With a pinch of salt: metagenomic insights into Namib Desert salt pan microbial mats and halites reveal functionally adapted and competitive communities

Martínez-Alvarez L[1], Ramond J-B[1,3], Vikram S[1], León-Sobrino C[1], Maggs-Kölling G[2], Cowan DA[1]

Centre for Microbial Ecology and Genomics (CMEG), Genomics Research Institute, Department of Biochemistry, Genetics and Microbiology, University of Pretoria, Pretoria, South Africa[1]. Gobabeb Namib Research Institute, Walvis Bay, Namibia[2]; Departamento Genética Molecular y Microbiología – Pontificia Universidad Católica de Chile – Chile[3].

Running title: Metagenomics of Namib Desert salt pan mats and halites

#Address correspondence to Laura Martínez-Alvarez, laura.martinez@bio.ku.dk or Don A. Cowan, don.cowan@up.ac.za

Abstract

Playas, saline-rich springs surrounded by halite evaporates (also called salt pans) have played an essential role in landscape erosion during the formation of the Namib Desert and are numerous in its central region. In this study, we used shotgun metagenomics to investigate the phylogenetic and functional capacities of the microbial communities in two salt pans (namely, Eisefeld and Hosabes) located in the Central Namib Desert. We studied the source and sink sediment mat communities of the saline streams, as well as those from two halites (crystallized structures on the stream margins). The microbial assemblages and potential functions were distinct in both niches. Independently from their localization (Eisfeld vs Hosabes and source vs sink), the sediment mat communities were dominated by members of the *Alphaproteobacteria* and *Gammaproteobacteria* classes, while halites were Archaea-dominated (*Euryarchaea* phylum) and also contained high abundances of the extremely halophilic bacterium *Salinibacter* sp. (phylum *Bacteroidetes*). Photoheterotrophy and chemoheterotrophy were the principal lifestyles in both niches, with halite communities having a reduced diversity of metabolic pathways. Intense microbial-virus interactions in both niches were implied by the widespread detection of CRISPR-Cas defense systems. We identified a putatively novel clade of type II CRISPR-Cas systems, as well as novel candidate viral lineages of the order *Caudovirales* and of *Euryarchaea*-infecting haloviruses. Putative gene transfer agent-like sequences within the *Alphaproteobacteria* were identified in the sediment mat communities. These horizontal gene transfer elements have the potential to drive genome plasticity and evolution of the *Alphaproteobacteria* in the Namib Desert salt pan microbiomes.

Importance

The hyperarid Namib Desert has is one of the oldest deserts on Earth. It contains multiple clusters of playas which are saline-rich springs surrounded by halite evaporites. Playas are of great ecological importance and the indigenous microorganisms that inhabit them are potentially involved in precipitation of minerals such as carbonates and sulfates. While there has been a considerable amount of research on the diversity and ecology of the microbiomes in the different edaphic niches of the Namib Desert, little is known about the microbial communities inhabiting its multiple playas. In this work, we provide a comprehensive taxonomic and functional characterization of the microbial and viral communities of sediment mats and halites from two salt pans in the central region of the Namib Desert, contributing towards the understanding of the ecology of this biome.

Introduction

Saline inland waters account for 5% of dryland surfaces globally (Bryant and Frigaard, 1996) and represent approximately 0.008% of the world's water. This is almost equivalent to the total amount of freshwater, estimated to be of 0.009% (Waiser and Robarts, 2009). Among these different saline inland waters, salt pans or playas are terrain depressions occurring frequently in arid ecosystems where underground water surfaces and evaporates, leading to the formation of a salt-crust over the ground sediment (Eckardt and Drake, 2010). The Namib Desert is one of the oldest deserts in the world, estimated to have been hyper-arid for the last 5 million years (Ward, 2010), and is characterized by the presence of numerous playas (Eckardt and Drake, 2010). Although salt pans occupy less than 5% of the central Namib Desert gravel plains, they are a major water source for the desert fauna and play an important role in the Namib Desert's geomorphology via gypsum ($CaSO_4*2H_20$) deposition and landscape erosion through salt weathering (Day and Seely, 2004; Eckardt and Drake, 2010; Eckardt et al., 2001). Furthermore, they produce some of the most saline inland waters in southern Africa, reaching up to 160 g/L (Day, 1993; Day and Seely, 2004).

The microbial diversity of salt pans worldwide is dependent on their particular geochemical characteristics, including salinity, pH and oxygen levels. Globally, microbial diversity decreases with increasing salinity, and it is accompanied by increasing proportions of *Euryarchaea* of the order *Halobacteriales* (Benlloch et al., 2002; Fernandez et al., 2016; Vera-Gargallo et al., 2019). In such environments, microbial communities develop into microbial mats with a vertical-layered structure in which each layer harbors different microorganisms with distinct metabolic capacities (Fourçans et al., 2006). The taxonomic diversity of saline microbial mat communities fluctuates in response to the constant changes in their surrounding physicochemical conditions; particularly in response to salinity, oxygen levels and the metabolic activity of the community (Dupraz and Visscher, 2005). While many studies have addressed the taxonomic diversity of saline microbial mats (Benlloch et al., 2002; Fernandez et al., 2016; Vera-Gargallo et al., 2019), their functional capacities and the ecological roles of their viral communities remain largely unexplored. Nevertheless, hypersaline environments are known to be rich in novel viruses (e.g., (Adriaenssens et al., 2016; Crits-Christoph et al., 2016; Santos et al., 2012; Sime-Ngando, 2014)) that are likely to modulate microbial community compositions and biogeochemical cycling functions (Brum and Sullivan, 2015). Although previous studies have reported the viral diversity of some edaphic niches of the Namib Desert, showing a wealth of novel viruses (Adriaenssens et al., 2015, 2016; Hesse et al., 2017; Zablocki et al., 2017), little information on virus-host pairs is available.

In order to investigate the microbial and viral communities in desert saline springs and associated saline niches, we investigated ten shotgun metagenomes from microbial mats and halites obtained from two central Namib Desert playas belonging to two different saline springs clusters (Day, 1993; Eckardt and Drake, 2010; Eckardt et al., 2001). We noted that mat microbial and viral communities from both salt pans were highly similar in their taxonomic distribution and functional potential but, not surprisingly, highly distinct from non-saline water-limited edaphic niches in the Namib Desert (e.g. hypolith, gravel plain soil, sand dune). Conversely, crystalline halites were dominated by taxa which are typically adapted to hypersaline conditions and low water availability, with relatively lower phylogenetic diversity. Novel viral taxa dominated both mat and halite communities, with a higher phylogenetic diversity in the former being consistent with the greater microbial diversity of mat microbiomes. Analyses of the defense systems used by the community against mobile genetic elements revealed abundant type I and type III CRISPR-Cas systems, as well as putative novel subtype of type II systems. Additionally, a cluster of gene transfer agents with the potential to mediate horizontal gene transfer events was identified in the mat communities. Overall, we suggest that the saline spring microbial mat and halite niches represent 'hotspots' of microbial

and viral diversity, characterized by diverse functional capacities, high inter-taxon competition and high capacities for genomic evolution and adaptation.

Results

Eight metagenomes were generated from microbial salt pan mats collected from the source and sink of the water stream in the Namib Desert Hosabes and Eisfeld playas. Additionally, two more metagenomes were produced from crystalline halites collected in the vicinity of the Hosabes playa stream; named as "dark"- and "red"-halite due to the surface color of the rock (Figure 1). The 10 metagenomes comprised between 1.04 (Hosabes source 2016) and 9.57 (Red halite) Gbp of sequencing data, which, after data processing and assembly, resulted in 4 527 165 contigs over 500 bp that were retained for further analyses (Supplementary Table 1). Nonpareil analyses clearly showed that that the sequence depth was high, covering between 66% (Eisfeld source 2017) to 90 % (Hosabes sink 2017) of the microbial communities in all samples (Supplementary Figure 1).

Niche-specific microbial assemblages in Namib Desert hypersaline environments

Around 65.5% of the coding sequences predicted from the metagenomic data were taxonomically assigned at Order level (Supplementary Table 1) and were used to profile the diversity of the microbial communities. Halite and mat metagenomes displayed clear distinct taxonomic composition as shown by their clear separation on PCA plots (Figure 2A). The variance between Eisfeld and Hosabes mat metagenomes was comparable to the variance between the source and the sink communities (Figure 2A), with an average 5% difference in the taxonomic diversity. Furthermore, both halite communities were more dissimilar than those of the stream mats (independently from year and sample type [source vs sink]).

The halite microbial communities were clearly Archaea-dominated when compared to the stream mats (Figure 2B). *Euryarchaeota* represented 55.4% of the dark halite. The *Halobacteriales* order dominated the archaeal fraction, with relative abundances ranging from 30.3% (red halite) to 55.1% (dark halite) (Figure 2B and Supplementary Table 5). The halite bacterial fraction was also dominated by salt-tolerant/halophilic genera which were not abundant in the saline mat assemblages, particularly *Salinibacter sp.* (Bacteroidetes phylum), *Halothece* sp. and *Dactylococcopsis sp*. (*Cyanobacteria* phylum; Supplementary Table 5). We highlight a 60-fold enrichment of the *Bacteroidetes Rhodothermaceae,* the *Cyanobacteria Aphanothecaceae* and the *Gammaproteobacteria Wenzhouxiangellaceae* in the halite bacterial fraction in comparison to the mat assemblages (Table 1).

Stream mat microbial communities comprised a total of 11 bacterial phyla and were dominated by members of the Proteobacteria phylum, and particularly of the Alpha- and *Gamma-Proteobacteria* classes with relative abundances of 14.4-26.8% and 12.8-21.8%, respectively. The other dominant mat bacterial phyla were *Bacteroidetes* (5.1-10%), *Cyanobacteria* (2-13%) and *Planctomycetes* (3.4-11.8%). *Actinobacteria*, *Verrucomicrobia*, *Firmicutes*, *Balneolaeota* and *Acidobacteria* comprised 1% to 3.6% (Figure 2B). We noted that the *Cyanobacteria Aphanothecaceae* and *Leptolyngbyaceae* families, and the alphaproteobacterial *Methylocystaceae* were slightly enriched in Hosabes salt pan metagenomes (4-fold), while the Eisfeld salt pan metagenomes were enriched in the cyanobacterial *Coleofasciculaceae* and *Alphaproteobacteria Hyphomicrobiaceae* family members (3-4 fold) (Supplementary Table 3). Furthermore, members of the alphaproteobacterial *Rhodobacteraceae, Methylocystaceae* and *Parvularculaceae* taxa, and of the *Bacteroidetes Flavobacteriaceae families,* were 2-3 fold richer in mat sources, and members of

the the *Nitrospinae, Hydrogenedentes* and *Balneolaeota* phyla were 2-3 fold more abundant in mat sink communities (Supplementary Table 4).

Altogether, our results clearly show that, despite a separation of over 120 km and belonging to 2 different stream mat "clusters" (Figure 1), the Eisfeld and Hosabes mat communities are highly similar. Conversely, those of the dark and red halite bacterial communities, which were sited only 50 m apart, were very different (Figure 2).

The functional capacities of the Namib Desert saline communities differ in halite and salt pans stream mats

An average of 39 % of the metagenomic open reading frames (ORFs) could be assigned to KEGG Ortholog (KO) terms (Supplementary Table 1) with around 35% belonging to the metabolism category, and predominantly to the amino acid and carbohydrate metabolism subcategories (Supplementary Figure 2). In the environmental information processing category genes from the membrane transport subcategory stand out, accounting for approximately 0.1% of the total ORFs (Supplementary Figure 2). Genes related to signal transduction and glycan biosynthesis were one-third to half less abundant in the halite than in mat communities (Supplementary Figure 2).

Transport of osmoprotectant solutes (glycine/betaine/proline transport [M00208], osmoprotectant and polyamine transport systems [M00209]) orthologs were widespread in the stream mat metagenomes and encoded in sequences belonging to *Alphaproteobacteria, Gammaproteobacteria, Bacteroidetes, Cyanobacteria, Haloferacales*, and *Detaproteobacteria* (Supplementary Table 6). This suggests these organisms employ a "salt-out" strategy to balance osmotic stress and therefore can withstand a range of salinities (Oren, 2008). By contrast, the absence of these systems in the *Halobacteria* and *Planctomycetes* suggests that these taxa rather employ a "salt-in" strategy (Oren, 2008). As expected for biofilm communities, transport systems for lipopolysaccharide and other capsular polysaccharides were 3-8 times more abundant in the mat metagenomes than in the halite ones (Supplementary Table 6).

The nutrient (C, N, S) biogeochemical cycling capacities were also very different in both niches (Figure 3): Anaerobic C fixation, nitrification, denitrification, dissimilatory nitrate reductions, dinitrogen fixation and dissimilatory sulfate reduction (a form of anaerobic respiration performed by chemoorganoheterotrophic microbes; (Qian et al., 2019)) were exclusively found in the mat metagenomes.

