Archaea and bacteria with surprising microdiversity show shifts in dominance over 1000-year time scales in hydrothermal chimneys

William J. Brazelton\textsuperscript{a*}, Kristin A. Ludwig\textsuperscript{a†}, Mitchell L. Sogin\textsuperscript{b}, Ekaterina N. Andreishcheva\textsuperscript{b}, Deborah S. Kelley\textsuperscript{a}, Chuan-Chou Shen\textsuperscript{c}, R. Lawrence Edwards\textsuperscript{d}, and John A. Baross\textsuperscript{a}

\textsuperscript{a} School of Oceanography and Center for Astrobiology and Early Evolution, University of Washington, Seattle, WA 98195
\textsuperscript{b} Josephine Bay Paul Center, Marine Biological Laboratory at Woods Hole, 7 MBL Street, Woods Hole, MA 02543
\textsuperscript{c} Dept. of Geosciences, National Taiwan University, No. 1, Sec. 4, Roosevelt Rd., Taipei 106, Taiwan
\textsuperscript{d} Dept. of Geology and Geophysics, University of Minnesota, 310 Pillsbury Dr. SE, Minneapolis, MN 55455

\textsuperscript{†} Present address: Consortium for Ocean Leadership, 1201 New York Ave 4\textsuperscript{th} Floor, Washington, D.C. 20005

*corresponding author: School of Oceanography, Box 357940, University of Washington, Seattle, WA 98195; Phone: 206-221-5755; Fax: 206-543-6073. E-mail: braz@u.washington.edu

24 text pages, 1 table, 5 figures, 11 SI text pages, 4 SI tables, 6 SI figures

Abbreviations: LCHF, Lost City Hydrothermal Field; LCMS, Lost City \textit{Methanosarcinales}; OTU, operational taxonomic unit

Data deposition: Tag sequences are available in the VAMPS database (http://vamps.mbl.edu), and in the NCBI Short Read Archive under accession number SRP000912.
Abstract

The Lost City Hydrothermal Field, an ultramafic-hosted system located 15 km west of the Mid-Atlantic Ridge, has experienced at least 30,000 years of hydrothermal activity. Previous studies have shown that its carbonate chimneys form by mixing of ~90°C, pH 9-11 hydrothermal fluids and cold seawater. Flow of methane and hydrogen-rich hydrothermal fluids in the porous interior chimney walls supports archaeal biofilm communities dominated by a single phylotype of *Methanosarcinales*. In this study, we have extensively sampled the carbonate-hosted archaeal and bacterial communities by obtaining sequences of >200,000 amplicons of the 16S rRNA V6 region and correlated the results with isotopic (^230^Th) ages of the chimneys over a 1200 year period. Rare sequences in young chimneys were often more abundant in older chimneys, indicating that members of the rare biosphere can become dominant members of the ecosystem when environmental conditions change. These results suggest that a long history of selection over many cycles of chimney growth has resulted in numerous closely related species at Lost City, each of which is pre-adapted to a particular set of re-occurring environmental conditions. Due to the unique characteristics of the Lost City Hydrothermal Field, these data offer an unprecedented opportunity to study the dynamics of a microbial ecosystem's rare biosphere over a thousand-year time scale.

Introduction

Mucilaginous biofilms coat the porous carbonate mineral matrix of actively venting chimneys of the Lost City Hydrothermal Field (LCHF), an ultramafic-hosted system located 15 km west of the spreading axis of the Mid-Atlantic Ridge at a depth of ~750 m [1](#). Previous studies have
shown that the biofilms contain $10^6$-$10^9$ cells per gram of carbonate mineral and that >80% of the cells belong to a single phylotype of archaea known as Lost City Methanosarcinales (LCMS) (2). It is presumed that the abundant LCMS cells have adapted to the ~90°C, pH 9-11 fluids and that they utilize high concentrations of dissolved hydrogen (H$_2$) and methane (CH$_4$) in the vent fluids (3, 4), but their physiology is unknown. As chimneys become less active, changes in mineralogy and fluid chemistry cause other organisms to become abundant (5). For example, chimneys with very little visible hydrothermal venting do not contain the LCMS biofilm.

Instead, a single phylotype belonging to the ANME-1 group of anaerobic methanotrophic archaea is the dominant archaeon (5). The outer walls of carbonate chimneys, where mineralogy and fluid chemistry can be substantially different compared to the chimney interiors (2, 6), harbor bacteria with high similarity to sulfur- and methane-oxidizers, but archaea are much more abundant in chimneys venting up to 90°C, high pH fluid (2).

Unlike typical black smoker systems that are fueled by heat from underlying magmatic activity, hydrothermal flow at the LCHF is driven by cooling of ultramafic rocks and lesser gabbroic material that underlie the field (1). Fluid chemistry is governed by exothermic geochemical reactions in the subsurface known as serpentinization (1). These reactions commonly occur in ultramafic environments on Earth where water reacts with the mineral olivine, and they are expected to occur on other planetary bodies hosting aqueous fluids (7). Serpentinization of subsurface ultramafic minerals is one of the most likely sources of CH$_4$ on Mars (8). At the LCHF, serpentinization reactions produce alkaline (pH 9-11) fluids that are rich in calcium (up to 30 mmol/kg), low in metals, and have near-zero concentrations of CO$_2$ (1, 4, 6).

Serpentinization also provides an abiogenic source of H$_2$ (14 mmol/kg), CH$_4$ (2 mmol/kg), and
low molecular weight hydrocarbons (3, 4). The ability of Lost City-like environments to generate reduced organic compounds exothermically and abiotically makes them plausible settings for the origin and early evolution of life (9).

Lost City also differs from most magma-driven systems in that it is very long-lived (10). Carbon isotopic measurements indicate that venting has been ongoing for at least 30,000 years, with individual chimneys active for at least 300 years, and modeling results indicate the system could remain active for up to 1 million years (10). Therefore, conditions within Lost City chimneys could have been conducive to the growth of LCMS for tens or hundreds of thousands of years, and it is possible that it has remained the dominant member of the ecosystem throughout that time. To our knowledge, biology has no precedent for such a low diversity, long-lived ecosystem. Characterizations of microbial community dynamics with respect to time are typically performed within the context of habitat monitoring. Numerous studies have described seasonal changes of particular communities (11, 12), documented initial colonizations of substrates (13), and observed short-term responses to environmental changes (14, 15). The time scales of these studies are on the order of months or at most years, practical time periods for observational studies. No direct evidence exists for changes in microbial communities occurring over longer time scales.

Technological advances in nucleic acid sequencing (16) are increasing the sensitivity at which changes in microbial communities can be detected. Recent high-sensitivity surveys of diverse environmental samples including those from shallow and deep ocean waters (17), hydrothermal fluids (17, 18), arctic tundra (19), and various soils (19, 20) have revealed that the vast majority
of a microbial habitat's diversity is comprised by taxa present at very low abundances – so low that they were previously undetectable. The most abundant organisms represent only a fraction of the total diversity. The high diversity of rare sequences indicates that they have been evolving over long periods of time and that at least some of them belong to organisms that are pre-adapted to environmental conditions that have occurred in the past and may re-occur in the future. A fundamental prediction of the 'Rare Biosphere' model (17) is that when environmental conditions change, some of these rare, pre-adapted taxa can rapidly exploit the new conditions, increase in abundance, and out-compete the once abundant organisms that were adapted to the past conditions. No studies have tested this prediction by examining a shift in species composition involving extremely rare taxa occurring during a known time interval.

The longevity and extremely low diversity of the LCHF carbonate chimneys provide a unique test of the Rare Biosphere model. In this study, we utilize pyrosequencing technology to generate an extensive sequence dataset of the 16S rRNA V6 region from four carbonate chimney samples collected from structures varying from a massive edifice venting 88°C, high pH fluid to a small, inactive chimney bathed in seawater. We investigate whether these low diversity communities harbor highly diverse but rare species and attempt to correlate our biological data with chimney sample ages derived from co-registered $^{230}$Th isotopic measurements. We show that rare sequences in young chimneys samples are often much more abundant in older samples, suggesting that organisms can remain rare for long time periods until environmental changes allow them to proliferate.
RESULTS

230Th dating of carbonate chimney samples

Uranium-thorium ($^{238}$U-$^{234}$U-$^{230}$Th-$^{232}$Th or U-Th or $^{230}$Th) disequilibrium dating is a powerful geochronological tool used to date inorganic and biogenic materials ranging in age from modern to 600 kyrs (21). In this study, four discrete chimney samples (Fig. 1) were dated using $^{230}$Th age-dating techniques. Sample 1 ($^{230}$Th age = 34.3 ± 8.1 yrs) was collected from a site known as Marker 3 or Poseidon where emitting fluids reach 88ºC (4). Samples 2 and 3 are from a flange named Marker C that is bathed by fluids exhibiting a temperature gradient of 9-70ºC due to mixing of hydrothermal fluids with ambient seawater. Samples 2 and 3 were located 20 cm apart (Fig. 1) but yielded very different ages ($^{230}$Th ages = 42.9 ± 23.3 yrs and 128.4 ± 57.2 yrs, respectively). Sample 4 ($^{230}$Th age = 1245 ± 257 yrs) was collected from a chimney with no visible venting and contained coral polyps on its exterior, typical of inactive chimneys at the LCHF (6). Complete descriptions of the samples and the $^{230}$Th age-dating technique are available in the Supplementary Information.

