Living to the high extreme: unraveling the composition, structure, and functional insights of bacterial communities thriving in the arsenic-rich Salar de Huasco – Altiplanic ecosystem.

Castro-Severyn, J¹; Pardo-Esté, C¹-²; Mendez, KN³; Fortt, J¹; Marquez, S³; Molina, F⁴; Castro-Nallar, E³; Remonsellez, F¹,⁵* and Saavedra, CP²*.

¹ Laboratorio de Microbiología Aplicada y Extremófilos, Facultad de Ingeniería y Ciencias Geológicas, Universidad Católica del Norte, Antofagasta, Chile.
² Laboratorio de Microbiología Molecular, Facultad de Ciencias de la Vida, Universidad Andres Bello, Santiago, Chile.
³ Center for Bioinformatics and Integrative Biology, Facultad de Ciencias de la Vida, Universidad Andres Bello, Santiago, Chile.
⁴ Sys2Diag, UMR9005 CNRS ALCEDIAG, Montpellier, France
⁵ Centro de Investigación Tecnológica del Agua en el Desierto-CEITSAZA, Universidad Católica del Norte, Antofagasta, Chile.

± Juan Castro-Severyn and Coral Pardo-Esté have contributed equally to this work. Author order was determined both by contribution and in order of increasing seniority.

*Correspondence: Francisco Remonsellez, fremonse@ucn.cl and Claudia P. Saavedra, csaavedra@unab.cl.

Running title: Altiplanic communities living under arsenic stress.
Microbial communities inhabiting extreme environments like Salar de Huasco (SH) are adapted to thrive while exposed to several abiotic pressures and the presence of toxic elements like arsenic (As). Hence, we aimed to uncover the role of arsenic in shaping bacterial composition, structure, and functional potential in five different sites in this Altiplanic wetland using a shotgun metagenomic approach. The sites exhibit wide gradients of arsenic (9 to 321 mg/kg), and our results showed highly diverse communities and a clear dominance exerted by the Proteobacteria and Bacteroidetes phyla. Functional potential analyses showed broadly convergent patterns, contrasting with their great taxonomic variability. Arsenic-related metabolism is different among the five communities, as well as other functional categories like those related to the CH₄ and S cycles. Particularly, we found that the distribution and abundance of As-related genes increase, following along the As concentration gradient. Approximately 75% of the detected genes for As-metabolism belong to expulsion mechanisms, being arsJ and arsP pumps related to sites with higher As concentrations and present almost exclusively in Proteobacteria. Furthermore, taxonomic diversity and functional potential are reflected in the 12 reconstructed high-quality MAGs (Metagenome Assembled Genomes) belonging to the Bacteroidetes (5), Proteobacteria (5), Cyanobacteria (1) and Gemmatimonadota (1) phyla. We conclude that SH microbial communities are diverse and possess a broad genetic repertoire to thrive under extreme conditions, including increasing concentrations of the highly toxic As. Finally, this environment represents a reservoir of unknown and undescribed microorganisms, with a great metabolic versatility, which needs further study.

KEY WORDS: Extremophiles, Metagenomics, Arsenic, MAGs, Altiplano.

IMPORTANCE

Microbial communities inhabiting extreme environments are fundamental for maintaining the ecosystems; however, little is known about their potential functions and interactions
among them. We sampled the microbial communities in Salar de Huasco (SH) in the Chilean Altiplano, a fragile and complex environment that comprises several stresses. We found that microbes in SH are taxonomically diverse; nonetheless, their functional potential seems to have an important convergence degree, suggesting high adaptation levels. Particularly, arsenic metabolism showed differences associated with increasing concentrations of the metalloid throughout the area, and it is effectively exerting a clear and significant pressure over these organisms. Thus, this research’s significance is that we described highly specialized communities thriving in little-explored environments under several pressures, considered analogous of early Earth and other planets, and can have the potential for unraveling technologies to face climate change repercussions in many areas of interest.

INTRODUCTION

Extreme environments like high-altitude wetlands select for adaptations in bacterial communities that enable them to thrive. This particular and fragile environment resemble life before the oxygenation of Earth and could serve as models for studying life in other planets. Moreover, microbial communities are critical for maintaining biogeochemical cycles, particularly in extreme environments where there is little presence of other life forms. Therefore, microorganisms are crucial for ecosystems’ health and functioning (1).

Salar de Huasco (SH), a high-altitude wetland located in the Chilean Altiplano (20°18’18’’S; 68°50’22’’W, Chile) at 3,800 m.a.s.l. is a Ramsar protected site, considered a hotspot for microbial life (2,3). This area is labeled as extreme due to very particular confluence of physicochemical and environmental conditions, like negative water balance, high daily temperature variation, very arid conditions, high salinity, low atmospheric pressure, high solar radiation, presence of arsenic, among other stressors (4, 5, 6, 7, 8, 9).

In high altitude wetlands, arsenic concentration, moisture availability, and salt concentrations model communities at a small scale (10, 11, 12, 13). The presence of toxic metal(oids), such as arsenic, is one of the main drivers of microbial communities’
composition (14), as an important selection pressure originated from natural geochemical processes or human activities. Also, previous studies have determined that the presence of metal(oids) can influence biogeochemical cycles, namely C, N and S, by promoting specific chemical reactions and the enrichment of chemoautotrophs, for example As(III)-oxidizing bacteria that can couple this process with nitrate reduction (15, 16, 17); thus, making relevant to assess the metabolic potential of indigenous microbes and communities. In the North of Chile, the relationship between volcanism activities and the presence of arsenic is a known feature, this geological process has been attributed to hydrothermal conditions such as geysers and fumaroles in the pre-range and high plateau (5). Besides, bacteria that thrive in these environments exhibit arsenic-related genes, including genes associated with methylation (ArsM), oxidation (AioAB, ArxAB), and dissimilatory reduction (ArrAB) (18). The most common resistance mechanism is based on As(III) extrusion from the cell by efflux pumps, which is commonly coupled with As(V) reduction (19). Thus, making the genes from the ars operon the most abundant arsenic resistance markers; the basic requirements comprises the arsenite efflux (ArsB, Acr3) and arsenate reductase (ArsC) (20).

Arsenic is a crucial element for the microbial community composition and metagenomic approaches have helped to shed some light on this subject. For instance, in Socompa’s stromatolites arsenic resistance is achieved mainly through reduction and expulsion of As(V) via Acr3 efflux pumps reductases (21, 22). Additionally, others have found that the presence of arsenic influenced the microbial communities, and the Rhodococcus genus was significantly enriched in elevated levels of the metalloid in groundwater (14). Furthermore, arsenic mobilization is distributed across a broad phylogenetic lineage in similar ecosystems, as arrA was detected in Betaproteobacteria, Deltaproteobacteria, and Nitrospirae MAGs (23). Also, although arsenic-related genes are widespread, they are not universal, in particular As metabolism genes such as aioA, arrA, arsM, arxA are less common in the environment compared with acr3, arsB, arsC and arsD (24).
In this study, we tested different established sites in the SH regarding taxonomic and functional heterogeneity. We focused on As-related genetic elements as well as other relevant metabolic functions. Evermore, such analyses enable the identification of relatively small subsets of markers associated with a particular ecologically important function. Besides, identifying those and the changes of bacterial communities are potentially useful bio-indicator for monitoring ecosystem health (25, 26). Therefore, considering the aforementioned we set up to describe and characterize the composition, structure and functional potential of bacterial communities from the sediments of five different Salar de Huasco sites along an arsenic gradient, among other environmental pressures.

