Chapter 11
Statistical Evaluation of Bacterial 16S rRNA Gene Sequences in Relation to Travertine Mineral Precipitation and Water Chemistry at Mammoth Hot Springs, Yellowstone National Park, USA

Héctor García Martín, John Veysey, George T. Bonheyo, Nigel Goldenfeld, and Bruce W. Fouke

It is possible that common earth-surface geological features can arise as a result of bacteria interacting with purely physical and chemical processes. The ability to distinguish ancient and modern mineral deposits that are biologically influenced from those that are purely abiotic in origin will advance our ability to interpret microbial evolution from the ancient rock record on earth and potentially other planets. As a step toward deciphering biotic from abiotic processes, we have combined

H.G. Martín, J. Veysey, and N. Goldenfeld
Department of Physics, University of Illinois at Urbana-Champaign, 1110 West Green Street, Urbana, IL 61801-3080, USA

G.T. Bonheyo and B.W. Fouke (✉)
Department of Geology, University of Illinois at Urbana-Champaign, 1301 West Green Street, Urbana, IL 61801-2938, USA
e-mail: fouke@illinois.edu

B.W. Fouke
Department of Microbiology, University of Illinois Urbana-Champaign, 601 S. Goodwin Avenue, Urbana, IL 61801, USA

N. Goldenfeld and B.W. Fouke
University of Illinois Urbana-Champaign, Institute for Genomic Biology, 1206 W. Gregory Drive, Urbana, IL 61801, USA

H.G. Martín
Joint BioEnergy Institute, Emeryville, CA 94608, USA
and
Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94710, USA

G.T. Bonheyo
Pacific Northwest Laboratory, Marine Sciences Laboratory, 1529 W. Sequim Bay Rd, Sequim, WA 98382, USA

L.L. Barton et al. (eds.), *Geomicrobiology: Molecular and Environmental Perspective*,
DOI 10.1007/978-90-481-9204-5_11,

carbonate mineralogical and geochemical analyses together with community-based microbial genetic analyses in hot spring drainage systems at Mammoth Hot Springs in Yellowstone National Park. Previously (Fouke et al. 2000, 2003), we reported the shape and chemistry of carbonate mineral deposits (*travertine*), which have formed along the hot spring outflow. This travertine exhibits five distinct ecological zonations (termed sedimentary depositional *facies*) even though most physical and chemical attributes of the spring water change smoothly and continuously over the course of the drainage outflow path. Here, we document an unexpectedly sharp correlation between microbial phylogenetic diversity and travertine facies, which suggests that changes in bacterial community composition are a sensitive indicator of environmental conditions along the spring outflow. These results provide an environmental context for constraining abiotic and biotic theories for the origin of distinct crystalline structures and chemistries formed during hot spring travertine precipitation.

In order to quantitatively track and identify the mechanisms and products of microbial fossilization during calcium carbonate precipitation, we have initiated a biocomplexity study of microbe-mineral-environmental interactions at Mammoth Hot Springs in Yellowstone National Park. The goal is to determine whether microbial community structure and activity directly influence the precipitation of high-temperature terrestrial calcium carbonate mineral deposits, called *travertine* (Ford and Pedley 1996; Pentecost 2005). Initially, we determined that the Mammoth spring drainage systems are composed of five ecological partitions (called *sedimentary facies*), an analysis based solely on the shape, structure and chemical composition of the travertine mineral deposits (Fouke et al. 2000). Analyses of spring water chemistry were then integrated with the travertine facies along the drainage system to establish water-mineral precipitation baselines defining the bulk system-level chemical evolution of the spring outflow system (Fouke et al. 2000). The next phase of the research was to spatially map the ecological distribution of bacteria with respect to the travertine facies in which they live, using what have now become standard culture-independent molecular analyses of 16S rRNA gene sequences (Fouke et al. 2003).

The present study is a rigorous statistical evaluation of these bacterial 16S rRNA gene sequences, which is the first such analysis completed within a natural environmental setting. We had not yet developed nor applied these statistical approaches at the time of the publication of the original clone libraries (Fouke et al. 2003). However, the results presented in the present paper are essential in that they provide the first quantitative validation of: (1) the completeness and randomness of these types of molecular microbial analyses in the environment; and (2) the system-scale correlation of microbial community composition with calcium carbonate mineral precipitation, which is the critical ecological relationship required to begin to identify the dynamics of microbial influence on carbonate mineral precipitation. These statistical approaches are universally applicable to other molecular studies of microbial ecology and will therefore be a valuable tool for studying biocomplex dynamics in the environment.

