Running Head: Correlation of Microbial Communities with Travertine Mineral Precipitation

Correlation of Microbial Communities with Travertine Mineral Precipitation at Mammoth Hot Springs, Yellowstone National Park, USA

George T. Bonheyo[1], Jorge Frias-Lopez[1], Héctor García Martín[2], John Veysey[2], Nigel Goldenfeld[2] and Bruce W. Fouke[1]

[1]*Department of Geology, University of Illinois at Urbana-Champaign, 1301 West Green Street, Urbana, Illinois 61801-2938, USA.*

[2]*Department of Physics, University of Illinois at Urbana-Champaign, 1110 West Green Street, Urbana, Illinois 61801-3080, USA.*

Competing interests: The authors have no competing interests that might be perceived to influence the results and discussion reported in this paper.

Correspondence and request for materials should be addressed to Bruce W. Fouke (Email: fouke@uiuc.edu).

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. We have combined carbonate mineralogical and geochemical environments together with community-based microbial genetic analyses in hot spring drainage systems at Mammoth Hot Springs in Yellowstone National Park. Previously (7), we reported the shape and chemistry of carbonate mineral deposits (*travertine*), that have formed along the hot spring pathways. This travertine exhibits five distinct ecological zonations (termed sedimentary *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.

We have initiated a biocomplexity study at Mammoth Hot Springs in Yellowstone National Park to determine whether microbial community structure and activity can influence the chemistry and morphology of travertine mineral deposits (1, 2). Carbonates are ideal for this type of study because they are precipitated at life-permitting temperatures, are sensitive to environmental conditions, and are the most ubiquitous sedimentary rocks at the Earth's surface (3 - 7). Subsurface waters erupt at Mammoth Hot Springs to precipitate terraced crystalline travertine deposits composed of aragonite and calcite (2, 7, 8, 9). Our studies were conducted at Spring AT-1 (7, 9), 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 and can exceed 1.5 m per year. The rapid precipitation partially seals the vents and reroutes surface flow paths, 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 precipitation (whether biotic or abiotic) is complex and not yet understood. The hot springs harbor diverse communities of microorganisms, representing at least 21 divisions of bacteria (7).

In order to analyze the physical, geological, and biological aspects of this rapidly changing hydrothermal system, it is necessary to first subdivide the spring drainage system into a series of recognizable sub-environments along the flow path. These sub-environments are known as sedimentary *facies*. A facies is defined by the sum of the physical, chemical, geological, and biological attributes of an environment of sedimentary 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 depositional facies model for Spring AT-1 based on physical characteristics (e.g. temperature and pH), appearance, mineralogy, and limited microscopic observations of the microbiology (Fig. 1). These 5 facies are the vent, apron and channel, pond, proximal slope, and distal slope (7, 9). 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 (7). Although any given travertine facies may be as much as 10s of meters long and cover 100s 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).

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 (7, 13). Field photographs and detailed diagrams depicting aerial and cross-sectional views of Spring AT-1, and sampling positions, have previously been published (7, 9). As a brief summary, samples were collected from the interior of each of the five facies, with each sample occurring

within the continuous flow path of the primary hot spring drainage outflow. 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 (7, 9), 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 (7, 14). 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 (15).

Nucleotide sequence accession numbers. The GenBank accession numbers for the 16S rRNA gene sequences analyzed in this study have previously been reported (7).

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 (16, 17, 18) and the 3% difference is a typical OTU definition (19). 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[24] curves were not convincing and non-parametric methods[25, 26] 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% and 93% (respectively) of the sequences partitioning to a single facies. Figure 2 graphically represents the distribution of sequences between the 5 facies using a 1% OTU definition. The plots for the 3% and 0.5% definitions are similar in appearance; however, under the 3% rule, 2 sequences may be found in all 5 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 Figure 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.

