Prokaryotic and viral community of the sulfate-rich crust from Peñahueca ephemeral lake, an astrobiology analogue

Ana-Belen Martin-Cuadrado[1], Ece Senel[1,2], Manuel Martínez-García[1], Ana Cifuentes[3], Fernando Santos[1], Cristina Almansa[4], Mercedes Moreno-Paz[5], Yolanda Blanco[5], Miriam García-Villadangos[5], M. Ángeles García del Cura [6], M. Esther Sanz-Montero[7], J. Pablo Rodríguez-Aranda[8], Ramon Rosselló-Móra[2], Josefa Antón*[1], Víctor Parro [3]

[1] Department of Physiology, Genetics and Microbiology, University of Alicante, Alicante, Spain

[2] Department of Biology, Graduate School of Sciences, Eskisehir Technical University, Yunusemre Campus, Eskisehir 26470, Turkey

[3] Department of Ecology and Marine Resources, Marine Microbiology Group, Mediterranean Institute for Advanced Studies (IMEDEA, CSIC-UIB), Esporles, Spain

[4] Research Technical Services (SSTTI), Microscopy Unit, University of Alicante, Alicante, Spain

[5] Department of Molecular Evolution, Centro de Astrobiología (INTA-CSIC), Madrid, Spain.

[6] Institute of Geosciences IGEO (CSIC, UCM), Madrid, Spain

[7] Department of Mineralogy and Petrology, Faculty of Geology, Complutense University, Madrid, Spain

[8] Department of Didactics of Sciences, Faculty of Education, Complutense University, Madrid, Spain.

*Correspondence to Josefa Antón: anton@ua.es

Originality-Significance Statement

Here, we have characterized the viral and microbial assemblages inhabiting the salt crust from Peñahueca lake, an ephemeral lake originated under semiarid conditions in central Spain. The analyzed crust is very rich in sulfate minerals highly dehydrated and it is thus an analogue of extraterrestrial bodies such as Mars and Jupiter´s moon Europa. The astrobiological exploration of the ocean worlds is a priority for NASA, and the Europa-Lander mission is on the schedule for the coming years. Therefore, characterizing the geo-microbiology and the biomarkers associated with the sulfate-rich Peñahueca salt crusts can contribute to support and interpret the results from current (NASA's MSL) and coming (ESA's ExoMars and NASA's Mars 2020) planetary exploration missions to investigate the feasibility of life elsewhere.

SUMMARY

Peñahueca is an athalassohaline hypersaline inland ephemeral lake originated under semiarid conditions in the central Iberian Peninsula (Spain). Its chemical composition makes it extreme for microbial life as well as a terrestrial analogue of other planetary environments. To investigate the persistence of microbial life associated with sulfate-rich crusts, we applied cultivation-independent methods (optical and electron microscopy, 16S rRNA gene profiling and metagenomics) to describe the prokaryotic community and its associated viruses. The diversity for Bacteria was very low and was vastly dominated by endospore formers related to *Pontibacillus marinus* of the *Firmicutes* phylum. The archaeal assemblage was more diverse and included taxa related to those normally found in hypersaline environments. Several "metagenome assembled genomes" were recovered, corresponding to new species of *Pontibacillus*, several species from the Halobacteria and one new member of the Nanohaloarchaeota. The viral assemblage, although composed of the morphotypes typical of high salt systems, showed little similarity to previously isolated/reconstructed halophages. Several putative prophages of *Pontibacillus* and haloarchaeal hosts were identified. Remarkably, the Peñahueca sulfate-rich metagenome contained CRISPR-associated proteins and repetitions which were over 10-fold higher than in most hypersaline systems analyzed so far.

INTRODUCTION

Hypersaline environments are found around the globe (Oren, 2002; Oren, 2002; Andrei *et al.*, 2012) and are models of life in extreme (i.e. harsh) environments (Rothschild and Mancinelli, 2001). Among hypersaline systems, the microbial communities of those of marine origin (thalassohaline), mainly coastal solar salterns, have mainly been largely described using different approaches. High-throughput culture-independent methods (based on 16S rRNA gene sequencing or direct sequencing of the microbial DNA) described brines as habitats with low-species richness, often populated by dense communities of halophilic prokaryotes (from 10^6 to above 10^8 cells mL^{-1}) (Oren, 2002; Ghai *et al.*, 2011; Boujelben *et al.*, 2012; Podell *et al.*, 2013; Gomariz *et al.*, 2015; Di Meglio *et al.*, 2016; Naghoni *et al.*, 2017). Furthermore, hypersaline thalassohaline waters harbor the highest numbers of virus-like particles (VLPs) reported so far for aquatic systems (Santos *et al.*, 2010). These viral communities have also been analyzed through metagenomics and showed a high diversity although common traits have been found among viral metagenomes from hypersaline systems around the world (Santos *et al.*, 2007; Rodriguez-Brito *et al.*, 2010; Santos *et al.*, 2010; Sime-Ngando *et al.*, 2010; Santos *et al.*, 2011; Boujelben *et al.*, 2012; Emerson *et al.*, 2012; Garcia-Heredia *et al.*, 2012).

However, non-marine hypersaline or athalassohaline environments (with salt proportions different from those of seawater), despite being more diverse and abundant than those of marine origin, remain vastly unexplored in terms of their microbial community composition. Using culture-dependent and molecular approaches, several athalassohaline aquatic systems have been described; e.g., inland haloalkaline lakes (Sorokin *et al.*, 2011; Vavourakis *et al.*, 2016); ephemeral water reservoirs such as the Salton Sea, California, USA (Hawley *et al.*, 2014); two hypersaline meromictic lakes in the Transylvanian Basin, Romania (Andrei *et al.*, 2015); the sulfated-halophilic Tirez inland lake in Central Spain (Prieto-Ballesteros *et al.*, 2003; Montoya *et al.*, 2013), the magnesium sulfate- rich Keke Salt Lake in China(Han *et al.*, 2017), and the lithium and magnesium rich salt lake Salar de Uyuni (Haferburg *et al.*, 2017; Rubin *et al.*, 2017; Ramos-Barbero *et al.*, 2019). Non-aquatic saline formations have also been subjected to microbial composition analysis, such as the Great Salt Plains of Oklahoma, USA (Crisler *et al.*, 2012);several halite-rich deposits of the Atacama desert in Chile (halite,

evaporite domes, microbial mats and crusts) (Crits-Christoph *et al.*, 2016; Rasuk *et al.*, 2016); salt-crusts from evaporation ponds in Eilat, Israel (Oren *et al.*, 2009), and saline soils (Vera-Gargallo and Ventosa, 2018). In general, these non-marine hypersaline habitats have more diverse microbial communities than marine brines. Regarding the viruses associated with these environments, only a few genomic fragments assembled from a metagenome collection from Salar Grande (Atacama desert, Chile) (Crits-Christoph *et al.*, 2016) and viral metagenomes from the Salar de Uyuni solar lake (Ramos-Barbero *et al.*, 2019) have been described so far.

In the present study, we have characterized the prokaryotic and viral communities from a sulfate-rich salt crust (PH1-SC) formed during the dry season in the Peñahueca lake (Toledo, Spain) (Fig. 1). This is a shallow saline lake with a maximum area of 150 ha and a semi-permanent hydric regime that leads to desiccation in the months of highest water deficit when salt precipitation occurs, leaving the characteristic polygonal landscape of a salar. The lake lies on Mesozoic terrains (Triassic, Keuper facies) enriched in sulfates and halites that emerge to the surface after dissolving in subterranean waters (Macau and Riba, 1965).

In addition to its interest as an uncharacterized extreme microbial habitat, Peñahueca lake has relevant implications in astrobiology as a terrestrial analogue to assess the feasibility of life and the preservation of biomarkers on Mars and Jupiter´s moon Europa. Large halite and sulfate-containing mineral deposits have already been detected on Mars remotely (Gendrin *et al.*, 2005; Gaillard *et al.*, 2013) as well as sulfate veins at crater Galeby via *in situ* analysis (Milliken *et al.*, 2014; Schwenzer *et al.*, 2016). The enrichment in sulfate minerals with different hydration degrees makes Peñahueca lake also a good model to study the magnesium sulfate-rich brines detected on the surface of Europa by the Galileo's near infrared mapping spectrometer (McCord *et al.*, 1998; Prieto-Ballesteros *et al.*, 2003; Orlando, 2005) or inferred through theoretical models (Zolotov and Shock, 2001). The astrobiological exploration of the ocean worlds (Europa or Saturn's moon Enceladus) is a priority for NASA, and the Europa-Lander mission is on the schedule for the coming years (Hand *et al.*, 2017). Therefore, characterizing the geo-microbiology and the biomarkers associated with the sulfate-rich Peñahueca ephemeral lake can contribute to support and interpret the

results from current (NASA's MSL) and coming (ESA's ExoMars and NASA's Mars 2020) planetary exploration missions to investigate the feasibility of life elsewhere.

Here, we carried out an in-depth characterization of the microbial community present in the sulfate-rich saline-crust in the Peñahueca athalassic lake. We used a multi-phasic approach including *in field* immunological detection with the antibody microarray Life Detector Chip (LDChip) (Rivas *et al.*, 2008; Parro *et al.*, 2011), microscopy analyses, cellular and viral metagenome characterization, as well as 16S rRNA gene amplicon sequencing. To delve into an understanding of the effect of environmental parameters on microbial and viral assemblages, we compared the community profiles with those from other hypersaline habitats around the world. We reconstructed several draft genomes of new microbial species and explored their metabolic potential in biogeochemical cycles. The virus-host interactions were also investigated, particularly through the prokaryotic CRISPR systems.

RESULTS AND DISCUSSION

Geochemical characterization

To perform a geo-microbiological characterization of Peñahueca lake, different samples were collected during the dry season in July 2011 (Supporting Information Table S1). A minimal part of the lake still retained some liquid water in the more depressed parts. We sampled the top surface (<1 cm depth; PH1-SC and PH2-SC), and below (2-5 cm depth; PH1-2cm and PH2-2cm and PH3-core) along a transect from the driest shore to the wetter central area (Figure 1). Mineralogical analyses (Supporting Information Table S1) showed a predominance of sulfate-bearing minerals, in agreement with previous work revealing that this area lays on evaporitic materials from the Triassic (Yébenes *et al.*, 1977). Particularly, sample PH1-SC was the only one containing the most dehydrated magnesium sulfate minerals we found: 5% (w/w) kieserite ($MgSO_4 \bullet (H_2O)$) and 70% (w/w) starkeyite ($MgSO_4 \bullet 4(H_2O)$), two of the most important in understanding the Martian water cycle (Chou and Seal Li, 2007). Therefore, we selected this driest, brownish, and most superficial and marginal sample for in-depth analysis of the prokaryotic and viral assemblages because: i) it covers a relatively large area of the lake every dry season, which highlights its relevance in the

system; ii) it showed the highest concentration of magnesium sulfate minerals at the lowest hydration stage, which can be considered a good terrestrial analogue for the sulfate-rich mineral regions on Mars (Vaniman *et al.*, 2004; Chou and Seal Li, 2007), and finally; iii) very few studies have been reported about the metagenome and viral communities of sulfate-rich salt settings. The chemical composition and the 16S rRNA based microbial characterization of the rest of the samples have been published elsewhere (Mora-Ruiz *et al.*, 2018).

On site detection of microbial biomarkers

During the campaign, several samples were analyzed *on site* with an antibody microarray (*Life Detector Chip*, LDChip) with more than 300 antibodies for detecting microbial markers (Rivas *et al.*, 2008; Parro *et al.*, 2011). Positive immuno-detections were detected in PH1 samples, indicating the presence of microbial material, mostly bacterial cells and exopolymeric substances from *Gammaproteobacteria* subclass related to sulfur oxidation metabolism (Supporting Information Fig. S1). Interestingly, the LDChip results clearly showed two well-differentiated layers: the superficial crust (PH1-SC), characterized by the absence of positive immunodetections of microbes usually inhabiting NaCl-enriched brines (e.g. *Salinibacter* sp.); and the lower sample PH1-2cm, where these microbes were detected (*Salinibacter*, *Colwellia* and *Streptomyces*), together with some markers related to strictly anaerobic archaea (*Methanobacterium* spp.). This might indicate the presence of anoxic micro-environments, where methanogenesis might be fueled by the acetate and formiate detected by ion chromatography in the 2 cm below the surface sample (data not shown). The absence of immunodetection against "typical" halophilic prokaryotes in the PH1-SC might be due to either their absence or their occurrence at a concentration below the limit of detection of the immunoassay. Additionally, we detected oligomers containing the disaccharide NAG-NAM (N-acetylglucosamine-β-(1-4)-N-acetylmuramic acid) in PH1-SC while no signal was found in the lower layer (Supporting Information Fig. S1). This could be the result of peptidoglycan degradation (Reith and Mayer, 2011) from abundant Gram-positive Firmicutes strains (see below). In any case, the presence of two different communities inhabiting the two layers of the PH1 sample was in agreement with the mineralogical analysis.

Microbial and viral numbers in Peñahueca salt crust

Counts of DAPI-stained cells showed a value of $9.64\pm0.9\cdot10^{6}$ cells per gram of PH1 salt crust (thereafter PH1-SC), which is within the range observed (as cells/ml) in hypersaline environments (Antón *et al.*, 2000; Burns *et al.*, 2004; Mutlu *et al.*, 2008; Di Meglio *et al.*, 2016). The proportion of Archaea and Bacteria in the crust was measured by FISH (Supporting Information Fig. S2). The results indicated that long thick rods stained with the bacterial probe constituted more than 70% of the total cells detected while the archaeal probe gave a very poor signal. Although normally Archaea outnumber Bacteria in many coastal hypersaline environments worldwide, this is not the case in other inland salt lakes (Di Meglio *et al.*, 2016). However, the low detection of Archaea by FISH might be due to a low level of metabolic activity in this part of the prokaryotic community (Amann *et al.*, 1995; Pernthaler and Amann, 2004; Gomariz *et al.*, 2015).

Viruses were recovered from the saline-crust and SYBR gold-stained. Counting showed $3.5\pm0.22\cdot10^{9}$ virus-like particles (VLPs) per gram of salt, a value of the same order as those described in brines from marine salterns, which harbor the highest viral densities reported so far for aquatic systems (Santos *et al.*, 2010; Baxter, 2011; Boujelben *et al.*, 2012; Ventosa *et al.*, 2015). Considering the number of cells detected by DAPI staining, the virus-host ratio in the PH1-SC sample resulted in more than 100 free VLPs per cell, a value among the highest described so far and very similar to those observed in the Argentinian salterns of Guatraché, a hypersaline system that also has a high sulfate concentration (Di Meglio *et al.*, 2016).

Transmission electron microscopy was used to ascertain the different viral morphotypes and their relative abundances (Fig. 2). A total of 1131 VLPs were analyzed and three main morphologies were observed: tailed viruses from the order *Caudovirales* (25% long-tailed and 7.5% short-tailed), spindle viruses (fusiform or lemon-shaped, 33%), and polyhedrical (27%). These morphologies have been observed in several salt-saturated systems (Guixa-Boixereu, 1996; Oren *et al.*, 1997; Sime-Ngando *et al.*, 2010; Garcia-Heredia *et al.*, 2012); specifically, the lemon-shaped viruses are believed to infect haloarchaea. Other less abundant morphologies, such as

filamentous VLPS, were also observed. Given their high morphology similarities with VLPs (Gill *et al.*, 2018), the possibility exists that some of the extracellular particles found in PH1-SC were indeed extracellular vesicles, or gene transfer agents (GTAs). No cells were observed in the TEM preparations.

Prokaryotic community structure in Peñahueca salt crust: 16S rRNA gene profiling

The microbial community of PH1-SC was analyzed by 16S rRNA gene profiling using both amplicon and whole metagenome sequencing. The first step in the characterization was the description of the community structure by means of Operational Phylogenetic Units (OPUs) assigned to the amplicon sequences and metagenomic reads containing 16S rRNA gene fragments. Results based on Good's coverage values (close to 1) from amplicon sequences indicated that our study recovered the existing diversity in PH1-SC (Table 1). Bacterial sequences obtained from amplicons were grouped into 5 OPUs with more than 87% of the reads related to the genus *Pontibacillus*, within the *Firmicutes* phylum.

