



Extreme Niche Partitioning and Microbial Dark Matter in a Mauna Loa Lava Tube

C B Fishman, J G Bevilacqua, A S Hahn, C Morgan-lang, N Wagner, O Gadson, A C Mcadam, J Bleacher, C Achilles, C Knudson, et al.

► To cite this version:

C B Fishman, J G Bevilacqua, A S Hahn, C Morgan-lang, N Wagner, et al.. Extreme Niche Partitioning and Microbial Dark Matter in a Mauna Loa Lava Tube. *Journal of Geophysical Research. Planets*, 2023, 128 (6), pp.e2022JE007283. <10.1029/2022je007283>. <hal-04272765>

HAL Id: hal-04272765

<https://hal.science/hal-04272765v1>

Submitted on 6 Nov 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

JGR Planets

RESEARCH ARTICLE

10.1029/2022JE007283

C. B. Fishman and J. G. Bevilacqua contributed equally to this work.

Special Section:

Exploring planetary caves as windows into subsurface geology, habitability, and astrobiology

Key Points:

- In this undisturbed environment, communities were characterized by multiple microniches, varying widely over the scale of meters
- The lava tube contained high amounts of “microbial dark matter” or uncultivated microbes that remain unclassified in taxonomic databases
- Although the mineralogy of many of the samples we studied was sulfate-rich, metabolisms unrelated to sulfur dominated the metagenome

Supporting Information:

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

Correspondence to:

S. S. Johnson,
sarah.johnson@georgetown.edu

Citation:

Fishman, C. B., Bevilacqua, J. G., Hahn, A. S., Morgan-Lang, C., Wagner, N., Gadson, O., et al. (2023). Extreme niche partitioning and microbial dark matter in a Mauna Loa lava tube. *Journal of Geophysical Research: Planets*, 128, e2022JE007283. <https://doi.org/10.1029/2022JE007283>

Received 1 MAR 2022

Accepted 29 JAN 2023

Author Contributions:

Conceptualization: C. B. Fishman, A. C. McAdam, S. S. Johnson

Data curation: C. B. Fishman

Formal analysis: C. B. Fishman, J. G. Bevilacqua, A. S. Hahn, O. Gadson, A. C. McAdam, C. Achilles

Funding acquisition: A. C. McAdam, J. Bleacher, S. S. Johnson

Extreme Niche Partitioning and Microbial Dark Matter in a Mauna Loa Lava Tube

C. B. Fishman¹, J. G. Bevilacqua¹, A. S. Hahn², C. Morgan-Lang² , N. Wagner¹, O. Gadson¹, A. C. McAdam³ , J. Bleacher^{3,4} , C. Achilles⁵ , C. Knudson^{3,5,6} , M. M. Millan^{1,3} , D. M. Bower⁶, M. Musilova^{7,8} , and S. S. Johnson^{1,9} 

¹Department of Biology, Georgetown University, Washington, DC, USA, ²Koonkie Inc., Menlo Park, CA, USA, ³NASA Goddard Space Flight Center, Greenbelt, MD, USA, ⁴NASA Headquarters, Washington, DC, USA, ⁵CRESST, University of Maryland, College Park, MD, USA, ⁶Department of Astronomy, University of Maryland, College Park, MD, USA, ⁷Institute of Robotics and Cybernetics, Bratislava, Slovakia, ⁸Slovak Organisation for Space Activities (SOSA), Bratislava, Slovakia, ⁹Science, Technology, and International Affairs Program, Georgetown University, Washington, DC, USA

Abstract Lava tubes are key targets in the search for life on Mars. Their basaltic walls provide protection from radiation and changing environmental conditions, which could enable life or preservation of previous life in an otherwise harsh environment. We can understand the potential for Martian life in lava tubes by studying the habitability of analog environments on Earth. In this study, we present the first characterization of the microbial life inside a pristine Mauna Loa lava tube. This study is the first to combine 16S SSU rRNA sequencing and whole genome shotgun sequencing to map the taxonomic makeup and functional potential of any lava tube community in Hawaii, enabling a deep understanding of the types of microbes that thrive in this unique environment and the metabolisms they use. We find a surprisingly high degree of niche partitioning over small spatial scales and discuss implications for life detection strategies. Based on recent bioinformatic advancements in metagenomics, we also assemble dozens of high-quality metagenome assembled genomes from the microbes living in the lava tubes, including several novel species.

Plain Language Summary In the search for life on Mars, lava tubes are targets for astrobiology. Lava tubes provide an underground shelter, protected from radiation and the harsh climate. Analog studies here on Earth can help us understand what types of microbes live in these terrains and how best to find them. In this study, we investigate the microbial communities in a previously unexplored lava tube on the Mauna Loa volcano in Hawaii. We set out to characterize the microbes that live there and what capabilities enable them to survive, such as how they acquire nutrients from their environment to produce energy. We collected mineral samples in the lava tube and sequenced the DNA that we found in the minerals. We used DNA sequences to identify which microbes are present as well as which metabolisms they can use. We found that there are significant differences in the microbial make-up of each of the samples, even though some are taken within several centimeters of each other, with implications for how to assay a region of interest on Mars. We also found several novel species.

1. Introduction

Beneath the terrain of Mauna Loa in Hawaii, the largest active volcano on the planet (United States Geological Survey, 2022), lies a maze of basaltic lava tubes that are home to uncharacterized microbial communities and novel species. This subsurface basaltic environment provides a unique environment for microbial life largely sheltered from the sun, wind, temperature fluctuations, plants, and animals. On Mauna Loa, lava tubes are coated in colorful secondary minerals, such as gypsum, thenardite, and calcite that form as a result of weathering (White, 2010). These mineral deposits can provide additional protection and substrates for life, including water and ions involved in their metabolisms (Boston et al., 2001; Huang et al., 2020; King & Nanba, 2008; Panieri et al., 2010; Sakai et al., 1982).

Lava tubes have also been identified on Mars from orbit and are of great interest in our search for signs of life on Mars (Léveillé & Datta, 2010). To find life or a record of previous life, it is advantageous to look underground, where the protective nature of subsurface environments may have shielded life or its traces from the subsequent alteration of surface conditions on Mars. Current scientific consensus posits that Martian subsurface environments could sustain life in an otherwise harsh environment (Cushing et al., 2007; Léveillé & Datta, 2010;

Investigation: C. B. Fishman, N. Wagner,

A. C. McAdam, C. Achilles, D. M. Bower

Methodology: C. B. Fishman, N.

Wagner, M. M. Millan, S. S. Johnson

Project Administration: A. C. McAdam,

J. Bleacher, C. Knudson, M. Musilova, S.

S. Johnson

Resources: S. S. Johnson

Software: C. B. Fishman, J. G.

Bevilacqua, A. S. Hahn, C. Morgan-Lang

Supervision: S. S. Johnson

Visualization: C. B. Fishman, J. G.

Bevilacqua, A. S. Hahn, C. Morgan-Lang

Writing – original draft: C. B. Fishman,

J. G. Bevilacqua, A. S. Hahn

Writing – review & editing: C. B.

Fishman, J. G. Bevilacqua, A. S. Hahn, C.

Morgan-Lang, N. Wagner, O. Gadson, A.

C. McAdam, J. Bleacher, C. Achilles, C.

Knudson, M. M. Millan, D. M. Bower, M.

Musilova, S. S. Johnson

Michalski et al., 2018). The formation of lava tubes in the Martian subsurface coincided with a marked period of Earth-like, habitable climate conditions (Hughes et al., 2019). The volcanic activity on Mars peaked in the late Noachian through Hesperian epochs (~3.7 billion years ago), and during this period, Mars also maintained a dense atmosphere and temperatures high enough to host oceans on the planet's surface (Baker et al., 1991; Carr & Head, 2010; Carter et al., 2015; Jakosky et al., 2017; Wordsworth, 2016; Zuber, 2001). Emerging microbial life could have found a home in lava tubes, whether life developed on the surface and subsequently adapted to live underground, washed into the subsurface via water circulation, or developed underground originally (Westall et al., 2015).

The lava tubes of Mauna Loa, Hawaii, offer a basaltic subsurface environment analogous in terms of volcanism and geochemistry to those visible from orbit on Mars. Investigating life in lava tubes can lend insights into habitability and the range of energy-harvesting possibilities available to microbial organisms in basaltic substrates. The lava tubes explored in this study were formed by geologically young flows (~200 years ago), mirroring the conditions in which pioneer microbes could have emerged in Martian lava tubes (Lockwood & Lipman, 1987).

Mauna Loa is intermittently exposed to acidic aerosols from degassing of surrounding, more recently active volcanoes such as Kilauea (Sakai et al., 1982). Sulfuric acid aerosols can weather basalt and cause sulfate mineral deposits in Mauna Loa's lava tubes, which modeling indicates are likely to have formed on Mars as well (Renggli et al., 2019; White, 2010; Zolotov & Mironenko, 2016). Although there is limited research on Martian subsurface mineralogy, scientists have identified minerals similar to those found on Mauna Loa on the planet's surface (Benison & Karmanocky, 2014; Ming et al., 2008; Vaniman et al., 2018), and geochemical modeling of evaporate processes on Martian basalt suggests that similar secondary mineral formations may exist in Martian lava tubes (Kah et al., 2018; Tosca et al., 2005). Minerals formed through microbial mediation, even if life were no longer present in Martian lava tubes, could also retain a record of its existence (Oschmann, 2000; Tang et al., 2014; Wilson & Bish, 2011). The variety of secondary minerals and microbe-mineral interactions create a range of potential niches in which microbial life may thrive.

In this investigation, we used 16S small subunit (SSU) ribosomal ribonucleic acid (rRNA) gene sequencing and whole genome shotgun (WGS) sequencing on mineral samples to elucidate the taxonomic composition and functional potential of the microbial communities in a Mauna Loa lava tube characterized by limited light, water, and external chemical energy inputs. Since little is known about the community structure and diversity of Hawaiian lava tubes, we aimed to discover which types of metabolic pathways enable microbial survival.

2. Materials and Methods

2.1. Sample Collection

Samples were collected from various sections of a lava tube on the northern flank of the Mauna Loa volcano (Figure 1). The lava tube is located at N19° 36.392' W155° 28.993' near the Hawaii Space Exploration Analog and Simulation (HI-SEAS) station at ~2,400 m elevation. Secondary minerals coated the walls and ceiling of the lava tube and were accessible by walking through with minimal climbing. Although these lava tubes are closed to the public, samples were chosen strategically (e.g., above ground-level, in crevices) to avoid contamination from HI-SEAS scientists or animals that could enter the lava tubes. Samples were collected along a gradient of light conditions, including high light, dim light, and darkness, as indicated by white, black, and gray stars (Figure 1a). Within each sampling area, samples were collected within ~10 cm to 3 m of each other. The samples ranged in color, texture, and morphology, including flakes and coralloids (Table S2 in Supporting Information S1). Each sample was collected and stored in cryotubes packed in sterile Whirl-Paks (Nasco, St. Petersburg, FL). Sterile sampling techniques were used, including manipulating samples with DNA-free utensils while wearing nitrile gloves. Samples were preserved in a cryoshipper at –80°C and on ice in transit from Hawaii to Georgetown University.

