Supporting Information

Denef et al. 10.1073/pnas.0907041107

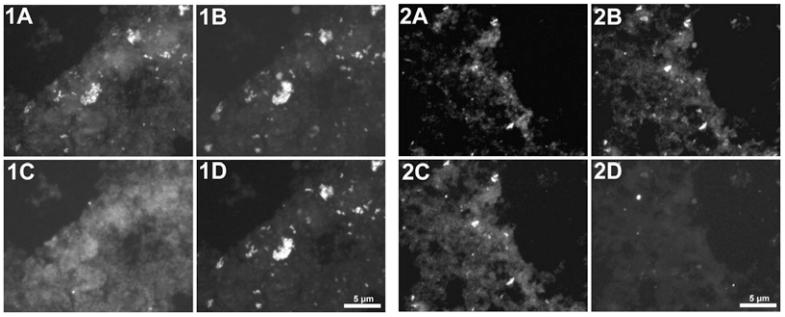

Fig. S1. Control experiments for fluorescent in situ hybridizations with *Leptospirillum* group II type–specific probes at optimal 35% formamide stringency. 1, *Leptospirillum* group II isolate CF-1 with UBA type rRNA locus; 2, *Leptospirillum* group II 5w02 isolate with five-way CG type rRNA locus, with DAPI stain (*A*), hybridized with Cy5-LF655 probe (*B*), FITC-L2CG353 probe (*C*), and Cy3-L2UBA353 probe (*D*). Background in control experiments originates from metal precipitates formed during growth of isolates.

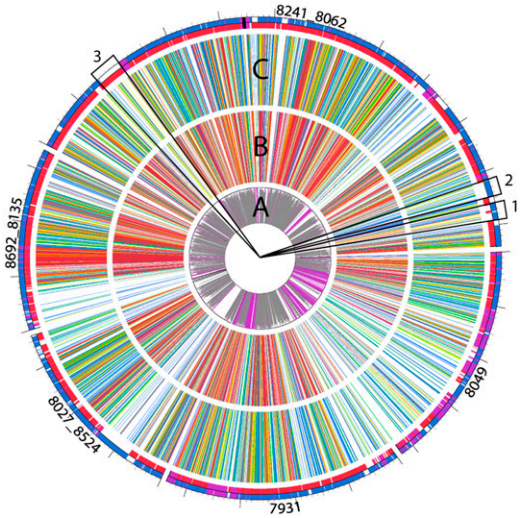

Fig. S2. Summary of the community proteomics data for *Leptospirillum* group II showing the average protein abundance (circle, C) and each protein's (unique or ortholog pair) relative occurrence (fraction of all samples, B). Outer circle represents the aligned UBA (blue) and five-way CG (red) genomes, with white space indicating the absence of an ortholog in the respective genome, purple indicating 100% identical proteins, and black indicating the CRISPR region (near the top of the circle). Markers indicate the gene number based on the aligned genomes of both organisms as in Table S5 (large, 100 genes; medium, 50 genes; small, 10 genes). Inner circle (A) indicates the percent amino acid identity between the orthologs, represented by the height of the bar (50–100%). Purple bars indicate 100% amino acid identity. Type-specific proteins were identified in considerably fewer samples than orthologs. Notable exceptions are two UBA type–specific regions on scaffold 8241_8062 between gene markers 1,060–1,090 {1} and to a lesser extent between 1,020–1,050 {2}, and a large five-way CG type–specific region of putative phage origin, between the gene markers 150 and 200 on scaffold 8241_8062 {3}.

