Convergent community assembly among globally separated acidic cave biofilms

Daniel Jones[a,b#], Irene Schaperdoth[c], Diana E. Northup[d], Rodolfo Gómez-Cruz[e], and Jennifer L. Macalady[c]

[a]Earth and Environmental Science, New Mexico Institute of Mining and Technology, Socorro, NM, USA

[b]National Cave and Karst Research Institute, Carlsbad, NM, USA

[c]Department of Geosciences, The Pennsylvania State University, University Park, PA 16802

[d]Department of Biology, University of New Mexico, Albuquerque, NM, USA

[e]Biological Sciences Academic Division, Autonomous University Juárez de Tabasco (UJAT), Villahermosa, Centro, Tabasco 86150, Mexico

#Address correspondence to Daniel S. Jones, daniel.s.jones@nmt.edu

Running head: Extreme acidophile community biogeography

Abstract

Acidophilic bacteria and archaea inhabit extreme geochemical 'islands' that can tell us when and how geographic barriers affect the biogeography of microorganisms. Here we describe microbial communities from extremely acidic (pH 0-1) biofilms known as "snottites" from hydrogen sulfide-rich caves around the world. Given the extreme acidity and subsurface location of these biofilms, and in light of earlier work showing strong geographic patterns among snottite *Acidithiobacillus* populations, we investigated their structure and diversity in order to understand how geography might impact community assembly. We used 16S rRNA gene cloning and fluorescence *in situ* hybridization (FISH) to investigate 26 snottite samples from four sulfidic caves in Italy and Mexico. All samples had very low biodiversity and were dominated by sulfur-oxidizing bacteria in the genus *Acidithiobacillus*. *Ferroplasma* and other archaea in the *Thermoplasmatales* ranged from 0 to 50% of total cells, and relatives of the bacterial genera *Acidimicrobium* and *Ferrimicrobium* were up to 15% of total cells. Rare phylotypes included *Sulfobacillus* spp. and members of the *Dependentiae* and *Saccharibacteria* (formerly TM6 and TM7). Although the same genera of acidophiles occurred in snottites on separate continents, most members of those genera represent substantially divergent populations with 16S rRNA genes that are only 95-98% similar. Our findings are consistent with a model of community assembly where sulfidic caves are stochastically colonized by microorganisms from local sources, which are strongly filtered through selection for extreme acid tolerance, and these different colonization histories are maintained by dispersal restrictions within and among caves.

Importance

Microorganisms that are adapted to extremely acidic conditions, known as extreme acidophiles, are catalysts for rock weathering, metal cycling, and mineral formation in naturally acidic environments. They are also important drivers of large-scale industrial processes such as biomining and contaminant remediation. Understanding the factors that govern their ecology and distribution can help us better predict and utilize their activities in natural and engineered systems. However, extremely acidic habitats are unusual in that they are almost always isolated within circumneutral landscapes. So where did their acid-adapted inhabitants come from, and how do new colonists arrive and become established? In this study, we took advantage of a

49 unique natural experiment in Earth's subsurface to show how isolation may have played a role in
50 the colonization history, community assembly, and diversity of highly acidic microbial biofilms.
51
52

53 **Introduction**

54 Extreme acidophiles drive important biogeochemical processes in natural and engineered
55 systems. Acidophilic sulfur- and iron-oxidizing bacteria and archaea speed up the weathering of
56 sulfide minerals in natural ores and mine waste (1, 2), control metal cycling in acidic drainages
57 (3-5), and create secondary mineral deposits in surface and subsurface environments (6, 7).
58 Acidophilic consortia are used for large-scale bioremediation of acidic mine drainage and other
59 heavy metal contaminated environments (8-11), and for industrial processes such as biomining
60 and sulfuric acid production (12-14). Understanding the ecology of organisms in naturally acidic
61 environments can inform microbial behavior in engineered systems and provide ideas for new
62 bioremediation strategies (e.g., 11, 15). Furthermore, acidic environments are useful model
63 systems for studying microbial evolution and ecology because they often contain low
64 biodiversity (16) and represent extreme geochemical "islands" in a sea of neutrophilic habitats
65 (17).

66 Studies conducted at local scales in acidic systems have identified geochemical factors
67 that influence microbial distribution, including pH, conductivity, and metal concentrations (e.g.,
68 18-25). However, microbial spatial distribution might not only be a result of natural selection by
69 contemporary environmental factors, but also a legacy of past environmental conditions and
70 stochastic colonization events (26). Dispersal barriers limit the movement of new species into an
71 environment, and as a result, communities from the same geographic region are more likely to
72 contain the same organisms because they share a similar biotic history (27-29).

73 Extremely acidic microbial biofilms form on the walls and ceilings of sulfide-rich caves,
74 in areas where hydrogen sulfide degasses from springs and streams to the cave atmosphere
75 (Figure 1). These biofilms, commonly known as "snottites," have pH values between 0 and 1.5,
76 and are found hanging from the surface of gypsum corrosion residues that isolate them from
77 buffering by the parent limestone. Snottites have been consistently observed where $H_2S(g)$
78 concentrations are between 0.2 to 25 parts-per-million by volume (ppmv), and rarely occur
79 where concentrations are outside of this range (30, 31). Previous research has shown that
80 snottites are inhabited by very low biodiversity communities that contain abundant
81 *Acidithiobacillus* spp., as well as representatives or close relatives of the genera *Acidimicrobium*,
82 *Sulfobacillus*, *Ferroplasma*, and *Cuniculiplasma* (30-35).

83 Sulfidic cave snottites present an exceptional opportunity to test how geographic barriers
84 affect microbial communities. Because snottites occur in stable subsurface environments and
85 only form on overhanging surfaces within a specific range of $H_2S(g)$ concentrations,
86 geochemical variability among snottite locations is relatively limited (Figure 1a). Furthermore,
87 their extremely low pH is a potential barrier to microbial dispersal in addition to the physical
88 barrier imposed by their subsurface location. In previous research, Jones et al. (36) used high-
89 resolution genetic analyses to resolve relationships among geographically separated populations
90 of *Acidithiobacillus spp.* from cave snottites. That work showed that *Acidithiobacillus*
91 populations from separate cave locations are divergent, and identified a distance-decay
92 relationship among the most abundant *Acidithiobacillus* spp. that was consistent with restricted
93 dispersal among cave locations. We therefore hypothesized that geographic isolation would also
94 affect other organisms in the snottite biofilms, and thereby impact the structure and diversity of
95 the whole biofilm community. We tested this hypothesis by sampling snottites from four caves
96 on two continents, and quantitatively comparing community structure and composition using 16S
97 rRNA gene sequencing and fluorescence *in situ* hybridization (FISH).
98
99
100 **Results**
101
102 **Field observations and geochemistry.** Snottites were collected from multiple locations
103 in four sulfidic caves: Cueva de Villa Luz and Cueva Luna Azufre in Tabasco, Mexico, and the
104 Grotte del Fiume (Frasassi) and Grotta Nuova di Rio Garrafo (Acquasanta Terme) in the Marche
105 region, Italy. $H_2S(g)$ concentrations at sampling sites ranged from 0.2 to 24 ppmv (Table 1),
106 except for samples GS07-35 and AS08-5 where $H_2S(g)$ was below detection. Snottites were also
107 observed (but not collected) in the Yellow Roses room of Cueva de Villa Luz, at $H_2S(g)$
108 concentrations between 30-50 ppm. $H_2S(g)$ commonly exceeds 80 ppm in the atmosphere of this
109 enclosed room, and snottites are not observed under these conditions (30, 35). Similarly, no
110 snottites were observed near a thermal cave stream in Acquasanta, where $H_2S(g)$ concentrations
111 commonly exceed 100 ppm (37, 38). We have observed that snottites form on a time scale as
112 short as months to weeks in Frasassi and Acquasanta, and at Villa Luz they have been observed
113 to form in days (39).

114 **16S rRNA gene libraries.** We cloned archaeal and bacterial 16S rRNA genes from six
115 snottite samples (Table S1). *Acidithiobacillus* phylotypes (Figure 2) were the most abundant
116 clones in all bacterial libraries, and were the only bacterial sequences retrieved from samples
117 PC05-1, AS07-3, and AS08-5. The other three libraries contained relatives of the bacterial
118 genera *Acidimicrobium* and *Ferrimicrobium* (Figure 3). Libraries from samples RS05-2 and
119 VL07-20 also contained phylotypes of the genus *Sulfobacillus* and candidate phylum
120 *Dependentiae* (formerly TM6 (40)) (Figure S1, Figure S2). Samples VL07-20 and LA07-10 also
121 contained members of the *Saccharibacteria* (formerly candidate phylum TM7 (41)).
122 *Ferroplasma* relatives are the most abundant archaeal clones in universal and archaeal-specific
123 clone libraries (Figure 4). Other archaeal clones are related to "C-plasma" and *Cuniculiplasma*
124 (formerly "G-plasma" (3, 42)) in the order *Thermoplasmatales* (Figure 4).
125 Although many of the same microbial genera were present in snottites from both Italy
126 and Mexico, snottite clones from the two countries have less than 97% 16S rRNA gene sequence
127 identity (Table S1). *Acidithiobacillus* clones are <95% similar (Figure 2), clones affiliated with
128 *Acidimicrobium/Ferrimicrobium* are <95% similar (Figure 3), *Sulfobacillus* clones are <92%
129 similar (Figure S1), and *Ferroplasma* sequences are <97% similar (Figure 4). The only exception
130 are C-plasma clones VLSa30 and DSJa51, which share 99.6% similarity (Figure 3, Table S1),
131 and clone DSJb30 in the *Dependentiae*, which is 97.9-98.1% similar to Villa Luz clones VLSb7,
132 VLSb29, and VLSb87 (Figure S2).
133 All of the *Acidithiobacillus* clones from Italy are members of the genus *Acidithiobacillus*
134 *thiooxidans*, while all *Acidithiobacillus* clones from the Mexican caves represent a separate
135 species of *Acidithiobacillus* (referred to as "*At.* group II" in Figure 2, after Jones et al. (2016))
136 that is more closely related to *At. caldus* than *At. thiooxidans*. In order to see if deeper
137 sequencing would reveal the presence of *At. thiooxidans* in Cueva de Villa Luz, we generated
138 much larger rRNA gene libraries from samples collected during a later expedition to Villa Luz.
139 *At. thiooxidans* sequences were not detected among 143,944 and 102,229 rRNA gene amplicon
140 sequences from two snottite samples. Like the clone sequences, *Acidithiobacillus* spp. in these
141 samples formed a single operational taxonomic unit (OTU) that was similar to clone sequences
142 in *At.* group II. Other abundant OTUs in the deeper amplicon datasets include members of
143 groups that were identified in clone libraries (*Acidimicrobiaceae*, *Sulfobacillus*,
144 *Saccharibacteria*, *Dependentiae*), as well as other members of the *Gamma*- and

145 *Alphaproteobacteria* that were not identified in the clones libraries, including an abundant OTU
146 of *Metallibacterium* and multiple OTUs from the *Acidimicrobiaceae* (Figure 2, 3, S1, S2, Table
147 S2).