2.2. Mineralogy

For X-ray diffraction analysis, sample fines were prepared by crushing to a fine powder via mortar and pestle, without sieving to preserve as much sample as possible (estimated <45 µm), or drilling material from the surfaces of samples using a Dremel tool when necessary. XRD patterns were acquired from 2 to 70° 2Theta (Cu Kα

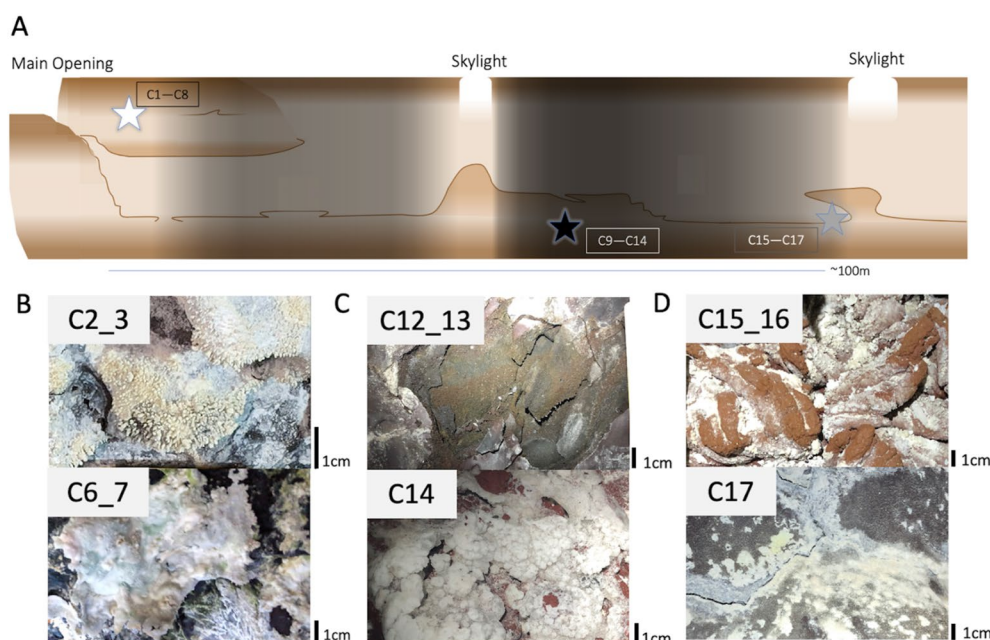


Figure 1. (a) Schematic of the lava tube with sampling locations and their topography and light conditions. Samples were collected in three concentrated areas with varying light exposure, denoted by a star symbol (left to right: light, dark, dim). Samples denoted with two numbers separated by an underscore (e.g., C2_3) are replicates. The total investigated region of this lava tube extends ~100 m. For illustrative purposes, samples C2_3 and C6_7 are shown in (b), samples C12_13 and C14 in (c), and samples C15_16 and C17 in (d). The black scale bar to the right of each image represents ~1 cm.

radiation, $k = 1.54059 \text{ \AA}$), at $0.01^\circ/\text{step}$ and at least 2 s/step using a Bruker D8 Discover diffractometer. Quantitative mineralogy was obtained by Rietveld refinement methods using the MDI (Materials Data Incorporated) Jade™ analysis software.

2.3. DNA Extraction

DNA was extracted with the ZymoBIOMICS DNA/RNA Miniprep Kit (Zymo Research, Irvine, CA) using high-density bashing beads and spin-filter binding chemistry. DNA was extracted from 100 to 250 mg of starting material and duplicates were extracted if sufficient sample mass was present. Cells were physically lysed by bead-beating in the FastPrep-24 Classic Instrument (MP Biomedicals, Irvine, CA) at a speed of 5.5 m/s for 1 min. DNA was purified following manufacturer's protocol, including multiple spin-filter washing steps. Resulting DNA was eluted in 35 μL of DNase/RNase-free water that had been heated to 70°C , and DNA was then stored at -80°C . DNA quantification was performed on a Qubit fluorometer using 1X dsDNA HS (High-Sensitivity) Assay Kit (Thermo Fisher Scientific, Waltham, MA). All samples used for analysis contained over 0.1 ng/ μL of DNA.

2.4. DNA Sequencing

Library preparation and paired-end sequencing for 16S SSU rRNA gene V4 region (515F–806 R) sequencing and WGS metagenome paired-end sequencing were performed at the University of Illinois at Chicago Sequencing Core. The V4 region within the 16S SSU rRNA gene has conserved sequences, used for primer binding, surrounding a variable region, which is used for taxonomic identification (Caporaso et al., 2011). This region was amplified from genomic DNA using PCR (5ul MyTaq HS Mix, 0.5ul 10 μM forward and reverse primers, 3 μL pure water, 1 μL template containing 0.1–15 ng/ μL DNA). PCR conditions were 95°C for 5 min, followed by 28 cycles of 95°C for 30 s, 55°C for 45 s and 72°C for 60 s. Amplicons were indexed and pooled as a library before being sequenced on an Illumina MiniSeq.

WGS sequencing captures the entire metagenome of a sample, including the 16S SSU rRNA regions, as well as the other genes of the community. Biological replicate samples were sequenced separately for 16S rRNA SSU

gene sequencing and were pooled to increase biomass for WGS sequencing. For example, samples C2 and C3 are replicates that were prepared separately and in parallel, kept distinct for 16S rRNA SSU sequencing and averaged in subsequent analyses or combined (denoted as C2_3) for WGS sequencing. Five samples (C1, C5, C7, C11, and C17) were prepped for 16S rRNA SSU sequencing but did not yield sufficient reads to include in analyses. Samples C5 and C7 were pooled with their biological replicates for WGS sequencing (and so are included in WGS samples C4_5 and C6_7), and results from the other three samples (C1, C11, C17) were only available for WGS sequencing, which has greater sequencing depth and the potential to gain more taxonomic and metabolic information. An Illumina DNA Prep kit was used to generate the DNA library for WGS sequencing, which was performed on a NovaSeq6000 SP flow cell to give 150 bp paired-end reads.

2.5. Bioinformatics Analyses

2.5.1. 16S SSU rRNA Gene Sequence Analysis

Massively parallel analyses of SSU rRNA genes highlight the genotypic diversity of a community and can be used to determine how this diversity changes in response to environmental change or disturbance (Lozupone and Knight, 2005). While these provide taxonomic data and function can sometimes be inferred from taxonomy (Douglas et al., 2020), they do not provide direct evidence of metabolic or functional potential. Reads generated from 16S SSU rRNA gene sequences were demultiplexed and trimmed to remove primers using Trimmomatic (Bolger et al., 2014). Amplicon sequence variants (ASVs) were selected for each of the 13 samples (including four replicates) using default settings of the DADA2 workflow version 1.16, and ASVs were assigned taxonomy using the Ribosomal Database Project database through DADA2 (Callahan et al., 2016). Samples with fewer than 5,000 reads were excluded from downstream analysis, and singleton ASVs (appearing once in only one sample) were filtered out for all analyses except for alpha diversity. Community composition and alpha diversity (Shannon diversity index) were assessed in R using the following packages: phyloseq, ggplot2, and pvcult (McMurdie & Holmes, 2013; Shannon, 1948; Suzuki & Shimodaira, 2006; Wickham, 2016). ASV distributions were assessed in R using UpSetR (Conway et al., 2017).

2.5.2. WGS Metagenome Analysis

WGS metagenomic sequencing analysis was conducted using the ATLAS pipeline (Kieser et al., 2020), which is used for the quality control, assembly, binning and annotation of metagenomes. In brief, ATLAS used clumpify (Bushnell, 2019) to remove PCR duplicates and used BBTools utilities (Bushnell, 2019) to remove adapters, low-quality bases, and short reads, then merged paired-end reads, filtered out PhiX contamination, and performed error correction using k-mer coverage. Read assemblies were generated using MetaSPAdes (Nurk et al., 2017) and BBDMap was used to map reads back to the assemblies (Bushnell, 2019). Taxonomy was assigned using Kaiju (Menzel et al., 2016). We used default settings for Kaiju wherein matches are filtered by a minimum length (11), score (65), and E-value (similar to blastp) (0.01). Prodigal (Hyatt et al., 2012) was used to predict open reading frames (ORFs) and EggNog Mapper (Cantalapiedra et al., 2021; Huerta-Cepas et al., 2019) was used for annotation. Figures were generated using pvcult (Suzuki & Shimodaira, 2006) and ggplot (Wickham, 2016). Finally, we used Tree-based Sensitive and Accurate Phylogenetic Profiler (TreeSAPP) (Morgan-Lang et al., 2020) to annotate ORFs related to assimilatory and dissimilatory sulfur metabolism as well as sulfur oxidation (SOX) genes. Reference sequences were sourced from EggNOG and UniProt databases, and resulting reference packages are available at <https://github.com/hallamlab/RefPkgs>.

2.5.3. Metagenome-Assembled Genomes

Metagenomic sequences contain both taxonomic and metabolic information about the great diversity of microbes in metagenomic data that have not yet been isolated and cultured (Steen et al., 2019; Wooley et al., 2010). Comparative genomic analyses allow the investigation of the metabolic potential of an organism or community. Because high throughput reads are short, the data must be assembled, or stitched together to create contiguous sequences (contigs). These contigs can then be binned or grouped together based on coverage, GC content, codon usage and other genomic features to define a “metagenome assembled genome” or MAG. MAGs are intended to be a single-taxon assembly and a close, but imperfect, representation of an individual genome. According to the ATLAS pipeline (Kieser et al., 2020), MAGs were generated using two binning tools, MetaBat2 and MaxBin2 (Kang et al., 2019; Wu et al., 2016). These bins were then combined using the DAS tool (C. Sieber, 2017; C. M. K. Sieber et al., 2018). Unique genomes were identified using dRep (Olm et al., 2017), which filters genomes by

size and quality before clustering using Mash and MUMmer (Delcher et al., 2003; Marçais et al., 2018; Ondov et al., 2016). The quality of the MAGs was quantified using Check-M (Parks et al., 2015). MAG taxonomy was assigned using GTDB-Tk, which places genomes according to protein reference trees (Chaumeil et al., 2019). Following ATLAS (Kieser et al., 2020), contig quality control was performed on all MAGs using MetaPathways (Konwar et al., 2015) on default settings, which then draws on MetaProdigal (Hyatt et al., 2012) for open reading frame prediction. ORFs were translated into amino acids and run against several reference databases: MetaCyc (2020-08-10) (Caspi et al., 2020; Karp et al., 2002), CAZy (2020-06-01) (Lombard et al., 2014), RefSeq-NR (2020-08-12-rel 201) (O'Leary et al., 2016), and Uniprot-Sprot (2020-08-12) (The UniProt Consortium, 2019), using FAST (Kim et al., 2016) to assess similarity between protein sequences. Taxonomy was assigned using Kaiju (Menzel et al., 2016) on default settings.

