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, J[1,±]; Pardo-Esté, C[1,2,±]; Mendez, KN[3]; Fortt, J[1]; Marquez, S[3]; Molina, F[4]; Castro-Nallar, E[3]; Remonsellez, F[1,5 *] and Saavedra, CP[2 *].

[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.

ABSTRACT

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 CH_4 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.

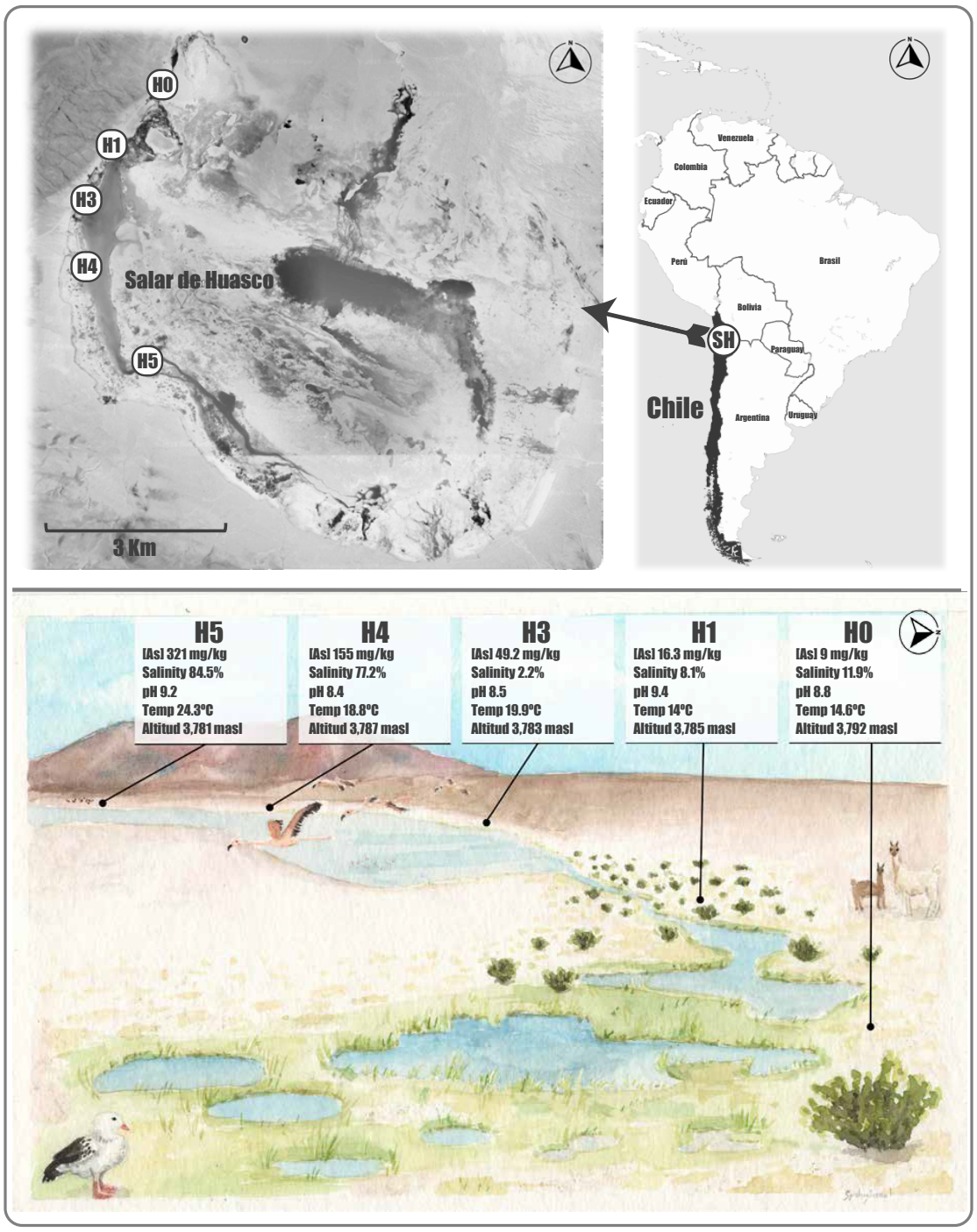


Figure 1. Salar de Huasco study area. Top panel: a map showing the five sampling sites (H0: 20º15'48.8"S - 68º52'28.4"W; H1: 20º16'27.7"S - 68º53'3"W; H3: 20º16'59.2"S - 68º53'16.7"W; H4: 20º17'40.9"S - 68º53'17.3"W and H5: 20º18'37"S - 68º52'42"W) investigated in this study. SH is located between the 68º47' - 68º54' W and 20º15' - 20º20' 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.

144

145 **The bacterial communities of SH are highly heterogeneous and rich in unknown taxa**

146 The contrasting differences presented among the SH sampling sites are reflected by
147 the taxonomic composition of bacterial communities, which showed contrasting patterns
148 and variation between sites (Figure 2). Overall, our results indicate that the *Proteobacteria*
149 and *Bacteroidetes* are the most prevalent phyla in the sampled sediments, which are
150 particularly enriched in the H3, H4 and H5 communities, accounting together for >60% of
151 all observed taxa (Figure 2A). In turn, these phyla represent ~50% in the H0 and H1
152 communities. Interestingly, the H0 site is dominated by the *Cyanobacteria* phylum with a
153 34% of the total community; also, this community has the highest abundance of *Firmicutes,*
154 *Patescibacteria* and *Spirochaetes* phyla as well as significantly lower proportion of
155 *Actinobacteria.* Moreover, H1 community profile presents the highest abundance of
156 *Chloroflexi, Actinobacteria, Verrucomicrobia, Planctomycetes* and *Acidobacteria* phyla. On
157 the other hand, H3, H4 and H5 communities are more similar to each other, despite the vast
158 differences in low-abundance taxa; many of which are exclusively present in only one
159 community (Supplementary Table S1).

160

161 The most abundant bacteria at the lowest (available) taxonomic rank reflect the
162 same pattern, with a great abundance of *Proteobacteria* genera *Roseovarius* and
163 *Desulfotignum* in the H3, H4 and H5 communities (Supplementary Figure S1). Moreover,
164 *Halomonas, Thiobacillus, Luteolibacter* and *Truepera* genera are more widespread, while
165 *Rhodohalobacter, Marinobacter, Psychroflexus* and *Brumimicrobium* are also concentrated
166 in H3, H4 and H5. Contrary, H0 and H1 are enriched in *Cyanobacteria* like *Arthrospira* and
167 three members of the *Chloroplast* order, as well as genera with various metabolism types
168 such as *Methylibium, Hydrogenophaga* and *Desulfomicrobium.* In addition, it is worth
169 mentioning that from the 3,801 bacterial ASVs detected, 6.5% and 58.3% could not be
170 classified within any known phylum nor genus respectively. Particularly, *Halomonas* and
171 *Marinobacter* are culturable bacteria and seem to be recurrent in the SH, as we reported
172 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). It is

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.

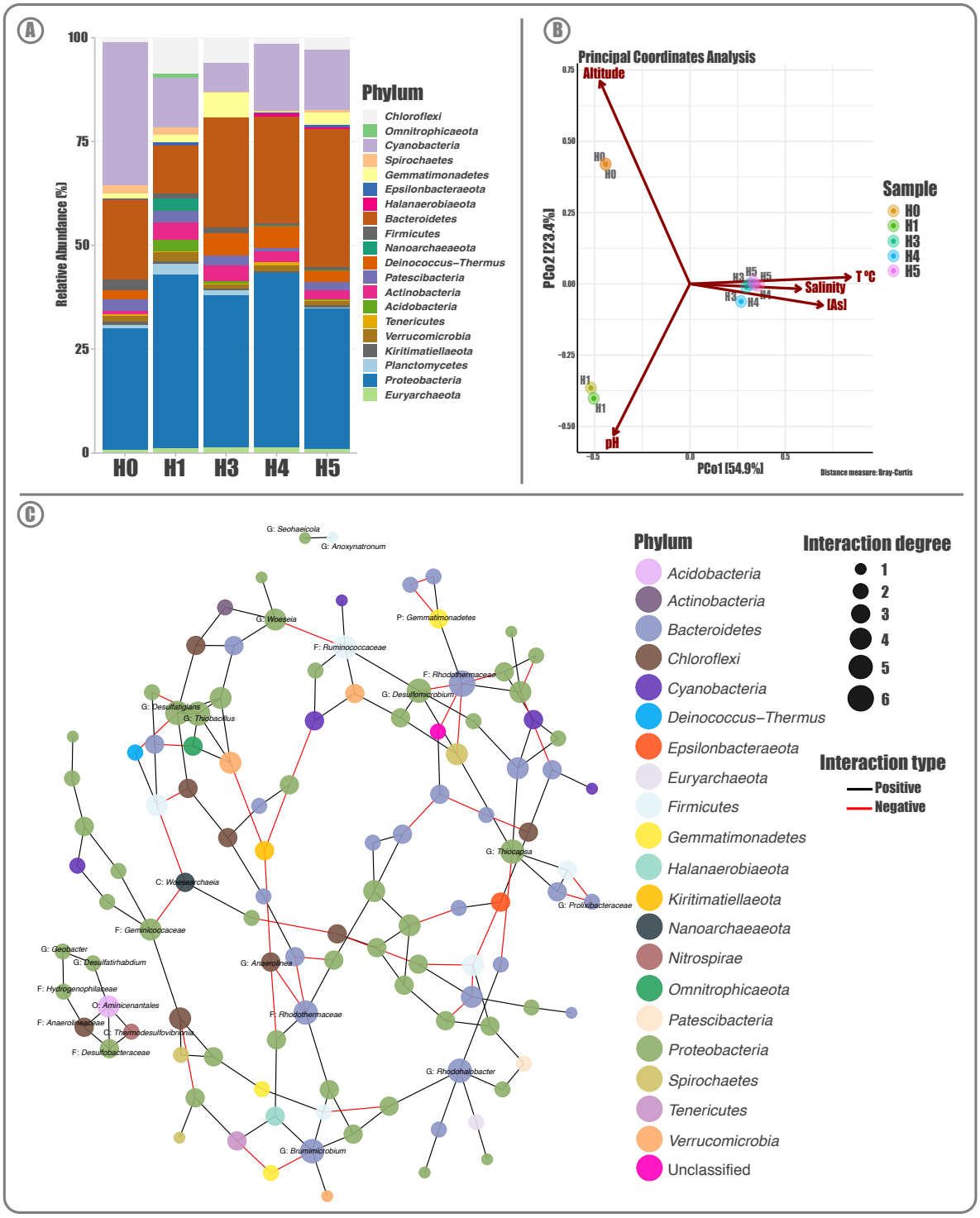


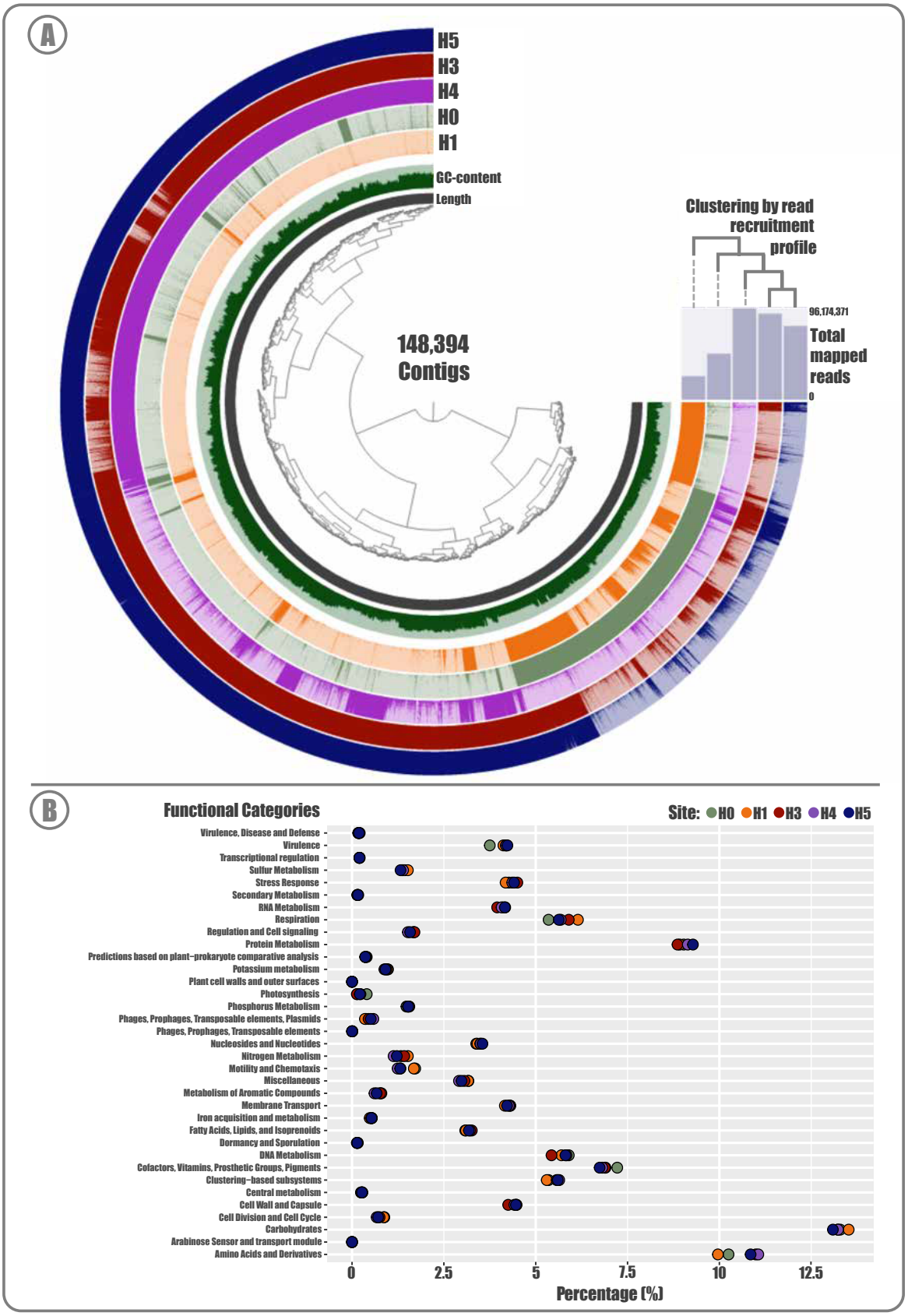
Figure 2. Salar de Huasco bacterial communities. A) Taxonomic composition and relative abundance of SH microbial communities in the five studied sites; Stacked-bars show the top20 most abundant bacteria at the phylum taxonomic level. **B)** Beta diversity by Principal Component Analysis (PCoA) on Hellinger 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

214 samples. The arrows indicate the explanatory power of the statistically significant
215 environmental parameters with regards to the observed variation in community
216 composition. **C)** Co-occurrence network analysis of the SH bacterial communities, the size
217 of each node (representing ASVs) is proportional to the number of connections (degrees),
218 the color of the edges connecting nodes represent the interaction type, the node color
219 indicates the taxonomic affiliation at phylum level and node labels are at the lowest
220 available taxonomic classification.
221

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

223 The five SH metagenomes we analyzed represent the surface sediment of 5 different
224 sites with a great variation in arsenic content (from 9 to 321 mg/Kg). The 337.5 million
225 quality-controlled reads were co-assembled and yielded a total of 994,545 contigs (≥ 1,000
226 bp) with 2.38 millions genes, which were used to generate abundance/distributions and
227 functional profiles of the 5 studied communities (Figure 3; Supplementary Table S2). Of
228 those contigs, 148,394 (≥ 2,500 bp) were hierarchically clustered to then be profiled by
229 read-recruitment of the data from the five communities (Figure 3A). The resulting patters
230 are highly variable, considering the contigs detection in each metagenome. In particular,
231 metagenomes from H1 and H5 are the most contrasting ones, as they respectively have
232 18.78% and 78.40% presence of all available contigs in the SH. Moreover, the rest of the
233 samples also have diverse percentages of representation (H0: 25.00%; H3: 72.08% and H4:
234 48.33%), and their clustering by the read recruitment profiles correlates with the taxonomic
235 profiles presented previously; being H3, H4 and H5 more similar to each other, and H0 and
236 H1 more distanced from each other and to the others as well. Contrary, the functional
237 profiles seem to have a higher level of convergence between samples in a broad way (Figure
238 3B). Furthermore, the more enriched categories (SEED Subsystems 1) are associated with
239 metabolism (amino acid derivatives, carbohydrates, protein metabolism, DNA metabolism,
240 cofactor, vitamins, prosthetic groups and pigments) which could be evidence of the needed
241 versatility and adaptability for bacteria to thrive in these harsh environments. Also, the
242 stress response, membrane transporter, cell wall and capsule categories, which are related
243 to the ability to thriving capabilities as well.
244

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.



