The ISME Journal (2010) 4, 520–530

www.nature.com/ismej

ORIGINAL ARTICLE

# Cultivation and quantitative proteomic analyses of acidophilic microbial communities

Christopher P Belnap[1,6], Chongle Pan[2,6], Nathan C VerBerkmoes[2], Mary E Power[3], Nagiza F Samatova[4], Rudolf L Carver[5], Robert L Hettich[2] and Jillian F Banfield[1,6]

[1]*Microbiology Graduate Group, Department of Plant and Microbial Biology, University of California, Berkeley, CA, USA;* [2]*Chemical Sciences Division, Oak Ridge National Laboratory, Oak Ridge, TN, USA;* [3]*Department of Integrative Biology, University of California, Berkeley, CA, USA;* [4]*Computer Science and Mathematics Division, Oak Ridge National Laboratory, Oak Ridge, TN, USA;* [5]*Iron Mountain Superfund Facility, Redding, CA, USA and* [6]*Microbiology Graduate Group, Department of Environmental Science, Policy and Management, University of California, Berkeley, CA, USA*

**Acid mine drainage (AMD), an extreme environment characterized by low pH and high metal concentrations, can support dense acidophilic microbial biofilm communities that rely on chemoautotrophic production based on iron oxidation. Field determined production rates indicate that, despite the extreme conditions, these communities are sufficiently well adapted to their habitats to achieve primary production rates comparable to those of microbial communities occurring in some non-extreme environments. To enable laboratory studies of growth, production and ecology of AMD microbial communities, a culturing system was designed to reproduce natural biofilms, including organisms recalcitrant to cultivation. A comprehensive metabolic labeling-based quantitative proteomic analysis was used to verify that natural and laboratory communities were comparable at the functional level. Results confirmed that the composition and core metabolic activities of laboratory-grown communities were similar to a natural community, including the presence of active, low abundance bacteria and archaea that have not yet been isolated. However, laboratory growth rates were slow compared with natural communities, and this correlated with increased abundance of stress response proteins for the dominant bacteria in laboratory communities. Modification of cultivation conditions reduced the abundance of stress response proteins and increased laboratory community growth rates. The research presented here represents the first description of the application of a metabolic labeling-based quantitative proteomic analysis at the community level and resulted in a model microbial community system ideal for testing physiological and ecological hypotheses.**

*The ISME Journal* (2010) **4,** 520–530; doi:10.1038/ismej.2009.139; published online 24 December 2009
**Subject Category:** integrated genomics and post-genomics approaches in microbial ecology
**Keywords:** proteomics; acid mine drainage; biofilm

## Introduction

Extremophilic microorganisms are found in a variety of conditions considered to be severely inhospitable for much of the biosphere (Brock, 1978; Rothschild and Mancinelli, 2001). Extensive research has documented species composition and distribution of extremophilic communities (Ferris and Ward, 1997; Pace, 1997; Ward *et al.*, 1998) and the specific adaptations that enable growth at extreme conditions, such as specialized lipid membranes and use of novel enzymes (Thomas and Dieckmann, 2002; van den Burg, 2003; Macalady *et al.*, 2004). However, it has also been shown that extremophiles must invest a fraction of their energy and metabolic resources to maintain correctly folded proteins, osmotic homeostasis, and to detoxify radical chemical species (Kawakami *et al.*, 2004; Baker-Austin *et al.*, 2005; Ram *et al.*, 2005). Few studies have measured autotrophic primary production within extreme environments (Giovannoni *et al.*, 1987) or evaluated metabolic responses in extreme conditions (Ram *et al.*, 2005) to assess how well these communities are adapted to their surroundings.

Extremophilic microbial communities that thrive in acid mine drainage (AMD) within the Richmond Mine system (Iron Mountain, CA, USA) have been the focus of extensive cultivation-independent studies (Bond *et al.*, 2000a, b; Tyson *et al.*, 2004; Baker *et al.*, 2006; Simmons *et al.*, 2008). These

Correspondence: JF Banfield, Department of Environmental Science, Policy and Management, University of California, Earth and Planetary Sciences Department, Berkeley, CA 94720, USA.
E-mail: jbanfield@berkeley.edu
[6]These authors contributed equally to this work.
Received 7 September 2009; revised 11 November 2009; accepted 11 November 2009; published online 24 December 2009

acidophilic biofilm-associated microbial communities flourish underground where AMD forms as the result of pyrite ($FeS_2$) oxidation (Baker and Banfield, 2003). Biofilms grow on the surfaces of hot (30–56 °C), extremely acidic (pH 0.5–1.2) pools and underground streams that contain sub-molar concentrations of iron and millimolar concentrations of zinc, copper, arsenic and other metals (Druschel *et al.*, 2004). Energy for microbial growth is conserved via aerobic oxidation of dissolved $Fe^{2+}$ and used to fix $N_2$ and $CO_2$ supplied from the air to support chemoautotrophic production (Bond *et al.*, 2000a; Tyson *et al.*, 2004).

Richmond Mine communities are dominated by chemoautotrophic iron-oxidizing bacteria belonging to *Leptospirillum* group II (Bond *et al.*, 2000b; Tyson *et al.*, 2004), including types related to *Leptospirillum ferriphilum* (Coram and Rawlings, 2002) and *Leptospirillum rubarum* (Goltsman *et al.*, 2009). *Leptospirillum* group II typically comprises the majority of cell biomass in natural samples, and are predicted to be the key primary producers (Tyson *et al.*, 2004). A related group of organisms, referred to as *Leptospirillum* group III, are capable of nitrogen fixation and comprise approximately 10% of biofilm communities (Tyson *et al.*, 2005). In addition, biofilms contain many low abundance archaeal species (Baker and Banfield, 2003; Baker *et al.*, 2006) that likely use mixotrophic and heterotrophic metabolisms (Dopson *et al.*, 2004). Other low abundance bacterial taxa from the Firmicutes and Actinobacteria lineages have also been characterized in AMD metagenomic data sets (Tyson *et al.*, 2004; Dick *et al.*, 2009).

Laboratory isolates have been paramount for detailed analysis of unique biological traits and industrial applications of extremophilic microorganisms (Brock and Freeze, 1969; Fiala and Stetter, 1986), including those cultured from Richmond Mine samples (Edwards *et al.*, 2000; Tyson *et al.*, 2005; Goltsman *et al.*, 2009). However, many extremophilic taxa remain un-cultured because of the difficulties associated with laboratory isolation. In recent years, attempts to isolate and culture environmental microorganisms have been aided by cultivation conditions that more closely resemble the natural environment (Connon and Giovannoni, 2002), including co-cultivation of two or more organisms (Kaeberlein *et al.*, 2002). Although cultivation of mixed communities is perhaps a more accurate representation of natural systems, functional analysis of individual microbial taxa within mixed consortia is challenging.

