

Supplementary Information for Brazelton & Baross:

“Abundant transposases encoded by the metagenome of a hydrothermal chimney biofilm”

DNA extraction and metagenomic sequencing

The carbonate chimney sample (H03_072705_R0424) was collected from the central 'Poseidon' edifice of the Lost City Hydrothermal Field (depth, 735 m; latitude, 30.12; longitude, -42.12) on 27 July 2005 by the DSV *Hercules* during the 2005 Lost City Expedition aboard the R/V *Ronald H. Brown*. Chimney material was frozen at -80°C immediately after collection and remained frozen until onshore analysis. DNA was extracted according to a protocol modified from previous reports (Brazelton *et al.*, 2006; Barton *et al.*, 2006) and summarized here. After crushing a frozen carbonate sample with a sterile mortar and pestle, approximately 0.25 – 0.5 g of chimney material were placed in a 2 mL microcentrifuge tube containing 250 µL of 2x buffer AE (200 mM Tris, 50 mM EDTA, 300 mM EGTA, 200 mM NaCl, pH 8) and 2 µg of poly-dIdC (Sigma-Aldrich) and incubated at 4°C overnight to allow chelation of salts and binding of DNA to poly-dIdC. Between 36-72 replicate tubes were processed in parallel, and a total of ~1 kg of carbonate minerals were processed. This protocol involves no size fractionation: no organisms, regardless of size, contained in the carbonate chimney sample were excluded prior to cell lysis. Proteinase K (final concentration 1.2 mg/mL) and 10 µL of 20% SDS were added to each tube before incubation at 37°C for at most 30 min. A further 150 µL of 20% SDS and 500 µL of phenol:chloroform:isoamyl alcohol (25:24:1 ratio by volume) were added to each tube before centrifugation at 12,000 g for 10 min. Supernatants were transferred to clean tubes for a second phenol:chloroform:isoamyl alcohol extraction. After centrifugation, supernatants were pooled into SnakeSkin dialysis tubing (Pierce) and dialyzed against 20 mM EGTA overnight at 4°C. This large scale dialysis step proved to be very efficient in removing inorganic minerals and organic inhibitors. After dialysis, DNA was precipitated by adding 0.1 vol 3M sodium acetate and 1 vol isopropanol and

stored at -20°C for 2-4 hours. Pellets were collected by centrifugation at 16,000g for 20 min at 8°C, washed once in 70% ethanol, dried in a vacuum centrifuge, and resuspended in TE (10 mM Tris, 1mM EDTA, pH 8). Typical yield was ~35 mg of DNA per g of carbonate chimney material. Library construction and end-sequencing of pUC18 inserts was conducted according to the standard protocols available at the DOE Joint Genome Institute website:

http://www.jgi.doe.gov/sequencing/protocols/protos_production.html. Two libraries were constructed from two subsamples of the same carbonate chimney sample. Reads from both libraries were combined for assembly and for analyses described here. Mean read length for the 46 361 reads from both libraries was 755 bp, and the mean length of all 6324 contigs was 1583 bp.

Quantification of transposases

Quantitative comparisons among assembled metagenomes are not straightforward when the assemblies contain contigs and scaffolds of various numbers, sizes, and sequence coverages. Therefore, our study utilized only unassembled reads from the Lost City dataset and those in the CAMERA database for purposes of comparing transposase abundances. Results of TBLASTN searches conducted via the CAMERA website were downloaded as .csv files, and a Python script was used to count the number of reads with at least one TBLASTN match (of E value better than 10^{-5}) to one of 852 protein sequences representing the seed alignments for 29 transposase families in the PFAM database. Our results are reported as percentage of total unassembled reads to account for variations in the sequencing coverage of different metagenomic studies. By making use of the PFAM seed sequences as the query, this approach is more exhaustive and less error prone than simply counting keywords in BLAST results, which relies on the accuracy of the BLAST hits' annotations.

