1   **Phylogenetic and Functional Substrate Specificity for**

2   **Endolithic Microbial Communities in hyper-arid environments**

3

4   Alexander Crits-Christoph[1], Courtney K. Robinson[1], Bing Ma[2], Jacques Ravel[2], Jacek

5   Wierzchos[3], Carmen Ascaso[3], Octavio Artieda[4], Virginia Souza-Egipsy[5], M. Cristina

6   Casero[3] and Jocelyne DiRuggiero[1§]

7   *[1]Biology Department, The Johns Hopkins University, Baltimore, MD, USA; [2]Institute for*

8   *Genome Sciences, University of Maryland School of Medicine, Baltimore, MD, USA;*

9   *[3]Museo Nacional de Ciencias Naturales, MNCN - CSIC, Madrid, Spain; [4]Universidad de*

10   *Extremadura, Plasencia, Spain; [5]Instituto de Ciencias Agrarias, ICA - CSIC, Madrid,*

11   *Spain.*

12

13   Running Title: Microbial communities in Atacama endoliths

14   Keywords: endoliths, cyanobacteria, extreme environment, Atacama Desert, arid

15   environment, metagenomics

16

17   [§]Corresponding author:

18   jdiruggiero@jhu.edu

19

ABSTRACT

Under extreme water deficit, endolithic (inside rock) microbial ecosystems are considered environmental refuges for life in cold and hot deserts, yet their diversity and functional adaptations remain vastly unexplored. The metagenomic analyses of the communities from two rock substrates, calcite and ignimbrite, revealed that they were dominated by *Cyanobacteria*, *Actinobacteria*, and *Chloroflexi*. The relative distribution of major phyla was significantly different between the two substrates and biodiversity estimates, from 16S rRNA gene sequences and from the metagenome data, all pointed to a higher taxonomic diversity in the calcite community. While both endolithic communities showed adaptations to extreme aridity and to the rock habitat, their functional capabilities revealed significant differences. ABC transporters and pathways for osmoregulation were more diverse in the calcite chasmoendolithic community. In contrast, the ignimbrite cryptoendolithic community was enriched in pathways for secondary metabolites, such as non-ribosomal peptides (NRPS) and polyketides (PKS). Assemblies of the metagenome data produced population genomes for the major phyla found in both communities and revealed a greater diversity of *Cyanobacteria* population genomes for the calcite substrate. Draft genomes of the dominant *Cyanobacteria* in each community were constructed with more than 93% estimated completeness. The two annotated proteomes shared 64% amino acid identity and a significantly higher number of genes involved in iron update, and NRPS gene clusters, were found in the draft genomes from the ignimbrite. Both the community-wide and genome-specific differences may be related to higher water availability and the colonization of large fissures and cracks in the calcite in contrast to a harsh competition for colonization space and nutrient resources in the narrow pores of the ignimbrite. Together, these results indicated that the habitable architecture of both lithic substrates- chasmoendolithic versus cryptoendolithic - might be an essential element in determining the colonization and the diversity of the microbial communities in endolithic substrates at the dry limit for life.

1 Introduction

2   The rate of desertification across our planet is accelerating as the result of human activity
3   and climate change. Ecosystems of arid and hyperarid environments are highly
4   vulnerable, including the microbial communities supporting these biomes. Under extreme
5   water deficit and high solar radiation, endolithic (inhabiting rock) habitats are considered
6   environmental refuges for life (Friedmann, 1982;Cary et al., 2010;Pointing and Belnap,
7   2012; Wierzchos et al., 2012). In these microbial ecosystems, the rock substrate provides
8   protection from incident UV, excessive solar radiation and freeze–thaw events while
9   providing physical stability, and enhanced moisture availability (Walker and Pace,
10 2007;Chan et al., 2012). Endolithic microbial communities are photosynthetic-based with
11 primary producers supporting a diversity of heterotrophic microorganisms (Walker and
12 Pace, 2007;Wierzchos et al., 2012).

13 Habitats as different as the underside of quartz rocks (Cowan et al., 2011;Chan et al.,
14 2013;de los Ríos et al., 2014), sandstones (Friedmann, 1982;de la Torre et al.,
15 2003;Pointing et al., 2009), granite boulders (Wei et al., 2015), the inside of gypsum
16 (Dong et al., 2007;Wierzchos et al., 2011;Wierzchos et al., 2015) and halite evaporites
17 (Wierzchos et al., 2006;Robinson et al., 2015), carbonaceous (DiRuggiero et al., 2013)
18 and volcanic rocks (Wierzchos et al., 2013;Cámara et al., 2014) among others, have
19 shown that life has found innovative ways to adapt to the extreme conditions of hyper-
20 arid deserts.

21 The Atacama Desert is one of the driest deserts on Earth and its hyper-arid core has been
22 described as "the most barren region imaginable" (McKay et al., 2003). In the Yungay
23 area of the hyperarid zone of the Atacama Desert, with decades between rainfall events
24 and extremely low air relative humidity (mean yr^{-1} values <35%), ancient halite crusts of
25 evaporitic origin have been shown to provide sufficient moisture to sustain
26 cryptoendolithic (within pores) microbial communities (Wierzchos et al., 2006;de los
27 Rios et al., 2010;Robinson et al., 2015). Another strategy for survival in this hyper arid
28 desert is the colonization of cracks and fissures of rock substrates, also called
29 chasmoendolithic colonization (DiRuggiero et al., 2013). Our work in the hyper-arid core
30 of the Atacama Desert revealed that such communities inhabit gypsum covered rhyolite
31 and calcite rocks, where the complex network of cracks and fissures of the rocks promote
32 water retention (DiRuggiero et al., 2013). Calcite is highly translucent, with a network of
33 large connected fissures providing abundant space for microbial colonization within a
34 few cm of the rock surface (DiRuggiero et al., 2013). We found that the diversity of the
35 calcite community was significantly increased when the water budget of the ecosystem
36 was potentially enhanced by dew formation on the rock in contrast to the rhyolite
37 community, where the unique source of water were scarce rainfalls (DiRuggiero et al.,
38 2013). More recently, we discovered communities colonizing ignimbrite, a pyroclastic
39 rock composed of plagioclase and biotite embedded in a matrix of weakly-welded glass
40 shards (Wierzchos et al., 2013). In the ignimbrite the colonization is cryptoendolithic,
41 with microorganisms colonizing small pores within the rock and only within a few mm of
42 the rock surface (Wierzchos et al., 2013;Cámara et al., 2014).

43 From previous work on microbial communities in hyper arid deserts, it is clear that the
44 climate regime play a major role in determining the habitability of a given substrate.
45 However, the susceptibility of rocks to colonization also depends on the architecture of

the lithic substrate (Wierzchos et al., 2015), including factors such as translucence, which allows transmission of photosynthetically active radiation (PAR), thermal conductivity, the presence of a network of pores and/or fissures connected to the rock surface, which is linked to the capacity at retaining water, and chemical composition. These factors are likely to impact greatly microbial colonization and diversity (Herrera et al., 2009;Cockell et al., 2011; Walker and Pace, 2007;Wierzchos et al., 2012).

Nevertheless, very little information is available about what rock properties are driving endolithic communities functioning and diversity. Here we used a metagenomic analysis of two types of endolithic colonization in calcite (chasmoendolithic) and ignimbrite (cryptoendolithic) rocks collected in the extreme environment of the hyper-arid core of the Atacama Desert. The taxonomic and functional composition of the communities were characterized and interpreted within the context of the structural properties of each rock type.

Material and Methods

Sampling and site characterization

Calcite (composed mainly of calcium carbonate) rocks were collected in 2013 near Valle de la Luna (VL and CB) (DiRuggiero, 2013) and ignimbrite (pyroclastic volcanic glass) rocks were collected in the Lomas de Tilocalar area (IG and MIG) (Wierzchos et al., 2013) of the Atacama Desert in northern Chile. Additional ignimbrite rocks were also collected from a nearby area 25 km to the North-East of Lomas to Tilocalar (PING). Microclimate data for the Lomas de Tilocalar area was collected *in situ* from April 2011 to 2013 using an Onset HOBO® Weather Station Data Logger (H21-001) as previously described (Wierzchos et al., 2013). Solar flux was measured using a PAR sensor for wavelengths of 400–700 nm (Wierzchos et al., 2013). Microclimate data for Valle de la Luna was extracted from historical records for the village of San Pedro de Atacama located 5 km east of our sampling site, as reported by DiRuggiero et al. (DiRuggiero et al., 2013).

Microscopy analyses

Colonized rock samples were processed for scanning electron microscopy in backscattered electron mode (SEM-BSE) observation and/or energy dispersive X-ray spectroscopy (EDS) microanalysis according to methods by Wierzchos et al. (Wierzchos and Ascaso, 1994;Wierzchos et al., 2011). SEM-BSE was used in combination with EDS to characterize the lithic substrates. Rock samples were observed using a scanning electron microscope (DSM960 Zeiss; Carl Zeiss) equipped with a solid-state, four diodes BSE detector plus an auxiliary X-ray EDS microanalytical system (Link ISIS Oxford, UK).

Fluorescence microscopy (FM) in structural illumination microscopy mode (SIM) using DAPI nucleic acids stain was performed on cell aggregates gently isolated from the chasmoendolithic habitat (Wierzchos et al., 2011). The samples were examined using a fluorescence microscope (AxioImager M2, Carl Zeiss, Germany) in SIM mode with a ApoTome (commercial SIM by Zeiss) system for 3 dimensional (3D) visualization of cell aggregates (Wierzchos et al., 2011).

