OPEN

The ISME Journal (2016) 10, 2879–2891

www.nature.com/ismej

ORIGINAL ARTICLE

Biogeography of sulfur-oxidizing *Acidithiobacillus* populations in extremely acidic cave biofilms

Daniel S Jones[1,2], Irene Schaperdoth[1] and Jennifer L Macalady[1]
[1]*Department of Geosciences, Penn State University, University Park, PA 16802, USA and* [2]*BioTechnology Institute and Department of Earth Sciences, University of Minnesota, Minneapolis, MN 55455, USA*

Extremely acidic (pH 0–1.5) *Acidithiobacillus*-dominated biofilms known as snottites are found in sulfide-rich caves around the world. Given the extreme geochemistry and subsurface location of the biofilms, we hypothesized that snottite *Acidithiobacillus* populations would be genetically isolated. We therefore investigated biogeographic relationships among snottite *Acidithiobacillus* spp. separated by geographic distances ranging from meters to 1000s of kilometers. We determined genetic relationships among the populations using techniques with three levels of resolution: (i) 16S rRNA gene sequencing, (ii) 16S–23S intergenic transcribed spacer (ITS) region sequencing and (iii) multi-locus sequencing typing (MLST). We also used metagenomics to compare functional gene characteristics of select populations. Based on 16S rRNA genes, snottites in Italy and Mexico are dominated by different sulfur-oxidizing *Acidithiobacillus* spp. Based on ITS sequences, *Acidithiobacillus thiooxidans* strains from different cave systems in Italy are genetically distinct. Based on MLST of isolates from Italy, genetic distance is positively correlated with geographic distance both among and within caves. However, metagenomics revealed that *At. thiooxidans* populations from different cave systems in Italy have different sulfur oxidation pathways and potentially other significant differences in metabolic capabilities. In light of those genomic differences, we argue that the observed correlation between genetic and geographic distance among snottite *Acidithiobacillus* populations is partially explained by an evolutionary model in which separate cave systems were stochastically colonized by different ancestral surface populations, which then continued to diverge and adapt *in situ*.
The ISME Journal (2016) **10,** 2879–2891; doi:10.1038/ismej.2016.74; published online 17 May 2016

Introduction

Biogeography is the study of the spatial distribution of life forms across landscapes. For macroorganisms, global biogeographic patterns are strongly influenced by geographic barriers to dispersal. Dispersal barriers limit gene flow between regions, and local evolutionary processes such as drift or adaptation cause endemism in geographically restricted areas. However, because single-celled organisms are very small, highly abundant, metabolically plastic, and disperse easily, whether and how microorganisms are affected by geography has been the subject of debate (Finlay, 2002). Traditionally, microbial ecologists have assumed that the effects of dispersal limitations are minimal, and have considered the biogeographical distribution of microorganisms largely a result of environmental selection (Baas Becking, 1934; de Wit and Bouvier, 2006). However, improvements in molecular genetic methods and lower DNA sequencing costs have provided the necessary tools to detect genetic divergence among microbial populations with much greater resolution. As a result, there is now evidence that physical isolation is important in shaping biogeographical distributions and in facilitating speciation in microorganisms as well as macroorganisms (Papke and Ward, 2004; Martiny *et al.*, 2006; Whitaker, 2006; Hanson *et al.*, 2012).

In recent years, population genetic studies using high-resolution techniques have suggested that microorganisms can exhibit a high degree of endemism (Cho and Tiedje, 2000; Papke *et al.*, 2003). In particular, pioneering research on hot spring microbiota by Whitaker *et al.* (2003) showed that the genetic distance among thermophilic archaeal populations was correlated with the geographic distance that separated them. Evidence for endemism as well as similar distance–decay patterns have been identified for other microbial taxa, including both extremophilic and non-extremophilic populations (Escobar-Páramo *et al.*, 2005; Vos and Velicer, 2008; Hahn *et al.*, 2015; Raymond and Alsop, 2015). These findings suggest that allopatric speciation is

Correspondence: DS Jones, BioTechnology Institute and Department of Earth Sciences, University of Minnesota, Minneapolis, MN 55455, USA.
E-mail: dsjones@umn.edu
or JL Macalady, Department of Geosciences, Penn State University, University Park, PA 16802, USA.
E-mail: jlm80@psu.edu
Received 3 January 2016; revised 17 March 2016; accepted 24 March 2016; published online 17 May 2016

an important and underappreciated force in microbial evolution (Whitaker, 2006). However, not all microorganisms seem to be geographically restricted over large spatial scales (e.g., Roberts and Cohan, 1995; Sikorski and Nevo, 2005; van Gremberghe *et al*., 2011; de Rezende *et al*., 2013; Ryšánek *et al*., 2014), and evidence for long distance microbial dispersal is present even in studies that identified geographic clustering using high-resolution genetic analyses. For example, Vos and Velicer (2008) found a robust pattern of increasing genetic divergence among populations of the spore forming soil bacterium *Myxococcus xanthus*, but also found that identical genotypes were present at distant sites and that some globally separated populations were not significantly genetically differentiated. Foti *et al.* (2006) found that, despite geographic clustering, identical genotypes of *Thioalkalavibrio* were present in soda lakes on different continents.

Here we report on the biogeography of extremely acidic biofilms known as snottites from hydrogen sulfide (H_2S)-rich caves. Sulfidic caves are fed by anoxic springs that degas H_2S(g) into the cave atmosphere and provide energy for sulfur-oxidizing microorganisms on the cave walls and ceilings. In most caves, snottites are observed where H_2S(g) concentrations in the cave air are between 0.2 and 25 parts-per-million by volume (ppmv), and rarely occur where concentrations are outside of this range (Figure 1) (Hose *et al*., 2000; Macalady *et al*., 2007). Previous research has shown that snottites are inhabited by low-diversity communities containing a few species of acidophilic bacteria and archaea (Hose *et al*., 2000; Vlasceanu *et al*., 2000; Macalady *et al*., 2007; Jones *et al*., 2012, 2014). The most abundant microorganism is *Acidithiobacillus thiooxidans*, which is a sulfide-oxidizing autotroph and likely responsible for forming the snottite biofilms (Jones *et al*., 2012).

Snottites therefore present an opportunity to test whether and how geographic barriers affect microbial biogeography. Because sulfidic caves are found around the world, and because snottites are found in multiple chambers within each cave system, they can be sampled at spatial scales ranging from meters to 1000 s of kilometers. Because snottites are extremely acidic (pH 0–1.5), whereas the surrounding areas are circumneutral, snottite microorganisms seem unlikely to survive transit between caves. Geochemical isolation constitutes a potential barrier to microbial dispersal in addition to the physical barrier already imposed by their location in the terrestrial subsurface, which presumably has reduced microbial transport by wind and large animal vectors compared with surface environments.

We therefore hypothesized that this physical and geochemical isolation would result in different caves harboring genetically distinct snottite populations, and that genetic distance among the *Acidithiobacillus* spp. would be correlated with the geographic distance separating cave locations. We used a population genetics approach to compare snottite *Acidithiobacillus* populations at different levels of genetic resolution, and by combining results from neutrally evolving markers with functional gene information from metagenomics, we propose a model in which isolation and stochastic colonization sculpt the biogeographic relationships among snottite *Acidithiobacillus* spp.

Materials and methods

Sample collection

Snottite biofilm samples were collected from four sulfidic caves: Cueva de Villa Luz and Cueva Luna Azufre in Tabasco, Mexico (Hose and Pisarowicz, 1999; Hose *et al*., 2000), and le Grotte di Frasassi (Galdenzi and Maruoka, 2003) and Grotta Nuova di Rio Garrafo (hereafter Acquasanta, following Jones *et al*., 2010) in the Marche region, Italy (Supplementary Table S1 and Supplementary Figure S1). Samples are summarized in Table 1, and location information is provided in

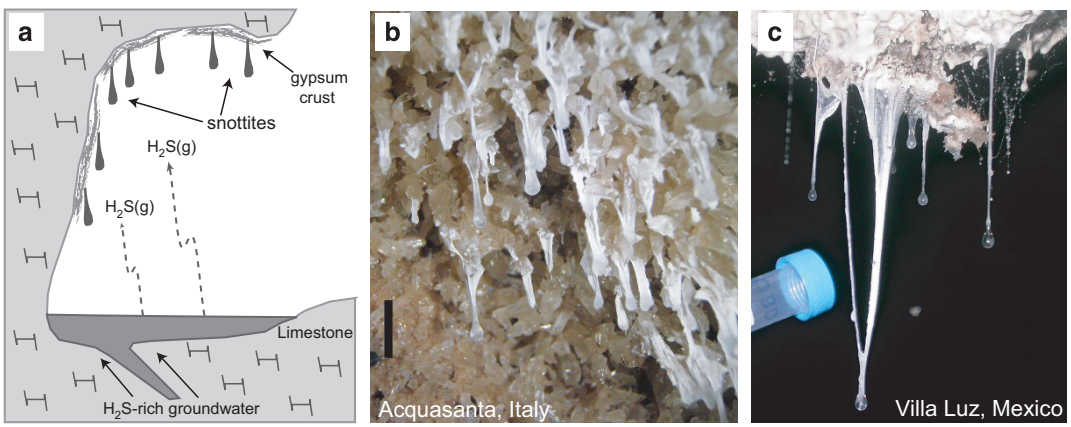

Figure 1 (**a**) Schematic depicting the snottite habitat in sulfidic caves. Snottites form in close proximity to H_2S-degassing cave streams, and typically occur where cave air H_2S(g) concentrations are between 0.2 and 25 ppm. (**b**, **c**) Representative photographs of snottites sampled in this study, from (**b**) Acquasanta, Italy and (**c**) Villa Luz, Mexico. Black scale bar in (**b**) is 1 cm.