A similar snapshot of the bacterial diversity was obtained when taxonomic microbial assignment was performed with the whole metagenome (Fig. 3): 96% of the bacterial metagenomic reads were affiliated to *Firmicutes*, 88% of which had their best match to the above mentioned *P. marinus* DSM 16465.This genus of endospore forming rods encompasses seven moderately halophilic species that grow at NaCl concentrations from 1 to 20% (Lim *et al.*, 2005; Chen *et al.*, 2009; Chen *et al.*, 2010; Huang *et al.*, 2015), although some strains of *Pontibacillus marinus* are able to grow up to 30% of salinity (Sass *et al.*, 2008). Strains of this genus have been isolated from salt lakes (Daas MS et al., 2017, Hua et al., 2008, Jeon et al., 2008), saline soils (Yang *et al.*, 2011; Lee *et al.*, 2015), and deep-sea hypersaline anoxic sediments (DHALs) of the Mediterranean Sea (Sass *et al.*, 2008). DHALs also harbor a high concentration of magnesium and sulfate and are considered to be one of the most extreme hypersaline systems in the planet (La Cono *et al.*, 2011; Yakimov *et al.*, 2013; La Cono *et al.*, 2019). Other *Firmicutes* have been found in several hypersaline environments accounting for different percentages of the bacterial 16S rRNA sequences: e.g. 15% in

Punta Cormorant (PC6), 6% in the medium-salinity pond SS19 from a marine saltern, 0.25% in the Salar Grande collection or between 2.8-0.18% in the Lake Mayghan samples (Ghai *et al.*, 2011; Crits-Christoph *et al.*, 2016; Naghoni *et al.*, 2017). Remarkably, the bacterial community of Keke Salt Lake (China), rich in magnesium and sulfate, was also dominated by *Firmicutes* that accounted for 74-81% of the 16S rRNA gene sequences, from which 51-58% belonged to *Bacillus* species (Han R *et al.*, 2017). To our knowledge, none of the saline metagenomes analyzed so far presented the dominance of *Pontibacillus* sp. found in the Peñahueca sulfate-rich crust, indicative of its peculiar microbial community composition.

The Shannon-Weiner index for the partial 16S rRNA archaeal amplicons showed a higher diversity than that of Bacteria (Table 1). A total of 168 representative archaeal sequences were grouped into 29 OPUs, 27 of which belonged to the *Euryarchaeota* phylum and the Halobacteria class. Metagenomics reads also pointed to this class as the predominant archaeal group (98% of the archaeal reads). Both 16S rRNA gene amplicons and metagenomics data showed that the archaeal community was dominated by the following genera: *Halorussus* (17%), *Natronoarchaeum* and *Haloterrigena* (each 16% of the sequences), *Halohasta* and *Halolamina* (10% each) and *Halorubrum* (4%). About 9% of the sequences were *Nanohaloarchaeota*.

Here, it is relevant to recall that, in order to perform DAPI and SYBR counts and to separate cells from viruses prior to DNA extraction, the salt crust had to be previously dissolved (see Experimental Procedures). Although we do not expect this sample manipulation to have a strong impact on the description of the community in PH1-SC, some biases regarding prokaryotic and virus diversity cannot be ruled out. In any case, FISH analyses were performed in samples that were fixed within 1 hour from beginning the dissolution process and their results were compatible with the dominance of *Pontibacillus* sp. discussed above, which validates the results of the DNA-based analyses of the community.

Metagenomic analysis of the Peñahueca salt crust: general features and metabolic profiling

A Meta-fast comparison (see Experimental Procedures) of PH1-SC metagenome with metagenomes from different hypersaline environments highlighted its singular microbial community composition. As shown in Supporting Information Fig. S3, Peñahueca appears in a separate branch within a cluster including the athalassohaline environments Salar Grande (Crits-Christoph *et al.*, 2016), "red" and "white" samples from Meyghan Lagoon (Naghoni *et al.*, 2017) as well as the marine saltern at Isla Cristina (Huelva, Spain) (Fernandez *et al.*, 2013). BLASTN comparisons indicated that only the samples from the Meyghan Lagoon and Salar Grande metagenome presented some similarities (always below 20%) with PH1-SC.

GC-content of the assembled contigs and the metagenomic reads presented a bimodal curve with a higher peak centered at 67% and a lower peak at 38.5% (Fig. 4 and Supporting Information Fig. S4). As found in other halophilic metagenomes, the high-GC peak was associated to the haloarchaea (by their best match hit against the nr-database), while most of the ORFs corresponding to the 38.5% GC peak had high similarities with *Pontibacillus* sp. genes, in good agreement with 16S rRNA gene data. Other hypersaline systems, such as the Santa Pola marine saltern metagenomes SP-37, SP-19 and SP13 (Alicante, Spain) (Ghai *et al.*, 2011); the intermediate saline pond (IC21) from another marine saltern (Huelva, Spain) (Fernandez *et al.*, 2013); the metagenome from Salar Grande (Chile) (Crits-Christoph *et al.*, 2016); and the Kulunda soda brines samples Tanatar-5, Picturesque Lake, Tanatar trona crystallizer and Bitter-1, (Vavourakis *et al.*, 2016)) also displayed this bimodal profile. However, the PHSC1 low GC peak was unusually low.

Regarding the metabolic profile, KEGG and COG annotation data showed that heterotrophy was the dominant life strategy, given that genes involved in glycolysis, amino acid transport (*livF*), sugar transport (saccharide and polyols; *msmX, msmK, malK, sugC, ggtA* and *msiK*), and peptide and oligopeptides transport systems (*dppC-D* and *appF*) were found. Furthermore, the complete pathways of gluconeogenesis, the complete tricarboxylic acid cycle, and oxidative phosphorylation were reconstructed (Supporting Information Fig. S5) while genes related to CO_2 fixation were not detected in the metagenome. Archaeal light-driven retinal-binding proteins (rhodopsins) such as the outward proton pump bacteriorhodopsin, the inward chloride pump halorhodopsin

(related to salt-in strategy, see below) and sensory rhodopsins were also found in high numbers in the PH1-SC metagenome. These membrane proteins are widespread within the Archaea and provide a way of transducing light into metabolic energy.

To cope with osmotic stress, microorganisms in hypersaline environments have developed two types of strategies: "salt-in" and "salt-out", that can be combined in the same organism (Oren, 2013). Salt-in strategists accumulate high concentrations of inorganic ions (mainly K^+) and have their intracellular machinery adapted to function in a high salt environment. This implies that salt-in strategists have acidic proteomes. On the other hand, salt-out strategists accumulate in their cytoplasm compatible solutes, which are organic compounds (such as glutamine, proline, glycine and betaine) compatible with cell metabolism that counterbalance external osmotic pressure (Oren, 2008). In general, extremely halophilic Archaea and Bacteria belonging to the Haloanaerobiales, the *Salinibacte*r assemblage (Oren, 2008) use the salt-in strategy while other extreme and most moderate halophiles accumulate compatible solutes. Other organisms, like members of the genus *Halorhodospira* (Deole *et al.*, 2013), have acidic proteomes and accumulate inorganic ions only when grown in high salt medium.

Metagenomic analysis indicated that both salt-in and salt-out strategists were present in the PH1-SC microbiome, as expected based on its phylogenetic composition. Genes involved in the transport and synthesis of compatible solutes were detected, including several coline/glycine/proline transporter, genes involved in the synthesis of glycine-betaine osmolyte (i.e. *putA*-like genes), and genes involved in proline production for osmotic purposes, like the crotonobetainyl-Coa reductase CaiA (Belitsky *et al.*, 2001; Saum and Müller, 2007). These two genes were also detected in the *P. marinus* genome. Furthermore, data from the isoelectric point of the translated metaproteome indicated that about 60% of the proteins had a pI of 4 (Supporting Information Fig. S6). This value is one unit lower than the pI calculated for the microbial community inhabiting a salt-saturated crystallizer pond from Santa Pola marine salterns, an environment which is typically rich in salt-in strategists (Ghai *et al.*, 2011). Furthermore, traits related with general osmo-response were also found, such as the presence of Na^+/H^+ antiporters (*nhaC* family) and voltage-gated potassium channels (*kch, trkA, mthK, pch*). Finally, other gene transporters dealing with the potential toxicity

of metals and other compounds, such as *zntA* and ABC CcmA and MdlB-like multidrug transporters (Kertesz and Wietek, 2001; Pilsyk and Paszewski, 2009), were also detected.

Metagenome assembled genomes

A total of 5122 contigs larger than 5 kb were assembled (Table 1), 73% of which had a GC-content between 60 to 70%, in good agreement with the GC profile discussed above. After classifying the contigs to the top-level taxa, only around one fifth could be clearly assigned and more than 50% of them (550 contigs) belonged to *Pontibacillus* sp.

In an attempt to recover metagenome assembled genomes (MAGs), contigs were binned as described in Experimental Procedures. A total of 7 bins (Table 2, Supporting Information Table S2) were retrieved with different degrees of quality according to (Konstantinidis *et al.*, 2017). These authors established that, among other characteristics, a metagenome assembled genome or MAG should only be considered good when its completeness is above 80% and presents less than 5% of contamination. ANI values of the recovered bins to their closest relatives indicated that all except one (identified as *Halolamina rubra*) may correspond to new species of haloarchaea. One of the bins was similar to members of the *Nanohaloarchaea*, which is a major uncultured group widely spread in hypersaline ecosystems (Ghai *et al.*, 2011; Narasingarao *et al.*, 2012). Although only 37% of the genome was recovered in this bin, it may represent a new genus (ANI lower than 68%, Supporting Information Fig. S7A). The phylogenomic affiliation using seven conserved genes (Supporting Information Fig. S7B), placed this bin in a clade with *Candidatus* Nanosalinarum and in the same cluster as some of the *Nanohaloarchaea* assembled from soda lakes (Vavourakis *et al.*, 2016).

Most recovered bins had a relatively low contribution to the overall community in PH1-SC although one of them represented a quantitatively relevant population since it recruited 5.9% of the metagenomics reads. *My_Taxa_Scan,* MIGA-tool, (Rodriguez *et al.*, 2018) together with an in-depth phylogenomic analysis of 163 concatenated

proteins indicated that this bin may represent a new *Halobacteriaceae* genus placed in a separate branch with members of *Haladaptatus* (Fig. 5, Haloarchaeon PH1-SC).

In spite of the abundance of *Pontibacillus*-like sequences, no bin was obtained for this species with the above-mentioned approach. This phenomenon has been previously described in hypersaline environments and may be due to a high intraspecific diversity which hampers correct assembly (Ramos-Barbero *et al.*, 2018). BLASTX comparisons of the ORFs from the contigs against the *P. marinus* reference genome allowed the detection of at least two close subpopulations of this genus (A and B, Supporting Information Fig. S8A) with ANI values of 79% and 91.3%, respectively, to the reference *P. marinus* DSM 16465 genome. According to the proposed and generally accepted species boundary of 95-96% for ANI (Richter and Rossello-Mora, 2009), these populations A and B could represent two new species of *Pontibacillus*. Their inferred phylogeny indicated that both were related to *P. marinus* DSM 16465 but separated from *P. yanchengensis* Y32 (Supporting Information Fig. S8B).

Interestingly, the analysis of the pI of both subpopulation proteomes was typical of "salt-in" strategists (Supporting Information Fig. S8B). The comparison of the available *Pontibacillus* sp. proteomes showed that, except for *P. chungwhensis* and *P. yanchengensis*, all pIs were less than 5 in more than 40% of their proteins. Among the *Pontibacillus* contigs from PH1-SC, several genes for Na^+/H^+ antiporters and Trk potassium transporters were found. In parallel, genes for the synthesis of osmolytes such as glycine and betaine, together with genes for their transport, were also detected. Thus, the possibility exists that these *Pontibacillus* subpopulations use a mixed strategy to regulate their osmolarity, as previously described for other prokaryotes (Oren, 2008). In any case, some data indicate that an acidic proteome is more a requirement than a proof of salt-in strategy (Deole *et al.*, 2013). Indeed, there are known salt-out strategists with acidic proteomes, such as *Spiribacter salinus* M19-40 (Lopez-Perez *et al.*, 2013) (Supporting Information Fig. S8B). The question of whether *P. marinus* and the *Pontibacillus* populations in Peñahueca are indeed (partly) salt-in strategists definitely deserves further research.

Viral Community Structure – **Metavirome**

Virus-like particles (VLPs) were separated from the cellular fraction (Supporting Information Fig. S9) and their DNA extracted, purified and sequenced as described below, yielding a total of 99 Mb, 7% of which could be assembled in a total of 9065 contigs with only 87 of them larger than 5 kb (Table 1). These sizes are distant from the PFGE results which showed that most abundant viral genomes in PH1-SC ranged from 23 to 150 kb with a more intense band at 40 kb (Supporting Information Fig. S9), which is the typical genome size observed in PFGE preparations of halovirus communities.

As expected, PH-SC1 cell and virus metagenomes had a strong correlation in terms of GC content (Supporting Information Fig. S4, with a slight shift to lower GC values for the viral sequences. This deviation (around to 4%) between the GC-content of viral sequences and those corresponding to their microbial hosts has been previously observed (Rocha and Danchin, 2002) also in hypersaline environments (Santos *et al.*, 2010).

Overall, PH1-SC virus metagenome (hereinafter metavirome) showed a high degree of novelty since less than 0.01% of the PH1-SC metavirome reads matched viral assemblages from other hypersaline environments worldwide, such as Lake Tyrrell, in Victoria, Australia, (Emerson *et al.*, 2013), Lake Retba in Senegal (Sime-Ngando *et al.*, 2010), coastal salterns in San Diego in the USA (Rodriguez-Brito *et al.*, 2010), Santa Pola in Spain (Santos *et al.*, 2010), and Sfax in Tunisia (Boujelben *et al.*, 2012).

Metavirome assembled contigs were submitted to VirSorter (Roux *et al.*, 2015) and only 178 were classified as reliable viral genomes, probably due to their short length and the consequent lack of viral-signal. Accordingly, only a small fraction of the ORFs in contigs (5.7%) found a viral homolog in previously described virus by BLASTX comparisons against the nr database (NCBI). A large proportion (86%) of the ORFs in contigs coded for unknown proteins, as frequently found when analyzing viral metagenomes (Brum *et al.*, 2016). Some of these unknown proteins had their best matches with specific genomic regions of different strains of *Halorubrum* (11.7%), *Haloferax* (9.7%) and *Bacillus* (7.5%), which indicated these viral contigs may belong either to contamination with cellular genomes (unlikely given the annotation data, see below), to temperate phages or just to viral contigs for which a "viral best hit" does not

exist in the databases yet. In good agreement with this second possibility, 341 contigs had integrases.

Since only 7% of the reads from the metavirome were assembled into contigs, the ORFs from the metavirome reads were also extracted and functionally classified accordingly to the phage-orthologous-clusters (POGs, (Kristensen *et al.*, 2013)). Confirming the novelty of the Peñahueca virus, only 434 proteins were classified. The best-represented categories were integrases (8.5%) together with transposases (3.4%). Other abundant proteins were terminases (6.7%), followed by structural proteins, such as portal (4.7%) and tail-like proteins. Finally, the annotation of the extracellular metavirome unveiled the intriguing presence of a high number of genes coding fosphosphoadenosine phosphosulfate reductases (PAPS reductase, COGs 0175 and 3969), which were around 4-fold more abundant in the PH1-SC metavirome than in the metagenome (Tables 3 and 4). This enzyme participates in the sulfur cycle and is thought to impart selective advantages to cells by facilitating inorganic sulfate assimilation (by reducing PAPS to phosphoadenosine phosphate (PAP) and liberating sulfite). Viral PAPS reductases have been found in some viral genomes as well as in prophages of *Bacillus* related species (Summer *et al.*, 2007; Garcia *et al.*, 2008; Farlow *et al.*, 2018). In the PH1-SC metavirome, PAPS reductases were around 6-7-fold more abundant than in other hypersaline viral metagenomes (data not shown). The open question is whether this "auxiliary metabolic gene" actually plays a role in the host fitness.