262

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.

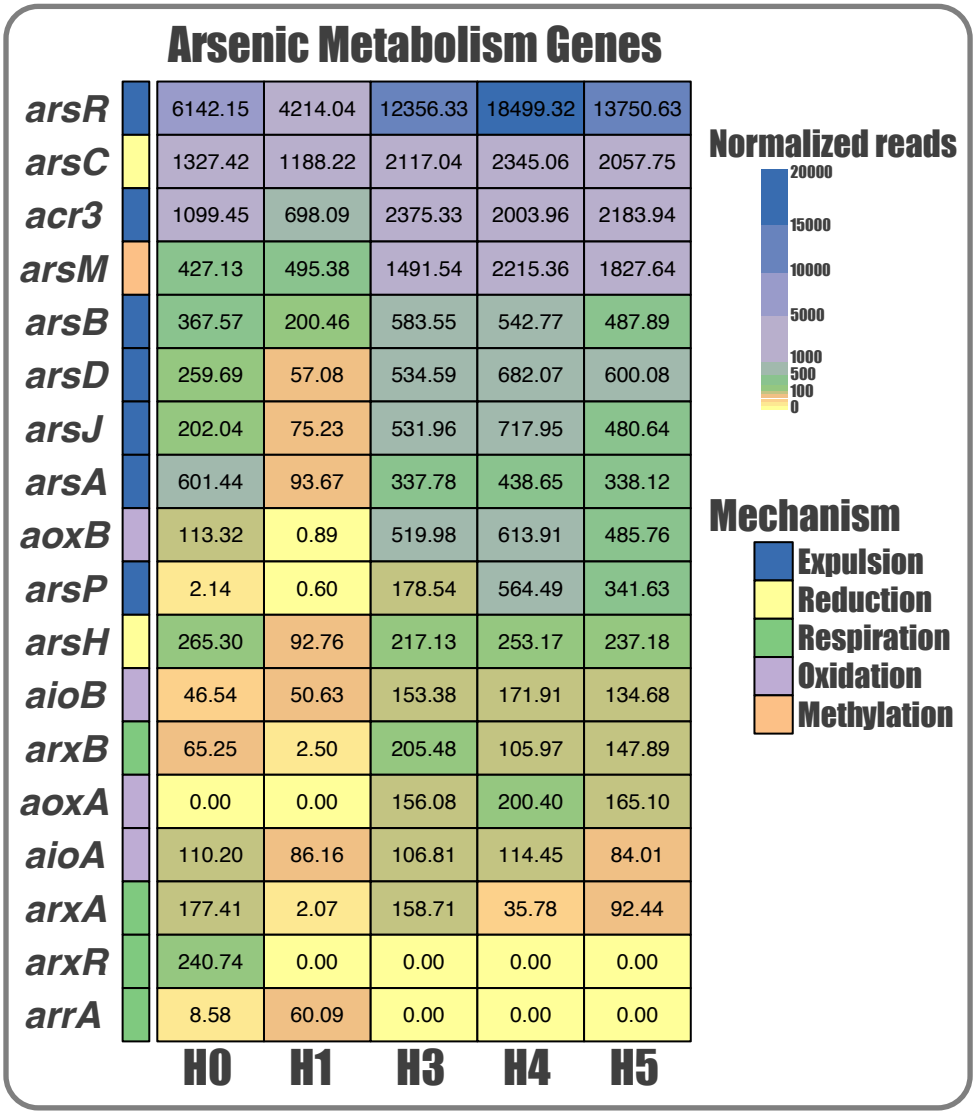
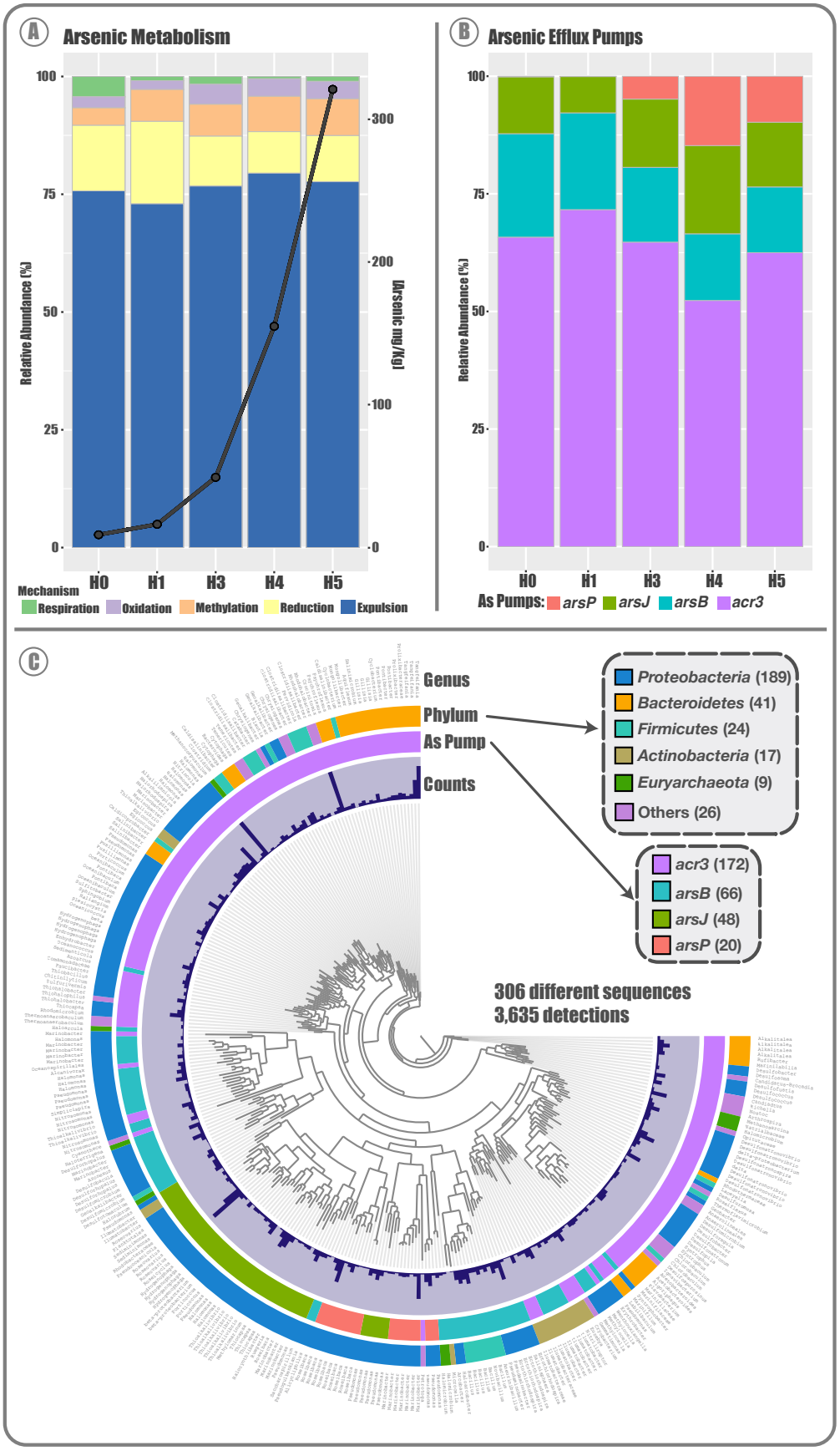


Figure 4. Distribution and abundance of As metabolism genes in the Salar de Huasco. Heatmap shows the (normalized) number of reads that aligned against the corresponding protein identified in each sample, according to the color scale. Genes are grouped by colors, representing the 5 As response mechanisms.

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

303 seems to be independent of the amount of detected sequences and the arsenic
304 concentration of the site. Furthermore, comparing the abundance of particular arsenic
305 efflux pumps (from the expulsion mechanism) we can observe that, indeed, most of the
306 sequences correspond to *acr3* (Figure 5B); followed by *arsB* and *arsJ*. Notably, *arsP* which is
307 an efflux permease that confers resistance to organic arsenics (roxarsone and
308 methylarsenite) was more abundant in the sites with higher arsenic concentration.
309 Moreover, the higher *acr3* variants and abundance could be due this pump is present in a
310 wider number of bacteria phyla, covering most of the found diversity (Figure 5C). Also, this
311 efflux pump seems to be more ancestral, regarding *arsP* and *arsJ* which are only present in
312 *Proteobacteria* and in more recent branches.
313



315 **Figure 5**. **Arsenic metabolism in the Salar de Huasco. A)** Distribution and abundance of
316 detected genes related to arsenic (As) metabolism in the five SH metagenomes; Stacked
317 bars represent the proportion of all the genes grouped by mechanism type based on relative
318 abundance (%), and the line represents total As concentration in each corresponding site
319 (mg/Kg of sediment). **B)** Stacked bars represent the proportion of all detected As efflux
320 pumps in the five metagenomes, based on genes relative abundance (%). **C)** Phylogenetic
321 analysis of 306 non-redundant sequences of arsenic efflux pumps detected in the SH, layers
322 surrounding the phylogenomic tree indicate: detection level, type of efflux pump,
323 taxonomy, the phylum level, and the species.
324

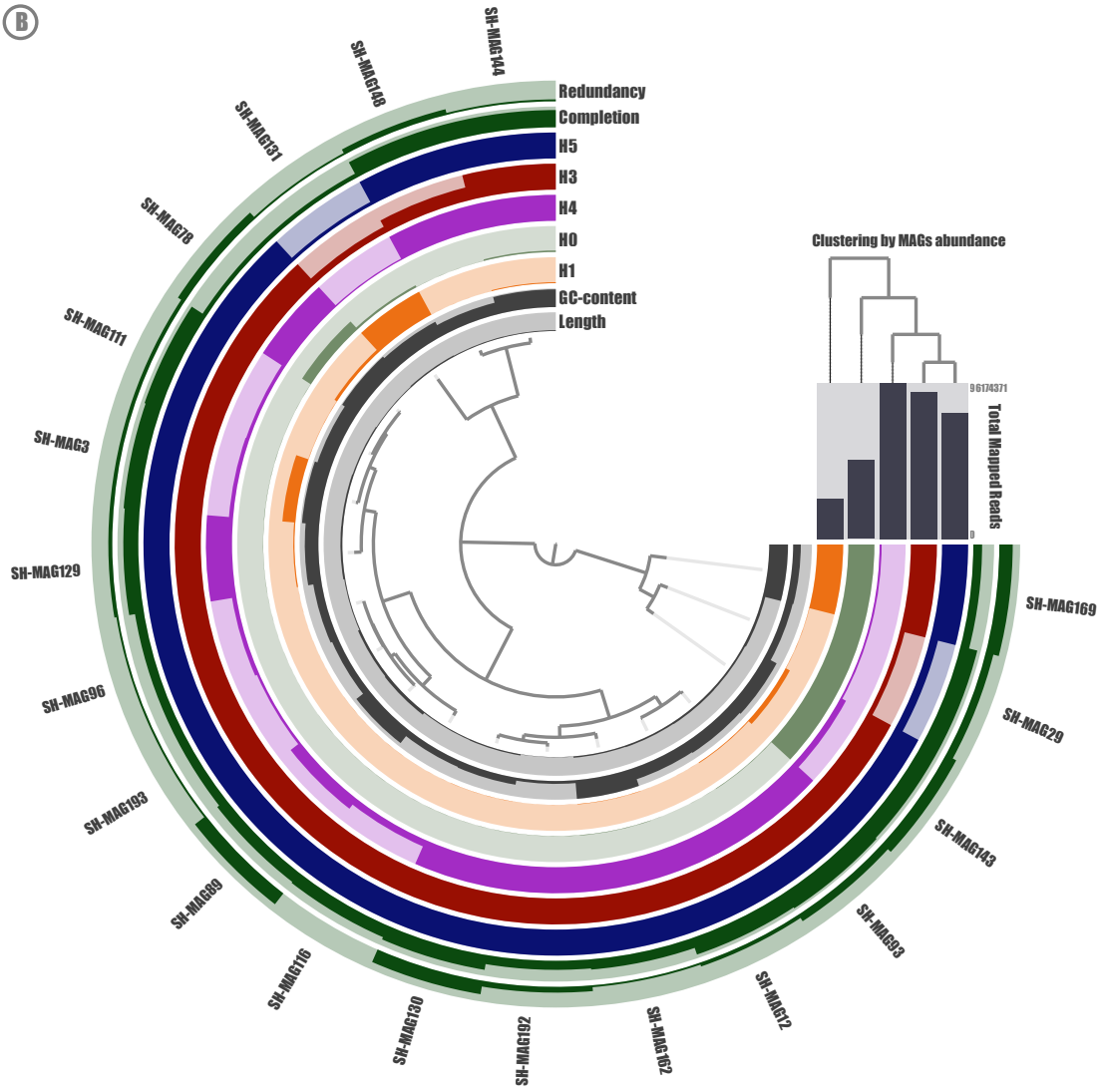
325 **Novel genomes from SH belong undescribed genera**

326 The binning the process reconstructed 195 bins (which were manually curated and
327 evaluated for completion and redundancy) to gain insights of non-culturable bacteria. This
328 resulted in 19 metagenome-assembled genomes (MAGs) that met the completion ≥ 80%
329 and redundancy ≤ 10% criteria (Figure 6); that clustered 4.99% of the contigs in the
330 metagenome profile database. The MAGs taxonomic affiliation resulted in 1/19 belonging
331 to the Eukaryota domain and the 18/19 remaining to the Bacteria domain, distributed in 4
332 different phyla (9 *Proteobacteria*, 7 *Bacteroidota*, 1 *Cyanobacteria* and 1
333 *Gemmatimonadetes* (Figure 6A). We only were able to assign 22% (4/18) of the Bacteria
334 MAGs to previously described genera, suggesting a great diversity of novel species that are
335 yet to be described. Also, MAGs contigs number was very variable from 208 to 626 among
336 the Bacteria (Eukaryota being substantially larger: 4,174), the same tendency was observed
337 for the GC-content (36.1% - 69.4%). On the other hand, most of the reconstructed MAGS
338 were represented or detected in the H3, H4 and H5 metagenomes, showing differential
339 abundance patterns in each one (Figure 6B; Supplementary Table S4).
340