Capacity for carbon fixation via the Calvin cycle was present in the mat and halite *Cyanobacteria* and *Alphaproteobacteria*, and less abundantly in members of the *Betaproteobacteria*. Photosystem I and II modules were complete for the *Cyanobacteria* in both stream mats and halites, and this was the only phylum with capacity of oxygenic photosynthesis. Nevertheless, phototrophy was widespread in *Alphaproteobacteria,* where the anoxygenic photosystem II module was complete in mat metagenomes (Figure 3A and Supplementary Table 6). Members of the mat *Gammaproteobacteria* and *Betaproteobacteria* may also have the functional capacity for anoxygenic phototrophy since the *pufLM* genes coding for subunits of the photosynthetic reaction centre were found in these taxa (see Supplementary Table 7). Evidence for the presence of the anaerobic Arnon-Buchanan and Wood-Ljungdahl pathways of carbon fixation was restricted to the salt pan *Deltaproteobacteria* and *Planctomycetes*, respectively (Figure 3A). Additionally, the capacity to obtain energy from CO oxidation (carboxydovory) was widespread in both mat and halite communities, implying that the use of alternative energy sources may assist these microorganisms to survive under oligotrophic conditions. No evidence of the capacity for methanogenesis was found in any of the metagenomes. Taken together, these data suggest that

Cyanobacteria and *Alphaproteobacteria* as the predominant primary producers in the Namib Desert hypersaline ecosystems.

It is noted that alternatives to the TCA cycle, which is thought to be employed to avoid carbon loss (Cronan and Laporte, 2005), were widespread in all metagenomic datasets. The glyoxylate cycle was found in *Alphaproteobacteria, Gammaproteobacteria, Cyanobacteria, Bacteroidetes* and *Halobacteriales*. The ethylmalonyl pathway was found exclusively in the salt pan *Alphaproteobacteria*, an expected finding given the high abundance of the family *Rhodobacteraceae* (Peyraud et al., 2009). The archaeal methylaspartate cycle, which has been described as an adaptation to halophilic conditions and carbon starvation (Borjian et al., 2016) was detected in the archaeal-dominated halite metagenomes, concurrent with the high presence of gene *phaC,* coding for the polyhydroxyalkanoate (PHA) synthase, in the *Halobacteraceae* and *Haloferacaceae*, as well as the *Alphaproteobacteria* (Supplementary Table 7). PHAs are major storage compounds in prokaryotes, and as the ethylmalonyl pathway is interrelated with the synthesis of PHAs, this cycle together with the methylaspartate pathway is linked to the capacity to adapt to environmental stresses (Petushkova et al., 2021).

Only *Cyanobacteria* from the stream mat communities showed the potential capacity to fix atmospheric nitrogen, albeit the genetic markers of this capacity (i.e. the *nif* genes) were in low abundance in the metagenomes. Assimilatory nitrate reduction capacity was found predominantly in the archaeal *Halobacteriales* and *Haloferacales* orders in halites, and in *Planctomycetes*, *Cyanobacteria* and *Alphaproteobacteria* in all datasets (Figure 3B). This step was the only inorganic nitrogen incorporation reaction detected in the halite microbial community, suggesting an incomplete nitrogen cycle in these communities. Complete modules for this pathway were detected in all main phyla *Alphaproteobacteria*, *Gammaproteobacteria*, *Planctomycetes*, *Bacteroidetes* and *Deltaproteobacteria* (Figure 3B). Metabolic capacity for denitrification was found in the salt pan *Alphaproteobacteria, Gammaproteobacteria* and *Bacteroidetes* (Figure 3B). Unlike the dissimilatory nitrate reduction, denitrification does not conserve nitrogen in the system, which is lost as volatile nitrogen forms (N_2). Some capacity for nitrification was detected for the salt pan *Betaproteobacteria Nitrosomonadales, Planctomycetes,* the *Gammaproteobacteria Methylococcales, Chromatiales* and *Oceanospirillales,* and the *Alphaproteobacteria Rhizobiales* and *Rhodobacterales* (Supplementary Table 7). No evidence of anaerobic ammonium oxidation (annamox) was detected in neither the salt pan nor the halite samples.

Assimilatory sulfate reduction was the only sulfur incorporation step identified in the halite communities. Conversely, salt pan *Gammaproteobacteria* possessed the capacity for anaerobic respiration through the dissimilatory sulfate reduction pathway. Similarly, the presence of the genes for thiosulfate oxidation via the SOX complex suggested a capacity for chemolithoautotrophy for the salt pan *Alpha-* and *Gammaproteobacteria* (Figure 3C). These results suggest that the halite microbial communities possess a simpler sulfur cycle than mat communities, reliant on environmental sulfate assimilation, whereas a few orders of the *Alpha- and Gamma-proteobacteria* in the salt pans are potentially capable of using sulfate as electron acceptor for respiration.

Defense mechanisms against mobile genetic elements are abundant, diverse and novel in Namib Desert saline microbial community metagenomes

To gain further insights into the dynamics of virus-host interactions in the Namib salt pan and halite microbial communities, we assessed the presence of KO terms of the category of Prokaryotic Defense Systems in our metagenomic data. An average of 0.94 % of the total KO counts belonged to this category,

from which the majority belonged to toxin-antitoxin systems (42.7 %), followed by restriction-modification systems (27.75 %), CRISPR-Cas systems (23.1 %) and DNA phosphorothioation systems (6.41 %) (Supplementary Table 8).

CRISPR-Cas adaptive immune systems consist of an array of short sequences (spacers) originating from mobile genetic elements and CRISPR-associated (Cas) proteins required for the acquisition and utilization of spacer sequences and targeting of the invading mobile genetic element (Marraffini, 2015). Given that the discovery of new CRISPR-Cas systems with unique capabilities is important for the development of new tools with biotechnological application (Donohue, 18), we further investigated the diversity of the salt pan CRISPR-Cas systems. The identification of CRISPR-Cas systems in the Namib salt pan and halite metagenomes (Supplementary Table 9) revealed that type III (which can target both DNA and RNA) and type I CRISPR systems (which specifically degrade DNA) were dominant; representing 54% and 18% of all identified CRISPR systems, respectively (Figure 4A). Type II CRISPR systems (which target DNA) also represented 10.9 % of the dataset. We further note that 17% of the CRISPR-Cas loci identified remained unclassified, suggesting that desert (hyper)saline niches can provide new sources of CRISPR systems (Figure 4A).

Most of the type II CRISPR-marker gene Cas9 sequences belonged to the *Planctomycetes (36%) and Verrucomicrobia* (14.1%) phyla as well as the *Alphaproteobacteria (18.4%), Gammaproteobacteria* (4.4%) and *Acidithiobacillia* (3.2%) proteobacterial classes (Supplementary Table 10). The high representation of *cas9* genes from the recently defined sulfur-oxidizing autotrophs *Acidithiobacillia* (Hudson et al., 2014) was remarkable, since this taxon represented from only 0.02% (dark halite) to 0.142% (Hosabes sink) of the total sequences. While the taxonomic distribution of CRISPR-Cas systems is highly patchy, particularly in bacteria, this result suggests an enrichment of type II CRISPR-Cas systems in the *Acidithiobacillia*. The majority of the Cas9 protein sequences were found to branch deeply within the II-C subtype, with no affiliations to the II-A and II-B subtypes (Figure 5B). These results suggest that a large proportion of the saline sample metagenomic Cas9 sequences may correspond to a novel subtype.

Viral diversity of the Namib salt pan and halite metagenomes

The VirSorter tool (Roux et al., 2015) was used to extract putative viral genomic content from the assembled metagenomic data. A total of 3448 contigs were predicted to be of viral origin, of which 857 were over 10 kb in length (Supplementary Table 11). This dataset is subsequently referred to as *mVir*. To compare the similarity of mVir viral populations to viruses in the RefSeq database and previously studied viruses from the Namib (named as NamibVir), a genome-based gene-sharing network was constructed using the vContact 2.0 pipeline which employs a distance-based hierarchical clustering approach to classify viral sequences into clusters that are equal to viral genera (Bin Jang et al., 2019). The organization of sequences in a network implies a common phylogenetic origin, as occurs for the *Caudovirales* network (Bin Jang et al., 2019).

This analysis resulted in the identification of 145 clusters containing 371 mVir sequences, plus 255 sequences classified as outliers (with only weak connections to any given cluster and with insufficient information to accurately assign them to genus level) and 197 as singletons (sequences with no similarity to any other, thus not included in the network). The resulting network of clustered viruses is shown in Figure 4. A total of 84% of the mVir clusters (122) were constituted exclusively by mVir sequences, constituting up to 138 putative new viral genera and highlighting the very high genetic diversity of the Namib salt pan viruses (Figure 5 and Supplementary Table 12-14).

Taxonomy assignment of the mVir-containing clusters (*n* ≥3 sequences*)* revealed that 53 clusters belong to the order *Caudovirales* (Figure 5, inset and Supplementary Table 15), confirming the already observed dominance of this viral order in many Namib Desert niches (Adriaenssens et al., 2016; Hesse et al., 2017; Zablocki et al., 2017). Only 11 clusters could be classified at family level and 4 at genus level (Supplementary Table 15); three halite virus clusters belonging to the *Euryarchaea*-infecting *Betapleolipovirus* and one to an unclassified halovirus family (Figure 5). Additionally, 23 sequences from the halite mVir metagenomic data formed a cluster together with 17 *Halobacteriales*-infecting viruses from the RefSeq database and 3 NamibVir sequences (Figure 5, *Halovirus* network). Nevertheless, the majority of could not be clustered at genus level, suggesting they belong to novel viral taxa infecting archaeal *Halobacteriales.* Other halite mVir sequences were connected to archaeal viruses of the families *Alphasphaerolipoviridae* and *Alphapleolipoviridae*, to the *Caudovirales* main network and to a group of *Cyanobacteria*-infecting *Podoviridae* known to infect freshwater and thermophilic members of this bacterial phylum (Figure 5).

We also compared our mVir dataset to the viral fraction of halite nodules from the hyperarid Atacama Desert (denoted as Hvir; (Crits-Christoph et al., 2016)). Only two Hvir sequences clustered together with mVir data, specifically with cluster VC_469 which contained 3 halite mVir and one NamibVir contigs (Supplementary Figure 3 and Supplementary Table 16). Despite the small number of sequences available, the formation of clusters comprised exclusively of metagenomic sequences show that halite rocks harbor novel viral diversity.

Virus-host interactions reveal novel *Planctomycetes*-infecting viruses

To investigate virus-host associations, we used an established *in silico* approach based on CRISPR spacer matches between the cellular and viral sequences (Edwards et al., 2016; Paez-Espino et al., 2016). A total of 1431 CRISPR spacers originating from the salt pan metagenomic data were matched to viral sequences retrieved from the RefSeq, NamibVir and mVir databases (Supplementary Figure 5 and Supplementary Table 17). The majority of the virus-host matches (88.4%) targeted mVir viruses, while 6.6% matches belonged to sequences from one of the NamibVir soil viromes (Hesse et al., 2017) and 4.9 % to RefSeqABV or RefSeq virus databases. Surprisingly, no hits to contigs with taxonomy assignment were found to viral sequences from a previously sequenced metavirome from the same Namib salt pans (Adriaenssens et al., 2016).

A total of 72% of the matches arose from the abundant *Proteobacteria* phylum, of which 43% originated from the *Gammaproteobacteria,* followed by a 7.5% from *Bacteroidetes* and 5.5 % to *Planctomycetes.* Matches to other low abundance taxa were found for the *Cyanobacteria, Lentisphaerae, Deltaproteobacteria, Gemmatimonadetes* and few hits to the halite *Halobacteria* (Supplementary Figure 5 and Supplementary Table 17).

An alternative approach to establish virus-host linkages is the prediction of proviral sequences within cellular contigs (Edwards et al., 2016). 201 sequences were identified as proviruses of which 119 (59.2%) could be phylogenetically assigned, particularly to *Alphaproteobacteria* (40%)*, Planctomycetes* 11% and *Euryarchaea* (10%) contigs (Supplementary Table 18). We note that proviruses were also identified in members of the *Verrucomicrobia* phylum, where this host linkage was not identified through CRISPR spacer matches. Furthermore, 7 viral clusters (VC_204, VC_69, VC_350, VC_180, VC_469, VC_409 and VC_406), including the largest mVir cluster VC_204, contain prophage sequences, implying a temperate infection mode for these viruses (Table 2).

With the results from CRISPR spacer matches and provirus prediction we were able to identify viral host-interactions for 8 of the largest *Caudovirales* and 4 *Euryarchaea*-infecting viral clusters. The clusters VC_69 and VC_74 are linked to the alphaproteobacterial *Rhodobacterales* order, and VC_ 350 to the *Rhizobiales* and *Rhodospirillales* ones. VC_204 and VC_180 are connected to the Planctomycetes, and VC_242, to the phylum *Lentisphaerae*. Most halovirus-host linkages were found through prophage identification, linking clusters VC_357, VC_390, VC_406 and VC_469 to members of the *Halobacteriales*, with cluster VC_469 specifically linked to the euryarchaeal *Halorubrum* genus (Table 2).