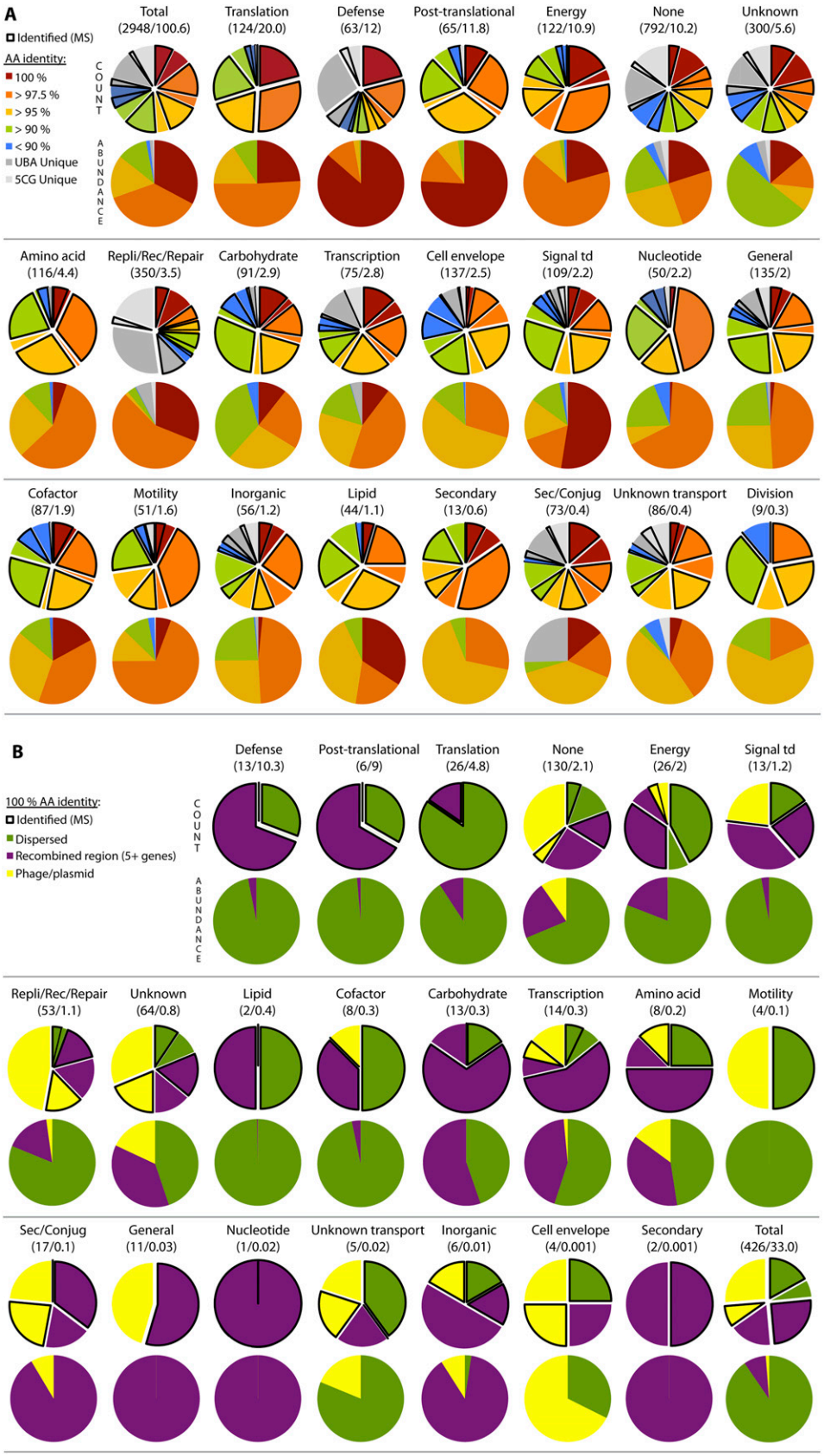

Fig. S3. (*A*) Biases in number and expression of conserved, divergent, and genotypic group–specific proteins among functional categories. Number of genes (count) and abundance of the corresponding proteins (abundance) is presented for each manually curated (as described in *Materials and Methods*) functional category and amino acid identity class (colors). In the count representation, the fraction of all genes for which the encoded protein was identified by pro-

Legend continued on following page

teomics is indicated by a black outline. In the abundance representation, the sum of the average normalized spectral abundance factors (NSAF) throughout all samples of all identified proteins for each functional category and identity class is presented. The number of genes in each category and the total abundance (sum of NSAFs) of the identified proteins are presented between brackets. (*B*) Similar to *A*, but limited to only those proteins that are 100% identical between the two *Leptospirillum* group II types.

