Novel Microbial Diversity Retrieved by Autonomous Robotic Exploration of the World’s Deepest Vertical Phreatic Sinkhole

Jason W. Sahl,1 Nathaniel Fairfield,2 J. Kirk Harris,3 David Wettergreen,2 William C. Stone,4 and John R. Spear1

Abstract

The deep phreatic thermal explorer (DEPTHX) is an autonomous underwater vehicle designed to navigate an unexplored environment, generate high-resolution three-dimensional (3-D) maps, collect biological samples based on an autonomous sampling decision, and return to its origin. In the spring of 2007, DEPTHX was deployed in Zacatón, a deep (~318 m), limestone, phreatic sinkhole (cenote) in northeastern Mexico. As DEPTHX descended, it generated a 3-D map based on the processing of range data from 54 onboard sonars. The vehicle collected water column samples and wall biomat samples throughout the depth profile of the cenote. Post-expedition sample analysis via comparative analysis of 16S rRNA gene sequences revealed a wealth of microbial diversity. Traditional Sanger gene sequencing combined with a barcoded-amplicon pyrosequencing approach revealed novel, phylum-level lineages from the domains Bacteria and Archaea; in addition, several novel subphylum lineages were also identified. Overall, DEPTHX successfully navigated and mapped Zacatón, and collected biological samples based on an autonomous decision, which revealed novel microbial diversity in a previously unexplored environment. Key Words: Autonomous—Robotics—Microbial diversity—16S rRNA. Astrobiology 10, 201–213.

1. Introduction

In the search for extraterrestrial microbial life, there is emergent interest in the field of exploratory autonomous robotics (Díaz-Calderón et al., 2007). To search for microbial life on planetary bodies that contain liquid water, such as the jovian moon Europa (Squyres et al., 1983), a robotic vehicle would be required to navigate through space, land on the moon’s surface, penetrate the ice sheet, explore the alien ocean for microscopic life, and transmit its findings back to Earth. In an effort to develop tools for autonomous navigation and exploration of unexplored water bodies, a vehicle known as the deep phreatic thermal explorer (DEPTHX) was designed (Krajick, 2007; Stone, 2007). The deployment goals of DEPTHX were to map and navigate autonomously through an unexplored environment and collect biological samples through autonomous decisions based on a set of programmed criteria.

Located in northeastern Mexico (Fig. 1A), Zacatón, the deepest water-filled vertical sinkhole (cenote) in the world (Gary and Sharp, 2006), provided the site for deployment of the DEPTHX vehicle (Fig. 1B). Zacatón (Fig. 1C) is fed by subsurface hydrothermal input from nearby volcanic activity, which provides a spatially and temporally stable temperature profile of ~30°C throughout the year, and a sulfur input that supplies the surrounding area its name: Rancho la Azufrosa (azufrosa is Spanish for sulfurous). The cenote is thought to have been created as hydrothermal water supplied from nearby volcanic activity was acidified with carbon dioxide or hydrogen sulfide, or both, to dissolve the host limestone as it moved upward through the subsurface (Gary et al., 2008). In a preliminary equipment test for DEPTHX, a sonar sonde dropped into Zacatón in May 2005 revealed that the cenote was at least 290 m deep and potentially much deeper. Zacatón was chosen as the test site for the deployment of DEPTHX because the cenote was deep, likely mappable, and unexplored.

2. Methods and Materials

2.1. Site description

Sistema Zacatón is located in a localized karst region of northeastern Mexico (22°59′35.10″N, 98°09′55.96″W). The...
geological formations of this system are part of a larger system of travertine-capped sinkholes, dry karst cave passages, and dolines in the vicinity (Gary et al., 2008). Zacatón is a 105 m diameter, ~318 m deep sinkhole with a 17 m deep limestone rim/cliff band above the water. Two meter-thick, round, floating reed islands, known as zacates, float and are blown around on the surface of the cenote. In direct sunlight, the clear blue water can turn milky-white as sulfur clouds develop in the top 5 m of the surface water and disappear when the Sun sets.

2.2. Vehicle engineering

A detailed description of the DEPTHX vehicle has been reported previously (Stone, 2007). In short, DEPTHX is approximately 1.5 m tall and 1.9 m in length and width and is powered by lithium-ion batteries. The vehicle has a full suite of underwater navigation sensors, including a Honeywell HG2001 Inertial Measurement Unit (IMU), two Parascientific Digiquartz depth sensors, and an RDI Navigator 600 Doppler Velocity Log (DVL). There is also a conductivity, temperature, and depth sensor for measuring the speed of sound for corrections to sonar-based measurements. The specifications for the inertial navigation system are roll/pitch 0.2° 2σ, yaw 0.4° 2σ, for the DVL velocities 0.3 cm/s 1σ, and for the depth sensors 0.01% of full range (10 cm for our 1000 m rated sensor). Fifty-four narrow-beam long-range sonar sensors allow the vehicle to detect subsurface features and generate maps in 3-D. The vehicle is equipped with a science package for microbiological sampling (see below).

The System Executive (i.e., the multiprocessor autonomous “brain” of the robot) allocates out steps in a dive specification based on the state of the system and its location, which is estimated using a combination of dead-reckoning and sonar-based localization. In the DEPTHX system, there is no onboard planning; rather, the Executive chooses actions from a primary or cascading set of contingency plans. Dive specifications are composed into mission plan files that encode, in an interpreted command language, the sequence of behaviors to execute. These behaviors may initiate sensing, navigate to locations, trigger on conditions, perform maneuvers, or acquire samples. The primary plan file is backed up by a number of contingency plans that are triggered on conditions encoded in the currently executing plan description.

