environmental microbiology

Environmental Microbiology (2021) **23**(7), 3477–3498 doi:10.1111/1462-2920.15630

Ancient saltern metagenomics: tracking changes in microbes and their viruses from the underground to the surface

Mª Dolores Ramos-Barbero,[1] Tomeu Viver,[2] Ane Zabaleta,[3] Ece Senel,[1,4] María Gomariz,[1] Iñaki Antigüedad,[3] Fernando Santos,[1] Manuel Martínez-García,[1] Ramon Rosselló-Móra[2] and Josefa Antón[1*]

[1]*Department of Physiology, Genetics and Microbiology, University of Alicante, 03690 San Vicent del Raspeig, Alicante, Spain.*

[2]*Marine Microbiology Group, Department of Animal and Microbial Diversity, Mediterranean Institute of Advanced Studies (IMEDEA; CSIC-UIB), Esporles, Illes Balears, 07190, Spain.*

[3]*Hydro-Environmental Processes Group, Geology Department, Science and Technology Faculty, University of the Basque Country UPV/EHU, Leioa, 48940, Spain.*

[4]*Department of Biology, Institute of Graduate Programs, Eskisehir Technical University, Yunusemre Campus, Eskisehir, 26470, Turkey.*

Summary

Microbial communities in hypersaline underground waters derive from ancient organisms trapped within the evaporitic salt crystals and are part of the poorly known subterranean biosphere. Here, we characterized the viral and prokaryotic assemblages present in the hypersaline springs that dissolve Triassic-Keuper evaporite rocks and feed the Añana Salt Valley (Araba/Alava, Basque Country, Spain). Four underground water samples (around 23% total salinity) with different levels of exposure to the open air were analysed by means of microscopy and metagenomics. Cells and viruses in the spring water had lower concentrations than what are normally found in hypersaline environments and seemed to be mostly inactive. Upon exposure to the open air, there was an increase in activity of both cells and viruses as well as a selection of phylotypes. The underground water was inhabited by a rich community harbouring a diverse set of genes coding for retinal binding proteins. A total of 35 viral contigs from 15 to 104 kb, representing partial or total viral genomes, were assembled and their evolutionary changes through the spring system were followed by SNP analysis and metagenomic island tracking. Overall, both the viral and the prokaryotic assemblages changed quickly upon exposure to the open air conditions.

Received 7 April, 2021; revised 26 May, 2021; accepted 6 June, 2021. *For correspondence. E-mail anton@ua.es; Tel. +34-965903870; Fax. +34 965909569.

Introduction

The study of microbial communities present in hypersaline springs has the potential to unveil the diversity of life in subterranean biomes. This underground world accounts for up to one-fifth of Earth's biomass but constitutes only around 8% of 16S rRNA genes in public databases (Anantharaman *et al*., 2016). Hyperhalophilic underground communities likely derive from ancient organisms trapped within the evaporitic salt crystals (Vreeland *et al*., 2000; Baati *et al*., 2010). When the hypersaline spring water is used to feed solar salterns, the system can be used to track the changes experienced by the underground community when first exposed to open-air and sunlight irradiation conditions. This monitoring is especially interesting in the case of extreme environments, in which ecological and evolutionary tempos overlap (Iranzo *et al*., 2017, 2019). In fact, microbial communities have been proposed to evolve faster in extreme than in 'mild' environments (Li et al., 2014). Finally, given the high concentrations of viruses normally found in hypersaline environments and the role of viruses in controlling their microbial hosts (Guixa-Boixareu *et al*., 1996), studies of microbial diversity and evolution in extreme environments should be extended to encompass the viral fraction.

However, despite the unique ecology of underground systems, only a few hypersaline underground springs have been studied. Examples of such environments are the Andean hypersaline waters feeding Maras salterns (Maturrano *et al*., 2006) or Antarctic hypersaline

oligotrophic springs (Colangelo-Lillis *et al*., 2016). Here, we have characterized the microbial and viral assemblages of the Añana Ancient Salt Valley hypersaline springs.

The Añana Salt Valley (Araba/Alava, Basque Country, Spain) consists of several saline water springs. This water has been used to produce salt continuously for at least the last 7000 years (Plata and Erkiaga, 2018). The salt valley lies on the northern edge of an active diapiric structure (19 km^2), where Triassic-Keuper evaporite rocks (gypsum and halite) slowly ascend (Frankovic *et al*., 2016) cutting through Lower Cretaceous-Tertiary rocks, and incorporate carbonate and ophitic rocks within a clayey mass. Consequently, the hydrogeological context of the diapir and, in particular, of the salt valley is quite complex. The salt valley is the main discharge area for the diapir's groundwater. There are two types of water in the springs of the valley, saline water and brackish water.

The saline water, which feeds the salt production crystallizers, is drained by five springs, of which the three most important (Hontana, El Cautivo and Santa Engracia) are the subject of this article; in the latter two, the outlet of the water is conditioned as a pool (Fig. 1), so that the water does not flow directly to the outside. Their combined discharge is about 2.5 L s^{-1}, remaining very constant throughout the year, indicative of the aquifer's high regulating capacity. All the saline springs have practically identical physico-chemical characteristics, indicating the same deep origin (−260 m, according to Iribar and Ábalos, 2011) and the same access route to the surface. The electrical conductivity of these waters is 180–220 mS cm^{-1} (with very little fluctuation and no annual cyclicity), and the temperature is 15–18 °C (average air temperature, 11 °C), all indicative of deep and slow flows (as also indicated by Tritium values). The waters are high in chloride-sodium (Cl: 140–160 g L^{-1}) but with a significant presence of sulfates (4–4.8 g L^{-1}) (unpublished data provided by the Hydro-Environmental Processes Group of the University of the Basque Country).

Our work using samples from underground saline waters showed that these waters harbour a diverse community of prokaryotes and viruses, some of which were selected upon exposure to open air conditions, which likely triggered the activities of bacterial cells and their viruses.

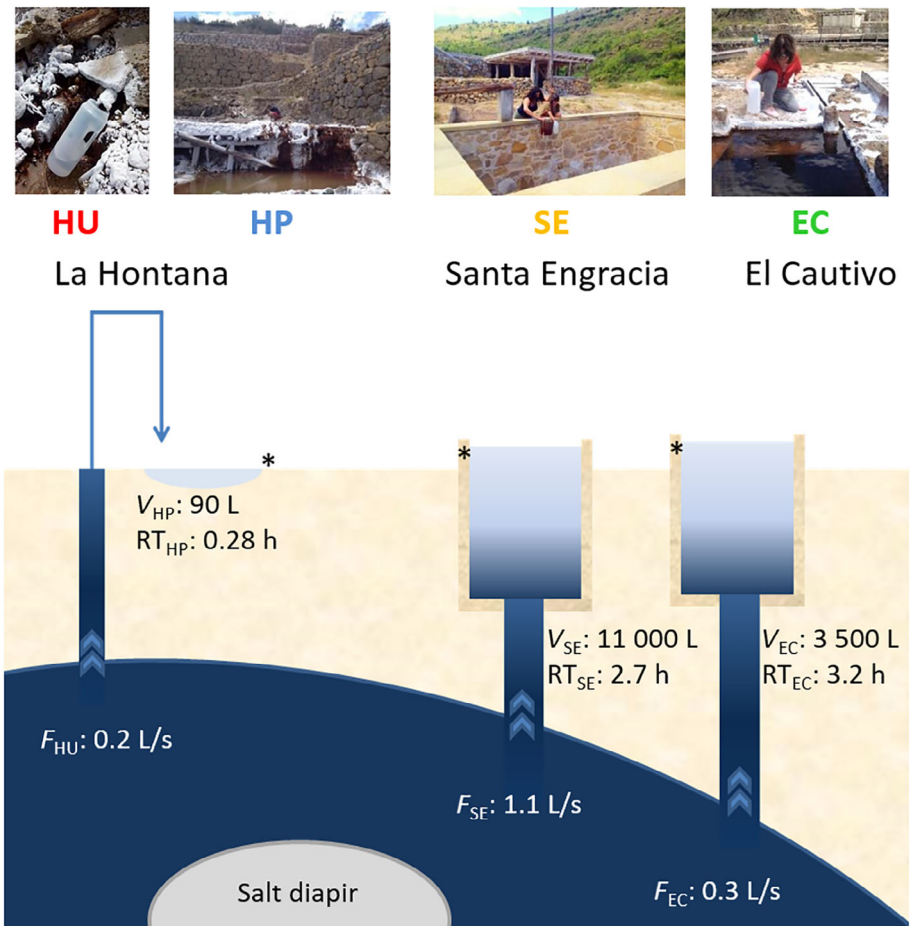

Fig 1. Samples analysed in this work. Brine samples were taken directly from the spring (HU), a pool fed from above with HU water (HP) and two pools (SE, EC) fed from the bottom; all the points with underground water had the same origin. The volume of the water in the pools is kept constant by allowing the water outflow by outlets (asterisks) connected to crystallizer ponds through a series of pipes. F: flux, V: volume, RT: calculated residence times. Subindexes refer to the names of the sampling points. [Color figure can be viewed at wileyonlinelibrary.com]

Results and discussion

The system and general microbial features

Four water samples were taken for this study: three from the springs La Hontana (HU, 'Hontana upwelling'), Santa Engracia (SE) and El Cautivo (EC), and one from the shallow puddle fed by the underground water emerging at HU, which in turn feeds some crystallizer ponds used for salt production (Fig. 1). Thus, the same underground brine emerging at HU also emerges at the bottom of SE and EC pools where it is temporarily stored and used for feeding the crystallizers. As shown in Fig. 1, these pools' walls are covered with concrete and isolated from soil. Brine input and output flows are kept equal, so the volumes of these pools remain constant. Volumes, flow rates and calculated residence times are shown in Fig. 1. Note that, from a hydrogeological point of view, the three springs have the same 'underground' origin and the mention of 'surface' or 'open air' in this work refers only to the exposure of these waters to the conditions of the pools where they were temporarily stored.