148 **Fluorescence *in situ* hybridization (FISH).** We sampled 26 snottite communities from
149 14 cave locations using FISH (Table 2). Representative FISH photomicrographs are shown in
150 Figures 5 and 6. All samples contain over 50% THIO1-positive cells (*Acidithiobacillus*), with the
151 exception of AS05-5. In sample AS05-5, over 25% of DAPI-strained cells did not hybridize with
152 any FISH probes, possibly indicating that a large proportion of the community was inactive at
153 the time of collection. ACM732-positive cells (*Acidimicrobium*/*Ferrimicrobium*) are present in
154 most but not all samples, and make up to 13.7% of total cells. In most samples, bacterial cells
155 that do not hybridize with either THIO1 or ACM732, hereafter referred to as "other bacteria,"
156 are <3% of cells. However, samples RS08-31, LA07-10, and VL07-26 have between 5.9-6.9%
157 other bacteria, and sample VL07-20 contains 15% other bacteria. At least some archaea are
158 present in all samples, although some samples contain less than 2% ARCH915-positive cells.
159 Twelve samples contain >2% FER656-positive cells (*Ferroplasma*), including four Acquasanta
160 samples with between 25-49% FER656. However, *Ferroplasma* are absent from some samples
161 that have abundant ARCH915-positive cells. Non-*Ferroplasma* archaea make up 15-30% of total
162 cells in several samples, including AS08-6 and several Frasassi RS samples.

163 We used principal component analysis (PCA) to explore the major axes of variance in
164 community composition based on the FISH dataset (Figure 7). For this and other statistical
165 analyses using FISH data (Table 2), our "species" were cell counts of THIO1, ACM732,
166 FER656, "other bacteria" (EUBMIX – (THIO1 + ACM732)), and "other archaea" (ARCH915 –
167 FER656). Principal components 1 and 2 combined explain more than 75% of the total variance.
168 The first principal component in Figure 7 appears to be related to location, with Acquasanta
169 samples plotting opposite Villa Luz samples. Species loadings show that the taxa driving this
170 clustering are FER656-positive cells, which were most abundant in the Acquasanta samples, and
171 ACM732 and "other bacteria" that were more abundant in Villa Luz.

172 ANalysis Of SIMilarities (ANOSIM; McCune and Grace, 2002) of FISH data indicates
173 that communities from different caves are significantly different (ANOSIM R = 0.268, p =
174 0.013), and this result remains significant if Villa Luz and Luna Azufre samples are pooled
175 (ANOSIM R = 0.287, p = 0.006). Mantel tests show that there is a statistically significant

176 correlation between community dissimilarity and geographical distance, both for all samples and
177 for Italian samples alone, if rank correlation is used (Mantel test with Spearman's *rho*: all
178 samples, *rho*=0.174, p=0.045; Italian samples only, *rho*=0.358, p<0.001; Frasassi samples only,
179 *rho*=0.158, p=0.065).
180 We also examined whether community composition is related to the concentration of
181 gases in the cave air. When these variables are fitted to the ordination analysis, CO_2 and SO_2
182 concentration align with the first ordination axis in the direction of the Acquasanta samples
183 (Figure 7), consistent with the high gas concentrations in that cave. These two variables are
184 statistically significantly correlated with the ordination axes ($r^2 = 0.33$, $p = 0.02$ for CO_2; $r^2 =$
185 0.31, $p = 0.02$ for SO_2), while H_2S concentration and temperature are not ($r^2 = 0.15$, $p = 0.22$ for
186 H_2S; $r^2 = 0.10$, $p = 0.34$ for temperature).
187
188
189 **Discussion**
190
191 **Snottite microbial communities.** *Acidithiobacillus* spp. were more than 50% of total
192 cells in all except one of the 26 snottite samples analyzed with FISH, and they were frequently
193 more than 80% of total cells (Table 2). *Acidithiobacillus* was also the only group present in all of
194 the biofilms sampled. These findings are consistent with other studies of snottite biofilms from
195 Frasassi and Cueva de Villa Luz (30, 31, 34, 35). Based on previous metagenomic and culture-
196 based analyses, *Acidithiobacillus thiooxidans* in Italian snottites are autotrophic sulfide-oxidizers
197 with the capacity to form extracellular polymeric substances, and were described as possible
198 'architects' of the biofilms (32, 36). The predominance of *Acidithiobacillus* spp. across the large
199 number of samples analyzed here is further evidence that *Acidithiobacillus* are responsible for
200 forming the acidic biofilms, and that they are important organisms for sulfide oxidation and
201 sulfuric acid production in sulfidic caves.
202 Close relatives or members of the bacterial genera *Acidimicrobium* and *Ferrimicrobium*
203 were present in most samples. Based on metagenomic analyses of Frasassi snottites, these
204 populations do not have the capacity to fix inorganic carbon, and are likely
205 chemoorganoheterotrophs or chemolithoheterotrophs (32, 33). However, the coverage of these
206 populations in previous snottite metagenomes was very low, and 16S rRNA genes recovered

207 from snottite populations are divergent from known isolates within the *Acidimicrobiaceae* family
208 (Figure 2), so inferences about their physiology are necessarily limited.
209 Other bacterial members included *Sulfobacillus* and members of the candidate phyla
210 *Dependentiae* (formerly TM6) and *Saccharibacteria* (formerly TM7). We hypothesize that,
211 based on their closest relatives in phylogenetic analyses and other members of their genera, the
212 *Sulfobacillus* represent another inorganic sulfur-oxidizing or organoheterotrophic population
213 (e.g., 43, 44) and the *Dependentiae* and *Saccharibacteria* are likely obligate symbionts or
214 parasites (40, 45, 46) (Figure 2; Figure S1). The protists and fungi that are commonly observed
215 in snottites (Figures 5, 6, S3; (31)) could be hosts for symbiotic *Dependentiae* and
216 *Saccharibacteria*. We observed nematodes and arthropods in Villa Luz and Luna Azufre
217 snottites that are additional potential hosts for symbionts (Figure S3). These likely represent the
218 extremely acid-tolerant bacteriophagous nematode (*Mesorhabditis acidophila*) and nematode-
219 eating mites previously described from Villa Luz snottites (47).
220 Archaea were present in all of the snottites sampled (Table 2). Members of the genus
221 *Ferroplasma* were detected in 18 out of the 26 samples analyzed using FISH, and were up to
222 50% of total cells. However, we also observed significant numbers of ARCH915-positive cells
223 that did not hybridize with FER656. Other archaeal populations identified in this and previous
224 studies of snottites include 'C-plasma' and *Cuniculiplasma* (formerly 'G-plasma') in the
225 *Thermoplasmatales* (Figure 4) (32, 33). In previous metagenomic analysis, *Ferroplasma* from
226 the Acquasanta caves had the genetic potential for sulfur oxidation, and *Cuniculiplasma* from
227 Frasassi had a homologue of sulfide quinone oxidoreductase (33), so these archaeal members
228 may have some capacity for lithotrophy in addition to organotrophic growth.
229 Villa Luz snottites were more diverse than snottites from other caves (Table S1). The
230 abundance of "other bacteria" in Villa Luz (Figure 7, Table 2) indicates that additional diversity
231 in these samples remains to be explored, and the deeper rRNA gene amplicon libraries from this
232 cave contain several taxa that haven't previously been described from snottites (see also D'Auria
233 et al. (35)). More work is needed to reveal the diversity of extreme acidophiles in this cave
234 community.
235 **Implications for community assembly.** Sulfidic cave snottites from across the globe are
236 populated by extreme acidophiles from the same few genera and families. Within those clades,
237 however, populations are divergent. The *Acidithiobacillus*, *Sulfobacillus*, and *Acidimicrobiaceae*

238 from snottites in the Mexican caves are less than 95% similar to those from the Italian caves, and
239 *Ferroplasma* spp. are less than 97% similar (Figure 2, 3, 4, S1). Furthermore, based on FISH cell
240 counts, differences among snottite communities are statistically significantly related to cave
241 location (Figure 7), so communities are site specific.

242 This pattern is consistent with a model of snottite community assembly in which caves
243 are stochastically colonized by microorganisms from local sources, followed by selective
244 filtering for those organisms that can survive in the extremely acidic subaerial cave environment.
245 If colonization events are sufficiently rare, then differences among snottite communities could
246 reflect differences in the founding community and/or later colonizers. Dispersal rates of extreme
247 acidophiles into these caves should be very low: snottites are analogous to extremely acidic
248 "islands" in a "sea" of alkaline limestone bedrock, so survival of extreme acidophiles during
249 transport into the cave environment would be rare. Consistent with this constraint, previous
250 research on the population structure of *Acidithiobacillus* spp. provides strong evidence for
251 restricted dispersal within and among sulfidic caves (36), and studies of other subsurface
252 communities have shown that they retain a site-specific "fingerprint" of their founding
253 communities (e.g., 28). If this model is correct, then differences in the composition of snottite
254 communities are partially the result of historical contingency. Whatever extreme acidophiles
255 happened to enter the cave were those that went on to form and colonize the extremely acidic
256 biofilms. Restricted dispersal could lead to novel diversity in this situation as those populations
257 are free to occupy new niche space and evolve and adapt *in situ*.

258 Differences among snottite communities could also result from environmental selection.
259 Earlier research found that *Acidithiobacillus* spp. from Frasassi and Acquasanta had differences
260 in their sulfur oxidation pathways and other functional genes (36), which indicates distinct
261 metabolic capabilities that could confer selective advantages in the environment. These two
262 *Acidithiobacillus* populations from Frasassi and Acquasanta had identical 16S rRNA genes.
263 Therefore, the far more divergent *Acidithiobacillus*, *Ferroplasma*, *Sulfobacillus,* and
264 *Acidimicrobiaceae* in Mexico versus Italy likely have substantial physiological differences that
265 the cave environment could select for. Snottites were chosen for this study because they occur in
266 a very specific geochemical niche in sulfidic caves (overhanging subaerial cave surfaces, $H_2S(g)$
267 between 0.2-25 ppm, extremely low pH; Figure 1a), so environmental variability among them
268 should be limited. However, we measured subtle but important geochemical differences: Frasassi

269 was substantially cooler than Acquasanta, Villa Luz, and Luna Azufre, $SO_2(g)$ was present in the
270 atmosphere of Acquasanta but not the other caves at the time of sampling (Figure 7, Table 1),
271 and there are likely other differences in nutrient or energy sources that we did not measure.
272 Villa Luz snottites were dominated by *Acidithiobacillus* spp. "group 2" that are more
273 closely related to *At. caldus* than *At. thiooxidans*. If *At. thiooxidans* populations were rare but
274 present in Villa Luz snottites, that would be evidence that they were being selected against.
275 However, deep amplicon sequencing did not identify any *At. thiooxidans* out of 246,173
276 amplicon sequences from Villa Luz snottites, although that does not exclude the possibility that
277 *At. thiooxidans* have had other opportunities to colonize the cave or that they exist elsewhere in
278 the cave. Interestingly, this is not true of the *Acidimicrobiaceae*. The most abundant
279 *Acidimicrobiaceae* OTU from the Villa Luz amplicon libraries clusters with the Villa Luz clones
280 (24.8% and 13.7% of VL13-1 and VL13-2 libraries, respectively), but there is a second rare
281 OTU that is more similar to the Frasassi clones that represents 1.1% and 0.003% of total
282 sequences from these libraries (OTU_9 in Figure 3). Nevertheless, this rare OTU is still >3%
283 different from the Frasassi *Acidimicrobiaceae* clones. Evolutionary rates for 16S rRNA genes are
284 0.025 to 0.091% per million years (48), so these differences represent at least tens of millions of
285 years of evolution, which highlights how different the Villa Luz and Luna Azufre populations
286 are compared to those in Frasassi and Acquasanta. The formation and colonization of snottites is
287 apparently an adaptation to the sulfidic cave environment that can be exploited by substantially
288 different organisms.
289
290
291 **Methods**
292
293 **Sample collection and geochemistry.** Snottite biofilm samples were collected from four
294 sulfidic caves: Cueva de Villa Luz and Cueva Luna Azufre in Tabasco, Mexico, and the Grotte
295 del Fiume (Frasassi) and Grotta Nuova di Rio Garrafo (Acquasanta Terme) in the Marche region,
296 Italy. Villa Luz is located 2 km south of Tapijulapa in Tabasco, Mexico, and contains roughly 2
297 km of passage in a Cretaceous limestone. The entrance to Luna Azufre is located 320 meters
298 southeast from Villa Luz, and Luna Azufre has over 500 m of passage in the same formation.
299 Both caves are fed by phreatic water containing up to 15 mM $H_2S(aq)$ that could originate either

