environmental microbiology

Environmental Microbiology (2011) **13**(1), 226–240 doi:10.1111/j.1462-2920.2010.02324.x

A comparative molecular analysis of water-filled limestone sinkholes in north-eastern Mexico

Jason W. Sahl,[1] Marcus O. Gary,[2] J. Kirk Harris[3] and John R. Spear[1]*

[1]*Environmental Science and Engineering, Colorado School of Mines, Golden, CO 80401, USA.*

[2]*Department of geological sciences, University of Texas, Austin, TX, USA.*

[3]*Pediatrics, University of Colorado, Denver, CO 80045, USA.*

Summary

Sistema Zacatón in north-eastern Mexico is host to several deep, water-filled, anoxic, karstic sinkholes (cenotes). These cenotes were explored, mapped, and geochemically and microbiologically sampled by the autonomous underwater vehicle deep phreatic thermal explorer (DEPTHX). The community structure of the filterable fraction of the water column and extensive microbial mats that coat the cenote walls was investigated by comparative analysis of small-subunit (SSU) 16S rRNA gene sequences. Full-length Sanger gene sequence analysis revealed novel microbial diversity that included three putative bacterial candidate phyla and three additional groups that showed high intra-clade distance with poorly characterized bacterial candidate phyla. Limited functional gene sequence analysis in these anoxic environments identified genes associated with methanogenesis, sulfate reduction and anaerobic ammonium oxidation. A directed, barcoded amplicon, multiplex pyrosequencing approach was employed to compare ~100 000 bacterial SSU gene sequences from water column and wall microbial mat samples from five cenotes in Sistema Zacatón. A new, high-resolution sequence distribution profile (SDP) method identified changes in specific phylogenetic types (phylotypes) in microbial mats at varied depths; Mantel tests showed a correlation of the genetic distances between mat communities in two cenotes and the geographic location of each cenote. Community structure profiles from the water column of three neighbouring cenotes showed distinct variation; statistically significant differences in the concentration of geochemical constituents suggest that the variation observed in microbial communities between neighbouring cenotes are due to geochemical variation.

Received 30 September, 2009; accepted 6 July, 2010. *For correspondence. E-mail jspear@mines.edu; Tel. (+1) 303 273 3497; Fax (+1) 303 273 3413.

Introduction

Cultivation-independent molecular methods have revolutionized our view of the diversity, geographical distribution and metabolic potential of environmental microbes. New pyrosequencing methods have allowed researchers to more thoroughly survey microbial communities in complex systems (Dinsdale *et al*., 2008). Pyrosequence analysis has changed our view of microbial diversity in both ocean (Huber *et al*., 2007) and soil (Roesch *et al*., 2007) samples. Comparative methods that exploit the large amount of sequence data generated by pyrosequencing can be used to answer fundamental questions that concern microbial community structure, distribution and function.

Sistema Zacatón in north-eastern Mexico (Gary *et al*., 2006a) is host to several water-filled, limestone sinkholes (cenotes). This system is part of a larger complex of water-filled cenotes, travertine-capped cenotes, and dry cave passages in the region. The five water-filled cenotes in Sistema Zacatón include Zacatón, Verde, Azufrosa, La Pilita and Caracol (Fig. 1). These five water bodies include diverse morphologies and source waters that range from hydrothermal to meteoric. The anoxic chemistry of the cenotes, which is thought to preclude the presence of grazers, combined with nutrients provided by the hydrothermal input, provides conditions amenable to the colonization of thick and intricate microbial mats.

An initial molecular survey into the microbial diversity of microbial mats in cenote Zacatón identified a diverse microbial assemblage that included two novel candidate phyla (Sahl *et al*., 2010). The current study extends this analysis to investigate the microbial community structure of water column and mat samples from five neighbouring cenotes. Additionally, this study explores the composition of microbial functional genes, and develops a new method to compare the taxonomic composition between samples. Samples from the five cenotes were compared with this method to determine the controls on microbial distribution between comparable environments in close

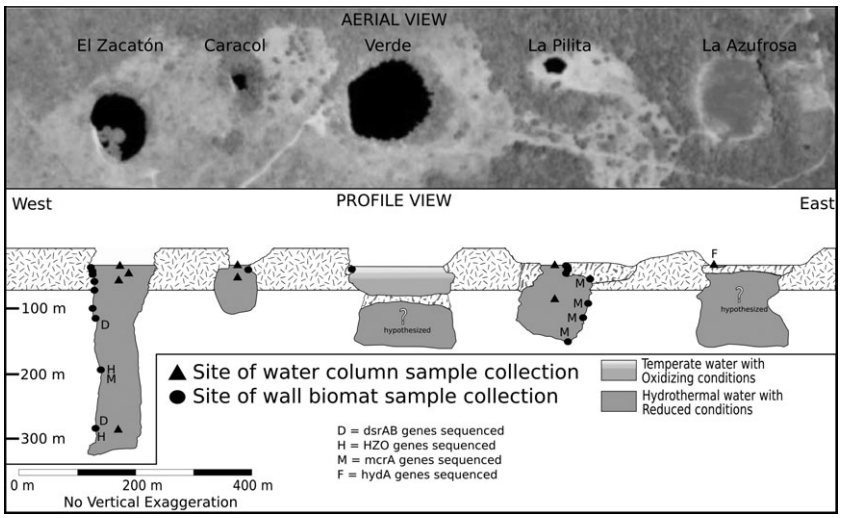

Fig. 1. A profile view of Sistema Zacatón cenotes. Sites where biological samples (water column and microbial mat) were collected and samples from which functional genes were amplified are shown.

geographical proximity. The samples were obtained either by divers or by the autonomous underwater vehicle deep phreatic thermal explorer (DEPTHX) (Krajick, 2007; Stone, 2007; Gary *et al*., 2008). DEPTHX not only collected samples inaccessible to divers, but also created 3-D maps of the cenotes through a sonar array and a simultaneous localization and mapping method (SLAM) (Fairfield *et al*., 2006; 2007); this method revealed the complex and variable morphology of these water-filled, karstic structures.

Results

Cenote morphology

Sonar mapping capabilities on the DEPTHX vehicle allowed for a high-resolution visualization of the morphologies of Sistema Zacatón cenotes. No significant discrete inlets or outlets were identified in cenotes Caracol and La Pilita. Caracol and La Pilita demonstrate a 'whiskey-jug' morphology (Gary, 2009), with secondary travertine deposition observed in the more shallow regions (Fig. 1). The morphology of cenote Zacatón is a straight shaft with remarkably flat, planar wall faces below depths of 85 m that reflect speleogenic structural control from pre-existing fractures in the Creaceous limestone matrix. Verde, which is hypothesized to overly a travertine-capped cenote based on geophysical investigations (Gary *et al*., 2009), has a morphology unlike the other cenotes of Sistema Zacatón (Fig. 1).

Water chemistry comparisons

Water column samples were collected at various depths in the cenotes (Fig. 1) and analysed for chemical constituents such as anions, cations and organic carbon. Overall, the chemical constituents of cenotes La Pilita and Zacatón were similar in concentration (Table 1) with variance in Eh, NH_4^+ and Zn; a multiple factor analysis of 18 water chemistry parameters, which sums eigenvectors created from the differences in each parameter, showed that most of the variation could be explained by the first factor (Fig. 2). However, the second factor showed a small, but reproducible difference in parameters between 14 water samples from La Pilita and Zacatón. Because of sample availability, only chemical constituents from La Pilita and Zacatón were compared.

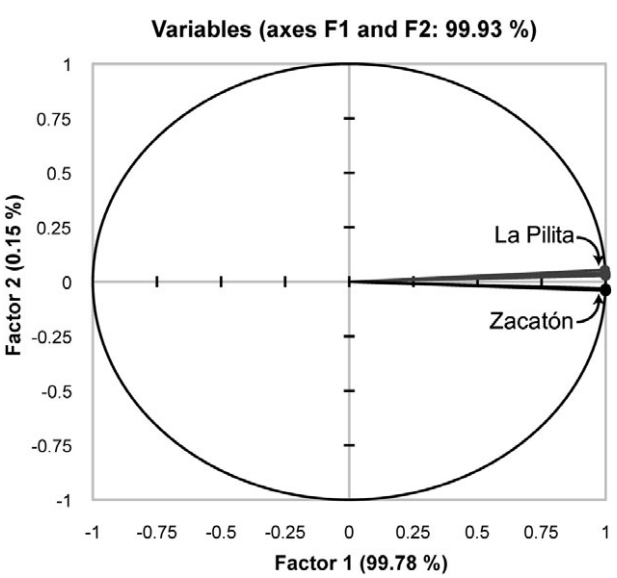

Fig. 2. A multiple factor analysis of 18 chemical parameters between 14 water column samples taken from La Pilita and Zacatón. The axes indicate the factor and the per cent of variation explained by that factor.

Table 1. Water chemistry parameters in Sistema Zacaton cenotes.

