Supporting Information

Baker et al. 10.1073/pnas.0914470107

SI Materials and Methods

Phylogenetic Analyses. Phylogenetic trees were constructed using the ARB software package and a Kimura/Dayhoff correction (1). *Giardia lamblia* (ATCC 50803) was used as the outgroup in all trees. Bootstrapping analyses were done using TREE_PUZZLE quarteting with defaults settings (2), and these values were superimposed on the FastDNAML (3) and Protein_ML trees shown in Figs. S1 and S2. The same 26 strains were used to generate the RadA, EF-2, and EF-1a trees shown in Fig. S2.

Fluorescent in Situ Hybridization. New FISH probes were designed to target ARMAN (archaeal Richmond Mine acidophilic nanoorganisms)-4 and -5 using methods described previously (4). The signal from these two probes was slight. To strengthen the signal, we used four unique probes targeting different regions of the ribosome. These probe sequences are as follows; ARM331 5′-GTCTCAGTACCCTTCTGGGG-3′, ARM329 5′-CCGYAGTGCTAGGGTCCTTC-3′, ARM424 5′-GGCAAAAGTTCCCTTCCGG-3′, and ARM927 5′-ACCCGTTTTTGTGCTCCC-3′. Probes were optimized for hybridization at 20% formamide concentration with a 46 °C incubation and washing at 48 °C.

Cryo-Electron Tomographic Imaging. Samples were removed from liquid nitrogen and loaded into the electron microscope sample holder for data acquisition. Images were acquired on a JEOL–3100 transmission electron microscope equipped a FEG electron source operating at 300 kV, an Omega energy filter, a Gatan 795 2K×2K CCD camera, and cryo-transfer stage. The stage was cooled using liquid nitrogen to 80 K.

To have a statistically relevant survey of cell sizes and morphologies, over 600 images or 2D projections were recorded using magnifications of 36K×, 30K×, and 25K× at the CCD, giving a pixel size of 0.83 nm, 1.0 nm, or 1.2 nm at the specimen, respectively. Underfocus values ranged between 8 ± 0.5 and 14 ± 0.5 μm, and energy filter widths were typically around 22 ± 2 eV. The survey of the grids and the selection of suitable targets for tilt-series acquisition were done in low-dose diffraction mode through the acquisition of hundreds of images.

Tomographic tilt series were acquired under low dose conditions, typically over an angular range between +65° and −65°, ± 5° with increments of 1° or 2°. Between 70 and 124 images were recorded for each series. A total of 69 3D data sets were obtained: 11 tilt series were acquired manually with the program Digital Micrograph (Gatan, Inc.), and 58 tilt series were acquired semi-automatically with the program Serial-EM (http://bio3d.colorado.edu/) adapted to JEOL microscopes. All images were recorded using a magnification of 36K×, 30K×, or 25K× at the CCD giving a pixel size of 0.833 nm, 1.0 nm, or 1.2 nm at the specimen, respectively. Underfocus values ranged between 9 ± 0.5 μm and 16 ± 0.5 μm, depending on the goal of the dataset, and energy filter widths ranged between 22 and 28 eV, also depending on the dataset. For all datasets, the maximum dose used per complete tilt series was ≈150 e-/Å^2, with typical values of ≈100 e-/Å2.

Image Processing and Analyses. All tomographic reconstructions were obtained with the program Imod (http://bio3d.colorado.edu/). The program ImageJ (National Institutes of Health, http://rsb.info.nih.gov/ij/) was used for analysis of the 2D image projections. Volume rendering and image analysis of tomographic reconstructions was done using the program VisIt (http://www.llnl.gov/visit). All movies were done with the package ffmpeg (www.ffmpeg.org).

Sequence Assembly, Annotation, and Analyses. Sequencing reads were assembled using Phred/Phrap software. Multiple assemblies were run with various parameters and the best assembly was chosen. Assemblies were manually curated to resolve errors using standard approaches (5). Initial binning of contigs used coverage, GC content, and blastp matches with GenBank and genomic datasets for acid mine drainage (AMD) biofilm organisms (National Center for Biotechnology Information). Comparison with AMD datasets indicated that, in addition to ARMAN-4 and -5, genomic fragments of Archaea, G-plasma (∼5×), A-plasma (∼5×), ARMAN-2 (∼2×), and *Ferroplasma* Type II (∼2×) were present in the filtrate dataset.

Genome fragments of ARMAN-4 and -5 are highly syntenous. Refinement of the binning made use of synteny between fragments in these assemblies. Finally, binning and assembly completeness were evaluated using tetranucleotide frequency sequence signatures, analyzed using emergent self-organizing maps, as previously described (5). Tetranucleotide-emergent self-organizing maps generated clearly defined clusters for each ARMAN group, providing strong confirmation that all fragments deriving from each ARMAN populations were included in the analysis (6).

It is known that multiple displacement amplification of genomic DNA results in the formation of chimeras (7). On average, about 5% of the filtrate library reads in this study were chimeric (typically chimeras involved sequences derived from genomic locations a few kilobases apart). The chimera incidence was estimated based on a direct comparison of *Thermoplasmatales* reads mapped onto fragments from nonamplified community libraries. All of the ARMAN contigs used in this study were manually checked for chimeras and all instances were removed from the consensus sequences of the genomes.

The annotations of ARMAN-2, and the ARMAN-4 and -5 genomes used the Microbial Genome Annotation pipeline at Oak Ridge National Laboratory. Three gene modeling programs were run on all contigs using default settings that permit overlapping genes. Generation, (ORNL, http://compbio.ornl.gov/generation/) Glimmer, and Critica (v1.05) were used to predict ORFs. tRNAScanSE was used to located tRNAs (8). These annotations were then manually curated, which mainly involved checking for artifacts such as genes split into two frames because of rare frame-shift sequencing errors. Gaps were searched for missing ORFs, and missing ORFs were identified at the ends of contigs.

The BLASTP e-value threshold was set at 10^{-5}. The Pfam and TIGRFAMs threshold was set at the trusted cutoff for each model. Enzyme Catalog references were obtained from PRIAM using a search cutoff of $1e^{-20}$. Clusters of orthologous groups were run using the default cutoff. Identification of tRNAs was done using tRNAscan-SE, run using the prokaryotic default settings.

1. Ludwig W, et al. (2004) ARB: A software environment for sequence data. *Nucleic Acids Res* 32:1363–1371.
2. Schmidt HA, Strimmer K, Vingron M, von Haeseler A (2002) TREE-PUZZLE: Maximum likelihood phylogenetic analysis using quartets and parallel computing. *Bioinformatics* 18:502–504.
3. Felsenstein J (1981) Evolutionary trees from DNA sequences: A maximum likelihood approach. *J Mol Evol* 17:368–376.
4. Baker BJ, et al. (2006) Lineages of acidophilic archaea revealed by community genomic analysis. *Science* 314:1933–1935.
5. Lo I, et al. (2007) Strain-resolved community proteomics reveals recombining genomes of acidophilic bacteria. *Nature* 446:537–541.
6. Dick GJ, et al. (2009) Community-wide analysis of microbial genome sequence signatures. *Genome Biol* 10:R85.