A

		Anvi'o pipeline					CheckM		GTDB-Tk				
		Percent Completion	Percent Redundancy	Num Contigs	N50	GC content	Percent Completion	Percent Contamination	Phylum	Class	Order	Family	Genus
*	SH-MAG111	95.77%	2.82%	332	11,127	60.76%	81.95%	1.54%	Bacteroidota	---	---	---	---
*	SH-MAG116	90.14%	0.00%	208	14,823	43.49%	83.04%	0.00%	Bacteroidota	Bacteroidia	Chitinophagales	Saprospiraceae	
*|**	SH-MAG12	95.77%	2.82%	413	11,471	54.64%	92.06%	3.28%	Bacteroidota	---	---	---	---
*	SH-MAG130	92.96%	9.86%	395	10,529	44.80%	93.50%	11.30%	Bacteroidota	Bacteroidia	Bacteroidales	---	---
	SH-MAG192	85.92%	4.23%	314	4,317	36.06%	68.88%	4.52%	Bacteroidota	Ignavibacteria	Ignavibacteriales	Melioribacteraceae	---
**	SH-MAG193	85.92%	1.41%	561	10,013	44.34%	92.35%	3.33%	Bacteroidota	Bacteroidia	Cytophagales	Cyclobacteriaceae	---
	SH-MAG96	85.92%	1.41%	234	14,512	47.74%	80.17%	1.49%	Bacteroidota	Bacteroidia	Chitinophagales	Saprospiraceae	---
*|**	SH-MAG29	95.77%	5.63%	346	20,718	44.60%	97.71%	0.66%	Cyanobacteria	Cyanobacteriia	Cyanobacteriales	Phormidiaceae	Arthrospira
*	SH-MAG3	94.37%	2.82%	436	10,739	66.26%	86.11%	6.17%	Gemmatimonadota	Gemmatimonadetes	---	---	---
	SH-MAG89	88.73%	9.86%	498	14,477	62.15%	82.81%	4.83%	Proteobacteria	Deltaproteobacteria	Bradymonadales	---	---
	SH-MAG131	80.28%	1.41%	626	5,244	63.23%	77.20%	5.18%	Proteobacteria	Deltaproteobacteria	Desulfobacterales	Desulfosarcinaceae	---
*	SH-MAG143	95.77%	8.45%	601	7,367	58.76%	84.68%	2.47%	Proteobacteria	Deltaproteobacteria	Desulfobacterales	Desulfosarcinaceae	Desulfatitalea
*	SH-MAG148	95.77%	4.23%	384	7,277	55.17%	87.42%	0.86%	Proteobacteria	Deltaproteobacteria	Desulfobacterales	---	---
*|**	SH-MAG93	97.18%	5.63%	550	9,043	56.28%	94.35%	3.10%	Proteobacteria	Deltaproteobacteria	Desulfobacterales	Desulfosarcinaceae	Desulfosarcina
*|**	SH-MAG129	92.96%	2.82%	390	14,010	58.58%	91.11%	3.40%	Proteobacteria	Gammaproteobacteria	---	---	---
*	SH-MAG144	95.77%	1.41%	315	8,476	67.34%	87.51%	1.67%	Proteobacteria	Gammaproteobacteria	Ectothiorhodospirales	Thioalkalivibrionaceae	Thioalkalivibrio
	SH-MAG162	87.32%	1.41%	431	8,147	69.40%	81.81%	3.19%	Proteobacteria	Gammaproteobacteria	Pseudomonadales	---	---
	SH-MAG78	80.28%	5.63%	371	6,424	67.34%	73.01%	2.21%	Proteobacteria	Gammaproteobacteria	---	---	---
	SH-MAG169	85.54%	9.64%	4176	8,047	42.02%	72.44%	39.84%	---	---	---	---	---

B



341

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 ≥ 90% and redundancy/contamination ≤ 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 ≥90% and redundancy/contamination ≤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.

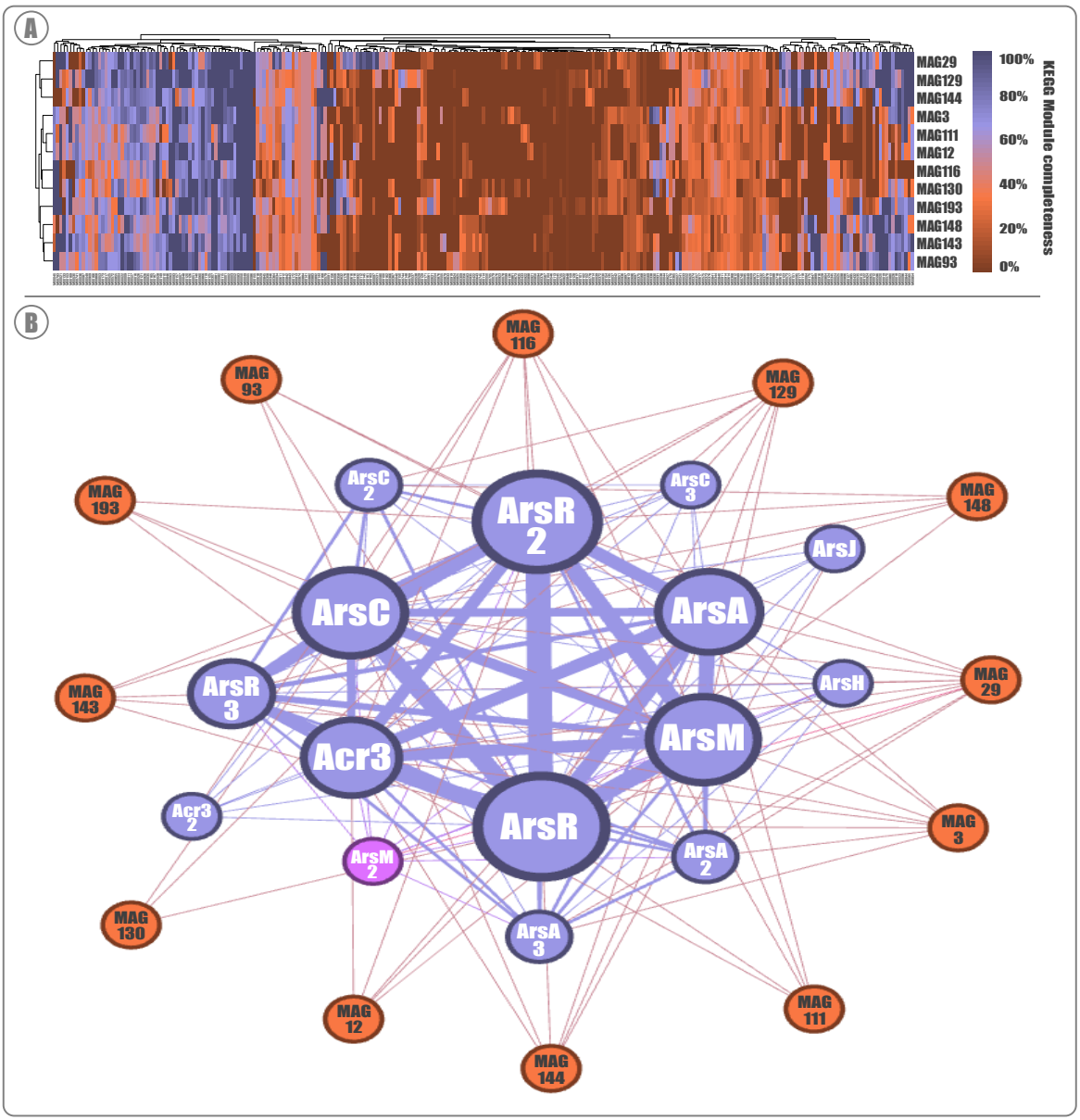


Figure 7. **MAGs functional potential profiles: global and arsenic specific. A)** Metabolic capabilities of the 12 selected MAGs according to KOfam (KEGG Orthologs database); the heatmap shows the KEGG Modules completition displayed by the color gradient for given module in each corresponding MAG. **B)** 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

385 The Salar de Huasco has an important spatial variation of arsenic in sediments, which
386 we have described as a gradient from north to south (Figure 1) (9); being consistent with
387 the widely reported heterogeneity for this area and the impact of extreme conditions over
388 the inhabitant organisms (30). Previous studies have detected variations in conductivity,
389 organic matter, and dissolved oxygen even in geographically closed areas (6, 7, 31, 32, 33).
390 In aquatic ecosystems, the quantity of trace elements and physical-chemical states of some
391 elements, are related to their circulation/interaction, and with the mechanisms of
392 corresponding living organism (34). Salar flats as SH have the peculiarity of containing more
393 than 50% of $CaCO_3$ and high levels of arsenic, which is very available and highly mobile. This
394 is because a fraction of the arsenic in sediments lies as soluble salts (Na_2HAsO_4); it also could
395 be associated to the calcium carbonate ($CaHAsO_4$) and, adsorbed into Fe and Al oxides (35).
396 The metabolism of inhabiting microbial communities or the activity of primary producers
397 could be the source of the diversity of abiotic properties (36). It is hypothesized that this
398 spatial and functional heterogeneity is the cornerstone of resilience in these extreme
399 ecosystems. We must consider the presence of high concentrations of arsenic as a selective
400 pressure, which could have a modulating effect on the composition of the H3, H4 and H5
401 communities (grouping them together and separating them from the rest).
402
403 The five communities described from the SH metagenomes are dominated by
404 *Proteobacteria, Bacteroidetes,* and *Cyanobacteria* (Figure 2), similar to previous reports for
405 other Altiplanic environments (Socompa volcanic Lake in Argentina and La Brava Lagoon -
406 Salar de Atacama in Chile) mostly carried out by 16S rRNA amplicon sequencing (37, 38).
407 Specifically, the same pattern has been reported previously for the SH (3, 32). However, the
408 proportions are different, probably because the shotgun metagenomic sequencing used in
409 this study is more sensitive, which is evidenced by the alpha diversity indices´ magnitude.
410 Nevertheless, these values are also higher than those reported by other metagenomic
411 studies from similar areas (39).
412

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 *Seohaeicola* 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).

442 At the lower available taxonomic rank, we found some bacteria that are typical for
443 extreme environments, such as *Thiobacillus*, a well-known sulfur-oxidizing bacterium, and
444 *Desulfotignum* an anaerobic group known to be sulphate-reductors, as well as other less
445 studied groups. The most abundant belong to the *Roseovarius* genus, an aerobic, non-
446 photoautotroph bacterium; *Rhodohalobacter* is a facultative anaerobic, moderately
447 halophilic group, and *Truepera* comprises aerobic quimio-organotropic alkaliphile
448 bacterium. This is evidence of the lifestyle diversity that these microorganisms have,
449 suggesting a cooperative and specialized community. Likewise, this is in coherence with our
450 other findings, the enrichment of photosynthetic organisms in sites H0 and H1, as well as
451 *Proteobacteria* as the most abundant group in all the other sampled sites. Other similarities
452 are evidenced, as the genus *Methylibium* is only found on the H0 site, where the methane-
453 related metabolism function was found to be enriched in the MEBS analyses. In addition,
454 the importance of some of these abundant taxa is highlighted, as they are structural parts
455 in the co-occurrence network, namely *Desulfomicrobium, Thiobacillus,* and
456 *Brumimicrobium*.
457
458 The functional potential of the five SH communities showed similar patterns in
459 general, with particular differences in some categories following the same association
460 pattern described before (Figure 3). Comparing each site, we found significant differences
461 in the enrichment of some important functional categories related to the maintenance of
462 critical geochemical cycles, including carbon, nitrogen, and phosphate metabolism, stress
463 resistance, transport of Zinc, Nickel, Cobalt, Iron, Manganese and osmotic stress. Notably,
464 the two functions with statistical differences between almost all communities were the
465 Circadian Clock in *Cyanobacteria* and the Calvin Benson cycle, suggesting differences in the
466 primary production and CO_2 fixation. The Circadian Rhythm is directly influenced by several
467 abiotic conditions, such as temperature and atmospheric pressure (52). Moreover, the
468 Calvin Benson cycle was the only metabolic pathway involved in CO_2 fixation reported in an
469 endolithic halite metagenome analysis, underlining its importance for microbial
470 communities living in these extreme conditions (53). This agrees with the MEBS analysis

471 results, where the carbon cycle relevant reactions presented a greater variation between
472 the metagenomes. Particularly, carbon flux control in these communities seems to be
473 different especially in regard to methanogenesis processes, as the far greater detection of
474 specific markers in H0 and H1 could account for a high presence of methanogenic species
475 in these two communities in comparison to the other metagenomes (54). This could be
476 related to the S decrease in H1, implying that methanogens outcompete sulfate reducers in
477 this community (55).
478
479 Furthermore, among these significant differences, arsenic resistance is enriched in
480 the metagenomes from the sites with the highest As concentration, which was expected
481 and has been reported previously (56, 57). Particularly, the distribution and abundance of
482 arsenic-related genes in the five SH communities showed significant differences among the
483 communities (Figure 4), as the presence of arsenic affects the dynamics of inhabiting
484 microorganisms (18, 58). Therefore, arsenic would promote the selection and abundance
485 of tolerant microorganisms (59, 60). Our results agree with this premise since the
486 abundance of all these genes increases following the gradient of arsenic concentration
487 among the five sites, indicating a different dynamic in each niche, even though there are
488 geographically closely located.
489
490 Overall, we found that most of the detected genes related to arsenic belong to
491 expulsion and reduction mechanisms (*arsR, arsC,* and *acr3*), which correlates with what was
492 reported in the Argentinian Altiplano (22). The transcriptional repressor ArsR was the one
493 with the highest abundance detected; as this protein participates in the regulation of
494 different arsenic pumps, namely ArsB, ACR3, and ArsP (61, 62); and is also reported to be in
495 operon conformations along with ArsC or ArsM (63). Hence, it was expected to be the most
496 abundant and broadly distributed. Moreover, to better assess the proportion of the arsenic
497 mechanisms found in SH, we separated the genes related to As expulsion out of the cell
498 from those related to reduction mechanism (usually reported together). This is due to the
499 fact that gene clusters including *arsB, arsP, arsK*, *arsJ,* and *acr3* are commonly

found/reported independent of arsenate reductase *arsC* (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 (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 (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 (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 *Exiguobacterium* genomes using a combination of bioinformatic tools (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 (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 *arsP*

529 was detected in the higher arsenic concentration sites. Therefore, this may be a
530 counteraction against a possible greater production of these highly toxic organic species at
531 these sites, where consistently the Arsenite Methyltransferase (*arsM*) gene was enriched
532 as well. Thus, these markers could be proposed as an arsenic contamination bioindicator
533 like *acr3* (73). Additionally, the distribution and phylogenetic relationships of arsenic pumps
534 among the metagenomes revealed distinct groups, which is consistent with those
535 previously reported: I) ArsB - Acr3 [ion/BART], II) ArsP [Permease]; III) ArsK - ArsJ [MFS] (65).
536 It has been hypothesized that the divergence presented by the arsenic resistance
537 mechanisms originated from Earth's geological changes to adapt for a particular function or
538 As emergent species (74). Besides, we found that *Proteobacteria* possess a broader gene
539 repertoire for arsenic response, which has also been described before (22).
540
541 The twelve reconstructed high-quality MAGs provide information about the
542 abundance and distribution of undescribed or unknown microorganisms (Figure 6), as well
543 as genomic insights into widespread arsenic metabolism. Most of these are affiliated within
544 *Proteobacteria* and *Bacteroidetes* phyla, reflecting what we described at the community
545 level. As stated before, the recovery of MAGs complements decades of cultivation and PCR-
546 survey efforts by providing information about taxa missing in culture collections (Candidate
547 Phyla Radiation), improving our understanding of microbial communities (75); particularly
548 those inhabiting extreme environments. Nonetheless, four MAGs were classified at the
549 genus rank: *Arthrospira*, a *Cyanobacteria* with great commercial interest as a pigment
550 source and to make spirulina supplements; *Thioalkalivibrio:* an aerobic sulfur-oxidizer;
551 *Desulfatitalea* and *Desulfosarcina*: two anaerobic sulfate reducers (76, 77, 78).
552
553 In the same way that these MAGs only cover the most abundant phyla of SH, we
554 only detected genes that belong to the most prevalent arsenic resistance mechanisms on
555 them (Figure 7). Thus, supporting expulsion, reduction, and methylation as the most
556 recurrent widespread mechanisms in the SH to cope with the high arsenic concentrations.
557 Conversely, no arsenic oxidation and/or respiration-related gene was detected on the