Archaeal community structure

Of 167,031 total archaeal V6 amplicons (‘tags’) sequenced from the four carbonate chimney samples (Fig. 1, Table S1), we identified 2,635 unique sequence types that clustered into 817 operational taxonomic units (OTUs) at the 3% distance threshold. Sample 4 ($^{230}$Th age = 1245 ± 257 yrs) contained significantly fewer OTUs than samples 1-3 ($^{230}$Th ages = 34-128 yrs), as shown by the rarefaction curves of observed OTUs versus the total number of tags sequenced (Fig. 2). The near-asymptotic shape of the rarefaction curves as well as the small difference between the observed numbers of OTUs and the values from two richness estimators (Tables 1
and S1) indicate that our study has nearly completely sampled the archaeal V6 amplicon libraries from these samples.

Figure 3 presents normalized profiles of archaeal OTUs for each sample. A single OTU corresponding to the previously published 16S rRNA sequence of the putative methanogen Lost City *Methanosarcinales* (LCMS) (2, 5) comprised >90% of all tag sequences from samples 1-3. An additional 533 OTUs, none of which contained >5% of any sample’s total tags, were also assigned to order *Methanosarcinales* (Fig. 3b) by the GAST automated taxonomic assignment process (22). The relative abundance of variant sequences generally decreased with increasing distance from the dominant sequence, a trend consistent with that expected from artificial diversity generated by pyrosequencing error. Several sequences occur much more frequently than predicted by random error, however, and probably represent genuine diversity. Furthermore, we have recently reported that variation in the V6 region of these *Methanosarcinales* sequences is correlated with greater variation in the intergenic transcribed spacer (ITS) region, a marker that often reflects physiological variation (23). Therefore, the number of OTUs measured in the V6 region by this study may underestimate the total genetic diversity.

The oldest sample (1245 ± 257 yrs) was also dominated by a single OTU, but it corresponds to the previously published (5) 16S rRNA clone LC1133A-9, which is phylogenetically similar to the ANME-1 group of anaerobic methanotrophic archaea (Fig. 3a). Interestingly, the same OTU that dominates sample 4 is also present at low abundance in sample 1 (Fig. 3b), indicating that ANME-1 organisms are present in chimneys as young as 34 yrs but are only dominant in much older chimneys. A third major archaeal group was represented by a set of sequences primarily
present only in sample 1 and matching 16S rRNA clone LC1408cren-55, which is 97% similar over its full length to the ammonia-oxidizing crenarchaeon *Nitrosopumilus maritimus* (24). The presence of these sequences in the youngest, hottest chimney sample and in hot chimney fluids (5) indicates that they probably represent thermophilic organisms, though their physiology is unclear because ammonia oxidation is not expected to be thermodynamically favorable in the anoxic fluids of LCHF chimneys.

The patterns visualized in Figure 3 are quantified in two Venn diagrams illustrating the percentage of tags and numbers of OTUs shared among samples (Fig. S2). We also calculated overall community similarities (Fig. S3) based on the relative abundance of each OTU (Bray-Curtis similarity coefficient) as well as the presence/absence of each OTU (Jaccard similarity coefficient). As expected from Fig. 3, the three samples aged 34-128 yrs were all very similar to each other (>85% Bray-Curtis, >48% Jaccard) and very different than sample 4 (<0.3% Bray-Curtis, <2% Jaccard). Remarkably, the archaeal community similarity between samples 2 and 3 (95% Bray-Curtis, 56% Jaccard) was much higher than the similarity between sample 1 replicates, so the small differences between samples 2 and 3 may not reflect any natural variation. See Supplementary Information for more details on estimation of variation between replicates.

**Bacterial community structure**

Our study identified most of the diversity present in the bacterial V6 amplicon library from the youngest chimney sample (34.3 ± 8.1 yrs), as shown by the nearly asymptotic rarefaction curve of OTUs discovered in sample 1 (Fig. 2). The rarefaction curves of samples 3 and 4 (128.4 ±
57.2 and 1245 ± 257 yrs), in contrast, have very steep slopes and are therefore likely to contain much more diversity than we were able to discover with >7000 tag sequences from each sample. A few, highly abundant sequence types dominated the bacterial communities in each sample with rare sequences accounting for most of the observed diversity (Fig. 4). This pattern is very similar to that seen with the archaeal communities (Fig. 3b), although the bacterial communities have a more even distribution (reflected in the much higher Simpson's reciprocal index of diversity, Tables 1 and S1) and have greater diversity, especially at large taxonomic distances (Fig. S1).

The taxonomic groups to which V6 tag sequences were assigned (Fig. 4) were consistent with previous work utilizing different techniques (5). Notably, sequences assigned to *Gammaproteobacteria*, *Alphaproteobacteria*, *Firmicutes*, and *Chloroflexi* were the most abundant and diverse. The most abundant phylotypes in each group (Table S3) probably represent non-photosynthetic organisms oxidizing reduced sulfur species (eg. *Thiotrichales*, *Rhodobacterales*) or H₂ (eg. *Clostridia*, *Dehalococcoidetes*) from the vent fluids. The small but significant quantities of abiotically-derived organic carbon found in Lost City hydrothermal fluids (4) may be an important carbon source for some of the bacteria, particularly *Clostridia* with high sequence similarity to *Desulfotomaculum* species, which occur in the youngest, most hydrothermally-influenced sample (Table S3). Many *Clostridia* utilize low molecular weight organic acids, often in syntrophic association with methanogens (25, 26, 27). The possible ecological roles of these organisms in LCHF chimneys have been discussed previously (5).

Although the shifts in bacterial OTU composition are not as obvious as those in the archaeal communities, there are clear differences among samples with different ages (Fig. 4). Samples 1
and 2 (34.3 ± 8.1 and 42.9 ± 23.3 yrs) have high overall community similarity (45% Bray-Curtis, 23% Jaccard), and samples 3 and 4 (128.4 ± 57.2 yrs and 1245 ± 257 yrs) are very similar to each other (34% Bray-Curtis, 23% Jaccard; see Fig. S3). Note that samples 2 and 3, which were only ~20 cm apart in situ (Fig. 1) and had nearly identical archaeal communities, have only moderate bacterial community similarity (18% Bray-Curtis, 17% Jaccard). Interestingly, the distribution of OTUs among samples was roughly constant regardless of whether all OTUs are included in the analysis or if only abundant or if only rare OTUs are considered (Figures S4 and S5). In short, samples of similar ages contain similar bacterial communities, and with few exceptions, this trend is true for dominant as well as rare organisms.

**Diversity within *Thiomicrospira***

Previous work has shown that sequences most closely related to genus *Thiomicrospira* (order Thiotrichales, Table S3) are widely distributed throughout the LCHF (5). A single OTU assigned to genus *Thiomicrospira* corresponding to previously published full length 16S rRNA clone LC1537B-12 comprised 37% of all bacterial V6 tags from the youngest sample (34.3 ± 8.1 yrs). The same OTU also occurred in the other three samples, but at lower relative abundance (28% at 43 yrs, 3% at 128 yrs, and 8% at 1245 yrs). An additional 23 OTUs (group 'C' in Fig. 5), none of which comprised more than 2% of any sample's total tags, were at least 93% similar to clone LC1537B-12. Similarly, group 'A' contained two dominant and several rare OTUs, all of which were at least 96.7% similar to the V6 region of clone LC1537B-49. The high occurrence of these OTUs in the younger, higher pH samples and absence in the two older, more neutral pH samples is consistent with the physiology of their closest relative (Fig. 5a), the alkaliphile *Thioalkalimicrobium cyclicum* (28). Group 'B' OTUs, in contrast, were the most dominant
Thiomicrospira-like sequences in sample 4, which showed no signs of venting and was likely the same temperature as the background seawater. All of these OTUs were at least 95% similar to a clone with 95.3% similarity over the full gene length to the psychrophile Thiomicrospira psychrophila (29).

Although all samples have several Thiomicrospira OTUs, samples of different ages are dominated by different OTUs (Fig. 5). The most abundant Thiomicrospira OTU in the oldest sample, for example, is extremely rare in the youngest sample, and the overall similarity of the Thiomicrospira community between these two samples is low (24% Bray-Curtis, 22% Jaccard). As seen with the whole bacterial community, the two youngest samples have the most similar Thiomicrospira OTUs (61% Bray-Curtis, 32% Jaccard), and the two oldest samples are also very similar (39% Bray-Curtis, 31% Jaccard).