300 from nearby volcanic or petroleum sources (30, 49, 50). The Frasassi Caves are located 3 km
301 southeast of Genga in the Marche region, Italy, and contain over 25 km of passage in a Jurassic
302 platform limestone. Sulfidic waters entering Frasassi contain up to 600 μM $H_2S(aq)$ that likely
303 originates from sulfate reduction in an underlying evaporite formation (51, 52). Grotta Nuova di
304 Rio Garrafo is located 2 km south of Acquasanta Terme, and 80 km to the southeast of Frasassi.
305 The Acquasanta cave consists of over 1 km of passage in a marly Eocene limestone, and is fed
306 by thermal sulfidic waters that reach temperatures up to 50°C with over 800 μM $H_2S(aq)$ (37,
307 53).
308 Samples were collected from 2-16 locations within each cave. At each sampling site,
309 multiple snottites (approximately 0.5-1.5 grams of biofilm) were collected into sterile centrifuge
310 tubes from an area of cave wall no larger than one square meter. At the time of collection, the pH
311 values of at least 5-10 snottites from the collection area were measured with pH paper (range 0-
312 2.5). The pH of all samples was between 0-1.5.
313 $H_2S(g)$ concentration was measured at the time of collection using Dräger tubes (range
314 0.2-60 ppm) (Dräger Safety Inc, Germany) and by an EN-MET MX2100 portable gas detector
315 (range 1-120 ppm) (ENMET corp, USA). $SO_2(g)$ and $CO_2(g)$ concentrations were also measured
316 by ENMET MX2100 (lower detection limits 0.1 ppm and 0.1 %, respectively), and $CO_2(g)$ was
317 also measured by Dräger tubes (100 to 6000 ppm) at some sites. In Villa Luz and Luna Azufre,
318 samples for CO_2(g) concentration were collected in sealed 60 mL serum bottles (2 bottles per
319 site, flushed 10 times each with 1 volume cave air) and measured with a SRI 310 gas
320 chromatograph (GC) equipped with a thermal conductivity detector and a PoraPak Q packed
321 column (3' x 1/8") with helium carrier gas at 80°C (SRI instruments, USA). $NH_3(g)$ and $N_2O(g)$
322 were measured using Dräger tubes at Acquasanta and Frasassi in 2005-2007, but those gases
323 were never detected (lower detection limit 0.25 and 0.5 ppm, respectively). The total range of
324 multiple gas measurements at each site is given in Table 1.
325 **16S rRNA gene libraries.** For DNA extraction, approximately 1 gram of snottite biofilm
326 was preserved in 5 parts RNAlater (Ambion, USA) within 4 hours of collection, and stored at -
327 20°C. DNA extraction of samples RS05-2 and PC05-1 were described in Macalady et al. (2007).
328 For samples RS05-24, VL07-20, LA07-10, AS07-3, AS08-5, and RS09-1, DNA was extracted
329 via bead beating and phenol-chloroform extraction as described in (32, 54). RNAlater-preserved
330 samples were first diluted with 3 parts 1x phosphate-buffered saline (PBS). To remove

331 polysaccharides from the DNA extract, prior to final elution the DNA was resuspended in 200
332 μL tris (200 mM, pH 8.0), 100 μL NaCl (5 M), and 600 μL ethanol (100%), incubated at -20°C
333 for 30 minutes, and pelleted by centrifugation.
334 Bacterial, universal, and archaeal 16S rRNA gene clone libraries were constructed from
335 environmental DNA templates. For bacterial and universal libraries, 16S rRNA genes were
336 amplified by polymerase chain reaction (PCR), ligated into pCR4-TOPO vector, and transformed
337 into OneShot Mach1 *E. coli* as described in (31). Archaeal clones were sequenced from sample
338 VL07-20 using forward primer 344f (ACG GGG YGC AGC AGG CGC GA) (55) and reverse
339 primer 1492r (GGT TAC CTT GTT ACG ACT T) (56). Archaeal clones were sequenced from
340 samples RS05-24 and RS09-1 using 344f and reverse primer deg1392r (ACR GGC GGT GTG
341 TRC) (32). Amplification, ligation, and transformation of the archaeal sequences was performed
342 as in Jones et al. (32). Some additional 16S rRNA gene clones were recovered from Frasassi
343 samples GS08-6, GB08-30, and RS08-31 using 27F and primer Sag2 for the *Acidithiobacillus*
344 16S-23S intergenic transcribed spacer region, as described in (36). Inserts were extracted from
345 RS05-2 and PC05-1 via plasmid extraction as described in (31). Inserts for all other samples
346 were amplified by colony PCR as described in (57). Clones were sequenced at the Penn State
347 University Nucleic Acid Facility.
348 High-throughput rRNA gene "amplicon" libraries were generated from two snottite
349 samples from Villa Luz collected in 2013 (58). Samples were preserved in sucrose lysis buffer
350 (59), and DNA was extracted using the MoBio PowerSoil™ DNA Isolation Kit (MoBio,
351 Carlsbad, CA). Amplicon libraries were created and sequenced by Molecular Research LP (MR
352 DNA) (Shallowater, TX), using bacterial-specific primers 46f (5′-
353 GCYTAAYACATGCAAGTCG-3′) and 1409r (5′-GTGACGGGCRGTGTGTRCAA-3′) (59,
354 60).
355 **Fluorescence *in situ* hybridization (FISH).** Approximately 0.25-0.5 grams of biofilm
356 was fixed for FISH analysis. Samples were either fixed for 3-4 hours in 3% (w/v)
357 paraformaldehyde within 10 hours of collection, or stored in RNAlater (Ambion, USA) as
358 described above and later fixed in paraformaldehyde. After fixation, samples were stored in 1:1
359 PBS/ethanol at -20°C. Fluorescence *in situ* hybridization followed the procedure outlined by
360 Hugenholtz et al. (61). Environmental samples were applied to 10-well Teflon coated slides,
361 dried, and dehydrated for 3 minutes each in 50%, 80%, and 90% ethanol washes. Hybridization

362 buffer contained 0.9 M NaCl, 20 mM of Tris/HCl pH 7.4, 0.01% sodium dodecyl sulfate (SDS),
363 25 ng of each fluorescently labeled probe (Table 3), and variable formamide concentrations
364 depending on the probe (Table 3). Slides were incubated at 46ºC for 2 hours, followed by a 15
365 minute incubation at 48ºC in wash buffer containing 20 mM Tris/HCl pH 7.4, 0.01% SDS, and
366 variable NaCl concentrations following Lathe (62). Slides were rinsed with deionized water,
367 counterstained with 4',6'-diamidino-2-phenylindole (DAPI), dried, and mounted with
368 Vectashield (Vector Laboratories Inc, Burlingame, CA, USA). Slides were viewed at 1000x
369 magnification with a Nikon Eclipse 80i epifluorescence microscope, and photomicrographs were
370 taken with a monochrome Photometrics CoolSNAP ES^2 CCD camera using NIS-Elements AS
371 3.0 software.

372 Between 800-1200 DAPI-stained cells were counted for each probe combination. Cells
373 were counted from at least six photos, each representing a different microscope field, from at
374 least two slide wells. Standard deviation was calculated from differences in proportion of probe-
375 labeled cells among each photo. To ensure objectivity in counts, photos for counting were taken
376 without prior examination of the field of view under more than one filter (EUBMIX whenever
377 possible).

378 ***Phylogenetic, bioinformatic, and statistical analyses.*** Clone sequences were assembled
379 and edited using CodonCode Aligner v.2.0.4 (CodonCode Corp., USA). Chimeric sequences
380 were identified and removed using Bellerophon 3 (63). Contiguous sequences were aligned using
381 the SINA aligner (64) (https://www.arb-silva.de) and loaded into an ARB (65) database with
382 SILVA database v. 132 (66, 67). Alignments were manually refined using the ARB_Edit4
383 aligner.

384 For phylogenetic analysis, alignments were trimmed to equalize sequence length, and
385 positions with <50% gaps were masked. Maximum likelihood phylogenies of full-length
386 sequences were calculated with RAxML v.8.2.12 (68), using the general time reversible
387 nucleotide substitution model, gamma distributed rates, the proportion of invariant sites and base
388 frequencies estimated from the data, and 100 rapid bootstrap replicates. Percent nucleotide
389 sequence similarity was calculated in ARB.

390 High throughput amplicon libraries were processed using the pipeline described in (69),
391 except that only the first 200 bp of the R1 (forward) read was used for OTU clustering and
392 taxonomic identification. Taxonomic classification of representative OTU sequences was

393 determined with mothur v.1.3.2 using the SILVA database v.132 and a confidence cutoff of 50
394 (70). Short amplicon sequences were then placed into phylogenetic trees using the evolutionary
395 placement algorithm (EPA) with default parameters (71).
396 For statistical analyses of FISH cell counts, we used average proportions of THIO1,
397 ACM732, FER656, "other bacteria" (total EUBMIX – (THIO1 + ACM732)), and "other
398 archaea" (ARCH915 – FER656). Principal component analyses (PCA), ANOSIM and Manel
399 tests (72) were performed using the vegan package in R (73, 74). The geographic distance matrix
400 used for Mantel tests included linear distances among individual caves and linear distances
401 between sampling locations within the same cave system, either measured at the time of
402 sampling or from cave maps (Table S3). Mantel statistics were evaluated with Spearman's rank
403 correlation and 1000 randomizations. For PCA analyses, cross-product matrices were calculated
404 with Pearson's correlation coefficients.
405 **Accession number(s).** 16S rRNA gene clones referenced in this study are available in the
406 GenBank database under accession numbers KU341124-KU341208, OP269604-OP269649,
407 KC582420-KC582530, and DQ499162-DQ499330. Amplicon libraries are available in NCBI's
408 Sequence Read Archive under project PRJNA878455.
409
410
411 **Acknowledgements**
412
413 Thanks to A. Montanari for logistical support and the use of facilities and laboratory space at the
414 Osservatorio Geologico di Coldigioco (Italy). S. Mariani, S. Galdenzi, S. Cerioni, and M.
415 Mainiero provided expert advice and field assistance with research in Italy, and F. Baldoni, S.
416 Recanatini, S. Dattagupta, and members of the Gruppo Speleologico CAI di Fabriano and the
417 Gruppo Speleologico Marchigiano CAI di Ancona assisted with sample collection and provided
418 additional excellent field support. We thank L. Rosales-Lagarde, L. Hose, and S. Dattagupta for
419 logistical support, insightful discussion, and field assistance with sample collection in Mexico.
420 Thanks to the municipal government of Tacotalpa, Tabasco, Mexico, and to C. Rogers Morales
421 Mendez and L. Felino Arevalo Gallegos, for granting permits for cave access and sample
422 collection. C. Alberto Cordero Martinez provided lodging and logistical support in Villa Luz
423 Park. The two 2013 Villa Luz samples were collected during an expedition supported by

National Geographic grant #EC0644-13 to P. Boston, and sequencing of these two libraries was supported by a Vehslage Grant from the National Speleological Foundation. This work was supported by a graduate research fellowship to DSJ from the Cave Conservancy Foundation, and grants to JLM from the National Science Foundation (EAR 0311854 and EAR 0527046) and NASA NAI (NNA04CC06A).