Interestingly, the two most abundant salt pan taxa, i.e., *Alphaproteobacteria* and *Gammaproteobacteria,* show different virus-host linkage profiles: the *Gammaproteobacteria* virus host linkages were mainly identified through CRISPR hits, while those of the *Alphaproteobacteria* were identified through the prediction of prophages. Moreover, viruses linked to the *Gammaproteobacteria* did not belong to the major mVir clusters in the gene-sharing network, while 3 of the 10 largest mVir clusters infected the *Alphaproteobacteria.* Overall, this suggested that stream mat *Gammaproteobacteria* viruses have predominantly a lytic infection mode while the *Alphaproteobacteria* viruses are rather lysogenic.

The low number of CRISPR spacer matches to a previously published Hosabes and Eisfeld saltern metaviromes was unexpected (Adriaenssens et al., 2016). To better explain this finding, an extended network including all the NamibVir sequences regardless of contig length was generated (Supplementary Figure 5). 90% of the metaviromic saltern contigs clustered together in a group of ssDNA viruses of the *Microviridae* family, as reported previously for this dataset (Adriaenssens et al., 2016) (Supplementary Figure 5 and Supplementary Table 19). Given the small size of these viruses (around 5 kb), they were not included in the previous protein-sharing network that only incorporated sequences over 10 kb. Moreover, the *Microviridae* cluster included only one mVir contig. Very few mVir contigs grouped with other metaviromic saltern contigs of putative dsDNA viruses. This demonstrates that the two datasets of viral sequences from the Namib salt pans, the metagenomic mVir and the metaviromic saltern fraction of the NamibVir, represent different populations of viruses with distinct taxonomic affiliations.

Alphaproteobacterial gene transfer agent-like islands are present in the Namib Desert saline stream mats

From the 43100 mVir ORFs, 44% were functionally annotated; the majority (646/7130) corresponding to proteins involved in DNA metabolism and viral structural proteins (Supplementary Table 20). Transposases and integrases were also especially abundant, accounting for 1.5% and 1.6% of the total ORFs, respectively.

Interestingly, 77 ORFs were annotated as gene transfer agent-like (GTA-like) structural proteins. Gene transfer agents are virus-like particles encoded and produced by their prokaryote hosts and containing random fragments of the host's genome. Consequently, they are considered to be viable vectors for horizontal gene transfer (Lang et al., 2017). Given that GTAs are well-documented elements of *Alphaproteobacteria* genomes (Lang et al., 2017), the presence of GTAs in the mVir data was investigated by performing a protein blast between mVir ORFs and reference GTAs (Supplementary Table 21). A total of 437 mVir contigs matched known *Alphaproteobacteria* GTA proteins. Although 76% of hits were to unclustered contigs (i.e., contigs unassigned to a viral cluster), 42 clusters matched GTA proteins, including all contigs from clusters VC_69 and VC_89 (Figures 5 and 6). We particularly note that VC_69 harbors 13.8% of the identified prophages that are taxonomically assigned to a viral genus. As GTAs are thought to derive from lysogenic viruses (Lang et al., 2017), the lysogenic nature of VC_69 supports a link between these viral elements and GTAs.

Protein homology to a GTA ORF is not sufficient to classify a viral sequence as a GTA (Lang et al., 2017). To distinguish a GTA from a prophage, it should not contain a viral replication module or the small subunit terminase andshould have a size of 13-15kb (Lang et al., 2017; Paul, 2008). These features were used to screen mVir clusters with a GTA-like signal (Supplementary Table 12; Figure 6 and Supplementary Figure 6). VC_69 exhibited all the necessary GTA-like features (Figure 6A). Moreover, the flanking regions of VC_69 sequences showed a high level of conservation, which would not be expected for "junk" sequences such as defective prophages (Supplementary Figure 7). In the gene-sharing network (Figure 5), VC_69 was connected to the mVir clusters VC_225, VC_309, VC_333 and VC_89, all of which have hits to GTA proteins, and to *Roseobacter phage RDJL1,* a virus phylogenetically related to the *Rhodobacteri capsulatus* GTA (RcGTA) (Lang et al., 2017). While no replication module was identified in the sequences of cluster VC_69, replication and structural modules were clearly present in VC_225, VC_309 and VC_89 (Figure 6A-C), and the average contig length of these clusters was 22 kb. We note that these clusters contain both GTA-like sequences and sequences that resemble true viruses, contrary to VC_69. Our conclusion is that the latter is better described as a prophage remnant putatively converted into a gene-transfer agent, thus having a potential implication in driving gene exchange in the Namib salt pan *Alphaproteobacteria*.

Discussion

Alpha- and *Gamma-Proteobacteria* dominate the Namib salt pans while *Euryarchaea* and *Salinibacter* spp. prevail in the halites

The Namib Desert salt pans perform a crucial role in landscape weathering and deposition of abundant evaporites such as gypsum (Day and Seely, 2004; Eckardt and Drake, 2010; Eckardt et al., 2001). However, the ecological role of their microbial communities has not been addressed in great detail. The microbial stream mats from Hosabes and Eisfeld showed similar taxonomic and functional profiles, despite being 124 km distant (Figure 2). This is in agreement with a previous metaviromic analysis reporting that virus communities of both stream mats were also closely related (Adriaenssens et al., 2016). The similarities in the geological and physicochemical composition of the Hosabes and Eisfeld playas, as well as the existence of an underground water system connecting them (Adriaenssens et al., 2016; Day, 1993; Eckardt et al., 2001) may explain the taxonomic and functional resemblance of both stream mat microbiomes. Conversely, halite communities were markedly different from the stream mat communities, despite being separated by only a few meters, the disparity being the consequence of the contrasting physicochemical features of each niche: i.e., the mildly saline aquatic mat habitat vs. hypersaline and water-limiting halite.

Proteobacteria, in particular *Alpha-* and *Gammaproteobacteria* were abundant in the salt pan mats (Figure 2B). A dominance of the *Proteobacteria* has been previously reported before as a characteristic of saline mat microbiome (Bolhuis et al., 2014). The taxonomic diversity profile at phylum level of the stream microbial mats is similar to other saline environments, such as marine water or saline ponds (Allen et al., 2009; Baumgartner et al., 2009; Benlloch et al., 2002; Fernandez et al., 2016; Kimbrel et al., 2018; Sunagawa et al., 2015; Zhang et al., 2019). However, saline environments can vary widely in their microbial diversities. While *Cyanobacteria, Proteobacteria, Bacteroidetes* and *Chloroflexi* are common members of phototrophic microbial mats (Prieto-Barajas et al., 2018), the Namib saline spring mats harbor a high percentage of *Planctomycetes*. This phylum has been identified in hypersaline mats from Shark Bay, Australia, Eleuthera, Bahamas and Tebenchique lake, Chile (Allen et al., 2009; Baumgartner et al., 2009; Fernandez et al., 2016), but not in salt pan mats from the Kalahari Desert in southern Africa (Genderjahn et al., 2018) .

Halite microbial communities were markedly different to those inhabiting the surrounding stream mats (Hosabes). Specifically, the triad of microorganisms dominating the halites (*Halobacteriales, Salinibacter* and *Halothece*) is almost absent from the mats (Supplementary Tables 3-4). Enrichment of "salt-in" strategists such as the *Halobacteriales* and *Salinibacter* has been reported in hypersaline environments (Benlloch et al., 2002; Kimbrel et al., 2018; Oren, 2008; Vera-Gargallo et al., 2019) as this characteristic makes them especially adapted to these habitats. Furthermore, the predominance of the cyanobacterium *Halothece* has also been reported in halite microbial communities from the Atacama Desert and the Boneville Salt Flats (Crits-Christoph et al., 2016; Davila et al., 2015; Finstad et al., 2017; Gómez-Silva et al., 2019; de Los Ríos et al., 2010; McGonigle et al., 2019). Overall, the cosmopolitan distribution of these three genera points to highly specialized functional adaptations to saline extremes and possibly also to interactions between them.

Contrast between the complex biogeochemical cycles of the Namib salt pan mats vs. the simple cycles of the oligotrophic halites

Proteobacteria are key to the functioning of Namib salt pan mat Namib as these were found able to perform several steps of the biogeochemical cycles. This metabolic diversity has been proposed as a feature that allows them to occupy different trophic niches and survive under fluctuating extreme conditions (Bolhuis et al., 2014). Although less abundant than the *Proteobacteria,* the mat *Planctomycetes* possess a diverse functional profile (e.g., potential to carry on carbon fixation through the Wood-Ljungdahl pathway, fermentation, nitrification and dissimilatory nitrate reduction) (Figure 3), positioning them as a core taxon for biogeochemical cycling in the community.

The analysis of the stream mat functional capacity positions the *Cyanobacteria* and the *Alpha-* and *Gammaproteobacteria* as the main primary producers of the community via their photosynthetic capacity, with the additional contribution of the *Deltaproteobacteria* and the *Planctomycetes* (Figure 3). By contrast, carbon fixation in the halite was almost an exclusive capacity of the *Cyanobacteria*, which belonged mainly to the genus *Halothece,* absent from the salt pan mats and adapted to hypersaline conditions (Figure 3). Additionally, the capacity to obtain energy from light (phototrophy) and from CO oxidation (carboxydovory) was widespread in both mat and halite communities, implying that the use of alternative energy sources may help these microorganisms to survive under oligotrophic conditions.

The limited apparent capacity for nitrogen fixation in the Hosabes and Eisfeld microbial mat assemblages suggests the Namib salt pan microorganisms may rely on the assimilation of nitrate compounds. By contrast, nitrogen fixation was absent from both halite samples, and environmental nitrate assimilation was the main step of the nitrogen cycle for the whole community (Figure 3). Altogether, this hints to an abundance of nitrogen compounds in the environment, in spite of the oligotrophic nature of desert biomes. A probable source of nitrogen to sustain the nitrogen cycle in the Namib playas may be humberstonite ($K_3Na_7Mg_2(SO_4)_6(NO_3)_2*H_2O$), a sulfate-nitrogen mineral that has been only identified in the Atacama and the Namib Deserts (Eckardt et al., 2001).

Similarly, the sulfur cycle of the Namib salt pan mats and halites was potentially dependent on assimilatory sulfate reduction, although capacity for anaerobic sulfate respiration was detected for members of the *Proteobacteria.* In this regard, the abundant presence of gypsum deposits of the Namib salt pans may represent the source of sulfate for the salt pan microbial mat and halite communities (Eckardt et al., 2001).

The main contrast between the stream mat and halite biogeochemical cycles resides in the nitrogen and sulfur cycles, with halite communities having simplified functional capacity relying on assimilation of compounds. These differences could arise from the decrease in diversity associated to the hypersaline conditions of halite minerals. Interestingly, halite minerals from the Atacama Desert have similar taxonomic and functional profiles to the Namib halites. In particular, the absence of nitrogen-fixing *Cyanobacteria* has been reported by several studies (Crits-Christoph et al., 2016; Finstad et al., 2017; Gómez-Silva et al., 2019) and recent work has described simple nitrogen and sulfur cycles limited to the uptake of inorganic nitrogen and sulfur (Gómez-Silva et al., 2019). The overall global taxonomic and functional resemblance of halite microbial communities points out to the selection of universal specialists adapted to the oligotrophic, hypersaline conditions of halites.

Putative novel type II CRISPR-Cas systems in the Namib salt pans

The relative abundance of defense systems in the metagenomic data is in accordance with previous reports of the abundance of these systems in bacterial and archaeal genomes, where toxin-antitoxin and restriction modification systems are the most widely distributed and occupy the largest fraction of the genome (Koonin et al., 2017; Puigbò et al., 2017). By contrast, the proportion of CRISPR-Cas systems in the salt pan microbial population was unusual: type III systems comprised over 50% of the CRISPR-Cas systems identified, a percentage that doubles the proportion of type III *loci* in the currently sequenced prokaryotic genomes. Type III *loci* represent around 25% of all CRISPR-Cas systems and type I CRISPR-Cas systems were the most widespread, with a relative abundance of approximately 60% (Makarova and Koonin, 2015).

Although the fraction of type II CRISPR-Cas systems in the salt pan metagenomic data was similar to the previous estimation of type II *loci* abundance in bacterial genomes (Makarova and Koonin, 2015), a phylogenetic analysis of the Cas9 protein, an effector and marker gene of type II systems, reveals novel diversity of sequences branching deeply within Cas9 II-C subtype, putatively constituting a novel subtype of Cas9 proteins (Figure 4). These new sequences belong to phyla where few type II CRISPR-Cas systems have been previously described, such as the *Planctomycetes* and *Verrucomicrobia,* underlying the importance of studying uncultured microorganisms of diverse environments. Given the importance of type II CRISPR-Cas systems for genome editing applications (Lau, 2018), these results suggest that the Namib Desert salt pans may represent a resource for identification of new CRISPR systems.