RESULTS

The Salar de Huasco comprises an important level of variation (27, 28), evidenced within a relatively small area (sampled area spans a distance of 5.9 km); including daily oscillations in a wide range of environmental parameters (temperature and humidity) and other which vary in spatial gradients, i.e., arsenic and salinity increasing from north to south (Figure 1) (9). In this context, shotgun metagenomic sequencing of the sediment samples representing the five SH sites yielded an average of 87.2 million reads (150 bp length) per sample, with a quality score ≥ 30 presented by ~95% of those obtained reads.
Figure 1. Salar de Huasco study area. Top panel: a map showing the five sampling sites (H0: 20°15’48.8"S - 68°52’28.4"W; H1: 20°16’27.7"S - 68°53’3"W; H3: 20°16’59.2"S - 68°53’16.7"W; H4: 20°17’40.9"S - 68°53’17.3"W and H5: 20°18’37"S - 68°52’42"W) investigated in this study. SH is located between the 68°47' - 68°54' W and 20°15' - 20°20' S in the Tarapacá region Northern Chile (Source: Google Earth). Bottom panel: illustration of the SH landscape seen from the H0 site to the South-West; the signs show some particular characteristics of each site (according to those reported in Castro-Severyn et al., 2020 (9)). The painting is the work of the illustrator Florence Gutzwiller @spideryscrawl.
The bacterial communities of SH are highly heterogeneous and rich in unknown taxa

The contrasting differences presented among the SH sampling sites are reflected by the taxonomic composition of bacterial communities, which showed contrasting patterns and variation between sites (Figure 2). Overall, our results indicate that the Proteobacteria and Bacteroidetes are the most prevalent phyla in the sampled sediments, which are particularly enriched in the H3, H4 and H5 communities, accounting together for >60% of all observed taxa (Figure 2A). In turn, these phyla represent ~50% in the H0 and H1 communities. Interestingly, the H0 site is dominated by the Cyanobacteria phylum with a 34% of the total community; also, this community has the highest abundance of Firmicutes, Patescibacteria and Spirochaetes phyla as well as significantly lower proportion of Actinobacteria. Moreover, H1 community profile presents the highest abundance of Chloroflexi, Actinobacteria, Verrucomicrobia, Planctomycetes and Acidobacteria phyla. On the other hand, H3, H4 and H5 communities are more similar to each other, despite the vast differences in low-abundance taxa; many of which are exclusively present in only one community (Supplementary Table S1).

The most abundant bacteria at the lowest (available) taxonomic rank reflect the same pattern, with a great abundance of Proteobacteria genera Roseovarius and Desulfotignum in the H3, H4 and H5 communities (Supplementary Figure S1). Moreover, Halomonas, Thiobacillus, Luteolibacter and Truepera genera are more widespread, while Rhodohalobacter, Marinobacter, Psychroflexus and Brumimicrobium are also concentrated in H3, H4 and H5. Contrary, H0 and H1 are enriched in Cyanobacteria like Arthrospira and three members of the Chloroplast order, as well as genera with various metabolism types such as Methylibium, Hydrogenophaga and Desulfomicrobium. In addition, it is worth mentioning that from the 3,801 bacterial ASVs detected, 6.5% and 58.3% could not be classified within any known phylum nor genus respectively. Particularly, Halomonas and Marinobacter are culturable bacteria and seem to be recurrent in the SH, as we reported before (29). Nonetheless, H4 community is by far the most diverse one, according to the
Observed, Shannon, Chao and Simpson diversity indices, which were quite heterogeneous between samples (Supplementary Figure S2). Specifically, the phylogenetic diversity showed that the H0 community has a higher number of distant taxa regarding the other communities under study.

As a whole, the beta diversity analysis shows that the dispersion patterns among the communities correlates with what was observed in terms of composition. Hence, the taxonomic composition of H3, H4 and H5 metagenomes are more similar between each other and distinctive from H0 and H1, producing three very well-defined groups (Figure 2B). This is contrasting to the obtained alpha diversity results, suggesting that their members could differ in abundance. Thus, confirming the great level of structure among the communities, considering the total bacterial abundance and taxa diversity. Moreover, the distribution of most abundant taxa among the five communities could explain this segregation. Furthermore, we investigated to what extent the microbial community structure was explained by the environmental factors and found that many variables exert a significant influence over the structure and distribution/grouping, among which arsenic appears to be a main driving force shaping these communities. Besides, salinity along with arsenic separate the H3, H4 and H5 communities, while H0 and H1 segregation is driven by altitude and pH respectively. Nonetheless, for the temperature we have to consider that the observed variation may be the result of the normal daily cycle with respect to the time when samples were taken.

The co-occurrence analysis allowed us to infer possible interactions among complex microbiomes. The network was composed by 112 ASVs having at least one significant correlation, graphically represented as a link between taxa. Total correlations were 155, being 115 positive and 40 negatives. Furthermore, the co-occurrence analyses showed three distinct separated networks within these microbial assembly. However, there is a big main group that comprises most taxa. As expected, the dominant phyla, *Proteobacteria* and *Bacteroidetes*, are among the nodes with a highest degree of interactions (Figure 2C).
worth mentioning that most of the relevant (highly connected) taxa identified in the networks are little described or studied. Overall, there is the prevalence of direct or indirect relationships among taxa, and this is suggestive of complementing functions in order to maintain the ecosystem. Also, there are few negative correlations that would suggest little niche overlap or competition among taxa.
Figure 2. Salar de Huasco bacterial communities. A) Taxonomic composition and relative abundance of SH microbial communities in the five studied sites; Stacked-bars show the top20 most abundant bacteria at the phylum taxonomic level. B) Beta diversity by Principal Component Analysis (PCoA) on Hellinger transformed amplicon sequence variant (ASV) relative abundances. Each point corresponds to a community from the different sites (represented colors), and its relative distance indicates the level of similarity to all other
samples. The arrows indicate the explanatory power of the statistically significant environmental parameters with regards to the observed variation in community composition. C) Co-occurrence network analysis of the SH bacterial communities, the size of each node (representing ASVs) is proportional to the number of connections (degrees), the color of the edges connecting nodes represent the interaction type, the node color indicates the taxonomic affiliation at phylum level and node labels are at the lowest available taxonomic classification.

**Functional approach of SH communities reveals metabolic specialization**

The five SH metagenomes we analyzed represent the surface sediment of 5 different sites with a great variation in arsenic content (from 9 to 321 mg/Kg). The 337.5 million quality-controlled reads were co-assembled and yielded a total of 994,545 contigs (≥ 1,000 bp) with 2.38 millions genes, which were used to generate abundance/distributions and functional profiles of the 5 studied communities (Figure 3; Supplementary Table S2). Of those contigs, 148,394 (≥ 2,500 bp) were hierarchically clustered to then be profiled by read-recruitment of the data from the five communities (Figure 3A). The resulting patterns are highly variable, considering the contigs detection in each metagenome. In particular, metagenomes from H1 and H5 are the most contrasting ones, as they respectively have 18.78% and 78.40% presence of all available contigs in the SH. Moreover, the rest of the samples also have diverse percentages of representation (H0: 25.00%; H3: 72.08% and H4: 48.33%), and their clustering by the read recruitment profiles correlates with the taxonomic profiles presented previously; being H3, H4 and H5 more similar to each other, and H0 and H1 more distanced from each other and to the others as well. Contrary, the functional profiles seem to have a higher level of convergence between samples in a broad way (Figure 3B). Furthermore, the more enriched categories (SEED Subsystems 1) are associated with metabolism (amino acid derivatives, carbohydrates, protein metabolism, DNA metabolism, cofactor, vitamins, prosthetic groups and pigments) which could be evidence of the needed versatility and adaptability for bacteria to thrive in these harsh environments. Also, the stress response, membrane transporter, cell wall and capsule categories, which are related to the ability to thriving capabilities as well.
Nonetheless, statistical testing showed that the differences presented by the metagenomes in most categories were significant (Supplementary Table S3). Particularly, the five SH communities showed particular differences in some categories such as arsenic resistance (SEED Subsystems 3), which follows the same tendency previously mentioned; where H0/H1 metagenomes are more similar regarding H3/H4/H5, with 0.21 - 0.22% and 0.26 - 0.27% of reads recruited by this category, respectively. Other significant categories among the communities were carbon, nitrogen, and phosphate metabolism; Zinc, Nickel, Cobalt, Iron and Manganese transport; osmotic stress; circadian clock in Cyanobacteria and the Calvin Benson cycle. Whereas, when assessing the ability of these bacterial communities to carry out necessary reactions to sustain some biogeochemical cycles, we found that S, N and CH$_4$ showed significant scores (MEBS analysis) for all five metagenomes; implying that the necessary metabolic pathways and machinery is present and with proper completeness (Supplementary Figure S3). Nevertheless, a wide difference is present in the CH$_4$ cycle, which is much more enriched in the H1 and H0 communities; especially in the Methane oxidation, Methanol, Methanogenesis and $mcrABC$ (markers) pathways. Also, a small decrease in H1 sulfur is also observed.
**Figure 3. Salar de Huasco metagenomes.** A) Circular map representing the universe of contigs detected in the five SH sites: the central tree represents the contigs organization on Ward’s linkage with Euclidean distances, the seven circle layers (from bottom-up) represent, for the corresponding contig: its length, GC-content, and presence on the five metagenomes. The top-right bars represent the number of mapped reads for the corresponding metagenome and the dendrogram their clustering by read-recruitment profile. B) Patterns of functional potential for each metagenome, according to the presence and abundance of the SEED database of metabolic pathways and functions: subsystems at level 1. The circles represent the percentage value for the corresponding category in each metagenome (defined by colors).