7. Lasken RS, Stockwell TB (2007) Mechanism of chimera formation during the multiple displacement amplification reaction. *BMC Biotechnol* 7:19.

8. Schattner P, Brooks AN, Lowe TM (2005) The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. *Nucleic Acids Res* 33:W686–W689.

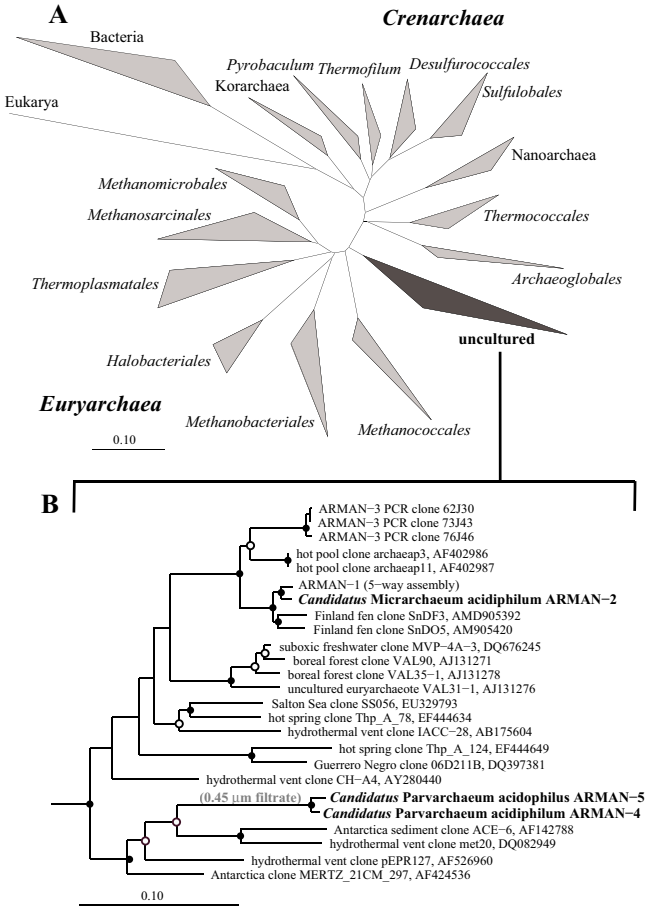


Fig. S1. Phylogenetic trees (maximum likelihood) of ARMAN group 16S rRNA genes. (*A*) The placement of ARMAN clade relative to cultured lineages. (*B*) A detailed tree of the ARMAN clade with the most closely related sequences (>800 bp in length) from other environments. Bootstrap values (TREE_PUZZLE method) labeled at the nodes with circles: >75 are solid circles and >50 are unfilled circles. Scale bars indicate changes per site.

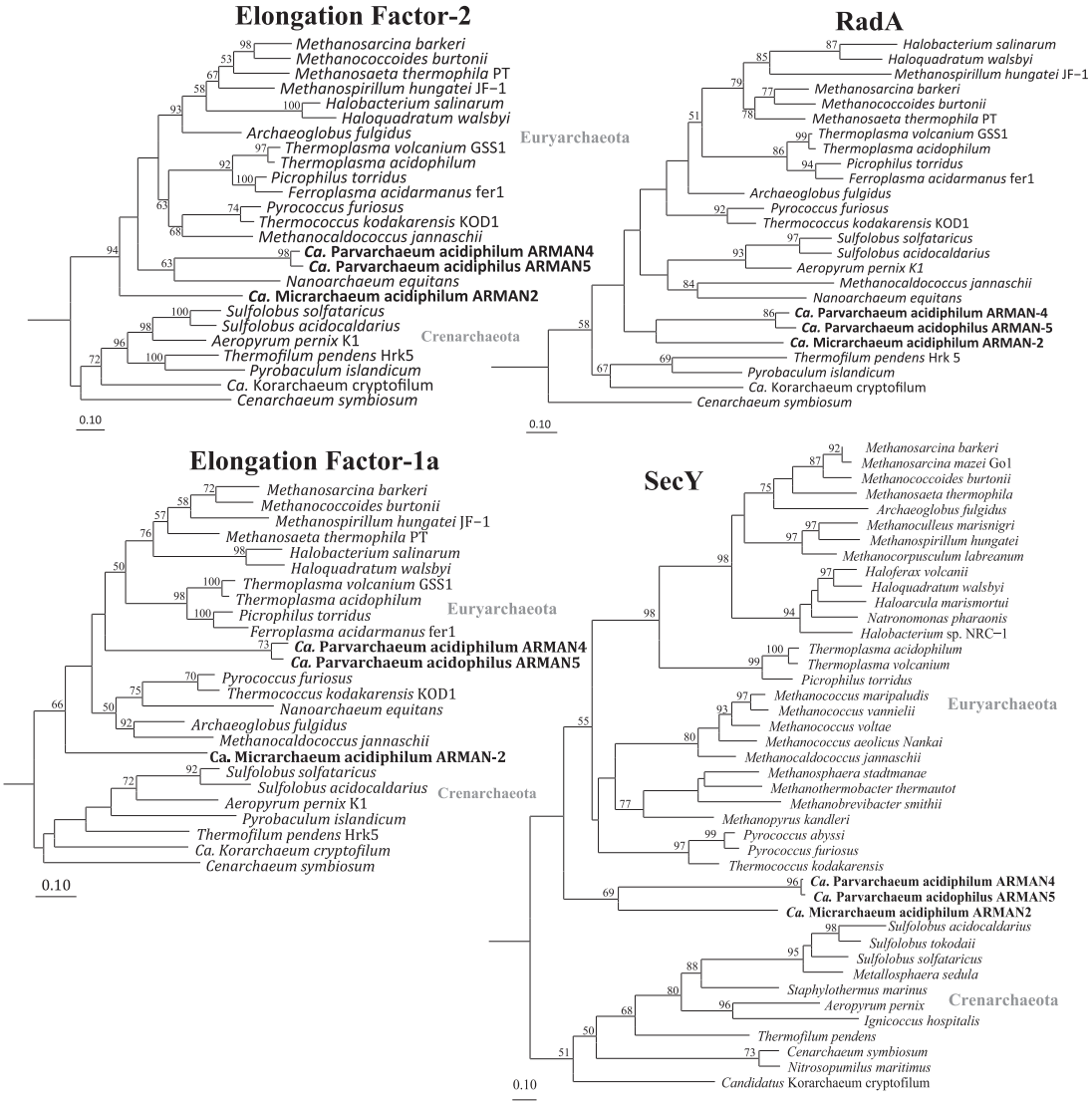


Fig. S2. Phylogenetic analyses (generated using maximum likelihood (Protein_ML) in the ARB software package) of predicted proteins from the ARMAN genomes. Numbers at the nodes are bootstrap statistical supports for each of the branchings.