Geological Setting of Mammoth Hot Springs

Subsurface waters erupt at Mammoth Hot Springs to precipitate terraced crystalline deposits, called travertine, which is composed of aragonite and calcite (Friedman 1970; Sorey 1991). Our studies were conducted at Spring AT-1 (Fouke et al. 2000), located on Angel Terrace, in the upper terrace region of the Mammoth Hot Springs complex. Spring AT-1 is typical of the hot springs found at Mammoth Hot Springs, in that as the spring water flows away from the subsurface vent, the water cools, degases CO_2, increases in pH, and precipitates travertine that steadily changes composition from nearly 100% aragonite to nearly 100% calcite. Precipitation rates are rapid (Kandianis et al. 2008) and can reach 5 mm/day. The rapid precipitation partially seals the vents and reroutes surface outflow, causing the spring flow path to regularly change in direction and intensity, which in turn influences subsequent travertine precipitation. The dynamical interplay between fluid flow and travertine precipitation, be it primarily biotic or abiotic in origin is complex and not yet understood. The hot springs harbor diverse communities of microorganisms, representing at least 21 divisions of bacteria (Fouke et al. 2003).

In order to analyze the physical, geological, and biological aspects of this rapidly changing hydrothermal system, we first subdivided the spring drainage system into a series of recognizable sub-environments or ecological partitions along the flow path. These sub-environments are known as sedimentary depositional *facies*. A facies is defined as a sedimentary rock deposit that represents the sum total of physical, chemical, geological, and biological processes active in a natural environment of sediment deposition and mineral accumulation. Each facies has its own distinct mineralogical and hydrological features and may therefore be readily identified, even if the overall drainage system significantly changes and migrates. Our previous work defined a five-component travertine facies model for Spring AT-1. This is based on physical and chemical characteristics of the spring water (temperature, pH, elemental and isotopic chemistry) and associated travertine (crystalline growth form and fabric, mineralogy, elemental and isotopic chemistry), quantitative modeling of this aqueous and solid data, and limited microscopic observations of the microbiology (Fig. 11.1). The following is a brief summary of these five facies, called the vent, apron and channel, pond, proximal slope, and distal slope (Fouke et al. 2000).

The facies model allows equivalent ecological locations in the spring drainage systems to be analyzed over time, despite nearly constant changes in the rate or direction of spring flow, and thus allows comparisons to be made between springs in different geographic locations and of different geological ages. Remarkably, we find that the physical structures characteristic of each facies develop sharp boundaries instead of gradual transitional zones (Fouke et al. 2000). Although any given travertine facies may be as much as tens of meters long and cover hundreds of square meters in area, the boundary between facies is relatively abrupt, occurring over as little as 1 cm in distance between the pond (1–3 m in length along the spring flowpath) and proximal slope (10–15 m in length) or up to 10 cm between the proximal slope and distal slope (10–15 m in length).

