

# UC Berkeley

## UC Berkeley Previously Published Works

### Title

Community transcriptomics reveals unexpected high microbial diversity in acidophilic biofilm communities

### Permalink

<https://escholarship.org/uc/item/8c09d8hg>

### Journal

The ISME Journal: Multidisciplinary Journal of Microbial Ecology, 9(4)

### ISSN

1751-7362

### Authors

Goltsman, Daniela S Aliaga

Comolli, Luis R

Thomas, Brian C

et al.

### Publication Date

2015-04-01

### DOI

10.1038/ismej.2014.200

Peer reviewed

# Community transcriptomics reveals unexpected high microbial diversity in acidophilic biofilm communities

Daniela S Aliaga Goltsman<sup>1,3</sup>, Luis R Comolli<sup>2</sup>, Brian C Thomas<sup>1</sup> and Jillian F Banfield<sup>1</sup>

<sup>1</sup> Environmental Science, Policy, and Management, University of California, Berkeley, CA, USA and <sup>2</sup> Lawrence Berkeley National Laboratory, Berkeley, CA, USA

Correspondence: DSA Goltsman or JF Banfield, Environmental Science, Policy, and Management, University of California, 113 Hilgard Hall, Policy and Management, Berkeley 94720, CA, USA. E-mail: [dgolts@eps.berkeley.edu](mailto:dgolts@eps.berkeley.edu) or [jbanfield@berkeley.edu](mailto:jbanfield@berkeley.edu)

<sup>3</sup> Current address: Stanford University, Stanford, CA, USA

## Abstract

A fundamental question in microbial ecology relates to community structure, and how this varies across environment types. It is widely believed that some environments, such as those at very low pH, host simple communities based on the low number of taxa, possibly due to the extreme environmental conditions. However, most analyses of species richness have relied on methods that provide relatively low ribosomal RNA (rRNA) sampling depth. Here we used community transcriptomics to analyze the microbial diversity of natural acid mine drainage biofilms from the Richmond Mine at Iron Mountain, California. Our analyses target deep pools of rRNA gene transcripts recovered from both natural and laboratory-grown biofilms across varying developmental stages. In all, 91.8% of the ~254 million Illumina reads mapped to rRNA genes represented in the SILVA database. Up to 159 different taxa, including Bacteria, Archaea and Eukaryotes, were identified. Diversity measures, ordination and hierarchical clustering separate environmental from laboratory-grown biofilms. In part, this is due to the much larger number of rare members in the environmental biofilms. Although *Leptospirillum* bacteria generally dominate biofilms, we detect a wide variety of other *Nitrospira* organisms present at very low abundance. Bacteria from the *Chloroflexi* phylum were also detected. The results indicate that the primary characteristic that has enabled prior extensive cultivation-independent 'omic' analyses is not simplicity but rather the high dominance by a few taxa. We conclude that a much larger variety of organisms than previously thought have adapted to this extreme environment, although only few are selected for at any one time.

## Introduction

Microbial ecology and evolution studies in model natural ecosystems can greatly advance understanding the ecology of more complex natural environments. The acid mine drainage (AMD) environment has been established as a model ecosystem, due in part to microbial compositional

characteristics that make the biofilms tractable for cultivation-independent molecular analyses (reviewed by Deneff et al., 2010b). AMD and other acidic environments have also been extensively studied due to their importance in acid and metal contamination, and application in the biomining industry (reviewed by Rawlings and Johnson, 2007; Gadd, 2010).

AMD systems are generally dominated by few acidophilic Bacteria belonging to the *Nitrospira* and *Proteobacteria* phyla, whereas lower abundance members include *Firmicutes*, *Actinobacteria* and *Acidobacteria* phyla, Archaea and Eukaryotes (for example, Druschel et al., 2004; Golyshina and Timmis, 2005; Baker et al., 2009; Amaral-Zettler et al., 2011; Garcia-Moyano et al., 2012; and reviewed by Johnson, 2012). Cultivation-independent techniques such as community genomics, proteomics, and microarrays have been successfully applied to study the physiology and ecology of acidophilic organisms in their natural environments. Results have provided new insight into the environmental and biological factors that influence community structure (Deneff et al., 2010a; Mueller et al., 2011). For example, protein expression of the most abundant bacteria in AMD biofilms from the Richmond Mine at Iron Mountain, California, is influenced by the community composition, whereas protein expression in lower abundance members may be controlled by abiotic factors such as pH, temperature, drainage flow and sample collection site (Mueller et al., 2010). In addition, pH and ferrous iron concentrations promote changes in the dominant community members in AMD biofilms (Shufen Ma, personal communication). Furthermore, fluorescence microscopy has shown that the community composition and organization of AMD biofilms change as biofilms mature (Wilmes et al., 2009). All analyses to date have focused exclusively on relatively abundant organisms (more than a few percent of each community). Lack of information about less abundant organisms has limited our understanding of the biology of extremely acidic environments and factors that impact community structure.

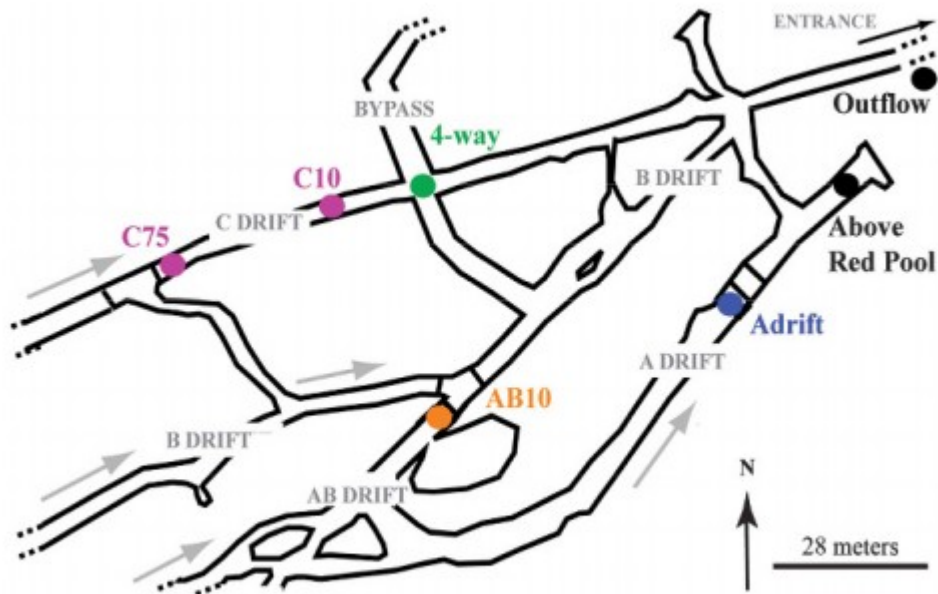
Deep-sequencing the small subunit ribosomal gene (SSU ribosomal DNA) has been widely applied in surveys of the community composition of natural environment (for example, reviewed by Rajendhran and Gunasekaran, 2011; Cox et al., 2013). However, this technique does not provide information about which organisms are actively growing. For example, Hiibel et al. (2010) used fingerprinting methods to analyze the SSU ribosomal RNA (rRNA) and ribosomal DNA from a mixed microbial assemblage, and demonstrated that analyses of the ribosomal DNA gene do not accurately represent the active members of the community. Therefore, studies of the total RNA pool using deep-sequencing technologies to evaluate the diversity of acidophilic communities have the potential to deeply explore microbial community composition and the proportional activity of community members, and provide extensive data sets suitable for diversity analyses.

