UC Davis
UC Davis Electronic Theses and Dissertations

Title

Microbial Biogeography of Basaltic Lava Caves

Permalink

https://escholarship.org/uc/item/86h3q0zb

ISBN

9798193266242

Author

Williams, Caden Caroline

Publication Date

2026-09-08

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

Microbial Biogeography of Basaltic Lava Caves

By

CADEN WILLIAMS

THESIS

Submitted in partial satisfaction of the requirements for the degree of

MASTER OF SCIENCE

in

Earth and Planetary Sciences

in the

OFFICE OF GRADUATE STUDIES

of the

UNIVERSITY OF CALIFORNIA

DAVIS

Approved:

Christen Grettenberger, Chair

Eliot Atekwana

Kari Cooper

Joanne Emerson

Committee in Charge

2026

1 Abstract

As lava cools, it provides new habitats and resources for life. Microbial colonization of newly exposed basalt contributes to weathering and ecosystem development, but we do not know which environmental and spatial factors structure these communities through time. To understand controls of microbial community assembly, we sampled 11 caves formed by basaltic lava flows spanning approximately 34 kyr across northern California and southern Idaho. In each cave, we sampled microbial biofilms and used 16S rRNA gene and internal transcribed spacer (ITS) region sequencing to identify bacterial and fungal communities, respectively. Here we show that microbial community composition in basaltic lava caves is structured hierarchically across spatial scales, with individual cave explaining more variation than lava flow age or biofilm color, while caves within the same lava flow are more similar to one another than caves from different flows. The strong influence of individual cave, together with greater similarity among caves sharing a lava flow, indicates that local spatial structure and shared flow history constrain community assembly. Community composition, particularly in fungi, varied with lava flow age despite no relationship between age and alpha diversity, suggesting taxonomic turnover through time. Most taxa were consistent with neutral model predictions, while stronger distance-decay in fungi suggests greater dispersal limitation than in bacteria. Together, these findings suggest that basaltic lava cave microbial community composition is consistent with assembly via stochastic processes and dispersal limitation. More broadly, they demonstrate how microbial communities can follow distinct assembly trajectories even among caves that share a common lava flow and eruptive history.

2 Acknowledgements

This work was financially supported by a 2025 Philip M. Smith Graduate Student Research Grant for Cave and Karst Research from the Cave Research Foundation, an award from the University of California, Davis, Department of Earth & Planetary Sciences Cordell Durrell Fund, and a 2024 Jumpstart Graduate Award from the National Science Foundation (NSF S-STEM Program), administered through the UC Davis Department of Earth & Planetary Sciences. In addition to providing funding, I wish to express my appreciation to the Earth & Planetary Sciences Department for providing numerous opportunities to explore and study California's natural beauty, including teaching assistantships for three classes involving field geology.

We thank personnel at Craters of the Moon National Monument, Lava Beds National Monument, Lassen National Forest, Shasta-Trinity National Forest, and Klamath National Forest for assistance with field work and sample permitting.

I am endlessly grateful to my thesis advisor, Dr. Christy Grettenberger, who has always encouraged me to aim high and keep going. I also appreciate my committee members, Drs. Eliot Atekwana, Kari Cooper, and Joanne Emerson, for their supervision of this thesis and helpful feedback.

Special thanks to Dr. Christopher Mulligan, who created Figures 1 and 2 for this thesis and provided countless moments of support, valuable scientific discussion, and cups of coffee. Thank you to my family and friends, who have now heard entirely too much about microbes and rocks (and will be hearing more). Lastly, I am grateful to everyone who has had an impact on my journey as a scientist, though they are far too many to enumerate here.

3 Introduction

3.1 Rock weathering

Rock weathering is a fundamental process in global biogeochemical cycling, playing a major role in the redistribution of elements. Weathering mobilizes elements from rocks, initiating their cycling among mineral, dissolved, and biological reservoirs through both abiotic and biologically-mediated transformations. Physical and chemical weathering acts by breaking apart crystal cleavages and chemical bonds, releasing elements like calcium and iron to the surrounding environment [1–4]. Weathering is important ecologically because elements become biologically available and new substrates are created. These processes open up new niches, allowing greater diversity and complexity to take hold in the developing ecosystem [5, 6].

3.2 Microbial roles in weathering

Ecosystem development begins as microbial communities rapidly colonize freshly erupted lava and establish biofilms on rock surfaces [7]. These biofilms form the foundation of developing ecosystems and influence rock weathering by changing local physicochemical conditions and accumulating nutrients [1, 3, 8–12]. Photosynthetic microbes fix carbon from the atmosphere, store it as biomass, and make the carbon bioavailable to other organisms. Over time, microbial metabolisms concentrate bioavailable forms of carbon and nutrients within the local environment on the nanometer to centimeter scale [1, 12–14]. Microbes also dissolve old minerals and form new ones [11, 15, 16].

The primary mechanism by which microbes weather rock is through the production of organic acids, which chelate metal cations and enhance mineral dissolution [1, 11, 17–20]. In addition to acid production, microbial activity can alter redox conditions [3, 19, 21–23], further accelerating the breakdown of rocks by promoting oxidation-reduction reactions that mobilize elements like iron and manganese. An example of this feedback between microbial redox activity and mineral dissolu-

tion/precipitation is seen through the potential of dissimilatory iron reduction to remobilize mineral-bound iron [23].

Many of these local geochemical changes are associated with the reaction space created by extracellular polymeric substances (EPS), which are produced by some microbes to adhere to rock surfaces. EPS enhance local reactions by maintaining moisture and forming a solution at the rock interface that encourages ion diffusion [1, 18, 22]. EPS make up most of a biofilm's organic matter (OM) [24], providing binding sites for the liberated metal cations [1, 18]. These processes collectively contribute to rock breakdown and the transformation of primary minerals into clay minerals, crucial for soil formation (Fig. 1).

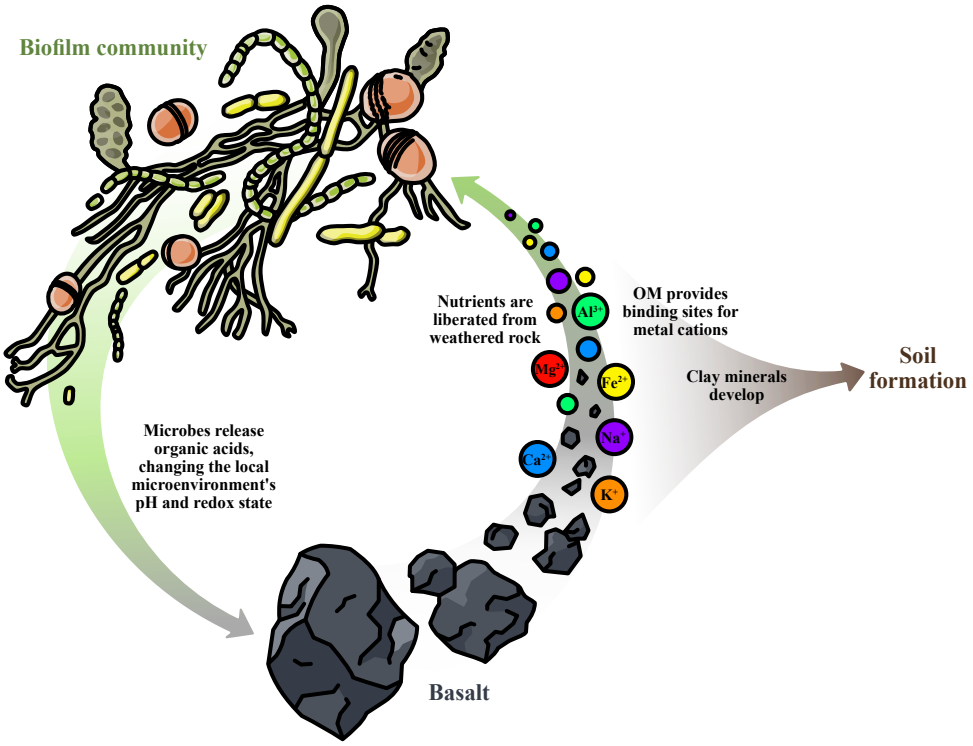

Figure 1: Biogeochemical interactions in biofilm communities on basalt.

Metabolic and chemical processes differ among microbial taxa, so changes in community composition can alter which weathering processes occur and which elements are mobilized or retained. Predicting how biogeochemical weathering activity changes through time thus requires understanding of how these microbial communities assemble.

3.3 Microbial community assembly

Understanding how microbial communities assemble is a central objective of microbial ecology because community assembly influences ecosystem functioning, biodiversity, and responses to environmental change [25–27]. Community assemblages reflect stochastic processes, including ecological drift and dispersal limitation, and deterministic processes such as environmental selection and diversification [7, 25, 28, 29]. These four drivers of community assembly constitute a simple conceptual framework [30] that allows ecologists to sort the predominant forces shaping each community.

Environmental filters, such as lava geochemistry and climate, influence deterministic processes [31–33], while the two major forces of stochasticity are drift, produced by random births and deaths [34], and dispersal limitations [35–39]. An important feature of stochastic assembly is the balance between extinction and recolonization from the regional species pool. Immigration can be quantified by the parameter m, representing the probability that a vacancy in the local community is filled by an immigrant from the regional pool rather than by local reproduction [36]. A low m value means that local communities are effectively closed, so composition changes via ecological drift rather than being replenished by outside colonizers. While microbes are less prone to dispersal limitations than macroorganisms [40–43], these limitations still matter for microbial community assembly [44–46]. When immigration between local communities is restricted, stochastic processes such as ecological drift become increasingly important, reducing the influence of interspecific competition on local community assembly [36, 47].

Community assembly is not a strict dichotomy between determinism and stochasticity. A meta-analysis of 158 metacommunity datasets found that disregarding dispersal limitation would have missed detectable structure in 37% of communities [48]. This reflects a broader shift in the field toward integration of deterministic and neutral processes as endpoints of a continuum rather than competing explanations [49], with stochastic and deterministic mechanisms often operating jointly on the same commu-

nity [50] and the relative importance of each varying across systems and over time [51].

As environmental conditions become increasingly restrictive or heterogeneous, deterministic filtering is expected to play a larger role in structuring communities [52]. Progressive changes in basalt weathering create a changing resource inventory for the biological community [53, 54]. The altered resource availability and physicochemical conditions can lead to changing niches and competition over time [55], changing the strength of environmental filtering and therefore community composition.

3.4 Primary succession

Ecological succession is the result of changing community composition, driven by the reciprocal action of organisms and their environment [5, 56]. Classical models frame succession as proceeding through facilitation, tolerance, or inhibition among successive colonizers [57], building on early descriptions of plant communities developing through predictable stages [58]. Primary succession is ecosystem development on an initially sterile substrate, such as lava flows.

A primary target for studying both primary ecological succession and rock weathering is basaltic lava, a globally common rock type [11, 59–63]. Microbial biofilms on the surfaces of basaltic rocks from lava flows host entire ecosystems of pioneer species [7, 64–69]. Microbial community composition and metabolic activity change the local physical and chemical landscape [69–73]. Directly observing microbial succession in volcanic environments is challenging because these processes often unfold over hundreds to thousands of years. One approach to studying long-term succession is to compare lava flows spanning a range of ages. Although such comparisons cannot completely isolate the effects of time from other site-specific factors, they provide a way to assess whether microbial community composition changes systematically with substrate age and successional stage.

Previous studies on microbial succession in volcanic environments, including lava flows from Iceland [68, 69, 72], Hawai'i [70, 73, 74], and various stratovolcanoes

[71, 75–77] have shown that microbial colonization begins within months of lava deposition (if not earlier) and that community composition changes significantly over timescales of years to centuries [69, 70]. This work has focused almost exclusively on open lava-flow surfaces. Basaltic eruptions, however, also commonly produce lava tubes and caves (Fig. 2) that offer a more controlled system for studying microbial community assembly. The walls and ceilings of lava caves are relatively sheltered from wind and rain, reducing dispersal that can homogenize microbial communities across sites. Additionally, limited soil and sunlight support predominantly microbially-driven ecosystems, allowing investigation of succession and rock colonization in the relative absence of vascular plants and other large organisms that dominate surface ecosystems.

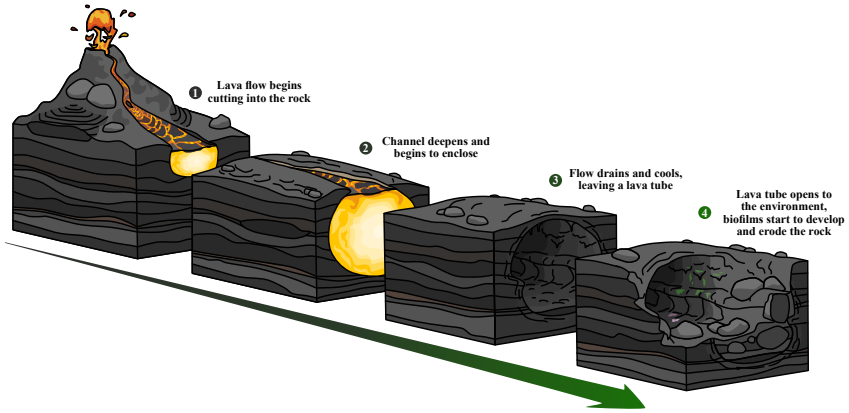

Figure 2: Development of lava tubes, from eruption to microbial colonization.

Lava caves specifically have their own, more limited body of microbiological work [64, 66, 67, 74, 78–83]. However, this cave-specific literature has largely focused on characterizing community structure and taxonomic novelty at single sites or timepoints, rather than tracking compositional change across a gradient of substrate ages.

3.5 Knowledge gap

Long-term microbial compositional change across lava caves has received little attention. Many studies have focused on flows younger than 500 years [69, 70, 73], marine-influenced environments with distinct weathering processes [73], or later successional sites dominated by plant-driven ecosystem changes [71, 76]. Earlier work

on microbial communities may have underestimated diversity due to the use of clone libraries [67–70, 75, 84], which have sequencing limitations relative to more recent developments in sequencing technology [85]. Collectively, these limitations leave important gaps in our understanding of microbial succession across the centuries-to-millennia timescales over which basalt weathers [86]. Few studies have explicitly tested whether long-term microbial succession in volcanic systems is driven primarily by deterministic environmental filtering or stochastic assembly processes. Because the balance between community assembly drivers is expected to change as ecosystems develop [25], a range of lava flow ages provides a natural laboratory for investigating shifts in microbial assembly processes.

Beyond their value as natural laboratories for studying microbial community assembly, lava caves remain among the least characterized terrestrial microbial habitats [80, 87]. Access restrictions and conservation concerns necessarily limit sampling and disturbance, making baseline characterization of their microbial diversity intrinsically valuable. Such surveys provide important reference data for future conservation efforts for these fragile subsurface ecosystems [88].

3.6 Goals

This study investigates the diversity, composition, and assembly of bacterial and fungal communities inhabiting basaltic lava caves to better understand microbial succession within the context of long-term rock weathering.

We proposed three hypotheses: (1) that microbial diversity increases with lava flow age; (2) that caves within the same lava flow are more compositionally similar than caves from different lava flows; and (3) that community composition is governed primarily by stochastic assembly processes.

4 Methods

4.1 Study sites

We collected 36 samples from 12 basaltic lava caves. These caves were located within Craters of the Moon National Monument and Preserve (CRMO) in Idaho, Lava Beds National Monument (LABE), and Klamath, Lassen, and Shasta-Trinity National Forests in northern California (Fig. 3). Caves were selected to cover a range of parameters, including lava flow age, geochemistry, and aridity (Table 1).

Lava flow geochemistry was retrieved from published bulk rock geochemistry data [89–94]. With the exception of caves in Lassen and Shasta-Trinity National Forests, all study sites are located in semiarid environments. The majority (9/12) of caves are located in high-elevation desert regions (those in CRMO and LABE), characterized by extensive exposed basaltic lava flows and minimal vegetation. These conditions reduce hydrological weathering, allowing for a clearer assessment of biological influences on rock alteration. Regional climatic context is included below for each set of caves, but the caves create microenvironments with generally lower temperatures and insulation from sun, rain, and snow [95, 96].