DNA extraction and sequencing

1. For 16S rRNA gene sequencing, total genomic DNA was extracted from rock powder
2. using the PowerSoil DNA Isolation kit (MoBio Laboratories Inc., Solana Beach, CA)
3. from four calcite and four ignimbrite rocks. DNA was amplified using the barcoded
4. universal primers 338F and 806R for the V3–V4 hypervariable region. Amplicons from
5. at least 3 amplification reactions were pooled together, purified, and sequenced using the
6. Illumina MiSeq platform by the Genomics Resource Center (GRC) at the Institute for
7. Genome Sciences (IGS), University of Maryland School of Medicine. For metagenomes,
8. the DNA was extracted from 4 rocks for each substrate from above (VLC and IG),
9. pooled, and sequencing libraries were prepared using the Nextera XT DNA sample
10. preparation kit (Illumina, San Diego, CA), with an average insert size of 400 bp, and
11. sequencing was performed on the Illumina HiSeq2500 platform by the GRC at IGS.

12. *16S rRNA gene sequences analysis*

13. Illumina paired-end sequences were processed using the QIIME package (v1.6.0)
14. (Caporaso et al., 2010) as previously described (DiRuggiero et al., 2013). Alpha and beta
15. diversity metrics were calculated based on OTUs at the 0.03% cutoff (OTUs0.03) in
16. QIIME. Statistical testing using Nonparametric Mann-Whitney tests were performed to
17. indicate confidence in similarities/differences observed.

18. *Metagenome analysis and assembly*

19. Illumina sequencing of the metagenomic libraries produced 126,820,576 and
20. 100,086,740 paired-end reads for the calcite and ignimbrite samples, respectively.
21. Ribosomal RNA sequence reads were removed using Bowtie (v1) (Langmead et al.,
22. 2009) and by mapping to SILVA reference database (Pruesse et al., 2007). Both paired-
23. end reads were filtered out if one read mapped to the rRNA database. Raw reads with
24. low-quality bases as determined by based calling (phred quality score of 20 that
25. corresponds to an error probability of 1%) were trimmed from the end of the sequence,
26. and the sequence longer than 75% of the original read length were retained. The reads
27. were assigned taxonomy content using PhyloSift (Darling et al., 2014) and functional
28. content using HUMAnN (Abubucker et al., 2012) and the Metagenomics analysis server
29. MG-RAST (Meyer et al., 2008). An assembly was produced for the metagenome using
30. the IDBA-UD assembler for metagenomic sequencing data (Peng et al., 2012). We use a
31. *k* range of 20-100 with a pre-correction step before assembly. Assembled contigs were
32. grouped into potential draft genomes using MaxBin (Wu et al., 2014), an expectation-
33. maximization (EM) algorithm using tetranucleotide frequencies, abundance levels, and
34. single-copy gene analysis. Taxonomy was assigned to each bin using both PhyloSift and
35. Kraken (Darling et al., 2014;Wood and Salzberg, 2014).

36. *Phylogenetic Analysis*

37. The phylogenetic position of cyanobacteria metagenomic bins was determined using
38. marker genes extracted by Phylosift and that were conserved across metagenomic bins
39. with at least 50% estimated genome completeness from MaxBin (Wu et al., 2014). The
40. marker genes were aligned individually to reference genes from the UniProt database
41. using MUSCLE (Edgar, 2004) and the individual alignments were concatenated. Low-
42. quality regions of the alignment were removed using Gblocks (Castresana, 2000) and a
43. maximum-likelihood phylogenetic tree was created using FastTree (Price et al., 2010).

5

1 *Population genome annotation*

2 Cyanobacterial genomic bins Ca9 and Ig12 were selected for annotation as draft genomes
3 based on both estimated genome completeness and n50 values of the binned contigs.
4 RAST server (Aziz et al., 2008) was used to annotate and compare the draft genomes to
5 known references. Putative phage contigs incorrectly binned were identified using the
6 annotation and removed from the draft genomes. The antibiotics and Secondary Metabolite
7 Analysis Shell antiSMASH 2 (Blin et al., 2013) was used for the prediction of non-
8 ribosomal peptides (NRPS) structures and synthetic gene clusters. The NORINE database
9 (http://bioinfo.lifl.fr/norine/) was used to determine chemical structures for putative
10 products of NRPS and synthetic gene clusters. The draft genomes were compared using
11 protein products predicted with Prodigal (Hyatt et al., 2010) and BLASTP (Gish and
12 States, 1993) with a cutoff of 80% amino acid identity between matches. Results from
13 the draft genomes comparisons were visualized using Circos (Krzywinski et al., 2009).

14 *Sequence data and availability*

15 All sequences were deposited at the National Center for Biotechnology Information
16 Sequence Read Archive under Bioproject ID PRJNA285514 and accession numbers
17 SAMN03754650 (ignimbrite) and SAMN03754649 (Calcite). The MG-RAST report for
18 the   data   is   available   under   ID   4600829.3   and   4600830.3
19 (http://metagenomics.anl.gov/metagenomics.cgi?page=MetagenomeOverview&metageno
20 me=4600829.3   and
21 http://metagenomics.anl.gov/metagenomics.cgi?page=MetagenomeOverview&metageno
22 me=4600830.3). Completed assemblies, annotation, and phylogenetic trees are available
23 at
24 http://figshare.com/articles/Phylogenetic_and_Functional_Substrate_Specificity_for_End
25 olithic_Microbial_Communities_from_the_Atacama_Desert/1606253.
26

27 # Results

28 We characterized at the molecular level the endolithic microbial communities from two
29 rock types from the Atacama Desert, calcite from Valle de La Luna and ignimbrite from
30 Lomas de Tilocalar. Both areas are extremely dry with less than ~ 25 mm rainfall each
31 year and average air RH between 18.3 and 40.5%, (Table 1). In both locales,
32 temperatures reached values above 32°C and maximum PAR values were about 2500
33 μmol s^{-1}m^{-2}.

34 The calcite samples were composed of laminated calcite layers with a thickness of
35 several centimeters. The calcite mineralogical composition and petrographic study were
36 previously reported by DiRuggiero et al. (2013). The calcite rock surface was covered by
37 a hardened semitransparent layer (up to 5 mm thick) with microrills forming small and
38 short sinuous channels on the rock surface (Fig. 1a). The most characteristic feature of
39 the calcium carbonate rock was the presence of irregular narrow fissures and cracks that
40 extended perpendicular and/or parallel to the rock surface (Fig. 1a and 1b). A large
41 number of these fissures and cracks, up to dozens of millimeters in length, were
42 colonized by chasmoendolithic microorganisms detected upon fracturing the rock (Fig.

1a). Additional fractures, parallel to the rock surface, also revealed similar chasmoendolithic colonization (black arrow on Fig. 1a). Using SEM-BSE, we observed an almost continuous colonization of these fissures by chasmoendoliths (Figs 1b-e) and the connection of the fissures to the rock surface (Figs 1). We observed mostly well-developed cyanobacteria cells surrounded by one or two sheaths of capsular-like extracellular polysaccharidic substances (EPSs) and heterotrophic bacteria present within some of the cyanobacteria aggregates (arrow in Fig. 1d). Closer to the rock surface, cyanobacteria cells appeared deteriorated with collapsed cells and remains of microbial colonization (arrows in Fig. 1e). The 3-Dimensional reconstruction of the cyanobacteria aggregates shown characteristic microbial aggregates inside sack-like structures formed by EPSs (green signal in Fig. 1f).

In contrast to the chasmoendolithic habitat found within the calcite fissures, the cryptoendolithic colonization within the volcanic ignimbrite presented a very different picture. The ignimbrite has a structure of vitrified foam identical to pumice rock with a large number of connected and non-connected bottle-shaped pores. The mineralogical, petrographic, and structural characteristics of the ignimbrite were previously reported by Wierzchos et al. (Wierzchos et al., 2013). The colonization zone of ignimbrite rocks was in the form of a narrow (1-3 mm thick) intensively green layer, parallel to the ignimbrite surface (Fig. 2a), and within 2-3 mm of the rock surface. This brown-colored varnish rock covering the surface was composed of allochthonous clay minerals and filling some of the pores at the surface (white arrow in Fig. 2b). Just below the ignimbrite surface cover, a cryptoendolithic colonization zone was observed (artificially cyan-colored spots in Fig. 2b, black arrow) with many small pores. The microbial assemblage within the bottle-shaped pores was composed of cyanobacteria and heterotrophic bacterial cells, seemingly not associated with EPSs (Fig. 2c-d). Some of the pores were not colonized whereas others were almost totally filled by cryptoendoliths (Fig. 2d). Cyanobacteria formed aggregates and, in places, cells were associated with heterotrophic bacteria (arrow in Fig. 2e). In some pores, we also observed both damaged in healthy cyanobacteria together with heterotrophic bacteria, probably also damaged (Fig. 2f).

Taxonomic diversity of the rock communities

We obtained a total of 315,334 16S rRNA gene paired-end sequences, with an average read length of 408 bp, for four individual rocks from each of the two endolithic substrates, calcite and ignimbrite. Taxonomic assignments of the sequence reads revealed that the dominant phyla, in both samples, were *Cyanobacteria*, *Actinobacteria*, and *Chloroflexi*, and that their relative distribution was significantly different between the two substrates (Fig. S1). Diversity metrics at a maximum sequencing depth of 2400 sequences per sample support the finding that the calcite substrate harbors a more diverse community in term of richness and phylogenetic diversity (Table S1; Fig. S2). Nonparametric Mann-Whitney tests reported that this difference is significant (p<0.05) across all diversity metrics. Although the statistical power of this test is weak due to the small sample size of this comparison (n=8), the ranges for each substrate of all diversity metrics are non-overlapping and distinct (Table S1). Photosynthetic cyanobacteria dominated both communities and greater proportional abundances were found in all the ignimbrite samples than in the calcite samples; this was also supported by a nonparametric Mann-Whitney test (p=0.03). Fourteen and 12 major OTUs of cyanobacteria (>1% abundance)

1   were identified in the calcite and ignimbrite communities, respectively, and none were
2   shared across the two communities.