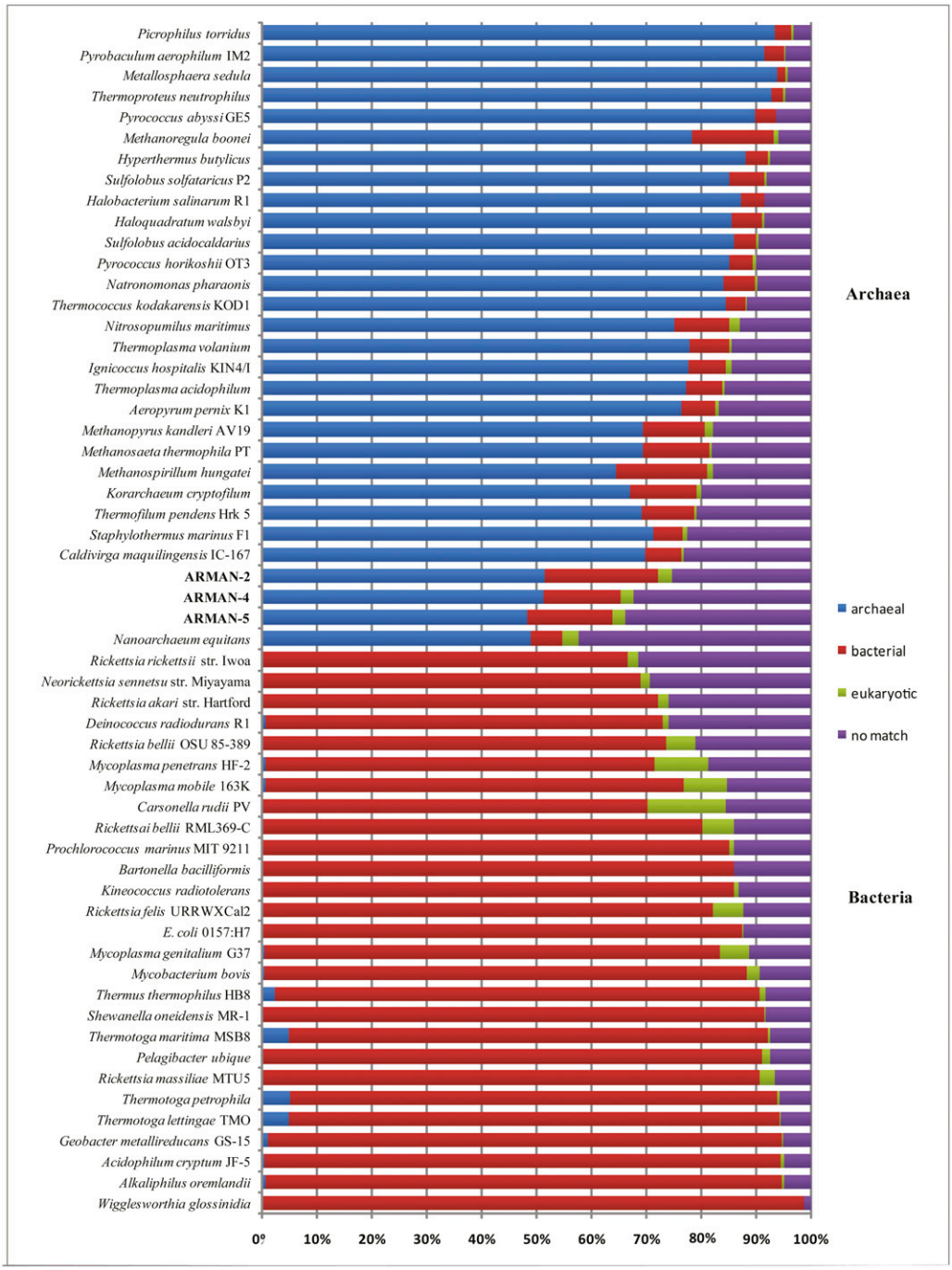


Fig. S3. Chart showing the percentage of top NCBI hits (blastp) to predicted proteins from each of the three domains for a set of sequenced bacterial and archaeal genomes. The percentage of "no hits" is high for ARMAN compared with other genomes.

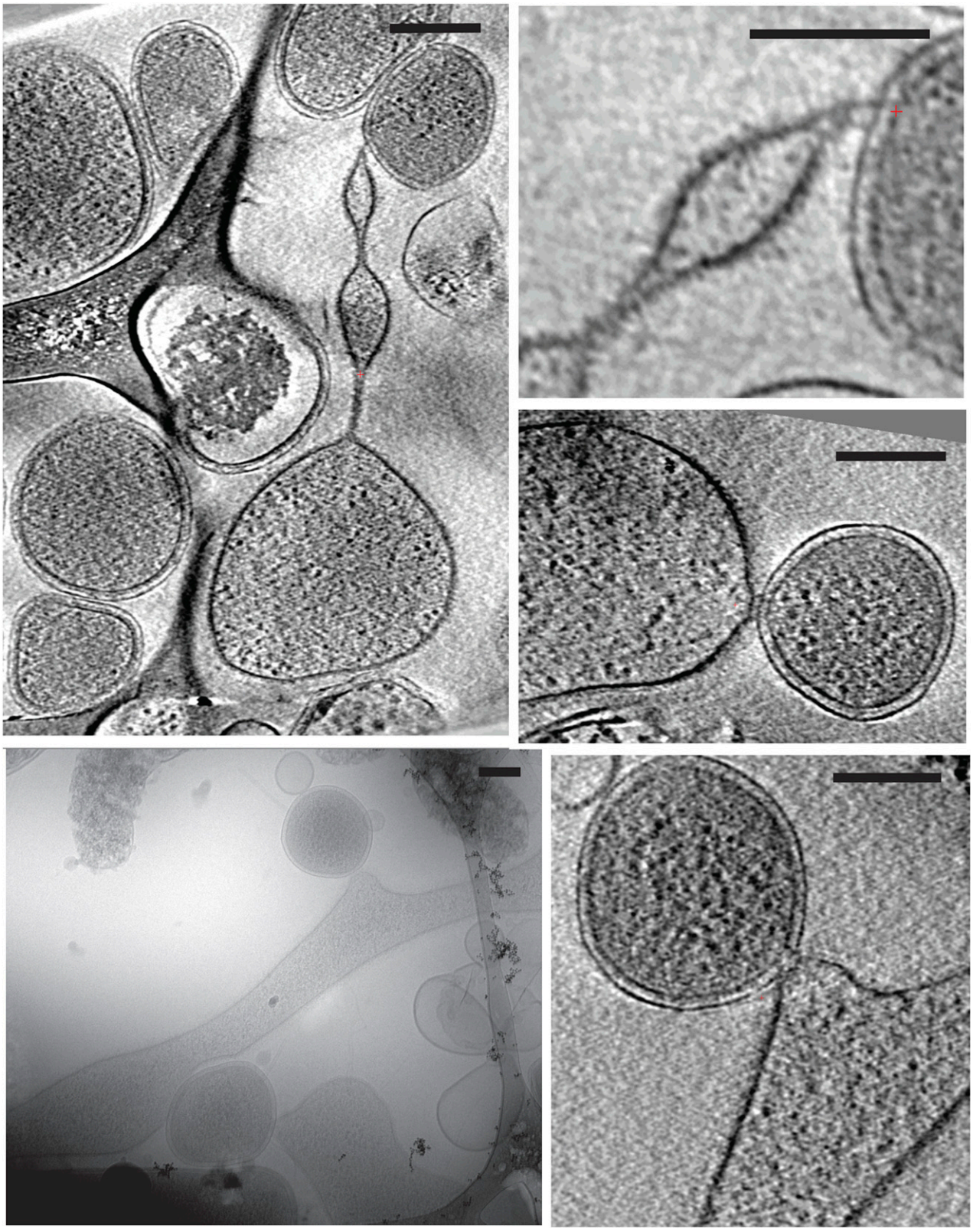


Fig. S4. One voxel-thick slices through tomographic reconstructions of ARMAN cells with associations with *Thermoplasma*-like cells. (Scale bars, 250 nm.)