BLASTN comparisons (see methods) revealed that 86% of the viral reads and 10% of the viral contigs (*cut-off*: 90% id in 100% of their length) were also present in the PH1-SC metagenome. Conversely, 13% of the cellular metagenomic reads and 6% of contigs (>5 kb) were present in the extracellular metavirome (Supporting Information Fig. S10). This could be due to (i) cross contamination between the cellular and the viral fraction or (ii) the presence of viruses in the cellular fraction, either attached to the cells or as replicating or integrated genomes. Although it cannot be ruled out, the lack (or very low levels) of typical microbial housekeeping genes among the viral reads indicated low contamination by cellular DNA (Table 3). More specifically, only 3 rRNAs (two 16S rRNAs and one partial 23S rRNA, i.e. 0.0016%) were found in reads (and

none in contigs) using RNA-scan. This is way below the threshold of 0.02% stablished by (Roux *et al.*, 2013) to consider the amount of cellular sequences "as very low and likely to be negligible".

To improve the retrieval of viral sequences from PH1-SC, contigs assembled from the metagenome that were also present in the extracellular metavirome were considered to be of viral origin. However, to increase the stringency of the analysis, only contigs that had a recruitment 3x higher in the metavirome than in the metagenome (Supporting Information Fig. S10) were selected (Supporting Information Table S2). Using this approach, 371 contigs (ranging from 5 to 42.5 kb) corresponding to viruses were retrieved from the metagenome. The viral nature of those contigs was further confirmed by the COG functional classification of their ORFs, which was very similar to that of the contigs assembled from the metavirome (Supporting Information Fig. S11).

These "cell-fraction associated" viruses included viruses infecting haloharchaeal genera as well as several *Bacillus* phages (Supporting Information Fig. S12), some of them carrying integrases. In fact, *Bacillus* spp. genomes frequently carry several prophages (Kunst *et al.*, 1997). Interestingly, five of these viral contigs related to *Bacillus* phages contained an acetylmuramoyl-L-alanine amidase with more than 70% similarity with the one present in the chromosome of *P. marinus* DSM 16465 (Fig. 6). This gene, together with their neighbor holins and the transmembrane proteins, resemble the lysis cassettes described in *Bacillus* (Nakonieczna *et al.*, 2015). Therefore, due to the high specificity of these lytic enzymes, it is very likely that these five contigs belong to temperate phages of *Pontibacillus*, for which no virus has been reported yet in the literature.

CRISPR-Cas systems

Since CRISPR-Cas systems can be used to investigate viral and host encounters in nature (Emerson *et al.*, 2013; Jackson *et al.*, 2017), we searched for the presence of these repeats within PH1-SC metagenome and metavirome.

First, we looked for the presence of CRISPR-like repetitions and Cas related proteins among a subset of 10^5 reads of the metagenome (Table 4). Surprisingly, the PH1-SC

metagenome, together with the Meygham samples, had the largest number of reads containing CRISPR-like repetitions (Fig. 7 and Table 4) within all the analyzed hypersaline environments. From this subset, a total of 75 different repeats and 188 unique spacers were obtained after clustering (*cut-off*: 97% identity in 100% of the length). When the metagenome reads containing CRISPR-like repetitions were compared against the nt database (NCBI), best hits of the most abundant repeats included several haloarcheal plasmids; i.e, *Halobacterium* sp. DL1 plasmid (97% identity), PL131 from *Natronomonas pharaonis* DSM 2160 (96%), plasmid pSTJ001 from *Halobacterium hubeiense* JI20-1 (91%) or plasmid pHmuk01 of *Halomicrobium mukohataei* DSM 12286 (88%). This finding supports the previous observation that dissemination of the CRISPR systems through plasmids is a widespread strategy (Godde and Bickerton, 2006).

Around 20% of the CRISPR-spacers (n=185) present in contigs assembled from the metagenome (n=22) had matches in contigs from the PH1-SC metavirome. In all cases but one, the targeted genes were either archaeal or from viruses infecting archaea. Therefore, as previously described for other natural environments (Burstein *et al.*, 2016), CRISPR mediated viral immunity in PH1-SC seemed more relevant for Archaea than for Bacteria. Indeed, *Pontibacillus* sp., the most abundant bacterium in this environment, lacks CRISPR systems.

The presence of CRISPR-associated (Cas) proteins (Koonin *et al.*, 2017) was analyzed as described in the Experimental Procedures. Following the same trend as the CRISPR tandem repeats, PH1-SC metagenome had one of the highest rates of Cas proteins found among several saline environments (Table 4). However, in spite of the high numbers of Cas-containing reads, Cas proteins were present in merely 39 contigs (> 5 kb) and only 15 of them had an associated CRIPSR array (with at least 3 tandem repeats), likely due to the difficulty of assembling repeated sequences. About half of these 15 fragments could be assigned to a microbe (mainly *Halobiforma, Haloferax* and *Halorubrum*) and 8 of them had spacers which targeted a contig in the PH1-SC metavirome (over 90% identity in at least 90% of the spacer-length (Supporting Information Table S4).

Finally, since the presence of complete CRISPR-Cas systems has also been described in viruses (Seed *et al.*, 2013), we also looked for these systems in PH1-SC metavirome. When the same size sub-set of reads (10^5, see above) was studied, 47 viral reads contained more than 3 CRISPR-like repeats and 293 Cas-like proteins were also found (Table 3). One possible explanation for this large amount of Cas-like proteins is that many of these proteins were viral endonucleases identified as Cas-like proteins. In this sense, it is noticeable that the largest COG-category represented among the ORFs extracted directly from the metavirome reads were related with "DNA-replication, recombination and repair" (Supporting Information Fig. S11), suggesting a plethora of proteins for DNA manipulation by the viral community.

In conclusion, these findings suggest that CRISPR-Cas systems in the Peñahueca crust microbial community are substantially over-represented compared with other hypersaline environments.

Concluding remarks

The microbial community of Peñahueca salt crust is characterized by a considerable degree of novelty. This is not unexpected given the peculiar mineralogy of the sample that very likely constitutes a harsh environment due to its marginal and superficial position in the lake, with a high degree of desiccation. These conditions and the changing nature of the lake may explain why an endospore former, *Pontibacillus* sp., is the dominant genus in the bacterial assemblage. However, there is no clear explanation for the high presence of CRISPR-*cas* systems (mainly in Archaea), or the high amount of putative temperate phages. Very likely, the study of the dynamics of this ephemeral lake will shed light on these points.

The peculiar mineralogical composition of Peñahueca salt crust makes this environment a good terrestrial analogue for the sulfate-rich mineral regions on Mars. This study provides new insights on the habitability of such analogues and the types of biomarkers that can be targeted in the search for life in extraterrestrial systems. If prokaryotic life exists elsewhere, the role of associated viruses on the configuration of the extraterrestrial microbial community and the fingerprint it can leave have to be considered in future planetary exploration.

EXPERIMENTAL PROCEDURES

Field site and sampling. Sampling was carried out along a precipitation gradient of the superficial sulfate-rich layer present in the Peñahueca lake (40º32'31''N 3º42'45''W // 39.5175849, -3.3390885; Toledo, Spain) in July 2011. Three different sampling areas were distinguished in the dry lake (Fig. 1): (i) Extensive areas made up of a thin layer of brownish salts (PH1) covering the surface from the shore towards the deeper and still wet sites; (ii) white salty and polygonal crusts (PH2) from the limit of the previous one to the wet deeper zones; (iii) and the *deeper parts* (PH3), still with some liquid water(brines). Sampling was done in all three areas at the top superficial layers from 0-1 cm (PH1-SC, PH2-SC) and from 2-5 cm below (dark material, PH1-2cm, PH2-2cm and PH3-core). Samples were collected with sterilized tools into sterile tubes and bottles and stored at 4ºC in a field truck laboratory and brought to the laboratory for further processing.

Mineralogical characterization. Mineralogical characterization of finely ground samples was done by X-ray diffraction (XRD) using a Philips PW-1710XRD system operating at 40 kV and 30 mA, at 2º/ min, with monochromated CuKα radiation. XRD spectra were obtained from 2 to 65º, 2 θ. Mineral phases were identified and quantified with DIFFRAC plus EVA diffraction software.

Crude extract preparation. Extracts were prepared from 40 g of each sample with GuHCl buffer (4 M guanidine-hydrochloride, 0.5 M EDTA, 0.5 M Tris, pH 7.6) following the procedure described by (Tuross and Stathoplos, 1993). The extracts were dialyzed (>1200 Da cut-off) against distilled water and lyophilized. The final extracts were used for total protein quantification using Quant-iT™ Protein Assay kit (Invitrogen, Oregon, USA) according to manufacturer's instructions and read on Qubit™ Fluorometer (Invitrogen, Oregon, USA). Total sugar content in the extracts was determined by using the phenol-sulphuric method (Dubois *et al.*, 1956). Crude liquid extracts from 0.5 g of different samples along a precipitation gradient were analyzed by a fluorescence sandwich microarray immunoassay with the LDChip

Microarray immunoassay with LDChip300. The LDChip300 is an antibody microarray initially designed to detect any trace of microbial molecular biomarkers

either for planetary exploration or environmental monitoring (Parro *et al.*, 2008; Rivas *et al.*, 2008). It contains about 300 antibodies against different molecules including those from cell membranes of archaea and bacteria, extracellular polymers, environmental extracts, proteins, DNA, peptides, exopolysaccharides (EPS) and amino acids. Samples were screened for microbial markers in the field laboratory by fluorescence sandwich microarray immunoassay as previously described (Parro *et al.*, 2011; Blanco *et al.*, 2017). Briefly, the potential biomarkers present in 0.5 g of sample were extracted by ultrasonication into 2 mL saline buffer (0.1 M Tris-HCl pH 8, 0.1 M NaCl, 0.1% Tween 20), filtered through 10 microns, and then 50 µL of the extract incubated with one out of the 9 printed antibody microarrays per microscope slide. After 1 h incubation, 5 min washing, and a second 1 h incubation with fluorescent tracer antibodies, the positive antigen-antibody reactions were read by scanning for fluorescence in a GenePix 4100A microarray scanner with a 635 nm laser. Blank samples containing only buffer were run in parallel. The final fluorescent intensity (F) for each antibody spot was calculated as: $F = (F_{sample} - F_{blank} - 3Fav_{control\ spots})$, being F the fluorescent intensity at 635 nm minus the local background as quantified by the software (GenePix Pro.) and where F_{sample} is the total fluorescence signal of the sample, F_{blank} the total fluorescence signal of the blank, and $Fav_{control\ spots}$ the average fluorescent signal of the control spots.

Cell counts and FISH. In order to avoid the osmotic disruption of the cells, prior to cell recovery, a 20% (w/v) salt solution was prepared by dissolving five grams of PH1S in 20 mL of sterile MQ water and filtering through 0.22 µm pore size filters to obtain a cell free salt solution. Then, 1 g of PH1-SC was completely (approx. 1 hour) dissolved in 4 ml of the salt solution. Large particles were removed by centrifugation at low speed (500 rpm for 4 min). An optimized FISH protocol was performed with the PH1-SC dissolved sample as previously described by (Antón *et al.*, 1999). The sample was fixed by adding formaldehyde at a final concentration of 6.7% and then filtered (0.2 µm GTTP filters, Millipore) to collect the cells. FISH was performed at 46°C for 2 h in the hybridization buffer (0.9 M NaCl, 20 mM Tris-HCl, 0.01% SDS and 35% formamide) with 5 ng/ul of the probes ARC915 and EUB338 to detect Archaea and Bacteria respectively (Amann *et al.*, 1995). Total cells were determined by staining with DAPI (4′,6-diadimino-2-phenylindole) as described before (Snaidr *et al.*, 1997). After

hybridization and DAPI staining, samples were examined with an Axioplan microscope (Leica DMLA) and stained cells were counted in more than 20 different microscopic fields to estimate the cell concentration in the sample.

Virus-like particles (VLPs) counts and transmission electron microscopy (TEM). VLPs in PH1-SC were determined by epifluorescence microscopy after staining with SYBR Gold (Invitrogen) as described before (Chen *et al.*, 2001; Ortmann and Suttle, 2009; Suttle and Fuhrman, 2010). Briefly, 300 g of PH1S sample were slowly (48 h total time) dissolved by adding small amounts of sample to 2 L of 20% SW (Rodriguez-Valera *et al.*, 1985) and filtered through sterile filter papers to remove large particles. 100 µl of this solution was fixed during 30 min at 4ºC with formaldehyde (final concentration 4%), brought to 1 ml with virus-free sterile MQ water and filtered throughin 0.02 µm pore size Anodisc Al_2O_3 filters (Whatman) to retain viruses. Anodisc filters were air dried and stained with SYBR Gold for 15 min at dark, mounted with antifade solution (Citifluor) and VLPs were examined by epifluorescence microscopy (Leica) with 1000X magnification.

The dissolved PH1-SC samples were concentrated as follows: after removing the microbial cells through filtration (0.22 µm filters, Durapore, Millipore), the filtrate was first concentrated using tangential flow filtration through a filter cassette with molecular weight *cut-off* 30 kDa (Vivaflow 200, Sartorius Stedim Biotech) followed by a second concentration with Amicon Ultra-15 centrifugal filter units (Millipore) up to 500 µl of final volume. From the viral concentrate, 5 µl was eluted to 1 ml with sterile 0.02 µm filtered MQ water and the VPLs recovered by ultracentrifugation (186 000 g; 2 h at 20ºC, Beckman Coulter Optima Max-xp with TLA-55 rotor). The virus pellet was suspended in 15 µl sterile 0.02 µm filtered MQ water. 5 µl of the viral suspension was stained for 45 seconds with uranyl acetate (2%) on formvar coated carbon copper grids (Electron Microscopy Sciences). Washing was performed on the grid with sterile 0.02 µm filtered MQ water. VLPs were observed in Jeol JEM-1400plus transmission electron microscope (JEOL Manufacturer, Tokyo, Japan) operating at 120 kV.

Cellular DNA extraction. The microbial biomass corresponding to 45 g of salt dissolved in 300 ml of sterile 20% SW was collected on 0.22 µm Durapore (Millipore)

filters using negative pressure. The microbial DNA was extracted from the filter using the DNeasy Blood & Tissue Kit (QIAGEN) according to the manufacturer's instructions. A total of 18.6 µg of DNA (measured with Nanodrop) was obtained.

Viral DNA extraction*.* Viral DNA was extracted from a virus concentrate containing 4.43 ± 0.11 10^{10} VLPs. The viral DNA was extracted using the RNeasy Mini Kit (QIAGEN) according to the manufacturer's instructions but using RNAse instead of DNAse. A total of 500 ng of DNA (measured with Nanodrop) was obtained. In order to determine genome sizes of viral assemblage in PH1-SC sample, pulsed-field gel electrophoresis (PFGE) was performed as previously described (Santos *et al.*, 2007) using approximately 10^8 VLPs per plug.

16S rRNA PCR amplicons. The V1-V4 region of the 16S-rRNA gene was amplified using primers universal to both Archaea (21F-PS and 907-PS2) and Bacteria (GM3-PS and 907-PS1) (Supporting Information Table S5). Primer sequences were modified by the addition of the 454-adapter sequences. The PCR reaction mix was performed in a final volume of 50 µl which contain 5 µl of 10X Buffer (Takara), 1 µl of 10 µl of forward and reverse primer, 25 µM each of dNTP, 0,1% BSA, 0,25 µl of Taq polymerase Takara Ex Taq (5 units/µl) and 2 µl of template DNA (<500 ng) from environmental samples. The amplification reactions were performed in an Applied Biosystem thermocycler using an initial denaturing cycle at 94°C (5 min) and 30 cycles of: 94°C (1 min), 55°C (30 sec), 72°C (2 min), with and final extension at 72°C (10 min). A secondary PCR was performed to incorporate barcodes and linkers into the previously obtained amplicons. This was done using 1 µl of a 1:10 dilution of the original PCR as a template, under a PCR with the same conditions but only 5 cycles. This PCR was done by triplicate, and the final products for each sample were mixed and purified using MSB® Spin PCRapace (Invitek), following the manufacturer's instructions.