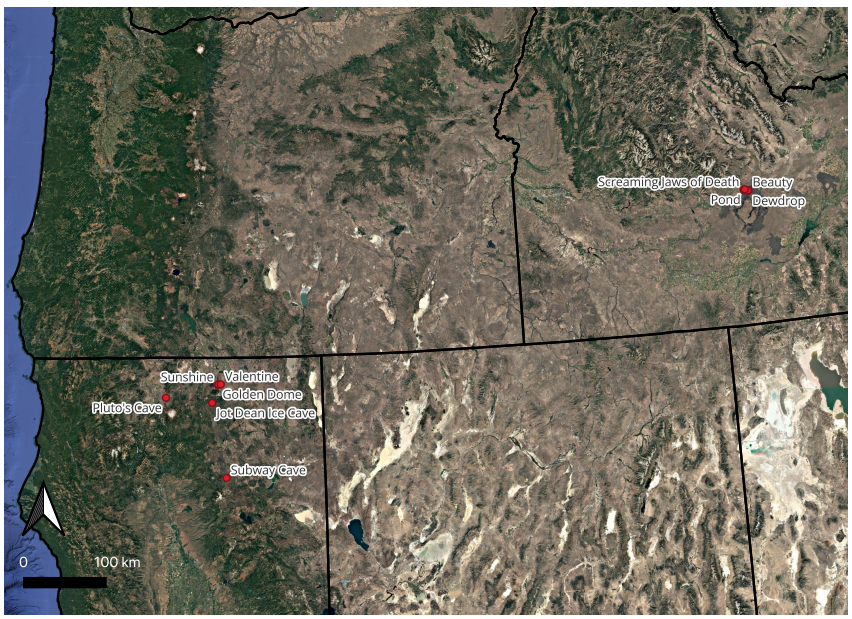

Figure 3: Map of cave sampling sites across northern California and southern Idaho.

Table 1: List of sampling site information.

State	Agency	Cave Name	Geologic Unit	Age	Age Source
ID	Craters of the Moon National Monument and Preserve	Beauty	Blue Dragon Flow	2.076 ± 0.045 kyr BP	[100]
ID	Craters of the Moon National Monument and Preserve	Dewdrop	Blue Dragon Flow	2.076 ± 0.045 kyr BP	[100]
ID	Craters of the Moon National Monument and Preserve	Indian Tunnel	Blue Dragon Flow	2.076 ± 0.045 kyr BP	[100]
ID	Craters of the Moon National Monument and Preserve	Pond	Grassy Cone Flow	7.360 ± 0.060 kyr BP	[100]
ID	Craters of the Moon National Monument and Preserve	Spongy Floor	Grassy Cone Flow	7.360 ± 0.060 kyr BP	[100]
ID	Craters of the Moon National Monument and Preserve	Screaming Jaws of Death	Grassy Cone Flow	7.360 ± 0.060 kyr BP	[100]
CA	Lava Beds National Monument	Valentine	Basalt of Valentine Cave	12.260 kyr BP	[97]
CA	Shasta-Trinity National Forest	Jot Dean	Basalt of Giant Crater	12.430 kyr BP	[97]
CA	Lassen National Forest	Subway	Basalt of Hat Creek	24 ± 6 ka	[91]
CA	Lava Beds National Monument	Sunshine	Basalt of Mammoth Crater	36 ± 16 ka	[98]
CA	Lava Beds National Monument	Golden Dome	Basalt of Mammoth Crater	36 ± 16 ka	[98]
CA	Klamath National Forest	Pluto's Cave	Pluto's Cave Basalt	116 ± 13 ka	[93]

4.1.1 Sampled caves and lava flows

In September 2023, we sampled three caves spanning two flows in Lava Beds National Monument, northern California: Valentine Cave, within the basalt of Valentine Cave (12.26 kyr BP [97]), and Sunshine and Golden Dome caves, both within the older basalt of Mammoth Crater (36 ± 16 ka [98]). In September 2024, we sampled six caves within Craters of the Moon National Monument and Preserve, Idaho, whose lava flows are associated with the Great Rift volcanic zone [99]. Beauty, Dewdrop, and Indian Tunnel caves formed within the Blue Dragon flow (2.076 ± 0.045 kyr [100]), while Pond, Screaming Jaws of Death, and Spongy Floor caves formed within the older Grassy Cone flow (7.360 ± 0.060 kyr [100]). Three other caves each came from a different national forest and a different flow: Subway Cave, in the basalt of Hat Creek (24 ± 6 ka [91]) within Lassen National Forest; Jot Dean Ice Cave, in the basalt of Giant Crater (12.430 kyr BP [97]), Shasta-Trinity National Forest; and Pluto's Cave, in the Pluto's Cave basalt (116 ± 13 ka [93]), Klamath National Forest. We sampled the latter two National Forests in May 2025 and Lassen National Forest in September 2025.

4.2 Sample collection

Samples were collected from cave walls and ceilings by using a rock hammer and a chisel sterilized with 70% ethanol to separate protruding rock with biofilm from the cave wall. Biofilms were identified by color, which can reflect different microbial metabolisms and communities [64, 66]. Designated "color" groups were: light green, dark green, green (generic), teal, pink, orange, "sheen", white, brown, and black.

Caves were selected in coordination with the National Park Service and Forest Service management authorities of each respective site. All sampling was carried out under permits issued by the respective land management agency: Craters of the Moon National Monument and Preserve under permit #CRMO-2024-SCI-0002, Lava Beds National Monument under permit #LABE-2023-SCI-0016, Shasta-Trinity National Forest under permit #062267, and Lassen National Forest under Authorization ID #ALM137001. Due to cave conservation concerns and permit restrictions, sampling was generally limited to three samples per cave, with one sample collected from each biofilm color present, with two exceptions. Only two samples were collected from Spongy Floor (black, dark green), while four samples were collected from Screaming Jaws of Death (brown, sheen, and two "green"). Biological replicates were not collected.

Each sample was captured directly into a Whirl-Pak® Standard Sterilized Sampling Bag, which was sealed after collection. Samples were stored on ice for transport and then stored at -20°C. Approximately $2 \times 2 \times 2$ cm rock chip samples were collected to include the microbial biofilm, the underlying rock, and the un-weathered rock below it (i.e., through the weathering rind). A total of 36 samples were collected from 12 caves; a list of all samples is presented in Table 4.

Figure 4: Table of samples and locations, excluding Pluto's Cave samples, which failed to yield sequence data.

Craters of the Moon National Monument and Preserve (ID)	
Blue Dragon Flow	
Beauty Cave	
light green	B-LG
dark green	B-DG
orange	B-O
Dewdrop Cave	
light green	DD-LG
dark green	DD-DG
teal	DD-T
Indian Tunnel Cave	
brown	IT-B
green	IT-G
pink	IT-P
Grassy Cone Flow	
Pond Cave	
light green	P-LG
dark green	P-DG
brown	P-B
Spongy Floor Cave	
black	SF-B
dark green	SF-DG
Screaming Jaws of Death Cave	
brown	SJD-B
green	SJD-G1
green	SJD-G2
sheen	SJD-S
Lava Beds National Monument (CA)	
Basalt of Mammoth Crater	
Golden Dome Cave	
light green	GD-LG
dark green	GD-DG
pink	GD-P
Sunshine Cave	
light green	S-LG
dark green	S-DG
pink	S-P
Basalt of Valentine Cave	
Valentine Cave	
light green	V-LG
dark green	V-DG
pink	V-P
Shasta-Trinity National Forest (CA)	
Basalt of Giant Crater	
Jot Dean Ice Cave	
light green	JD-LG
orange	JD-O
pink	JD-P
Lassen National Forest (CA)	
Basalt of Hat Creek	
Subway Cave	
orange	SC-O
black	SC-B
white	SC-W

Samples were prepared for DNA extraction by grinding the rock into a powder using a mortar and pestle, sterilized with 70% ethanol. DNA was extracted from each sample, along with extraction blanks, using a Qiagen DNeasy PowerBiofilm Kit (MoBio, Carlsbad, USA) following the manufacturer's instructions.

4.3 DNA sequencing

DNA was sequenced at SeqCenter (Pittsburgh, PA, USA). Community composition was determined by sequencing the 16S rRNA gene, for bacteria, and the internal transcribed spacer (ITS) region, for fungi. Amplification for each gene or region was performed using Zymo Research's Quick-16S Kit or Quick-ITS Plus NGS Library Prep Kit. The primer pair 341F-805R (5'CCTACGGGNGGCWGCAG3' and 5'GAC-

TACHVGGGTATCTAATCC3', respectively) was used to target the V3/V4 regions of the 16S rRNA gene and ITS3F-ITS4R (5'GCATCGATGAAGAACGCAG3' and 5'TCCTCCGCTTATTGATATGC3', respectively) targeted the ITS2 region. Samples were cleaned and normalized, then sequenced on a P1 or P2 600cyc NextSeq2000 Flowcell, generating 2x301bp paired-end reads. SeqCenter performed quality control and adapter trimming using bcl-convert (v4.2.4) [101]. Sequences identified as PCR artifacts (primer dimers) were filtered for read length > 150bp, PolyN strings < 10 sequential Ns, and PolyG strings < 150 sequential Gs. Sequence data were processed in RStudio using `DADA2` (v.1.38.0), an R package that denoises amplicon sequence data to produce amplicon sequence variants (ASVs) [102]. Relative to other methods, ASV methods offer improved sensitivity and resolution [103] and more accurately identify contaminants, which is especially important in low-biomass communities [104], like cave biofilms.

4.4 16S rRNA gene workflow

16S rRNA data were processed following "A `DADA2` workflow for Big Data: Paired-end (1.4 or later)" (`https://benjjneb.github.io/dada2/bigdata_paired.html`). Using "filterAndTrim", reads were trimmed for quality, with a maximum of 245 forward and 245 reverse base pairs. Reads with undetermined base calls, sequences associated with PhiX (an internal control), and primers (17 basepairs from the beginning of the forward reads and 21 basepairs from the beginning of the reverse reads) were removed. The error allowance was 2 errors per sequence. Reads were truncated at the first instance of a quality score less than or equal to 2. Error rates associated with the sequencing run were modeled using "learnErrors" and applied to the filtered and trimmed sequence data using "dada". Forward and reverse reads were assembled into sequence contigs. Chimeras were removed using the function "removeBimeraDenovo", where chimera identification was performed through the consensus procedure.

4.4.1 Taxonomic assignment

We used DADA2's "assignTaxonomy" function (a native implementation of the naive Bayesian classifier method for taxonomic assignment [105]) to compare sample sequences to sequences from known organisms in the SILVA non-redundant database (v138.2) [106]. Sequence data were subsequently filtered using the "subset_taxa" function to exclude chloroplast and mitochondrial sequences and to include only sequences matching Kingdom Bacteria.

4.4.2 Predicted functional annotation

Putative functional profiles of bacterial ASVs were predicted using FAPROTAX (v1.2.12), a taxonomic database comparison method [107]. Additional bacterial functional prediction was performed using PICRUSt2 (v2.6.3), a phylogenetic inference method [108]. ASVs were filtered to include only sequences with low phylogenetic distance to the reference database, and therefore higher prediction accuracy within PICRUSt2. This was quantified as nearest sequenced taxon index (NSTI) values lower than 0.5. MetaCyc ID pathway identifiers were mapped to descriptions, then we used a targeted keyword search for "nitrogen | nitrate | nitrite | ammonia", "sulfur | sulfate | sulfite | thiosulfate", "siderophore | iron", "carbon fixation | Calvin | reductive TCA", and "organic acid | gluconate | citrate" to identify pathways associated with each of these processes for descriptive reporting. Given the documented inconsistency in differential abundance results across methods for PICRUSt2-predicted data [108], pathway abundances were reported descriptively rather than tested statistically.

4.5 ITS region workflow

ITS region data were processed in RStudio using DADA2 (v.1.38.0) following "DADA2 ITS Pipeline Workflow (1.8)" (https://benjjneb.github.io/dada2/ITS_workflow.html). Reads were pre-filtered using "filterAndTrim" (maxN = 0) to remove any sequences containing ambiguous base calls (Ns), with no additional filtering performed at this step. Primer orientation and presence were then verified by generating all

four possible orientations (forward, complement, reverse, and reverse-complement) of the forward and reverse primer sequences using `Biostrings` (v.2.78.0) and counting primer occurrences across the pre-filtered reads.

Primers were removed using `cutadapt` (v.5.2) [109]. The `cutadapt` tool was run separately on each sample with linked-adapter trimming, with up to two adapter-removal passes per read to account for potential primer read-through. Per SeqCenter's documentation, the ITS3F/ITS4R primers were preceded by variable-length spacers (phased), which SeqCenter's own quality-control step removed prior to delivery along with part of the primer sequence. Inspection of the raw fastq files showed that the forward primer consistently lacked its first base, so it was matched using two Ns preceding the 16bp core, while the reverse primer was intact and matched at its full length.

Following primer removal, the trimmed reads were filtered using "filterAndTrim" (maxN = 0, maxEE = 2, truncQ = 2, rm.phix = TRUE). A minimum length filter (minLen = 50) was applied to remove spurious, very low-length sequences. Error rates associated with the sequencing run were modeled using "learnErrors" and applied to the filtered and trimmed sequence data using "dada". Forward and reverse reads were assembled into sequence contigs, and chimeras were removed using "removeBimeraDenovo", with chimera identification performed through the consensus procedure.

4.5.1 Taxonomic assignment

`DADA2`'s "assignTaxonomy" function compared sample sequences to sequences from known organisms in the UNITE ITS database (v10.0) [110] to assign each ASV a taxonomic identity. Sequence data were subsequently filtered using the "subset_taxa" function to include only sequences matching Kingdom Fungi. In addition to the ASVs categorized as "Unassigned" by the UNITE database, some taxonomic ranks were classified as *incertae sedis*. This term refers to ASVs with uncertain taxonomic placement in existing literature; frequently these uncertainties occur at species, genus, and family ranks, but may also occur at the phylum level, particularly in the understudied en-

vironmental fungi [111]. *Incertae sedis* is preserved as a distinct category from "Unassigned" because the former label reflects a known group of organisms (e.g., species, genera, etc.) with uncertain taxonomic placement, whereas "Unassigned" taxa may be assigned to a clade through phylogenetic analysis of ASV sequences.

4.5.2 Predicted functional annotation

Fungal ASVs with unambiguous genus-level taxonomic assignments were assigned putative lifestyle traits using the FungalTraits database [112].

4.6 Community data integration and diversity estimation

Once filtered by taxonomy, the 16S rRNA and ITS datasets were each analyzed in the R packages `phyloseq` (v.1.56.0) and `vegan` (v.2.7.3) [113, 114]. Rarefaction curves, generated using the `vegan` function "rarecurve", and Good's coverage calculations were used to assess whether sequencing depth was sufficient to capture microbial diversity within samples. After generating rarefaction curves for both datasets, the `phyloseq` package was used to integrate sequence count data, taxonomic assignments, and associated DNA sequences into a single object for each dataset. Using `phyloseq` functions, sequence data were transformed from sample counts to proportions with respect to the total using "transform_sample_counts". These relative abundance data were then aggregated to higher taxonomic levels using "tax_glom". We quantified taxonomic resolution across ranks for each dataset by calculating the proportion of reads that could not be assigned beyond each taxonomic level, including both missing (NA) designations (both datasets) and *incertae sedis* designations (ITS dataset only), identified via string-matching on the taxonomy table.

4.7 Statistical analysis

To characterize microbial community composition and its relationship to lava flow age, biofilm color, and geographic location, we used a combination of multivariate, univariate, and neutral-model-based statistical approaches, following general

practices for microbiome community data [115]. Analyses primarily targeted three questions: whether alpha diversity increased with lava flow age (H1), whether caves within the same lava flow were more compositionally similar to one another than to caves from different lava flows (H2), and whether observed community patterns were consistent with stochastic, neutral assembly processes (H3). All analyses were conducted separately for bacterial (16S rRNA gene) and fungal (ITS region) datasets.

P-values were adjusted for multiple comparisons using Benjamini-Hochberg (BH) correction [116] for each Kruskal-Wallis diversity test and each pairwise PERMANOVA comparison. The single variable, sequential, and marginal PERMANOVAs test distinct, specified hypotheses rather than a family of comparable tests, and thus were not corrected for multiple comparisons.

4.7.1 Spatial and temporal structuring of community composition

The differences in composition between communities were calculated as Bray-Curtis dissimilarity [117, 118] from the relative abundance data using `phyloseq`'s "distance" function. Community dissimilarity was visualized using principal coordinate analysis (PCoA) [119]. Convex hulls were fitted around samples grouped by cave or flow unit where at least three samples were present per group, using the `grDevices` package's "chull" function.

