Multiple scales of diversification within natural populations of archaea in hydrothermal chimney biofilms

William J. Brazelton[1*], Mitchell L. Sogin[2], and John A. Baross[1]

[1] School of Oceanography and Center for Astrobiology and Early Evolution, University of Washington

[2] Josephine Bay Paul Center, Marine Biological Laboratory at Woods Hole

[*] corresponding author: Box 357940, School of Oceanography, University of Washington, Seattle, WA 98195. telephone: 206-221-5755. fax: 206-543-6073. braz@u.washington.edu

running title: Microdiversity within a single-species biofilm

Abstract

Corroborative data collected from 16S rRNA clone libraries, intergenic transcribed spacer (ITS) region clone libraries, and 16S rRNA hypervariable region tag pyrosequencing demonstrate microdiversity within single-species archaeal biofilms of the Lost City Hydrothermal Field. Both 16S rRNA clone libraries and pyrosequencing of the V6 hypervariable region show that Lost City Methanosarcinales (LCMS) biofilms are dominated by a single sequence, but the pyrosequencing dataset also reveals the presence of an additional 1654 rare sequences. Clone libraries constructed with DNA spanning the V6 hypervariable region and ITS show that multiple ITS sequences are associated with the same dominant V6 sequence. Furthermore, ITS variability differed among three chimney samples, and the sample with the highest ITS diversity also contained the highest V6 diversity as measured by clone libraries as well as tag pyrosequencing. These results indicate that the extensive microdiversity detected in V6 tag sequences is an underestimate of genetic diversity within the archaeal biofilms.

INTRODUCTION

Biofilms coating carbonate chimneys of the Lost City Hydrothermal Field (Kelley *et al*., 2005) are dominated by a single 16S rRNA phylotype referred to as Lost City *Methanosarcinales* (LCMS; Schrenk *et al.,* 2004; Brazelton *et al*., 2006). Previous studies have shown that >80% of all cells in carbonate chimneys venting 20-90ºC, pH 9-11 fluids hybridize to a fluorescent *in situ* hybridization (FISH) probe specific to LCMS (Schrenk *et al*., 2004). LCMS has resisted laboratory cultivation, but it is presumed to subsist on the high concentrations of hydrogen and/or methane gas venting from the carbonate chimneys (Kelley *et al.*, 2005).

Previous studies have shown ecologically relevant genetic and physiological diversity within natural populations of archaea and bacteria that initially seemed to contain very little genetic diversity based on 16S rRNA sequences. For example, bacterioplankton with >99% similar 16S rRNA sequences can harbor extensive genomic variation (Thompson *et al.,* 2005) and comprise many ecologically distinct strains (Hunt *et al.,* 2008). Variation in the intergenic transcribed spacer (ITS) region, which is less conserved than 16S rRNA, is often a better predictor of genomic and ecological variation. ITS sequence variation delineates cyanobacterial 'ecotypes' that have substantial differences in genomic content (Rocap *et al*., 2003) and physiological differences linked to distinct localizations within water columns (West *et al*., 2001) or microbial mats (Ferris *et al.,* 2003). Environmental sequencing of the ITS region has also proved useful in resolving genetically distinct clusters within uncultivated organisms belonging to the *Thermococcales* group of thermophilic archaea (Huber *et al.*, 2006), the Group I *Crenarchaeota* (Schleper *et al*., 1998, Nicol *et al*., 2006), and the SAR11 group of marine bacteria (Garcia-Martinez & Rodriguez-Valera, 2000).

In this paper we test whether the LCMS phylotype consists of genetically distinct subpopulations by

thoroughly exploring the sequence diversity in the 16S rRNA gene as well as the ITS region, utilizing both Sanger sequencing of clone libraries and tag pyrosequencing of the V6 hypervariable region.

RESULTS AND DISCUSSION

16S rRNA clone library

An archaeal 16S rRNA clone library was constructed (by the DOE Joint Genome Institute) from a single carbonate chimney collected from the main chimney structure at Lost City known as Poseidon (sample LC0424). Sequences were obtained from 486 clones (GenBank accession numbers FJ791302-FJ791787), all of which showed high sequence similarity to the previously published (Schrenk *et al*., 2004) 16S rRNA sequence of LCMS. After screening for length and quality, 200 clone sequences each containing at least 1250 bp were selected for further analysis.

All 200 clones were at least 98.8% similar over the 1253 bp alignment, but 163 unique sequences were detected (Figure 1). Although the evenness of unique sequences is high (Table 1) because the most common sequence was shared by only 36 clones, no other sequence was shared by more than two clones. Most of the variations among sequences were substitutions; insertions and deletions were comparatively rare (Table 1). Similar results are achieved if only the V6 hypervariable region is considered (where the V6 is defined by the primers used for V6 tag pyrosequencing described below). Of the 200 clones, 179 have identical V6 sequences, and the 21 variant clones represent 19 additional sequences.

Comparing the variant sequences to the most common sequence yields a mutation rate of 0.15% for the nearly full-length gene and 0.16% for the V6 region. Because the sequence differences are rare and mostly unique, it is possible that they could be caused by DNA polymerase error. A *Taq* DNA

polymerase error rate of 2.3 x 10^{-5} per base per cycle (Li *et al*., 2006), however, would only contribute 0.046% sequence variation after 20 cycles of amplification (JGI Standard Protocol) during the polymerase chain reaction. Therefore, polymerase error is unlikely to account for all of the diversity observed in our clone libraries.

V6 hypervariable region tag sequences

We obtained 16,260 tag sequences of the V6 hypervariable region of the archaeal 16S rRNA gene from another sample (LC1408) of the same chimney used for the 16S rRNA clone library. More than 91% of these tags were assigned to the family *Methanosarcinaceae* by GAST (Huse *et al*. 2008) and showed an extremely uneven abundance distribution. Of the 14,869 *Methanosarcinaceae* tags, 75% were identical to the corresponding V6 region of 179 of the 200 full-length 16S rRNA clones. The remaining 25% (3695 tags) comprised 622 different sequences clustering into 235 operational taxonomic units (OTUs) at 97% sequence similarity (Figure 1).

The second most common V6 tag sequence (representing ~5% of all tags) differs from the dominant sequence by lacking the final GAG at the 3' end. The deletion was not caused by premature truncation of pyrosequencing extension because in each case the distal primer was accurately sequenced. The sequence GAGAG at the 3' end of the V6 region is highly conserved in archaeal rRNA, but 0.8% of archaeal sequences, including many methanogens, in the RefHVR_v6 database (http://vamps.mbl.edu) lack the final GAG (S. Huse, personal communication). Because this database is derived from traditional Sanger sequencing of clones, the GAG deletion in our data is unlikely to be caused by pyrosequencing error. The lack of this deletion in our clone libraries, however, is puzzling.

Two additional samples (LC1404 and LC1443) collected from a different chimney showed very similar

distributions, being dominated by the same sequence with a large diversity of very rare sequences (Figure 2a). The temperature and fluid chemistry at this chimney was similar to the chimney from which sample LC1408 was collected, although samples LC1404 and LC1443 had much higher cell densities (Table S1). The three samples together contained 72,577 tags assigned to the family *Methanosarcinaceae* representing 1654 different sequences and 536 operational taxonomic units at 97% sequence similarity. The extreme rarity of the diverse sequences raises questions regarding the effect of pyrosequencing error. Tag abundances decreased substantially with increasing distance from the most dominant sequence, a trend that is consistent with the expected effect of random sequencing error from one dominant template. Some sequences, however, appeared much more frequently than others with the same number of substitutions and indels (prominent peaks in Figure 2a), so these may represent genuine diversity above a background error rate.

Three additional features of our data argue against a significant contribution from pyrosequencing error to the observed diversity. Firstly, the amount of sequence variation was too high to be generated by pyrosequencing error alone. Comparing all variant V6 tag sequences to the one dominant sequence yielded mutation rates of 0.55-0.71% for the three samples (Table 1), while the error rate associated with the pyrosequencing technique and quality-filtering procedure used in this study should not exceed 0.16% (Huse *et al*. 2007). Most of the mutations were insertions and deletions, whose pyrosequencing-associated rates can vary depending on the template sequence, but the substitution rates (0.15-0.20%) were also much higher than the maximum expected from pyrosequencing error (0.03%, Huse *et al*., 2007).