DNA sequencing. Microbial DNA was sequenced with Illumina technology in a Miseq sequence (Genomic Center of Universidad Autónoma de Barcelona). Nextera kit (Illumina) was used to prepare the metagenomic libraries following manufacture´s protocol. Sequencing yielded a total of ≈8 Gb of sequences (more than 25 million reads) of 242 pb median size (see Table 1 for further sequencing details). Viral DNA

was pyrosequenced in a Roche-454-GS-FLX Titanium sequencer (Lifesequencing Company, Valencia, Spain) yielding a total of 99 Mb (186046 reads) of 533 pb median size*.*

Community structure. (i) Community structure from the 16S-rRNA amplicons. Data was processed following the Mothur pipeline (Schloss *et al.*, 2009). Briefly, low-quality sequences were removed (sequences of <500 bp with an average quality score of <25, no ambiguities allowed, and no mismatches in primers and barcodes). The final average length of the amplicons was 857 and 807 pb for Bacteria and Archaea respectively. Chimeras were detected and removed with the application of Chimera Uchime implemented in Mothur. Sequences were clustered into Operational Taxonomic Units (OTUs) at 97% level using UCLUST tool included in QIIME (Caporaso *et al.*, 2010). After singletons and doubletons removal, representative sequences of the bacterial and archaeal OTUs were incorporated separately into the non-redundant SILVA Ref 132 (Quast *et al.*, 2013) database using the ARB package (Ludwig *et al.*, 2004). Alignment was performed with the SINA tool (SILVA Incremental Aligner; (Pruesse *et al.*, 2012)) using the LTP 111 database as template and manually improved following the reference alignment in ARB-editor (Yarza *et al.*, 2010). The closest relative non-type strain sequence of the SILVA REF 132 affiliating with the OTU representatives were selected and merged with the LTP 111 type-strain sequence database. For the final tree reconstruction, the selected representative sequences with an additional set of about 750 supporting sequences (highest quality in the LTP) and covering a balanced representation of all major phyla of both Bacteria and Archaea domains were used for a neighbor joining reconstruction (Munoz *et al.*, 2016). To this final topology, all OTU representatives were inserted using the parsimony tool, and clustered in OPUs (Operational Phylogenetic Units; (Franca *et al.*, 2014; Mora-Ruiz Mdel *et al.*, 2015) based on the visual inspection of the final tree. The PAST software v1.82b (Hammer *et al.*, 2001) was used to compute the diversity indexes in the archaeal and bacterial sequence datasets. The rarefaction curves were calculated (Hurlbert, 1971) and plotted as the number of Operational Taxonomic Units (OTUs) versus the number of reads. (ii) Community structure from the metagenomic reads. Individual reads of both datasets (metagenome and metavirome) were compared against the nr-NCBI collection using Diamond (Buchfink *et al.*, 2015). Based on

sequence similarity against the NCBI-nr database, the best hit for each gene was determined and used to bin the top-level taxa. The data was analyzed using MEGAN6 (Huson *et al.*, 2007).

Assembly, gene prediction and annotation of sequence data. The sequenced reads from the bacterial collection were quality trimmed and assembled using IDBA-UD (Peng *et al.*, 2012) with the following parameters: -maxk 124 –mink 70 –step 20 – pre_correction. Otherwise indicated, contigs larger than 5 Kb were considered for the metagenome. For the metavirome, we used CLC Genomics workbench (Quiagen) assembler and only the contigs above 500 pb were considered. Gene predictions on the individual reads and the assembled contigs were carried out using Prodigal (Hyatt *et al.*, 2010). Only proteins bigger than 50 amino acids were taken into account. tRNAs were screened using tRNA-SCAN (Lowe and Eddy, 1997). rRNAs were screened using meta_rna and RNAscan (Huang *et al.*, 2009). Functional annotation of the putative ORFs was performed comparing the predicted protein sequences against the NCBI-nr database by BLAST (Altschul *et al.*, 1990), Pfam (Bateman *et al.*, 2004) (cut-off E-value 10^{-5}), COGs (Galperin *et al.*, 2015) and POGs (Makarova *et al.*, 2015). Domain predictions in the sequences were performed using the HMMER package (Eddy, 2008) (against the Pfam database) and the HHpred server (Soding *et al.*, 2005). Housekeeping genes present in the metagenome and metavirome were searched for using the ORFs (>50 amino acids) coded in the reads from the complete collections or subsets of 10^6 reads. Those contigs with more than half of their length with their best hit to a single genus were assigned taxonomically. The isoelectric point (pI) of the coding sequences (>50 amino acids) were calculated using the program IEP from the EMBOSS package (Rice *et al.*, 2000). For Fig. S3, subsets of 10^5 proteins from each dataset were separated and their pI calculated (Punta Cormorant Lagoon (PC6; (Rusch *et al.*, 2007), Santa Pola saltern collections (SP-13, SP-19 and SP-37; (Ghai *et al.*, 2011), Isla Cristina saltern (IC21; (Fernandez *et al.*, 2013), Salar Grande (Crits-Christoph *et al.*, 2016) and the Kulunda Soda Lakes samples (Tc, T5, PL and B1; (Vavourakis *et al.*, 2016). Metagenome and metavirome collections were compared among them using Meta-Fast (Ulyantsev *et al.*, 2016). For the detection of the reads containing CRISPR-like repetitions, the CRT v.1.2. program (Bland *et al.*, 2007) was used with the following parameters: -minNR 3, -minRL 15, -maxRL 50, -minSL 15, -

maxSL 60, and -searchWL 8. CRISPR-related proteins were detected using HHMER (Eddy, 2008) against the COGs database. For the BLASTN comparisons for the repeats and spacers from the CRISPR systems, the word_size parameter was adjusted to 5.

Reconstruction and analysis of draft genomes. Phylogenomic trees. Further binning of the contigs was performed by differential coverage, tetranucleotide frequencies and GC-content. Sequence bins were checked with the MaxBin2 program (Wu *et al.*, 2016). Possible bin contamination and strain heterogeneity was estimated with CheckM v0.9.7 (Parks *et al.*, 2014). Taxonomic affiliation of the selected bins was determined via phylogenetic trees based on concatenates of all proteins identified within Clusters of Orthologous Groups (COG; (Tatusov *et al.*, 2001; Galperin *et al.*, 2015) and shared with complete, or near-complete, reference genomes. The concatenated proteins were aligned using Kalign (Lassmann and Sonnhammer, 2005) and maximum-likelihood trees were performed using MEGA v.7 (Kumar *et al.*, 2016) and FastTree (Price *et al.*, 2010) with local support values using Shimodaira-Hasegawa test. Completeness of the reconstructed bacterial and archaeal genomes was estimated by the presence of essential/core genes using HMMER (35, 112 and 53 genes respectively; (Raes *et al.*, 2007; Narasingarao *et al.*, 2012; Albertsen *et al.*, 2013). HMMER was also used for the detection of the Cas proteins, COG1203, 1343, 1468, 1518, 1583, 1688 and 3649 present in the COG-database. My_taxa_scan, from the MIGA tool was used also to infer also the taxonomy and calculate de AAI (Rodriguez *et al.*, 2018). An E_value $<10^{-5}$ and an alignment coverage >65% were used as cut-off to define homologs of the essential/core genes. Average nucleotide identity (ANI) and conserved DNA fraction between reconstructed and /or reference genomes were calculated based on the whole genome sequence as described by (Goris *et al.*, 2007). GC-content was calculated using the "*geecee*" tool from the emboss package (Rice *et al.*, 2000). Cd-hit program (Huang *et al.*, 2010) was used to obtain the non-redundant proteins databases, (cut-off: 90% identity, 70% coverage) and to eliminate the redundancy from the CRISPR tandems repeats and proto-spacers (cut-off: 97% identity, 100% coverage).

Recruitment plots. Fragment recruitments of the PH1-SC metagenome and metavirome were performed against the recovered bins using BLASTN. Only those reads with more than 95% of identity in a minimum alignment of 50 bp were taken into account. The number of hits was normalized against the size of the genomes and the database and expressed as "reads per kilobase of sequenced reads per gigabase of mapped reads" (rpkg).

Comparison with other environmental datasets. Peñahueca datasets were compared with other metagenomes using BLASTN and MetaFast (Ulyantsev *et al.*, 2016). The following halophilic collections were used in this analysis: three microbial metagenomes from Santa Pola saltern (Spain) (SP13, SP19 and SS37, (Ghai *et al.*, 2011), one from Isla Cristina saltern (Spain) (IC-21, (Fernandez *et al.*, 2013)); the metagenome of the endolithic community from Salar-Grande (Chile) (Crits-Christoph *et al.*, 2016), the hypersaline Lake Meyghan (Iran) (Naghoni *et al.*, 2017) - and six metagenomes from alkaline lakes (Russia) (Vavourakis *et al.*, 2016). The low-salinity (6%) dataset collection from Punta Cormorant from the GOS datasets (PC6, (Rusch *et al.*, 2007) was also included in the analysis. A metagenome from the fresh waterlake of Baikal (5 m) (Cabello-Yeves *et al.*, 2017) was used as an outgroup. The metaviromes used in this analysis were those from Tyrrell Lake (Victoria Australia), (Emerson *et al.*, 2013)), three datasets from the saltern of San Diego, California (USA); (Rodriguez-Brito *et al.*, 2010), one dataset from Santa Pola (Alicante, Spain (Santos *et al.*, 2010), another from the Sfax saltern (Tunisia) (Boujelben *et al.*, 2012) and the metavirome from Lake Retba (Senegal, (Sime-Ngando *et al.*, 2010)). A hit was counted using the criteria of minimum 95% identity and alignment length of 50 bp. To compare the results among different data sets, the number of reads was normalized (to the collection size and the sequence length) and expressed as rpkg.

Virome Net. The ORFs from the metavirome contigs (>5 kb) were compared among themselves using a reciprocal best-hit approximation (BLASTP). ORFs from known halophages (Tang *et al.*, 2002; Porter *et al.*, 2005; Pagaling *et al.*, 2007; Santos *et al.*, 2007; Pietila *et al.*, 2010; Pina *et al.*, 2011; Garcia-Heredia *et al.*, 2012; Crits-Christoph *et al.*, 2016) were used as references. Homology relationship was plotted using Cytoscape (Shannon *et al.*, 2003).

Nucleotide sequence accession number. The raw sequences of the two metagenomics datasets have been deposited with the NCBI under the Bioproject accession number: PRJNA421101 (Biosample for the metagenome: SUB3297431 and Biosample for the metavirome: SUB3297453). The assembled contigs have been deposited as a Whole Genome Shotgun project, Metagenome MG_PH1-SC assembled contigs: SDGW00000000; Metavirome MV_PH1-SC assembled contigs: SDGX00000000. 16S rRNA sequences have been deposited in the SRA file SRR8881907 and SRR8882225.

Acknowledgments

Conflict of Interest

The authors declare no conflict of interest.

TABLES

Table 1. General features of the Peñahueca lake salt-crust (PH1-SC) sample. Summary of pyrosequenced 16S rRNA amplicons and statistical indexes. Genomic features of the metagenome and metavirome obtained, raw sequence reads and number of assembled contigs.

Table 2. General features of the reconstructed genomes from the Peñahueca saline crust.

Table 3. Number of housekeeping genes and CRISPR related proteins found in PH1-SC metavirome.

Table 4. Number of housekeeping genes and CRISPR related proteins found in metagenomes from several hypersaline environments.

FIGURE LEGENDS

Fig. 1. The Peñahueca ephemeral lake and sampling. Maps showing the Peñahueca lagoon in central Spain (40º32'31''N 3º42'45''). The squared area is amplified at the bottom to show the sampling sites.

Fig. 2. A: Transmission electron micrographs showing the viral morphotypes found in Peñahueca salt crust. Examples of micrographs used for counting (scale bars are indicated in images). Different viral morphotypes are indicated by arrows: (a) fusiform (spindle or lemon-shaped), (b) caudovirus medium-length tail, (c) polyhedral, (d) short-tailed caudovirus, (e) long-tailed caudovirus, (f-i) other morphologies. Examples of putative extracellular vesicles are marked with asterisks. **B:** Distribution (in percentages) of the morphotypes shown in panel A.

Fig. 3. Taxonomic profiles of the microbiota from Peñahueca salt crust. Taxonomic distribution of the metagenomics reads assigned to top level taxa obtained with MEGAN.

Fig. 4. GC distribution of the Peñahueca salt crust metagenome. Relative frequency of the GC-content of the assembled contigs from the PH1-SC metagenome is indicated in red. The ORFs of the contigs were compared against the nr-NCBI (grey dots) and the media of the similarity for each of the DNA fragments was plotted on the right *y-axe*. Standard deviation for each media is shown.

Fig. 5. Metagenome assembled genomes. Maximum-likelihood phylogenetic tree based on 163 proteins shared between the draft haloarchaeal genomes reconstructed from the PH1-SC metagenome and reference genomes. Bars indicate the amount of sequence difference.

Fig. 6. Genomes of *Pontibacillus* phages. Genome comparison of the *Pontibacillus* sp. putative phages assembled from the Peñahueca metagenome containing a recognizable lysis cassette.

Fig. 7. CRISPR-Cas systems. Number of CRISPR-like repetitions and CRISPR related proteins found in PH1-SC metagenome compared with other halophilic environments (subsets of 10^5 reads were used).

SUPPORTING INFORMATION

TABLES

Table S1. Mineral composition of the Peñahueca samples.

Table S2. List of contigs of each of the draft genomes identified in the Peñahueca crust metagenome (MG_PH1-SC).

Table S3. List of viral contigs identified in the PH1-SC metagenome (only those with a ratio rpkg-metavirome (MV)/rpkg-metagenome(MG) >3 are listed).

Table S4. Description of metagenomic contigs with CRISPR-Cas systems.

Table S5. Primers used in this study. Primers used in the second PCR include linkers and barcodes into the amplicons. In italics, the sequence of 454-FLX specific linkers. In lowercase, the sequences of the different barcodes used in this study. Universal 907 primer sequence is from (Muyzer and Smalla, 1998).

FIGURES

Fig. S1. *In situ* analysis with LDChip300 detected microbial cell and exopolymeric material. I, Cellular and extracellular polymeric substances (EPS) from microbes whose natural habitat is enriched in sulfate salts and sediments; II, Cellular extracts and EPS from sulfur oxidizing Gammaproteobacteria; III, different protein and peptides, two of them related to bacterial ferritins; 1 and 3, steroid compounds similar to cortisol (1, polyclonal antibody and 3, monoclonal antibody); 2, oligomers of *N*-acetylglucosamine (NAG) and some bacterial cell wall oligosaccharides containing *N*-acetylglucosamine

and *N*-acetylmuramic acid (NAG-NAM); 4, Humic substances; 5,12, Nucleotide derivative; 6, *Colwellia sp.* (Gammaproteobacteria)*; 7, 8, *Salinibacter ruber* MB and *Salinibacter ruber* PR1; 9, *Streptomyces* spp.; 10, *Methanobacterium formicicum*; 11, Bacterial LPS (Lipopolysaccharide); 13, 14, 15, aromatic compounds (benzoic acid, steroid-like, tetracycline).

Fig. S2. DAPI (right) and FISH (left) staining cells using the ARC915 (up) and EUB338 (bottom) probes. Scale bar: 3 µm.

Fig. S3. MetaFast graphical output, a dissimilatory heatmap with dendrogram for several haline metagenomes. The hierarchical clustering is performed using average linkage and Bray-Curtis metric. Similarity was measured by Spearman correlation. Nomenclature of the samples is as follows. Baikal lake (5m)-(1) (Cabello-Yeves *et al.*, 2017); Santa Pola saltern SP13-(3), SP33-(4), SP19-(5) and SP37-(6) (Ghai *et al.*, 2011); Isla Cristina saltern-(7) (Fernandez *et al.*, 2013); Red-(8), White-(9) and Green-(2) samples from Meyghan Lagoon (Naghoni *et al.*, 2017); Peñahueca lake PH1-SC-(10) (this work); Salar Grande-(11) (Crits-Christoph *et al.*, 2016); Picturesque Lake-(12,18), Lake Bitter-(13,14) and Lake Tanatar-(15,16) (Vavourakis *et al.*, 2016); Punta Cormoran (Rusch *et al.*, 2007).