Parameter	La Pilita					Caracol		Zacatón				Verde	Azufrosa
Depth	0 m	30 m	50 m	80 m	100 m	0 m	17 m	0 m	22 m	90 m	273 m	0 m	0 m
Temperature	31.6	31.6	31.6	31.6	31.6	29.5	29.6	30.1	NM	30.0	30.1	26.0	NM
pH	6.64	6.87	6.87	6.87	6.87	6.62	6.67	6.67	6.58	6.72	6.53	6.63	6.60
Eh (mV)		90	97	86	49	–284	–136	–175	–240	–200	–282	0	–207
TOC (mM)	0.14	0.12	0.12	0.15	0.11	0.10	NM	0.16	NM	0.18	0.14	0.80	NM
SC ($\mu S\ cm^{-1}$)		879	879	880	880	877	853	759	726	767	761	722	873
Fe-T (μM)	0.01	BDL	BDL	BDL	BDL	0.15	0.74	0.01	0.27	0.38	0.20	0.16	0.13
NH_4^+ (mM)	0.11	NM	NM	NM	NM	0.14	1.32	0.05	1.26	NM	NM	BDL	0.12
H2S (mM)	0.008	NM	NM	NM	NM	0.008	0.005	0.01	0.006	NM	NM	BDL	0.020
S (mM)	0.20	0.20	0.20	0.20	0.20	0.12	NM	0.18	NM	0.17	0.18	0.17	NM
Major anions (mM)													
PO_4^{3-}	BDL	BDL	BDL	BDL	BDL	BDL	BDL	BDL	BDL	4.E-04	5.E-04	BDL	BDL
CL^-	1.11	1.1	1.09	1.09	1.06	1.07	1.18	1.01	1.15	0.95	0.95	0.99	0.97
NO_3^-	3.E-04	1.E-03	5.E-04	1.E-03	1.E-03	1.E-03	3.E-03	1.E-03	5.E-03	2.E-03	2.E-03	6.E-03	3.E-03
Br	0.003	0.004	0.003	0.003	0.003	0.003	0.007	0.002	0.008	0.002	0.002	0.005	0
HCO_3^-	6.23	6.25	6.35	6.22	6.07	6.23	6.06	6.24	6.32	6.17	6.29	6.18	6.26
SO_4^{2-}	0.17	0.17	0.15	0.17	0.16	0.11	0.17	0.15	0.17	0.15	0.16	0.17	0.15
Major cations (mM)													
Ca	3.11	3.17	3.15	3.13	3.04	3.18	3.40	3.11	2.99	2.80	2.96	2.95	3.13
K	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.08	0.08	0.08	0.08	0.08	0.08
Mg	0.75	0.76	0.75	0.75	0.73	0.72	0.82	0.49	0.53	0.46	0.49	0.53	0.56
Na	1.33	1.35	1.34	1.34	1.30	1.49	1.50	1.38	1.12	1.31	1.32	1.10	1.17
Si	0.42	0.43	0.42	0.42	0.41	0.49	0.42	0.45	0.39	0.42	0.45	0.41	0.40
Sr	0.008	0.008	0.008	0.008	0.008	0.008	0.007	0.006	0.006	0.006	0.006	0.007	0.007
Trace elements (μM)													
Zn	BDL	0.21	0.23	0.41	0.08	2.91	0.61	13.30	2.00	6.40	4.30	0.46	0.89
Mo	0.67	0.13	0.10	0.23	0.09	0.02	0.01	0.01	0.01	0.03	0.01	0.03	0.04
As	BDL	BDL	BDL	BDL	0.019	BDL	BDL	BDL	BDL	BDL	BDL	BDL	BDL
Cd	0.01	0.01	0.01	0.02	0.02	0.03	0.03	0.03	0.02	0.03	0.03	0.03	0.03
Charge imbalance													
Sum cations (meq)	9.15	9.31	9.24	9.20	8.94	9.39	10.04	8.66	8.24	7.91	8.30	8.14	8.62
Sum anions (meq)	7.51	7.52	7.59	7.48	7.29	7.41	7.42	7.40	7.65	7.27	7.40	7.35	7.38
Charge imbalance (%)	9.8	10.6	9.8	10.3	10.1	11.8	15.0	7.8	3.7	4.2	5.7	5.1	7.8

BDL = below detection limit, NM = not measured.

Water column community comparisons

Two litres water samples at a range of depths from the cenotes Zacatón, La Pilita, Caracol, Verde and Azufrosa were filtered through 0.2 μm filters and then frozen for microbial community analysis. The filters collected from Zacatón and Caracol were red following filtration, regardless of depth collected; filters collected from La Pilita at all depths were green (0–100 m). An analysis of 16S rRNA gene sequences (both Sanger and pyrosequences) amplified from DNA extracted from filters identified phylogenetic groups with expected pigmentation consistent with these observations. Sequence distribution profiles (SDPs) created from sequence libraries obtained from Zacatón and Caracol were dominated by *Epsilonproteobacteria* and *Chlorobi* sequences (Fig. 3), whereas libraries from La Pilita showed an absence of *Epsilonproteobacteria* and *Chlorobi* sequences. La Pilita contained sequences related to *Chlorophycea* chloroplasts (> 98% sequence identity) and *Betaproteobacteria* sequences (> 96% sequence identity); the *Betaproteobacteria* sequences were related to those amplified from a deep subsurface sample (Gihring *et al*., 2006). Overall, the diversity in the water column of each cenote was low (Table 2) compared with the wall- and surface-associated microbial mats.

Comparisons of water column microbial communities across the photic zone

The SDP method developed in this study creates a fingerprint of taxonomic pyrosequence classifications and was used to compare the microbial communities from samples across the photic zone of three cenotes. The results show that in each cenote, there appears to be a subtle shift in the phylogenetic type (phylotype) distribution between shallow and deep water-column samples (Fig. 3). For example, in La Pilita, the surface sample and a sample collected at the bottom of the photic zone (44 m) are both dominated by *Gammaproteobacteria* sequences (Fig. 3). However, as can be clearly seen in the SDP, an additional *Gammaproteobacteria* phylotype is enriched in the deep sample and is absent in the surface sample. The surface sample sequences group with *Rhodocyclales* and the SBR1001 clade (Fig. 4), which is only recognized under the Hugenholtz taxonomic system in Greengenes (DeSantis *et al*.,

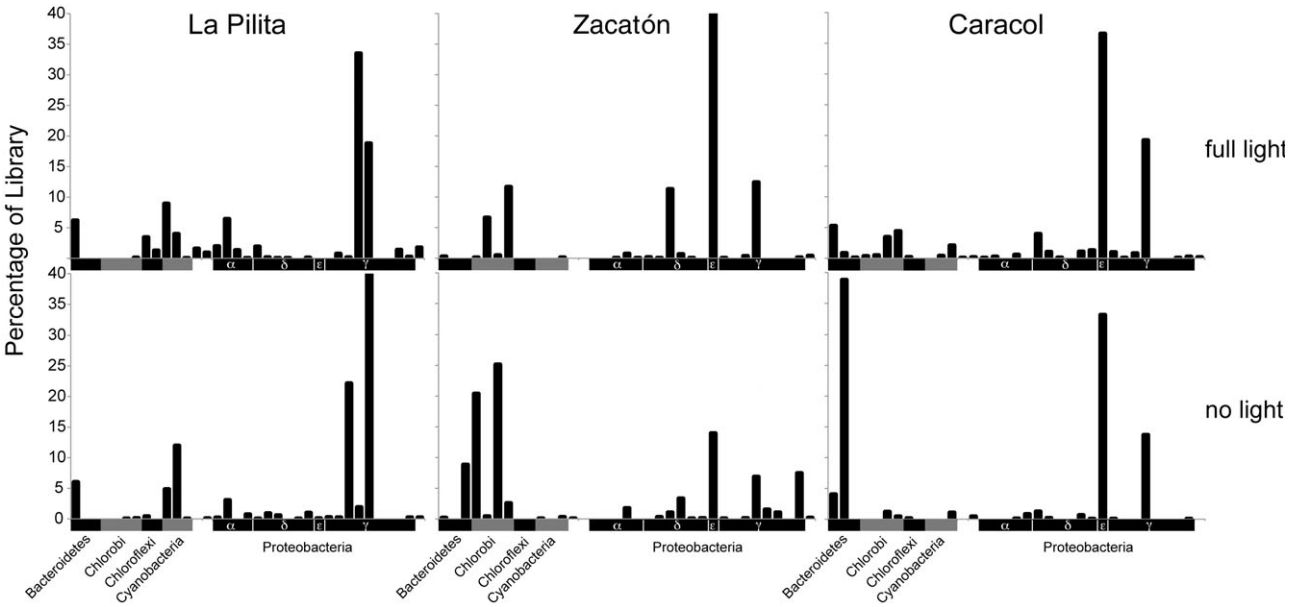

Fig. 3. Sequence distribution profiles of microbial phylotype distribution between water column samples from three cenotes as determined by pyrosequencing bacterial 16S rRNA genes. Peaks on the *x*-axis represent individual phylotypes, while the *y*-axis represents the percentage of the total library. The bars on the *x*-axis represent relevant phylogenetic groups.

2006a). The sample collected at the bottom of the photic zone contained sequences that group with SBR1001 and *Neisserales*, which are absent from the surface sample. Identical near-full-length sequences from the *Neisserales* clade were also amplified from the cenote water column (Accession No. FJ484989) and show 95% relatedness to sequences from a deep subsurface gold mine (M.M. Milleson and T.L. Kieft, unpublished, AY762628). Despite differences in specific phylotypes across the photic zone between the three cenotes, a significance test of SDPs between samples showed that the differences between bulk community structure were not statistically significant ($P > 0.10$).

In Zacatón and Caracol, an increase in phylotype abundance from the phylum *Bacteroidetes* was observed with water column sample depth (Fig. 3). Near-full-length sequences of this phylotype, which were amplified from the water column of Zacatón, La Pilita and Caracol, are related (~94%) to sequences from anoxic, freshwater lake sediment (Schwarz *et al*., 2007).

Phylogenetic classification of water and mat communities

16S rRNA phylogenetic methods were used to understand how sequences amplified from both microbial mat and water column samples in the cenotes surveyed are related to sequences previously reported from comparable environments. Near-full-length sequences were used to infer phylogenetic relationships and to identify novel lineages. Pyrosequences, which have been shown to provide comparable results with longer sequences (Liu *et al*., 2007; 2008), were used for more in-depth surveys of microbial diversity (Huber *et al*., 2009).

The microbial community structure of Zacatón microbial mats has previously been shown to be diverse, but relatively homogenous below 20 m (Sahl *et al*., 2010). In this study, the same trend of homogeneity within mat communities in deeper regions of La Pilita was also observed (Fig. 5). In each cenote, *Deltaproteobacteria* sequences are abundant at depth within microbial mats and are represented by one dominant phylotype; a representative near-full-length sequence showed only 96% identity to sequences in public databases (lake sediment sample, Accession No. DQ787719) (Fig. 4) (Nelson *et al*., 2007). *Chloroflexi* sequences are dominant in La Pilita microbial mats (9–32%) and are present in lower abundance in Zacatón (6–20%) and Caracol (3–7.3%).

The archaeal diversity within the cenotes was high, with a reciprocal Simpson diversity index ($1/D = 447$) greater than that observed with the *Bacteria* ($1/D = 205$) (Sanger sequences only). A group of archaeal sequences were closely related to sequences (> 97% sequence identity) from public databases that were amplified from marine sediments (Heijs *et al*., 2008) and hydrothermal environments (Spear *et al*., 2005); however, there is no closely related cultured representative. Additional archaeal sequences from wall-associated microbial mats were related to known methanogens (*Methanomicrobia*) and anaerobic methane oxidizers (ANME-1); dissolved

Table 2. Sample information, including pyrosequencing coverage, microbial diversity and richness, of microbial mat and water samples analyzed in this study.