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 (21). Although the pond through distal slope facies have temperature profiles that would support both autotrophic and heterotrophic lifestyles, we actually find a reduction in the number of species represented in the proximal slope and distal slope. Although unproven, we hypothesize that such variation may result from differences in the environmental stability of each facies with regards to temperature, pH, and water flow. 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 under-sampling 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 Figure 3 (22). 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[23]. 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 Figure 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 theories for the origin of travertine terraces: either the origin is biotic, or else the microbial ecology is strongly coupled to the geochemistry through mechanisms presently unknown.

ACKNOWLEDGEMENTS

This research was supported by grants from the NSF Biocomplexity in the Environment Program, NSF Geosciences Postdoctoral Research Fellowship Program, Petroleum

Research Fund of the American Chemical Society Starter Grant Program, and the University of Illinois Urbana-Champaign Critical Research Initiative. Discussions with A. Salyers and C. Woese are gratefully acknowledged.

REFERENCES

1. Northrup, D, Lavoie, K.H. 2001. Geomicrobiology of caves: a review. *Geomicrobiology J.* **18**, 199-221.
2. Ford, T.D., Pedley, H.M. 1996. A review of tufa and travertine deposits of the world. *Earth Sci. Rev.* **41**, 117-175.
3. 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.
4. Given, R.K. & Wilkinson, B.H. 1985. Kinetic control of morphology, composition, and mineralogy of abiotic sedimentary carbonates. *J. Sed. Petrology*, **55**, 109-119.
5. Chafetz, H.S., Ruxh, P.F., Utech, N.M. 1991. Microenvironmental controls on mineralogy and habit of $CaCO_3$ precipitates: an example from an active travertine system. *Sedimentology*, **38**, 107-126.
6. Burton, E.A. 1993. Controls on marine carbonate cement mineralogy: review and reassessment. *Chem. Geol.*, **105**,163-179.
7. Fouke, B.W., Bonheyo, G.T., Sanzenbacher, B.L., Frias-Lopez, J. 2003. Partitioning of bacterial communities between travertine depositional facies at Mammoth Hot Springs, Yellowstone National Park, U.S.A. *Can. J. Earth Sci.*, 40, 1531-1548 p.

8. Sorey, M.L. 1991. Effects of potential geothermal development in the Corwin Springs known geothermal resources area, Montana, on the thermal features of Yellowstone National Park. 91-4052 (U.S. Geological Survey, Menlo Park, CA).

9. Fouke, B.W. Farmer, J.D., Des Marais, D.J., Pratt, L., Sturchio, N.C., Burns, P.C., Discipulo, M.K. 2000. Depositional facies and aqueous-solid geochemistry of travertine-depositing hot springs (Angel Terrace, Mammoth Hot Springs, Yellowstone National Park, USA). *J. Sed. Res*., **70**, 265-285.

10. Farmer, J.D. 2000. Hydrothermal systems: doorways to early biosphere evolution. *GSA Today*, **10**, 1-8.

11. Benson, L., White, L.D., Rye, R. 1996. Carbonate deposition, Pyramid Lake subbasin, Nevada: 4. Comparison of stable isotope values of carbonate deposits (tufas) and the Lahontan lake-level record. *Palaeogeography, Palaeoclimatology, and Palaeoecology*, **122**, 45-76.

12. Boston, P.J. 2001. Cave biosignature suites: microbes, minerals and Mars. *Astrobiology*, **1**, 25-55.

13. The sampling regime included filtering 2 liters of water collected from each facies, sampling microbial masses, and sampling from the uppermost 1 cm of travertine deposits.

14. Frias-Lopez, J., Zerkle, A.L., Bonheyo, G.T., Fouke, B.W. 2002. Cyanobacteria diversity associated with coral black band disease in Caribbean and Indo-Pacific reefs. *Appl. Environ. Microbiol.*, **69**, 2409-2413.

15. A complete file of all of the sequences identified and their GenBank accession numbers is available as Tables 1 and 2 in the supplementary material, to be made available online.

16. Sequence and PCR error rate: Repetitive single-stranded sequencing and editing of the same sequences gave a predicted error rate of 0.32% and the error rate for Taq polymerase is 10-4 (17, 18).