Overall, residence times in the water reservoirs are rather short (from less than 1 h in the La Hontana puddle to around 3 h in EC and SE pools). However, these are average values and small amounts of cells or viruses may remain for longer times in the reservoirs. Note that: (i) HP is directly exposed to the environment (both air and clay soil) since it is fed from the source HU; (ii) HU has never been exposed to the exterior conditions as its waters flow directly from the underground spring; and (iii) SE and EC pools, which are fed from the bottom, contain a mixture of fresh and stored (for an average of ca. 3 h) groundwater. Therefore, water in SE and EC has been less exposed to the open air/surface conditions than that from HP. As shown in Table 1 the main component of the four analysed brines is NaCl, in good agreement with previous results (see above).

Total cell and CARD-FISH counts with general probes for *Bacteria* and *Archaea* were determined for the four samples (Supplementary Fig. S1; Table 1). Overall, the analysed samples, except HP, harboured around 10^6 cells ml^{-1}, which was lower than what is normally found in coastal hypersaline waters (Santos *et al*., 2012; Boujelben *et al*., 2012b; Di Meglio *et al*., 2016; Roux *et al*., 2016). La Hontana spring water (HU), freshly emerged from the underground, harboured 2.68×10^6 cells ml^{-1}, which is greater than the density of cells in an inland hypersaline spring feeding the Peruvian salterns of Maras, which had around 100 cells ml^{-1} (Maturrano *et al*., 2006), a fact that may be related to the relatively high temperature of Añana underground saline water, which is considered to be higher than that recorded at the emerging points (Iribar and Ábalos, 2011). Not unexpectedly, the fraction of cells detected with CARD-FISH was very low in the HU sample and increased when the underground brine was exposed to the open air conditions. However, the number of cells detected by CARD-FISH never surpassed 50% of the total cell counts. The

Table 1. Characteristics of the samples from Añana Ancient Salt Valley used in this work.

	Samples			
Parameters	La Hontana upwelling (HU)	La Hontana pond (HP)	Santa Engracia (SE)	El Cautivo (EC)
pH	7.34	6.68	6.68	6.69
%Salt[a]	24	23	23.8	23
Cl^-	156.98	163.60	156.59	150.19
Br^-	0.10	0.09	0.09	0.09
NO_3^-	0.04	0.04	0.04	0.03
SO_4^{2-}	4.36	4.48	4.32	4.23
Na^+	100.09	95.92	96.04	92.18
K^+	0.47	0.45	0.44	0.43
Mg^{2+}	0.81	0.37	0.25	0.28
Ca^{2+}	1.19	1.29	1.55	1.35
VLP ml^{-1}[b]	$7.00 \times 10^4 \pm 2.16 \times 10^4$	$2.09 \times 10^6 \pm 9.30 \times 10^5$	$4.40 \times 10^5 \pm 1.73 \times 10^5$	$2.24 \times 10^5 \pm 1.01 \times 10^5$
Archaea ml^{-1}[c]	$1.60 \times 10^5 \pm 1.26 \times 10^5$	$1.23 \times 10^7 \pm 2.10 \times 10^6$	$2.10 \times 10^6 \pm 5.77 \times 10^5$	$1.18 \times 10^6 \pm 4.05 \times 10^5$
Bacteria ml^{-1}[c]	$3.20 \times 10^5 \pm 2.53 \times 10^5$	$4.80 \times 10^6 \pm 2.00 \times 10^5$	$1.30 \times 10^6 \pm 3.46 \times 10^5$	$1.06 \times 10^6 \pm 9.85 \times 10^5$
Cells ml^{-1}[d]	$2.68 \times 10^6 \pm 8.27 \times 10^5$	$2.57 \times 10^7 \pm 1.96 \times 10^6$	$6.75 \times 10^6 \pm 3.00 \times 10^5$	$3.43 \times 10^6 \pm 1.72 \times 10^6$
% Archaea[c]	6.0	47.9	31.1	34.4
% Bacteria[c]	11.9	18.7	19.3	30.9
% ND[e]	82.1	33.5	49.6	34.7
VLP/cell	0.03	0.08	0.07	0.07

[a]Measured *in situ* with a hand refractometer.
[b]Epifluorescence counts following Sybr-Gold staining.
[c]CARD-FISH.
[d]DAPI.
[e]Percentage of cells not detected by FISH with Archaea/Bacteria probes. Concentrations of ions are given in g L^{-1}.

number of 16S rRNA molecules (i.e. ribosomes) per cell required to detect a bacterial cell by CARD-FISH is around a few tens, depending on the sample conditions (Hoshino *et al*., 2008), which is very low even for active bacteria which frequently harbour more ribosomes (Amann and Fuchs, 2008). Therefore, the low detectability levels in the samples could indicate that a large proportion of the cells might be inactive, dormant or dead, especially for the emerging brine (HU).

The number of viruses (counted as virus-like particles, VLPs) was remarkably low in the HU, with a virus to cell ratio of ~0.03. This ratio is extremely low compared to aboveground high salt systems, which normally harbour from 10 to up to hundreds of viruses per cell (Di Meglio *et al*., 2016). In the case of Arctic hypersaline spring brines, the ratio is lower than 10 and viruses always outnumber microbes (Colangelo-Lillis *et al*., 2016). However, in other underground freshwaters, this ratio is around 0.1, with around 10^4 VLPs ml^{-1} (Roudnew *et al*., 2013; Parikka *et al*., 2017). This could again be related to the low metabolic activity or dormancy of the available hosts, since viruses rely on their host energy resources for their replication (Mahmoudabadi *et al*., 2017). Accordingly, viral particle concentrations in HP, SE and EC increased compared to HU. As with the cells, HP presented the highest VLP concentration while SE and EC, filled with a mix of freshly upwelled and stored water, presented intermediate concentrations.

Metagenomic analysis of cell and virus assemblages

Overall characteristics of viral and cellular metagenomes

The microbial (cells and viruses) communities in the four Añana water samples were analysed by metagenome sequencing. Characteristics of the metagenomes are shown in Supplementary Table S1. All metagenomes could be considered of good quality according to the criteria of Rodriguez-R and Konstantinidis (2014), with sequencing coverages above 60% as calculated by Nonpareil (Rodriguez-R and Konstantinidis, 2014). The viral metagenomes had a very low 16S rRNA gene signal (below 0.02% of 16S rRNA reads) and thus cellular contamination could be considered negligible (Roux *et al*., 2013; Enault *et al*., 2017).

Prior to assembly and annotation, unassembled reads from all metagenomes were compared 'all versus all', in order to provide an overall picture of changes that the underground viral and cell communities underwent when exposed to the air. In addition, comparisons using unassembled data overcome biases introduced by metagenomic assembly since in many cases assembly does not capture the whole extant diversity of cell and virus populations in a sample (Martinez-Hernandez *et al*., 2017). Changes experienced by the cellular and viral assemblages upon emerging from the brine were analysed by calculating the percentage of shared sequences (Fig. 2A and B) as well as the pairwise overall ANIs among the corresponding metagenomes (Fig. 2C). Overall, both considering the sequences shared and the similarity of the shared sequences, the cellular assemblage seemed to change more than the viral fraction. As shown in Fig. 2B, SE and EC were the closest cell metagenomes, as could be expected from their similar environmental conditions and in agreement with 16S rRNA genes and MAG analyses (see below), whereas HP and HU were the closest viral metagenomes.

As also shown in Fig. 2D, between 1.4% and 3.5% of the viral metagenome reads were also present in the corresponding cell metagenomes. Although this viral signal could also correspond to viruses attached to the cells or to proviruses, it can be taken a proxy of viruses replicating inside cells (Hallam *et al*., 2004; Ghai *et al*., 2010; López-Pérez *et al*., 2017) and thus as a proxy of viral activity. The higher presence of viral reads in HP compared to HU samples could thus be indicating that viruses are more active in the puddle than in the spring brine, as further discussed below.

Prokaryotic diversity: the underground seed bank

To gain insight into the taxa present in the microbial communities of the samples, reads corresponding to 16S rRNA gene fragments were retrieved from the four metagenomes (Supplementary Dataset 1). As shown in Fig. 3A, more sequences were assigned to *Archaea* (mainly of the class *Halobacteria*) than to *Bacteria*, though the bacterial assemblage was more diverse; both of these results are consistent with other studies of hypersaline environments (Antón *et al*., 2002; Baker *et al*., 2010; Ventosa *et al*., 2015; Mora-Ruiz *et al*., 2018). BLAST analyses (data not shown) of operational phylogenetic units (OPUs), accounting for >1% of the sequences in each sample, indicated that the OPUs were closely related to sequences retrieved from other hypersaline environments worldwide, either terrestrial or aquatic. Very recently, a new species of *Alpha-proteobacteria* (*Altererythrobacter muriae*) has been isolated from the SE spring (Muniozguren *et al*., 2021). In our dataset this bacterium corresponded to OPU 114 (Supplementary Dataset 1) that was detected, albeit at very low abundance, in HU, as also indicated by recruiting the reads from the four cellular metagenomes generated in this work (data not shown) to *A. muriae* genomic sequence.