The most widely used indices for exploring microbial diversity are the species richness (which gives equal weight to all phylotypes present in a

community), the Shannon–Wiener index (which measures the information content of a community) and the Simpson’s index (which measures the probability that two phylotypes are the same; reviewed by Bent and Forney, 2008). In addition, Hill’s diversity indices summarize different properties of a community, depending on the presence and relative abundance of taxa. Here we apply these indices to transcriptomic data to survey acidophilic microbial communities in depth, and to evaluate the importance of biofilm maturation stage, location-dependent environmental factors, and growth in laboratory bioreactors in shaping AMD microbial diversity. The results greatly expand our understanding of the diversity of AMD biofilms, and confirm that these are truly complex ecosystems. The findings clarify that it is dominance by a few taxa, not lack of complexity, that makes AMD environments good model systems for studying community physiology and ecology.

### Materials and methods

Eight biofilms were collected from the A-drift, C-drift, AB-drift and 4-way locations within the Richmond Mine at Iron Mountain Mines, California (40°40′ 38.42″N and 122° 31′ 19.90″W, elevation of ~900 m; Figure 1, Table 1, and Supplementary Table S1). In addition, biofilms were grown at pH 1 and 37 °C in laboratory bioreactors using inocula from two locations within the A-drift, and mine outflow, as previously described (Belnap et al., 2010). The growth stage of the biofilms was estimated visually as described earlier (Wilmes et al., 2009). Briefly, very thin, pink biofilms were labeled growth stage 0 (GS 0); thicker, light brownish biofilms were labeled GS 1; and very thick, dark-colored biofilms were labeled GS 2. Biofilms were snap-frozen in liquid nitrogen upon collection and stored at –80 °C.



**Figure 1** Map of the Richmond Mine showing the locations from where biofilms and inocula for bioreactors were collected.

Total RNA was obtained using two acid phenol-chloroform-isoamyl alcohol (Ambion, Grand Island, NY, USA) extractions, and immediately purified using the RNEasy MinElute kit (Qiagen, Valencia, CA, USA). Integrity and concentration of the RNA was assessed using a Bioanalyzer 2100 (Agilent Technologies, Santa Clara, CA, USA). Good quality (RNA integrity number >7) total RNA was converted to cDNA as described by Parkhomchuk et al. (2009) in order to keep the strand specificity of the transcriptome. Resulting cDNA was fragmented using a Covaris S-system (Covaris Inc., Woburn, MA, USA) to an average fragment size of 200 bp. Fragmented cDNA was sent to the University of California Davis for library preparation and sequencing. Illumina single-end GAII sequencing (75 bp reads) was obtained for five samples (Supplementary Table S1) and Illumina paired-end HiSeq sequencing (100 bp reads) was done for the other eight biofilms (Table 1). An important consideration for sequence-based 'omic' studies is that the depth of sampling is large enough that organisms at very low abundance levels can be detected. It is notable that we do not detect a wide variety of organisms typically encountered in the laboratory environment (for example, typical human or room microbiome organisms, such as skin-associated bacteria) or sequences that would suggest contamination from an isolate or other metagenomic project. Therefore, we are confident that rare community members, detected in both laboratory-grown bioreactors and field-collected samples, were intrinsic to those samples.

**Table 1** Description of samples collected for this study

Sample ID	Date collected	Location	GS	Type	T (°C)	pH	No. of reads (M) <sup>a</sup>	% rRNA <sup>b</sup>
AB10_GS0	11.02.10	AB-drift	0	Env	39	0.8	27.87	96.22
AB10_GS1	11.02.10	AB-drift	1	Env	39	0.8	30.24	96.46
C10_GS0	11.02.10	C-drift	0	Env	42	0.8	27.12	94.12
C10_GS05	11.02.10	C-drift	0.5	Env	42	0.8	30.91	96.16
C10_GS1	11.02.10	C-drift	1	Env	42	0.8	28.60	95.49
4-way_GS2	07.15.11	4-way	2	Env	39	0.7	26.68	85.73
R3_GS0	09.28.10	A-drift <sup>c</sup>	0	BR	37	1.31	31.56	95.00
R3_GS1	10.06.10	A-drift <sup>c</sup>	1	BR	37	1.74	30.49	94.08

Abbreviations: BR, bioreactor-grown biofilms; Env, environmental biofilms; GS, growth stage; rRNA, ribosomal RNA.

<sup>a</sup>Total No. of reads obtained for that sample.

<sup>b</sup>Percentage of reads mapping to rRNA from the SILVA database.

<sup>c</sup>For bioreactor-grown biofilms, location refers to the place where the inoculum was obtained from within the Richmond Mine.

Sample ID indicates the location and growth stage of the biofilm at the time of collection (e.g., sample AB10\_GS0 was an early growth-stage biofilm collected at the AB-drift, while sample R3\_GS1 was a late growth stage biofilm collected from bioreactor #3).

Low-quality bases were trimmed from the sequencing reads using the `fastx_trimmer` script ([http://hannonlab.cshl.edu/fastx\\_toolkit/](http://hannonlab.cshl.edu/fastx_toolkit/)) and the `sickle` trimmer script with default parameters (<https://github.com/najoshi/sickle>). Reads >40 bp were kept for further analyses. Trimmed reads were mapped to the SSU and large subunit rRNA gene SILVA databases (Quast et al., 2013) using Bowtie (Langmead, 2010) with default parameters to separate ribosomal from non-ribosomal reads. Ribosomal reads were then mapped using Bowtie with default parameter to the SSU\_Ref\_108 rRNA SILVA database. Mapped reads were assembled using a reference-based approach with Cufflinks (Martin and Wang, 2011; Roberts et al., 2011), and assembled transcripts were clustered at 97% identity using Uclust (Edgar, 2010). Abundance measures (normalized count values generated by Cufflinks) are reported in Supplementary Table S2.

In order to support the results obtained by the Cufflinks pipeline, we reconstructed full-length SSU rRNA gene sequences from lower-abundance community members. Sequencing reads belonging to the abundant *Leptospirillum* groups II and III rRNA genes were removed from the ribosomal RNA data files by mapping reads using Bowtie (Langmead, 2010). SSU rRNA gene sequences were then reconstructed from the remainder *Leptospirillum*-subtracted reads using EMIRGE (Miller et al., 2011) with parameters `-l 101 -i 300 -s 100 -phred33`, and run until 40 iterations. Sequences classified as chimeras by `uchime` (Edgar et al., 2011) and `DECIPHER` (Wright et al., 2012)

were excluded. Relative abundances and taxonomic classification are reported in Supplementary Table S3.