To examine spatial patterns of community dissimilarity, we used the `vegan` package's "mantel" function to test the Spearman rank correlations between Bray-Curtis community dissimilarity values and pairwise geographic distances among sample sites (999 permutations). Distance was calculated with UTM coordinates of each cave using the "dist" function in R. Samples within the same cave were assigned a distance of zero. To evaluate whether the geographic Mantel result reflected genuine spatial autocorrelation rather than the compositional distinctness of individual caves, we also tested the distance-decay relationship after excluding same-cave comparisons. Sample-level dissimilarity and distance matrices were collapsed to cave-level means using only cross-cave sample pairs for this purpose.

A set of Mantel correlograms [120] was used to examine how community dissimilarity varied across flow age and geographic distance. One set of Mantel correlograms used discrete age-difference classes, with significance assessed via progressive Holm correction across classes, while the other set used geographic distance classes to identify the spatial scale at which the isolation-by-distance signal identified by the overall Mantel test is strongest.

The distance-decay relationship was visualized by fitting an asymptotic regression (using the "SSasymp" function from R's `stats` package) of Bray–Curtis dissimilarity against log-transformed geographic distance:

$$y = R_0 + (A - R_0)(1 - e^{-kx}), \tag{1}$$

An asymptotic curve was chosen because Bray–Curtis dissimilarity has a maximum value of 1 (the most dissimilar two communities can be, with zero overlapping taxa).

4.7.2 Alpha diversity and lava flow age

To test whether diversity increased with lava flow age, we calculated Shannon and Simpson diversity indices [121] for each sample from raw read counts using `phyloseq`'s "estimate_richness" function. We compared diversity across caves, lava flow units, and biofilm colors using Kruskal-Wallis tests, a non-parametric approach appropriate given the small sample size within each group. We additionally modeled the relationship between each diversity index and lava flow age directly using linear regression.

4.7.3 Compositional differences among caves, lava flow units, and biofilm colors

We computed permutational analysis of variance (PERMANOVA) using `vegan`'s `adonis2` function (999 permutations) to test whether microbial community composition differed among caves, lava flow units, and biofilm colors, and whether it was associated with lava flow age. Cave, lava flow unit, biofilm color, and lava flow age were each first evaluated individually as single-predictor PERMANOVAs.

Cave is structurally nested within lava flow unit, as each cave belongs to exactly one flow unit, and multiple samples were collected within each cave. This has two implications for statistical testing: (i) single-predictor cave model cannot explain less variance than the corresponding unit model, regardless of effect size; R^2_{cave} and R^2_{unit} are therefore not compared directly as evidence for H2, and (ii) samples collected within the same cave are not independent with respect to a between-unit comparison. To avoid pseudoreplication [122], the single-predictor flow unit PERMANOVAs were run on reduced datasets in which each cave's samples were collapsed to a single mean community composition. Bray-Curtis distances were then recomputed from these cave-level mean compositions and used in the flow unit PERMANOVAs.

To test directly whether caves within the same lava flow unit were more compositionally similar to one another than caves from different units, we used analysis of similarity (ANOSIM; `vegan`'s "anosim" function, 999 permutations) [123] on the same cave-level Bray-Curtis distance matrix. The ANOSIM approach compares the ranks of within- and between-group Bray-Curtis dissimilarities using permutation tests.

Sequential PERMANOVAs were used to test how much additional variance was explained by flow unit after accounting for cave. Biofilm color was analyzed using both single-predictor PERMANOVA and sequential PERMANOVAs with both variable orders (color before cave and cave before color) to assess whether its contribution to explained variance depended on model order. Lava flow age was tested as an additional single-variable PERMANOVA predictor of community composition, distinct from its use as a predictor of alpha diversity.

PERMANOVA is sensitive to differences in within-group dispersion as well as differences in group centroids. Thus, we tested the homogeneity of multivariate dispersion for each grouping variable (cave, flow unit, and color) using `vegan`'s "betadisper" function. As `betadisper` cannot meaningfully estimate within-group dispersion from a single sample, we additionally report dispersion results after excluding such groups. For the 16S rRNA dataset, all 11 represented caves and all 6 flow units had $\geq$2 samples (no exclusions), whereas 4 of 10 color categories (brown, sheen, teal, and

white) were represented by single samples and excluded. For the ITS dataset, 1 of 10 represented caves (Jot Dean), 1 of 6 flow units (Basalt of Giant Crater), and 4 of 10 color categories (black, sheen, teal, and white) were represented by a single sample and excluded, leaving 9 caves, 5 flow units, and 6 color categories. We note where PERMANOVA results for a given variable may be at least partly attributable to differences in dispersion rather than differences in community composition. Dispersion test p-values were derived using `vegan`'s "permutest" function; a fixed random seed (19990724) was set prior to all permutation-based analyses to ensure reproducibility.

Following the global PERMANOVAs, pairwise PERMANOVAs were used to identify significant pairs of lava flow units. Cave-level pairwise tests were not run because the number of caves did not meet minimum sample size for permutation-based testing. Where pairwise PERMANOVA identified significant differences between specific flow units, similarity percentage (SIMPER; `vegan`'s "simper" function; 999 permutations) analysis [123] was used to identify the ASVs contributing most to the observed Bray-Curtis dissimilarity between pairs, with each ASV's genus-level taxonomic identity (where resolved) reported to aid interpretation. To reduce noise from rare taxa, ASVs were filtered to those present in at least 10% of samples, resulting in 750 bacterial ASVs and 142 fungal ASVs included in the SIMPER analyses.

Univariate Kruskal-Wallis testing assessed the significant relationship between `FAPROTAX`'s taxon-associated metabolism and biofilm color for each metabolic capability. To test whether the overall functional profile (rather than any single capability) was associated with biofilm color, a Bray-Curtis distance matrix [117, 118] was constructed from the relative abundances of the represented functional groups and analyzed using PERMANOVA (`vegan`'s "adonis2", 999 permutations), with permutations restricted within cave ("strata" argument) due to the nested sampling design.

4.7.4 Neutral community models

To evaluate the degree to which community assembly is consistent with stochastic (neutral) processes, we fit neutral community models to ASV and genus-level rela-

tive abundance distributions for both fungi and bacteria using the `neutral.fit` function in the `MicEco` R package (v. 0.10.0) [124], based on Sloan's neutral community model for prokaryotes [125]. A close fit to the neutral model indicates that observed relative abundance distributions are consistent with the stochastic processes of ecological drift and dispersal limitation. However, this model approach cannot rule out niche-based selection, which may produce abundance distributions that are difficult to distinguish from purely neutral processes [126].

4.8 Metagenomics

In addition to 16S rRNA gene and ITS region sequencing, one sample (S-P) was selected for metagenomic sequencing, which provides information on functional potential, including specific genes, metabolic pathways, and enzyme capabilities. Metagenomic data were further processed using standard bioinformatics tools on the KBase platform [127]. The quality of the raw sequence data was checked using `FastQC` (v0.12.1) [128]. Reads were assembled using `MEGAHIT` (v1.2.9) [129] under the "meta-sensitive" parameter preset to form contigs. Contigs longer than 2,500 bp were binned with `MetaBAT2` (v.1.7) [130], then the bins were filtered using `CheckM` (v1.0.18) [131] to keep bins that were $\geq$ 90% complete and $\leq$ 5% contaminated. Coverage and relative abundance of metagenome-assembled bins were estimated by mapping quality-filtered reads back to each bin using `Bowtie2` (v2.3.2) [132].

The bins meeting quality thresholds were classified with `GTDB-Tk` (v2.3.2; GTDB release R214) [133]. Taxonomy was assigned to each bin using a minimum alignment threshold of 10%, and genomes were placed into the GTDB reference tree based on domain-specific marker genes. Bins were extracted using `Extract Bins as Assemblies from BinnedContigs` (v1.0.2) for annotation. Both the overall metagenome (assembly file) and individual bins were annotated with `Prokka` (v1.14.5) [134] using default parameters, with more detailed functional annotation provided by `DRAM` (v0.1.2) [135]. The Model SEED tool `Build Metagenome Metabolic Model` (v0.1) [136] was used to reconstruct two draft metagenome-scale community metabolic models from the `Prokka`-

annotated assembly. One model was gapfilled using the default "complete media" to identify the smallest set of additional reactions required for biomass production, and the other model was produced without gapfilling.

5 Results

Of the 36 samples collected from 12 caves, all three samples from Pluto's Cave yielded insufficient DNA for sequencing and were excluded. This left 33 samples from 11 caves for downstream analyses. Two samples (P-B and SJD-B) failed to yield usable 16S rRNA gene amplicon sequence data, leaving 31 samples for bacterial analyses. Eight samples (IT-P, JD-LG, JD-P, P-B, all Sunshine Cave samples [S-DG, S-LG, S-P], and SC-B) failed to yield usable ITS sequence data, leaving 25 samples for fungal analyses. Several biofilm color categories were represented by single samples in the final dataset, including sheen, teal, and white (both bacteria and fungi), as well as brown (bacteria only) and black (fungi only).

In addition to the flattened rarefaction curves (Fig. 5), Good's coverage estimates of $\geq$ 99% for each sample indicated sequencing depth was sufficient to capture community diversity.

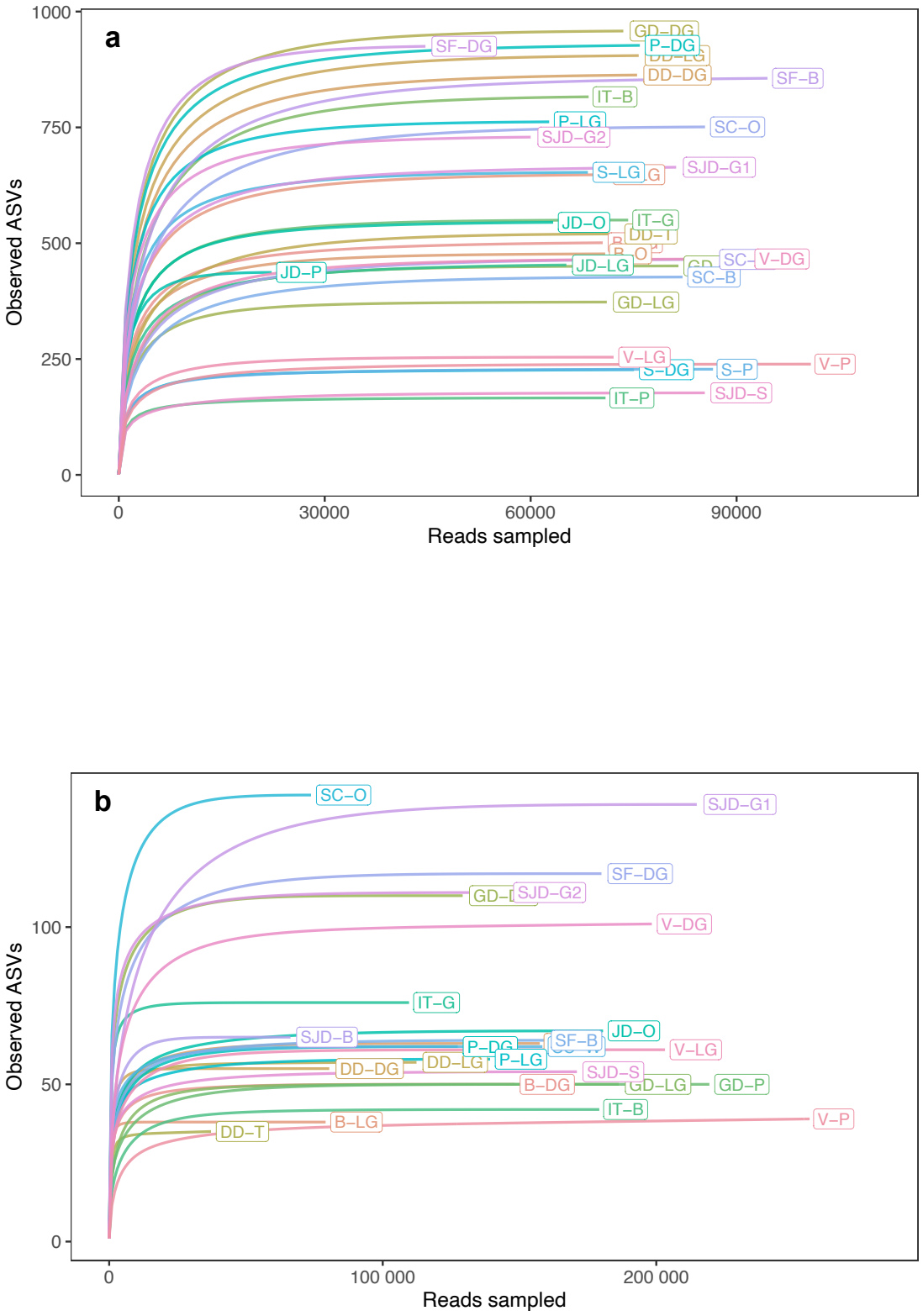

Figure 5: Rarefaction curves for bacterial (a) and fungal (b) datasets. Curves approaching plateaus indicate sufficient sequencing depth to capture observed diversity.

5.1 Bacterial communities

There were 10,708 unique bacterial ASVs and 34 phyla. The ten most abundant phyla across all samples were Actinomycetota (present in 100% of samples), Bacteroidota (100%), Chloroflexota (100%), Gemmatimonadota (100%), Planctomycetota (100%), Pseudomonadota (100%), Acidobacteriota (90.3%), Cyanobacteriota (83.9%), Verrucomicrobiota (77.4%), and Rhodothermota (64.5%). Some phyla were highly prevalent despite low abundance in each community, such as Patescibacteria (present in 96.8% of samples) and Deinococcota (87.1%).

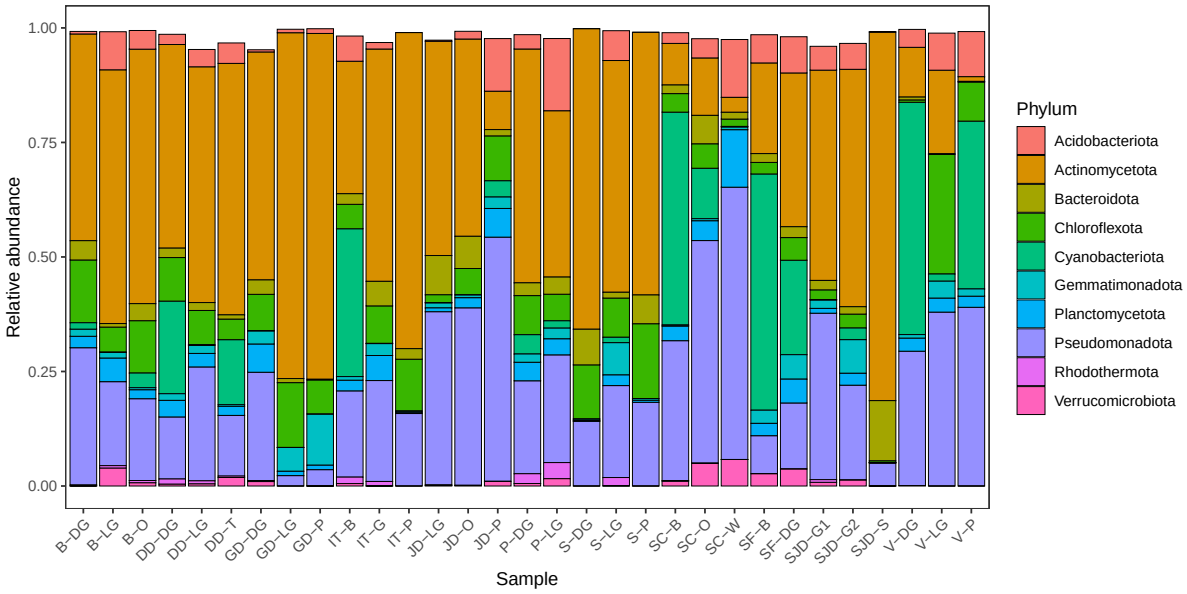

Figure 6: Ten most abundant bacterial phyla across all samples.

A minority (0.21%; 0.02% of total reads) of bacterial ASVs were taxonomically unresolved at the phylum level. There was less taxonomic resolution at the genus level, with 45.9% of ASVs (35.3% of total reads) unable to be matched to genera in the SILVA database. Of the 687 genera that were identified, only 2 (0.3%; *Rubrobacter* and *Pseudonocardia*) were detected in 100% of samples, while the majority showed variable occurrence across samples (standard deviation = 0.25); 24 genera (3.5%) were detected in $\geq$ 75% of samples.