3. Data and Results

3.1. Mineralogy Results

Most samples were composed of >50% sulfate minerals, which were predominantly gypsum (CaSO_4) and thenardite (Na_2SO_4) (See Table S1 in Supporting Information S1). Exceptions included calcite-rich samples (C2_3, C9), a sample with a large (81%) X-ray amorphous component (C8, with 16% calcite and 3% gypsum), and a sample representative of the underlying basalt (C12_13).

3.2. 16S SSU rRNA Gene Sequence Results

16S SSU rRNA gene analysis was used to characterize the community composition and structure of nine samples collected along this Mauna Loa lava tube. After sequencing, processing, and filtering out samples with low read coverage, the read counts per sample ranged from 12,787 to 180,276 total reads per sample. In terms of taxonomic assignments, after filtering out singleton ASVs and before averaging across replicate samples, and excluding 1 ASV, which did not have a taxonomic assignment at the domain level, we found that 49.2% of ASVs were assigned down to the genus level, while 21.3% of ASVs could not be resolved below the class level. Within this set of highly unresolved ASVs, we observed that 7.5% of total ASVs could not be assigned below the domain level. These taxonomic assignments indicate a prevalence of novel taxa within the tube and/or underrepresented taxa in databases. The most abundant phylum in all samples except one (C12_13) is *Actinobacteria*, and other dominant phyla across most samples in this set are *Proteobacteria* and *Chloroflexi* (Figure 2b). Sample C12_13 is dominated by *Proteobacteria* and is the one sample representative of the underlying basaltic mineralogy rather than secondary minerals. ASV analysis also identified several candidate divisions and poorly characterized phyla in our samples, including BRC1, WPS-1, *Candidatus Saccharibacteria*, *Parcubacteria*, and *Ignavibacteriae*. This analysis also indicated the presence of bacterial and archaeal phyla known to contain extremophiles, including *Deinococcus-Thermus* and *Euryarchaeota*. Given the small sample set and the differences between the few samples we have, there are likely more uncaptured taxa in this environment.

In terms of ASV distribution across samples, we noted that very few ASVs were shared among samples, even samples taken within a small spatial range. Over one quarter (28.3%) of ASVs were found in only one sample, and only 18.4% of ASVs were found in five or more of the nine samples studied here. 49.1% of ASVs were found in three or more samples. We found that Sample C12_13 had the highest number (322) of unique ASVs that were not shared by any others, as well as an additional 134 ASVs shared only by sample C9 (Figure 2c).

Both C12_13 and C9 were samples collected from the dark portion of the lava tube, where a lack of light exposure may drive communities toward more specialized taxa that are either not supported or are out-competed in environments with access to light. The underlying uniqueness of ASVs in samples C9 and C12_13 may also be driven by mineralogy. C12_13 was collected from the basalt crust and is primarily composed of olivine and plagioclase, while C9 is a secondary mineral sample composed primarily of calcite as well as amorphous minerals (Table S1 in Supporting Information S1). C9 also shares many ASVs with samples C2_3, C4, and C8, which are in the light yet contain calcite and/or amorphous mineral contents, as does sample C9.

The unique populations identified in these samples, however, do not indicate higher diversity. Instead, we found that samples in darkness or dim light had lower diversity, on average, than samples in the light (Table 1). Alpha diversity was assessed using Shannon diversity, which varied across samples collected from the two sections

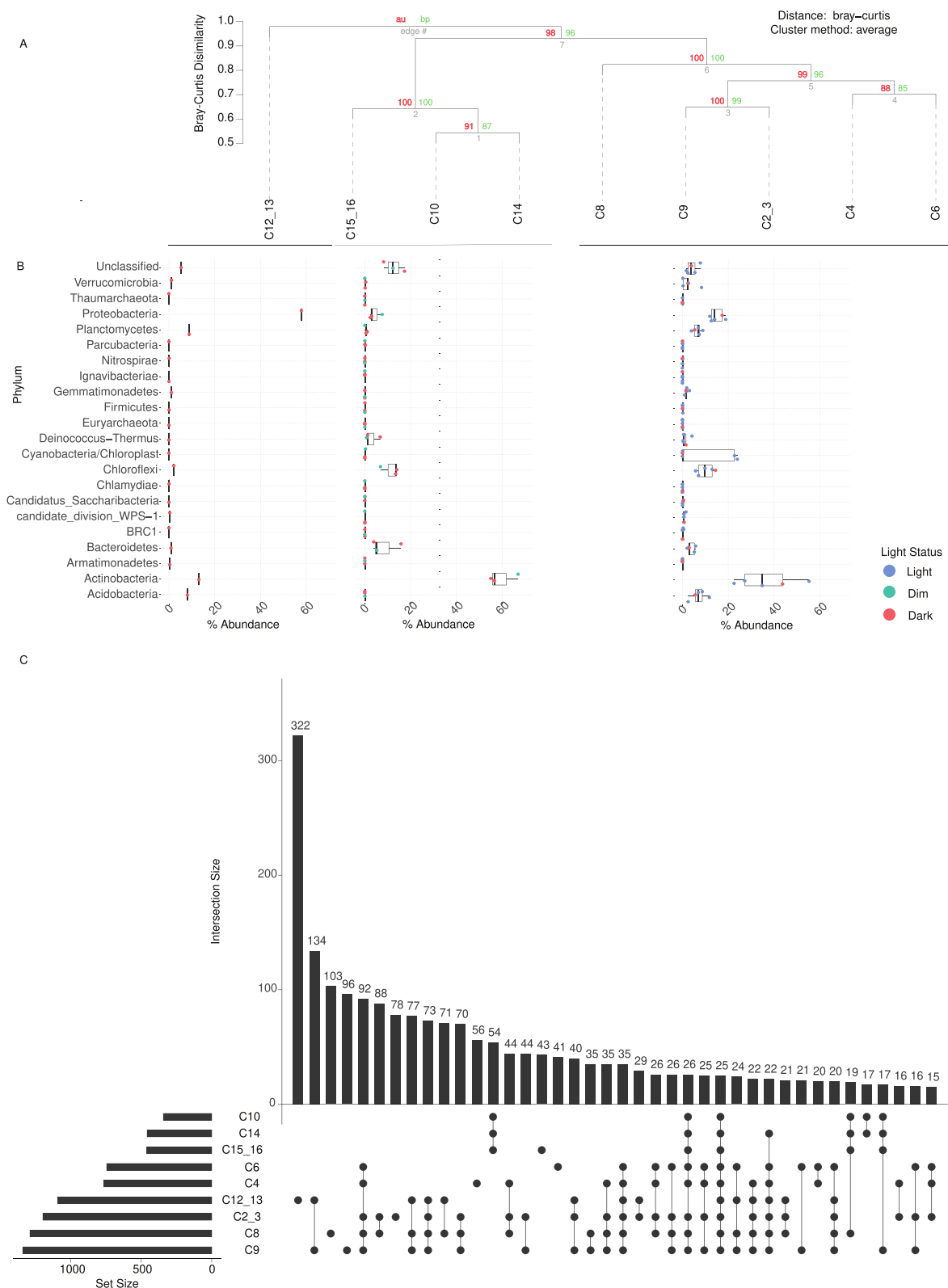


Figure 2. Dendro-box plot of microbial communities identified by 16S SSU rRNA sequencing across the samples taken within the lava tube. (a) Clustering shows the relatedness of communities, with replicate samples averaged. (b) Box plots reveal the distribution of amplicon sequence variants (ASVs) in the phyla per sample. These are separated by the clusters of the dendrogram and annotated with the light condition of each sample. (c) UpSet plot of ASVs shared among samples. Samples with two numbers separated by an underscore are averaged replicates.

Table 1
Shannon Diversity for the Amplicon Sequence Variants Across the Lava Tube

Sample ID	Shannon diversity
C2_3	5.49
C4	4.98
C6	4.91
C8	5.19
C9	5.03
C10	3.57
C12_13	4.74
C14	3.63
C15_16	3.68

of the lava tube (Welch Two Sample t -test $p = 0.029$). Samples in direct light had a mean Shannon diversity of 5.14 (± 0.26), while samples in either dim light or darkness had mean Shannon diversity of 4.13 (± 0.70). Despite the high numbers of unique ASVs found only in samples in the dark, the lower overall Shannon diversity scores of samples in dim light or darkness suggest that community size is reduced and more specialized taxa thrive in these conditions (Table 1). This variation could also be explained by the mineralogical differences between samples as the samples in the light are primarily composed of gypsum. However, sample C10 has the lowest diversity, was collected in the dark, and is 98% gypsum, suggesting that the lower diversity in samples C10, C14, and C15_16 is due to differences in light. The exception to this diversity trend is sample C9, which has a diversity score of 5.03. As noted previously, sample C9 has unique mineralogical makeup, which may support a more diverse microbial community despite being in the dark. It is likely that the microbial communities are shaped by all the characteristics of their niche. More work with an increased sample size is needed to capture robust trends or reveal nuanced relationships between the environment, mineralogy, and microbial communities.

3.3. WGS Metagenome Result

To understand the metabolic potential of these communities, we analyzed WGS data for 12 samples. As described above, the WGS analysis included several samples that were excluded from the 16S analysis due to low sequencing quality and read abundance. The WGS metagenomes were assembled from an average of 6,739,369 reads per sample, with an average of 6,276,358 reads assembled per sample.

First, we assessed the taxonomic composition of these samples using Kaiju (Menzel et al., 2016) (Figure 3). Archaea account for five of the identified phyla. The taxa we found represent a range of aerobic and anaerobic microbes. Many of these phyla are known to contain extremophilic members and members utilizing chemolithotrophy (Fujitani et al., 2020; Slobodkin et al., 2019; Thompson et al., 2017; Waite et al., 2020). Of note, all samples contain *Deinococcus radiodurans*, an extremophilic species known to withstand high levels of radiation, desiccation, and cold temperatures (Ho et al., 2016). On average, 83.1% of the samples could not be taxonomically assigned, suggesting abundant “microbial dark matter,” which is consistent with these being rare taxa and/or an understudied environment (Bowman, 2018; Marcy et al., 2007).

Next, we predicted an average of 122,381 high-quality ORFs per sample, 33% of which matched annotations in the Kyoto Encyclopedia of Genes and Genomes (KEGG), to identify a total of 10,550 unique KEGG orthologs (KOs). A high number (957,840) of ORFs could not be annotated using the EggNog Mapper (Cantalapiedra et al., 2021; Huerta-Cepas et al., 2019) (Figure 4). While this level of ambiguity is expected given the unique environment (Howe et al., 2014), the high proportion of ORFs that are not well characterized adds uncertainty to the functional composition of the microbial communities. Top KO categories include transcription; translation, ribosomal structure, and biogenesis; replication, recombination, and repair; and amino acid transport and metabolism. Additionally, in the dim and dark, there is less metabolic variance within samples than in the light, suggesting a narrower metabolic distribution than that of samples taken in direct light (Figure 4b). This is consistent with a more diverse community within the “light” samples and suggests that light enables greater breadth of metabolic capacity when the same or similar chemical nutrients are available to both light and dim/dark communities. Clustering of samples based on KOs does not show distinct groups, suggesting that there is little similarity in the annotated functional profiles and a high degree of niche partitioning (Figure 4a). Indeed, a permutational multivariate analysis of variance analysis revealed a trend in the abundance of the functional categories between light and dim and dark samples (group together) ($p = 0.075$). Still, given the small sample set and high number of unannotated ORFs, this analysis could fail to capture relationships or variation between samples based on light, mineralogy, or community structure.