In good agreement with the analyses mentioned above (Fig. 2C), HU was the most divergent and most diverse

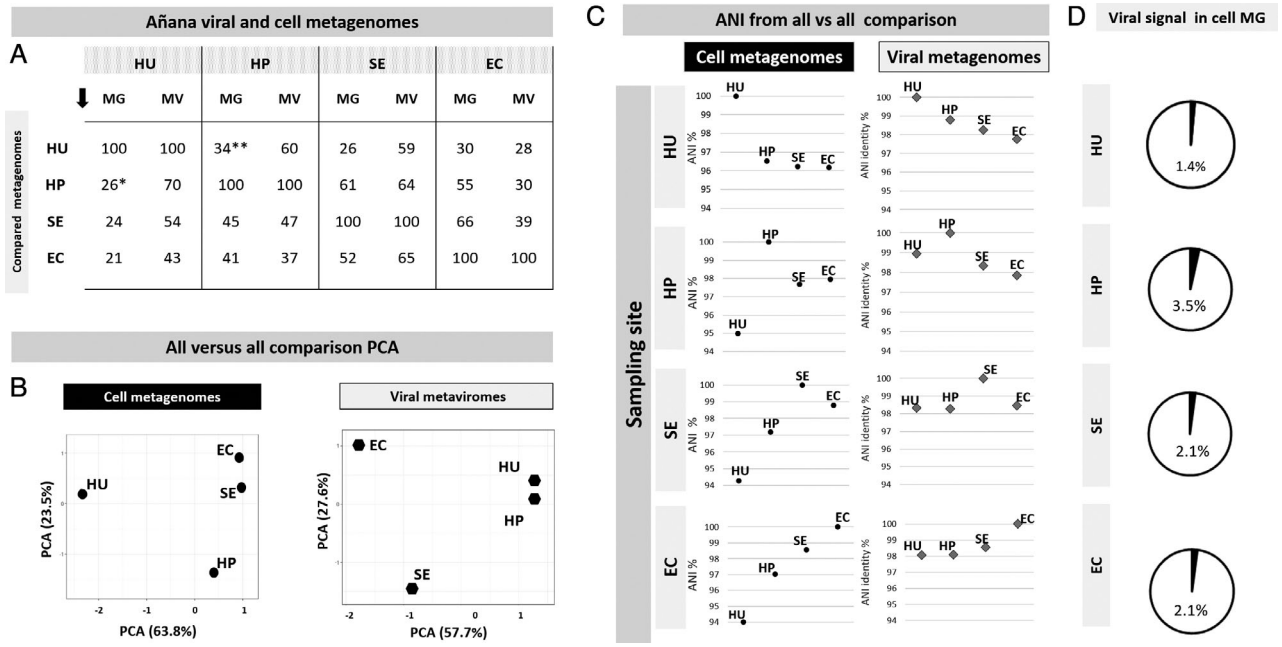

Fig 2. A Comparison of Añana metagenomes.
A. Comparison 'all versus all' among cellular (MG) and viral (MV) metagenomic reads. Numbers indicate the percentage of shared sequences. The direction of the comparison is from top to the bottom. For instance, the cellular metagenome of HU shares 26% of its sequences with the cellular metagenome of HP (*), while the cell metagenome of HP shares 34% of its sequences with the cellular metagenome of HU (**).
B. 'All versus all' PCA (principal component analyses) of Añana viral and cellular metagenomes based on % of reads shared between metagenomes.
C. Graph representation of ANI (average nucleotide identity from shared reads) from comparison all versus all viromes and cell metagenomes of Añana saltern. Only hits over 70% of coverage were considered.
D. Viral signal in cell metagenomes in Añana samples. The pie charts show the percentage of sequences shared between viral-cellular metagenome pairs for each sampling point.

sample according to the composition of its microbiota (Fig. 3B). The community structure in HU was also different in terms of the presence of dominant phylotypes, since HU had fewer phylotypes that dominated. Furthermore, the freshly emerged underground water (HU) harboured the highest proportion (56%) of sample-specific OPUs (i.e. OPUs not present in the rest of the samples, Fig. 3C). These sample-specific taxa could correspond to the dead or dormant cells that did not contribute to the active and viral infection-susceptible fraction of the microbiome. The subsurface water seemed thus to be acting as a seed bank of OPUs (i.e. phylotypes) for the surface system since most OPUs present in the three sampled water reservoirs were also present in the spring (Figs 3C and 4). Thus, there were 'underground' and 'surface' OPUs, as well as some pool-specific ones, albeit these were normally not very abundant in the corresponding metagenomes. As highlighted in Fig. 4, 'underground' phylotypes, which dropped sharply or even disappeared in the samples exposed to the surface conditions, included OPUs 342, 459, 397, 398, 425, corresponding respectively to the candidate phylum *Parcubacteria*, the *Nanohaloarchaeota*, *Haloplanus* sp., *Halolamina* sp. and *Halomicroarcul*a sp. Supplementary dataset S1 provides a detailed description of OPU abundances and changes through the Añana pools. Especially remarkable was the abundance of members of the candidate phylum *Parcubacteria* which seem to be autochthonous to groundwaters (He, *et al*. 2021). Many of these HU-specific OPUs could correspond to members of the ancient groundwater dissolving the salt diapir, given their low abundances and their similarities to mesophilic bacteria (see Supplementary dataset 1). In any case, most of these specific OPUs were present at very low abundances in HU.

Among the 'surface' OPUs the most outstanding case was that of *Salinibacter* sp. (represented by OPU 218), which experienced a fold-change of over 32 in the pools compared to the upwelling. Finally, the presence in HU of a relatively high number (close to 9%) of OPUs related to the chloroplast of the green algae *Dunaliella* sp. was intriguing (Fig. 3A). As expected, the proportion of this OPU was higher in the other pools, where *Dunaliella* could constitute the most important primary producer (Oren, 2014; Keerthi *et al*., 2018), feeding the heterotrophic bacteria, and responsible for the at least a ten-fold increase of cells in HP in comparison with the HU. However, the presence of a phototrophic eukaryotic organism in the dark underground

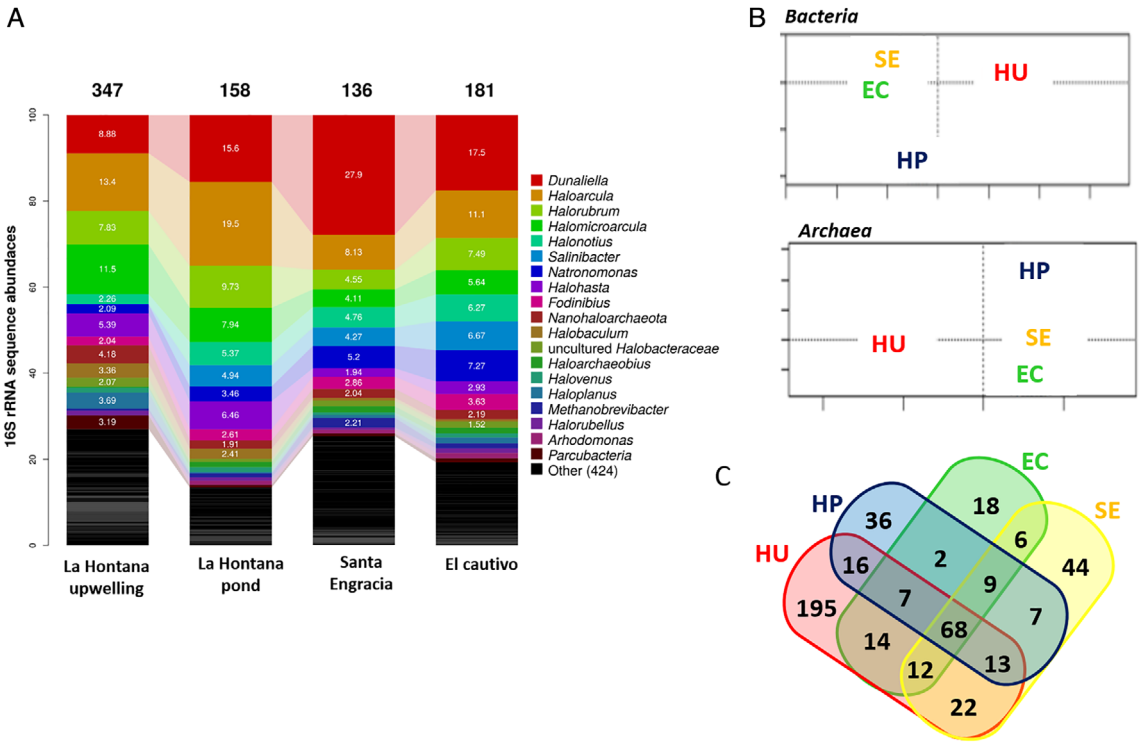

Fig 3. Characterization of the microbial communities based on 16S rRNA reads retrieved from the cellular metagenomes.
A. Taxonomic distribution at the genus level.
B. Communities clustered using coordinated analysis of the weighted UniFrac distance matrix.
C. Venn diagrams showing the genera common to the different metagenomes. [Color figure can be viewed at wileyonlinelibrary.com]

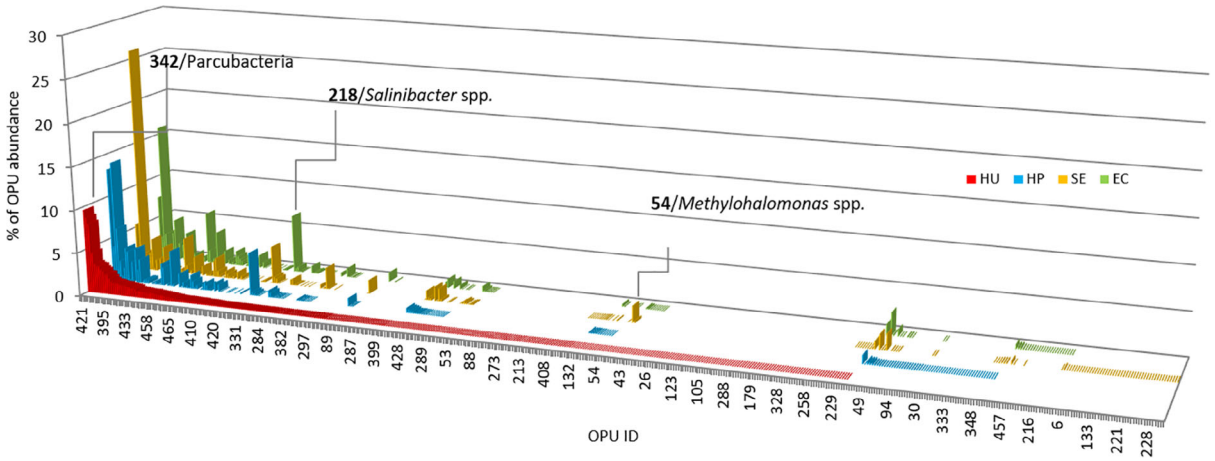

Fig 4. OPU distribution in the four metagenomes (*see Supplementary dataset 1), ordered according to their abundances in the HU sample. Some OPUs are marked as examples. HU is shown in red, HP in blue, SE in yellow and EC in green. [Color figure can be viewed at wileyonlinelibrary.com]

water was rather difficult to explain and could be related to either a dormant state recovered from the dissolution of the evaporite or to a potential for a mixotrophic lifestyle of *Dunaliella* under certain environmental conditions. Whatever the case, the presence of phototrophic microbes in underground systems has been reported previously, such as for viable cyanobacteria in deep continental subsurface (Reisser, 2007; Puente-Sánchez *et al*., 2018).