**DISCUSSION**

**Dominance by a few species**

Over the past >30,000 years of venting at the Lost City Hydrothermal Field (10), extreme environmental conditions created by mixing of <150°C, pH 11 fluids enriched in dissolved H₂ and CH₄ with oxygen-rich seawater have selected for an unusual assemblage of microorganisms. Previous studies have shown that >80% of detectable cells in the warm, anoxic interior zones of carbonate chimneys are dominated by biofilms formed by Lost City Methanosarcinales (LCMS, 2). This study highlights the extremely low archaeal diversity of Lost City chimneys: >90% of all V6 tags in each sample clustered into a single OTU. Such extreme dominance of single phylotypes within hydrothermal chimneys is unprecedented; in other hydrothermal environments
where methanogens dominate, diverse species adapted to various metabolic substrates are typically present (30, 31).

A small number of species also dominate the bacterial communities, and the identity of the dominant species often differs among chimney samples (Fig. 4). This trend is true even within narrow taxonomic groups. For example, chimney samples varied with respect to which OTUs representing genus *Thiomicrospira* dominated the community (Fig. 5). These organisms probably require reduced sulfur species from hydrothermal fluid, but they also need access to oxygen (the electron acceptor for all known *Thiomicrospira*) from seawater. Inorganic carbon is unavailable from high pH hydrothermal fluid, so organisms in chimneys with little exposure to seawater may be carbon-limited, as evidenced by the unprecedented enrichment of $^{13}$C in lipids extracted from Lost City chimneys (32). Thus the diversity of *Thiomicrospira* sequences detected in this study (Fig. 5) probably resulted from many similar species adapting to slightly different niches within concentration gradients of reduced sulfur species, oxygen, and inorganic carbon and the associated temperature and pH gradients.

**Species shift between rare and dominant**

The most notable differences in the archaeal community involve the shift to a community completely dominated by ANME-1 sequences in sample 4 ($1245 \pm 257$ yrs) as well as the relative lack of the Marine Group I *Crenarchaeota* in samples 2-4. Although ANME-1 sequences are nearly absent from samples 1-3, the identical sequence that dominates sample 4 is present, but rare, in sample 1 ($34.3 \pm 8.1$ yrs), which was collected from a chimney 100 meters distant and shallower in depth. Deeper sequencing of samples 2 and 3 would most likely also recover
this sequence at low abundance. Previous work has shown that ANME-1 sequences are present in multiple chimneys throughout the LCHF (5) including one chimney that has been dated to 10,500 years (33). Therefore, ANME-1 organisms are well dispersed throughout the vent field but are able to thrive only in chimneys that are not venting high-temperature, high-pH fluids, conditions typical of old chimneys (6). It is likely that these chimneys previously vented high-temperature, high-pH fluids and were therefore inhabited by microbial communities more similar to that of the younger chimneys reported here.

Many more examples of sequences that are rare in young samples but abundant in older samples are evident in the bacterial V6 tag pyrosequences (Fig. S6). Taxonomic differences among chimney samples of varying ages are most likely caused by differences in mineralogy and fluid chemistry, and both of these factors are expected to change with somewhat predictable trends during the cycle of chimney formation, growth, and senescence (6). Although fluid chemistry is less directly correlated to chimney sample age than mineralogy, it is clear that over long time scales fluid chemistry strongly determines the mineralogy of chimneys (6). On short time scales, fluctuations in chemistry may affect microbial communities independently of mineralogy and chimney sample age, and it can be expected that these factors will have varying influences on different organisms.

Further work is necessary to confidently distinguish the influences of mineralogy, fluid chemistry, and age on each archaeal and bacterial species, but in general, our results show that chimney sample age can be a useful indicator of bacterial community composition (Fig. 4). For example, samples 2 and 3 were collected from the same chimney flange, separated by
approximately 20 cm in situ, but have remarkably different bacterial communities (18% Bray-Curtis similarity). Sample 2 (42.9 ± 23.3 yrs) was closer to the vent flange opening and is composed of aragonite (CaCO₃) and brucite (Mg(OH)₂) minerals that are stable in vent fluid-dominated environments. In contrast, sample 3 is older (128.4 ± 57.2 yrs), contains more calcite than aragonite (calcite is a polymorph of metastable aragonite) and little brucite, and was more exposed to ambient seawater (6). Clearly, the distribution of bacterial species correlates more strongly with sample age and mineralogy than with physical distance since sample 2 has high similarity (45% Bray-Curtis) with sample 1, which is from a different chimney but of similar mineralogy and age (34.3 ± 8.1 yrs). Moreover, sample 3 is dominated by the same bacterial sequences as sample 4 (1245 ± 257 yrs), even though those two samples were separated by at least 100 meters laterally and in depth. The four samples examined in this study do not constitute strong statistical evidence, but the initial data from this study, together with previous surveys of microbial diversity and geochemistry of additional samples (5, 6), are consistent with a correlation of mineralogy and age with microbial community composition.

The archaeal communities, in contrast, appear to be more strongly influenced by fluid chemistry when comparing samples younger than ~150 yrs. Although samples 2 and 3 have different ages and bacterial communities, their archaeal populations are practically identical (95% Bray-Curtis similarity). Both samples were exposed to the same hydrothermal fluid, which may have fueled the same archaeal population while bacteria in the older sample adapted over time to the increased influence of cold seawater that also resulted in mineralogical and porosity changes (6). Both samples have high cell densities and high proportions of archaea (Table S4), indicating that the archaeal populations are strongly stimulated by the conditions at this chimney flange. Thus,
the archaeal community may not be influenced by mineralogical changes as long as fluid chemistry (e.g., H₂ and CH₄ concentrations) remains constant. Indeed, minor differences in fluid chemistry (3, 4) between chimneys may explain the presence of *Crenarchaeota* in sample 1 but not sample 2, which is of nearly identical age (Table 1).

**The Lost City Rare Biosphere**

The archaeal and bacterial communities inhabiting carbonate chimneys at the LCHF are not ‘frozen in time’ and do not record the history of the chimneys. Nevertheless, many chimneys must have formed, grown, and become inactive during the >30,000 years that the LCHF has been active (10). Furthermore, it is clear that mineralogy and fluid chemistry are correlated over long time scales with chimney growth stages (6) and chimney age (33). Therefore, we infer that old chimneys formerly had environmental characteristics typical of younger chimneys. There is no reason to expect important differences between young chimneys active today and young chimneys that were active thousands of years ago. If environmental conditions broadly determine microbial community composition, microbial communities inhabiting chimneys today should resemble those that inhabited chimneys of similar ages hundreds or thousands of years ago.

Although we have been able to collect V6 tag pyrosequences from only four samples, previous studies of LCHF chimneys have shown that archaeal and bacterial communities change when chimneys become less active (5). Furthermore, ²³⁰Th ages of additional LCHF chimney samples are consistent with the trend that chimney samples of similar ages have similar mineralogy and fluid chemistry (33). In conclusion, the results from this study and others (1, 5, 6, 10, 33) support a connection between microbial community composition and the geochemical conditions characteristic of a chimney sample's age over long time scales.
The shifts in community composition observed in this study do not reflect new speciation events. Our results show that organisms favorably selected by new conditions already existed at low levels before the environmental change occurred (Fig S6). Therefore, the genomes of the favorably selected organisms must have already encoded the necessary adaptations prior to the change. They were preadapted to the new conditions. Alternatively, every community shift may have resulted from a rare organism acquiring a useful gene via lateral gene transfer. Another possibility is that mutation and/or recombination occurred much more quickly in genes required for the adaptation but not in the V6 hypervariable region measured in this study. These processes may occur, but a more parsimonious explanation of the many shifts detected in this study where a rare organism rapidly increased its relative abundance is that these organisms had been preadapted for hundreds or thousands of years to the particular niche created by the environmental change. Therefore, the rare biosphere of the Lost City microbial community represents a large repository of genetic memory created during a long history of past environmental changes that selected for new species within a small pool of organisms that originally colonized the extreme environment. The rare organisms were able to rapidly exploit new niches as they arose because they had been previously selected for the same conditions in the past.

The Lost City Hydrothermal Field has been active for at least 30,000 years and probably much longer (10). Ultramafic environments such as Lost City have surely existed throughout Earth's history and were probably much more widespread on the early Earth when ultramafic lavas were more common (1, 34). Indeed, the large quantities of hydrogen generated at ultramafic-hosted systems like Lost City may have supported the earliest ecosystems on Earth (35, 36). Therefore,
the ecological dynamics described in this study may have been occurring in similar environments for most of Earth's history, and the large pool of rare organisms present today may reflect that long history.