558 recovered MAGs. This could indicate that the different ways used to expel the toxic out of
559 the cell would be more straightforward and direct as a first line of defense against arsenic;
560 while the oxidation and respiration would involve a coupling with the bacteria central
561 metabolism and a larger machinery to benefit from this compound, which seems to be
562 associated with less abundant and highly adapted or specialized bacteria. Also, arsenic
563 oxidation and respiration have been reported in association with nitrate and sulfate
564 reduction, respectively (10, 79). Nonetheless, we need to take into consideration the MAGs
565 completion levels as a possible source for missing genes.
566
567 Overall, our results reveal that populations of *Proteobacteria, Bacteroidetes* and
568 *Cyanobacteria* are abundant across wide ecological niches in the SH spanning a challenging
569 ensemble of environmental conditions and physico-chemical parameters, among which
570 arsenic is highly relevant. As the arsenic cycle and the bacterial contribution to it, has yet to
571 be completely understood, more environmental studies are needed, while metagenomic
572 approaches have been shedding light on the possible role of unknown or undescribed
573 microorganisms additional transcriptomic, metabolomic and cultivation approaches will be
574 essential to define these phenomena and its relevance in a global context.
575
576
577 **CONCLUSION**
578 The Altiplano array of ecological niches is a reservoir for microbial diversity, showing
579 great richness in adapted organisms capable of facing these challenging conditions.
580 Particularly, Salar de Huasco is a highly diverse ecosystem, where salinity and As
581 concentration contribute to shaping the community composition, mainly represented by
582 *Proteobacteria* and *Bacteroidetes*. Also, the interaction networks within these communities
583 showed three distinct groups of related taxa, but no “key stone” nodes were found.
584 Nonetheless, little niche overlap was determined. Altogether, these indicate that SH studied
585 niches harbor highly diverse communities, being H1 and H5 the most contrasting ones.
586 Moreover, the most abundant arsenic-related genes found in these communities indicate

587 that the As(V) reduction and subsequent As(III) expulsion would be the most common
588 strategy used to detoxify the cell of arsenic; furthermore, regarding to the expulsion pumps,
589 the most abundant were Acr3, followed by ArsB*;* however, in sites with high As
590 concentration ArsP begins to be enriched. In addition, 12 high-quality, non-redundant
591 MAGs were reconstructed from the metagenomes; those represented the dominant
592 diversity detected across the communities as well as the metabolic variability and the
593 presence of marker-genes related to the most recurrent As resistance mechanisms
594 (expulsion, reduction and methylation). Finally, in order to further elucidate the strategies
595 and relationships between the microbial taxa and among biotic and abiotic components
596 within the ecosystem, further multidisciplinary studies are required, as well as the use of
597 the ever evolving NGS approaches with the increasing database information to better
598 understand the evolutionary process of adaptation to the extreme conditions presented by
599 these unique ecosystems.
600
601
602 **MATERIALS AND METHODS**
603 **Study Area and Sampling**
604 Salar de Huasco National Park (SH) is an area located on the Chilean Altiplano that is
605 known for its spatial heterogeneity, as well as great biodiversity and physicochemical
606 characteristics. This ecosystem is mostly composed of streams, salt crusts, peatlands, and
607 shallow (permanent and non-permanent) lakes with salinity and arsenic gradients from
608 north to south (9). We sampled surface sediments (to a depth of 5 cm) in sterile tubes from
609 five different sites (H0 to H5 previously described in (6)) by duplicate during fieldwork done
610 in June 2018; those samples were kept and transported in a cooler until stored at -20ºC for
611 subsequent DNA extraction. Physicochemical parameters like temperature, salinity, and pH
612 were recorded (HI 98192 and HI 2211 - HANNA Instruments) *in situ* (Figure 1).
613
614 **DNA extraction and high throughput shotgun sequencing**

615 Total DNA was extracted from the sediment samples from each SH site using the
616 DNeasy PowerSoil Kit (Qiagen Inc., Hilden, Germany) following manufacturer's instructions.
617 DNA integrity, quality, and quantity were verified through 1% agarose gel electrophoresis,
618 $OD_{260/280}$ ratio, and fluorescence using a Qubit® 3.0 Fluorometer along with the Qubit dsDNA
619 HS Assay Kit (Thermo Fisher Scientific, MA, USA). Next, Paired-end (150bp) libraries were
620 constructed for each sample in duplicates, at the Centro de Biotecnología Vegetal,
621 Universidad Andrés Bello (Santiago, Chile) using the TruSeq Nano DNA Kit (Illumina Inc., CA,
622 USA.) following the TruSeq Nano DNA Sample Preparation Guide 15041110 Rev. D. Libraries
623 were sent for sequencing at Macrogen Inc. (Seoul, Korea) on a HiSeq 4000 platform
624 (Illumina Inc., CA, USA). Then, raw data was evaluated using FastQC v0.11.8 (80) for quality
625 control, adapters were removed from the reads of all samples using Trimmomatic v0.30
626 (81) and the filtering and trimming (length ≤ 100bp, Ns = 0, and Q ≤ 30 thresholds) was
627 performed with PRINSEQ v0.20.4 (82). The whole raw data sets have been deposited at
628 DDBJ/ENA/GenBank under the Bioproject: PRJNA573913.
629
630 **Taxonomic profiling analysis**
631 Quality-controlled reads for each sample were profiled using the phyloFlash pipeline
632 (83) to obtain all reads that align with the bacterial SSU rRNA (small-subunit rRNA) SILVA
633 v138 database (84). Later, these sequences (FASTQ files) were processed using R v3.5.2 and
634 RStudio v1.1.463 (85, 86) following the DADA2 v1.16.0 R package pipeline (87) to infer
635 amplicon sequence variants (ASVs) present in each sample. Briefly, after dereplication,
636 denoising, and paired reads merge steps, the ASV table was built with 97% clustering, the
637 chimeras removed, and the taxonomic was assigned against the Silva v138 database (84)
638 using DADA2 Ribosomal Database Project's (RDP) naive Bayesian classifier (88). Then, data
639 was normalized by variance stabilizing transformation using the R package DESeq2 v1.28.1
640 (89). Also, a multiple sequence alignment was created using the R package DECIPHER
641 v2.16.1 (90) to infer a phylogeny with FastTree v2.1.10 (91). Furthermore, a phyloseq-object
642 (containing the ASVs, taxonomy assignation, phylogenetic tree, and samples meta-data)
643 was created using the R package Phyloseq v1.32.0 (92); in order to calculate the alpha

644 diversity indexes, along with btools v0.0.1 R package. Also, beta diversity (PCoA - Bray Curtis
645 distance with environmental variables fit) was calculated using the R package ampvis2
646 v2.4.5 (93), and visualizations were generated with ggplot2 v2.2.1 (94) R package.
647
648 **Co-occurrence networks**
649 We used the same phyloseq object, which was agglomerated by best hit using the
650 microbiomeutilities v1.00.11 R package (95); filtered by tax abundance (0.5% in at least one
651 sample) using the Genefilter v 1.72.0 (96) and Phyloseq v1.32.0 (92) R packages. Then, the
652 co-occurrence network was estimated using the SPIEC-EASI v0.1.4 R package (SParse
653 InversE Covariance Estimation for Ecological Association Inference) (97), using
654 neighborhood selection model (lambda.min.ratio=1e-2, nlambda=20, and 50 replicates
655 parameters). Finally, the network was visualized using the ggnet2 function of GGally v1.5.0
656 R package (a ggplot2 extension) (98).
657
658 **Functional profiling analysis**
659 The patterns of functional potential as subsystems with different specificity levels
660 for each community were determined by the presence/absence and relative abundance of
661 the quality-controlled reads that aligned against the metabolic pathways and functions in
662 the SEED database (99) with SUPER-FOCUS v0.35 software (100), which uses DIAMOND
663 v2.0.6 (101) for fast and efficient alignment. Statistical analysis was carried out in STAMP
664 v2.3.1 (Software package for Analyzing Taxonomic or Metabolic Profiles) (102) using
665 Welch's t-test to compare all samples.
666
667 **Metagenome co-assembly and read mapping**
668 As we need for subsequent analyses: i) a co-assembly (the reads from all samples
669 assembled together: .fasta file), ii) its annotations (.gff, .ffn and .faa files) and iii) an
670 alignment per samples (all the reads from each sample mapped against the co-assembly:
671 .bam files); we proceeded to assemble the quality-controlled reads from all samples using
672 MEGAHIT v1.1.3 (103), with the –presets meta-large option and a minimum contig length

673 of 1 kb. The co-assembly was then evaluated with MetaQUAST v5.0.2 (104) and annotated
674 with Prokka v1.11 (105) using the metagenome mode. Furthermore, we mapped the
675 quality-controlled reads from each sample against the co-assembly using Bowtie2 v2.3.4
676 (106) and stored the recruited reads (sorted and indexed) as BAM files using SAMtools v1.3
677 (107).
678
679 **Metagenome profiling**
680 We followed the anvi'o v7 pipeline (108). First, we created a contig database with
681 the co-assembed contigs, which uses Prodigal v2.6.3 (109), HMMER v3.3.1 (110) and NCBI
682 COGs (111) to identify genes calls and functionally annotate them (anvi-gen-contigs-
683 database, anvi-run-hmms, anvi-run-ncbi-cogs). Secondly, we profiled each sample's BAM
684 file against the contigs database to estimate the detection and coverage statistics for each
685 contig (anvi-profile). Then, we combined the profiles in a single merged metagenomic
686 profile database, which uses all individual statistics to compute hierarchical clustering of
687 the contigs (anvi-merge). Finally, we visualized the merged profile on the anvi'o interactive
688 interface, which allows easy exploration and curation of the metagenomes (anvi-
689 interactive). On the other hand, we used MEBS (Multigenomic Entropy Based Score) v1.0
690 package (112) to evaluate and compare the S, N, O, CH_4, and Fe biogeochemical cycles on
691 each metagenome using the predicted proteins (annotated .faa file from each individual
692 assembly).
693
694 **Metagenome target gene search**
695 To estimate the abundance of genes related to arsenic metabolism, the read counts
696 for each predicted gene of each sample were obtained using the corresponding BAM file
697 and the co-assembly annotated GFF file, with HTSeq-Counts v0.13.5 (113). Also, we
698 constructed a database from the GenBank – "Identical Protein Groups" with all available
699 prokaryotic proteins of arsenic metabolism (Acr3, AioA, AioB, AioR, AioS, AioX, AoxA, AoxB,
700 AoxC, AoxD, ArrA, ArrB, ArrC, ArsA, ArsB, ArsC, ArsD, ArsH, ArsJ, ArsK, ArsM, ArsN, ArsO,
701 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 KOfam - 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-

731 metabolism). The MAGs module completition was visualized using pheatmap v1.0.12 R
732 package (115).
733
734 **MAGs target gene search**
735 All MAGs were queried against the same previously constructed database (with the
736 arsenic metabolism genes) using CRB-BLAST v0.6.6 (E-value ≥$1E^{-05}$, Identity ≥70% and Query
737 Coverage ≥70%) (114). Furthermore, the resulting hits matrix was compared and
738 supplemented with the detected KOfams related to arsenic, and the final matrix was used
739 to generate a functional network using Gephi v0.9.2 (124) to connect MAGs and detected
740 genes.
741
742
743 **DATA AVAILABILITY**
744 The whole raw data sets have been deposited at DDBJ/ENA/GenBank under the Bioproject:
745 PRJNA573913.
746
747
748 **AUTHOR CONTRIBUTIONS**
749 JC-S, CP-E, EC-N, FR, and CPS conceived and designed the study. JC-S, JF, FM, FR, and CPS
750 performed the field work. JC-S and CP-E processed the samples, performed the
751 experimental procedures and carried out the bioinformatics analyzes. KM, SM and EC-N
752 contributed with critical bioinformatics advice. CPS, FR, and EC-N contributed with reagents,
753 materials, and analysis tools. JC-S and CP-E interpreted the results and wrote the first
754 manuscript draft. All authors read and approved the final manuscript.
755
756
757 **ACKNOWLEDGEMENTS**
758 We would like to thank the illustrator Florence Gutzwiller for the SH landscape painting
759 (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

1. Bahram M, Netherway T, Frioux C, Ferretti P, Coelho LP, Geisen S, Bork P, Hildebrand F. 2020. Metagenomic assessment of the global diversity and distribution of bacteria and fungi. Env Microbiol 23(1):316-326
2. Albarracín V, Kurth D, Ordoñez O, Belfiore C, Luccini E, Salum G, Piacentini R, Farías, ME. 2015. High-Up: A remote reservoir of Microbial Extremophiles in Central Andean Wetlands. Front Microbiol 6, 1404.

3. Dorador C, Vila I, Witzel KP, Imhoff JF. 2013. Bacterial and archaeal diversity in high altitude wetlands of the Chilean Altiplano. Fundam Appl Limnol 182, 135–159.
4. Aceituno P. 1997. Aspectos generales del clima en el Altiplano Sudamericano. In R Charrier, P Aceituno, M Castro, A Llanos, LA Raggi (Eds.), *El Altiplano: ciencia y conciencia de los Andes, Actas del 21 simposio internacional de estudios altiplánicos* (pp. 63–69). Santiago, Chile: Universidad de Chile.
5. Risacher F, Alonso H, Salazar C. 2003. The origin of brines and salts in Chilean salars: A hydrochemical review. Earth Sci Rev 63, 249–293.
6. Dorador C, Vila I, Imhoff JF, Witzel KP. 2008. Cyanobacterial diversity in Salar de Huasco, a high altitude saline wetland in northern Chile: an example of geographical dispersion?. FEMS microbiol 64(3), 419-432.
7. Hernández KL, Yannicelli B, Olsen LM, Dorador C, Menschel EJ, Molina, V, Remonsellez F, Hengst M, Jeffrey WH. 2016. Microbial activity response to solar radiation across contrasting environmental conditions in Salar de Huasco, Northern Chilean Altiplano. Front Microbiol 7, 1857.
8. Molina V, Hernández K, Dorador C, Eissler Y, Hengst M, Pérez V, Harrod C. 2016. Bacterial active community cycling in response to solar radiation and their influence on nutrient changes in a high- altitude wetland. Front Microbiol 7, 1–15.
9. Castro-Severyn J, Pardo-Esté C, Mendez KN, Morales N, Marquez SL, Molina F, Remonsellez F, Castro-Nallar E, Saavedra CP. 2020. Genomic Variation and Arsenic Tolerance Emerged as Niche Specific Adaptations by Different *Exiguobacterium* Strains Isolated From the Extreme Salar de Huasco Environment in Chilean–Altiplano. Front Microbiol 11, 1632.
10. Demergasso CS, Guillermo CD, Lorena EG, Mur JJP, Pedrós-Alió C. 2007. Microbial precipitation of arsenic sulfides in Andean salt flats. Geomicrobiol J 24, 111–123.
11. Escalante G, Campos V, Valenzuela C, Yañez J, Zaror C, Mondaca M. 2009. Arsenic resistant bacteria isolated from arsenic contaminated river in the Atacama Desert (Chile). Bull Environ Contam Toxicol 83(5):657H661.
12. Escudero LV, Casamayor EO, Chong G, Pedrós-Alió C, Demergasso C. 2013. Distribution of microbial arsenic reduction, oxidation and extrusion genes along a wide range of environmental arsenic concentrations. PLoS One 8(10), e78890.
13. Finstad KM, Probst AJ, Thomas BC, Andersen GL, Demergasso C, Echeverría A, Amundson R, Banfield JF. 2017. Microbial community structure and the persistence of cyanobacterial populations in salt crusts of the hyperarid Atacama Desert from genome-resolved metagenomics. Front Microbiol 8, 1435.
14. Wang L, Zhipeng Y, Jing C. 2020. Metagenomic insights into microbial arsenic metabolism in shallow groundwater of Datong basin, China. Chemosphere. 245. 125603.
15. Berhe A, Barnes R, Six J, Marin-Spiotta E. 2018. Role of soil erosion in biogeochemical cycling of essential elements: carbon, nitrogen, and phosphorus. Annu Rev Earth Pl Sci 46, 521–548.
16. Zhang J, Shi Q, Fan S, Zhang Y, Zhang M, Zhang J. 2020. Distinction between Cr and other heavy-metal-resistant bacteria involved in C/N cycling in contaminated soils of copper producing sites. J Hazard Mater 402, 123454.