431 REFERENCES
432

433 1. Vera M, Schippers A, Sand W. 2013. Progress in bioleaching: fundamentals and
434 mechanisms of bacterial metal sulfide oxidation—part A. Appl Microbiol Biotechnol
435 97:7529-7541.
436 2. Johnson DB. 2014. Biomining—biotechnologies for extracting and recovering metals
437 from ores and waste materials. Curr Opin Biotechnol 30:24-31.
438 3. Baker BJ, Banfield JF. 2003. Microbial communities in acid mine drainage. FEMS
439 Microbiol Ecol 44:139-152.
440 4. Rowe OF, Sánchez-España J, Hallberg KB, Johnson DB. 2007. Microbial communities
441 and geochemical dynamics in an extremely acidic, metal-rich stream at an abandoned
442 sulfide mine (Huelva, Spain) underpinned by two functional primary production systems.
443 Environ Microbiol 9:1761-1771.
444 5. Santelli CM, Pfister DH, Lazarus D, Sun L, Burgos WD, Hansel CM. 2010. Promotion of
445 Mn(II) oxidation and remediation of coal mine drainage in passive treatment systems by
446 diverse fungal and bacterial communities. Appl Environ Microbiol 76:4871-4875.
447 6. Mansor M, Harouaka K, Gonzales MS, Macalady JL, Fantle MS. 2018. Transport-
448 induced spatial patterns of sulfur isotopes ($\delta^{34}S$) as biosignatures. Astrobiology 18:59-72.
449 7. Burgos WD, Borch T, Troyer LD, Luan F, Larson LN, Brown JF, Lambson J, Shimizu
450 M. 2012. Schwertmannite and Fe oxides formed by biological low-pH Fe(II) oxidation
451 versus abiotic neutralization: Impact on trace metal sequestration. Geochim Cosmochim
452 Acta 76:29-44.
453 8. Grettenberger CL, Pearce AR, Bibby KJ, Jones DS, Burgos WD, Macalady JL. 2017.
454 Efficient low-pH iron removal by a microbial iron oxide mound ecosystem at Scalp Level
455 Run. Appl Environ Microbiol:e00015-17.
456 9. Johnson DB, Hallberg KB. 2005. Acid mine drainage remediation options: a review.
457 Science of the Total Environment 338:3-14.
458 10. Kumar RN, Nagendran R. 2007. Influence of initial pH on bioleaching of heavy metals
459 from contaminated soil employing indigenous *Acidithiobacillus thiooxidans*.
460 Chemosphere 66:1775-1781.
461 11. Larson LN, Sánchez-España J, Kaley B, Sheng Y, Burgos WD. 2014. Thermodynamic
462 controls on the kinetics of microbial low-pH Fe(II) oxidation. Environ Sci Technol
463 48:9246-9254.
464 12. Johnson DB. 2018. The evolution, current status, and future prospects of using
465 biotechnologies in the mineral extraction and metal recovery sectors. Minerals 8:343.
466 13. Pakostova E, Johnson DB. 2019. Microbial generation of sulfuric acid from granular
467 elemental sulfur in laboratory-scale bioreactors. Hydrometallurgy 190:105152.
468 14. Rawlings DE, Johnson DB. 2007. The microbiology of biomining: development and
469 optimization of mineral-oxidizing microbial consortia. Microbiology 153:315-324.
470 15. Rowe OF, Johnson DB. 2008. Comparison of ferric iron generation by different species
471 of acidophilic bacteria immobilized in packed-bed reactors. Syst Appl Microbiol 31:68-
472 77.
473 16. Tyson GW, Chapman J, Hugenholtz P, Allen EE, Ram RJ, Richardson PM, Solovyev
474 VV, Rubin EM, Rokhsar DS, Banfield JF. 2004. Community structure and metabolism
475 through reconstruction of microbial genomes from the environment. Nature 428:37-43.

17. Whitaker RJ, Banfield JF. 2005. Population dynamics through the lens of extreme environments. Rev Mineral Geochem 59:259-277.
18. Palacios C, Zettler E, Amils R, Amaral-Zettler L. 2008. Contrasting microbial community assembly hypotheses: a reconciling tale from the Rio Tinto. PLoS ONE 3:e3853.
19. Lear G, Niyogi D, Harding J, Dong Y, Lewis G. 2009. Biofilm bacterial community structure in streams affected by acid mine drainage. Appl Environ Microbiol 75:3455-3460.
20. Kuang J-L, Huang L-N, Chen L-X, Hua Z-S, Li S-J, Hu M, Li J-T, Shu W-S. 2013. Contemporary environmental variation determines microbial diversity patterns in acid mine drainage. ISME J 7:1038-1050.
21. Sheng Y, Bibby K, Grettenberger C, Kaley B, Macalady JL, Wang G, Burgos WD. 2016. Geochemical and temporal influences on the enrichment of acidophilic iron-oxidizing bacterial communities. Appl Environ Microbiol 82:3611-3621.
22. Jones DS, Kohl C, Grettenberger C, Larson LN, Burgos WD, Macalady JL. 2015. Geochemical niches of iron-oxidizing acidophiles in acidic coal mine drainage. Appl Environ Microbiol 81:1242-1250.
23. Bond PL, Druschel GK, Banfield JF. 2000. Comparison of acid mine drainage microbial communities in physically and geochemically distinct ecosystems. Appl Environ Microbiol 66:4962-4971.
24. Edwards KJ, Gihring TM, Banfield JF. 1999. Seasonal variations in microbial populations and environmental conditions in an extreme acid mine drainage environment. Appl Environ Microbiol 65:3627-3632.
25. Schrenk MO, Edwards KJ, Goodman RM, Hamers RJ, Banfield JF. 1998. Distribution of *Thiobacillus ferrooxidans* and *Leptospirillum ferrooxidans*: Implications for generation of acid mine drainage. Science 279:1519-1522.
26. Martiny JBH, Bohannan BJ, Brown JH, Colwell RK, Fuhrman JA, Green JL, Horner-Devine MC, Kane M, Krumins JA, Kuske CR. 2006. Microbial biogeography: putting microorganisms on the map. Nature Reviews Microbiology 4:102-112.
27. Langenheder S, Lindström ES, Tranvik LJ. 2006. Structure and function of bacterial communities emerging from different sources under identical conditions. Appl Environ Microbiol 72:212-220.
28. Marshall IPG, Ren G, Jaussi M, Lomstein BA, Jørgensen BB, Røy H, Kjeldsen KU. 2019. Environmental filtering determines family-level structure of sulfate-reducing microbial communities in subsurface marine sediments. ISME J 13:1920-1932.
29. Takacs-Vesbach C, Mitchell K, Jackson-Weaver O, Reysenbach A-L. 2008. Volcanic calderas delineate biogeographic provinces among Yellowstone thermophiles. Environ Microbiol 10:1681-1689.
30. Hose LD, Palmer AN, Palmer MV, Northup DE, Boston PJ, DuChene HR. 2000. Microbiology and geochemistry in a hydrogen-sulphide-rich karst environment. Chem Geol 169:399-423.
31. Macalady JL, Jones DS, Lyon EH. 2007. Extremely acidic, pendulous microbial biofilms from the Frasassi cave system, Italy. Environ Microbiol 9:1402-1414.
32. Jones DS, Albrecht HL, Dawson KS, Schaperdoth I, Freeman KH, Pi Y, Pearson A, Macalady JL. 2012. Community genomic analysis of an extremely acidophilic sulfur-oxidizing biofilm. ISME J 6:158-170.

522 33. Jones DS, Schaperdoth I, Macalady JL. 2014. Metagenomic evidence for sulfide
523 oxidation in extremely acidic cave biofilms. Geomicrobiol J 31:194-204.
524 34. Vlasceanu L, Sarbu SM, Engel AS, Kinkle BK. 2000. Acidic cave wall biofilms located
525 in the Frasassi Gorge, Italy. Geomicrobiol J 17:125-139.
526 35. D'Auria G, Artacho A, Rojas RA, Bautista JS, Méndez R, Gamboa MT, Gamboa JR,
527 Gómez-Cruz R. 2018. Metagenomics of bacterial diversity in Villa Luz Caves with sulfur
528 water springs. Genes 9:55.
529 36. Jones DS, Schaperdoth I, Macalady JL. 2016. Biogeography of sulfur-oxidizing
530 *Acidithiobacillus* populations in extremely acidic cave biofilms. ISME J 10:2879.
531 37. Galdenzi S, Cocchioni F, Filipponi G, Morichetti L, Scuri S, Selvaggio R, Cocchioni M.
532 2010. The sulfidic thermal caves of Acquasanta Terme (central Italy). J Cave Karst Stud
533 72:43-58.
534 38. Jones D, Tobler D, Schaperdoth I, Mainiero M, Macalady J. 2010. Community structure
535 of subsurface biofilms in the thermal sulfidic caves of Acquasanta Terme, Italy. Appl
536 Environ Microbiol 76:5902-5910.
537 39. Hose LD, Macalady JL. 2006. Observations from active sulfidic karst systems: is the
538 present the key to understanding Guadalupe Mountain speleogenesis, p 185-194. *In* Land
539 L, Lueth V, Raatz B, Boston P, Love D (ed), New Mexico Geological Society
540 Guidebook, 57th Annual Fall Field Conference, Caves and Karst of Southeastern New
541 Mexico, vol 57.
542 40. Yeoh YK, Sekiguchi Y, Parks DH, Hugenholtz P. 2015. Comparative genomics of
543 candidate phylum TM6 suggests that parasitism is widespread and ancestral in this
544 lineage. Mol Biol Evol 33:915-927.
545 41. Albertsen M, Hugenholtz P, Skarshewski A, Nielsen KL, Tyson GW, Nielsen PH. 2013.
546 Genome sequences of rare, uncultured bacteria obtained by differential coverage binning
547 of multiple metagenomes. Nat Biotechnol 31:533-538.
548 42. Golyshina OV, Lünsdorf H, Kublanov IV, Goldenstein NI, Hinrichs K-U, Golyshin PN.
549 2016. The novel extremely acidophilic, cell-wall-deficient archaeon *Cuniculiplasma*
550 *divulgatum* gen. nov., sp. nov. represents a new family, *Cuniculiplasmataceae* fam. nov.,
551 of the order *Thermoplasmatales*. Int J Syst Evol Microbiol 66:332-340.
552 43. Norris PR, Clark DA, Owen JP, Waterhouse S. 1996. Characteristics of *Sulfobacillus*
553 *acidophilus* sp. nov. and other moderately thermophilic mineral-sulphide-oxidizing
554 bacteria. Microbiology 142:775-783.
555 44. Panyushkina AE, Babenko VV, Nikitina AS, Selezneva OV, Tsaplina IA, Letarova MA,
556 Kostryukova ES, Letarov AV. 2019. *Sulfobacillus thermotolerans*: new insights into
557 resistance and metabolic capacities of acidophilic chemolithotrophs. Sci Rep 9:15069.
558 45. Murugkar PP, Collins AJ, Dewhirst FE. 2019. Complete genome sequence of strain
559 PM004, a novel cultured member of the human oral microbiome from the candidate
560 phylum *Saccharibacteria* (TM7). Microbiology Resource Announcements 8:e01159-19.
561 46. Deeg CM, Zimmer MM, George EE, Husnik F, Keeling PJ, Suttle CA. 2019.
562 *Chromulinavorax destructans*, a pathogen of microzooplankton that provides a window
563 into the enigmatic candidate phylum *Dependentiae*. PLoS Path 15:e1007801.
564 47. Borgonie G, Dierick M, Houthoofd W, Willems M, Jacobs P, Bert W. 2010. Refuge
565 From predation, the benefit of living in an extreme acidic environment? The Biological
566 Bulletin 219:268-276.