2.3. Vehicle navigation and SLAM

The high-grade IMU had negligible yaw drift over time, and under nominal conditions the onboard sensors provided dead-reckoned navigation on the order of 0.5% of distance traveled. Under certain conditions [e.g., when the vehicle gets too close to a wall (< ~2 m)], the DVL cannot provide velocity measurements, at which point the quality of the dead-reckoned solution degrades significantly (~ 10% of distance traveled). To estimate velocity, the DVL requires that each of its four sonars receive a reflected signal from either the water column or a surface. At times of transition, however, as when moving away from a surface into open water or when the orientation of the surface deflects the return away, the unit can “lose lock” and become unable to estimate velocity. As a result, dead-reckoning accuracy is reduced until velocity lock is re-established. Underwater, there is no absolute position information (like GPS), so this dead-reckoning position error is cumulative.

The vehicle can bound this position error by performing its own localization estimate (using the onboard sonars) relative to a map. However, exploration implies that there is no map available, thus the vehicle has to build a map with new data and simultaneously localize itself using that map; this is known as the simultaneous localization and mapping (SLAM) problem. The approach we describe below builds highly accurate 3-D maps; but, due to the coupled nature of localization and mapping, it is susceptible to unavoidable error (or drift) over long distances, except when the vehicle can close loops and return to previously mapped areas. We developed a real-time probabilistic SLAM algorithm that provides an estimate of the vehicle trajectory as well as a 3-D metric map of the environment [see Fairfield et al. (2006, 2007) for a more complete description of our SLAM approach].

For SLAM, the DEPTHX vehicle has an array of 54 narrow-beam (24 at 300 kHz, 1.75° beam width; and 32 at 600 kHz, 2° beam width with ranges of 200 and 100 m, respectively) sonars that provide a constellation of range measurements around the vehicle at about 1 Hz. This array is in the shape of three great circles, a configuration that was arrived at after studying the suitability of various sonar geometries for 3-D mapping (Fairfield et al., 2006). The sonars have long ranges (100 or 200 m), and the accuracy of the range measurements
is approximately 10 cm; however, the low resolution, update rate, and point density make the mapping problem under-water significantly more difficult than it is with ranging sensors, like a laser scanner, that provide fast, accurate, high-resolution ranges in typical air-filled environments.

2.4. Sample collection and preparation

Wall biomat samples were collected with an onboard spring-loaded sampling arm and stored in a stainless-steel borer with a sealable door; once a single wall sample was taken, the door closed on the collection borer. Immediately after the vehicle surfaced, wall biomat samples were ejected into 2 ml cryovials and frozen in liquid nitrogen. Between sampling trips, the borer was cleaned with a dilute bleach solution (3%) and thoroughly rinsed with sterile, double-distilled water.

Prior to water column sample collection, DEPTHX ran its pump for 2 min, which allowed for the sample collection tubing to be completely flushed with in situ water. Samples were then collected into sterile urinary collection bags (2 L capacity). Upon surfacing, the collection bags were connected to a peristaltic pump, and water was filtered directly onto 0.2 μm polyethersulfone filters (Pall Corporation, East Hills, NY) housed in a stainless steel high-pressure filter manifold (Millipore, Billerica, MA). One quarter of each filter was cut with flame-sterilized scissors and frozen in liquid nitrogen. New collection bags were used for each sampling trip.

2.5. Water and wall biomat chemistry

Values for pH, temperature, dissolved oxygen, and specific conductivity were recorded with a modified Hydrotech HT6 meter (Hydrotech, Hutto, TX) mounted on the DEPTHX vehicle. Sulfide concentrations were calculated with a modified laboratory sensor mounted on the vehicle. For shallow water column samples, concentrations of ammonia and hydrogen sulfide were manually recorded with CHEMetrics kits (CHEMetrics, Calverton, VA) and a field photometer. Prior to filtering for microbiological analysis, water for the DEPTHX sampling missions was collected for laboratory water analyte analysis. Fifteen milliliters of water was syringe filtered with a sealable door; once a single wall sample was taken, the door closed on the collection borer. Immediately after the vehicle surfaced, wall biomat samples were ejected into 2 ml cryovials and frozen in liquid nitrogen. Between sampling trips, the borer was cleaned with a dilute bleach solution (3%) and thoroughly rinsed with sterile, double-distilled water.

2.6. DNA extraction, PCR, cloning, sequencing

DNA was extracted from wall samples and filter sections with the Mobio Powersoil Kit (Mobio, Carlsbad, CA) and eluted in a volume of 30 μl TE. Polymerase chain reaction (PCR) of the 16S rRNA gene was performed on all samples by using universal forward primer 515F (5'-GTG CCA GCM GCC GCG GTA A-3') (Lane, 1991), bacterial-specific forward primer 8F (5'-AGA GTT TGA TCC TGG CTG AG-3') (Lane, 1991), and archaeal-specific forward primers 4Fa (5'-TCC GTG TGA TCC TGC CCG-3') (Hershberger et al., 1996) and 21Fa (5'-TTC CGG TTG ATC ATG CGC GA-3') (Reysenbach and Pace, 1995). Reverse primers included the nominally universal 1391R (5'-GAC GGG CCG TGW GTR CA-3') and 1492R (5'-GGT TAC CTT GTT ACG ACT T-3') (Lane, 1991).