3   We generated over 22 Gbp of high quality, paired-end, metagenomic shotgun sequences
4   for the calcite and ignimbrite substrates (Table 2). Taxonomic assignment of the
5   metagenomic reads was performed with PhyloSift using all extracted marker genes (Fig.
6   3). When compared to the 16S rRNA gene taxonomic distribution, averaged for each
7   rock substrate, we found a general consensus between the two datasets (Fig. 3), which
8   supported a higher taxonomic diversity for the calcite community. In both substrates the
9   metagenome analysis showed higher proportional abundance of rare taxa, probably as the
10   result of variation in genome sizes across all taxa. Reconstruction of 16S rRNA gene
11   sequences from the metagenomic dataset with EMIRGE (Miller et al., 2011) provided
12   full-length genes for all the major phyla observed in the 16S rRNA gene dataset (Fig. S3)
13   and indicated again a higher taxonomic diversity in the calcite community.

14   ***Functional diversity of the rock communities***

15   The functional composition of the calcite and ignimbrite metagenomes was analyzed with
16   HUMAnN and MG-RAST using total sequence reads. The HUMAnN analysis showed
17   that at the KEGG Super-pathway level, the two communities maintained a strong degree
18   of similarity in functional composition (Fig. S4). Functional assignment of sequence
19   reads to the SEED or COG database by MG-RAST (Fig. 4) also showed high similarity
20   between the two communities. When using the SEED database to classify sequence reads
21   into metabolic functions, categories with the most reads were for metabolisms related to
22   "carbohydrates", "proteins", "amino acids and derivatives", and "cofactors, vitamins,
23   prosthetic groups and pigments" (Fig. 4).

24   The metabolic potential for both the calcite and ignimbrite communities was further
25   analyzed using the MG-RAST server. Genes encoding for pathways for photosynthesis
26   and auxotrophic CO_2 fixation were found in both the calcite and ignimbrite metagenomes
27   and included biosynthetic genes for photosystems I and II (PSI and PSII), phycobilisome
28   (light harvesting antennae of PSII in cyanobacteria), the Calvin-Benson cycle,
29   carboxyzomes, and photorespiration. Genes for nitrogen fixation were not detected,
30   despite an intensive search using BLASTP and HMMER (Finn et al., 2011) on all
31   predicted proteins from the sequence reads and from the assembled metagenomes. Using
32   the SEED functional classification, we found that the major pathways for nitrogen
33   acquisition in both communities were ammonia assimilation (direct uptake of ammonia
34   and assimilation into glutamine and glutamate) and nitrate and nitrite ammonification
35   (direct uptake of nitrate or nitrite and subsequent reduction to ammonia); minor pathways
36   included allantoin utilization and cyanate hydrolysis (Fig. S5).

37   Two major pathways for phosphorus acquisition were identified in both communities
38   from the abundance of their protein encoding genes. One extensive suite of genes
39   constituted a putative Pho regulon and another set of genes encoded for proteins involved
40   in the transport and metabolism of phosphonates. With respect to iron transport, the
41   calcite and ignimbrite metagenomes contained mostly siderophore/heme iron transporters.
42   These ABC transporters are known to mediate the translocation of iron-siderophore
43   complexes across the cytoplasmic membrane (Köster, 2001). We also found ABC
44   transporters of the ferric iron type and, to a lesser extent, of the ferrous iron type and

8

1   known to be induced by low pH (Köster, 2001).

2   The functional analysis of the metagenomes from both communities also uncovered a
3   diversity of stress response pathways, with most notably genes involved in carbon
4   starvation, cold shock, oxidative stress, osmotic stress/desiccation tolerance, and
5   secondary metabolites. In both the calcite and ignimbrite metagenomes, we found genes
6   encoding for polyol transport (ABC transport systems), in particular glycerol, choline,
7   betaine, and glycine betaine uptake systems (ABC transport systems), betaine and ectoine
8   biosynthesis, and synthesis of glucans (Table S3). A gene encoding an aquaporin Z water
9   channel, mediating rapid flux of water across the cellular membrane in response to abrupt
10   changes in osmotic pressure (Calamita, 2000), was also found in high abundance. Also
11   notable, the large number of genes dedicated to trehalose biosynthesis and uptake (Table
12   S3). Secondary metabolites included a wide variety of compounds including non-
13   ribosomally made peptides (NRPS), polyketides (PKS), and alkaloids. The majority of
14   genes encoding for NRPS and PKS represented a wide range of siderophores, including
15   pyoverdine, pyochelin, and yersiniabactin (Table S2). Biosynthetic pathways for other
16   secondary metabolites included those for alkaloids, phenylpropanoids, phenazines, and
17   phytoalexins.

18   The two SEED subsystems of greatest difference between the two communities were
19   phosphorous metabolism and membrane transport, with both functional categories
20   showing higher gene abundance in the calcite community (Fig. 4a). Functional diversity
21   was also higher in the calcite community for genes involved in iron acquisition and
22   transport, Mn transport, and the synthesis of compatible solutes such as betaine, ectoine,
23   and glucans (Table S3). One notable exception was a ferrichrome iron receptor, which
24   was about 3 times more abundant in the ignimbrite than in the calcite community (Table
25   S3); the encoding genes were all annotated to cyanobacteria taxa. Overall, the pathways
26   involved in osmoregulation were more diverse in the calcite than in the ignimbrite
27   community (Table S3). In addition, the functional categories most enriched in the calcite
28   community were not from the most abundant taxa; for example, genes involved in the
29   synthesis of osmoregulated periplasmic glucans, and most genes for the ectoine
30   biosynthesis pathways, were from *Proteobacteria*, *Actinomycetales*, and *Firmicutes*
31   rather than from *Cyanobacteria* or *Rubrobacter*. In contrast, when looking at gene
32   abundance rather than diversity, we found considerably more NRPS and PKS related
33   genes, especially siderophores, in the ignimbrite metagenome (Table S2). The COG
34   category for secondary metabolite biosynthesis/transport also shows similar higher levels
35   of these genes in the ignimbrite metagenome (Fig. 4b).

36   *Sequence assembly*

37   The 12.8 Gbp of metagenome sequence data for the calcite community was assembled
38   into a total of 350 Mbp of contigs (344,986) with an n50 value of 1,530 bp and a
39   maximum contig length of 101,619 bp. The 10.1 Gbp metagenome sequence data for the
40   ignimbrite community were assembled into 220 Mbp of contigs (166,509) with an n50
41   value of 2,631 bp and a maximum length of 119,637 bp (Table 2). Binning of contigs into
42   population genomes using MaxBin generated 42 and 34 putative population genome bins,
43   for the calcite and ignimbrite communities, respectively. All bins were individually
44   assigned taxonomy using both Kraken and PhyloSift, and bins mostly composed of single
45   taxa were further analyzed (Table 3). We found that the bins' G+C content closely

1   matched that of the reference species, with G+G below 50% for *Cyanobacteria* and
2   above 65% for *Rubrobacter* species. Interestingly, the relative abundance ratio of a bin
3   did not necessarily match its estimated completeness. Rather, population genomes with
4   the highest coverage produced lower quality assemblies, i.e. more contigs of smaller sizes
5   (compare for example Ca01 with Ca09, and Ig01 with Ig12). This likely resulted from a
6   higher representation of micro-heterogeneity at high coverage, for a given population
7   genome, impairing the assembler functionality.

8   The most abundant population genomes found in the calcite and ignimbrite metagenomes
9   were related to *Chroococcidiopsis* and *Gloeocapsa*, both members of the *Chroococcales*.
10 Cyanobacteria of the *Acaryochloris* genus were found only in the calcite community. The
11 *Actinobacteria* were dominated by *Rubrobacter xylanophilus*, together with *Conexibacter*
12 *woesei* and a number of other *Actinomycetales* species, the later differing between the
13 two communities. *Chloroflexi* populations composed of *Thermomicrobia*, *Chloroflexus*,
14 and *Thermobaculum* species were also found in the ignimbrite and calcite communities.
15 The dominant *Proteobacteria* was *Sphingomonas*, and in the ignimbrite community, the
16 low G+C% *Bacteroidetes* species *Segetibacter* was represented (Table 3). The
17 phylogenetic analysis of the metagenomic bins indicated that there was significantly
18 greater diversity within the calcite *Cyanobacteria* community, with bins distributed in
19 two distinct clades, whereas only one *Cyanobacteria* clade was found within the
20 ignimbrite dataset (Fig. S6). The ignimbrite *Cyanobacteria* population genomes were
21 distinct at the phylogenetic level to that of the calcite, but one of the calcite metagenomic
22 bins (Ca02) was consistently more closely related to the ignimbrite than the others calcite
23 bins, further supporting the extent of the taxonomic diversity of the calcite community
24 (Fig. S6).
25 A candidate draft genome for the dominant *Cyanobacteria* population genome in each
26 community was selected based on estimated completeness, calculated from single gene
27 copy analysis (MaxBin). The estimated completeness was 94.4% for *Gloeocapsa sp.*
28 Ca09, from the calcite metagenome, and 93.5% for *Chroococcidiopsis sp.* Ig12, from the
29 ignimbrite metagenome (Table 3). The genomic bins were annotated using the RAST
30 server, and the two genomes were found to share similar hits for 64% of their total gene
31 products. Three contigs from Ca09 and five contigs from Ig12 were found to be entirely
32 composed of phage-related and hypothetical proteins, making them likely to be
33 cyanophage genomic fragments binned together with the host genome. This co-binning
34 was probably due to the tendency of bacteriophages to share tetranucleotide frequencies,
35 genes, and codon usage patterns with their host (Soueidan et al., 2014). These contigs
36 were removed from the draft genomes for the analysis of annotation features and gene
37 products. On average the two annotated proteomes had 69.4% amino acid identity, and
38 only 27% and 24% of the gene products could be assigned a SEED category for Ca09 and
39 Ig12, respectively. The draft genomes had multiple genes involved in choline and betaine
40 biosynthesis subsystem and Aquaporin Z, both related to osmotic stress. A Mycosporine
41 synthesis cluster, MysA, MysB, and MysC, was found in the Ca09 but not in Ig12 draft
42 genome; a BLAST search against the entire metagenomes found these genes to be absent
43 from the ignimbrite metagenome. The Ca09 draft genome had six copies of the
44 ferrichrome-iron receptor and the Ig12 draft genome had 12, while the closest reference
45 genome in RAST, *Nostoc punctiforme PCC 73102*, only had one copy of this gene.
46 Another iron related gene, the Iron(III) dicitrate transport system FecB gene, had multiple

1 duplications in both Ca09 and Ig12, but is absent in the *Nostoc punctiforme* RAST
2 annotation.