Table S1. List of clusters of orthologous groups matches for *Ca. M. acidiphilum* ARMAN-2, *Ca. P. acidiphilum* ARMAN-4, and *Ca.* P. acidophilus ARMAN-5 genes that are represented in the bacterial domain, but have not been seen previously in other Archaea

Gene	COG	Function	Top BLASTP hit	E-value
Ca. M. acidiphilum ARMAN-2				
UNLARM2_2	COG3545	Esterase of the α/β hydrolase fold	Hypothetical protein LPC_0082 (*Legionella pneumophila* str. Corby)	2.00E-36
UNLARM2_115	COG0635	Coproporphyrinogen III oxidase and related Fe-S oxidoreductase	Coproporphyrinogen III oxidase (*Cyanothece* sp. CCY0110)	1.00E-41
UNLARM2_160	COG4191	Signal transduction histidine kinase regulating C4-dicarboxylate transport system	Putative two-component sensor kinase (*Azoarcus* sp. BH72)	1.00E-64
UNLARM2_341	COG1434	Conserved hypothetical protein	Protein of unknown function DUF218 (candidate division TM7 genomosp GTL1)	2.00E-14
UNLARM2_359	COG3694	ABC-type uncharacterized transport system permease component	Conserved hypothetical protein (*Streptococcus agalactiae* 515)	3.00E-14
UNLARM2_360	COG4587	ABC-type uncharacterized transport system permease component	Permease component (*Clostridium perfringens*)	6.00E-21
UNLARM2_361	COG4586	ABC-type uncharacterized transport system ATBJBARM5_e component	ABC transporter related (*Clostridium phytofermentans* ISDg)	1.00E-56
UNLARM2_368	COG0722	3-deoxy-D-arabino-heptulosonate 7-phosphate (DAHP) synthase	Phospho-2-dehydro-3-deoxyheptonate aldolase (*Geobacter lovleyi* SZ)	1.00E-100
UNLARM2_414	COG2814	Aabinose efflux permease	Major facilitator superfamily MFS_1 (*Rhodoferax ferrireducens* T118)	5.00E-24
UNLARM2_573	COG3545	Esterase of the α/β hydrolase fold	protein of unknown function DUF1234 (*Clostridium beijerinckii* NCIMB 8052)	2.00E-21
UNLARM2_701	COG3118	Thioredoxin domain-containing protein	Thioredoxin (*Marinobacter aquaeolei* VT8)	7.00E-28
UNLARM2_876	COG0669	Phosphopantetheine adenylyltransferase	Phosphopantetheine adenylyltransferase (Pantetheine-phosphate adenylyltransferase) (PPAT)	1.00E-42
Ca. P. acidiphilum ARMAN-4				
BJBARM4_72	COG2227	2-polyprenyl-3-methyl-5-hydroxy-6-metoxy-1 4-benzoquinol methylase	Hypothetical protein B14911_25735 (*Bacillus* sp. NRRL B-14911)	3.00E-21
BJBARM4_403	COG0173	Aspartyl-tRNA synthetase	Aspartyl-tRNA synthetase (*Enterococcus faecalis* V583)	2.00E-94
BJBARM4_562	COG3118	Thioredoxin domain-containing protein	TrxA protein (*Mannheimia succiniciproducens* MBEL55E)	5.00E-24
BJBARM4_626	COG4874	Uncharacterized protein containing a pentein-type domain	Hypothetical protein PBAL39_17419 (*Pedobacter* sp. BAL39)	3.00E-76
BJBARM4_807	COG3727	DNA G:T-mismatch repair endonuclease	DNA mismatch endonuclease Vsr (*Sulfurovum* sp. NBC37-1)	6.00E-25
Ca. P. acidophilus ARMAN-5				
BJBARM5_224	COG2814	Arabinose efflux permease	Transporter permease (*Propionibacterium acnes* KPA171202)	6.00E-51
BJBARM5_589	COG3118	Thioredoxin domain-containing protein	Thioredoxin (*Pseudoalteromonas atlantica* T6c)	3.00E-24
BJBARM5_799	COG2227	2-polyprenyl-3-methyl-5-hydroxy-6-metoxy-1 4-benzoquinol methylase	Hypothetical protein BH0355 (*Bacillus halodurans* C-125)	1.00E-22
BJBARM5_990	COG3118	Thioredoxin domain-containing protein	TrxA protein (*Mannheimia succiniciproducens* MBEL55E)	5.00E-24

Table S2. Percentage of crenarchaeal and euryarchaeal core genes

Organism	*Crenarchaeota*: 201 core genes	*Euryarchaeota*: 52 core genes
Ca. M. acidiphilum ARMAN-2	44%	31%
Ca. P. acidiphilum ARMAN-4	31%	25%
Ca. P. acidophilus ARMAN-5	33%	25%
Korarchaeota	84%	63%
Nanoarchaeum	36%	17%
Euryarchaeota		
Thermoplasma acidophilum	65%	100%
Methanococcus jannaschii	59%	100%
Crenarchaeota		
Sulfolobus tokodaii	100%	83%
Pyrococcus abysii	100%	77%

Percentage of crenarchaeal and euryarchaeal core genes, as described by Elkins et al. [Elkins JG, et al. (2008) A korarchaeal genome reveals insights into the evolution of the Archaea. *Proc Natl Acad Sci USA* 105:8102–8107. in ARMAN groups that were assigned to archaeal clusters of orthologous groups and data for other Archaea.

Table S3. List of 18 clusters of orthologous groups

Cluster of ortholgous group	Description	ARMAN-4	ARMAN-5
COG1855	ATBJBARM5_e (PilT family)	No	No
COG1711	DNA replication initiation complex subunit, GINS family	No	No
COG1761	DNA-directed RNA polymerase, subunit L	No	No
COG2016	Predicted RNA-binding protein (contains PUA domain)	No	No
COG2260	Predicted Zn-ribbon RNA-binding protein	No	No
COG2888	Predicted Zn-ribbon RNA-binding protein with a function in translation	No	No
COG0130	Pseudouridine synthase	Present	Present
COG0255	Ribosomal protein L29	No	No
COG2097	Ribosomal protein L31E	Present	Present
COG1631	Ribosomal protein L44E	Present	Present
COG1383	Ribosomal protein S17E	No	No
COG2051	Ribosomal protein S27E	Present	Present
COG1718	Serine/threonine protein kinase involved in cell cycle control	Present	No
COG0681	Signal peptidase I	No	No
COG0195	Transcription elongation factor	Present	No
COG2092	Translation elongation factor EF-1beta	No	No

List of 18 clusters of orthologous groups that are present in all 34 archaeal genomes used to make the archaeal clusters of orthologous groups database, but missing in ARMAN-2. Columns to the right indicate whether these genes are present or absent in ARMAN Groups 4 and 5.