48. Kuo CH, Ochman H. 2009. Inferring clocks when lacking rocks: the variable rates of molecular evolution in bacteria. Biol Direct 4:35.
49. Hose LD, Rosales-Lagarde L. 2017. Sulfur-rich caves of Southern Tabasco, Mexico, p 803-814. *In* Klimchouk A, N. Palmer A, De Waele J, S. Auler A, Audra P (ed), Hypogene Karst Regions and Caves of the World. Springer International Publishing, Cham. (doi:10.1007/978-3-319-53348-3_54)
50. Rosales Lagarde L, Boston PJ, Campbell AR, Hose LD, Axen G, Stafford KW. 2014. Hydrogeology of northern Sierra de Chiapas, Mexico: a conceptual model based on a geochemical characterization of sulfide-rich karst brackish springs. Hydrogeol J 22:1447-1467.
51. Galdenzi S, Jones DS. 2017. The Frasassi Caves: A "classical" active hypogenic cave, p 143-159. *In* Klimchouk A, N. Palmer A, De Waele J, S. Auler A, Audra P (ed), Hypogene Karst Regions and Caves of the World. Springer International Publishing, Cham. (doi:10.1007/978-3-319-53348-3_8)
52. Jones DS, Polerecky L, Galdenzi S, Dempsey BA, Macalady JL. 2015. Fate of sulfide in the Frasassi cave system and implications for sulfuric acid speleogenesis. Chem Geol 410:21-27.
53. Galdenzi S. 2017. The thermal hypogenic caves of Acquasanta Terme (Central Italy), p 169-182. *In* Klimchouk A, N. Palmer A, De Waele J, S. Auler A, Audra P (ed), Hypogene Karst Regions and Caves of the World. Springer International Publishing, Cham. (doi:10.1007/978-3-319-53348-3_10)
54. Bond PL, Smriga SP, Banfield JF. 2000. Phylogeny of microorganisms populating a thick, subaerial, predominantly lithotrophic biofilm at an extreme acid mine drainage site. Appl Environ Microbiol 66:3842-3849.
55. Raskin L, Stromley JM, Rittmann BE, Stahl DA. 1994. Group-specific 16S rRNA hybridization probes to describe natural communities of methanogens. Appl Environ Microbiol 60:1232-1240.
56. Lane D. 1991. 16S/23S rRNA sequencing, p 115-175. *In* Stackebrandt E GM (ed), Nucleic Acid Techniques in Bacterial Systematics. Wiley, New York.
57. Macalady JL, Dattagupta S, Schaperdoth I, Jones DS, Druschel GK, Eastman D. 2008. Niche differentiation among sulfur-oxidizing bacterial populations in cave waters. ISME J 2:590-601.
58. Kimble JC. 2018. Ph.D. thesis. The University of New Mexico.Comparison of bacterial and archaeal communities in the subsurface versus surface: Implications for nitrogen cycling.
59. Northup DE, Melim LA, Spilde MN, Hathaway JJM, Garcia MG, Moya M, Stone FD, Boston PJ, Dapkevicius MLNE, Riquelme C. 2011. Lava cave microbial communities within mats and secondary mineral deposits: Implications for life detection on other planets. Astrobiology 11:601-618.
60. Kimble JC, Winter AS, Spilde MN, Sinsabaugh RL, Northup DE. 2018. A potential central role of Thaumarchaeota in N-Cycling in a semi-arid environment, Fort Stanton Cave, Snowy River passage, New Mexico, USA. FEMS Microbiol Ecol 94.
61. Hugenholtz P, Tyson GW, Blackall LL. 2002. Design and evaluation of 16S rRNA-targeted oligonucleotide probes for fluorescence in situ hybridization, p 29-42. *In* de Muro MA, Rapley R (ed), Gene Probes: Principles and Protocols. Humana Press, Totowa, NJ. (doi:10.1385/1-59259-238-4:029)

613 62. Lathe R. 1985. Synthetic oligonucleotide probes deduced from amino acid sequence data:
614 Theoretical and practical considerations. J Mol Biol 183:1-12.
615 63. Huber T, Faulkner G, Hugenholtz P. 2004. Bellerophon; a program to detect chimeric
616 sequences in multiple sequence alignments. Bioinformatics 20:2317-2319.
617 64. Pruesse E, Quast C, Knittel K, Fuchs BM, Ludwig W, Peplies J, Glöckner FO. 2007.
618 SILVA: a comprehensive online resource for quality checked and aligned ribosomal
619 RNA sequence data compatible with ARB. Nucleic Acids Res 35:7188-7196.
620 65. Ludwig W, Strunk O, Westram R, Richter L, Meier H, Yadhukumar, Buchner A, Lai T,
621 Steppi S, Jobb G, Förster W, Brettske I, Gerber S, Ginhart AW, Gross O, Grumann S,
622 Hermann S, Jost R, König A, Liss T, Lüßmann R, May M, Nonhoff B, Reichel B,
623 Strehlow R, Stamatakis A, Stuckmann N, Vilbig A, Lenke M, Ludwig T, Bode A,
624 Schleifer K-H. 2004. ARB: a software environment for sequence data. Nucleic Acids Res
625 32:1363-1371.
626 66. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glöckner FO.
627 2012. The SILVA ribosomal RNA gene database project: improved data processing and
628 web-based tools. Nucleic Acids Res 41:D590-D596.
629 67. Yilmaz P, Parfrey LW, Yarza P, Gerken J, Pruesse E, Quast C, Schweer T, Peplies J,
630 Ludwig W, Glöckner FO. 2013. The SILVA and "All-species Living Tree Project (LTP)"
631 taxonomic frameworks. Nucleic Acids Res 42:D643-D648.
632 68. Stamatakis A. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses
633 with thousands of taxa and mixed models. Bioinformatics 22:2688-2690.
634 69. Jones DS, Lapakko KA, Wenz ZJ, Olson MC, Roepke EW, Sadowsky MJ, Novak PJ,
635 Bailey JV. 2017. Novel microbial assemblages dominate weathered sulfide-bearing rock
636 from copper-nickel deposits in the Duluth complex, Minnesota, USA. Appl Environ
637 Microbiol 83:e00909-17.
638 70. Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, Lesniewski RA,
639 Oakley BB, Parks DH, Robinson CJ. 2009. Introducing mothur: open-source, platform-
640 independent, community-supported software for describing and comparing microbial
641 communities. Appl Environ Microbiol 75:7537-7541.
642 71. Berger SA, Krompass D, Stamatakis A. 2011. Performance, accuracy, and web server for
643 evolutionary placement of short sequence reads under maximum likelihood. Syst
644 Biol:syr010.
645 72. McCune B, Grace JB, Urban DL. 2002. Analysis of Ecological Communities, vol 28.
646 MjM software design Gleneden Beach, OR.
647 73. R Development Core Team. 2007. R: A language and environment for statistical
648 computing., R Foundation for Statistical Computing, Vienna, Austria. http://www.R-
649 project.org.
650 74. Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, McGlinn D, Minchin PR,
651 O'Hara R, Simpson GL, Solymos P, Szoecs E, Wagner H. 2017. vegan: Community
652 Ecology Package. R package version 2.4-3. https://CRAN.R-project.org/package=vegan.
653
654

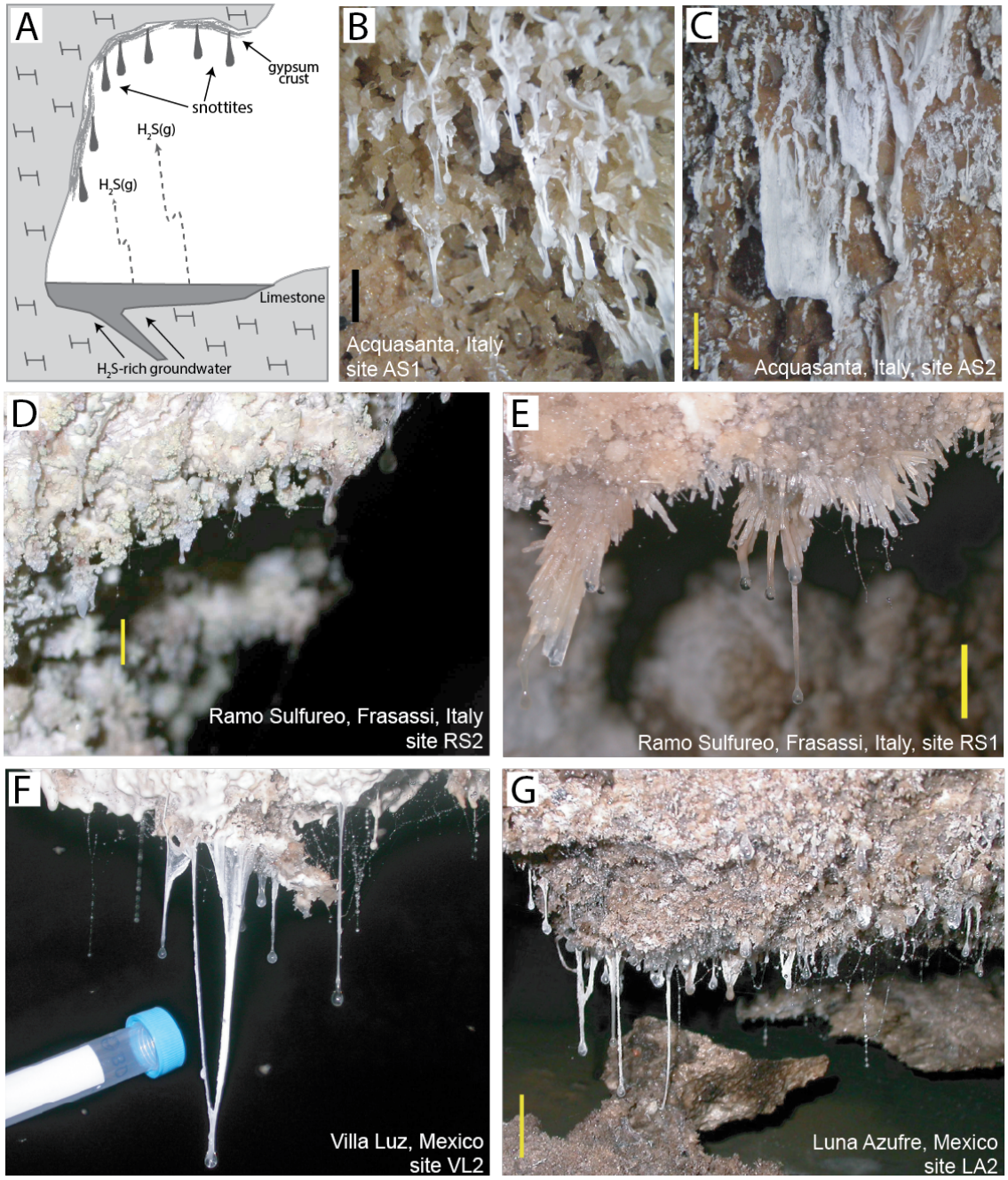

Figure 1. (A) Schematic depicting the snottite niche in sulfidic caves. Snottites hang from gypsum wall crusts on cave walls and ceilings in close proximity to sulfidic cave streams. (B-G) Representative photographs of snottite communities described in this study. Yellow scale bars are 2 cm, black bar in (b) is 1 cm, and the collection tube in (f) is 1.5 cm in diameter.