We hypothesized that the high availability of sulfur, coupled with the low light and oligotrophic conditions of the lava tube, would drive the microbial communities toward sulfur metabolisms that could be used for energy production. We investigated sulfur metabolisms using a gene-centric approach where metagenomic query

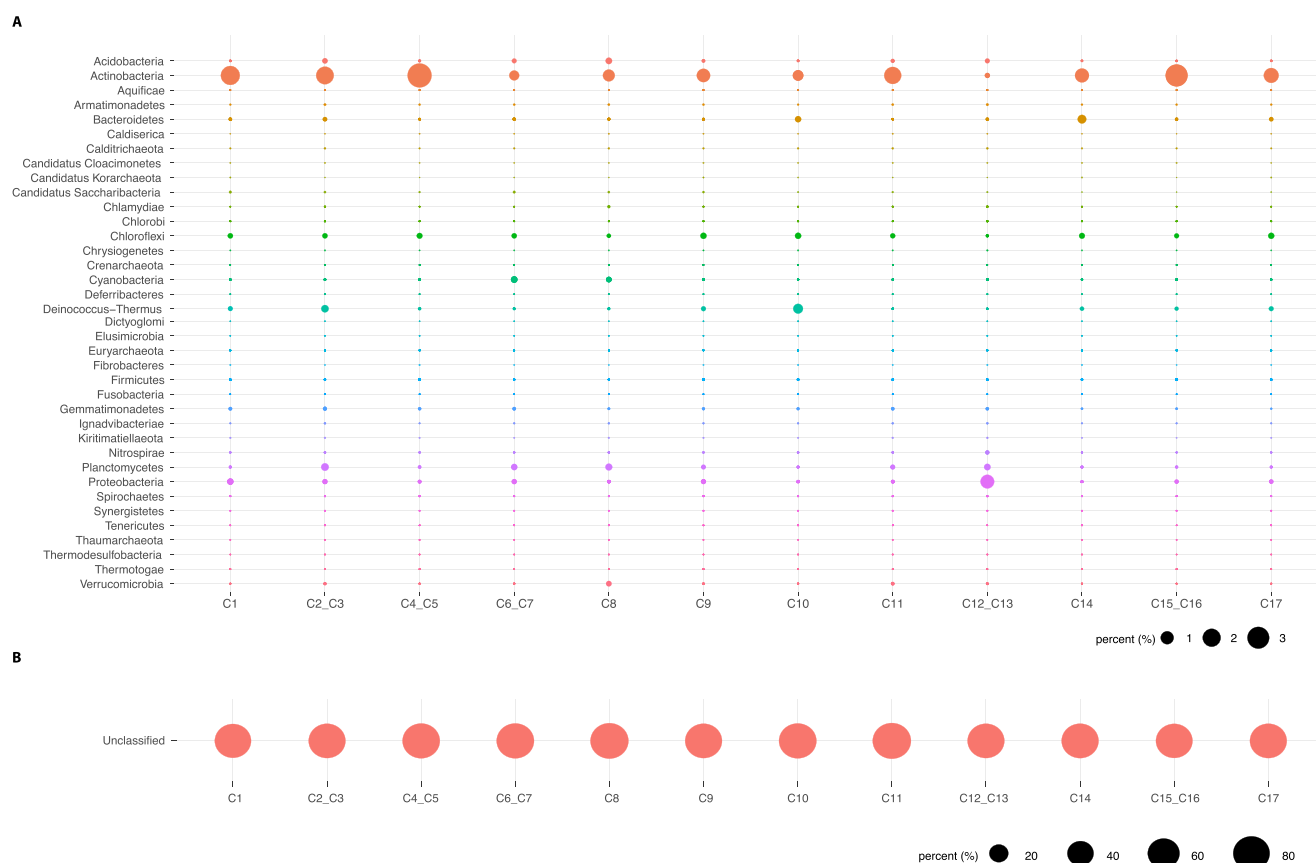


Figure 3. (a) Taxonomic annotation of classified reads from the whole genome shotgun sequencing from 12 microbial samples taken along a lava tube transect. (b) Unclassified reads from the 12 samples plotted separately to enable better visualization of the differences among the classified reads at the phylum level.

sequences were placed onto reference trees of proteins mediating most steps in assimilatory and dissimilatory sulfate reduction, and SOX. Hundreds of hits to *cysD*, *cysN*, *cysH*, *cysI*, and *cysJ* indicate that the potential for assimilatory sulfate reduction, a pathway for the production of biomolecules containing sulfur, is widespread in our samples (Figure 5a) (Kushkevych et al., 2020). The potential for dissimilatory sulfate reduction, a pathway by which anaerobes can use sulfate as a terminal electron acceptor, primarily represented by 583 hits to DsrA and DsrB proteins, was also encoded by most communities (Figure 5b) (Marietou et al., 2018). A signal was found for the SOX system, a pathway that can be used for SOX by some chemolithotrophs, with 402 hits to the *soxB* reference gene and 24 hits each to *soxY* and *soxZ* (Figure 5c) (Grabarczyk & Berks, 2017). However, the signal for sulfur metabolism genes, though present, was weak overall in these samples, suggesting that it is not a primary driver of metabolism in the community. This low level of sulfur metabolism may be a result of the availability of other sources of energy and the oxygen in the lava tube. Because of these conditions, there may not be environmental pressures to drive the system toward sulfur energy metabolism.

3.4. Metagenome Assembled Genomes

Because Mauna Loa lava tubes have not previously been characterized with genomics, they are of particular interest for the identification of novel taxa and their functional potential. To that end, we investigated many of the poorly characterized taxa present in our samples via the assembly of MAGs, which allow us to directly link taxonomy and functional potential to reveal the mechanisms of survival in these environments.

We assembled 72 high- (>90% completeness and <5% contamination)- and medium-quality (>50% completeness and <10% contamination) MAGs from the WGS metagenome data (Table S2 in Supporting Information S1). Thirty-eight MAGs were high-quality (Bowers et al., 2017) and 34 were identified as medium-quality. MAG completeness ranged from 57.27% to 100%, with an average of 89.31% \pm 10.09% (SD). Contamination ranged

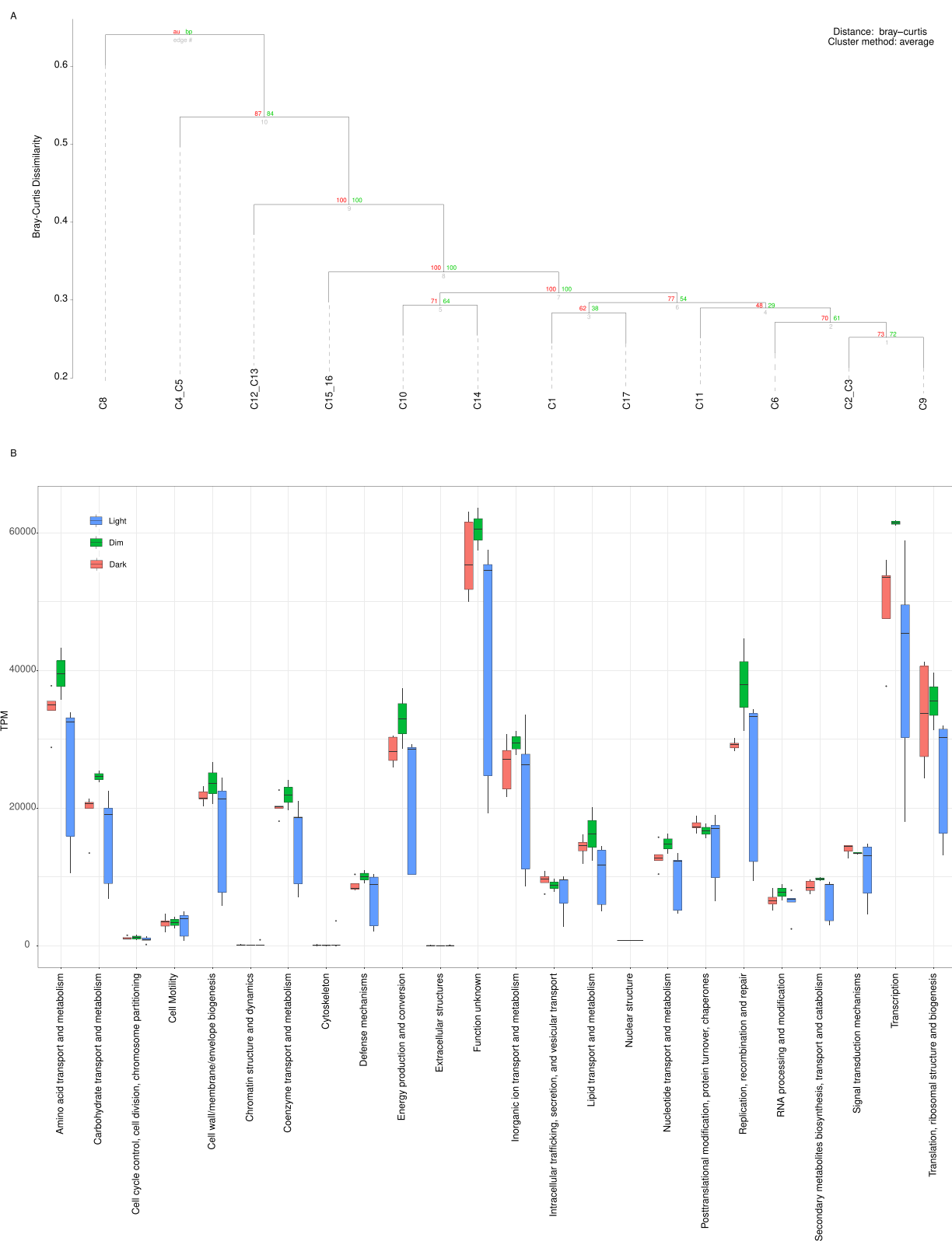


Figure 4. (a) Dendrogram of the beta diversity of Kyoto Encyclopedia of Genes and Genomes (KEGG) orthologs (KOs) found within the microbial communities along the lava tube. (b) Boxplots of KEGG categories for all samples with whole genome shotgun data along the lava tube, annotated by light condition.