There are several methods for functional analysis of microbial communities. These include stable isotope probing, in which stable isotopes are used as a tracer, and amplification and sequencing (or microarray detection) of functional genes and messenger RNAs to monitor specific activities (for example, nitrogen cycle studies). Analysis of microbial communities via shotgun-proteomics (Ram *et al.*, 2005) and transcriptomic methods (Frias-Lopez *et al.*, 2008; Urich *et al.*, 2008) can provide related but distinct comprehensive snapshots of the activities of organisms in communities. For genomically well-characterized microbial communities, protein and transcript identifications can be linked to organisms and community structure. Recently, significant progress has been made using mass spectrometry-based proteomic analyses for natural microbial systems in which metagenomic information is available, such as communities in AMD (Ram *et al.*, 2005; Lo *et al.*, 2007; Denef *et al.*, 2009) and other environments (Sowell *et al.*, 2008; Verberkmoes *et al.*, 2008; Wilmes *et al.*, 2008).

In this study, a laboratory system was designed to cultivate biofilm consortia with species composition and metabolism comparable to natural AMD biofilms. A metabolic labeling-based quantitative proteomic analysis was used to verify that laboratory-cultivated, $^{15}N$-labeled biofilms are functionally similar to a natural biofilm from the Richmond Mine. Laboratory communities included active, low abundance and previously uncultivated organisms. An important product of this work is an approach to simultaneously and quantitatively analyze the metabolic behavior of microorganisms within communities and under a range of environmental conditions. The use of $^{15}N$-labeled biofilms provides a degree of quantification not approached in previous mass spectrometry-based microbial proteomics analyses of microbial communities.

## Materials and methods

*Environmental sampling*

Natural biofilm communities were sampled within the Richmond Mine. Biofilms collected for growth and production measurements were termed PP01 and PP02. Sample PP01 was collected at the AB site during November 2004 (Supplementary Figure 1). This site had been washed clear 13 days before sampling, and hence the biofilm was $\leqslant$13 days old. Three 25 $cm^2$ sections of biofilm were removed from randomly distributed locations of a broad (10 $m^2$), shallow (10–20 cm) pool (Figure 1a). Sample PP02 was collected approximately 50 m within the C drift (Supplementary Figure 1). At this site, three plastic enclosures were anchored within the pyrite sediment to form an isolated $15 \times 15$ cm area at the solution surface in which the biofilm forms (Figure 1b). Enclosures extended $\sim$10 cm below the surface of a 30 cm deep pool to prevent lateral growth of surface biofilm into the study areas, ensuring recolonization from planktonic cells (Supplementary Figure 2a). Newly grown biofilms were collected after 19 days. In the laboratory, production samples were centrifuged to remove AMD solution, stored at −80 °C, and lyophilized for 24–36 h. For production analysis, percent carbon of the lyophilized samples was measured using a

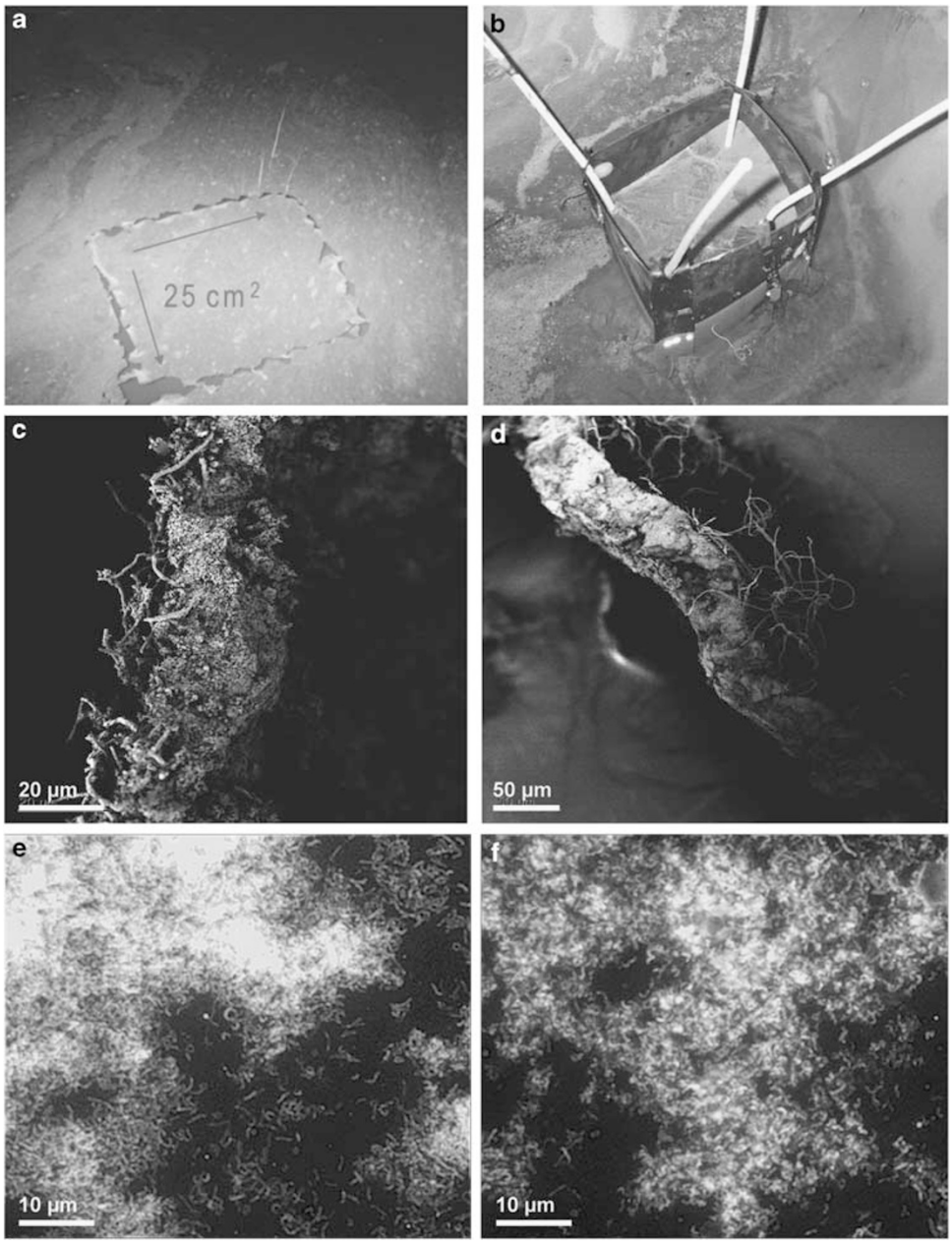


**Figure 1** Highly productive Richmond Mine biofilm communities. (**a**) Biofilm section sampled from a 13-day old floating, surface biofilm community (sample PP01, 1 of 3). Surrounding biofilm was covering the surface of a $\sim 10\,m^2$ pool. (**b**) 19-day old biofilm enclosure ($15 \times 15$ cm) anchored in the middle of a rapidly flowing stream (stream was $\sim 0.5$ m wide). Biofilm growth on the solution surface within the enclosure (shown) was sampled for growth and production measurements (sample PP02, 1 of 3). Before placement of the enclosure, there was no visible biofilm growth (for schematic, see Supplementary Figure 2). Carbon weight percentages from newly formed biomass (**a** and **b**) were used to calculate *in situ* production rates. Cross-section images of sample PP01 biofilm (**c**), and sample PP02 biofilm (**d**) captured using scanning electron microscopy illustrate biofilm thickness and density. Larger structures extending away from the biofilm are fungal filaments. Community composition was observed using fluorescence *in situ* hybridization (FISH) of sample PP01 (**e**) and sample PP02 (**f**) identifying *Leptospirillum* species (yellow), additional bacterial species (red) and archaeal species (blue). FISH images were taken at $630\times$ magnification. The color reproduction of this figure is available on the html full text version of the manuscript.