PCRs were performed with Promega PCR Master Mix (Promega, Madison, WI) and a primer concentration of 1 μM. PCR thermal cycling conditions consisted of an initial denaturation at 94°C for 2 min, followed by 30 cycles of denaturation (94°C for 1 min), annealing (50–55°C for 1 min) and elongation (72°C for 1 min). PCR amplicons were gel purified (Millipore, Billerica, MA), TOPO TA (Invitrogen, Carlsbad, CA) cloned and sequenced on a MegaBACE 1000 dye-terminating sequencer.

2.7. Sanger sequence analysis

Sequences were processed by PHRED and PHRAP with use of the XplorSeq interface (Frank, 2008). Multi-direction reads were performed on each sample, and contiguous reads were compiled. Sequences were screened for chimeras with Mallard (Ashelford et al., 2006), and non-chimeric sequences were aligned with the NAST aligner (DeSantis et al., 2006a). Sequences were inserted into the ARB dendogram by parsimony insertion with use of the Lanemask filter (Lane, 1991). A global sequence comparison was made by building a distance matrix between the query and target sequence in ARB; the sequence distance was then converted to a percent relatedness.

2.8. Phylogenetic analyses

Sequences that did not group with known phyla in ARB with use of the current Greengenes (DeSantis et al., 2006b) database release were analyzed in a phylogenetic tree that contained representative sequences from currently recognized bacterial phyla and candidate phyla. Sequences were exported unaligned and subsequently aligned with the Infernal aligner (Nawrocki et al., 2009). Covariance model (Eddy and Durbin, 1994) files were created with a seed alignment downloaded from the comparative RNA web site (Cannone and Subramanian, 2002), and a consensus secondary structure was calculated with RNAalifold (Hofacker, 2003). Alignment confidence estimates were calculated with Infernal, with use of the--p flag, for each residue of each sequence. Columns with high alignment confidence values (~100%) were considered to be conserved, and a filter mask, which masks out variable columns, was created with custom Perl script. The conserved column alignment mask, along with all aligned sequences, was imported back into ARB. Manual adjustments were made to the alignment, and sequences were exported with use of the custom conserved filter mask. For the Archaea, a phylogenetic tree was inferred via the RAXML web server (Stamatakis et al., 2008). For the Bacteria, the number of taxa...
analyzed (~36,000) exceeded the capabilities of the maximum-likelihood algorithm, and a neighbor-joining method, which uses a minimum evolution principle (Price et al., 2009), was used to infer the tree topology.

2.9. Pyrosequencing and analysis

Polymerase chain reactions were conducted with Promega PCR Master Mix and the bacterial-specific primers 8F and 338R (5'-CAT GCT GCC TCC CGT AGG AGT-3') (Amann et al., 1990); these priming sites span the hypervariable V1 and V2 regions of the 16S rRNA gene (Van de Peer et al., 1996). The reverse primer contained a Hamming barcode, which allowed for the pooling of multiple samples in a single high-throughput sequencing run (Hamady et al., 2008). PCR amplicons were normalized, quantified on a NanoDrop ND-1000 (NanoDrop, Wilmington, DE), diluted in molecular-grade water to a concentration of ~0.01 ng/µl, and sequenced on a 454 GS FLX machine (454 Life Sciences, Branford, CT). Sequences were sorted by barcode and quality filtered to remove sequences with ambiguous characters (Ns) and sequences that were of an insufficient length (<2 standard deviations of the mean) with custom Java script. The barcode and primer sequence were trimmed prior to analysis.

Sequences were aligned with use of the ribosomal database project (RDP) pyrosequencing pipeline (Cole et al., 2009), which uses the Infernal aligner in conjunction with a consensus secondary structure alignment. The pyrosequencing pipeline was also used to cluster sequences based on a furthest neighbor algorithm; operational taxonomic units (OTUs) were then formed based on a 97% sequence identity threshold. A representative sequence from each OTU was aligned with the NAST aligner and classified with the Greengenes classifier (DeSantis et al., 2006b). Taxonomic classifications were made according to the Hugenholtz taxonomic system (DeSantis et al., 2006b).

2.10. qPCR

The total number of 16S rRNA gene copies in extracted samples was estimated with quantitative PCR (qPCR) on a Lightcycler 480 (Roche). Degenerate primers for conserved regions flanking the hypervariable V4 region of the 16S rRNA gene (Raue et al., 1988) were forward primer (5'-AYT GGG YDT AAA GNG-3') and reverse primer (5'-TAC NVG GGT ATC TAA TCC-3') (Cole et al., 2009); these primers cover >94.5% of long (>1200 nucleotides), high-quality sequences in the RDP database, using the probe match tool (Cole et al., 2009). Primers were also designed specifically for the C2 Crenarchaeota; these include forward primer 303F (5'-GGG AGC CCG GAG ATG-3') and reverse primer 1069R (5'-ACC TCA CGG AAC GAA-3') (Dr. Bradley Stevenson, unpublished). A calibration curve was calculated by using seven standards, in triplicate, and purified and serial diluted amplicons. Gradient PCR revealed an optimum annealing temperature for each primer of 48°C.

The reverse primer 494R (5' -CTG AGG GTA CCG TCA ATA-3') was also designed for the putative candidate phylum Azufrosa group 1 (AG1). The primer was compared against the RDP database with the probe match tool, which only returned six positive matches; all these sequences grouped with AG1 and only two were >1200 nucleotides in length. This reverse primer was paired with the bacterial primer 8F to determine the distribution of this clade throughout the depth profile of Zacatón.