The annotation of the assembled metagenomes confirmed the dominance of *Archaea* (Supplementary dataset 2) and allowed the retrieval of 18 (almost) complete 16S rRNA gene sequences that could be used for a more accurate phylogenetic analysis (Supplementary Fig. S2). It is worth mentioning that some of the 16S rRNA gene sequences recovered in this work were over 99% identical to sequences obtained in a preliminary study by denaturing gradient gel electrophoresis (DGGE) from Santa Engracia spring in 2011 (in purple in Supplementary Fig. S2 and Supplementary dataset 1) indicating the persistence of the corresponding species in the system.

Metagenome annotation indicated that COG distribution was similar among the three springs and HP. However, it also confirmed that the community of HU was the most divergent of the four analysed samples (Supplementary Fig. S3 A). These results indicated that not only was the phylogenetic diversity wider in HU compared to the rest of samples, but there were genes that were specifically present in the freshly upwelled underground water while others were enriched in the waters exposed to the surface. Among the metabolic categories detected in the samples, the case of retinal binding proteins (RBP, also known as rhodopsins) was especially remarkable. RBPs retrieved from Añana metagenomes (Supplementary Fig. S4) included the three types normally found in hypersaline systems: (i) the outward-directed proton pumps, such as the archaeal bacteriorhodopsin (BR), the bacterial proteorhodopsin (PR), and *Salinibacter* xanthorhodopsin (XR), (ii) the inward-directed chloride pump halorhodopsin (HR), and (iii) the sensory rhodopsins (SRI and SRII) involved in light sensing for phototaxis (Oren, 2002). The presence of light-related genes in the underground water was very intriguing (as was the presence of the green algae *Dunaliella* mentioned above), but even more intriguing was the high diversity of RBPs. As shown in Supplementary Fig. S4, RBPs in HU were not only abundant but were also more diverse than in the rest of samples. Furthermore, many of the RBP coding genes were specific to HU and were not found in the rest of samples, except for XR-related genes. The lower presence of XR in HU was not unexpected since it is produced by *Salinibacter* spp. (and close relatives) which were present only in very low levels in HU. Overall, the underground water harboured diverse novel RBP coding genes as indicated by their distances from previously described rhodopsins. The spring is thus also acting as an RBP seed bank (Baati *et al*., 2010). We can speculate that this seed bank originated in the ancient community trapped in the underground evaporite rocks.

Recovery of microbial genomes from metagenomes: changes in the underground community when exposed to the surface conditions

After 16S rRNA gene read recovery and metagenome annotation, the next step was the assembly and analysis of microbial genomes or metagenome-assembled genomes (MAGs). A total of nine MAGs of different quality (according to Konstantinidis *et al*., 2017) were recovered by co-assembly of the four metagenomes (with the exception of MAG A2_MAG005, which was recovered from EC). Taken together, the MAGs did not constitute a dominant proportion of the total metagenomic sequences (from 1.96% of the reads in HU to 18.1%, 17.8% and 23.03% in HP, SE and EC respectively) but were still useful for tracking genome dynamics across the system. The main characteristics of the MAGs are shown in Table 2 and Supplementary Fig. S5. All MAGs represented new species of known genera or of unclassified higher taxonomic categories. A close look at the annotation of the MAGs (Supplementary dataset 3) indicated interesting features of ecological relevance such as the presence of many genes involved in detoxification of arsenate or other metals, capability of anaerobic nitrate respiration, use of phosphonates, presence of different types of RBPs (see above), among other traits.

In an attempt to track the community changes associated with the exposure of spring water to surface conditions, we recruited metagenome reads to the MAGs in the different pools and calculated ANIr at 95% identity as a proxy of the overall intra-population diversity of the MAG-related genomes (Starnawski *et al*., 2017). Coverage, abundance and single nucleotide polymorphism (SNP) rates were also calculated (Table 2).

Overall, except for MAG021 that was specific to SE, the behaviour of the MAGs in the two pools SE and EC was similar, indicating that the hydrogeologic context of the system exerted some deterministic effects on the populations represented by the MAGs. Most MAGs seemed especially successful in the water bodies exposed to the surface. As expected, the MAG corresponding to a new genus within the *Salinibacteraceae* was especially abundant in the HP, SE and EC samples, where they reached densities above 3.8×10^4 cells ml^{-1} (calculated as in Starnawski *et al*. 2017), as shown in Table 2.

Quantifying the number of SNPs per Mb showed that some populations (MAG002 and MAG015) seemed to have diversified upon exposure to surface conditions, whereas other populations were more homogeneous despite their increased abundances (e.g. MAG005). The high homogeneity of the population represented by MAG013 (*Ectothiorhospiraceae*), which was present in the HP, SE and EC (see also Supplementary Fig. S6), was demonstrated by this MAG having zero SNPs in its

Table 2. Statistics of MAGs recovered from cellular metagenomes.

	MAG002 (A)[a]	MAG003 (A)	MAG004 (A)	MAG005 (A)	MAG013 (B)	MAG015 (A)	MAG021 (B)[a]	MAG022 (A)	A2_MAG005 (B)
No. contigs	233	379	249	516	575	605	776	632	974
No. bases	2 329 966	3 073 719	1 117 659	2 039 487	2 226 472	1 323 474	2 544 609	2 774 985	2 272 857
%GC	64.5	66.9	62.7	62	64.7	65	56.5	66.4	65.3
Completeness	100	92.99	61.5	50	85.8	23.1	82.1	53.8	34.9
Contamination	3.8	5.51	0	3.8	4.7	0	1.9	0	3.8
Closest relative genome by MiGA	*Natronomonas moolapensis*	*Haloarcula* sp. JP L23 NZ CP050014	*Halorubrum* sp. PV6 NZ CP030064	*Halorubrum lacusprofundi* ATCC 49239 NC	*Spiribacter* sp. 2438 NZ CP046046	*Haloarcula hispanica* ATCC 33960	*Thiohalobacter thiocyanaticus* NZ AP018052	*Haloarcula* sp. JP L23 NZ CP050014	*Salinivenus lutea* NZ NCRC00000000
ANI	83.6	ND	ND	ND	ND	ND	ND	ND	ND
AAI	81.0	74.8	71.7	63.0	52.4	54.6	48.6	74.9	79.9
Taxonomy	*Natronomonas* sp. new species	*Haloarcula* sp. new species	*Halobacteriaceae* new genus	*Halorubraceae* new genus	*Ectothiorhodospiraceae* new genus	*Halobacteriales* new family	γ-Proteobacteria new order	*Haloarculaceae* new genus	*Salinibacteraceae* new genus
Abundance in HU[b]	0.08%, 2.14×10^3	0.05%, $1.34 \cdot 10^3$	0.13%, 3.48×10^3	0.23%, 6.16×10^3	0.01%, 2.68×10^2	0.45%, 1.21×10^4	0.01%, 2.68×10^2	0.07%, 1.88×10^3	0.004%, 1.07×10^2
Abundance in HP[b]	0.81%, 2.08×10^5	0.52%, 1.34×10^5	1.90%, 4.88×10^5	1%, 2.57×10^5	0.52%, 1.34×10^5	0.45%, 1.16×10^5	0.02%, 5.14×10^3	2.3%, 5.91×10^5	0.15%, 3.86×10^4
Abundance in SE[b]	2.20%, 1.49×10^5	0.98%, 6.62×10^4	1.00%, 6.75×10^4	1%, 6.75×10^4	0.41%, 2.77×10^4	0.26%, 1.76×10^4	0.95%, 6.41×10^4	0.02%, 1.35×10^3	0.8%, 5.40×10^4
Abundance in EC[b]	2.67%, 9.16×10^4	1.91%, 6.55×10^4	1.89%, 6.48×10^4	1.4%, 4.80×10^4	0.42%, 1.44×10^4	0.29%, 9.95×10^3	0.03%, 1.03×10^3	0.02% 7.55×10^2	1.33%, 4.56×10^4
HU[c]	99.3, 0.7x, 9	97.6, 0.5x 39.6	99.0, 0.9x, 6.8	98.8, 1.7x, 271.6	99.4, 0.1x, 0	99.1, 3.3x, 5.2	99.1, 0.59x, 0	98.5, 0.59X,19	98.6, 0.03X, 0
HP[c]	99.4, 3.0x, 15.4	99.1, 1.9x, 11	99.1, 6.9x, 9.7	99.1, 4.6x, 70.1	99.7, 2.0x, 0	99.1, 1.9x, 7.5	99.1, 0.09x, 0	98.5, 9.9X, 1	97.9, 0.6X, 0
SE[c]	99.5, 17.9x, 207.2	99.5, 6.9x, 13	99.1, 6.7x, 20.9	99.2, 8.5x, 126.9	99.5, 3.1x, 0	99.0, 2.1x, 98.9	99.3, 7.3x, 13.36	96.4, 0.18X, 41	99.3, 6.4X, 2.1
EC[c]	99.4, 11.8x, 133.9	99.5, 8.3x, 6.1	99.1, 8.0x, 27.2	99.2, 7.1x, 54.9	99.5, 2.1x, 0	98.9, 1.4x, 68.7	99.2, 0.15x, 0	96.1, 0.15X, 24.5	99.3, 6.5X, 7

All MAGs were retrieved from co-assembly of the four metagenomes, except for A2_MAG005, which was retrieved from EC.
ND: not determined.
[a](A): Archaea; (B): Bacteria.
[b]Abundance in % nucleotides recruited/Mb followed by the calculated population size in cells ml^{-1} (as in Starnawski *et al*. 2017).
[c]For each of the four ponds the following parameters are provided in this order: ANIr, coverage of the bin, and SNPs Mb^{-1} of bin.

2.2 Mb genome, which may indicate that this bacterium was a recent colonizer of the pools (Meziti *et al*., 2019).

Viral diversity

Taxonomic composition of metagenomes from the free virus assemblages (Supplementary Figs S7–S11) indicated that these communities harboured viruses related to those identified in other hypersaline environments, although with a considerable level of novelty since only 3% of viral contigs could be assigned to known virus groups. The comparison of the functional categories in the four viral metagenomes (Supplementary Fig. S3B) did not show clear differences among them. However, the low proportion of annotated genes (from 10% to 19%, see Supplementary Table S1) in the four viral assemblages hampers any comprehensive comparison.