Additional analyses

We also investigated the presence of resolvase domains (PFAM entries PF07508 and PF00239) and inverted repeats in the Lost City metagenomic data. Many replicative transposons (those that replicate themselves during transposition) encode a resolvase enzyme that is required for replication. Only 48 Lost City metagenomic reads contained resolvase domains, indicating that replicative transposons are not abundant. Many transposons are associated with terminal inverted repeats as a result of their integration mechanisms. We searched for inverted repeats in the Lost City assembled contigs with the Inverted Repeat Finder (<http://tandem.bu.edu>, Warburton *et al.*, 2004). Of all 6324 contigs, 303 (4.8%) contained inverted repeats. Of the 690 contigs containing transposases, 63 (9.1%) contained inverted repeats. The scarcity of inverted repeats could indicate that the abundant transposases detected in this study have not been recently active, and the inverted repeats have degraded over time. Because most of the contigs containing transposases are small and often contain only fragments of the transposase itself, however, it is not surprising that few inverted repeats were found associated with transposases, and no firm conclusions can be made.

Phylogenetic diversity

Archaeal and bacterial 16S rRNA clone libraries were constructed (according to JGI standard protocols) from DNA extractions of the same two subsamples used for construction of the metagenomic libraries. No significant differences were detected between the two subsamples, so clones were pooled for further analyses. All 315 archaeal 16S rRNA clones showed high similarity to the previously published sequences for Lost City *Methanosarcinales* (Schrenk *et al.*, 2004). Accession numbers for archaeal clones include FJ791302-FJ791787. The metagenomic data contained a surprising lack of archaeal sequences considering the results from a previous study of similar chimney samples that showed much higher archaeal abundances compared to bacteria (Schrenk *et al.*, 2004). The discrepancy remains unexplained but is probably the result of the highly heterogeneous nature of the carbonate chimney samples (Kelly *et al.*, 2005; Ludwig *et al.*, 2006). Bacterial phylogenetic

diversity is shown in Supplementary Figure 1.

Legend for Supplementary Figure 1

A neighbor-joining tree was constructed with the 658 bacterial 16S rRNA clones, and the phylogenetic affiliations shown in the figure were determined by nearest-neighbor matches provided by Greengenes (greengenes.lbl.gov). Clones with nearest neighbors belonging to the *Thiomicrospira* genus make up 46% of all clones. Accession numbers for bacterial clones include FJ791828-FJ792485.

References

1. Barton HA, Taylor NM, Lubbers BR, Pemberton AC (2006) DNA extraction from low-biomass carbonate rock: An improved method with reduced contamination and the low-biomass contaminant database. *Journal of Microbiological Methods* **66**:21–31.
2. Brazelton WJ, Schrenk MO, Kelley DS, Baross JA (2006) Methane- and sulfur-metabolizing microbial communities dominate the Lost City hydrothermal field ecosystem. *Appl Environ Microbiol* **72**:6257–6270.
3. Kelley DS, Karson JA, Fruh-Green GL, Yoerger DR, Shank TM, Butterfield DA et al. (2005) A serpentinite-hosted ecosystem: the Lost City hydrothermal field. *Science* **307**:1428–1434.
4. Ludwig KA, Kelley DS, Butterfield DA, Nelson BK, Fruh-Green G (2006) Formation and evolution of carbonate chimneys at the Lost City Hydrothermal Field. *Geochim Cosmochim Acta* **70**:3625-3645.
5. Schrenk MO, Kelley DS, Bolton SA, Baross JA (2004) Low archaeal diversity linked to subsurface geochemical processes at the Lost City Hydrothermal Field, Mid-Atlantic Ridge. *Environ Microbiol* **6**:1086-1095.

6. Warburton PE, Giordano J, Cheung F, Gelfand Y, Benson G (2004) Inverted repeat structure of the human genome: the X-chromosome contains a preponderance of large, highly homologous inverted repeats that contain testes genes. *Genome Research* **14**:1861-1869.

Supplementary Figure 1
Brazelton & Baross

