23. We are grateful to the National Institute of Child Health and Human Development, National Institute of Neurological Disorders and Stroke, and National Institute of Dental and Craniofacial Research at the NIH for financial support of this work. This research was also supported by the Intramural Research Program of the NIH, National Center for Biotechnology Information.

Supporting Online Material
www.sciencemag.org/cgi/content/full/314/5807/1930/DC1
Materials and Methods
SOM Text
Figs. S1 to S6
References

13 June 2006; accepted 16 November 2006
10.1126/science.1131262

Lineages of Acidophilic Archaea Revealed by Community Genomic Analysis

Brett J. Baker,[1] Gene W. Tyson,[2] Richard I. Webb,[3] Judith Flanagan,[2]* Philip Hugenholtz,[2]† Eric E. Allen,[2]‡ Jillian F. Banfield[1,2]§

Novel, low-abundance microbial species can be easily overlooked in standard polymerase chain reaction (PCR)–based surveys. We used community genomic data obtained without PCR or cultivation to reconstruct DNA fragments bearing unusual 16*S* ribosomal RNA (rRNA) and protein-coding genes from organisms belonging to novel archaeal lineages. The organisms are minor components of all biofilms growing in pH 0.5 to 1.5 solutions within the Richmond Mine, California. Probes specific for 16*S* rRNA showed that the fraction less than 0.45 micrometers in diameter is dominated by these organisms. Transmission electron microscope images revealed that the cells are pleomorphic with unusual folded membrane protrusions and have apparent volumes of <0.006 cubic micrometer.

Our understanding of the variety of microorganisms that populate natural environments was advanced by the development of polymerase chain reaction (PCR)–based, cultivation-independent methods that target one or a small number of genes (*1*–*3*). Genomic analyses of DNA sequence fragments derived from multispecies consortia (*4*–*6*) and whole environments (*7*–*9*) have provided new information about diversity and metabolic potential. However, PCR-based methods have limited ability to detect organisms whose genes are significantly divergent relative to gene sequences in databases, and most cultivation-independent genomic sequencing approaches are relatively insensitive to organisms that occur at low abundance. Consequently, it is likely that low-abundance microorganisms distantly related to known species will be undetected members of natural consortia, even in low complexity systems such as acid mine drainage (AMD) (*10*).

An important way in which microorganisms affect geochemical cycles is by accelerating the dissolution of minerals. For example, microorganisms can derive metabolic energy by oxidizing iron released by the dissolution of pyrite (FeS_2). The ferric iron by-product promotes further pyrite dissolution, leading to AMD generation. AMD solutions forming underground in the Richmond Mine at Iron Mountain, California, are warm (30° to 59°C), acidic (pH ~0.5 to 1.5), metal-rich [submolar Fe^{2+} and micromolar As and Cu (*11*)] and host active microbial communities. Extensive cultivation-independent sequence analysis of functional and rRNA genes (*11*, *12*) revealed that biofilms contain a significant number of Archaea, but the diversity reported to date has been limited to the order Thermoplasmatales (*10*). Current models for AMD generation thus include only these species.

The genomes of the five dominant members of one biofilm community from the "5-way" region of the Richmond Mine (fig. S1) were largely reconstructed through the assembly of 76 Mb of shotgun genomic sequence (*4*). Previously unreported is a genome fragment that encodes part of the 16*S* rRNA gene of a novel archaeal lineage: Archaeal Richmond Mine Acidophilic Nanoorganism (ARMAN-1). Using an expanded data set that now comprises more than 100 Mb of genomic sequence, we reconstructed a contiguous 4.2-kb fragment adjacent to this gene. A second 13.2-kb genome fragment encoding a 16*S* rRNA gene from an organism that is related to ARMAN-1 (ARMAN-2) was reconstructed from 117 Mb of community genomic sequence derived from a biofilm from the A drift (fig. S1). Within the data sets from each site, results to date indicate that each ARMAN population is near-clonal.

Comparison of the ARMAN-1 and -2 DNA fragments revealed some gene rearrangements, insertions, and deletions (Fig. 1). Genes present in both organisms encode putative inorganic pyrophosphatases, a transcription regulator, and a gene shown to be an arsenate reductase (*13*). Comparative analysis of these genes with sequences in the public databases consistently

[1]Department of Earth and Planetary Sciences, University of California, Berkeley, CA 94720, USA. [2]Environmental Science, Policy, and Management, University of California, Berkeley, CA 94720, USA. [3]Centre for Microscopy and Microanalysis and Department of Microbiology and Parasitology, University of Queensland, Brisbane 4072, Australia.