Table S4. Summary of all identified *Ca. M. acidiphilum* ARMAN-2 proteins from all proteomic datasets, along with total spectral counts (as sorted by predicted functional classes)

Protein	Description	Total sum	No. of biofilms	Average spectra	Function
UNLARM2_969	Branched-chain amino acid aminotransferase	5	1	5.0	Amino acid metabolism
UNLARM2_682	Glu/Leu/Phe/Val dehydrogenase	108	4	27.0	Amino acid metabolism
UNLARM2_350	Histidine ammonia-lyase	36	3	12.0	Amino acid metabolism
UNLARM2_722	Aspartate/glutamate/uridylate kinase	20	1	20.0	Amino acid metabolism
UNLARM2_750	Aldehyde Dehydrogenase_	10	3	3.3	Amino acid metabolism
UNLARM2_658	Ornithine carbamoyltransferase	7	2	3.5	Amino acid metabolism
UNLARM2_992	Arginyl-tRNA synthetase	6	1	6.0	Amino acid metabolism
UNLARM2_104	Asparaginyl-tRNA synthetase	6	2	3.0	Amino acid metabolism
UNLARM2_78	Isoleucyl-tRNA synthetase	5	2	2.5	Amino acid metabolism
UNLARM2_403	Aldehyde Dehydrogenase_	4	1	4.0	Amino acid metabolism
UNLARM2_908	Cysteinyl-tRNA synthetase	4	2	2.0	Amino acid metabolism
UNLARM2_183	Valyl-tRNA synthetase	4	2	2.0	Amino acid metabolism
UNLARM2_351	Acetylornithine deacetylase or succinyl-diaminopimelate desuccinylase	2	1	2.0	Amino acid metabolism
UNLARM2_961	Aminotransferase class I and II	2	1	2.0	Amino acid metabolism
UNLARM2_155	Aminotransferase class V	2	1	2.0	Amino acid metabolism
UNLARM2_717	Aminotransferase class-III	2	1	2.0	Amino acid metabolism
UNLARM2_869	Chorismate mutase	2	1	2.0	Amino acid metabolism
UNLARM2_878	Glycine hydroxymethyltransferase	2	1	2.0	Amino acid metabolism
UNLARM2_489	Seryl-tRNA synthetase	2	1	2.0	Amino acid metabolism
UNLARM2_226	Threonyl-tRNA synthetase	2	1	2.0	Amino acid metabolism
UNLARM2_735	SMC domain protein	24	7	3.4	Cell division/Replication
UNLARM2_213	Cell division/Replication protein FtsZ	7	2	3.5	Cell division/Replication
UNLARM2_644	Inorganic diphosphatase	7	2	3.5	Cell division/Replication
UNLARM2_522	DNA ligase I, ATP-dependent Dnl1	4	2	2.0	Cell division/Replication
UNLARM2_189	DNA topoisomerase I	4	2	2.0	Cell division/Replication
UNLARM2_674	SMC domain protein	3	1	3.0	Cell division/Replication
UNLARM2_895	ATBJBARM5_e involved in chromosome partitioning-like	2	1	2.0	Cell division/Replication
UNLARM2_325	Mandelate racemase/muconate lactonizing protein	86	2	43.0	Cell wall/Membrane/Lipid biogenesis
UNLARM2_148	Oxidoreductase FAD/NAD(P)-binding domain protein	6	2	3.0	Cell wall/Membrane/Lipid biogenesis
UNLARM2_421	Geranylgeranylglyceryl phosphate synthase	6	3	2.0	Cell wall/Membrane/Lipid biogenesis
UNLARM2_418	Glycosyl transferase family 2	5	2	2.5	Cell wall/Membrane/Lipid biogenesis
UNLARM2_714	Glycosyl transferase group 1	2	1	2.0	Cell wall/Membrane/Lipid biogenesis
UNLARM2_776	Glycosyl transferase group 1	2	1	2.0	Cell wall/Membrane/Lipid biogenesis
UNLARM2_739	Geranylgeranyl reductase	2	1	2.0	Cell wall/Membrane/Lipid biogenesis
UNLARM2_039	3-hydroxy-3-methylglutaryl CoA synthase-like	8	2	4.0	Cell wall/Membrane/Lipid biogenesis
UNLARM2_544	3-hydroxybutyryl-CoA dehydrogenase	5	1	5.0	Cell wall/Membrane/Lipid biogenesis
UNLARM2_386	Diphosphomevalonate decarboxylase	2	1	2.0	Cell wall/Membrane/Lipid biogenesis
UNLARM2_204	H^+-transporting two-sector ATBJBARM5_e α/β subunit central region	165	8	20.6	Energy production
UNLARM2_838	NADP oxidoreductase coenzyme F420-dependent	130	5	26.0	Energy production
UNLARM2_203	H^+-transporting two-sector ATBJBARM5_e α/β subunit central region	111	8	13.9	Energy production
UNLARM2_769	Pyruvate flavodoxin/ferredoxin oxidoreductase domain protein	39	4	9.8	Energy production
UNLARM2_762	Aconitate hydratase 1	27	6	4.5	Energy production
UNLARM2_564	Succinate dehydrogenase or fumarate reductase, flavoprotein subunit	12	2	6.0	Energy production

Table S4. Cont.