The assembled fraction represented from 18% to 36% of the viral metagenome total reads (Table S1). This constituted a considerable fraction of the viral assemblage, compared to other viral metagenomes (Hayes *et al*., 2017; Villamor *et al*., 2017). After the assembly of the four viral metagenomes, the non-redundant contigs larger than 15 kb were selected for an in-depth analysis. A total of such 30 contigs (from 15 056 to 104 118 kb, average 30 329, see Table 3) were considered *bona fide* viral genomes based on the criteria described previously (Roux *et al*., 2015, Coutinho *et al*., 2019a; Coutinho *et al*., 2019; Silveira *et al*., 2020). The presence of these viral genomes in the different pools was analysed by contig recruitment as shown in Fig. 5A (see Supplementary Table S2 for abundances). Only four *bona fide* viral genomes were found in all samples (see Supplementary Fig. S12 for their annotation). In most cases, viral genomes, although present in different samples, were assembled only from one of them. This extensively reported 'metagenomics anomaly' (Martinez-Hernandez *et al*., 2017; Ramos-Barbero *et al*., 2019) is likely related to the microdiversity of viral genomes, as discussed below.

Is the viral fraction in the underground water infective?

The concentration of viruses in the HU water was very low, likely too low to allow for infection of host cells. A rough estimate of infection rates in HU, calculated as in Colangelo-Lillis *et al*. (2016), yields as little as 0.0001 contacts per microbe per day. Furthermore, the activity of the cell fraction, as shown by CARD-FISH, seemed also very low, which could also hamper viral propagation. Indeed, none of the viral genomes assembled from the four viral metagenomes were detected in the HU cellular metagenome, as shown in Fig. 5B, which indicated that likely these viruses were not replicating inside HU cells. This inference assumes that the presence of viral genomes in the cell metagenomes can be taken, with some limitations (i.e. presence of prophages, virus attached to the cells, etc.), as a proxy of active viral replication inside the cells. All these results reinforce the idea that some of the cells and viruses recovered in HU could have an ancient origin and were released by the dissolution of the salt crystals of the evaporitic rock primary source of the brines.

The question is thus whether the VLPs observed in HU corresponded to infective particles or were just remnants of ancient infection events which were preserved in the brine or were dissolved from the salt crust. In order to investigate this point, we can assume that infective viruses, once given the opportunity to find their hosts, will replicate within the host cytoplasm. Thus, the search for HU viruses in the cellular metagenomes of the Añana system would unveil whether or not they were replicating. As shown in Fig. 5B, some of the viruses from HU could be detected by contig recruitment in cellular metagenomes of HP, pointing to the infectivity of the viruses present in the underground brines.

To further explore this possibility, the reads shared between the HU viral metagenome and the cellular metagenomes from the rest of the pools were extracted and assembled, yielding a total of 10 viral contigs larger than 10 kb (Supplementary Table S3). We will refer to them as 'targeted replicating' HU viral genomes. The viral nature of these contigs was confirmed and their putative hosts determined. As shown in Fig. 6, these 'replicating' viral genomes were abundant in the free viral fraction of HU (except for viral genome h), despite their low abundances in the cell fraction. In contrast, most of these viruses were present in the cellular metagenomes of 'surface' samples, in abundances ranging from 0.01% to 0.12% of metagenome nucleotides (Supplementary Table S2). The presence of HU viruses replicating inside HP cells was especially remarkable, since it indicates that the activation of the corresponding viral particles was quite fast (notice that HU water falls directly on HP). Taken together, these results suggest that viral particles present in the underground water very likely corresponded to infective virus.

Dynamics of the viral communities: evolution in action

Given the hydrogeology of the system, we can assume that viral genomes present both in HU and an additional sample came from the underground brines. Thus, tracking the changes in viral genomes from HU to a surface pool can shed light on the tempo and mode of viral evolution. A possible approach to monitor evolution in action would be to track SNP changes in the core genome of these viral populations. Furthermore, the study of the viral

Table 3. Characteristics of viral genomes larger than 15 kb considered *bona fide* viral genomes including 'cosmopolitan' viral genomes. [Color table can be viewed at wileyonlinelibrary.com]

Viral genomes from MVs		RP ID	Genome size (bp)	% GC	No. ORFs	% viral ORFs	Detected in pond	BEST-HIT by BLASTp and BLASTx	Viral classification by VIRSORTER	Putative host
HU (5)	**HU_scaffold_0***	1	104 118	58.8	166	17.4	HU/HP/SE/EC	Halovirus HVTV-1	Prophage Category 2	*Halorubrum*
	HU_scaffold_1*	2	36 134	58.5	67	43.2	HU/HP/SE/EC	Environmental Halophage eHP-15	Category 2	*Haloferax*
	HU_contigs_100_1**	3	33 065	61.7	52	23.0	HU/HP/SE	Halovirus HCTV-2	Category 2	*Archaea*
	HU_contigs_100_3	4	28 468	58.8	29	24.1	HU/HP	Environmental halophage eHP-25/ *Salinibacter* virus M1EM1	Category 2	*Salinibacter*
	HU_contigs_100_11*	5	16 295	43.4	23	8.7	HU/HP/SE/EC	*Flavobacterium* phage 11b	Category 2	Marine bacteria
HP (16)	HP_contig-100_8	6	24 370	48.8	34	11.7	HP	Halovirus HRTV-8	Category 2	*Halorubrum*/ *Haloferax*
	HP_contigs_100_0	7	33 635	45.5	49	8.1	HU/HP	Environmental halophage eHP-13	Category 2	*known*
	HP_contigs_100_1	8	32 419	50.4	50	12.0	HU/HP	Environmental halophage eHP-32	–	*Halorubrum*
	HP_contigs_100_2	9	29 462	52.5	34	14.7	HU/HP	Halovirus HGTV-1	–	*Archaea*
	HP_contigs_100_3[1]	–	29 161	58.8	54	33.3	Not included	*Haloferax* tailed virus 1	Category 2	*Haloferax*
	HP_contigs_100_5**	10	27 942	58.8	28	25.0	HU/HP	*Salinibacter* virus M1EM-1	Category 2	*Salinibacter*
	HP_contigs_100_6**	11	27 705	59.7	50	16.0	HU/HP/SE	[Halovirus HVTV-1	Category 2	*Archaea*
	HP_contigs_100_10	12	23 357	57.2	23	17.3	HU/HP	Prokaryotic dsDNA virus sp.	–	*Salinibacter*
	HP_contigs_100_11**	13	23 248	43	23	26.0	HU/HP/SE	Prokaryotic dsDNA virus sp.	–	Marine bacteria
	HP_contigs_100_18[2]	–	41 501	62.1	29	13.7	Not included	[Halovirus HCTV-2]	Category 2	*Archaea*
	HP_contigs_100_13**	14	19 768	56.5	23	60.8	HU/HP/SE	Halovirus HHTV-1	Category 2	*Haloarcula*
	HP_contigs_100_21	15	16 367	62.3	16	12.5	HU/HP	Environmental halophage eHP-25	–	*Salinibacter*/ *Salinivenus*
	HP_contigs_100_22	16	15 834	40.2	28	14.2	HP/SE	Halovirus HRTV-8	–	*Archaea*
	HP_contigs_100_24[3]	–	15 393	63.9	20	25.0	Not included	Halovirus HCTV-1	Category 2	*Haloarcula*
	HP_contigs_100_25	17**	15 118	48.3	18	44.4	HU/HP/SE/	*Halorubrum* phage GNf2	–	*Halorubrum*
	HP_contigs_100_26	18	15 056	39.1	30	13.3	HU/HP	Halovirus HGTV-1	–	*Bacteria*
SE (6)	SE_contigs_100_1	19	44 096	65.1	79	12.6	SE/EC	Halovirus HVTV-1	Category 2	*Haloterrigena*/ *Halorubrum*
	SE_contigs_100_2***	20	42 194	59.4	72	12.5	SE	*Haloarcula californiae* icosahedral virus 1	Category 2	*Archaea*
	SE_contigs_100_3	21	41 671	67.6	65	12.3	SE/EC	Halovirus HHTV-1	Category 2	*Haloarcula*/ *Halorubrum*
	SE_contigs_100_4	22	41 124	62.9	76	10.5	HU/HP/SE	Halovirus HRTV-4	Category 2	*Haloarcula*
	SE_contigs_100_6**	23	18 244	64.6	25	68.0	SE/EC	Halovirus HHTV-2	Category 2	*Archaea*
	SE_contigs_100_8**	24	24 309	67.2	22	68.1	HP/SE/EC	*Haloarcula californiae* icosahedral virus 1	Category 2	*Haloarcula*
EC (3)	EC_contigs_100_0	25	51 168	54.9	93	11.8	SE/EC	*Halobacterium* phage phiH	Category 2	*Haloterrigena*/ *Halorubrum*
	EC_contig-100_3*	2526	18 028	35.3	33	9.0	HU/HP/SE/EC	Vibrio phage 1.215.A._10N.222.54.F7	Category 2	*Bacteria*
	EC_contigs_100_4	27	17 915	67.6	24	20.8	SE/EC	*Halobacterium* phage phiH	–	*Halorubrum*

Marked in grey genomes. HU sampling point in red, HP sampling point in blue, SE sampling point in yellow, EC sampling point in green. Additionally, the grey rows correpond with redundant viral genomes.
(1) Redundant genome 99.98ID% with HU_scaffold_1.
(2) Redundant genome 99.97%ID with HU_contig_1.
(3) Redundant genome 99.84% ID with SE_contigs_4.
*Cosmopolitan viruses (genomes present in all viral metagenome).
**Pseudo-cosmopolitan viruses (genomes present in three viral metagenome).
***Autochthonous viruses (genomes detected only in one viral metagenome).
() No. of viral genomes.
Category 2 (quite sure viral genome according with VirSorter tool criteria (Roux *et al*., 2015)).