17. Tindall, K.R., Kunkel, T.A. 1988. Fidelity of DNA synthesis by the Thermus aquaticus DNA polymerase: *Biochemistry*, **27**, 6008-6013.

18. Barnes, W.M. 1992. The fidelity of Taq polymerase catalyzing PCR is improved by an N-terminal deletion. *Gene*, **112,** 29-35.

19. 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.

20. Plots for 3% and 0.5% OTU definitions are included as Figure 1 in the supplementary material, to be made available online.

21. Miller, S.R., Castenholz, R.W. 2000. Evolution of thermotolerance in hot spring cyanobacteria of the genus Synechococcus. *Appl. Environ. Microbiol*. **66**, 4222-4229.

22. Additional plots for the other facies and OTU definitions are found in Figures 2-4 of the supplementary material, to be made available online.

23. Hughes, J.B., Hellmann, J.J., Ricketts, T.H., Bohannan, B.J.M. 2001. Counting the uncountable: statistical approaches to estimating microbial diversity. *Applied and Environmental Microbiology* **67**, 4399-4406.

24. Colwell, R. K., Coddington, J.A. 1994. Estimating terrestrial biodiversity through extrapolation. *Philosophical Transactions of the Royal Society (Series B)* **345**, 101.

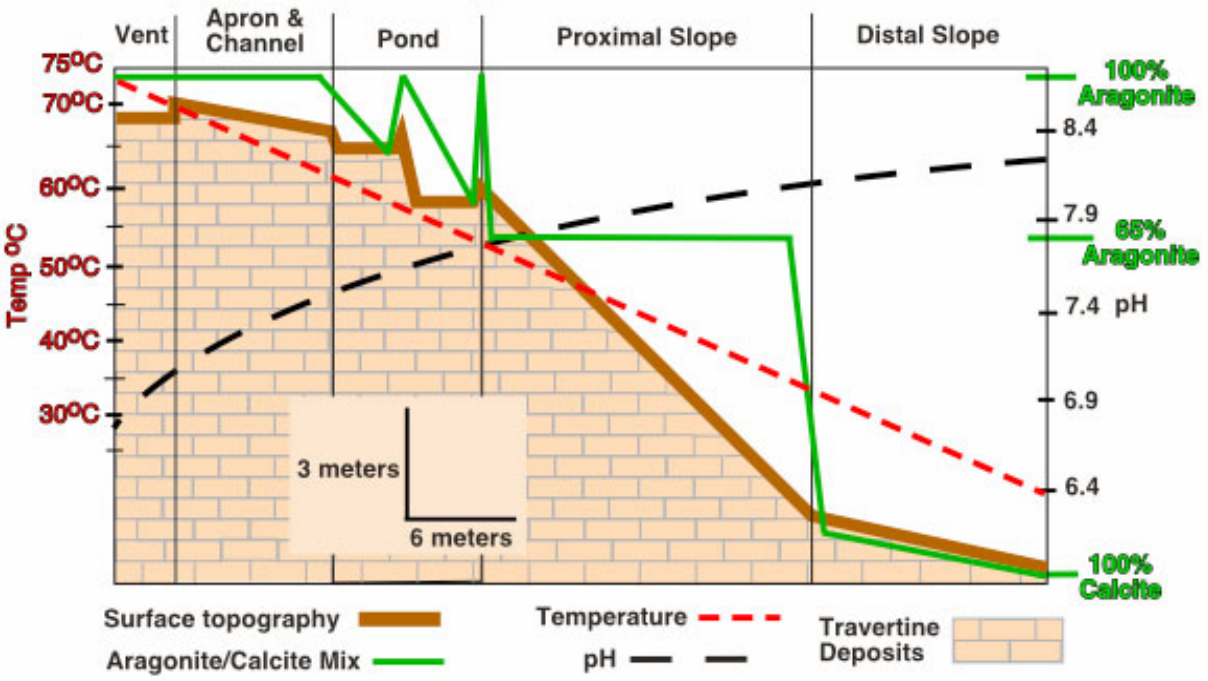
25. Krebs, C. 1989. *Ecological methodology* (Harper and Row, New York).