**Arsenic expulsion is the main mechanism to thrive in the SH**

To gain a broader view on the communities’ functional potential related to arsenic metabolism or resistance/tolerance we determined the abundance of the genes belonging to the known mechanisms (Figure 4). Again, the abundance pattern of these genes correlates with the observed tendency, being more related those H3, H4 and H5 sites, distant from H0 and H1. Moreover, although marker genes of arsenic methylation, reduction, oxidation, respiration, and expulsion mechanisms are present in all sites, most are significantly more abundant in the H3, H4 and H5 sites. This may be due to the much higher arsenic concentration in the sediments of these sites. The *arsR* regulator, the *arsC* reductase and the *acr3* pump were the most abundant ones, which would indicate that the As(V) reduction and subsequent As(III) expulsion would be the most common strategy used by the SH inhabitant bacteria. Moreover, *arsM* was also among que most abundant genes, particularly enriched in the H3, H4 and H5 metagenomes. Whereas, genes related to oxidation and respiration mechanisms are less abundant and present an interesting distribution; being those related to oxidation (*aoxA, aoxB, aioB*) are enriched in H3, H4 and H5, contrary to those related to respiration (*arrA, arxR*) are enriched in H0 and H1.
Figure 4. Distribution and abundance of As metabolism genes in the Salar de Huasco.

Heatmap shows the (normalized) number of reads that aligned against the corresponding protein identified in each sample, according to the color scale. Genes are grouped by colors, representing the 5 As response mechanisms.

Overall, the proportional distribution pattern of the mechanisms in the 5 metagenomes had some similarity and is quite constant (Figure 5). Particularly, the expulsion mechanism as a whole was the most abundant one for all the sites covering around 75% of all the sequences (Figure 5A); observing the greatest differences in oxidation and respiration mechanisms, as we stated before. Additionally, respiration genes are more present in H0 and reduction in H1. Interestingly, the pattern between the five metagenomes...
seems to be independent of the amount of detected sequences and the arsenic concentration of the site. Furthermore, comparing the abundance of particular arsenic efflux pumps (from the expulsion mechanism) we can observe that, indeed, most of the sequences correspond to *acr3* (Figure 5B); followed by *arsB* and *arsJ*. Notably, *arsP* which is an efflux permease that confers resistance to organic arsenics (roxarsone and methylarsenite) was more abundant in the sites with higher arsenic concentration. Moreover, the higher *acr3* variants and abundance could be due this pump is present in a wider number of bacteria phyla, covering most of the found diversity (Figure 5C). Also, this efflux pump seems to be more ancestral, regarding *arsP* and *arsJ* which are only present in *Proteobacteria* and in more recent branches.
Figure 5. Arsenic metabolism in the Salar de Huasco. A) Distribution and abundance of detected genes related to arsenic (As) metabolism in the five SH metagenomes; Stacked bars represent the proportion of all the genes grouped by mechanism type based on relative abundance (%), and the line represents total As concentration in each corresponding site (mg/Kg of sediment). B) Stacked bars represent the proportion of all detected As efflux pumps in the five metagenomes, based on genes relative abundance (%). C) Phylogenetic analysis of 306 non-redundant sequences of arsenic efflux pumps detected in the SH, layers surrounding the phylogenomic tree indicate: detection level, type of efflux pump, taxonomy, the phylum level, and the species.

Novel genomes from SH belong undescribed genera

The binning the process reconstructed 195 bins (which were manually curated and evaluated for completion and redundancy) to gain insights of non-culturable bacteria. This resulted in 19 metagenome-assembled genomes (MAGs) that met the completion ≥ 80% and redundancy ≤ 10% criteria (Figure 6); that clustered 4.99% of the contigs in the metagenome profile database. The MAGs taxonomic affiliation resulted in 1/19 belonging to the Eukaryota domain and the 18/19 remaining to the Bacteria domain, distributed in 4 different phyla (9 Proteobacteria, 7 Bacteroidota, 1 Cyanobacteria and 1 Gemmatimonadetes (Figure 6A). We only were able to assign 22% (4/18) of the Bacteria MAGs to previously described genera, suggesting a great diversity of novel species that are yet to be described. Also, MAGs contigs number was very variable from 208 to 626 among the Bacteria (Eukaryota being substantially larger: 4,174), the same tendency was observed for the GC-content (36.1% - 69.4%). On the other hand, most of the reconstructed MAGS were represented or detected in the H3, H4 and H5 metagenomes, showing differential abundance patterns in each one (Figure 6B; Supplementary Table S4).
Figure 6. Salar de Huasco Metagenome-Assembled Genomes. A) Genomic features summary of the SH 19 MAGs, quality indices, and taxonomic affiliation (the results of *Anvi’o pipeline and/or **CheckM meet the quality standard of completion \( \geq 90\% \) and redundancy/contamination \( \leq 10\% \) (according to 125). B) Presence and abundance of the 19 MAGs on the five SH metagenomes: the central tree represents the MAGs organization on Ward’s linkage with Euclidean distances, the seven circle layers (from bottom-up) represent, for the corresponding MAG: its length, GC-content, and abundance on the five metagenomes. The top-right bars represent the total number of mapped reads for the corresponding metagenome and the dendrogram their clustering by absence/presence profile.

**Novel genomes from SH are mostly arsenic reducers**

The functional potential of reconstructed MAGs was evaluated globally and particularly in relation to their repertoire of arsenic associated genes (Figure 7). For these evaluations, only the 12 MAGs that meet the quality standard of completion \( \geq 90\% \) and redundancy/contamination \( \leq 10\% \) were considered. Therefore, a total of 267 KEGG modules were detected among the MAGs, presenting variable absence/presence and completion patterns (Figure 7A; Supplementary Table S5). This variation reflects the MAGs taxonomic affiliations, being all the *Bacteroidota* (MAG111, MAG12, MAG116, MAG130 and MAG193) clustered together, as well as the *Deltaproteobacteria* (MAG148, MAG143 and MAG93) and *Gammaproteobacteria* (MAG129 and MAG144). On the other hand, the only *Cyanobacteria* (MAG29) is close to the *Gammaproteobacteria* and the only *Gemmatimonadota* (MAG3) is close to the *Bacteroidota*. Particularly, the MAGs presented a considerable number of genes regarding arsenic metabolism, where the same tendency shown before was observed; with a dominant presence of genes related to As expulsion (*acr3, arsA, arsJ* and *arsR*) throughout all taxa, followed by reduction (*arsC* and *arsH*) and methylation (*arsM*) (Figure 7B, Supplementary Table S6). Interestingly, *arsJ* was only detected in one *Proteobacteria*, which agrees with its lower abundance found at the metagenomic level in the SH. While, *arsH* was only detected in the *Cyanobacteria*, which could also reflect the abundance of this taxa and this gene globally. In addition, the absence/presence patterns among the MAGs seems to be group-specific, broadly associated by phyla.
Figure 7. MAGs functional potential profiles: global and arsenic specific. A) Metabolic capabilities of the 12 selected MAGs according to KOfam (KEGG Orthologs database); the heatmap shows the KEGG Modules completion displayed by the color gradient for given module in each corresponding MAG. B) Functional network analysis of the 12 MAGs according to the presence and abundance of detected arsenic metabolism genes; size of gene nodes represents the level of detection, and the size of the edges represents the correlation level. Blue nodes represent genes, and orange nodes represent MAGs.

DISCUSSION
The Salar de Huasco has an important spatial variation of arsenic in sediments, which we have described as a gradient from north to south (Figure 1) (9); being consistent with the widely reported heterogeneity for this area and the impact of extreme conditions over the inhabitant organisms (30). Previous studies have detected variations in conductivity, organic matter, and dissolved oxygen even in geographically closed areas (6, 7, 31, 32, 33). In aquatic ecosystems, the quantity of trace elements and physical-chemical states of some elements, are related to their circulation/interaction, and with the mechanisms of corresponding living organism (34). Salar flats as SH have the peculiarity of containing more than 50% of CaCO₃ and high levels of arsenic, which is very available and highly mobile. This is because a fraction of the arsenic in sediments lies as soluble salts (Na₂HAsO₄); it also could be associated to the calcium carbonate (CaHAsO₄) and, adsorbed into Fe and Al oxides (35). The metabolism of inhabiting microbial communities or the activity of primary producers could be the source of the diversity of abiotic properties (36). It is hypothesized that this spatial and functional heterogeneity is the cornerstone of resilience in these extreme ecosystems. We must consider the presence of high concentrations of arsenic as a selective pressure, which could have a modulating effect on the composition of the H3, H4 and H5 communities (grouping them together and separating them from the rest).