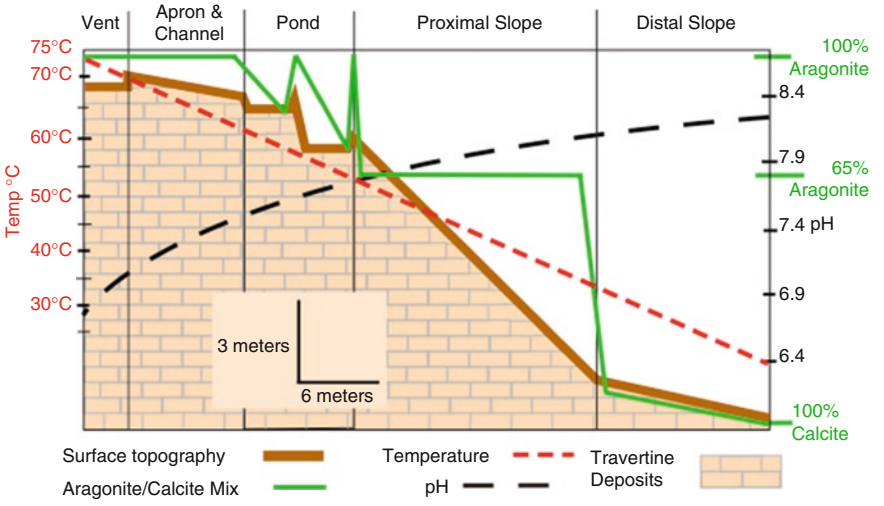

Fig. 11.1 Facies model. Cross-sectional view of Spring AT-1 with 2 × vertical exaggeration to highlight the topography of the spring features. Trends in pH, temperature and travertine aragonite/calcite mineralogical ratios overlay the structural representation to show how these attributes change with increasing distance from the spring outflow source vent

The aqueous chemistry of the hot spring drainage system is dominated by CO_2 degassing and dropping temperature as proven by Rayleigh-type fractionation calculations of spring water dissolved inorganic carbon (DIC) and its associated $\delta^{13}C$ versus (Fouke et al. 2000). While these physical factors help drive the rapid precipitation of carbonate crystals to deposit travertine at rates as high as 5 mm/day at the pond lip, they are not the exclusive controls on precipitation. Significant biological controls on travertine crystal form and isotope chemistry have been identified where travertine crystals entomb and preserve the shape of filamentous *Aquificales* bacteria, and by quantitative subtraction of degassing and temperature effects on $\delta^{13}C$ and $\delta^{18}O$ isotopic fractionation in the spring water and the travertine. These robust disequilibrium signatures may be biologically mediated and systematically increase in magnitude from the high (73°C) to the low (≤25°C) temperature portions of the Spring AT-1 outflow.

We then conducted a culture-independent molecular survey of the bacterial communities, which are distinctly partitioned between travertine depositional facies in the surface drainage system of Spring AT-1 (Fouke et al. 2003). PCR amplification and sequencing of 16S rRNA genes with universally conserved bacterial primers has identified over 553 unique partial and 104 complete gene sequences (derived from more than 14,000 clones) affiliated with 221 unique species that represent 21 bacterial divisions. These sequences exhibited less than 12% similarity in bacterial community composition between each of the travertine depositional facies. This implies that relatively little downstream bacterial transport and colonization take place despite the rapid and continuous flow of spring water from the high-temperature to

low-temperature facies. These results suggest that travertine depositional facies, which are independently determined by the physical and chemical conditions of the hot spring drainage system, effectively predict bacterial community composition as well as the morphology and chemistry of travertine precipitation.

Materials and Methods

Field work and sample collection. We collected multiple samples from within each of the five travertine facies at Spring AT-1 for the purpose of conducting the first direct correlation of bacterial 16S rRNA gene sequence identifications with travertine mineral precipitation in the context of sedimentary depositional facies (Fouke et al. 2003). Field photographs and detailed diagrams depicting aerial and cross-sectional views of Spring AT-1, and sampling positions, have previously been published (Fouke et al. 2000, 2003). As a brief summary, samples were collected from the interior of each of the five facies, with each sample occurring within the continuous primary flow path of the primary hot spring drainage outflow (Veysey et al. 2008). The sampling strategy for the present study was to conduct an initial characterization of the microbial communities inhabiting each travertine facies. Therefore, each sample was collected from the middle of each facies and was thus laterally separated from the next sample by as much as a few meters. With the results presented in this study and our previous work (Fouke et al. 2003), our ongoing microbiological analyses of Spring AT-1 is currently focused on detailed mm-scale sampling across the boundaries between facies, as well as a correlation of specific crystal morphologies and chemistries with microbial phylogenetic and functional diversity. However, these next progressive and strategic stages of our analysis of Spring AT-1 would not be possible without the synthesis of the data presented in the present paper.