Lysogenic viruses infect the main microbial taxa in the Namib salt pan and halite communities

The application of an *in silico* approach to study the Namib salt pan virus population allowed the identification of 138 putative novel viral genera, almost exclusively belonging to the order *Caudovirales,* the largest viral order of prokaryotic viruses to date (Bin Jang et al., 2019) (Supplementary Table 15). Comparisons of the mVir dataset obtained from this study to other viruses from the Namib Desert reveals a niche-dependent viral taxonomic diversity (Figure 5, Supplementary Figure 5), in agreement with the taxonomic differences in the microbial populations inhabiting each type of niche (Johnson et al., 2017).

Novel viruses of the euryarchaea *Halobacteriales* were also identified (clusters VC_357, VC_390, VC_406 and VC_460, Figure 5). Addition of mVir data and halite viruses from the Atacama Desert to the known Refseq haloviruses produced a rearrangement in the taxonomic affiliation of some RefSeq viruses, a phenomenon that indicates haloarchaeal viruses are under-sampled (Bin Jang et al., 2019) and that further sampling of these viral populations is necessary to better chart these euryarchaeal viruses.

Virus-host linkages were identified to eleven different prokaryotic phyla, especially for the most abundant *Alpha-* and *Gammaproteobacteria* (Table 2). Interestingly, linkages to *Alphaproteobacteria* were mainly through the identification of proviruses, while linkages to the *Gammaproteobacteria* were principally through CRISPR spacer hits. This could reflect a divergent infection mode of the viruses infecting each taxon: primarily lysogenic viruses infecting *Alphaproteobacteria* hosts and lytic viruses infecting *Gammaproteobacteria*. The targeting of an integrated element by the CRISPR system would be strongly selected against and could explain the lack of virus-spacer hits in the *Alphaproteobacteria*.

The profuse host associations between the most abundant novel mVir viral genera in the salt pan mat, *Proteobacteria, Planctomycetes* and *Lentisphaerae,* as well as to the halite *Halobacteriales* and *Haloferacales* (Table 2), hints at an important role of viruses in nutrient recycling in the Namib salt pan communities through the infection and lysis of the abundant host taxa. Additionally, the identification of putative lysogenic viral lineages that include the largest mVir cluster identified in this study (VC_204 and VC_180) and infect members of the *Planctomycetes* (Supplementary Tables 17-18) suggests that these viruses could impact microbial mat function and structure, since members of this phylum are among the most abundant in the salt pan mats studied and possess unique metabolic capacities within their community.

One surprising observation of this study is the dissimilar viral taxonomic profiles of the mVir data from this work and a previous metaviromic study of the Hosabes and Eisfeld microbial mat viral populations. We conjecture that this could be the result of the different methodological approaches employed to analyze the viral fractions in these communities, where mVir sequences mined from metagenomic data may be enriched in proviruses while extracellular viruses used to produce metaviromes are enriched in lytic virus progeny, as has been suggested by previous work with soil viromes (Emerson et al., 2018; Trubl et al., 2018). For example, lytic archaeal viruses of the *Salterprovirus* genus (of which virus His1 is the reference strain) (Bath et al., 2006) were present in the metavirome of the salt pan but absent in the mVir data, which instead contained several proviruses (Table 2). Additionally, methods associated to metavirome library preparation, specifically the use of multiple displacement amplification (MDA), introduce a strong bias in favor of the amplification of ssDNA, which may explain the overwhelming presence of ssDNA *Microviridae* genomes in the metaviromic salt pan study (Roux et al., 2016). Taken together, these observations argue strongly in favor of using multiple different methods to obtain complementary information to characterize the virus diversity of any community.

Virus domestication impacts horizontal gene transfer in the Namib salt pan *Alphaproteobacteria*

It is hypothesized that *in silico* tools used to predict viruses from bulk metagenomic data may include other elements of the mobilome such as plasmids or relic phages (e.g., gene transfer agents and provirus remnants inserted in the microbial genomes) (Emerson et al., 2018; Roux et al., 2015; Shakya et al., 2017). Within this mobilome, GTAs are of special interest. These small virus-like particles are highly abundant in marine environments, where they have been shown to mediate HGT-events at very high frequencies (i.e., 10^{-2} to 10^{-4} (McDaniel et al., 2010, 2012)). GTAs arise from the incorporation lysogenic viruses that become inactive and are recruited or "domesticated" by the cell as tools for HGT. As a result, it is difficult to differentiate them from true virus sequences in environmental metagenomes (Lang et al., 2017).

Surprisingly, cluster mVir VC_69 was found to correspond to a gene-transfer agent instead of an authentic viral taxon. This cluster had the highest similarity to RcGTA in the mVir data and displayed all features characteristic of GTAs (Figure 6, Supplementary Figure 7), suggesting that the Namib salt pan *Rhodobacterales* have the capacity to produce GTAs, as other members of the *Alphaproteobacteria* (Lang et

al., 2017). Conversely, clusters VC_89, VC_255 and VC_309 contained both small GTA-like sequences together with *bona fide* viral sequences over 30 kb and with replication modules, suggesting that they correspond to true, active viruses. Although the true impact of GTA-mediated HGT is not known, it is hypothesized to be crucial for cellular adaptation and evolution, driving the diversification and adaptation of the alphaproteobacterial clades containing them to different environments (Shakya et al., 2017). This has important implications in the adaptability of the Namib salt pan *Rhodobacterales*.

Conclusions

Deserts are polyextreme environments and their indigenous microbial communities are particularly subjected to water limitation and oligotrophy. Consequently, deserts are among the global ecosystems with the lowest microbial diversities and abundances. In deserts, playas/salars/saltpans represent microniches with constant water availability and are characterized by high salt concentration. We show that the microbial mats of these saline springs constitute a large assemblage of microbial lineages with a vast metabolic genetic versatility, potentially enabling them to cope with fluctuating and extreme environmental conditions. Analyses of the viral fraction of the salt pan microbiome suggests that these habitats are a hub of novel viruses and viral activity.

Materials and Methods

Sample collection

Sediment and microbial mat samples were collected under sterile conditions in Whirlpack® bags from two salt pans in the Namib Desert in April 2016 and 2017: Hosabes (S 23°30'425'', E 15°04'309'') and Eisfeld (S 22°29'002'', E 14°34'363'') salt pans (Figure 1A-C). For each sampling campaign, samples were collected at the 'source' and 'sink' of each salt pan. Temperature and conductivity measurements were taken on site at the time of sampling. Samples from two halites close to the stream of the Hosabes salt pan were also collected in 2017: a red (S 23°30´25.1´´, E 15°04´17.1´´; Figure 1D) and a black halite (S 23°30´25.1´´, E 15°04´17.5´´; Figure 1E). The samples were stored at room temperature prior to their transport the CMEG laboratory, where they were stored at -20°C until metagenomic DNA (mDNA) extraction.

DNA extraction and sequencing

Samples were thawed on ice and 6-10 aliquots of approximately 0.25 g each were subjected to DNA extraction using the PowerLyzer® PowerSoil® DNA Isolation kit (QIAGEN) following the manufacturer´s instructions. Prior to mDNA extraction, the halite samples were pulverized with a sterile mortar and dissolved in a sterile 20% NaCl solution and the biomass was recovered by centrifugation at 10 000 x g for 15 min, 4°C. Aliquots of approximately 0.2 g of biomass pellets were further used for mDNA extraction also with the PowerLyzer® PowerSoil® DNA Isolation kit (QIAGEN).

DNA samples from the same location were pooled, concentrated using ethanol precipitation, and further purified using the DNeasy PowerClean CleanUp kit (QIAGEN). Consequently, ten composite mDNA preparations (from 8 salt pan and 2 halites, respectively) underwent library construction and sequencing by Admera Health, LCC (NJ, USA). The PCR-free KAPA Hyper prep kit was used for library preparation and paired-end reads of 150 bp were sequenced in a single lane using the Illumina HiSeq X platform.

Quality control, filtering and assembly of the sequenced data

Between 53 (Eisfeld sink 2017) and 132 (Hosabes sink 2016) million paired-end reads were obtained for each sample (Supplementary Table 1). The quality of the reads was checked using FastQC (2015). The BBTools package (BBMap – Bushnell B.- sourceforge.net/projects/bbmap) was used for read filtering: adapter trimming was done with BBDuk using the recommended settings and the read quality trimming was done at the Q10 quality level, with a minimum average quality of Q15 and a minimum length of 100 bp. The dedupe script (BBTools package) was used with the recommended settings to remove exact duplicates. The post-QC reads were assembled with SPAdes v3.9.0 (Bankevich et al., 2012) using the 'meta' flag. Only contigs over 500 bp were retained for further analysis. Assembly statistics were calculated with the Stats script implemented in the BBTools package and BBMap was used for the calculation of the sequencing depth per contig. reads were mapped to the assembly using Bowtie2 (Langmead and Salzberg, 2012). The quality filtering and assembly results are shown in Supplementary Table 1.

The 10 resulting contig files were uploaded to the IMG/M system (Chen et al., 2017) for functional and phylogenetic annotation and are available under the Study ID Gs0133438, Analysis Projects Ga0248485, Ga0248504, Ga0254891, Ga0255014, Ga0255825, Ga0256419, Ga0256679 and Ga0256680.

Shotgun metagenome phylogenetic and functional analysis

To gain insight into the genetic potential of the Namib salt pan and halite microbial communities, two approaches were combined. KOs belonging to the most abundant taxa (>5 %) were screened for the presence of complete KEGG modules (see Materials and Methods and Supplementary Table 6). Secondly, pathway hallmark genes were used as markers to assess the functional potential of the community, as well as the taxa possessing such genetic potential

The taxonomic assignment of the genes of each dataset was retrieved from the IMG/M annotation pipeline (Chen et al., 2017). The reads from each metagenome were mapped to the predicted genes and the relative abundance of each gene was calculated as described previously (Wagner et al., 2012). A raw count matrix of taxa at family level was created by combining the gene counts for each family in each metagenome. The resulting matrix was subjected to differential statistical analysis with the R package *DESeq2* (Love et al., 2014) using the Walden test with a $Padj \leq 0.05$. Only taxa with a log2FoldChange (log2FC) ≥ 1 were further considered.

For functional analysis, the assignment of KO terms to the predicted open reading frames (ORFs) was retrieved from the IMG/M pipeline (Chen et al., 2017). Phyla with more than 5% of abundance were selected for taxon-based functional analysis. Consequently, only genes assigned to the phyla *Proteobacteria*, *Bacteroidetes*, *Cyanobacteria*, *Planctomycetes* and *Euryarchaea* were retrieved, and the KO terms associated to each taxonomic group submitted to KEGG Mapper (Kanehisa, 2017) for the reconstruction of complete KEGG functional modules. A matrix of KO counts was constructed relating the KEGG modules to the taxa in each metagenome and normalized by the total KO terms of each dataset (Supplementary Table 6). Additionally, metabolic processes were investigated using specific KO terms as markers, and their relative abundances were calculated as described above (Supplementary Table 7).

CRISPR-Cas classification

PSI-BLAST (Altschul and Koonin, 1998) to the Conserved Domains database (CDD) (Marchler-Bauer et al., 2017) (downloaded on March 2017) with an E value ≤ 10^{-6} was used to identify Cas genes present in the salt pan metagenomic data (Makarova and Koonin, 2015). Multigenic *cas* loci were defined as two or more *cas* genes within 5 genes up- or down-stream of each other. Clusters were classified using weighted consensus of all members, assigning a value to each depending on their subtype specificity. For type II loci subtype

classification, protein sequences of the hallmark Cas9 gene larger than 500 amino acids were extracted and aligned together with reference Cas9 sequences from the Swissprot (2019) and CDD databases using MAFFT (Katoh et al., 2019) with default parameters. Duplicates were removed by CD-HIT (Huang et al., 2010) and the alignment was refined using MaxAlign (Gouveia-Oliveira et al., 2007). A phylogenetic Neighbor-Joining tree was built using the WAG substitution model and 500 bootstrap resamplings for all gap-free sites.

Identification of viral genomic from the Namib Desert metagenomic datasets

The metagenomic contig datasets were each processed with VirSorter v1.0.3 (Roux et al., 2015) through the Cyverse Discovery Environment using both the RefSeqABVir and the Virome databases, and as recommended only categories 1, 2, 4 and 5 were kept for further analysis. A total of 3485 predicted viral sequences were obtained from the complete metagenomic data after removal of duplicated entries. Among these, 201 sequences were identified as prophages of which 44 (i.e., 20%) were circular (Supplementary Table 11). Manual curation of the sequences further revealed the presence of 37 small contigs (shorter than 5 kb in length) and of high coverage only composed of genes encoding for integrases, transposases and recombinases. These most probably represent repetitive regions and were not considered for further analysis. The resulting viral database is named as mVir.