Cenote	Sample type	Sample depth	# pyrosequences	# OTUs (97%)	Chao1 richness estimate	Best parametric estimate	Coverage %	Simpson index (1/*D*)
Zacatón	Water	0	2662	140	305	563 (se134)	25	5.51
	Mat	1	2490	584	1187	1552 (se128)	38	41.88
	Mat	4	3680	1209	2659	4062 (se309)	30	147.94
	Mat	7	2697	718	1689	2656 (se267)	27	6.83
	Mat	10	3306	1444	3228	4670 (se259)	31	129.95
	Water	17	1810	204	527	674 (se 94)	30	10.50
	Mat	20	3948	1753	4003	5389 (se270)	33	578.90
	Mat	32	2705	901	2293	4011 (se507)	22	55.90
	Mat	58	1459	923	2280	5482 (se754)	17	623.09
	Mat	100	1761	655	1654	3022 (se430)	22	55.55
	Mat	114	1898	1129	3038	4984 (se475)	23	712.97
	Mat	195	2119	1056	2303	4305 (se460)	25	348.34
	Mat	273	2338	1186	2696	4485 (se395)	26	607.78
	Water	273	2396	392	962	1697 (se248)	23	12.66
La Pilita	Water	0	2109	192	437	569 (se 84)	34	8.13
	Mat	1	2177	268	626	849 (se104)	32	5.69
	Mat	6	2161	458	1051	1431 (se164)	32	9.31
	Mat	11	3271	1081	2480	3939 (se343)	27	55.33
	Mat	11	1766	635	1306	1836 (se206)	36	152.96
	Rock	11	2120	775	1746	2567 (se199)	30	115.90
	Green mat	12	2639	741	1807	2862 (se324)	26	20.53
	Red mat	12	2299	548	1092	1520 (se128)	36	23.80
	Mat	15	2624	604	975	2057 (se314)	29	26.18
	Mat	18	2949	558	1311	1942 (se184)	29	10.13
	Mat	20	2389	1036	2391	3550 (se275)	29	209.20
	Mat	22	2446	752	2255	3531 (se487)	21	17.50
	Mat	22	3370	1220	2766	3955 (se168)	31	194.95
	Mat	30	1818	873	2132	3643 (se417)	24	343.37
	Water	45	2695	182	543	581 (se 98)	31	4.96
	Mat	50	1956	992	2619	4935 (se605)	20	469.77
	Mat	80	4165	1612	3836	5276 (se311)	31	386.01
	Sediment	108	1777	616	1709	2377 (se318)	26	262.45
Caracol	Water	0	2375	519	1909	2478 (se336)	21	9.87
	Orange mat	7	2689	960	2353	3262 (se243)	29	127.67
	Green mat	7	3150	324	760	1097 (se140)	30	3.08
	Water	32	2422	157	327	429 (se 79)	37	7.30
Verde	Mat	1	2588	689	1400	1935 (se164)	36	20.30
Azufrosa	Water	1	3022	611	1586	2287 (se230)	27	13.07

methane was found throughout the water column of La Pilita and Caracol, ranging from 31 to 84 μM.

Bacterial and archaeal sequences were also amplified from a rock that protrudes horizontally from the wall in La Pilita at a depth of 11 m. Comparative sequence analysis suggests that similar types of phylotypes were obtained from the community in the rock core (endolith) and wall microbial mats at a similar depth (data not shown). Some bacterial phylotypes amplified from the endolith were not associated with microbial mat or water samples; however, all of the archaeal phylotypes were also observed in mat samples.

Wall microbial mat comparisons

Sequence distribution profiles from La Pilita and Zacatón mats identified changes in specific phylotype abundance with depth (Fig. 5). For example, the community from shallow regions of each cenote was dominated by cyanobacterial sequences, while the communities at depth were dominated by sequences from *Deltaproteobacteria*, *Chloroflexi* and *Nitrospirae*. The microbial mat samples from La Pilita showed significant percentages (≤ 7.9%) of sequences from the candidate phylum OP8, which were absent or present in lower abundance in Zacatón (≤ 1.1%) and Caracol (≤ 2.5%). These sequences are phylogenetically related to sequences amplified from marine sediments (Takeuchi *et al*., 2009).

When all samples from the five cenotes were clustered using the SDP method, water column samples from Zacatón, La Pilita, Caracol and Azufrosa clustered together (Fig. 6). Deep, surface-associated microbial mat samples from La Pilita and Zacatón clustered together, although there was separation by cenote (Fig. 6). The endolithic community from La Pilita clustered apart from both the water and mat samples.

A Mantel test (a test for statistical correlation between two matrices) was performed on both shallow and deep

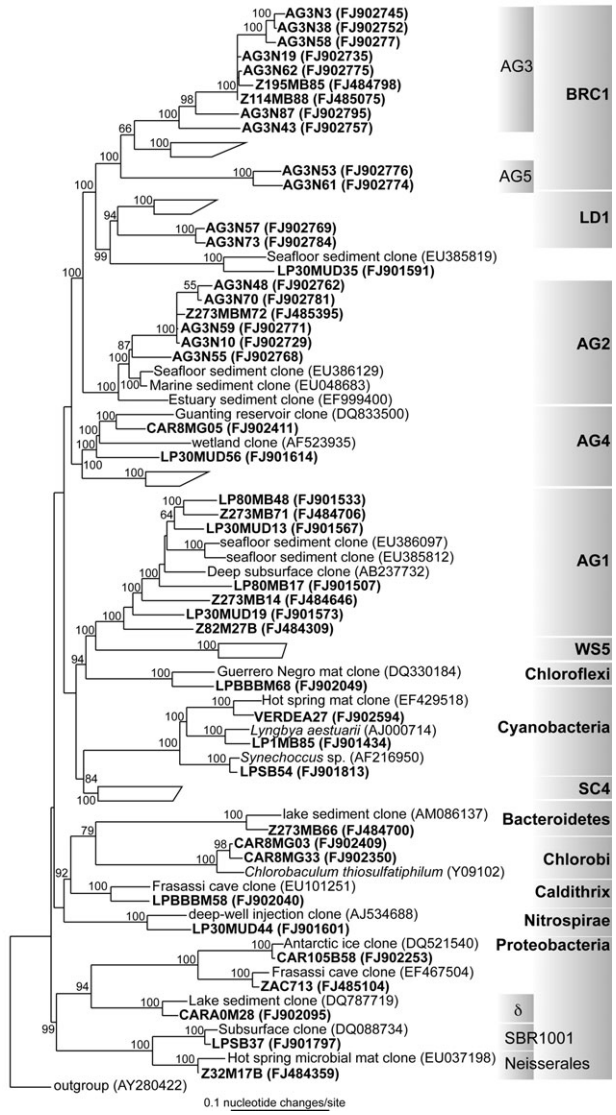

Fig. 4. A phylogenetic tree of the bacterial domain constructed with Mr Bayes. Numbers at nodes indicate posterior probabilities (only values > 50% shown). Bold sequences were amplified from Sistema Zacatón.

microbial mats to determine the controls on microbial community structure. The geographic location of the cenote was shown to significantly ($P = 0.001$) correlate with SDPs in deep mat samples; shallow samples, however, were not correlated with a specific cenote ($P = 0.376$) or depth ($P = 0.899$).

Fine resolution mat differences

One limitation to the DEPTHX sampling method is that the vehicle collects a 3–4 cm core of the wall-associated microbial mat material, which homogenizes any fine laminations. To determine the composition of laminated mats, one sample obtained by divers in Caracol was manually sorted and two separate dissected layers (orange and green) were frozen in separate tubes. Comparative analysis of pyrosequences amplified from these two samples showed that the green mat was dominated by sequences that group with *Chlorobi* (83% of library) while orange mat sequences were represented by *Deltaproteobacteria* (19%); *Chlorobi* sequences were much less dominant (13%) in this sample. Correspondingly, the diversity of the green mat was much lower than that of the orange mat (Table 2), and they did not cluster together with the SDP clustering method (Fig. 6).

Novel lineages

An initial survey of the microbial diversity in El Zacatón revealed the presence of two well-supported candidate phyla (Sahl *et al*., 2010). A phylogenetic analysis of sequences from the other cenotes of Sistema Zacatón identified additional sequences that group with Azufrosa Group 1 (AG1) (Fig. 4). An additional putative candidate phylum, Azufrosa Group 4 (AG4) (Fig. 4), was identified with additional sequence analysis. Polymerase chain reactions (PCR) with a new reverse primer (1427R), designed from two sequences from Azufrosa Group 2 (AG2), and the bacterial-specific forward primer 8F identified an additional 50 near-full-length sequences from AG2 and Azufrosa Group 3 (AG3). Although AG2 and AG3 are not closely related phylogenetically (~80%), the priming sites of the reverse primer differ only by one nucleotide on the 3′ end.

The 8F and 1427R primer pair also amplified sequences that showed high (~20%) intra-clade distance (ICD) (Robertson *et al*., 2009) to candidate divisions BRC1 (Derakshani *et al*., 2001) and LD1 (Freitag and Prosser, 2003) (Fig. 4). Additional sequencing will determine whether sequences from these clades (AG3, AG5, AG6) represent novel candidate phyla or expand the poorly characterized candidate phyla BRC1 and LD1.

Diversity, richness and coverage estimates

Coverage estimates based on parametric species richness estimates ranged from 17% to 38%. A Spearman correlation test identified a positive (0.837 for Zacatón and 0.929 for La Pilita) and significant correlation of the reciprocal Simpson diversity index with depth ($P = 0.001$ for Zacatón and $P = 0.002$ for La Pilita). The diversity of water column samples was significantly lower than microbial mat samples at every depth.