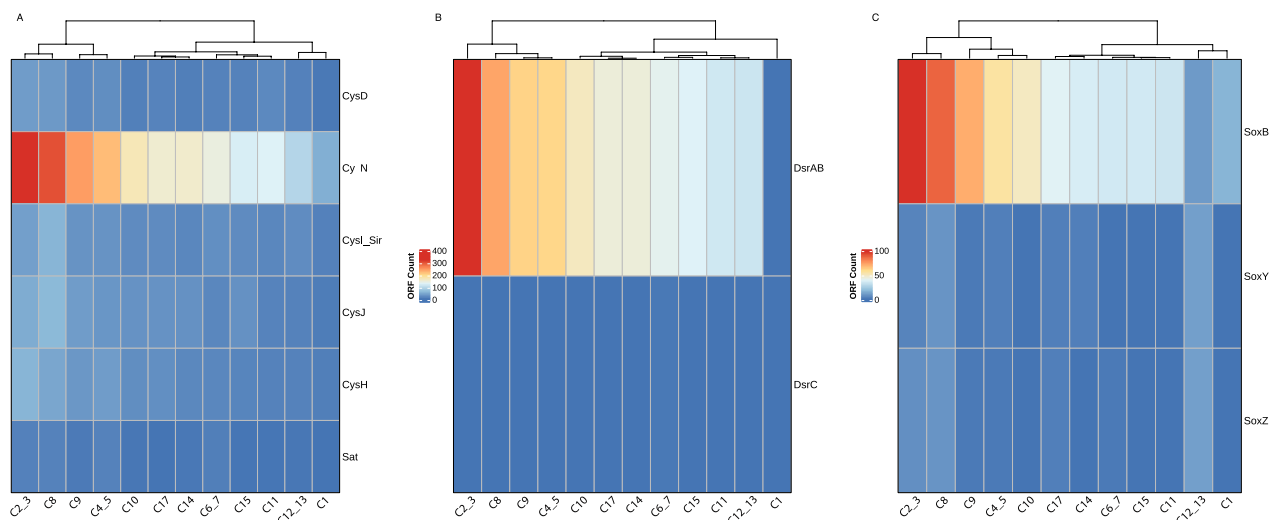


Figure 5. Abundance of representative genes for assimilatory (a) and dissimilatory sulfate reduction (b), and sulfur oxidation pathways (c). The sample dendrogram was generated by hierarchical cluster analysis with Euclidean distances. Heatmaps were generated with the ComplexHeatmap R package (Gu et al., 2016).

from 0.0% to 9.73% with an average of 2.71% \pm 2.38% (SD). The average N50, a statistic that represents the length of the shortest read within an assembly above which all larger sequences together make up 50% of the nucleotides in the assembly, was 21,325.5 (\pm 20,601.5). We predicted an average of 3,591 ORFs per MAG. All MAGs could be annotated to at least the phylum level. Dominant phyla among the MAGs were *Actinobacteriota* (51.39%), *Chloroflexi* (12.5%), *Proteobacteria* (9.72%) and *Acidobacteria* (6.94%) (Figure S1 in Supporting Information S1). This reflects the high-level taxonomic composition observed in the 16S SSU rRNA analyses, suggesting that the MAGs are a good representation of the overall community. Several MAGs represented previously uncharacterized taxa, including putative novel taxa within *Solirubrobacteraceae* and *Gaiellaceae* (Fishman et al., 2022; Gadson et al., 2022).

Next, we assessed whether MAGs were found in multiple samples by mapping the reads from each sample to each MAG using Burrows-Wheeler Aligner, a tool that maps low-divergent sequences against a reference, such as an assembly (Li, 2013) and normalizing the recruitment over 1000 bp regions to calculate median coverage (Figure 6). Akin to both the 16S SSU rRNA analyses and the KEGG-based functional profiling, we find very little similarity among samples, despite some samples being collected within centimeters of one another, suggesting once again a high degree of niche partitioning among samples. Only 22.2% of the 72 MAGs identified were found in six or more of our 12 samples, and only two MAGs were found in at least 75% of the samples (9/12 samples). These two MAGs were both *Actinobacteriota*: one was annotated to the family *SIRW01*, and one was annotated to the family *Rubrobacteraceae*. This result is consistent with the 16S rRNA gene analysis, which found that *Actinobacteria* was the

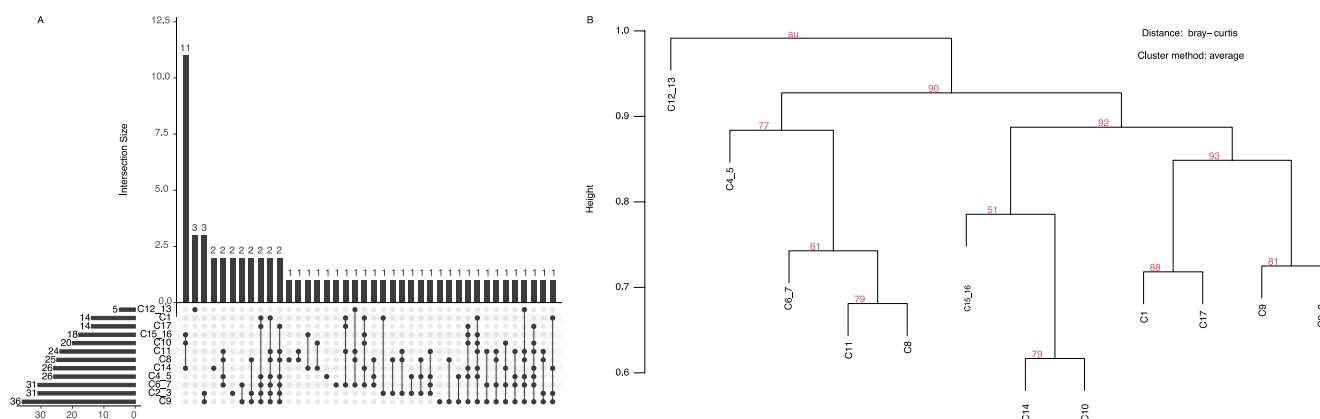


Figure 6. (a) UpSet plots of metagenome-assembled genomes (MAGs) shared among samples. (b) Beta diversity of MAGs across samples based on median abundance of all 72 high- and medium-quality MAGs.

dominant phylum in most samples. 15.3% of the MAGs appeared in only a single sample. 56.9% of MAGs were found in three samples or fewer (present in 25% of samples or less). Following clustering of the median coverage of all 72 MAGs across the 12 samples, we found three clusters (Figure 6a). This clustering is very similar to the 16S clustering (Figure 2), with the exception of C8 now closer to C4_5 and C6_7. Finally, we assessed pathway-level similarity among MAGs and samples (Figure S2 in Supporting Information S1). As many of the ORFs within these samples could not be annotated, we predicted pathways in order to assess higher level functions among the MAGs (Figure 6b) (Oltvai & Barabási, 2002). A total of 669 unique pathways were identified within the 72 high- and medium-quality MAGs. We found that in general the MAGs are clustered by taxonomy rather than sample of origin, suggesting that taxa of similar lineage have similar metabolic function (Figure S2 in Supporting Information S1). While MAGs represent a subset of the community, this suggests that while taxa are distributed among samples and there are distinct spatial distribution patterns, each community may have the ability to cycle nutrients and function independently.

4. Conclusions

4.1. Habitability and Novelty of Lava Tubes

The punishing radiation on the surface of Mars may be prohibitive of life that relies on sunlight for energy. Instead of such phototrophs, Mars may be more likely to sustain chemotrophs that thrive in the protected subsurface. To sustain this metabolism on Earth, organic molecules are the most common source of electron donors. On Mars, where organic molecules are scarce, molecular hydrogen that is produced through the serpentinization of volcanic rock may be an available electron donor in the subsurface (Schulte et al., 2006). Notably, sample C12_13 clusters separately in the hierarchical analysis of ASVs (Figure 2), and this is the only sample predominantly constituted of olivine, plagioclase, and pyroxene (Table S1 in Supporting Information S1). These are primary minerals, whereas the minerals dominant in other samples (gypsum, calcite, thenardite) form secondary alteration products. The fact that C12_13 is dominated by primary basaltic minerals indicates that it is a sample of the volcanic rock foundation of the lava tube as opposed to a secondary deposit. These minerals may create a unique niche that is more suitable for certain microbial metabolisms that drive this difference and that could potentially function in a Martian lava tube. Prescott et al., 2022 also found extreme niche partitioning in other Hawaiian lava tubes with only 16S rRNA SSU sequencing, which they attribute to the competitive exclusion based on low nutrient availability in an oligotrophic environment.

The functional analysis also illustrates a wide array of metabolisms utilized by microbes in the tubes, with significant differences between samples collected in light versus those collected in dim light and darkness (Figure 4). Taxonomic and functional diversity can be linked to habitability, as a habitable environment with more diverse niches may be more likely to support a higher diversity of taxa and/or a wider set of functional capabilities that are specialized for various niches (Cockell et al., 2016). The moderate diversity found here (Table 1) suggests a moderate level of habitability in the lava tube despite it being a unique environment. *Actinobacteria*, *Proteobacteria* and *Chloroflexi* are the dominant phyla within these samples. This is consistent with studies of lava tubes in Spain (Gonzalez-Pimentel et al., 2018), lava tube ice in California (O'Connor et al., 2021), and several other studies conducted in Hawaii (Hathaway, 2010; Northup et al., 2011; Teehera et al., 2018). Similarly, poorly characterized phyla, including BRC1, have been found in other lava tube studies in Hawaii and New Mexico (Northup et al., 2011). While little is known about this candidate division, early work reconstructing draft genomes suggests that this phylum is widely distributed in oxic and anoxic, often extreme, environments (Kadnikov et al., 2019). Within the WGS data, over 80% of the samples could be classified as microbial dark matter (Figure 3). Still, all samples contain *Deinococcus radiodurans*, an extremophilic species which can tolerate conditions of desiccation and radiation, as well as low temperatures (Ho et al., 2016). Although these traits are not necessary for life on Mauna Loa, microbes with similar survival mechanisms could inhabit the Martian subsurface.

Finally, while the MAGs represent a subset of the total community, the similarity to the high-level taxonomic composition observed in the 16S SSU rRNA analyses suggests that the MAGs are a good representation of the overall community. Deeper functional analysis of the ORFs and pathways within these 72 high- and medium-quality genomes may lend further insight into the adaptation necessary for thriving within these lava tube environments.