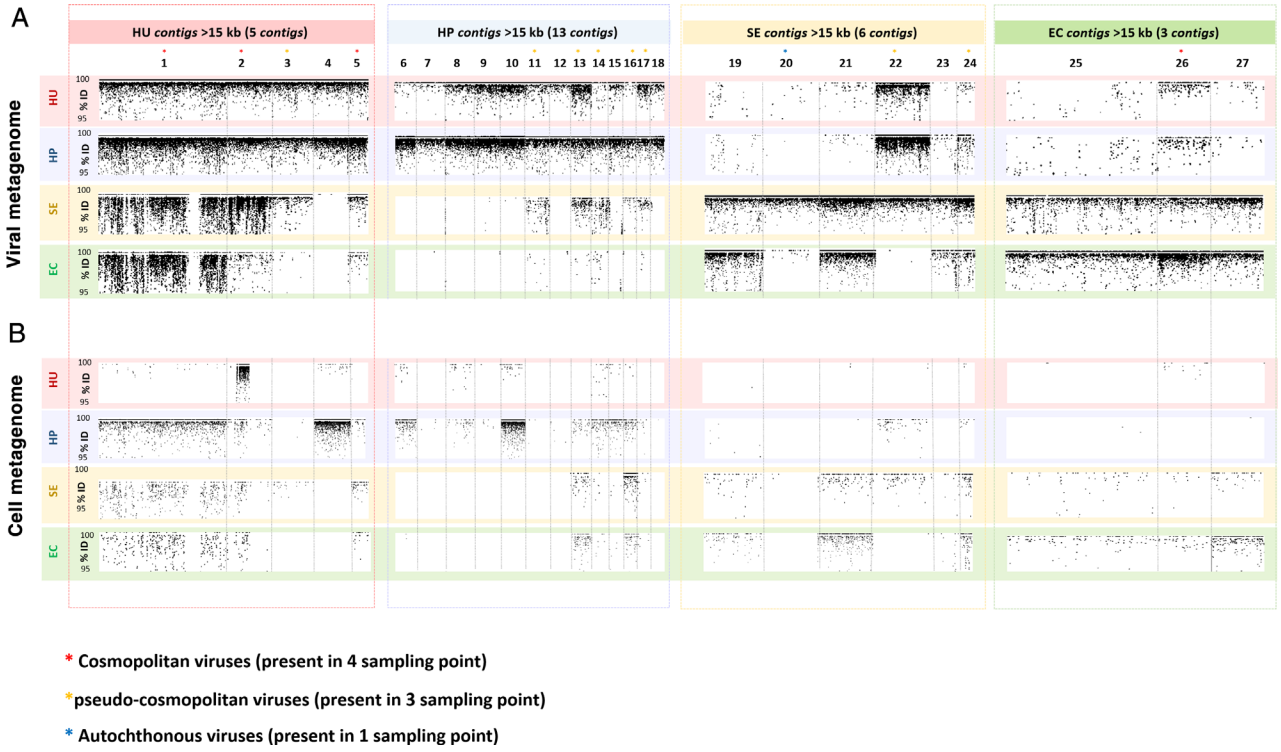

Fig 5. Graphs showing recruitment of Añana viral and cell metagenome reads to viral genomes. Shown along the top are viral contigs recovered by assembly of the (A) four viral metagenomes or (B) four cell metagenomes (along the vertical axis). The description and characteristics of the viral genomes are detailed in Table 3. [Color figure can be viewed at wileyonlinelibrary.com]

contigs also allows for the identification and tracking of genomic islands, which normally derive from lateral gene transfer and are part of the accessory genome (Coleman, 2006; Mizuno *et al*., 2013, 2014; Rodriguez-Valera *et al*., 2014). This last approach, however, cannot be undertaken with the co-assembled MAGs described above since, as discussed elsewhere (Starnawski *et al*., 2017; Wilkins *et al*., 2019), they normally represent the core genome of an environmental population.

A 6 kb genomic island in the viral genome represented by HU_Scaffold_0 was detected by visual inspection of the recruitment plots shown in Fig. 7. This 104-kb contig represented a complete viral genome and was present in the four viral assemblages analysed. The plots showing recruitment of this viral contig against the four viral metagenomes unveiled the presence in the viral population of a metagenomic island that was recovered in HU and HP, but that was below the detection limit in the two pools EC and SE. The abundance of HU_Scaffold_0 decreased with the (temporal) distance from the spring, where it accounted for 0.0087% nt/kb of the viral metagenome (Fig. 7B). Such hypervariable regions (i.e. islands) in viral genomes have been suggested to underlie adaptations to new hosts (Garcia-Doval and van Raaij, 2012; Mizuno *et al*., 2014; Rodriguez-Valera *et al*., 2014). However, in the case of our data, this hypothesis could not be explored since only hypothetical proteins were annotated for the island (Supplementary Fig. S12).

As mentioned above, viral evolution can also be approached by tracking mutations in the viral genomes. For this purpose, first we looked at the general mutation rate by identifying SNPs and then determining whether each change was subjected to positive or purifying selection as estimated using pN/pS values for the protein-coding genes. The pN/pS metric is similar to dN/dS (rate of nonsynonymous to synonymous mutations) and is valid for metagenomics analyses since it does not rely on the identification of allelic variants. These analyses were carried out both with the viral genomes directly assembled from the viral metagenomes and from the HU viruses assembled from cellular metagenomes. For each viral genome, the genes considered and the calculated % SNP and pN and pS values are shown in Supplementary datasets 4 and 5.

Overall, the proportion of SNPs (percentage of polymorphisms) in the viral genomes that could be assembled directly from the viral metagenomic reads was significantly lower than that of the 'replicating' viruses (2.27 vs. 8.01%; p-value 2.22×10^{-16} using the Mann–Whitney–Wilcox test). This is likely the reason why, as explained above, these 'replicating' viruses (assembled

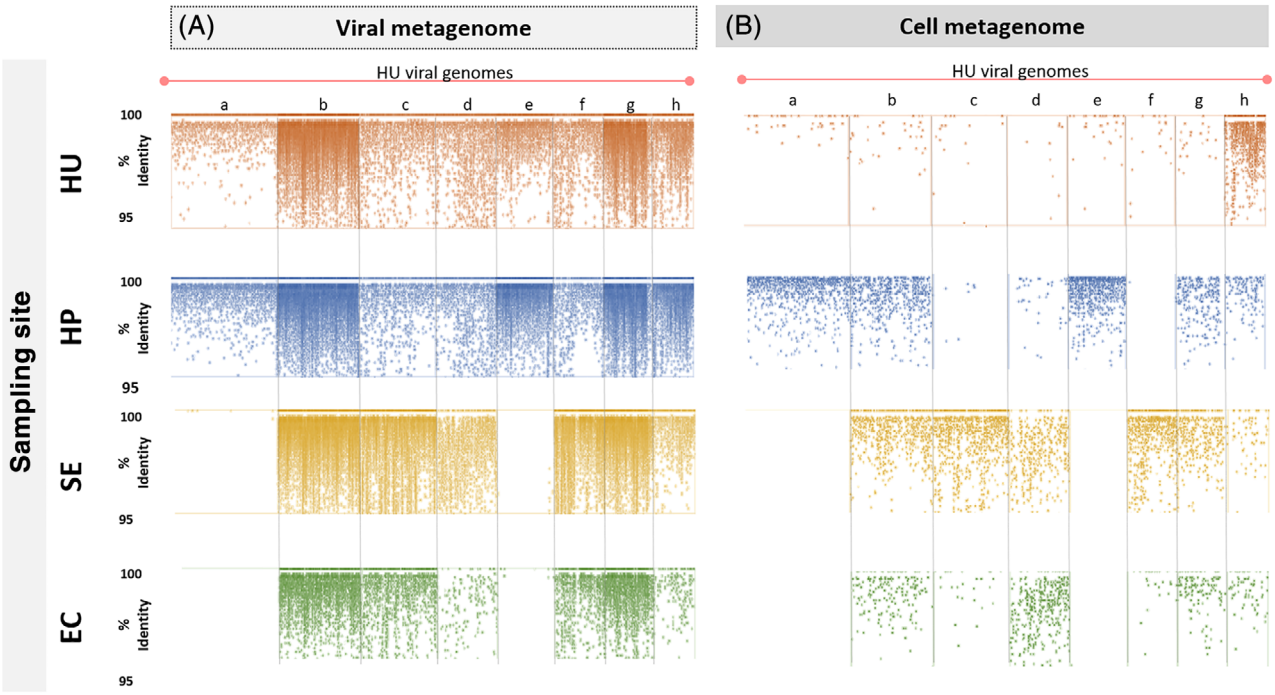

Fig 6. (A) Plot of viral metagenome read recruitment to viral genomes from HU.
B. Plot of cell metagenome read recruitment to viral genomes from HU. Taxonomic affiliation based on BLASTx best-hit (a) EMM1 *Salinibacter* sp. virus, (b) *Halorubrum* phage, (c) *Rhodobacter siphingovirus* phage, (d) *Halorubrum* phage, (e) *Haloarcula* phage, (f) *Pseudomonas* phage, (g) *Halorubrum* phage, (h) *Halobacterium* phage. Supplementary Table S3 shows the viral genome details. [Color figure can be viewed at wileyonlinelibrary.com]

from cell metagenomic reads), could not be directly assembled from the viral metagenomes in spite of their high abundances in the pools (see Supplementary Table S2).

In order to check whether the transition from the underground to the surface exposed HP, SE or EC was accompanied by a change in SNPs in the viral genomes, all pairwise changes in SNP frequencies were calculated for each viral genome present in HU and in another sample (see Fig. 8). Whether the calculated frequency changes corresponded significantly to an increase or a decrease of SNP was estimated with the Wilcoxon test (Rubino *et al*., 2017) (see Supplementary dataset 5). Among all the 31 comparisons, in 21 cases (marked with black asterisks in Fig. 8) corresponding to 11 different viral genomes, there was a significant increase of SNPs in the viral genomes when moving from the underground to the surface. Only for viruses contig_1_MG and contig_6_MG (corresponding to *Halorubrum* phages, see Supplementary Table S3) were decreased in SNPs detected in any of the pools compared to HU. It is important to note that differences in SNPs, either among viral genomes or for a given genome in different pools, could not be explained only by differences in abundance measured as the normalized % of nucleotides recruited by the genome size in the extracellular viral metagenome, as shown in Supplementary Table 2 and Supplementary dataset 2. The correlation between SNP and abundance improved when the 'effective' viral population size was calculated for each viral genome taking into account the number of VLPs in each pool (Fig. 8B and C), although still there were some minority viruses displaying high SNP numbers [contig_2_MG (c in Fig. 8A) in SE, contig_3_MG (d in Fig. 8A) in HU, contig_4_MG (e in Fig. 8A) in HP, contig_5_MG (f in Fig. 8A) in SE]. These results, similar to what was found when analysing marine viruses (Coutinho *et al*., 2019a; Coutinho *et al*., 2019) or lake virophages, indicate that larger populations did not necessarily display more SNPs, as one could have expected. In any case, it may not be correct to consider the one-point abundance in the extracellular viral metagenome as a perfect metric of viral population size since it may be affected by other factors (stability of the viral particles, changes in the host population, lag between replication and viral production, among others).