Functional gene identity and analysis

The geochemistry of the cenotes, combined with lineages associated with anaerobic metabolisms, directed the search into the presence of specific functional genes. Sequences from a hydrazine-oxidizing enzyme (HZO)

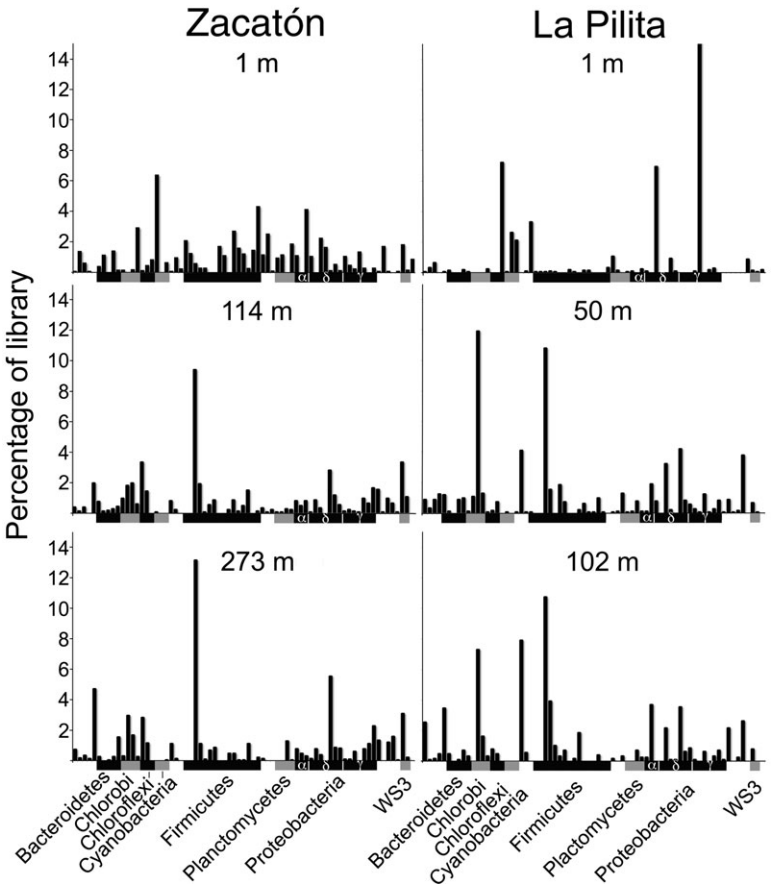

Fig. 5. Sequence distribution profiles of microbial mat samples taken from Zacatón and La Pilita at provided depths. The *x*-axis represents specific phylotypes and the *y*-axis is the percentage of the total library from each sample. Only phylotypes that represent more than 1% of any one library are shown. Sequences were obtained by pyrosequencing bacterial 16S rRNA genes.

gene were amplified from deep microbial mat samples in Zacatón (Fig. 1) and showed > 96% amino acid similarity with sequences from uncultured *Planctomycetes* bacteria in anammox enrichment cultures (Shimamura *et al*., 2007); 16S rRNA gene sequences were also amplified from deep microbial mat samples that grouped with *Planctomycetes*, although they failed to compose a significant percentage of the overall community. Methyl coenzyme M reductase alpha subunit (*mcrA*) gene sequences were amplified from sediment and wall mat samples in Zacatón and La Pilita (Fig. 1). Two clusters of *mcrA* sequences were formed through phylogenetic analysis (not shown); the sequences were associated with methanogenesis in a deep-sea mud volcano (Kormas *et al*., 2008), and sequences associated with anaerobic methane oxidation in river sediments (Hallam *et al*., 2004). When the methanogenic *mcrA* gene sequences were compared with sequenced genomes, they shared relatively low (84%) sequence similarity with *Methanothermobacter thermautotrophicus*, with only a partial local alignment.

Sequences that grouped with *Deltaproteobacteria*, which are commonly associated with sulfate reduction, directed the search into dissimilatory sulfite reductase alpha and beta subunit (*dsrAB*) gene sequences from Sistema Zacatón microbial mats. Amplified sequences were distantly (60–93% sequence identity) related to *dsrAB* gene sequences from deep aquifers (Bagwell *et al*., 2006), hot spring microbial mats (Dillon *et al*., 2007) and deep-sea sediments (Kaneko *et al*., 2007).

Sequences were also amplified using primers specific for the iron–iron hydrogenase (*hydA*) gene, but were only successfully amplified from shallow samples (< 10 m) in La Pilita, Zacatón and Azufrosa. A BLASTX (Altschul *et al*., 1990) homology search against the GenBank non-redundant protein database showed that many of the amplified sequences were most closely related (76–85% amino acid identity) to sequences from the Guerrero Negro hypersaline mat (Spear *et al*., 2003; Ley *et al*., 2006) using the same *hydA* primer pair (Boyd *et al*., 2009). However, additional sequences were related to cultured *Bacteria*, although remotely (< 70% identity), such as *Moorella thermoacetica*, *Desulforudis audaxviater* and *Thermodesulfovibrio yellowstonii*.

Discussion

Cenote morphology

Secondary calcium carbonate precipitation due to photosynthesis and/or sulfate reduction may explain the narrow

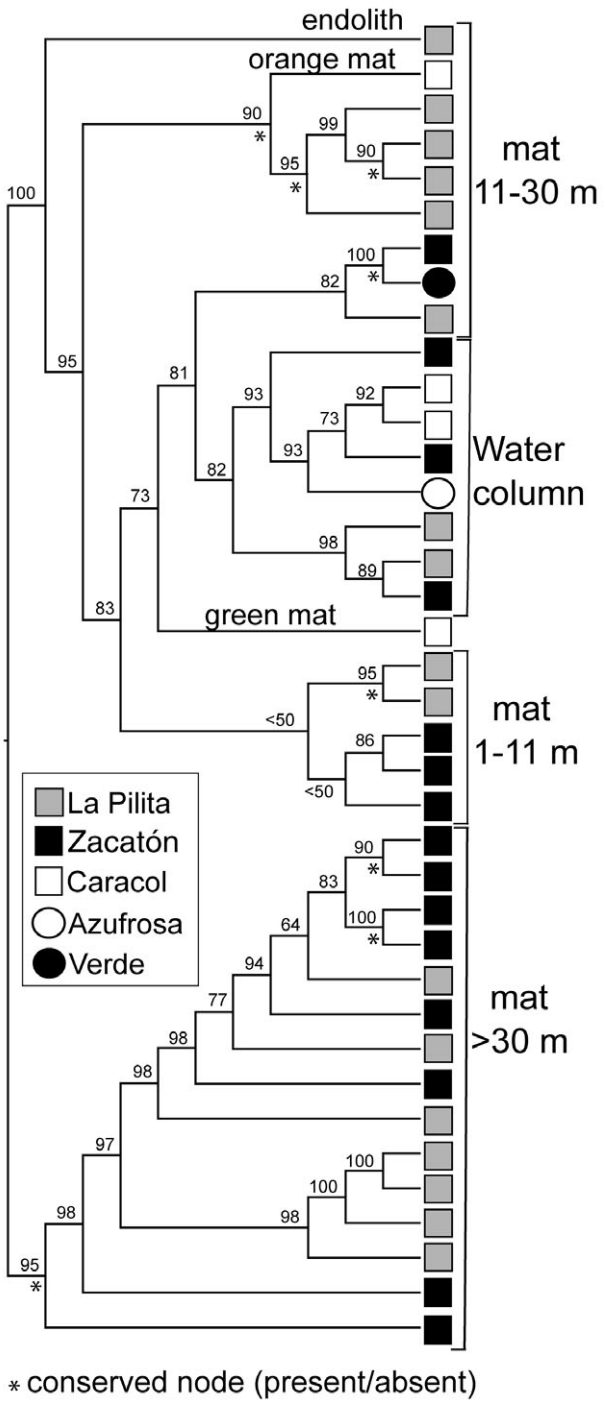

Fig. 6. A UPGMA tree of clustered environments from Sistema Zacatón. Numbers at nodes indicate jackknife support values.

neck morphology of both La Pilita and Caracol (Fig. 1). 16S rRNA sequence analysis of the subsurface endolithic community in the neck feature of La Pilita identified similar phylotypes with surrounding microbial mat communities, which further suggests that the surface-associated microbial mats are either (i) being lithified by calcium carbonate precipitation, or (ii) responsible for active biomineralizations that are in turn responsible for the gross morphology of travertine-filled sinkhole closure. This would then have relevance to the microbial influence on globally distributed travertine deposits. Thus, nearby travertine-filled sinkholes indicate that calcium carbonate precipitating cenotes, such as La Pilita and Azufrosa, may be excellent study systems for the closure of microbially dominated, travertine-filled sinkholes (Gary *et al*., 2006b).

Variation in water column microbial communities

Differences in the microbial community composition in the filterable fraction of the water column from cenotes Zacatón, La Pilita and Caracol were suspected by visual observation and validated by 16S rRNA comparative sequence analysis. Observed visual pigmentation differences may be due to red carotenoids (Zacatón and Caracol), which are common in *Chlorobi* (Takaichi *et al*., 1997), and green chlorophyll pigments (La Pilita) common in green algae. Small, but statistically significant differences in the geochemistry between Zacatón and La Pilita may help to explain the variation in microbial community structure in neighbouring cenotes.

One additional abiotic difference between La Pilita compared with Zacatón and Caracol is the amount of available light. Caracol is almost completely tree shaded and Zacatón has a 17 m limestone rim that limits the amount of available sunlight; La Pilita, however, has no tree shade, only a 1 m rim, and receives full, unshaded sunlight all year. The PAR values confirm this trend, with values at the surface of La Pilita (258 μmol photons s^{-1} m^{-2}) greater than those observed in Caracol (65 μmol photons s^{-1} m^{-2}) and Zacatón (132 μmol photons s^{-1} m^{-2}). One study showed that *Chlorobium phaeobacteroides* frequently prefers lower light conditions in the environment (Rimmer *et al*., 2008); sequences phylogenetically related to *C. phaeobacteroides* are prevalent in Caracol and Zacatón and absent in La Pilita. Additional experiments are needed to determine if light intensity is selecting for the dominance of unicellular algae, such as the observed *Chlamydomonas* spp. and corresponding chloroplast sequences, in La Pilita. The presence of *Cyanobacteria* sequences (Fig. 5) in shallow and deep regions of La Pilita indicate active mixing throughout the cenote even where light is unlikely to penetrate (Fig. 1).