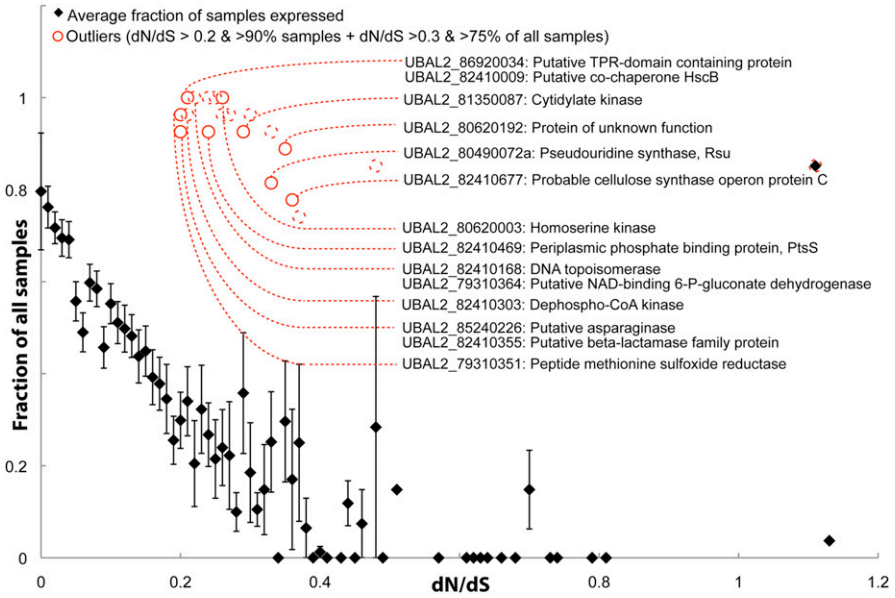

Fig. S4. Relationship between dN/dS and the fraction of all samples the protein was identified in. Average (and SE) fractions for all proteins with a dN/dS = x are presented, as well as the outliers in red. Dotted red circles already presented in Fig. 4.

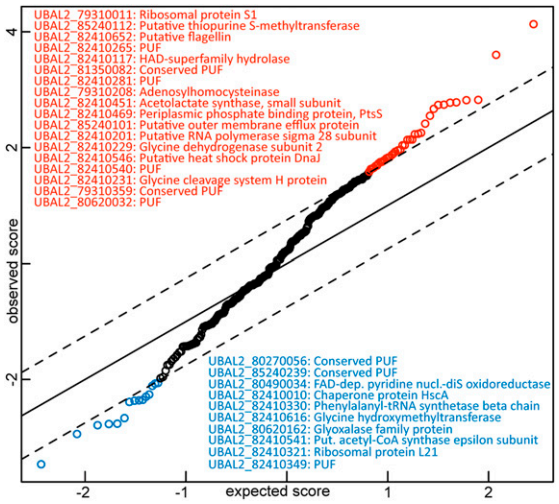

Fig. S5. SAM analysis at a false discovery rate <5% (delta = 0.68), with proteins more abundant in the five-way CG type in red, and those more abundant in the UBA type marked in blue. Gene numbers and product description are listed for those proteins that are significantly more abundant in either type and display at least a 1.5-fold difference between the UBA and five-way CG type proteomic datasets (based on clustering in Fig. 5). PUF, protein of unknown function.