The five communities described from the SH metagenomes are dominated by Proteobacteria, Bacteroidetes, and Cyanobacteria (Figure 2), similar to previous reports for other Altiplanic environments (Socompa volcanic Lake in Argentina and La Brava Lagoon - Salar de Atacama in Chile) mostly carried out by 16S rRNA amplicon sequencing (37, 38). Specifically, the same pattern has been reported previously for the SH (3, 32). However, the proportions are different, probably because the shotgun metagenomic sequencing used in this study is more sensitive, which is evidenced by the alpha diversity indices’ magnitude. Nevertheless, these values are also higher than those reported by other metagenomic studies from similar areas (39).
Moreover, there is evidence of *Proteobacteria* enrichment in places with different arsenic concentrations (12), as many microorganisms described in this group are capable of interacting and tolerate arsenic. Some examples are *Acidithiobacillus* and *Desulfovibrio* genera, which can solubilize arsenic from solid compounds or precipitate it by coupling arsenic and sulfur reduction, respectively (10, 40). The phylum *Bacteroidetes* is widespread and commonly found in hyper-saline wetlands and microbial mats (38, 41, 42, 43, 44, 45). Furthermore, *Cyanobacteria* have fundamental roles in any community they belong to, as primary producers and participating in bio-weathering and matrix transformation processes (46). Also, they are usually abundant in places exposed to sunlight (47). Besides, previous reports have proposed the cyanobacterial communities in SH are unique (6). In addition, many recurrent representatives of these communities (*Marinobacter* and *Halomonas*) have been cultured and isolated in laboratory conditions (29). Particularly, we isolated *Exiguobacterium* strains from these communities to describe the mechanisms responsible for their high arsenic resistance (9, 48, 49).

On the other hand, a high percentage of unidentified genera have also been reported in the Chilean Atacama Desert, where those reach 66% putatively novel taxa belonging to that bacterial “dark matter” (50), this is a direct consequence of database shortage, reinforcing the need to keep exploring these unique environments and describe the novel, highly adapted microorganisms that remain unknown to science and that are of fundamental importance to the maintenance of these ecosystems. The co-occurrence networks show clear evidence of overlapping among the bacterial taxonomic composition of the metagenomes or “core microbiome”, mainly composed of unknown and little-studied/described taxa. As an interaction example, we have *Seohaenicola* and *Anoxynatronum*, belonging to an isolated in the network involved in nitrogen cycling as autotrophic and heterotrophic taxa. From the co-occurrence networks, we can infer that this is a very fragile and susceptible ecosystem as there are no keystone nodes detected (51).
At the lower available taxonomic rank, we found some bacteria that are typical for extreme environments, such as *Thiobacillus*, a well-known sulfur-oxidizing bacterium, and *Desulfotignum* an anaerobic group known to be sulphate-reductors, as well as other less studied groups. The most abundant belong to the *Roseovarius* genus, an aerobic, non-photoautotroph bacterium; *Rhodohalobacter* is a facultative anaerobic, moderately halophilic group, and *Truepera* comprises aerobic quimio-organotropic alkaliphile bacterium. This is evidence of the lifestyle diversity that these microorganisms have, suggesting a cooperative and specialized community. Likewise, this is in coherence with our other findings, the enrichment of photosynthetic organisms in sites H0 and H1, as well as *Proteobacteria* as the most abundant group in all the other sampled sites. Other similarities are evidenced, as the genus *Methylibium* is only found on the H0 site, where the methane-related metabolism function was found to be enriched in the MEBS analyses. In addition, the importance of some of these abundant taxa is highlighted, as they are structural parts in the co-occurrence network, namely *Desulfomicrobium*, *Thiobacillus*, and *Brumimicrobium*.

The functional potential of the five SH communities showed similar patterns in general, with particular differences in some categories following the same association pattern described before (Figure 3). Comparing each site, we found significant differences in the enrichment of some important functional categories related to the maintenance of critical geochemical cycles, including carbon, nitrogen, and phosphate metabolism, stress resistance, transport of Zinc, Nickel, Cobalt, Iron, Manganese and osmotic stress. Notably, the two functions with statistical differences between almost all communities were the Circadian Clock in *Cyanobacteria* and the Calvin Benson cycle, suggesting differences in the primary production and CO₂ fixation. The Circadian Rhythm is directly influenced by several abiotic conditions, such as temperature and atmospheric pressure (52). Moreover, the Calvin Benson cycle was the only metabolic pathway involved in CO₂ fixation reported in an endolithic halite metagenome analysis, underlining its importance for microbial communities living in these extreme conditions (53). This agrees with the MEBS analysis.
results, where the carbon cycle relevant reactions presented a greater variation between
the metagenomes. Particularly, carbon flux control in these communities seems to be
different especially in regard to methanogenesis processes, as the far greater detection of
specific markers in H0 and H1 could account for a high presence of methanogenic species
in these two communities in comparison to the other metagenomes (54). This could be
related to the S decrease in H1, implying that methanogens outcompete sulfate reducers in
this community (55).

Furthermore, among these significant differences, arsenic resistance is enriched in
the metagenomes from the sites with the highest As concentration, which was expected
and has been reported previously (56, 57). Particularly, the distribution and abundance of
arsenic-related genes in the five SH communities showed significant differences among the
communities (Figure 4), as the presence of arsenic affects the dynamics of inhabiting
microorganisms (18, 58). Therefore, arsenic would promote the selection and abundance
of tolerant microorganisms (59, 60). Our results agree with this premise since the
abundance of all these genes increases following the gradient of arsenic concentration
among the five sites, indicating a different dynamic in each niche, even though there are
geographically closely located.

Overall, we found that most of the detected genes related to arsenic belong to
expulsion and reduction mechanisms (arsR, arsC, and acr3), which correlates with what was
reported in the Argentinian Altiplano (22). The transcriptional repressor ArsR was the one
with the highest abundance detected; as this protein participates in the regulation of
different arsenic pumps, namely ArsB, ACR3, and ArsP (61, 62); and is also reported to be in
operon conformations along with ArsC or ArsM (63). Hence, it was expected to be the most
abundant and broadly distributed. Moreover, to better assess the proportion of the arsenic
mechanisms found in SH, we separated the genes related to As expulsion out of the cell
from those related to reduction mechanism (usually reported together). This is due to the
fact that gene clusters including arsB, arsP, arsK, arsJ, and acr3 are commonly
found/reported independent of arsenate reductase \textit{arsC} \cite{9, 62, 64, 65}. Also, arsenic methyltransferase (ArsM) was the next most abundant protein, and particularly enriched in H3, H4, and H5, suggesting this protein is part of a complementary mechanism used to thrive in sites with high arsenic concentrations \cite{66}.

Even though proteins related to oxidation and respiration of As are less abundant; their distribution is interesting, which could be due to niche-specific conditions. Despite these mechanisms are somewhat known and recurrent, much remains unclear, mostly regarding their function and interconnection with the bacterial central metabolism \cite{67, 68}. This problem is exacerbated when working in unexplored environments, where unknown or unclassified organisms abound. Besides, the amount of misclassified and/or unclassified protein sequences in databases is a major blockage for these investigations. Nonetheless, this is fully acknowledged and is being addressed; thus, helping to generate important findings \cite{69, 70, 71, 72}. For example, our group was able to manually identify the ArsP and ArsK transporters, as well as a possible arsenic respiration system, additionally a missing arsenic methylase was evidenced in \textit{Exiguobacterium} genomes using a combination of bioinformatic tools \cite{9}. Nevertheless, database shortage and the classification by homology problems must always be considered when gene- or metabolism- detection is being carried out.

Broadly, the most abundant mechanism found in the SH was arsenic expulsion, with around 75\% of all the arsenic-related sequences (Figure 5). Nonetheless, by merging the expulsion and reduction mechanisms, we would get results consistent with previous reports for paddy, contaminated, and natural arsenic-rich soils (around 85\%) for the reduction mechanism \cite{56}. Likewise, here we report that mechanism proportions are relatively equivalent between the five SH sites, and those are independent of the total sequence abundance as well as the arsenic concentration. Whereas respiration and oxidation mechanisms showed some variation, which could be reflecting the taxonomic composition of the communities. Furthermore, we also found that the Organoarsenical Permease \textit{arsP}
was detected in the higher arsenic concentration sites. Therefore, this may be a counteraction against a possible greater production of these highly toxic organic species at these sites, where consistently the Arsenite Methyltransferase (arsM) gene was enriched as well. Thus, these markers could be proposed as an arsenic contamination bioindicator like acr3 (73). Additionally, the distribution and phylogenetic relationships of arsenic pumps among the metagenomes revealed distinct groups, which is consistent with those previously reported: I) ArsB - Acr3 [ion/BART], II) ArsP [Permease]; III) ArsK - ArsJ [MFS] (65).