5.1.1 Putative functional annotation

PICRUSt2 analysis assessed that 98% of the bacterial ASVs were within the quality threshold and the median NSTI = 0.12. The quality threshold was exceeded by 212 ASVs (2%) and only one ASV was excluded by PICRUSt2's default NSTI $\geq$ 2.0 cutoff.

Predicted MetaCyc pathways included the Calvin-Benson-Bassham cycle, nitrate and sulfate reduction, and the reductive TCA cycle. Phototrophic sulfur bacteria were predicted to be present in 90.3% of samples, with a maximum relative abundance (bacteria) of 0.0476%. FAPROTAX analysis identified 35 putative functional groups with non-zero presence out of 92 total FAPROTAX groups. Four metabolisms were predicted to be present in 100% of samples: chemoheterotrophy, aerobic chemoheterotrophy, nitrate reduction, and aromatic compound degradation. Taxa known to be capable of these functions had a mean relative abundance per sample of 34%, 33%, 8%, and 2.5%, respectively, and a maximum relative abundance of 56%, 49%, 27%, and 7.3%, respectively. Fermentation-capable taxa were also present in nearly every sample (30 of 31), with a mean relative abundance range of 0.38% to 2.6%. Taxa capable of phototrophy and associated functions (oxygenic photoautotrophy, photoautotrophy, photosynthetic cyanobacteria, and, to a lesser extent, photoheterotrophy) comprised around a fifth of each sample's community. These taxa were present in 84% to 87% of samples and each had a maximum relative abundance of 21%, with the exception of photoheterotrophy-associated taxa, which were present in 32% of samples with a maximum relative abundance of 0.2%.

Predicted functional profile differed significantly by biofilm color (PERMANOVA; $R^2 = 0.55$, $p = 0.007$) when considering all 35 FAPROTAX groups together as one Bray-Curtis distance matrix, stratified by cave. However, no individual functional group was associated with a biofilm color ($p < 0.05$) using Kruskal-Wallis testing with BH correction, so no individual function could be identified as the specific driver of significance after multiple-comparison correction.

5.2 Fungal data

Fungal communities were less diverse than bacterial communities. There were 1,171 total ASVs, though only 838 ASVs were identifiable to a named phylum through the UNITE database. There were 7 identifiable phyla. Ascomycota (73.2% mean abundance per sample) accounted for 86.2% of the classified ASVs. Additional contributions were from Basidiomycota, Chytridiomycota, Rozellomycota, Mortierellomycota, Sanchytriomycota, and Mucoromycota. However, these phyla were not evenly distributed among samples. Besides Ascomycota, the other 6 phyla were only present in $\geq$ 1% relative abundance in one or two sample and Mucoromycota was only detected in one sample (SJD-G2) at 0.0048% relative abundance. Five of the 25 fungal samples (B-DG, B-LG, DD-LG, IT-G, and P-LG) had less than 50% of their total relative abundance classified at the phylum level (Fig. 7).

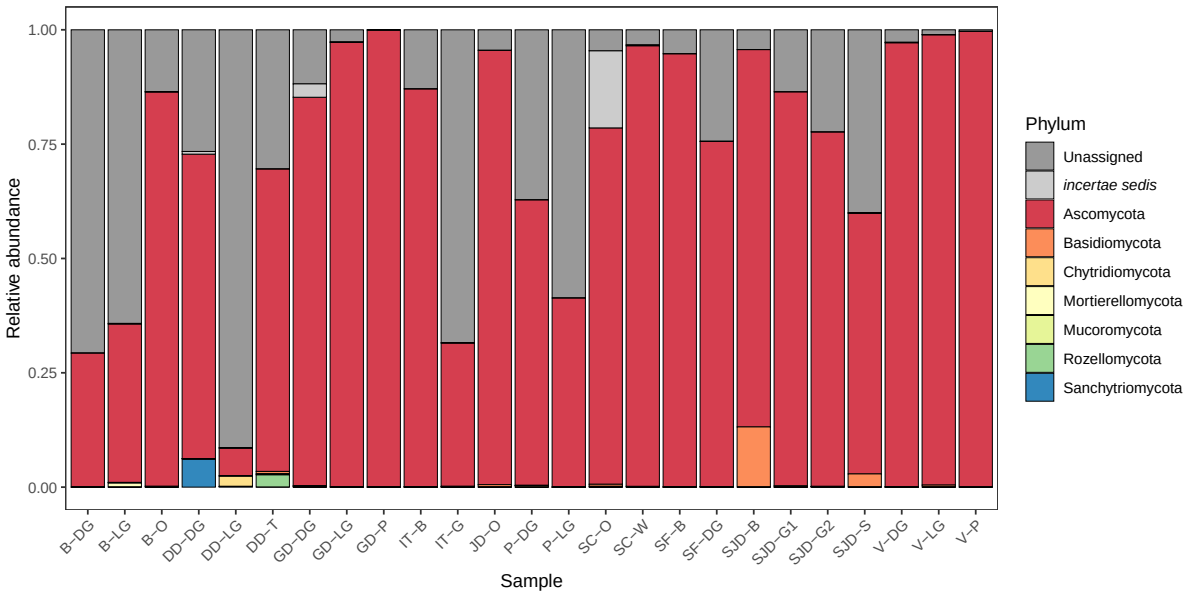

Figure 7: Relative abundances of fungal phyla across samples.

A minority of ASVs (0.48% of total reads) were identifiable only to the Kingdom level (i.e., Fungi). Excluding these completely unresolved ASVs, 59% of ASVs with *incertae sedis* designations first encounter uncertainty at the family level. *Incertae sedis* classifications were generally rare at the phylum level, with notable proportions observed only in SC-O (16.9% of the sample's community) and GD-DG (2.9%). Of the 256

fungal genera identified using the UNITE database, only one (0.39%; *Cladosporium*) occurred in $\geq$75% of samples. The standard deviation of the frequency distribution was 0.14. Between the FungalTraits and UNITE databases, there were 3 unique order-level and 9 unique family-level disagreements among the 460 ASVs identified to the genus level.

5.2.1 Functional profiles

Of the 460 fungal ASVs identified to the genus level (39.3% of fungal ASVs), 437 were assigned to at least a "primary lifestyle" using the FungalTraits database. The remaining 23 ASVs corresponded to genera included in FungalTraits but for which no primary lifestyle annotation was available. The most common primary lifestyle was saprotrophy (50.34% of genus-identified ASVs), followed by lichenized fungi (22.4%) and plant pathogens (13.3%).

5.3 Relationships

5.3.1 Compositional differences: cave, flow unit, and color

Community composition at the ASV level differed significantly among caves (PERMANOVA; bacteria: $R^2 = 0.484$, $p < 0.001$; fungi: $R^2 = 0.503$, $p < 0.001$). In single-variable PERMANOVA tests, lava flow unit also explained a moderate proportion of variation in community composition (bacteria: $R^2 = 0.28$; fungi: $R^2 = 0.29$; each $p = 0.001$). However, after accounting for cave identity in sequential tests, lava flow unit's contribution dropped to $R^2 = 0$ for both bacterial and fungal communities. In correcting for lava flow unit-cave pseudoreplication [122], we additionally tested lava flow unit on cave-level centroids (one mean composition per cave; bacteria: $n = 11$ caves; fungi: $n = 10$ caves). Flow unit differentiated these cave-level centroids (bacteria: $R^2 = 0.58$, $p = 0.006$; fungi: $R^2 = 0.61$, $p = 0.003$). This R^2 is not on the same scale as the sample-level estimate above. Since cave-centroid aggregation removes within-cave variance from total SS, only the p-value is comparable.

Within-flow cave pairs were significantly more similar in composition than

between-flow cave pairs (ANOSIM; bacteria: $R = 0.619$, $p = 0.003$; fungi: $R = 0.581$, $p = 0.023$). Same-flow cave pairs (e.g., Beauty–Dewdrop, Golden Dome–Sunshine) cluster more closely in PCoA visualization (Fig. 8) than cross-flow pairs, consistent with the ANOSIM results. Mean pairwise Bray-Curtis dissimilarity was lower within flows than between flows for both datasets (bacteria: 0.85 vs 0.95; fungi: 0.91 vs 0.99). Community composition did not differ significantly between biofilm color categories (bacteria: $p = 0.56$; fungi: $p = 0.64$).

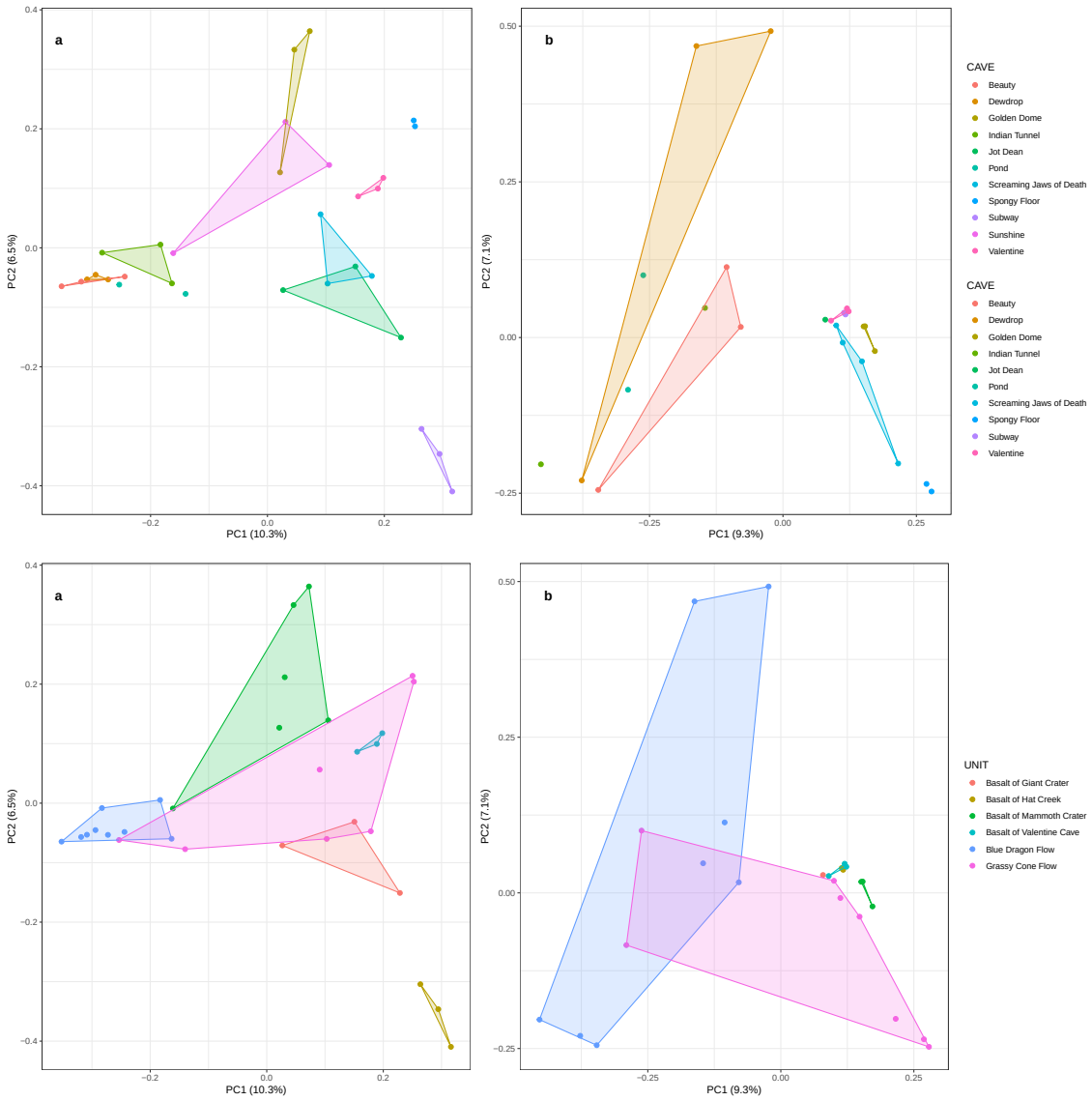

Figure 8: Principal coordinate analysis (PCoA) of Bray-Curtis dissimilarity for bacterial (a) and fungal (b) communities, with convex hulls drawn around samples grouped by cave identity (top) and lava flow unit (bottom).

5.3.2 Dispersion analysis

For the 16S rRNA gene dataset, all 11 caves and all 6 flow units had ≥2 samples, so there were no exclusions in these categories. In the fungal dataset, 1 of 10 caves (Jot Dean) and 1 of 6 flow units (Basalt of Giant Crater) were represented by a single sample and so excluded from the filtered dispersion tests. For both datasets, 4 of 10 color categories were represented by single samples: sheen, teal, and white, plus black (ITS dataset) or brown (16S rRNA dataset).

Color showed the strongest dispersion effect of the three variables. Before excluding single-sample color categories, `betadisper` indicated strong dispersion differences for both datasets ($p = 0.001$; permutest). This lessened but remained significant after exclusion (bacteria: $p = 0.011$; fungi: $p = 0.006$; permutest), indicating heterogeneity within color categories beyond what the single-sample groups alone could explain. To investigate this further, subcategories of green biofilm (light green and dark green) were combined into a single "green" category; however, the merged green category showed greater within-group variability, so the original ten-category scheme was retained.

Cave dispersion showed no significant difference for bacteria (betadisper $p = 0.101$; permutest). For fungal data, cave dispersion was significant both when unfiltered ($p = 0.007$) and after excluding the single-sample cave ($p = 0.02$). Lava flow unit showed a significant dispersion effect for both datasets (bacteria: $p = 0.022$; fungi: $p = 0.001$; permutest). Excluding single-sample groups did not change the determination of significance for cave or unit in either dataset.

5.3.3 Pairwise PERMANOVA testing

Pairwise testing between lava flow units (Fig. 9) showed that 11 of 15 pairs (73.3% of unit pairs) for bacterial data and 1 of 15 (6.7%) for fungal data were significant after BH correction ($p_{adjusted} < 0.05$). Blue Dragon Flow was significantly different in every pairwise comparison for bacterial data (5 of 11 pairs; $p_{adjusted} < 0.03$), but only in comparison to Grassy Cone Flow ($p_{adjusted} = 0.03$) for fungal data. Among

cave pairs, 0 of 45 (fungi) and 0 of 55 (bacteria) pairs remained significant after correction.

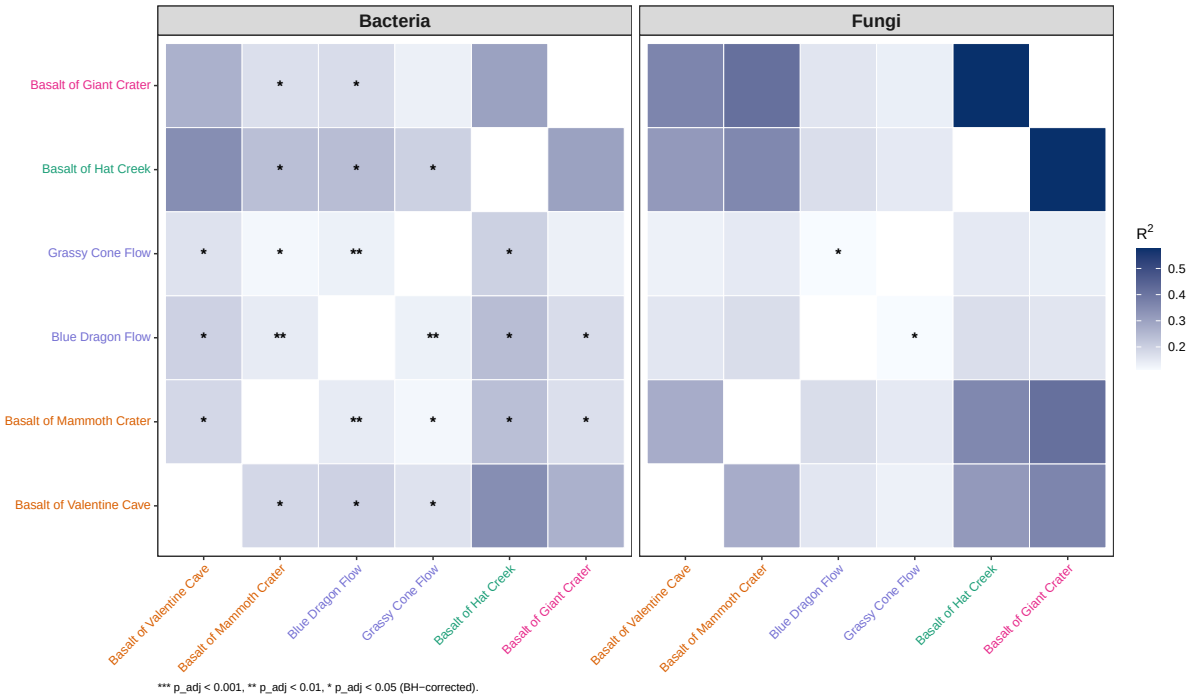

Figure 9: Heatmap of pairwise unit comparisons through PERMANOVA testing. Asterisks indicate BH-corrected p-value magnitude. Label color indicates location: Lava Beds NM in orange, Craters of the Moon NMP in purple, Lassen NF in green, and Shasta-Trinity NF in pink.