DNA extraction, PCR amplification, cloning and sequencing. The DNA extraction protocols and 16S rRNA gene sequence PCR amplification protocols employed have been optimized to avoid biases and have previously been described (Frias-Lopez et al. 2002; Fouke et al. 2003). DNA amplified using universal bacterial primers was then cloned in order to isolate the individual 16S rRNA gene sequences. To maximize the number of unique sequences identified (thus better characterizing the total diversity of the spring system) we chose to avoid sequencing identical clones derived from a single PCR reaction. Of the greater than 14,000 clones generated, approximately 5,000 clones were screened by RFLP analyses and 1,050 potentially unique clones were selected for sequencing. Ultimately, 657 partial 16S rRNA gene sequences were obtained, and 108 of these were sequenced as contigs to completion.

Nucleotide sequence accession numbers. The GenBank accession numbers for the 16S rRNA gene sequences analyzed in this study have previously been reported (Fouke et al. 2003).

Statistical analyses. We analyzed our sequences using three Operational Taxonomic Unit (OTU) definitions, defined by sequence differences of 0.5%, 1%,

and 3%, to determine whether our interpretations of environmental partitioning could be affected by such variation. The lower bound is due to our PCR and sequence derived error rate (Tindall and Kunkel 1988; Barnes 1992) and the 3% difference is a typical OTU definition (Stackebrandt and Goebel 1994). In our accumulation curves, a straight line would indicate that we have sampled only a small subset of the total biodiversity: new OTUs are found at a constant rate with each additional new sample analyzed. If a facies is well sampled, however, the curve will flatten asymptotically when the number of samples, n, is large, because novel OTU sequences are detected with decreasing frequency.

To quantitatively estimate how well each facies has been sampled, accumulation curves were fitted to analytical curves obtained by modeling the sampling process. We assume that in each environmental sample collected, there is a maximum of N possible bacterial cells that could be detected, and that each of these cells would be present and detected in the sample with a probability p, regardless of the cell's identity. The factor p includes the combined probability of the cell being captured and detected through the process of DNA extraction and amplification of the 16S rRNA gene sequences via PCR. Thus, we use multiple methods of DNA extraction to eliminate cell durability biases and amplify the 16S rRNA gene via PCR. Finally, we screen the resultant clone library in an attempt to sequence only unique clones within that sample, as opposed to repeatedly sequencing identical clones. In this manner we increase the likelihood that an OTU will be detected even if it is not numerically dominant in the clone library (which may be due to extraction, amplification, and cloning biases rather than environmental population abundance).

The likelihood that each sequence we analyze will represent a new OTU is approximated as $(1-S/S_o)$, where S is the number of different OTUs already identified and S_o is the total number of different OTUs present in the environment. For each sequence, the probability that the number of different OTUs will increase is $p(1-S/S_o)$. This leads to an accumulation curve of the type $S = S_m\,(1-exp(-Kt))$, where t is the maximum number of individuals that would be found if $p = 1$ and K is a constant related to the sampling procedure. This is not quite what was represented in the accumulation curves, since we only have information about samples rather than individuals, as explained above. Nonetheless, the number of samples n is simply $n = t/N$, so $S = S_m\,(1-exp(-Kn))$. The parameters K and S_m were determined from a linear fit of $log(dS/dn)$ versus $-n$. Estimates through other methods were also attempted: fits to hyperbolic accumulation curves (Colwell and Coddington 1994) were not convincing and non-parametric methods (Krebs 1989; Chao and Lee 1994) yielded variances that were too large to be trustworthy.

Results

We identified 193 OTUs using the 3% cutoff and found that 90% of these could be identified in only one of the facies (partitioned between facies). There were 237 OTUs using the 1% cutoff and 331 OTUs using the 0.5% cutoff with 91%