Table 1 Summary of *Acidithiobacillus* isolates and environmental sequences

Cave system	*Sample*	*Year*	*Site*	*No. of isolates*	*No. of environmental ITS sequences*	*No. of environmental 16S rRNA sequences*
Frasassi, Italy	GS6	2008	GS1	6	24	24
	GS3	2007	GS2	2		
	RS2	2005	RS2	1		68
	RS20	2007	RS1	2		
	RS30	2008	RS2	7	15	15
	RS31	2008	RS3	3	8	8
	RS9[a]	2009	RS2		37	37
	GB30	2008	GB1	6	28	28
	GB31	2008	GB2	2		
	PC1	2005	PC1			71
	PC30	2007	PC1	3		
Acquasanta, Italy	AS1	2007	AS2	3		
	AS3	2007	AS4	3	29	29
	AS4	2007	AS3	2		
	AS5[a]	2008	AS4			
Villa Luz, Mexico	VL20	2007	VL1			17
Luna Azufre, Mexico	LA10	2007	LA1	4		28
	LA13	2007	LA2	2		

[a]Samples for metagenomic sequencing.

Supplementary Figure S1. Permits for access to Villa Luz were granted by C Rogers Morales Mendez and L Felino Arevalo Gallegos. Sampling and personal safety equipment was thoroughly washed between cave trips, and separate gear was used for Italian and Mexican cave expeditions.

Sampling locations were selected to encompass multiple spatial scales as well as a range of H_2S(g) concentrations. Cave air concentrations of H_2S(g), CO_2(g) and SO_2(g) were measured using Dräger tubes (Dräger Safety, Pittsburg, PA USA) and/or an ENMET MX2100 portable gas detector (ENMET Corp., Ann Arbor, MI USA). Gas concentrations ranged from below detection to 30 ppmv for H_2S(g), from 700 to 6000 ppmv for CO_2(g) and from below detection to 36 ppmv for SO_2(g) (Supplementary Table S2). At the time of collection, the pH values of 5–10 snottites in the area were measured with pH paper (range 0–2.5), and all values were between 0 and 1.5. (pH paper is necessary because the small size and viscous texture of the biofilms prohibit measurement by pH electrodes.) Biofilm aliquots for DNA extraction were preserved in RNAlater (Thermo Fisher Scientific, Waltham, MA, USA) and stored at −20 °C. Aliquots for enrichment culturing were stored at 4°C until inoculation.

Enrichment and isolation of Acidithiobacillus *strains*
Acidithiobacillus strains were enriched in liquid media with either thiosulfate or elemental sulfur as the electron donor (ATCC 1353 *Thiobacillus albertis* medium or 125 *Thiobacillus* medium, http://www.atcc.com) (Harrison, 1982). Isolation was achieved by plating enrichments onto solid thiosulfate media (3% agar, 2 × 1353 thiosulfate media, pH 4). Colonies were re-plated, picked and re-grown in liquid 1353 medium prior to DNA extraction. Initially, successful isolation was assessed by uniform colony morphology on solid media and microscopically by uniform cell morphology. All strains were passaged and plated the same number of times, except for strain RS2a, which was isolated in 2005 and passaged weekly for 1 year. Strains from samples RS2, GB30 and GS6 all grew similarly on liquid elemental sulfur media at initial pH values from 5 down to 0.2.

DNA extraction, 16S rRNA gene and ITS sequencing
DNA was extracted from 46 isolates (Table 1) using the Mo Bio UltraClean Microbial DNA Isolation Kit, according to the manufacturers instructions (Mo Bio Laboratories, Inc., Carlsbad, CA, USA). Partial 16S rRNA genes and complete ITS sequences were amplified from each strain via PCR using forward primer 27f (AGA GTT TGA TCC TGG CTC AG) and reverse primer Sag2 (TGG CTG GGT TGC CCC ATT C), modified from Sagredo *et al.* (1992). PCR was performed with 5 min initial denaturation at 95 °C, followed by 25 cycles of 60 s denaturation at 95 °C, 30 s annealing at 50 °C, 2 min elongation at 72 °C, and with 12 min final elongation at 72 °C. Products were purified with the QIAquick PCR purification kit (Qiagen Inc., Valencia, CA, USA) and directly sequenced using primers 27f, Sag2 and 1392r (ACG GGC GGT GTG TRC) at the Penn State Core Genomics Facility with an ABI Hitachi 3730XL DNA analyzer and BigDye fluorescent terminator chemistry v3.1 (Applied Biosystems, Foster City, CA, USA).

For environmental biofilm samples, DNA extraction, amplification and cloning of environmental 16S

rRNA sequences from samples RS2 and PC1 was described in Macalady *et al.* (2007). For all other samples, DNA was extracted from snottite biofilms exactly as in Jones *et al.* (2012). *Acidithiobacillus* 16S rRNA gene sequences were cloned from Villa Luz and Luna Azufre snottites using bacterial-specific primers 27f and 1492r (GGT TAC CTT GTT ACG ACT T) following the amplification, cloning and colony PCR procedures of Macalady *et al.* (2008). The 16S rRNA gene+ITS sequences were amplified from Frasassi and Acquasanta snottites using primers 27f and Sag2 as described above for isolates. PCR products were purified using the Qiaex II gel extraction kit (Qiagen Inc.). Cloning was performed by ligating the purified product into the pCR4-TOPO plasmid and used to transform either One Shot Mach1 T1 or TOP10 chemically competent *Escherichia coli* (Invitrogen Corp., Carlsbad, CA USA). Inserts were extracted by colony PCR as in Macalady *et al.* (2008) with M13 primers (Invitrogen Corp.) and sequenced at the Penn State Core Genomics Facility as described above.

Multi-locus sequence typing

MLST was performed by amplifying and sequencing six loci (*recA*, *rpoB*, *atpD*, *ileS*, *rpl1p* and *rps2p*) from each isolate (Table S2). Each of these loci are from housekeeping genes typically represented by a single copy per genome. Primers for each locus were designed based on nucleotide sequences of full-length genes from *Acidithiobacillus ferrooxidans* strains ATCC 23270 and ATCC 53993 and using assembled *At. thiooxidans* sequences from preliminary metagenomic sequencing (below, and Jones *et al.*, 2012). Primers for *recA* and *atpD* were modified from Amouric *et al.* (2011). Loci were amplified using primers and PCR conditions provided in Table S2. PCR products were purified with the Omega E.Z.N.A. Cycle-Pure Kit (Omega Bio-Tek, Inc., Norcross, GA, USA) and sequenced at the Penn State Core Genomics Facility.

Phylogenetic and statistical analyses

Raw DNA sequences were assembled and manually checked for quality with CodonCode Aligner v.2.0 (CodonCode Corporation, Dedham, MA, USA). rRNA gene sequences were aligned using the NAST aligner at Greengenes (DeSantis *et al.*, 2006), chimera checked with Bellerophon 3 (Huber *et al.*, 2004) and imported and further aligned in ARB (Ludwig *et al.*, 2004). Sequences from the six loci for MLST were concatenated, and then aligned using ClustalX v.2.1 (Larkin *et al.*, 2007).