It has been hypothesized that the divergence presented by the arsenic resistance mechanisms originated from Earth's geological changes to adapt for a particular function or as emergent species (74). Besides, we found that Proteobacteria possess a broader gene repertoire for arsenic response, which has also been described before (22).

The twelve reconstructed high-quality MAGs provide information about the abundance and distribution of undescribed or unknown microorganisms (Figure 6), as well as genomic insights into widespread arsenic metabolism. Most of these are affiliated within Proteobacteria and Bacteroidetes phyla, reflecting what we described at the community level. As stated before, the recovery of MAGs complements decades of cultivation and PCR-survey efforts by providing information about taxa missing in culture collections (Candidate Phyla Radiation), improving our understanding of microbial communities (75); particularly those inhabiting extreme environments. Nonetheless, four MAGs were classified at the genus rank: Arthrospira, a Cyanobacteria with great commercial interest as a pigment source and to make spirulina supplements; Thioalkalivibrio: an aerobic sulfur-oxidizer; Desulfatitalea and Desulfosarcina: two anaerobic sulfate reducers (76, 77, 78).

In the same way that these MAGs only cover the most abundant phyla of SH, we only detected genes that belong to the most prevalent arsenic resistance mechanisms on them (Figure 7). Thus, supporting expulsion, reduction, and methylation as the most recurrent widespread mechanisms in the SH to cope with the high arsenic concentrations. Conversely, no arsenic oxidation and/or respiration-related gene was detected on the
recovered MAGs. This could indicate that the different ways used to expel the toxic out of the cell would be more straightforward and direct as a first line of defense against arsenic; while the oxidation and respiration would involve a coupling with the bacteria central metabolism and a larger machinery to benefit from this compound, which seems to be associated with less abundant and highly adapted or specialized bacteria. Also, arsenic oxidation and respiration have been reported in association with nitrate and sulfate reduction, respectively (10, 79). Nonetheless, we need to take into consideration the MAGs completion levels as a possible source for missing genes.

Overall, our results reveal that populations of *Proteobacteria*, *Bacteroidetes* and *Cyanobacteria* are abundant across wide ecological niches in the SH spanning a challenging ensemble of environmental conditions and physico-chemical parameters, among which arsenic is highly relevant. As the arsenic cycle and the bacterial contribution to it, has yet to be completely understood, more environmental studies are needed, while metagenomic approaches have been shedding light on the possible role of unknown or undescribed microorganisms additional transcriptomic, metabolomic and cultivation approaches will be essential to define these phenomena and its relevance in a global context.

**CONCLUSION**

The Altiplano array of ecological niches is a reservoir for microbial diversity, showing great richness in adapted organisms capable of facing these challenging conditions. Particularly, Salar de Huasco is a highly diverse ecosystem, where salinity and As concentration contribute to shaping the community composition, mainly represented by *Proteobacteria* and *Bacteroidetes*. Also, the interaction networks within these communities showed three distinct groups of related taxa, but no “key stone” nodes were found. Nonetheless, little niche overlap was determined. Altogether, these indicate that SH studied niches harbor highly diverse communities, being H1 and H5 the most contrasting ones. Moreover, the most abundant arsenic-related genes found in these communities indicate
that the As(V) reduction and subsequent As(III) expulsion would be the most common strategy used to detoxify the cell of arsenic; furthermore, regarding to the expulsion pumps, the most abundant were Acr3, followed by ArsB; however, in sites with high As concentration ArsP begins to be enriched. In addition, 12 high-quality, non-redundant MAGs were reconstructed from the metagenomes; those represented the dominant diversity detected across the communities as well as the metabolic variability and the presence of marker-genomes related to the most recurrent As resistance mechanisms (expulsion, reduction and methylation). Finally, in order to further elucidate the strategies and relationships between the microbial taxa and among biotic and abiotic components within the ecosystem, further multidisciplinary studies are required, as well as the use of the ever evolving NGS approaches with the increasing database information to better understand the evolutionary process of adaptation to the extreme conditions presented by these unique ecosystems.

MATERIALS AND METHODS

Study Area and Sampling

Salar de Huasco National Park (SH) is an area located on the Chilean Altiplano that is known for its spatial heterogeneity, as well as great biodiversity and physicochemical characteristics. This ecosystem is mostly composed of streams, salt crusts, peatlands, and shallow (permanent and non-permanent) lakes with salinity and arsenic gradients from north to south (9). We sampled surface sediments (to a depth of 5 cm) in sterile tubes from five different sites (H0 to H5 previously described in (6)) by duplicate during fieldwork done in June 2018; those samples were kept and transported in a cooler until stored at -20°C for subsequent DNA extraction. Physicochemical parameters like temperature, salinity, and pH were recorded (HI 98192 and HI 2211 - HANNA Instruments) in situ (Figure 1).

DNA extraction and high throughput shotgun sequencing
Total DNA was extracted from the sediment samples from each SH site using the DNeasy PowerSoil Kit (Qiagen Inc., Hilden, Germany) following manufacturer's instructions. DNA integrity, quality, and quantity were verified through 1% agarose gel electrophoresis, OD$_{260/280}$ ratio, and fluorescence using a Qubit® 3.0 Fluorometer along with the Qubit dsDNA HS Assay Kit (Thermo Fisher Scientific, MA, USA). Next, Paired-end (150bp) libraries were constructed for each sample in duplicates, at the Centro de Biotecnología Vegetal, Universidad Andrés Bello (Santiago, Chile) using the TruSeq Nano DNA Kit (Illumina Inc., CA, USA.) following the TruSeq Nano DNA Sample Preparation Guide 15041110 Rev. D. Libraries were sent for sequencing at Macrogen Inc. (Seoul, Korea) on a HiSeq 4000 platform (Illumina Inc., CA, USA). Then, raw data was evaluated using FastQC v0.11.8 (80) for quality control, adapters were removed from the reads of all samples using Trimmomatic v0.30 (81) and the filtering and trimming (length ≤ 100bp, Ns = 0, and Q ≤ 30 thresholds) was performed with PRINSEQ v0.20.4 (82). The whole raw data sets have been deposited at DDBJ/ENA/GenBank under the Bioproject: PRJNA573913.

**Taxonomic profiling analysis**

Quality-controlled reads for each sample were profiled using the phyloFlash pipeline (83) to obtain all reads that align with the bacterial SSU rRNA (small-subunit rRNA) SILVA v138 database (84). Later, these sequences (FASTQ files) were processed using R v3.5.2 and RStudio v1.1.463 (85, 86) following the DADA2 v1.16.0 R package pipeline (87) to infer amplicon sequence variants (ASVs) present in each sample. Briefly, after dereplication, denoising, and paired reads merge steps, the ASV table was built with 97% clustering, the chimeras removed, and the taxonomic was assigned against the Silva v138 database (84) using DADA2 Ribosomal Database Project’s (RDP) naive Bayesian classifier (88). Then, data was normalized by variance stabilizing transformation using the R package DESeq2 v1.28.1 (89). Also, a multiple sequence alignment was created using the R package DECIPHER v2.16.1 (90) to infer a phylogeny with FastTree v2.1.10 (91). Furthermore, a phyloseq-object (containing the ASVs, taxonomy assignation, phylogenetic tree, and samples meta-data) was created using the R package Phyloseq v1.32.0 (92); in order to calculate the alpha
diversity indexes, along with btools v0.0.1 R package. Also, beta diversity (PCoA - Bray Curtis distance with environmental variables fit) was calculated using the R package ampvis2 v2.4.5 (93), and visualizations were generated with ggplot2 v2.2.1 (94) R package.

Co-occurrence networks

We used the same phyloseq object, which was agglomerated by best hit using the microbiomeutilities v1.00.11 R package (95); filtered by tax abundance (0.5% in at least one sample) using the Genefilter v 1.72.0 (96) and Phyloseq v1.32.0 (92) R packages. Then, the co-occurrence network was estimated using the SPIEC-EASI v0.1.4 R package (SParse InversE Covariance Estimation for Ecological Association Inference) (97), using neighborhood selection model (lambda.min.ratio=1e-2, nlambdas=20, and 50 replicates parameters). Finally, the network was visualized using the ggnetwork function of GGally v1.5.0 R package (a ggplot2 extension) (98).