*Present address: University of California, San Francisco, CA 94143, USA.
†Present address: Department of Energy Joint Genome Institute, Walnut Creek, CA 94598, USA.
‡Present address: University of California, San Diego, La Jolla, CA 92093, USA.
§To whom correspondence should be addressed. E-mail: jill@eps.berkeley.edu

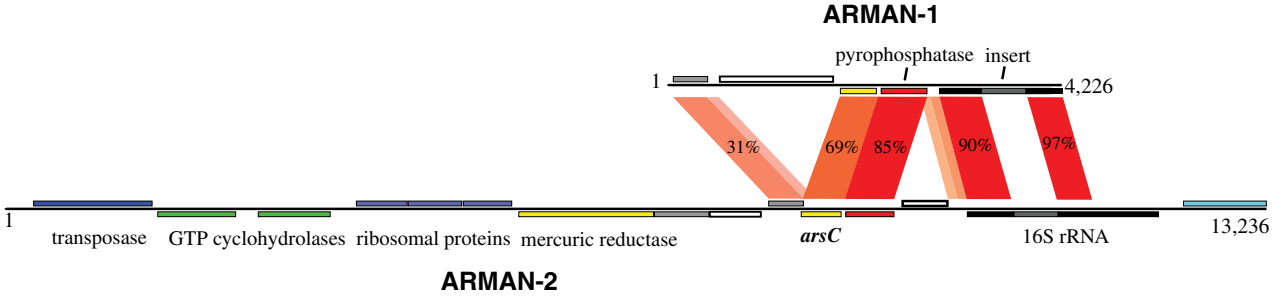

Fig. 1. Comparison of syntenous genomic regions of ARMAN-1 [from the "5-way" (CG) community (*4*)] and ARMAN-2 (from the UBA community). Orthologs and their protein identity are indicated by the red bands. The percentage similarity for the 16*S* rRNA gene sequences is also shown. Numbers at ends indicate length (number of nucleotides); predicted open reading frames for hypothetical proteins are indicated by boxes.

indicated that ARMAN-1 and -2 are representatives of a deeply branching lineage within Euryarchaeota with no cultivated representatives. The only other identified members of this lineage are uncultivated organisms from a hot pool (78°C, pH 7.5) with 85% similarity to ARMAN-2 (Fig. 2) (*14*), and organisms represented by clones from an acidic (pH 4.2 to 4.8) west Siberian peat bog [88% similarity to ARMAN-2 (*15*)]. In both of these studies, atypical 16*S* rRNA gene primers were used.

The 16*S* rRNA genes of ARMAN-1 and -2 have three to five mismatches with commonly used broad-specificity primers (four mismatches with Arc-23F and Univ-1492R; table S1), which explains why these organisms were missed in previous PCR-based surveys (*10*–*12*). The 16*S* rRNA genes both have 519–base pair inserts that encode predicted proteins. The inserts (1102–1103 *E. coli* numbering) share protein-level similarity to each other (*13*) and have weak homology with a functionally characterized homing endonuclease encoded within the 16*S* rRNA gene of *Aeropyrum pernix* (*16*). Excision of the inserts, probably at the bulge-helix-bulge (fig. S2), leaves complete and fully aligned rRNA genes. Insertion sequences (e.g., introns) in archaeal and bacterial rRNA genes are uncommon, although other instances have been reported in pathogens, symbionts, and the crenarchaeal orders Thermoproteales and Desulfurococcales (*17*).

To further explore the diversity and distribution of ARMAN-like microorganisms, we designed primers that target the 16*S* rRNA and putative *arsC* genes of ARMAN-1 and -2, and applied them to samples collected from throughout the Richmond Mine (*13*) (fig. S1). The primers amplified 16*S* rRNA genes from a third previously undetected lineage, ARMAN-3, that are 18% divergent from ARMAN-1 and -2 (Fig. 2). The ARMAN-3 organisms do not have inserts in their 16*S* rRNA genes. This observation, in combination with the relatively low similarity between inserts in the ARMAN-1 and -2 genes, suggests that insertion sequences were acquired after the divergence of the three groups. Both the *arsC* and 16*S* rRNA genes from an ARMAN-lineage organism were recovered from all samples tested, indicating that these organisms are present in most biofilms growing at the site. ARMAN group organisms are present in samples that differ significantly in habitat type (e.g., subaerial and subaqueous biofilms) and geochemical conditions (pH 0.5 to 1.5, temperature 30° to 47°C).