Phylogenetic analyses were performed in PAUP* v.4b10 (Swofford, 2000). Maximum likelihood analysis of the concatenated MLST data set was performed using the Tamura-Nei nucleotide substitution model and parameters selected by the Bayesian information criterion with jModelTest v.2.1.1 (Posada, 2008). Maximum likelihood analysis of 16S rRNA genes was also performed with the Tamura-Nei substitution model, selected by the Bayesian information criterion in jModelTest. Maximum parsimony bootstrap analyses (2000 replicates) were performed with 25 random addition replicates and tree-bisection reconnection branch swapping, and neighbor joining bootstrap analyses (2000 replicates) with Jukes-Cantor distance. Bayesian analysis was implemented in MrBayes v.3.0b4 (Huelsenbeck and Ronquist, 2001) with six substitution rate categories and gamma-distributed rate variation. Bayesian analysis was run for 500 000 generations, trees were saved every 100 generations and posterior probabilities calculated after discarding the first 20% of trees. Prior to analysis, 16S rRNA gene sequences and MLST sequences were clipped to equal length (1371 and 5777 characters, respectively). No positions were masked.

ITS sequences were aligned in a custom ARB database. Phylogenetic analyses of ITS regions were accomplished in three ways. (i) First, all positions with $>50\%$ gaps were masked, and sequences were analyzed by neighbor joining and maximum parsimony as described above (367 positions, 73 variable sites, 52 parsimony informative sites). Positions with $<50\%$ gaps were included to increase the number of variable sites for phylogenetic analysis. (ii) Second, a separate analysis was performed with all gapped positions masked from the alignment, and sequences were analyzed by both neighbor joining and maximum parsimony analyses. A strict consensus tree was constructed from equally parsimonious topologies (339 positions, 59 variable positions, 38 parsimony informative sites). (iii) A third analysis was performed in which gaps in the ITS alignment were treated as characters for parsimony analysis. Alignment gaps represent insertion or deletion (indel) events, and contain additional phylogenetic information that is not included in the analysis of nucleotide characters alone. Gaps were quantified by simple indel coding (Simmons and Ochoterena, 2000) using the software package GapCoder (Young and Healy, 2003) (Supplementary Figure S2 and Supplementary Table S3). Simple indel coding is a conservative approach in which indels with different 5′ and 3′ termini are considered separate characters, and smaller gaps that occur completely within larger gaps are treated as missing data for the sequences with the larger gaps. The presence–absence indel matrix was subjected to maximum parsimony analysis in PAUP* and a strict consensus tree was calculated. Prior to phylogenetic analysis, ITS clones with one nucleotide difference were grouped using the pre.cluster command in Mothur v.1.24 (Schloss *et al.*, 2009). The identity of the tRNA sequences in the ITS region was confirmed with tRNAscan-SE v.1.21 (Schattner *et al.*, 2005).

Mantel and partial Mantel tests were performed in R v.2.6.1 (R Core Development Team, 2007) with the Vegan package (Oksanen *et al.*, 2008). All Mantel tests were performed using Pearson's

product–moment correlation. For Mantel tests of genetic distance versus environmental distance among sites, geochemical variables were first relativized to the maximum value for each variable in all samples ('relativization by maximum', McCune and Grace, 2002), and environmental dissimilarity among samples was calculated with Euclidean distance. Partial mantel tests of genetic versus environmental distance were calculated while controlling for environmental dissimilarity based on Euclidean distance as described above.

MLST, ITS and rRNA gene sequences have been submitted to GenBank under the following accession numbers: KU249220-KU249459 (MLST); KU249460-KU249649 and KU341209-KU341239 (ITS and rRNA gene sequences from isolate and environmental samples); and DQ499162-DQ499330 and KU341124-KU341208 (other environmental rRNA gene sequences).

Metagenomic analysis

Two metagenomic data sets were generated from snottite biofilm samples AS08-5 and RS09-1 (hereafter AS5 and RS9) from the Acquasanta and Frasassi cave systems, respectively. DNA was extracted from the biofilm samples as described above, and metagenomic data sets were generated by submitting the environmental DNA extracts for pyrosequencing on a Roche GS 454 FLX with FLX Titanium chemistry (454 Life Sciences, Branford, CT, USA). Raw metagenomic data sets from AS5 and RS9 included 112.7 and 104.7 megabase pairs (Mbp) of sequence data, respectively. Preliminary analysis and binning of these data sets were reported in Jones *et al.* (2014). Briefly, AS5 and RS9 were assembled with the Newbler assembler (gsAssembler) version 2.6 (454 Life Sciences) using default parameters, except with minimum overlap identity 95 and minimum overlap length 60. Sample AS5 assembled into 4855 contigs longer than 500 bp, with an N50 value of 1453 bp and longest contig of 17 kilobase pairs (kbp) length. Sample RS9 assembled into 3424 contigs longer than 500 bp, with an N50 value of 3085 and longest contig 37.6 kbp. *At. thiooxidans* sequences were identified based on coverage, tetranucleotide frequency and sequence similarity to other *Acidithiobacillus* genome sequences as described in Jones *et al.* (2014).

Homologs of genes involved in the oxidation of reduced inorganic sulfur compounds were identified in the *At. thiooxidans* bins from Jones *et al.* (2014). To further compare genomic differences among the *At. thiooxidans* populations in AS5 and RS9, we compared the two metagenomes against the genomes of three publicly available *At. thiooxidans* strains. Predicted protein-coding genes from *At. thiooxidans* ATCC 19377 (Valdes *et al.* 2011), *At. thiooxidans* Licanantay (Travisany *et al.*, 2014) and *At. thiooxidans* A01 (Yin *et al.*, 2014a,b) were downloaded from the NCBI database (http://www.ncbi.nlm.nih.gov/bioproject/) under the accessions PRJNA36587 (ATCC 19377), PRJNA245008 (Licanantay) and PRJNA230432 (A01). In order to determine homologs that were shared between the metagenomes and each of these three genomes, we used TBLASTN (Altschul *et al.*, 1997) to compare predicted protein-coding genes from strains ATCC 19377, Licanantay and A01 against the AS5 and RS9 assemblies in turn.

Metagenomic data are available at the Sequence Read Archive (SRA, http://www.ncbi.nlm.nih.gov/sra) under accession numbers SRX225604 (RS09-1) and SRX225750 (AS08-5), and assembled contigs can be accessed at the integrated microbial genomes (IMG) system (http://img.jgi.doe.gov) under Taxon Object IDs 3300000824 and 3300000825.

Results

16S rRNA gene sequence analysis of Acidithiobacillus *populations*

We isolated 46 strains of *Acidithiobacillus* spp., 40 from Italy and 6 from Mexico (Table 1). We sequenced 16S rRNA genes from the isolates, and also cloned *Acidithiobacillus* 16S rRNA genes from 10 environmental samples (Table 1). *Acidithiobacillus* spp. from Italy are strains of *At. thiooxidans*, while strains from the Mexico group are a sister clade to *At. caldus*, hereafter referred to as *At.* group II (Figure 2). All *Acidithiobacillus* 16S rRNA gene sequences from Italy share >99% identity, as do all 16S rRNA genes from Mexico. Further phylogenetic relationships could not be resolved among *Acidithiobacillus* sequences from caves on the same continent based on 16S rRNA genes.

ITS sequence analysis

Because analysis of environmental and isolate 16S rRNA genes indicated that *Acidithiobacillus* from Italy and Mexico represent separate species (Figure 2), we only analyzed the 16S–23S ITS region from Italian acidithiobacilli, and did not attempt to isolate additional strains from Mexico. We successfully sequenced the ITS region from 25 of the *At. thiooxidans* isolates, and cloned the ITS region from six environmental samples (Table 1). However, for isolates from samples RS20, RS30 and RS31, as well as isolates GS6d, GS6f and GS3b from samples GS6 and GS3, the ITS region amplified but the result was a mixture of two sequences that could not be resolved. Therefore, we cloned six ITS sequences each from isolates RS30a and RS31a, and identified two distinct ITS sequences from each strain (dashed box in Figure 3 and Supplementary Figure S2b). The presence of different ITS sequences in these strains likely represents intragenomic divergence among multiple copies of the *rrn* operon (Sagredo *et al.*, 1992; Stewart and Cavanaugh, 2007).