Construction and curation of a Namib Desert virome database

The Namib Desert viruses database (namely, NamibVir) was constructed by retrieving viral sequences from all the Namib Desert metaviromic studies available (i.e., from hypolith, salt pan and soil samples; (Adriaenssens et al., 2016; Hesse et al., 2017; Zablocki et al., 2017). To remove contaminant microbial sequences, the metaviromic contigs were processed with VirSorter using the virome decontamination mode (Roux et al., 2015). Ultimately, the NamibVir database contained 57740 sequences of which only 101 had a size of 10 kb or longer.

Identification of virus-host pairs using CRISPR spacer matches

Taxonomic classification of the CRISPR *loci*-containing scaffolds was obtained using the Contig Annotation Tool (Cambuy et al., 2016). Taxonomic classification of the viral sequences was inferred based on the consensus of BLASTP hits (E value ≤ 10^{-5}) to the RefSeq virus proteins (Brister et al., 2015) (downloaded on June, 2018), using MEGAN4 (Huson et al., 2011) to assign contigs to taxa.

A total of 4821 CRISPR *loci,* containing 38126 spacer sequences, were retrieved from the IMG/M annotation (Chen et al., 2017) of the salt pan and halite metagenomic datasets. These spacers were used to create a spacer database. To identify virus-host pairs, the spacer database was compared to four different virus databases: RefSeq viral genomes (Brister et al., 2015), RefSeqABV (Roux et al., 2015), NamibVir and the predicted metagenomic viral sequences (mVir database) as described previously (Paez-Espino et al., 2016). Briefly: spacers were aligned to viral sequences using blastn (blastn-short task, E-value of 10^{-10}, percent identity of 95% and max_target_sequences = 1). This resulted in 1542 virus-spacer linkages of which only the 1145 matches with contig taxonomy assignment were used to construct a map of virus-host interactions that was visualized using Cytoscape (https://cytoscape.org/ v 3.7.0) (Supplementary Table 17).

Construction of a genome-based viral network

Proteins in the 857 mVir genomes over 10 kb were clustered with proteins from 3747 viral genomes in RefSeq (June 2018) and the 101 genomes (>10 kb) from NamibVir database using an all-versus-all BLASTP (E-value 0.00001) followed by the aggrupation into protein clusters as previously described (Bolduc et al.,

2017). A similarity score was calculated using vContact v2.0 (Bin Jang et al., 2019) and the resulting network was visualized with Cytoscape ("https://cytoscape.org/" v 3.7.0) using an edge-weighted spring model. Taxonomy assignment of the mVir viral clusters was done following three criteria: if the cluster included one or more RefSeq viruses their taxonomic affiliation from the NCBI taxonomy was assigned to the viral cluster at the lowest taxonomy rank in common (using a 75% cut-off value). If the cluster consisted exclusively of mVir and NamibVir sequences, the lowest taxonomy rank in common to all sequences (using a 70% cut-off value) obtained from the consensus of BLASTP hits as described above was selected as putative taxonomy of the cluster. Finally, network topology was also used to assign taxonomy. If the sequences belonged to a network of sequences containing RefSeq viruses with taxonomic consensus at order level they were automatically assigned to that order.

Annotation of viral sequences

The mVir genes were annotated by using a combination of the functional annotation retrieved from the IMG/M pipeline with the result of matching the viral ORFs to the Prokaryotic Virus Orthologous Groups (pVOG) database (Grazziotin et al., 2017) using hmmsearch (Johnson et al., 2010) with an E-value threshold of 10^{-5}.

BLASTP vs GTAs

BLASTP similarity searches were carried out on all mVir ORFs using the genes of *Rhodobacter capsulatus* (GenBank: AF181080.3) and *Bartonella australis* GTAs (Guy et al., 2013), retaining matches with E-values ≤ 10^{-5}. Results are reported in Supplementary Table 21.

Accession numbers

Metagenomic data generated in this work can be accessed through the IMG/M database (https://www.img.jgi.doe.gov) under GOLD Sequencing Project ID: Gp0293142 and IMG Genome IDs: 3300023218, 3300023197, 3300022725, 3300023214, 3300023202, 3300022723, 3300022777, 3300022719, 3300022719 and 3300022719.

Acknowledgements

The authors acknowledge funding support from the University of Pretoria and the National Research Foundation (NRF grant 113308). The authors declare that no conflicts of interest exist.

Author contributions

LMA designed the study, performed the experimental work and bioinformatic analysis of the data and drafted the manuscript. SV was involved in the taxonomic analysis of the contigs and CLS carried out the classification of CRISPR-Cas systems. JBR participated in sample collection and critical revisions of the manuscript. GMK provided logistical support in the Namib Desert. DAC provided funding and analysis tools, and assisted with manuscript revisions. All authors have read and approved the final manuscript.

References

Adriaenssens, E.M., Van Zyl, L., De Maayer, P., Rubagotti, E., Rybicki, E., Tuffin, M., and Cowan, D.A. (2015). Metagenomic analysis of the viral community in Namib Desert hypoliths. Environ. Microbiol. *17*, 480–495.

Adriaenssens, E.M., van Zyl, L.J., Cowan, D.A., and Trindade, M.I. (2016). Metaviromics of Namib Desert Salt Pans: A Novel Lineage of Haloarchaeal Salterproviruses and a Rich Source of ssDNA Viruses. Viruses *8*.

666 Allen, M.A., Goh, F., Burns, B.P., and Neilan, B.A. (2009). Bacterial, archaeal and eukaryotic diversity of
667 smooth and pustular microbial mat communities in the hypersaline lagoon of Shark Bay. Geobiology *7*, 82–
668 96.

669 Altschul, S.F., and Koonin, E.V. (1998). Iterated profile searches with PSI-BLAST--a tool for discovery in
670 protein databases. Trends Biochem. Sci. *23*, 444–447.

671 Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., Lesin, V.M., Nikolenko, S.I.,
672 Pham, S., Prjibelski, A.D., et al. (2012). SPAdes: a new genome assembly algorithm and its applications to
673 single-cell sequencing. J. Comput. Biol. *19*, 455–477.

674 Bath, C., Cukalac, T., Porter, K., and Dyall-Smith, M.L. (2006). His1 and His2 are distantly related, spindle-
675 shaped haloviruses belonging to the novel virus group, Salterprovirus. Virology *350*, 228–239.

676 Baumgartner, L.K., Dupraz, C., Buckley, D.H., Spear, J.R., Pace, N.R., and Visscher, P.T. (2009). Microbial
677 species richness and metabolic activities in hypersaline microbial mats: insight into biosignature formation
678 through lithification. Astrobiology *9*, 861–874.

679 Benlloch, S., López-López, A., Casamayor, E.O., Øvreås, L., Goddard, V., Daae, F.L., Smerdon, G., Massana,
680 R., Joint, I., Thingstad, F., et al. (2002). Prokaryotic genetic diversity throughout the salinity gradient of a
681 coastal solar saltern. Environ. Microbiol. *4*, 349–360.

682 Bin Jang, H., Bolduc, B., Zablocki, O., Kuhn, J.H., Roux, S., Adriaenssens, E.M., Brister, J.R., Kropinski, A.M.,
683 Krupovic, M., Lavigne, R., et al. (2019). Taxonomic assignment of uncultivated prokaryotic virus genomes is
684 enabled by gene-sharing networks. Nat. Biotechnol. *37*, 632–639.

685 Bolduc, B., Jang, H.B., Doulcier, G., You, Z.-Q., Roux, S., and Sullivan, M.B. (2017). vConTACT: an iVirus tool
686 to classify double-stranded DNA viruses that infect Archaea and Bacteria. PeerJ *5*, e3243.

687 Bolhuis, H., Cretoiu, M.S., and Stal, L.J. (2014). Molecular ecology of microbial mats. FEMS Microbiol. Ecol.
688 *90*, 335–350.

689 Borjian, F., Han, J., Hou, J., Xiang, H., and Berg, I.A. (2016). The methylaspartate cycle in haloarchaea and its
690 possible role in carbon metabolism. ISME J *10*, 546–557.

691 Brister, J.R., Ako-adjei, D., Bao, Y., and Blinkova, O. (2015). NCBI Viral Genomes Resource. Nucleic Acids Res
692 *43*, D571–D577.

693 Brum, J.R., and Sullivan, M.B. (2015). Rising to the challenge: accelerated pace of discovery transforms
694 marine virology. Nat. Rev. Microbiol. *13*, 147–159.

695 Bryant, D.A. and Frigaard, Niels-Ulrik (1996). Validated linear mixture modelling of Landsat TM data for
696 mapping evaporite minerals on a playa surface: methods and applications. International Journal of Remote
697 Sensing *17*, 315–330.

698 Cambuy, D.D., Coutinho, F.H., and Dutilh, B.E. (2016). Contig annotation tool CAT robustly classifies
699 assembled metagenomic contigs and long sequences. BioRxiv 072868.

700 Chen, I.-M.A., Markowitz, V.M., Chu, K., Palaniappan, K., Szeto, E., Pillay, M., Ratner, A., Huang, J.,
701 Andersen, E., Huntemann, M., et al. (2017). IMG/M: integrated genome and metagenome comparative
702 data analysis system. Nucleic Acids Res. *45*, D507–D516.

703 Crits-Christoph, A., Gelsinger, D.R., Ma, B., Wierzchos, J., Ravel, J., Davila, A., Casero, M.C., and DiRuggiero,
704 J. (2016). Functional interactions of archaea, bacteria and viruses in a hypersaline endolithic community.
705 Environ. Microbiol. *18*, 2064–2077.

706 Cronan, J.E., and Laporte, D. (2005). Tricarboxylic Acid Cycle and Glyoxylate Bypass. EcoSal Plus *1*.

Davila, A.F., Hawes, I., Araya, J.G., Gelsinger, D.R., DiRuggiero, J., Ascaso, C., Osano, A., and Wierzchos, J. (2015). In situ metabolism in halite endolithic microbial communities of the hyperarid Atacama Desert. Front Microbiol *6*, 1035.

Day, J.A. (1993). The major ion chemistry of some southern African saline systems. In Saline Lakes V, S.H. Hurlbert, ed. (Springer Netherlands), pp. 37–59.

Day, J., and Seely, M. (2004). Physical and chemical conditions in an hypersaline spring in the Namib Desert. Hydrobiologia.

Dupraz, C., and Visscher, P.T. (2005). Microbial lithification in marine stromatolites and hypersaline mats. Trends Microbiol. *13*, 429–438.

Eckardt, F., and Drake, N. (2010). Introducing the Namib Desert Playas. pp. 19–25.

Eckardt, F.D., Drake, N., Goudie, A.S., White, K., and Viles, H. (2001). The role of playas in pedogenic gypsum crust formation in the Central Namib Desert: a theoretical model. Earth Surface Processes and Landforms *26*, 1177–1193.

Edwards, R.A., McNair, K., Faust, K., Raes, J., and Dutilh, B.E. (2016). Computational approaches to predict bacteriophage-host relationships. FEMS Microbiol. Rev. *40*, 258–272.

Emerson, J.B., Roux, S., Brum, J.R., Bolduc, B., Woodcroft, B.J., Jang, H.B., Singleton, C.M., Solden, L.M., Naas, A.E., Boyd, J.A., et al. (2018). Host-linked soil viral ecology along a permafrost thaw gradient. Nat Microbiol *3*, 870–880.

Fernandez, A.B., Rasuk, M.C., Visscher, P.T., Contreras, M., Novoa, F., Poire, D.G., Patterson, M.M., Ventosa, A., and Farias, M.E. (2016). Microbial Diversity in Sediment Ecosystems (Evaporites Domes, Microbial Mats, and Crusts) of Hypersaline Laguna Tebenquiche, Salar de Atacama, Chile. Front Microbiol *7*, 1284.

Finstad, K.M., Probst, A.J., Thomas, B.C., Andersen, G.L., Demergasso, C., Echeverría, A., Amundson, R.G., and Banfield, J.F. (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.

Fourçans, A., Solé, A., Diestra, E., Ranchou-Peyruse, A., Esteve, I., Caumette, P., and Duran, R. (2006). Vertical migration of phototrophic bacterial populations in a hypersaline microbial mat from Salins-de-Giraud (Camargue, France). FEMS Microbiology Ecology *57*, 367–377.

Genderjahn, S., Alawi, M., Mangelsdorf, K., Horn, F., and Wagner, D. (2018). Desiccation- and Saline-Tolerant Bacteria and Archaea in Kalahari Pan Sediments. Front Microbiol *9*, 2082.

Gómez-Silva, B., Vilo-Muñoz, C., Galetović, A., Dong, Q., Castelán-Sánchez, H.G., Pérez-Llano, Y., Sánchez-Carbente, M.D.R., Dávila-Ramos, S., Cortés-López, N.G., Martínez-Ávila, L., et al. (2019). Metagenomics of Atacama Lithobiontic Extremophile Life Unveils Highlights on Fungal Communities, Biogeochemical Cycles and Carbohydrate-Active Enzymes. Microorganisms *7*.

