**Title:**

Environmental factors directing spatial diversity and adaptation in halophilic desert microbiome

**Authors:**

Gherman Uritskiy, Adam Munn, Micah Dailey, Diego Gelsinger, Samantha Getsin, Alfonso Davila, Peter McCullough, James Taylor, Jocelyne DiRuggiero

**Stages of manuscript preparation:**

1. Finalize story and conclusions
2. Make main and supplementary figures
3. Write results
4. Write discussion
5. Write introduction
6. Write methods
7. Edit results, revise figures
8. Edit discussion, add references
9. Edit introduction, add references
10. Edit methods
11. Read entire paper

**INTRODUCTION**

Understanding the relationship between environment and composition of microbial communities is the first step to being able to make more robust predictions of microbiome dynamics. The composition of microbial communities is dictated by a wide variety of environmental abiotic and biotic factors. Variables such as temperature, pH, nutrient availability, and metabolite availability have deterministic effects on the composition and functions of microbiomes, as these factors dictate which organisms will thrive or be out-competed. In particular, factors that are the most limiting for survival in each unique environment will have the greatest impact in directing community assembly. Most microbial communities are nutrient-limited, meaning the ability to utilize the available metabolites will be one of the stronger forces selection. However, in more extreme environments other abiotic factors can become more important. Desert ecosystems, where life is typically limited by water availability, make for interesting model systems for investigating the core principles of microbial community assembly. In this study, we utilize a unique model desert ecosystem found in the Atacama Desert.

Surviving at the dry limit for life means organisms better equipped to survive desiccation will have a competitive advantage even if they would be outcompeted in any other setting. Microorganisms in halite microbiomes need to endure extreme humidity swings and even complete dehydration. Surviving at the dry limit for life means that any changes in water availability will likely have a great impact on their overall composition and structure. A previous study of halite microbiomes across different regions of the Atacama Desert has shown drastic differences in community structure, and linked these differences to the relative humidity of each region. In another study, a rapid addition of excess water from a rare rain event in 2015 resulted in a dramatic re-composition of the halite communities.

Beyond deterministic factors, stochastic factors can also have an effect on community assembly. In particular, gene drift in combination with physical isolation of individual communities can result in diverging community composition despite similar environmental conditions. This is particularly true at finer scales of community composition (genus and below), as a high degree of functional redundancy between closely-related taxa can result in different communities that essentially function in a near-identical way. A previous study found that fine-scale composition of microbial communities was more similar in halites next to each other than between different sites (but still at the same relative humidity), suggesting that strain dispersal rates also shape the community composition in this environment.

Microbial communities living inside halite (salt rocks) formations in the Atacama Desert, Chile make for an ideal model for studying microbial community assembly principles. Isolated under the surface of individual halites, these communities have minimal exchange of biomass and nutrients, allowing each community to develop independently and allowing us to glean insight into the stochastic and deterministic elements of microbial community structure. Despite an average annual precipitation of less than 1mm, life in these microbiomes evolved to survive on humidity from the air. Living at the dry limit for life causes their composition particularly sensitive to minor changes in environment, making them a compelling model to investigate early effects of climate change on microbiome composition and function.

Halite nodules form exclusively in evaporitic NaCl salt basins in Atacama Desert such as Salar Grande and Salar Llamara, where large salt deposits near the surface of the salars interface with the unique climate of Northern Atacama Desert. The proximity of the ocean and the presence of unique wind patterns result in extreme relative humidity swings during the diurnal cycle – from 30% during the day to 90% during the night. The deliquescent properties of salt allow it to draw on the moisture from the air when the humidity rises above 75% RH, which produces small amounts of liquid water. As this liquid evaporates during the day, capillary action moves more brine toward the surface, resulting in an overall displacement of the salt. Over several years, these hydration and dehydration cycles move enough salt is moved toward the surface to produce intricate porous formations known as halites.

The hydration cycles cause these halites to continually evolve in shape and structure, changing gradually over the course of years. During their dehydration during the day, capillary action moves liquid brine from the center of the nodule toward the peripheries, moving organic molecules and even live cells along. This is evidenced by the accumulation of scytonemin – a natural pigment produced by Cyanobacteria – at the surface of the halite nodules.

The halite microbial communities are comprised entirely of highly adapted halophiles, as they need to survive and thrive in saturated salt brines. The two dominant heterotrophic taxa found in this community are a halophilic archaeon Halobacteria (a Euryarcaheota) and a halophilic bacteria Salinibacter (a Bacteroidetes). Both these taxonomic groups are salt-in strategists, which means that they selectively import potassium ions and use them to counteract the external osmotic pressure from sodium ions. Other heterotrophs in the community include halophilic Proteobacteria, Actinobacteria, and a ecto-parasite of Halobacteria - Nanohaloarchaea. The biologically available carbon in the community is fixed by several species of Cyanobacteria and green algae. Previous characterization of the community metatranscriptome revealed that all of these community members are transcriptionally active, however the algae were particularly active. All phototrophs heavily prioritized the photosynthetic and carbon fixation pathways in their transcriptomes, while Halobacteria transcribed bactereorhodopsins – modified rhodopsins that in the presence of light can be used by this taxon to create a proton gradient and thus generate supplementary ATP.