Secondly, many of the bases with the highest substitution rate in the V6 tags were also the most variable bases in the clone library sequences. Positions outlined with a black box in Figure 2b were the site of at

least two substitutions in clone libraries (including the full-length library described above and the three V6-ITS libraries described below). All of these positions also had greater than average substitution rates in the V6 tag dataset (indicated by orange and red shading in Figure 2b). It is highly unlikely that error introduced by both Sanger sequencing of clone libraries and tag pyrosequencing could cause this correspondence in site-specific substitution rates. Furthermore, the transition/transversion ratios associated with substitutions in the V6 tags were very similar to that found in the full-length clone libraries (Table 1).

Finally, pyrosequencing error alone cannot account for the high similarity between the V6 tag distributions of the two samples from the same chimney (LC1404 and LC1443) compared to that of LC1408, which was collected from a different chimney. Although all three samples are very similar in their complement of abundant sequences (Figure 2a), only a small proportion of the total sequences were shared among samples (Jaccard similarities of 22-26%) due to the large number of rare sequences. Interestingly, LC1404 and LC1443 both contained fewer unique sequences than sample LC1408 (Table 1), and the Bray-Curtis community similarity between the two samples from the same chimney was higher than the community similarity between samples from different chimneys (see Supplementary Information for details). Although this comparison involves only three samples and thus is not strong statistical evidence, it is suggestive that small differences in rare V6 tag sequences reflect environmental variation.

V6-ITS clone libraries

To directly compare the diversity of the V6 region within the LCMS biofilms to a marker known to be more variable in other organisms (Rocap *et al*., 2002), we constructed clone libraries of ~1071 bp DNA fragments spanning the 3' end of the 16S rRNA gene including the V6 hypervariable region and the

intergenic transcribed spacer (ITS) region between the 16S and 23S rRNA genes. Approximately 150-200 clones were sequenced from each of the same three carbonate chimney samples used for V6 tag pyrosequencing. As expected, nearly all 197 V6-ITS clones from sample LC1408 shared the same V6 sequence that dominated the pyrosequencing dataset. Only 7 clones had variant V6 sequences (Figure 1), and each of these were unique and the result of transitions. V6-ITS clones from samples LC1404 and LC1443 were also dominated by a single V6 sequence with only a few variants mostly caused by transitions. Sequencing error cannot be discounted as a source for such a small number of V6 variants.

Although the ITS regions of all 516 V6-ITS clones were of nearly identical size (360 bp) and >98% similar to each other, 104 different sequences were detected among the three samples. For samples LC1408 and LC1443 the mutation rate within the ITS region (0.24% and 0.16%) was higher than the mutation rate within the V6 hypervariable region (0.06% and 0.11%), but in sample LC1404 the ITS region exhibited even less variation (0.02%) than in the V6 (0.04%) (Table 1). The variation in the V6 regions of the V6-ITS clones were substantially lower than that observed for the V6 region of the 16S rRNA clones, even though 34-38 cycles were required for amplification of the V6-ITS clones, compared to 20 cycles for the 16S rRNA clones. We conclude that error introduced during amplification and cloning does not appear to greatly affect the observed trends in ITS sequence variation.

The ITS region of LCMS encodes an Ala-tRNA and shows sequence homology with the ITS regions of several methanogens (Figure S1). Sequence variations were most commonly associated with two predicted stem-loop structures in the region upstream of the tRNA gene (Figure 3a). The five most common variations were present in 10-30 clones per library; most positions were variable in only 0-1 clones (Figure 3a). The highly non-random distribution of sequence variation along the length of the ITS argues strongly against a large contribution of variation from sequencing error.

Sample LC1408 contained 47 different ITS sequences (Figure 1), more than LC1404 (23 sequences) or LC1443 (43 sequences). The evenness of sample LC1408 was higher than that of the other samples (Table 1), as the most common sequence comprised just 37.8% of all clones. The greater evenness in LC1408 ITS sequences may be due, in part, to the higher number of cycles required for sufficient PCR amplification of this sample, but this effect is not expected to be large for reasons described above and in the Supplementary Information. Furthermore, it is intriguing that the ITS clone libraries as well as the V6 tag datasets showed the highest diversity and evenness in sample LC1408 and the least diversity and evenness in sample LC1404 (Table 1). This correspondence between genetic markers and sequencing technologies supports the observed trends as reliable indicators of biological diversity and not artifacts of the methodology.

The ITS region appears to reveal a scale of diversity that is not reflected in 16S rRNA sequences. Compared to the 16S rRNA clone libraries and V6 tag pyrosequencing datasets, the ITS clones showed a more even abundance distribution of sequences (as shown in the higher evenness values in Table 1 and in Figure S2). Of all 516 V6-ITS clones, 231 contained ITS sequence variations, and eight of these variants occurred more than twice. In contrast, none of the 16S rRNA variants occurred more than twice, so it is possible that many of these variants were generated by sequencing error. Of the 231 clones with variant ITS sequences, 221 clones had identical V6 sequences. The 10 exceptions involved 9 different V6 sequences and 6 different ITS sequences. Thus nearly all of the observed ITS variation is associated with the same dominant V6 sequence, and it is likely that a tag pyrosequencing study of the LCMS biofilm with primers targeting the ITS region would reveal even more microdiversity than the thousands of V6 sequence types found in this study.

Multiple studies have shown that large genomic differences are possible among organisms with only small variations in 16S rRNA sequence (Beja *et al*., 2002; Welch *et al*., 2002; Rocap *et al*., 2003; Thompson *et al*., 2005), but further work is necessary to determine if the microdiversity reported in our study is associated with larger scale genomic variations leading to important physiological and ecological consequences. The V6 tag dataset alone does not compel rejection of a null hypothesis of ecologically-neutral genetic drift within a clonal population because it is possible for the many extremely rare V6 tag sequences to reflect 'background' mutations not yet affected by selection and speciation. The highly non-random nature of the ITS variation, however, provides stronger evidence for ecologically relevant diversity. The markedly different distributions of ITS genotypes among chimney samples (Figure 3b) may be an indication that the biofilm community contains several distinct subpopulations represented by different ITS genotypes. Determining whether these subpopulations represent physiologically and ecologically distinct units (*i.e.* ecotypes or species) will require further genomic and physiological experiments. In particular, these experiments should test the hypothesis that differentiation within this one group of archaea is the result of subpopulations colonizing multiple niches within the chimney to maximize utilization of resources that are unavailable to other organisms due to the extreme conditions of Lost City chimneys (Kelley *et al*., 2005; Brazelton *et al*., 2006).

The detection of so many rare V6 sequences was only technically feasible in this study due to the extremely low diversity of the Lost City carbonate chimneys. As sequencing technology continues to improve in sensitivity, fidelity, and read length, measurements of even finer scale microdiversity and comparisons of variation across multiple genomic markers will become possible for systems with greater diversity. This near-future technology could be used to test whether the rare microdiversity reported here is a natural feature of microbial populations or an unusual characteristic unique to this extremophilic archaeal community.

Acknowledgements

We thank Deborah Kelley, chief scientist, and the scientific parties of the 2003 and 2005 Lost City Expeditions as well as the crews of the R/V *Atlantis,* DSV *Alvin,* R/V *Ronald H. Brown*, and DSV *Hercules*. We also appreciate technical assistance and helpful discussions with Ekaterina Andreishcheva, Sue Huse, and Julie Huber. This research was supported by the W.M. Keck Foundation to MLS, the NASA Astrobiology Institute through the Carnegie Institution for Science to JAB and through the MBL to MLS.

References

Beja, O., Koonin, E.V., Aravind, L., Taylor, L.T., Seitz, H., Stein, J.L. *et al*. (2002) Comparative genomic analysis of archaeal genotypic variants in a single population and in two different oceanic provinces. *Appl Environ Microbiol* **68**: 335-345.

Brazelton, W.J., Schrenk, M.O., Kelley, D.S., and Baross, J.A. (2006) Methane- and sulfur-metabolizing microbial communities dominate the Lost City hydrothermal field ecosystem. *Appl Environ Microbiol* **72**: 6257–6270.