METHODS

Pyrosequencing and diversity calculations

Carbonate chimney samples were collected from the LCHF with DSV Alvin during cruise AT07-34 aboard the R/V Atlantis in April/May 2003 (http://www.lostcity.washington.edu). Detailed descriptions of $^{230}$Th dating techniques and the DNA extraction protocol are available in the Supporting Information. V6 amplicon libraries were constructed and sequenced as in (17) and (18) with a 454 Life Sciences GS20 pyrosequencer and a 454 Life Sciences FLX pyrosequencer for the archaeal data. Tag sequences were screened for quality as recommend by (37). Archaeal and bacterial sequence alignments were constructed by submitting to the NAST aligner (http://greengenes.lbl.gov) all unique sequences pooled from all four samples, including primers to ensure full-length alignment. Obvious alignment errors were manually corrected, and primers were trimmed, resulting in final archaeal and bacterial alignments in which most sequences contained 60-65 basepairs. Similar results as those reported here were achieved by aligning with MUSCLE (38). The distance matrix for each alignment was calculated with quickdist as described in (17). Sequences were clustered into OTUs, and rarefaction curves and diversity estimators were calculated with DOTUR (39). Some of the rare sequences detected in this study may be artifacts of pyrosequencing technology (40, 41), but this possibility does not affect the main conclusions of our study. Because many of the rare sequences are much more abundant in other samples, these sequences are clearly not artifacts. Furthermore, we have recently reported
that diversity in V6 tag pyrosequences is correlated to diversity at another marker and may actually underestimate the total genetic diversity (23). Additional methods for pyrosequencing and error determination are available in the Supplementary Information.

Comparison of OTU membership among samples

After sequences were clustered into OTUs with DOTUR, the program SONS (42) was used to determine the relative abundance distribution of each OTU in each sample. The SONS output informed creation of the plots in Figs. 3-5 and S2-S6. Samples 1-4 yielded 16260, 32345, 25471, and 21983 archaeal tags and 21582, 5567, 7162, and 8716 bacterial tags, respectively. To normalize relative abundances of each OTU among samples, tags were randomly resampled down to the sample with the fewest tags using Daisy-Chopper (available from P. Swift, J. Gilbert, and D. Field) after clustering OTUs. Consequently, normalized numbers of OTUs reported in Table 1 and Figs. 3-5 are lower than the total non-normalized values reported in Table S1 and Fig. 2. Bray-Curtis and Jaccard similarities between samples were calculated with Primer 6 (http://www.primer-e.com) without any further data transformation. Taxonomies were assigned to each tag by the GAST process (22) via the VAMPS website.

Data Availability

All V6 sequence data is available at the VAMPS database, http://vamps.mbl.edu, under dataset names ICM_LCY_Av6 and ICM_LCY_Bv6 and in the NCBI Short Read Archive under submission number SRP000912. See Supplementary Information for sample names in the databases.
ACKNOWLEDGEMENTS

We would like to express our appreciation to the crews of the R/V *Atlantis* and DSV *Alvin* and the scientific party of the 2003 Lost City Expedition. We thank J. Huber and S. Huse for helpful discussions. 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, NSF Grant OCE0137206 and NOAA Ocean Exploration support to DSK, and grants 96-2116-M002-003 and 97-2752-M004-PAE to C.-C. Shen.

REFERENCES


FIGURE LEGENDS

**Figure 1.** Location within the Lost City Hydrothermal Field of carbonate chimney samples from which sequences were collected. Sample 1 was obtained from the pinnacle of the main edifice known as Poseidon, 50 m above the flange from which samples 2 and 3 were collected. Sample 4 is from a small, isolated chimney with no apparent venting activity found near the bottom of a cliff face. Scale bars in all photographs are 10 cm.

**Figure 2.** Rarefaction analysis of archaeal and bacterial V6 tag sequences for each of the four chimney samples. OTUs: operational taxonomic units defined by clustering sequences with a 3% pairwise distance threshold. Bacterial diversity (curves outlined in black) is clearly greater than archaeal diversity in all samples. The rarefaction curves for archaeal sequences are near-asymptotic, indicating nearly complete sampling of archaeal V6 amplicon libraries. Dashed line indicates the point to which data were subsampled in order to compare diversity values among samples in Table 1.

**Figure 3.** Relative abundance distribution of archaeal OTUs among samples. (a) Bootstrapped phylogenetic tree of full-length archaeal 16S rRNA clones showing the relationship of three Lost City clones (in bold) to previously published close relatives. (b) Relative normalized abundance
(as percentage of total tag sequences per sample) of each archaeal OTU (clustered at 3% distance threshold) in each of the four chimney samples. The OTUs are labeled according to which of the three groups they belong according to pairwise similarity with the V6 region of the full-length 16S rRNA clones. LCMS: Lost City *Methanosarcinales*, ANME-1: anaerobic methanotrophic archaea, MGI: Marine Group I *Crenarchaeota*.

**Figure 4.** Relative abundance distribution of bacterial OTUs among samples. OTUs are labeled according to their taxonomic assignment by the GAST process (22). Bact: *Bacteroides*, Cflx: *Chloroflexus*, Firm: *Firmicutes*, α: *Alphaproteobacteria*, γ: *Gammaproteobacteria*, δ: *Deltaproteobacteria*, ε: *Epsilonproteobacteria*.

**Figure 5.** Relative abundance distribution of *Thiomicrospira*-related OTUs among samples. (a) Bootstrapped phylogenetic tree of full-length 16S rRNA clones showing the relationship of Lost City clones (in bold) to previously published *Thiomicrospira* clones. (b) Relative normalized abundance of each OTU assigned to genus *Thiomicrospira* by GAST (22) in each of the four chimney samples. Pairwise similarities between sequences representative of each OTU and the V6 region of each clone were calculated, and OTUs were grouped according to which clone they were most closely related, as indicated by the A-E labels. In groups A-D, similarities between sequences and clones ranged from 93-100%; in group E, 86-100%. OTUs within each group that have the highest similarity to the clone are sorted to the left.
<table>
<thead>
<tr>
<th>Chimney Sample&lt;sup&gt;1&lt;/sup&gt;</th>
<th>Max Fluid Temp (ºC)</th>
<th>Age (yrs)</th>
<th>OTUs&lt;sup&gt;3&lt;/sup&gt;</th>
<th>ACE richness estimator&lt;sup&gt;4&lt;/sup&gt;</th>
<th>Chao1 richness estimator&lt;sup&gt;4&lt;/sup&gt;</th>
<th>Simpson's Reciprocal Index&lt;sup&gt;4&lt;/sup&gt;</th>
<th>OTUs&lt;sup&gt;3&lt;/sup&gt;</th>
<th>ACE richness estimator&lt;sup&gt;4&lt;/sup&gt;</th>
<th>Chao1 richness estimator&lt;sup&gt;4&lt;/sup&gt;</th>
<th>Simpson's Reciprocal Index&lt;sup&gt;4&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>88</td>
<td>34.3 ± 8.1</td>
<td>342</td>
<td>448 - 568</td>
<td>441 - 599</td>
<td>2.04 - 2.13</td>
<td>177</td>
<td>283 - 388</td>
<td>217 - 319</td>
<td>5.6 – 5.9</td>
</tr>
<tr>
<td>2</td>
<td>70</td>
<td>42.9 ± 23.3</td>
<td>265</td>
<td>336 - 438</td>
<td>310 - 405</td>
<td>1.56 - 1.61</td>
<td>270</td>
<td>458 - 587</td>
<td>360 - 522</td>
<td>8.3 – 9.1</td>
</tr>
<tr>
<td>3</td>
<td>70</td>
<td>128.4 ± 57.2</td>
<td>265</td>
<td>322 - 410</td>
<td>301 - 386</td>
<td>1.64 - 1.69</td>
<td>428</td>
<td>859 – 1,061</td>
<td>648 - 925</td>
<td>20.0</td>
</tr>
<tr>
<td>4</td>
<td>nd</td>
<td>1,245.4 ± 257.2</td>
<td>135</td>
<td>131 - 170</td>
<td>129 - 191</td>
<td>1.22 - 1.25</td>
<td>307</td>
<td>582 - 744</td>
<td>434 - 630</td>
<td>16.7</td>
</tr>
</tbody>
</table>

<sup>1</sup> DSV Alvin numbers for samples 1-4 are 3881-1408, 3869-1404, 3869-1443, and 3876-1133, respectively.

<sup>2</sup> Archaeal diversity is reported for OTU relative abundances normalized down to 16,260 tags; bacterial diversity is reported for OTU relative abundances normalized down to 5,567 tags. Table S1 compares both archaeal and bacteria diversity at 5,567 tags.

<sup>3</sup> OTUs (operational taxonomic units) defined at 3% sequence difference.

<sup>4</sup> Ranges of diversity indices include 95% confidence intervals as calculated by DOTUR (39).