Gouveia-Oliveira, R., Sackett, P.W., and Pedersen, A.G. (2007). MaxAlign: maximizing usable data in an alignment. BMC Bioinformatics *8*, 312.

Grazziotin, A.L., Koonin, E.V., and Kristensen, D.M. (2017). Prokaryotic Virus Orthologous Groups (pVOGs): a resource for comparative genomics and protein family annotation. Nucleic Acids Res *45*, D491–D498.

Guy, L., Nystedt, B., Toft, C., Zaremba-Niedzwiedzka, K., Berglund, E.C., Granberg, F., Näslund, K., Eriksson, A.-S., and Andersson, S.G.E. (2013). A gene transfer agent and a dynamic repertoire of secretion systems hold the keys to the explosive radiation of the emerging pathogen Bartonella. PLoS Genet. *9*, e1003393.

748 Hesse, U., van Heusden, P., Kirby, B.M., Olonade, I., van Zyl, L.J., and Trindade, M. (2017). Virome Assembly
749 and Annotation: A Surprise in the Namib Desert. Front Microbiol *8*, 13.

750 Huang, Y., Niu, B., Gao, Y., Fu, L., and Li, W. (2010). CD-HIT Suite: a web server for clustering and comparing
751 biological sequences. Bioinformatics *26*, 680–682.

752 Hudson, C.M., Williams, K.P., and Kelly, D.P. (2014). Definitive assignment by multigenome analysis of the
753 gammaproteobacterial genus Thermithiobacillus to the class Acidithiobacillia. Pol. J. Microbiol. *63*, 245–
754 247.

755 Huson, D.H., Mitra, S., Ruscheweyh, H.-J., Weber, N., and Schuster, S.C. (2011). Integrative analysis of
756 environmental sequences using MEGAN4. Genome Res *21*, 1552–1560.

757 Johnson, L.S., Eddy, S.R., and Portugaly, E. (2010). Hidden Markov model speed heuristic and iterative HMM
758 search procedure. BMC Bioinformatics *11*, 431.

759 Johnson, R.M., Ramond, J.-B., Gunnigle, E., Seely, M., and Cowan, D.A. (2017). Namib Desert edaphic
760 bacterial, fungal and archaeal communities assemble through deterministic processes but are influenced by
761 different abiotic parameters. Extremophiles *21*, 381–392.

762 Kanehisa, M. (2017). Enzyme Annotation and Metabolic Reconstruction Using KEGG. Methods Mol. Biol.
763 *1611*, 135–145.

764 Katoh, K., Rozewicki, J., and Yamada, K.D. (2019). MAFFT online service: multiple sequence alignment,
765 interactive sequence choice and visualization. Brief Bioinform *20*, 1160–1166.

766 Kimbrel, J.A., Ballor, N., Wu, Y.-W., David, M.M., Hazen, T.C., Simmons, B.A., Singer, S.W., and Jansson, J.K.
767 (2018). Microbial Community Structure and Functional Potential Along a Hypersaline Gradient. Front
768 Microbiol *9*, 1492.

769 Koonin, E.V., Makarova, K.S., and Zhang, F. (2017). Diversity, classification and evolution of CRISPR-Cas
770 systems. Curr. Opin. Microbiol. *37*, 67–78.

771 Lang, A.S., Westbye, A.B., and Beatty, J.T. (2017). The Distribution, Evolution, and Roles of Gene Transfer
772 Agents in Prokaryotic Genetic Exchange. Annu Rev Virol *4*, 87–104.

773 Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods *9*, 357–
774 359.

775 Lau, C.-H. (2018). Applications of CRISPR-Cas in Bioengineering, Biotechnology, and Translational Research.
776 The CRISPR Journal *1*, 379–404.

777 de Los Ríos, A., Valea, S., Ascaso, C., Davila, A., Kastovsky, J., McKay, C.P., Gómez-Silva, B., and Wierzchos, J.
778 (2010). Comparative analysis of the microbial communities inhabiting halite evaporites of the Atacama
779 Desert. Int. Microbiol. *13*, 79–89.

780 Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-
781 seq data with DESeq2. Genome Biol. *15*, 550.

782 Makarova, K.S., and Koonin, E.V. (2015). Annotation and Classification of CRISPR-Cas Systems. Methods
783 Mol. Biol. *1311*, 47–75.

784 Marchler-Bauer, A., Bo, Y., Han, L., He, J., Lanczycki, C.J., Lu, S., Chitsaz, F., Derbyshire, M.K., Geer, R.C.,
785 Gonzales, N.R., et al. (2017). CDD/SPARCLE: functional classification of proteins via subfamily domain
786 architectures. Nucleic Acids Res. *45*, D200–D203.

787 Marraffini, L.A. (2015). CRISPR-Cas immunity in prokaryotes. Nature *526*, 55–61.

McDaniel, L.D., Young, E., Delaney, J., Ruhnau, F., Ritchie, K.B., and Paul, J.H. (2010). High frequency of horizontal gene transfer in the oceans. Science *330*, 50.

McDaniel, L.D., Young, E.C., Ritchie, K.B., and Paul, J.H. (2012). Environmental factors influencing gene transfer agent (GTA) mediated transduction in the subtropical ocean. PLoS ONE *7*, e43506.

McGonigle, J.M., Bernau, J.A., Bowen, B.B., and Brazelton, W.J. (2019). Robust Archaeal and Bacterial Communities Inhabit Shallow Subsurface Sediments of the Bonneville Salt Flats. MSphere *4*.

Oren, A. (2008). Microbial life at high salt concentrations: phylogenetic and metabolic diversity. Saline Syst. *4*, 2.

Paez-Espino, D., Eloe-Fadrosh, E.A., Pavlopoulos, G.A., Thomas, A.D., Huntemann, M., Mikhailova, N., Rubin, E., Ivanova, N.N., and Kyrpides, N.C. (2016). Uncovering Earth's virome. Nature *536*, 425–430.

Paul, J.H. (2008). Prophages in marine bacteria: dangerous molecular time bombs or the key to survival in the seas? ISME J *2*, 579–589.

Petushkova, E., Mayorova, E., and Tsygankov, A. (2021). TCA Cycle Replenishing Pathways in Photosynthetic Purple Non-Sulfur Bacteria Growing with Acetate. Life (Basel) *11*, 711.

Peyraud, R., Kiefer, P., Christen, P., Massou, S., Portais, J.-C., and Vorholt, J.A. (2009). Demonstration of the ethylmalonyl-CoA pathway by using 13C metabolomics. PNAS *106*, 4846–4851.

Prieto-Barajas, C.M., Valencia-Cantero, E., and Santoyo, G. (2018). Microbial mat ecosystems: Structure types, functional diversity, and biotechnological application. Electronic Journal of Biotechnology *31*, 48–56.

Puigbò, P., Makarova, K.S., Kristensen, D.M., Wolf, Y.I., and Koonin, E.V. (2017). Reconstruction of the evolution of microbial defense systems. BMC Evol. Biol. *17*, 94.

Qian, Z., Tianwei, H., Mackey, H.R., van Loosdrecht, M.C.M., and Guanghao, C. (2019). Recent advances in dissimilatory sulfate reduction: From metabolic study to application. Water Research *150*, 162–181.

Roux, S., Enault, F., Hurwitz, B.L., and Sullivan, M.B. (2015). VirSorter: mining viral signal from microbial genomic data. PeerJ *3*, e985.

Roux, S., Solonenko, N.E., Dang, V.T., Poulos, B.T., Schwenck, S.M., Goldsmith, D.B., Coleman, M.L., Breitbart, M., and Sullivan, M.B. (2016). Towards quantitative viromics for both double-stranded and single-stranded DNA viruses. PeerJ *4*, e2777.

Santos, F., Yarza, P., Parro, V., Meseguer, I., Rosselló-Móra, R., and Antón, J. (2012). Culture-independent approaches for studying viruses from hypersaline environments. Appl. Environ. Microbiol. *78*, 1635–1643.

Shakya, M., Soucy, S.M., and Zhaxybayeva, O. (2017). Insights into origin and evolution of α-proteobacterial gene transfer agents. Virus Evol *3*, vex036.

Sime-Ngando, T. (2014). Environmental bacteriophages: viruses of microbes in aquatic ecosystems. Front Microbiol *5*, 355.

Sunagawa, S., Coelho, L.P., Chaffron, S., Kultima, J.R., Labadie, K., Salazar, G., Djahanschiri, B., Zeller, G., Mende, D.R., Alberti, A., et al. (2015). Ocean plankton. Structure and function of the global ocean microbiome. Science *348*, 1261359.

Trubl, G., Jang, H.B., Roux, S., Emerson, J.B., Solonenko, N., Vik, D.R., Solden, L., Ellenbogen, J., Runyon, A.T., Bolduc, B., et al. (2018). Soil Viruses Are Underexplored Players in Ecosystem Carbon Processing. MSystems *3*.

827 Vera-Gargallo, B., Chowdhury, T.R., Brown, J., Fansler, S.J., Durán-Viseras, A., Sánchez-Porro, C., Bailey, V.L.,
828 Jansson, J.K., and Ventosa, A. (2019). Spatial distribution of prokaryotic communities in hypersaline soils. Sci
829 Rep *9*, 1769.

830 Wagner, G.P., Kin, K., and Lynch, V.J. (2012). Measurement of mRNA abundance using RNA-seq data: RPKM
831 measure is inconsistent among samples. Theory Biosci. *131*, 281–285.

832 Waiser, M.J., and Robarts, R.D. (2009). Saline Inland Waters. In Encyclopedia of Inland Waters, G.E. Likens,
833 ed. (Oxford: Academic Press), pp. 634–644.

834 Ward, J.D. (2010). On the antiquity of the Namib. 9.

835 Zablocki, O., Adriaenssens, E.M., Frossard, A., Seely, M., Ramond, J.-B., and Cowan, D. (2017). Metaviromes
836 of Extracellular Soil Viruses along a Namib Desert Aridity Gradient. Genome Announc *5*.

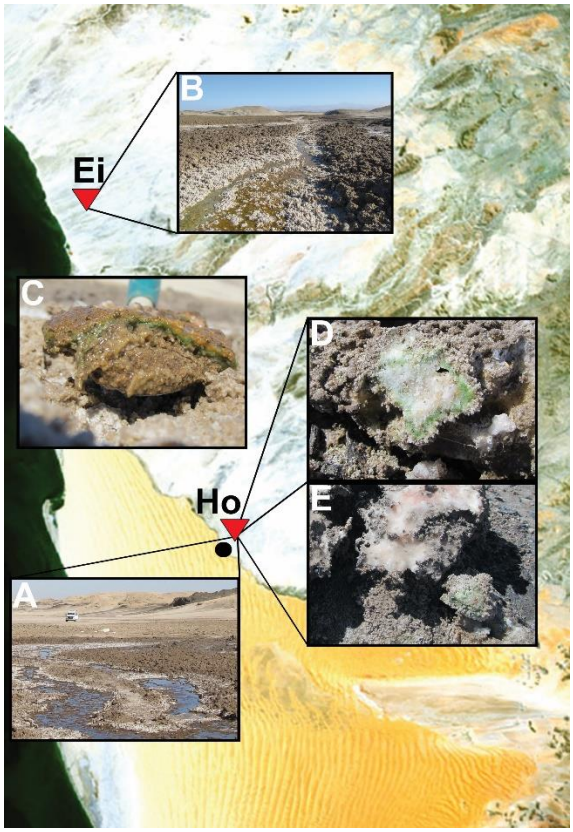
837 Zhang, W., Ding, W., Li, Y.-X., Tam, C., Bougouffa, S., Wang, R., Pei, B., Chiang, H., Leung, P., Lu, Y., et al.
838 (2019). Marine biofilms constitute a bank of hidden microbial diversity and functional potential. Nat
839 Commun *10*, 517.

840 (2015). FastQC.