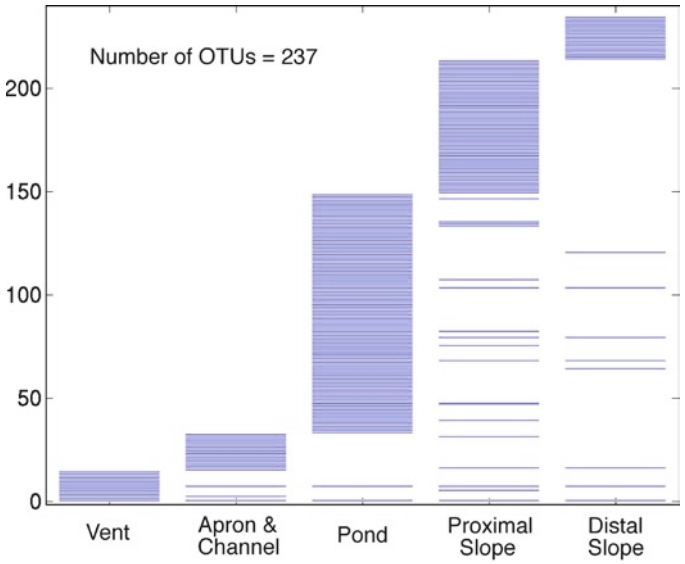

Fig. 11.2 Species present in each facies: 1% OTU definition. Each OTU is numbered sequentially, starting with OTUs that first appear in the Vent facies, followed by OTUs that first appear in the Apron and Channel, then the Pond, the Proximal Slope, and lastly the Distal Slope facies. The figure provides a graphical representation of where each OTU (*y*-axis) is found (*x*-axis)

and 93% (respectively) of the sequences partitioning to a single facies. Figure 11.2 graphically represents the distribution of sequences between the five facies using a 1% OTU definition. The plots for the 3% and 0.5% definitions are similar in appearance; however, under the 3% rule, two sequences may be found in all five facies (20). Finally, the total number of sequences that can be found in more than one facies remains low under each OTU definition: 19 OTUs under the 3% definition, 20 OTUs under the 1% definition, and 24 OTUs under the 0.5% definition.

Accumulation curves were generated for the three different OTU definitions (3%, 1% and 0.5%) for each facies and the results for the pond facies are shown in Fig. 11.3b. The curve from each OTU definition collapses into the same curve, giving some confidence in the robustness of the sampling procedure and the validity of the assumption of random sampling used to derive the exponential accumulation curve. We see this pattern no matter which OTU definition is used. In the model above, all of the OTUs were assumed equally likely to appear (hence the factor $1-S/S_o$). In a more realistic model the probability of finding each OTU should be proportional to its abundance. However, the approximations used above describe the data well and provide a tractable expression for the accumulation curve.