3 We used antiSMASH to identify biosynthesis loci for secondary metabolism in the four
4 major cyanobacteria population genomes from the calcite and ignimbrite metagenomes
5 (Table 3). Similarly to the whole metagenome, we found an increased abundance of
6 NRPS and PKS secondary metabolite genes and gene clusters in the Ig12 ignimbrite
7 genome. Searches in the NORINE database (http://bioinfo.lifl.fr/norine/) revealed that the
8 closest chemical structures for the putative products of these gene clusters were often
9 siderophores (Fig. 6). A figure linking genes products with greater than 80% amino acid
10 identity between the complement of contigs for the Ca9 and Ig12 draft genomes
11 illustrates the close relatedness between these strains and also the greater abundance of
12 NRPS/PKS and Ferrichrome-iron receptor proteins in the draft genome from the
13 ignimbrite community (Fig. 7).

14

Discussion

16 By linking taxonomy to function, our study of endolithic systems in the Atacama Desert
17 provides an understanding of microbial adaptation strategies to extreme desiccation and
18 solar irradiance. It also provides an integrated view of the biotic and abiotic factors
19 shaping the functioning of these highly specific ecosystems. Functional metagenomic
20 analysis was carried out on two rock substrates, calcite and ignimbrite, and the context
21 for the interpretation of the molecular data was provided by the examination of water
22 availability, the mineralogical composition, and the type of colonization (chasmo- versus
23 cryptoendolithic) for the two types of rock.

24 The Valle de la Luna (calcite rocks) and Lomas de Tilocalar (ignimbrite rocks) sampling
25 sites are located in the hyper-arid zone of the Atacama Desert and, as such, both locations
26 experience rare precipitations (DiRuggiero et al., 2013;Wierzchos et al., 2013). It is also
27 the site of the highest solar irradiance recorded on Earth (Wierzchos et al., 2015).
28 Moisture on the rock surface at Lomas de Tilocalar was only detected during and shortly
29 after a rain event and the air RH is one of the lowest reported for the Atacama Desert
30 (DiRuggiero et al., 2013;Wierzchos et al., 2013;Wierzchos et al., 2015). In contrast, the
31 yearly average of air RH at Valle de la Luna is about twice that in Lomas de Tilocalar
32 and the presence of a hardened layer and microrills on the calcite rock surface was
33 indicative of dew formation (DiRuggiero et al., 2013). This is important because dew
34 formation, together with higher RH, can significantly increase the atmospheric water
35 budget for the calcite microbial community (Kidron, 2000;Büdel et al., 2008;DiRuggiero
36 et al., 2013). Higher maximum temperatures in Lomas de Tilocalar could have the
37 opposite effect by potentially increasing the rate of evapotranspiration, limiting moisture
38 availability within the ignimbrite (Houston and Hartley, 2003).

39 In addition to atmospheric water availability, the structure and geochemical composition
40 of a rock substrate are also essential determinants of its bioreceptivity (Walker and Pace,
41 2007;Nienow, 2009). While the mineralogy between the two substrates was significantly
42 different, a laminated calcite layered rock and a pyroclastic rock composed of plagioclase
43 and biotite embedded in a matrix of weakly-welded glass shards, we argue here that the
44 major differences lied within the physical properties of the rocks (DiRuggiero et al.,

11

1 2013;Wierzchos et al., 2013;Cámara et al., 2014). The calcite rocks displayed large
2 colonized cracks and fissures, dozens of millimeters deep, all connected to the surface
3 and providing direct access to surface water and increased light illumination. In contrast,
4 the bottle-shaped pores found in the ignimbrite represented a limited colonization space.
5 Not all the pores were colonized, as observed by SEM-BSE, probably because not
6 connected to the surface. In addition, the dark varnish cover of the ignimbrite
7 significantly decreases the incoming PAR radiation, which explains the very narrow
8 colonization zone right underneath the rock's surface we observed.

9 Taken together, these observations point to a calcite environment far less extreme than
10 that of the ignimbrite. This is substantiated by biodiversity estimates from 16S rRNA
11 gene sequences, and from the metagenome data, showing a higher taxonomic diversity
12 for the calcite than for the ignimbrite community. While both rock communities were
13 dominated by members of the *Cyanobacteria*, we found a significantly higher proportion
14 of chlorophototrophs (>20% difference) in the ignimbrite than the calcite community. In
15 a previous study of halite nodules, we reported a higher fraction of the cyanobacterium
16 *Halothece* in samples collected in the driest and more extreme area of the desert than in
17 Salars exposed to periodic fog events (Robinson et al., 2015). It was suggested that a
18 higher proportion of photoautotrophs was required to support the associated heterotrophic
19 populations because of lower primary productivity under water stress (Robinson et al.,
20 2015). This might also be the case for the differences in chlorophototrophs abundance we
21 observed between the calcite and ignimbrite rocks.

22 It is important to point out that the metagenome for each community was obtained with
23 DNA extracted from four individual rocks each. The 16S rRNA gene data was obtained
24 for each of the individual rocks and supported the taxonomic assignments of the
25 metagenomes. Using the metagenomic data, we identified the dominant cyanobacteria in
26 the calcite substrate as *Gleocapsa* whereas previous work assigned them to the
27 *Chroococcidiopsis* (DiRuggiero et al., 2013). This discrepancy arose because the
28 *Gleocapsa* and *Chroococcidiopsis* taxa are closely related, we use metagenome data, and
29 different taxonomic assignment algorithms produces different data. It is also possible that
30 these belong to a new genus that is neither *Gleocapsa* or *Chroococcidiopsis*.

31 The identification of photosynthesis and carbon fixation pathways from the metagenomes
32 confirmed that primary production was mainly carried out via photosynthesis. However,
33 in contrast to Antarctic cryptoendolithic communities, we did not detect any genes for
34 chemolithoautotrophy (Wei et al., 2015). Genes for diazotrophy were also absent from
35 both metagenomes and the major sources of nitrogen for the communities were
36 potentially ammonia and nitrates. These later nitrogen sources might not be limiting in
37 the Atacama Desert where high nitrate concentration from atmospheric origin (Michalski
38 et al., 2004) have been reported. This might explain the absence of nitrogen fixation in
39 endoliths from this desert in contrast to endolithic niches in the McMurdo Dry Valleys
40 (Chan et al., 2013;Wei et al., 2015).

41 The metabolic genes and pathways annotated from the calcite and ignimbrite
42 metagenomes revealed that the rock substrate is a stressful environment with its own
43 challenges. Nutrient limitation was suggested by a number of stress adaptation strategies,
44 including an abundance of protein encoding genes for two major pathways for phosphate
45 acquisition. In cyanobacteria, and a number of other bacteria, the Pho regulon, involved

1 in transport and metabolism of phosphonates, is controlled by environmental inorganic
2 phosphate levels and mediates an adaptive response to phosphate starvation (Monds et al.,
3 2006;Adams et al., 2008). Additionally, phosphonates can also be used as an alternative
4 source of phosphorus under limiting conditions (Dyhrman et al., 2006;Adams et al.,
5 2008). The large number of genes involved in carbon starvation found in the
6 metagenomes is also indicative of periods of low-nutrient stress in the rock environment.
7 These genes play a role in long-term survival under extended periods of starvation
8 (Kolter et al., 1993), provide resistance to additional stresses such as oxidizing agent and
9 near-UV radiation (Watson et al., 1998), and might give a competitive advantage to the
10 cells once growth resumes (Zambrano et al., 1993).

11 Adaptation to desiccation in the rock environment involved a number of strategies for
12 desiccation avoidance and protection via the biosynthesis and accumulation of organic
13 osmolytes. The metagenome analyses revealed metabolic pathways for a broad spectrum
14 of compatible solutes that have been found across bacterial phyla (da Costa et al.,
15 1998;Roberts, 2005). These solutes are thought to contribute to desiccation tolerance
16 because in high concentrations they generate low water potentials in the cytoplasm
17 without incurring metabolic damage (Yancey, 2005;Oren, 2007). Of note, is the high
18 abundance of genes for polyol metabolism, which in algae, fungi, plants, and a few
19 bacteria, is a key part of a biochemical protective strategy against water loss (da Costa et
20 al., 1998;Holzinger and Karsten, 2013). Functional categories for osmoregulation were
21 more diverse in the calcite community but were not from the most abundant taxa,
22 suggesting that the calcite functional diversity was driven by the taxonomic diversity of
23 the less abundant taxa in the community. We were not able to identified genes for EPS
24 biosynthesis in the metagenomes but microscopic observations of the endolithic
25 communities revealed abundant EPS in the calcite but not in the ignimbrite microbial
26 assemblages. In the calcite environment with more liquid water available, and frequent
27 wetting-drying cycles, EPS might play a significant role in water retention, as
28 demonstrated in similarly dry environments (Hu et al., 2012). In contrast, in the
29 ignimbrite, where the unique source of liquid water are very scarce rain events and
30 complete dehydration periods lasting as long as 9 months, the environmental conditions
31 might too harsh for EPS effectiveness and for the energy expenditure required for their
32 biosynthesis.