Fig. S4. GC-content of reads from the Peñahueca crust datasets (MG_PH1-SC, metagenome and MV_PH1-SC, metavirome) compared with other saline habitats. GC-content was calculated in intervals of bin width 1 and compared to the GC-distribution of other metagenomes from saline environments; the crystallizer pond SP-37 (salinity, S=37%), intermediate saline pond SP-19 (S=19%) and low saline pond SP13 (S=13%) of the solar saltern in Santa Pola, Spain (Ghai *et al.*, 2011); the intermediate saline pond IC21 (S=21%) from the solar salter Isla Cristina, Spain (Fernandez *et al.*, 2013); the metagenome from Salar Grande, Chile (Crits-Christoph *et al.*, 2016) and the Kulunda soda brines samples (Tanatar-5 (T5, S=17%), Picturesque Lake (PL, S=25%), Tanatar trona crystallizer (Tc, S=30%) and Bitter-1 (B1, S=40%), (Vavourakis *et al.*, 2016). Subsets of 10^5 reads were used to calculate the GC-content of the Salar Grande and the soda brines samples. The large peak in the GC-profile of SP-37 was attributed

to the predominance of the square archaeon *Haloquadratum walsbyi* with a low average GC-content (47.9%) compared to other members of the *Halobacteria*.

Fig.S5. COG (A) and KEGG (B) classification of the ORFs coded by PH1-SC metagenome reads.

Fig.S6. Predicted isoelectric points (pI) for the coding sequences of the sequence reads from the Peñahueca crust collections (MG-PH1-SC, metagenome and MV-PH1-SC, metavirome) compared with the crystallizer pond SP-37 (salinity=37%) of the solar saltern in Santa Pola, Spain (Ghai *et al.*, 2011) and the sea-water collection from the Mediterranean DCM-2007 (Ghai *et al.*, 2010).

Fig.S7. A: Matrix of the ANI values of the reconstructed Nanohaloarchaeon PH1-SC from Peñahueca metagenome compared with other previously described reference genomes. **B:** Maximum-likelihood phylogenetic tree based on 7 shared concatenated COGs between the reconstructed draft genome of the Nanohaloarchaeon PH1-SC and other close reference genomes. Bars indicate the amount of sequence difference.

Fig. S8. A: Coverage and similarity between the two subpopulations (A and B) of *Pontibacillus* reconstructed from the PH1-SC metagenome. **B:** Maximum-likelihood phylogenetic tree based on 213 shared proteins between the draft genome of the two subpopulations A and B-*Pontibacillus* reconstructed from the PH1-SC metagenome and relative *Bacillus* reference genomes. Bars indicate the amount of sequence difference. Heat-map on the right indicates the predicted isoelectric points (pI) in a window from 4 to 13 demonstrating the effect of the 'salt-in' strategy for osmotic adaptation. The pI profiles of *Escherichia coli* K12 (bacterial 'salt-out'), *Spiribacter salinus* M19-40 (bacterial 'salt-in'/mixed strategy) and *Salinibacter ruber* M31 (bacterial 'salt-in') were also plotted for comparison.

Fig. S9. A: Photomicrograph of SYBR-Gold stained VLPs along the virus purification protocol. **B:** Pulsed-field gel electrophoresis (PFGE) of low range PFG marker from (lane 1) NEB (a mixture of lambda ladder and lambda digested with *Hind* III) and (lane 2) the viral DNA obtained from the PH1-SC sample (step 4 on the left). DNA sizes are given in kb.

Fig. S10. Number of contigs in the metavirome (A) and the metagenome (B) (*y*-axis) *versus* the ratio of the rpkgs (*x*-axis) of the contigs assembled from the metavirome and the metagenome.

Fig. S11. A: COG functional classification of the PH1-SC metavirome reads. A total of 12% of the proteins showed significant homology to the COGs database (22934 ORFs). **B:** COG functional classification of the ORFs from PH1-SC metagenome contigs with more than > 3-fold coverage in the metavirome than in the metagenome.

Fig. S12. Coverage (in rpkgs) of the putative viral assembled contigs >5 kb from the metagenome of Peñahueca (stars) and the *Bacilli*-like phage contigs (bars). The coverage in the metavirome is indicated in red and that in the metagenome, in black.

REFERENCES

Albertsen, M., P. Hugenholtz, A. Skarshewski, K. L. Nielsen, G. W. Tyson and P. H. Nielsen (2013) Genome sequences of rare, uncultured bacteria obtained by differential coverage binning of multiple metagenomes. *Nat Biotechnol* **31**: 533-8.

Altschul, S. F., W. Gish, W. Miller, E. W. Myers and D. J. Lipman (1990) Basic local alignment search tool. *J Mol Biol* **215**: 403-10.

Amann, R. I., W. Ludwig and K. H. Schleifer (1995) Phylogenetic identification and in situ detection of individual microbial cells without cultivation. *Microbiol Rev* **59**: 143-69.

Andrei, A. S., H. L. Banciu and A. Oren (2012) Living with salt: metabolic and phylogenetic diversity of archaea inhabiting saline ecosystems. *FEMS Microbiol Lett* **330**: 1-9.

Andrei, A. S., M. S. Robeson, 2nd, A. Baricz, C. Coman, V. Muntean, A. Ionescu*, et al.* (2015) Contrasting taxonomic stratification of microbial communities in two hypersaline meromictic lakes. *ISME J* **9**: 2642-56.

Antón, J., E. Llobet-Brossa, F. Rodriguez-Valera and R. Amann (1999) Fluorescence in situ hybridization analysis of the prokaryotic community inhabiting crystallizer ponds. *Environ Microbiol* **1**: 517-23.

Antón, J., R. Rossello-Mora, F. Rodriguez-Valera and R. Amann (2000) Extremely halophilic bacteria in crystallizer ponds from solar salterns. *Appl Environ Microbiol* **66**: 3052-7.

Bateman, A., L. Coin, R. Durbin, R. D. Finn, V. Hollich, S. Griffiths-Jones*, et al.* (2004) The Pfam protein families database. *Nucleic Acids Res* **32**: D138-41.

Baxter, B. K., Mangalea, M.R., Willcox, S., Sabet, S., Nagoulat, M.-N., and Griffith, J.D. (2011) Haloviruses of Great Salt Lake: a model for understanding viral diversity. *In A. Ventosa, A. Oren, and Y. Ma (ed.), Halophiles and hypersalyne environments*: **173**:190.

Belitsky, B. R., J. Brill, E. Bremer and A. L. Sonenshein (2001) Multiple genes for the last step of proline biosynthesis in Bacillus subtilis. *J Bacteriol* **183**: 4389-92.

Blanco, Y., M. Moreno-Paz and V. Parro (2017) Experimental protocol for detecting cyanobacteria in liquid and solid samples with an antibody microarray chip. *J Vis Exp.* **120**: e54994

Bland, C., T. L. Ramsey, F. Sabree, M. Lowe, K. Brown, N. C. Kyrpides and P. Hugenholtz (2007) CRISPR recognition tool (CRT): a tool for automatic detection of clustered regularly interspaced palindromic repeats. *BMC Bioinformatics* **8**: 209.

Boujelben, I., M. Gomariz, M. Martinez-Garcia, F. Santos, A. Pena, C. Lopez, *et al.* (2012) Spatial and seasonal prokaryotic community dynamics in ponds of increasing salinity of Sfax solar saltern in Tunisia. *Antonie Van Leeuwenhoek.* **101**:845-57

Boujelben, I., P. Yarza, C. Almansa, J. Villamor, S. Maalej, J. Anton and F. Santos (2012) Virioplankton community structure in Tunisian solar salterns. *Appl Environ Microbiol* **78**: 7429-37.

Brum, J. R., B. L. Hurwitz, O. Schofield, H. W. Ducklow and M. B. Sullivan (2016) Seasonal time bombs: dominant temperate viruses affect Southern Ocean microbial dynamics. *ISME J* **10**: 437-49.

Buchfink, B., C. Xie and D. H. Huson (2015) Fast and sensitive protein alignment using DIAMOND. *Nat Methods* **12**: 59-60.

Burns, D. G., H. M. Camakaris, P. H. Janssen and M. L. Dyall-Smith (2004) Combined use of cultivation-dependent and cultivation-independent methods indicates that members of most haloarchaeal groups in an Australian crystallizer pond are cultivable. *Appl Environ Microbiol* **70**: 5258-65.

Burstein, D., L. B. Harrington, S. C. Strutt, A. J. Probst, K. Anantharaman, B. C. Thomas, *et al.* (2016) New CRISPR-Cas systems from uncultivated microbes. *Nature* **542**: 237-241.

Cabello-Yeves, P. J., T. I. Zemskaya, R. Rosselli, F. H. Coutinho, A. S. Zakharenko, V. V. Blinov and F. Rodriguez-Valera (2017) Genomes of novel microbial lineages assembled from the sub-ice waters of Lake Baikal. *Appl Environ Microbiol* **8: 84**: e02132-17.

Caporaso, J. G., J. Kuczynski, J. Stombaugh, K. Bittinger, F. D. Bushman, E. K. Costello, *et al.* (2010) QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* **7**: 335-6.

Crisler, J. D., T. M. Newville, F. Chen, B. C. Clark and M. A. Schneegurt (2012) Bacterial growth at the high concentrations of magnesium sulfate found in martian soils. *Astrobiology* **12**: 98-106.

Crits-Christoph, A., D. R. Gelsinger, B. Ma, J. Wierzchos, J. Ravel, A. Davila, *et al.* (2016) Functional interactions of archaea, bacteria and viruses in a hypersaline endolithic community. *Environ Microbiol* **18**: 2064-77.

Chen, F., J. R. Lu, B. J. Binder, Y. C. Liu and R. E. Hodson (2001) Application of digital image analysis and flow cytometry to enumerate marine viruses stained with SYBR gold. *Appl Environ Microbiol* **67**: 539-45.

Chen, Y. G., Y. Q. Zhang, H. D. Xiao, Z. X. Liu, L. B. Yi, J. X. Shi, *et al.* (2009) *Pontibacillus halophilus* sp. nov., a moderately halophilic bacterium isolated from a sea urchin. *Int J Syst Evol Microbiol* **59**: 1635-9.

Chen, Y. G., Y. Q. Zhang, L. B. Yi, Z. Y. Li, Y. X. Wang, H. D. Xiao, *et al.* (2010) *Pontibacillus litoralis* sp. nov., a facultatively anaerobic bacterium isolated from a sea anemone, and emended description of the genus *Pontibacillus. Int J Syst Evol Microbiol* **60**: 560-5.

Chou, I. M. and R. R. Seal Li (2007) Magnesium and calcium sulfate stabilities and the water budget of Mars. *J Geophys Res E: Planets* **112**.

Deole, R., J. Challacombe, D. W. Raiford and W. D. Hoff (2013) An extremely halophilic proteobacterium combines a highly acidic proteome with a low cytoplasmic potassium content. *J Biol Chem* **288**: 581-8.

Di Meglio, L., F. Santos, M. Gomariz, C. Almansa, C. Lopez, J. Anton and D. Nercessian (2016) Seasonal dynamics of extremely halophilic microbial communities in three Argentinian salterns. *FEMS Microbiol Ecol* **92**.

Dubois, M., K. A. Gilles, J. K. Hamilton, P. A. Rebers and F. Smith (1956) Colorimetric method for determination of sugars and related substances. *Anal Chemy* **28**: 350-356.

Eddy, S. R. (2008) A probabilistic model of local sequence alignment that simplifies statistical significance estimation. *PLoS Comput Biol* **4**: e1000069.

Emerson, J. B., K. Andrade, B. C. Thomas, A. Norman, E. E. Allen, K. B. Heidelberg and J. F. Banfield (2013) Virus-host and CRISPR dynamics in Archaea-dominated hypersaline Lake Tyrrell, Victoria, Australia. *Archaea* **2013**: 370871.

Emerson, J. B., B. C. Thomas, K. Andrade, E. E. Allen, K. B. Heidelberg and J. F. Banfield (2012) Dynamic viral populations in hypersaline systems as revealed by metagenomic assembly. *Appl Environ Microbiol* **78**: 6309-20.

Emerson, J. B., B. C. Thomas, K. Andrade, K. B. Heidelberg and J. F. Banfield (2013) New approaches indicate constant viral diversity despite shifts in assemblage structure in an Australian hypersaline lake. *Appl Environ Microbiol* **79**: 6755-64.

Farlow, J., D. Bolkvadze, L. Leshkasheli, I. Kusradze, A. Kotorashvili, N. Kotaria*, et al.* (2018) Genomic characterization of three novel Basilisk-like phages infecting *Bacillus anthracis. BMC Genomics* **19**: 685.

Fernandez, A. B., M. J. Leon, B. Vera, C. Sanchez-Porro and A. Ventosa (2013) Metagenomic sequence of prokaryotic microbiota from an intermediate-salinity pond of a saltern in Isla Cristina, Spain. *Genome Announc* **2**.

Franca, L., A. Lopez-Lopez, R. Rossello-Mora and M. S. da Costa (2014) Microbial diversity and dynamics of a groundwater and a still bottled natural mineral water. *Environ Microbiol* **17**: 577-93.

Gaillard, F., J. Michalski, G. Berger, S. M. McLennan and B. Scaillet (2013) Geochemical reservoirs and timing of sulfur cycling on Mars. *Space Sci Rev* **174**: 251–300.

Galperin, M. Y., K. S. Makarova, Y. I. Wolf and E. V. Koonin (2015) Expanded microbial genome coverage and improved protein family annotation in the COG database. *Nucleic Acids Res* **43**: D261-9.

Garcia-Heredia, I., A. B. Martin-Cuadrado, F. J. Mojica, F. Santos, A. Mira, J. Antón and F. Rodriguez-Valera (2012) Reconstructing viral genomes from the environment using fosmid clones: the case of haloviruses. *PLoS One* **7**: e33802.

Garcia, P., C. Monjardin, R. Martin, C. Madera, N. Soberon, E. Garcia*, et al.* (2008) Isolation of new *Stenotrophomonas* bacteriophages and genomic characterization of temperate phage S1. *Appl Environ Microbiol* **74**: 7552-60.

Gendrin, A., N. Mangold, J. P. Bibring, Y. Langevin, B. Gondet, F. Poulet*, et al.* (2005) Sulfates in Martian layered terrains: the OMEGA/Mars Express view. *Science* **307**: 1587-91.

Ghai, R., A. Martin-Cuadrado, A. Gonzaga, I. Garcia-Heredia, R. Cabrera, J. Martin*, et al.* (2010) Metagenome of the Mediterranenan deep Chlorophyl Maximum studied by direct and fosmid library 454 pyrosequencing. *The ISME Journal.* **4**:1154-66

Ghai, R., L. Pašić, A. B. Fernández, A.-B. Martin-Cuadrado, C. M. Mizuno, K. D. McMahon*, et al.* (2011) New abundant microbial groups in aquatic hypersaline environments. *Sci. Rep.* **1**: 135.

Gill, S., R. Catchpole and P. Forterre (2018) Extracellular membrane vesicles (EVs) in the three domains of life and beyond. *FEMS Microbiol Rev.* fuy042.

Godde, J. S. and A. Bickerton (2006) The repetitive DNA elements called CRISPRs and their associated genes: evidence of horizontal transfer among prokaryotes. *J Mol Evol* **62**: 718-29.

Gomariz, M., M. Martinez-Garcia, F. Santos, F. Rodriguez, S. Capella-Gutierrez, T. Gabaldon*, et al.* (2015) From community approaches to single-cell genomics: the discovery of ubiquitous hyperhalophilic Bacteroidetes generalists. *ISME J* **9**: 16-31.

Goris, J., K. T. Konstantinidis, J. A. Klappenbach, T. Coenye, P. Vandamme and J. M. Tiedje (2007) DNA-DNA hybridization values and their relationship to whole-genome sequence similarities. *Int J Syst Evol Microbiol* **57**: 81-91.

Guixa-Boixereu, N. (1996) Viral lysis and bacterivory as prokaryotic loss factors along a salinity gradient. *Aquat Microb Ecol* **11**: 213-227.

Haferburg, G., J. A. D. Groning, N. Schmidt, N. A. Kummer, J. C. Erquicia and M. Schlomann (2017) Microbial diversity of the hypersaline and lithium-rich Salar de Uyuni, Bolivia. *Microbiol Res* **199**: 19-28.

Hammer, O., D. Harper and P. Ryan (2001) PAST: Paleontological Statistics software package for education and data analysis. *Paleontologia Electronica* **9**: 9.

Han R, Zhang X, Liu J, Long Q, Chen L, Liu D and Z. D (2017) Microbial community structure and diversity within hypersaline Keke Salt Lake environments. *Can J Microbiol* **63**:895-908.