To further explore the evolution of the different viral genomes, pN and pS values were calculated for all their genes in each of the pools (as described in the Material and methods section). In order to avoid arbitrary elimination of zero values, all calculated pN and pS values were

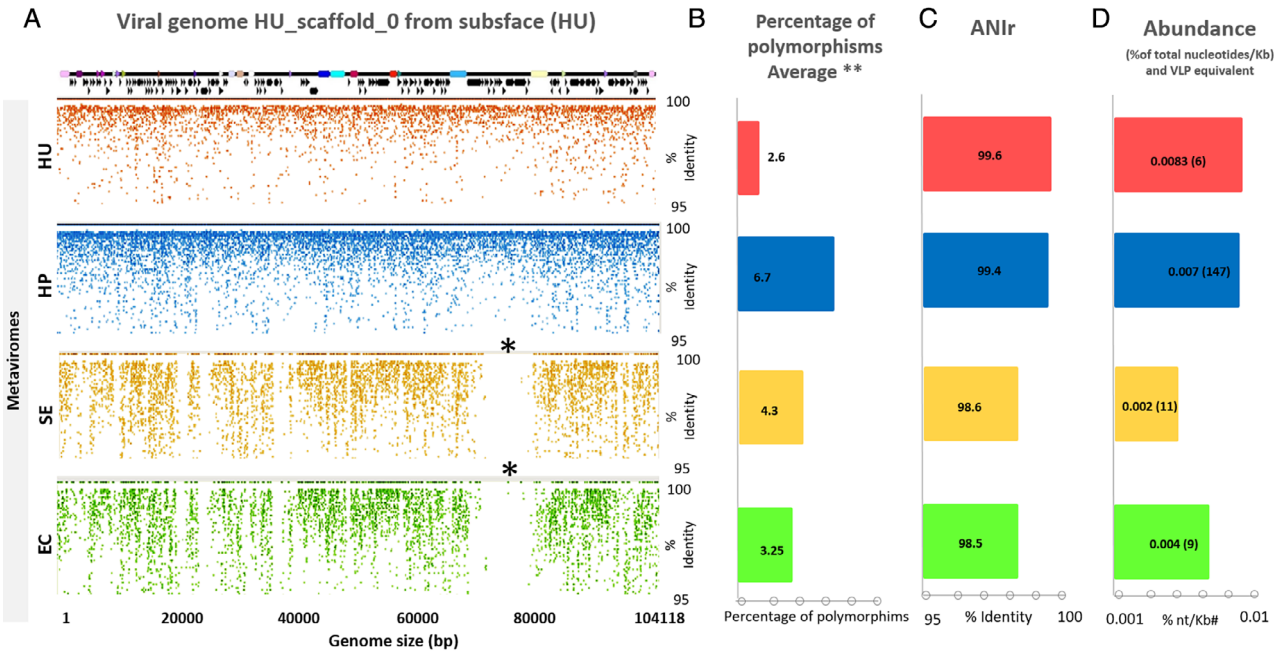

Fig 7. A. Recruitment plot of cosmopolitan HU_scaffold_0 in all viral metagenomes. The asterisk (*) indicates the hypervariablility island detected in this genome.
B. Percentage of polymorphism normalized by sequencing depth of PU_scaffold_0 in Añana's viral metagenomes (**) only considering shared regions.
C. Average nucleotide identity of reads (ANIr) of viral genome HU_scaffold_0 in Añana's viral metagenomes.
D. HU_scaffold_0 abundances in all viral metagenomes; only hits over 70% of coverage and 95% of identity were considered. # % of total nucleotides/viral genome size (kb). Parentheses show VLP equivalents. [Color figure can be viewed at wileyonlinelibrary.com]

included in the analysis and, in cases where pS = 0, the proportion and distribution of pN and pS were analysed for each genome. Detailed results are provided in Supplementary datasets 4 and summarized in Supplementary Table S4 and Supplementary dataset 6. A common trend was not found for the 16 viral genomes analysed nor when the genomes were grouped by viruses recovered from the extracellular viral metagenomes (nine viral genomes) or 'targeted replicating' viruses (seven viral genomes). Indeed, for both groups we detected viral genomes undergoing strong purifying selection in all the samples, as well as others, experiencing diversifying selection.

These differences can be exemplified with viral genomes contig_1_MG and contig_6_MG (b and g in Fig. 8A), both tentatively assigned to *Halorubrum* hosts and assembled from the cellular metagenomes (see above). Both viral genomes displayed extremely high SNP values, up to 20%, and most SNPs were evenly distributed along the genomes. However, despite both viral genomes having similarly high values of SNPs and similar abundances in all the pools, they likely were exposed to different evolutionary constraints since, while viral genome contig_1_MG (b in Fig. 8A) was undergoing strong purifying selection in all pools except EC (pN/pS values below 0.6 for most of its genes, Supplementary dataset 5 and 6), viral genome contig_6_MG (g in Fig. 8A) was under positive selection, with pN/pN values above 1 in all pools.

Overall, the two types of evolution detected in the viral genomes (horizontal and vertical, islands and mutations respectively) were driving evolution of the viral assemblages when moving from the spring to the open-air reservoirs. However, as exemplified above, it seems that different viruses followed different evolutionary trajectories. This was not unexpected since the pressure exerted by the change in the environmental conditions was different for different members of the microbial community who host different viruses.

Concluding remarks

Overall, our results suggest that microbial and viral assemblages in the underground brines experienced different changes when exposed to the open air. A plausible explanation of the important differences between HU and the rest of the brines was that part of the microbial population in the underground brines was inactive, probably dead or dormant, and had been released from the rock dissolution. Part of the viral community of HU, that we show to be infective, could also have been ancient but viable. Previous studies showed that the fast gene

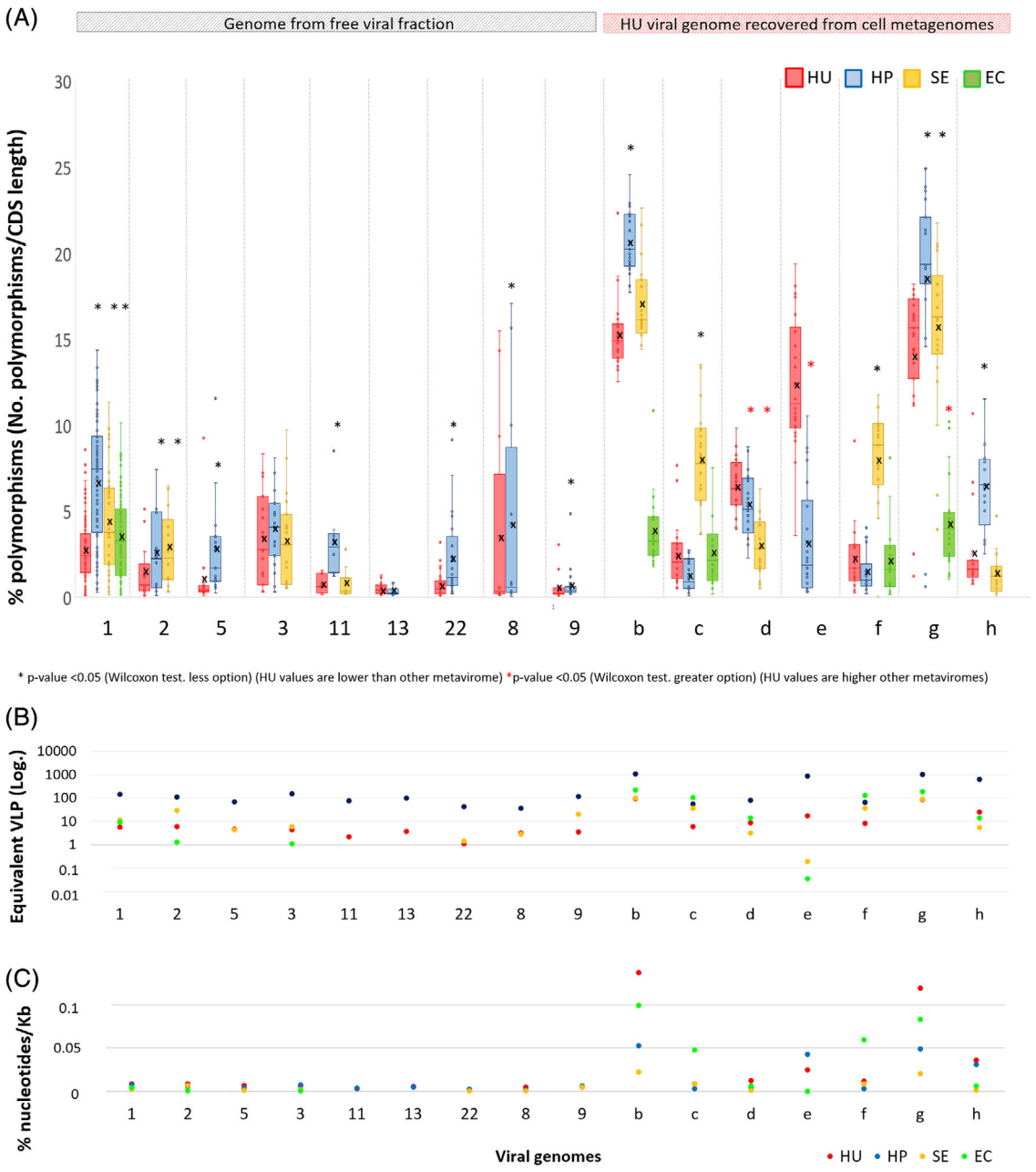

Fig 8. A. Box plots showing the percentage of polymorphism (No. Polymorphisms/CDS length (nt)) of analysed bonafide viral contigs in all Añana viral metagenomes. Median percentage of polymorphic sites of each viral genome is indicated as (**X**) in the boxplot. Additionally, black asterisks (*) indicate significant differences between HU and the rest of the viral metagenomes using the Wilcoxon test less option. Red asterisks (*) indicate significant differences between HU and the rest of the viral metagenomes using the Wilcoxon test high option. Only genes present in all viral metagenomes were included in the analyses. Table 3 and Supplementary Table S3 show the viral genomes details.
B. Equivalent VLP (Log).
C. Viral abundances in all Añana viral metagenomes. [Color figure can be viewed at wileyonlinelibrary.com]