Lachat QC 8000 element analyzer (Lachat Instruments, Loveland, CO, USA). The natural sample used for quantitative proteomics analysis, termed AB1106, was collected in November 2006 near the site of sample PP01 (Supplementary Figure 1) and frozen at $-80\,^{\circ}C$ on site.

*Laboratory cultures*

Acid mine drainage biofilm communities were cultured in laboratory reactors stored in a dark, humid chamber heated to $40\,^{\circ}C$ (Figures 2a and b). Reactors (30 cm in length) contained a 0.5 cm deep bed of fine-grained pyrite sediment and 2 cm of

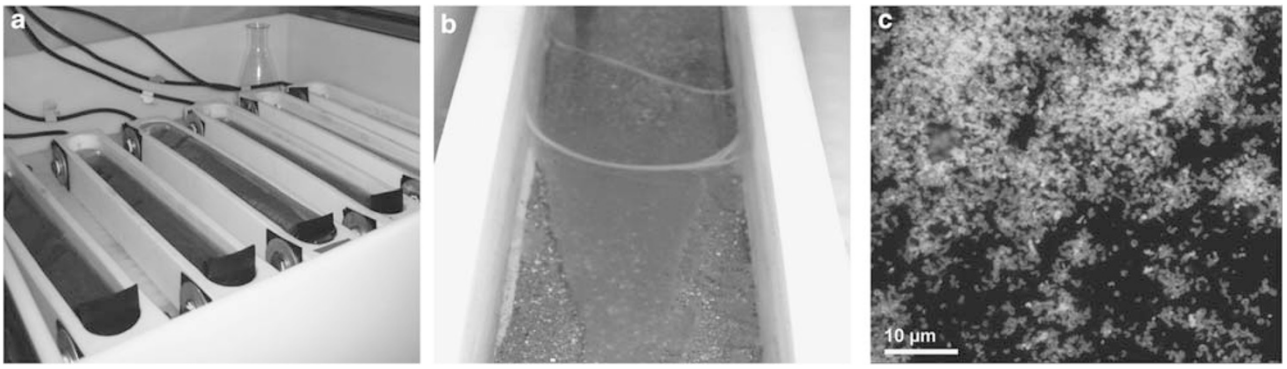


**Figure 2** Laboratory cultivation of acid mine drainage (AMD) biofilm communities. (**a**) 30 cm long laboratory flow through reactors were used to replicate environmental conditions (for schematic, see Supplementary Figure 2). (**b**) Thin, floating, laboratory biofilm community formed on the surface of laboratory reactor solutions, similar to natural biofilms. Reactor channel width is 3.5 cm. (**c**) Fluorescence *in situ* hybridization (FISH) indicated laboratory biofilm communities were dominated by *Leptospirillum* species (yellow), with low abundance archaea (blue) and other bacteria (red) present. Image is 630 × magnification. The color reproduction of this figure is available on the html full text version of the manuscript.

overlying, flowing solution. In all, 2 l of 9K medium (Supplementary Table 1) was continually recycled through each reactor, and 1 l of this reservoir was replaced with fresh medium every 7 days (Supplementary Figure 2b). No organic carbon substrate was provided. For 9 K-bio-reactor medium (BR) experiments, a modified 9 K medium was used and replenished every 5 days (Supplementary Table 1). Flow rates were between 0.2 and 0.3 ml per minute, yielding a residence time of 12–18 h. In both 9 K and 9 K-BR experiments, three reactors were inoculated by addition of 0.5 ml of diluted and dispersed community sample derived from the Richmond Mine (same inoculum for 9 K and 9 K-BR reactors). For both 9 K and 9 K-BR laboratory communities, production rates were measured as described above.

For proteomics experiments, labeled biofilm was grown using $^{15}N$-ammonium sulfate (Isotec, Miamisburg, OH, USA). Biofilm samples were collected from reactors, centrifuged at 7150 *g* for 3 min at 4 °C to remove the supernatant solution, and immediately frozen at −80 °C.

*Biofilm characterization*

Scanning electron microscopy was performed on biofilm samples PP01 and PP02 fixed in 4% paraformaldehyde and resuspended in 1:1 ethanol:-phosphate buffer solution. Samples were prepared using a critical point drier, and examined using a Hitachi S-5000 scanning electron microscope (Hitachi, Tokyo, Japan).

For community growth calculations, cells were stained with 4′,6-diamidino-2-phenylindole and counted at 630 times magnification using a Leica DMRX epifluorescence microscope. Planktonic cells were sampled at the PP01 and PP02 sites and assumed to be the initial number of cells. AMD solution was collected in 500 ml bottles and filtered using 0.2 µm pore size membrane filters (Hach, Loveland, CO, USA). Filters were then stained with 4′,6-diamidino-2-phenylindole, and cells were counted in three replicate fields of view per sample. To estimate final cell numbers, a known volume of fixed biofilm sample from each of the sites was physically disrupted in 9 K medium, and a dilute aliquot was placed on welled microscope slides. Slides were stained with 4′,6-diamidino-2-phenylindole and total cells in 10–20 replicate fields of view per sample were counted. Average fields of view counts were converted to total cells per milliliter. A similar procedure was used to quantify cell concentrations in laboratory reactors. Production rates of PP01, PP02 and laboratory samples were calculated using the amount of carbon in grams per $m^2$ of biomass divided by the growth period in days ($g\,C\,m^{-2}\,day^{-1}$). For all samples, three biological replicates were used to calculate production rates.

For proteomics samples, species abundances were estimated using fluorescence *in situ* hybridization (FISH) with lineage-specific probes as described earlier (Amann, 1995; Bond and Banfield, 2001) for the dominant community members in both environmental and laboratory biofilms. FISH cell counts from three replicate fields of view were used to calculate species percentages from total cell counts using 4′,6-diamidino-2-phenylindole (>1000 total cells counted per sample).

*Proteomics sample preparation, liquid chromatography–tandem mass spectrometry analysis and protein identification*

Whole-cell lysates were prepared from the field sample AB1106 and $^{15}N$-labeled laboratory biofilm samples as described earlier (Ram *et al.*, 2005; Lo *et al.*, 2007). For laboratory protein extraction, biofilm samples from two separate reactors were combined to yield greater biomass and reduce the influence of individual reactor variability. The whole-cell lysates of natural and laboratory samples were equally mixed before liquid chromatography–

tandem mass spectrometry (LC-MS/MS) analysis according to protein concentration (9 K) or sample weight (9 K-BR) (Supplementary Figure 2c). Protein mixtures were digested with trypsin and cleaned up for LC-MS/MS analysis as described earlier (Ram *et al.*, 2005; Lo *et al.*, 2007).