17. Li Y, Zhang M, Xu R, Lin H, Sun X, Xu F, Gao P, Kong T, Xiao E, Yang N, Sun W. 2021. Arsenic and antimony co-contamination influences on soil microbial community composition and functions: Relevance to arsenic resistance and carbon, nitrogen, and sulfur cycling. Env Int 153, 106522.
18. Andres J, Bertin PN. 2016. The microbial genomics of arsenic. FEMS Microbiol Rev. 40:299–322.
19. Desoeuvre A, Casiot C, Héry M. 2016. Diversity and distribution of arsenic-related genes along a pollution gradient in a river affected by acid mine drainage. Microb Ecol 71(3), 672-685.
20. Rosen BP. 2002. Biochemistry of arsenic detoxification. FEBS Lett 529:86–92.
21. Sheik CS, Mitchell TW, Rizvi FZ, Rehman Y, Faisal M, Hasnain S, McInerney M, Krumholz, LR. 2012. Exposure of soil microbial communities to chromium and arsenic alters their diversity and structure. PloS one 7(6), e40059.
22. Kurth D, Amadio A, Ordoñez O, Albarracín V, Gärtner W, Farías ME. 2017. Arsenic metabolism in high altitude modern stromatolites revealed by metagenomic analysis. Sci Reports. 7: 1024.
23. Danczak RE, Johnston MD, Kenah C, Slattery M, Wilkins MJ. 2019. Capability for arsenic mobilization in groundwater is distributed across broad phylogenetic lineages. PloS one 14(9), e0221694.
24. Dunivin TK, Yeh SY, Shade A. 2019. A global survey of arsenic-related genes in soil microbiomes. BMC biology 17(1), 1-17.
25. Aykanat T, Lindqvist M, Pritchard VL, Primmer CR. 2016. From population genomics to conservation and management: A workflow for targeted analysis of markers identified using genome-wide approaches in Atlantic salmon Salmo salar. J Fish Biol 89(6), 2658-2679.
26. Ezeokoli OT, Bezuidenhout CC, Maboeta MS, Khasa DP, Adeleke RA. 2020. Structural and functional differentiation of bacterial communities in post-coal mining reclamation soils of South Africa: bioindicators of soil ecosystem restoration. Sci Rep 10(1), 1-14.
27. Dorador C, Vila I, Remonsellez F, Imhoff JF, Witzel KP. 2010. Unique clusters of Archaea in Salar de Huasco, an athalassohaline evaporitic basin of the Chilean Altiplano. FEMS Microbiol 73(2), 291-302.
28. Eissler Y, Dorador C, Kieft B, Molina V, Hengst M. 2020. Virus and Potential Host Microbes from Viral-Enriched Metagenomic Characterization in the High-Altitude Wetland, Salar de Huasco, Chile. Microorganisms 8(7), 1077.
29. Remonsellez F, Castro-Severyn J, Pardo-Esté C, Aguilar P, Fortt J, Salinas C, Barahona S, León J, Fuentes B, Areche C, Hernández K, Saavedra CP. 2018. Characterization and salt response in recurrent halotolerant *Exiguobacterium* sp. SH31 isolated from sediments of Salar de Huasco, Chilean Altiplano. Front Microbiol 9, 2228.
30. Vásquez-Dean J, Maza F, Morel I, Pulgar R, González M. 2020. Microbial communities from arid environments on a global scale. A systematic review. Biol Res 53(1), 1-12.
31. Cabrol NA, Grin EA, Chong G, Minkley E, Hock AN, Yu Y, Bebout L, Fleming E, Häder D, Demergasso C, Gibson J, Escudero L, Dorador C, Lim D, Woosley C, Morris R,

879 Tambley T, Gaete V, Galvez M, Smith E, Uskin-Peate I, Salazar C, Dawidowicz G,
880 Majerowicz J. 2009. The high-lakes project. J Geophys Res 114(G2).
881 32. Aguilar P, Acosta E, Dorador C, Sommaruga R. 2016. Large differences in bacterial
882 community composition among three nearby extreme waterbodies of the high
883 Andean plateau. Front Microbiol 7, 976.
884 33. Cortés-Albayay C, Silber J, Imhoff JF, Asenjo JA, Andrews B, Nouioui I, Dorador C.
885 2019. The polyextreme ecosystem, Salar de Huasco at the Chilean Altiplano of the
886 Atacama Desert houses diverse Streptomyces spp. with promising pharmaceutical
887 potentials. Diversity 11(5), 69.
888 34. Bourg ACM. 1989. Adsorption of Trace Inorganic and Organic Contaminants by Solid
889 Particulate Matter, in Aquatic Ecotoxicology: Fundamental Concepts and
890 Methodologies. Boudou A and Ribeyre F. eds. CRC Press, Inc. Boca Raton, Florida;
891 pp.107-148.
892 35. Herrera V, De Gregori I, Pinochet H. 2009. Assessment of trace elements and
893 mobility of arsenic and manganese in lagoon sediments of the salars of Huasco and
894 Coposa, Chilean Altiplano. J Chil Chem Soc 54(3), 282-288.
895 36. Oren A. 2013. Life at high salt concentrations, intracellular KCl concentrations, and
896 acidic proteomes. Front Microbiol 4, 315.
897 37. Farías ME, Rascovan N, Toneatti DM, Albarracín VH, Flores MR, Poiré DG, Collavino
898 M, Aguilar M, Vazquez M, Polerecky L. 2013. The discovery of stromatolites
899 developing at 3570 m above sea level in a high-altitude volcanic lake Socompa,
900 Argentinean Andes. PloS one 8(1), e53497.
901 38. Farías ME, Contreras M, Rasuk MC, Kurth D, Flores MR, Poire DG, Novoa F, Visscher
902 PT. 2014. Characterization of bacterial diversity associated with microbial mats,
903 gypsum evaporites and carbonate microbialites in thalassic wetlands: Tebenquiche
904 and La Brava, Salar de Atacama, Chile. Extremophiles 18(2), 311-329.
905 39. Saona LA, Soria M, Durán-Toro V, Wörmer L, Milucka J, Castro-Nallar E, Meneses C,
906 Contreras M, Farías ME. 2021. Phosphate-Arsenic Interactions in Halophilic
907 Microorganisms of the Microbial Mat from Laguna Tebenquiche: from the
908 Microenvironment to the Genomes. Microb Ecol 1-13.
909 40. Ko MS, Park HS, Kim KW, Lee JU. 2013. The role of Acidithiobacillus ferrooxidans and
910 Acidithiobacillus thiooxidans in arsenic bioleaching from soil. Environ Geochem
911 HLTH 35(6), 727-733.
912 41. Demergasso C, Casamayor EO, Chong G, Galleguillos P, Escudero L, Pedros-Alió C.
913 2004 Distribution of prokaryotic genetic diversity in athalassohaline lakes of the
914 Atacama Desert, Northern Chile. FEMS Microbiol Ecol 48:57–69
915 42. Demergasso C, Escudero L, Casamayor EO, Chong G, Balague V, Pedros-Alió C. 2008
916 Novelty and spatio-temporal heterogeneity in the bacterial diversity of hypersaline
917 Lake Tebenquiche (Salar de Atacama). Extremophiles 12:491–504
918 43. Demergasso C, Dorador C, Meneses D, Blamey J, Cabrol N, Escudero L, Chong G.
919 2010 Prokaryotic diversity pattern in high-altitude ecosystems of the Chilean
920 Altiplano. J Geophys Res 115:G00D09
921 44. Dorador C. 2007. Microbial communities in high altitude altiplanic wetlands in
922 northern Chile: phytogeny, diversity and function. Doctoral dissertation, Christian-

Albrechts-Universität, Kiel, Germany, p 166

45. Jiang H, Dong H, Yu B, Liu X, Li Y, Ji S, Zhang CL .2007. Microbial response to salinity change in Lake Chaka, a hypersaline lake on Tibetan plateau. Environ Microbiol 9:2603–2621

46. Jung P, Baumann K, Lehnert LW, Samolov E, Achilles S, Schermer M, Wraase L, Eckhardt KU, Bader M, Leinweber P, Karsten U, Bendix J, Büdel B. 2020. Desert breath—How fog promotes a novel type of soil biocenosis, forming the coastal Atacama Desert's living skin. Geobiology 18(1), 113-124.

47. Samolov E, Baumann K, Büdel B, Jung P, Leinweber P, Mikhailyuk T, Karsten U, Glaser K. 2020. Biodiversity of algae and cyanobacteria in biological soil crusts collected along a climatic gradient in Chile using an integrative approach. Microorganisms 8(7), 1047.

48. Castro-Severyn J, Remonsellez F, Valenzuela SL, Salinas C, Fortt J, Aguilar P, Pardo-Esté C, Dorador C, Quatrini R, Molina F, Aguayo D, Castro-Nallar E, Saavedra CP. 2017. Comparative genomics analysis of a new *Exiguobacterium* strain from Salar de Huasco reveals a repertoire of stress-related genes and arsenic resistance. Front Microbiol 8, 456.

49. Castro-Severyn J, Pardo-Esté C, Sulbaran Y, Cabezas C, Gariazzo V, Briones A, Morales N, Séveno M, Decourcelle, Salvetat N, Remonsellez F, Castro-Nallar E, Saavedra CP. 2019. Arsenic response of three altiplanic *Exiguobacterium* strains with different tolerance levels against the metalloid species: a proteomics study. Front Microbiol 10, 2161.

50. Idris H, Goodfellow M, Sanderson R, Asenjo JA, Bull AT. 2017. Actinobacterial rare biospheres and dark matter revealed in habitats of the Chilean Atacama Desert. Sci Rep 7(1), 1-11.

51. Agler MT, Ruhe J, Kroll S, Morhenn C, Kim ST, Weigel D, Kemen EM. 2016. Microbial hub taxa link host and abiotic factors to plant microbiome variation. PLoS biology 14(1), e1002352.

52. Kitahara R, Oyama K, Kawamura T, Mitsuhashi K, Kitazawa S, Yasunaga K, Sagara N, Fujimoto M, Terauchi, K. 2019. Pressure accelerates the circadian clock of cyanobacteria. Sci Rep 9(1), 1-8.

53. Crits-Christoph A, Gelsinger DR, Wierzchos J, Ravel J, Davila A, Casero MC, DiRuggiero J. 2016. Functional interactions of archaea, bacteria and viruses in a hypersaline endolithic community. Env Microbiol 18(6), 2064-2077.

54. Luton PE, Wayne JM, Sharp RJ, Riley PW. 2002. The *mcrA* gene as an alternative to 16S rRNA in the phylogenetic analysis of methanogen populations in landfill. Microbiology 148(11), 3521-3530.

55. Sela-Adler M, Ronen Z, Herut B, Antler G, Vigderovich H, Eckert W, Sivan O. 2017. Co-existence of methanogenesis and sulfate reduction with common substrates in sulfate-rich estuarine sediments. Front Microbiol 8, 766.

56. Xiao KQ, Li LG, Ma LP, Zhang SY, Bao P, Zhang T, Zhu YG. 2016. Metagenomic analysis revealed highly diverse microbial arsenic metabolism genes in paddy soils with low-arsenic contents. Env Poll 211, 1-8.

57. Zhai W, Qin T, Li L, Guo T, Yin X, Khan MI, Hashmi MZ, Liu X, Tang X, Xu, J. 2020. Abundance and diversity of microbial arsenic biotransformation genes in the sludge of full-scale anaerobic digesters from a municipal wastewater treatment plant. Env Int 138, 105535.
58. Slyemi D, Bonnefoy V. 2012. How prokaryotes deal with arsenic. Env Microbiol Rep 4(6), 571-586.
59. Jia Y, Huang H, Chen Z, Zhu YG. 2014. Arsenic uptake by rice is influenced by microbe-mediated arsenic redox changes in the rhizosphere. Env Sci Tech 48(2), 1001-1007.
60. Zhang SY, Zhao FJ, Sun GX, Su JQ, Yang XR, Li H, Zhu YG. 2015. Diversity and abundance of arsenic biotransformation genes in paddy soils from southern China. Env Sci Tech 49(7), 4138-4146.
61. Osman D, Cavet JS. 2010. Bacterial metal-sensing proteins exemplified by ArsR–SmtB family repressors. Nat Prod Rep 27(5), 668-680.
62. Shen Z, Luangtongkum T, Qiang Z, Jeon B, Wang L, Zhang Q. 2014. Identification of a novel membrane transporter mediating resistance to organic arsenic in *Campylobacter jejuni*. Antimicrob Agents Chemother 58(4), 2021-2029.
63. Fekih BI, Zhang C, Li YP, Zhao Y, Alwathnani HA, Saquib Q, Rensing C, Cervantes C. 2018. Distribution of Arsenic Resistance Genes in Prokaryotes. Front Microbiol 9, 2473.
64. Mateos LM, Villadangos AF, Alfonso G, Mourenza A, Marcos-Pascual L, Letek M, Pedre B, Messens J, Gil JA. 2017. The arsenic detoxification system in corynebacteria: basis and application for bioremediation and redox control. Adv Appl Microbiol 99, 103-137.
65. Shi K, Li C, Rensing C, Dai X, Fan X, Wang G. 2018. Efflux transporter ArsK is responsible for bacterial resistance to arsenite, antimonite, trivalent roxarsone, and methylarsenite. App Env Microbiol 84(24).
66. Xue S, Jiang X, Wu C, Hartley W, Qian Z, Luo X, Li W. 2020. Microbial driven iron reduction affects arsenic transformation and transportation in soil-rice system. Env Poll 260, 114010.
67. Oremland RS, Saltikov CW, Wolfe-Simon F, Stolz JF. 2009. Arsenic in the evolution of earth and extraterrestrial ecosystems. J Geomicrobiol 26(7), 522-536.
68. Mazumder P, Sharma SK, Taki K, Kalamdhad AS, Kumar M. 2020. Microbes involved in arsenic mobilization and respiration: a review on isolation, identification, isolates and implications. Environ Geochem HLTH 1-27.
69. Lobb B, Kurtz DA, Moreno-Hagelsieb G, Doxey AC. 2015. Remote homology and the functions of metagenomic dark matter. Front Genet 6, 234.
70. da Costa WLO, Araújo CLDA, Días LM, Pereira LCDS, Alves JTC, Araujo FA, Folador EL, Henriques I, Silva A, Folador ARC. 2018. Functional annotation of hypothetical proteins from the *Exiguobacterium antarcticum* strain B7 reveals proteins involved in adaptation to extreme environments, including high arsenic resistance. PloS one 13(6), e0198965.
71. Antczak M, Michaelis M, Wass MN. 2019. Environmental conditions shape the nature of a minimal bacterial genome. Nat comm 10(1), 1-13.