Phylogenetic analysis of the ITS region was complicated by multiple gaps in the sequence alignment, presumably due to insertion or deletion events (indels) (Supplementary Figure S2). The

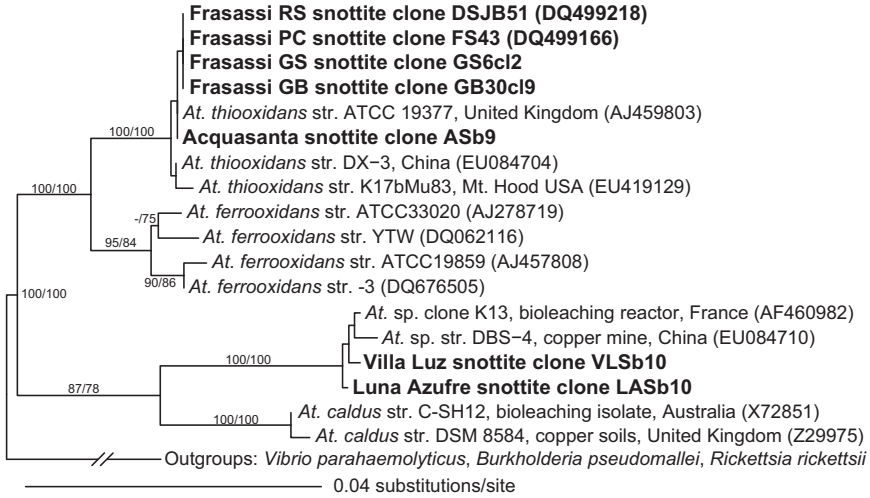

Figure 2 Maximum likelihood phylogram of 16S rRNA gene sequences from the genus *Acidithiobacillus*. Representative sequences from this study are shown in bold. Numbers indicate bootstrap support by neighbor joining and maximum parsimony, in that order (only values >75% shown).

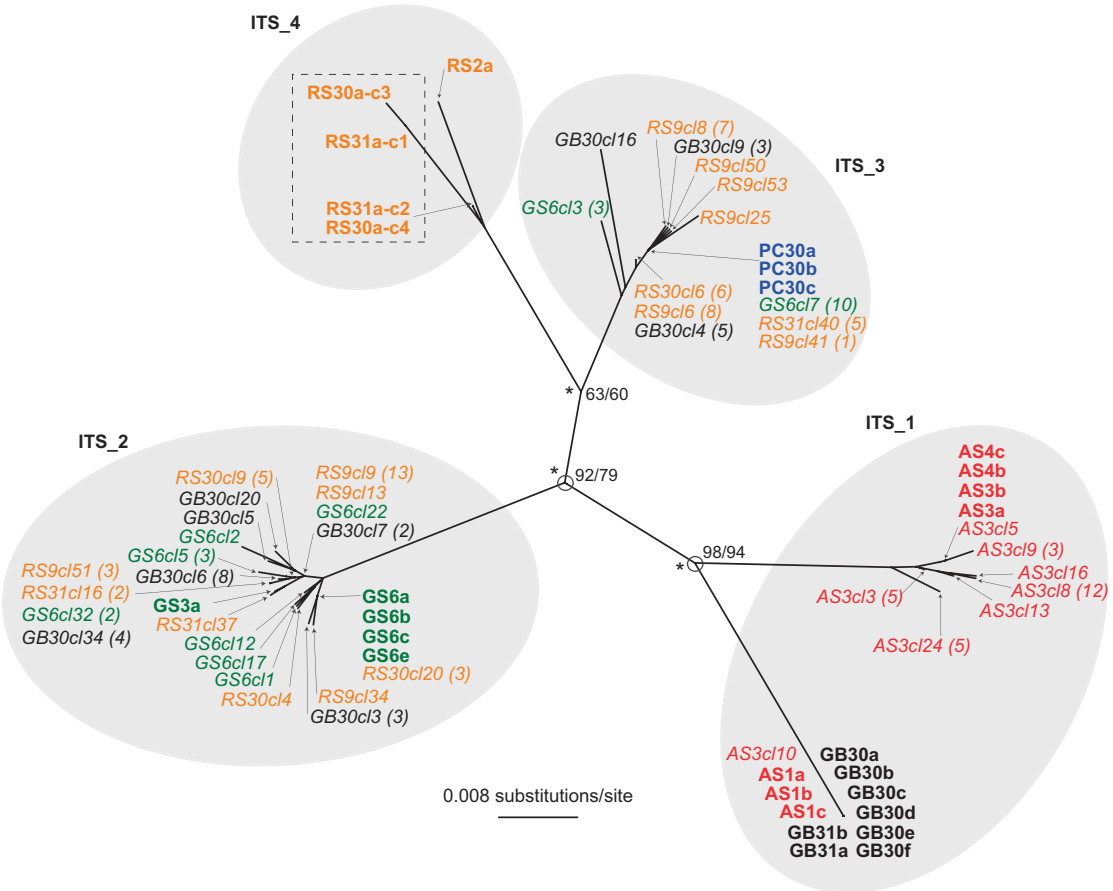

Figure 3 Phylogenetic analysis of ITS sequences from environmental and isolate *At. thiooxidans* strains. Sequence names are colored by cave location, names in bold indicate isolates and italicized names indicate environmental clones. The numbers of clones from the same sample represented by each sequence are given in parentheses. The base tree is a neighbor joining phylogram constructed after excluding alignment positions with >50% gaps, and numbers indicate maximum parsimony and neighbor joining bootstrap support for nodes connecting the four major clades. Stars indicate nodes compatible with both neighbor joining analysis and maximum parsimony consensus trees after excluding all gapped positions, and the solid circles indicate nodes compatible with maximum parsimony analysis of indels (Supplementary Figure S2). The dashed box in clade ITS_4 indicates ITS regions that were cloned from isolates RS30a and RS31a.

beginning and end of the ITS region is conserved, and the start includes both a $tRNA^{ala}$ and a $tRNA^{ile}$ while the end contains a box A antiterminator sequence (Venegas *et al.*, 1988; Sagredo *et al.*, 1992) (Supplementary Figure S2a). Most of the variable positions in the alignment are in a heavily gapped middle region. Phylogenetic analysis of isolate and environmental ITS sequences produced four major clades, designated ITS_1, ITS_2, ITS_3 and ITS_4 (Figure 3). All phylogenetic techniques were in agreement, except that maximum parsimony analysis of indels did not reproduce group ITS_4 (Figure 3 and Supplementary Figure S2c).

In the consensus phylogeny (Figure 3), clade ITS_1 includes environmental and isolate sequences from Acquasanta, and Frasassi isolates from samples GS30 and GB31. Clades ITS_2 and ITS_3 contain all Frasassi environmental ITS sequences as well as most Frasassi isolates. Clade ITS_4 contains Frasassi isolates from samples RS30, RS31 and RS2. Because ITS sequences from isolates RS30a, RS31a, GB30a–f and GB31a–b occur in different clades than environmental ITS sequences cloned from the same biofilm samples (Figure 3), those isolates do not represent the dominant snottite populations in their samples.

Environmental ITS sequences cloned from samples RS, GS and GB occur in clades ITS_2 and ITS_3 (Figure 3). As with isolates RS30a and RS31a, we attribute this to intragenomic divergence among multiple *rrn* operons in the dominant *At. thiooxidans* populations.

MLST of Acidithiobacillus *isolates*

For MLST, we sequenced six loci from all 40 Italian *At. thiooxidans* isolates. Total aligned length of the six concatenated loci is 5777 bp, which includes 627 variable sites. Phylogenetic analysis of the concatenated sequences produced three major clades (Figure 4): (i) clade MLST_1, which includes all isolates from Acquasanta, as well as all isolates from Frasassi samples GB30 and GB31; (ii) clade MLST_2, which contains Frasassi isolates from samples RS30, RS31 and RS2; and (iii) clade MLST_3, which contains all other isolates from Frasassi. *At. thiooxidans* isolates from Frasassi sites GB30 and GB31 are identical to isolates from Acquasanta site AS1 based on MLST.