In this study we build upon previous research to further characterize the drivers of halite microbiome diversity and composition. Beyond looking at broad regions of the Atacama Desert, we gradually zoom into smaller distance scales, from kilometers, to meters, and finally, to centimeters. Taking our observations of community structure, relative abundance of major taxonomic groups, species diversity, and total biomass, we characterize the abiotic factors governing halite microbiome assembly at each of the distance scales.

**METHODS**

**Sample collection and DNA extraction**

Halites were harvested across Salar Grande, a saltern flat in the Northern part of the Atacama Desert, Chile (see Supp. Data for coordinates of all samples). For the large distance scales, broad regions (~500m2) were sampled along transects in the north (SG1) and south (SG2) ends of the salar. For medium scales, halite nodules were harvested within 50m3 areas at the top and bottom of a hill in the SG1 locale. For inter-nodule and intra-nodule comparisons in the small distance scales, halites were harvested in a 50m2 are at the top of the SG1 hill. Halite nodules were collected by breaking them open with a sterilized hammer, and collecting colonized small pieces of broken halite (1-10cm across) from the center of the nodules. These pieces were stored in dark in dry conditions until they could be ground and used for cell and DNA extraction in the lab. For the intra-halite experiments, entire nodules were wrapped in plastic and transported to be cut in the lab.

**Cell and DNA extraction**

The colonized halite pieces were ground into fine powder for cells and gDNA extraction. For intra-halite sampling, the intact nodules were vertically sliced with a mechanical saw, exporing the colonization areas within. The sampling location of interest (top, middle, and bottom) were scraped with a sterilized knife for obtain sufficient material. Cells were extracted from the ground halite powder as previously described [Robinson paper], and the DNAeasy Powersoil DNA extraction kit (QIAGEN) was used to extract gDNA from the resulting cell pellet.

**Cell count estimates**

To estimate the total biomass of halite samples, 0.5g of ground halite was reserved for cell counting using fluorescence microscopy. The halite powder was gradually dissolved in 6.5ml of a solution containing 20% NaCl solution and 1% TWEEN. The solution was gently shaken for 30 minutes to break cell clumps, and was then DAPI was added to a total concentration of 0.5μg/ml. After a 10-minute incubation 2ml of the solution was put through a 25mm wide XX black polycarbonate filter filter by using a vacuum filtration tower (3 filter replicates in total). Each filter was then fixed to a glass slide using 1 drop of non-fluorescent immersion oil and imaged through a DAPI (blue) fluorescent light filter at 400X magnification. For each membrane, 5 images were taken with XX Zeiss camera (15 images total for each halite sample). The total number of visible cells was counted in each image using an automate *CellProfiler* v2.1 pipeline, in which the *CorrectIlluminationCalculate* function was used to normalize the background light levels, and *IdentifyPrimaryObjects* function was used to find and count unique nuclei (exact pipeline parameters can be found in *CellProfiler* pipeline file in Supp. Data). The number of cells per gram of halite was then then calculated from the number of cells in each image by accounting for the *eFOV* of the camera at that magnification (0.203mm2), the total area of the filter (226.98mm2), and the original amount of halite going into each membrane. The reusting simplified formula was , where *n* is the number of cells in the image, and *N* is the estimated number of cells per gram of halite. Cell count estimates outside of 2 standard deviations of the mean were discarded for each biological replicate.

**16S rDNA amplicon library preparation and sequencing**

The communities’ 16S rDNA was amplified with a 2-step amplification and barcoding PCR strategy as previously described by amplifying the hypervariable V3-V4 region with 515F and 926R primers. PCR was done with the Phusion High-Fidelity PCR kit (New England BioLabs) with 40ng of gDNA. Barcoded samples were quantified with the Qubit dsDNA HS Assay Kit (Invitrogen), pooled and sequenced on the Illumina MiSeq platform with 250 bp paired-end reads at the Johns Hopkins Genetic Resources Core Facility (GRCF).

**16S rDNA gene amplicon sequence variant pre-processing**

The de-multiplexed and quality trimmed 16S amplicon reads from the MiSeq sequencer were processed with Qiime2 2018.8.0. The three major comparison experiments (large, medium, and small distance scales) were processed independently from each other. Dada2 was used to call ASV variants using only the forward reads of the amplicon data (options: --denoise-single, --p-trunc-len 230 --p-chimera-method consensus). Alignment mafft was used to create a multiple alignment of ASV sequences and phylogeny fasttree was used to construct the phylogeny tree. The sampling depth used for the core-metrics-phylogenetic generation was chosen independently in each experiment based on the sample with the lowest read coverage. For ASV taxonomy assignment, a feature-classifier was first build using the SILVA 16S rRNA gene v128 database and the sequence of the 515F universal primer. For a complete list of commands used in this analysis, see additional documentation at <https://github.com/ursky/spatial_paper>.

**Statistical comparisons of sampling sites**

All comparisons between sites were made with built-in statistical packages within Qiime2 2018.8.0. Alpha and beta-diversity metrics were calculated for each distance scale experiment by using the core-metrics-phylogenetic command. Alpha diversity between different sample groups was compared with the alpha-group-significance with both the Faith\_PD and Evenness diversity metrics, and significance between beta diversity between sites was computed with the beta-group-significance command using the Weighted Unifrac dissimilarity matrices. The PCoA projection of the Weighted Unifrac dissimilarity matrices were also imported into custom visualization scripts. Enrichment for specific taxa at each taxonomic rank was tested using the ANCOM statistical test. The taxonomy of each ASV was estimated with the classify-sklearn command using a custom classifier (see above), and the relative abundance of major taxonomic groups was imported into custom scripts for plotting and statistical analysis. Differential abundance significance of each taxon was tested using an independent two-sided t-test. For the intra-halite sample comparison, the relative abundances were also standardized to account for high inter-nodule and inter-slice variability. The relative abundance of each taxon in each sample was standardized to its average relative abundance in that slice. The correlation of these normalized abundances with the distance to the nodule surface was calculated with a paired sample two-sided T-test.