Fluorescently labeled oligonucleotide probes targeting the 16*S* rRNA (ARM980 and ARM1357) of all known ARMAN-lineage microorganisms enabled optical microscope–based visualization of the cells in the suite of biofilm samples (Fig. 3A). The specificity of probe binding was verified using an archaeal-ARMAN (ARC915-ARM915) competitor probe set (*13*). Notably, the cells labeled with the ARM probes are significantly smaller than other bacterial and archaeal cells in the samples. A filtration-based method was used to concentrate cells for ultrastructural characterization.

A biofilm from the A drift (fig. S1) was homogenized, and the fraction of the biomass that passed through 0.45-μm filters was collected (*13*). To verify that ARMAN-lineage organisms dominated the filtrate, we extracted genomic DNA and amplified the 16*S* rRNA genes with the ARMAN-, archaeal-, and bacterial-specific primers. Significant amplification product was obtained only with the ARMAN primers (fig. S3). Cloning and sequencing of this amplification product (24 clones) revealed that the filtrate was dominated by ARMAN-2 organisms but contained some ARMAN-1 organisms. Cells in the filtrate were labeled with ARMAN-specific fluorescent probes (Fig. 3B), and the resulting images were compared with those obtained from environmental samples (Fig. 3A). The results confirm that ARMAN organisms are highly enriched by filtration (Fig. 3B).

The filtrate was frozen under high pressure and cryo-substituted for transmission electron microscope (TEM)–based imaging of cell size and morphology (*13*). In addition to ruptured cell membranes, we detected some morphologies typical of *Leptospirillum* cells known to be present at low abundance in the filtrate, as well as rounded objects of highly variable size (average diameter ~85 nm;

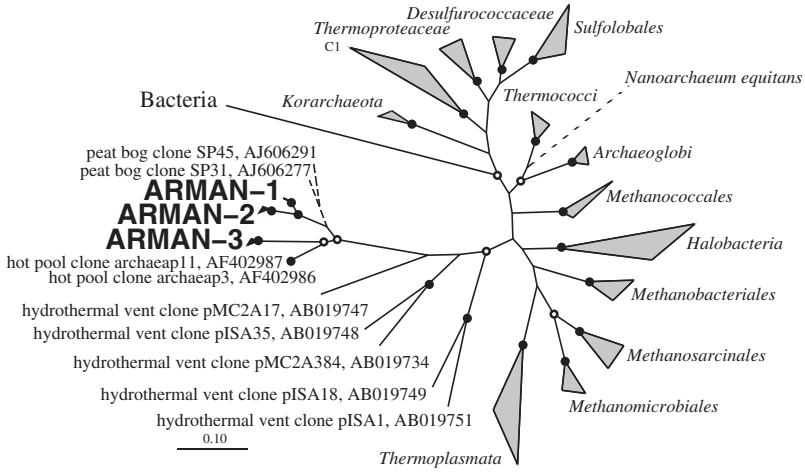

Fig. 2. Phylogeny of the ARMAN groups and several other archaeal phyla, based on 16*S* rRNA gene sequence analysis (maximum likelihood method). Bootstrap values are indicated at the nodes (solid circles, >75%; open circles, >50%). The dashed branch to *Nanoarchaeum equitans* and the uncultured peat bog clones SP31 and SP45 (*15*) signifies that we were unable to resolve these positions in our analyses.

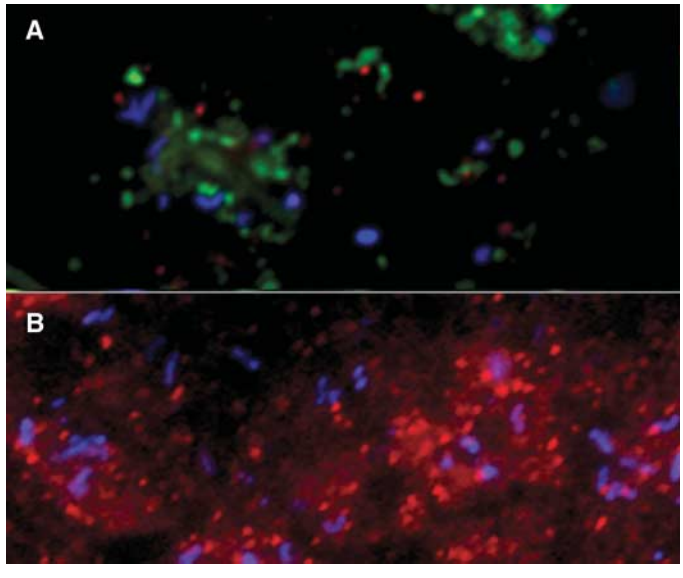

Fig. 3. Fluorescence in situ hybridization micrographs of Archaea (ARC915, fluorescein isothiocyanate in green), Bacteria (EUB338, Cy5 in blue), and ARMAN (ARM980 and ARM1357, Cy3 in red) in (**A**) an A-drift biofilm sample and (**B**) the 0.45-μm filtrate. Magnifications, 630×.