Protein	Description	Total sum	No. of biofilms	Average spectra	Function
UNLARM2_559	Malate dehydrogenase (oxaloacetate-decarboxylating) [NADP(+)]	11	3	3.7	Energy production
UNLARM2_768	Pyruvate ferredoxin/flavodoxin oxidoreductase, β subunit	5	1	5.0	Energy production
UNLARM2_561	Succinate dehydrogenase and fumarate reductase iron-sulfur protein	4	2	2.0	Energy production
UNLARM2_143	Pyruvate dehydrogenase (acetyl-transferring) E1 component, α subunit	3	1	3.0	Energy production
UNLARM2_146	Dihydrolipoamide dehydrogenase	3	1	3.0	Energy production
UNLARM2_852	Aldo/keto reductase	2	1	2.0	Energy production
UNLARM2_493	Hypothetical protein	480	5	96.0	Function unknown
UNLARM2_752	Hypothetical protein	97	4	24.3	Function unknown
UNLARM2_592	Hypothetical protein	87	2	43.5	Function unknown
UNLARM2_689	Hypothetical protein	80	3	26.7	Function unknown
UNLARM2_594	Hypothetical protein	56	2	28.0	Function unknown
UNLARM2_425	Hypothetical protein	52	4	13.0	Function unknown
UNLARM2_700	Hypothetical protein	51	4	12.8	Function unknown
UNLARM2_918	Conserved hypothetical protein	45	5	9.0	Function unknown
UNLARM2_558	Hypothetical protein	41	2	20.5	Function unknown
UNLARM2_521	Hypothetical protein	36	4	9.0	Function unknown
UNLARM2_118	Hypothetical protein	31	4	7.8	Function unknown
UNLARM2_695	3′ exoribonuclease	21	4	5.3	Function unknown
UNLARM2_291	Hypothetical protein	15	4	3.8	Function unknown
UNLARM2_778	Nucleic acid binding OB-fold tRNA/helicase-type	15	4	3.8	Function unknown
UNLARM2_308	Protein of unknown function DUF124	14	3	4.7	Function unknown
UNLARM2_151	Putative circadian clock protein, KaiC	13	1	13.0	Function unknown
UNLARM2_978	RNA-binding S4 domain protein	11	1	11.0	Function unknown
UNLARM2_261	Hypothetical protein	8	1	8.0	Function unknown
UNLARM2_830	Protein of unknown function DUF124	8	2	4.0	Function unknown
UNLARM2_367	Band 7 protein	8	1	8.0	Function unknown
UNLARM2_024	Hypothetical protein	7	1	7.0	Function unknown
UNLARM2_93	Hypothetical protein	7	3	2.3	Function unknown
UNLARM2_696	3′ exoribonuclease	7	1	7.0	Function unknown
UNLARM2_161	Conserved hypothetical protein	6	3	2.0	Function unknown
UNLARM2_127	TOPRIM domain protein	6	3	2.0	Function unknown
UNLARM2_241	Hypothetical protein	5	1	5.0	Function unknown
UNLARM2_020	Hypothetical protein	4	1	4.0	Function unknown
UNLARM2_280	Hypothetical protein	4	1	4.0	Function unknown
UNLARM2_586	Hypothetical protein	4	2	2.0	Function unknown
UNLARM2_720	Hypothetical protein	4	2	2.0	Function unknown
UNLARM2_219	Protein of unknown function DUF1621	4	1	4.0	Function unknown
UNLARM2_169	Protein of unknown function DUF87	4	2	2.0	Function unknown
UNLARM2_446	dTMP kinase	4	2	2.0	Function unknown
UNLARM2_610	Nonspecific serine/threonine protein kinase	4	1	4.0	Function unknown
UNLARM2_073	Conserved hypothetical protein	3	1	3.0	Function unknown
UNLARM2_962	Conserved hypothetical protein	3	1	3.0	Function unknown
UNLARM2_429	Protein of unknown function Met10	3	1	3.0	Function unknown
UNLARM2_987	DEAD/H associated domain protein	3	1	3.0	Function unknown
UNLARM2_836	Proliferating cell nuclear antigen PcnA	3	1	3.0	Function unknown
UNLARM2_352	Rossmann fold nucleotide-binding protein	3	1	3.0	Function unknown
UNLARM2_43	HAD-superfamily hydrolase, subfamily IIB	2	1	2.0	Function unknown
UNLARM2_718	CoA-binding domain protein	2	1	2.0	Function unknown
UNLARM2_019	Hypothetical protein	2	1	2.0	Function unknown
UNLARM2_58	Hypothetical protein	2	1	2.0	Function unknown
UNLARM2_126	Hypothetical protein	2	1	2.0	Function unknown
UNLARM2_150	Hypothetical protein	2	1	2.0	Function unknown
UNLARM2_84	Hypothetical protein	2	1	2.0	Function unknown
UNLARM2_167	Hypothetical protein	2	1	2.0	Function unknown
UNLARM2_230	Hypothetical protein	2	1	2.0	Function unknown
UNLARM2_310	Hypothetical protein	2	1	2.0	Function unknown

Table S4. Cont.

Protein	Description	Total sum	No. of biofilms	Average spectra	Function
UNLARM2_355	Hypothetical protein	2	1	2.0	Function unknown
UNLARM2_448	Hypothetical protein	2	1	2.0	Function unknown
UNLARM2_464	Hypothetical protein	2	1	2.0	Function unknown
UNLARM2_341	Protein of unknown function DUF218	2	1	2.0	Function unknown
UNLARM2_973	Protein of unknown function DUF373	2	1	2.0	Function unknown
UNLARM2_194	Protein of unknown function DUF655	2	1	2.0	Function unknown
UNLARM2_366	Chlorite dismutase	2	1	2.0	Function unknown
UNLARM2_795	DEAD/DEAH box helicase domain protein	2	1	2.0	Function unknown
UNLARM2_715	TATA-box binding family protein	2	1	2.0	Function unknown
UNLARM2_406	Dihydroxy-acid dehydratase	24	3	8.0	General metabolism
UNLARM2_725	Deoxyhypusine synthase	9	2	4.5	General metabolism
UNLARM2_555	Carboxymethylenebutenolidase	4	1	4.0	General metabolism
UNLARM2_654	GTP cyclohydrolase II	2	1	2.0	General metabolism
UNLARM2_842	Nicotinamide phosphoribosyltransferase	2	1	2.0	General metabolism
UNLARM2_279	ATP–cobalamin adenosyltransferase	2	1	2.0	General metabolism
UNLARM2_465	Thymidylate kinase	43	4	10.8	Nucleotide metabolism
UNLARM2_297	Ribonucleoside-diphosphate reductase, adenosylcobalamin-dependent	32	4	8.0	Nucleotide metabolism
UNLARM2_585	Nucleoside-diphosphate kinase	2	1	2.0	Nucleotide metabolism
UNLARM2_780	Thermosome	204	5	40.8	PTMs, protein folding and turnover
UNLARM2_459	Heat shock protein Hsp20	45	4	11.3	PTMs, protein folding and turnover
UNLARM2_453	Proteasome endopeptidase complex	44	5	8.8	PTMs, protein folding and turnover
UNLARM2_456	Chaperone protein DnaK	39	5	7.8	PTMs, protein folding and turnover
UNLARM2_701	Thioredoxin	31	4	7.8	PTMs, protein folding and turnover
UNLARM2_692	Proteasome endopeptidase complex	27	2	13.5	PTMs, protein folding and turnover
UNLARM2_177	Radical SAM domain protein	26	6	4.3	PTMs, protein folding and turnover
UNLARM2_120	AAA family ATBJBARM5_e, CDC48 subfamily	16	4	4.0	PTMs, protein folding and turnover
UNLARM2_736	Peptidase M50	11	1	11.0	PTMs, protein folding and turnover
UNLARM2_301	Peptidase S16, Lon-like protease	10	1	10.0	PTMs, protein folding and turnover
UNLARM2_47	Peptidase M1 membrane alanine aminopeptidase	9	2	4.5	PTMs, protein folding and turnover
UNLARM2_455	Chaperone protein DnaJ	8	3	2.7	PTMs, protein folding and turnover
UNLARM2_858	Peptidase U61 LD-carboxypeptidase A	4	1	4.0	PTMs, protein folding and turnover
UNLARM2_591	ATBJBARM5_e, AAA+ superfamily	3	1	3.0	PTMs, protein folding and turnover
UNLARM2_1005	Radical SAM domain protein	2	1	2.0	PTMs, protein folding and turnover
UNLARM2_886	AAA ATBJBARM5_e	2	1	2.0	PTMs, protein folding and turnover
UNLARM2_246	AAA ATBJBARM5_e central domain protein	2	1	2.0	PTMs, protein folding and turnover
UNLARM2_964	AAA ATBJBARM5_e central domain protein	2	1	2.0	PTMs, protein folding and turnover
UNLARM2_684	Alkyl hydroperoxide reductase/ Thiol specific antioxidant/ Mal allergen	2	1	2.0	PTMs, protein folding and turnover
UNLARM2_224	Putative signal transduction protein with CBS domains	26	3	8.7	Signal transduction/Intracellular trafficking
UNLARM2_917	Type II secretion system protein E	11	3	3.7	Signal transduction/Intracellular trafficking
UNLARM2_915	Signal recognition particle-docking protein FtsY	7	2	3.5	Signal transduction/Intracellular trafficking