SSU rRNA gene sequences assembled via Cufflinks and those reconstructed with EMIRGE were classified using the Ribosome Database Project web server with an 80% cutoff for the lowest classification level (Wang et al., 2007). Unclassified sequences (generally Eukaryotic) were searched against the SILVA database (Pruesse et al., 2012), the Protist Ribosomal Reference Database (Guillou et al., 2013) and the non-redundant NCBI database using BLASTN. All sequences were then aligned using the SINA Aligner (Pruesse et al., 2012). Phylogenetic tree construction was done using FastTree with parameters `-gtr -nt -gamma` (Price et al., 2009), and the tree was visualized using the iTol website (Letunic and Bork, 2007). The *Nitrospira* phylum phylogenetic tree was built using Phyml with 100 bootstrap (Guindon et al., 2010), with sequences aligned using the RDP aligner (Cole et al., 2014).

Supplementary Table S2 was used as input file for diversity analyses. Specifically, rank abundance curves and non-metric multidimensional scaling analyses were done in R (Team, 2008) using the R packages Picante (Kembel et al., 2010) and BiodiversityR (Kindt and Coe, 2005). Net relatedness index (NRI) and nearest taxon index (NTI) values were estimated using the vegan R package (Oksanen et al., 2013), and principal coordinates analyses were estimated using the Fast Unifrac website (Hamady et al., 2010). Diversity profiles were calculated as presented by Leinster and Cobbold (2012) and Doll et al. (2013).

Cryogenic transmission electron microscopy specimen preparation and imaging were done as described by Comolli et al. (2009) and Baker et al. (2010).

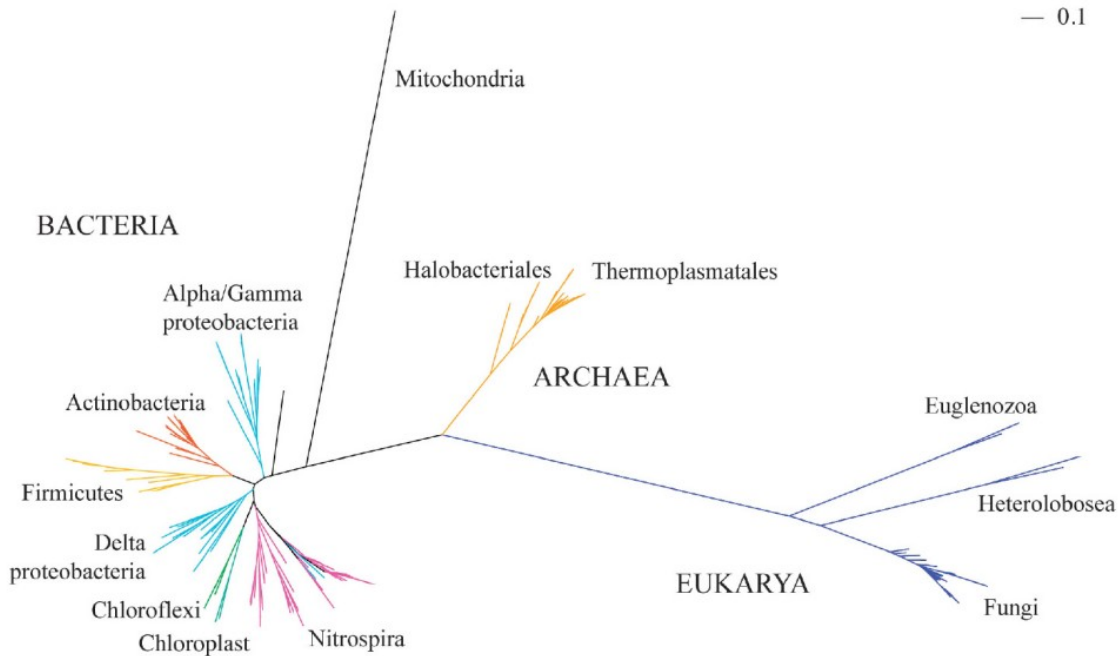
Transcriptomics reads have been submitted to the NCBI sequence read archive under the accession number SRP026490. Assembled SSU rRNA sequences analyzed in this work are available in fasta format as Supplementary Materials.

## Results and discussion

### Community transcriptomics

Biofilms at increasing stages of development were obtained from five locations within the Richmond Mine for community transcriptomic analyses (Table 1, Supplementary Table S1 and Figure 1). Biofilms were also grown in laboratory bioreactors in order to compare the diversity of environmental and bioreactor transcriptomes. An average 91.8% of the total sequencing reads mapped to 16S/18S and 23S/28S rRNA genes from the SILVA database, and Archaea, Bacteria and Eukaryotes were identified (Figure 2, Table 2 and Supplementary Table S4). In total, 1773 SSU rRNA assembled transcripts were clustered at 97% identity using Uclust into 462 operational taxonomic units (OTUs). Of these, 159 OTUs longer than 140 bp, which were also present in at least two transcriptomics data sets, were classified by The

Ribosome Database Project (Wang et al., 2007) and by BLAST searches against the SILVA database (Supplementary Table S2). Assembled transcripts were searched against a database of hypervariable regions of the SSU rRNA gene (Huse et al., 2008) and support the taxonomic classifications (Supplementary Table S5). Community diversity analyses were done with these 159 OTUs.



**Figure 2** Phylogenetic tree of the SSU rRNA genes identified in transcriptomic samples. Actino: Actinobacteria. Assembled SSU rRNA sequences were aligned using the SINA aligner, and phylogenetic tree reconstruction was done using FastTree.

**Table 2** Relative abundance of SSU rRNA gene transcripts assembled via Cufflinks (%)

	<i>Bacteria</i>	<i>Archaea</i>	<i>Eukaryota</i>	<i>Nitrospira</i>				
				<i>Group I</i>	<i>Group II</i>	<i>Group III</i>	<i>Group IV</i>	<i>Other</i>
AB10 GS0	7.24	0.82	0.13	2.32	68.12	16.90	4.37	0.10
AB10 GS1	12.61	1.32	1.02	4.44	68.39	11.74	0.46	0.02
C10 GS0	8.25	0.66	0.01	2.79	80.38	7.48	0.40	0.02
C10 GS05	8.62	0.81	0.11	2.37	80.41	7.40	0.27	0.01
C10 GS1	7.84	0.80	0.60	2.62	79.80	7.83	0.50	0.02
4-way GS2	5.35	5.30	3.58	2.25	71.90	10.90	0.71	0.02
R3 GS0	4.44	0.06	0.00	0.37	11.17	59.91	23.48	0.58
R3 GS1	4.44	0.51	0.15	0.29	6.67	60.86	26.10	0.97

Abbreviations: GS, growth stage; rRNA, ribosomal RNA; SSU, small subunit.