72. Makarova KS, Wolf YI, Koonin EV. 2019. Towards functional characterization of archaeal genomic dark matter. Biochem Soc Trans 47(1), 389-398.
73. Cai L, Liu G, Rensing C, Wang G. 2009. Genes involved in arsenic transformation and resistance associated with different levels of arsenic-contaminated soils. BMC Microbiol 9(1), 1-11.
74. Chen SC, Sun GX, Yan Y, Konstantinidis KT, Zhang SY, Deng Y, Li XM, Cui HL, Musat F, Popp D, Rosen B, Zhu YG. 2020. The Great Oxidation Event expanded the genetic repertoire of arsenic metabolism and cycling. PNAS 117(19), 10414-10421.
75. Delmont TO, Quince C, Shaiber A, Esen ÖC, Lee ST, Rappé MS, McLellan S, Lücker S, Eren AM. 2018. Nitrogen-fixing populations of Planctomycetes and Proteobacteria are abundant in surface ocean metagenomes. Nat Microbiol 3(7), 804-813.
76. Leema JM, Kirubagaran R, Vinithkumar NV, Dheenan PS, Karthikayulu S. 2010. High value pigment production from *Arthrospira* (Spirulina) platensis cultured in seawater. Bioresour Technol 101(23), 9221-9227.
77. Sorokin DY, Muntyan MS, Panteleeva AN, Muyzer G. 2012. *Thioalkalivibrio sulfidiphilus* sp. nov., a haloalkaliphilic, sulfur-oxidizing gammaproteobacterium from alkaline habitats. Int J Syst Evol Microbiol 62(8), 1884-1889.
78. Purcell AM, Mikucki JA, Achberger AM, Alekhina IA, Barbante C, Christner BC, Ghosh D, Michaud A, Mitchell A, Priscu J, Scherer R, Skidmore M, Vick-Majors TJ. 2014. Microbial sulfur transformations in sediments from Subglacial Lake Whillans. Front Microbiol 5, 594.
79. Zargar K, Hoeft S, Oremland R, Saltikov CW. 2010. Identification of a novel arsenite oxidase gene, *arxA*, in the haloalkaliphilic, arsenite-oxidizing bacterium *Alkalilimnicola ehrlichii* strain MLHE-1. J Bacteriol 192(14), 3755-3762.
80. Andrews S. 2010. FastQC a quality-control tool for high-throughput sequence data http://www. Bioinformaticsbabraham. ac. uk/projects/fastqc.
81. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30(15), 2114–2120.
82. Schmieder R, Edwards R. 2011. Quality control and preprocessing of metagenomic datasets. Bioinformatics 27(6), 863-864.
83. Gruber-Vodicka HR, Seah BK, Pruesse E. 2020. phyloFlash: Rapid Small-Subunit rRNA Profiling and Targeted Assembly from Metagenomes. Msystems, 5(5).
84. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glöckner FO. 2012. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res 41(D1), D590-D596.
85. R Core Team. 2018. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.
86. RStudio Team. 2016. RStudio: Integrated development for R. Boston, MA: RStudio Inc. Retrieved from http://www.rstudio.com/
87. Callahan B, McMurdie P, Rosen M, Han A, Johnson A, Holmes S. 2016. DADA2: High-resolution sample inference from Illumina amplicon data. Nat Methods 13, 581–583.

88. Wang Q, Garrity GM, Tiedje JM, Cole JR. 2007. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. Appl Environ Microbiol 73(16), 5261-5267.
89. Love M, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 15, 550.
90. Wright ES. 2016. Using DECIPHER v2. 0 to analyze big biological sequence data in R. R Journal 8(1).
91. Price MN, Dehal PS, Arkin AP. 2009. FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. Mol Biol Evol 26(7), 1641-1650.
92. McMurdie PJ, Holmes S. 2013. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. PloS one 8(4), e61217.
93. Andersen KS, Kirkegaard RH, Karst SM, Albertsen M. 2018. ampvis2: an R package to analyse and visualise 16S rRNA amplicon data. BioRxiv 299537.
94. Wickham H. 2016. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York.
95. Lahti L, Shetty S, Blake T, Salojarvi J. 2017. Tools for microbiome analysis in R. Version, 1, 10013.
96. Gentleman R, Carey V, Huber W, Hahne F. 2011. Genefilter: Methods for filtering genes from microarray experiments. R package version, 1(0).
97. Kurtz ZD, Müller CL, Miraldi ER, Littman DR, Blaser MJ, Bonneau RA. 2015. Sparse and compositionally robust inference of microbial ecological networks. PLoS Comput Biol 11(5), e1004226.
98. Schloerke B, Crowley J, Cook D, Briatte F, Marbach M, Thoen E, Elberg A, Toomet O, Crowley J, Hofmann H, Wickman H. 2018. GGally: Extension to 'ggplot2'. R package version 1.4.0. https://CRAN.R-project.org/package=GGally.
99. Overbeek R, Olson R, Pusch GD, Olsen GJ, Davis JJ, Disz T, Edwards R, Gerdes S, Parrello B, Shukla M, Vonstein V, Wattam A, Xia F, Stevens R. 2014. The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). Nucleic Acids Res 42(D1), D206-D214.
100. Silva GGZ, Green KT, Dutilh BE, Edwards RA. 2016. SUPER-FOCUS: a tool for agile functional analysis of shotgun metagenomic data. Bioinformatics 32(3), 354-361.
101. Buchfink B, Xie C, Huson DH. 2015. Fast and sensitive protein alignment using DIAMOND. Nat Methods 12(1), 59-60.
102. Parks DH, Tyson GW, Hugenholtz P, Beiko RG. 2014. STAMP: statistical analysis of taxonomic and functional profiles. Bioinformatics 30(21), 3123-3124.
103. Li D, Luo R, Liu CM, Leung CM, Ting HF, Sadakane K, Yamashita H, Lam TW. 2016. MEGAHIT v1.0: a fast and scalable metagenome assembler driven by advanced methodologies and community practices. Methods 102, 3-11.
104. Mikheenko A, Saveliev V, Gurevich A. 2016. MetaQUAST: evaluation of metagenome assemblies. Bioinformatics 32(7), 1088-1090.
105. Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics 30(14), 2068-2069.
106. Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. Nat Methods 9(4), 357.

107. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. 2009. The sequence alignment/map format and SAMtools. Bioinformatics 25(16), 2078-2079.
108. Eren AM, Esen ÖC, Quince C, Vineis JH, Morrison HG, Sogin ML, Delmont TO. 2015. Anvi'o: an advanced analysis and visualization platform for 'omics data. PeerJ 3, e1319.
109. Hyatt D, Chen GL, LoCascio PF, Land ML, Larimer FW, Hauser LJ. 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics 11(1), 119.
110. Eddy SR. 2011. Accelerated profile HMM searches. PLoS Comput Biol 7(10), e1002195.
111. Tatusov RL, Galperin MY, Natale DA, Koonin EV. 2000. The COG database: a tool for genome-scale analysis of protein functions and evolution. Nucleic Acids Res 28(1), 33-36.
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.
113. Anders S, Pyl PT, Huber W. 2015. HTSeq—a Python framework to work with high-throughput sequencing data. Bioinformatics 31(2), 166-169.
114. Aubry S, Kelly S, Kümpers BM, Smith-Unna RD, Hibberd JM. 2014. Deep evolutionary comparison of gene expression identifies parallel recruitment of trans-factors in two independent origins of C4 photosynthesis. PLoS Genet 10(6), e1004365.
115. Kolde R, Kolde MR. 2015. Package 'pheatmap'. R Package 1(7), 790.
116. Katoh K, Misawa K, Kuma KI, Miyata T. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res 30(14), 3059-3066.
117. Alneberg J, Bjarnason BS, De Bruijn I, Schirmer M, Quick J, Ijaz UZ, Lahti L, Loman N, Andersson A, Quince C. 2014. Binning metagenomic contigs by coverage and composition. Nat Methods 11(11), 1144-1146.
118. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. 2015. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res 25(7), 1043-1055.
119. Chaumeil PA, Mussig AJ, Hugenholtz P, Parks DH. 2020. GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. Bioinformatics 36(6), btz848.
120. Parks DH, Chuvochina M, Chaumeil PA, Rinke C, Mussig AJ, Hugenholtz P. 2020. A complete domain-to-species taxonomy for Bacteria and Archaea. Nat Biotech 1-8.
121. Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M. 2016. KEGG as a reference resource for gene and protein annotation. Nucleic Acids Res 44(D1), D457-D462.
122. Aramaki T, Blanc-Mathieu R, Endo H, Ohkubo K, Kanehisa M, Goto S, Ogata H. 2020. KofamKOALA: KEGG ortholog assignment based on profile HMM and adaptive score threshold. Bioinformatics 36(7), 2251-2252.

123. Muto A, Kotera M, Tokimatsu T, Nakagawa Z, Goto S, Kanehisa M. 2013. Modular architecture of metabolic pathways revealed by conserved sequences of reactions. J Chem Inf Model 53(3), 613-622.
124. Bastian M, Heymann S, Jacomy M. 2009. Gephi: an open-source software for exploring and manipulating networks. ICWSM 2, 361–362.
125. Bowers RM, Kyrpides NC, Stepanauskas R, Harmon-Smith M, Doud D, Reddy TBK, Schulz F, Jarett J, Rivers A, Eloe-Fadrosh EA, Tringe SG, Ivanova NN, Copeland A, Clum A, Becraft ED, Malmstrom RR, Birre B, Podar M, Bork P, Weinstock GM, Garrity G, Dodsworth JA, Yooseph S, Sutton G, Glöckner F, Gilbert J, Nelson W, Hallam S, Jungbluth SP, Ettema TJ, Tighe S, Konstantinidis KT, Liu WT, Baker BJ, Rattei T, Eisen JA, Hedlund B, McMahon KD, Fierer N, Knight R, Finn R, Cochrane G, Karsch-Mizrachi I, Tyson GW, Rinke C, Lapidus A, Meyer F, Yilmaz P, Parks D, Eren AM, Schriml L, Hugenholtz P, Woyke T. 2017. Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. Nat Biotechnol 35(8), 725-731.

SUPPLEMENTARY MATERIAL

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

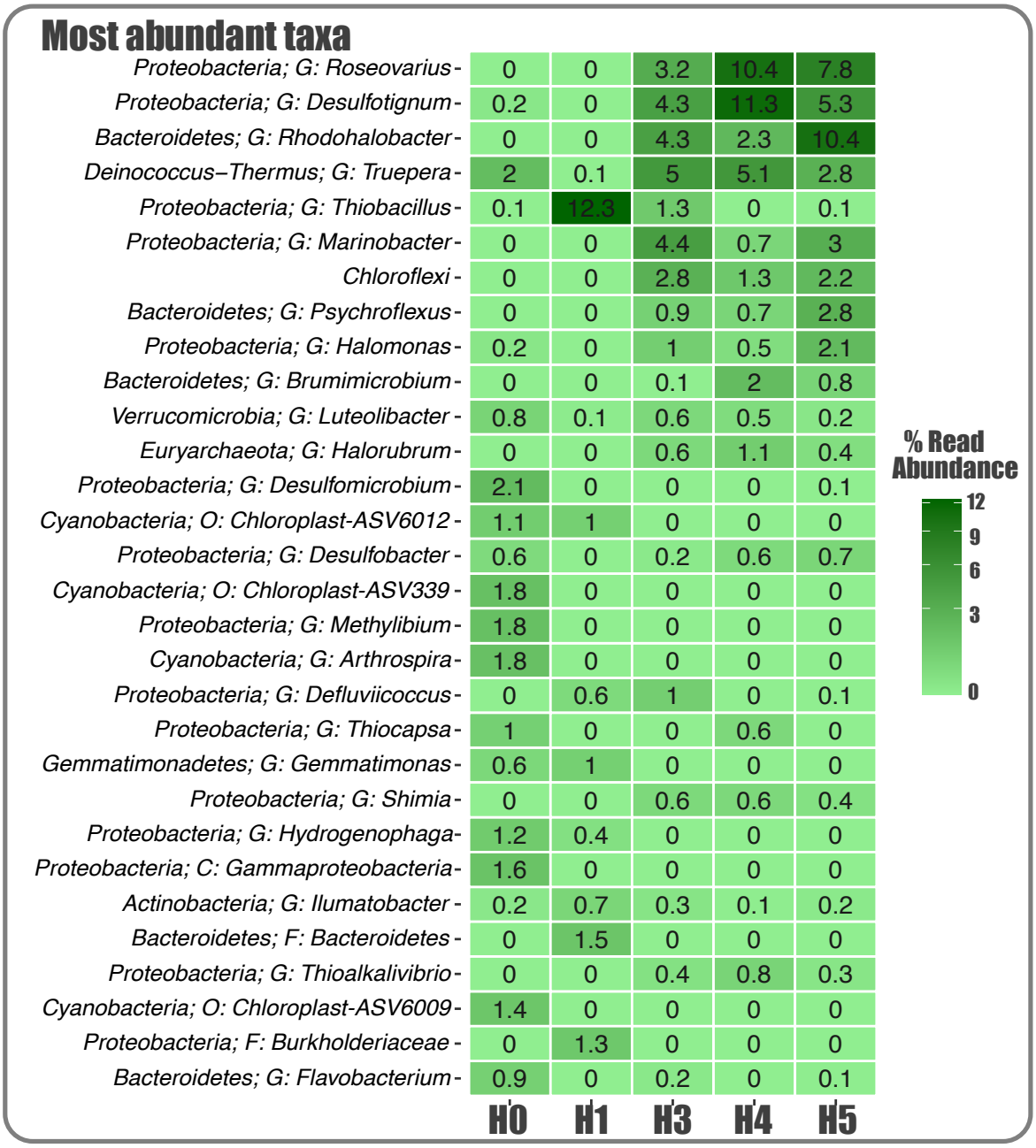
Supplementary Table S2. Statistical values of the SH co-assembly, representing the five metagenomes.

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.

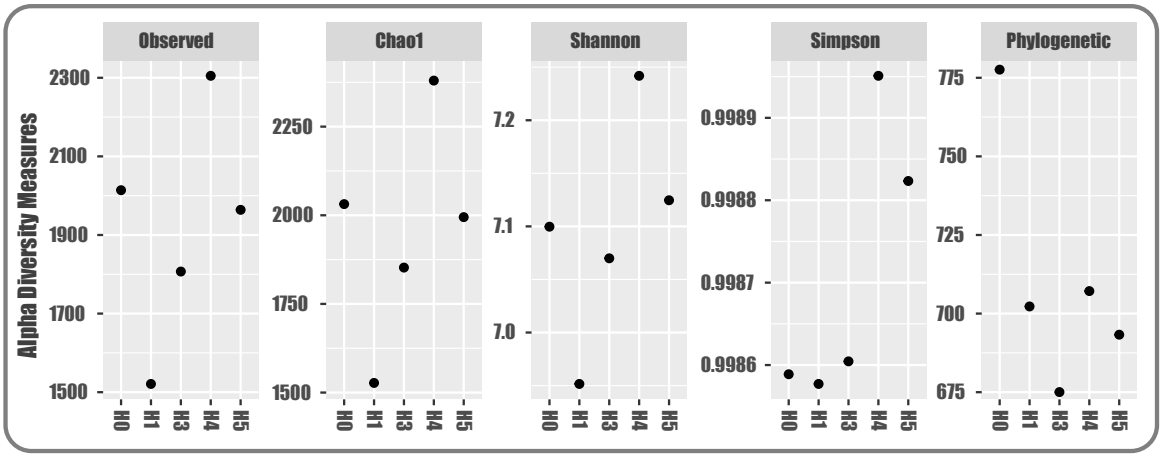
Supplementary Table S4. Indexes of relative abundance and detection of each recovered MAG across the five SH metagenomes.

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

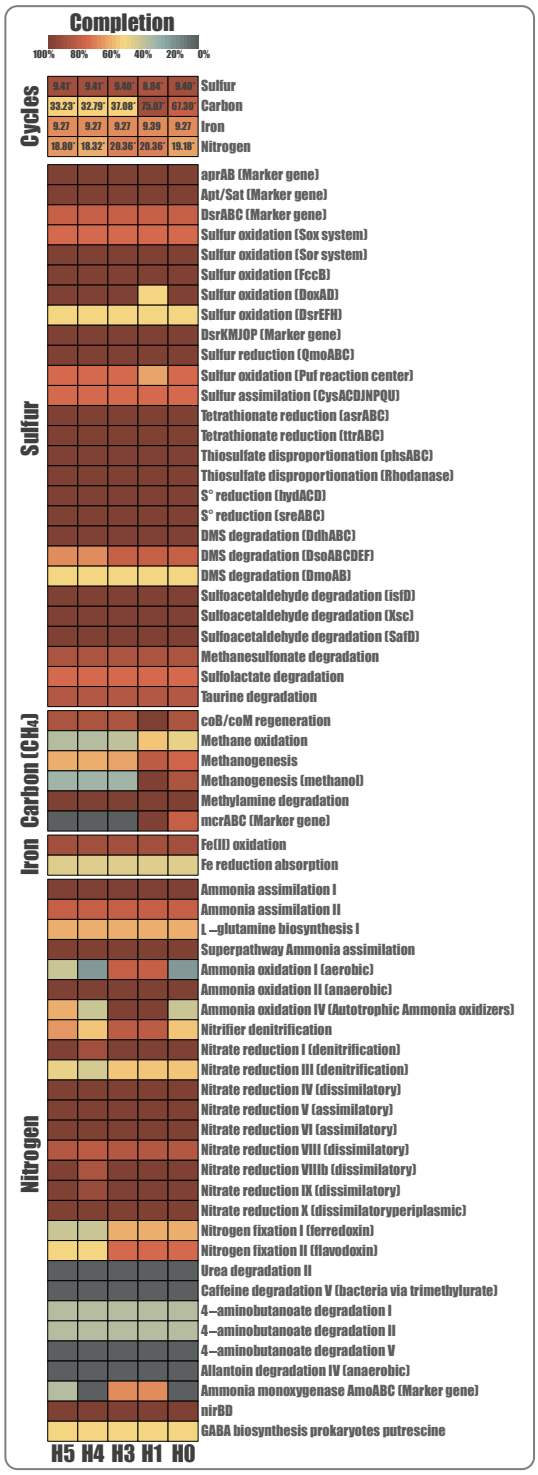
Supplementary Table S6. Presence and copy number of arsenic resistance related genes in each analyzed MAG.