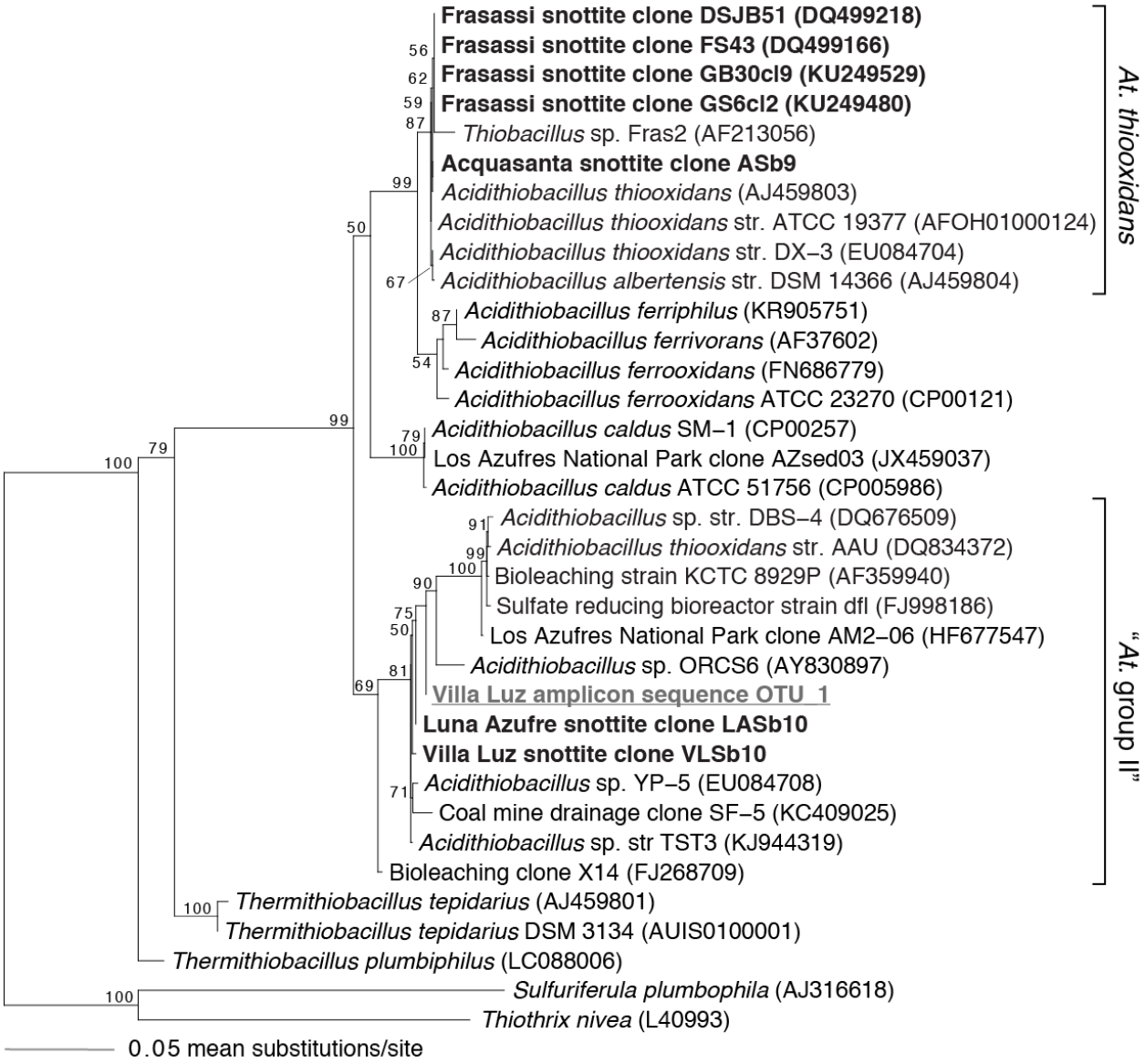

Figure 2. Phylogenetic analysis of 16S rRNA gene sequences from the genus *Acidithiobacillus*. The base tree is a maximum likelihood phylogram created with nearly full-length sequences, with shorter amplicon sequence (gray underline) placed using the EPA algorithm. Representative sequences from 16S rRNA gene clones are in bold black. Bootstrap values >50% are provided for each node.

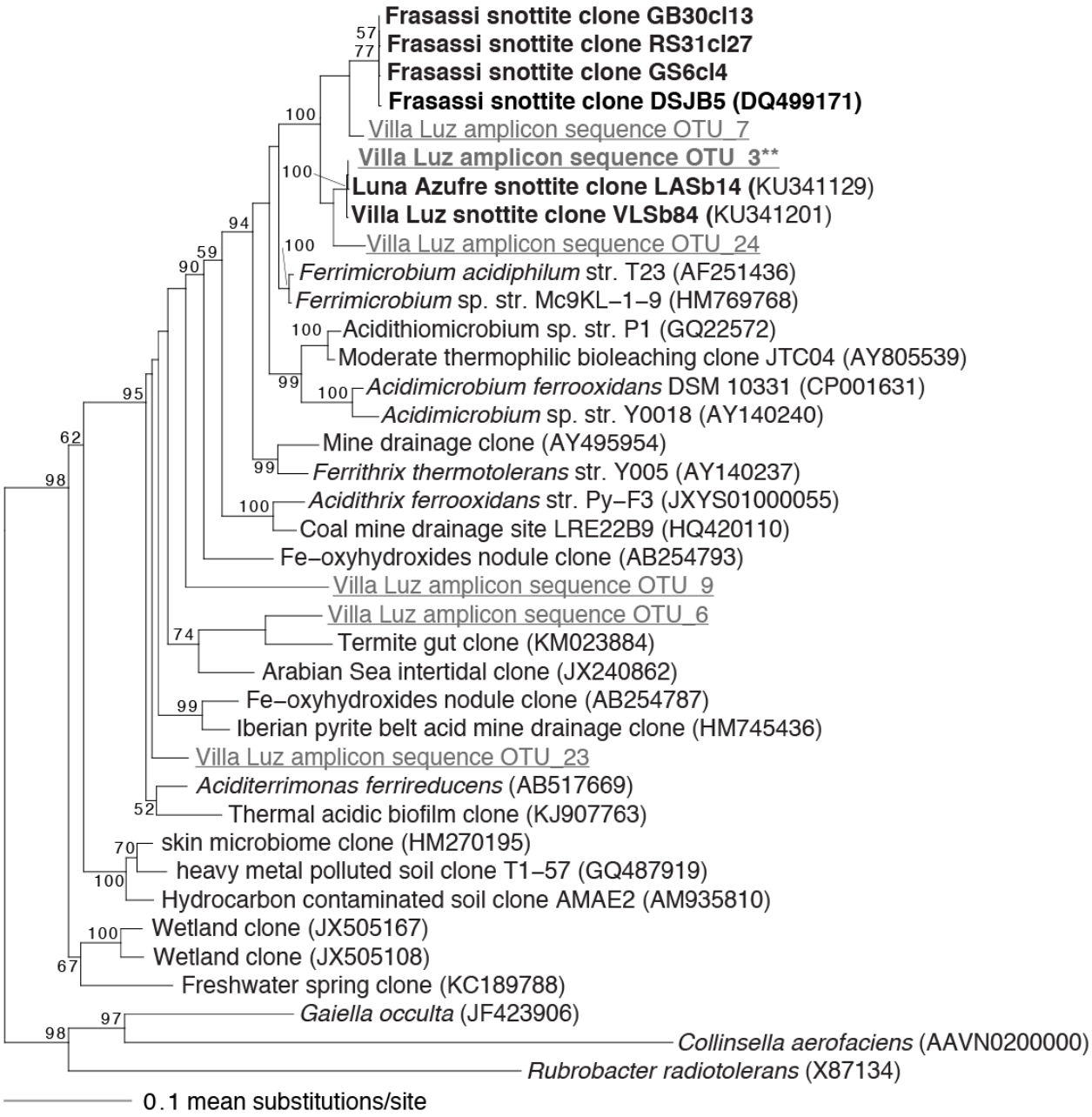

Figure 3. Phylogenetic analysis of 16S rRNA gene sequences from the *Acidimicrobiaceae* family. The base tree is a maximum likelihood phylogram created with nearly full-length sequences, with shorter amplicon sequences (gray underline) placed after the fact using the EPA algorithm. Representative sequences from 16S rRNA gene clones are in bold black. The amplicon sequence in bold gray indicated by two asterisks (**) represents that most abundant *Acidimicrobiaceae* OTU (24.8 and 13.7% of the VL13-1 and VL13-2 libraries, respectively); other *Acidimicrobiaceae* OTUs are <2% relative abundance (Table S2). Bootstrap values >50% are provided for each node.