841

842

843 ## Figures

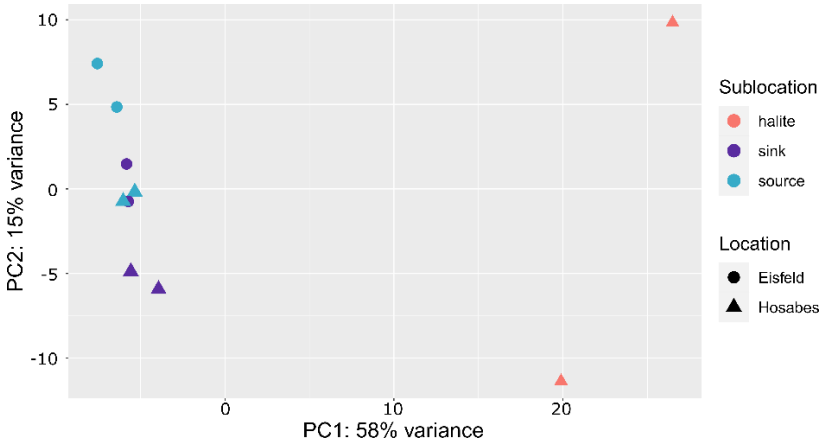


844

845 **Figure 1. Sampling location distribution.** The red triangles indicate the two sampled playas Hosabes (Ho)
846 and Eisfeld (Ei) and their location from the Gobabeb Research Centre (black dot). Insets depict the salt pan

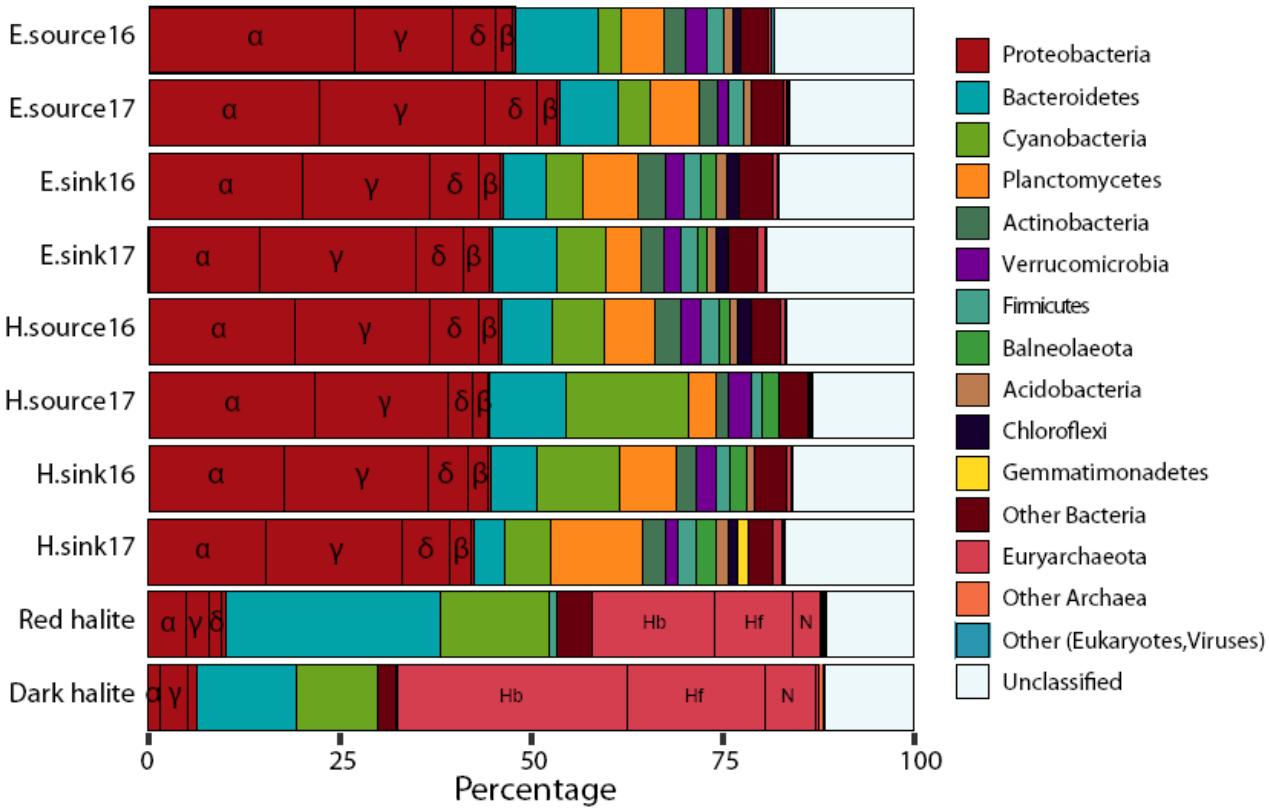
streams of Hosabes **(A)** and Eisfeld **(B)**, and a close-up of collected samples: Hosabes mat and **(C)**, and the dark **(D)** and red **(E)** halite rocks. Image adapted from ESA, CC BY-SA 3.0 IGO.

A



B

867



868 **Fig 2. Taxonomic profile of the microbial mat and halite samples. A)** Principal component plot of the
869 Namib stream mat and halite samples based on taxonomy at Order level. Taxonomic diversity was
870 compared by location (e.g., Eisfeld or Hosabes) and sublocation (e.g., source, sink or halite). **B)** Taxonomic
871 classification of the metagenomic open reading frames. *Proteobacteria* dominate the salt pan samples,
872 while members of the *Euryarchaea* are predominant in both halites. H – Hosabes, E – Eisfeld, α –
873 *Alphaproteobacteria*, β – *Betaproteobacteria*, γ – *Gammaproteobacteria*, δ – *Deltaproteobacteria*, Hb –
874 *Halobacteriales*, Hf – *Haloferacales*, N – *Natrialbales*.

875

876

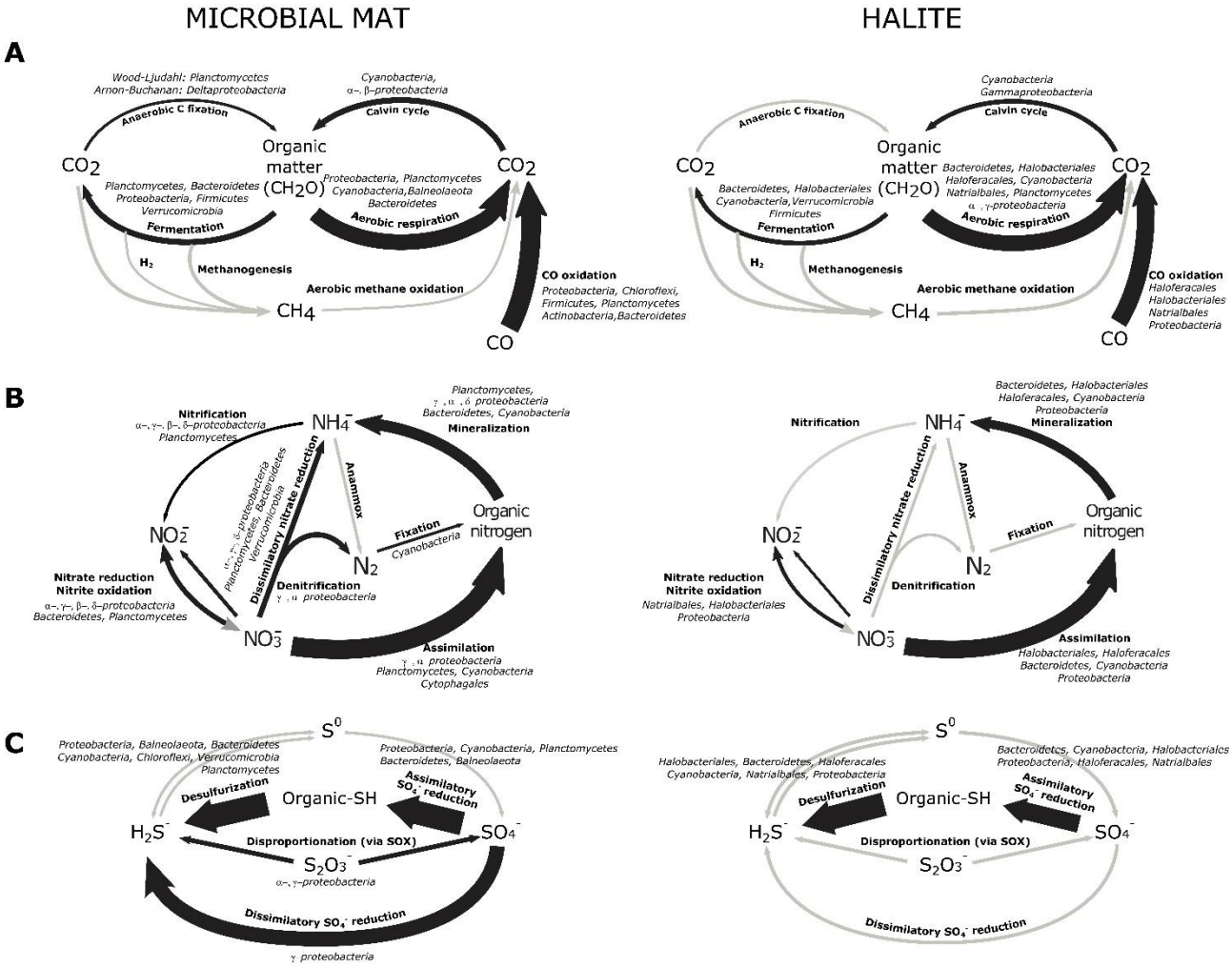
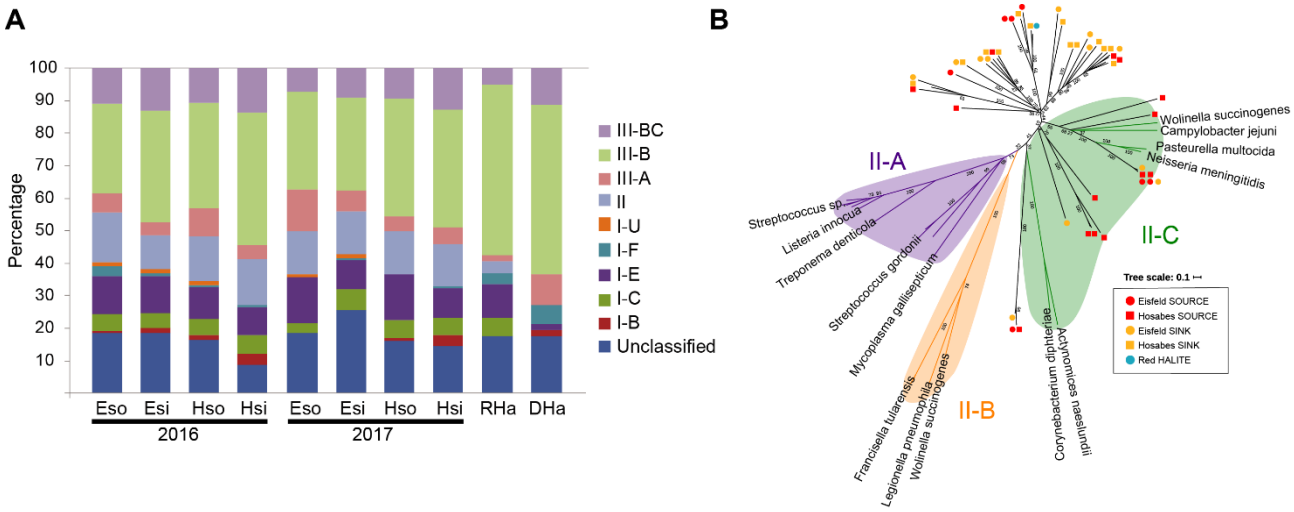


Figure 3. Functional potential of the Namib salt pan and halite microbial communities. Panels depict the carbon **(A)**, nitrogen **(B)** and sulfur **(C)** cycles from the stream mat (left) and halite (right) metagenomic data. Black arrows represent steps of the cycle present in the community, while grey arrows represent pathways not detected in the metagenomes. Thickness of arrows is proportional to the marker gene counts for each pathway. Main taxa with the genetic potential for each metabolic step are shown in italics.



886

887

888 **Figure 4. CRISPR-Cas type I and type III systems are widespread in the Namib Desert salt pan microbial**
889 **communities. A:** Relative abundance of the CRISPR-Cas systems in each of the metagenomic datasets. E –
890 Eisfeld; H – Hosabes; so – source; si – sink; RHa – red halite; DHa – dark halite. **B:** Phylogenetic tree of the
891 Cas9 protein sequences present in the Namib salt pan metagenomes. Colored branches indicate a
892 reference protein sequence obtained from the RefSeq protein database, while black branches indicate
893 protein sequences obtained from the metagenomic data. Sequences belonging to the three described Cas9
894 categories are shaded in purple (II-A), green (II-C) or orange (II-B). The symbols at the end of the branch
895 indicate the location within the salt pan from where the metagenomic sequence was obtained (red: Eisfeld,
896 yellow: Hosabes, blue: halite circle: source, square: sink).