33 Cyanobacteria are known to produce a wide range of secondary metabolites including
34 PKS, NRPS, and antibiotics (Leao et al., 2012;Puglisi et al., 2014). One of the major
35 differences, at the functional level, between the calcite and ignimbrite metagenomes was
36 for secondary metabolites pathways. We found a considerably higher number of NRPs
37 and PKS related genes in the ignimbrite metagenome, and the draft genome of some of
38 the most abundant cyanobacteria from the ignimbrite community, than in the calcite
39 community. Biosynthetic pathways for other secondary metabolites included those for
40 alkaloids, involved in host fitness and protection (Leflaive and Ten-Hage, 2007),
41 phenylpropanoids, which have been shown to enhance antioxidant activity and tolerance
42 to stress in cyanobacteria (Singh et al., 2014), phenazines known to contribute to
43 behavior and ecological fitness (Pierson III and Pierson, 2010), and phytoalexins with
44 antimicrobial and often antioxidative activities (Patterson and Bolis, 1997). The synthesis
45 of these secondary metabolites is often directed toward protection from radiation,
46 allelopathy, resource competition, and signaling (Leao et al., 2012;Puglisi et al., 2014)

1   and, in the case of the ignimbrite, might be an indication of the fierce competition for
2   scarce resources and limited colonization space in the small pores of this substrate rock.
3   Indeed the production of small antimicrobial compounds has been reported as a strategy
4   for eliminating prior residents before colonization of a new space by a number of
5   microorganisms (Hibbing et al., 2010). Additionally, SEM-BSE images of colonized
6   pores from the ignimbrite rock revealed cell decayed cells in some of the pores,
7   potentially reflecting the effects of antimicrobial compounds or resource competition. In
8   contrast the chasmoendolithic environment of the calcite provided more physical space
9   and therefore may promote interactions between members of the community for
10   metabolic cooperation. In addition to antimicrobials, database searches revealed that the
11   closest chemical structures for some of the NRPs and PKS gene clusters found in both
12   communities were for siderophores. Iron is an essential element for most organisms
13   because of its essential role in redox enzymes, membrane-bound electron transport chains,
14   but also in photosynthesis (Andrews et al., 2003). The abundance of siderophores in the
15   rock communities, and the enrichments in genes for iron acquisition in the cyanobacteria
16   draft genome when compared to the reference genome from *N. punctiforme*, strongly
17   suggest iron starvation. This was emphasized in the ignimbrite community with a large
18   number of sequence reads assigned to gene clusters for siderophores and a ferrichrome-
19   iron receptor. Furthermore, the mycosporine gene cluster found in the calcite
20   cyanobacteria, but not the ignimbrite, may indicate differences in UV radiation stress
21   experienced by the two communities (Garcia-Pichel et al., 1993; Oren, 1997).

22   Environmental stress is a major driver of microbial diversity and this is reflected in the
23   functional annotation of metagenomes from Atacama Desert endoliths. Our analysis is
24   reflective of adaptations to desiccation, osmotic stress, low nutrients, iron deficiency, and
25   in the case of the ignimbrite, of a fierce battle for colonization space and resources among
26   community members. The relative abundance of phototrophs in each community is likely
27   indicative of the *in situ* level of primary production, a hypothesis that might be tested by
28   careful field measurements of microbial activity.

29

Acknowledgements

34

References

36   Abubucker, S., Segata, N., Goll, J., Schubert, A.M., Izard, J., Cantarel, B.L., Rodriguez-
37   Mueller, B., Zucker, J., Thiagarajan, M., Henrissat, B., White, O., Kelley, S.T.,
38   Methe, B., Schloss, P.D., Gevers, D., Mitreva, M., and Huttenhower, C. (2012).
39   Metabolic reconstruction for metagenomic data and its application to the human
40   microbiome. *PLoS Comput Biol* 8, e1002358. doi: 10.1371/journal.pcbi.1002358.
41   Adams, M.M., Gomez-Garcia, M.R., Grossman, A.R., and Bhaya, D. (2008). Phosphorus
42   deprivation responses and phosphonate utilization in a thermophilic

Synechococcus sp. from microbial mats. *J Bacteriol* 190, 8171-8184. doi: 10.1128/JB.01011-08.

Andrews, S.C., Robinson, A.K., and Rodriguez-Quinones, F. (2003). Bacterial iron homeostasis. *FEMS Microbiol Rev* 27, 215-237.

Aziz, R.K., Bartels, D., Best, A.A., DeJongh, M., Disz, T., Edwards, R.A., Formsma, K., Gerdes, S., Glass, E.M., Kubal, M., Meyer F, Olsen, G.J., Olson, R., Osterman, A.L., Overbeek, R.A., McNeil, L.K., D., P., Paczian, T., Parrello, B., Pusch, G.D., Reich, C., Stevens, R., Vassieva, O., Vonstein, V., Wilke, A., and Zagnitko, O. (2008). The RAST Server: rapid annotations using subsystems technology. *BMC Genomics* 9, 75.

Blin, K., Medema, M.H., Kazempour, D., Fischbach, M.A., Breitling, R., Takano, E., and Weber, T. (2013). antiSMASH 2.0--a versatile platform for genome mining of secondary metabolite producers. *Nucleic Acids Res* 41, W204-212. doi: 10.1093/nar/gkt449.

Büdel, B., J., B., Bicker, F.R., and Allan Green, T.G. (2008). Dewfall As a Water Source Frequently Activates the Endolithic Cyanobacterial Communities in the Granites of Taylor Valley, Antarctica. *J Phycol* 44, 1415-1424.

Calamita, G. (2000). The Escherichia coli aquaporin-Z water channel. *Mol Microbiol* 37, 254-262.

Cámara, B., Suzuki, S., Nealson, K.H., Wierzchos, J., Ascaso, C., Artieda, O., and de los Ríos, A. (2014). Ignimbrite textural properties as determinants of endolithic colonization patterns from hyper-arid Atacama Desert. *Inter Microbiol* 17, 235-247.

Caporaso, J.G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F.D., Costello, E.K., Fierer, N., Pena, A.G., Goodrich, J.K., Gordon, J.I., Huttley, G.A., Kelley, S.T., Knights, D., Koenig, J.E., Ley, R.E., Lozupone, C.A., McDonald, D., Muegge, B.D., Pirrung, M., Reeder, J., Sevinsky, J.R., Turnbaugh, P.J., Walters, W.A., Widmann, J., Yatsunenko, T., Zaneveld, J., and Knight, R. (2010). QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* 7, 335-336. doi: nmeth.f.303 [pii]10.1038/nmeth.f.303.

Cary, S.C., McDonald, I.R., Barrett, J.E., and Cowan, D.A. (2010). On the rocks: the microbiology of Antarctic Dry Valley soils. *Nat Rev Microbiol* 8, 129-138. doi: nrmicro2281 [pii]10.1038/nrmicro2281.

Castresana, J. (2000). Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol Biol Evol* 17, 540-552.

Chan, Y., Lacap, D.C., Lau, M.C., Ha, K.Y., Warren-Rhodes, K.A., Cockell, C.S., Cowan, D.A., McKay, C.P., and Pointing, S.B. (2012). Hypolithic microbial communities: between a rock and a hard place. *Environ Microbiol* 14, 2272-2282. doi: 10.1111/j.1462-2920.2012.02821.x.

Chan, Y., Van Nostrand, J.D., Zhou, J., Pointing, S.B., and Farrell, R.L. (2013). Functional ecology of an Antarctic Dry Valley. *Proc Natl Acad Sci U S A* 110, 8990-8995. doi: 10.1073/pnas.1300643110.

Cockell, C.S., Pybus, D., Olsson-Francis, K., Kelly, L., Petley, D., Rosser, N., Howard, K., and Mosselmans, F. (2011). Molecular characterization and geological microenvironment of a microbial community inhabiting weathered receding shale cliffs. *Microb Ecol* 61, 166-181. doi: 10.1007/s00248-010-9730-6.

Cowan, D.A., Pointing, S.B., Stevens, M.I., Cary, S.C., Stomeo, F., and Tuffin, I.M. (2011). Distribution and abiotic influences on hypolithic microbial communities in an Antarctic Dry Valley. *Polar Biol* 34, 307-311.

da Costa, M.S., Santos, H., and Galinski, E.A. (1998). An overview of the role and diversity of compatible solutes in Bacteria and Archaea. *Adv Biochem Eng Biotechnol* 61, 117-153.

Darling, A.E., Jospin, G., Lowe, E., Matsen, F.A.t., Bik, H.M., and Eisen, J.A. (2014). PhyloSift: phylogenetic analysis of genomes and metagenomes. *PeerJ* 2, e243. doi: 10.7717/peerj.243.

de la Torre, J.R., Goebel, B.M., Friedmann, E.I., and Pace, N.R. (2003). Microbial diversity of cryptoendolithic communities from the McMurdo Dry Valleys, Antarctica. *Appl Environ Microbiol* 69, 3858-3867.

de los Ríos, A., Cary, C.S., and Cowan, D. (2014). The spatial structures of hypolithic communities in the Dry Valleys of East Antarctica. *Polar Biol* 37, 1823-1833.

de los Rios, A., Valea, S., Ascaso, C., Davila, A.F., Kastovsky, J., McKay, C.P., Gomez-Silva, B., and Wierzchos, J. (2010). Comparative analysis of the microbial communities inhabiting halite evaporites of the Atacama Desert. *Int Microbiol* 2, 79-89.