Functional profiling analysis

The patterns of functional potential as subsystems with different specificity levels for each community were determined by the presence/absence and relative abundance of the quality-controlled reads that aligned against the metabolic pathways and functions in the SEED database (99) with SUPER-FOCUS v0.35 software (100), which uses DIAMOND v2.0.6 (101) for fast and efficient alignment. Statistical analysis was carried out in STAMP v2.3.1 (Software package for Analyzing Taxonomic or Metabolic Profiles) (102) using Welch’s t-test to compare all samples.

Metagenome co-assembly and read mapping

As we need for subsequent analyses: i) a co-assembly (the reads from all samples assembled together: .fasta file), ii) its annotations (.gff, .ffn and .faa files) and iii) an alignment per samples (all the reads from each sample mapped against the co-assembly: .bam files); we proceeded to assemble the quality-controlled reads from all samples using MEGAHIT v1.1.3 (103), with the –presets meta-large option and a minimum contig length
of 1 kb. The co-assembly was then evaluated with MetaQUAST v5.0.2 (104) and annotated with Prokka v1.11 (105) using the metagenome mode. Furthermore, we mapped the quality-controlled reads from each sample against the co-assembly using Bowtie2 v2.3.4 (106) and stored the recruited reads (sorted and indexed) as BAM files using SAMtools v1.3 (107).

**Metagenome profiling**

We followed the anvi’o v7 pipeline (108). First, we created a contig database with the co-assembled contigs, which uses Prodigal v2.6.3 (109), HMMER v3.3.1 (110) and NCBI COGs (111) to identify genes calls and functionally annotate them (anvi-gen-contigs-database, anvi-run-hmms, anvi-run-ncbi-cogs). Secondly, we profiled each sample’s BAM file against the contigs database to estimate the detection and coverage statistics for each contig (anvi-profile). Then, we combined the profiles in a single merged metagenomic profile database, which uses all individual statistics to compute hierarchical clustering of the contigs (anvi-merge). Finally, we visualized the merged profile on the anvi’o interactive interface, which allows easy exploration and curation of the metagenomes (anvi-interactive). On the other hand, we used MEBS (Multigenomic Entropy Based Score) v1.0 package (112) to evaluate and compare the S, N, O, CH₄, and Fe biogeochemical cycles on each metagenome using the predicted proteins (annotated .faa file from each individual assembly).

**Metagenome target gene search**

To estimate the abundance of genes related to arsenic metabolism, the read counts for each predicted gene of each sample were obtained using the corresponding BAM file and the co-assembly annotated GFF file, with HTSeq-Counts v0.13.5 (113). Also, we constructed a database from the GenBank – “Identical Protein Groups” with all available prokaryotic proteins of arsenic metabolism (Acr3, AioA, AioB, AioR, AioS, AioX, AoxA, AoxB, AoxC, AoxD, ArrA, ArrB, ArrC, ArsA, ArsB, ArsC, ArsD, ArsH, ArsJ, ArsK, ArsM, ArsN, ArsO, ArsP, ArsR, ArsT, ArxA, ArxB, ArxR, ArxS and ArxX) and this was targeted with the co-
assembly predicted proteins (.faa file) using CRB-BLAST v0.6.6 (E-value ≥1E⁻⁰⁵, Identity ≥70% and Query Coverage ≥70%) (114). Finally, matching the hits with the gene counts (normalized by the target gene length and the corresponding library size), we calculated the relative abundance of the interest genes for the five metagenomes. Visualizations were made with R packages ggplot2 v2.2.1 (94) and pheatmap v1.0.12 (115). Furthermore, all the detected sequences that corresponded to arsenic efflux pumps were extracted from the .ffn file and aligned using MAFFT v7 (116). As phylogeny inference was calculated with FastTree v2.1.10 (91) and visualized with the anvi’o v7 interactive interface (108).

**Metagenomic binning**

The contigs in the metagenome profile were clustered through anvi’o v7 (108) using CONCOCT v1.1.0 (117) binning program, which adds a bins collection the profile (anvi-cluster-contigs). The resulting bins were evaluated for completion/redundancy, and manual curation/refinement was carried out in the interactive interface (anvi-estimate-genome-completeness and anvi-refine). Finally, the refined bins were displayed in the interactive interface and summarized to obtain all the statistics and files for downstream analysis (anvi-interactive and anvi-summarize).

**MAGs evaluation, taxonomy, and functional estimation**

We define MAGs (Metagenome Assembles Genomes) as bins with a completion >80% and redundancy <10%. Then, we used CheckM v1.0.13 (118) for a more robust evaluation. Later, to infer MAGs taxonomy, we used GTDB-Tk v0.3.2 (119) along with The Genome Taxonomy Database (120). Moreover, we estimated the MAGs metabolic potential by evaluating their gene content with anvi’o v7 (108); First, functions and metabolic pathways were annotated to the MAGs using HMM hits from KOfiles - KEGG Orthologs (KO) database (121, 122) (anvi-run-kegg-kofams). Secondly, relying on these KO annotations, the metabolic pathways were predicted considering those defined by KOs in the KEGG MODULES resource (123), where a KO represents a gene function, and a module represent a group of KOs that together carry out the reactions in a metabolic path (anvi-estimate-
metabolism). The MAGs module completion was visualized using pheatmap v1.0.12 R package (115).

**MAGs target gene search**

All MAGs were queried against the same previously constructed database (with the arsenic metabolism genes) using CRB-BLAST v0.6.6 (E-value ≥1E⁻⁰⁵, Identity ≥70% and Query Coverage ≥70%) (114). Furthermore, the resulting hits matrix was compared and supplemented with the detected KOfams related to arsenic, and the final matrix was used to generate a functional network using Gephi v0.9.2 (124) to connect MAGs and detected genes.

**DATA AVAILABILITY**

The whole raw data sets have been deposited at DDBJ/ENA/GenBank under the Bioproject: PRJNA573913.

**AUTHOR CONTRIBUTIONS**

JC-S, CP-E, EC-N, FR, and CPS conceived and designed the study. JC-S, JF, FM, FR, and CPS performed the field work. JC-S and CP-E processed the samples, performed the experimental procedures and carried out the bioinformatics analyzes. KM, SM and EC-N contributed with critical bioinformatics advice. CPS, FR, and EC-N contributed with reagents, materials, and analysis tools. JC-S and CP-E interpreted the results and wrote the first manuscript draft. All authors read and approved the final manuscript.

**ACKNOWLEDGEMENTS**

We would like to thank the illustrator Florence Gutzwiller for the SH landscape painting (https://spideryscrawl-illustration.webnode.com/). Also, we would like to thank
Universidad Andres Bello’s high-performance computing cluster, Dylan (http://www.castrolab.org/), for providing data storage, support, and computing power for bioinformatic analyses. Also, We would also like to thank the MerenLab group (https://merenlab.org/) for their reachability in providing help, advice and solutions related to the use of Anvi’o. In addition, we’d like to thank to Dra. Valerie de Anda (https://valdeanda.github.io/) for the guidance and advice to implement the MEBS software.

FUNDING
This research was sponsored by ANID (Agencia Nacional de Investigación y Desarrollo de Chile) grants. CPS was funded by ANID-FONDECYT regular 1210633 and ECOS-ANID 170023. EC-N was funded by ANID-FONDECYT Regular 1200834 and ANID-PIA-Anillo INACH ACT192057. JC-S was founded by ANID 2021 Post-Doctoral FONDECYT 3210156. CP-E was founded by Universidad Católica del Norte 2021 Post-Doctoral fellowship. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

CONFLICT OF INTEREST STATEMENT
The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

REFERENCES


90. Wright ES. 2016. Using DECIPHER v2.0 to analyze big biological sequence data in R. R Journal 8(1).


112. De Anda V, Zapata-Peñasco I, Poot-Hernandez AC, Eguiarte LE, Contreras-Moreira B, Souza V. 2017. MEBS, a software platform to evaluate large (meta) genomic collections according to their metabolic machinery: unraveling the sulfur cycle. GigaScience 6(11), gix096.


SUPPLEMENTARY MATERIAL

Supplementary Table S1. Relative abundance of all detected phyla in the five SH communities.

Supplementary Table S2. Statistical values of the SH co-assembly, representing the five metagenomes.

Supplementary Table S3. Statistical evaluation of enriched functional categories (SEED subsystem 1) in each of the five communities regarding the rest according to Welch’s t-test.

Supplementary Table S4. Indexes of relative abundance and detection of each recovered MAG across the five SH metagenomes.

Supplementary Table S5. Completion index of all detected KEGG modules each analyzed MAG.