4.2. Drivers of Niche Partitioning

Each sample along the tubes hosted a unique microbial community (Figure 2). This taxonomic diversity among samples, combined with variation in genomic potential (Figures 4 and 6), suggests that communities are well-adapted to the specific minerals and conditions on which they live, thereby creating distinct spatial

distribution patterns. Such niche partitioning may help ensure environmental change and disturbance have disparate effects across community members, thereby tempering the effects of localized changes and enhancing the resiliency of life within the broader lava tube environment. Indeed, few ASVs and MAGs were shared among samples, and there was little similarity among functional annotations within the community or within pathways of the MAGs. Previous work found that weathered rocks have closely associated microbes adapted to the specific niche of the mineral composition (Riquelme et al., 2015), consistent with the community level differences observed across the samples. It is also possible that colonizing microbes can serve as loci for mineral precipitation or sediment accumulation (Van Driessche et al., 2019). Within the dark and dim samples taken in the lava tube, there is generally less metabolic variance across the samples, suggesting there is a narrower metabolic distribution under these conditions compared to samples under the light (Table 1). This may be due to light enabling greater breadth of metabolism, such as the ability to perform photosynthesis. Historical evidence suggests that the Mauna Loa lava tubes sustain a sulfur flux from the nearby active volcano. Sulfur cycling is of particular astrobiological interest as Martian surface and meteorite analyses indicate a sulfur-rich lithosphere, which could provide redox reagents for sulfur-metabolizing microbes (Franz et al., 2018). Without a Martian atmosphere to provide oxygen and protection from UV radiation, selective pressures may favor microbes that can subsist in the dark anaerobically using sulfur for energy. If these microbes were present at a low level when Mars was more Earth-like, they would likely become dominant in the community as Mars shifted to harsher conditions. In this lava tube, members of these microbial communities can oxidize and reduce sulfur ions (Figure 5). Large evolutionary distances separated the majority of the annotated protein sequences from references, suggesting uncharacterized organisms were facilitating these pathways in lava tubes. Although not every gene from the main sulfur metabolism pathways was represented, these findings are a promising start for future investigation of sulfur cycling in these lava tube environments. Nonetheless, it is important to note that with metagenomic analyses alone, we cannot determine whether these pathways are currently being utilized/expressed, particularly given the aerobic conditions of the lava tube as well as the presence of other chemical nutrients. However, the presence of phyla containing sulfur-reducing bacteria (Figure 3) shows that some microbes are utilizing these pathways. Previous terrestrial studies have also shown that sulfates like gypsum can also preserve biosignatures for geologically long periods (billions of years), making them a doubly promising target for astrobiological investigation (Allwood et al., 2013; Aubrey et al., 2006).

4.3. Exploration of Extreme Environments

In applying our findings to potential future exploration of similar environments on Earth or on Mars, studies could compare surface soil samples with samples from the subsurface to investigate which microbes found in these samples are unique to the subsurface or present in the general environment. This analysis could be paired with water analyses to investigate how microbes could be transported via water flows into the lava tubes over time. These sampling strategies each have implications for our understanding of whether life could have emerged on the early, more habitable Martian surface and subsequently been transported to the subsurface, potentially allowing it to survive the shift to harsher environmental conditions (Westall et al., 2015). Due to the niche partitioning we observed in this environment, further studies in this lava tube or similar environments should account for differences in the communities inhabiting primary versus secondary minerals.

Additionally, including RNA sequencing for transcriptomics would reveal the differences between an organism's metabolic potential and the metabolic functions that the organism actively utilizes. This comparison would allow investigations to identify which genes, and the functions they confer, might be necessary for survival within the lava tubes. For example, the upregulation of DNA repair genes across communities sampled under high light exposure conditions could indicate radiation tolerance (Li et al., 2015; Rainey et al., 2005).

Microbiological data could also be analyzed in conjunction with additional mineralogical and geochemical data, as well as organic composition through gas chromatography-mass spectroscopy and Raman spectroscopy. This would enable the definition of mineralogical niches, indicate bioavailable nutrients, and suggest the potential for biosignature preservation in this environment. Linking types of Mars-relevant microbial life, including sulfur-reducing microbes, to geological features will enable targeting of Martian exploration to geologically similar sites that may have once hosted life. Additionally, among the minerals identified as supportive to microbial life, analyses can inform which are most likely to preserve biosignatures of past life, and thus target exploration to those minerals. Further, some mineralogic weathering is of biotic origin (Barton et al., 2001), and understanding potential microbe-mineralogical interactions could help map mineralogical features to the presence of microbes.

4.4. Summary

Mauna Loa lava tubes represent a novel and understudied source for examining microbial life and represent a unique analog for studying the possibility of past and/or present life on Mars. Both 16S SSU rRNA sequencing and WGS sequencing revealed a plethora of microbial taxa thriving in this unique environment. We found evidence of high levels of “microbial dark matter” in both taxonomic and functional analyses of the communities. We also discovered that despite proximity in a protected environment, there was a high degree of niche partitioning, suggesting that mineralogical or fine-scale environmental variations drive community composition in the lava tube. Ultimately, our findings indicate that careful sampling of distinct and specific mineralogical features must be conducted to adequately assess the possibility of life on Mars.

Data Availability Statement

All 16S rRNA gene sequences have been deposited in the FigShare repository (Hahn, 2022a). Both forward (_R1) and reverse (_R2) reads are available. WGS reads have also been deposited in the FigShare repository (Hahn, 2023). Metagenome Assembled Genomes have been deposited in the FigShare repository (Hahn, 2022b).

Acknowledgments

Support for this research was provided by NASA's PSD Research Program through the GSFC ISFM Goddard Instrument Field Team program and the GSFC ISFM Fundamental Laboratory Research program. We acknowledge that the sample collection for this research was conducted at Mauna Loa on the Island of Hawaii, the homeland stewarded by Native Hawaiians. This land was accessible to us for investigation because it was ceded by the Hawaiian Kingdom to the United States through coercion. The knowledge we gain from our study of life in this pristine lava tube is a testament to the long history of care by Native Hawaiians for this land.

References

- Allwood, A. C., Burch, I. W., Rouchy, J. M., & Coleman, M. (2013). Morphological biosignatures in gypsum: Diverse formation processes of messinian (~6.0 Ma) gypsum stromatolites. *Astrobiology*, 13(9), 870–886. <https://doi.org/10.1089/ast.2013.1021>
- Aubrey, A., Cleaves, H. J., Chalmers, J. H., Skelley, A. M., Mathies, R. A., Grunthaner, F. J., et al. (2006). Sulfate minerals and organic compounds on Mars. *Geology*, 34(5), 357. <https://doi.org/10.1130/G22316.1>
- Baker, V. R., Strom, R. G., Gulick, V. C., Kargel, J. S., Komatsu, G., & Kale, V. S. (1991). Ancient oceans, ice sheets and the hydrological cycle on Mars. *Nature*, 352(6336), 589–594. <https://doi.org/10.1038/352589a0>
- Barton, H. A., Spear, J. R., & Pace, N. R. (2001). Microbial life in the underworld: Biogenicity in secondary mineral formations. *Geomicrobiology Journal*, 18(3), 359–368. <https://doi.org/10.1080/01490450152467840>
- Benison, K. C., & Karmanocky, F. J. (2014). Could microorganisms be preserved in Mars gypsum? Insights from terrestrial examples. *Geology*, 42(7), 615–618. <https://doi.org/10.1130/G35542.1>
- Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, 30(15), 2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>
- Boston, P. J., Spilde, M. N., Northup, D. E., Melim, L. A., Soroka, D. S., Kleina, L. G., et al. (2001). Cave biosignature suites: Microbes, minerals, and Mars. *Astrobiology*, 1(1), 25–55. <https://doi.org/10.1089/153110701750137413>
- Bowers, R. M., Kyrpides, N. C., Stepanauskas, R., Harmon-Smith, M., Doud, D., Reddy, T. B. K., et al. (2017). Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. *Nature Biotechnology*, 35(8), 725–731. <https://doi.org/10.1038/nbt.3893>
- Bowman, J. S. (2018). Identification of microbial dark matter in Antarctic environments. *Frontiers in Microbiology*, 9, 3165. <https://doi.org/10.3389/fmicb.2018.03165>
- Bushnell, B. (2019). BBTools. Retrieved from <https://sourceforge.net/projects/bbmap/>
- Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J. A., & Holmes, S. P. (2016). DADA2: High-resolution sample inference from Illumina amplicon data. *Nature Methods*, 13(7), 581–583. <https://doi.org/10.1038/nmeth.3869>
- Cantalapiedra, C. P., Hernández-Plaza, A., Letunic, I., Bork, P., & Huerta-Cepas, J. (2021). eggNOG-mapper v2: Functional annotation, orthology assignments, and domain prediction at the metagenomic scale. *Molecular Biology and Evolution*, 38(12), 5825–5829. <https://doi.org/10.1093/molbev/msab293>
- Caporaso, J. G., Lauber, C. L., Walters, W. A., Berg-Lyons, D., Lozupone, C. A., Turnbaugh, P. J., et al. (2011). Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proceedings of the National Academy of Sciences of the United States America*, 108(1), 4516–4522. <https://doi.org/10.1073/pnas.100080107>
- Carr, M. H., & Head, J. W. (2010). Geologic history of Mars. *Earth and Planetary Science Letters*, 294(3–4), 185–203. <https://doi.org/10.1016/j.epsl.2009.06.042>
- Carter, J., Loizeau, D., Mangold, N., Poulet, F., & Bibring, J.-P. (2015). Widespread surface weathering on early Mars: A case for a warmer and wetter climate. *Icarus*, 248, 373–382. <https://doi.org/10.1016/j.icarus.2014.11.011>
- Caspi, R., Billington, R., Keseler, I. M., Kothari, A., Krummenacker, M., Midford, P. E., et al. (2020). The MetaCyc database of metabolic pathways and enzymes—A 2019 update. *Nucleic Acids Research*, 48(D1), D445–D453. <https://doi.org/10.1093/nar/gkz862>
- Chaumeil, P.-A., Mussig, A. J., Hugenholtz, P., & Parks, D. H. (2019). GTDB-Tk: A toolkit to classify genomes with the genome taxonomy database. *Bioinformatics*, btz848. <https://doi.org/10.1093/bioinformatics/btz848>
- Cockell, C. S., Bush, T., Bryce, C., Direito, S., Fox-Powell, M., Harrison, J. P., et al. (2016). Habitability: A review. *Astrobiology*, 16(1), 89–117. <https://doi.org/10.1089/ast.2015.1295>
- Conway, J. R., Lex, A., & Gehlenborg, N. (2017). UpSetR: An R package for the visualization of intersecting sets and their properties. *Bioinformatics*, 33(18), 2938–2940. <https://doi.org/10.1093/bioinformatics/btx364>
- Cushing, G. E., Titus, T. N., Wynne, J. J., & Christensen, P. R. (2007). THEMIS observes possible cave skylights on Mars. *Geophysical Research Letters*, 34(17), L17201. <https://doi.org/10.1029/2007GL030709>
- Delcher, A. L., Salzberg, S. L., & Phillippy, A. M. (2003). Using MUMmer to identify similar regions in large sequence sets. *Current Protocols in Bioinformatics*, 00(1). <https://doi.org/10.1002/0471250953.bi1003s00>
- Douglas, G. M., Maffei, V. J., Zaneveld, J. R., Yurgel, S. N., Brown, J. R., Taylor, C. M., et al. (2020). PICRUSt2 for prediction of metagenome functions. *Nature Biotechnology*, 38(6), 685–688. <https://doi.org/10.1038/s41587-020-0548-6>
- Fishman, C. B., Bevilacqua, J. G., Gadson, O., Hahn, A. S., McAdam, A. C., Bleacher, J., & Johnson, S. S. (2022). Basaltic lava tubes host a putative novel genus in the family Solirubrobacteraceae. *Microbial Resource Announcements*, 11(10). <https://doi.org/10.1128/mra.00499-22>