Whole-cell protein fractions were analyzed in duplicate by two-dimensional LC-MS/MS using a linear trap quadrupole (LTQ)-Orbitrap mass spectrometer (Thermo Fisher, San Jose, CA, USA). In all, 100 μg per sample was loaded onto a 150 μm ID back column packed with 2–3 cm of strong cation exchange resin (Luna, Phenomonex, Torrance, CA, USA) and 2–3 cm of C18 reverse-phase resin (Aqua, Phenomonex). A 100 μm ID PicoFrit front column (New Objective, Woburn, MA, USA) was packed with 15 cm of C18 reverse-phase resin. Two dimensional 24 h LC separation was carried out by the same method as Ram *et al.* (2005). Full scans (400–1700 *m/z*) were acquired in the Orbitrap at resolution of 30 000. Each full scan was followed by three data-dependent MS/MS scans acquired in the LTQ linear ion trap at 35% normalized collision energy. Protein identifications were made by reference to genomic data from similar communities (Lo *et al.*, 2007). All MS/MS scans were searched with the SEQUEST program (Eng *et al.*, 1994) against a protein sequence database composed from AMD genomic data (Supplementary Methods). The DTA-Select 1.9 program (Tabb *et al.*, 2002) was used to filter peptide identifications and assemble peptides into proteins using a minimum of two identified peptides per protein (Supplementary Methods). Functional categories were assigned to *Leptospirillum* group II strain five-way CG as described earlier (Denef *et al.*, 2009).

*Quantitative proteomic analysis*

ProRata was used for protein abundance ratio determination and confidence interval evaluation (Pan *et al.*, 2006a, b). Selected ion chromatograms were extracted for the light and heavy isotopologues of all identified peptides (Supplementary Figure 3). The mass-to-charge window was set to be 0.1 Da around the expected mass-to-charge ratios calculated from peptide sequence. Peptides were quantified and filtered with a minimum profile signal-to-noise ratio cutoff of two. Proteins were quantified based on both unique peptides that can be attributed to a single protein and non-unique peptides that are shared among multiple proteins. Using reverse database searching methods (Peng *et al.*, 2003), previous studies (Ram *et al.*, 2005; Lo *et al.*, 2007) have shown a false-positive rate of 1–2% at the protein and peptide level for this community. Only peptides that could be accurately quantified were retained, thus driving down the false-positive rate even further. In all, 90% confidence intervals were calculated for each protein and used as the range of error of the $\log_2$ abundance ratio. Relative protein abundances for individual species were normalized by centering the median $\log_2$-ratio value to zero. Proteins with $\log_2$-ratios $>1$ or $<-1$ ($>2\times$ difference) and confidence intervals excluding zero were considered different between samples. A one-sample statistical test distinguished functional category distributions that were significantly different from zero for *Leptospirillum* group II strain five-way CG. $\text{Log}_2$-ratios for each category were analyzed using the Shapiro–Wilk test of normality, and a *t*-test or Wilcoxon Rank-Sum test (for categories that did not follow a normal distribution) was used to identify categories that were significantly different in either the environmental or laboratory sample ($P<0.05$). All statistical calculations were performed using the R software environment (http://www.r-project.org/).

## Results

*Natural growth and production*

Production samples PP01 and PP02 were tens of microns thick, multi-species biofilm communities dominated by *Leptospirillum* spp. (Figures 1c–f). Environmental conditions, cell counts, averaged carbon and nitrogen concentrations, and growth and production rates for environmental samples are

**Table 1** Richmond mine and laboratory sample characterization

| *Sample* | *PP01* | *PP02* | *Laboratory (9 K)* | *Laboratory (9 K-BR)* |
|---|---|---|---|---|
| Temperature (°C) | 37 | 42 | 40 | 40 |
| pH[a] | 0.9 | 0.7 | 1.1 | 1.0 |
| Growth period (days) | 13 | 19 | 12.3 ± 1.3 | 7 |
| Initial planktonic cell number (cells ml$^{-1}$) | $9.16\times10^{3}$ | $1.53\times10^{4}$ | $2.01\times10^{5}$ | $4.04\times10^{4}$ |
| Biofilm cell number (cells ml$^{-1}$) | $9.26\times10^{11}$ | $1.60\times10^{11}$ | $2.06\times10^{10}$ | $1.21\times10^{9}$ |
| Community doubling time (hours) | 11.7 ± 0.2 | 19.3 ± 1.1 | 17.8 ± 2.5 | 11.3 ± 0.4 |
| Measured growth area (m$^2$) | 0.063 | 0.023 | 0.01 | 0.01 |
| Measured dry weight (g) | 1.07 ± 0.19 | 0.70 ± 0.09 | 0.031 ± 0.007 | 0.010 ± 0.004 |
| Carbon weight (%) | 37.9 ± 0.53 | 30.4 ± 0.34 | 34.6 ± 4.5 | 34.0 ± 2.1 |
| Calculated production rate (g C m$^{-2}$ day$^{-1}$) | 0.50 ± 0.15 | 0.49 ± 0.12 | 0.10 ± 0.07 | 0.07 ± 0.05 |

For all conditions $n=3$, and standard error is shown for averages.
[a]Laboratory pH values measured at the time of sampling.

shown in Table 1. Community doubling times for the *in situ* biofilm samples are estimated to be ~12–19 h (Table 1). This calculation assumes exponential growth over the entire time period, and thus may underestimate maximum growth rates (especially for sample PP02, for which a longer growth period was sampled). Production rates determined from carbon concentrations in the field study areas were 0.50 and 0.49 g C $m^{-2}$ $day^{-1}$ for biofilm samples PP01 and PP02, respectively (Table 1).

*Laboratory growth and production*

Using the original 9 K acidophilic growth medium, laboratory biofilm formation occurred in 11–15 days, with an average development time of 12.3 days ($n = 3$) after the planktonic cell density in each reactor reached ~$10^5$ cells $ml^{-1}$ (Table 1). Average doubling time of the laboratory biofilm communities was estimated to be 17.8 ± 2.5 h (Table 1). Laboratory-grown biofilm communities resembled field-collected biofilms in morphology and color (Figure 2b), and FISH analysis indicated similar species composition and abundance to natural communities from the Richmond Mine (Figure 2c and Supplementary Table 2) (Bond and Banfield, 2001). Laboratory communities were dominated by *Leptospirillum* group II, with *Leptospirillum* group III present at lower abundance (Supplementary Table 2). In addition, many low abundance members were identified using FISH including several archaeal species such as *Ferroplasma* types I and II and G-plasma (Supplementary Table 2), as well as Actinobacteria lineages (data not shown). Production rates calculated for laboratory communities were 0.1 g C $m^{-2}$ $day^{-1}$ (Table 1).