5.3.4 Age

At the sample level, lava flow age explained 6-7% of variance in both bacterial and fungal community composition ($p = 0.002$ for bacteria, $p = 0.001$ for fungi). Age is constant within a cave, so the relationship between age and compositional variance was re-tested on cave-level centroids. Cave-level centroid ($n = 11$ caves, bacteria; $n = 10$, fungi) testing revealed that, after correction for pseudoreplication, fungal communities still had a significant relationship with age ($R^2 = 0.135$, $p = 0.030$) while bacterial communities did not ($R^2 = 0.118$, $p = 0.137$).

Alpha diversity showed no relationship with lava flow age. Shannon and Simpson diversity index values for bacterial and fungal ASVs showed no significant association ($p > 0.4$) with either linear or logarithmic lava flow age models. This pattern held whether tested at the individual sample level (Fig. 10) or aggregated lava flow

level (Fig. 11).

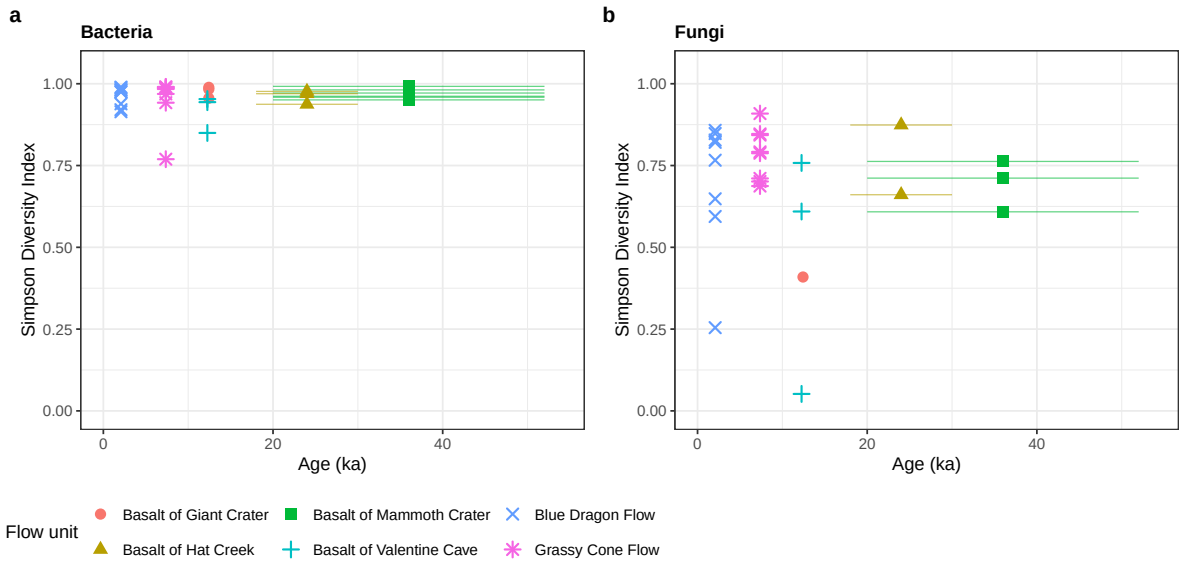

Figure 10: Simpson diversity index versus lava-flow age for (a) bacteria and (b) fungi, one point per sample. Horizontal error bars show age uncertainty.

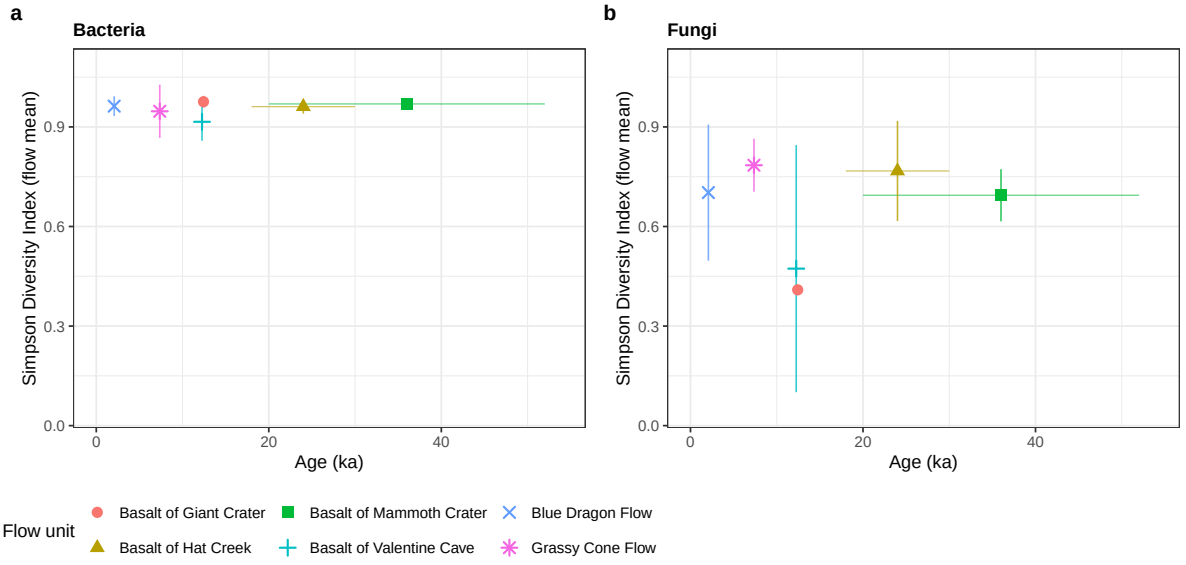

Figure 11: Simpson diversity index versus flow age for (a) bacteria and (b) fungi, averaged within each flow unit. Points show the flow mean. Vertical error bars show ± 1 SD across each flow's samples; horizontal error bars show age uncertainty.

5.3.5 Spatial relationships

Geographic distance was positively correlated with community dissimilarity for bacterial ($r = 0.25$, $p = 0.002$) and fungal ($r = 0.48$, $p = 0.001$) datasets when all

pairwise comparisons were pooled (Fig. 12). Community dissimilarity was substantially lower among samples from the same cave than among samples from different caves. This pattern was found to be significant ($p = 0.001$) for both bacteria (mean within = 0.842, between = 0.965; ANOSIM $R = 0.699$) and fungi (mean within = 0.884, between = 0.987; ANOSIM $R = 0.852$). When testing samples only from different caves, the significant distance-decay relationship was retained for fungi ($r = 0.379$, $p = 0.020$), but not bacteria ($r = 0.128$, $p = 0.185$).

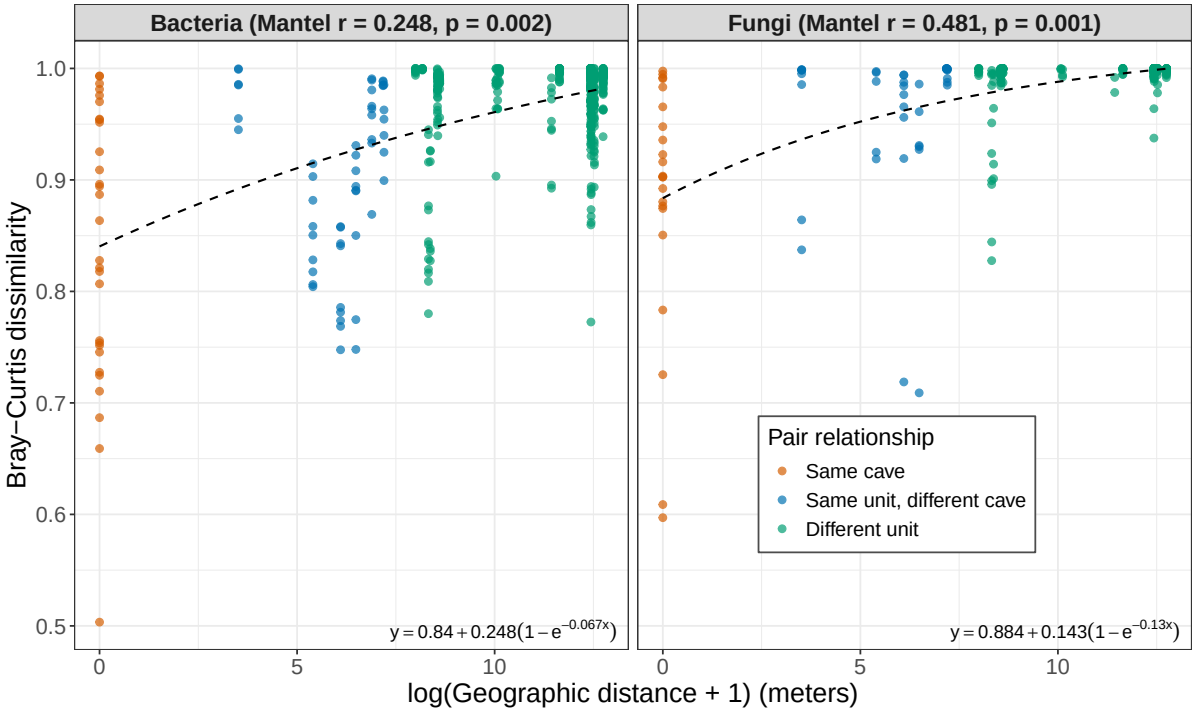

Figure 12: Relationship between geographic distance and Bray–Curtis community dissimilarity for bacterial (left) and fungal (right) communities. Points represent pairwise comparisons among caves, with color indicating pair relationship as described in the legend. Lines show fitted relationships between geographic distance and community dissimilarity. Mantel correlation coefficients and p-values are shown for each dataset. Geographic distance was log-transformed as $\log(\text{distance} + 1, \text{m})$ for visualization only; the 1-m constant was added to accommodate zero-distance comparisons within caves.

To characterize these relationships across spatial and temporal scales, we constructed two sets of Mantel correlograms: one binning cave pairs by geographic distance, the other by flow-age difference.

The shortest geographic distance class (approximately 17 km) showed significant positive correlation ($p_{corrected} = 0.001$) between geographic proximity and com-

munity similarity for both bacteria ($r = 0.29$) and fungi ($r = 0.40$). Flow-age differences showed a similar pattern: caves with similar flow ages had significantly more similar communities (bacteria: $r = 0.39$; fungi: $r = 0.46$; both $p_{corrected} = 0.001$), whereas pairs differing by approximately 12 ka showed significant dissimilarity (bacteria: $r = -0.32$, $p_{corrected} = 0.003$; fungi: $r = -0.29$, $p_{corrected} = 0.003$). Other distance and age-difference classes were either empty because of the clustered sampling design or were not significant after correction.

Pairwise geographic distance and pairwise flow-age difference among caves were correlated (Pearson $r = 0.51$, Spearman $\rho = 0.53$, $n = 11$ caves), indicating that geographic separation and differences in flow age were partially confounded. This confound was disproportionately driven by two neighboring, similarly-aged flows: Blue Dragon Flow and Grassy Cone Flow together accounted for 45 of 55 cave pairs (81.8%) in the bacterial dataset and 39 of 45 cave pairs (86.7%) in the fungal dataset. Excluding caves belonging to either flow, the correlation was weak and negative for both datasets (bacteria: $r = -0.212$, $\rho = -0.226$, $n = 10$ pairs; fungi: $r = -0.324$, $\rho = -0.429$, $n = 6$ pairs), compared to $r = 0.510$–0.526 and $\rho = 0.538$–0.582 with all caves included.

5.4 Neutral models

The majority of bacterial (94.9%) and fungal (95.6%) ASVs fit within the 95% confidence interval envelope around the neutral model prediction (Fig. 13). Approximately 4% of ASVs (bacterial or fungal) occurred more frequently than predicted, and less than 1% occurred less frequently than predicted. At the genus level, 67% of bacterial genera fit within the neutral model prediction envelope, compared to 82% of fungal genera. Fitted immigration parameter (m) values ranged from 10^{-3} to 10^{-6} across ASV and genus fits for both bacteria and fungi.

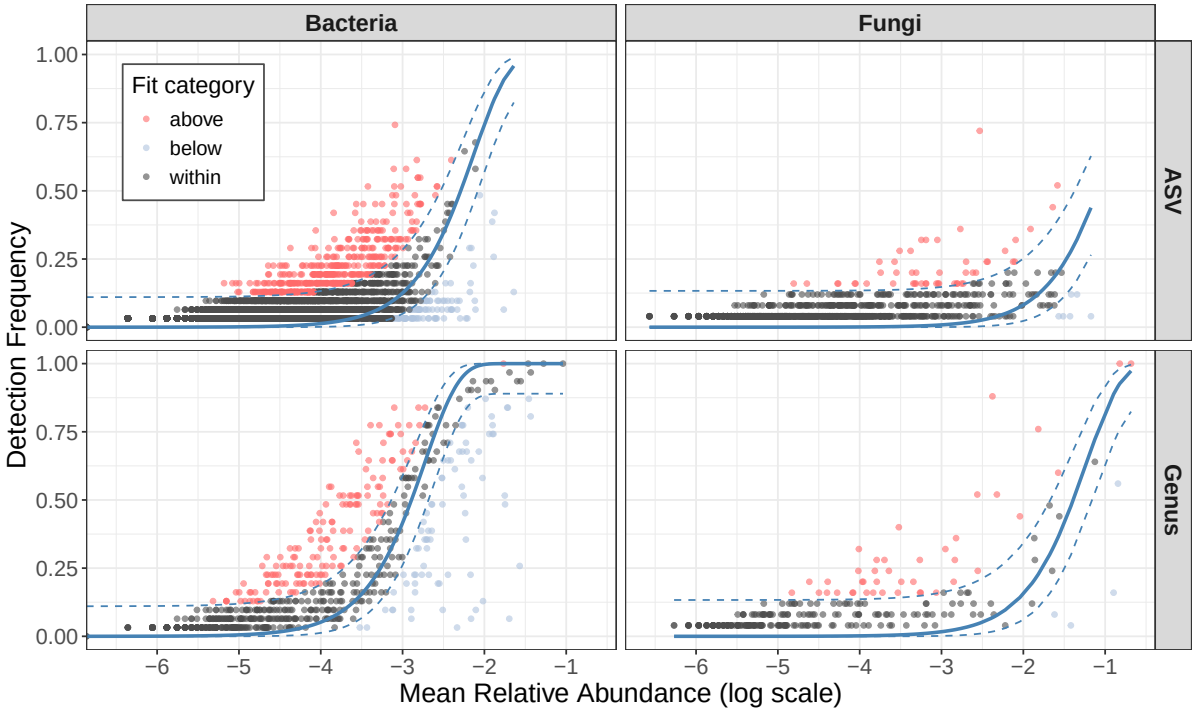

Figure 13: Neutral model fits at ASV (top) and genus (bottom) taxonomic resolutions for the bacterial (left) and fungal (right) datasets. Solid lines represent neutral model predictions and dashed lines indicate 95% confidence intervals.

5.5 Metagenomics

The assembled metagenome from Sunshine Cave's pink biofilm sample (S-P) was 250.98 Mb in total size (42,934 contigs, N_{50} = 7,759 bp). `Prokka` identified 262,378 protein-coding sequences and 262,378 total genes. Contigs were sorted into 55 bins using `MetaBAT2`; of these, 10 metagenome-assembled genomes (MAGs) met the completeness ($\geq$ 90%) and contamination ($\leq$ 5%) filtering thresholds (Fig. 14). All MAGs were assigned to the domain Bacteria, spanning four phyla: Chloroflexota (4 MAGs), Pseudomonadota (3 MAGs), Bacteroidota (2 MAGs), and Actinomycetota (1 MAG).

Figure 14: Summary of properties of the metagenome-assembled genomes (MAGs) from a pink biofilm taken from Sunshine Cave, Lava Beds National Monument, CA.