2.11. Species richness estimators

Parametric species richness estimates were calculated with methods reported previously (Jeon et al., 2006). Six different parametric models [single-point Poisson, negative binomial, inverse Gaussian, lognormal, Pareto, mixture of two exponentials (M.E.)] were used to fit the observed data based on the frequency of sequences at the designated percent identity level. The best performing model estimate was chosen based on a biologically relevant standard error [-<estimate/2], the highest goodness-of-fit value, and a high right truncation point. For comparison, the non-parametric species richness estimates Chao1 (Chao, 1987) and ACE (Chao and Lee, 1992) were calculated with EstimateS (Colwell, 2005). Coverage estimates were calculated by dividing the number of observed OTUs by the best parametric estimate.

2.12. Secondary structure analysis

A comparative secondary structure analysis was performed by aligning a consensus sequence from AG1 against a 16S rRNA sequence from E. coli (accession #J01695). The sequence from AG1 was then manually mapped onto the well-established E. coli 16S rRNA secondary structure (Woese et al., 1980) to identify regions of conservation and divergence.

2.13. Nucleotide accession numbers

The 1588 Sanger sequences generated in this study were submitted to Genbank under accession numbers FJ484010-FJ485598. Pyrosequences were submitted to the Genbank short read archive under accession number SRA008128.3.

3. Results

3.1. Vehicle navigation and SLAM

With SLAM, we built a three-dimensional 0.5 m resolution metric map of Zacatón. Without ground truth, it is difficult to make strong statements about localization and map accuracy, but the maps showed consistent findings independently produced across many dives. Our method was able to use sparse, noisy, low-resolution, and low-rate sonar data to build 3-D maps, localize, run SLAM on board the DEPTHX vehicle, and generate the first ever maps of this large-scale, complex, natural formation.

3.2. Water chemistry

As DEPTHX maneuvered through the cenote, the vehicle recorded thousands of water chemistry values with the Hydrotech sensor. The data reveal an incredibly homogeneous water body, with nearly identical values for pH and temperature throughout the depth profile of the cenote (Table 1). Major ion analyses confirmed this trend, with uniform concentrations of sulfate, nitrate, chloride, and sodium. However, statistically significant (P < 0.03) negative Spearman correlations (ρ = -0.7) of zinc, iron, and manganese with depth were observed. No clear chemistry boundaries were detectable; for example, even the surface water sample showed reducing conditions, with no clear oxic/anoxic boundary or chemocline observed. Sulfate was the most abundant common terminal
<table>
<thead>
<tr>
<th>Parameter</th>
<th>Surface</th>
<th>22 m</th>
<th>30 m</th>
<th>60 m</th>
<th>90 m</th>
<th>150 m</th>
<th>198 m</th>
<th>273 m</th>
<th>285 m</th>
<th>4 m</th>
<th>32 m</th>
<th>114 m</th>
<th>273 m</th>
</tr>
</thead>
<tbody>
<tr>
<td>Temp (°C)</td>
<td>30.10</td>
<td>30.03</td>
<td>30.02</td>
<td>30.03</td>
<td>30.00</td>
<td>30.04</td>
<td>30.10</td>
<td>30.10</td>
<td>30.06</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pH</td>
<td>6.67</td>
<td>6.58</td>
<td>6.73</td>
<td>6.72</td>
<td>6.72</td>
<td>6.52</td>
<td>6.54</td>
<td>6.53</td>
<td>6.50</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eh (Mv)</td>
<td>−175</td>
<td>−168</td>
<td>−177</td>
<td>−185</td>
<td>−200</td>
<td>−274</td>
<td>−282</td>
<td>−282</td>
<td>−263</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DO (mM)</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TOC (mM)</td>
<td>0.16</td>
<td>0.15</td>
<td>0.18</td>
<td>0.16</td>
<td>0.18</td>
<td>0.24</td>
<td>0.14</td>
<td>0.14</td>
<td>0.39</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spec Cond (µS/cm)</td>
<td>759</td>
<td>767</td>
<td>767</td>
<td>767</td>
<td>767</td>
<td>767</td>
<td>761</td>
<td>761</td>
<td>761</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fe - T (µM)</td>
<td>0.01</td>
<td>1.63</td>
<td>0.52</td>
<td>1.04</td>
<td>0.38</td>
<td>10.20</td>
<td>10.20</td>
<td>0.20</td>
<td>0.08</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NH₄⁺ (mM)</td>
<td>0.05</td>
<td>0.07</td>
<td>NM</td>
<td>NM</td>
<td>NM</td>
<td>NM</td>
<td>NM</td>
<td>NM</td>
<td>NM</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PO₄³⁻ (mM)</td>
<td>BDL</td>
<td>BDL</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>BDL</td>
<td>BDL</td>
<td>0</td>
<td>BDL</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cl⁻</td>
<td>1.01</td>
<td>0.97</td>
<td>0.98</td>
<td>0.97</td>
<td>0.95</td>
<td>0.97</td>
<td>0.96</td>
<td>0.98</td>
<td>0.95</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NO₃⁻</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Br</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SO₄²⁻</td>
<td>0.15</td>
<td>0.15</td>
<td>0.15</td>
<td>0.15</td>
<td>0.15</td>
<td>0.16</td>
<td>0.16</td>
<td>0.14</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Major cations (mM)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ca</td>
<td>3.11</td>
<td>3.37</td>
<td>2.79</td>
<td>3.06</td>
<td>2.80</td>
<td>2.94</td>
<td>2.98</td>
<td>2.96</td>
<td>2.71</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>K</td>
<td>0.08</td>
<td>0.07</td>
<td>0.08</td>
<td>0.08</td>
<td>0.08</td>
<td>0.08</td>
<td>0.08</td>
<td>0.08</td>
<td>0.09</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mg</td>
<td>0.49</td>
<td>0.44</td>
<td>0.44</td>
<td>0.46</td>
<td>0.46</td>
<td>0.49</td>
<td>0.47</td>
<td>0.49</td>
<td>0.47</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Na</td>
<td>1.38</td>
<td>1.20</td>
<td>1.27</td>
<td>1.30</td>
<td>1.31</td>
<td>1.72</td>
<td>1.21</td>
<td>1.32</td>
<td>2.30</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>H₂S</td>
<td>0.01</td>
<td>0.01</td>
<td>NM</td>
<td>NM</td>
<td>NM</td>
<td>NM</td>
<td>NM</td>
<td>NM</td>
<td>NM</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trace elements (µM)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zn</td>
<td>13.31</td>
<td>37.17</td>
<td>7.80</td>
<td>12.39</td>
<td>14.84</td>
<td>4.13</td>
<td>7.04</td>
<td>1.84</td>
<td>1.07</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ni</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mn</td>
<td>0.73</td>
<td>0.73</td>
<td>0.73</td>
<td>1.09</td>
<td>1.46</td>
<td>0.55</td>
<td>0.55</td>
<td>0.55</td>
<td>0.36</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Σ cations (meq)</td>
<td>8.27</td>
<td>8.20</td>
<td>0.23</td>
<td>0.25</td>
<td>0.24</td>
<td>0.26</td>
<td>0.24</td>
<td>0.25</td>
<td>0.28</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Σ anions (meq)</td>
<td>7.44</td>
<td>7.81</td>
<td>0.13</td>
<td>0.13</td>
<td>0.13</td>
<td>0.13</td>
<td>0.13</td>
<td>0.13</td>
<td>0.13</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Charge imbalance (%)</td>
<td>5.28</td>
<td>2.44</td>
<td>1.70</td>
<td>5.96</td>
<td>3.23</td>
<td>7.65</td>
<td>3.94</td>
<td>4.58</td>
<td>7.43</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