However, the comparison between environmental and isolate ITS regions described above indicates

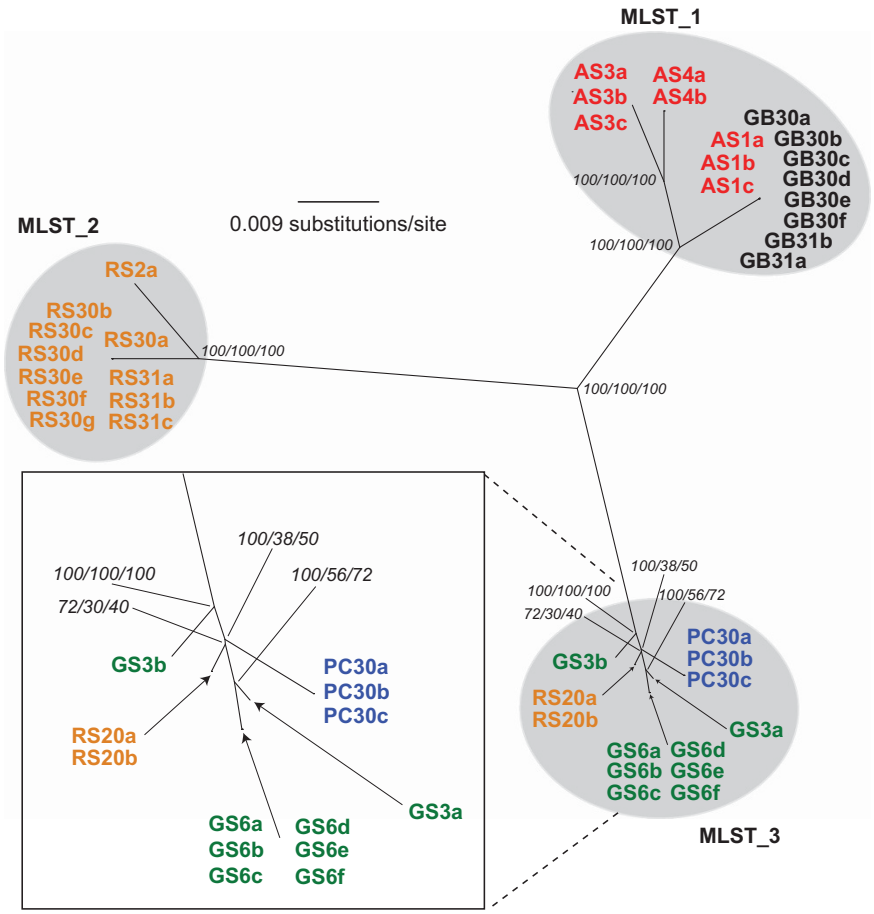

Figure 4 Maximum likelihood analyses of *At. thiooxidans* isolates based on MLST. Numbers at each node indicate posterior probabilities and bootstrap support from Bayesian, maximum parsimony and neighbor joining analyses, in that order. Sequence names are colored by cave location.

that isolates from samples RS30, RS31, RS2, GB30 and GB31 do not represent the dominant *At. thiooxidans* populations at those sites (Figure 3). Therefore, we conclude that those isolates represent 'weeds' that were selected for *in vitro*. Sequences in clade MLST_2 and isolates GB30a–f and GB31a–b were thus excluded from subsequent statistical analyses.

For the remaining 21 isolates, pairwise genetic distance is positively correlated with geographic distance (Mantel test, $r=0.967$, $P<0.001$) (Figure 5a). Among Frasassi isolates alone, pairwise genetic distance is also positively correlated with geographic distance (Mantel test, $r=0.51$, $P=0.005$) (Figure 5b). Genetic distance among isolates is also positively correlated with environmental conditions, based on concentrations of H_2S(g), CO_2(g) and SO_2(g) in the cave air (Supplementary Table S4) (Mantel tests: all isolates, $r=0.67$, $P<0.001$; Frasassi only, $r=0.39$, $P=0.02$). However, partial mantel tests indicate that genetic distance among isolates remains positively correlated with genetic distance after controlling for environmental variables (partial Mantel tests: all isolates, $r=0.94$, $P<0.001$; Frasassi only, $r=0.35$, $P=0.02$).

Metagenomic analysis of Acidithiobacillus *metabolic potential*

Because ITS sequence analysis and MLST indicated that *At. thiooxidans* populations from Acquasanta and Frasassi were genetically divergent, we generated metagenomic data sets from one biofilm sample from each cave system (samples RS9 and AS5) to compare functional genetic potential that could impact physiological attributes of the populations. The *At. thiooxidans* population in sample RS9 falls in clade MLST_3, while strains in sample AS5 fall in group MLST_1 (Figure 4). Coverage of *Acidithiobacillus* sequences in each metagenome is roughly 20×. AS5 also contains a second *Acidithiobacillus* population at roughly 10× coverage (see Jones *et al*., 2014), consistent with two populations identified by ITS cloning (clone AS3cl10 versus other AS3 clones in Figure 3). Additional rare *At. thiooxidans* populations are likely present in both samples at <5× coverage (Jones *et al*., 2014). Other taxa in the metagenomes include archaea (*Ferroplasma* and G-plasma) and *Acidimicrobium*-like organisms (Jones *et al*., 2014), consistent with previous work (Macalady *et al*., 2007; Jones *et al*., 2012).

Identification of putative sequences encoding sulfur oxidation enzymes revealed that the main *At. thiooxidans* populations in AS5 and RS9 have a difference in their sulfur oxidation pathways. The *At. thiooxidans* in AS5 has a homolog of sulfur oxygenase reductase (SOR), while the RS9 population does not (Figure 6). Based on coverage and phylogenetic analysis, the SOR in AS5 unambiguously belongs to the most abundant *At. thiooxidans* population in that sample (Jones *et al*., 2014). Coverage of *At. thiooxidans* in the AS5 and RS9 data sets is sufficient to ensure that the complete genetic complement of the dominant *At. thiooxidans* populations is present in the metagenomes, and this result is also consistent with an earlier report that *At. thiooxidans* from the site RS lack a SOR homolog (Jones *et al*., 2012). Accordingly, a SOR is present in two of the three *At. thiooxidans* isolates whose genomes have been sequenced to date: *At. thiooxidans* strains Licanantay (Travisany *et al*., 2014) and A01 (Yin *et al*., 2014a) have a SOR, while ATCC 19377 does not (Valdes *et al*., 2011). The absence of SOR in some snottite *At. thiooxidans* confirms that the absence of SOR in strain ATCC 19377 is likely a real biological difference, rather than an artifact of incomplete sequencing (Yin *et al*., 2014a).

Other than the presence of SOR, the AS and RS populations appear to have similar sulfur oxidation capabilities. Both populations encode homologs for the major enzyme complexes in partial SOX systems (SoxAX, SoxB and SoxYZ but not SoxCD), four structurally distinct sulfide quinone reductases (SqrA, SqrC, SqrE and SqrF), thiosulfate dehydrogenase (DoxD) and tetrathionate dehydrogenase

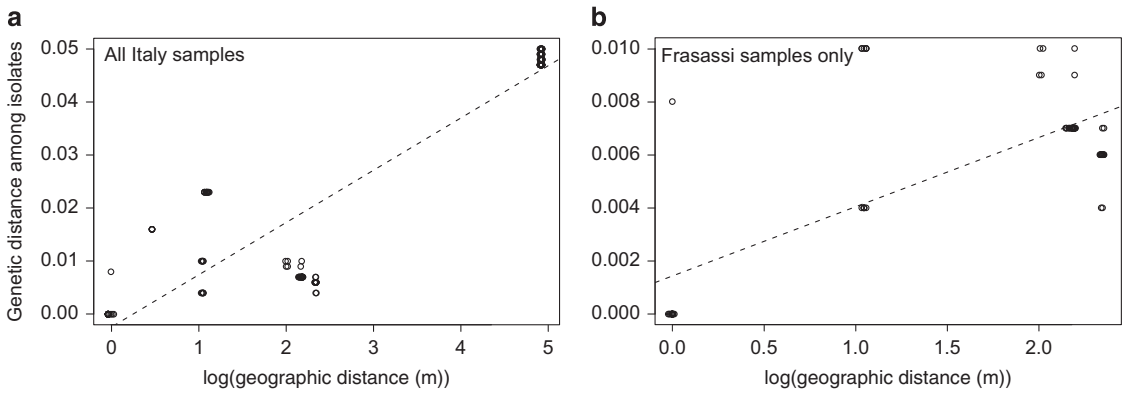

Figure 5 Genetic distance versus geographic distance for *At. thiooxidans* isolates from (**a**) the Frasassi and Acquasanta cave systems, and (**b**) the Frasassi cave system only. Isolates from RS30, RS31, RS2, GB30 and GB31 are excluded (see text for details). Genetic distance among strains was determined from MLST (Figure 4). Addition of a small random number was used to spread out overlapping points along the *x*-axis. Dashed lines are least squares regression lines, and Pearson's correlation for the depicted relationships are statistically significant (**a**: $r=0.93$, $P\ll0.001$; **b**: $r=0.75$, $P\ll0.001$).

(TetH) (Figure 6). They both have homologs of *hdrABC*, which has been proposed to be involved in the oxidation of S^0 in *Acidithiobacillus* spp. based on transcriptomic evidence (Quatrini *et al.*, 2009). Both data sets also contain homologs of all three of the proposed sulfur dioxygenases identified by Yin *et al.* (2014a), although experimental evidence will be required to confirm the sulfur dioxygenase function of these proteins.