Microbial mat community variation

Differences were observed in the abundance and distribution of phylotypes between deep microbial mat samples from La Pilita and Zacatón (Fig. 5). The results of a Mantel test showed that the genetic distance between samples from deeper mats from both cenotes correlate with the

geographic location of the cenote; samples from shallow mats do not correlate with cenote geography or depth, likely due to photosynthetic dependence. This suggests that variable geochemistry, particularly at depth, selects for different mat-associated microorganisms in systems in close geographical proximity. Indeed, the multiple factor analysis (Fig. 2) shows a clear distinction between 18 chemical parameters in La Pilita and Zacatón.

Phylogenetic classification of water and mat communities

Sequences amplified in this study were phylogenetically related to sequences from sulfurous lakes in Spain (Casamayor *et al*., 2000; Lliros *et al*., 2008), freshwater (Schwarz *et al*., 2007) and marine sediment (Martinez *et al*., 2006) samples, and sulfurous caves (Macalady *et al*., 2008). A truly comparable system to Sistema Zacatón in terms of geochemistry and depth has not been reported. Cuatro Ciénegas in northern Mexico contains shallow, travertine, hydrothermal water bodies (Breitbart *et al*., 2009), but the phylogenetic composition of microbes is much different than that of Sistema Zacatón cenotes (not shown).

SDP analysis and clustering methods

To better determine and relate phylogenetic sequence classifications between cenotes and with those obtained from other locales (above), we developed a SDP approach. In the comparison of microbial community structure between samples with similar phylum-level classifications, high-resolution taxonomic information can identify changes in specific phylotype abundance. Comparative analysis software can provide statistical significance to changes in bulk microbial community composition, but changes in specific lineages need to be observed with methods that consider taxonomic information. The Hugenholtz taxonomic system (DeSantis *et al*., 2006a), which recognizes the largest number of bacterial phyla, was found to provide the taxonomic detail needed to differentiate samples with similar phylum-level composition.

The SDP method was designed to test the relatedness of the taxonomic structure of multiple communities. The comparison of SDPs between communities does not rely on operational taxonomic unit (OTU)-based methods, and small sequencing errors common to pyrosequencing (Kunin *et al*., 2009; Huse *et al*., 2010) are silent if they do not change the taxonomic classification. As with any hypothesis testing method, the appropriate tool, such as the SDP method, UniFrac (Lozupone *et al*., 2006), mothur (Schloss *et al*., 2009) or metastats (White *et al*., 2009), should be chosen based on the specific nature of the hypothesis.

In this study, the comparison of SDPs identified changes in relative phylotype abundance across the photic zone of three cenotes. Furthermore, distinct differences in the sequence distribution in wall-associated microbial mat samples were observed across the depth profile of cenotes Zacatón and La Pilita. This comparative SDP method is a useful tool for the detailed comparison of samples from similar environments. SDPs can also be used for the identification of specific phylotype variation across geochemical gradients. One limitation of the SDP method is that it relies on the relative abundance of amplicons from PCR amplification. Numerous biases exist in the PCR amplification of 16S rRNA gene sequences; the most notable bias is that microorganisms contain variable numbers of the *rrn* operon (Klappenbach *et al*., 2001).

Pyrosequences from each sample of this study were clustered with the SDP clustering method. This method is advantageous for short sequences because they are anchored phylogenetically by near-full-length sequences, no phylogenetic tree is required, non-overlapping sequences can be compared, a multiple sequence alignment is not required, and there is no maximum allowable number of environments or sequences. One limitation of this method compared with the commonly used UniFrac clustering method (Lozupone *et al*., 2006), which incorporates phylogenetic information, is that it fails to take into account shared ancestry from related sequences between communities. Many nodes were not recovered when phylotypes were clustered with just presence/absence data (Fig. 6); this suggests that abundance information for each phylotype is necessary for community comparisons with the SDP method.

The SDP jackknife analysis of the Azufrosa sample cluster suggests that the structure of the cluster is stable (Fig. 6). The parsimony test implemented in mothur (Schloss *et al*., 2009) identified a significant correlation ($P < 0.001$) of the four primary groups shown in Fig. 6. An analysis of tree topologies with METATREE (Nye, 2008) from trees constructed with a variable number of omitted taxonomic classifications suggested that omitting 10% of all classifications over 1000 iterations generally retained the unmasked tree topology, but also identified cluster relationships with low support values based on only a few shared taxonomic classifications.

Functional gene analysis

The geochemistry of water analysed from the cenotes showed the presence of electron acceptors and donors, such as sulfate and methane, prevalent in anoxic environments. Furthermore, 16S rRNA gene sequences showed phylogenetic relatedness to sequences from organisms shown to facilitate anaerobic processes such as sulfate reduction, anaerobic ammonium oxidation and methano-

genesis. To validate these two observations, a targeted amplification and sequencing of published functional genes was performed. Sequences from genes associated with anaerobic metabolism were successfully amplified from deep-wall microbial mat samples. *mcrA* and *dsrA* gene sequences were most closely related to other environmental sequences and showed only low relatedness to sequences from cultured representatives. These results correlate well with 16S rRNA gene sequence analysis, which also showed only low relatedness to sequences from cultured methanogens and sulfate reducers respectively.

Sequences were also amplified from microbial lineages known to anaerobically oxidize methane and ammonium. HZO gene sequences were successfully amplified from deep mat samples in Zacatón, which suggests that anaerobic ammonium oxidation is a potential metabolism in deep mat samples. Additionally, the presence of 16S rRNA gene sequences phylogenetically related to known anaerobic methane oxidizers, the identification of *mcrA* genes related to those associated with anaerobic methane oxidation, and the presence of methane throughout the water column of the cenotes suggests this might be an important mechanism in the cycling of reduced organic material.

Iron–iron hydrogenase sequences, which have been associated with anaerobic processes such as fermentation and sulfate reduction in Bacteria (Vignais *et al*., 2001), were successfully amplified from shallow samples in Zacatón, La Pilita and Azufrosa. A comparative analysis of 16S rRNA gene sequences identified phylotypes present in samples with positive *hydA* amplification as well as phylotypes absent in samples with no *hydA* amplification. The low relatedness of these hydrogenase sequences to genome sequences in public databases, however, prevents any attempt to associate a hydrogenase sequence with a particular organism or clade, and further reflects the paucity of information that regards the extant of environmental hydrogenases in general.

The results of a partial functional gene analysis show that microbes in deep microbial mat samples of the cenotes contain the genetic machinery to conduct metabolisms, such as methanogenesis and sulfate reduction, common in anoxic environments. The results further indicate that DNA samples collected from these cenotes subjected to metagenomic and genomic approaches will likely provide revealing insight into the complex geobiological dynamic of the subsurface environment.

Conclusions

A comparative sequence analysis of Sistema Zacatón cenote samples identified novel microbial diversity. Full-length 16S rRNA Sanger gene sequence analysis identified three novel bacterial candidate phyla as well as other groups that showed high ICD with poorly characterized candidate phyla; additional sequencing is needed for groups with poor sequence representation to enhance the reconstruction of phylogenetic relationships. A comparative analysis of pyrosequences, combined with a new SDP method, identified abundance changes of specific phylotypes across the photic zone of the water column, and the depth profile of cenote wall-associated microbial mats. Differences in microbial communities between cenotes in close geographical proximity were attributed to slight, but distinct, geochemical variation.

Functional genes associated with anaerobic metabolisms were successfully amplified from deep microbial mat samples. These genes are associated with microbial lineages whose 16S rRNA gene sequences were successfully amplified from wall mats, although a direct correlation of genes with microbes could not be made. Amplified functional gene sequences generally showed low identity to previously sequenced genes, which correlates well with the generally low identity of sequences from Sistema Zacatón to those in public databases based on comparative 16S rRNA gene sequence analysis. The presence of these metabolic genes provides insight into the types of potential microbial metabolisms in these systems. This information combined with the geochemistry and SSU gene sequence analyses allows us to better understand the microbial ecology in these previously unexplored environments.

Experimental procedures

Site location

The cenotes surveyed in this study are located in a localized karst region of north-eastern Mexico (22°59′35.10″N, 98°09′55.96″W) known as Sistema Zacatón. Five water-filled sinkholes (Zacatón, La Pilita, Caracol, Verde, Azufrosa) are located in the underwater cave system (Fig. 1), with additional cenotes and travertine-filled sinkholes nearby.

Sample collection

Water column and wall-microbial mat samples from both Zacatón and La Pilita were collected by a spring-loaded sampling arm on DEPTHX in the Spring of 2007. Sample collection and preparation have been discussed previously (Sahl *et al*., 2010). Mat samples from a hydrothermal vent in cenotes Verde and Caracol were manually collected by divers in sterile 50 ml falcon tubes in November 2008. Divers also collected a rock from 11 m deep in La Pilita to determine if SSU rRNA genes could be successfully amplified and sequenced. The rock was pared aseptically (Kieft *et al*., 2007; Sahl *et al*., 2008) to ensure that only the centre was isolated for DNA extraction for further amplification and sequencing.

Additional samples were collected to determine the transition in microbial communities across the photic zone with a 2 l Niskin bottle (Ben Meadows); photic zone samples were all

Table 3. Polymerase chain reaction (PCR) primers used in this study.