**Table 1.** Comparison of $^{230}$Th ages and V6 tag sequence diversity measurements for carbonate chimney samples from the Lost City Hydrothermal Field.
Sample Names

Samples 1-4 in this study refer to DSV *Alvin* sample names 3881-1408, 3869-1404, 3869-1443, and 3876-1133, respectively. The NCBI Short Read Archive and names of 16S rRNA clones from previous studies (Brazelton et al., 2006) refer to samples 1-4 as LC1408, LC1404, LC1443, and LC1133, respectively. The VAMPS database (http://vamps.mbl.edu) identifies archaeal V6 tag sequences from this study as ICM_LCY_Av6 and bacterial V6 tag sequences as ICM_LCY_Bv6. Samples 1-4 are designated in VAMPS as LCY_0016_2003_05_16, LCY_0014_2003_05_04, LCY_0018_2003_05_04, and LCY_0012_2003_05_11, respectively, and bacterial V6 tag sequences as LCY_0005_2003_05_16, LCY_0003_2003_05_04, LCY_0007_2003_05_04, and LCY_0001_2003_05_11, respectively.

Sample Descriptions

Previous studies have used field observations to classify the LCHF chimney samples based on observable hydrothermal activity (where “active” samples are those egressing hydrothermal fluids and “inactive” chimneys have no apparent venting activity) (Ludwig et al., 2006). Active structures are typically very friable, porous (up to 50%) and are composed of a mix of aragonite (CaCO$_3$), brucite (Mg(OH)$_2$), and some calcite (CaCO$_3$) whereas inactive samples are typically more lithified and are composed of mostly aragonite and calcite (Ludwig et al., 2006). In this study, three samples from actively venting chimneys (samples 1-3) and one sample from an inactive chimney (sample 4) were selected to obtain a set of samples ranging from young to old and ranging from close to distant physical proximity. All samples were collected with DSV *Alvin* and stored in a closed 'biobox' until shipboard retrieval. Samples 2 and 3 were collected on the same dive; other samples were collected on separate
dives. Previously published microbiological and biochemical characteristics of these samples are
summarized in Table S4.

Samples 2 (42.9 ± 23.3 yrs) and 3 (128.4 ± 57.2 yrs) are from a structure named “Marker C” and were
located 20 cm apart (Fig. 1). Marker C is a ~50 cm wide flange structure with several small
(chromometers tall) chimneys growing on the top of the flange (Fig. 1). Sample 2 was collected from the
front of the flange and sample 3 was a small spire collected from the top (Fig. 1). Both structures were
cream white with a reddish discoloration that remains unexplained (Ludwig et al., 2006). These
chimneys were visibly venting, and the mixing of emitting fluids with ambient seawater created a
temperature gradient of 9°C – 70°C from the interior of the flange outward. Porosity of the LCHF
active chimneys ranges from <5-50% and is determined quantitatively using petrographic thin sections
(Ludwig et al., 2006). Only one thin section of a sample in this study (sample 3) was made, which had
a porosity of 33% (Ludwig et al., 2006). Samples representing earlier growth stages generally have
higher porosities than samples representing later growth stages (Ludwig et al., 2006), so it is expected
that samples 1-2 have >33% porosity, and sample 4 has <33% porosity. Visual examination of the
samples supports this hypothesis, but quantitative measurements were not conducted.

Sample 1 (34.3 ± 8.1 yrs) was collected from a site known as “Marker 3” or “Poseidon” (Kelley et al.,
2005). Poseidon is the landmark of the LCHF, towering 60 m above the seafloor, emitting fluids at
temperatures ranging from 55-88°C. Sample 1 minerals appeared bright white in color, very friable,
and not lithified. A mucilaginous substance, most likely a biofilm, coated the sample.

Sample 4 (1,245 ± 257.2 yrs) was collected from a small spire growing from the basement rock on the
eastern side of the field (Fig. 1). This structure was not visibly venting, though it was observed only
once before sampling. It was gray in color, contained some biolitic material and had visible fluid flow
channels, suggesting that some venting may have occurred in the recent past. Coral polyps were found on the exterior of the structure after recovery, which is typical of inactive chimneys at the LCHF (Ludwig et al., 2006). Temperature and chemistry conditions at this site were presumed to be that of ambient seawater, and no measurements were performed.

Fluid chemistry and temperature data from the chimneys at Marker 3 (Sample 1) and Marker C (samples 2-3) do not show any obviously important differences (Proskurowski et al., 2006; 2008). The chimney at Marker 3 appears slightly hotter (88ºC max temp compared to 70ºC max temp), but the in situ temperature of the minerals collected in this study would have been determined by their distance from the fluid source. Sample 3 was most likely at a much cooler temperature than sample 2 because it was further from the chimney flange opening. Concentrations of H2 and carbon species are highly similar at markers 3 and C (Proskurowski et al., 2006; 2008). Chimney mineralogy and 230-Th age, therefore, are expected to be much more important drivers of microbial community composition than are temperature and fluid chemistry, and the data presented in this study is consistent with that hypothesis.

**U/Th dating of carbonate chimneys**

Numerous questions remain concerning the development, formation conditions, and microbial colonization of this novel hydrothermal system. Although previous work using radiocarbon techniques shows that hydrothermal activity has been on-going for at least 30 kyrs, modeling suggests that the field may be even older (Früh-Green et al., 2003). The carbonate composition of the LCHF chimneys (Kelley et al., 2005; Ludwig et al., 2006) is amenable to both radiocarbon and U-series dating, allowing important constraints to be placed on the history of hydrothermal activity and the timescales over which the chimneys form. Uranium-thorium (\(^{238}\text{U}, {^{234}\text{U}, {^{230}\text{Th}, {^{232}\text{Th or U-} {^{230}\text{Th)}} disequilibrium dating is a powerful geochronological tool used to date inorganic and biogenic materials ranging in age
from modern to 600 kyrs (Edwards et al., 1986/87; 1987; 2003). In this study, four discrete chimney samples were dated using $^{238}\text{U}-^{234}\text{U}-^{232}\text{Th}-^{230}\text{Th}$ age dating techniques to compare these ages to the microbial community structure of co-registered samples.

For U-Th shore-based analyses, chimney sub-samples were chemically prepared in the Minnesota Isotope Lab (MIL) using methods described in Edwards et al. (1986/7) and Shen et al. (2003). Approximately 0.2 g of carbonate was weighed in acid-cleaned Teflon beakers, dissolved in HNO$_3$, then spiked (Chen and Wasserburg, 1981; Chen et al., 1986). After adding 5 drops of HClO$_4$, the samples were capped and heated for 4-6 hours to remove organics and equilibrate the spike with the sample. Uranium and Th aliquots were separated using Fe co-precipitation and ion chromatography, dissolved in 1% HNO$_3$ + 0.005 N HF, and then stored in acid-cleaned plastic ICP-vials. Procedural blanks were measured regularly and three-month average values were 0.02 ± 0.01 pmol $^{238}\text{U}$, 0.003 ± 0.003 pmol $^{232}\text{Th}$, and 0.0006 ± 0.0005 fmol $^{230}\text{Th}$.

All samples were analyzed on an ICP-SF-MS using methods described by Shen et al. (2002). Data reduction was completed off-line as described by Cheng et al. (2000) and Shen et al. (2002). Ages were calculated iteratively as described by Edwards et al. (1987) and Shen et al., (2002). Values for decay constants $\lambda_{230}$ ($9.1577 \times 10^{-6}$ yr$^{-1}$) and $\lambda_{234}$ ($2.8263 \times 10^{-6}$ yr$^{-1}$) are from Cheng et al. (2000) and $\lambda_{238}$ ($1.5513 \times 10^{-10}$ yr$^{-1}$) is from Jaffey et al. (1971). Measured errors of isotopic and concentration are given as 2 standard deviation of the mean ($2\sigma_m$) and age precisions are reported as 2 standard deviations ($2\sigma$). Ages of subsamples from the same large sample were highly similar (K. Ludwig doctoral dissertation; manuscript in preparation), indicating that ages of the four samples in this study are representative of the bulk material used for DNA extraction.
DNA extraction from carbonate chimneys

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: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.

Generation and sequencing of V6 amplicons

Archaeal and bacterial V6 amplicons were generated as previously described (Sogin et al., 2006; Huber et al., 2007) and summarized here. PCR reactions utilized Platinum Taq HiFi polymerase (Invitrogen) to reduce errors introduced during amplification and involved the following cycling conditions: 94°C
for 2 min; 30 cycles of (94°C for 30s, 50°C for 20s, 72°C for 1min); and 72°C for 2min. A cocktail of degenerate primers were used to maximize taxonomic coverage, as described by Huber et al. (2007). Amplicon concentrations were quantitated with a BioAnalyzer 2100 (Agilent Technologies) and normalized among samples before sequencing. Amplicons from each sample were attached to beads and sequenced with a 454 Life Sciences GS20 pyrosequencer according to previously published protocols (Margulies et al., 2005; Sogin et al., 2006; Huber et al., 2007). Archaeal data was also generated with a 454 Life Sciences FLX pyrosequencer because the number of archaeal sequences generated by the GS20 were highly variable among samples. Sequences generated by both machines were pooled together to improve OTU clustering robustness, but comparisons of archaeal diversity among samples considered only FLX data. Sequences were screened for quality as recommended by Huse et al. (2007). Specifically, all sequences satisfying any of the following criteria were not considered for further analysis: those with one or more Ns, those without an exact match to one of the forward primers, or those shorter than 50 nts after removal of primers. Automated taxonomic assignments of V6 tag sequences were performed by the GAST process, which compares V6 tags to a reference database of V6 regions from full-length 16S rRNA clones of known taxonomy (Huse et al., 2008). GAST requires two-thirds of the sequences of a given taxonomy to match the query tag sequence in order to make an assignment. If no taxonomic category fulfills the two-thirds rule, it is reported as “unassigned” (Tables S2 and S3). As the taxonomy of uncultured organisms is inconsistently reported in GenBank, GAST is frequently unable to reliably assign taxonomy to groups of organisms without cultured representatives (e.g. ANME-1 in Table S2).