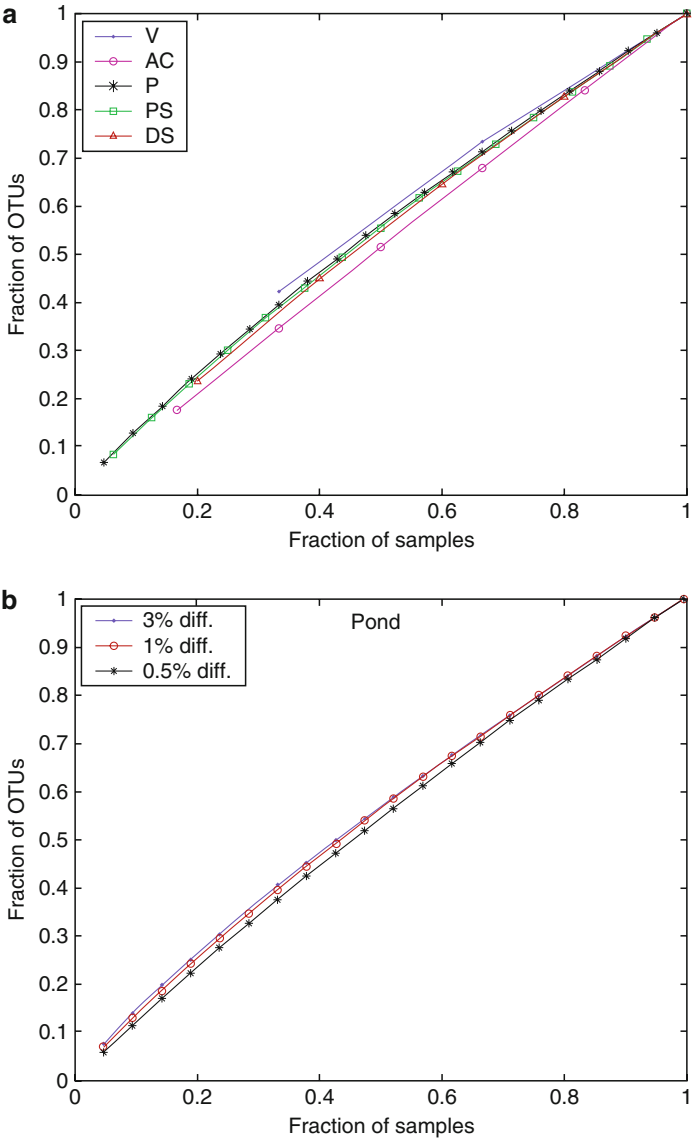

Fig. 11.3 Accumulation curves and exponential fits. (**a**) Accumulation curve generated for each facies using a 1% OTU definition. (**b**) Accumulation curves generated for the pond facies using a 3%, 1%, and 0.5% OTU definitions. Accumulation curves for different OTU definitions collapse into the same curve when the *x* and *y*-axis are properly scaled by the total number of OTUs and samples, respectively

Discussion

Although different microbial species have specific growth requirements and preferred temperature and pH ranges, the tight partitioning with respect to the travertine facies is nonetheless remarkable. First, it is surprising that very few of the upstream sequences were not also detected downstream. We initially expected that the rapid flow of the spring would result in downstream transport of microbial cells, and thus we thought that many sequences would also being identified downstream of their point of initial detection. Consequently, we performed most of our analyses on the first four facies extending from the vent. Surprisingly, the sequences detected in the water column of one facies, which are presumably most susceptible to being flushed downstream, were not typically detected downstream of their original facies. Secondly, because bacterial species have a preferred range of environmental growth conditions, we expected that many sequences would be found across facies boundaries, coinciding with gradual temperature and pH changes. However, the facies boundaries proved to be nearly absolute boundaries with respect to detected bacterial 16S rRNA gene sequences. Although we observed particular sequences over a range of conditions within each travertine facies, with very few exceptions, OTUs were not found to traverse the facies boundaries.

Inferred metabolic activity of the identified bacteria, derived from comparison of our sequences to GenBank, indicates that the bacterial communities found in the spring drainage system change from primarily chemolithotrophic in the vent facies, to photoautotrophic and ultimately to heterotrophic in the distal slope facies. Associated with this transition is an observed increase in the total number of OTUs and their associated bacterial divisions from the vent to the pond facies. The number of OTUs decreases, however, with down flow progression into the proximal slope and distal slope facies. These trends in our data can be interpreted as follows: fewer OTUs and bacterial divisions would be expected at the upper temperature limits of the spring where little organic matter is available for heterotrophy and the temperature is at the upper limit for photosynthesis (Miller and Castenholz 2000). Although the pond, proximal-slope and distal-slope facies have temperature profiles that would support both autotrophic and heterotrophic lifestyles (Fouke et al. 2000), we actually find a reduction in the number of species represented in the proximal-slope and distal-slope facies. This variation results from differences in the environmental stability of each facies with regards to temperature, pH, and water flow (Veysey et al. 2008). Ponds, for example, have the widest temperature and pH range of any facies and show greater fluctuations in flow direction and intensity.

To validate our interpretation of facies partitioning, we need to determine what proportion of the total community in each facies we have identified. Severe undersampling might prevent us from identifying OTUs that actually do occur in multiple facies. Estimates for the total number of OTUs in each facies are made using an exponential fit to the accumulation curve in Fig. 11.3. The accumulation curve plots the number of different OTUs, S, found in a given number of samples versus this number of samples, n. Since all of the samples are assumed to be equivalent, this

graph is an average over all possible permutations of these samples. Accumulation curves are traditionally made using the number of individuals as the x-axis instead of the number of samples (Hughes et al. 2001). However, our samples amalgamate large numbers of individuals: we have information regarding which OTUs are present in each sample, but not the OTU identity for every individual in the sample. The abundance of unique gene sequences in the clone libraries are not representative of the abundances in the environmental sample due to the inherent DNA extraction and PCR biases. Therefore, the clone library data cannot be used to make accumulation curves.

Accumulation curves were generated based on the three different OTU definitions (3%, 1% and 0.5%) for each facies, with the results for the pond facies shown as an example in Fig. 11.3b. The curve from each OTU definition collapses into the same curve, giving confidence in the robustness of the sampling procedure and the validity of the assumption of random sampling used to derive the exponential accumulation curve. Thus, since all of the individual cells are captured with equal probability, we expect that the observed OTUs represent the most numerically abundant bacteria in each facies. Consequently, we conclude that these species (and therefore most of the bacterial consortia) are partitioned according to the travertine facies model. This finding constrains abiotic and biotic theories for the origin of travertine terraces (Kandianis et al. 2008; Veysey and Goldenfeld 2008).