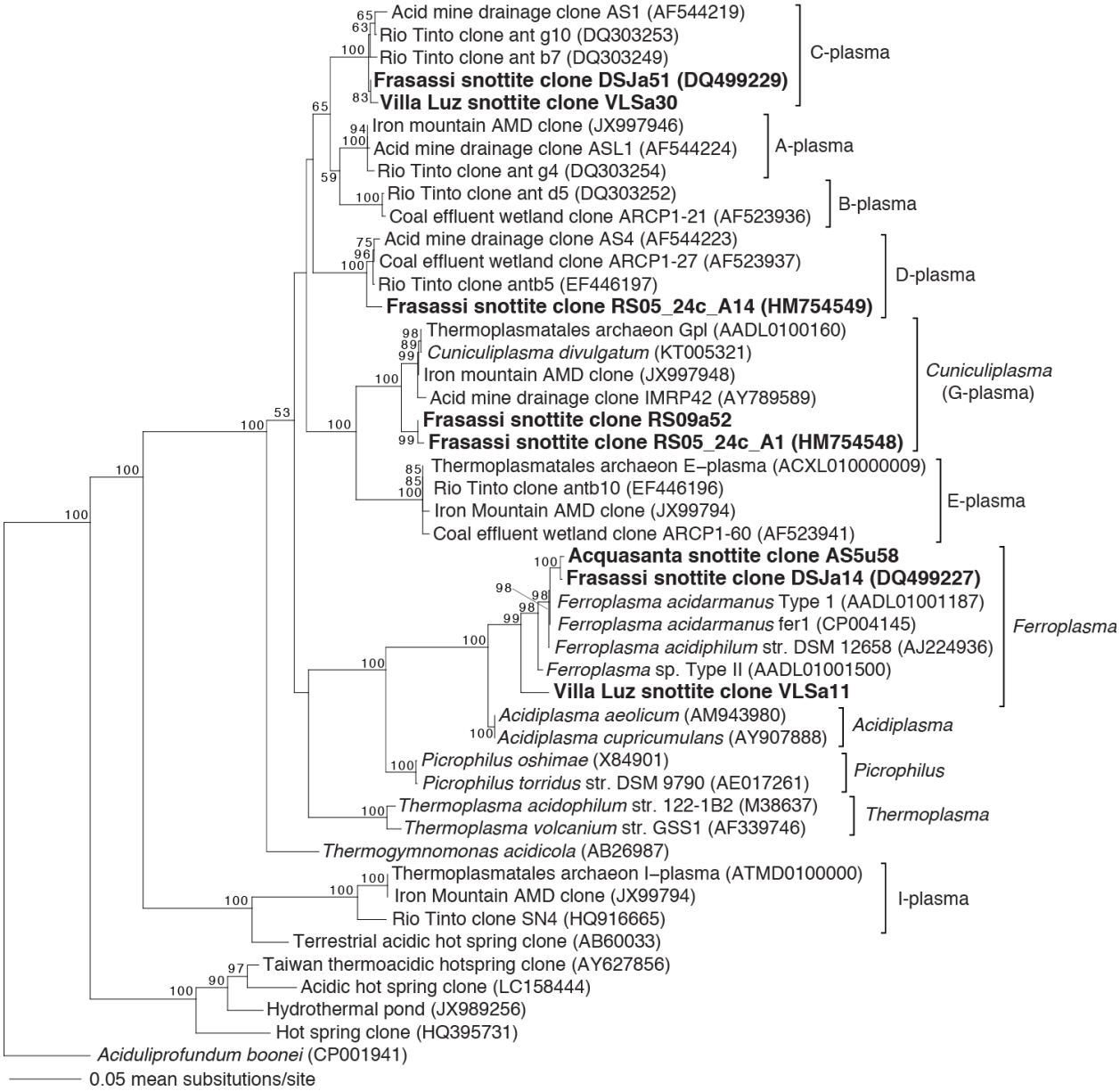

Figure 4. Neighbor joining phylogram of 16S rRNA gene sequences from the Archaeal family *Thermoplasmatales*. Snottite clones are in bold black. Bootstrap values >50% are provided for each node. Group names for the 'alphabet plasmas' are after Baker and Banfield (3).

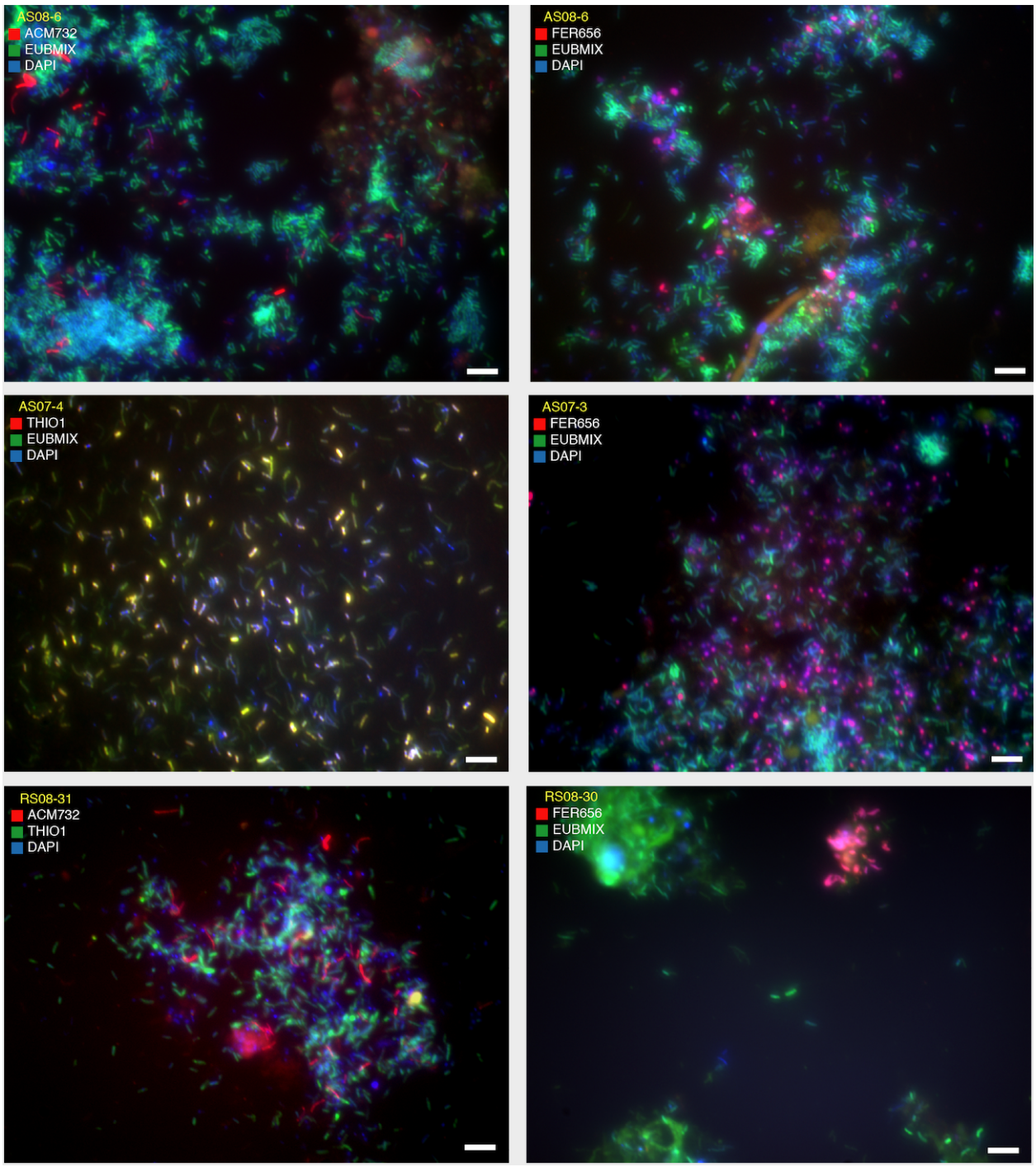

Figure 5. Representative fluorescence *in situ* hybridization (FISH) photomicrographs. The sample name and probes used for each photo are provided in the legend. White scale bars are 5 µM. Abundant DAPI-labeled cells in AS08-6 are nearly all archaea, as are abundant DAPI-labeled cells in RS08-31.

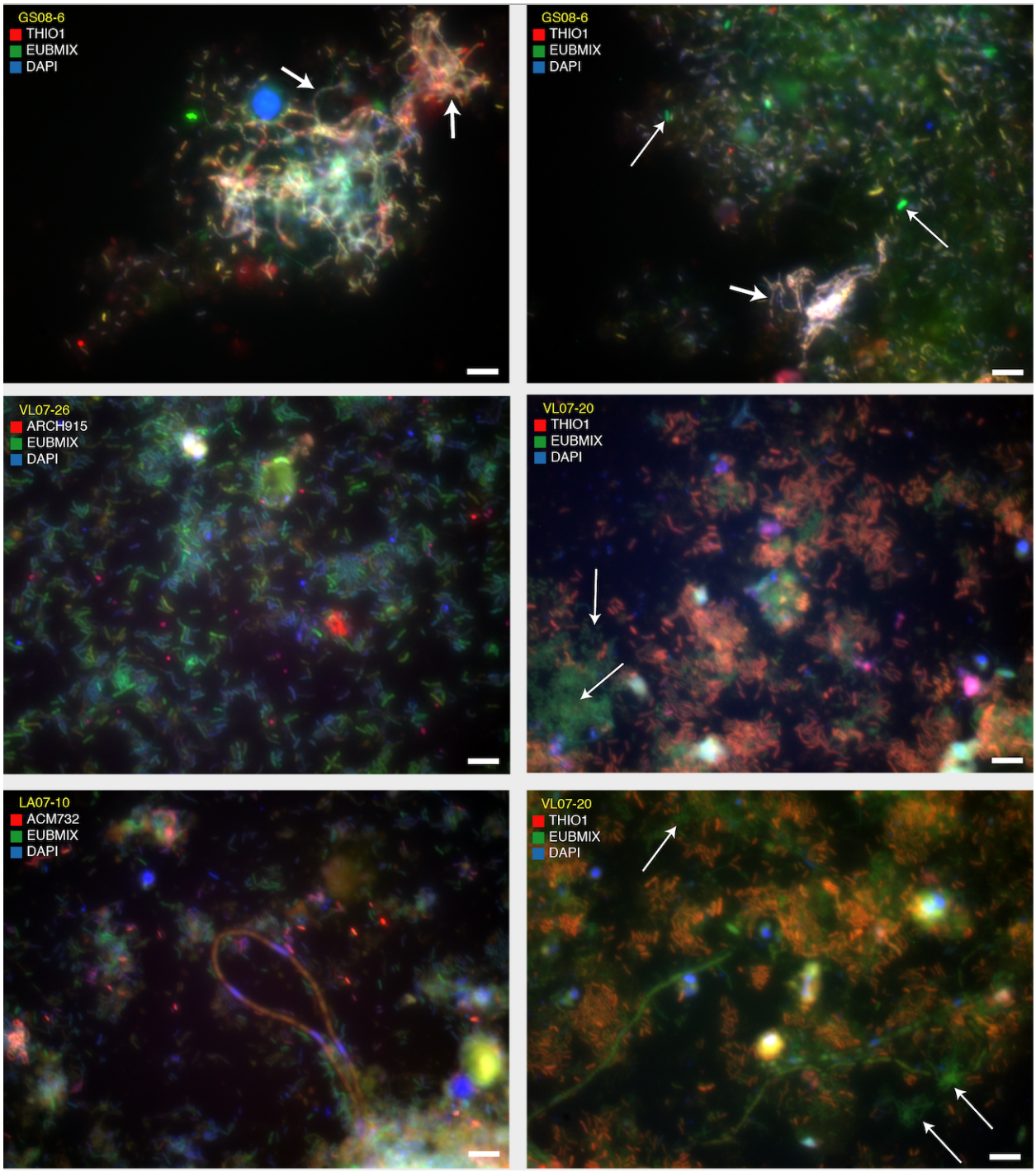

Figure 6. Representative fluorescence *in situ* hybridization (FISH) photomicrographs. The sample name and probes used for each photo are provided in the legend. White scale bars are 5 µM. Note the abundant non-THIO1 labeled cells (white arrows) in VL07-20. Fat arrow in GS08-6 indicates THIO1-labeled cells linked into filaments, and thin arrows indicate non-THIO1-labeled bacterial cells.