Estimation of pyrosequencing error

Each of the three extremely dominant sequence types representing the main archaeal groups was one member of a large, unexpectedly diverse pool of mostly rare sequences (Figure 3b). For example, the four chimney samples contained 1,242 different V6 sequence types that were 3% distant or less from
the dominant sequence, representing nearly half of all 2,635 sequence types identified in this study. A further 530 sequence types were between 3% and 10% distant from the dominant sequence. Only 63 sequence types were more than 10% distant to the dominant sequence. (Together, these 1,835 sequence types clustered into the 534 OTUs displayed in Fig. 3). Similar patterns were seen with sequences related to the ANME-1 and Marine Group I clones: 403 different V6 sequence types are 3% distant or less from the dominant ANME-1 sequence, and 125 sequence types are 3% distant or less from the dominant Marine Group I sequence. Although pyrosequencing error may have generated some of the variant sequences within clusters, pyrosequencing error alone cannot account for all of the observed diversity. If the variant sequences within each major cluster are compared to that cluster's dominant sequence, a nucleotide substitution rate of 0.18% is calculated for the LCMS cluster, 0.08% for the ANME-1 cluster, and 0.76% for the Marine Group I cluster. (The much higher substitution rate for the Marine Group I cluster is due to the relatively high abundance of two variant sequences compared to the dominant sequence.) These values are much higher than the substitution rate of 0.03% attributed to pyrosequencing error of quality-screened tags by Huse et al. (11). Therefore, some of the most abundant variant sequences are extremely unlikely to be generated by pyrosequencing error and most likely represent genuine diversity within the phylogenetic group.

**Determination of variability due to amplicon generation and sequencing**

Because the three young chimney samples share the same dominant LCMS V6 sequence type, nearly all community differences in the presence/absence of OTUs among samples involve rare sequences. It is possible that differences involving rare sequences could be a trivial result due to inherent variability of the technique. We tested this possibility by generating and sequencing replicate amplicon libraries from the same DNA preparation of sample 1 in order to test whether the rare OTU membership was reproduced. Because the replicate sequencing runs were performed months apart, with independent
amplifications from template, and on different pyrosequencing machines (454 Life Sciences GS20 and FLX), we consider this a very conservative test of replication. The GS20 replicate yielded 17,425 tag sequences and 343 OTUs. The FLX replicate yielded 16,260 tag sequences and 342 OTUs. The GS20 OTUs were randomly resampled down to 16,260 sequences, reducing the number of OTUs to 286. The Bray-Curtis similarity between the OTUs of the sample 1 replicates was high (89%) due to the nearly identical relative abundances of the most common OTUs. As expected, incomplete sampling of the amplicon library caused differences in the presence/absence of the many rare OTUs so that the Jaccard similarity between replicates was just 50%. (When sequence types represented by only one tag ('singletons') among all samples were removed from analysis before normalization and clustering, the Jaccard similarity improved to 73%.) Because the similarity between replicates was higher than the similarity between sample 1 and the other samples, the differences among samples are greater than that expected from the technique itself. A remarkable exception is the archaeal community similarity between samples 2 and 3 (95% Bray-Curtis, 56% Jaccard). The similarity was much higher than the similarity between sample 1 replicates, so the small differences between these samples may not reflect any natural variation. Therefore, the archaeal communities of samples 2 and 3 can be considered identical. It should be noted, however, that samples 2 and 3 were both sequenced on the same machine (FLX), while the sample 1 replicates were sequenced on different machines, so the actual experimental error between samples sequenced on the same machine is probably much lower than that represented by the sample 1 replicates.

Comparison with clone libraries

The nearly full-length sequences of 16S rRNA clones to which V6 tag sequences were compared have been previously published (Brazelton et al., 2006; Brazelton and Baross, 2009). Global alignment sequence similarities between V6 tags and the corresponding region of 16S rRNA clones were calculated with MatGat (Campanella 2003). Phylogenetic trees of 16S rRNA clones were constructed
with TreePuzzle (Version 5.2, Schmidt 2002) using default options and exact parameter estimation. GenBank accession numbers for sequences in Fig. 3 tree, from top to bottom: M59133, FJ791573, M59140, AY817738, AM746096, NC_010085, FJ795512, U51469. GenBank accession numbers for sequences in Fig. 5 tree, from top to bottom: AF329082, AF016046, DQ270609, FJ792098, AF013974, DQ270608, DQ270607, FJ792484, AJ404732, AB166731, L40810.

References


17. Margulies, M. *et al.* (2005) Genome sequencing in microfabricated high-density picolitre


<table>
<thead>
<tr>
<th></th>
<th>Archaea</th>
<th>Bacteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>V6 tag sequences</td>
<td>167,031</td>
<td>43,027</td>
</tr>
<tr>
<td>unique sequence types</td>
<td>2,635 (1,163)</td>
<td>2,082</td>
</tr>
<tr>
<td>OTUs (3% distance)</td>
<td>817 (444)</td>
<td>1,135</td>
</tr>
<tr>
<td>ACE (3% distance)</td>
<td>980 – 1,024 (543 - 642)</td>
<td>2,203 – 2,362</td>
</tr>
<tr>
<td>Chao1 (3% distance)</td>
<td>943 - 1078 (524 - 641)</td>
<td>1,692 – 2,048</td>
</tr>
<tr>
<td>Simpson's Reciprocal Index (3%</td>
<td>2.3 (2.3)</td>
<td>12.5</td>
</tr>
<tr>
<td>distance)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>OTUs (10% distance)</td>
<td>116 (81)</td>
<td>653</td>
</tr>
<tr>
<td>ACE (10% distance)</td>
<td>142 - 191 (85 - 111)</td>
<td>1,215 – 1,294</td>
</tr>
<tr>
<td>Chao1 (10% distance)</td>
<td>130 - 246 (83 - 124)</td>
<td>922 - 1189</td>
</tr>
<tr>
<td>Simpson's Reciprocal Index (10%</td>
<td>2.04 (2.04)</td>
<td>11.1</td>
</tr>
<tr>
<td>distance)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Table S1.** Sequencing depth (as number of V6 tag sequences), observed number of OTUs (operational taxonomic units), richness estimators (ACE and Chao1), and Simpson’s Reciprocal Index of diversity for all archaeal and bacterial OTUs in this study. Values in parentheses are derived from randomly resampling the 167,031 archaeal tags down to 43,027 tags in order to compare archaeal diversity with bacterial diversity at equal sequencing efforts. The 167,031 archaeal tags includes 139,086 tags from the 454 Life Sciences FLX pyrosequencer and 27,945 tags from the 454 Life Sciences GS20 pyrosequencer.
<table>
<thead>
<tr>
<th>Phylum</th>
<th>Class</th>
<th>Order</th>
<th>Sample 1 (34 yrs)</th>
<th>Sample 2 (43 yrs)</th>
<th>Sample 3 (128 yrs)</th>
<th>Sample 4 (1245 yrs)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Tags</td>
<td>OTUs</td>
<td>Tags</td>
<td>OTUs</td>
</tr>
<tr>
<td>Crenarchaeota</td>
<td>MBGA</td>
<td>unassigned</td>
<td>3</td>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Crenarchaeota</td>
<td>MGI</td>
<td>unassigned</td>
<td>5</td>
<td>3</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Crenarchaeota</td>
<td>Thermoprotei</td>
<td>unassigned</td>
<td>45</td>
<td>11</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Crenarchaeota</td>
<td>unassigned</td>
<td>unassigned</td>
<td>259</td>
<td>27</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Euryarchaeota</td>
<td>Methanomicrobia</td>
<td>Methanosarcinales</td>
<td>14,689</td>
<td>281</td>
<td>16,258</td>
<td>264</td>
</tr>
<tr>
<td>Euryarchaeota</td>
<td>Methanomicrobia</td>
<td>Methanosetaeaceae</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Euryarchaeota</td>
<td>misc.</td>
<td></td>
<td>4</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Euryarchaeota</td>
<td>unassigned</td>
<td>unassigned</td>
<td>53</td>
<td>15</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Korarchaeota</td>
<td>unassigned</td>
<td>unassigned</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>unassigned</td>
<td>unassigned</td>
<td>unassigned</td>
<td>1020</td>
<td>1</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Archaea total</td>
<td></td>
<td></td>
<td>16,260</td>
<td>342</td>
<td>16,260</td>
<td>265</td>
</tr>
</tbody>
</table>