Members of the phylum *Nitrospira* represent >85% of the actively growing community in all biofilms, while other Bacteria include members of the phyla *Proteobacteria* (of the classes Alpha, Delta, Epsilon, Gamma and T18), *Actinobacteria* (generally of the *Acidimicrobiales* class), *Acidobacteria*, *Firmicutes* (class *Clostridia*) and *Chloroflexi* (*Caldilineae* class; Figure 2 and Supplementary Table S2). Protists, fungi and red algae have been observed previously in microscopy-based studies and SSU rRNA gene surveys of



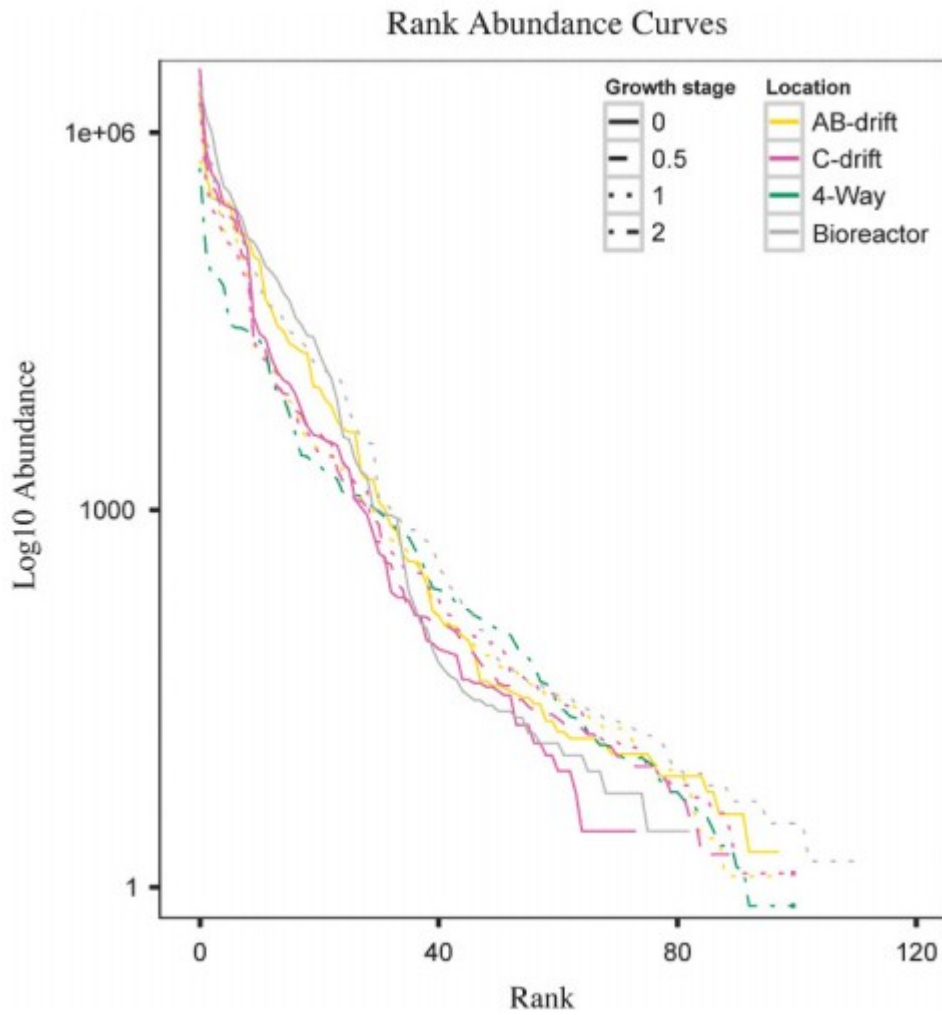
acidophilic environments (Baker et al., 2009; Prasanna et al., 2011; Zirnstein et al., 2012). In this study, OTUs belonging to the phyla *Excavata*, *Heterolobosea*, *Rhodophyta* and *Opisthokonta* were generally abundant in the highest developmental-stage environmental biofilm (Supplementary Tables S2 and S3). *Thermoplasmatales* Archaea A-, E- and G-plasma, *Ferroplasma* Type I and the Nanoarchaea ARMAN-1 and ARMAN-4 were observed in all biofilms, but were most active in high developmental-stage environmental biofilms (Supplementary Table S2).

### Community diversity

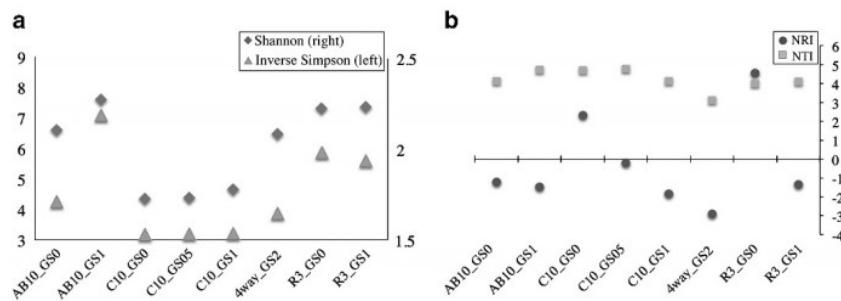
Rank abundance curves show that biofilms reach a mean species richness of <100 taxa in each sample and organismal abundance curves fall quickly, indicating that samples are largely dominated by a few OTUs (Figure 3). C-drift curves indicate lower richness and evenness compared with other biofilms, likely due to more extreme conditions found in this location than at other sites (Table 1 and Supplementary Table S1). Biofilms grown in bioreactors show comparable richness and evenness to environmental samples (Figure 3).

The taxonomy-based Shannon–Wiener diversity index and Inverse Simpson’s index of diversity show similar values for biofilms from the same site, and a slight increase in diversity is observed as growth stage increases, indicating that taxonomic diversity varies more with location than with developmental stage of the biofilm (Figure 4a). Biofilms collected at the C-drift show the lowest diversity indices, consistent with the low richness observed in rank abundance curves.

The NRI and NTI are measures of phylogenetic clustering or overdispersion closer to the root (NRI) and at the tips (NTI) of a phylogeny (Vamosi et al., 2009). NRI values are generally negative (except for two early growth-stage biofilms that show the lowest taxonomic richness), indicating overdispersion closer to the root of the tree (that is, taxa are more distantly related to each other at higher taxonomic levels; Figure 4b). Positive NTI values show phylogenetic clustering in all samples, indicating that taxa are more closely related at the tips of the phylogeny (Figure 4b). Positive NTI and negative NRI values have been observed in Rio Tinto AMD samples (Amaral-Zettler et al., 2011). Overall, the results suggest that the AMD system is characterized by a high diversity at the phylum level, but with relatively few closely related members of each phylum.



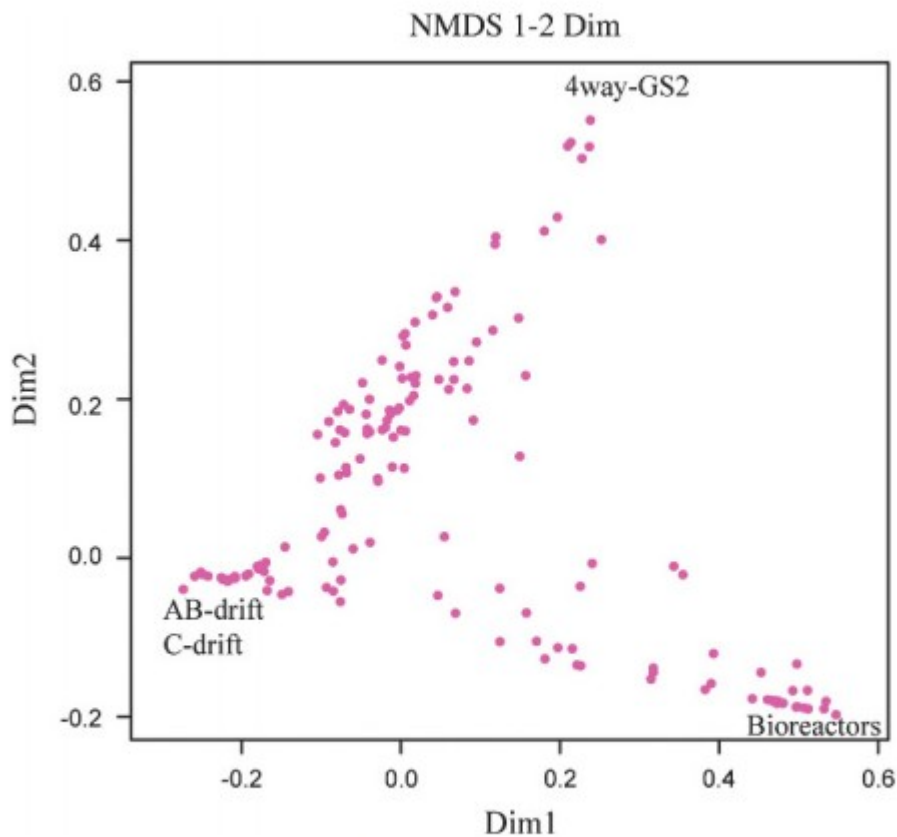
**Figure 3** Rank abundance curves categorized by growth stage and location. Y-axis is represented in logarithmic scale.