Ferris, M.J., Kuhl, M., Wieland, A., and Ward, D.M. (2003) Cyanobacterial ecotypes in different optical microenvironments of a 68 degrees C hot spring mat community revealed by 16S-23S rRNA internal transcribed spacer region variation. *Appl Environ Microbiol* **69**: 2893–2898.

Garcia-Martinez, J., and Rodriguez-Valera, F. (2000) Microdiversity of uncultured marine prokaryotes: the SAR11 cluster and the marine *Archaea* of Group I. *Mol Ecol* **9**: 935–948.

Huber, J.A., Butterfield, D.A., and Baross, J.A. (2006) Diversity and distribution of subseafloor *Thermococcales* populations in diffuse hydrothermal vents at an active deep-sea volcano in the northeast Pacific Ocean. *J Geophys Res Biogeosciences* **111**: G04016.

Hunt, D.E., David, L.A., Gevers, D., Preheim, S.P., Alm, E.J., and Polz, M.F. (2008) Resource partitioning and sympatric differentiation among closely related bacterioplankton. *Science* **320**: 1081-1085.

Huse, S.M., Huber, J.A., Morrison, H.G., Sogin, M.L., and Welch, D.M. (2007) Accuracy and quality of massively parallel DNA pyrosequencing. *Genome Biol* **8**: R143.

Kelley, D.S., Karson, J.A., Fruh-Green, G.L., Yoerger, D.R., Shank, T.M., Butterfield, D.A. *et al*. (2005) A serpentinite-hosted ecosystem: the Lost City hydrothermal field. *Science* **307**: 1428–1434.

Li, M., Diehl, F., Dressman, D., Vogelstein, B., and Kinzler, K.W. (2006) BEAMing up for detection and quantification of rare sequence variants. *Nat Methods* **3**: 95–97.

Lowe, T.M. and Eddy, S.R. (1997) tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res* **25:** 955-964.

Markham, N. R. and Zuker, M. (2005) DINAMelt web server for nucleic acid melting prediction. *Nucleic Acids Res* **33:** W577-W581.

Nicol, G.W., Tscherko, D., Chang, L., Hammesfahr, U., and Prosser, J.I. (2006) Crenarchaeal community assembly and microdiversity in developing soils at two sites associated with deglaciation. *Environ Microbiol* **8**: 1382–1393.

Rocap, G., Distel, D.L., Waterbury, J.B., and Chisholm, S.W. (2002) Resolution of *Prochlorococcus* and *Synechococcus* ecotypes by using 16S-23S ribosomal DNA internal transcribed spacer sequences. *Appl Environ Microbiol* **68**: 1180-1191.

Rocap, G., Larimer, F.W., Lamerdin, J., Malfatti, S., Chain, P., Ahlgren, N.A. *et al.* (2003) Genome divergence in two *Prochlorococcus* ecotypes reflects oceanic niche differentiation. *Nature* **424**: 1042-1047.

Schleper, C., DeLong, E.F., Preston, C.M., Feldman, R.A., Wu, K.Y., and Swanson, R.V. (1998) Genomic analysis reveals chromosomal variation in natural populations of the uncultured

psychrophilic archaeon *Cenarchaeum symbiosum*. *J Bacteriol* **180**: 5003-5009.

Schrenk, M.O., Kelley, D.S., Bolton, S.A., and Baross, J.A. (2004) Low archaeal diversity linked to subseafloor geochemical processes at the Lost City Hydrothermal Field, Mid-Atlantic Ridge. *Environ Microbiol* **6**: 1086-1095.

Thompson, J.R., Pacocha, S., Pharino, C., Klepac-Ceraj, V., Hunt, D.E., Benoit, J. *et al.* (2005) Genotypic diversity within a natural coastal bacterioplankton population. *Science* **307**: 1311–1313.

Welch, R.A., Burland, V., Plunkett, G.r., Redford, P., Roesch, P., Rasko, D. *et al.* (2002) Extensive mosaic structure revealed by the complete genome sequence of uropathogenic *Escherichia coli*. *Proc Natl Acad Sci USA* **99**: 17020–17024.

West, N.J., Schonhuber, W.A., Fuller, N.J., Amann, R.I., Rippka, R., Post, A.F., and Scanlan, D.J. (2001) Closely related *Prochlorococcus* genotypes show remarkably different depth distributions in two oceanic regions as revealed by in situ hybridization using 16S rRNA-targeted oligonucleotides. *Microbiology* **147**: 1731–1744.

Figure Captions

Figure 1. Comparison of tag pyrosequencing and clone library data from the same carbonate chimney. All sequences from 200 nearly full-length 16S rRNA clones obtained from sample LC0424 were more than 97% similar to each other (A). Collection of 14,869 tag pyrosequences of the V6 hypervariable region from a different sample (LC1408) of the same chimney revealed much greater diversity (B). A clone library constructed with DNA from sample LC1408 spanning the V6 hypervariable region and the intergenic transcribed spacer (ITS) region showed more diversity in the ITS region (C).

Figure 2. Tag pyrosequences of the V6 hypervariable region reveal a wide range of highly similar, rare

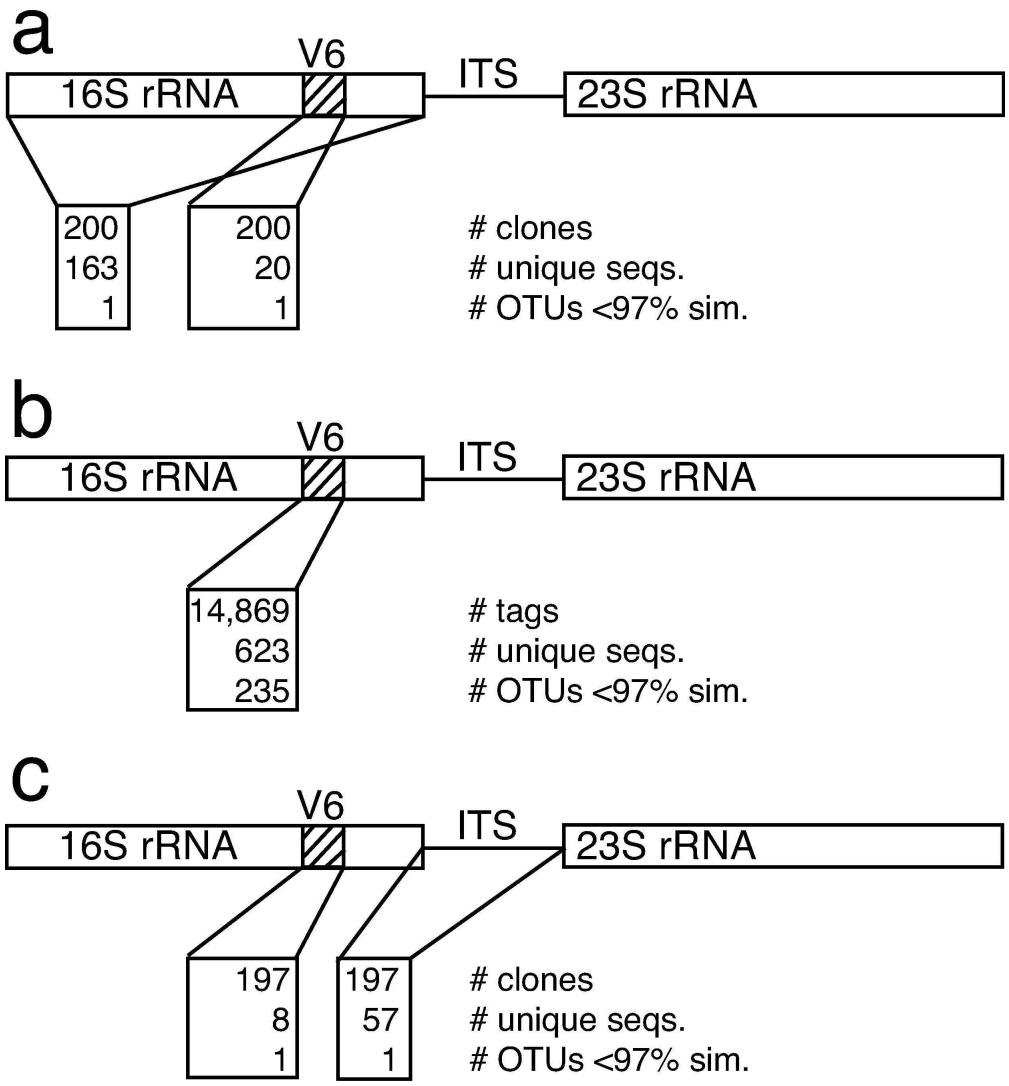
sequences. The relative abundance distribution (A) of 1654 different V6 sequences among the three samples (LC1408, LC1404, LC1133) shows the extreme dominance of one sequence and the diversity of rare sequences. Differences (and similarities) among samples are more easily seen when the one dominant sequence and sequences observed only once in the total dataset ('singletons') are omitted to show only the 483 most common variants (inset). Sequences are sorted along the X axis by distance to the dominant sequence. The predicted secondary structure of the V6 region (B) was slightly modified from the archaeal structure on the Comparative RNA Web Site (http://www.rna.ccbb.utexas.edu) to fit the dominant sequence. Those bases that are variable in at least two clones among all the clone libraries in this study (in boxes) are among the most highly variable (orange and red shading) in the V6 tag dataset as well. All V6 sequence data is available at the VAMPS database, http://vamps.mbl.edu., under dataset name ICM_LCY_Av6 and in the NCBI Short Read Archive under submission number SRP000912.