DiRuggiero, J., Wierzchos, J., Robinson, C.K., Souterre, T., Ravel, R., Artieda, O., Souza-Egipsy, V., and Ascaso, C. (2013). Microbial Colonization of Chasmoendolithic Habitats in the Hyper-arid Zone of the Atacama Desert. *Biogeosciences* 10, 2439-2450.

Dong, H., Rech, J.A., Jiang, H., Sun, H., and Buck, B.J. (2007). Endolithic cyanobacteria in soil gypsum: occurrences in Atacama (Chile), Mojave (USA), and Al-Jafr Basin (Jordan) Deserts. *J. Geophys. Res.* 112, G02030, doi:02010.01029/02006JG000385.

Dyhrman, S.T., Chappell, P.D., Haley, S.T., Moffett, J.W., Orchard, E.D., Waterbury, J.B., and Webb, E.A. (2006). Phosphonate utilization by the globally important marine diazotroph Trichodesmium. *Nature* 439, 68-71. doi: 10.1038/nature04203.

Edgar, R.C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32, 1792-1797.

Finn, R.D., Clements, J., and Eddy, S.R. (2011). HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res* 39, W29-37. doi: 10.1093/nar/gkr367.

Friedmann, E.I. (1982). Endolithic Microorganisms in the Antarctic Cold Desert. *Science* 215, 1045-1053. doi: 215/4536/1045 [pii]10.1126/science.215.4536.1045.

Garcia-Pichel, F., Wingard, C.E., and Castenholz, R.W. (1993). Evidence Regarding the UV Sunscreen Role of a Mycosporine-Like Compound in the Cyanobacterium Gloeocapsa sp. *Appl Environ Microbiol* 59, 170-176.

Gish, W., and States, D.J. (1993). Identification of protein coding regions by database similarity search. *Nature Genetics* 3, 266-272.

Herrera, A., Cockell, C.S., Self, S., Blaxter, M., Reitner, J., Thorsteinsson, T., Arp, G., Dröse, W., and Tindle, A.G. (2009). A cryptoendolithic community in volcanic glass. *Astrobiology* 9, 369-381.

Hibbing, M.E., Fuqua, C., Parsek, M.R., and Peterson, S.B. (2010). Bacterial competition: surviving and thriving in the microbial jungle. *Nat Rev Microbiol* 8, 15-25. doi: 10.1038/nrmicro2259.

Holzinger, A., and Karsten, U. (2013). Desiccation stress and tolerance in green algae: consequences for ultrastructure, physiological and molecular mechanisms. *Front Plant Sci* 4, 327. doi: 10.3389/fpls.2013.00327.

Houston, J., and Hartley, A. (2003). The central Andean west-slope rainshadow and its potential contribution to the origin of hyperaridity in the Atacama Desert. *Int J Climatol* 23, 1453-1464.

Hu, C., Gao, K., and Whitton, B.A. (2012). "Semi-arid regions and deserts," in *Ecology of cyanobacteria II: their diversity in space and time,* ed. B.A. Whitton. (Netherlands: Springer), 345-370.

Hyatt, D., Chen, G.L., Locascio, P.F., Land, M.L., Larimer, F.W., and Hauser, L.J. (2010). Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 11, 119. doi: 10.1186/1471-2105-11-119.

Kidron, G.J. (2000). Analysis of dew precipitation in three habitats within a small arid drainage basin, Negev Highlands, Israel. *Atmosph Res* 55, 257–270.

Kolter, R., Siegele, D.A., and Tormo, A. (1993). The stationary phase of the bacterial life cycle. *Annu Rev Microbiol* 47, 855-874. doi: 10.1146/annurev.mi.47.100193.004231.

Köster, W. (2001). ABC transporter-mediated uptake of iron, siderophores, heme and vitamin B12. *Res Microbiol* 152, 291-301.

Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., Jones, S.J., and Marra, M.A. (2009). Circos: an information aesthetic for comparative genomics. *Genome Res* 19, 1639-1645. doi: 10.1101/gr.092759.109.

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* 10, R25. doi: gb-2009-10-3-r25 [pii]10.1186/gb-2009-10-3-r25.

Leao, P.N., Engene, N., Antunes, A., Gerwick, W.H., and Vasconcelos, V. (2012). The chemical ecology of cyanobacteria. *Nat Prod Rep* 29, 372-391. doi: 10.1039/c2np00075j.

Leflaive, J.P., and Ten-Hage, L. (2007). Algal and cyanobacterial secondary metabolites in freshwaters: a comparison of allelopathic compounds and toxins. *Freshwater Biol* 52, 199-214.

McKay, C.P., Friedmann, E.I., Gomez-Silva, B., Caceres-Villanueva, L., Andersen, D.T., and Landheim, R. (2003). Temperature and moisture conditions for life in the extreme arid region of the Atacama desert: four years of observations including the El Niño of 1997-1998. *Astrobiology* 3, 393-406.

Meyer, F., Paarmann, D., D'Souza, M., Olson, R., Glass, E.M., Kubal, M., Paczian, T., Rodriguez, A., Stevens, R., Wilke, A., Wilkening, J., and Edwards, R.A. (2008). The metagenomics RAST server - a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC Bioinformatics* 9, 386. doi: 10.1186/1471-2105-9-386.

Michalski, G., Böhlke, J.K., and Thiemens, M. (2004). Long term atmospheric deposition as the source of nitrate and other salts in the Atacama Desert, Chile: New

evidence from mass independent oxygen isotopic compositions. *Geochim Cosmochim Acta* 68, 4023–4038.

Miller, C.S., Baker, B.J., Thomas, B.C., Singer, S.W., and Banfield, J.F. (2011). EMIRGE: reconstruction of full-length ribosomal genes from microbial community short read sequencing data. *Genome Biol* 12, R44. doi: 10.1186/gb-2011-12-5-r44.

Monds, R.D., Newell, P.D., Schwartzman, J.A., and O'Toole, G.A. (2006). Conservation of the Pho regulon in Pseudomonas fluorescens Pf0-1. *Appl Environ Microbiol* 72, 1910-1924. doi: 10.1128/AEM.72.3.1910-1924.2006.

Nienow, J.A. (2009). "Extremophiles: Dry Environments (including Cryptoendoliths)," in *Encyclopedia of Microbiology,* ed. S.M. Schaechter. 3rd edition ed (Oxford: Elsevier), 159-173.

Oren, A. (1997). Mycosporine‐like amino acids as osmotic solutes in a community of halophilic cyanobacteria. *Geomicrobiol J* 14, 231-240

Oren, A. (2007). "Diversity of organic osmotic compounds and osmotic adaptation in cyanobacteria and algae," in *Algae and cyanobacteria in extreme environments,* ed. J. Seckback. (Berlin: Springer).

Patterson, G.M.L., and Bolis, C.M. (1997). Fungal cell-wall polysaccharides elicit an antifungal secondary metabolite (Phytoalexin) in the cyanobacterium Scytonema ocellatum. *J Phycol* 33, 54-60

Peng, Y., Leung, H.C., Yiu, S.M., and Chin, F.Y. (2012). IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics* 28, 1420-1428.

Pierson III , L.S., and Pierson, E.A. (2010). Metabolism and function of phenazines in bacteria: impacts on the behavior of bacteria in the environment and biotechnological processes. *Appl Microbiol Biotechnol* 86, 1659–1670.

Pointing, S.B., and Belnap, J. (2012). Microbial colonization and controls in dryland systems. *Nat Rev Microbiol* 10, 551-562. doi: 10.1038/nrmicro2831.

Pointing, S.B., Chan, Y., Lacap, D.C., Lau, M.C., Jurgens, J.A., and Farrell, R.L. (2009). Highly specialized microbial diversity in hyper-arid polar desert. *Proc Natl Acad Sci U S A* 106, 19964-19969. doi: 0908274106 [pii]10.1073/pnas.0908274106.

Price, M.N., Dehal, P.S., and Arkin, A.P. (2010). FastTree 2--approximately maximum-likelihood trees for large alignments. *PLoS One* 5, e9490. doi: 10.1371/journal.pone.0009490.

Pruesse, E., Quast, C., Knittel, K., Fuchs, B., Ludwig, W., Peplies, J., and Glöckner, F.O. (2007). SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nuc Acids Res* 35, 7188-7196.

Puglisi, M.P., Sneed, J.M., Sharp, K.H., Ritson-Williams, R., and Paul, V.J. (2014). Marine chemical ecology in benthic environments. *Nat Prod Rep* 31, 1510-1553.

Roberts, M.F. (2005). Organic compatible solutes of halotolerant and halophilic microorganisms. *Saline Systems* 1, 5. doi: 10.1186/1746-1448-1-5.

Robinson, C.K., Wierzchos, J., Black, C., Crits-Christoph, A., Ma, B., Ravel, J., Ascaso, C., Artieda, O., Valea, S., Roldan, M., Gomez-Silva, B., and DiRuggiero, J. (2015). Microbial diversity and the presence of algae in halite endolithic

communities are correlated to atmospheric moisture in the hyper-arid zone of the Atacama Desert. *Environ Microbiol* 17, 299-315. doi: 10.1111/1462-2920.12364.

Singh, D.P., Prabha, R., Meena, K.K., Sharma, L., and Sharma, A.K. (2014). Induced accumulation of polyphenolics and favonoids in Cyanobacteria under salts tress protects organisms through enhanced antioxidant activity. *Am J Plant Sci* 5, 726-735.

Soueidan, H., Schmitt, L.A., Candresse, T., and Nikolski, M. (2014). Finding and identifying the viral needle in the metagenomic haystack: trends and challenges. *Front Microbiol* 5, 739. doi: 10.3389/fmicb.2014.00739.