MAG Name	GTDB classification	Size (Mb)	No. of contigs	N_{50} (bp)	GC content (%)	Coverage (×)	Completeness (%)	Contamination (%)	No. of protein-coding genes	Coding density (%)
bin.007	Bacteria; Chloroflexota; Chloroflexia; Thermomicrobiales; UBA6265	3.64	239.00	20156.00	58.76	67.04	94.45	1.42	3445.00	86.43
bin.014	Bacteria; Chloroflexota; Chloroflexia; Thermomicrobiales; UBA6265; *JACCVC01*	4.22	132.00	50130.00	63.01	25.48	93.08	2.36	3956.00	85.66
bin.018	Bacteria; Bacteroidota; Bacteroidia; Cytophagales; Cyclobacteriaceae	5.69	326.00	24031.00	37.64	35.29	96.43	3.57	4882.00	79.36
bin.025	Bacteria; Chloroflexota; Chloroflexia; Thermomicrobiales; CFX8; *JACCZH01*	3.46	216.00	21217.00	59.56	80.69	95.33	2.34	3474.00	89.12
bin.031	Bacteria; Actinomycetota; Acidimicrobiia; Acidimicrobiales; JAAYBP01; *Iamia*	3.72	376.00	13735.00	71.45	13.99	92.59	1.28	3850.00	88.79
bin.033	Bacteria; Pseudomonadota; Alphaproteobacteria; Caulobacterales; Caulobacteraceae; *Brevundimonas*	2.34	190.00	15287.00	67.04	124.48	92.60	1.96	2450.00	89.00
bin.034	Bacteria; Chloroflexota; Chloroflexia; Thermomicrobiales; UBA6265; *CADCWI01*	3.94	114.00	53928.00	64.29	94.14	95.91	0.94	3526.00	84.23
bin.038	Bacteria; Pseudomonadota; Alphaproteobacteria; Caulobacterales; Caulobacteraceae; *Brevundimonas*	2.72	87.00	47793.00	68.89	394.45	94.71	1.43	2705.00	89.47
bin.042	Bacteria; Pseudomonadota; Alphaproteobacteria; Sphingomonadales; Sphingomonadaceae; *Allosphingosinicella*	2.91	87.00	52063.00	65.79	26.79	97.39	2.16	3006.00	89.86
bin.050	Bacteria; Bacteroidota; Bacteroidia; Cytophagales; Cyclobacteriaceae; *ANT-B6*	5.10	324.00	21721.00	41.83	27.55	90.33	3.13	4636.00	82.48

Metabolic potential encoded in this biofilm community was assessed using DRAM annotation of the 10 recovered MAGs (Fig. 15); the 48 absent functions are not shown. Notable missing functions include photosynthesis, nitrogen and sulfur metabolisms, methanogenesis and methanotrophy, CAZy, the second stage of arsenate reduction, and conversions of butyrate, propionate, and acetate.

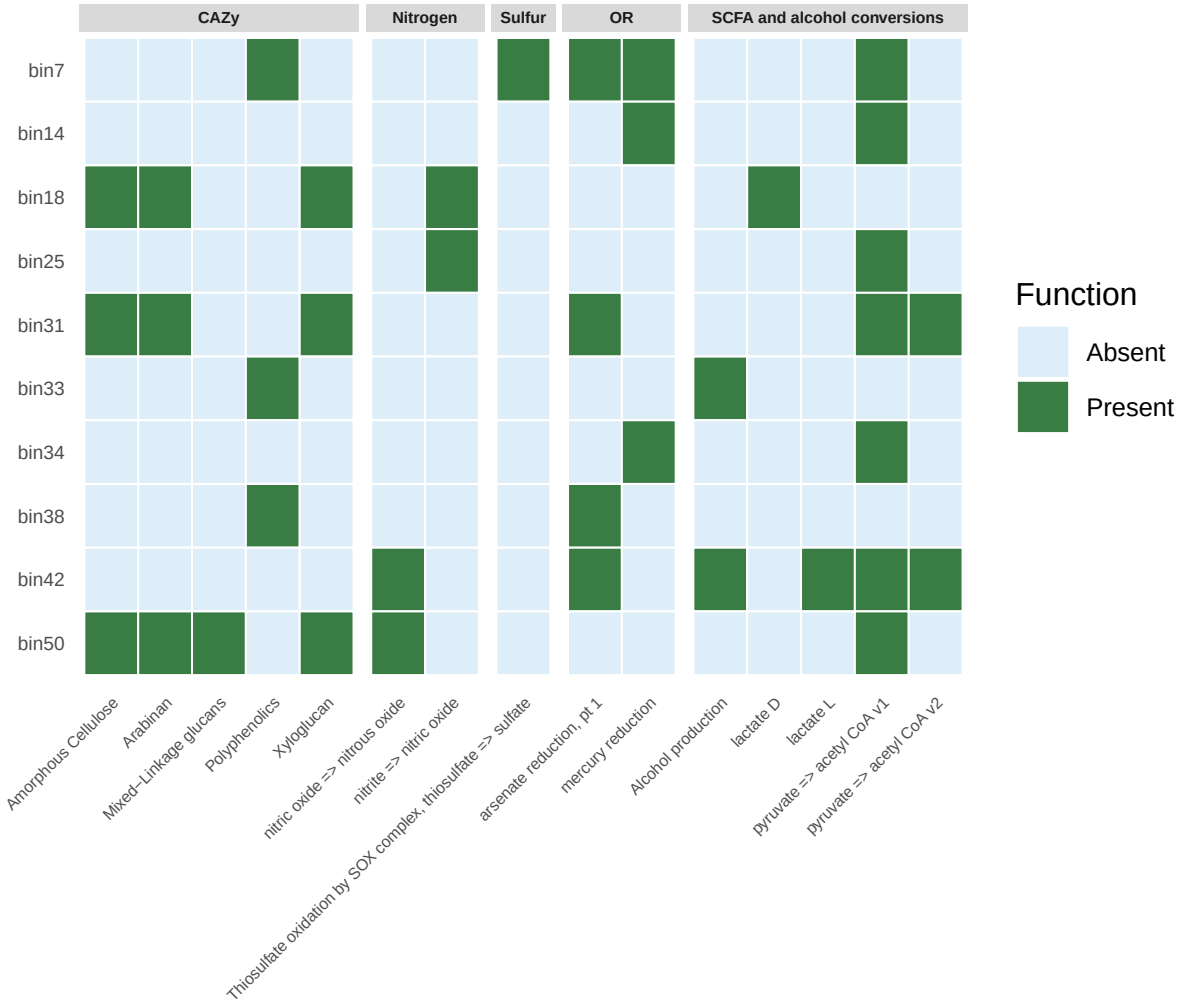

Figure 15: Heatmap of each DRAM-annotated MAG, showing presence or absence of relevant metabolic functions: CAZy (carbohydrate-active enzymes), nitrogen and sulfur metabolisms, OR (other reductases), and SCFA (short-chain fatty acids) and alcohol conversions.

The non-gapfilled draft metagenome-scale community metabolic model (MCMM) constructed from the unbinned metagenome yielded 85 biochemical reactions and 214 metabolites connected by those reactions. With gapfilling, there were 368 reactions and 478 metabolites.

6 Discussion

6.1 Cave identity best explained community structure

Cave identity explained approximately half of compositional variation in both bacterial and fungal amplicon datasets, indicating strong cave-level structuring. How-

ever, because caves are nested within lava flows, the sequential PERMANOVA could not estimate an independent effect of flow unit once cave identity was included. We therefore evaluated flow-level structure using cave-level centroids and ANOSIM. Both types of analyses showed that caves within the same lava flow were significantly more similar in community composition than caves from different flows, demonstrating that the observed flow-level structure is not an artifact of repeated within-cave sampling. Together, these results suggest that caves within the same lava flow share characteristics that influence community assembly, although the specific mechanisms (e.g., shared geochemistry, dispersal limitation, etc.) cannot be distinguished with the collected data alone.

The lack of an association between biofilm color and community composition supports cave and flow as the primary levels of community structure. Biofilms of the same color occurred in different caves but did not necessarily harbor similar communities. Although biofilm color can consistently predict community composition in some cave systems [84, 137], this is not always the case. Imprecision in assigning color categories through field observation is one cause for inconsistency, as apparent macroscopic similarity does not necessarily translate to microscopic structure [64]. Color may be instead indicative of metabolites or other chemical phenomena, rather than a characteristic of specific community members [67]. Consistent with this possibility, we found few genus-level similarities among pink biofilms. Color is an immediately visible and therefore seemingly useful characteristic for distinguishing biofilms, but other aspects of their physical appearance, such as filament structure, may be similarly diverse yet unreliable predictors of community composition [64].

Interpretation of these group-level effects also requires consideration of within-group spread. Within-group spread differed by group for lava flow unit, in both datasets, and for cave, in fungi. These spread differences are visible as uneven hull sizes in Fig. 8. Part of the unit effect and the fungal cave effect may reflect how tightly samples cluster within a group, rather than differences in typical composition between groups. The cave effect in bacterial communities, with homogeneous spread,

reflects only centroid separation and therefore discernible differences in composition by group.

6.1.1 Taxonomic contributors to cave-level differences

Several taxa contributed substantially to pairwise compositional differences among caves. Bacterial ASVs assigned to *Rubrobacter* (Actinomycetota: Rubrobacteriaceae) were disproportionately represented among taxa driving pairwise compositional differences, with ten distinct *Rubrobacter* ASVs recurring among the top SIMPER contributors across a majority of cave-pair comparisons. *Rubrobacter*'s known tolerance to desiccation [138, 139] is consistent with its prominence in the high desert study sites.

In some cases, community distinctiveness was attributable to a single ASV. The bacterial ASV *Leptolyngbya* sp. (ASV1) alone accounted for 25-65% of the pairwise Bray-Curtis dissimilarity between Spongy Floor and every other cave in the bacterial dataset. Similarly, a single lichenized fungal ASV (*Verrucaria othmarbreussii*, ASV2) accounted for 27–48% of the Bray-Curtis dissimilarity between Indian Tunnel and every other cave in the fungal dataset.

Besides *Verrucaria othmarbreussii*, the two ASVs contributing most consistently to compositional differences in the fungal dataset (present among the top SIMPER contributors in 25/45 and 18/45 cave-pair comparisons, respectively) could not be resolved beyond Kingdom Fungi in the UNITE database. The lack of taxonomic resolution for these ASVs could reflect novel taxa and/or database underrepresentation. Consequently, the contribution of these ASVs to cave-level differentiation could not be interpreted taxonomically.

6.1.2 Predicted functional profiles

Phototrophy-associated taxa, present in 84–87% of samples, were an expected majority, given the sampling locations' proximity to cave entrances and skylights. The prevalence of taxa associated with chemoheterotrophy suggests that communi-

ties are capable of maximizing the limited organic input [140]. This mix of chemoheterotrophic and photosynthetic functional groups is consistent with findings from other basaltic lava tube communities [83].

6.2 Assembly processes underlying cave-level structure

Community composition in these caves was broadly consistent with stochastic assembly and dispersal limitation, supported by three lines of evidence. First, the majority of bacteria and fungi fit within neutral model predictions regardless of taxonomic rank, though fit was greater at the ASV level than at the genus level (Fig. 13). Second, communities also appear to be dispersal-limited, as calculated immigration probability values were close to zero. Low values indicate that local vacancies are filled primarily by local reproduction rather than external immigration. Dispersal limitation appears to be stronger in fungi than bacteria: only 1 of 256 fungal genera (0.4%), compared with 24 of 687 bacterial genera (3.5%), occurred in $\geq$75% of sampled communities. Community dissimilarity also increased with geographic distance, but after excluding within-cave comparisons, the distance-decay pattern persisted only for fungi. Lastly, cave identity explained more variance than any measured deterministic covariate (age, biofilm color).

While these results are consistent with stochastic community assembly, there are important considerations. Unmeasured cave-specific conditions such as moisture or airflow could provide a deterministic explanation for the strong relationship between cave identity and community composition [29]. Community composition can also vary substantially even when diversity metrics remain relatively stable [25, 141]. The link between function and taxonomy can be tenuous for microbes [142], and as such, relying on taxonomy (i.e., using ASV- or genus-level occurrence frequency) to interpret patterns of microbial community assembly is likely to erroneously imply high stochasticity. Thus, our results are consistent with predominantly stochastic assembly, but taxonomy-based methods cannot fully distinguish stochastic variation in taxonomic identity from deterministic selection acting on function.

6.3 Community composition changes with age, but diversity does not

Flow age was significantly associated with community composition in fungi and bacteria, though it explained little of the variation. In contrast to community composition, flow age was not significantly associated with diversity. This is likely due to two reasons: (i) Simpson's Index of Diversity has a maximum value of 1; most sampled bacterial communities were close to this value (Fig. 10) and (ii) even the youngest lava flows sampled are too old to observe low-diversity communities. Colonization and initial community assembly occur soon after lava emplacement, even before cooling is completed [7]. We hypothesize that communities may quickly reach a relatively stable level of diversity that subsequently changes little across the age range sampled here.

Our results are consistent with this process, suggesting that lava flow age is associated with shifts in which taxa are present (beta diversity) without a corresponding change in overall diversity or evenness (alpha diversity). Microbial succession in these cave communities therefore could involve taxonomic turnover rather than a simple, progressive increase in diversity. For fungi, this pattern of compositional change without a corresponding change in diversity is consistent with taxonomic turnover rather than progressive accumulation of diversity [143]. The bacterial cave-level age effect did not hold, so this succession pattern is not clearly supported for bacteria in this dataset.

However, taxonomic turnover does not necessarily imply functional turnover, raising the possibility that the compositional changes observed across the age gradient reflect changes in community membership without corresponding changes in ecological function. Bacterial communities with different taxonomic compositions can nonetheless possess highly similar functional compositions [142]. Functional equivalence may therefore preserve ecosystem processes despite taxonomic turnover or stochastic taxonomic assembly.

The difference in ASV- and genus-level neutral model fit could reflect genus-level identity capturing functional traits and/or core metabolic strategies needed for establishment, while ASVs within a genus vary in other, non-core traits [144]. Environ-

mental filtering may therefore act more strongly on genus-level functional strategies than on finer-scale variation among ASVs within genera. This combination of filtering and stochasticity among functionally equivalent options is consistent with the competitive lottery hypothesis [145, 146]. The competitive lottery hypothesis would predict the rank-dependent pattern observed here, although distinguishing environmental filtering from competition remains difficult using observational data alone [147].

Under this interpretation, an apparently random taxonomic swap could therefore represent deterministic replacement of one organism by a functional equivalent. Taxonomic identity can mask such deterministic signals; niche differentiation may only be detectable when within-species trait variability is included [144]. Similarly, in a cross-habitat meta-analysis of bacterial primary succession [148], communities that diverged sharply in taxonomic composition were more similar in functional gene content, indicating functional convergence among taxonomically distinct assemblages. Phylogenetic or taxonomic relatedness may be a weak proxy for trait similarity, particularly for microbes, since a single ecological function can be performed by unrelated taxa [149]. Together, these findings suggest that the age-associated taxonomic differences observed in this study may reflect changes in the identities of community members over time while maintaining broadly similar metabolic capabilities. As this study inferred function primarily from taxonomic composition, additional genome-resolved metagenomic analyses across multiple caves are necessary to determine whether functional turnover accompanies taxonomic succession.

6.3.1 Metagenomics

The available metagenomic data from one sample provide an opportunity to explore whether taxonomic composition is a reliable proxy for functional potential in our dataset. All four MAG phyla were also detected in the 16S rRNA gene dataset, consistent with the MAGs representing a small subset of the broader community rather than a taxonomically distinct one. Overlap at the family and genus level was limited, which is expected given that only 10 of 55 recovered bins passed quality filtering and

that genus-level taxonomy is more database-dependent than phylum-level taxonomy. *Iamia*, for example, was detected by both methods but assigned to different families (GTDB: JAAYBP01; SILVA: Iamiaceae).

The two draft community metabolic models further illustrate this incompleteness. The non-gapfilled model built directly from `Prokka`-annotated genes in the assembly supported only 85 reactions and 214 metabolites. However, a gapfilled model suggested a four-fold increase was necessary to meet the minimum set of additional reactions for the community to produce biomass on a complete medium. This discrepancy suggests that most of the reactions required for a self-consistent, growth-capable metabolic network may be present in MAGs that either were not recovered or did not meet quality thresholds. The non-gapfilled model thus serves as a conservative estimate of this community's metabolic potential, and the gapfilled model as an upper bound on what may be biologically present but undetected.