Owing to relatively slow growth rates compared with sample PP01, and relatively lower production rates compared with both natural samples PP01 and PP02, cultivation of laboratory communities was repeated using a modified 9 K medium (referred to as 9 K-BR). The 9 K-BR medium contained lowered nitrogen and phosphorus salt concentrations to more closely resemble natural AMD solutions from the Richmond Mine (Supplementary Table 1), and solution recharge rate was increased. Using 9 K-BR medium, biofilm development time decreased to 7 days and community doubling time was 11.3 ± 0.4 h ($n = 3$) (Table 1), significantly faster than simultaneous growth using the original 9 K medium (*t*-test, $P = 0.0446$, $n = 3$). However, community production rates were not significantly different compared with 9 K laboratory samples (Table 1).

*Quantitative proteomic comparison of AB1106 and 9 K laboratory communities*

The AB1106 biofilm sample used for quantitative proteomics comparison was examined using FISH to confirm general species similarity to $^{15}N$-labeled 9 K laboratory samples. Dominant species were in similar abundance for the AB1106 and 9 K laboratory samples, with the exception of *Leptospirillum* group III, which was present at lower abundance in the laboratory sample (Supplementary Table 2). Although fungi are present in laboratory communities, incomplete genomic data sets for eukaryotic organisms in Richmond Mine samples preclude proteomic characterization. Mass spectral analysis confirmed that the nitrogen composition of bacterial and archaeal proteins in laboratory-grown biofilms was >95% $^{15}N$. In total, the relative abundances of 2687 proteins were quantified between AB1106 and the 9 K laboratory sample (Supplementary Table 3). The abundances of the majority of *Leptospirillum* group II proteins (63 and 68% for strain types five-way CG and UBA, respectively) were within a twofold cutoff in the field and laboratory biofilms (Supplementary Table 3). Included within this group are proteins that perform essential functions such as energy conversion, nucleotide metabolism, transcription, translation and coenzyme metabolism. Examples are many of the NADH dehydrogenase subunits, ATPase synthase proteins, most ribosomal proteins and transfer RNA synthetases, and proteins involved in coenzyme-A, cobalamin, riboflavin and biotin synthesis (Supplementary Table 6). The majority of proteins for low abundance community members, such as *Leptospirillum* group III and G-plasma, showed a similar trend (Supplementary Tables 3 and 6). In addition, proteomic data confirmed the presence of other low abundance bacterial and archaeal species in laboratory communities such as Actinobacteria, Firmicutes and several archaea of the Thermoplasmatales lineage (Supplementary Tables 3 and 6).

Functional categories and individual proteins found to be more abundant in either sample were examined to identify metabolic differences that may have existed between natural and laboratory communities. Many more proteins were detected for the dominant bacterium, *Leptospirillum* group II strain five-way CG; hence, we focused on this organism to identify key differences between the field and laboratory samples. *Leptospirillum* group II five-way CG functional categories found to be significantly more abundant in the natural community were energy production and conversion ($P = 0.0098$, $n = 69$), cell wall/membrane/envelope biogenesis ($P = 0.0022$, $n = 40$), cell motility ($P = 0.0071$, $n = 21$), and intracellular trafficking, secretion and vesicular transport ($P = 0.0406$, $n = 25$). In contrast, functional categories significantly more abundant in the 9 K laboratory community included defense mechanisms ($P = 0.0089$, $n = 25$), and transcriptional proteins ($P = 0.0098$, $n = 29$).

*Leptospirillum* group II proteins involved in energy conversion pathways found to be highly abundant in AB1106 included several c-type cytochromes. In particular, the earlier characterized *Leptospirillum* c-type cytochrome 572 (Jeans *et al.*, 2008) implicated in iron oxidation was 5.3–6.0 times

more abundant in the natural sample (Supplementary Table 4). Several core energetic proteins, such as two NADH dehydrogenase and ATP synthase subunits, were also more abundant in the AB1106 sample (Supplementary Table 4). Proteins involved in cell envelope/membrane biogenesis processes identified to be more abundant in the AB1106 sample included those assigned to peptidoglycan and lipopolysaccharide biosynthetic pathways (Supplementary Table 4). For example, undecaprenyl-phosphate galactose phosphotransferase and glucose-1-phosphate thymidyltransferase, proteins putatively involved in biosynthesis of polysaccharide components of lipopolysaccharide, were 68.5 and 5.3 times more abundant in the AB1106 sample, respectively (Supplementary Table 4). Highly abundant cell motility proteins included three proteins involved in flagellar assembly that were 2.5–12.1 times more abundant in the AB1106 sample (Supplementary Table 4). Intracellular trafficking, secretion and vesicular transport proteins more abundant in AB1106 included proteins putatively belonging to general secretion pathways (Supplementary Table 6), although significance for this category was not strong.

*Leptospirillum* group II defense proteins found to be over-represented in the laboratory included radical repair enzymes such as a peroxiredoxin and a thioredoxin peroxidase (Supplementary Table 5). In laboratory samples, highly abundant proteins in other functional categories were also indicative of oxidative stress, such as a peptide methionine sulfoxide reductase and proteins involved in iron–sulfur cluster assembly (Supplementary Table 5). In addition to oxidative stress response proteins, osmotic shock response proteins such as ectoine synthase, a protein likely involved in the biosynthetic pathway of the compatible solute molecule ectoine (Louis and Galinski, 1997), was 9.8 times more abundant in 9 K laboratory-associated *Leptospirillum* group II cells. Other proteins involved in general defense pathways, such as heat shock chaperones and a phage shock protein, were also more abundant in the laboratory sample (Supplementary Table 5). Although transcription proteins were significantly increased in the 9 K laboratory community, several of these were identified as transcriptional regulators, a possible indication of environmental response (Supplementary Table 6).

Low abundance species were characterized at less detail because of fewer quantified proteins (Supplementary Table 3). However, protein abundance ratios for several archaeal species indicate a trend similar to that observed for *Leptospirillum* group II. For example, several chaperone proteins belonging to *Ferroplasma* type I and A-plasma were ~4–20 times more abundant in the laboratory community (Supplementary Table 6), suggesting high stress levels for these organisms. In addition, several archaeal proteins indicative of higher production rates were observed to be more abundant in the natural sample, similar to the response of *Leptospirillum* group II. Specifically, G-plasma and E-plasma proteins involved in fatty acid metabolism such as acetyl CoA acetyltransferase, acyl-CoA synthetase and enoyl-CoA hydratase were more abundant in AB1106 (Supplementary Table 6).

*Quantitative proteomic comparison of AB1106 and 9 K-BR laboratory communities*

Quantitative proteomics experiments were repeated using the identical AB1106 sample compared with 9 K-BR laboratory communities (Supplementary Table 3). In total, 2588 proteins were quantified between AB1106 and 9 K-BR laboratory communities. *Leptospirillum* group II five-way CG functional categories that were significantly more abundant in AB1106 compared with the 9 K-BR community included energy production and conversion ($P = 0.0069$, $n = 69$), and amino acid transport and metabolism ($P = 0.0001$, $n = 79$). Signal transduction mechanisms were more abundant in the 9 K-BR community, although the level of significance was low ($P = 0.0322$, $n = 18$).