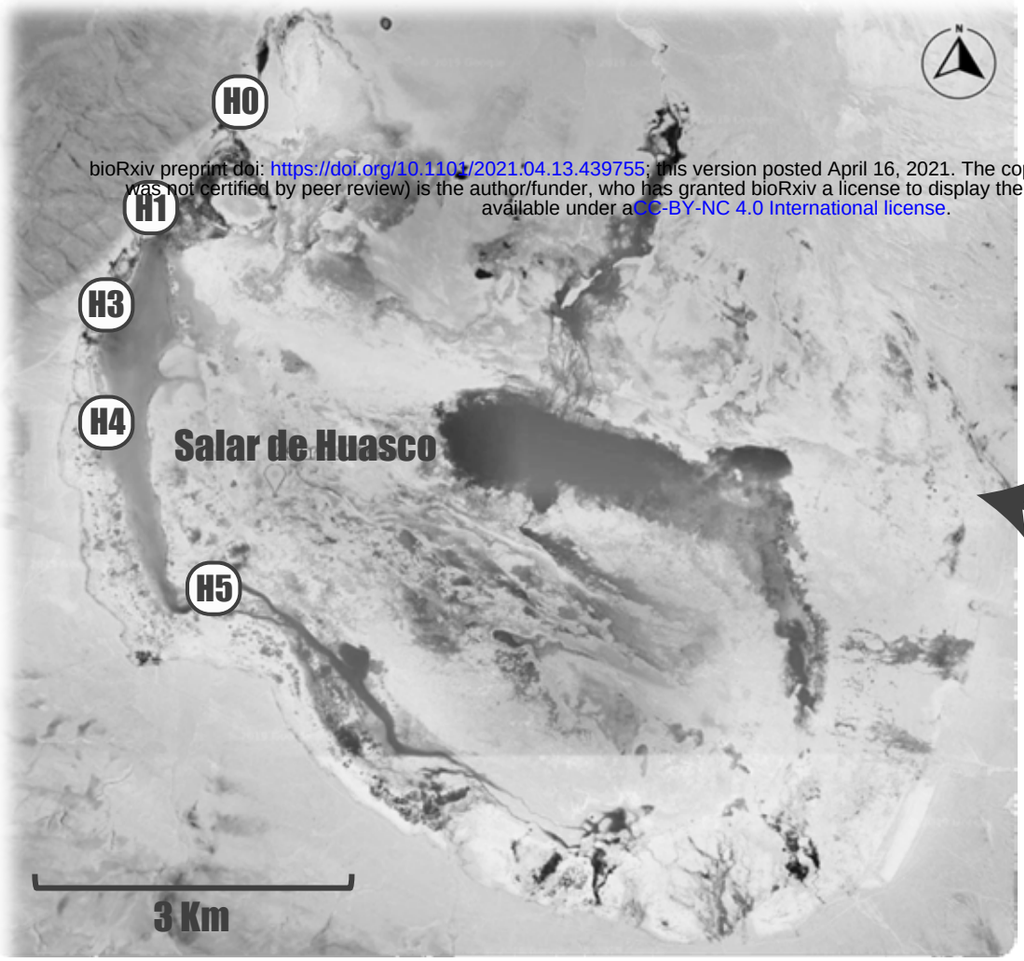
Supplementary Figure S1. Composition and structure of SH bacterial communities. Heatmap showing the 30 most abundant ASVs (bacterial lineages). Taxonomic is showed at the phylum rank plus the best hit available (C: class; O: order; F: family; G: genus).

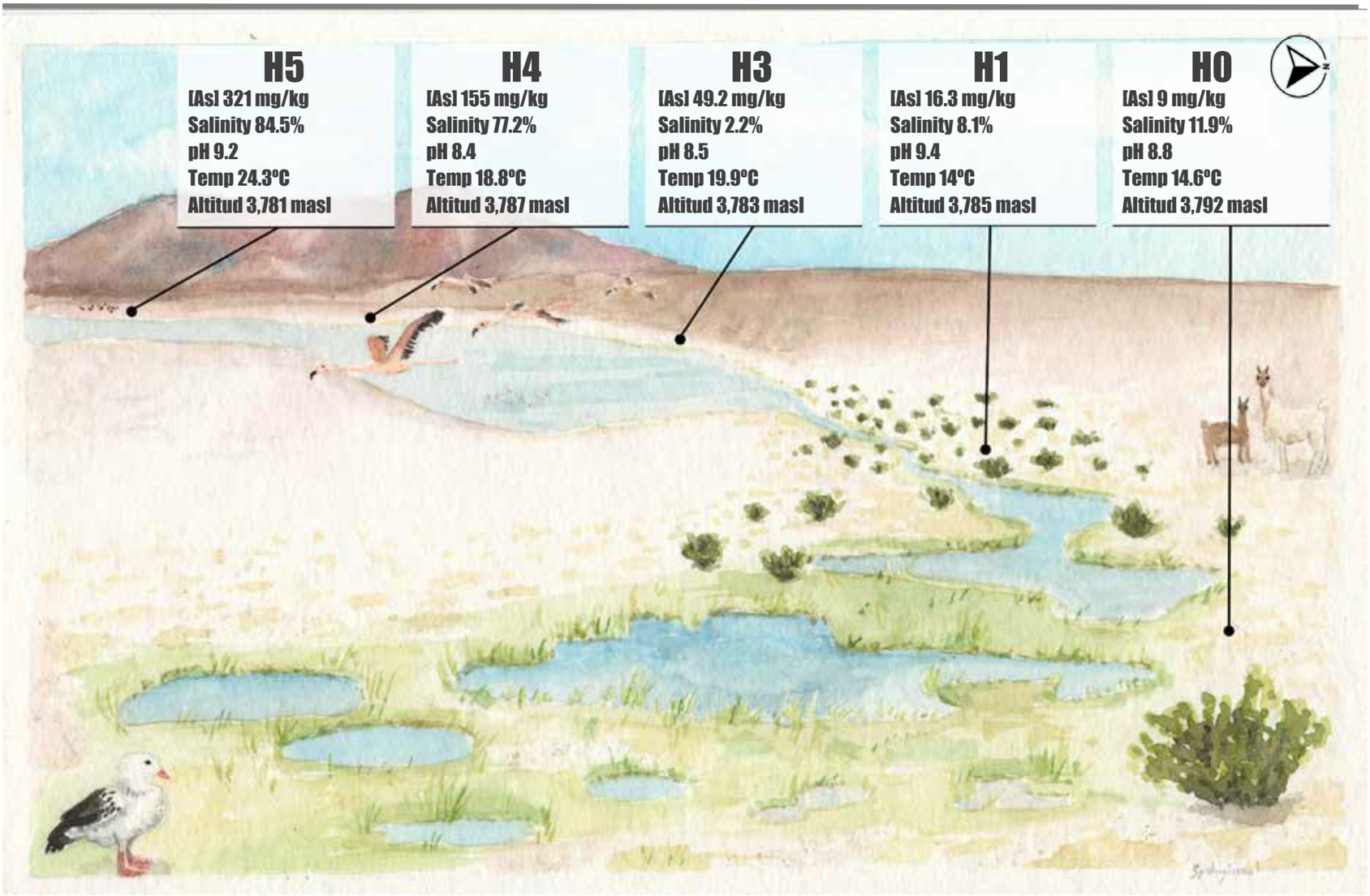


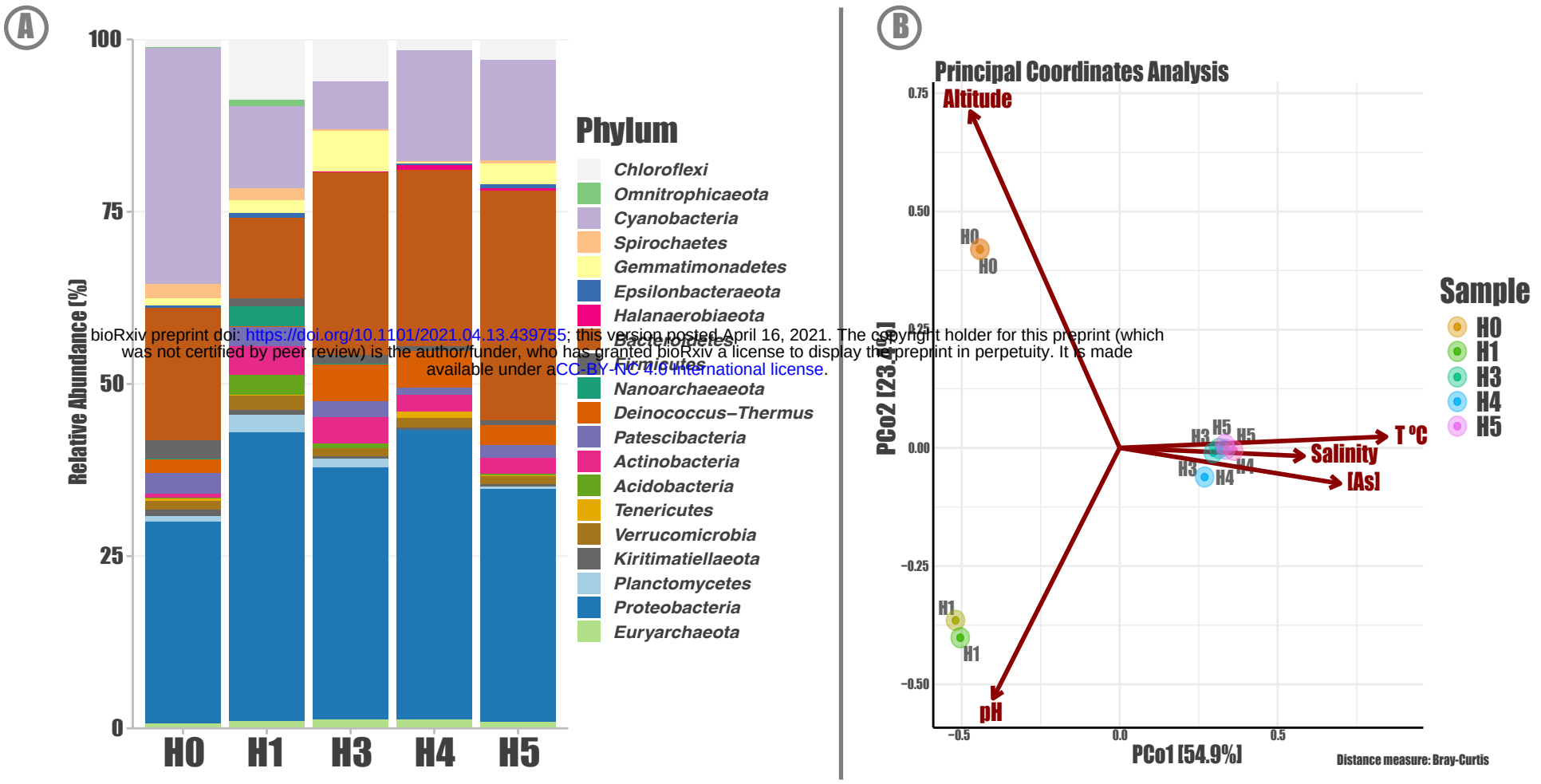
Supplementary Figure S2. Alpha diversity indices for the five SH communities.