Several functions predicted from this same sample's 16S rRNA gene data (e.g., nitrate reduction, aromatic compound degradation) were consistent with `DRAM` annotations of the recovered MAGs, while others (nitrogen and sulfur metabolism, absent from the `DRAM`-annotated MAGs) were not. These differences, even within the same sample, underline the gap between taxonomy-based functional prediction from sequence data and genome-resolved annotation.

6.4 Limitations

Necessitated by cave conservation and permitting constraints, all tested experimental design categories had small sample sizes. A lack of biological replicates limited statistical power and prevented assessment of within-biofilm variability. Confidence in comparisons among color-defined communities is limited due to the qualitative assignment of color categories in the field. Additionally, several color categories (e.g., sheen, teal, white, brown, and black) were represented by single samples. None of Pluto's Cave samples yielded sufficient DNA to sequence. Of the samples with sufficient DNA, sequencing failures further reduced sample sizes available for down-

stream analyses, as only 86% of samples passed 16S rRNA gene sequencing and 69% passed ITS region sequencing. Because of the small sample sizes and non-parametric nature of the datasets, permutation-based tests were used, but there remains the problem of limited statistical power.

Caves were nested within lava flow units, so cave identity and flow unit were not independent predictors. This was addressed by (i) testing flow-unit differences using cave-level centroids rather than treating individual samples within caves as independent replicates and (ii) using the "strata" argument in the "adonis2" function during FAPROTAX PERMANOVA testing. However, with approximately 3 samples per cave, strata-restricted permutations are limited (6 within-cave arrangements per cave), so permutation tests have limited statistical power. *P*-values should thus be interpreted as indicative rather than highly precise.

Fungal taxa were less diverse at every taxonomic rank, so it is harder to draw conclusions from these data (beyond ASV) than from the bacteria. Additionally, taxonomic resolution was generally poor at all ranks for the fungal data. Two major sources of taxonomic uncertainty were present in the form of (i) sequences unclassified by the UNITE database, and (ii) hierarchical ambiguity classified as *incertae sedis*. Of the 1,171 fungal ASVs, 309 (26.4% of all fungal ASVs and 20.7% of reads) were unclassified at the phylum level using the UNITE database. This means that a substantial fraction of fungal sequences belong to taxa that cannot be assigned to a fungal phylum. This reflects limitations of reference database coverage for environmental fungal diversity; future in-depth phylogenetic analysis could further identify lineages associated with ASVs. Sequence data from this study will be uploaded to a public repository. This is particularly valuable because cave conservation inherently limits sampling efforts, so our sampling campaign offers opportunities for other researchers in this field.

The primary use of 16S rRNA gene and ITS region sequencing, rather than additional metagenomics, was chosen due to cost and logistic advantages, but this hinders the ability to interpret true functional roles. A trait-based approach would allow ex-

amination of actual functional diversity, which better reflects the underlying ecology rather than the primarily taxonomic view provided here.

7 Conclusions

Our results indicate that microbial community assembly in basaltic lava caves was shaped by both spatial structure and stochastic processes, including dispersal limitation and ecological drift. Lava flow age was not associated with alpha diversity, although it was associated with community composition, and cave identity was the strongest measured predictor of community composition, explaining more variance than biofilm color, lava flow unit, or lava flow age.

Although cave identity explained more compositional variance than lava flow unit, caves within the same flow unit were significantly more similar to one another than caves from different flow units for both bacteria and fungi. Thus, community composition retains a signal of shared lava flow history despite substantial differentiation among individual caves. The observed patterns were broadly, though not unambiguously, consistent with stochastic assembly, particularly given the evidence for dispersal limitation and the fit of most taxa to neutral-model predictions.

The strong cave-level effect indicates that individual caves, even within the same lava flow, represent distinct microbial microenvironments. The mechanisms underlying this differentiation cannot be resolved from the variables measured here. Fine-scale environmental properties such as geochemistry, mineralogy, porosity, and hydrology were not characterized but may provide important explanations for the observed cave-specific community structure. To link biogeochemical processes with assembly patterns, our future work will investigate weathering signatures using spatially-resolved element mapping of the biofilm-rock interface.

The fungal relationship between lava flow age and community composition, despite the absence of a relationship with diversity, could reflect taxonomic turnover among functionally similar taxa [142]. This interpretation is also compatible with the dispersal limitation identified for fungi. A combination of dispersal limitation and

functional equivalence could therefore buffer ecosystem processes through succession, although functional equivalence was not directly tested here. More generally, taxonomic and functional turnover need not coincide, meaning that taxonomy-based evidence for stochastic assembly cannot rule out deterministic selection operating at the level of function rather than taxonomic identity. Trait- or gene-resolved analyses will be needed to distinguish these possibilities in future work.

This multi-cave, multi-flow-age assessment of microbial community assembly in basaltic lava caves demonstrates the importance of accounting for nested sampling structure when evaluating the relative contributions of cave and lava flow. The genome-resolved data yield an initial functional and genomic baseline for these sensitive and still largely under-characterized habitats. These results highlight individual caves as important units of microbial biogeographic variation while providing a framework for investigating how local environmental heterogeneity, lava flow history, and stochastic processes interact to shape microbial communities in subsurface basaltic environments.

References

1. Banfield, J. F., Barker, W. W., Welch, S. A. & Taunton, A. Biological impact on mineral dissolution: Application of the lichen model to understanding mineral weathering in the rhizosphere. *Proceedings of the National Academy of Sciences* **96,** 3404–3411 (1999).
2. Barker, W. W., Welch, S. A., Chu, S. & Banfield, J. F. Experimental observations of the effects of bacteria on aluminosilicate weathering. *American Mineralogist* **83,** 1551–1563 (1998).
3. Stranghoener, M., Schippers, A., Dultz, S. & Behrens, H. Experimental Microbial Alteration and Fe Mobilization From Basaltic Rocks of the ICDP HSDP2 Drill Core, Hilo, Hawaii. English. *Frontiers in Microbiology* **9** (2018).
4. Jackson, T. A. & Keller, W. D. A comparative study of the role of lichens and "inorganic" processes in the chemical weathering of Recent Hawaiian lava flows. *American Journal of Science* **269** (1970).
5. Odum, E. P. in *The Ecological Design and Planning Reader* (ed Ndubisi, F. O.) 203–216 (Island Press/Center for Resource Economics, 1969).
6. Jones, A. A. & Bennett, P. C. Mineral Microniches Control the Diversity of Subsurface Microbial Populations. *Geomicrobiology Journal* **31,** 246–261 (2014).
7. Hadland, N., Hamilton, C. & Duhamel, S. Young volcanic terrains are windows into early microbial colonization. *Communications Earth and Environment* **5** (2024).
8. Wild, B., Gerrits, R. & Bonneville, S. The contribution of living organisms to rock weathering in the critical zone. *npj Materials Degradation* **6,** 1–16 (2022).
9. Rogers, J. R. & Bennett, P. C. Mineral stimulation of subsurface microorganisms: release of limiting nutrients from silicates. *Chemical Geology* **203,** 91–108 (2004).
10. Dong, H. *et al.* A critical review of mineral–microbe interaction and co-evolution: mechanisms and applications. *National Science Review* **9,** nwac128 (2022).
11. Wu, L., Jacobson, A. D., Chen, H.-C. & Hausner, M. Characterization of elemental release during microbe–basalt interactions at $T = 28$ °C. *Geochimica et Cosmochimica Acta* **71,** 2224–2239 (2007).
12. Wang, P.-L. *et al.* Microbial communities modulate chemical weathering and carbon dioxide cycling in an active orogen in Taiwan. *Communications Earth & Environment* **5,** 1–10 (2024).
13. Selensky, M. J., Masterson, A. L., Blank, J. G., Lee, S. C. & Osburn, M. R. Stable Carbon Isotope Depletions in Lipid Biomarkers Suggest Subsurface Carbon Fixation in Lava Caves. *Journal of Geophysical Research: Biogeosciences* **126,** e2021JG006430 (2021).
14. Wierzchos, J. & Ascaso, C. Application of back-scattered electron imaging to the study of the lichen-rock interface. *Journal of Microscopy* **175,** 54–59 (1994).
15. Zhu, H.-Z., Jiang, C.-Y. & Liu, S.-J. Microbial roles in cave biogeochemical cycling. English. *Frontiers in Microbiology* **13** (2022).

16. Doc Richardson, C. *et al.* Secondary sulfate mineralization and basaltic chemistry of craters of the Moon National Monument, Idaho: Potential martian analog. *Planetary and Space Science* **65,** 93–103 (2012).

17. Adamo, P. & Violante, P. Weathering of rocks and neogenesis of minerals associated with lichen activity. *Applied Clay Science* **16,** 229–256 (2000).

18. Welch, S. A., Barker, W. W. & Banfield, J. F. Microbial extracellular polysaccharides and plagioclase dissolution. *Geochimica et Cosmochimica Acta* **63,** 1405–1419 (1999).

19. Welch, S. A. & Banfield, J. F. Modification of olivine surface morphology and reactivity by microbial activity during chemical weathering. *Geochimica et Cosmochimica Acta* **66,** 213–221 (2002).

20. Benzerara, K. *et al.* Experimental Colonization and Alteration of Orthopyroxene by the Pleomorphic Bacteria Ramlibacter tataouinensis. *Geomicrobiology Journal* **21,** 341–349 (2004).

21. Podia, M., Yadav, P., Hooda, S., Diwan, P. & Gupta, R. K. in *Weathering and Erosion Processes in the Natural Environment* 265–302 (John Wiley & Sons, Ltd, 2023).

22. E. Northup Diana, K. H. L. Geomicrobiology of Caves: A Review. *Geomicrobiology Journal* **18,** 199–222 (2001).

23. Lovley, D. R. & Phillips, E. J. P. Novel Mode of Microbial Energy Metabolism: Organic Carbon Oxidation Coupled to Dissimilatory Reduction of Iron or Manganese. *Applied and Environmental Microbiology* **54,** 1472–1480 (1988).

24. Characklis, W. G. & Marshall, K. C. eng. in *Biofilms* 93–130 (Wiley, 1990).

25. Nemergut, D. R. *et al.* Patterns and Processes of Microbial Community Assembly. *Microbiology and Molecular Biology Reviews* **77,** 342–356 (2013).

26. Prosser, J. I. *et al.* The role of ecological theory in microbial ecology. *Nature Reviews Microbiology* **5,** 384–392 (2007).

27. Hanson, C. A., Fuhrman, J. A., Horner-Devine, M. C. & Martiny, J. B. H. Beyond biogeographic patterns: processes shaping the microbial landscape. *Nature Reviews Microbiology* **10,** 497–506 (2012).

28. Woodcock, S. *et al.* Neutral assembly of bacterial communities. *FEMS Microbiology Ecology* **62,** 171–180 (2007).

29. Stegen, J. C., Lin, X., Konopka, A. E. & Fredrickson, J. K. Stochastic and deterministic assembly processes in subsurface microbial communities. *The ISME Journal* **6,** 1653–1664 (2012).

30. Vellend, M. Conceptual Synthesis in Community Ecology. *The Quarterly review of biology* **85,** 183–206 (2010).

31. Kraft, N. J. B. *et al.* Community assembly, coexistence and the environmental filtering metaphor. *Functional Ecology* **29,** 592–599 (2015).

32. Cadotte, M. W. & Tucker, C. M. Embracing the Nonindependence of the Environmental Filter: A Reply to Responses. *Trends in Ecology & Evolution* **32,** 886–887 (2017).

33. Thakur, M. P. & Wright, A. J. Environmental Filtering, Niche Construction, and Trait Variability: The Missing Discussion. *Trends in Ecology & Evolution* **32,** 884–886 (2017).

34. Fodelianakis, S., Valenzuela-Cuevas, A., Barozzi, A. & Daffonchio, D. Direct quantification of ecological drift at the population level in synthetic bacterial communities. *The ISME Journal* **15,** 55–66 (2021).

35. Caswell, H. Community Structure: A Neutral Model Analysis. *Ecological Monographs* **46,** 327–354 (1976).

36. Hubbell, S. P. A unified theory of biogeography and relative species abundance and its application to tropical rain forests and coral reefs. *Coral Reefs* **16,** S9–S21 (1997).

37. Chave, J. Neutral theory and community ecology. *Ecology Letters* **7,** 241–253 (2004).

38. Bell, G. The Distribution of Abundance in Neutral Communities. *The American Naturalist* **155,** 606–617 (2000).

39. Woodcock, S., Curtis, T. P., Head, I. M., Lunn, M. & Sloan, W. T. Taxa–area relationships for microbes: the unsampled and the unseen. *Ecology Letters* **9,** 805–812 (2006).

40. O'Malley, M. A. The nineteenth century roots of 'everything is everywhere'. *Nature Reviews Microbiology* **5,** 647–651 (2007).

41. O'Malley, M. A. '*Everything is everywhere*: but *the environment selects*': ubiquitous distribution and ecological determinism in microbial biogeography. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences* **39,** 314–325 (2008).

42. Cockell, C. S. Are microorganisms everywhere they can be? *Environmental Microbiology* **23,** 6355–6363 (2021).

43. Baas Becking, L. *Geobiologie of Inleiding tot de Milieukunde* (W.P. Van Stockum & Zoon: The Hague, The Netherlands, 1934).

44. Telford, R. J., Vandvik, V. & Birks, H. J. B. Dispersal Limitations Matter for Microbial Morphospecies. *Science* **312,** 1015–1015 (2006).

45. Zhou, J. & Ning, D. Stochastic Community Assembly: Does It Matter in Microbial Ecology? *Microbiology and Molecular Biology Reviews* **81,** 10.1128/mmbr.00002–17 (2017).

46. Archer, S. D. J. *et al.* Airborne microbial transport limitation to isolated Antarctic soil habitats. *Nature Microbiology* **4,** 925–932 (2019).

47. Hubbell, S. P. Neutral Theory and the Evolution of Ecological Equivalence. *Ecology* **87,** 1387–1398 (2006).

48. Cottenie, K. Integrating environmental and spatial processes in ecological community dynamics. *Ecology Letters* **8,** 1175–1182 (2005).

49. Leibold, M. A. & McPeek, M. A. Coexistence of the Niche and Neutral Perspectives in Community Ecology. *Ecology* **87,** 1399–1410 (2006).

50. Tilman, D. Niche tradeoffs, neutrality, and community structure: A stochastic theory of resource competition, invasion, and community assembly. *Proceedings of the National Academy of Sciences* **101,** 10854–10861 (2004).

51. Lortie, C. J. *et al.* Rethinking plant community theory. *Oikos* **107,** 433–438 (2004).

52. Van Der Gast, C. J., Ager, D. & Lilley, A. K. Temporal scaling of bacterial taxa is influenced by both stochastic and deterministic ecological factors. *Environmental Microbiology* **10,** 1411–1418 (2008).

53. Blanckenburg, F. v. *et al.* Rock Weathering and Nutrient Cycling Along an Erodosequence. *American Journal of Science* **321** (2021).

54. Frings, P. J., Schubring, F., Oelze, M. & Blanckenburg, F. v. Quantifying Biotic and Abiotic Si Fluxes in the Critical Zone With Ge/Si Ratios Along a Gradient of Erosion Rates. *American Journal of Science* **321** (2021).

55. De Wit, R. Does the Environment "Filter" or "Select" Species? Bridging the Ecologies of Microbes and Macro-Organisms for a Common Niche Assembly Theory. *Environments* **12,** 350 (2025).

56. Cowles, H. C. The Ecological Relations of the Vegetation on the Sand Dunes of Lake Michigan. Part I.-Geographical Relations of the Dune Floras. *Botanical Gazette* **27,** 95–117 (1899).

57. Connell, J. H. & Slatyer, R. O. Mechanisms of Succession in Natural Communities and Their Role in Community Stability and Organization. *The American Naturalist* **111,** 1119–1144 (1977).

58. Clements, F. E. *Plant Succession: An Analysis of the Development of Vegetation* (Carnegie Institution of Washington, 1916).

59. Thorseth, I., Furnes, H. & Heldal, M. The importance of microbiological activity in the alteration of natural basaltic glass. *Geochimica et Cosmochimica Acta* **56,** 845–850 (1992).

60. Dessert, C., Dupré, B., Gaillardet, J., François, L. M. & Allègre, C. J. Basalt weathering laws and the impact of basalt weathering on the global carbon cycle. *Chemical Geology* **202,** 257–273 (2003).

61. Navarre-Sitchler, A. & Brantley, S. Basalt weathering across scales. *Earth and Planetary Science Letters* **261,** 321–334 (2007).