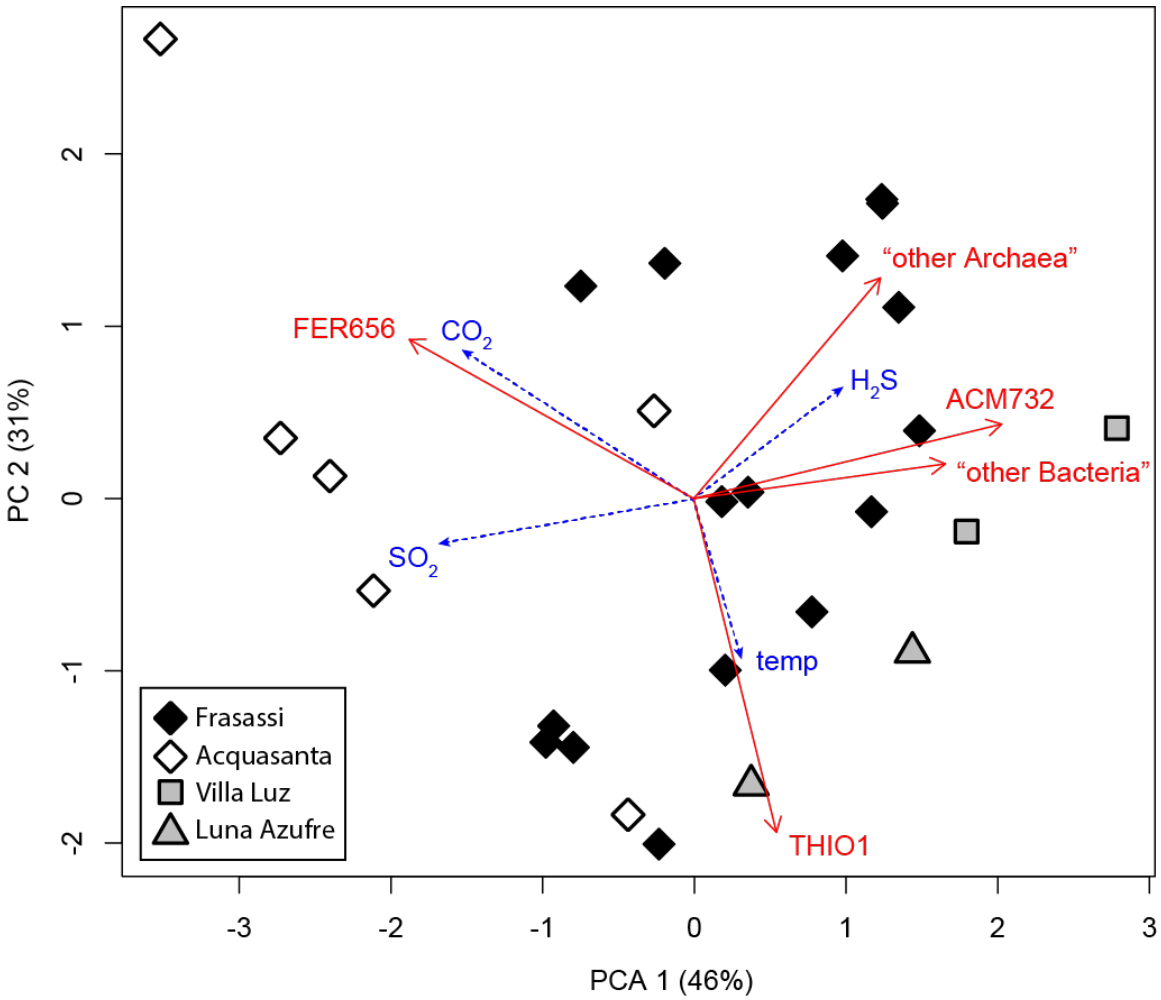

Figure 7. Principal component analysis (PCA) of FISH cell counts (Table 3). Sample scores are coded by cave (communities from different caves are significantly different by ANOSIM). Fitted vectors of environmental variables are included as an overlay; CO_2 and SO_2 gas concentrations are statistically significantly correlated with the ordination axes ($p < 0.05$), H_2S gas and temperature are not ($p > 0.1$).

Table 1. Summary of snottite samples and associated environmental data.

Sample	Cave	Site	Wall minerology	Temp (°C)	[$H_2S(g)$] (ppmv)	[$CO_2(g)$]	[$SO_2(g)$] (ppmv)
GS07-35	Frasassi	GS2	gypsum	13-14	bd[b]	700 ppm	bd
GS08-6	Frasassi	GS1	gypsum	13-14	0.8 (0.6-1)	1400 ppm	nm
RS05-2	Frasassi	RS2	gypsum and S^o	13-14	nm	nm	nm
RS05-24	Frasassi	RS2	gypsum and S^o	13-14	16 (8-24)	>3000 ppm	nm
RS05-1	Frasassi	RS1	gypsum	13-14	nm	nm	nm
RS05-20	Frasassi	RS1	gypsum	13-14	3.4	>3000 ppm	nm
RS07-21	Frasassi	RS3	gypsum and S^o	13-14	21 (17-25)	0.2 %	bd
RS07-22	Frasassi	RS2	gypsum and S^o	13-14	14.5 (11-18)	0.2 %	bd
RS08-30	Frasassi	RS2	gypsum and S^o	13-14	11.5 (11-12)	2500 ppm	nm
RS08-31	Frasassi	RS3	gypsum and S^o	13-14	14 (12-16)	3000 ppm	nm
RS09-1	Frasassi	RS2	gypsum and S^o	13-14	8 (5-11)	0.25%	bd
RS09-10	Frasassi	RS2	gypsum and S^o	13-14	10.5 (8-13)	0.3%	bd
RS09-11	Frasassi	RS3	gypsum and S^o	13-14	21 (20-22)	0.4%	bd
PC05-1	Frasassi	PC1	gypsum	13-14	nm	nm	nm
PC05-13	Frasassi	PC1	gypsum	13-14	1.1 (0.2-2)	1200 ppm	nm
GB08-30	Frasassi	GB1	gypsum	13-14	4 (3-5)	1400 ppm	nm
AS05-5	Acquasanta	AS3	gypsum	19-21	nm	nm	nm
AS07-1	Acquasanta	AS3	gypsum	19-21	2 (1-3)	0.4 %	4.5 (2-36)
AS07-3	Acquasanta	AS4	sediment[a]	19-21	2.9 (1-4.8)	0.4 %	5 (3-7)
AS07-4	Acquasanta	AS2	sediment[a]	25-35	24 (18-30)	0.6 %	11.5 (10-13)
AS08-5	Acquasanta	AS3	gypsum	19-21	bd[c]	0.45 %	5 (4-6)
AS08-6	Acquasanta	AS2	sediment[a]	25-35	3	0.55 %	4.5 (4-5)
VL07-20	Villa Luz	VL1	gypsum	25-28	12 (6-18)	1200 ppm	bd
VL07-26	Villa Luz	VL2	gypsum	25-28	3.5 (3-4)	1100 ppm	bd
LA07-10	Luna Azufre	LA1	gypsum	28-33	3	1250 ppm	bd
LA07-13	Luna Azufre	LA2	gypsum	28-33	2	1250 ppm	bd

[a]Sediment is a mixture of gypsum, clay, and chert particles
[b]Measured by ENMET MX2100, lower detection limit (ldl) 1 ppm.
[c]Measured by Draeger tube, ldl 0.2 ppm.

723 **Table 2.** FISH cell counts for snottite communities

Sample	FISH probe abundance				
	EUBMIX	THIO1	ACM732	ARCH915	FER656
GS07-35	88.9 (±4.4)	78.1 (±2.6)	8.2 (±1.5)	9.8 (±4.7)	0.6 (±1.1)
GS08-6	95.6 (±2.7)	91.5 (±2.0)	1.7 (±1.6)	1.8 (±0.9)	nd
RS05-2	79.8 (±8.4)	68.7 (±7.9)	8.6 (±5.8)	17.7 (±16.5)	11.4 (±7.3)
RS05-24	75.3 (±8.2)	68.8 (±8.0)	5.8 (±2.0)	16.5 (±8.60	nd
RS05-1	61.9 (±4.8)	56.0 (±5.6)	3.5 (±1.7)	38.6 (±13.9)	22.5 (±4.6)
RS05-20	61.0 (±7.0)	54.4 (±7.1)	5.0 (±1.5)	34.9 (±10.2)	14.3 (±9.1)
RS07-21	66.3 (±7.7)	53.8 (±5.0)	10.7 (±3.4)	31.9 (±6.8)	4.4 (±1.4)
RS07-22	66.2 (±6.6)	57.5 (±8.9)	6.7 (±4.8)	31.2 (±6.6)	1.3 (±1.7)
RS08-30	85.5 (±7.8)	83.0 (±7.5)	0.3 (±0.4)	10.0 (±8.0)	9.3 (±4.8)
RS08-31	70.6 (±7.2)	57.2 (±8.9)	7.5 (±5.8)	22.6 (±5.8)	1.4 (±1.0)
RS09-1	63.9 (±6.3)	51.3 (±6.4)	7.0 (±2.8)	30.9 (±5.6)	4.5 (±1.6)
RS09-10	87.5 (±5.1)	81.3 (±6.6)	4.6 (±4.60	9.4 (±4.2)	nd
RS09-11	82.7 (±7.5)	68.5 (±11.5)	9.5 (±5.4)	18.6 (±4.4)	0.5 (±0.6)
PC05-1	76.3 (±7.0)	75.9 (±7.3)	nd	1.76 (±4.7)	nd
PC05-13	79.3 (±8.6)	77.8 (±8.8)	0.7 (±1.8)	0.5 (±0.8)	nd
GB08-30	86.9 (±4.7)	74.4 (±8.5)	11.4 (±6.4)	15.6 (±3.8)	0.4 (±4.8)
AS05-5	22.7 (±12.6)	22.4 (±12.0)	nd	49.0 (±11.6)	49.0 (±11.6)
AS07-1	66.9 (±11.4)	66.9 (±11.4)	nd	30.2 (±11.0)	30.2 (±11.0)
AS07-3	61.6 (±3.7)	61.6 (±3.7)	nd	38.4 (±3.7)	38.4 (±3.7)
AS07-4	91.3 (±5.9)	88.7 (±4.0)	1.5 (±1.0)	3.1 (±1.8)	nd
AS08-5	72.2 (±4.3)	72.2 (±4.3)	nd	25.8 (±4.7)	25.8 (±4.7)
AS08-6	69.4 (±4.7)	65.8 (±4.1)	2.4 (±0.8)	29.2 (±5.3)	8.7 (±1.7)
VL07-20	87.2 (±7.2)	61.9 (±7.3)	10.3 (±5.0)	10.8 (±5.1)	2.0 (±2.8)
VL07-26	90.3 (±2.7)	70.6 (±3.3)	13.7 (±2.8)	5.4 (±1.4)	0.5 (±0.5)
LA07-10	96.1 (±2.1)	79.5 (±6.3)	9.7 (±1.9)	3.1 (±3.5)	nd
LA07-13	96.8 (±0.9)	87.7 (±2.7)	6.4 (±2.2)	0.9 (±0.7)	nd

Numbers for each FISH probe represent percentage of total DAPI stained cells. Numbers in parentheses represent ± one standard deviation. nd = none detected. FISH probe specificity (Table 3) is as follows: EUMBIX = Bacteria, THIO1 = *Acidithiobacillus,* AMC732 = *Acidimicrobium,* ARCH915 = Archaea, FER656 = *Ferroplasma.*

724
725
726
727
728
729 **Table 3.** FISH probes used.

Probe name	Sequence (5'-3')	% formamide	specificity	label	source
EUB338	GCT GCC TCC CGT AGG AGT	0-50%	most Bacteria	FITC	Amann et al., 1990
EUB338-II	GCA GCC ACC CGT AGG TGT	0-50%	Planctomycetales	FITC	Daims et al., 1999
EUB338-III	GCT GCC ACC CGT AGG TGT	0-50%	Verrucomicrobiales	FITC	Daims et al., 1999
ARCH915	GTG CTC CCC CGC CAA TTC CT	20%	most Archaea	CY3, FITC	Stahl and Amann, 1991
THIO1	GCG CTT TCT GGG GTC TGC	35%	Acidithiobacillus spp.	CY3, CY5	Gonzalez-Toril et al. 2003
ACM732	GTA CCG GCC CAG ATC GCT G	35%	Acidimicrobium spp.	CY3	Bond et al. 2000
FER656	CGT TTA ACC TCA CCC GAT C	25%	Ferroplasma spp.	CY3	Bond et al. 2000

EUBMIX is a mixture of EUB338, EUB338-II, and EUB338-III

730