**16S rRNA gene OTU abundance comparisons**

Quality trimmed 16S amplicon reads from were also processed with MacQIIME v1.9.1 to assign amplicon clusters at the OTU level. The reads were clustered at a 97% similarity cutoff with the pick\_open\_reference\_otus.py function (with --suppress\_step4 option), using the SILVA 123 database [37] release as reference and USEARCH v6.1.554 [38]. The OTUs were filtered with filter\_otus\_from\_otu\_table.py (-n 2 option), resulting in a total of 472 OTUs for site 1 and 329 OTUs for site 2. The beta diversity metrics of samples from the two sites were compared by normalizing the OTU tables with normalize\_table.py (default options), and then running beta\_diversity.py (-m unweighted\_unifrac, weighted\_unifrac). The OTU tables were exported for custom visualization with the Seaborn clustermap function. For a complete list of commands used in this analysis, see additional documentation at <https://github.com/ursky/spatial_paper>.

**Climate data collection and processing**

Relative humidity and temperature measurements in each sampled location were made over the course of 3 months at hourly (large and medium distance scales) or 5-minute (intra-nodule measurements) intervals. Time under deliquescence was estimated for each sampling site by calculating the number of hours that the relative humidity of the air was above 75%.

**Intra-nodule light measurements**

To measure light transmittance inside halite nodules, halites were placed under controlled lighting conditions with a 500 W halogen lamp 44 cm above the nodule as the only source of elimination. A handheld fiber optic spectrometer was inserted into a tight hole drilled in the underside of the halite nodule to the desired distance from the top. The relative effective light transmission at each wavelength was estimated by dividing each measurement to the respective intensity measured from the unfiltered light source (the broad spectrum lamp). A cosine-corrector was used to homogenize the fiber optic cable’s angular response. Only 500nm – 900nm wavelengths were considered. To account for inter-halite variance in the comparison of light transmittance to the top and middle positions of the halites, the light transmittance measurements were standardized to the mean of the “top” measurements made in all three nodules.

**RESULTS**

**Section 1: Sampling scheme and scales of diversity**

To investigate the relationship of halite microbial community structure with abiotic factors, we conducted a robust sampling survey of the halite nodules in Salar Grande of Atacama, Chile. To address factors universally important for community assembly, community structure was interrogated across varying scales – ranging from major regions of the salar to micro-niches present within a single halite nodule (Table 1).

To address adaptation differences between halite communities subject to drastically different climate condition, we performed a wide sampling of the North and South ends of the Salar Grande in Atacama Desert, Chile (Fig. 1A). Samples were taken from 500m2 areas at each end of the salar, with 39 samples from SG1 (North) and 46 samples from SG2 (South). There regions were more than 19km apart, and were expected to have drastically different climate conditions due to the varying heights of the mountains that shield the salar from the ocean, and thus humidity. The SG1 location had tall mountain formations (450m-804m above salar level) immediately to the West of the sampling collation, while the SG2 location had much shorted mountains and hills (58m-205m above salar level) between it and the ocean to the West. Both locations were relatively similar distances to the ocean to the West (11.6km for SG1 and 9.9km for SG2).

To investigate adaptations of halite communities subject to much more similar, but still distinct conditions, samples were taken at the top and bottom of a prominent hill at SG1. The hill had 32m of elevation gain over 330m. 19 samples were taken from the top of the hill, and 12 from the bottom (Fig. 1B). Halite nodules covered the ground throughout the sampling area. To look at communality differences between individual halites subject to similar external conditions, we performed a robust sampling of 6 halite nodules in a 10m2 area at the top of the SG1 hill. 6-9 replicates were sampled from each individual halite.

Finally, to look at community structure adaptations within different niches found within a single halite, we sampled 6 halite nodules in more detail. The nodules were sliced with a saw in three locations, and in each slice samples were taken near the top, bottom, and middle of the halite (Fig 1C). This sampling scheme allows us to interrogate community structure differences and similarities across the vertical and horizontal components, as well as correlate the structure to the distance to the nodule’s surface. The top, middle, and bottom sampling points were chosen arbitrarily down the center of the slice, and the white substrate (NaCl) was harvested avoiding the large dark deposits of insoluble minerals.

**Section 2: Factors influencing halite communities**

To determine differences in climate conditions between the samples locations, HOBO weather probes were left for prolonged periods of time (3-15 months) to record relative humidity and temperature. Data was collected at SG2 in a flat area, as well as the top and bottom of the hill at SG1.