To further compare potential metabolic differences among the *At. thiooxidans* populations in AS5 and RS9, we compared the two metagenomes against three isolate *At. thiooxidans* strains whose genomes are publically available. RS9 has homologs with 39–78 of the predicted protein-coding genes in the three isolate genomes that are absent in AS5 (Table 2). AS5 shares 172–360 homologs with the isolates that are absent in RS9, and the three isolates contain 111–540 protein-coding genes with no homolog in either AS5 or RS9 (Table 2). Most of the apparent genomic differences between the AS5 and RS9 populations encode hypothetical proteins. However, the presence and absence of homologs for genes involved in motility (e.g., *pilT*) and transport functions (e.g., peptide and sugar transporters) indicate further differences in metabolic capabilities between the AS5 and RS9 *At. thiooxidans* (Supplementary Table S5), in addition to their sulfur oxidation pathways (Figure 6). More detailed characterization of the physiological differences among the snottite

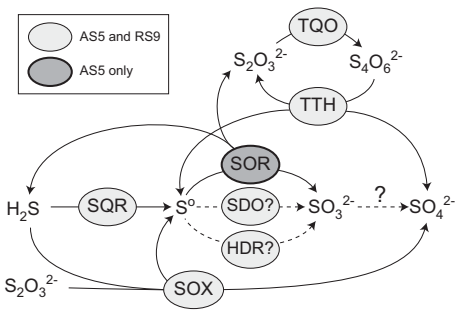

Figure 6 Cartoon of enzymatic sulfur transformations involved in the oxidation of reduced inorganic sulfur compounds by snottite *At. thiooxidans*, inferred from metagenomic analysis. Abbreviations: HDR, heterodisulfide reductase; SDO, sulfur diooxygenase; SOR, sulfur oxygenase reductase; SOX, multicomponent sulfur oxidation pathway; SQR, sulfide:quinone oxidoreductase; TQO, thiosulfate:quinone oxidoreductase; TTH, tetrathionate hydrolase.

At. thiooxidans strains is beyond the scope of this study. However, it seems clear that the *At. thiooxidans* strains from Frasassi encode metabolic capabilities not present in the Acquasanta population, and vice versa.

Discussion

Biogeography of snottite Acidithiobacillus *populations*
Acidithiobacillus 16S rRNA gene sequences from the Italy and Mexico locations share less than 95% nucleotide identity, and therefore should be considered separate species, *At. thiooxidans* and '*At.* group II' (Figure 2). The presence of separate *Acidithiobacillus* spp. in Italy and Mexico is not likely due to geographic isolation of those clades, because *At. thiooxidans* sequences have been found around the world, as have sequences from the *At.* group II (Figure 2). Therefore, the geographic pattern we observe is either due to environmental selection for *At. thiooxidans* in Italy and *At.* group II in Mexico, or due to stochastic colonization events that led to the establishment of *At. thiooxidans* and *At.* group II in separate cave systems. We are unable to distinguish between these two hypotheses using our existing data. However, the ecological success of both species suggests that snottite biofilm formation is an adaptation to the sulfidic cave environment, rather than a strategy associated with a single species of *Acidithiobacillus*.

The distance–decay relationship revealed by MLST suggests that evolutionary relationships among the dominant snottite *At. thiooxidans* populations in Italy result from restricted dispersal within and among caves (Figure 5). However, in addition to genetic differences based on housekeeping genes (Figure 4), populations in Acquasanta sample AS5 encode a sulfur oxidation enzyme that is absent from *At. thiooxidans* in RS9. This difference, while subtle, clearly indicates a potential for different sulfur oxidation capabilities among the *At. thiooxidans* populations that did not likely result from genetic drift. It is possible that the Acquasanta and Frasassi populations originated from the same ancestral population, and that the observed difference in sulfur oxidation pathways resulted from gene loss or horizontal gene transfer (Figure 7, dashed arrow). However, a more parsimonious explanation is that

Table 2 Total protein-coding genes in publicly available *At. thiooxidans* isolate genomes with homologs in either the AS5 or the RS9 snottite metagenome

	At. thiooxidans ATCC 19377	*At. thiooxidans* Licanantay	*At. thiooxidans* A01
Total protein-coding genes in isolate genome	2831	3785	3699
Total protein-coding genes with homologs in RS9 (but not AS5)	40	77	78
Total protein-coding genes with homologs in AS5 (but not RS9)	171	360	267
Total protein-coding genes with no homologs in either metagenome	111	540	539
Total protein-coding genes with homologs in both RS9 and AS5	2509	2808	2815

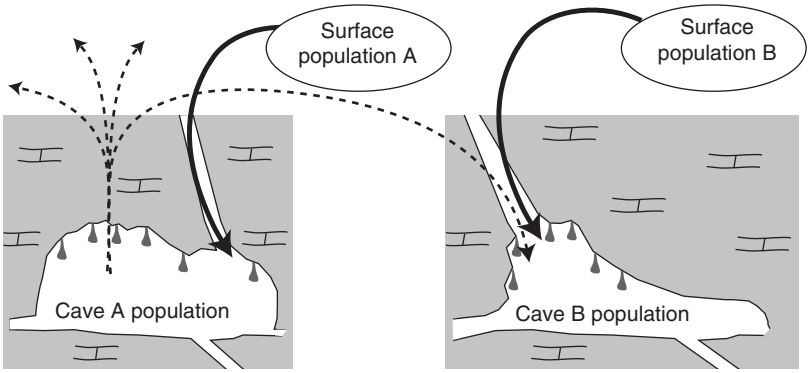

Figure 7 Proposed evolutionary histories of snottite populations in Italy. The caves could have been colonized from a single ancestral population (dashed arrow) that diverged by genetic drift or adaptation. Alternatively, the different cave populations could have originated from separate surface-dwelling populations with slight physiological differences (bold arrows), which then continued to diverge *in situ*.

the snottite populations in AS5 and RS9 originated from two ancestral populations with slightly different sulfur oxidation capabilities (Figure 7, solid arrows). The distance–decay relationship in Figure 5a could therefore partially result from genetic differences among the ancestral populations (founder effects), rather than purely from local adaptation or drift. Differences in functional attributes among populations from within the Frasassi cave system were not explored in our study, so we cannot distinguish whether radiation within caves or multiple colonization events shaped the genetic relationships among Frasassi populations (Figures 4 and 5b). It is worth noting, however, that snottite populations in the Acquasanta, Frasassi and Mexico caves all seem to reflect colonization by different strains or species of *Acidithiobacillus*.

Evidence for recent colonization events

The presence of rare *At. thiooxidans* strains in snottites that do not conform to the distance–decay relationship in Figure 5 makes it clear that the biogeography of sulfidic cave snottites is complex (Supplementary Figure S3). We were initially surprised that identical MLST genotypes could be cultured from Acquasanta site AS1 and Frasassi site GB (Figure 4), as these two cave sites are separated by more than 80 km. It is extremely unlikely that this result is due to lab contamination. In addition to stringent protocols designed prior to the culturing study and aimed at eliminating laboratory cross contamination, the AS1 inoculum was collected and used to produce isolates in 2007, while GB30 and 31 samples were collected and cultured in 2008. Moreover, if lab contamination were the source of the matching AS and GB isolates, we would expect to see the same contaminant in other 2008 cultures, and this did not occur (Table 1).

The GB cave site is closer to the surface than any of our other sites, and sample locations GB30 and GB31 are less than 30 m from the cave entrance (Supplementary Figure S1). Site GB is therefore subject to frequent tourist and recreational caving traffic. It is possible that a recent colonization event is responsible for the presence of identical rare strains at Frasassi site GB and Acquasanta site AS1. Human traffic into caves has provided a mechanism for microbial transport that did not exist until recently. Dust, lint and other aerosols from tourist caves advect into more distant passages (Michie, 1999), and in Frasassi, lint can be found cemented into speleothems in remote passages far removed from the heavily trafficked show cave (S Montanari, personal communication). More knowledge of the mechanisms of microbial transport and colonization in cave systems could inform cave management and protection strategies (e.g., Dupont *et al.*, 2007; Bastian and Alabouvette, 2009).

Despite these likely recent colonization events, the genetic distance among the dominant Acquasanta and Frasassi populations may be maintained because low abundance strains have not become established. Microbial biogeographic patterns may be determined by how effectively transplanted strains can take hold and compete with entrenched endemic populations (Gorbushina *et al.*, 2007; Hervàs *et al.*, 2009; Hanson *et al.*, 2012). A history of stochastic colonization and establishment of *Acidithiobacillus* populations from different surface sources is consistent with our finding that snottites in Acquasanta, Frasassi and Mexico caves are dominated by different strains or species of *Acidithiobacillus* (Figure 7).