Figure 3. The variability of specific bases within the ITS region of Lost City Methanosarcinales differs among chimney samples. The secondary structure of the ITS (A) was predicted by UNAFOLD (Markham and Zuker, 2005) and modified to match the tRNA structure predicted by tRNAscan-SE (Lowe and Eddy, 1997). Bases are color-coded to indicate the number of clones (out of 517 total) that differed from the dominant sequence at that position, and the five most variable positions are numbered and compared among samples in (B). The most frequent variation, a C to T transition, occurred in 22 clones in sample LC1408, in 0 clones in LC1404, and 21 clones in LC1443. Accession numbers for clones including the V6 and ITS regions include GQ272945-GQ273460.

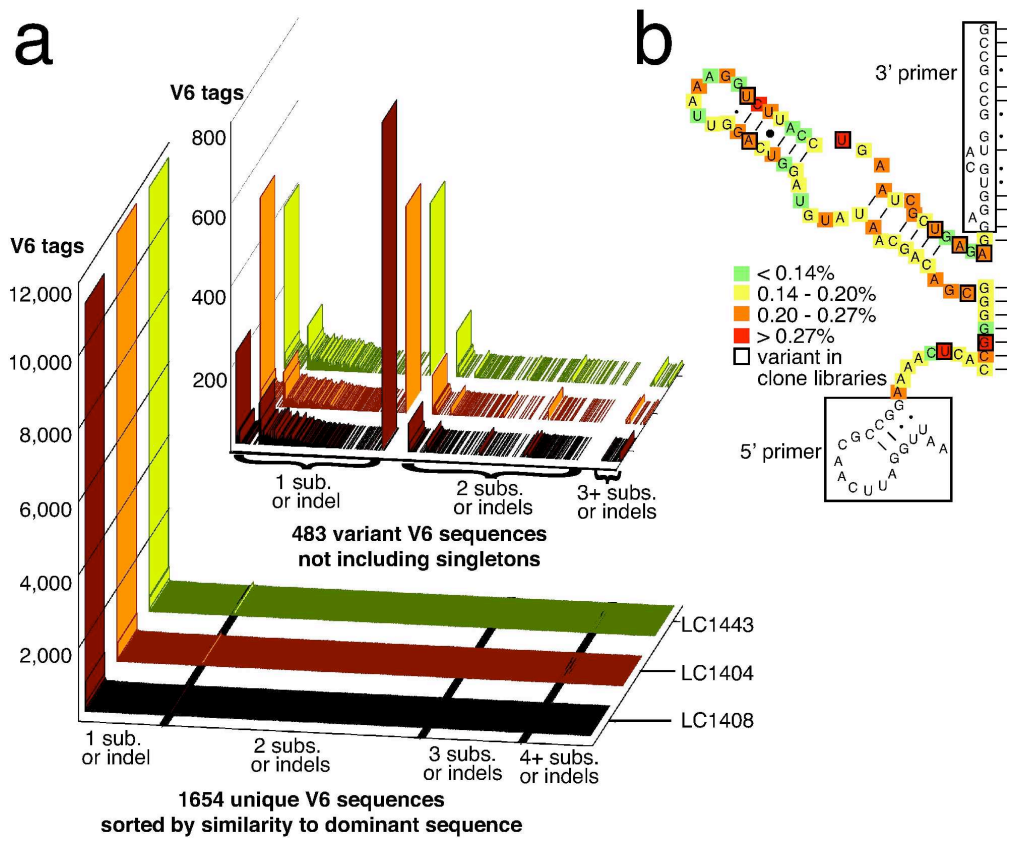
	sample[1]	clones or tags	length (bp)	unique sequences[2]	total mutation rate	insertion rate	deletion rate	substitution rate	Ti/Tv ratio[3]	evenness[4]
16S rRNA clones	LC0424 (full length)	200	1253	163	0.15%	0.01%	0.01%	0.13%	269/45	0.91 ± 0.04
	LC0424 (V6 region)	200	65	20	0.16%	0%	0%	0.16%	18/3	0.21 ± 0.07
V6 pyrosequencing tags	LC1408	14,869	65	623	0.71%	0.12%	0.38%	0.20%	6.10	0.267 ± 0.007
	LC1404	32,340	65	472	0.55%	0.16%	0.24%	0.16%	4.47	0.236 ± 0.007
	LC1443	25,368	65	487	0.58%	0.16%	0.22%	0.20%	8.78	0.247 ± 0.007
V6-ITS clones (V6 region)	LC1408	196	65	8	0.06%	0%	0%	0.06%	8/0	0.11 ± 0.06
	LC1404	132	65	4	0.04%	0%	0%	0.04%	6/0	0.09 ± 0.09
	LC1443	189	65	12	0.11%	0%	0%	0.11%	12/1	0.16 ± 0.07
V6-ITS clones (ITS region)	LC1408	196	360	57	0.24%	0.03%	0.06%	0.16%	103/7	0.65 ± 0.06
	LC1404	132	360	32	0.02%	0.0005%	0.0005%	0.02%	41/0	0.44 ± 0.10
	LC1443	189	360	51	0.16%	0%	0.0001%	0.16%	99/9	0.56 ± 0.07

[1] Full sample names: LC0424, H03_072705_R0424; LC1408, 3881-1408; LC1404, 3869-1404; LC1443, 3869-1443.
[2] Unique sequences for V6 pyrosequencing tags were calculated after normalizing samples down to 14,869 total tags.
[3] Ti/Tv ratio is shown as numbers of transitions/transversions for clones and as decimal fraction for tags.
[4] Evenness derived from the Shannon-Weaver index and its standard deviation (calculated by DOTUR, Schloss 2005).