The climate patterns in SG1 and SG2 were significantly different. During the tested time period SG2 was consistently cooler (by 5.2 degrees on average; Fig. S1A) and more humid (by 11% on average; Fig. S1B) that SG1. These differences appeared to be minimal during the night and early morning, when winds in both locations were minimal. After the winds picked up around noon, the differences in temperature and humidity began to be more apparent, with the greatest differences being around 2pm, when the temperature difference was as high as 7 degrees and the humidity difference was 15%. The differences in relative humidity also meant that the halite nodules at SG2 spent more time in conditions permitting for deliquescence (relative humidity needs to be >75% for salt to absorb water from the air). On average, SG1 nodules had 9.1 hours to re-hydrate during the night, while SG2 nodules had 12.5 hours (Fig S1C).

The differences in climate conditions at the top and bottom of the hill present in SG1 were subtler. The top location was significantly cooler (2-3 degrees; Fig. S1D) during the afternoon hours, and consistently slightly more humid (4-6%; Fig. S1E). While statistically significant (paired t-test, and Pearson correlations), these differences do not obviously consequential for halite nodule microbiota. However, when considering the total time under deliquescence, the top location provided significantly more time for nodule rehydration (10.2 hours) than the bottom (9.1 hours; Fig. S1F).

Micro-climate measurements were also taken inside a selected model halite at the top of the SG1 hill. Probes in the top, center, and bottom of the halite recorded humidity and temperature over the course of 3 months, and were contrasted with corresponding measurements made outside the halite (Fig. 2). The interior temperature of the halite closely traced that of the outside air. Shortly after sunrise (7:50-7:58am during the sampling time period) the outside temperature sharply increased from increased, while the internal temperature lagged behind by an hour while the nodule heated up. Toward the afternoon, the internal temperatures significantly surpassed that of the outside air, with the temperatures at the top of the halite reaching as high as 40 degrees. During the night, the relative humidity of the nodule interior peaked at 80%, as values above 75% allow for condensation of liquid water through deliquescence. Complementing the temperature, the humidity of the outside air dropped sharply after sunrise, however the interior of the nodule remained fully hydrated (~75%, or the point of deliquescence) for up to 3 hours longer. The outer locations (top and bottom) of the halite interior dehydrated significantly during the afternoon, and gradually rehydrated toward the evening when the outside humidity increased. Strikingly, however, the center of the halite (middle) remained fully saturated with water throughout the day, never dipping below 75% relative humidity.

To investigate differences in halite topology across the larger scales, measurements were taken comparing the halite density at SG2 and the top and bottom of SG1. Randomized transects were used to estimate the fraction of the ground covered by halites in different stages of development, starting from a flat polygon and ending in a mature, tall nodule (Fig. S2). The halite density in the flat regions of SG1 and SG2 were relatively similar, with ~30-40% of the ground covered by nodules (Fig. S3). The top of the hill in SG1, however, was significantly more densely covered in nodules, with ~60% of the area being constituted by nodules. Furthermore, a large fraction of these halites were large, fully mature nodules that could no longer be associated with their original polygons.

To investigate levels of available light for photosynthesis within the halites, we used a handheld fiber optic spectrometer to measure spectra transmission at different positions within the halites. From the underside of the halite nodule, a hole was drilled to a specific depth corresponding to a particular distance from the illuminated (top) surface of the halite nodule. A fiber optic probe was inserted into the hole. Due to the practical difficulties of operating the spectrometer and the drill in the field, the measurements were instead obtained in the lab, using a 500W broad-spectrum halogen lamp positioned 44 cm above the halites to simulate the solar illumination. The lamp’s spectrum was measured and used for normalization of the data to obtain the effective transmission inside the halite, thus nullifying any major differences between the spectra from the lamp and the sun. We used a cosine-corrector (translucent screen that scatters light but does not affect the spectrum) to homogenize the fiber optic cable’s angular response. For the bottom positions within the halites, we found that our direct transmission measurements in the lab could not reproduce realistic conditions from the desert. In the field, light scatters around and even underneath the halite nodules, which it not practical to reproduce in laboratory conditions. In all three tested nodules, we found that the center positions received less than 10% of the light available at the top positions (Fig. 3). This difference was even greater in the primary excitation wavelength of chlorophyll (680nm), with only 1-5% of the light usable for photosynthesis reaching the center of the nodules. This absorption of red light by chlorophyll was what led to the green interior halite color.

To investigate the relationship between internal halite positioning and total microbial biomass, cell density was estimated for some of the samples coming from the top, middle, and bottom positions of their respective halites. Cell nuclei were stained with DAPI and counted threw a microscope using an automated cell-counting pipeline. In order to get a more robust cell count estimate in each biological replicate, 3 technical replicates were performed for each sample, and 5 fields of view were counted per technical replicate. The absolute cell counts per gram of halite were then back-calculated from the cell counts per field of view. Taking cell density (million cells per gram) as a proxy for biomass, this provided a reasonable estimate for the total biomass of each sample.

Comparing sample biomass at each relative internal position within the 5 sampled halites revealed a great degree of biomass variety (Fig. S4). Some of the halites, particularly halite #6, had significantly more biomass than the others (5.6M/g compared to the overall 1.9M/g average in all samples). There was also a high variance in the average biomass of vertical slices coming from the same halite nodule. To account for this variance when comparing the biomass differences between the top, middle, and bottom positions within each slice, the average cell density estimates of each biological replicate were standardized to the maximum value in that slice (Fig. S5). Doing so highlighted that there was no position in the halites that had a generally higher or lower biomass. Out of the 12 slices, 2 had the highest biomass at the top, 4 at the middle, and 6 at the bottom.