BDL, below detection limit; NM, not measured.
DO, dissolved oxygen; TOC, total organic carbon.
electron acceptor, which, combined with lineages associated with sulfate reduction (see below), suggests that this may be an important mode of metabolism throughout the depth profile of the cenote.

Inductively coupled plasma atomic emission spectroscopy data were only obtained for four wall samples in Zacatón (Table 1). Only zinc and nickel showed a correlation ($r = 1.0$) of concentration with depth (Table 1). Approximately 8% of the biomat mass consisted of extractable metals; the remaining mass of the mat is likely non-digestible minerals. Calcium comprised the highest mass percent of the mats.

### 3.3. 16S rRNA gene sequence analysis

To sample the microbiology of Zacatón, the DEPTHX vehicle was equipped with five sterile collection bags for aqueous samples and a spring-loaded extendable arm designed to sample the extensive biomats that coat the walls of the cenote (Fig. 1D, 1E, 1F).

Culture-independent gene sequence analysis (Pace, 1997; Spear et al., 2005) allowed for the analysis of microbial community structure in 3-D space, all in the context of the geochemistry of the cenote. Visualization of the mats by an onboard camera on DEPTHX revealed thick, complex photosynthetic mats in the shallow subsurface (<30 m) (Fig. 1D, 1F) and delicate, striated mats 273 m deep in the cenote (Fig. 1E). The initial survey of 16S rRNA gene sequences resulted in 1588 Sanger sequence reads that guided a more comprehensive sampling of the microbial community with a barcoded-amplicon pyrosequencing approach (Hamady et al., 2008).

Results of the 16S rRNA gene sequence analysis of the filterable fraction of the water column suggest that the community is dominated by a few phylogenetic types (phylotypes) such as *Arcobacter* spp. and *Sulfuromales* spp., which group with the Epsilonproteobacteria. Sequences related to these *Arcobacter* and *Sulfuromales* phylotypes have been isolated from other sulfidic caves (Macalady et al., 2008; Porter and Engel, 2008) and are thought to be important microbes in the cycling of carbon, nitrogen, and sulfurous compounds (Campbell et al., 2006). Even within the Epsilonproteobacteria, clear differences were observed with depth; for example, deep in the water column (273 m), the epsilon sequences were phylogenetically related (98% global identity) to *Sulfuricurvum kujiense*, a chemoheterotrophic microbe capable of anaerobic sulfide oxidation (Kodama and Watanabe, 2004).

Other sequences from the water column grouped with the Gammaproteobacteria and were related (96% global identity) to sequences amplified from deep subsurface mine fluids in South Africa (Gihring et al., 2006). In general, the water column microbial diversity was low (Table 2); in fact, out of ~7500 full- and partial-length gene sequences screened from a depth profile of the water column, only 850 OTUs were compiled with a 97% sequence identity threshold. Even with low observed diversity and dominance of a few phylotypes, clear trends in phylotype distribution and diversity were observed in the depth profile of the water column (Table 2).