Name	Sequence (5′–3′)	Specificity	Reference
4Fa	TCC GGT TGA TCC TGC CRG	Archaeal SSU	Hershberger *et al*. (1996)
8F	AGA GTT TGA TCC TGG CTC AG	Bacterial SSU	Lane (1991)
21Fa	TTC CGG TTG ATC CYG CCG GA	Archaeal SSU	Reysenbach and Pace (1995)
338R	CAT GCT GCC TCC CGT AGG AGT	Bacterial SSU	Amann (1990)
515F	GTG CCA GCM GCC GCG GTA A	Universal SSU	Lane (1991)
1195RE	GCA TCA CAG ACC TG	Eukaryal SSU	Dawson and Pace (2002)
1427R	CAC TCA ACT CAG GTC ATG	AG2	This study
1492R	GGT TAC CTT GTT ACG ACT T	Universal SSU	Lane (1991)
Ana-hzo1F	TGT GCA TGG TCA ATT GAA AG	*hzo*	Quan *et al*. (2008)
Ana-hzo2R	ACC TCT TCW GCA GGT GCA T	*hzo*	Quan *et al*. (2008)
mcrAF	GGT GGT GTM GGA TTC ACA CAR TAY GCW ACA GC	*mcrA*	Luton *et al*. (2002)
mcrAR	TTC ATT GCR TAG TTW GGR TAG TT	*mcrA*	Luton *et al*. (2002)
Fe-Fe 272F	GCH GAY MTB ACH ATW ATG GAR GA	*hydA*	Boyd *et al*. (2009)
Fe-Fe 427R	GCH GCY TCC ATD ACD CCD CCN GT	*hydA*	Boyd *et al*. (2009)
DSR2F	CAT GAA GGA YGA CAT CAA	*dsrAB*	Wagner *et al*. (1998)
DSR4R	GTG TAG CAG TTA CCG CA	*dsrAB*	Wagner *et al*. (1998)
1F1	CAG GAY GAR CTK CAC CG	*dsrAB*	Dhillon *et al*. (2003)
1R1	CCC TGG GTR TGR AYR AT	*dsrAB*	Dhillon *et al*. (2003)

collected on the same day in November 2008. Light intensity (μmol photons s^{-1} m^{-2}) was recorded with a LI-COR LI-193 submersible quantum sensor (LI-COR, Lincoln, NE). When the light intensity was no longer detectable, a sample was collected with the Niskin sampler. The Niskin bottle was connected directly to a peristaltic pump with flexible tubing and filtered through a 0.2 μm polyethersulfone filter (Pall Corporation, East Hills, NY) stored in a high-pressure stainless steel manifold (Millipore, Billerica, MA). Filters were cut with flame-sterilized scissors, placed in cryovials and immediately frozen in liquid nitrogen.

Water chemistry

Water chemistry sampling methods have been described previously (Sahl *et al*., 2010). Briefly, 15 ml of water was filtered (0.45 um) in the field and acidified for inductively coupled plasma-atomic emission spectroscopy (ICP-AES) analysis. An additional sample of field water was stored on ice in the field and analysed in the lab for total organic carbon, ions and alkalinity. Other water column parameters, such as pH, temperature, specific conductivity and redox potential, were calculated with an onboard sensor on the DEPTHX vehicle (Sahl *et al*., 2010).

For dissolved methane analysis, water samples were collected in 20 ml syringes. A small volume of air was then injected into the syringe and the sample was left to equilibrate. The syringe headspace was injected into a field gas chromatograph and the mass of methane was calculated by comparison to known standards. The dissolved methane concentrations were then calculated with Henry's Law.

A multiple factor analysis was performed in XLSTAT by the comparison of 18 water chemistry variables in 14 samples from La Pilita and Zacatón (seven from each). The eigenvalues were plotted for the first two factors, which explained > 99.78% of the observed variation.

DNA extraction, PCR, cloning

Bulk community DNA was extracted using the Mobio PowerSoil kit (Mobio, Carlsbad, CA), with a 1 min bead-beating step substituted for the 10 min vortex step. The La Pilita rock sample was prepared and DNA was extracted as reported previously (Sahl *et al*., 2008). PCR reactions were performed with a suite of SSU primers (Table 3) including bacterial-specific forward primer 8F (Lane, 1991), universal forward primer 515F (Lane, 1991), archaeal-specific primers 4Fa (Hershberger *et al*., 1996) and 21Fa (Reysenbach and Pace, 1995), and universal reverse primers 1391R (Lane, 1991) and 1492R (Lane, 1991), using protocols described elsewhere (Sahl *et al*., 2010). An additional reverse primer, 1427R, was designed in ARB (Ludwig *et al*., 2004) from near-full-length gene sequences from AG2, a novel candidate phylum previously identified in Zacatón (Sahl *et al*., 2010); this primer was used to obtain additional, near-full-length sequences from this group. PCR amplicons were gel-purified (Millipore, Billerica, MA), TOPO TA cloned using electrocompetent *Escherichia coli* (Invitrogen), and sequenced on a MegaBACE 1000 dye-terminating sequencer.

Sanger sequence analysis

Sanger sequence reads were assembled with XplorSeq (Frank, 2008), chimera-checked with Mallard (Ashelford *et al*., 2006) and aligned with the NAST aligner (DeSantis *et al*., 2006b). Sequence alignments were manually corrected using ARB on columns unmasked with the Lanemask filter (Lane, 1991). Bacterial and archaeal sequences were then inserted into the ARB dendogram using the latest Greengenes (DeSantis *et al*., 2006a) database release. Relevant bacterial and archaeal sequences, including those showing closely related phylogeny, were then exported with the Lanemask filter. Phylogenetic trees were constructed with Mr Bayes (Huelsenback and Ronquist, 2001).

Pyrosequencing and sequence analysis

To further explore the microbial community composition of cenote microbial mats, a barcoded primer amplicon approach was employed for sample multiplexing (Sogin *et al*., 2006) on

the 454 FLX platform (454 Life Sciences) with bacterial-specific 16S rRNA primers 8F (Lane, 1991) and 338R (Amann *et al*., 1995). Custom Java script was used to sort the raw sequences, which also deleted sequences with ambiguous nucleotide characters (Ns) and sequences that were shorter than the mean minus two standard deviations of the normal distribution of sequence length (Huse *et al*., 2007). Sequences were aligned with the RDP pyrosequencing pipeline (Cole *et al*., 2009), which uses the Infernal aligner (Nawrocki *et al*., 2009) and a consensus secondary structure model (Cannone and Subramanian, 2002). Sequences were clustered into OTUs using the RDP complete linkage clustering algorithm (Cole *et al*., 2009) at a distance of 3%. One representative sequence from each cluster was NAST aligned and classified using the online Greengenes classification programme; the subsequent taxonomic classification was then correlated with the OTU abundance information.

SDP analysis and clustering method

Identical taxonomic classifications were grouped with custom Java script and the abundance of each classification was converted to a percentage and plotted. The plot, which represents each taxonomic classification on the *x*-axis and relative abundance on the *y*-axis, constitutes the SDP (see Figs 2 and 5). This method differs from other community similarity calculators in that it is comparing the shape of pairwise SDP curves and not differences in OTU membership. Custom scripts to run the SDP method are included as Supporting File S1.

A clustering method was developed from the sequence distribution profiles. The number of sequences from each taxonomic classification from each environment was first converted to a percentage to normalize variable sequence numbers. A pairwise comparison was then made from the taxonomic classification and abundance information for each sample. The smallest and largest percentage for each classification from each pairwise comparison, which represents the actual amount and maximum possible amount of shared abundance, respectively, was then separately summed; the summed minimum percentages between two environments (A and B) were then divided by the summed maximum percentages and a ratio (R_{AB}) was calculated (Eq. 1):

$$R_{AB} = \frac{\sum_{i=1}^{n} (\text{min percentage}_{AB})}{\sum_{i=1}^{n} (\text{max percentage}_{AB})} \quad (1)$$

Two samples which share all their taxonomic classifications in the same abundance would have a maximum R_{AB} value of 1, while those that shared no classifications would have an R_{AB} value of 0. The pairwise comparisons were converted to a distance value (D) by subtracting the R_{AB} value from 1, and an Unweighted Pair Group Method with Arithmetic Mean (UPGMA) cluster was calculated with the neighbour-joining method in Phylip (Felsenstein, 1989). A distance matrix was made for each pairwise D value and the distance matrix was compared with a proximity matrix in a Mantel test (Mantel, 1967). Correlations of distance with either depth or the geographic location of the cenote were tested.

A jackknife analysis was performed with the clustering method by randomly omitting a defined number of taxonomic classifications and calculating a distance matrix; this process was repeated 1000 times. The number of classifications to omit was empirically determined by the point at which the topology of the cluster changes from the unmasked tree; multiple tree topologies were then compared with METATREE (Nye, 2008). When 10% of the classifications were omitted, the topology of the cluster changed from the original topology, but still retained the major nodes of the cluster topology. One thousand UPGMA trees were generated and the consensus tree was calculated with CONSENSE from the Phylip package (Felsenstein, 1989). The cluster was also calculated with phylotype presence or absence, to test the effect of PCR bias on the cluster topology; nodes conserved following presence/absence analysis are shown in Fig. 6.

A significance test was also conducted using the SDP method. The taxonomic classifications were randomly shuffled between each pairwise comparison 1000 times and a distance value ($D_{shuffled}$) was calculated. A statistical *P*-value was calculated by the comparison of the number of times that $D_{shuffled}$ was greater than or equal to the original ($D_{original}$) divided by the total number of permutations (Eq. 2). The Bonferroni correction, which was obtained by multiplying the raw *P*-value by the number of samples, was used to correct for multiple comparisons.

$$P = \frac{1}{1000} \sum_{i=1}^{1000} \#(D_{shuffled} > D_{original}) \quad (2)$$

Diversity and coverage estimates

The de-replicated data from the RDP clustering tool were processed with parametric species richness estimators (Hong *et al*., 2006; Jeon *et al*., 2006) to find the best estimate with the lowest standard error and highest goodness of fit. Coverage was calculated by dividing the number of observed OTUs by the best parametric estimate. For comparison, the non-parametric species richness estimator Chao1 was calculated with mothur (Schloss *et al*., 2009). The reciprocal Simpson diversity estimate (Simpson, 1949) was calculated by EstimateS (Colwell, 2005).

Functional genes

Polymerase chain reaction primers specific for functional genes relevant to geochemical cycles in anoxic environments (Table 3) were used on genomic DNAs. These included the primers Ana-hzo1F and Ana-hzo2R (Quan *et al*., 2008), which amplify HZO genes necessary for anaerobic ammonium oxidation; primers mcrAF and mcrAR (Luton *et al*., 2002), which amplify the methyl coenzyme M reductase alpha subunit (*mcrA*) gene; and primers Fe-Fe 272F and Fe-Fe 427R which amplify the gene encoding the [FeFe]-hydrogenase large subunit (*hydA*) (Boyd *et al*., 2009), which is associated with, but not limited to, anaerobic processes in *Bacteria*, such as fermentation and sulfate reduction (Vignais *et al*., 2001; Vignais and Billoud, 2007). PCR primers DSR2F and DSR4 were selected for the amplification of the

dissimilatory sulfite reductase alpha and beta subunit (*dsrAB*) gene (Wagner *et al*., 1998). Primers 1FI and 1R1 (Dhillon *et al*., 2003) were used to capture additional environmental *dsrAB* gene diversity.