**Section 3: Taxonomic structure differences**

Comparing the overall community structure at the ASV level also revealed distinct differences between sampling location at every scale of diversity. The Weighted Unifrac dissimilarity index was applied to normalized ASV abundance tables to compute inter-sample distances in taxonomic composition. This metric considered not only organism abundance, but also appropriately weighed ASV phylogenetic similarity, which reduced the SNP-level noise introduced by ASV clustering. These dissimilarity matrices were then used to compute difference significance between different locations in the desert.

At the large distance scale, we found samples from SG1 to be significantly different than those from SG2 (*PERMANOVA: pval=* 0.001, test statistic= 28.36; Fig. 4A). At the medium distance scale, samples collected at the top of the SG1 hill were significantly different from those at the bottom of the SG1 hill (*PERMANOVA: pval=* 0.001, test statistic= 22.5; Fig. 4B). The average inter-sample dissimilarity between sites at these two distance scales also differed. The dissimilarity distance between SG1 and SG2 (0.96), was significantly higher than that between SG1-top and SG1-bottom (0.91; Student’s T-test, pval<0.001).

Comparing the community structure differences samples from collected from neighboring nodules at the same site (SG1-top; small distance scale) and from samples samples collected from different positions within the nodules (tiny distance scale) also revealed major differences (Fig. 5, S6). Replicate samples from different nodules were significantly different from each other (PERMANOVA: pval=0.012, test statistic= 2.15), as were samples collected from different vertical slices (PERMANOVA: pval<0.001, test statistic= 2.77). Finally, the samples collected between the top, middle, and bottom positions within the nodules were also signiffcantly different (PERMANOVA: pval<0.013, test statistic=2.59).

Observing the first and second principal components of the dissimilarity matrix PCoA projection revealed that only the vertical slice and halite identifiers clearly visually separated along these components (Fig. S6). The top-middle-bottom spatial separation was only clearly evident in the third and fourth principal components, where the samples separated based on the samples’ distances to the nodule’s surface (Fig. 5). This is important to node as the first and second principal components explained a much greater degree of inter-sample variance (49% and 11%) compared to that of the third and fourth components (7% and 6%).

**Section 4: Phyla relative abundances**

To evaluate high-level differences in microbial community structure between the sampled locations, we compared the relative abundance compositions at the phylum level. The halite microbiome is comprised of 6 phyla – *Euryarchaeota* (almost exclusively comprised of *Halobacteria*), *Bacteroidetes*, *Cyanobacteria*, *Proteobacteria*, *Actinobacteria*, and *Nanohaloacrchaea*. Green algae (*Chlorophyta*) could also be indirectly detected with Chloroplast 16S rRNA gene, however they were only consistently present at the SG2 location.

The high-level taxonomic composition of SG1 and SG2 differed dramatically (Fig. 6A). It should be noted that for the comparison of these general regions samples were collected in several transects over a large 500m2 areas to capture the general microbiome diversity in each side of the salar. Despite the composition variance introduced by this sampling scheme, several phyla were consistently more abundant in one of the locations. In general, *Euryarchaeota* was more abundant at the SG1 location, constituting the majority of the community. On the other hand, *Chlorophyta* (estimated with chloroplast sequences) and *Proteobacteria* were significantly more abundant at the SG2 location, and were almost absent in the SG1 (Student T-tests, pval<0.0001). Strikingly, we found that the *Chlorophyta* (chloroplast) 16S rRNA gene relative abundances in SG2 were nearly equal, and sometimes greater than that of *Cyanobacteria*.

Evaluating taxon differences with the ANCOM enrichment test also revealed similar trends. At the D2 level, *Chlorophyta* and *Proteobacteria* were more abundant in SG2 (ANCOM W=8 and 10, respectively). At the D1 (domain) level, Archaea was significantly more relatively abundant in SG1 (ANCOM, W=2). Interestingly, no significantly differential taxa were detected at the ASV (amplicon sequence variant) level.

Comparing the relative taxonomic composition of halite microbial communities at the top and bottom of the SG1 hill also revealed major differences in phyla abundances (Fig. 6B). These samples were collected within 50m2 areas at the top and bottom, so the inter-replicate composition variance is significantly lower. *Cyanobacteria* were more relatively abundant at the top of the hill than that bottom (Student T-test pval<0.001), while *Euryarchaeota* (constituted almost exclusively by *Halobacteria*), Proteobacteria, and Actinobacteria were more abundant at the top (Student T-test pval<0.0001). Chlorophyca chloroplast sequences were only detected at low abundances in a few samples.

Reevaluating taxa enrichment with the ANCOM significance test produced slightly different results than that from the T-tests on the total relative phyla abundances. At the D2 level, *Nanohaloarchaea*, *Proteobacteria*, and *Actinobacteria* were found to be significantly more abundant at the top of the hill than the bottom (ANCOM W=6,9,8, respectively), and *Cyanobacteria* was more abundant at the bottom of the hill (ANCOM W=7). No significance was observed for *Euryarchaeota*.

The relative phyla abundances were also different throughout different positions within the halites nodules. Several halites were collected at the top of the SG1 hill and sliced open. Samples were collected along the slices in locations roughly corresponding to the, middle, and bottom of the nodules. Because of high inter-nodule and inter-slice variability of the microbial community composition, the relative abundance of each taxa in each sample was standardized to its average relative abundance in that slice. This standardization resulted in a relative abundance average of 1, and highlighted differences in phyla spatial distribution along the top, middle, and bottom position of the nodules (Fig. S9).