**Table S2.** Numbers of archaeal tag sequences and OTUs clustered with a 3% distance threshold for the most frequently detected phyla, classes, and orders. Taxonomies were assigned with the GAST algorithm (Huse et al. 2008). Numbers of tag sequences have been normalized down to a total of 16,260 sequences per sample by random resampling after OTU clustering. Pre-normalized total numbers of tag sequences were 16,260 (sample 1); 32,345 (sample 2); 25,471 (sample 3); and 21,983 (sample 4). Most of the OTUs labeled ‘LCMS’ in Fig. 3 were assigned to order Methanosarcinales; most ANME-1 OTUs were assigned to Euryarchaeota unassigned; and most MGI OTUs were assigned to ‘Crenarchaeota unassigned’ or ‘unassigned’.
<table>
<thead>
<tr>
<th>Phylum</th>
<th>Class</th>
<th>Order</th>
<th>Sample 1 (34 yrs)</th>
<th>Sample 2 (43 yrs)</th>
<th>Sample 3 (128 yrs)</th>
<th>Sample 4 (1245 yrs)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acidobacteria</td>
<td>Acidobacteria</td>
<td>Acidobacteriales</td>
<td>0 0</td>
<td>0 0</td>
<td>10 7</td>
<td>3 2</td>
</tr>
<tr>
<td>Acidobacteria</td>
<td>unassigned</td>
<td>unassigned</td>
<td>0 0</td>
<td>0 0</td>
<td>1 1</td>
<td>0 0</td>
</tr>
<tr>
<td>Actinobacteria</td>
<td>Actinobacteria</td>
<td>misc.</td>
<td>0 0</td>
<td>0 0</td>
<td>12 5</td>
<td>23 9</td>
</tr>
<tr>
<td>Actinobacteria</td>
<td>unassigned</td>
<td>unassigned</td>
<td>0 0</td>
<td>0 0</td>
<td>3 1</td>
<td>14 2</td>
</tr>
<tr>
<td>Bacteroidetes</td>
<td>Bacteroidia</td>
<td>Bacteroidales</td>
<td>0 0</td>
<td>0 0</td>
<td>3 1</td>
<td>1 1</td>
</tr>
<tr>
<td>Bacteroidetes</td>
<td>Flavobacteria</td>
<td>Flavobacteriales</td>
<td>27 7</td>
<td>103 18</td>
<td>301 37</td>
<td>40 10</td>
</tr>
<tr>
<td>Bacteroidetes</td>
<td>Sphingobacteria</td>
<td>Sphingobacteriales</td>
<td>42 4</td>
<td>29 8</td>
<td>39 21</td>
<td>11 9</td>
</tr>
<tr>
<td>Bacteroidetes</td>
<td>unassigned</td>
<td>unassigned</td>
<td>36 2</td>
<td>11 4</td>
<td>48 12</td>
<td>3 1</td>
</tr>
<tr>
<td>BRC1</td>
<td>unassigned</td>
<td>unassigned</td>
<td>0 0</td>
<td>0 0</td>
<td>89 5</td>
<td>163 9</td>
</tr>
<tr>
<td>Chloroflexi</td>
<td>Anaerolineae</td>
<td>unassigned</td>
<td>2 1</td>
<td>0 0</td>
<td>0 0</td>
<td>0 0</td>
</tr>
<tr>
<td>Chloroflexi</td>
<td>Caldilineae</td>
<td>Caldilineales</td>
<td>1 1</td>
<td>3 1</td>
<td>11 1</td>
<td>1 1</td>
</tr>
<tr>
<td>Chloroflexi</td>
<td>Dehalococoidetes</td>
<td>unassigned</td>
<td>0 0</td>
<td>0 0</td>
<td>1083 17</td>
<td>518 8</td>
</tr>
<tr>
<td>Deferribacteres</td>
<td>Deferribacteres</td>
<td>Deferribacteriales</td>
<td>5 1</td>
<td>0 0</td>
<td>7 2</td>
<td>1 1</td>
</tr>
<tr>
<td>Deferribacteres</td>
<td>unassigned</td>
<td>unassigned</td>
<td>0 0</td>
<td>0 0</td>
<td>1 1</td>
<td>0 0</td>
</tr>
<tr>
<td>Deinococcus-Thermus</td>
<td>Deinococci</td>
<td>misc.</td>
<td>31 4</td>
<td>0 0</td>
<td>0 0</td>
<td>0 0</td>
</tr>
<tr>
<td>Firmicutes</td>
<td>Bacilli</td>
<td>misc.</td>
<td>0 0</td>
<td>1 1</td>
<td>2 1</td>
<td>0 0</td>
</tr>
<tr>
<td>Firmicutes</td>
<td>Clostridia</td>
<td>misc.</td>
<td>1073 42</td>
<td>426 31</td>
<td>228 21</td>
<td>57 9</td>
</tr>
<tr>
<td>Nitrospira</td>
<td>Nitrospira</td>
<td>Nitrospirales</td>
<td>0 0</td>
<td>3 1</td>
<td>601 8</td>
<td>83 3</td>
</tr>
<tr>
<td>OD1</td>
<td>unassigned</td>
<td>unassigned</td>
<td>0 0</td>
<td>8 2</td>
<td>62 10</td>
<td>10 5</td>
</tr>
<tr>
<td>Proteobacteria</td>
<td>Alphaproteobacteria</td>
<td>misc.</td>
<td>15 3</td>
<td>24 10</td>
<td>14 9</td>
<td>9 7</td>
</tr>
<tr>
<td>Proteobacteria</td>
<td>Alphaproteobacteria</td>
<td>Rhizobiales</td>
<td>121 3</td>
<td>15 6</td>
<td>64 6</td>
<td>45 13</td>
</tr>
<tr>
<td>Proteobacteria</td>
<td>Alphaproteobacteria</td>
<td>Rhodobacteriales</td>
<td>273 19</td>
<td>1294 40</td>
<td>545 39</td>
<td>2146 50</td>
</tr>
<tr>
<td>Proteobacteria</td>
<td>Alphaproteobacteria</td>
<td>unassigned</td>
<td>102 9</td>
<td>108 7</td>
<td>149 25</td>
<td>140 16</td>
</tr>
<tr>
<td>Phylum</td>
<td>Class</td>
<td>Order</td>
<td>Sample 1</td>
<td>Sample 2</td>
<td>Sample 3</td>
<td>Sample 4</td>
</tr>
<tr>
<td>-------------------------------</td>
<td>-------------------------------</td>
<td>------------------------------------</td>
<td>----------</td>
<td>----------</td>
<td>----------</td>
<td>----------</td>
</tr>
<tr>
<td>Proteobacteria</td>
<td>Betaproteobacteria</td>
<td>Burkholderiales</td>
<td>15</td>
<td>3</td>
<td>11</td>
<td>5</td>
</tr>
<tr>
<td>Proteobacteria</td>
<td>Betaproteobacteria</td>
<td>misc.</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>Proteobacteria</td>
<td>Deltaproteobacteria</td>
<td>Desulfobacterales</td>
<td>5</td>
<td>2</td>
<td>52</td>
<td>9</td>
</tr>
<tr>
<td>Proteobacteria</td>
<td>Deltaproteobacteria</td>
<td>Desulfovibrionales</td>
<td>15</td>
<td>5</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Proteobacteria</td>
<td>Deltaproteobacteria</td>
<td>misc.</td>
<td>0</td>
<td>0</td>
<td>14</td>
<td>6</td>
</tr>
<tr>
<td>Proteobacteria</td>
<td>Epsilonproteobacteria</td>
<td>Campylobacterales</td>
<td>158</td>
<td>9</td>
<td>29</td>
<td>9</td>
</tr>
<tr>
<td>Proteobacteria</td>
<td>Epsilonproteobacteria</td>
<td>Nautiliales</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Proteobacteria</td>
<td>Gammaproteobacteria</td>
<td>Methylcoccales</td>
<td>0</td>
<td>0</td>
<td>54</td>
<td>3</td>
</tr>
<tr>
<td>Proteobacteria</td>
<td>Gammaproteobacteria</td>
<td>Thiotrichales</td>
<td>3483</td>
<td>42</td>
<td>3122</td>
<td>56</td>
</tr>
<tr>
<td>Proteobacteria</td>
<td>Gammaproteobacteria</td>
<td>misc.</td>
<td>114</td>
<td>10</td>
<td>153</td>
<td>21</td>
</tr>
<tr>
<td>Proteobacteria</td>
<td>unassigned</td>
<td>unassigned</td>
<td>2</td>
<td>2</td>
<td>52</td>
<td>11</td>
</tr>
<tr>
<td>Tenericutes</td>
<td>Mollicutes</td>
<td>Acholeplasmatales</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Thermomicrobia</td>
<td>unassigned</td>
<td>unassigned</td>
<td>2</td>
<td>1</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>misc. assigned¹</td>
<td></td>
<td></td>
<td>42</td>
<td>6</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>unassigned</td>
<td></td>
<td></td>
<td>3</td>
<td>1</td>
<td>36</td>
<td>6</td>
</tr>
<tr>
<td>Bacteria Total</td>
<td></td>
<td></td>
<td>5567</td>
<td>177</td>
<td>5567</td>
<td>270</td>
</tr>
</tbody>
</table>