- Franz, H. B., King, P. L., & Gaillard, F. (2018). Sulfur on Mars from the atmosphere to the core. In J. Filiberto & S. P. Schwenzer (Eds.), *Volatiles in the Martian crust* (pp. 119–183). Elsevier.
- Fujitani, H., Momouchi, K., Ishii, K., Nomachi, M., Kikuchi, S., Ushiki, N., et al. (2020). Genomic and physiological characteristics of a novel nitrite-oxidizing Nitrospira strain isolated from a drinking water treatment plant. *Frontiers in Microbiology*, 11, 545190. <https://doi.org/10.3389/fmicb.2020.545190>
- Gadson, O., Bevilacqua, J. G., Fishman, C. B., Hahn, A. S., McAdam, A. C., Bleacher, J., & Johnson, S. S. (2022). Metagenome-assembled genome of a putative chemoheterotroph from volcanic terrain in Hawaii. *Microbiology Resource Announcements*, 11(11). <https://doi.org/10.1128/mra.00556-22>
- Gonzalez-Pimentel, J. L., Miller, A. Z., Jurado, V., Laiz, L., Pereira, M. F. C., & Saiz-Jimenez, C. (2018). Yellow coloured mats from lava tubes of La Palma (Canary Islands, Spain) are dominated by metabolically active Actinobacteria. *Scientific Reports*, 8(1), 1944. <https://doi.org/10.1038/s41598-018-20393-2>
- Grabarczyk, D. B., & Berks, B. C. (2017). Intermediates in the Sox sulfur oxidation pathway are bound to a sulfane conjugate of the carrier protein SoxYZ. *PLoS One*, 12(3), e0173395. <https://doi.org/10.1371/journal.pone.0173395>
- Gu, Z., Eils, R., & Schlesner, M. (2016). Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics*, 32(18), 2847–2849. <https://doi.org/10.1093/bioinformatics/btw313>
- Hahn, A. (2022a). 16S SSU rRNA sample [Dataset]. Figshare. <https://doi.org/10.6084/m9.figshare.19253372.v1>
- Hahn, A. (2022b). Metagenome assembled genomes [Dataset]. Figshare. <https://doi.org/10.6084/m9.figshare.19253363>
- Hahn, A. (2023). WGS data [Dataset]. figshare. <https://doi.org/10.6084/M9.FIGSHARE.22696237>
- Hathaway, J. J. M. (2010). *Molecular phylogenetic investigation of microbial diversity and nitrogen cycling in lava tubes*. University of New Mexico Retrieved from https://digitalrepository.unm.edu/biol_etds/48/
- Ho, J., Adeolu, M., Khadka, B., & Gupta, R. S. (2016). Identification of distinctive molecular traits that are characteristic of the phylum “Deinococcus-Thermus” and distinguish its main constituent groups. *Systematic & Applied Microbiology*, 39(7), 453–463. <https://doi.org/10.1016/j.syapm.2016.07.003>
- Howe, A. C., Jansson, J. K., Malfatti, S. A., Tringe, S. G., Tiedje, J. M., & Brown, C. T. (2014). Tackling soil diversity with the assembly of large, complex metagenomes. *Proceedings of the National Academy of Sciences*, 111(13), 4904–4909. <https://doi.org/10.1073/pnas.1402564111>
- Huang, W., Ertekin, E., Wang, T., Cruz, L., Dailey, M., DiRuggiero, J., & Kisailus, D. (2020). Mechanism of water extraction from gypsum rock by desert colonizing microorganisms. *Proceedings of the National Academy of Sciences of the United States of America*, 117(20), 10681–10687. <https://doi.org/10.1073/pnas.2001613117>
- Huerta-Cepas, J., Szklarczyk, D., Heller, D., Hernández-Plaza, A., Forslund, S. K., Cook, H., et al. (2019). eggNOG 5.0: A hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Research*, 47(D1), D309–D314. <https://doi.org/10.1093/nar/gky1085>
- Hughes, S. S., Haberle, C. W., Kobs Nawotniak, S. E., Sehlke, A., Garry, W. B., Elphic, R. C., et al. (2019). Basaltic terrains in Idaho and Hawai‘i as planetary analogs for Mars geology and astrobiology. *Astrobiology*, 19(3), 260–283. <https://doi.org/10.1089/ast.2018.1847>
- Hyatt, D., LoCascio, P. F., Hauser, L. J., & Ueberbacher, E. C. (2012). Gene and translation initiation site prediction in metagenomic sequences. *Bioinformatics*, 28(17), 2223–2230. <https://doi.org/10.1093/bioinformatics/bts429>
- Jakosky, B. M., Slipski, M., Benna, M., Mahaffy, P., Elrod, M., Yelle, R., et al. (2017). Mars’ atmospheric history derived from upper-atmosphere measurements of $^{38}\text{Ar}/^{36}\text{Ar}$. *Science*, 355(6332), 1408–1410. <https://doi.org/10.1126/science.aai7721>
- Kadnikov, V. V., Mardanov, A. V., Beletsky, A. V., Rakitin, A. L., Frank, Y. A., Karnachuk, O. V., & Ravin, N. V. (2019). Phylogeny and physiology of candidate phylum BRC1 inferred from the first complete metagenome-assembled genome obtained from deep subsurface aquifer. *Systematic & Applied Microbiology*, 42(1), 67–76. <https://doi.org/10.1016/j.syapm.2018.08.013>
- Kah, L. C., Stack, K. M., Eigenbrode, J. L., Yingst, R. A., & Edgett, K. S. (2018). Syndepositional precipitation of calcium sulfate in Gale Crater, Mars. *Terra Nova*, 30(6), 431–439. <https://doi.org/10.1111/ter.12359>
- Kang, D. D., Li, F., Kirtan, E., Thomas, A., Egan, R., An, H., & Wang, Z. (2019). MetaBAT 2: An adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. *PeerJ*, 7, e7359. <https://doi.org/10.7717/peerj.7359>
- Karp, P. D., Riley, M., Paley, S. M., & Pellegrini-Toole, A. (2002). The MetaCyc database. *Nucleic Acids Research*, 30(1), 59–61. <https://doi.org/10.1093/nar/30.1.59>
- Kieser, S., Brown, J., Zdobnov, E. M., Trajkovski, M., & McCue, L. A. (2020). ATLAS: A snakemake workflow for assembly, annotation, and genomic binning of metagenome sequence data. *BMC Bioinformatics*, 21(1), 257. <https://doi.org/10.1186/s12859-020-03585-4>
- Kim, D., Hahn, A. S., Hanson, N. W., Konwar, K. M., & Hallam, S. J. (2016). FAST: Fast annotation with synchronized threads. In *2016 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB)*. IEEE.
- King, G. M., & Nanba, K. (2008). Distribution of atmospheric methane oxidation and methanotrophic communities on Hawaiian volcanic deposits and soils. *Microbes and Environments*, 23(4), 326–330. <https://doi.org/10.1264/jsm.2.ME08529>
- Konwar, K. M., Hanson, N. W., Bhatia, M. P., Kim, D., Wu, S.-J., Hahn, A. S., et al. (2015). MetaPathways v2.5: Quantitative functional, taxonomic and usability improvements. *Bioinformatics*, 31(20), 3345–3347. <https://doi.org/10.1093/bioinformatics/btv361>
- Kushkevych, I., Abdulina, D., Kováč, J., Dordević, D., Vítězová, M., Iutynska, G., & Rittmann, S. K. M. (2020). Adenosine-5′-phosphosulfate- and sulfite reductases activities of sulfate-reducing bacteria from various environments. *Biomolecules*, 10(6), 921. <https://doi.org/10.3390/biom10060921>
- Léveillé, R. J., & Datta, S. (2010). Lava tubes and basaltic caves as astrobiological targets on Earth and Mars: A review. *Planetary and Space Science*, 58(4), 592–598. <https://doi.org/10.1016/j.pss.2009.06.004>
- Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. ArXiv:1303.3997 [q-Bio] Retrieved from <http://arxiv.org/abs/1303.3997>
- Li, L., Chen, Z., Ding, X., Shan, Z., Liu, L., & Guo, J. (2015). Deep sequencing analysis of the Kineococcus radiotolerans transcriptome in response to ionizing radiation. *Microbiological Research*, 170, 248–254. <https://doi.org/10.1016/j.micres.2014.10.003>
- Lockwood, J. P., & Lipman, P. W. (1987). Holocene eruptive history of Mauna Loa volcano. *US Geological Survey Professional Papers*, 1350(1), 509–535.
- Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P. M., & Henrissat, B. (2014). The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Research*, 42(D1), D490–D495. <https://doi.org/10.1093/nar/gkt1178>
- Lozupone, C., & Knight, R. (2005). UniFrac: A new phylogenetic method for comparing microbial communities. *Applied and Environmental Microbiology*, 71(12), 8228–8235. <https://doi.org/10.1128/AEM.71.12.8228-8235.2005>
- Marçais, G., Delcher, A. L., Phillippy, A. M., Coston, R., Salzberg, S. L., & Zimin, A. (2018). MUMmer4: A fast and versatile genome alignment system. *PLoS Computational Biology*, 14(1), e1005944. <https://doi.org/10.1371/journal.pcbi.1005944>