In contrast, sequences obtained from the walls of the cenote showed high microbial diversity within the bacterial and archaeal domains (Figures 2 and 3); for example, out of 105 bacterial phyla currently recognized by the Hugenholtz taxonomic system (DeSantis et al., 2006b), 74% were found in the mats of Zacatón (more than 10 representative sequences from each clade) based on ~39,000 pyrosequences, as determined by the Greengenes classifier. Furthermore, a clear depth profile of bacterial phylotypes was observed, with novel cyanobacterial sequences (95% global identity with nearest neighbor) dominant in shallow regions of the cenote and sequences associated with sulfate reduction (Deltaproteobacteria, Thermodesulfovibrio) dominant below the photic zone (~40 m) (Fig. 4). While some Deltaproteobacterial sequences showed relatedness to cultured taxa, such as *Syntrophobacter pfeffernii* (93%) (Wallrabenstein et al., 1995) and *Desulfomonile tiedjei* (95%) (DeWeerd et al., 1990), many Deltaproteobacterial sequences (~45%) in deeper samples (>100 m) belonged to 15 OTUs that were not closely related (<95%) to Deltaproteobacterial sequences in public databases from environmental samples. Sequences associated with

### Table 2. Microbial Diversity and Coverage Estimates for Samples in Zacatón

<table>
<thead>
<tr>
<th>Matrix</th>
<th>Sample depth (m)</th>
<th>Sequences</th>
<th>OTUs (97%)</th>
<th>Best parametric estimate</th>
<th>coverage (%)</th>
<th>ACE</th>
<th>Chao 1</th>
<th>Shannon index</th>
<th>Simpson index (1/D)</th>
</tr>
</thead>
<tbody>
<tr>
<td>mat</td>
<td>1</td>
<td>2490</td>
<td>584</td>
<td>1552 (se 128)</td>
<td>38</td>
<td>1187</td>
<td>1187</td>
<td>5.1</td>
<td>41.88</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>3680</td>
<td>1209</td>
<td>4062 (se 309)</td>
<td>30</td>
<td>2659</td>
<td>2659</td>
<td>6.23</td>
<td>147.94</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>2697</td>
<td>718</td>
<td>2656 (se 267)</td>
<td>27</td>
<td>1689</td>
<td>1700</td>
<td>4.32</td>
<td>6.83</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>3306</td>
<td>1444</td>
<td>4670 (se 259)</td>
<td>31</td>
<td>3228</td>
<td>3237</td>
<td>6.51</td>
<td>129.95</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>3948</td>
<td>1753</td>
<td>5389 (se 270)</td>
<td>33</td>
<td>4149</td>
<td>4003</td>
<td>6.94</td>
<td>578.9</td>
</tr>
<tr>
<td></td>
<td>32</td>
<td>2705</td>
<td>901</td>
<td>4011 (se 507)</td>
<td>22</td>
<td>2280</td>
<td>2293</td>
<td>5.64</td>
<td>55.9</td>
</tr>
<tr>
<td></td>
<td>58</td>
<td>1459</td>
<td>923</td>
<td>5482 (se 754)</td>
<td>17</td>
<td>2991</td>
<td>3011</td>
<td>6.54</td>
<td>623.09</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>1761</td>
<td>655</td>
<td>3022 (se 430)</td>
<td>22</td>
<td>1642</td>
<td>1654</td>
<td>5.45</td>
<td>55.55</td>
</tr>
<tr>
<td></td>
<td>114</td>
<td>1898</td>
<td>1129</td>
<td>4984 (se 475)</td>
<td>23</td>
<td>3025</td>
<td>3038</td>
<td>6.72</td>
<td>712.97</td>
</tr>
<tr>
<td></td>
<td>195</td>
<td>2119</td>
<td>1056</td>
<td>4305 (se 460)</td>
<td>25</td>
<td>2295</td>
<td>2303</td>
<td>6.52</td>
<td>348.34</td>
</tr>
<tr>
<td></td>
<td>273</td>
<td>2338</td>
<td>1186</td>
<td>4485 (se 395)</td>
<td>26</td>
<td>2686</td>
<td>2696</td>
<td>6.69</td>
<td>607.78</td>
</tr>
<tr>
<td>water</td>
<td>0</td>
<td>2662</td>
<td>140</td>
<td>563 (se 134)</td>
<td>25</td>
<td>297</td>
<td>305</td>
<td>2.29</td>
<td>5.51</td>
</tr>
<tr>
<td></td>
<td>17</td>
<td>1810</td>
<td>204</td>
<td>674 (se 94)</td>
<td>30</td>
<td>513</td>
<td>527</td>
<td>3.24</td>
<td>10.5</td>
</tr>
<tr>
<td></td>
<td>273</td>
<td>2396</td>
<td>392</td>
<td>1697 (se 248)</td>
<td>23</td>
<td>951</td>
<td>962</td>
<td>3.69</td>
<td>12.66</td>
</tr>
</tbody>
</table>
photosynthetic anoxygenic sulfide oxidizers (Chlorobi) were found throughout the depth profile of the cenote, even in zones well below the photic zone (Fig. 4). At depths of >100m, the community was dominated by sequences that grouped with Deltaproteobacteria and Chloroflexi (Fig. 4). Epsilonproteobacterial sequences, which are ubiquitous in the water column, were almost entirely absent from all wall biomat samples, regardless of depth.

Archaeal diversity was dominated (58% of archaeal sequences) by the C2 Crenarchaeota (Fig. 4), which has been recognized as a monophyletic group by some taxonomic curators (DeLong and Pace, 2001; Robertson et al., 2005; DeSantis et al., 2006b); this clade is also associated with the so-called miscellaneous Crenarchaeota group (Nunoura et al., 2005). Sequences that group with this clade have been amplified from deep subsurface gold mines (Nunoura et al., 2005), geothermal wells (Rogers and Amend, 2005), deep marine anoxic basins (Jeon et al., 2008), and a hydrothermal vent in the Mariana Trough (unpublished data).