Hand, K. P., A. E. Murray, J. B. Garvin, W. B. Brinckerhoff, B. C. Christner, K. S. Edgett*, et al.* (2017) Report of the Europa Lander Sciwence Definition Team.

Hawley, E. R., W. Schackwitz and M. Hess (2014) Metagenomic sequencing of two Salton Sea microbiomes. *Genome Announc* **2**: e01208-13.

Huang, J., Z. X. Qiao, J. W. Tang and G. Wang (2015) High quality draft genome sequence of the moderately halophilic bacterium *Pontibacillus yanchengensis* Y32(T) and comparison among Pontibacillus *genomes. Stand Genomic Sci* **10: 93.**

Huang, Y., P. Gilna and W. Li (2009) Identification of ribosomal RNA genes in metagenomic fragments. *Bioinformatics* **25**: 1338-40.

Huang, Y., B. Niu, Y. Gao, L. Fu and W. Li (2010) CD-HIT Suite: a web server for clustering and comparing biological sequences. *Bioinformatics* **26**: 680-2.

Hurlbert, S. (1971) The non-concept of species diversity: a critique and alternative parameters. *Ecology* **52**: 577-586.

Huson, D. H., A. F. Auch, J. Qi and S. C. Schuster (2007) MEGAN analysis of metagenomic data. *Genome Res* **17**:377-86.

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

Jackson, S. A., R. E. McKenzie, R. D. Fagerlund, S. N. Kieper, P. C. Fineran and S. J. Brouns (2017) CRISPR-Cas: Adapting to change. *Science* **7**:356(6333).

Kertesz, M. A. and C. Wietek (2001) Desulfurization and desulfonation: applications of sulfur-controlled gene expression in bacteria. *Appl Microbiol Biotechnol* **57**: 460-6.

Konstantinidis, K. T., R. Rosselló-Móra and R. Amann (2017) Uncultivated microbes in need of their own taxonomy. *ISME J* **11**: 2399-2406.

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

Kristensen, D. M., A. S. Waller, T. Yamada, P. Bork, A. R. Mushegian and E. V. Koonin (2013) Orthologous gene clusters and taxon signature genes for viruses of prokaryotes. *J Bacteriol* **195**: 941-50.

Kumar, S., G. Stecher and K. Tamura (2016) MEGA7: Molecular evolutionary genetics analysis Version 7.0 for bigger datasets. *Mol Biol Evol* **33**: 1870-4.

Kunst, F., N. Ogasawara, I. Moszer, A. M. Albertini, G. Alloni, V. Azevedo*, et al.* (1997) The complete genome sequence of the gram-positive bacterium *Bacillus subtilis. Nature* **390**: 249-56.

La Cono, V., G. Bortoluzzi, E. Messina, G. La Spada, F. Smedile, L. Giuliano*, et al.* (2019) The discovery of Lake Hephaestus, the youngest athalassohaline deep-sea formation on Earth. *Sci Rep* **9**: 1679.

La Cono, V., F. Smedile, G. Bortoluzzi, E. Arcadi, G. Maimone, E. Messina*, et al.* (2011) Unveiling microbial life in new deep-sea hypersaline Lake Thetis. Part I: Prokaryotes and environmental settings. *Environ Microbiol* **13**: 2250-68.

Lassmann, T. and E. L. Sonnhammer (2005) Kalign--an accurate and fast multiple sequence alignment algorithm. *BMC Bioinformatics* **6**: 298.

Lee, J. C., Y. S. Kim, B. S. Yun and K. S. Whang (2015) *Pontibacillus salicampi* sp. nov., a moderately halophilic bacterium isolated from saltern soil. *Int J Syst Evol Microbiol* **65**: 375-80.

Lim, J. M., C. O. Jeon, D. J. Park, H. R. Kim, B. J. Yoon and C. J. Kim (2005) *Pontibacillus marinus* sp. nov., a moderately halophilic bacterium from a solar saltern, and emended description of the genus *Pontibacillus. Int J Syst Evol Microbiol* **55**: 1027-31.

Lopez-Perez, M., R. Ghai, M. J. Leon, A. Rodriguez-Olmos, J. L. Copa-Patino, J. Soliveri, *et al.* (2013) Genomes of "Spiribacter", a streamlined, successful halophilic bacterium. *BMC Genomics* **14**: 787.

Lowe, T. M. and S. R. Eddy (1997) tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res* **25**: 955-64.

Ludwig, W., O. Strunk, R. Westram, L. Richter, H. Meier, Yadhukumar, *et al.* (2004) ARB: a software environment for sequence data. *Nucleic Acids Res* **32**: 1363-71.

Macau, F. and O. Riba (1965) Situación, características y extensión de los terrenos yesíferos en España I. Col. internac. Las obras públicas en los terrenos yesíferos. **0-1**: 1-28.

Makarova, K. S., Y. I. Wolf and E. V. Koonin (2015) Archaeal Clusters of Orthologous Genes (arCOGs): An update and application for analysis of shared features between Thermococcales, Methanococcales, and Methanobacteriales. *Life (Basel)* **5**: 818-40.

McCord, T. B., G. B. Hansen, F. P. Fanale, R. W. Carlson, D. L. Matson, T. V. Johnson, *et al.* (1998) Salts on Europa's surface detected by Galileo's near infrared mapping spectrometer. The NIMS Team. *Science* **280**: 1242-5.

Milliken, R. E., R. C. Ewing, W. W. Fischer and J. Hurowitz (2014) Wind-blown sandstones cemented by sulfate and clay minerals in Gale Crater, Mars. *Geophys. Res. Lett.* **41**: 1149–1154.

Montoya, L., C. Vizioli, N. Rodriguez, M. J. Rastoll, R. Amils and I. Marin (2013) Microbial community composition of Tirez lagoon (Spain), a highly sulfated athalassohaline environment. *Aquat Biosyst* **9**: 19.

Mora-Ruiz Mdel, R., F. Font-Verdera, C. Diaz-Gil, M. Urdiain, G. Rodriguez-Valdecantos, B. Gonzalez, *et al.* (2015) Moderate halophilic bacteria colonizing the phylloplane of halophytes of the subfamily Salicornioideae (Amaranthaceae). *Syst Appl Microbiol* **38**: 406-16.

Mora-Ruiz, M. D. R., A. Cifuentes, F. Font-Verdera, C. Perez-Fernandez, M. E. Farias, B. Gonzalez, *et al.* (2018) Biogeographical patterns of bacterial and archaeal communities from distant hypersaline environments. *Syst Appl Microbiol* **41**: 139-150.

Munoz, R., R. Rossello-Mora and R. Amann (2016) Revised phylogeny of Bacteroidetes and proposal of sixteen new taxa and two new combinations including Rhodothermaeota phyl. nov. *Syst Appl Microbiol* **39**: 281-96.

Mutlu, M. B., M. Martinez-Garcia, F. Santos, A. Pena, K. Guven and J. Anton (2008) Prokaryotic diversity in Tuz Lake, a hypersaline environment in Inland Turkey. *FEMS Microbiol Ecol* **65**: 474-83.

Muyzer, G. and K. Smalla (1998) Application of denaturing gradient gel electrophoresis (DGGE) and temperature gradient gel electrophoresis (TGGE) in microbial ecology. *Antonie Van Leeuwenhoek* **73**: 127-41.

Naghoni, A., G. Emtiazi, M. A. Amoozegar, M. S. Cretoiu, L. J. Stal, Z. Etemadifar, *et al.* (2017) Microbial diversity in the hypersaline Lake Meyghan, Iran. *Sci Rep* **7**: 11522.

Nakonieczna, A., C. J. Cooper and R. Gryko (2015) Bacteriophages and bacteriophage-derived endolysins as potential therapeutics to combat Gram-positive spore forming bacteria. *J Appl Microbiol* **119**: 620-31.

Narasingarao, P., S. Podell, J. A. Ugalde, C. Brochier-Armanet, J. B. Emerson, J. J. Brocks, *et al.* (2012) De novo metagenomic assembly reveals abundant novel major lineage of Archaea in hypersaline microbial communities. *Isme J* **6**:81-93.

Oren, A. (2002) Diversity of halophilic microorganisms: environments, phylogeny, physiology, and applications. *J Ind Microbiol Biotechnol* **28**: 56-63.

Oren, A. (2002) Molecular ecology of extremely halophilic Archaea and Bacteria. *FEMS Microbiol Ecol* **39**: 1-7.

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

Oren, A. (2013) Life at high salt concentrations, intracellular KCl concentrations, and acidic proteomes. *Front Microbiol* **4**: 315.

Oren, A., G. Bratbak and M. Heldal (1997) Occurrence of virus-like particles in the Dead Sea. *Extremophiles* **1**: 143-9.

Oren, A., K. Sørensen, D. E. Canfield, A. Teske, D. Ionescu, A. Lipski and K. Altendorf (2009) Microbial communities and processes within a hypersaline gypsum crust in a saltern evaporation pond (Eilat, Israel). *Hydrobiologia.* **626**: 15–26.

Orlando, T. M., McCord, T.B., Grieves, G.A. (2005) The chemical nature of Europa surface material and the relation to a subsurface ocean. *Icarus* **177**: 528-533.

Ortmann, A. C. and C. A. Suttle (2009) Determination of virus abundance by epifluorescence microscopy. *Methods Mol Biol* **501**: 87-95.

Pagaling, E., R. D. Haigh, W. D. Grant, D. A. Cowan, B. E. Jones, Y. Ma*, et al.* (2007) Sequence analysis of an archaeal virus isolated from a hypersaline lake in Inner Mongolia, China. *BMC Genomics* **8**: 410.

Parks, D. H., M. Imelfort, C. T. Skennerton, P. Hugenholtz and G. W. Tyson (2014) CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res* **25**: 1043-55.

Parro, V., G. de Diego-Castilla, J. A. Rodriguez-Manfredi, L. A. Rivas, Y. Blanco-Lopez, E. Sebastian*, et al.* (2011) SOLID3: a multiplex antibody microarray-based optical sensor instrument for in situ life detection in planetary exploration. *Astrobiology* **11**: 15-28.

Parro, V., P. Fernandez-Calvo, J. A. Rodriguez Manfredi, M. Moreno-Paz, L. A. Rivas, M. Garcia-Villadangos*, et al.* (2008) SOLID2: an antibody array-based life-detector instrument in a Mars Drilling Simulation Experiment (MARTE). *Astrobiology* **8**: 987-99.

Peng, Y., H. C. Leung, S. M. Yiu and F. Y. Chin (2012) IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics* **28**: 1420-8.

Pernthaler, A. and R. Amann (2004) Simultaneous fluorescence in situ hybridization of mRNA and rRNA in environmental bacteria. *Appl Environ Microbiol* **70**: 5426-33.

Pietila, M. K., S. Laurinavicius, J. Sund, E. Roine and D. H. Bamford (2010) The single-stranded DNA genome of novel archaeal virus halorubrum pleomorphic virus 1 is enclosed in the envelope decorated with glycoprotein spikes. *J Virol* **84**: 788-98.

Pilsyk, S. and A. Paszewski (2009) Sulfate permeasesphylogenetic diversity of sulfate transport. *Acta Biochim Pol* **56**: 375-84.

Pina, M., A. Bize, P. Forterre and D. Prangishvili (2011) The archeoviruses. *FEMS Microbiol Rev* **35**: 1035-54.

Podell, S., J. A. Ugalde, P. Narasingarao, J. F. Banfield, K. B. Heidelberg and E. E. Allen (2013) Assembly-driven community genomics of a hypersaline microbial ecosystem. *PLoS One* **8**: e61692.

Porter, K., P. Kukkaro, J. K. Bamford, C. Bath, H. M. Kivela, M. L. Dyall-Smith and D. H. Bamford (2005) SH1: A novel, spherical halovirus isolated from an Australian hypersaline lake. *Virology* **335**: 22-33.

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

Prieto-Ballesteros, O., N. Rodriguez, J. S. Kargel, C. G. Kessler, R. Amils and D. F. Remolar (2003) Tirez lake as a terrestrial analog of Europa. *Astrobiology* **3**: 863-77.

Pruesse, E., J. Peplies and F. O. Glockner (2012) SINA: accurate high-throughput multiple sequence alignment of ribosomal RNA genes. *Bioinformatics* **28**: 1823-9.

Quast, C., E. Pruesse, P. Yilmaz, J. Gerken, T. Schweer, P. Yarza*, et al.* (2013) The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res* **41**: D590-6.

Raes, J., J. O. Korbel, M. J. Lercher, C. von Mering and P. Bork (2007) Prediction of effective genome size in metagenomic samples. *Genome Biol* **8**: R10.

Ramos-Barbero, M. D., A. B. Martin-Cuadrado, T. Viver, F. Santos, M. Martinez-Garcia and J. Anton (2019) Recovering microbial genomes from metagenomes in hypersaline environments: The Good, the Bad and the Ugly. *Syst Appl Microbiol* **42**: 30-40
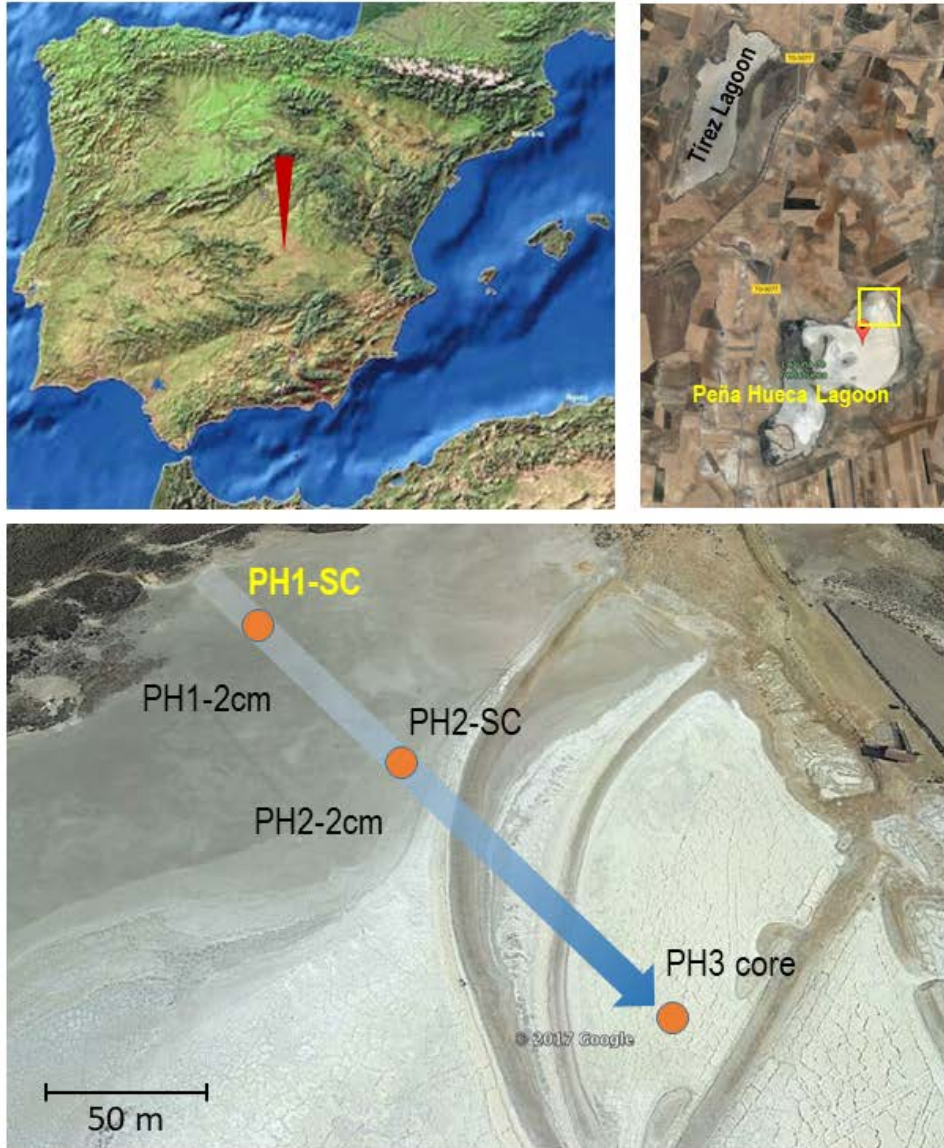
Ramos-Barbero, M. D., J. M. Martínez, C. Almansa, N. Rodríguez, J. Villamor, M. Gomariz*, et al.* (2019) Prokaryotic and viral community structure in the singular chaotropic salt lake Salar de Uyuni. *Environ. Microbiol. In press.*

Rasuk, M. C., A. B. Fernandez, D. Kurth, M. Contreras, F. Novoa, D. Poire and M. E. Farias (2016) Bacterial Diversity in Microbial Mats and Sediments from the Atacama Desert. *Microb Ecol* **71**: 44-56.