Nucleotide accession numbers

A total of 1830 Sanger sequences were submitted to GenBank under Accession No. FJ901344–FJ903173. 16S rRNA pyrosequence reads were submitted to the GenBank short read archive under Accession No. SRA008306. Sanger 16S rRNA sequences were named according to the cenote from which they were amplified (Z = Zacatón, LP = La Pilita, C = Caracol, AZU = Azufrosa), the depth of the sample (m) and the type of sample (F = water column filter, M or no suffix indicates microbial mat). Sequences with the AG3N prefix were amplified with the reverse primer designed to target AG3 sequences.

Acknowledgements

The authors thank Alejandro Dávila for access to Rancho la Azufrosa. Many thanks go to the DEPTHX team including Bill Stone, John Kerr, George Cantor, Dom Jonak, David Wettergreen, Nathaniel Fairfield, Nora and Antonio Fregoso, Robin Gary, and the Southwest Research Institute. Thanks go to Philip Bennett for providing the dissolved methane data and Chase Williamson and Chuck Pepe-Ranney for their helpful comments. The authors would like to thank the Pace Lab at the University of Colorado for help operating the MegaBACE 1000 and the Consortium for Comparative Genomics at the University of Colorado, Denver, for operating the 454 pyrosequencer. All custom Java script used in sequence analysis was written by Erik Sahl. This work was funded by a NASA ASTEP Grant #NNG04GC09G.

References

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990) Basic local alignment search tool. *J Mol Biol* **215:** 403–410.

Amann, R.I., Ludwig, W., and Schleifer, K.H. (1995) Phylogenetic identification and in situ detection of individual microbial cells without cultivation. *Microbiol Rev* **59:** 143–169.

Amann, R., Binder, B.J., Olson, R.J., Chisholm, S.W., Devereux, R., and Stahl, D.A. (1990) Combination of 16S rRNA-targeted oligonucleotide probes with flow cytometry for analyzing mixed microbial populations. *Appl Environ Microbiol* **56:** 1919–1925.

Ashelford, K.E., Chuzhanova, N.A., Fry, J.C., Jones, A.J., and Weightman, A.J. (2006) New screening software shows that most recent large 16S rRNA gene clone libraries contain chimeras. *Appl Environ Microbiol* **72:** 5734–5741.

Bagwell, C.E., Liu, X., Wu, L., and Zhou, J. (2006) Effects of legacy nuclear waste on the compositional diversity and distributions of sulfate-reducing bacteria in a terrestrial subsurface aquifer. *FEMS Microbiol Ecol* **55:** 424–431.

Boyd, E.S., Spear, J.R., and Peters, J.W. (2009) [FeFe] hydrogenase genetic diversity provides insight into molecular adaptation in a saline microbial mat community. *Appl Environ Microbiol* **75:** 4620–4623.

Breitbart, M., Hoare, A., Nitti, A., Siefert, J., Haynes, M., Dinsdale, E., *et al*. (2009) Metagenomic and stable isotopic analyses of modern freshwater microbialites in Cuatro Cienegas, Mexico. *Environ Microbiol* **11:** 16–34.

Cannone, J.J., and Subramanian, S. (2002) The comparative RNA web (CRW) site: an online database of comparative sequence and structure information for ribosomal, intron, and other RNAs. *BMC Bioinformatics* **3:** 2.

Casamayor, E.O., Schafer, H., Baneras, L., Pedros-alio, C., and Muyzer, G. (2000) Identification of and spatio-temporal differences between microbial assemblages from two neighboring sulfurous lakes: comparison by microscopy and denaturing gradient gel electrophoresis. *Appl Environ Microbiol* **66:** 499–508.

Cole, J.R., Wang, Q., Cardenas, E., Fish, J., Chai, B., Farris, R.J., *et al*. (2009) The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. *Nucleic Acids Res* **37:** D141–D145.

Colwell, R.K. (2005) EstimateS: statistical estimation of species richness and shared species from samples. Version 8. *User's guide and application* [WWW document]. URL http://viceroy.eeb.uconn.edu/estimates.

Dawson, S.C., and Pace, N.R. (2002) Novel kingdom-level eukaryotic diversity in anoxic environments. *Proc Natl Acad Sci* **99:** 8324–8329.

Derakshani, M., Lukow, T., and Liesack, W. (2001) Novel bacterial lineages at the (sub)division level as detected by signature nucleotide-targeted recovery of 16S rRNA genes from bulk soil and rice roots of flooded rice microcosms. *Appl Environ Microbiol* **67:** 623–631.

DeSantis, T.Z., Hugenholtz, P., Larsen, N., Rojas, M., Brodie, E.L., Keller, K., *et al*. (2006a) Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl Environ Microbiol* **72:** 5069–5072.

DeSantis, T.Z., Hugenholtz, P., Keller, K., Brodie, E.L., Larsen, N., Piceno, Y.M., *et al*. (2006b) NAST: a multiple sequence alignment server for comparative analysis of 16S rRNA genes. *Nucleic Acids Res* **34:** W394–W399.

Dhillon, A., Teske, A., Dillon, J., Stahl, D.A., and Sogin, M.L. (2003) Molecular characterization of sulfate-reducing bacteria in the Guaymas Basin. *Appl Environ Microbiol* **69:** 2765–2772.

Dillon, J.G., Fishbain, S., Miller, S.R., Bebout, B.M., Habicht, K.S., Webb, S.M., and Stahl, D.A. (2007) High rates of sulfate reduction in a low-sulfate hot spring microbial mat are driven by a low level of diversity of sulfate-respiring microorganisms. *Appl Environ Microbiol* **73:** 5218–5226.

Dinsdale, E.A., Edwards, R.A., Hall, D., Angly, F., Breitbart, M., Brulc, J.M., *et al*. (2008) Functional metagenomic profiling of nine biomes. *Nature* **452:** 629–632.

Fairfield, N., Kantor, G.A., and Wettergreen, D. (2006) *IEEE Internal Conference on Robotics and Automation (ICRA)*, Orlando, FL, USA.

Fairfield, N., Kantor, G.A., and Wettergreen, D. (2007) Real-time SLAM with octree evidence grids for exploration in underwater tunnels. *J Field Robot* **24:** 3–21.

Felsenstein, J. (1989) Phylogeny inference package (version 3.2). *Cladistics* **5:** 164–166.

Frank, D.N. (2008) XplorSeq: a software environment for integrated management and phylogenetic analysis of metagenomic sequence data. *BMC Bioinformatics* **9:** 420.

Freitag, T.E., and Prosser, J.I. (2003) Community structure of ammonia-oxidizing bacteria within anoxic marine sediments. *Appl Environ Microbiol* **69:** 1359–1371.

Gary, M.O. (2009) *Karst Hydrogeology and Speleogenesis of Sistema Zacatón*, Bulletin 21. Austin, TX, USA: Association for Mexican Cave Studies.

Gary, M.O., Sharp, J.M., Jr, Harmon, R.S., and Wicks, C. (2006a) Volcanogenic karstification of Sistema Zacatón, Mexico. In *Perspectives on Karst Geomorphology, Hydrology, and Geochemistry – A Tribute Volume to Derek C Ford and William B White*. Geological Society of America Special Paper 404, pp. 79–89.

Gary, M.O., Halihan, T., Sharp, J.M., Jr, Mouri, S., and Thorstad, J. (2006b) Electrical resistivity imaging of travertine capped sinkholes: deep lakes with lids. In Philadelphia, pp. Paper No. 218-214-Paper No. 218-214.

Gary, M.O., Fairfield, N., Stone, W.C., Wettergreen, D., Kantor, G.A., and Sharp, J.M., Jr (2008) 3D mapping and characterization of sistema Zacatón from DEPTHX (DEep Phreatic THermal eXplorer). In *Proceedings of KARST08: 11th Sinkhole Conference ASCE*.

Gary, M.O., Halihan, T., and Sharp, J.M., Jr (2009) Detection of sub-travertine lakes using electrical resistivity imaging, Sistema Zacatón, Mexico. In *15th International Congress on Speleology Proceedings*. Kerrville, TX, USA, pp. 575–579.

Gihring, T.M., Moser, D.P., Lin, L.H., Davidson, M., Onstott, T.C., Morgan, L., *et al*. (2006) The distribution of microbial taxa in the subsurface water of the Kalahari Shield, South Africa. *Geomicrobiol J* **23:** 415–430.

Hallam, S.J., Putnam, N., Preston, C.M., Detter, J.C., Rokhsar, D., Richardson, P.M., and DeLong, E.F. (2004) Reverse methanogenesis: testing the hypothesis with environmental genomics. *Science* **305:** 1457–1462.

Heijs, S.K., Laverman, A.M., Forney, L.J., Hardoim, P.R., and van Elsas, J.D. (2008) Comparison of deep-sea sediment microbial communities in the Eastern Mediterranean. *FEMS Microbiol Ecol* **64:** 362–377.

Hershberger, K.L., Barns, S.M., Reysenbach, A.L., Dawson, S.C., and Pace, N.R. (1996) Wide diversity of Crenarchaeota. *Nature* **384:** 420–420.

Hong, S.H., Bunge, J., Jeon, S.O., and Epstein, S.S. (2006) Predicting microbial species richness. *Proc Natl Acad Sci USA* **103:** 117–122.

Huber, J.A., Welch, D.M., Morrison, H.G., Huse, S.M., Neal, P.R., Butterfield, D.A., and Sogin, M.L. (2007) Microbial population structures in the deep marine biosphere. *Science* **318:** 97–100.

Huber, J.A., Morrison, H.G., Huse, S.M., Neal, P.R., Sogin, M.L., and Welch, D.B.M. (2009) Effect of PCR amplicon size on assessments of clone library microbial diversity and community structure. *Environ Microbiol* **11:** 1292–1302.

Huelsenback, J.P., and Ronquist, F. (2001) MRBAYES: Bayesian inference of phylogeny. *Bioinformatics* **17:** 754–755.

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

Huse, S.M., Welch, D.B.M., Morrison, H.G., and Sogin, M.L. (2010) Ironing out the wrinkles in the rare biosphere through improved OTU clustering. *Environ Microbiol* **12:** 1889–1898.