Under the new growth conditions, defense mechanisms were no longer significantly more abundant in the laboratory community, and fewer *Leptospirillum* group II oxidative damage repair proteins were more abundant in the laboratory compared with the natural biofilm (Supplementary Tables 5 and 7). Several of the proteins involved in iron–sulfur cluster assembly and a thioredoxin peroxidase decreased to relative abundance levels comparable to those of AB1106 (within 2 ×; Supplementary Table 5). Other proteins involved in metabolic stress pathways also decreased in the modified growth medium. For example, the compatible solute biosynthesis protein ectoine synthase decreased nearly 10-fold in the 9 K-BR laboratory sample to within twofold of the AB1106 sample (Supplementary Table 5). However, an Hsp20 heat shock and osmotic shock protein increased in the 9 K-BR compared with the original 9 K medium indicating stress response was not completely eliminated (Supplementary Table 5).

As indicated above, proteins involved in energy conversion pathways remained more abundant in sample AB1106 compared with the 9 K-BR laboratory sample (Supplementary Table 4). In addition, several enzymes involved in cell envelope biosynthesis remained more abundant in AB1106 (Supplementary Table 4), although the cell envelope/membrane biogenesis category was not significantly different.

In general, fewer proteins from low abundance organisms were detected in the comparison between AB1106 and 9 K-BR, therefore interpretation of protein abundance changes caused by the modification of cultivation conditions was difficult for these organisms. However, the general increase in the fraction of proteins within two fold difference for

low abundance bacterial and archaeal species suggest an overall increase in community similarity between the laboratory and environmental biofilms (Supplementary Table 3).

## Discussion

In this study, we measured high rates of biomass production for naturally occurring AMD biofilm communities and calculated rates of primary production as an estimate of *in situ* chemoautotrophic production. To further analyze growth of the AMD community, we designed a laboratory culturing system to grow multi-species biofilms similar to those present in the natural environment. Laboratory culturing of microbial communities representative of natural counterparts is difficult; hence, we used a metabolic labeling-based quantitative proteomic method to compare natural and laboratory community metabolic profiles and confirm the presence and activity of low abundance members. We also compared field and laboratory biofilms to identify metabolic differences in the dominant bacterium, *Leptospirillum* group II strain five-way CG. Through culturing improvements, we decreased metabolic stress in the laboratory communities and increased community growth rates.

*Acid mine drainage chemoautotrophic production*

We estimate chemoautotrophic production within the Richmond Mine to be in the range of $0.5\,g\,C\,m^{-2}\,day^{-1}$ (Table 1). Production rates calculated using percent carbon from total biomass provides a basic estimate of *in situ* productivity under extreme conditions in which sophisticated instrumentation cannot be deployed. This method also considers exopolysaccharides that comprise the biofilm matrix, a potentially substantial sink of assimilated carbon. Owing to the occurrence of upstream communities at the field sample sites, input of cells and dissolved organic carbon may have contributed to total biomass accumulation within the study areas. However, given the experimental design, influx of cells and dissolved organic carbon was likely balanced by efflux (Supplementary Figure 2a). The calculated production rates for natural AMD communities are comparable to, or greater than, rates for other non-extreme microbial ecosystems (Table 2). To produce the observed amount of carbon biomass, we estimate the Richmond Mine communities oxidize more than 200 g of $Fe^{2+}\,m^{-2}\,day^{-1}$ (Supplementary Methods). Comparison of production rates across different environments (Table 2) indicates energy resource availability is perhaps far more important than geochemical parameters (for example, bacterial production rates in the nutrient-poor deep ocean are lower than production rates in an extremely low pH AMD environment). We conclude that, despite extremes of very low pH and high concentrations of toxic metals, production by natural AMD microbial communities occurs at a rate comparable to rates in many non-extreme environments. In fact, it is the extreme conditions (that is, large pH gradients coupled to the availability of electrons from dissolved ferrous iron), which drive the chemiosmotic energy production (Rawlings, 2005) that enables acidophiles to offset metabolic stress and achieve high production rates.

*Comparison of 9K laboratory communities to environmental samples*

After the establishment of initial planktonic communities, 9 K laboratory community doubling times were comparable to field sample PP02, but considerably slower than sample PP01 (Table 1). In addition, field production rates for both samples

**Table 2** General comparison of production rates across different environments

| *Environment* | *Production ($g\,C\,m^{-2}\,day^{-1}$)* | *Source* |
|---|---|---|
| Freshwater lake (epiliminion; annual avg) | 2.15 | Hadas *et al.* (2001) |
| Eutrophic ocean (NE Atlantic; June avg) | 1.77 | Morel *et al.* (1996) |
| Arctic ice shelf (primary; annual avg) | 1.54 | Mueller *et al.* (2005) |
| Mesotrophic ocean (NE Atlantic; June avg) | 1.18 | Morel *et al.* (1996) |
| Eutrophic Thermal lake | 1.10 | Forsyth and McColl (1974) |
| **AMD biofilm PP01** | **0.50** | **This study** |
| **AMD biofilm PP02** | **0.49** | **This study** |
| Sub-tidal cyanobacterial biofilm | 0.38 | Lugomela *et al.* (2005) |
| Freshwater lake (oxic/anoxic; annual avg) | 0.38 | Hadas *et al.* (2001) |
| Oligotrophic ocean (NE Atlantic; June avg) | 0.35 | Morel *et al.* (1996) |
| Antarctic lake benthic mat (annual avg) | 0.04 | Hawes *et al.* (2001) |
| Arctic ice shelf (bacterial; annual avg) | 0.002 | Mueller *et al.* (2005) |
| Deep ocean (bacterial) | 0.001 | Nagata *et al.* (2000) |

Abbreviations: AMD, acid mine drainage; avg, average. Production rates determined for AMD biofilms (in bold) are compared to previously reported values for other extreme and non-extreme environments. Many previously reported production rates are photosynthetically based. Rates specifically calculated for primary or bacterial production are denoted. All units were converted to grams of carbon $m^{-2}\,day^{-1}$.

PP01 and PP02 were approximately five times higher than those measured in the laboratory ($0.1\,g\,C\,m^{-2}\,day^{-1}$, Table 1). We attribute differences in community production to non-optimal laboratory growth conditions, a finding supported by the abundance of metabolic stress proteins observed in quantitative proteomic comparisons. Specifically, stress-related *Leptospirillum* group II proteins more abundant in the laboratory biofilms included the iron–sulfur cluster assembly proteins, which form the center of many important redox active enzymes (Supplementary Table 5). It has been suggested that upregulation of the iron–sulfur cluster assembly operon occurs during periods of oxidative stress in response to free radical damage to the redox active sites of iron–sulfur proteins (Alamuri *et al.*, 2006). Furthermore, a separate abundant oxidative stress protein in the laboratory sample, peptide methionine sulfoxide reductase, has been shown to have a direct role in protection from reactive oxygen species (Moskovitz *et al.*, 1995). The increased metabolic stress in the laboratory community likely limited biomass accumulation rates.