Reith, J. and C. Mayer (2011) Peptidoglycan turnover and recycling in Gram-positive bacteria. *Appl Microbiol Biotechnol* **92**: 1-11.

Rice, P., I. Longden and A. Bleasby (2000) EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet* **16**: 276-7.

Richter, M. and R. Rosello-Mora (2009) Shifting the genomic gold standard for the prokaryotic species definition. *Proc Natl Acad Sci U S A* **106**: 19126-31.

Rivas, L. A., M. Garcia-Villadangos, M. Moreno-Paz, P. Cruz-Gil, J. Gomez-Elvira and V. Parro (2008) A 200-antibody microarray biochip for environmental monitoring: searching for universal microbial biomarkers through immunoprofiling. *Anal Chem* **80**: 7970-9.

Rocha, E. P. and A. Danchin (2002) Base composition bias might result from competition for metabolic resources. *Trends Genet* **18**: 291-4.

Rodriguez-Brito, B., L. Li, L. Wegley, M. Furlan, F. Angly, M. Breitbart*, et al.* (2010) Viral and microbial community dynamics in four aquatic environments. *Isme J* **4**: 739-51.

Rodriguez-Valera, F., A. Ventosa, G. Juez and J. F. Imhoff (1985) Variation of environmental features and microbial populations with salt concentrations in a multi-pond saltern. *Microb Ecol* **11**: 107-115.

Rodriguez, R. L., S. Gunturu, W. T. Harvey, R. Rossello-Mora, J. M. Tiedje, J. R. Cole and K. T. Konstantinidis (2018) The Microbial Genomes Atlas (MiGA) webserver: taxonomic and gene diversity analysis of Archaea and Bacteria at the whole genome level. *Nucleic Acids Res* **46**: W282-W288.

Rothschild, L. J. and R. L. Mancinelli (2001) Life in extreme environments. *Nature* **409**: 1092-101.

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

Roux, S., M. Krupovic, D. Debroas, P. Forterre and F. Enault (2013) Assessment of viral community functional potential from viral metagenomes may be hampered by contamination with cellular sequences. *Open Biol* **3**: 130160.

Rubin, S. S., I. Marin, M. J. Gomez, E. A. Morales, I. Zekker, P. San Martin-Uriz*, et al.* (2017) Prokaryotic diversity and community composition in the Salar de Uyuni, a large scale, chaotropic salt flat. *Environ Microbiol* **19**: 3745-3754.

Rusch, D. B., A. L. Halpern, G. Sutton, K. B. Heidelberg, S. Williamson, S. Yooseph*, et al.* (2007) The Sorcerer II Global Ocean Sampling Expedition: Northwest Atlantic through Eastern Tropical Pacific. *PLoS Biol* **5**: e77.

Santos, F., A. Meyerdierks, A. Pena, R. Rossello-Mora, R. Amann and J. Antón (2007) Metagenomic approach to the study of halophages: the environmental halophage 1. *Environ Microbiol* **9**: 1711-23.

Santos, F., M. Moreno-Paz, I. Meseguer, C. Lopez, R. Rossello-Mora, V. Parro and J. Antón (2011) Metatranscriptomic analysis of extremely halophilic viral communities. *ISME J* **5**: 1621-33.

Santos, F., P. Yarza, V. Parro, C. Briones and J. Antón (2010) The metavirome of a hypersaline environment. *Environ Microbiol* **12**:2965-76.

Sass, A. M., B. A. McKew, H. Sass, J. Fichtel, K. N. Timmis and T. J. McGenity (2008) Diversity of *Bacillus*-like organisms isolated from deep-sea hypersaline anoxic sediments. *Saline Systems* **4**: 8.

Saum, S. H. and V. Müller (2007) Salinity-dependent switching of osmolyte strategies in a moderately halophilic bacterium: glutamate induces proline biosynthesis in *Halobacillus halophilus*. *J Bacteriol* **189**: 6968-75.
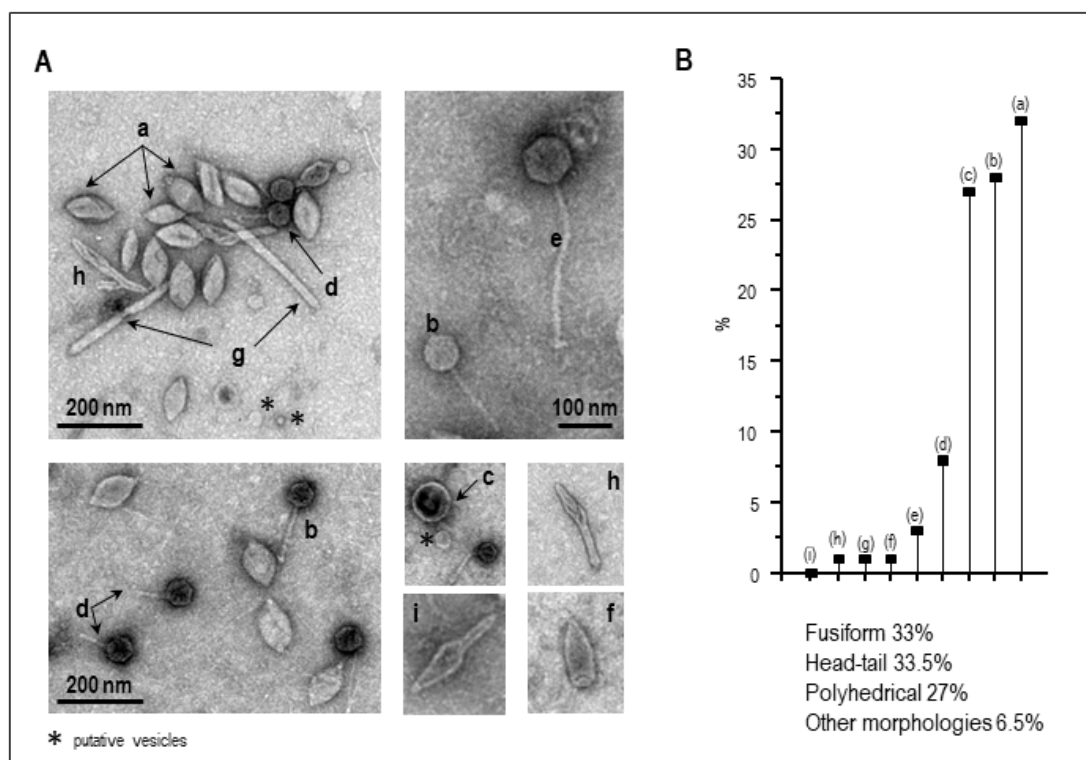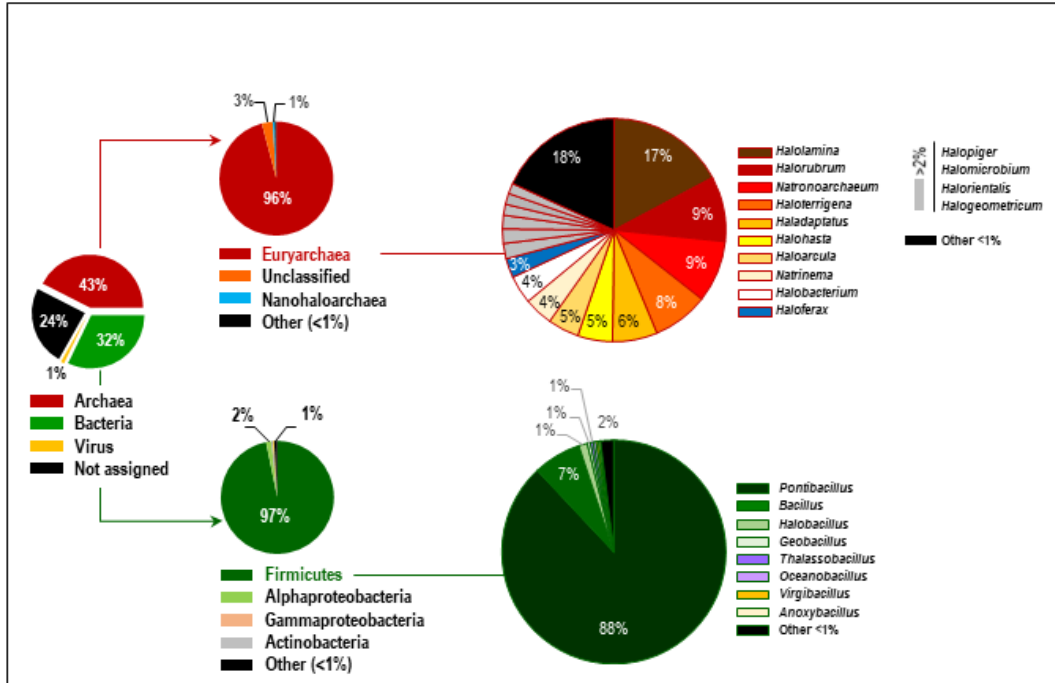
Schloss, P. D., S. L. Westcott, T. Ryabin, J. R. Hall, M. Hartmann, E. B. Hollister*, et al.* (2009) Introducing Mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl Environ Microbiol* **75**: 7537-41.

Schwenzer, S. P., J. C. Bridges, R. C. Wiens, P. G. Conrad, S. P. Kelley, R. Leveille*, et al.* (2016) Fluids during diagenesis and sulfate vein formation in sediments at Gale crater. *Mars. Meteoritics and Planetary Science* **51**: 2175-2202.

Seed, K. D., D. W. Lazinski, S. B. Calderwood and A. Camilli (2013) A bacteriophage encodes its own CRISPR/Cas adaptive response to evade host innate immunity. *Nature* **494**: 489-91.

Shannon, P., A. Markiel, O. Ozier, N. S. Baliga, J. T. Wang, D. Ramage*, et al.* (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* **13**: 2498-504.

Sime-Ngando, T., S. Lucas, A. Robin, K. P. Tucker, J. Colombet, Y. Bettarel*, et al.* (2010) Diversity of virus-host systems in hypersaline Lake Retba, Senegal. *Environ Microbiol* **13**:1956-72

Snaidr, J., R. Amann, I. Huber, W. Ludwig and K. H. Schleifer (1997) Phylogenetic analysis and in situ identification of bacteria in activated sludge. *Appl Environ Microbiol* **63**: 2884-96.

Soding, J., A. Biegert and A. N. Lupas (2005) The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res* **33**: W244-8.

Sorokin, D. Y., T. P. Tourova, T. V. Kolganova, E. N. Detkova, E. A. Galinski and G. Muyzer (2011) Culturable diversity of lithotrophic haloalkaliphilic sulfate-reducing bacteria in soda lakes and the description of *Desulfonatronum thioautotrophicum* sp. nov., *Desulfonatronum thiosulfatophilum* sp. nov., *Desulfonatronovibrio thiodismutans* sp. nov., and *Desulfonatronovibrio magnus* sp. nov. *Extremophiles* **15**: 391-401.

Summer, E. J., J. J. Gill, C. Upton, C. F. Gonzalez and R. Young (2007) Role of phages in the pathogenesis of Burkholderia, or 'Where are the toxin genes in *Burkholderia* phages?' *Curr Opin Microbiol* **10**: 410-7.

Suttle, C. A. and J. Fuhrman (2010) Enumeration of virus particles in aquatic or sediment samples by epifluorescence microscopy. *Manual of Aquatic Viral Ecology* **15**: 145-153.

Tang, S. L., S. Nuttall, K. Ngui, C. Fisher, P. Lopez and M. Dyall-Smith (2002) HF2: a double-stranded DNA tailed haloarchaeal virus with a mosaic genome. *Mol Microbiol* **44**: 283-96.

Tatusov, R. L., D. A. Natale, I. V. Garkavtsev, T. A. Tatusova, U. T. Shankavaram, B. S. Rao*, et al.* (2001) The COG database: new developments in phylogenetic classification of proteins from complete genomes. *Nucleic Acids Res* **29**: 22-8.

Tuross, N. and L. Stathoplos (1993) Ancient proteins in fossil bones. *Methods Enzymol* **224**: 121-9.

Ulyantsev, V. I., S. V. Kazakov, V. B. Dubinkina, A. V. Tyakht and D. G. Alexeev (2016) MetaFast: fast reference-free graph-based comparison of shotgun metagenomic data. *Bioinformatics* **32**: 2760-7.

Vaniman, D. T., D. L. Bish, S. J. Chipera, C. I. Fialips, J. W. Carey and W. C. Feldman (2004) Magnesium sulphate salts and the history of water on Mars. *Nature* **431**: 663-5.

Vavourakis, C. D., R. Ghai, F. Rodriguez-Valera, D. Y. Sorokin, S. G. Tringe, P. Hugenholtz and G. Muyzer (2016) Metagenomic insights into the uncultured diversity and physiology of microbes in four hypersaline soda lake brines. *Front Microbiol* **7**: 211.

Ventosa, A., R. R. de la Haba, C. Sanchez-Porro and R. T. Papke (2015) Microbial diversity of hypersaline environments: a metagenomic approach. *Curr Opin Microbiol* **25**: 80-7.

Vera-Gargallo, B. and A. Ventosa (2018) Metagenomic insights into the phylogenetic and metabolic diversity of the prokaryotic community dwelling in hypersaline soils from the Odiel saltmarshes (SW Spain). *Genes (Basel)* **9**.

Wu, Y. W., B. A. Simmons and S. W. Singer (2016) MaxBin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets. *Bioinformatics* **32**: 605-7.

Yakimov, M. M., V. La Cono, V. Z. Slepak, G. La Spada, E. Arcadi, E. Messina*, et al.* (2013) Microbial life in the Lake Medee, the largest deep-sea salt-saturated formation. *Sci Rep* **3**: 3554.

Yang, Y., Z. Zou, M. He and G. Wang (2011) *Pontibacillus yanchengensis* sp. nov., a moderately halophilic bacterium isolated from salt field soil. *Int J Syst Evol Microbiol* **61**: 1906-11.

Yarza, P., W. Ludwig, J. Euzeby, R. Amann, K. H. Schleifer, F. O. Glockner and R. Rossello-Mora (2010) Update of the All-Species Living Tree Project based on 16S and 23S rRNA sequence analyses. *Syst Appl Microbiol* **33**: 291-9.

Yébenes, A., R. Marfil, Y. Soriano, Y. A. De La Peña and M. Díaz Molina (1977) El Trías de Alcázar de San Juan (Región de La Mancha). *Cuadernos de Geologia Ibérica*, 4, 495-508.

Zolotov, M. Y. and E. L. Shock (2001) Composition and stability of salts on the surface of Europa and their oceanic origin. *Journal of Geophysical Research* **106**: 32815-32827.