**Figure 4** (a) Shannon–Wiener's diversity index (right Y-axis), and Simpson Index of Diversity (left Y-axis). (b) NRI and NTI graph for all samples.

Non-metric multidimensional scaling analyses were calculated using Bray–Curtis distance matrix and two dimensions were selected (a significant reduction in stress was observed from one to two dimensions, data not

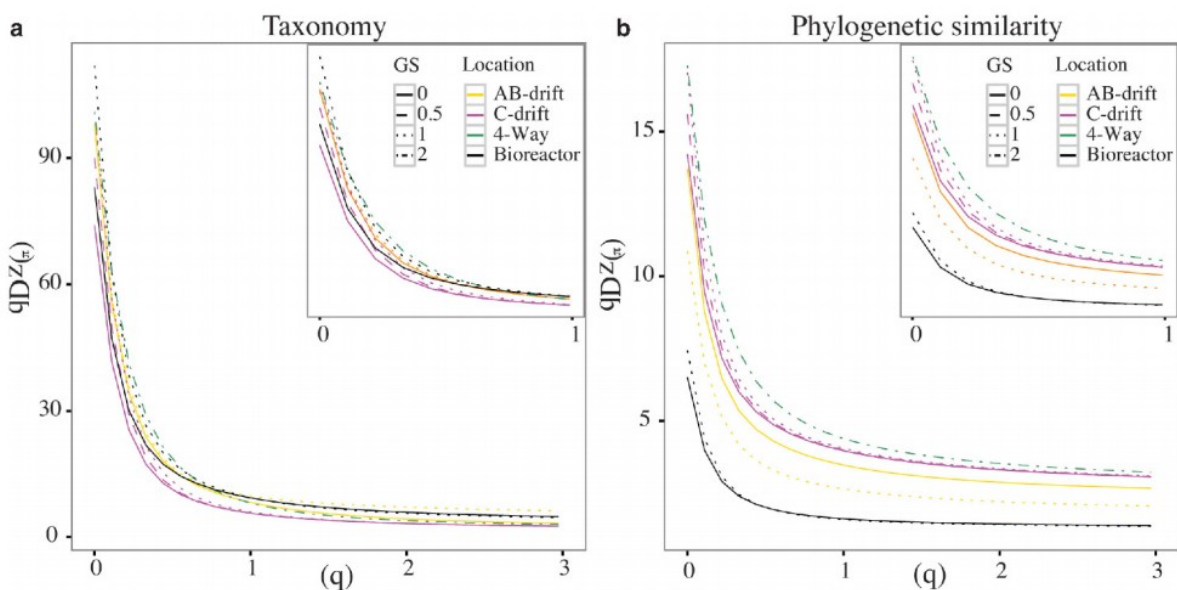
shown). Non-metric multidimensional scaling ordination separates bioreactors from environmental samples, while the mature biofilm (4-way GS2) separates from the other biofilms (Figure 5 and Supplementary Figure S1). Weighted Unifrac principal coordinates analyses and hierarchical cluster analyses support the separation between bioreactor and environmental biofilms, as well as late from early growth-stage biofilms (Supplementary Figure S2), suggesting again that microbial community composition in the AMD system is driven primarily by environmental factors and, to a lesser extent, by the developmental stage of the biofilm.



**Figure 5** Non-metric multidimensional scaling (NMDS) analyses of microbial communities. Each dot represents one taxon.

Diversity profiles (Leinster and Cobbold, 2012) were used to explore differences in community diversity while taking taxonomy and phylogenetic similarity into account (Figure 6 and Supplementary Figure S3). In diversity profile plots, the y axis represents a calculated effective diversity value and the x axis represents a sensitivity parameter ' $q$ ', where smaller ' $q$ ' gives higher weight to rare taxa and this weight decreases with increasing ' $q$ ' (Leinster and Cobbold, 2012; Doll et al., 2013). Overall, one community is considered more diverse than another if its diversity profile curve lies above the other curve. When considering only taxonomic identity (that is, just

OTUs), C-drift samples are the least diverse, and the most mature biofilm (4-way GS2) is the most diverse only at low values of 'q' (Figure 6a and Supplementary Figure S3). When adding phylogenetic similarity to the profiles, it becomes evident that the environmental samples are more diverse than bioreactor biofilms, and more mature biofilms are more diverse than early biofilms (Figure 6b and Supplementary Figure S3B). The 'open' nature and constant change in conditions of environmental samples likely contribute to higher community diversity. Additionally, as observed on microscopy-based analyses that considered relatively abundant taxa (Wilmes et al., 2009), larger community membership likely increases diversity as biofilms mature.