Clone libraries constructed with universal primers suggest that C2 sequences were prevalent at depth in the cenote and in lower abundance, or absent altogether, in shallow regions.

**FIG. 2.** Neighbor-joining phylogenetic tree of the bacterial domain. Numbers in parentheses indicate accession numbers. Dark wedges indicate phyla that are represented in either biomat or water column samples.
qPCR confirmed this trend, with an increased number of C2 16S rRNA gene copy numbers, compared to the total 16S rRNA gene copy numbers, with sample depth (Fig. 5). Two nucleotide mismatches were found between the V4 forward primer and sequences of the C2 Crenarchaeota on the 3' end. To allow for the potential of the primer not annealing to C2 gene templates, the percentage of C2 gene copy numbers was calculated by comparison of the C2 gene copy numbers to the V4 gene copy numbers plus the C2 gene copy numbers. This calculation is considered even more conservative, given the fact that bacteria generally contain more rrn operon copies than archaea (Klappenbach et al., 2001; Acinas et al., 2004).

Additional archaeal sequences grouped with known methanogens (Methanomicrobia) as well as anaerobic methane oxidizers (ANME-1). Singleton, non-chimeric, sequences were observed on the archaeal tree (Fig. 3) that showed low global identity to nearest neighbors (~ 85% relatedness). Additional sequencing with archaeal-specific primers may help to further define these novel lineages.

### 3.4. Azufrosa candidate phyla

Initial phylogenetic analyses using ARB parsimony insertion suggested that some phylotypes did not group with known bacterial phyla at a phylogenetic distance (~ 0.20) consistent with phylum-level delineations (Schloss and Handelsman, 2005) (Fig. 2). In total, three novel candidate phyla were identified, although two of these phyla were only represented by two sequences from Zacatón mats greater than 1250 nucleotides in length. Additional phylogenetic analyses by maximum likelihood (Stamatakis, 2006) and neighbor-joining algorithms (Price et al., 2009) support the presence of these novel lineages.

Azufrosa group 1 (AG1) was the most represented novel candidate phylum identified. Sequences from deep aquifers...
in Japan (Shimizu et al., 2006) and New York (Bakermans and Madsen, 2002) are also affiliated with AG1 (Fig. 3). A secondary structure analysis of AG1 sequences was compared to the established secondary structure of *E. coli*, and a truncation of Helix 17 was identified (Fig. 6), which has also been seen in archaea and other bacteria (Raué et al., 1988). The PCR primer 494R spans the unique truncation in the helix, which allows for high specificity to this group.

All sequences that group with AG1 were amplified from deep samples in Zacatón (>82m). A qPCR reaction was run to determine the relative abundance and distribution of this clade throughout the cenote water column; however, no clear trend in abundance was observed with depth (Fig. 5). The distribution of these sequences, based on these results, does not appear to be confined to deep environments.

### 3.5. Diversity and species richness

A positive and significant (*P* = 0.001) Spearman correlation (*ρ* = 0.83) of biomat microbial diversity (1/D) with depth was identified (Table 2). Parametric species richness estimators were higher than non-parametric estimators for all samples, which is consistent with other studies (Hong et al., 2006; Jeon et al., 2006). Species richness coverage estimates ranged from 17–38%.

However, richness estimates have been shown to scale with sequencing effort (Morales et al., 2009), which suggests that additional sequencing may reveal these estimates to be an overestimation of the calculated community coverage.

### 3.6. Autonomous sampling

One of the two primary missions of DEPTHX was to collect a biological sample based on an autonomous decision.

---

**FIG. 4.** Vertical profile of Zacatón and a heat map showing the relative distribution of selected phyla at various depths. Abbreviations include Proteob., Proteobacteria (α = alpha, δ = delta, γ = gamma, ε = epsilon); Bacterio., Bacteroidetes; Cyanob., Cyanobacteria; Nitrosp., Nitrospirae; Firmicu., Firmicutes; Spiroch., Spirochaetes; Acidoba., Acidobacteria; Chlorob., Chlorobi; Chlorof., Chloroflexi. “Others” represents the summed remainder of bacterial phyla.

**FIG. 5.** Relative abundance of C2 Crenarchaeota and Azufrosa group 1 (AG1) gene copies in the depth profile of Zacatón compared to the total gene copy number.
DEPTHX was programmed to detect chemical interfaces; we hypothesized that these interfaces provide sources of chemical disequilibrium and were the most likely location of varied microbial diversity and metabolism. However, due to the homogeneous nature of detectable chemical parameters in Zacatón, DEPTHX was unable to locate a clear chemical interface in the water column for biological sampling. In fact, only the sulfide concentration appeared to vary significantly with depth. Based on this observation, a voltage threshold that corresponded to an in situ sulfide concentration was programmed into DEPTHX. In one sampling mission, DEPTHX detected the programmed threshold, approached the wall of the cenote, collected a wall biomat sample, and returned to its origin based on the SLAM method. Although a sulfide calibration was never successfully calculated, the autonomous decision to sample was successfully made by DEPTHX.

4. Discussion

4.1. Water chemistry

The DEPTHX vehicle was programmed to detect chemical interfaces. Currently within the field of microbiology, the presence of known electron donors and acceptors may indicate that the energetics are amenable to microbial metabolism (Hand et al., 2007; Hoehler et al., 2007). However, an analysis of water chemistry parameters in Zacatón revealed an extremely homogeneous water body. This homogeneity is likely caused by convective mixing as hydrothermal groundwater from the deep subsurface moves upward through the water column. The hydrothermal source is likely deep, as no discernable differences in temperature with depth. Based on this observation, a voltage threshold that corresponded to an in situ sulfide concentration was programmed into DEPTHX. In one sampling mission, DEPTHX detected the programmed threshold, approached the wall of the cenote, collected a wall biomat sample, and returned to its origin based on the SLAM method. Although a sulfide calibration was never successfully calculated, the autonomous decision to sample was successfully made by DEPTHX.