Figure 1. Martin-Cuadrado et al. 2019

A

200 nm

100 nm

200 nm

* putative vesicles

B

Fusiform 33%
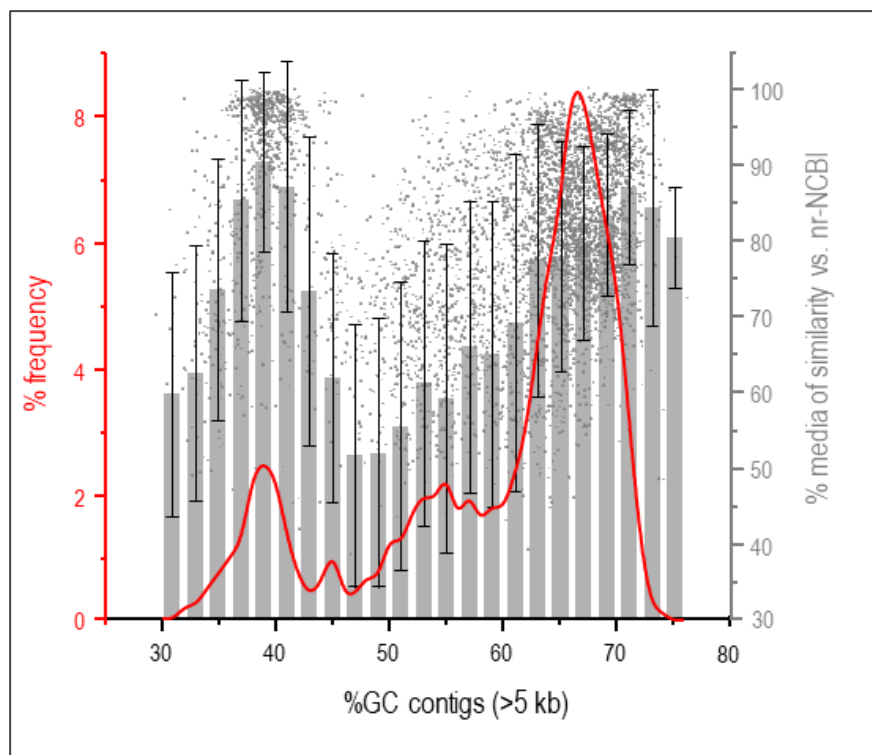Head-tail 33.5%
Polyhedrical 27%
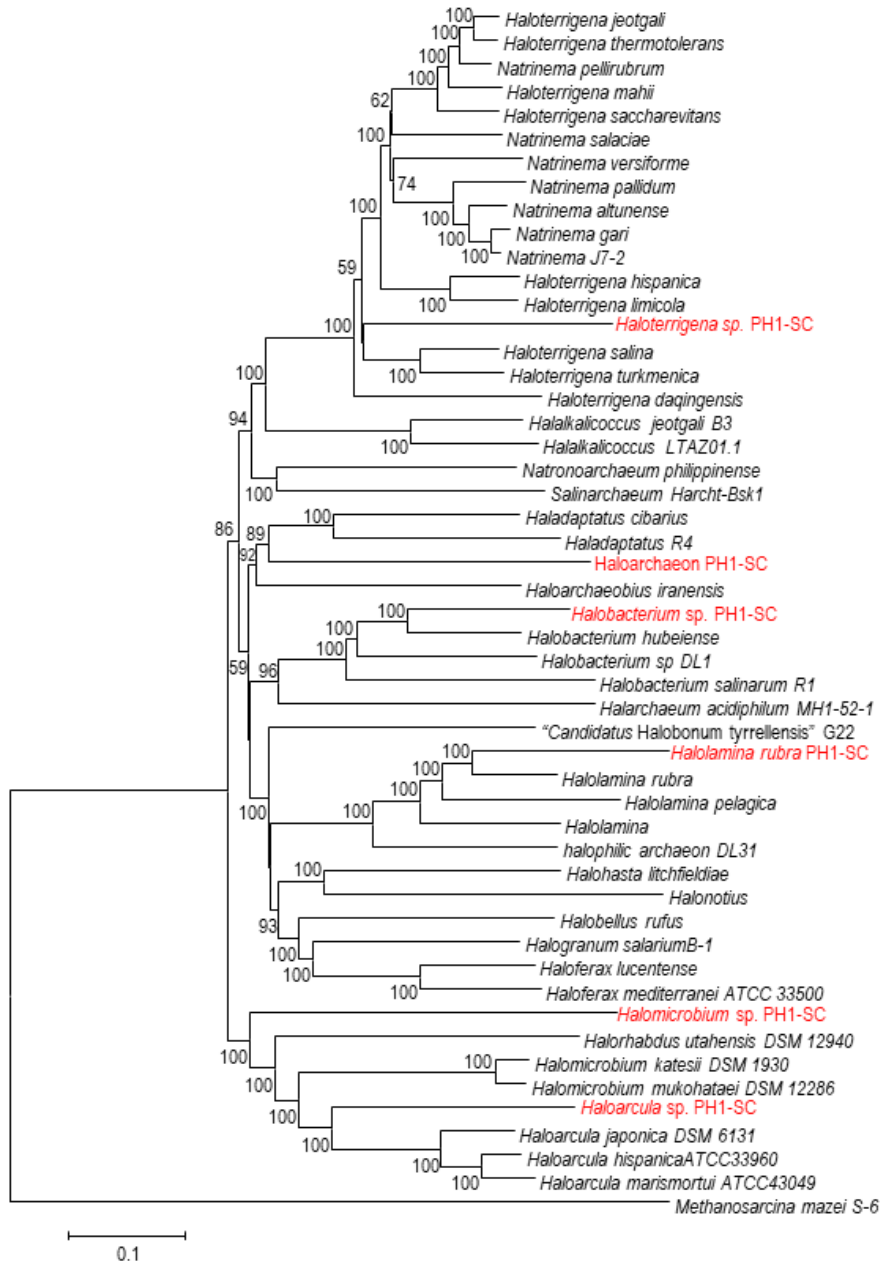Other morphologies 6.5%

Figure 5. Martin-Cuadrado et al. 2019
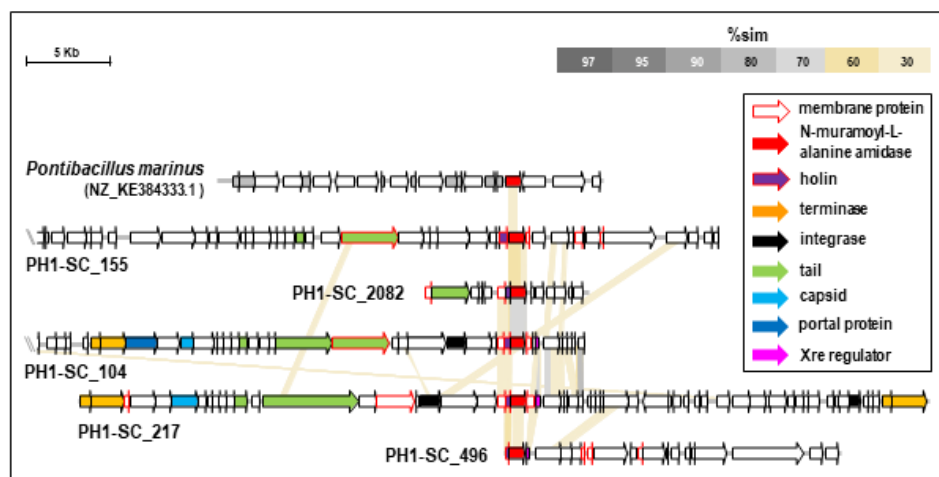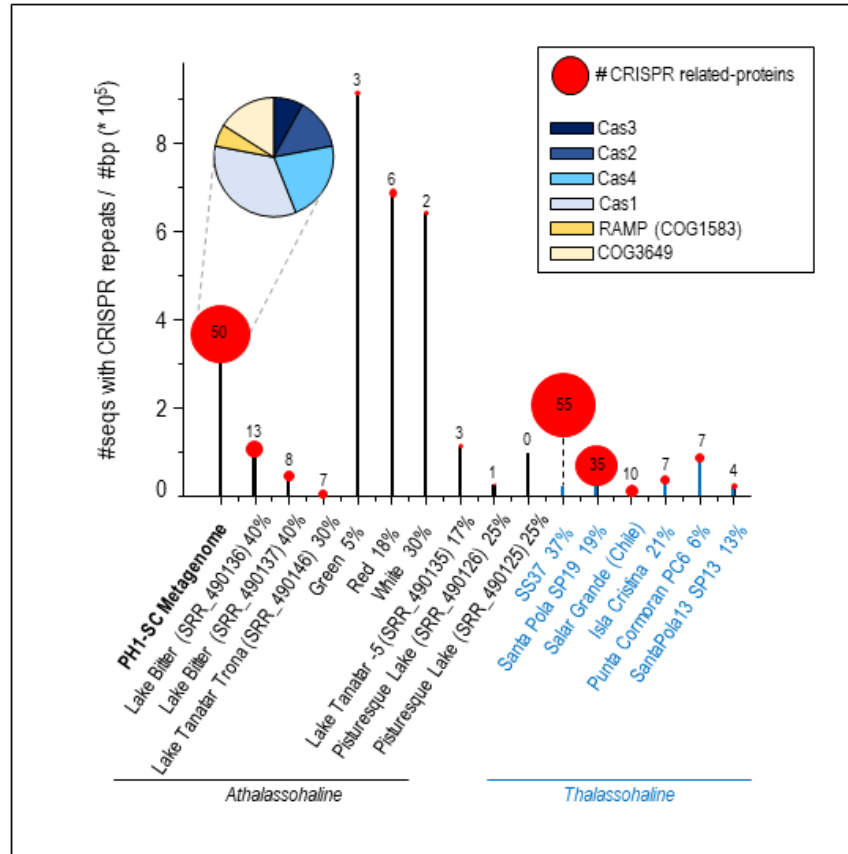
Table 1. General features of the Peñahueca lagoon salt-crust (PH1-SC) sample. Summary of pyrosequenced 16S-rRNA amplicons and statistical indexes. Genomic features of the metagenome and metavirome obtained, raw sequence reads and number of assembled contigs.

PH1-SC 16S-rRNA amplicons	Bacteria	Archaea
#raw reads/trimmed	4033/2329	14026/7654
OTUs[a]	41	456
OPUs	5	29
Dominance-D	0.84	0.05
Shannon-Wiener H	0.51	3.91
Evenness	0.04	0.12
Good´s coverage	0.99	0.97
PH1-SC	**Metagenome**	**Metavirome**
# sequences	28477984	186046
# Mb	6912	99.19
Read length (media. bp)	242.7	533.2
# assembled contigs (>500pb)	108281	9065
Total assembly length (pb)	188 029 879	10 051 707
#contigs ≥ 5 Kb	5122	87
Maximum contig length (pb)	113404	20686

Table 2. General features of the reconstructed genomes from the Peñahueca saline crust (PH1-SC)

BIN-Name	Haloarchaeon PH1-SC	*Halobacterium* sp. PH1-SC	*Halomicrobium* sp. PH1-SC	*Haloterrigena* sp. PH1-SC	Nanohaloarchaeon PH1-SC	*Haloarcula* sp. PH1-SC	*Halolamina rubra* PH1-SC
Taxonomic affiliation	*Haladaptatus*	*Halobacterium*	*Halomicrobium*	*Haloterrigena*	Nanohaloarchaea	*Haloarcula*	*Halolamina rubra*
length (pb)	2906552	2183499	2524619	1660287	592217	2640068	1838310
#contigs	271	174	255	84	59	217	188
N50	12057	14694	10782	30645	10743	14632	10314
ANI	73.51%	86.60%	76.30%	88.02%	68.20%	76.61%	95.44%
AAI	63.18% *Haladaptatus paucihalophilus*	69.07% *Halobacterium salinarum* R1	67.6% *Halomicrobium mukohataei* DSM 12286	86.31% *Haloterrigena pellirubrum* DSM 15624	54.59% Nanohaloarchaea archaeon SG9	70.65% *Haloarcula hispanica* N601	70.75% *Halolamina rubra*
%GC (+-deviation)	66.36 (1.73)	69.52 (1.30)	69.62 (1.22)	68.05 (0.65)	50.48 (0.97)	66.26 (1.46)	70.31 (1.57)
# reads recruited (%)	1706626 (5.9)	172834 (0.6)	175851 (0.6)	227549 (0.8)	25276 (0.08)	263051 (0.9)	297920 (1.1)
rpkg(s)	83.48	11.2	9.7	19.5	6.14	13.9	22.8
%genome (1) (2) (3)	94.2-66.67-92.5	88.2-55.9-84.9	80.1-57.6-83.1	51.4-38.7-62.3	40.1-33.3-47.2	48.5-51.3-60.4	51.4-44.1-60.3
CheckM completeness	71.54	83.01	67.83	53.39	36.63	79.23	59.52
CheckM contamination	0.00	2.98	4.27	4.09	0.00	1.80	1.95
DAStool completeness	81.58	73.68	71.05	68.42	52.63	65.79	65.79
DAStool redundancy	5.88	7.89	5.88	5.88	1.96	7.89	7.84
tRNAs	16	16	16	11	12	13	16
rRNAs	ND	ND	ND	108 pb-5S	252 pb - 23S	111 pb-5S	ND

Table 3. Number of housekeeping genes and CRISPR related proteins found in PH1-SC metavirome reads.

	median size read (pb)	Mb	reads with CRISPR arrays	seqs/Mb - %total	Size Repeat-protospacer (pb)	ORFs	CRISPR-related proteins	16S rRNA	RecA	RpoB	GyrA	GyrB	SecY	RP[a]	BR[b]	PAPS reductase[c]
complete	533.2	99.2	80	0.80 – 0.04	28-38	186046	454	2[d]	0	2	1	0	0	14	0	31-200
subset (10⁵ reads)	528.6	52.9	47	0.89 - 0.05	30-35	141590	293	ND	0	0	0	0	0	2	0	15-105

[a]RP: Ribosomal proteins; [b]BR: Bacteriorhodopsin; [c]COGS 3969 and 0175; [d]2 reads with best match with *Pontibacillus* sp. 16S rRNA gene.

Table 4. Number of housekeeping genes and CRISPR related proteins found in several hypersaline environments.

	Metagenomes - subset (10^5 sequences)															
Dataset	median size read (pb)	Mb	reads with CRISPR arrays	seqs/Mb - %total	Size Repeat-protospacer (pb)	ORFs	CRISPR-related proteins	RecA	RpoB	GyrA	GyrB	SecY	RP[a]	BR[b]	PAPS reductase[c]	Ref[c]
PH1-SC	242	24.3	73	3.01 - 0.07	29-36	99867	50	13	55	75	38	25	510	13	8-25	(1)
Green	148	14.8	135	9.1-0.17	25-33	81144	3	11	30	37	33	24	546	5	0-9	(2)
White	148	14.8	95	6.4-0.134	25-36	84041	2	10	36	24	29	16	615	19	1-2	(2)
Red	149	14.8	101	6.8-0.14	25-37	83884	6	10	36	23	15	14	585	14	0-8	(2)
Pisturesque Lake (25%) (SRR490125)	300	30.0	29	0.97 - 0.03	18-26	112575	0	23	59	60	54	35	537	25	0-11	(3)
Pisturesque Lake (25%) (SRR490125)	300	30.0	7	0.23 - 0.01	16-25	100031	1	9	44	33	40	24	325	13	0-2	(3)
Lake Tanatar -5 (17%) (SRR490135)	300	30.0	33	1.1 - 0.03	16-20	115492	3	36	84	98	105	30	621	0	0-10	(3)

Lake Tanatar Trona (30%) (SRR490146)	300	30.0	0	0	ND	115042	7	25	51	68	56	30	519	18	8-4	(3)
Lake Bitter (SRR490136)	300	30.0	27	0.9 - 0.03	16-21.5	108905	13	20	44	40	53	34	432	19	0-2	(3)
Lake Bitter (SRR490137)	300	30.0	11	0.37 - 0.01	16-23	107997	8	19	59	39	33	22	426	19	0-4	(3)
CR30-summer (37%)	438.9	43.9	9	0.21 - 0.01	23-35.5	114357	55	48	107	78	81	49	970	56	0-43	(4)
Santa Pola SP19	435.4	43.5	10	0.23 - 0.01	20.5-36.5	118929	35	57	175	143	170	56	1268	57	0-38	(4)
Santa Pola13 SP13	207	20.7	4	0.19 - (0)	15-27.5	93988	4	40	106	120	88	48	875	9	0-12	(4)
Salar Grande	101	10.1	0	0	ND	90443	10	6	44	22	42	18	242	11	1-5	(5)
Isla Cristina 21	297.6	29.8	8	0.27 - 0.01	24.5-35	99966	7	29	110	62	48	47	730	33	2-24	(6)
Punta Cormoran (6%)	1075.7	107.6	**82**	0.76 - 0.08	21.5-33	75269	**7**	27	82	105	93	40	1007	6	0-24	(7)

[a]RP: Ribosomal proteins; [b]BR: Bacteriorhodopsin; [c]COGS 3969 and 0175.; [c]Ref: (1) This work; (2) Naghoni et al. 2017; (3) Vavourakis et al 2016; (4) Ghai et al 2011; (5) Kris et al., 2016; (6) Fernández et al., 2013; (7) Rush et al 2006.