¹misc. assigned includes: Chlamydiae, Chlorobi, Cyanobacteria, Fibrobacteres, Fusobacteria, Gemmatimonadetes, Lentisphaerae, Planctomycetes, Spirochaetes, Thermotogae, TM7, and Verrucomicrobia

**Table S3.** Numbers of bacterial tag sequences and OTUs clustered with a 3% distance threshold for the most frequently detected phyla, classes, and orders. Taxonomies were assigned with the GAST algorithm (Huse et al. 2008). Numbers of tag sequences have been normalized down to a total of 5567 sequences per sample by random resampling after OTU clustering. Pre-normalized total numbers of tag sequences were 21,582 (sample 1); 5567 (sample 2); 7162 (sample 3); and 8716 (sample 4). All of the *Thiomicrospira* sequences in Fig. 6 are included under Thiotrichales here. Most of the ‘misc’ Clostridia sequences have high sequence similarity to previously published *Desulfotomaculum*-like clones (Brazelton et al. 2006).
Table S4. Previously published microbiological and biochemical characteristics of the four carbonate chimney samples. 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).

<table>
<thead>
<tr>
<th>Sample</th>
<th>Cells g⁻¹ dry weightᵃ</th>
<th>Archaeaᵇ</th>
<th>Bacteriaᵇ</th>
<th>LCMSᵇ</th>
<th>Total organic carbon (%)</th>
<th>δ¹³C_toe (‰ vs. VPDB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample 1 (3881-1408)</td>
<td>2.0 x 10⁸</td>
<td>25</td>
<td>14</td>
<td>18</td>
<td>n.d.</td>
<td>n.d.</td>
</tr>
<tr>
<td>Sample 2 (3869-1404)</td>
<td>1200 x 10⁸</td>
<td>41</td>
<td>8</td>
<td>32</td>
<td>0.20</td>
<td>-7.8</td>
</tr>
<tr>
<td>Sample 3 (3869-1443)</td>
<td>1600 x 10⁸</td>
<td>38</td>
<td>10</td>
<td>21</td>
<td>n.d.</td>
<td>n.d.</td>
</tr>
<tr>
<td>Sample 4 (3876-1133)</td>
<td>1.6 x 10⁸</td>
<td>24</td>
<td>19</td>
<td>12</td>
<td>0.15</td>
<td>-16.3</td>
</tr>
</tbody>
</table>

ᵃ Determined by DAPI-staining
ᵇ Percentage of DAPI-stained cells detected by FISH probe specific to each group
SUPPLEMENTARY INFORMATION FIGURE CAPTIONS

**Figure S1.** Archaeal diversity is much lower than bacterial diversity at higher genetic distances. The number of archaeal and bacterial operational taxonomic units (OTUs) clustered at different pairwise distance thresholds are plotted for sample 2. At small distances, similar numbers of archaeal and bacterial OTUs were discovered. At large distances, very few archaeal OTUs were discovered, indicating very low diversity at high taxonomic levels (*e.g.* order, class, phylum).

**Figure S2.** Venn diagrams of archaeal and bacterial OTUs clustered with a 3% distance threshold showing the percentage of all tag sequences and the numbers of OTUs shared by the four chimney samples.

**Figure S3.** Pairwise whole archaeal and bacterial community similarities between samples considering numbers of tag sequences in each OTU (Bray-Curtis) and considering only the presence/absence of each OTU (Jaccard).

**Figure S4.** Relative abundance of each bacterial OTU (clustered with a 3% distance threshold) in each of the four chimney samples and the resulting pairwise whole bacterial community similarities between samples. OTUs occurring at different relative abundance scales are displayed separately: (a) OTUs containing >2% of at least one sample's total tag sequences (b) OTUs containing 0.2% and 2% of a sample's total tag sequences (c) OTUs containing fewer than 0.2% of all samples' total tag sequences. Community
similarities between samples remain fairly constant regardless of whether abundant or very rare OTUs are examined. The highest community similarities are typically between the two youngest samples and the two oldest samples (highlighted in red when they represent the two highest similarity values). OTUs are labeled according to their taxonomic assignment by GAST (Huse et al., 2008). B: Bacteroides, C: Chloroflexus, F: Firmicutes, N: Nitrospira, A: Alphaproteobacteria, D: Deltaproteobacteria, E: Epsilonproteobacteria, G: Gammaproteobacteria, U: Unassigned.

**Figure S5.** Venn diagrams of bacterial OTUs (clustered with a 3% distance threshold) showing the percentage of all tag sequences and the numbers of OTUs shared by the four chimney samples. Separate Venn diagrams are shown for each of the relative abundance scales shown in Figure S4.

**Figure S6.** Examples of OTUs that are dominant in old chimney samples and present but rare in young chimney samples. OTUs are labeled according to their taxonomic assignment. Bact: Bacteroides, Firm: Firmicutes, Nitr: Nitrospira, α: Alphaproteobacteria, β: Betaproteobacteria, δ: Deltaproteobacteria, γ: Gammaproteobacteria.
Figure S2

% tags

Archaea

Bacteria

# OTUs

sample 4 (1245 yrs)
sample 3 (128 yrs)
sample 2 (43 yrs)
sample 1 (34 yrs)
**Figure S3**

**Archaeal V6 tag community similarities between samples**

<table>
<thead>
<tr>
<th></th>
<th>Bray-Curtis similarity</th>
<th></th>
<th></th>
<th></th>
<th>Jaccard similarity</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>128 yrs</td>
<td>43 yrs</td>
<td>34 yrs</td>
<td></td>
<td>sample 4 (1245 yrs)</td>
<td>128 yrs</td>
<td>43 yrs</td>
<td>34 yrs</td>
</tr>
<tr>
<td>sample 4 (1245 yrs)</td>
<td>0.2</td>
<td>0.1</td>
<td>0.2</td>
<td></td>
<td>sample 3 (128 yrs)</td>
<td>1.9</td>
<td>1.1</td>
<td>1.5</td>
</tr>
<tr>
<td>sample 3 (128 yrs)</td>
<td>95</td>
<td>87</td>
<td>85</td>
<td></td>
<td>sample 2 (43 yrs)</td>
<td>56</td>
<td>48</td>
<td>50</td>
</tr>
<tr>
<td>sample 2 (43 yrs)</td>
<td>85</td>
<td>85</td>
<td>85</td>
<td></td>
<td>sample 1 (34 yrs)</td>
<td>45.0</td>
<td>45.0</td>
<td>45.0</td>
</tr>
<tr>
<td>sample 1 (34 yrs)</td>
<td>36.1</td>
<td>37.7</td>
<td>12.1</td>
<td></td>
<td></td>
<td>23.0</td>
<td>16.7</td>
<td>8.8</td>
</tr>
</tbody>
</table>

**Bacterial V6 tag community similarities between samples**

<table>
<thead>
<tr>
<th></th>
<th>Bray-Curtis similarity</th>
<th></th>
<th></th>
<th></th>
<th>Jaccard similarity</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>128 yrs</td>
<td>43 yrs</td>
<td>34 yrs</td>
<td></td>
<td>sample 4 (1245 yrs)</td>
<td>128 yrs</td>
<td>43 yrs</td>
<td>34 yrs</td>
</tr>
<tr>
<td>sample 4 (1245 yrs)</td>
<td>36.1</td>
<td>37.7</td>
<td>12.1</td>
<td></td>
<td>sample 3 (128 yrs)</td>
<td>23.0</td>
<td>16.7</td>
<td>8.8</td>
</tr>
<tr>
<td>sample 3 (128 yrs)</td>
<td>17.8</td>
<td>6.5</td>
<td>45.0</td>
<td></td>
<td>sample 2 (43 yrs)</td>
<td>17.4</td>
<td>8.3</td>
<td>23.6</td>
</tr>
<tr>
<td>sample 2 (43 yrs)</td>
<td>45.0</td>
<td>45.0</td>
<td>45.0</td>
<td></td>
<td>sample 1 (34 yrs)</td>
<td>23.6</td>
<td>23.6</td>
<td>23.6</td>
</tr>
<tr>
<td>sample 1 (34 yrs)</td>
<td>17.4</td>
<td>17.4</td>
<td>17.4</td>
<td></td>
<td></td>
<td>23.6</td>
<td>23.6</td>
<td>23.6</td>
</tr>
</tbody>
</table>
Figure S4

% total tags per sample

21 OTUs

116 OTUs

998 OTUs

a

b

c

Jaccard

Bray-Curtis
Figure S5

>2% total tags per sample

0.2%-2% total tags per sample

<0.2% total tags per sample