Supplementary Table S6. Presence and copy number of arsenic resistance related genes in each analyzed MAG.
**Most abundant taxa**

<table>
<thead>
<tr>
<th>Taxa</th>
<th>H0</th>
<th>H1</th>
<th>H3</th>
<th>H4</th>
<th>H5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proteobacteria; G: Roseovarius-</td>
<td>0</td>
<td>0</td>
<td>3.2</td>
<td>10.4</td>
<td>7.8</td>
</tr>
<tr>
<td>Proteobacteria; G: Desulfitotignum-</td>
<td>0.2</td>
<td>0</td>
<td>4.3</td>
<td>11.3</td>
<td>5.3</td>
</tr>
<tr>
<td>Bacteroidetes; G: Rhodohalobacter-</td>
<td>0</td>
<td>0</td>
<td>4.3</td>
<td>2.3</td>
<td>10.4</td>
</tr>
<tr>
<td>Deinococcus–Thermus; G: Truepera-</td>
<td>2</td>
<td>0.1</td>
<td>5</td>
<td>5.1</td>
<td>2.8</td>
</tr>
<tr>
<td>Proteobacteria; G: Thiobacillus-</td>
<td>0.1</td>
<td>11.3</td>
<td>1.3</td>
<td>0</td>
<td>0.1</td>
</tr>
<tr>
<td>Proteobacteria; G: Marinobacter-</td>
<td>0</td>
<td>0</td>
<td>4.4</td>
<td>0.7</td>
<td>3</td>
</tr>
<tr>
<td>Chloroflexi-</td>
<td>0</td>
<td>0</td>
<td>2.8</td>
<td>1.3</td>
<td>2.2</td>
</tr>
<tr>
<td>Bacteroidetes; G: Psychrophilus-</td>
<td>0</td>
<td>0</td>
<td>0.9</td>
<td>0.7</td>
<td>2.8</td>
</tr>
<tr>
<td>Proteobacteria; G: Halomonas-</td>
<td>0.2</td>
<td>0</td>
<td>1</td>
<td>0.5</td>
<td>2.1</td>
</tr>
<tr>
<td>Bacteroidetes; G: Brumibacterium-</td>
<td>0</td>
<td>0</td>
<td>0.1</td>
<td>2</td>
<td>0.8</td>
</tr>
<tr>
<td>Verrucomicrobia; G: Luteolibacter-</td>
<td>0.8</td>
<td>0.1</td>
<td>0.6</td>
<td>0.5</td>
<td>0.2</td>
</tr>
<tr>
<td>Euryarchaeota; G: Halorubrum-</td>
<td>0</td>
<td>0</td>
<td>0.6</td>
<td>1.1</td>
<td>0.4</td>
</tr>
<tr>
<td>Proteobacteria; G: Desulfomicrobium-</td>
<td>2.1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.1</td>
</tr>
<tr>
<td>Cyanobacteria; O: Chloroplast-ASV6012-</td>
<td>1.1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Proteobacteria; G: Desulfobacter-</td>
<td>0.6</td>
<td>0</td>
<td>0.2</td>
<td>0.6</td>
<td>0.7</td>
</tr>
<tr>
<td>Cyanobacteria; O: Chloroplast-ASV339-</td>
<td>1.8</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Proteobacteria; G: Methylibium-</td>
<td>1.8</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Cyanobacteria; G: Arthrospira-</td>
<td>1.8</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Proteobacteria; G: Defluviicoccus-</td>
<td>0</td>
<td>0.6</td>
<td>1</td>
<td>0</td>
<td>0.1</td>
</tr>
<tr>
<td>Proteobacteria; G: Thiocapsa-</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0.6</td>
<td>0</td>
</tr>
<tr>
<td>Gemmatimonadetes; G: Gemmatimonas-</td>
<td>0.6</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Proteobacteria; G: Shimia-</td>
<td>0</td>
<td>0</td>
<td>0.6</td>
<td>0.6</td>
<td>0.4</td>
</tr>
<tr>
<td>Proteobacteria; G: Hydrogenophaga-</td>
<td>1.2</td>
<td>0.4</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Proteobacteria; C: Gammaproteobacteria-</td>
<td>1.6</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Actinobacteria; G: Ilumatobacter-</td>
<td>0.2</td>
<td>0.7</td>
<td>0.3</td>
<td>0.1</td>
<td>0.2</td>
</tr>
<tr>
<td>Bacteroidetes; F: Bacteroidetes-</td>
<td>0</td>
<td>1.5</td>
<td>0</td>
<td>0.1</td>
<td>0.3</td>
</tr>
<tr>
<td>Proteobacteria; G: Thioalkalivibrio-</td>
<td>0</td>
<td>0</td>
<td>0.4</td>
<td>0.8</td>
<td>0.3</td>
</tr>
<tr>
<td>Cyanobacteria; O: Chloroplus-ASV6009-</td>
<td>1.4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Proteobacteria; F: Burkholderiaceae-</td>
<td>0</td>
<td>1.3</td>
<td>0</td>
<td>0.1</td>
<td>0.1</td>
</tr>
<tr>
<td>Bacteroidetes; G: Flavobacterium-</td>
<td>0.9</td>
<td>0</td>
<td>0.2</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

**Supplementary Figure S1.** Composition and structure of SH bacterial communities. Heatmap showing the 30 most abundant ASVs (bacterial lineages). Taxonomic is showed at the phylum rank plus the best hit available (C: class; O: order; F: family; G: genus).
Supplementary Figure S2. Alpha diversity indices for the five SH communities.
Supplementary Figure S3. MEBS analysis heatmap displaying the completeness of N, Fe, S and CH₄ pathways: as a whole (first top section) and particular reactions (bottom four sections). The color gradient shows the percentage of completion for each pathway (from lowest to highest) and the values at the top section represent the corresponding MEBS score (* FDR ≤ 0.01).
<table>
<thead>
<tr>
<th>Location</th>
<th>As</th>
<th>Salinity</th>
<th>pH</th>
<th>Temp</th>
<th>Altitude</th>
</tr>
</thead>
<tbody>
<tr>
<td>H0</td>
<td>9 mg/kg</td>
<td>11.9%</td>
<td>8.8</td>
<td>14.6ºC</td>
<td>3,792 masl</td>
</tr>
<tr>
<td>H1</td>
<td>321 mg/kg</td>
<td>84.5%</td>
<td>9.2</td>
<td>24.3ºC</td>
<td>3,781 masl</td>
</tr>
<tr>
<td>H3</td>
<td>155 mg/kg</td>
<td>77.2%</td>
<td>8.4</td>
<td>18.8ºC</td>
<td>3,787 masl</td>
</tr>
<tr>
<td>H4</td>
<td>49.2 mg/kg</td>
<td>2.2%</td>
<td>8.5</td>
<td>19.9ºC</td>
<td>3,783 masl</td>
</tr>
<tr>
<td>H5</td>
<td>16.3 mg/kg</td>
<td>8.1%</td>
<td>9.4</td>
<td>14ºC</td>
<td>3,785 masl</td>
</tr>
</tbody>
</table>

Salar de Huasco is located in the Atacama Desert, Chile.
# Arsenic Metabolism Genes