We found that *Euryarchaeota* (constituted entirely of *Halobacteria*) was significantly more abundant at the bottom of the halite than the middle, while Bacteroidetes showed the reverse trend (Student’s T-test, pval<0.01; Fig. S9). However, the difference magnitudes of these differences were very small (<8% and <20%, respectively). Cyanobacteria was significantly and consistently more abundant at the tops of the nodules than the middles (Student’s T-test, pval<0.001) by nearly 80%. *Actinobacteria*, *Nanohaloarchaea*, and *Proteobacteria* were consistently most abundant in the middle of the halites and less abuntant at the top and bottom positions (Student’s T-test, pval<0.01). This preference for the center of the halites resulted in a major increase in relative abundance at the center – ~310% for *Actinobacteria*, ~70% for *Nanohaloarchaea*, and ~50% for *Proteobacteria*. *Chlorophyta* (chloroplast) sequences were only detected at low abundances is a few samples, and thus were not included in this analysis.

During the sampling of the various positions inside the nodules, the distance of each biological sample to the surface was recorded. This allowed for correlating distance to the surface with the standardized relative abundance of each taxon. Just as described above, the abundances of each taxon in each position were still standardized to their average abundance in each slice (Fig. 7). Importantly, while this projection allows for inference of more general trends between abundance and distance to the surface, the samples collected at the top and bottom of the nodules have equally close to the surface, and cannot be distinguished. The significance of the positive of negative trend was evaluated with a Spearman correlation test (pval<0.01). Similar to what was found in the categorical comparisons, *Actinobacteria*, *Nanohaloarchaea*, and *Proteobacteria* were significantly more abundant further away from the surface, reaching maximum relative abundances at 2-3cm away from the nodule surface. *Cyanobacteria* on the other hand, significantly decreased in relative abundance as the distance to the surface increased. Interestingly, both Euryarchaeota and Bacteroidetes relative abundances were not significantly correlated with the distance to the surface.

**Section 5: OTU diversity differences:**

Comparing alpha diversity metrics across the scales of diversity at the ASV clustering level did not reveal any significant differences between the locations as the heterogeneity between biological replicates from each site was too high, particularly when comparing SG1 with SG2 and SG1-top with SG1-bottom. To further investigate diversity differences between samples from dryer of more humid locations, we looked at the presence or absence statistics of organisms. To reduce some of the SNP-level noise present at the ASV clustering level, the 16S rRNA gene sequences were re-clustered into OTUs at the 97% similarity level. This allowed for easier comparison of species abundance between broad regions in the large and medium distance scales.

Comparing OTU abundances between SG1 and SG2 revealed that a significant percentage of OTUs were present at both sites, a subset of OTUs were only present at the SG2 location (Fig. 8A). This group of OTUs was not significantly enriched in any particular phylum (random subset simulation, 1 million iterations, pval>0.01). On the other hand, very few OTUs that were abundant in SG1 were not abundant in at least a few SG2 samples. Similarly, OTU abundances in top and bottom of the SG1 hill showed that the top location had a large group of OTUs that were not present in most samples at the bottom (Fig. 8B). On the other hand, almost no OTUs were only present at the bottom of the hill. Similarly, to the SG1 to SG2 comparison, the exclusive SG1-top OTU group was not significantly enriched for any particular phylum, but rather consisted of a wide variety of organisms (random subset simulation, 1 million iterations, pval>0.01).

While no taxa enrichment was found in the OTU presence and absence data analysis, we investigated the OTU distribution of some taxa of interest across these different locations. In particular, we considered the presence and absence of *Proteobacteria* OTUs, as this phylum was universally differentially abundant at the large and medium distance scales. We found that SG2 not only has a greater relative abundance of *Proteobacteria*, but also has a greater diversity of this taxon, containing a group of OTUs that are almost completely absent in SG1 (Fig. S10A). Similarly, several of the *Proteobacteria* OTUs at the SG1 location were only present in samples at the top, and not the bottom of the hill (Fig. S10B).

**Section 6: Community structure is more similar in samples that are closer together**

Inter-sample dissimilarity comparisons were also used to determine if samples collected farther apart (larger distance scales) were more dissimilar than those collected closer together (smaller distance scales). The Bray-Curtis dissimilarity was used to estimate inter-sample distance metric at the OTU-level, highlighting the raw community composition differences without considering inter-OTU phylogenetic similarity.

The average inter-sample dissimilarity was the highest between samples collected in SG1 and SG2 (large distance scale; Fig. 9). The dissimilarities at this level also showed the greatest level of variance, possibly due to samples at each of these sites originating from a wide range of locations. The dissimilarities between samples from SG1-top and SG1-bottom (medium distance scale) were significantly lower (Student’s T-test, pval<0.001). Zooming in further into the smaller distance scales, we found that the inter-sample dissimilarities between samples become significantly and progressively lower (Student’s T-tests, pval<0.001).

At the smallest distance scales (intra-halite samples), we found that samples coming from the same positions within the halites (e.g. top positions of halite 1) were less dissimilar than those coming from different positions (Student’s T-tests, pval<0.001). However, this trend was not observed when comparing the horizontal (slice) component – samples from the same slices were not more similar than samples from different slice (but the same halite).