Table S4. Cont.

Protein	Description	Total sum	No. of biofilms	Average spectra	Function
UNLARM2_94	Type II secretion system protein E	7	1	7.0	Signal transduction/Intracellular trafficking
UNLARM2_160	Signal transduction histidine kinase regulating C4-dicarboxylate transport system-like	2	1	2.0	Signal transduction/Intracellular trafficking
UNLARM2_471	RNA polymerase Rpb2 domain 6	34	7	4.9	Transcription
UNLARM2_472	RNA polymerase β subunit	33	4	8.3	Transcription
UNLARM2_450	Transcription factor TFIIB cyclin-related	23	5	4.6	Transcription
UNLARM2_470	RNA polymerase α subunit	22	4	5.5	Transcription
UNLARM2_473	RNA polymerase Rpb5	12	1	12.0	Transcription
UNLARM2_469	RNA polymerase Rpb1 domain 5	4	2	2.0	Transcription
UNLARM2_616	DNA-directed RNA polymerase	2	1	2.0	Transcription
UNLARM2_926	RNA polymerase insert	2	1	2.0	Transcription
UNLARM2_196	RNA polymerase Rpb4	2	1	2.0	Transcription
UNLARM2_250	Translation elongation factor EF-1, subunit α	194	5	38.8	Translation
UNLARM2_837	DNase/rho motif-related TRAM	28	2	14.0	Translation
UNLARM2_637	Ribosomal protein 60S	27	4	6.8	Translation
UNLARM2_921	Ribosomal protein S2	27	3	9.0	Translation
UNLARM2_730	Translation elongation factor aEF-2	16	4	4.0	Translation
UNLARM2_731	Ribosomal protein S7	11	1	11.0	Translation
UNLARM2_636	Ribosomal protein L10	8	3	2.7	Translation
UNLARM2_985	Ribosomal protein S5	8	2	4.0	Translation
UNLARM2_820	Ribosomal protein L3	5	2	2.5	Translation
UNLARM2_729	Ribosomal protein L7Ae/L30e/S12e/Gadd45	5	1	5.0	Translation
UNLARM2_660	30S ribosomal protein S8e	4	2	2.0	Translation
UNLARM2_427	Ribosomal protein L11	4	2	2.0	Translation
UNLARM2_652	Ribosomal protein S13	4	2	2.0	Translation
UNLARM2_651	Ribosomal protein S4	4	2	2.0	Translation
UNLARM2_121	Ribosomal protein S6e	4	2	2.0	Translation
UNLARM2_686	S1 IF1 family protein	4	1	4.0	Translation
UNLARM2_924	Ribosomal protein L13	3	1	3.0	Translation
UNLARM2_819	Ribosomal protein L4/L1e	3	1	3.0	Translation
UNLARM2_814	Ribosomal protein S3- domain protein	3	1	3.0	Translation
UNLARM2_76	Nucleotidyl transferase	2	1	2.0	Translation
UNLARM2_678	Ribosomal protein L16	2	1	2.0	Translation
UNLARM2_923	Ribosomal protein S9	2	1	2.0	Translation
UNLARM2_216	SSU ribosomal protein S24E	2	1	2.0	Translation

One hundred seventy-three ARMAN-2 proteins (~15% of the genome) were detected across all searched proteomes. PTM, posttranslational modification; SMC, structural maintenance of chromosomes.

Table S5. Summary of all identified *Ca. P. acidiphilum* ARMAN-4 proteins from all proteomic datasets, along with total spectral counts (as sorted by predicted functional classes)

Protein	Description	Total sum	No. of biofilms	Average spectra	Function
BJBARM4_984	Glycine hydroxymethyltransferase	6	2	3.0	Amino acid metabolism
BJBARM4_953	Alanine–tRNA ligase	2	1	2.0	Amino acid metabolism
BJBARM4_182	Phosphoadenosine phosphosulfate reductase	2	1	2.0	Amino acid metabolism
BJBARM4_946	DNA topoisomerase type IA central domain protein	2	1	2.0	Cell division/Replication
BJBARM4_758	Mandelate racemase/muconate lactonizing protein	2	1	2.0	Cell wall/Membrane/Lipid biogenesis
BJBARM4_086	H^+-transporting two-sector ATBJBARM5_e α/β subunit central region	79	4	19.8	Energy production
BJBARM4_570	Succinate dehydrogenase	33	3	11.0	Energy production
BJBARM4_902	Aldo/keto reductase	14	2	7.0	Energy production
BJBARM4_536	Aconitate hydratase 1	8	3	2.7	Energy production
BJBARM4_283	Fumarate lyase	4	1	4.0	Energy production
BJBARM4_495	Phosphoenolpyruvate synthase	4	2	2.0	Energy production
BJBARM4_761	Aldehyde Dehydrogenase	3	1	3.0	Energy production
BJBARM4_746	Succinate–CoA ligase (ADP-forming)	2	1	2.0	Energy production
BJBARM4_1007	Undefined product	21	3	7.0	Function unknown
BJBARM4_895	NADPH-dependent FMN reductase	6	1	6.0	Function unknown
BJBARM4_163	Hypothetical protein	4	2	2.0	Function unknown
BJBARM4_417	Hypothetical protein	4	2	2.0	Function unknown
BJBARM4_399	Hypothetical protein	3	1	3.0	Function unknown
BJBARM4_776	Hypothetical protein	3	1	3.0	Function unknown
BJBARM4_921	Cell Division/Replication factor C	3	1	3.0	Function unknown
BJBARM4_046	Hypothetical protein	2	1	2.0	Function unknown
BJBARM4_730	Hypothetical protein	2	1	2.0	Function unknown
BJBARM4_919	Hypothetical protein	2	1	2.0	Function unknown
BJBARM4_754	Shwachman-Bodian-Diamond syndrome protein	2	1	2.0	Function unknown
BJBARM4_452	Plasma-membrane proton-efflux P-type ATBJBARM5_e	2	1	2.0	General metabolism
BJBARM4_510	Histidine triad (HIT) protein	2	1	2.0	General metabolism
BJBARM4_025	Chaperone protein DnaK	15	2	7.5	PTMs, protein folding and turnover
BJBARM4_093	Microtubule-severing ATBJBARM5_e	13	3	4.3	PTMs, protein folding and turnover
BJBARM4_177	Thermosome	6	2	3.0	PTMs, protein folding and turnover
BJBARM4_343	Nucleotidyl transferase	2	1	2.0	Translation
BJBARM4_767	Ribosomal protein L3	2	1	2.0	Translation
BJBARM4_947	Translation elongation factor EF-1, subunit-α	2	1	2.0	Translation

Thirty-two ARMAN-4 proteins were detected across all searched proteomes.