fig. S4) interpreted to be membrane buds. The other abundant cell-like objects are pleomorphic and surrounded by an archaeal-like cell wall with an S-layer. These objects have one or two folded membrane protrusions (Fig. 4 and fig. S5) that somewhat resemble structures of unknown function observed in the crenarchaeon *Pyrodictium abyssi* (*18*). On the basis of 60 measurements (excluding the protrusions), the sectioned cell-like objects have a mean length of 244 nm (range, 193 to 299 nm) and a mean width of 175 nm (range, 129 to 207 nm). The colocalization of DNA stain and probes binding specifically to ARMAN rRNA, the existence of clearly contiguous external membranes, and the presence of an archaeal-type cell wall indicate that these objects are the ARMAN cells. The apparently dense packing of the ribosomes may account for the strength of the fluorescence signal in optical microscope images, despite the small cell size. The TEM data confirm the inference from the filtration experiment that the ARMAN-lineage organisms are extremely small.

It has not escaped our attention that if the average cell size estimated from TEM observations is accurate, the ARMAN cells have volumes of <0.006 μm^3 [calculated from the maximum size observed in ultramicrotomed sections (*13*)], making them smaller than any other known cellular life form. The smallest known Archaea, of the phylum Nanoarchaea, have cell volumes that range from 0.02 to 0.70 μm^3 (*19*). They are obligate parasites of other Archaea (*20*, *21*). The smallest Bacteria include those described from Greenland ice cores, with cell volumes of 0.04 to 0.10 μm^3 (*22*), and members of the SAR11 clade, which inhabit the oligotrophic open ocean and have cell volumes of 0.031 to 0.051 μm^3 (*23*, *24*). Note that these data do not rule out the possibility of larger cytoplasmic volumes if there are unobserved connections between the objects that appear to be cells.

If the ARMAN cells have volumes of <0.006 μm^3, and half of the cell volume is occupied by ribosomes, then there is enough room for at least 350 ribosomes of normal size. However, the cells are smaller than the minimum size expected on the basis of theoretic considerations for free-living cells (*25*). We have no evidence to suggest that the ARMAN cells are parasitic on other community members. If future work demonstrates that ARMAN cells are viable and have cell volumes of <0.006 μm^3, it may be necessary to reconsider existing paradigms for the minimum requirements for life.

The ARMAN groups expand the variety of archaea known to be associated with AMD. Further targeted genomic characterization of cells concentrated by filtration should help to elucidate the ecological roles of these tiny, enigmatic, uncultivated microorganisms. Our findings emphasize the possibility that novel organisms with unexpected characteristics remain to be discovered among the relatively low-abundance members of microbial communities.