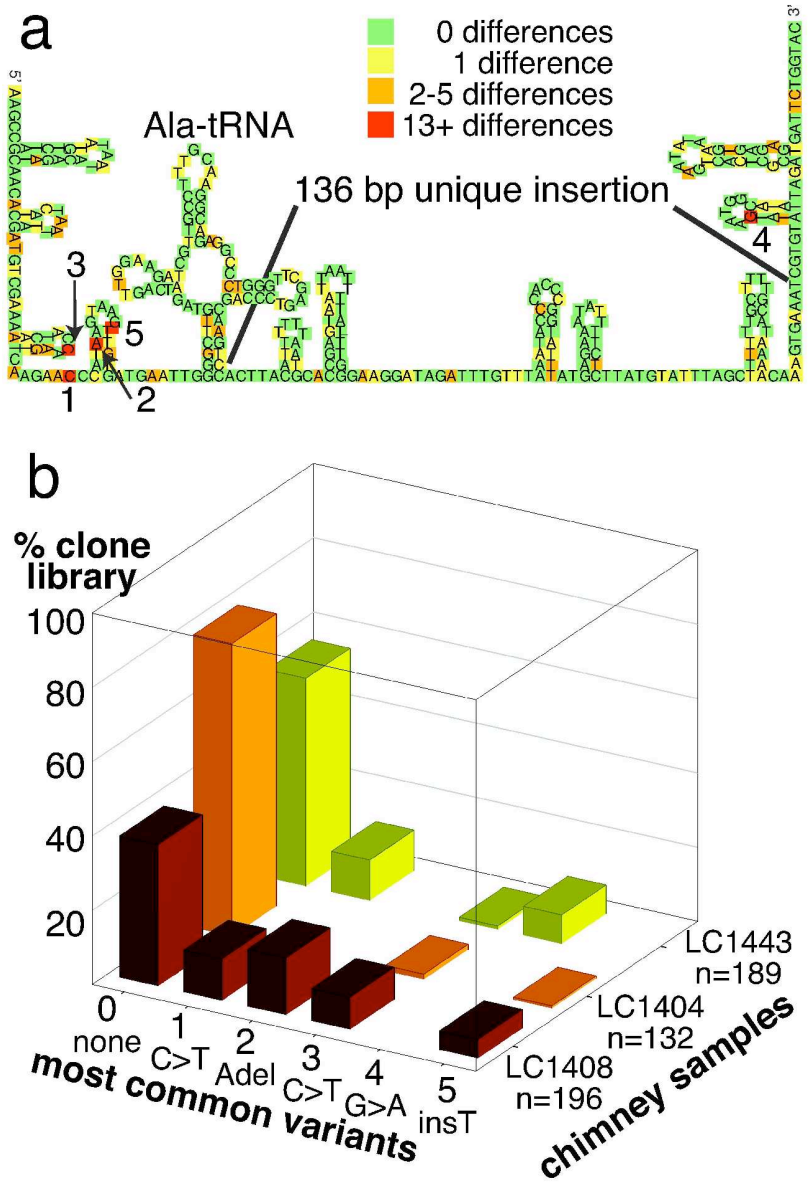
Table 1. Diversity comparison of the 16S rRNA, V6 hypervariable region, and ITS region among Lost City carbonate chimney samples.



Comparison of tag pyrosequencing and clone library data from the same carbonate chimney. All sequences from 200 nearly full-length 16S rRNA clones obtained from sample LC0424 were more than 97% similar to each other (A). Collection of 14,869 tag pyrosequences of the V6 hypervariable region from a different sample (LC1408) of the same chimney revealed much greater diversity (B). A clone library constructed with DNA from sample LC1408 spanning the V6 hypervariable region and the intergenic transcribed spacer (ITS) region showed more diversity in the ITS region (C).
105x120mm (600 x 600 DPI)



Tag pyrosequences of the V6 hypervariable region reveal a wide range of highly similar, rare sequences. The relative abundance distribution (A) of 1654 different V6 sequences among the three samples (LC1408, LC1404, LC1133) shows the extreme dominance of one sequence and the diversity of rare sequences. Differences (and similarities) among samples are more easily seen when the one dominant sequence and sequences observed only once in the total dataset ('singletons') are omitted to show only the 483 most common variants (inset). Sequences are sorted along the X axis by distance to the dominant sequence. The predicted secondary structure of the V6 region (B) was slightly modified from the archaeal structure on the Comparative RNA Web Site (http://www.rna.ccbb.utexas.edu) to fit the dominant sequence. Those bases that are variable in at least two clones among all the clone libraries in this study (in boxes) are among the most highly variable (orange and red shading) in the V6 tag dataset as well. All V6 sequence data is available at the VAMPS database, http://vamps.mbl.edu., under dataset name ICM_LCY_Av6 and in the NCBI Short Read Archive under submission number SRP000912.
167x141mm (600 x 600 DPI)



The variability of specific bases within the ITS region of Lost City Methanosarcinales differs among chimney samples. The secondary structure of the ITS (A) was predicted by UNAFOLD (Markham and Zuker, 2005) and modified to match the tRNA structure predicted by tRNAscan-SE (Lowe and Eddy, 1997). Bases are color-coded to indicate the number of clones (out of 517 total) that differed from the dominant sequence at that position, and the five most variable positions are numbered and compared among samples in (B). The most frequent variation, a C to T transition, occurred in 22 clones in sample LC1408, in 0 clones in LC1404, and 21 clones in LC1443. Accession numbers for clones including the V6 and ITS regions include GQ272945-GQ273460.
109x163mm (600 x 600 DPI)

Supplementary Information

Multiple scales of diversification within natural populations of archaea in hydrothermal chimney biofilms

William J. Brazelton[1*], Mitchell L. Sogin[2], and John A. Baross[1]

[1] School of Oceanography and Center for Astrobiology and Early Evolution, University of Washington

[2] Josephine Bay Paul Center, Marine Biological Laboratory at Woods Hole

Sample descriptions

Carbonate chimney samples LC1408 (full sample name 3881-1408), LC1404 (3869-1404), and LC1443 (3869-1443) were collected from the Lost City Hydrothermal Field (LCHF, depth, ~735 m; latitude, 30.12; longitude, -42.12) with DSV Alvin during cruise AT07-34 aboard the R/V Atlantis in April/May 2003 (http://www.lostcity.washington.edu). Sample LC0424 (H03_072705_R0424) was collected by DSV *Hercules* during the 2005 Lost City Expedition aboard the R/V *Ronald H. Brown*. LC0424 and LC1408 were collected from a site known as Marker 3 or 'Poseidon,' a 60 m tall edifice emitting fluids at temperatures ranging from 55-88°C (Kelley *et al*. 2005). LC1408 minerals appeared bright white in color, very friable, and not lithified. Samples LC1404 and LC1443 are from a structure named Marker C, a ~50 cm wide flange structure with several small (centimeters tall) chimneys growing on the top of the flange. LC1404 was collected from the front of the flange, and LC1443 was a small spire collected from the top. Both samples were cream white with a reddish discoloration that remains unexplained (Ludwig et al., 2006). Additional published characteristics of the samples are summarized in Table S1.

Shipboard, subsamples of chimney material were frozen immediately at -80°C and remained frozen until onshore analysis. DNA was extracted from carbonate chimney samples 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 approximately 15 g of carbonate minerals were processed for each sample. 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:choloroform: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.

Construction and sequencing of clone libraries

Two 16S rRNA clone libraries including a total of 486 clones (GenBank accession numbers FJ791302-FJ791787) from sample LC0424 were constructed by the DOE Joint Genome Institute according to the standard protocol published on their website: http://my.jgi.doe.gov/general/index.html. The V6-ITS clone libraries including a total of 516 clones from three samples (accession numbers GQ272945-GQ273460) were constructed from amplicons covering the 16S rRNA V6 region downstream through the intergenic transcribed spacer (ITS) region to the 23S rRNA. PCR amplification was conducted

according to the protocol of Huber *et al.* (2006). The forward primer (886F-LCMS: GAAGTACGGCCGCAAGGC) targets a region just upstream of the Lost City Methanosarcinales V6 region, and the reverse primer (58Ra: GCTTATCGCAGCTTGSCACG) targets the 5' end of the archaeal 23S rRNA gene (Huber *et al*. 2006). V6-ITS amplicons were reconditioned using the protocol of Thompson *et al.* (2002) and cloned using the TOPO-TA cloning kit (Invitrogen) according to the manufacturer's instructions. Cloned inserts were sequenced at the University of Washington High-Throughput Genomics Unit (www.htseq.org) with sequencing primers described by Huber *et al.* (2006). Because of inhibitors that could not be removed from the DNA preparations, PCR amplification of V6-ITS clones required 34-38 cycles of PCR amplification. It is possible that the higher evenness in LC1408 (Table 1) resulted from the higher number of cycles (38) used during PCR amplification of this sample compared to other two samples, which required only 34 cycles. The higher diversity in LC1408 and LC1443 compared to LC1404, however, is unlikely to be affected by cycle number or polymerase error, because only 34 cycles were used for both LC1443 and LC1404 and because of the high mutation rates in these libraries compared to that expected from polymerase and sequencing error, as described in the main text. More amplification cycles may have been required for sample LC1408 because it contained 100x lower archaeal density than the other two samples (Table X?) even though efforts were made to equalize DNA template concentrations. All alignments were calculated with MUSCLE (Edgar *et al*., 2004).