Jeon, S.O., Bunge, J., Stoeck, T., Barger, K.J.A., Hong, S.H., and Epstein, S.S. (2006) Synthetic statistical approach reveals a high degree of richness of microbial eukaryotes in an anoxic water column. *Appl Environ Microbiol* **72:** 6578–6583.

Kaneko, R., Hayashi, T., Tanahashi, M., and Naganuma, T. (2007) Phylogenetic diversity and distribution of dissimilatory sulfite reductase genes from deep-sea sediment cores. *Mar Biotechnol* **9:** 429–436.

Kieft, T.L., Phelps, T.J., Fredrickson, J.K., Hurst, C.J., Crawford, R.L., Knudsen, G.R., *et al*. (2007) Drilling, coring, and sampling subsurface environments. In *Manual of Environmental Microbiology*, 3rd edn. Hurst, C.J., Crawford, R.L., Garland, J.L., Lipson, D.A., Mills, A.L., and Stetzenbach, L.D. (eds). Washington, DC, USA: ASM Press, pp. 799–818.

Klappenbach, J.A., Saxman, P.R., Cole, J.R., and Schmidt, T.M. (2001) rrndb: the ribosomal RNA operon copy number database. *Nucleic Acids Res* **29:** 181–184.

Kormas, K.A., Meziti, A., Dahlmann, A., De Lange, G.J., and Lykousis, V. (2008) Characterization of methanogenic and prokaryotic assemblages based on *mcrA* and 16S rRNA gene diversity in sediments of the Kazan mud volcano (Mediterranean Sea). *Geobiology* **6:** 450–460.

Krajick, K. (2007) Robot seeks new life – and new funding – in the abyss of Zacatón. *Science* **315:** 322–324.

Kunin, V., Engelbrektson, A., Ochman, H., and Hugenholtz, P. (2009) Wrinkles in the rare biosphere: pyrosequencing errors can lead to artificial inflation of diversity estimates. *Environ Microbiol* **12:** 118–123.

Lane, D.J. (1991) 16S/23S rRNA sequencing. In *Nucleic Acid Techniques in Bacterial Systematics*. Stackebrandt, E., and Goodfellow, M. (eds). Chichester, UK: Wiley and Sons, pp. 329–329.

Ley, R.E., Harris, J.K., Wilcox, J., Spear, J.R., Miller, S.R., Bebout, B.M., *et al*. (2006) Unexpected diversity and complexity of the Guerrero Negro hypersaline microbial mat. *Appl Environ Microbiol* **72:** 3685–3695.

Liu, Z., Lozupone, C., Hamady, M., Bushman, F.D., and Knight, R. (2007) Short pyrosequencing reads suffice for accurate microbial community analysis. *Nucleic Acids Res* **35:** e120.

Liu, Z., DeSantis, T.Z., Andersen, G.L., and Knight, R. (2008) Accurate taxonomy assignments from 16S rRNA sequences produced by highly parallel pyrosequencers. *Nucleic Acids Res* **36:** e120.

Lliros, M., Casamayor, E.O., and Borrego, C. (2008) High archaeal richness in the water column of a freshwater sulfurous karstic lake along an interannual study. *FEMS Microbiol Ecol* **66:** 331–342.

Lozupone, C., Hamady, M., and Knight, R. (2006) UniFrac-an online tool for comparing microbial community diversity in a phylogenetic context. *Bioinformatics* **7:** 371.

Ludwig, W., Strunk, O., Westram, R., Richter, L., Meier, H., Yadhukumar, *et al*. (2004) ARB: a software environment for sequence data. *Nucleic Acids Res* **32:** 1363–1371.

Luton, P.E., Wayne, J.M., Sharp, R.J., and Riley, P.W. (2002) The *mcrA* gene as an alternative to 16S rRNA in the phylogenetic analysis of methanogen populations in landfill. *Microbiology* **148:** 3521–3530.

Macalady, J.L., Dattagupta, S., Schaperdoth, I., Jones, D.S., Druschel, G.K., and Eastman, D. (2008) Niche differentiation among sulfur-oxidizing bacterial populations in cave waters. *ISME J* **2:** 590–601.

Mantel, N. (1967) The detection of disease clustering and a generalized regression approach. *Cancer Res* **27:** 209–220.

Martinez, R.J., Mills, H.J., Story, S., and Sobecky, P.A. (2006) Prokaryotic diversity and metabolically active microbial populations in sediments from an active mud volcano in the Gulf of Mexico. *Environ Microbiol* **8:** 1783–1796.

Nawrocki, E.P., Kolbe, D.L., and Eddy, S.R. (2009) Infernal 1.0: inference of RNA alignments. *Bioinformatics* **25:** 1335–1337.

Nelson, D.M., Ohene-Adjei, S., Sheng Hu, F., Cann, I.K.O., and Mackie, R.I. (2007) Bacterial diversity and distribution in the Holocene sediments of a northern temperate lake. *Microb Ecol* **54:** 252–263.

Nye, T.M.W. (2008) Trees of trees: an approach to comparing multiple alternative phylogenies. *Syst Biol* **57:** 785–794.

Quan, Z.X., Rhee, S.K., Zuo, J.E., Yang, Y., Bae, J.W., Park, J.R., *et al*. (2008) Diversity of ammonium-oxidizing bacteria in a granular sludge anaerobic ammonium-oxidizing (anammox) reactor. *Environ Microbiol* **10:** 3130–3139.

Reysenbach, A.L., and Pace, N.R. (1995) Reliable amplification of hyperthermophilic archaeal 16S rRNA genes by the polymerase chain reaction. In *Archaea: A Laboratory Manual-Thermophiles*. Adams, M.W.W. (ed.). New York, USA: Cold Spring Harbour Laboratory Press, pp. 101–107.

Rimmer, A., Ostrovsky, I., and Yacobi, Y.Z. (2008) Light availability for *Chlorobium phaeobacteroides* development in Lake Kinneret. *J Plankton Res* **30:** 765–776.

Robertson, C.E., Spear, J.R., Harris, J.K., and Pace, N.R. (2009) Diversity and stratification of archaea in a hypersaline microbial mat. *Appl Environ Microbiol* **75:** 1801–1810.

Roesch, L.F.W., Fulthorpe, R.R., Riova, A., Casella, G., Hadwin, A.D.W., Daroub, S.H., *et al*. (2007) Pyrosequencing enumerates and contrasts soil microbial diversity. *ISME J* **1:** 283–290.

Sahl, J.W., Schmidt, R., Swanner, E.D., Mandernack, K.W., Templeton, A.S., Kieft, T.L., *et al*. (2008) Subsurface microbial diversity in deep-granitic-fracture water in Colorado. *Appl Environ Microbiol* **74:** 143–152.

Sahl, J.W., Fairfield, N., Harris, J.K., Wettergreen, D., Stone, W.C., and Spear, J.R. (2010) Novel microbial diversity retrieved by autonomous robotic exploration of the world's deepest vertical phreatic sinkhole. *Astrobiology* **10:** 201–213.

Schloss, P.D., Westcott, S.L., Ryabin, T., Hall, J.R., Hartmann, M., Hollister, E.B., *et al*. (2009) Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl Environ Microbiol* **75:** 7537–7541.

Schwarz, J.I.K., Eckert, W., and Conrad, R. (2007) Community structure of *Archaea* and *Bacteria* in a profundal lake sediment Lake Kinneret (Israel). *Syst Appl Microbiol* **30:** 239–254.

Shimamura, M., Nishiyama, T., Shigetomo, H., Toyomoto, T., Kawahara, Y., Furukawa, K., and Fujii, T. (2007) Isolation of a multiheme protein with features of a hydrazine-oxidizing enzyme from an anaerobic ammonium-oxidizing enrichment culture. *Appl Environ Microbiol* **73:** 1065–1072.

Simpson, E.H. (1949) Measurement of diversity. *Nature* **163:** 688–688.

Sogin, M.L., Morrison, H.G., Huber, J.A., Welch, D.M., Huse, S.M., Neal, P.R., *et al*. (2006) Microbial diversity in the deep sea and the underexplored 'rare biosphere'. *Proc Natl Acad Sci USA* **103:** 12115–12120.

Spear, J.R., Ley, R.E., Berger, A.B., and Pace, N.R. (2003) Complexity in natural microbial ecosystems: the Guerrero Negro experience. *Biol Bull* **204:** 168–173.

Spear, J.R., Walker, J.J., McCollom, T.M., and Pace, N.R. (2005) Hydrogen and bioenergetics in the Yellowstone Geothermal Ecosystem. *Proc Natl Acad Sci USA* **102:** 2555–2560.

Stone, W.C. (2007) *Conference on Un-manned, Un-tethered Submersable Technology*, 20–22 August, Durham, NH, USA.

Takaichi, S., Wang, Z.Y., Umetsu, M., Nozawa, T., Shimada, K., and Madigan, M.T. (1997) New carotenoids from the thermophilic green sulfur bacterium *Chlrobium tepidum*: 1′,2′-dihydro-gamma-carotene, 1′,2′-dihydrochlorobactene, and OH-chlorobactene glucoside ester, and the carotenoid composition of different strains. *Arch Microbiol* **168:** 270–276.

Takeuchi, M., Komai, T., Hanada, S., Tamaki, H., Tanabe, S., Miyachi, Y., *et al*. (2009) Bacterial and archaeal 16S rRNA genes in late pleistocene to holocene muddy sediments from the Kanto plain of Japan. *Geomicrobiol J* **26:** 104–118.

Vignais, P.M., and Billoud, B. (2007) Occurrence, classification, and biological function of hydrogenases: an overview. *Chem Rev* **107:** 4206–4272.

Vignais, P.M., Billoud, B., and Meyer, J. (2001) Classification and phylogeny of hydrogenases. *FEMS Microbiol Rev* **25:** 455–501.

Wagner, M., Roger, A.J., Flax, J.L., Brusseau, G.A., and Stahl, D.A. (1998) Phylogeny of dissimilatory sulfite reductases supports an early origin of sulfate respiration. *J Bacteriol* **180:** 2975–2982.

White, J.R., Nagarajan, N., and Pop, M. (2009) Statistical methods for detecting differentially abundant features in clinical metagenomic samples. *PLoS Comput Biol* **5:** e1000352.

Supporting information

Additional Supporting Information may be found in the online version of this article:

Supporting File. S1. Java scripts required to compute differences between sequence distribution profiles.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.