Many proteins characteristic of increased growth and production were relatively more abundant in the environmental sample AB1106. In particular, the abundance of several c-type cytochromes (Supplementary Table 4) implicated in iron oxidation suggests overall increased iron oxidation rates in field communities. In addition, several proteins involved in lipopolysaccharide production were several times more abundant in the field sample AB1106, and may have contributed toward increased biofilm formation (Supplementary Table 4). Other cell envelope biosynthesis proteins abundant in the environmental sample, such as those involved in peptidoglycan synthesis (Supplementary Table 4), suggest generally increased cell growth rates. The higher abundances of flagella biosynthetic proteins in the field community (Supplementary Table 4) may reflect more pronounced chemical gradients, likely because of more extensive biofilm development.

Archaea within the Richmond Mine system likely rely on bacterial production for a steady supply of organic substrates, such as carbohydrate and lipid polymers generated for biofilm formation. Higher abundances of archaeal fatty acid utilization proteins observed in the environmental sample probably correlate with increased heterotrophic production that may be sustained by faster growth and primary production rates associated with *Leptospirillum* group II. In the laboratory community, increased archaeal stress response proteins correlated with the stress response of *Leptospirillum* group II. As community production is likely dependent on all microbial species because of mutualistic and synergistic interactions (Hallmann *et al.*, 1992; Johnson, 1998), stress to the archaea may have further decreased net production and growth rates in laboratory reactors.

*Comparison of 9 K-BR laboratory communities to environmental samples*

Using the modified culture conditions, significant increases in growth rates were observed for the laboratory community (9 K-BR, Table 1). Quantitative proteomic results comparing laboratory communities cultured using the 9 K-BR medium with sample AB1106 indicated an overall increase in similarity of the laboratory and field sample. For *Leptospirillum* group II as well as low abundance bacterial and archaeal taxa, the fraction of proteins considered to be within similar abundance was greater compared with the original 9 K laboratory comparison (Supplementary Table 3). Furthermore, many functional categories identified as significantly higher in either AB1106 or the 9 K laboratory community were no longer differentially abundant (that is, cell wall/membrane/envelope biogenesis; cell motility; intracellular trafficking, secretion, and vesicular transport; defense mechanisms; transcription).

Although *Leptospirillum* group II energy production and conversion proteins remained significantly more abundant in the natural sample even after cultivation conditions were changed, several enzymes potentially involved in iron oxidation, including cytochrome 572, reached abundance levels comparable to those of natural biofilms (Supplementary Table 4). The reduction of stress response proteins in laboratory communities (Supplementary Table 5), as well as the increase of specific enzymes involved in energy generation, likely contributed to increased community doubling times in the modified laboratory solutions (Table 1). However, remaining metabolic stress (Supplementary Table 5), and the under representation of other core energetic proteins in 9 K-BR laboratory communities (Supplementary Table 4) may explain the low production rates of this sample compared with the environmental biofilm, similar to observations made for the 9 K laboratory community. We conclude that, despite important growth improvements, the low production rates for 9 K-BR laboratory communities suggest the need for further culture optimization in future experiments. If the laboratory communities become substrate limited as they reach high cell densities, production rates might be increased by increasing solution replenishment rates (less recycling). This would increase the supply of reduced iron, dissolved carbon dioxide and oxygen to laboratory communities, as well as reduce build up of inhibitory waste products.

There was an abundance of *Leptospirillum* group II proteins associated with amino acid metabolism in the AB1106 biofilm compared with the 9 K-BR laboratory biofilm. This difference was not observed earlier under the original cultivation conditions. Particularly abundant in AB1106 (or depleted in the 9 K-BR laboratory sample) were two putative urea cycle proteins (ornithine carbamoyltransferase and acetylornithine aminotransferase—Supplementary

Table 7), perhaps associated with the decrease in ammonia concentrations in 9 K-BR solutions. Lowered nitrogen concentrations in the laboratory did not appear to inhibit growth, but the high abundance of urea cycle proteins in the AB1106 sample suggest comparatively high levels of fixed nitrogen availability in the natural environment. Although the AB1106 had a higher representation of the potentially nitrogen-fixing organism *Leptospirillum* group III than either laboratory-grown community (Supplementary Table 2), we cannot attribute the high nitrogen availability to the presence of *Leptospirillum* group III in the environmental biofilm as nitrogen fixation proteins were not identified by proteomics. Nitrogen generated from upstream communities or released by decay of old sunken biofilms may be sufficient to support growth in the mine locations where samples were collected.

In summary, community cultivation and metabolic $^{15}$N-labeling enabled extensive quantitative proteomic analyses to compare relative protein abundances in laboratory and natural consortia. Thus, we could quantitatively compare protein abundance levels and monitor changes at the functional level for the dominant organism within multispecies AMD biofilms. To our knowledge, this is the first application of quantitative proteomics to functionally compare laboratory and natural microbial communities. The present study also illustrates the use of quantitative proteomics as a tool for optimization of laboratory growth of mixed communities. This research shows that it is possible to use changes in protein abundances to detect metabolic responses in mixed microbial systems. Similar methods could be adapted to optimize important enzymatic pathways associated with microbial communities involved in technologies such as lignin degradation, wastewater treatment and mineral bioleaching.

## Acknowledgements

Mr Ted Arman (President, Iron Mountain Mines) and Dr Richard Sugarek are thanked for site access and other assistance. We also thank Mike Zach, Andy Thompson and Becca Ryals for technical assistance. Vincent Denef and Paul Wilmes are thanked for helpful discussions. This research was supported by the US Department of Energy Genomics: GTL and ASCR programs, the NSF Biocomplexity Program and the NASA Astrobiology Institute.

## References

Alamuri P, Mehta N, Burk A, Maier RJ. (2006). Regulation of the *Helicobacter pylori* Fe-S cluster synthesis protein NifS by iron, oxidative stress conditions, and Fur. *J Bacteriol* **188**: 5325–5330.

Amann RI. (1995). Fluorescently labelled, rRNA-targeted oligonucleotide probes in the study of microbial ecology. *Mol Ecol* **4**: 543–554.

Baker-Austin C, Dopson M, Wexler M, Sawers RG, Bond PL. (2005). Molecular insight into extreme copper resistance in the extremophilic archaeon '*Ferroplasma acidarmanus*' Fer1. *Microbiology* **151**: 2637–2646.

Baker BJ, Banfield JF. (2003). Microbial communities in acid mine drainage. *FEMS Microbiol Ecol* **44**: 139–152.

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.

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. *Microbial Ecol* **2**: 149–161.

Bond PL, Druschel GK, Banfield JF. (2000a). Comparison of acid mine drainage microbial communities in physically and geochemically distinct ecosystems. *Appl Environ Microbiol* **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. *Appl Environ Microbiol* **66**: 3842–3849.

Brock TD. (1978). *Thermophilic Microorganisms and Life at High Temperatures*. Springer-Verlag: New York, pp 11–37.

Brock TD, Freeze H. (1969). *Thermus aquaticus* gen. n. and sp. n., a nonsporulating extreme thermophile. *J Bacteriol* **98**: 289–297.

Connon SA, Giovannoni SJ. (2002). High-throughput methods for culturing microorganisms in very-low-nutrient media yield diverse new marine isolates. *Appl Environ Microbiol* **68**: 3878–3885.