- Marcy, Y., Ouverney, C., Bik, E. M., Losekann, T., Ivanova, N., Martin, H. G., et al. (2007). Dissecting biological “dark matter” with single-cell genetic analysis of rare and uncultivated TM7 microbes from the human mouth. *Proceedings of the National Academy of Sciences of the United States of America*, 104(29), 11889–11894. <https://doi.org/10.1073/pnas.0704662104>
- Marietou, A., Røy, H., Jørgensen, B. B., & Kjeldsen, K. U. (2018). Sulfate transporters in dissimilatory sulfate reducing microorganisms: A comparative genomics analysis. *Frontiers in Microbiology*, 9, 309. <https://doi.org/10.3389/fmicb.2018.00309>
- McMurdie, P. J., & Holmes, S. (2013). phyloseq: An R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS One*, 8(4), e61217. <https://doi.org/10.1371/journal.pone.0061217>
- Menzel, P., Ng, K. L., & Krogh, A. (2016). Fast and sensitive taxonomic classification for metagenomics with Kaiju. *Nature Communications*, 7(1), 11257. <https://doi.org/10.1038/ncomms11257>
- Michalski, J. R., Onstott, T. C., Mojzsis, S. J., Mustard, J., Chan, Q. H. S., Niles, P. B., & Johnson, S. S. (2018). The Martian subsurface as a potential window into the origin of life. *Nature Geoscience*, 11(1), 21–26. <https://doi.org/10.1038/s41561-017-0015-2>
- Ming, D. W., Morris, R. V., & Clark, B. C. (2008). Aqueous alteration on Mars. In J. Bell (Ed.), *The Martian surface: Composition, mineralogy, and physical properties*. Cambridge University Press.
- Morgan-Lang, C., McLaughlin, R., Armstrong, Z., Zhang, G., Chan, K., & Hallam, S. J. (2020). TreeSAPP: The tree-based sensitive and accurate phylogenetic profiler. *Bioinformatics*, 36(18), 4706–4713. <https://doi.org/10.1093/bioinformatics/btaa588>
- Northup, D. E., Melim, L. A., Spilde, M. N., Hathaway, J. J. M., Garcia, M. G., Moya, M., et al. (2011). Lava cave microbial communities within mats and secondary mineral deposits: Implications for life detection on other planets. *Astrobiology*, 11(7), 601–618. <https://doi.org/10.1089/ast.2010.0562>
- Nurk, S., Meleshko, D., Korobeynikov, A., & Pevzner, P. A. (2017). metaSPAdes: A new versatile metagenomic assembler. *Genome Research*, 27(5), 824–834. <https://doi.org/10.1101/gr.213959.116>
- O'Connor, B. R. W., Fernández-Martínez, M. Á., Lévillé, R. J., & Whyte, L. G. (2021). Taxonomic characterization and microbial activity determination of cold-adapted microbial communities in lava tube ice caves from lava beds national monument, a high-fidelity Mars analogue environment. *Astrobiology*, 21(5), 613–627. <https://doi.org/10.1089/ast.2020.2327>
- O'Leary, N. A., Wright, M. W., Brister, J. R., Ciufu, S., Haddad, D., McVeigh, R., et al. (2016). Reference sequence (RefSeq) database at NCBI: Current status, taxonomic expansion, and functional annotation. *Nucleic Acids Research*, 44(D1), D733–D745. <https://doi.org/10.1093/nar/gkv1189>
- Olm, M. R., Brown, C. T., Brooks, B., & Banfield, J. F. (2017). dRep: A tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication. *The ISME Journal*, 11(12), 2864–2868. <https://doi.org/10.1038/ismej.2017.126>
- Oltvai, Z. N., & Barabási, A.-L. (2002). Life's complexity pyramid. *Science*, 298(5594), 763–764. <https://doi.org/10.1126/science.1078563>
- Ondov, B. D., Treangen, T. J., Melsted, P., Mallonee, A. B., Bergman, N. H., Koren, S., & Phillippy, A. M. (2016). Mash: Fast genome and metagenome distance estimation using MinHash. *Genome Biology*, 17(1), 132. <https://doi.org/10.1186/s13059-016-0997-x>
- Oschmann, W. (2000). Microbes and black shales. In R. E. Riding & S. M. Awramik (Eds.), *Microbial sediments* (pp. 137–148). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-662-04036-2_16
- Panieri, G., Lugli, S., Manzi, V., Roveri, M., Schreiber, B. C., & Palinska, K. A. (2010). Ribosomal RNA gene fragments from fossilized cyanobacteria identified in primary gypsum from the late Miocene, Italy. *Geobiology*, 8(2), 101–111. <https://doi.org/10.1111/j.1472-4669.2009.00230.x>
- Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P., & Tyson, G. W. (2015). CheckM: Assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Research*, 25(7), 1043–1055. <https://doi.org/10.1101/gr.186072.114>
- Prescott, R. D., Zamkovaya, T., Donachie, S. P., Northup, D. E., Medley, J. J., Monsalve, N., et al. (2022). Islands within islands: Bacterial phylogenetic structure and consortia in Hawaiian lava caves and fumaroles. *Frontiers in Microbiology*, 13, 934708. <https://doi.org/10.3389/fmicb.2022.934708>
- Rainey, F. A., Ray, K., Ferreira, M., Gatz, B. Z., Nobre, M. F., Bagaley, D., et al. (2005). Extensive diversity of ionizing-radiation-resistant bacteria recovered from sonoran desert soil and description of nine new species of the genus *Deinococcus* obtained from a single soil sample. *Applied and Environmental Microbiology*, 71(9), 5225–5235. <https://doi.org/10.1128/AEM.71.9.5225-5235.2005>
- Renggli, C. J., Palm, A. B., King, P. L., & Guagliardo, P. (2019). Implications of reactions between SO₂ and basaltic glasses for the mineralogy of planetary crusts. *Journal of Geophysical Research: Planets*, 124(10), 2563–2582. <https://doi.org/10.1029/2019JE006045>
- Riquelme, C., Marshall Hathaway, J. J., Enes Dapkevicius, M. L. N., Miller, A. Z., Kooser, A., Northup, D. E., et al. (2015). Actinobacterial diversity in volcanic caves and associated geomicrobiological interactions. *Frontiers in Microbiology*, 6. <https://doi.org/10.3389/fmicb.2015.01342>
- Sakai, H., Casadevall, T. J., & Moore, J. G. (1982). Chemistry and isotope ratios of sulfur in basalts and volcanic gases at Kilauea volcano, Hawaii. *Geochimica et Cosmochimica Acta*, 46(5), 729–738. [https://doi.org/10.1016/0016-7037\(82\)90024-2](https://doi.org/10.1016/0016-7037(82)90024-2)
- Schulte, M., Blake, D., Hoehler, T., & McCollom, T. (2006). Serpentinization and its implications for life on the early Earth and Mars. *Astrobiology*, 6(2), 364–376. <https://doi.org/10.1089/ast.2006.6.364>
- Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, 27(3), 379–423. <https://doi.org/10.1002/j.1538-7305.1948.tb01338.x>
- Siebert, C. (2017). *Dereplication, aggregation and scoring tool (DAS tool) v1.0*. Lawrence Berkeley National Laboratory (LBNL). <https://doi.org/10.11578/DC.20201001.19>
- Siebert, C. M. K., Probst, A. J., Sharrar, A., Thomas, B. C., Hess, M., Tringe, S. G., & Banfield, J. F. (2018). Recovery of genomes from metagenomes via a dereplication, aggregation and scoring strategy. *Nature Microbiology*, 3(7), 836–843. <https://doi.org/10.1038/s41564-018-0171-1>
- Slobodkin, A., Slobodkina, G., Allieux, M., Alain, K., Jebbar, M., Shadrin, V., et al. (2019). Genomic insights into the carbon and energy metabolism of a thermophilic deep-sea bacterium *Deferribacter autotrophicus* revealed new metabolic traits in the phylum *Deferribacteres*. *Genes*, 10(11), 849. <https://doi.org/10.3390/genes10110849>
- Steen, A. D., Crits-Christoph, A., Carini, P., DeAngelis, K. M., Fierer, N., Lloyd, K. G., & Cameron Thrash, J. (2019). High proportions of bacteria and archaea across most biomes remain uncultured. *The ISME Journal*, 13(12), 3126–3130. <https://doi.org/10.1038/s41396-019-0484-y>
- Suzuki, R., & Shimodaira, H. (2006). Pvcust: An R package for assessing the uncertainty in hierarchical clustering. *Bioinformatics*, 22(12), 1540–1542. <https://doi.org/10.1093/bioinformatics/btl117>
- Tang, M., Ehreiser, A., & Li, Y.-L. (2014). Gypsum in modern Kamchatka volcanic hot springs and the Lower Cambrian black shale: Applied to the microbial-mediated precipitation of sulfates on Mars. *American Mineralogist*, 99(10), 2126–2137. <https://doi.org/10.2138/am-2014-4754>
- Teehera, K. B., Jungbluth, S. P., Onac, B. P., Acosta-Maeda, T. E., Hellebrand, E., Misra, A. K., et al. (2018). Cryogenic minerals in Hawaiian lava tubes: A geochemical and microbiological exploration. *Geomicrobiology Journal*, 35(3), 227–241. <https://doi.org/10.1080/01490451.2017.1362079>
- The UniProt Consortium. (2019). UniProt: A worldwide hub of protein knowledge. *Nucleic Acids Research*, 47(D1), D506–D515. <https://doi.org/10.1093/nar/gky1049>

- Thompson, K. J., Simister, R. L., Hahn, A. S., Hallam, S. J., & Crowe, S. A. (2017). Nutrient acquisition and the metabolic potential of photoferrotrophic Chlorobi. *Frontiers in Microbiology*, 8, 1212. <https://doi.org/10.3389/fmicb.2017.01212>
- Tosca, N. J., McLennan, S. M., Clark, B. C., Grotzinger, J. P., Hurowitz, J. A., Knoll, A. H., et al. (2005). Geochemical modeling of evaporation processes on Mars: Insight from the sedimentary record at Meridiani Planum. *Earth and Planetary Science Letters*, 240(1), 122–148. <https://doi.org/10.1016/j.epsl.2005.09.042>
- United States Geological Survey. (2022). Mauna Loa: Geology and history. Retrieved from <https://www.usgs.gov/volcanoes/mauna-loa/geology-and-history>
- Van Driessche, A. E. S., Stawski, T. M., & Kellermeier, M. (2019). Calcium sulfate precipitation pathways in natural and engineered environments. *Chemical Geology*, 530, 119274. <https://doi.org/10.1016/j.chemgeo.2019.119274>
- Vaniman, D. T., Martínez, G. M., Rampe, E. B., Bristow, T. F., Blake, D. F., Yen, A. S., et al. (2018). Gypsum, bassanite, and anhydrite at Gale crater, Mars. *American Mineralogist*, 103(7), 1011–1020. <https://doi.org/10.2138/am-2018-6346>
- Waite, D. W., Chuvoshina, M., Pelikan, C., Parks, D. H., Yilmaz, P., Wagner, M., et al. (2020). Proposal to reclassify the proteobacterial classes Deltaproteobacteria and Oligoflexia, and the phylum Thermodesulfobacteria into four phyla reflecting major functional capabilities. *International Journal of Systematic and Evolutionary Microbiology*, 70(11), 5972–6016. <https://doi.org/10.1099/ijsem.0.004213>
- Westall, F., Foucher, F., Bost, N., Bertrand, M., Loizeau, D., Vago, J. L., et al. (2015). Biosignatures on Mars: What, where, and how? Implications for the search for martian life. *Astrobiology*, 15(11), 998–1029. <https://doi.org/10.1089/ast.2015.1374>
- White, W. (2010). Secondary minerals in volcanic caves: Data from Hawai'i. *Journal of Cave and Karst Studies*, 72(2), 75–85. <https://doi.org/10.4311/jcks2009es0080>
- Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis* (2nd ed.). Springer International Publishing. <https://doi.org/10.1007/978-3-319-24277-4>
- Wilson, S. A., & Bish, D. L. (2011). Formation of gypsum and bassanite by cation exchange reactions in the absence of free-liquid H₂O: Implications for Mars. *Journal of Geophysical Research*, 116(E9), E09010. <https://doi.org/10.1029/2011JE003853>
- Wooley, J. C., Godzik, A., & Friedberg, I. (2010). A primer on metagenomics. *PLoS Computational Biology*, 6(2), e1000667. <https://doi.org/10.1371/journal.pcbi.1000667>
- Wordsworth, R. D. (2016). The climate of early Mars. *Annual Review of Earth and Planetary Sciences*, 44(1), 381–408. <https://doi.org/10.1146/annurev-earth-060115-012355>
- Wu, Y.-W., Simmons, B. A., & Singer, S. W. (2016). MaxBin 2.0: An automated binning algorithm to recover genomes from multiple metagenomic datasets. *Bioinformatics*, 32(4), 605–607. <https://doi.org/10.1093/bioinformatics/btv638>
- Zolotov, M. Y., & Mironenko, M. V. (2016). Chemical models for martian weathering profiles: Insights into formation of layered phyllosilicate and sulfate deposits. *Icarus*, 275, 203–220. <https://doi.org/10.1016/j.icarus.2016.04.011>
- Zuber, M. T. (2001). The crust and mantle of Mars. *Nature*, 412(6843), 220–227. <https://doi.org/10.1038/35084163>