Other Supporting Information Files

Table S1. Environmental parameters for AMD -air interface bioflim samples

Standards used for pH measurement were pH 1.00 and pH 1.68 (Ricca Chemical Company) (map of Richmond Mine in Fig. 2). together with standards prepared from standardized sulfuric acid (Fisher Scientific) for the pH range 0 -2 (1). Samples for iron(II)/total iron were filtered in the mine with 0.1-μm syringe filters (Supor membrane, Pall Corporation), preserved immediately with 1% trace-metal grade 6N HCl (Fisher Optima) and stored at 4°C in amber HDPE bottles. Samples for total metal analysis were 0.1 μm filtered within 6 h of sample collection, preserved with 1% trace-metal grade nitric acid (Fisher Optima) and stored at 4°C until analysis. Ferrous and total iron were measured using a ferrozine method (2) within 1 week of sample collection. Metals were determined by inductively coupled plasma–optical emission spectroscopy using a Perkin-Elmer 5300 DV or Leeman Labs Direct Reading Echelle. Replicate samples run on both instruments agreed within 5% for the metals and within 10% for arsenic.

*Fluid dynamics parameter (the higher the more rapid flow and/or the narrower the flow path);

†for subsequent statistical analyses, time was coded incrementally based on the number of months elapsed since the first sampling time (March 2002 = 1);

‡biofilm developmental stage;

§flow measured at entrance of the Richmond Mine;

¶conductivity of the AMD solution. Conductivity standards (12.9 and 111 mS/cm, Thermo Scientific Orion), corrected to in situ temperature, were used to calibrate the conductivity probe on-site.

1. Nordstrom DK, Alpers CN, Ptacek CJ, Blowes DW (2000) Negative pH and extremely acidic mine waters from Iron Mountain, California. *Environ Sci Technol* 34:254–258.

Table S1 (DOC)

Table S2. Biological data for sample set

(*i*) FISH data with standard error between brackets;Per sample, five representative fields of view were imaged; one randomly selected quadrant from each of the five fields was counted. Standard error was calculated over these technical replicates, except for samples where biological replicates were available, in which case the error between biological replicates was reported.

(*ii*) when a proteomics data set was available, the fractions of the different genome types as determined by proteomics-inferred genome typing (2); and

(*iii*) number of *Leptospirillum* group II proteins identified.

1. Denef VJ, et al. (2009) Proteomics-inferred genome typing (PIGT) demonstrates inter-population recombination as a strategy for environmental adaptation. *Environ Microbiol* 11: 313–325.
2. To TB, Nordstrom DK, Cunningham KM, Ball JW, McCleskey RB (1999) New Method for the Direct Determination of Dissolved Fe(III) Concentration in Acid Mine Waters.. *Environ Sci Technol* 33:807–813.

Table S2 (DOC)

Table S3. Summary of divergent genes encoding proteins identified by proteomics (<90% amino acid identity), proteins with significant differences in abundance between the two genotypic groups, and outliers in dN/dS analysis

*Locus tag for *Leptospirillum* group II UBA (UBAL2_) or five-way CG (CGL2_). Unique proteins to either genome are listed if they were identified in 33% or more of the evaluated samples.

†Outlying values presented only.

‡ Score (d)/fold change determined by SAM for proteins significantly more abundant at FDR of 7% (d > 0 indicates protein more abundant in five-way CG).

§Located in recombinant region of recombinant types X.

¶Expressed divergent protein, 50% of samples.

PAmino acid identity between UBA and five-way CG orthologs.

**Average protein abundance (NSAF).

††Number of samples in which the protein was identified.

Table S3 (DOC)

Table S4. SAM analysis (FDR < 0.05 at delta = 0.68), sorted by functional category and ranked by score

Gray indicates < 1.5-fold change.

Table S4 (DOC)

Table S5. Annotation, role category, percent amino acid identity between orthologs and proteomics data summary for Leptospirillum group II UBA and five-way CG (aligned representation)

Table S5 (XLS)