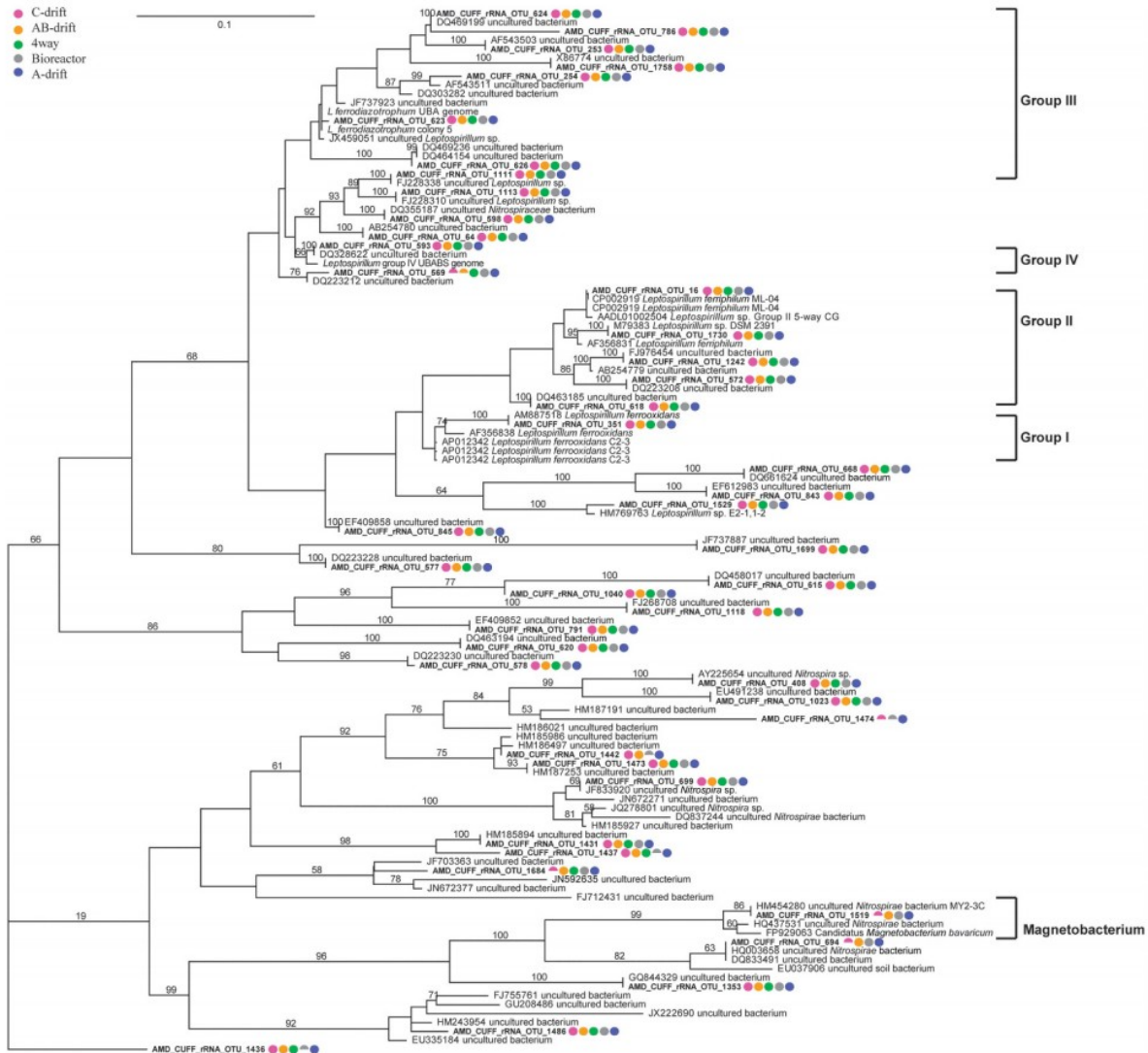


**Figure 6** Diversity profiles categorized by growth stage (GS) and location. (a) Taxonomy-based profiles; (b) phylogenetic similarity-based profiles.

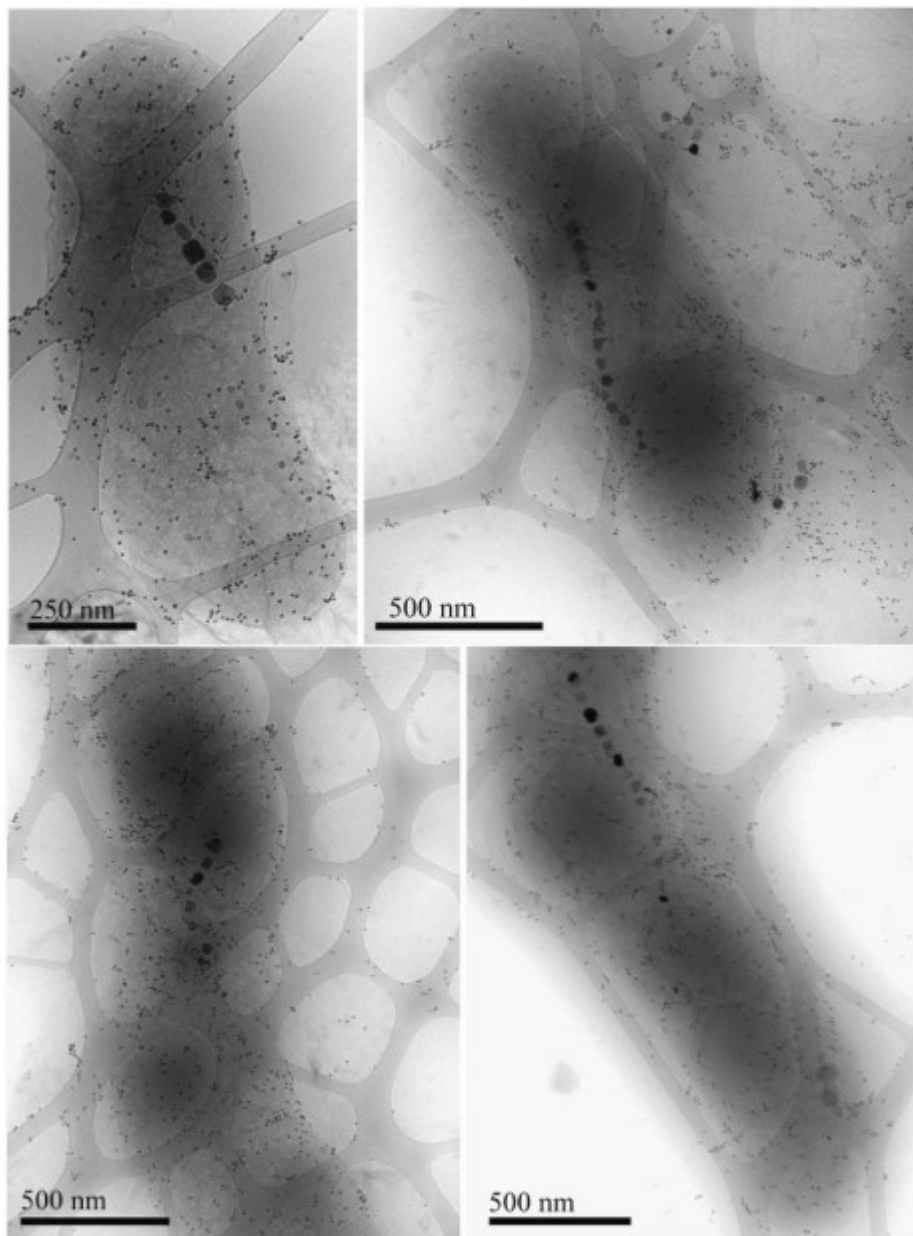
### *Nitrospira* phylum and low-abundance organisms

There were 45 OTUs identified as belonging to the *Nitrospira* phylum and 18 were confidently classified as *Leptospirillum* groups I-IV by the Ribosome Database Project, BLAST to the SILVA database and phylogenetic tree construction (Figure 7 and Supplementary Tables S2 and S6). *Leptospirillum* group II generally dominates environmental biofilms, whereas *Leptospirillum* group III dominates bioreactor-grown samples (Table 2 and Supplementary Table S4). A change in dominance from *Leptospirillum* group II to III has been observed in biofilms exposed to low  $Fe^{+2}/Fe^{+3}$  ratios and higher pH (Shufen Ma, personal communication). Reads from *Leptospirillum* groups I and IV are usually present at low abundance in all samples, and the activity of group IV increases in biofilms where *Leptospirillum* group III is the dominant member (Table 2). Other *Nitrospira* phylum OTUs observed in most transcriptomics samples include sequences related to the *Magnetobacterium* genus, as well as other uncultured and unclassified *Nitrospiraceae* (Figure 7, Supplementary Tables S2 and S5). Sequences belonging to *Magnetobacterium* spp. have not

yet been recovered from metagenomic data sets from AMD biofilms. In fact, magnetotactic bacteria from the *Nitrospira* phylum are typically found in freshwater lakes, lake sediments and hot springs (reviewed by Amann et al., 2007 and Lefevre and Bazyliński, 2013). Using cryogenic transmission electron microscopy, we have observed long vibrio-shaped cells with intracellular magnetosome-like structures in AMD biofilms (Figure 8). Bullet-shaped chains of magnetite, like those in Figure 8, have been reported in the magnetotactic Delta-Proteobacterium *Desulfovibrio magneticum* (reviewed by Komeili, 2012). Therefore, the cryogenic transmission electron microscopy images confirm the presence of magnetotactic bacteria in AMD systems and support the transcriptomics findings.