**DISCUSSION**

Minor changes in environmental conditions can have a major impact on the microbial community composition of extremophile microbiomes. This effect is amplified by the fact that the halite dessert communities survive at the dry limit for life, where water availability is extremely limited and is likely the limiting factor for survival. At large distance scales, where the climate conditions are distinct from each other, the respective microbial communities residing in these regions are very distinct in composition. Perhaps more surprising, however, was the fact that microbial composition was still very distinct between the top and bottom of the SG1 hill, even though the measured temperature and humidity differences appeared to be subtle. While the average relative humidity differences at the top and bottom of the hill were very minor, the effect these differences could have on the interior of the halite nodules could be accentuated by the fact that this minor difference affects the total amount of time that the halites have during the night to re-hydrate via deliquescence. With this in mind, the actual relative humidity of the outside air may not be as relevant to the halite microbiota as the time during which the halites can re-hydrate.

While larger-scale climate differences had a clear impact on halite microbiome composition, we observed the most extreme differences in environmental conditions within the interiors of individual nodules. Our relative humidity measurements inside the nodules revealed that while the outer areas of the halites get partially dehydrated during the day, and rehydrate during the night, the center of the halites remained fully hydrated throughout the day, rarely dipping below the deliquescence point. This has major implications on the adaptations that microbial communities need to have in these respective sites – the microbiota at the halite periphery needs to be able to survive routine desiccation, while the microbiota in the center does not. While we observed distinct differences in the composition of microbial communities in these different sites inside the halite, it is interesting to note that most of the major community members we still represented in the periphery as well as in the center. In addition, the overall taxonomic structure similarity was greater between these intra-nodule sites than between the top and bottom of the SG1 hill. This could be possibly explained by the hydration cycles within the nodules, with causes liquid water movement through capillary action, resulting in the mixing of the interior microbiota over time.

Comparing phyla relative abundance differences across all the considered distance scales revealed that *Proteobacteria*, *Nanohaloarchaea*, and *Actinobacteria* are generally more relatively abundant in locations with with higher water availability. This is especially evident inside the halite nodules, where the relative abundances of these taxa were much higher in the centers of the nodules. This suggests that these heterotrophs are not as well adapted to survive the desiccation cycles compared to the more resilient salt-in strategist halophiles *Halobacteria* and *Bacteroidetes*. *Nanohalobacteria* are also salt-in strategists, however are also parasitic, relying strongly on their Halobacteria hosts. Their adaptations to this lifestyle include a tiny physical size and a compact and streamlined genome, both of which could result in sacrificing some more advanced desiccation adaptations.

Green algae have been characterized with metagenomics and metatranscriptomics in the SG1 location, however our amplicon-based methods detected very low abundances in this site. At the significantly more humid SG2, however, algae seem to be very abundant, often being more relatively abundant than *Cyanobacteria*. Previous research on these algae have shown that their presence correlates with fog events (Robinson), so it is possible that the SG1 location was too dry at the time of sampling to support significant algae populations.

Cyanobacteria is the only phyla that seemed to consistently more relatively abundant in dryer sites. This is consistent with previous research in desert microbiomes, where Cyanobacteria was more resilient in surviving desiccated conditions. Inside the halite interior, Cyanobacteria particularly were more relatively abundant at the top of the nodules. Our light transmittance measurements inside the nodules showed that the tops of the nodules had an order magnitude more light available for photosynthesis compared to the center. While cyanobacteria are likely also limited by water availability, the need for sufficient light to perform photosynthesis likely means that they need to balance these two environmental variables to thrive. In the context of this halite community, it seems that it is more optimal to for Cyanobacteria to be near the surface of the nodules.

The absolute abundance, which we could infer by multiplying the relative abundance fraction by the absolute cells/g cell counts could allow for another level of understanding for spatial partitioning of the major taxa of halite microbial communities. However, we found that the total biomass of different locations within the halites were not correlated with the conditions found in these respective locations. In particular, we hypothesized that the centers of the halites would contain significantly more biomass than he periphery because of the favorable conditions stemming from stable water availability throughout the day. Alternatively, the tops of the halites could also harbor more biomass due to increased carbon fixation potential from increased light availability. Instead, we found that the biomass was seemingly randomly dispersed throughout the nodules, with some slices having the highest cell counts in the bottom, some in the middle, and some at the top positions. As such, we postulate that the biomass of each microbial niche in the nodules is largely dictated by stochastic abiotic factors stemming from the nodule structure itself. One of such factors could be the total available surface area found inside the salt pores of the halites, which are visibly highly variable in different locations within any given halite. This variable is difficult to measure in the scope of this study, however past studies in other lithic microbiomes have demonstrated that internal rock topology has a great impact on colonization success. While we predict that the total biomass is stochastically dispersed throughout the halite due to the random non-uniform structures of the halite’s interior, we found that the relative composition of the communities was correlated with the distance of the microbial niches to the surface, which has major implications on the stability of water availability.