Comparison of techniques for resolving biogeographic relationships among populations

Not surprisingly, we detected genetic divergence among isolates using ITS sequencing and MLST that we did not detect by 16S rRNA sequence analysis. This and other studies make it clear that the 16S rRNA gene is not sufficient to resolve all existing biogeographic relationships among microbial populations (Whitaker *et al.*, 2003; Vos and Velicer, 2008). However, the presence of highly similar or identical 16S rRNA sequences in extreme environments across the planet (e.g., Hollibaugh *et al.*, 2002; Whitaker *et al.*, 2003; Palacios *et al.*, 2008) suggests

that global dispersal may occur over long timescales, consistent with the million-year timescale for 16S rRNA gene evolution (Ochman *et al*., 1999; Itoh *et al*., 2002; Kuo and Ochman, 2009).

We found that there were both advantages and disadvantages to using ITS sequence analysis and MLST. Although the ITS region is more variable than the 16S rRNA gene, ITS sequence analysis provides less genetic resolution than MLST and is challenging for phylogenetic reconstruction due to the presence of large and frequent insertions and deletions. Alignment is unreliable because the ITS region has no secondary structure, contains multiple large gaps, and is subject to intragenomic divergence. We were able to use the ITS region to identify divergent populations from different cave systems, but genetic relationships among them are not completely resolved (Figure 3). Most significantly, environmental ITS cloning allowed us to evaluate which isolates were representative of the dominant snottite populations. MLST is a powerful technique for strain-level taxonomic identification, phylogenetic analysis and for identifying recombination events (e.g., Falush *et al*., 2003). However, as we found with isolates from sites RS and GB, MLST is subject to the biases inherent in culturing. A combination of techniques, like those applied here, is useful to balance the strengths, limitations and resolution of different measures.

Implications for microbial biogeography

This work represents a novel contribution to the field of microbial biogeography by applying a population genetics approach to investigate extremely acidophilic microorganisms in caves, an exceptional natural experiment in microbial isolation. We found that different species of *Acidithiobacillus* inhabit snottite biofilms on different continents, and, as hypothesized, we identified a distance–decay pattern among dominant snottite populations within Italy. However, this pattern emerged only after removing sequences representing rare *Acidithiobacillus* strains ('weeds'). The distance–decay relationship is maintained among the dominant populations in the presence of these rare strains either because the dominant populations are established and/or the rare strains are less fit. Furthermore, metagenomic analysis revealed differences in metabolic potential among cave populations, and we found statistically significant correlations between genetic distance and both geographic distance and environmental variables. We argue that the observed biogeographic relationships among snottite *Acidithiobacillus* populations are best explained by a scenario in which caves were initially colonized by distinct surface-dwelling *Acidithiobacillus* spp., which then continued to diverge and adapt *in situ*.

While distance–decay patterns are frequently observed in studies of geographically separated microbial populations, the underlying mechanisms behind the genetic relationships often remain elusive. Our findings underscore the role of stochastic colonization in creating or contributing to distance–decay patterns. This study also highlights the importance of using multiple genetic techniques for recognizing the biogeographic relationships among microbial populations, as well as the application of functional genomic information for confirming, rejecting or extending conclusions based on neutrally evolving markers alone. In this case, functional genomic information revealed metabolic differences that improved our interpretation of the processes behind the pattern (Figure 7). Nevertheless, and despite the extraordinarily low diversity and physical and chemical isolation of snottite microbial communities, the genetic relationships among cave *Acidithiobacillus* spp. are complex, and clearly indicate that multiple processes and timescales are relevant in explaining how microbial populations are distributed in the present day. Future experimental and observational studies will be required to build a more complete understanding of the the relative importance of physical and chemical barriers, dispersal, colonization, mutation, drift and other processes in controlling the genetic makeup of the diverse microbial species found in nature.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgements

This work was supported by a generous graduate research fellowship to DSJ from the Cave Conservancy Foundation, and grants to JLM from the Biogeosciences Program of the National Science Foundation (EAR 0311854 and EAR 0527046) and NASA NAI (NNA04CC06A). We thank A Montanari for logistical support and the use of facilities and laboratory space at the Osservatorio Geologico di Coldigioco (Italy), and C Alberto Cordero Martinez for lodging and logistical support in Villa Luz Park. Permits for access to Villa Luz were granted by C Rogers Morales Mendez and L Felino Arevalo Gallegos. We thank S Mariani, S Galdenzi, S Cerioni and M Mainiero for expert advice and field assistance with research in Italy. F Baldoni, S Recanatini, S Dattagupta and the members of the Gruppo Speleologico CAI di Fabriano and the Gruppo Speleologico Marchigiano CAI di Ancona provided additional assistance. We thank L Rosales-Lagarde, L Hose and S Dattagupta for field assistance in Mexico, T Jones and E Lyon for laboratory assistance at Penn State, and C House and C Grettenberger for insightful discussion.

References

Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W *et al.* (1997). Gapped BLAST and PSI-BLAST:

a new generation of protein database search programs. *Nucleic Acids Res* **25**: 3389–3402.

Amouric A, Brochier-Armanet C, Johnson DB, Bonnefoy V, Hallberg KB. (2011). Phylogenetic and genetic variation among Fe(II)-oxidizing acidithiobacilli supports the view that these comprise multiple species with different ferrous iron oxidation pathways. *Microbiology* **157**: 111.

Baas Becking LGM. (1934). *Geobiologie of inleiding tot de milieukunde*. WP Van Stockum and Zoon: The Hague, the Netherlands.

Bastian F, Alabouvette C. (2009). Lights and shadows on the conservation of a rock art cave: the case of Lascaux Cave. *Int J Speleol* **38**: 55–60.

Cho J-C, Tiedje JM. (2000). Biogeography and degree of endemicity of fluorescent *Pseudomonas* strains in soil. *Appl Environ Microbiol* **66**: 5448–5456.

de Rezende JR, Kjeldsen KU, Hubert CR, Finster K, Loy A, Jørgensen BB. (2013). Dispersal of thermophilic *Desulfotomaculum* endospores into Baltic Sea sediments over thousands of years. *ISME J* **7**: 72–84.

de Wit R, Bouvier T. (2006). 'Everything is everywhere, but, the environment selects'; what did Baas Becking and Beijerinck really say? *Environ Microbiol* **8**: 755–758.

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

Dupont J, Jacquet C, Dennetière B, Lacoste S, Bousta F, Orial G *et al.* (2007). Invasion of the French Paleolithic painted cave of Lascaux by members of the *Fusarium solani* species complex. *Mycologia* **99**: 526–533.

Escobar-Páramo P, Ghosh S, DiRuggiero J. (2005). Evidence for genetic drift in the diversification of a geographically isolated population of the hyperthermophilic archaeon *Pyrococcus*. *Mol Biol Evol* **22**: 2297.

Falush D, Stephens M, Pritchard JK. (2003). Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics* **164**: 1567.

Finlay BJ. (2002). Global dispersal of free-living microbial eukaryote species. *Science* **296**: 1061–1063.

Foti M, Ma S, Sorokin DY, Rademaker JLW, Kuenen JG, Muyzer G. (2006). Genetic diversity and biogeography of haloalkaliphilic sulphur oxidizing bacteria belonging to the genus *Thioalkalivibrio*. *FEMS Microbiol Ecol* **56**: 95–101.

Galdenzi S, Maruoka T. (2003). Gypsum deposits in the Frasassi Caves, central Italy. *J Cave Karst Stud* **65**: 111–125.

Gorbushina AA, Kort R, Schulte A, Lazarus D, Schnetger B, Brumsack HJ *et al.* (2007). Life in Darwin's dust: intercontinental transport and survival of microbes in the nineteenth century. *Environ Microbiol* **9**: 2911–2922.

Hahn MW, Koll U, Jezberová J, Camacho A. (2015). Global phylogeography of pelagic *Polynucleobacter* bacteria: restricted geographic distribution of subgroups, isolation by distance and influence of climate. *Environ Microbiol* **17**: 829–840.

Hanson CA, Fuhrman JA, Horner-Devine MC, Martiny JB. (2012). Beyond biogeographic patterns: processes shaping the microbial landscape. *Nat Rev Microbiol* **10**: 497–506.

Harrison AP. (1982). Genomic and physiological diversity amongst strains of *Thiobacillus ferrooxidans*, and genomic comparison with *Thiobacillus thiooxidans*. *Arch Microbiol* **131**: 68–76.