**Figure 7** Un-rooted, maximum likelihood phylogenetic tree of members of the phylum *Nitrospira*. Colored dots indicate the location where the taxon is present, and half dots indicate that the taxon was present in some of the samples collected at that location.



**Figure 8** Cryo-TEM images of cells recovered from AMD biofilms containing magnetosome-like bodies.

Sequences longer than 1000 bp related to *Acidithiobacillus caldus* and other uncultured *Acidithiobacilli* were assembled from the data, and are mostly observed in bioreactor samples (Supplementary Table S2). These bacteria are often encountered in higher pH and lower temperature environments such as downstream of AMD sites and bioleaching systems (reviewed by Johnson, 2012). Among the *Actinobacteria*, the full sequence of *Ferromicrobium* sp. Mc9KL-1-9 and sequences longer than 900 bp related to uncultured *Acidimicrobium* bacteria were recovered from bioreactors and

late growth-stage biofilms. Among the *Firmicutes*, full-length sequences related to *Sulfobacillus thermosulfidooxidans* and *Alicyclobacillus disulfidooxidans* were observed in all biofilms and bioreactor samples (Supplementary Table S2). Both of these organisms have been detected by cultivation-based and/or cultivation-independent methods in the AMD system (unpublished).

### SSU rRNA gene reconstruction

Using EMIRGE and a data set of transcriptomics reads after subtraction of reads mapping to *Leptospirillum* groups II and III, we reconstructed 41 SSU rRNA gene sequences, which range from 836 to 1862 bp in length. RDP classification and percent identity alignment to SSU rRNA genes from the SILVA and Greengenes databases indicate the presence of Archaea, Bacteria and Eukaryotes, and support the transcriptomics findings described above (Supplementary Table S3 and Supplementary Figure S4). RNA sequencing reads from the Archaea C-plasma, G-plasma and A-plasma, and the *Actinobacteria* and *Clostridia* were most abundant in environmental samples, whereas E-plasma and the *Proteobacteria* were most abundant in bioreactor biofilms. Full-length Eukaryotic SSU rRNA genes from the order *Capnodiales*, *Eurotiales*, *Cyanidiales* and *Schizopyrenida*, and from the *Dothideomycetes* class were reconstructed, and reads are most abundant in late growth-stage biofilms. Mitochondrial and chloroplast ribosomal sequences were only observed in the most mature 4-way GS2 biofilm (Supplementary Table S3).

### Conclusion

It is known that the RNA content and the number of ribosomes per cell are proportional to the rate of cellular protein synthesis and the rate of growth of the population (reviewed by Pace, 1973, Condon et al., 1995 and Elser et al., 2000). For example, it has been shown that the rate of rRNA synthesis in *Escherichia coli* is about twofold higher in rapid growth compared with slow growth, although the rate of mRNA synthesis does not decrease as much as rRNA synthesis under impaired growth conditions (Pace, 1973). An increase in the RNA to protein ratio (thus, an increase in total RNA and rRNA) with increasing growth rate was also observed in other organisms (Wagner, 1994; Karpinets et al., 2006). Therefore, a measure of the total RNA and rRNA content provides information on the diversity based on how actively organisms are growing in a population.

Another important consideration is the potential difference between total rRNA transcripts and the fraction that forms active ribosomes. As reviewed by Pace (1973), in *E. coli* growing at high growth rates most of the rRNA synthesized becomes stable (forming ribosomes), and a large amount of the rRNA synthesized at slower growth is degraded. Thus, assuming that the rRNA sampled from a natural environment at a specific time point corresponds primarily to mature, stable RNA, the amount of rRNA measured would reflect the proportion of ribosomes and the proportional activity of the cell.

Using deep sequencing of total RNA from AMD biofilms, we demonstrate that phylogenetic similarity-based diversity analyses are an appropriate method to compare levels of diversity and activity between different samples and developmental stages, and are an important addition to traditional diversity analyses studies. The findings indicate that AMD biofilm diversity is driven primarily by environmental factors rather than by the developmental stage of the biofilm. Within the same location, microbial diversity increases with biofilm maturation. Bacteria of the genus *Leptospirillum* dominate all samples studied, where *Leptospirillum* group II dominate environmental biofilms, whereas *Leptospirillum* group III dominate bioreactor samples. Bacteria of the *Magnetobacterium* genus, and of the *Chloroflexi* phylum have not been observed previously in community genomics or microscopy-based analyses of the well-studied Richmond Mine AMD model system. Our analyses indicate that, despite the low pH, elevated temperature and very high metal concentrations, AMD systems can be much more diverse than previously thought. The results suggest that, rather than just a few organisms being able to grow in this environment, many organisms can grow but only very few are selected for at one time, likely due to the low resource diversity in the system. We demonstrate that it is the dominance of a few taxa that makes acidophilic communities well suited for studies of ecology, physiology and diversity.

Accession codes

Accessions

Sequence Read Archive

SRP026490

References

Amann R, Peplies J, Schüler D . (2007) *Diversity and Taxonomy of Magnetotactic Bacteria*. Springer: Berlin Heidelberg, pp 25–36.

Amaral-Zettler LA, Zettler ER, Theroux SM, Palacios C, Aguilera A, Amils R . (2011). Microbial community structure across the tree of life in the extreme Rio Tinto. *ISME J* 5: 42–50.

Baker BJ, Tyson GW, Goosherst L, Banfield JF . (2009). Insights into the diversity of eukaryotes in acid mine drainage biofilm communities. *Appl Environ Microbiol* 75: 2192–2199.

Baker BJ, Comolli LR, Dick GJ, Hauser LJ, Hyatt D, Dill BD *et al.* (2010). Enigmatic, ultrasmall, uncultivated Archaea. *Proc Natl Acad Sci USA* 107: 8806–8811.

Belnap CP, Pan C, VerBerkmoes NC, Power ME, Samatova NF, Carver RL *et al.* (2010). Cultivation and quantitative proteomic analyses of acidophilic microbial communities. *ISME J* 4: 520–530.



Bent SJ, Forney LJ . (2008). The tragedy of the uncommon: understanding limitations in the analysis of microbial diversity. *ISME J* 2: 689–695.

Cole JR, Wang Q, Fish JA, Chai B, McGarrell DM, Sun Y *et al.* (2014). Ribosomal Database Project: data and tools for high throughput rRNA analysis. *Nucleic Acids Res* 42: D633–D642.

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.