The differences in community composition between major locations in the salar could also be explained by geographic isolation influencing gene drift and the gradual divergence of microbial composition over time. Contained in the rocks, the halite microbial communities have extremely limited capacity to exchange biomass between nodules, and increased distance makes this even less likely. The only way cells could travel between SG1 and SG2 would be extremely rare events of dormant cells being transported by the harsh winds of the desert, which could contribute to gene diversification, but ultimately is unlikely to have a major impact on general community composition. We observe however, that at both the large (SG1 vs SG2) and medium (SG1-top vs SG1-bottom) distance scales, the more humid location generally has a greater diversity at the OTU level. If the differences between these respective locations were due to only gene drift, we would not expect this to be the case. In addition to this, the high-level taxonomic differences between these locations do not appear to be stochastic. Instead, specific phyla (especially Proteobacteria) appear to be both more diverse and more relatively abundant in the more sites with higher water availability. Taken together, we conclude that water is the major driving factor influencing both the relative composition of the halite microbial communities, but also their respective diversity. Increased water availability enabled greater diversity at the OTU level.

While water appears to be the major abiotic factor dictating the relative composition of halite communities, light availability also has its impact on the community’s phototrophs. The total light availability was expected to be equal between the major sampled locations, however the light availability was not equal at different positions within the halite interiors. We found that as the light traveled deeper into the nodule, more light was absorbed by the salt, cells, and debris in the halites. This was especially true for light wavelengths strongly utilized by the chlorophyll at 680nm, meaning that the light penetrating to the nodule center was notably red-shifted, and had less usable energy for photosynthesis. Cyanobacteria, which together with green algae fix organic carbon for the community’s heterotrophs, were generally found to be more relatively abundant near the surface of the halite nodules, particularly at the top. Combined with our observations of increased light availability at the top of the halite, we conclude that the cyanobacteria have an advantage over other taxa at the surface of the nodules because they are able to utilize the light for photosynthesis. Cyanobacteria also require water for survival, and likely need to balance out water and light availability, which are mutually exclusive resources in the nodule interiors. Based on these observations, we suggest that even under these desiccated conditions, Cyanobacteria and light-limited as a well as water-limited.

The heterotrophs of the community are dependent on the biologically-available carbon fixed by the phototrophs, however. This means that they are also indirectly dependent on light availability, but based on our observations they are largely limited by water availability. The two most abundant groups in the halite microbiome – Halobacteria (the major class within Euryarchaeota) and Salinibacter (the major class within Bacteroidetes) – are also able to utilize light through modified rhodopsins, which they use to generate ATP. Metatranscriptomic interrogation of this community revealed that this method of additional energy income is especially upregulated in Halobacteria. Despite this, we were unable to detect significant preference of these taxa towards the nodule surface or interior. This suggests that Halobacteria and Salinibacter are likely also limited by both light and water availability, however favor these factors more evenly compared to Cyanobacteria.

BANFIELD SHOWED THIS TOO. Not surprisingly, we found that the relative compositions of halite microbial communities become more similar the closer they are in proximity to one another. This could be possibly explained by stochastic gene drift, which results in divergence of composition of segregated microbial communities. However, taken in light of our observations of more deterministic effects of water availability on the microbial composition across all the interrogated scales of diversity, these results point the alternate interpretation. The increased similarity of relative composition stems from increased similarity in environmental conditions. Gene drift also likely has an impact, particularly when looking at the inter- and intra-nodule distance scales. As discussed previously, the different positions within the nodules have drastically different abiotic conditions, much more so than between general climate differences between the top and bottom of the SG1 hill. The communities, however, are more different between the top and bottom of the hill than between he top and bottom of the nodule interiors. This suggests that the microbial composition similarity between any two location are driven by water availability and gene drift at the larger scales of diversity, where the physical segregation of communities by the halite rocks plays a role in community formation. At the smaller intra-nodule distance scales, however, gene drift plays a much smaller role due to the regulal inter-mixing of the internal microbial niches through micro-fluidic rotation of water during the daily hydration-dehydration cycles. This results in deterministic factors – water availability – being the major driver for community composition differences.

While this study interrogated the general taxonomic composition differences between sites, the location-specific adaptations are likely even more pronounced at the functional potential and functional levels. Future studies utilizing information from shotgun metagenomics to look at gene and pathway enrichment in response to environmental differences could provide a much more in-depth view at halite community adaptation. Furthermore, using metatranscriptomics to look at real-time transcriptional adaptations of these communities to different conditions would provide an even more complete picture of how halite microbial communities adapt to these extreme conditions.

In this study we investigated the abiotic factors governing microbiome composition across different scales of diversity and distance, allowing for identification of factors that govern halite colonization at large desert-wide scales, as well as smaller micro-niche specific scales. We found that while water availability and gene drift both govern microbiome assembly at the larger scales, deterministic abiotic factors (water and light) are most important at at smaller intra-nodule scales, where cell dispersion is not limited. This principle can be applied to a variety of non-desert communities. For example, differences in gut microbiomes between different people are produced by both resource differences (e.g. diet) and gene drift (stochastic colonization of each person). But differences between upper-lower intestine colonization are governed by only resource availability. We also report an unexpected stochastic compartmentalization of biomass within the nodules despite the deterministic relative abundance composition.

**CONCLUSIONS**

1. Small environmental parameter changes have large effects on community structure (\*\*\*)
2. Water availability affect relative abundance of major phyla (\*\*\*)
3. Higher water availability permits a greater number of (all) OTUs (\*\*)
4. Light availability influences *Cyanobacteria* relative abundance inside halites, and partially for *Halobacteria* and *Salinobacter* (\*)
5. The more similar to locations are in conditions, the more similar the communities residing in them (\*\*\*)