897

898

899

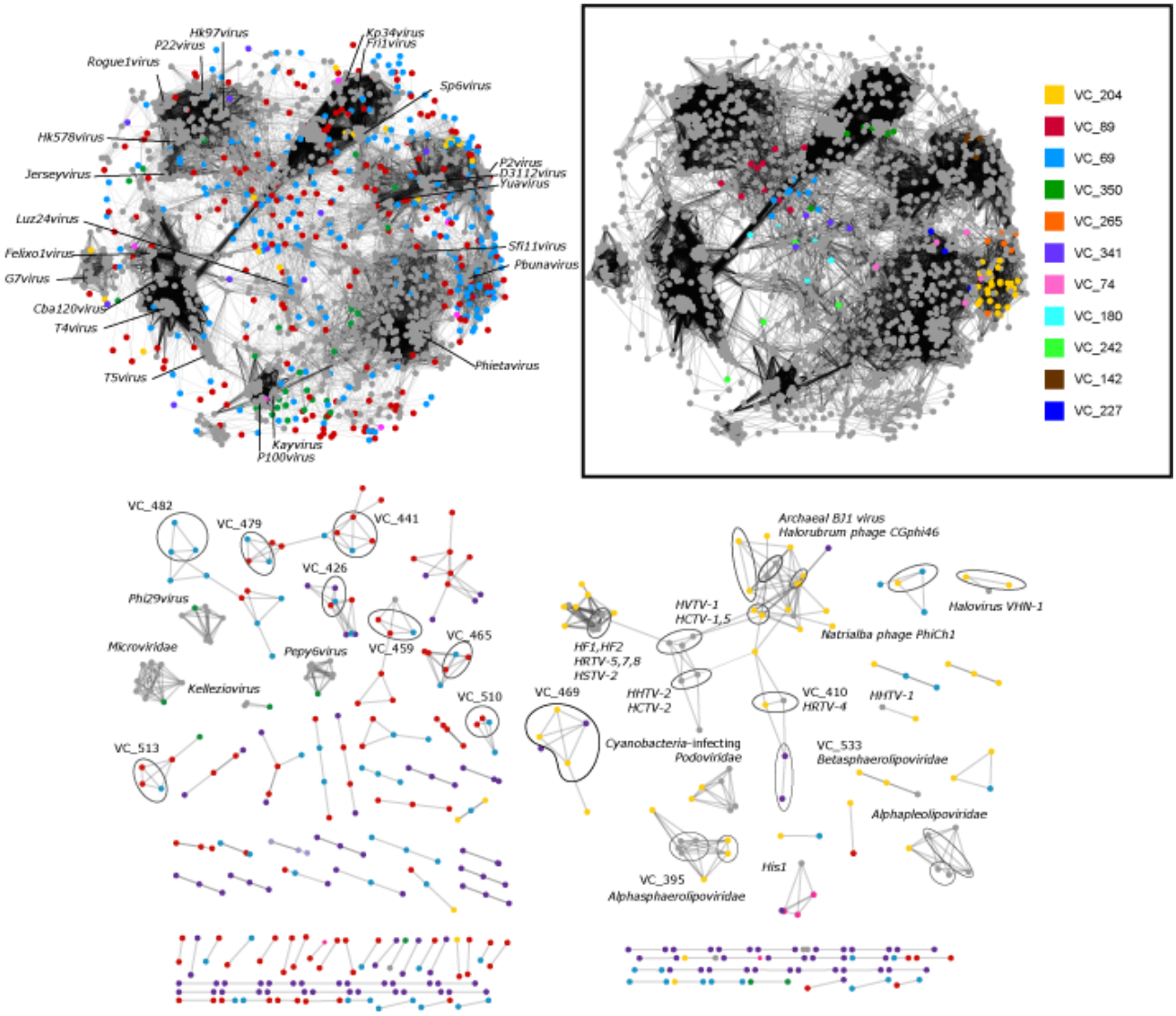


Figure 5. Genome-based network of shared protein content. Each node represents a viral genome and edges represent statistically significant relationships between the protein profiles of those viral genomes. Groups composed exclusively of RefSeq viruses were excluded for clarity. Viral clusters of interest are indicated with the prefix "VC" followed by their corresponding number, and/or circled in black. **Inset:** Network of *Caudovirales* indicating the largest viral clusters with mVir sequences.

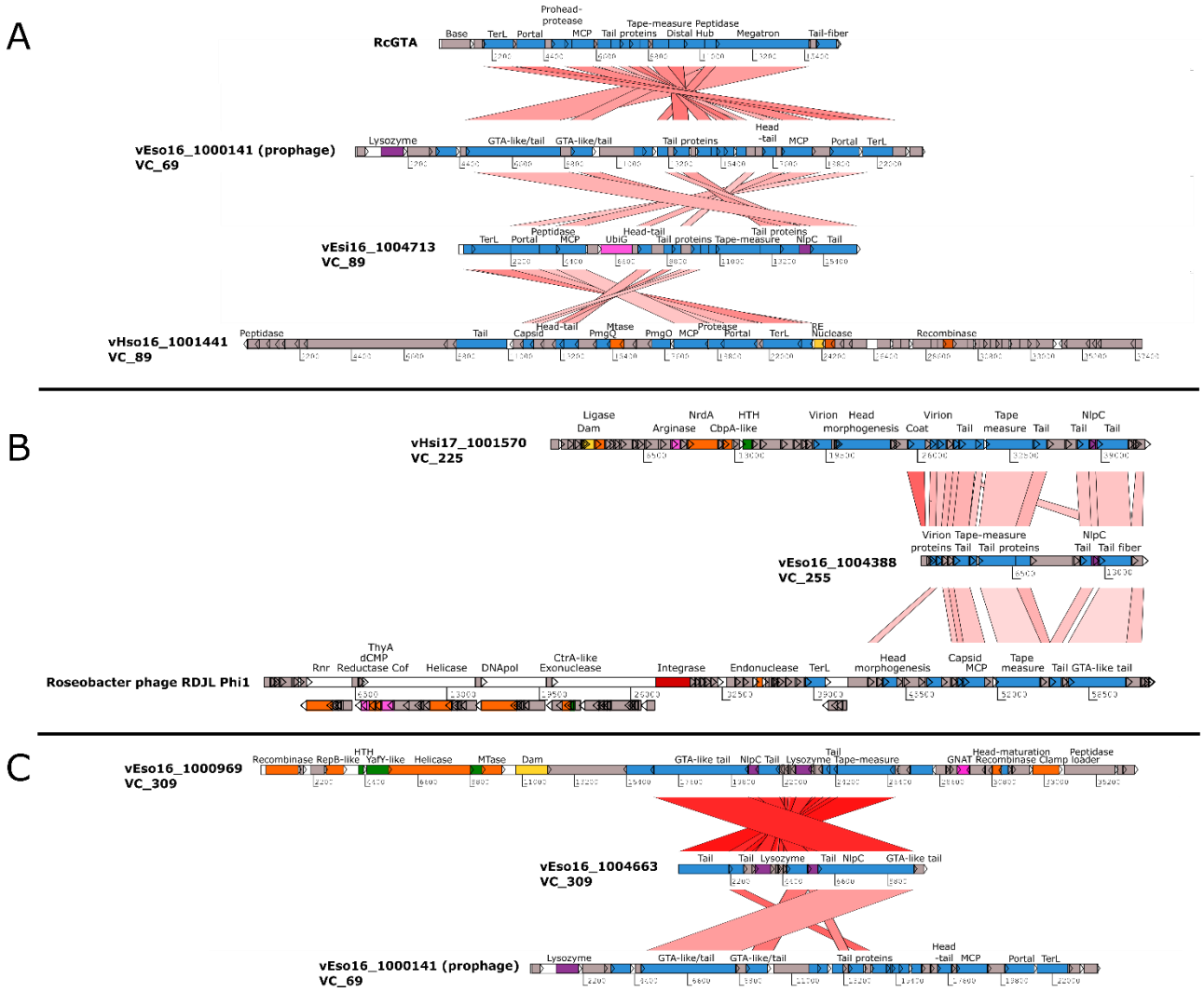


Figure 6. Overview of the genomic organization of GTA-like and virus-like contigs from mVir clusters with GTA-signal. A: Comparison of *Rhodobacter capsulatus* GTA (RcGTA) to the largest contig of VC_69 and GTA-like or virus-like contigs from VC_89. **B**: Comparison of VC_255 contigs to *Roseobacter* phage RDJL Phi1 virus. **C**: Genomic organization of VC_309 contigs. The shades of red represent pairwise protein similarity, with stronger red as the most similar. Genes are shown as boxes with an arrow indicating their orientation in the genome, and colored according to their assigned functional category: blue – virus structure and assembly, orange – DNA replication, green – transcription, purple – virus exit, yellow – defense mechanisms, red – integration, pink – metabolism.

927 **Table 1. Principal differences in taxonomic composition of salt pan and halite samples at family level.**
928 Fold change differences (log2FC) in taxon abundance between stream mat and halite samples. The column
929 "Type of sample" indicates in which community is the taxon enriched.

Phylum	Family	Log2FC	Type of sample
Proteobacteria – Alphaproteobacteria	*Rhodobacteraceae*	2.94	mat
	Rhodospirillaceae	1.78	halite
	Hyphomicrobiaceae	4.09	mat
	Parvularculaceae	4.34	mat
	Cohaesibacteraceae	2.84	mat
	Sneathiellaceae	2.88	mat
Proteobacteria – Betaproteobacteria	*Sterolibacteriaceae*	2.33	mat
Proteobacteria – Gammaproteobacteria	*Wenzhouxiangellaceae*	6.2	mat
	Halieaceae	3.92	mat
	Woesiaceae	2.55	mat
	Spongiibacteraceae	2.89	mat
Proteobacteria – Deltaproteobacteria	*Anaeromyxobacteraceae*	1.17	mat
Cyanobacteria	*Aphanothecaceae*	6.9	halite
	Synechococcaceae	4.79	halite
	Croococcidiopsidaceae	3.59	halite
	Hyellaceae	2.88	halite
	Chrococcaceae	2.82	halite
	Microcoleaceae	2.81	halite
	Sytonemataceae	2.78	halite
	Cyanothecaceae	2.54	halite
	Microcystaceae	2.51	halite
	Nostocaceae	2.5	halite
	Rivulariaceae	2.35	halite
	Hapalosiphonaceae	2.34	halite
	Merismopediaceae	2.16	mat
Planctomycetes	*Planctomycetaceae*	2.29	mat
Euryarchaea-Halobacteria	*Haloarculaceae*	10.22	halite
	Halobacteriaceae	9.84	halite
	Haloferacaceae	8.76	halite
	Halorubraceae	8.7	halite
	Halococcaceae	8.54	halite
	Natrialbaceae	8.47	halite
Euryarchaea-Methanomicrobia	*Methanosarcinaceae*	2.5	halite
Bacteroidetes	*Rhodothermaceae*	6.66	halite
Actinobacteria	*Acidimicrobiaceae*	3.13	mat
Verrucomicrobia	*Chthoniobacteraceae*	2.92	mat
	Verrucomicrobiaceae	2.35	mat
Acidobacteria	*Solibacteriaceae*	2.23	mat

930

931 **Table 2. Virus-host connectivity**. Total number of linkages (*n*) between mVir viral clusters and their
932 predicted hosts.

Host taxonomy		Viral cluster	Type of match	*n*
Actinobacteria-Actinobacteria-Streptomycetales		VC_650	CRISPR spacer	3
Bacteroidetes -unclassified		vHso16_1000846	CRISPR spacer	27
Cyanobacteria -unclassified		VC_254	CRISPR spacer	5
Deinococcus-Thermus-Deinococci-Thermales		VC_627	CRISPR spacer	3
Euryarchaeota-Halobacteria		VC_469, VC_604	CRISPR spacer	4
		VC_357, VC_390, VC_406, VC_469	Prophage	4
		Outliers/Singletons	Prophage	3
Gemmatimonadetes-Gemmatimonadetes-Gemmatimonadales		VC_480, VC_389	CRISPR spacer	5
		VC_650	Prophage	1
Lentisphaerae		VC_242	CRISPR spacer	1
Nitrospirae-Nitrospira-Nitrospirales		VC_426	CRISPR spacer	1
Planctomycetes		Outliers/Singletons	CRISPR spacer, Prophage	12, 6
		VC_471	CRISPR spacer	11
		VC_204, VC_180, VC_310	Prophage	6
Proteobacteria-Alphaproteobacteria	*Rhizobiales*	Outliers/Singletons	CRISPR spacer	3
		VC_350, VC_120	Prophage	3
	Rhodobacterales	Outliers/Singletons	CRISPR spacer, Prophage	4, 5
		VC_69, VC_143, VC_635, VC_309, VC_74	Prophage	10
	Rhodospirillales	VC_350, VC_120	Prophage	1
	unclassified	Outliers/Singletons	CRISPR spacer, Prophage	10, 11
		VC_69, VC_333, VC_391, VC_556, VC_653	Prophage	7
		VC_464, VC_573, VC_637, VC_89	CRISPR spacer	4
Proteobacteria-Betaproteobacteria-Burkholderiales		VC_89	CRISPR spacer	1
Proteobacteria-Deltaproteobacteria		VC_409	Prophage	2
		VC_530	CRISPR spacer	1
Proteobacteria-Gammaproteobacteria	*Alteromonadales*	VC_176, VC_341	CRISPR spacer	4
	Chromatiales	Outliers/Singletons	CRISPR spacer	14
		VC_387, VC_176, VC_509, VC_224	CRISPR spacer	10
		VC_649	Prophage	1
	unclassified	Outliers/Singletons	CRISPR spacer, Prophage	95, 3
		548, VC_387, VC_509, VC_176, VC_503, VC_515, V	CRISPR spacer	28
Verrucomicrobia-Opitutae		VC_337	Prophage	2

933

934