62. Nesbitt, H. W. & Wilson, R. E. Recent chemical weathering of basalts. *American Journal of Science* **292** (1992).

63. Babechuk, M. G., Widdowson, M. & Kamber, B. S. Quantifying chemical weathering intensity and trace element release from two contrasting basalt profiles, Deccan Traps, India. *Chemical Geology* **363,** 56–75 (2014).

64. Northup, D. *et al.* Lava Cave Microbial Communities Within Mats and Secondary Mineral Deposits: Implications for Life Detection on Other Planets. *Astrobiology* **11,** 601–618 (2011).

65. Herrera-Belaroussi, A. *et al.* Bacterial Colonization and Weathering of Terrestrial Obsidian in Iceland. *Geomicrobiology Journal* **25,** 25–37 (2008).

66. Lavoie, K. H. *et al.* Comparison of bacterial communities from lava cave microbial mats to overlying surface soils from Lava Beds National Monument, USA. *PLOS ONE* **12,** e0169339 (2017).

67. Hathaway, J. J. M. *et al.* Comparison of Bacterial Diversity in Azorean and Hawai'ian Lava Cave Microbial Mats. *Geomicrobiology Journal* **31,** 205–220 (2014).

68. Kelly, L. C. *et al.* Bacterial Diversity of Weathered Terrestrial Icelandic Volcanic Glasses. *Microbial Ecology* **60,** 740–752 (2010).

69. Kelly, L. C., Cockell, C. S., Thorsteinsson, T., Marteinsson, V. & Stevenson, J. Pioneer Microbial Communities of the Fimmvörðuháls Lava Flow, Eyjafjallajökull, Iceland. *Microbial Ecology* **68,** 504–518 (2014).

70. Gomez-Alvarez, V., King, G. M. & Nüsslein, K. Comparative bacterial diversity in recent Hawaiian volcanic deposits of different ages. *FEMS Microbiology Ecology* **60,** 60–73 (2007).

71. Yoshitake, S. *et al.* Successional changes in the soil microbial community along a vegetation development sequence in a subalpine volcanic desert on Mount Fuji, Japan. *Plant and Soil* **364,** 261–272 (2013).

72. Bergsten, P. *et al.* Basalt-Hosted Microbial Communities in the Subsurface of the Young Volcanic Island of Surtsey, Iceland. eng. *Frontiers in Microbiology* **12,** 728977 (2021).

73. Sterling, J. J. *et al.* Primary Microbial Succession in the Anchialine Ecosystem. *Integrative And Comparative Biology* **62,** 275–287 (2022).

74. Fishman, C. B. *et al.* Extreme Niche Partitioning and Microbial Dark Matter in a Mauna Loa Lava Tube. *Journal of Geophysical Research: Planets* **128,** e2022JE007283 (2023).

75. Mark Ibekwe, A., Kennedy, A. C., Halvorson, J. J. & Yang, C.-H. Characterization of developing microbial communities in Mount St. Helens pyroclastic substrate. *Soil Biology and Biochemistry* **39,** 2496–2507 (2007).

76. Hopkins, D. W. *et al.* Plant litter decomposition and microbial characteristics in volcanic soils (Mt Etna, Sicily) at different stages of development. *Biology and Fertility of Soils* **43,** 461–469 (2007).

77. Dragone, N. B. *et al.* The Early Microbial Colonizers of a Short-Lived Volcanic Island in the Kingdom of Tonga. *mBio* **14,** e03313–22 (2023).

78. Howarth, F. in, 318–336 (1981).

79. Prescott, R. D. *et al.* Islands Within Islands: Bacterial Phylogenetic Structure and Consortia in Hawaiian Lava Caves and Fumaroles. English. *Frontiers in Microbiology* **13** (2022).

80. Weng, M. M. *et al.* Life Underground: Investigating Microbial Communities and Their Biomarkers in Mars-Analog Lava Tubes at Craters of the Moon National Monument and Preserve. *Journal of Geophysical Research: Planets* **127,** e2022JE007268 (2022).

81. Nicolosi, G., Gonzalez-Pimentel, J. L., Piano, E., Isaia, M. & Miller, A. Z. First Insights into the Bacterial Diversity of Mount Etna Volcanic Caves. *Microbial Ecology* **86,** 1632–1645 (2023).

82. Gutierrez-Patricio, S. *et al.* Microbiological exploration of the Cueva del Viento lava tube system in Tenerife, Canary Islands. *Environmental Microbiology Reports* **16,** e13245 (2024).

83. Palma, V. *et al.* Connecting molecular biomarkers, mineralogical composition, and microbial diversity from Mars analog lava tubes. *Science of The Total Environment* **913,** 169583 (2024).

84. Riquelme, C. *et al.* Cave microbial community composition in oceanic islands: disentangling the effect of different colored mats in diversity patterns of Azorean lava caves. *FEMS Microbiology Ecology* **91,** fiv141 (2015).

85. Ashelford, K. E., Chuzhanova, N. A., Fry, J. C., Jones, A. J. & Weightman, A. J. New Screening Software Shows that Most Recent Large 16S rRNA Gene Clone Libraries Contain Chimeras. *Applied and Environmental Microbiology* **72,** 5734–5741 (2006).

86. Brantley, S. L. *et al.* Twelve testable hypotheses on the geobiology of weathering. *Geobiology* **9,** 140–165 (2011).

87. Riquelme, C. *et al.* Actinobacterial Diversity in Volcanic Caves and Associated Geomicrobiological Interactions. English. *Frontiers in Microbiology* **6** (2015).

88. Junker, R. R. & Farwig, N. Microbial conservation is essential for sustaining ecosystem functions and services. *Proceedings of the National Academy of Sciences* **122,** e2503013122 (2025).

89. Donnelly-Nolan, J. M. *et al.* The Giant Crater Lava Field: Geology and geochemistry of a compositionally zoned, high-alumina basalt to basaltic andesite eruption at Medicine Lake Volcano, California. *Journal of Geophysical Research: Solid Earth* **96,** 21843–21863 (1991).

90. Stout, M. Z., Nicholls, J. & Kuntz, M. A. Petrological and Mineralogical Variations in 2500-2000 yr B.P. Lava Flows, Craters of the Moon Lava Field, Idaho. *Journal of Petrology* **35,** 1681–1715 (1994).

91. Turrin, B. D., Muffler, L. J. P., Clynne, M. A. & Champion, D. E. Robust 24 ± 6 ka 40Ar/39Ar age of a low-potassium tholeiitic basalt in the Lassen region of NE California. *Quaternary Research* **68,** 96–110 (2007).

92. Putirka, K. D., Kuntz, M. A., Unruh, D. M. & Vaid, N. Magma Evolution and Ascent at the Craters of the Moon and Neighboring Volcanic Fields, Southern Idaho, USA: Implications for the Evolution of Polygenetic and Monogenetic Volcanic Fields. *Journal of Petrology* **50,** 1639–1665 (2009).

93. Christiansen, R. L., Calvert, A. T. & Grove, T. L. *Geologic field-trip guide to Mount Shasta Volcano, northern California* tech. rep. 2017-5022-K3 (U.S. Geological Survey, 2017).

94. Chadwick, J., Schwartz, M., Mclane, D., Collins, E. & Kamenov, G. Compositional heterogeneity of the 3.4 km3 Blue Dragon flow, Craters of the Moon Volcanic Field, Idaho. *Journal of Volcanology and Geothermal Research* **388,** 106690 (2019).

95. Barton, H. A. Introduction to cave microbiology: A review for the non-specialist. *Journal of Cave and Karst Studies* **68,** 43–54 (2006).

96. Forti, P. Genetic processes of cave minerals in volcanic environments: An overview. *Journal of Cave and Karst Studies* **67,** 3–13 (2005).

97. Nathenson, M., Donnelly-Nolan, J., Champion, D. & Lowenstern, J. *Chronology of postglacial eruptive activity and calculation of eruption probabilities for Medicine Lake volcano, northern California* tech. rep. (U.S. Geological Survey, 2007), 10.

98. Donnelly-Nolan, J. *Geologic map of Medicine Lake volcano, northern California* Scientific Investigations Map (2010).

99. Kuntz, M. A., Champion, D. E., Spiker, E. C. & Lefebvre, R. H. Contrasting magma types and steady-state, volume-predictable, basaltic volcanism along the Great Rift, Idaho. *Geological Society of America Bulletin* **97,** 579 (1986).

100. Kuntz, M. A., Spiker, E. C., Rubin, M., Champion, D. E. & Lefebvre, R. H. Radiocarbon Studies of Latest Pleistocene and Holocene Lava Flows of the Snake River Plain, Idaho: Data, Lessons, Interpretations. *Quaternary Research* **25,** 163–176 (1986).

101. *BCL-Convert: a proprietary Illumina software for the conversion of BCL files to basecalls.*

102. Callahan, B. J. *et al.* DADA2: High-resolution sample inference from Illumina amplicon data. *Nature Methods* **13,** 581–583 (2016).

103. Prodan, A. *et al.* Comparing bioinformatic pipelines for microbial 16S rRNA amplicon sequencing. *PLOS ONE* **15,** e0227434 (2020).

104. Caruso, V., Song, X., Asquith, M. & Karstens, L. Performance of Microbiome Sequence Inference Methods in Environments with Varying Biomass. *mSystems* **4,** 10.1128/msystems.00163–18 (2019).

105. Wang, Q., Garrity, G. M., Tiedje, J. M. & Cole, J. R. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. eng. *Applied and Environmental Microbiology* **73,** 5261–5267 (2007).

106. Quast, C. *et al.* The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Research* **41,** D590–D596 (2013).

107. Louca, S., Parfrey, L. W. & Doebeli, M. Decoupling function and taxonomy in the global ocean microbiome. *Science* **353,** 1272–1277 (2016).

108. Douglas, G. M. *et al.* PICRUSt2 for prediction of metagenome functions. *Nature Biotechnology* **38,** 685–688 (2020).

109. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17,** 10–12 (2011).

110. Abarenkov, K. *et al. UNITE general FASTA release for Fungi* 2025.

111. Hyde, K. D. *et al.* The 2024 Outline of Fungi and fungus-like taxa. *mycosphere* **15,** 5146–6239 (2024).

112. Põlme, S. *et al.* FungalTraits: a user-friendly traits database of fungi and fungus-like stramenopiles. *Fungal Diversity* **105,** 1–16 (2020).

113. McMurdie, P. J. & Holmes, S. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. eng. *PloS One* **8,** e61217 (2013).

114. Oksanen, J. *et al.* Vegan: Community Ecology Package. *R Package Version 2.2-1* **2,** 1–2 (2015).

115. Lutz, K. C. *et al.* A Survey of Statistical Methods for Microbiome Data Analysis. English. *Frontiers in Applied Mathematics and Statistics* **8** (2022).

116. Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society: Series B (Methodological)* **57,** 289–300 (1995).

117. Bray, J. R. & Curtis, J. T. An Ordination of the Upland Forest Communities of Southern Wisconsin. *Ecological Monographs* **27,** 325–349 (1957).

118. Ricotta, C. & Podani, J. On some properties of the Bray-Curtis dissimilarity and their ecological meaning. *Ecological Complexity* **31,** 201–205 (2017).

119. Borcard, D., Gillet, F. & Legendre, P. in *Numerical Ecology with R* (eds Borcard, D., Gillet, F. & Legendre, P.) 151–201 (Springer International Publishing, 2018).

120. Borcard, D. & Legendre, P. Is the Mantel correlogram powerful enough to be useful in ecological analysis? A simulation study. *Ecology* **93,** 1473–1481 (2012).

121. Borcard, D., Gillet, F. & Legendre, P. in *Numerical Ecology with R* (eds Borcard, D., Gillet, F. & Legendre, P.) 369–412 (Springer International Publishing, 2018).

122. Hurlbert, S. H. Pseudoreplication and the Design of Ecological Field Experiments. *Ecological Monographs* **54,** 187–211 (1984).

123. Clarke, K. R. Non-parametric multivariate analyses of changes in community structure. *Australian Journal of Ecology* **18,** 117–143 (1993).

124. Russel, J. & Oksanen, J. *Russel88/MicEco: v0.10.0* 2025.

125. Sloan, W. T. *et al.* Quantifying the roles of immigration and chance in shaping prokaryote community structure. *Environmental Microbiology* **8,** 732–740 (2006).

126. Gotelli, N. J. & McGill, B. J. Null Versus Neutral Models: What's The Difference? *Ecography* **29,** 793–800 (2006).

127. Arkin, A. P. *et al.* KBase: The United States Department of Energy Systems Biology Knowledgebase. *Nature Biotechnology* **36,** 566–569 (2018).

128. *Babraham Bioinformatics - FastQC A Quality Control tool for High Throughput Sequence Data*

129. Li, D., Liu, C.-M., Luo, R., Sadakane, K. & Lam, T.-W. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* **31,** 1674–1676 (2015).

130. Kang, D. D. *et al.* MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. *PeerJ* **7,** e7359 (2019).

131. Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P. & Tyson, G. W. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Research* **25,** 1043–1055 (2015).

132. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nature Methods* **9,** 357–359 (2012).

133. Chaumeil, P.-A., Mussig, A. J., Hugenholtz, P. & Parks, D. H. GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. *Bioinformatics* **36,** 1925–1927 (2020).

134. Seemann, T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30,** 2068–2069 (2014).

135. Shaffer, M. *et al.* DRAM for distilling microbial metabolism to automate the curation of microbiome function. *Nucleic Acids Research* **48,** 8883–8900 (2020).

136. Henry, C. S. *et al.* High-throughput generation, optimization and analysis of genome-scale metabolic models. eng. *Nature Biotechnology* **28,** 977–982 (2010).

137. Porca, E., Jurado, V., Žgur-Bertok, D., Saiz-Jimenez, C. & Pašić, L. Comparative analysis of yellow microbial communities growing on the walls of geographically distinct caves indicates a common core of microorganisms involved in their formation. *FEMS Microbiology Ecology* **81,** 255–266 (2012).

138. Albuquerque, L., Johnson, M. M., Schumann, P., Rainey, F. A. & da Costa, M. S. Description of two new thermophilic species of the genus *Rubrobacter*, *Rubrobacter calidifluminis* sp. nov. and *Rubrobacter naiadicus* sp. nov., and emended description of the genus *Rubrobacter* and the species *Rubrobacter bracarensis*. *Systematic and Applied Microbiology* **37,** 235–243 (2014).

139. Meier, D. V., Imminger, S., Gillor, O. & Woebken, D. Distribution of Mixotrophy and Desiccation Survival Mechanisms across Microbial Genomes in an Arid Biological Soil Crust Community. *mSystems* **6** (ed Lax, S.) e00786–20 (2021).

140. Vitousek, P. M. & Reiners, W. A. Ecosystem Succession and Nutrient Retention: A Hypothesis. *BioScience* **25,** 376–381 (1975).

141. Fierer, N., Nemergut, D., Knight, R. & Craine, J. M. Changes through time: integrating microorganisms into the study of succession. *Research in Microbiology* **161,** 635–642 (2010).

142. Burke, C., Steinberg, P., Rusch, D., Kjelleberg, S. & Thomas, T. Bacterial community assembly based on functional genes rather than species. *Proceedings of the National Academy of Sciences* **108,** 14288–14293 (2011).

143. White, E. *et al.* A comparison of the species–time relationship across ecosystems and taxonomic groups. *Oikos* **112,** 185–195 (2006).

144. Jung, V., Violle, C., Mondy, C., Hoffmann, L. & Muller, S. Intraspecific variability and trait-based community assembly. *Journal of Ecology* **98,** 1134–1140 (2010).

145. Munday, P. L. Competitive Coexistence of Coral-Dwelling Fishes: The Lottery Hypothesis Revisited. *Ecology* **85,** 623–628 (2004).

146. Abrams, P. A. Recruitment, Lotteries, and Coexistence in Coral Reef Fish. *The American Naturalist* **123,** 44–55 (1984).

147. Cadotte, M. W. & Tucker, C. M. Should Environmental Filtering be Abandoned? *Trends in Ecology & Evolution* **32,** 429–437 (2017).

148. Ortiz-Álvarez, R., Fierer, N., de los Ríos, A., Casamayor, E. O. & Barberán, A. Consistent changes in the taxonomic structure and functional attributes of bacterial communities during primary succession. *The ISME Journal* **12,** 1658–1667 (2018).

149. Gerhold, P., Cahill Jr, J. F., Winter, M., Bartish, I. V. & Prinzing, A. Phylogenetic patterns are not proxies of community assembly mechanisms (they are far better). *Functional Ecology* **29,** 600–614 (2015).