Walker, J.J., and Pace, N.R. (2007). Endolithic microbial ecosystems. *Annu Rev Microbiol* 61, 331-347. doi: 10.1146/annurev.micro.61.080706.093302.

Watson, S.P., Clements, M.O., and Foster, S.J. (1998). Characterization of the starvation-survival response of Staphylococcus aureus. *J Bacteriol* 180, 1750-1758.

Wei, S.T.S., Fernandez-Martinez, M., Chan , Y., Van Nostrand, J.D., de los Rios-Murillo, A., Chiu, J.M.Y., Ganeshram, A.M., Cary, S.C., Zhou, J., and Pointing, S.B. (2015). Diverse metabolic and stress-tolerance pathways in chasmoendolithic and soil communities of Miers Valley, McMurdo Dry Valleys, Antarctica. *Polar Biol* 38, 433-443.

Wierzchos, J., and Ascaso, C. (1994). Application of backscattered electron imaging to the study of the lichen–rock interface. *J Microsc* 175, 54–59.

Wierzchos, J., Ascaso, C., and McKay, C.P. (2006). Endolithic cyanobacteria in halite rocks from the hyperarid core of the Atacama Desert. *Astrobiology* 6, 415-422.

Wierzchos, J., Camara, B., de Los Rios, A., Davila, A.F., Sanchez Almazo, I.M., Artieda, O., Wierzchos, K., Gomez-Silva, B., McKay, C., and Ascaso, C. (2011). Microbial colonization of Ca-sulfate crusts in the hyperarid core of the Atacama Desert: implications for the search for life on Mars. *Geobiology* 9, 44-60. doi: 10.1111/j.1472-4669.2010.00254.xGBI254 [pii].

Wierzchos, J., Davila, A.F., Artieda, O., Cámara-Gallego, B., de los Ríos, A., Nealson, K.H., Valea, S., García-González, M.T., and Ascaso, C. (2013). Ignimbrite as a substrate for endolithic life in the hyper-arid Atacama Desert: Implications for the search for life on Mars. *Icarus* 224, 334–346.

Wierzchos, J., de los Ríos, A., and Ascaso, C. (2012). Microorganisms in desert rocks: the edge of life on Earth. *Inter Microbiol* 15, 173-183.

Wierzchos, J., DiRuggiero, J., Vítek, P., Artieda, O., Souza-Egipsy, V., Skaloud, P., Tisza, M., Davila, A.F., Vílchez, C., Garbayo, I., and Ascaso, C. (2015). Adaptation strategies of endolithic chlorophototrophs to survive the hyperarid and extreme solar radiation environment of the Atacama Desert. *Front Microbiol* 6, 934. doi: 910.3389/fmicb.2015.00934.

Wood, D.E., and Salzberg, S.L. (2014). Kraken: ultrafast metagenomic sequence classification using exact alignments. *Genome Biol* 15, R46. doi: 10.1186/gb-2014-15-3-r46.

Wu, Y.W., Tang, Y.H., Tringe, S.G., Simmons, B.A., and Singer, S.W. (2014). MaxBin: an automated binning method to recover individual genomes from metagenomes using an expectation-maximization algorithm. *Microbiome* 2, 26. doi: 10.1186/2049-2618-2-26.

1   Yancey, P.H. (2005). Organic osmolytes as compatible, metabolic and counteracting
2   cytoprotectants in high osmolarity and other stresses. *Exp Biol* 208, 2819–2830.
3   Zambrano, M.M., Siegele, D.A., Almiron, M., Tormo, A., and Kolter, R. (1993).
4   Microbial competition: Escherichia coli mutants that take over stationary phase
5   cultures. *Science* 259, 1757-1760.

1 **Tables**

2

3 Table 1: Microclimate data for the VL (calcite) and LdT (ignimbrite) sampling sites

Sampling sites	PAR µmol m^{-2} s^{-1}		Air temperature (°C)			Air relative humidity (%)			Rainfall (mm)
	ave	max.	min.	ave	max.	min.	ave	max.	ave
Valle de la Luna (VL)[a]	-	2600	-1.8	13.1	32.2	16.7	40.5	80	27.8
Lomas to Tilocalar (LdT)[b]	442	2553	- 4.6	15	44.8	1.2	18.3	100	25.6

Maximum (max.), minimum (min.) and yearly average (ave) values for photosynthetic active radiation (PAR), temperature, and relative humidity (RH); [a]data from DiRuggiero et al., 2013; [b]data collected by HOBO weather station from April 2011-2013; "-" no data.

4
5
6 Table 2: Number of sequence reads, Gbp of sequences, and contigs statistics for the
7 calcite and ignimbrite metagenomes

Sample	Number of paired-end reads[1]	Number of Gbp sequenced	Number of Gbp assembled	Number of contigs	Contig mean length (bp)	Contig maximum length (bp)
Ignimbrite (MTQ)	100,086,740	10.1	0.22	166,509	1,351	101,619
Calcite (VL)	126,820,576	12.8	0.35	344,986	1,016	119,637

[1]after quality control

8
9
10 Table 3: Most abundant species in the calcite and ignimbrite communities obtained by
11 binning the metagenome sequence data

Calcite							Ignimbrite						
Bin number	Taxa	%[a]	Estimated completeness %	Number of contigs	Size Mbp	G+C %	Bin number	Taxa	%[a]	Estimated completeness %	Number of contigs	Size Mbp	G+C %
Ca01	*Gloeocapsa sp.*	20.7	95.3	2329	7.7	48	Ig01	*Chroococcidiopsis sp.*	32.8	73.8	2031	5.1	45
Ca02	*Gloeocapsa sp.*	9.2	46.7	2388	5.6	46	Ig02	*Rubrobacter xylanophilus*	15.1	38.3	1215	2.2	68
Ca03	*Rubrobacter xylanophilus*	4.2	43	413	1.2	65	Ig04	*Rubrobacter xylanophilus*	5.9	31.8	1900	6	67
Ca04	*Rubrobacter xylanophilus*	3.9	35.5	330	1.5	69	Ig05	*Frankineae sp.*	5.1	74.8	460	4.5	75
Ca05	*Acaryochloris sp.*	3.6	68.2	1027	5.5	48	Ig06	*Conexibacter woesei*	2.7	87.9	357	3.9	75
Ca06	*Conexibacter woesei*	3.4	45.8	307	1.9	74	Ig07	*Frankineae sp.*	2.3	93.5	457	3.9	72
Ca07	*Thermomicrobia sp.*	3.2	83.2	1161	3.9	72	Ig09	*Thermobaculum sp.*	2.0	86	661	3.5	69
Ca09	*Gloeocapsa sp.*	2.8	94.4	800	6.5	45	Ig10	*Thermomicrobia sp.*	2.0	89.7	548	5.1	63
Ca17	*Conexibacter woesei*	1.9	86	483	3.3	72	Ig11	*Conexibacter woesei*	1.9	59.8	575	2.9	76
Ca19	*Microlunatus sp.*	1.6	91.6	538	4.3	71	Ig12	*Chroococcidiopsis sp.*	1.7	93.5	817	6.6	46

[a]Relative abundance ratios; defined as actual coverage levels divided by summed coverage levels of all genomes.

12
13

1   **FIGURE LEGENDS**

2   Fig. 1. Chasmoendolithic microbial communities within the fissures and cracks of calcite
3   rock. (a) Freshly fractured calcite show fissure wall with green spots – chasmoendoliths;
4   black arrow point to parallel to the rock surface fissure also colonized; white arrow,
5   hardened layer with microrills (asterisk). (b) SEM-BSE low magnification image of
6   transversal section showing the connected to the rock surface fissure full-fill by
7   chasmoendolithic colonization; the square, black and white arrows point to detailed (c),
8   (d) and (e) images respectively. (c) SEM-BSE image shows chasmoendolith's aggregates
9   colonizing fissure walls. (d) SEM-BSE image shows well developed cyanobacteria
10   surrounded by EPSs and heterotrophic bacteria (arrow). (e) SEM-BSE image shows
11   collapsed cyanobacteria cells and remains of microbial colonization (black arrows). (f)
12   SIM 3D image of cyanobacteria aggregates; red signal – cyanobacteria photosynthetic
13   pigments autofluorescence, blue signal – cyanobacteria nucleoids and green signal
14   sheaths of EPSs surrounding cyanobacteria aggregates and singular cells.