Coram NJ, Rawlings DE. (2002). Molecular relationship between two groups of the genus *Leptospirillum* and the finding that *Leptospirillum ferriphilum* sp. nov. dominates South African commercial biooxidation tanks that operate at 40 degrees C. *Appl Environ Microbiol* **68**: 838–845.

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. *Environ Microbiol* **11**: 313–325.

Dick GJ, Andersson A, Baker BJ, Simmons SS, Thomas BC, Yelton AP *et al.* (2009). Community-wide analysis of microbial genome sequence signatures. *Genome Biol* **10**: R85.

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. *Appl Environ Microbiol* **70**: 2079–2088.

Druschel G, Baker B, Gihring T, Banfield J. (2004). Acid mine drainage biogeochemistry at Iron Mountain, California. *Geochemical Transactions* **5**: 13.

Edwards KJ, Bond PL, Gihring TM, Banfield JF. (2000). An archaeal iron-oxidizing extreme acidophile important in acid mine drainage. *Science* **287**: 1796–1799.

Eng JK, McCormack AL, Yates III JR. (1994). An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J Am Soc Mass Spectrometry* **5**: 976–989.

Ferris MJ, Ward DM. (1997). Seasonal distributions of dominant 16S rRNA-defined populations in a hot spring microbial mat examined by denaturing gradient gel electrophoresis. *Appl Environ Microbiol* **63**: 1375–1381.

Fiala G, Stetter KO. (1986). *Pyrococcus furiosus* sp. nov. represents a novel genus of marine heterotrophic archaebacteria growing optimally at 100 °C. *Arch Microbiol* **145**: 56–61.

Frias-Lopez J, Shi Y, Tyson GW, Coleman ML, Schuster SC, Chisholm SW *et al.* (2008). Microbial community gene expression in ocean surface waters. *Proc Natl Acad Sci USA* **105**: 3805–3810.

Giovannoni SJ, Revsbech NP, Ward DM, Castenholz RW. (1987). Obligately phototrophic *Chloroflexus*: primary production in anaerobic hot spring microbial mats. *Arch Microbiol* **147**: 80–87.

Goltsman DSA, Denef VJ, Singer SW, VerBerkmoes NC, Lefsrud M, Mueller R *et al.* (2009). Community genomic and proteomic analysis of chemoautotrophic, iron-oxidizing '*Leptospirillum rubarum*' (Group II) and *Leptospirillum ferrodiazotrophum* (Group III) in acid mine drainage biofilms. *Appl Environ Microbiol* **75**: 4599–4615.

Hallmann R, Friedrich A, Koops HP, Pommerening-Roeser A, Rohde K, Zenneck C *et al.* (1992). Physiological characteristics of *Thiobacillus ferrooxidans* and *Leptospirillum ferrooxidans* and physicochemical factors influence microbial metal leaching. *Geomicrobiol J* **10**: 193–206.

Jeans C, Singer SW, Chan CS, VerBerkmoes NC, Shah M, Hettich RL *et al.* (2008). Cytochrome 572 is a conspicuous membrane protein with iron oxidation activity purified directly from a natural acidophilic microbial community. *ISME J* **2**: 542–550.

Johnson DB. (1998). Biodiversity and ecology of acidophilic microorganisms. *FEMS Microbiol Ecol* **27**: 307–317.

Kaeberlein T, Lewis K, Epstein SS. (2002). Isolating 'uncultivable' microorganisms in pure culture in a simulated natural environment. *Science* **296**: 1127–1129.

Kawakami R, Sakuraba H, Kamohara S, Goda S, Kawarabayasi Y, Ohshima T. (2004). Oxidative stress response in an anaerobic hyperthermophilic archaeon: presence of a functional peroxiredoxin in *Pyrococcus horikoshii*. *J Biochem* **136**: 541–547.

Lo I, Denef 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* **143**: 1141–1149.

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**: 1433–4909.

Moskovitz J, Rahman MA, Strassman J, Yancey SO, Kushner SR, Brot N *et al.* (1995). *Escherichia coli* peptide methionine sulfoxide reductase gene: regulation of expression and role in protecting against oxidative damage. *J Bacteriol* **177**: 502–507.

Pace NR. (1997). A molecular view of microbial diversity and the biosphere. *Science* **276**: 734–740.

Pan C, Kora G, McDonald WH, Tabb DL, VerBerkmoes NC, Hurst GB *et al.* (2006a). ProRata: a quantitative proteomics program for accurate protein abundance ratio estimation with confidence interval evaluation. *Anal Chem* **78**: 7121–7131.

Pan C, Kora G, Tabb DL, Pelletier DA, McDonald WH, Hurst GB *et al.* (2006b). Robust estimation of peptide abundance ratios and rigorous scoring of their variability and bias in quantitative shotgun proteomics. *Anal Chem* **78**: 7110–7120.

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.

Ram RJ, VerBerkmoes NC, Thelen MP, Tyson GW, Baker BJ, Blake II RC *et al.* (2005). Community proteomics of a natural microbial biofilm. *Science* **308**: 1915–1920.

Rawlings D. (2005). Characteristics and adaptability of iron- and sulfur-oxidizing microorganisms used for the recovery of metals from minerals and their concentrates. *Microbial Cell Factories* **4**: 13.

Rothschild LJ, Mancinelli RL. (2001). Life in extreme environments. *Nature* **409**: 1092–1101.

Simmons SL, DiBartolo G, Denef 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**: e177.

Sowell SM, Wilhelm LJ, Norbeck AD, Lipton MS, Nicora CD, Barofsky DF *et al.* (2008). Transport functions dominate the SAR11 metaproteome at low-nutrient extremes in the Sargasso Sea. *ISME J* **3**: 93–105.

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

Thomas DN, Dieckmann GS. (2002). Antarctic sea ice—a habitat for extremophiles. *Science* **295**: 641–644.

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.

Tyson GW, Lo I, Baker BJ, Allen EE, Hugenholtz P, Banfield JF. (2005). Genome-directed isolation of the key nitrogen fixer *Leptospirillum ferrodiazotrophum* sp. nov. from an acidophilic microbial community. *Appl Environ Microbiol* **71**: 6319–6324.

Urich T, Lanzen A, Qi J, Huson DH, Schleper C, Schuster SC. (2008). Simultaneous assessment of soil microbial community structure and function through analysis of the meta-transcriptome. *PLoS ONE* **3**: e2527.

van den Burg B. (2003). Extremophiles as a source for novel enzymes. *Curr Opinion Microbiol* **6**: 213–218.

Verberkmoes NC, Russell AL, Shah M, Godzik A, Rosenquist M, Halfvarson J *et al.* (2008). Shotgun metaproteomics of the human distal gut microbiota. *ISME J* **3**: 179–189.

Ward DM, Ferris MJ, Nold SC, Bateson MM. (1998). A natural view of microbial biodiversity within hot spring cyanobacterial mat communities. *Microbiol Mol Biol Rev* **62**: 1353–1370.

Wilmes P, Andersson AF, Lefsrud MG, Wexler M, Shah M, Zhang B *et al.* (2008). Community proteogenomics highlights microbial strain-variant protein expression within activated sludge performing enhanced biological phosphorus removal. *ISME J* **2**: 853–864.

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