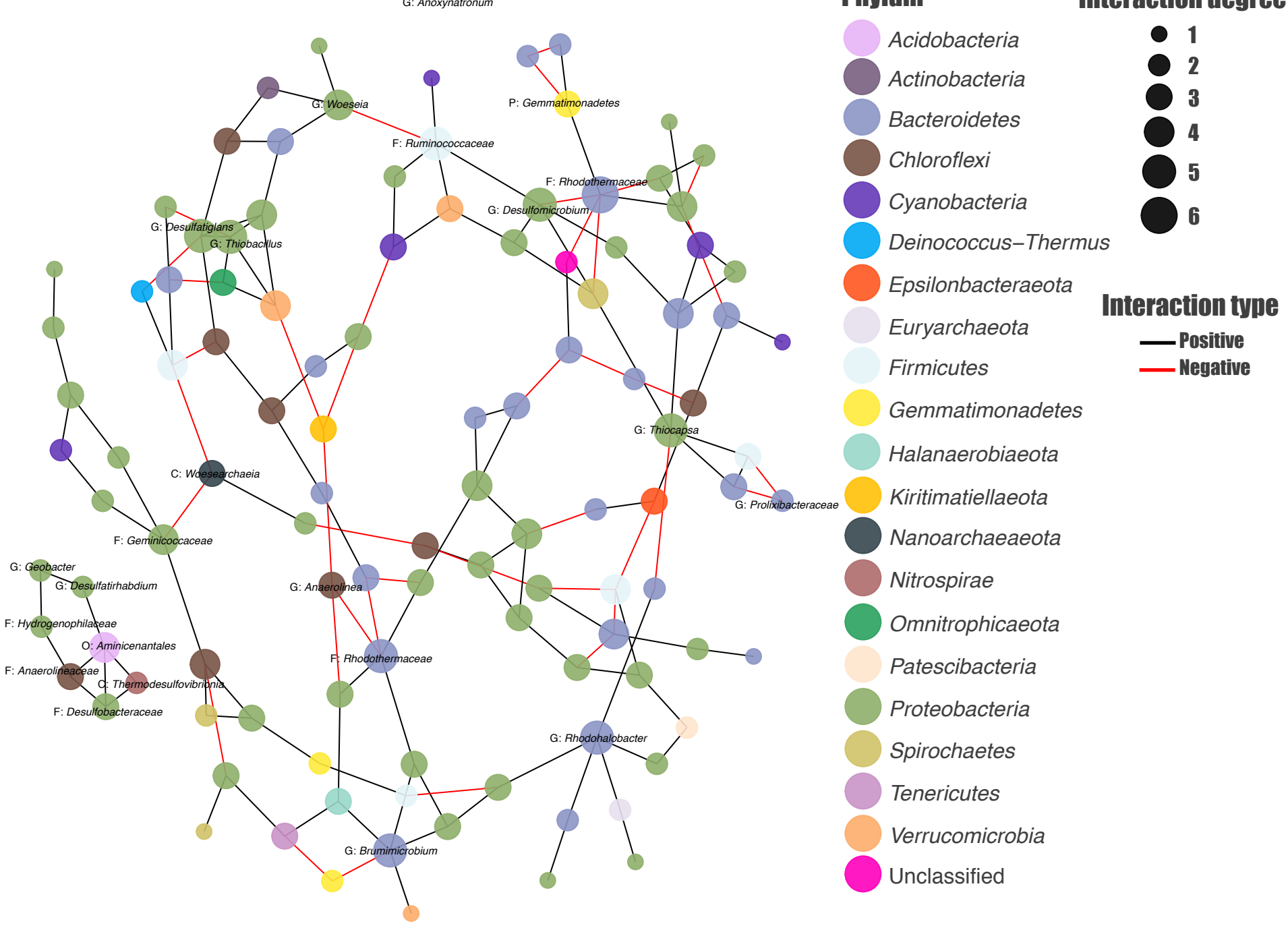


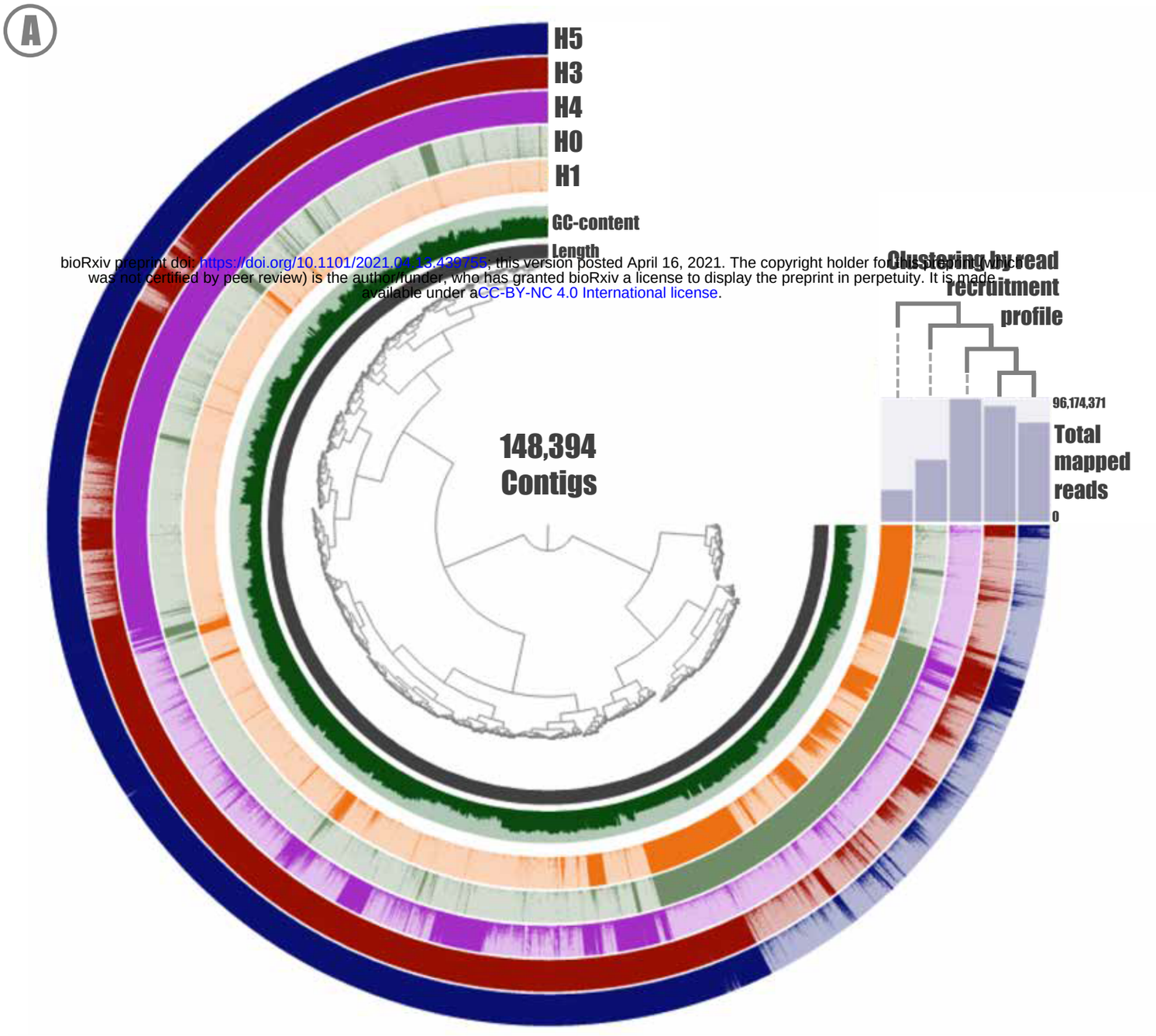
Supplementary Figure S3. MEBS analysis heatmap displaying the completeness of N, Fe, S and CH_4 pathways: as a whole (first top section) and particular reactions (bottom four sections). The color gradient shows the percentage of completition for each pathway (from lowest to highest) and the values at the top section represent the corresponding MEBS score (* FDR ≤ 0.01).

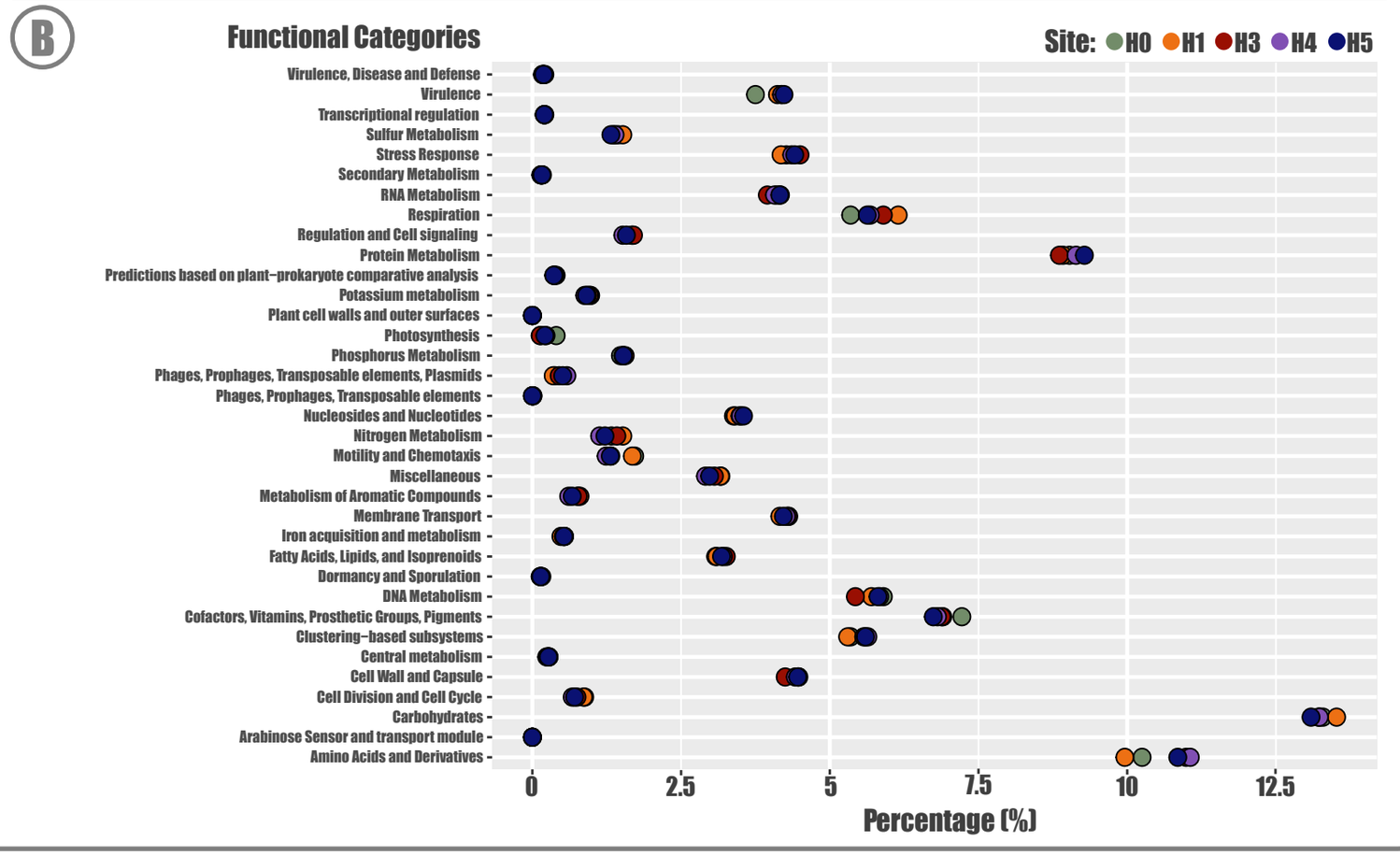








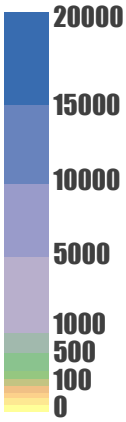




Arsenic Metabolism Genes

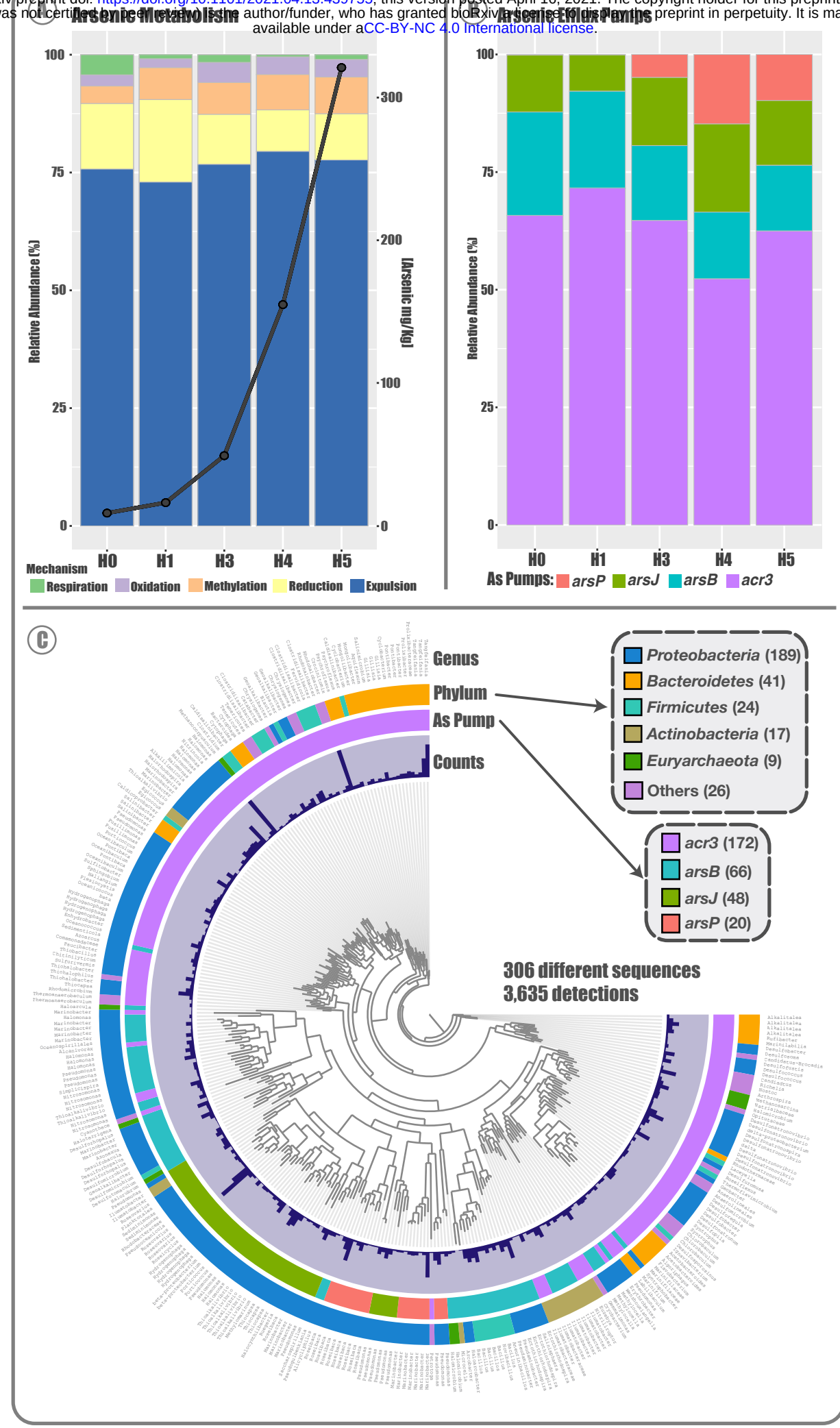
	H0	H1	H3	H4	H5
arsR	6142.15	4214.04	12356.33	18499.32	13750.63
arsC	1327.42	1188.22	2117.04	2345.06	2057.75
acr3	1099.45	698.09	2375.33	2003.96	2183.94
arsM	427.13	495.38	1491.54	2215.36	1827.64
arsB	367.57	200.46	583.55	542.77	487.89
arsD	259.69	57.08	534.59	682.07	600.08
arsJ	202.04	75.23	531.96	717.95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

Normalized reads



Mechanism

- Expulsion
- Reduction
- Respiration
- Oxidation
- Methylation



A

		Percent Completion	Percent Redundancy	Num Contigs	N50	GC content	Percent Completion	Percent Contamination	Phylum	Class	Order	Family	Genus
*	SH-MAG111	95.77%	2.82%	332	11,127	60.76%	81.95%	1.54%	Bacteroidota	---	---	---	---
*	SH-MAG116	90.14%	0.00%	208	14,823	43.49%	83.04%	0.00%	Bacteroidota	Bacteroidia	Chitinophagales	Saprospiraceae	
*|**	SH-MAG12	95.77%	2.82%	413	11,471	54.64%	92.06%	3.28%	Bacteroidota	---	---	---	---
*	SH-MAG130	92.96%	9.86%	395	10,529	44.80%	93.50%	11.30%	Bacteroidota	Bacteroidia	Bacteroidales	---	---
	SH-MAG192	85.92%	4.23%	314	4,317	36.06%	68.88%	4.52%	Bacteroidota	Ignavibacteria	Ignavibacteriales	Melioribacteraceae	---
**	SH-MAG193	85.92%	1.41%	561	10,013	44.34%	92.35%	3.33%	Bacteroidota	Bacteroidia	Cytophagales	Cyclobacteriaceae	---
	SH-MAG96	85.92%	1.41%	234	14,512	47.74%	80.17%	1.49%	Bacteroidota	Bacteroidia	Chitinophagales	Saprospiraceae	---
*|**	SH-MAG29	95.77%	5.63%	346	20,718	44.60%	97.71%	0.66%	Cyanobacteria	Cyanobacteriia	Cyanobacteriales	Phormidiaceae	Arthrospira
*	SH-MAG3	94.37%	2.82%	436	10,739	66.26%	86.11%	6.17%	Gemmatimonadota	Gemmatimonadetes	---	---	---
	SH-MAG89	88.73%	9.86%	498	14,477	62.15%	82.81%	4.83%	Proteobacteria	Deltaproteobacteria	Bradymonadales	---	---
	SH-MAG131	80.28%	1.41%	626	5,244	63.23%	77.20%	5.18%	Proteobacteria	Deltaproteobacteria	Desulfobacterales	Desulfosarcinaceae	---
*	SH-MAG143	95.77%	8.45%	601	7,367	58.76%	84.68%	2.47%	Proteobacteria	Deltaproteobacteria	Desulfobacterales	Desulfosarcinaceae	Desulfatitalea
*	SH-MAG148	95.77%	4.23%	384	7,277	55.17%	87.42%	0.86%	Proteobacteria	Deltaproteobacteria	Desulfobacterales	---	---
*|**	SH-MAG93	97.18%	5.63%	550	9,043	56.28%	94.35%	3.10%	Proteobacteria	Deltaproteobacteria	Desulfobacterales	Desulfosarcinaceae	Desulfosarcina
*|**	SH-MAG129	92.96%	2.82%	390	14,010	58.58%	91.11%	3.40%	Proteobacteria	Gammaproteobacteria	---	---	---
*	SH-MAG144	95.77%	1.41%	315	8,476	67.34%	87.51%	1.67%	Proteobacteria	Gammaproteobacteria	Ectothiorhodospirales	Thioalkalivibrionaceae	Thioalkalivibrio
	SH-MAG162	87.32%	1.41%	431	8,147	69.40%	81.81%	3.19%	Proteobacteria	Gammaproteobacteria	Pseudomonadales	---	---
	SH-MAG78	80.28%	5.63%	371	6,424	67.34%	73.01%	2.21%	Proteobacteria	Gammaproteobacteria	---	---	---
	SH-MAG169	85.54%	9.64%	4176	8,047	42.02%	72.44%	39.84%	---	---	---	---	---

B