<table>
<thead>
<tr>
<th>Gene</th>
<th>H0</th>
<th>H1</th>
<th>H3</th>
<th>H4</th>
<th>H5</th>
</tr>
</thead>
<tbody>
<tr>
<td>arsR</td>
<td>6142.15</td>
<td>4214.04</td>
<td>12356.33</td>
<td>18499.32</td>
<td>13750.63</td>
</tr>
<tr>
<td>arsC</td>
<td>1327.42</td>
<td>1188.22</td>
<td>2117.04</td>
<td>2345.06</td>
<td>2057.75</td>
</tr>
<tr>
<td>acr3</td>
<td>1099.45</td>
<td>698.09</td>
<td>2375.33</td>
<td>2003.96</td>
<td>2183.94</td>
</tr>
<tr>
<td>arsM</td>
<td>427.13</td>
<td>495.38</td>
<td>1491.54</td>
<td>2215.36</td>
<td>1827.64</td>
</tr>
<tr>
<td>arsB</td>
<td>367.57</td>
<td>200.46</td>
<td>583.55</td>
<td>542.77</td>
<td>487.89</td>
</tr>
<tr>
<td>arsD</td>
<td>259.69</td>
<td>57.08</td>
<td>534.59</td>
<td>682.07</td>
<td>600.08</td>
</tr>
<tr>
<td>arsJ</td>
<td>202.04</td>
<td>75.23</td>
<td>531.96</td>
<td>717.95</td>
<td>480.64</td>
</tr>
<tr>
<td>arsA</td>
<td>601.44</td>
<td>93.67</td>
<td>337.78</td>
<td>438.65</td>
<td>338.12</td>
</tr>
<tr>
<td>aoxB</td>
<td>113.32</td>
<td>0.89</td>
<td>519.98</td>
<td>613.91</td>
<td>485.76</td>
</tr>
<tr>
<td>arsP</td>
<td>2.14</td>
<td>0.60</td>
<td>178.54</td>
<td>564.49</td>
<td>341.63</td>
</tr>
<tr>
<td>arsH</td>
<td>265.30</td>
<td>92.76</td>
<td>217.13</td>
<td>253.17</td>
<td>237.18</td>
</tr>
<tr>
<td>aioB</td>
<td>46.54</td>
<td>50.63</td>
<td>153.38</td>
<td>171.91</td>
<td>134.68</td>
</tr>
<tr>
<td>arxB</td>
<td>65.25</td>
<td>2.50</td>
<td>205.48</td>
<td>105.97</td>
<td>147.89</td>
</tr>
<tr>
<td>aoxA</td>
<td>0.00</td>
<td>0.00</td>
<td>156.08</td>
<td>200.40</td>
<td>165.10</td>
</tr>
<tr>
<td>aioA</td>
<td>110.20</td>
<td>86.16</td>
<td>106.81</td>
<td>114.45</td>
<td>84.01</td>
</tr>
<tr>
<td>arxA</td>
<td>177.41</td>
<td>2.07</td>
<td>158.71</td>
<td>35.78</td>
<td>92.44</td>
</tr>
<tr>
<td>arxR</td>
<td>240.74</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>arrA</td>
<td>8.58</td>
<td>60.09</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
</tbody>
</table>

**Mechanism**

- **Expulsion**
- **Reduction**
- **Respiration**
- **Oxidation**
- **Methylation**

**Normalized reads**

- 20,000
- 15,000
- 10,000
- 5,000
- 1,000
- 500
- 100
- 0
<table>
<thead>
<tr>
<th>MAG</th>
<th>Percent Completion</th>
<th>Percent Reduction</th>
<th>Num Contigs</th>
<th>GC content</th>
<th>Percent Completion</th>
<th>Percent Contamination</th>
<th>Phylum</th>
<th>Class</th>
<th>Order</th>
<th>Family</th>
<th>Genus</th>
</tr>
</thead>
<tbody>
<tr>
<td>SH-MAG111</td>
<td>95.77%</td>
<td>2.82%</td>
<td>332</td>
<td>11,127</td>
<td>60.76%</td>
<td>81.95%</td>
<td>Bacteroidota</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>SH-MAG116</td>
<td>90.14%</td>
<td>0.00%</td>
<td>208</td>
<td>14,823</td>
<td>43.49%</td>
<td>83.04%</td>
<td>Bacteroidota</td>
<td>Bacteroidia</td>
<td>Chitinophagales</td>
<td>Sapropiraceae</td>
<td>---</td>
</tr>
<tr>
<td>SH-MAG12</td>
<td>95.77%</td>
<td>2.82%</td>
<td>413</td>
<td>11,471</td>
<td>54.64%</td>
<td>92.06%</td>
<td>Bacteroidota</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>SH-MAG130</td>
<td>92.96%</td>
<td>9.86%</td>
<td>395</td>
<td>10,529</td>
<td>44.80%</td>
<td>93.50%</td>
<td>Bacteroidota</td>
<td>Bacteroidia</td>
<td>Bacteroidales</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>SH-MAG192</td>
<td>85.92%</td>
<td>4.23%</td>
<td>314</td>
<td>4,317</td>
<td>36.06%</td>
<td>68.88%</td>
<td>Bacteroidota</td>
<td>Ignavibacteriales</td>
<td>Ignavibacteriales</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>SH-MAG193</td>
<td>85.92%</td>
<td>4.23%</td>
<td>561</td>
<td>10,013</td>
<td>44.34%</td>
<td>92.35%</td>
<td>Bacteroidota</td>
<td>Bacteroidia</td>
<td>Cytophagales</td>
<td>Cyclobacteriales</td>
<td>---</td>
</tr>
<tr>
<td>SH-MAG96</td>
<td>85.92%</td>
<td>4.23%</td>
<td>334</td>
<td>14,552</td>
<td>47.74%</td>
<td>80.17%</td>
<td>Bacteroidota</td>
<td>Bacteroidia</td>
<td>Chitinophagales</td>
<td>Sapropiraceae</td>
<td>---</td>
</tr>
<tr>
<td>SH-MAG29</td>
<td>95.77%</td>
<td>5.63%</td>
<td>346</td>
<td>20,718</td>
<td>46.60%</td>
<td>97.71%</td>
<td>Cyanobacteria</td>
<td>Cyanobacteriales</td>
<td>---</td>
<td>---</td>
<td>Arthospira</td>
</tr>
<tr>
<td>SH-MAG3</td>
<td>94.37%</td>
<td>2.82%</td>
<td>436</td>
<td>10,739</td>
<td>66.26%</td>
<td>86.11%</td>
<td>Gemmatimonadota</td>
<td>Gemmatimonadetes</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>SH-MAG89</td>
<td>88.73%</td>
<td>9.86%</td>
<td>498</td>
<td>14,477</td>
<td>62.15%</td>
<td>82.81%</td>
<td>Proteobacteria</td>
<td>Deltaproteobacteria</td>
<td>Bradymonadales</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>SH-MAG131</td>
<td>80.28%</td>
<td>1.41%</td>
<td>626</td>
<td>5,244</td>
<td>63.23%</td>
<td>77.20%</td>
<td>Proteobacteria</td>
<td>Delta-proteobacteria</td>
<td>Desulfobacteriales</td>
<td>Desulfosarcinae</td>
<td>---</td>
</tr>
<tr>
<td>SH-MAG143</td>
<td>95.77%</td>
<td>8.45%</td>
<td>601</td>
<td>7,367</td>
<td>58.76%</td>
<td>84.68%</td>
<td>Proteobacteria</td>
<td>Deltaproteobacteria</td>
<td>Desulfobacteriales</td>
<td>Desulfosarcinae</td>
<td>Desulfatelia</td>
</tr>
<tr>
<td>SH-MAG148</td>
<td>95.77%</td>
<td>8.45%</td>
<td>384</td>
<td>7,777</td>
<td>55.17%</td>
<td>87.42%</td>
<td>Proteobacteria</td>
<td>Deltaproteobacteria</td>
<td>Desulfobacteriales</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>SH-MAG93</td>
<td>97.18%</td>
<td>5.63%</td>
<td>550</td>
<td>9,043</td>
<td>56.28%</td>
<td>94.35%</td>
<td>Proteobacteria</td>
<td>Deltaproteobacteria</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>SH-MAG129</td>
<td>92.96%</td>
<td>2.82%</td>
<td>390</td>
<td>14,010</td>
<td>58.58%</td>
<td>91.11%</td>
<td>Proteobacteria</td>
<td>Gammaproteobacteria</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>SH-MAG144</td>
<td>95.77%</td>
<td>1.41%</td>
<td>315</td>
<td>8,476</td>
<td>67.34%</td>
<td>87.51%</td>
<td>Proteobacteria</td>
<td>Gammaproteobacteria</td>
<td>Etotihoeidospirales</td>
<td>Thioalkalivirionate</td>
<td>Thioalkalivibrio</td>
</tr>
<tr>
<td>SH-MAG162</td>
<td>87.32%</td>
<td>1.41%</td>
<td>431</td>
<td>8,147</td>
<td>69.40%</td>
<td>81.81%</td>
<td>Proteobacteria</td>
<td>Gammaproteobacteria</td>
<td>Pseudomonadales</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>SH-MAG78</td>
<td>80.28%</td>
<td>5.63%</td>
<td>371</td>
<td>6,424</td>
<td>67.34%</td>
<td>73.01%</td>
<td>Proteobacteria</td>
<td>Gammaproteobacteria</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>SH-MAG169</td>
<td>85.54%</td>
<td>9.64%</td>
<td>4176</td>
<td>8,047</td>
<td>42.02%</td>
<td>72.44%</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
</tbody>
</table>

**GC-content, Length, Completion, Redundancy, H0, H1, H2, H3, H4, H5, H6**

**Cluster by MAGs abundance**

**Total Mapped Reads**

---

*Note: The data represents the abundance and distribution of MAGs (Metagenome Assembled Genomes) across different phyla and classes.*