4.2. Microbial diversity

The geochemistry and morphology of Zacatón select for a unique and diverse microbial community. A comparison of microbial communities between Zacatón and other hydrothermal travertine systems (Breitbart et al., 2008), as well as freshwater sulfurous lakes (Casamayor et al., 2000), showed that the microbial communities in Zacatón are significantly different than these comparable systems.

Initial Sanger sequencing of wall biomat samples revealed a diverse microbial community. Subsequent directed pyrosequencing provided a higher-resolution analysis of the microbial community structure in Zacatón. This method has been used previously to identify the low-abundance microbial phylotypes in an environment, which constitute the “rare biosphere” (Sogin et al., 2006).

Overall, pyrosequencing of wall biomat DNA revealed higher microbial diversity than Sanger sequencing. Although this is likely a result of the greater sequencing effort, results of a recent study suggest that shorter sequences reveal higher microbial diversity than longer sequences due to more efficient amplification (Huber et al., 2009). In this study, longer Sanger sequences were used for accurate phylogenetic analysis and to anchor shorter pyrosequences on phylogenetic trees, while shorter sequences were used for assessments of microbial diversity and distribution. Such an approach may take full advantage of the benefits and limitations of short and long gene sequences.

A positive correlation of biomat microbial diversity with depth was observed in the depth profile of Zacatón. Higher microbial diversity with depth has also been observed in other anoxic systems (Madrid et al., 2001) and may be attributed to more varied metabolisms below the photic zone.

4.3. 16S rRNA gene sequence analysis

Water column samples showed low microbial diversity, with Epsilonproteobacteria and Chlorobi phylotypes dominating the clone libraries (Fig. 4). Both these groups have been associated with sulfate oxidation (Bergstein and Cavari, 1983; Wagner et al., 1998); their presence, combined with elemental sulfur clouds that develop in sunlight, suggests that these microbes may oxidize sulfide in Zacatón as well. Microbes typically associated with oxygenic photosynthesis, such as cyanobacteria and green algae, were absent from water column clone libraries. Epsilonproteobacteria sequences dominated water column samples and were absent or in low abundance in wall biomat samples; cultivated members of related Epsilonproteobacteria are motile (Miller et al., 2007; Ho et al., 2008), which suggests that the Epsilonproteobacteria in Zacatón prefer an unattached lifestyle.

Shallow biomat samples in the cenote contained sequences from known phototrophic organisms (e.g., Cyanobacteria) and photosynthetic sulfide oxidizers (e.g., Chlorobi). However, even deeper samples (>100 m) contained sequences that group with Chlorobi. Microorganisms that group with Chlorobi have been shown to oxidize sulfide in the presence of geothermal radiation (Beatty et al., 2005), which may help explain the presence of sequences associated with obligate from major cations, anions, or total organic carbon; these results correspond well with the homogeneous behavior of other abiotic parameters.
photosynthetically derived carbon. This type of metabolism may be applicable to deep, dark, extraterrestrial aqueous systems with a hydrothermal feed source.

4.4. Autonomous robotic exploration

During the course of the field campaign, DEPTHX was able to implement SLAM, detect a programmed chemical interface, sample both the water column and wall biomat, and return to its origin, all based on decisions executed on board without any human guidance. The sampling and subsequent molecular analysis of diver-inaccessible samples, with correlated spatial and geochemical information, have allowed for a more thorough understanding of microbial diversity and distribution in hydrothermal, phreatic sinkholes.

As robotic technology advances, the integration of the autonomous sampling and maneuvering methods discussed in this study with advanced onboard chemistry and molecular analyses tools will enable the development of a vehicle capable of searching for microbial life on planetary bodies that contain liquid water, such as Europa.

Acknowledgments

The authors would like to graciously thank Alejandro Dávila for access to Rancho la Azufrosa and the entire DEPTHX team including George Kantor, Antonio and Nora Fregoso, John Kerr, Marcus and Robin Gary, Dom Jonak, and the Southwest Research Institute. Additional thanks to the Pace Lab at the University of Colorado for their assistance operating the MegaBACE 1000 and to the Consortium for Comparative Genomics located at the University of Colorado Denver for operating the 454 GS FLX machine. Custom Java script to decode barcodes was written by Erik Sahl and the custom Perl infernal script was written by Chuck Peperanny. The authors would also like to thank Dr. Bradley Stevenson for providing the C2 primer set and Dr. Laura Beer for her valuable comments. Funding for this project was provided by a NASA ASTEP grant NNG04GC09G.

Abbreviations

AG1, Azufrosa Group 1; DEPTHX, deep phreatic thermal explorer; DVL, Doppler Velocity Log; IMU, Inertial Measurement Unit; OTUs, operational taxonomic units; PCR, polymerase chain reaction; qPCR, quantitative PCR; RDP, ribosomal database project; SLAM, simultaneous localization and mapping.

References


Address correspondence to: John R. Spear
Colorado School of Mines
Environmental Science and Engineering
Golden, CO 80401
E-mail: jspear@mines.edu

Submitted 22 April 2009
Accepted 28 December 2009