Condon C, Squires C, Squires CL . (1995). Control of rRNA transcription in *Escherichia coli*. *Microbiol Rev* 59: 623–645.

Cox MJ, Cookson WO, Moffatt MF . (2013). Sequencing the human microbiome in health and disease. *Hum Mol Genet* 22: R88–R94.

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 USA* 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.

Doll HM, Armitage DW, Daly RA, Emerson JB, Goltsman DS, Yelton AP *et al.* (2013). Utilizing novel diversity estimators to quantify multiple dimensions of microbial biodiversity across domains. *BMC Microbiol* 13: 259.

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

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.

Elser JJ, Sterner RW, Gorokhova E, Fagan WF, Markow TA, Cotner JB *et al.* (2000). Biological stoichiometry from genes to ecosystems. *Ecol Lett* 3: 540–550.

Gadd GM . (2010). Metals, minerals and microbes: geomicrobiology and bioremediation. *Microbiology* 156: 609–643.

Garcia-Moyano A, Gonzalez-Toril E, Aguilera A, Amils R . (2012). Comparative microbial ecology study of the sediments and the water column of the Rio Tinto, an extreme acidic environment. *FEMS Microbiol Ecol* 81: 303–314.

Golyshina OV, Timmis KN . (2005). *Ferroplasma* and relatives, recently discovered cell wall-lacking archaea making a living in extremely acid, heavy metal-rich environments. *Environ Microbiol* 7: 1277–1288.

Guillou L, Bachar D, Audic S, Bass D, Berney C, Bittner L *et al.* (2013). The Protist Ribosomal Reference database (PR2): a catalog of unicellular eukaryote small sub-unit rRNA sequences with curated taxonomy. *Nucleic Acids Res* 41: D597–D604.

Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O . (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol* 59: 307–321.

Hamady M, Lozupone C, Knight R . (2010). Fast UniFrac: facilitating high-throughput phylogenetic analyses of microbial communities including analysis of pyrosequencing and PhyloChip data. *ISME J* 4: 17–27.

Hiibel SR, Pruden A, Crimi B, Reardon KF . (2010). Active community profiling via capillary electrophoresis single-strand conformation polymorphism analysis of amplified 16S rRNA and 16S rRNA genes. *J Microbiol Methods* 83: 286–290.

Huse SM, Dethlefsen L, Huber JA, Mark Welch D, Relman DA, Sogin ML . (2008). Exploring microbial diversity and taxonomy using SSU rRNA hypervariable tag sequencing. *PLoS Genet* 4: e1000255.

Johnson DB . (2012). Geomicrobiology of extremely acidic subsurface environments. *FEMS Microbiol Ecol* 81: 2–12.

Karpinets TV, Greenwood DJ, Sams CE, Ammons JT . (2006). RNA:protein ratio of the unicellular organism as a characteristic of phosphorous and nitrogen stoichiometry and of the cellular requirement of ribosomes for protein synthesis. *BMC Biol* 4: 30.

Kembel SW, Cowan PD, Helmus MR, Cornwell WK, Morlon H, Ackerly DD *et al.* (2010). Picante: R tools for integrating phylogenies and ecology. *Bioinformatics* 26: 1463–1464.

Kindt R, Coe R . (2005) *Tree Diversity Analysis. A Manual and Software for Common Statistical Methods for Ecological and Biodiversity Studies*. World Agroforestry Centre (ICRAF): Nairobi (Kenya).

Komeili A . (2012). Molecular mechanisms of compartmentalization and biomineralization in magnetotactic bacteria. *FEMS Microbiol Rev* 36: 232–255.

Langmead B . (2010). Aligning short sequencing reads with Bowtie. *Curr Protoc Bioinformatics* Chapter 11: Unit 11, 17.

Lefevre CT, Bazyliński DA . (2013). Ecology, diversity, and evolution of magnetotactic bacteria. *Microbiol Mol Biol Rev* 77: 497–526.

Leinster T, Cobbold CA . (2012). Measuring diversity: the importance of species similarity. *Ecology* 93: 477–489.

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

- Martin JA, Wang Z . (2011). Next-generation transcriptome assembly. *Nat Rev Genet* 12: 671–682.
- 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.
- Mueller RS, Deneff VJ, Kalnejais LH, Suttle KB, Thomas BC, Wilmes P *et al.* (2010). Ecological distribution and population physiology defined by proteomics in a natural microbial community. *Mol Syst Biol* 6: 374.
- Mueller RS, Dill BD, Pan C, 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. *Environ Microbiol* 13: 2279–2292.
- Oksanen J, Blanchet FG, Kindt R, Legendre P, Minchin PR, O'Hara RB *et al.* (2013). vegan: Community Ecology Package. R package version 2.0-6 <http://cran.r-project.org/web/packages/vegan/index.html>.
- Pace NR . (1973). Structure and synthesis of the ribosomal ribonucleic acid of prokaryotes. *Bacteriol Rev* 37: 562–603.
- Parkhomchuk D, Borodina T, Amstislavskiy V, Banaru M, Hallen L, Krobitch S *et al.* (2009). Transcriptome analysis by strand-specific sequencing of complementary DNA. *Nucleic Acids Res* 37: e123.
- Prasanna R, Ratha SK, Rojas C, Bruns MA . (2011). Algal diversity in flowing waters at an acidic mine drainage 'barrens' in central Pennsylvania, USA. *Folia Microbiol (Praha)* 56: 491–496.
- Price MN, Dehal PS, Arkin AP . (2009). FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. *Mol Biol Evol* 26: 1641–1650.
- Pruesse E, Peplies J, Glockner FO . (2012). SINA: accurate high-throughput multiple sequence alignment of ribosomal RNA genes. *Bioinformatics* 28: 1823–1829.
- Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P *et al.* (2013). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res* 41: D590–D596.
- Rajendhran J, Gunasekaran P . (2011). Microbial phylogeny and diversity: small subunit ribosomal RNA sequence analysis and beyond. *Microbiol Res* 166: 99–110.
- Rawlings DE, Johnson DB . (2007). The microbiology of biomining: development and optimization of mineral-oxidizing microbial consortia. *Microbiology* 153: 315–324.

Roberts A, Trapnell C, Donaghey J, Rinn JL, Pachter L . (2011). Improving RNA-Seq expression estimates by correcting for fragment bias. *Genome Biol* 12: R22.

Team RDC . (2008). R: A language and environment for statistical computing. <http://www.r-project.org/>.

Vamosi SM, Heard SB, Vamosi JC, Webb CO . (2009). Emerging patterns in the comparative analysis of phylogenetic community structure. *Mol Ecol* 18: 572-592.

Wagner R . (1994). The regulation of ribosomal RNA synthesis and bacterial cell growth. *Arch Microbiol* 161: 100-109.

Wang Q, Garrity GM, Tiedje JM, Cole JR . (2007). Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl Environ Microbiol* 73: 5261-5267.

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.

Wright ES, Yilmaz LS, Noguera DR . (2012). DECIPHER, a search-based approach to chimera identification for 16S rRNA sequences. *Appl Environ Microbiol* 78: 717-725.

Zirnstien I, Arnold T, Krawczyk-Barsch E, Jenk U, Bernhard G, Roske I . (2012). Eukaryotic life in biofilms formed in a uranium mine. *Microbiologyopen* 1: 83-94.