15

16   Fig. 2. Cryptoendolithic microbial communities within the pores of ignimbrite. (a)
17   Freshly fractured ignimbrite show narrow green layer of cryptoendolithic
18   microorganisms beneath the rock surface (arrow). (b) SEM-BSE image showing
19   micromorphological features of rhyolitic ignimbrite composed of glass shards welded
20   together forming bottle-shaped vesicles (left image side) and irregular-shape pores (right
21   image side); white arrow point to the surface pore filled by varnish rock mineral
22   components; aggregates of cryptoendolithic microbial colonization were marked by cyan-
23   colour and black arrow points to aggregate in detail shown in fig. 2c. (c) SEM-BSE
24   image of cyanobacteria composed aggregate containing heterotrophic bacteria (white
25   arrows). (d) SEM-BSE image of empty and colonized bottle-shaped pores. (e) SEM-BSE
26   image of detail view of cyanobacteria aggregates containing heterotrophic bacteria (white
27   arrow). (f) Magnified central part of fig. 2e showing heterotrophic bacteria in close
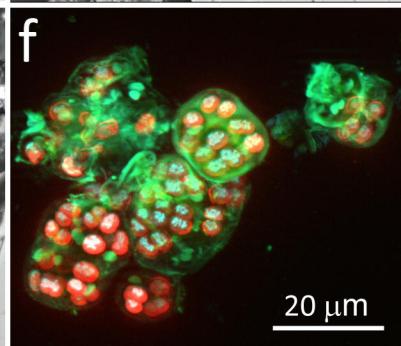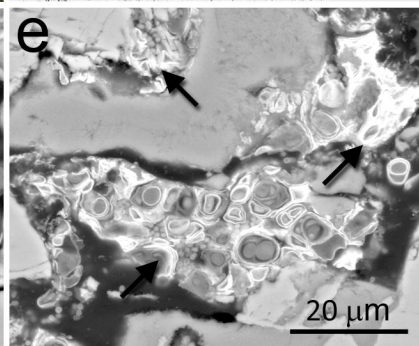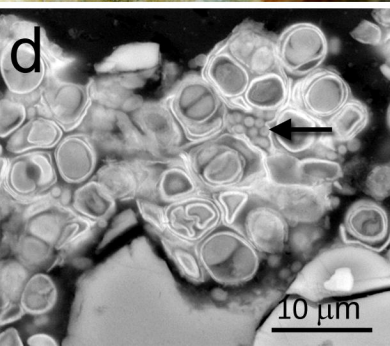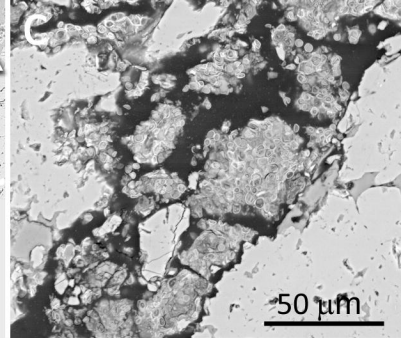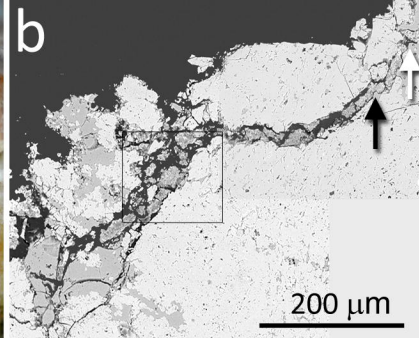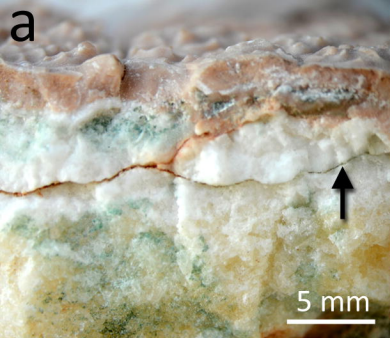28   position to the declined cyanobacteria and their remains.

29

30   Fig. 3. Taxonomic composition at the phylum level for the calcite and ignimbrite
31   communities using either 16S rRNA gene sequences (16S) or PhyloSift marker genes
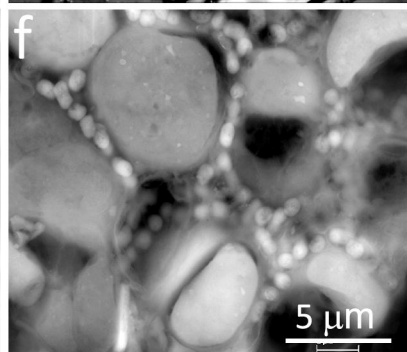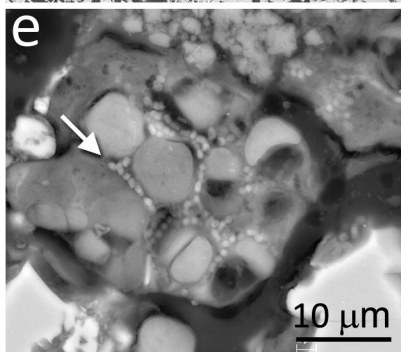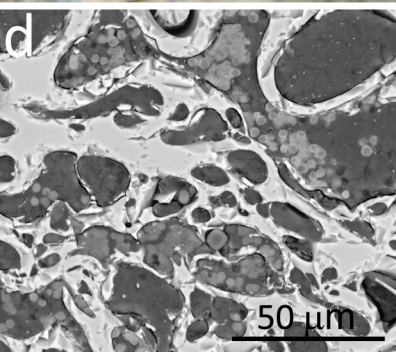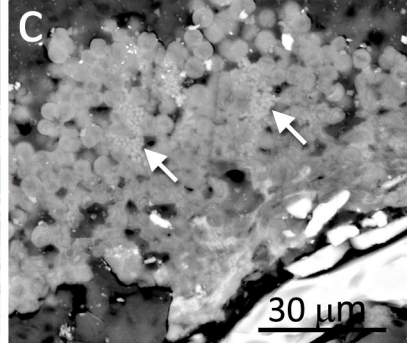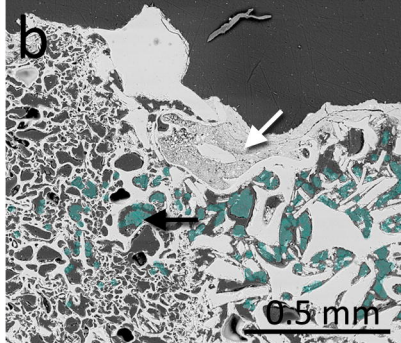32   from the metagenome data (Metag) (10% relative abundance cut-off).

33

34   Fig. 4. Normalized abundance of functional categories for the calcite and ignimbrite
35   community metagenome sequence reads for (a) SEED subsystems and (b) COG
36   generated with MG-RAST.

37

38   Fig. 5. Number of non-ribosomal peptides and polyketides genes per Mbp in
39   cyanobacteria draft genomes from the calcite (Ca1 and Ca9) and ignimbrite (Ig1 and
40   Ig12) communities using AntiSmash predictions.

41

42   Fig. 6. NORINE predictions for putative products of non-ribosomal peptides and
43   polyketides gene clusters found in the draft genomes of cyanobacteria from the calcite
44   (Ca9) and ignimbrite (Ig12) communities.

45

1   Fig. 7. Relationships between contigs from the calcite Ca09 and ignimbrite Ig12 draft
2   genomes. Links are between gene products with greater than 80% amino acid identity.
3   Highlighted in blue are secondary metabolite clusters identified by antiSMASH, and
4   marked in black are Ferrichrome-iron receptor proteins. Figure produced with Circos
5   (Krzywinski 2009).

a — 5 mm

b — 0.5 mm

c — 30 μm

d — 50 μm

e — 10 μm

f — 5 μm

Relative Abundance (y-axis, 0% to 100%)

Categories (x-axis): Calcite 16S, Calcite Metag., Ignimbrite 16S, Ignimbrite Metag.

Legend:
- Other Bacteria
- Euryarchaeota
- Verrucomicrobia
- Firmicutes
- Gemmatimonadetes
- Thermi
- Acidobacteria
- Bacteroidetes
- Proteobacteria
- Chloroflexi
- Actinobacteria
- Cyanobacteria

a

■ Calcite
■ Ignimbrite

Phosphorus Metabolism
Membrane Transport
Regulation and Cell Signaling
RNA Metabolism
Dormancy and Sporulation
Protein Metabolism
Nucleosides and Nucleotides
Miscellaneous
Secondary Metabolism
Virulence, Disease and Defense
DNA Metabolism
Phages, Prophages, Transposable elements, Plasmids
Motility and Chemotaxis
Clustering-based subsystems
Stress Response
Cofactors, Vitamins, Prosthetic Groups, Pigments
Cell Wall and Capsule
Fatty Acids, Lipids, and Isoprenoids
Carbohydrates
Amino Acids and Derivatives
Metabolism of Aromatic Compounds
Potassium metabolism
Sulfur Metabolism
Iron acquisition and metabolism
Respiration
Nitrogen Metabolism
Photosynthesis
Cell Division and Cell Cycle

b

Secondary metabolites biosynthesis, transport and catabolism
Replication, recombination and repair
Carbohydrate transport and metabolism
Translation, ribosomal structure and biogenesis
Transcription
Energy production and conversion
Amino acid transport and metabolism
Cell wall/membrane/envelope biogenesis
Signal transduction mechanisms
Posttranslational modification, protein turnover, chaperones
Defense mechanisms
Nucleotide transport and metabolism
Intracellular trafficking, secretion, and vesicular transport
Inorganic ion transport and metabolism
Cell cycle control, cell division, chromosome partitioning
Lipid transport and metabolism
Cell motility
RNA processing and modification
Coenzyme transport and metabolism
Cytoskeleton
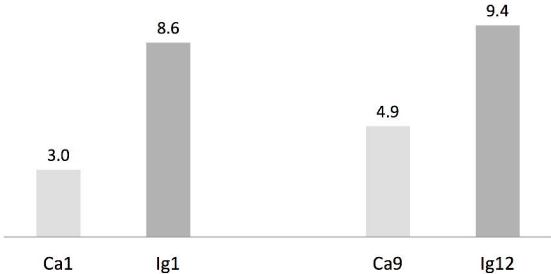Chromatin structure and dynamics
Extracellular structures

| Calcite | Ignimbrite |

Compound	Biosynthetic gene cluster (antiSMASH)	Family (NORINE)	Activity (NORINE)
a	NRPS	ACV	antibiotic
b	NRPS-T1PKS Microcystin (siderophore)	coronatine	toxin
c	NRPS-T1PKS Anabaenopeptin	largamide	protease inhibitor
d	NRPS Xenoamicins		
e	NRPS	pyoverdin	siderophore
f	NRPS-T1PKS	pyoverdin	siderophore
g	NRPS Nostophycin	coronatine	toxin
i	NRPS similar to bacteriocins		
I	NRPS siderophore	coronatine	toxin

Calcite

Ignimbrite