Analysis of tag pyrosequences

Protocols for construction and sequencing of V6 amplicon libraries have been described previously (Sogin *et al*., 2006; Huber *et al.*, 2007). Tag sequences were screened for quality as recommend by Huse *et al.* (2007). Sequences assigned to the family *Methanosarcinaceae* by GAST (Huse *et al*., 2008) were aligned with MUSCLE (Edgar *et al.*, 2004). Distance matrices were calculated with quickdist as described by Sogin *et al.* (2006) except that terminal gaps were penalized in our study because we

inspected the 3' ends to confirm that primers were acurately trimmed and that the most common 3' deletions were not the result of incomplete sequences. Evenness values were derived from the Shannon-Weaver index as calculated by DOTUR (Schloss *et al*., 2005), and 97% sequence similarity OTUs were calculated with DOTUR. To normalize relative abundances of each sequence among samples, tags were randomly resampled down to the sample with the fewest tags (LC1408: 14,869 tags) using Daisy-Chopper (available at http://www.genomics.ceh.ac.uk/GeneSwytch/Tools.html).

Community similarities among samples

The abundance distributions of tag sequences in the three samples were highly similar, though sample LC1404 is more similar to LC1443 (94% Bray-Curtis similarity), which was sampled ~20 cm away on the same chimney, than to LC1408 (90% Bray-Curtis similarity), which was collected from a different chimney. After removing the one dominant sequence (because the Bray-Curtis index is weighted toward dominant members) and sequences occurring only once in one sample (to decrease the number of heavily undersampled sequences), the abundance distributions of the 483 remaining sequences (Fig. 2a) yielded a greater Bray-Curtis similarity between samples from the same chimney (LC1404 and LC1443, 79%) than between samples from different chimneys, (70-71%). If only very rare sequences (represented by fewer than 10 tags in each sample after normalization) were considered in the similarity calculation, the same trend was observed: LC1404 and LC1443 were 46% similar but only 35-38% similar to LC1408 according to the Bray-Curtis index. Therefore, the abundances of dominant as well as rare sequences are more similar in samples from the same chimney than in samples from different chimneys.

References

Barton, H.A., Taylor, N.M., Lubbers, B.R., and Pemberton, A.C. (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.

Bradley, A.S., Hayes, J.M., Summons, R.E. (2009) Extraordinary ^{13}C enrichment of diether lipids at the Lost City Hydrothermal Field indicates a carbon-limited ecosystem. *Geochim. Cosmochim. Acta* 73:102–118.

Brazelton, W.J., Schrenk, M.O., Kelley, D.S., and Baross, J.A. (2006) Methane- and sulfur-metabolizing microbial communities dominate the Lost City hydrothermal field ecosystem. *Appl Environ Microbiol* **72**: 6257–6270.

Edgar, R.C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32**:1792-7.

Huber, J.A., Butterfield, D.A., and Baross, J.A. (2006) Diversity and distribution of subseafloor *Thermococcales* populations in diffuse hydrothermal vents at an active deep-sea volcano in the northeast Pacific Ocean. *J Geophys Res Biogeosciences* **111**: G04016.

Huber, J.A., Welch, D.B.M., Morrison, H.G., Huse, S.M., Neal, P.R., Butterfield, D.A., and Sogin, M.L. (2007) Microbial population structures in the deep marine biosphere. *Science* **5**:97-100.

Huse, S.M., Huber, J.A., Morrison, H.G., Sogin, M.L., and Welch, D.M. (2007) Accuracy and quality of massively parallel DNA pyrosequencing. *Genome Biol.* **8**:R143.

Huse, S.M., Huber, J.A., Morrison, H.G., Sogin, M.L., and Welch, D.M. (2008) Exploring microbial diversity and taxonomy using SSU rRNA hypervariable tag sequencing. *PLoS Genet* **4**:e1000255.

Kelley, D.S., Karson, J.A., Fruh-Green, G.L., Yoerger, D.R., Shank, T.M., Butterfield, D.A. *et al*. (2005) A serpentinite-hosted ecosystem: the Lost City hydrothermal field. *Science* **307**: 1428–1434.

Ludwig, K.A., Kelley, D.S., Butterfield, D.A., Nelson, B.K., Fruh-Green, G. (2006) Formation and

evolution of carbonate chimneys at the Lost City Hydrothermal Field. *Geochim Cosmochim Acta* **70**:3625–3645.

Proskurowski, G., Lilley, M.D., Kelley, D.S., Olson, E.J. (2006) Low temperature volatile production at the Lost City Hydrothermal Field, evidence from a hydrogen stable isotope geothermometer. *Chem Geol* **229**:331-343.

Proskurowski, G. *et al*. (2008) Abiogenic hydrocarbon production at lost city hydrothermal field. *Science* **319**:604-7.

Schloss, P.D. and Handelsman, J. (2005) Introducing DOTUR, a computer program for defining operational taxonomic units and estimating species richness. *Appl Environ Microbiol* **71**:1501-6.

Schrenk, M.O., Kelley, D.S., Bolton, S.A., Baross, J.A. (2004) Low archaeal diversity linked to subseafloor geochemical processes at the Lost City Hydrothermal Field, Mid-Atlantic Ridge. *Environ Microbiol* 6:1086–1095.

Sogin, M.L., Morrison, H.G., Huber, J.A., Welch, D.M., Huse, S.M., Neal, P.R., Arrieta, J.M., and Herndl, G.J. (2006) Microbial diversity in the deep sea and the underexplored "rare biosphere". *Proc Natl Acad Sci USA* **103**:12115-20.

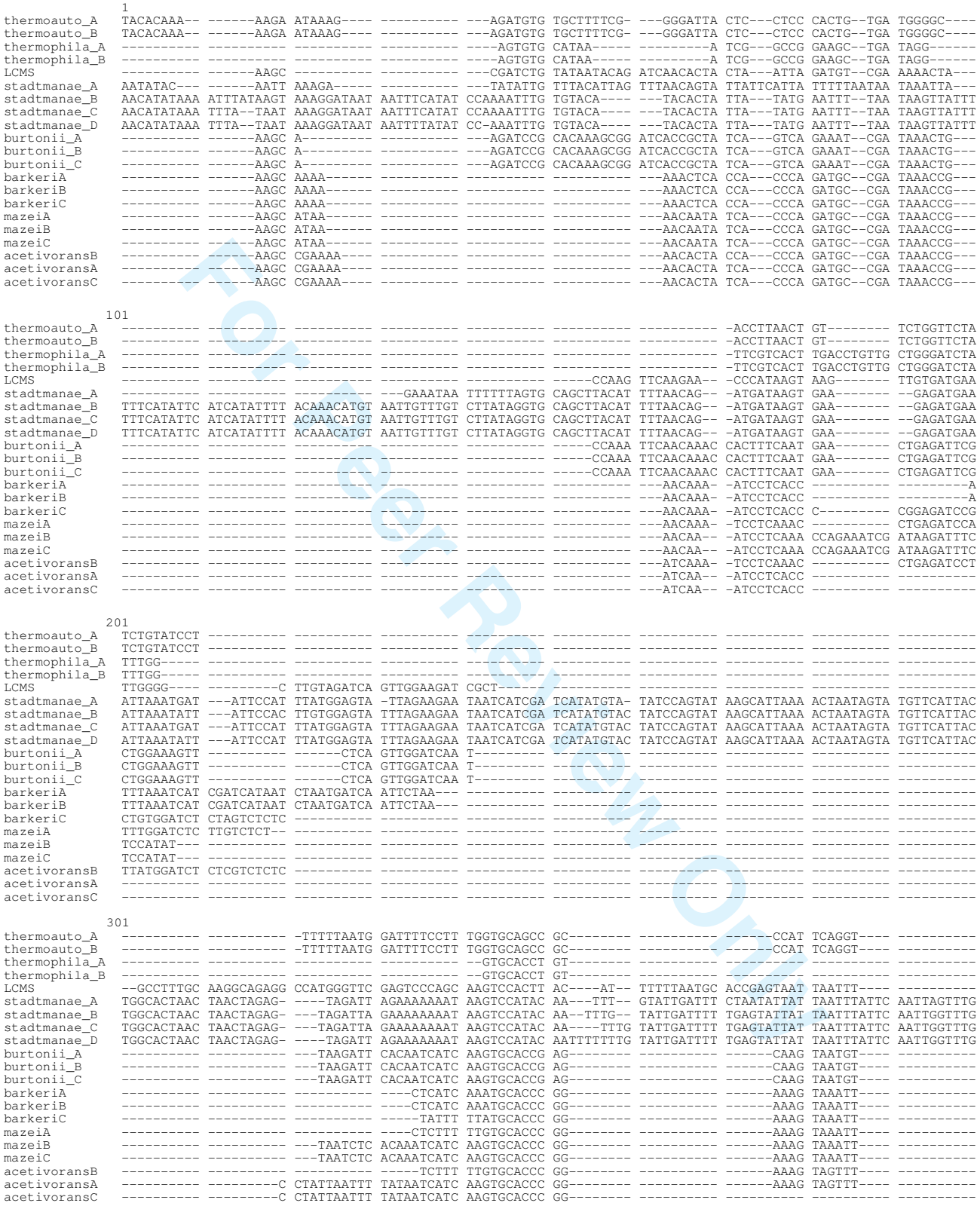
Thompson, J. R., Marcelino, L. A., and Polz, M.F. (2002) Heteroduplexes in mixed-template amplifications: formation, consequence and elimination by 'reconditioning PCR'. *Nucleic Acids Res* **30:**2083-2088.