Acknowledgements This research was supported by the National Science Foundation Biocomplexity in the Environment Coupled Biogeochemical Cycles Program (EAR 0221743), the National Science Foundation Geosciences Postdoctoral Research Fellowship Program (EAR-0000501), the Petroleum Research Fund of the American Chemical Society Starter Grant Program (34549-G2), and the University of Illinois Urbana-Champaign Critical Research Initiative. This work was completed under National Park Service research permit number 3060R. Conclusions in this study are those of the authors and do not necessarily reflect those of the funding or permitting agencies. Discussions with A. Salyers and C. Woese are gratefully acknowledged.

References

Barnes WM (1992) The fidelity of Taq polymerase catalyzing PCR is improved by an N-terminal deletion. Gene 112:29–35

Chao A, Lee S-M (1994) Estimating population size via sample coverage for closed capture-recapture models. Biometrics 50:88

Colwell RK, Coddington A (1994) Estimating terrestrial biodiversity through extrapolation. Phil Trans R Soc Lond Ser B 345:101–108

Ford TD, Pedley HM (1996) A review of tufa and travertine deposits of the world. Earth Sci Rev 41:117–175

Fouke BW, Farmer JD, Des Marais D, Pratt L, Sturchio NC, Burns PC, Discipulo MK (2000) Depositional facies and aqueous-solid geochemistry of travertine-depositing hot springs (Angel Terrace, Mammoth Hot Springs, Yellowstone National Park, USA). J Sediment Res 70:265–285

Fouke BW, Bonheyo GT, Sanzenbache BL, Frias-Lopez J (2003) Partitioning of bacterial communities between travertine depositional facies at Mammoth Hot Springs, Yellowstone National Park, USA. Can J Earth Sci 40:1531–1548

Frias-Lopez J, Zerkle AL, Bonheyo GT, Fouke BW (2002) Cyanobacteria diversity associated with coral black band disease in Caribbean and Indo-Pacific reefs. Appl Environ Microbiol 69:2409–2413

Friedman I (1970) Some investigations of the deposition of travertine from hot springs: I. The isotope chemistry of a travertine-depositing spring. Geochim Cosmochim Acta 34:1303–1315

Hughes JB, Hellmann JJ, Ricketts TH, Bohannan BJM (2001) Counting the uncountable: statistical approaches to estimating microbial diversity. Appl Environ Microbiol 67:4399–4406

Kandianis MT, Fouke BW, Veysey J, Johnson RW, Inskeep W (2008) Microbial biomass: a catalyst for $CaCO_3$ precipitation in advection-dominated transport regimes. GSA Bull 120:442–450

Krebs C (1989) Ecological methodology. Harper & Row, New York

Miller SR, Castenholz RW (2000) Evolution of thermotolerance in hot spring cyanobacteria of the genus *Synechococcus*. Appl Environ Microbiol 66:4222–4229

Pentecost A (2005) Travertine. Springer, Heidelberg

Sorey ML (1991) Effects of potential geothermal development in the Corwin Springs known geothermal resources area, Montana, on the thermal features of Yellowstone National Park. Water Resources Investigations Report No. 91-4052. US Geological Survey

Stackebrandt E, Goebel B (1994) Taxonomic Note: a place for DNA-DNA reassociation and 16S rRNA sequence analysis in the present species definition in bacteriology. Int J Syst Bacteriol 44:846–849

Tindall KR, Kunkel TA (1988) Fidelity of DNA synthesis by the Thermus aquaticus DNA polymerase. Biochemistry 27:6008–6013

Veysey J, Goldenfeld N (2008) Watching rocks grow. Nat Phys 3:1–5

Veysey J, Fouke BW, Kandianis MT, Schickel TJ, Johnson RW, Goldenfeld N (2008) Reconstruction of water temperature, pH, and flux of ancient hot springs from travertine depositional facies. J Sediment Res 78:69–76