1 26. Chao, A. and Lee, S-M. 1994. Estimating population size via sample coverage for
2 closed capture-recapture models, *Biometrics* **50**, 88.

Fig 1. Facies model. Cross-sectional view of Spring AT-1 with 2x 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.

Fig. 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).

Fig. 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.


Vent
Apron & Channel
Pond
Proximal Slope
Distal Slope
75°C
70°C
60°C
50°C
40°C
30°C
Temp °C
100% Aragonite
8.4
7.9
65% Aragonite
7.4 pH
6.9
6.4
100% Calcite
3 meters
6 meters
Surface topography
Aragonite/Calcite Mix
Temperature
pH
Travertine Deposits

1

2 Figure 1

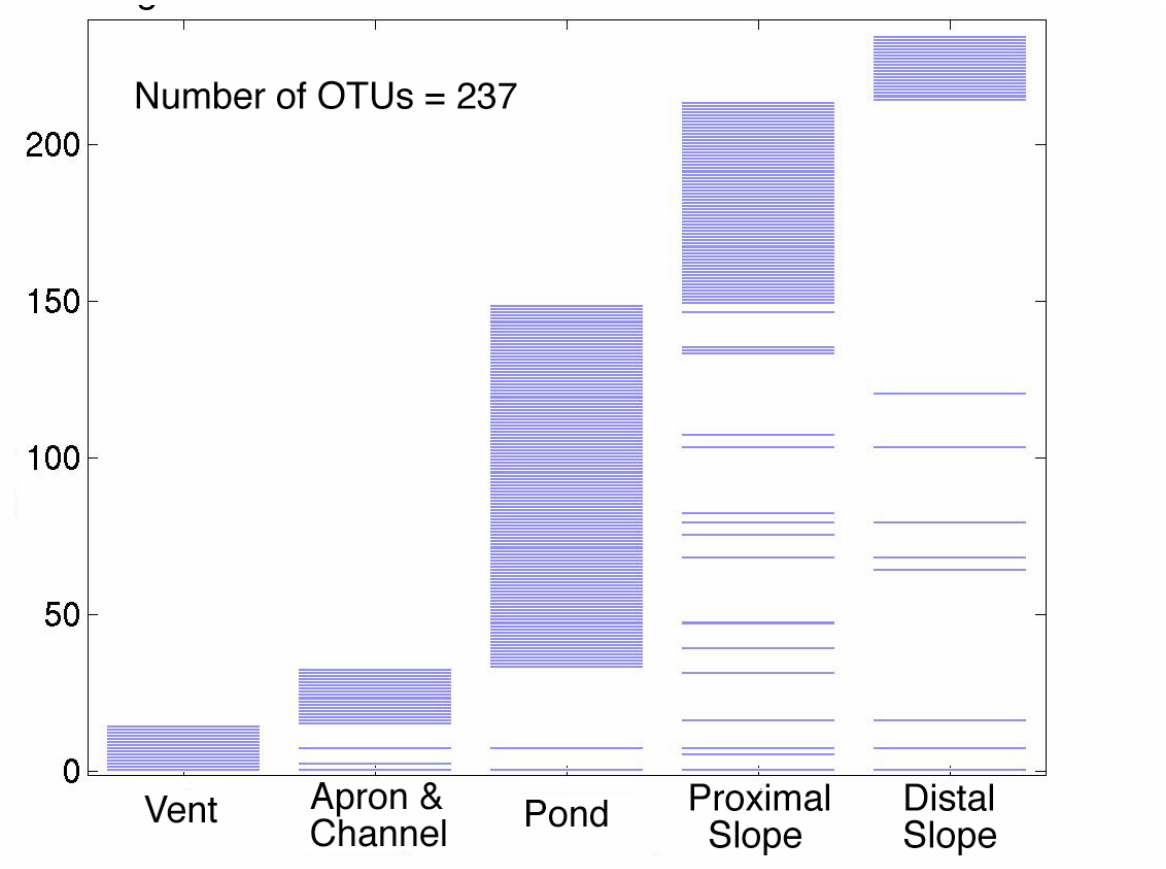
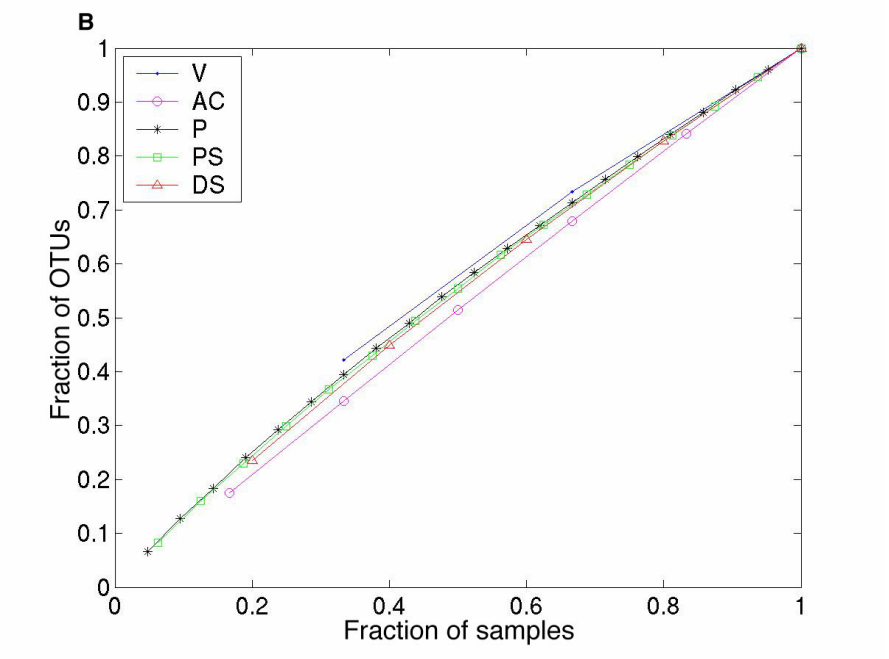
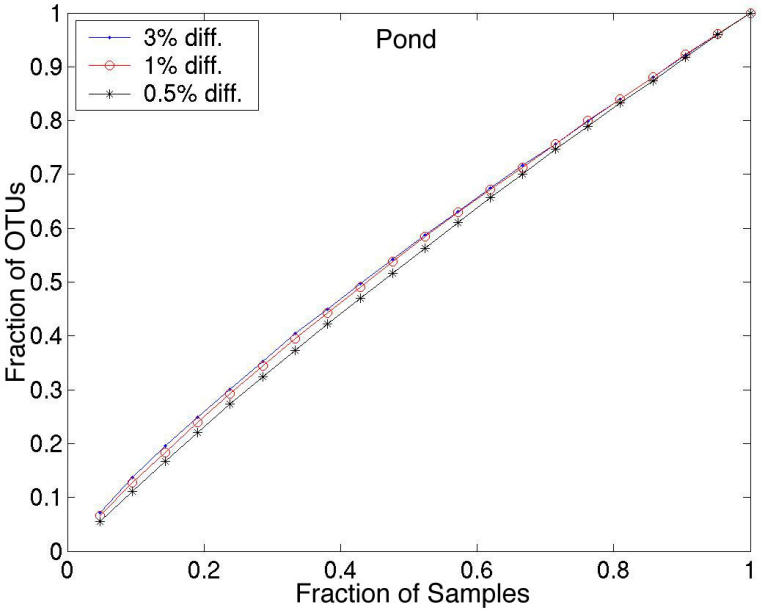

Number of OTUs = 237
200
150
100
50
0
Vent
Apron &
Channel
Pond
Proximal
Slope
Distal
Slope

Figure 2.



1 Fig 3A

2



3 Fig 3B

4