Chimney Sample	Chimney Location	Max fluid temp (ºC)	Max fluid H_2 (mmol kg^{-1})	Max fluid CH_4 (mmol kg^{-1})	Cells g^{-1} dry weight[a]	Archaea[b]	Bacteria[b]	LCMS[b]	Total organic carbon (%)	$\text{‰}^{13}C_{toc}$ (‰ vs. VPDB)
LC1408	Marker 3	88	13.26	1.55	2.0×10^{-8}	25%	14%	18%	n.d.	n.d.
LC1404	Marker C	70	14.38	1.98	1200×10^{-8}	41%	8%	32%	0.20	-7.8
LC1443	Marker C	70	14.38	1.98	1600×10^{-8}	38%	10%	21%	n.d.	n.d.

[a] Determined by DAPI-staining
[b] Percentage of DAPI-stained cells detected by FISH probe specific to each group

Table S1. Previously published characteristics of the three carbonate chimney samples from which V6 tags and V6-ITS clone libraries were sequenced. Fluid temperatures and concentrations of H_2 and CH_4 are maximum values reported by Proskurowski *et al.* (2006 & 2008). Cell densities and proportions of phylogenetic groups are from Schrenk et al. (2004) and M. Schrenk (doctoral dissertation, 2005). Organic carbon concentrations and isotopic measurements are from Bradley et al. (2009). Fluid temperature and chemistry are identical for samples LC1404 and LC1443 because these carbonate samples were collected from the same chimney.

Figure S1 with caption below:



```
             401
thermoauto_A   -------TGT GGG------- ---ATGTGGT GGTGAAGTTG GAATGATATG G--------- ---------- ---------- ---------- ----------
thermoauto_B   -------TGT GGG------- ---ATGTGGT GGTGAAGTTG GAATGATATG G--------- ---------- ---------- ---------- ----------
thermophila_A  ---------- ---------- ---GGAGAAC TGCTCCACAG GGAAGGGCTG AT--------- ---------- ---------- ---------- ----------
thermophila_B  ---------- ---------- ---GGAGAAC TGCTCCACAG GGAAGGGCTG AT--------- ---------- ---------- ---------- ----------
LCMS           -----TATTT GGG------- ----AAGGAT A--------- ----GATTTG TTTAAATACC CACCCGGTAT TTATGAG--- -----AAATA TTTTCTCTTA
stadtmanae_A   TTAGATGTTT GTG-TAATAT ATAATAGAAT AGTAATATTT CTATGGTTTA CTTAAATAAT ATAATAGTAT TAATACGTTT TAATAAGATT TGCTTATATA
stadtmanae_B   TTAGATGTTT GTGTAATAAT ATAATAGAAT AGTAATATTT CTATGGTTTA CTTAAATAAT ATAATAGTAT TAATACGTTT TAATAAGATT TGCTTATATA
stadtmanae_C   TTAGATGTTT GTGTAATAAT ATAATAGAAT AGTAATTCT- --ATGGTTTA CTTAAATAAT ATAATAGTAT TAATACGTTT TAATAAGATT TGCTTATATA
stadtmanae_D   TTAGATGTTT GTGTAATAAT ATAATAGAAT AGTAATTCT- --ATGGTTTA CTTAAATAAT ATAATAGTAT TAATACGTTT TAATAAGATT TGCTTATATA
burtonii_A     -----TGCTT GGG------- ----AAGGAT G--------- ----GATGTG CCTGA----- ---------- ---------- ---------- ----------
burtonii_B     -----TGCTT GGG------- ----AAGGAT G--------- ----GATGTG CCTGA----- ---------- ---------- ---------- ----------
burtonii_C     -----TGCTT GGG------- ----AAGGAT G--------- ----GATGTG CCTGA----- ---------- ---------- ---------- ----------
barkeriA       -------TTC GGG------- ---GAAGGGC GGATTGCCTG CGTTGACACG C--------- ---------- ---------- ---------- ----------
barkeriB       -------TTC GGG------- ---GAAGGGC GGATTGCCTG CGTTGACACG C--------- ---------- ---------- ---------- ----------
barkeriC       -------TTC GGG------- ---GAAGGGC GGATTGCCTG CGTTGACACG C--------- ---------- ---------- ---------- ----------
mazeiA         -------TTC GGG------- ---GAAGGAT GGATAGCCTG CGCGGAACCG C--------- ---------- ---------- ---------- ----------
mazeiB         -------TTC GGG------- ---GAAGGAT GGATAGCCTG CGCGGAACCG C--------- ---------- ---------- ---------- ----------
mazeiC         -------TTC GGG------- ---GAAGGAT GGATAGCCTG CGCGGAACCG C--------- ---------- ---------- ---------- ----------
acetivoransB   -------TTC GGG------- ---GAAGGAT GGATAGCCTG TGCTGAAGCT C--------- ---------- ---------- ---------- ----------
acetivoransA   -------TTC GGG------- ---GAAGGAT GGATAGCCTG TGCTGAAGCT C--------- ---------- ---------- ---------- ----------
acetivoransC   ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------

             501
thermoauto_A   ---------- ---------T GATTTCCACG CATAGGAGAA ACC------- ---------C GATTGTAATC --CAAAACTG GCATTAACTG ACCAGAGAGA
thermoauto_B   ---------- ---------T GATTTCCACG CATAGGAGAA ACC------- ---------C GATTGTAATC --CAAAACTG GCATTAACTG ACCAGAGAGA
thermophila_A  ---------- ---------- -------AAT CAAGATGAGG CCA------- ---------C GTATACGCTT TGCAGACCAG ACGCTCACTG ---------A
thermophila_B  ---------- ---------- -------AAT CAAGATGAGG CCA------- ---------C GTATACGCTT TGCAGACCAG ACGCTCACTG ---------A
LCMS           TGTATTTAGC TTTTTTGCTT TTGCATAAAA CAAAGTGAAA TCG------- ---------T GTATATGAAT GGCATATTAG ACGCTCACTG ---------A
stadtmanae_A   TATTTTTTGC TTTGTTTGAT TTGTTTTAAA TCTACTACAG ACAATATTAT TTTTGTCACT ATATATAATA GGAAAA--AA AAGAACACTG ------TATA
stadtmanae_B   TATTTTTTGC TTTGTTTGAT TTGTTTTAAA TCTACTACAG ACAATATTAT TTTTGTCACT ATATATAATA GGAAAA--AA AAGAACACTG ------TATA
stadtmanae_C   TATTTTTTGC TTTGTTTGAT TTGTTTTAAA TCTACTACAG ACAATATTAT TTTTGTCACT ATATATAATA GGAAAA--AA AAGAACACTG ------TATA
stadtmanae_D   TATTTTTTGC TTTGTTTGAT TTGTTTTAAA TCTACTACAG ACAATATTAT TTTTGTCACT ATATATAATA GGAAAA--AA AAGAACACTA ------TGTA
burtonii_A     ---------- ------TACC CCATATCAGG TACTATGAGA TCA------- ---------T GTATACATAT TACATATCAG ACGCTCACTG ---------G
burtonii_B     ---------- ------TACC CCATATCAGG TACTATGAGA TCA------- ---------T GTATACATAT TACATATCAG ACGCTCACTG ---------G
burtonii_C     ---------- ------TACC CCATATCAGG TACTATGAGA TCA------- ---------T GTATACATAT TACATATCAG ACGCTCACTG ---------G
barkeriA       ---------- ---------- -------AGG TACA-TGAAG TCA------- ---------T GTATAAGTGC TGTATACTGG ACGCTTTCTG ---------G
barkeriB       ---------- ---------- -------AGG TACA-TGAAG TCA------- ---------T GTATAAGTGC TGTATACTGG ACGCTTTCTG ---------G
barkeriC       ---------- ---------- -------AGG TACA-TGAAG TCA------- ---------T GTATAAGTGC TGTATACTGG ACGCTTTCTG ---------G
mazeiA         ---------- ---------- -------AGG CACA-TGAAG TCG------- ---------T GTATAGGTGC TGTATATTGA ACGCTAACTG ---------G
mazeiB         ---------- ---------- -------AGG CACA-TGAAG TCG------- ---------T GTATAGGTGC TGTATATTGA ACGCTAACTG ---------G
mazeiC         ---------- ---------- -------AGG CACA-TGAAG TCG------- ---------T GTATAGGTGC TGTATATTGA ACGCTAACTG ---------G
acetivoransB   ---------- ---------- -------AGG CATA-TGAAG TCG------- ---------T GTATATGTGC TGTATACTGG ACGCTTACTG ---------G
acetivoransA   ---------- ---------- -------AGG CATA-TGAAG TCG------- ---------T GTATATGTGC TGTATACTGG ACGCTTACTG ---------G
acetivoransC   ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------A

             601
thermoauto_A   AGGC-AGTTA AACCAAACCC TAGCTTA--- ---------- ---------- ----------
thermoauto_B   AGGC-AGTTA AACCAAACCC TAGCTTA--- ---------- ---------- ----------
thermophila_A  GATC-AGTGG GACGATTAAG CTGCT----- ---------- ---------- ----------
thermophila_B  GATC-AGTGG GACGATTAAG CTGCT----- ---------- ---------- ----------
LCMS           ATATAAGTGA GAGTGATTCT GGTAC----- ---------- ---------- ----------
stadtmanae_A   AAAT-GGTGA AATTTTGTAT AATAAAAAAT TTTTTTTCTT ---------- ----------
stadtmanae_B   AAAT-GGTGA AATTTTGTAT AATAAAAAAT TTTTTTTCTT ---------- ----------
stadtmanae_C   AAAT-GGTGA AATTTTGTAT AATAAAAAAT TTTTTTTCTT ---------- ----------
stadtmanae_D   AAAT-GGTGA AATTTTGTAT AATAAAAAAT TTTTTTTCTT ---------- ----------
burtonii_A     ACAA-AGTGA GATGGACTCT GGTA------ ---------- ---------- ----------
burtonii_B     ACAA-AGTGA GATGGACTCT GGTA------ ---------- ---------- ----------
burtonii_C     ACAA-AGTGA GATGGACTCT GGTA------ ---------- ---------- ----------
barkeriA       ACCT-GGTTA GGATACACAG GAA------- ---------- ---------- ----------
barkeriB       ACCT-GGTTA GGATACACAG GAA------- ---------- ---------- ----------
barkeriC       ACCT-GGTTA GGATACACAG GAA------- ---------- ---------- ----------
mazeiA         ACCT-GGTTA GGTATATAGG AAT------- ---------- ---------- ----------
mazeiB         ACCT-GGTTA GGTATATAGG AAT------- ---------- ---------- ----------
mazeiC         ACCT-GGTTA GGTATATAGG AA-------- ---------- ---------- ----------
acetivoransB   ACCT-GGTTA GGTAAT---- ---------- ---------- ---------- ----------
acetivoransA   ACCT-GGTTA GGTAATTAGG AATTATGCTA TCAGGTGGAT GGCTCGGCTC AAGAGCTTA
acetivoransC   AAGT-AGTT- ---------- ---------- ---------- ---------- ----------
```