Hervàs A, Camarero L, Reche I, Casamayor EO. (2009). Viability and potential for immigration of airborne bacteria from Africa that reach high mountain lakes in Europe. *Environ Microbiol* **11**: 1612–1623.

Hollibaugh JT, Bano N, Ducklow HW. (2002). Widespread distribution in polar oceans of a 16 S rRNA gene sequence with affinity to *Nitrosospira*-like ammonia-oxidizing bacteria. *Appl Environ Microbiol* **68**: 1478–1484.

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.

Hose LD, Pisarowicz JA. (1999). Cueva de Villa Luz, Tabasco, Mexico: reconnaissance study of an active sulfur spring cave and ecosystem. *J Cave Karst Stud* **61**: 13–21.

Huber T, Faulkner G, Hugenholtz P. (2004). Bellerophon; a program to detect chimeric sequences in multiple sequence alignments. *Bioinformatics* **20**: 2317–2319.

Huelsenbeck JP, Ronquist F. (2001). MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* **17**: 754–755.

Itoh T, Martin W, Nei M. (2002). Acceleration of genomic evolution caused by enhanced mutation rate in endocellular symbionts. *Proc Natl Acad Sci USA* **99**: 12944.

Jones D, Tobler D, Schaperdoth I, Mainiero M, Macalady J. (2010). Community structure of subsurface biofilms in the thermal sulfidic caves of Acquasanta Terme, Italy. *Appl Environ Microbiol* **76**: 5902–5910.

Jones DS, Albrecht HL, Dawson KS, Schaperdoth I, Freeman KH, Pi Y *et al.* (2012). Community genomic analysis of an extremely acidophilic sulfur-oxidizing biofilm. *ISME J* **6**: 158–170.

Jones DS, Schaperdoth I, Macalady JL. (2014). Metagenomic evidence for sulfide oxidation in extremely acidic cave biofilms. *Geomicrobiol J* **31**: 194–204.

Kuo CH, Ochman H. (2009). Inferring clocks when lacking rocks: the variable rates of molecular evolution in bacteria. *Biol Direct* **4**: 35.

Larkin M, Blackshields G, Brown N, Chenna R, McGettigan P, McWilliam H *et al.* (2007). Clustal W and Clustal X version 2.0. *Bioinformatics* **23**: 2947.

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

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.

Macalady JL, Jones DS, Lyon EH. (2007). Extremely acidic, pendulous microbial biofilms from the Frasassi cave system, Italy. *Environ Microbiol* **9**: 1402–1414.

Martiny JBH, Bohannan BJ, Brown JH, Colwell RK, Fuhrman JA, Green JL *et al.* (2006). Microbial biogeography: putting microorganisms on the map. *Nat Rev Microbiol* **4**: 102–112.

McCune B, Grace JB. (2002). Analysis of Ecological Communities, vol. 28. MjM Software Design, Gleneden Beach, OR.

Michie NA. (1999). An instrument and method for measurement of dust fall in caves. *National Cave and Karst Management Symposium*, pp 23–28.

Ochman H, Elwyn S, Moran NA. (1999). Calibrating bacterial evolution. *Proc Natl Acad Sci USA* **96**: 12638.

Oksanen J, Kindt R, Legendre P, O'Hara R, Simpson GL, Stevens MHH. (2008). Vegan: Community Ecology Package, R package version 1.11-0.

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.

Papke RT, Ramsing NB, Bateson MM, Ward DM. (2003). Geographical isolation in hot spring cyanobacteria. *Environ Microbiol* **5**: 650–659.

Papke RT, Ward DM. (2004). The importance of physical isolation to microbial diversification. *FEMS Microbiol Ecol* **48**: 293–303.

Posada D. (2008). jModelTest: phylogenetic model averaging. *Mol Biol Evol* **25**: 1253–1256.

Quatrini R, Appia-Ayme C, Denis Y, Jedlicki E, Holmes DS, Bonnefoy V. (2009). Extending the models for iron and sulfur oxidation in the extreme acidophile *Acidithiobacillus ferrooxidans*. *BMC Genomics* **10**: 394.

R Core Development Team (2007). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing: Vienna, Austria.

Raymond J, Alsop EB. (2015). Microbial evolution in extreme environments: microbial migration, genomic highways, and geochemical barriers in hydrothermal ecosystems. *Environ Syst Res* **4**: 1–12.

Roberts MS, Cohan FM. (1995). Recombination and migration rates in natural populations of *Bacillus subtilis* and *Bacillus mojavensis*. *Evolution* **49**: 1081–1094.

Ryšánek D, Hrčková K, Škaloud P. (2014). Global ubiquity and local endemism of free-living terrestrial protists: phylogeographic assessment of the streptophyte alga *Klebsormidium*. *Environ Microbiol* **17**: 689–698.

Sagredo B, Jedlicki E, Orellana O. (1992). Organization of the 16 S-23 S intergenic spacer region of the two rRNA operons from *Thiobacillus ferrooxidans*. *Geomicrobiol J* **10**: 239–247.

Schattner P, Brooks AN, Lowe TM. (2005). The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. *Nucleic Acids Res* **33**: W686.

Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB *et al.* (2009). Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl Environ Microbiol* **75**: 7537–7541.

Sikorski J, Nevo E. (2005). Adaptation and incipient sympatric speciation of *Bacillus simplex* under microclimatic contrast at "Evolution Canyons" I and II, Israel. *Proc Natl Acad Sci USA* **102**: 15924.

Simmons MP, Ochoterena H. (2000). Gaps as characters in sequence-based phylogenetic analyses. *Syst Biol* **49**: 369–381.

Stewart FJ, Cavanaugh CM. (2007). Intragenomic variation and evolution of the internal transcribed spacer of the rRNA operon in bacteria. *J Mol Evol* **65**: 44–67.

Swofford DL. (2000). *PAUP*: Phylogenetic Analysis using Parsimony and Other Methods (Software)*. Sinauer Associates: Sunderland, MA.

Travisany D, Cortés MP, Latorre M, Di Genova A, Budinich M, Bobadilla-Fazzini RA *et al.* (2014). A new genome of *Acidithiobacillus thiooxidans* provides insights into adaptation to a bioleaching environment. *Res Microbiol* **165**: 743–752.

Valdes J, Ossandon F, Quatrini R, Dopson M, Holmes DS. (2011). Draft genome sequence of the extremely acidophilic biomining bacterium *Acidithiobacillus thiooxidans* ATCC 19377 provides insights into the evolution of the *Acidithiobacillus* genus. *J Bacteriol* **193**: 7003–7004.

van Gremberghe I, Leliaert F, Mergeay J, Vanormelingen P, Van der Gucht K, Debeer A-E *et al.* (2011). Lack of phylogeographic structure in the freshwater cyanobacterium *Microcystis aeruginosa* suggests global dispersal. *PloS one* **6**: 1–12.

Venegas A, Hevia E, Sánchez H. (1988). Sequence of two tRNA genes from a *Thiobacillus ferrooxidans* ribosomal operon. *Nucleic Acids Res* **16**: 8179.

Vlasceanu L, Sarbu SM, Engel AS, Kinkle BK. (2000). Acidic cave wall biofilms located in the Frasassi Gorge, Italy. *Geomicrobiol J* **17**: 125–139.

Vos M, Velicer GJ. (2008). Isolation by distance in the spore-forming soil bacterium *Myxococcus xanthus*. *Curr Biol* **18**: 386–391.

Whitaker RJ. (2006). Allopatric origins of microbial species. *Phil Trans Royal Soc B* **361**: 1975.

Whitaker RJ, Grogan DW, Taylor JW. (2003). Geographic barriors isolate endemic populations of hyperthermophilic archaea. *Science* **301**: 976–978.

Yin H, Zhang X, Li X, He Z, Liang Y, Guo X *et al.* (2014a). Whole-genome sequencing reveals novel insights into sulfur oxidation in the extremophile *Acidithiobacillus thiooxidans*. *BMC Microbiol* **14**: 179.

Yin H, Zhang X, Liang Y, Xiao Y, Niu J, Liu X. (2014b). Draft genome sequence of the extremophile *Acidithiobacillus thiooxidans* A01, isolated from the wastewater of a coal dump. *Genome Announc* **2**: e00222–00214.

Young ND, Healy J. (2003). GapCoder automates the use of indel characters in phylogenetic analysis. *BMC Bioinformatics* **4**: 6.

Supplementary Information accompanies this paper on The ISME Journal website (http://www.nature.com/ismej)