DM, Wu C-H, Mehay S, Summons RE, Newman DK. (2012). Identification and characterization of *Rhodospirillum rubrum* TIE-1 hopanoid biosynthesis mutants. *Geobiology* 10:163–177.
- White RA, Blainey, P. C., Fan HC, Quake SR. (2009). Digital PCR provides sensitive and absolute calibration for high throughput sequencing. *BMC Genomics* 10.
- Wilmes P, Bowen BP, Thomas BC, Mueller RS, Denev VJ, VerBerkmoes NC, et al. (2010). Metabolome-proteome differentiation coupled to microbial divergence. *mBio* 1.
- Wilmes P, Remis JP, Hwang M, Auer M, Thelen MP, Banfield JF. (2009). Natural acidophilic

biofilm communities reflect distinct organismal and functional organization. *Isme J.* 3:266–270.

Winogradsky S. (1887). Über Schwefelbakterien. *Botanische Zeitung* 45:489–600.

Wood AP, Kelly DP. (1983). Autotrophic and mixotrophic growth of three thermoacidophilic iron-oxidizing bacteria. *FEMS Microbiol. Lett.* 20:107–112.

Wood AP, Kelly DP. (1984). Growth and Sugar Metabolism of a Thermoaddophilic Iron-oxidizing Mixotrophic Bacterium. *Microbiology* 130:1337–1349.

Wright ES, Yilmaz LS, Noguera DR. (2012). DECIPHER, a Search-Based Approach to Chimera Identification for 16S rRNA Sequences. *Applied and Environmental Microbiology* 78:717–725.

Wu M, Eisen JA. (2008). A simple, fast, and accurate method of phylogenomic inference. *Genome Biol.* 9:R151. <http://genomebiology.com/2008/9/10/R151>.

Yelton AP, Comolli LR, Justice NB, Castelle C, Denev VJ, Thomas BC, et al. (2013). Comparative genomics in acid mine drainage biofilm communities reveals metabolic and structural differentiation of co-occurring archaea. *BMC Genomics* 14:485.

Yelton AP, Thomas BC, Simmons SL, Wilmes P, Zemla A, Thelen MP, et al. (2011). A Semi-Quantitative, Synteny-Based Method to Improve Functional Predictions for Hypothetical and Poorly Annotated Bacterial and Archaeal Genes. *PLoS Comp Biol* 7.

Zhang H, Guo W, Xu C, Zhou H, Chen X. (2013). Site-specific mutagenesis and functional analysis of active sites of sulfur oxygenase reductase from Gram-positive moderate thermophile *Sulfobacillus acidophilus* ... *Microbiological research*.

Ziegler S, Dolch K, Geiger K, Krause S, Asskamp M, Eusterhues K, et al. (2013). Oxygen-dependent niche formation of a pyrite-dependent acidophilic consortium built by archaea and bacteria. 1–13.