Figure S1. The LCMS ITS region encodes a tRNA and shows sequence similarity to the ITS regions of several methanogens. The alignment includes sequences from: *Methanosaeta thermophila* (NC_008553), *Methanosarcina barkeri* (NC_007349), *Ms. acetivorans* (NC_003552), *Ms. mazei* (NC_003901), *Methanococcoides burtonii* (NC_007955), *Methanobacterium thermoautotrophicus* (NC_000916), *Methanosphaera stadtmanae* (NC_007681), and Lost City Methanosarcinales (GQ273207).

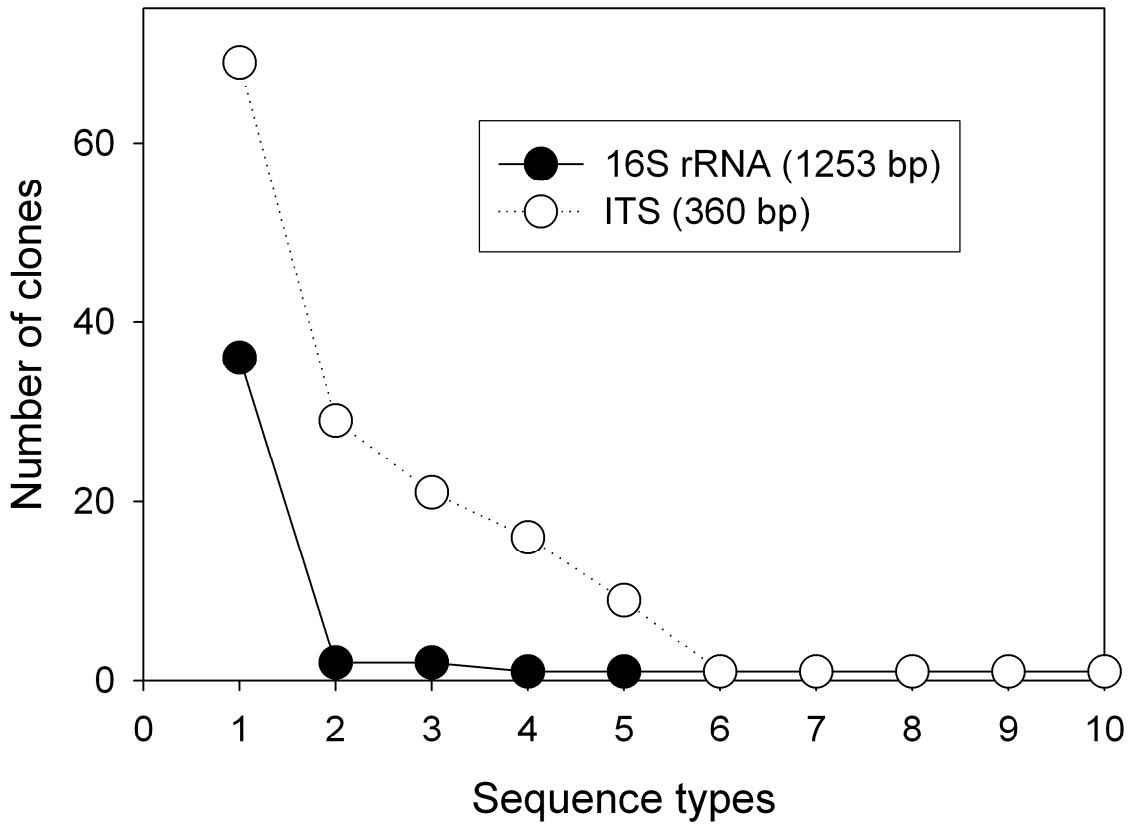


Figure S2. Rank-abundance plot showing the number of clones sharing the 10 most frequently occurring 16S rRNA and ITS sequences in samples LC0424 and LC1408, both of which were collected from the Poseidon chimney (Marker 3). Only one 16S rRNA sequence occurs more than twice, but five ITS sequences occur many times in this sample. As shown in Figure 3b, other samples contain different abundant ITS sequences.