Table S6. Summary of all identified *Ca. P. acidophilus* ARMAN-5 proteins from all proteomic datasets, along with total spectral counts (as sorted by predicted functional classes)

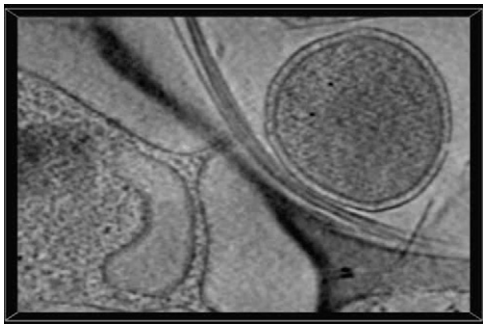
Protein	Description	Total sum	No. of biofilms	Average spectra	Function
BJBARM5_475	Zinc-binding alcohol dehydrogenase family protein	4	1	4.0	Amino acid metabolism
BJBARM5_756	Histone acetyltransferase, ELP3 family	8	2	4.0	Cell division/Replication
BJBARM5_827	DNA topoisomerase VI subunit B	2	1	2.0	Cell division/Replication
BJBARM5_656	H^+-transporting two-sector ATBJBARM5_e α/β subunit central region	80	4	20.0	Energy production
BJBARM5_821	H^+-transporting two-sector ATBJBARM5_e α/β subunit central region	78	4	19.5	Energy production
BJBARM5_597	Succinate dehydrogenase	39	3	13.0	Energy production
BJBARM5_951	H^+-transporting two-sector ATBJBARM5_e α/β subunit domain protein	30	3	10.0	Energy production
BJBARM5_415	Mercuric reductase	10	2	5.0	Energy production
BJBARM5_510	Aconitate hydratase 1	9	3	3.0	Energy production
BJBARM5_1022	Aconitate hydratase 1	9	3	3.0	Energy production
BJBARM5_160	Hypothetical protein	33	2	16.5	Function unknown
BJBARM5_1073	Hypothetical protein	33	2	16.5	Function unknown
BJBARM5_1106	Signal recognition particle receptor	7	2	3.5	Function unknown
BJBARM5_081	Hypothetical protein	4	2	2.0	Function unknown
BJBARM5_432	Hypothetical protein	4	2	2.0	Function unknown
BJBARM5_929	Hypothetical protein	4	2	2.0	Function unknown
BJBARM5_859	Hypothetical protein	3	1	3.0	Function unknown
BJBARM5_022	Hypothetical protein	2	1	2.0	Function unknown
BJBARM5_230	Hypothetical protein	2	1	2.0	Function unknown
BJBARM5_627	Hypothetical protein	2	1	2.0	Function unknown
BJBARM5_715	Hypothetical protein	2	1	2.0	Function unknown
BJBARM5_831	RNA-binding protein-like	2	1	2.0	Function unknown
BJBARM5_857	Chaperone protein DnaK	17	2	8.5	PTMs, protein folding and turnover
BJBARM5_694	RecA-superfamily ATBJBARM5_e implicated in signal transduction-like	2	1	2.0	Signal transduction/Intracellular trafficking
BJBARM5_877	RecA-superfamily ATBJBARM5_e implicated in signal transduction-like	2	1	2.0	Signal transduction/Intracellular trafficking
BJBARM5_1004	Ribosomal protein L32e	2	1	2.0	Translation
BJBARM5_089	Diphthine synthase	2	1	2.0	Translation

Twenty-seven ARMAN-5 proteins were detected across all searched proteomes.

Table S7. Number of proteins, peptides, and spectral counts attributable to ARMAN proteins across the biofilms

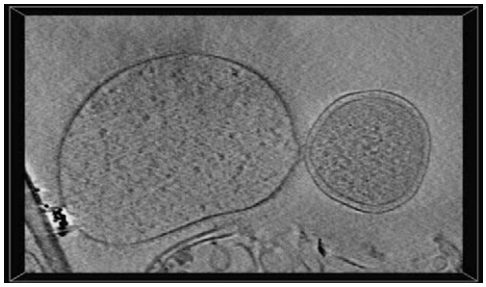
Sample type	Total spectra	ARMAN spectra	ARMAN spectra	Total peptides	ARMAN peptides	ARMAN peptides
ABend*	517,348	63	0.01%	190,600	40	0.02%
ABfront M2 lysis*	398,886	77	0.02%	166,738	44	0.03%
ABfront S lysis*	161,137	82	0.05%	82,139	61	0.07%
UBAultraback M2 lysis†	369,475	319	0.09%	178,519	277	0.16%
UBAultraback S lysis†	119,299	169	0.14%	48,891	156	0.32%
ABmuck*	396,544	97	0.02%	166,493	64	0.04%
ABmuck Friable*	475,026	128	0.03%	206,557	83	0.04%
UBA-BS M2 lysis†	205,596	944	**0.46%**	99,179	668	**0.67%**
UBA-BS S lyis†	113,991	1505	**1.32%**	56,242	1008	**1.79%**
UBA-BS2†	206,023	642	0.31%	106,376	544	0.51%

Note the highest abundance of ARMAN proteins in biofilms from the UBA BS location (in boldface). The S lysis buffer yielded higher percentages of ARMAN identifications than the M2 buffer (*Materials and Methods*).
*AMD_CoreDB_04232008, the core acid mine drainage database containing ARMAN-2 but not -4 or -5.
†amdv1allfrm_arman_AMD_CoreDB_04232008, supplemented with ARMAN-4 and -5.



Movie S1. Three-dimensional transmission cryo-electron microscopy reconstruction of ARMAN-*Thermoplasmatales* interaction. This is the same field of view shown in Fig. 4. The cell on the left is the archaeon from the order *Thermplasmatales* and the one on the right is an ARMAN.

Movie S1



Movie S2. Three-dimensional transmission cryo-electron microscopy reconstruction of ARMAN-*Thermoplasmatales* interaction. The cell on the left is the archaeon from the order *Thermplasmatales* and the one on the right is an ARMAN.

Movie S2

Movie S3. Three-dimensional transmission cryo-electron microscopy reconstruction of ARMAN-*Thermoplasmatales* interaction. The cell on the bottom is the archaeon from the order *Thermplasmatales* and the one on the top is an ARMAN.

Movie S3