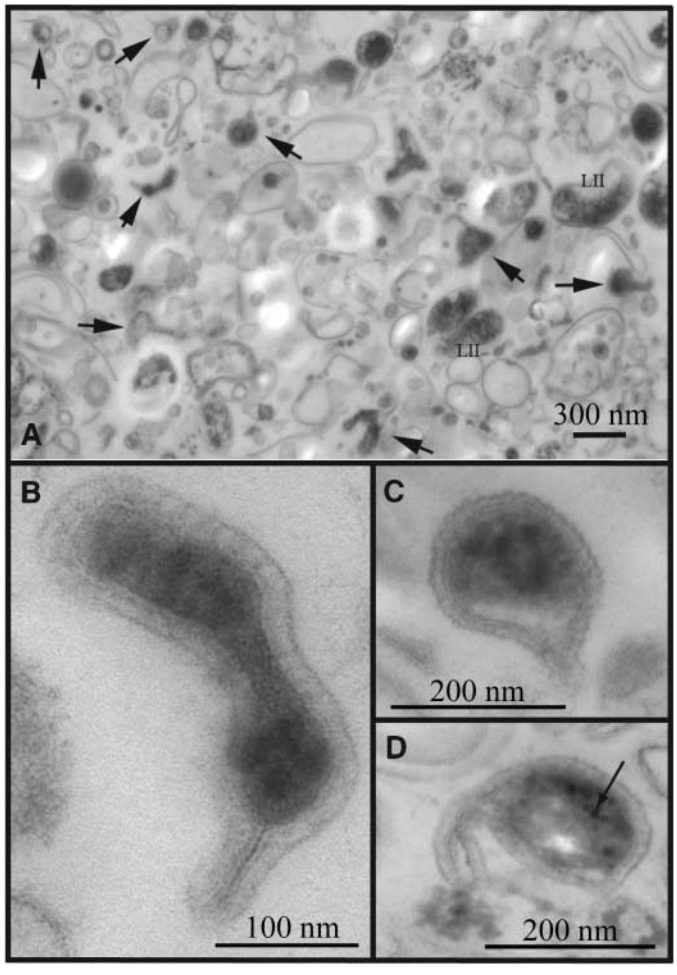

Fig. 4. Transmission electron microscope images of four cells inferred to belong to the ARMAN group. (**A**) A large area showing many cells of the ARMAN type, a subset of which are indicated by arrows. Also present are rounded objects, membrane debris, and a small number of *Leptospirillum* group II cells (LII). (**B**) Most cells exhibit one or two folded membrane-bounded protrusions (see fig. S5). (**C**) Most cells appear to be surrounded by an S-layer with periodicity in the cell surface. The dark internal contrast is consistent with densely packed ribosomes. (**D**) Some cells have very dark inclusions (arrow).

References and Notes

1. P. Hugenholtz *et al.*, *J. Bacteriol.* **180**, 4765 (1998).
2. B. J. Baker *et al.*, *Environ. Microbiol.* **5**, 267 (2003).
3. C. A. Francis *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 14683 (2005).
4. G. W. Tyson *et al.*, *Nature* **428**, 37 (2004).
5. O. Béjà *et al.*, *Science* **289**, 1902 (2000).
6. M. R. Rondon *et al.*, *Appl. Environ. Microbiol.* **66**, 2541 (2000).
7. E. F. DeLong *et al.*, *Science* **311**, 496 (2006).
8. S. G. Tringe *et al.*, *Science* **308**, 554 (2004).
9. J. C. Venter *et al.*, *Science* **304**, 66 (2004); published online 4 March 2004 (10.1126/science.1093857).
10. B. J. Baker, J. F. Banfield, *FEMS Microb. Ecol.* **44**, 139 (2003).
11. G. K. Druschel *et al.*, *Geochem. Trans.* **5**, 13 (2004).
12. P. L. Bond *et al.*, *Appl. Environ. Microbiol.* **66**, 3842 (2000).
13. See supporting material on *Science* Online.
14. A. Sunna, P. L. Bergquist, *Extremophiles* **7**, 63 (2003).
15. O. R. Kotsyurbenko *et al.*, *Environ. Microbiol.* **6**, 1159 (2004).
16. N. Nomura *et al.*, *J. Bacteriol.* **180**, 3635 (1998).
17. B. J. Baker *et al.*, *Appl. Environ. Microbiol.* **69**, 5512 (2003).
18. G. Rieger *et al.*, *J. Struct. Biol.* **115**, 78 (1995).
19. H. Huber *et al.*, *Res. Microbiol.* **154**, 165 (2003).
20. H. Huber *et al.*, *Nature* **417**, 63 (2002).
21. E. Waters *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 12984 (2003).
22. V. I. Miteva, J. E. Brenchley, *Appl. Environ. Microbiol.* **71**, 7806 (2005).
23. R. R. Malmstrom *et al.*, *Appl. Environ. Microbiol.* **70**, 4129 (2004).
24. S. J. Giovannoni *et al.*, *Science* **309**, 1242 (2005).
25. Steering Group for the Workshop on Size Limits of Very Small Microorganisms, National Research Council, *Size Limits of Very Small Microorganisms: Proceedings of a Workshop* (National Academies Press, Washington, DC, 1999).

26. Supported by the NSF Biocomplexity Program, the DOE Genomics:GTL program, and the NASA Astrobiology Institute. We thank T. Arman (owner, Iron Mountain Mines) and R. Carver for access to the site; K. Nealson, A. Driks, C. Wimpee, and B. Hedlund for their comments; K. MacDonald for assistance with preparation of sample for TEM analyses; J. R. Giska for assistance with *arsC* amplification, and J. C. Detter and P. Richardson for library construction and genome sequencing support. Genomic sequencing was carried out at the Joint Genome Institute. The fully annotated genomic fragment belonging to ARMAN-1 was deposited under GenBank accession number AY652726, and the ARMAN-2 fragment was deposited under number DQ848677.

Supporting Online Material

www.sciencemag.org/cgi/content/full/314/5807/1933/DC1
Materials and Methods
Figs. S1 to S5
Table S1
References

18 July 2006; accepted 14 November 2006
10.1126/science.1132690