turnover can provide ecological advantages to viral populations, such as the ability to carry on infecting hosts (Boucher *et al*., 2011; Smillie *et al*., 2011; Cordero and Polz, 2014; Mizuno *et al*., 2014; Rodriguez-Valera *et al*., 2014). 'Surface' microbiomes may be directly influenced by the presence of higher organic substrate availability due to an increased primary production (especially due to *Dunaliella* sp.) under sunlight irradiation. The increase of substrate availability may promote the activation of the metabolism of the cells and their growth, and therefore the activation of the infectivity of the accompanying viruses. The most puzzling aspect of all these

dynamics (both viral and microbial) was that it seems as though they occurred in a short time frame (i.e. hours), which is especially remarkable for the changes detected between HU and HP (Fig. 1). This is an open question of this work that deserves future studies.

Materials and methods

Study site and sample collection

Four brine samples were collected from Añana Salterns located in Araba/Álava, Spain (42° 48′ 7.22″; −2° 59′ 10.76″) in July 2015 (Fig. 1). Samples from one spring outlet (La Hontana, HU), two pools (El Cautivo EC and Santa Engracia SE) and one small puddle directly fed by the spring (La Hontana pool, HP) were collected over less than 1 h. Salinity was measured *in situ* with a hand refractometer (Eclipse). The average air temperature was 23 °C. Ionic composition of the samples was analysed by atomic absorption spectrometry at the Technical Services of the University of Alicante.

Cell and virus quantification by DAPI, FISH, CARD-FISH and Sybr gold

Fresh samples were fixed *in situ* with formaldehyde (F.A. Sigma-Aldrich) at 7% final concentration for 16 h at 4 °C. After fixation, 8.8 ml of PBS (1×) was added and the mixture was filtered through a 0.22 μm pore size with 16 mm diameter filters (Isopore GTTP, Millipore). Filters were stored at −20 °C until use. Double-staining 4′,6-diamidino-2-phenylindole (DAPI; 1 μg ml^{-1}) and fluorescence in situ hybridization (FISH) with EUB338 and Arc915 probes were performed as described by Anton *et al*. (1999). CARD-FISH was carried out with the probes Eub338-I-III and Arc915 as described by Gomariz *et al*. (2015).

For viral quantification, 100 μl of sample was fixed with formaldehyde at 4% final concentration (30 min, 4 °C). For storage, the sample was diluted in 887 μl of PBS (1×). Fluorescent Sybr-Gold strain was executed as in Boujelben *et al*. (2012a).

DNA extraction from cellular and viral fractions

Collected salt waters, 1.8 L from HP and HU samples and 3.6 L from EC and SE samples, were centrifuged at 30 000*g*, 40 min and 20 °C (Avanti J-30I, Beckman with JA 14 rotor) to separate viral and cell fractions. Supernatants were filtered (0.22 μm pore size) with the standard method of tangential flow filtration through Vivaflow 200-PES system (initial volume to 20 ml) with 30 kDa and reconcentrated to 500 μl by 100 kDa Amicon Ultra centrifugal filters (Millipore). Cell DNA was extracted from the pellet using the UltraClean® Soil DNA (isolation kit, Mo Bio lab) and viral DNA was extracted from the concentrated supernatants using the QIAamp® ultrasens epicentre® virus kit implemented with Turbo DNAse (Ambion) to remove dissolved DNA prior viral DNA extraction. Viral DNA quality and quantity were measured by QUIBIT® 2.0, using the highly sensitive fluorochrome (2–100 ng) kit. Additionally, the detection of cell contamination in viral DNA was performed by PCR using DGGE archaeal and bacterial primers; in the four viral DNAs the absence of cell contamination was confirmed.

Denaturing gradient gel electrophoresis analysis of prokaryotic diversity

A sample taken in July 2011 from the Santa Engracia pool was used for a preliminary characterization of the microbial community by means of DGGE. The extracted genomic DNA was used for PCR amplifications of bacterial and archaeal 16S rRNA genes by using the specific forward primers 27F and 21F respectively, and the universal reverse primer 1492R as described in Mutlu *et al*. (2008).. Then, the pooled products of 16S rRNA genes of Santa Engracia samples were PCR amplified for DGGE analysis with the following primer sets: 341F-GC and 907R for Bacteria; 344F-GC and 907R for Archaea (Muyzer *et al*., 1993; Schäfer *et al*., 2001). The PCR program for Bacteria and Archaea was used as described previously (Mutlu *et al*., 2008). PCR products were then purified with GeneJet PCR Purification Kit (Fermentas, Thermo Scientific) and 300 ng of Archaea and Bacteria samples were loaded to the DGGE gel. DGGE was performed with the DCode System (Bio-Rad, Hercules, CA). Each PCR product was loaded onto a 6% (wt./vol.) polyacrylamide gel (acrylamide:bis-acrylamide gel stock solution 37.5:1; Bio-Rad), with a linear gradient of denaturing agents from 40% to 70% (100% denaturing agents consists of 7 M urea and 40% formamide) in 1× TAE buffer (40 mM Tris, pH 8.0; 20 mM acetic acid; 1 mM EDTA). The gel was run at a constant voltage of 60 V for 18 h at 60 °C in 1× TAE running buffer. After electrophoresis the DGGE gel was stained with SYBR Green (Invitrogen) and photographed with a Typhoon 9410 (Amersham Biosciences) imaging system. For sequencing and identification of DGGE fragments, selected bands were excised with sterile razor blades from the DGGE gel and soaked overnight in 20 μl of sterile MilliQ water. Two microliters of each band were then reamplified, with the same primer set (without the GC clamp), and the resulting PCR products were then purified with GeneJet PCR Purification Kit (Fermentas,

Thermo Scientific) and 50 ng of samples were Sanger sequenced using the 907R primer.

Sequencing, read trimming and cleaning, read joining, community coverage and assembly

Sequencing of cell and viral DNAs was performed using an Illumina Mi-seq Nextera XT 300 × 2 bp paired-end run. Paired-end reads were joined using Fastq-join from the eatools suite (Aronesty, 2011) for cell metagenomes, and fq2fa from IDBA 1.1.1 assembler for the viral metagenomes (Peng *et al*., 2012). Reads were quality assessed and trimmed using PRINSEQ software (Schmieder and Edwards, 2011). Reads shorter than 50 bp and with a quality lower than 20 were discarded. Additionally, the viral metagenome reads were cleaned of contaminant sequences (Salter *et al*., 2014) (including sequences from human-associated bacteria, eukaryotes and humans); contaminant detection was performed using Diamond BLASTx (*e*-value 0.0001 (Buchfink *et al*., 2014) and Kaiju (query coverage >70%) (Menzel *et al*., 2016) both against the NCBI nr database (released on 01/01/2020). The most abundant sequence contaminant corresponded to *Propionibacterium acnes*, one of the most common Illumina kit contaminant sequences (Salter *et al*., 2014). Due to the high percentage of contamination from *Propionibacterium* spp. (from 0.1% to 5% of the total reads), virome data were cleaned using two steps. First, all virome reads were recruited separately against the *Propionibacterium acnes* NC008065.1 genome and matching reads were removed (query coverage >70% ID >90%). Second, as previously mentioned, contaminant sequences were also detected by DIAMOND (Buchfink *et al*., 2014) and Kaiju (Menzel *et al*., 2016). Finally, all contaminant sequences were extracted (from 1% to 19% of total reads) using FastA.filter.pl from the enveomics package (Rodriguez-R and Konstantinidis, 2016) to get the clean viromes used in the study. The Nonpareil tool (Rodriguez-R and Konstantinidis, 2014) was used to estimate the coverage of the community in each metagenome dataset with default parameters. *De novo* assemblies of trimmed reads were generated using the IDBA 1.1.1 assembler (Peng *et al*., 2012) with the '-pre_correction' option. In order to recover additional complete viral genomes from the assembly sequences, contigs were extended by the 'map to reference' option (minimum match identity 99%) of the Genious 6.1.8 platform (Kearse *et al*., 2012) according to the recommendation of Villamor *et al*. (2017)..

Classification of 'bona fide' viral sequences in viral metagenomes

Viral contigs larger than 15 kb were checked to guarantee their quality (they were considered viral when they harboured viral genes and did not align completely with a cellular genome). In this work, we have considered as a *bona fide* viral genomes those sequences which achieved the following requirements: the sequences should be included in one of the VirSorter categories (from 1 to 6) (Roux *et al*., 2015); they should have a viral hit by DIAMOND BLASTx against the NCBI nr database (*e*-value <0.00001); at least 8% of the predicted genes in a given contig should be annotated as viral genes by DIAMOND BLASTp; lastly, BLASTn searches were performed to discard any contigs that likely originated from cells. The redundant viral sequences were detected by Cd-hit (cdhit-est -c 0.9 –n 8) (Li and Godzik, 2006) and omitted by FastA.filter.pl (Rodriguez-R and Konstantinidis, 2016).

The *in silico* virus–host prediction was performed using multiple tools based on BLAST identification of virus–host homologous ORFs (Coutinho *et al*., 2017), CRISPR systems, tRNA identification by GtRNAdb (Chan and Lowe, 2009) and oligonucleotide frequencies (Edwards *et al*., 2016). In the set of platforms used for the virus–host pair identification, IMG/V3 (Roux *et al*., 2020) and PHISDetector (Zhang *et al*., 2019) were included. Additionally, virus–host pairs were manually assigned for each *bona fide* viral genome by checking gene similarities and genome synteny with their putative host using the RefSeq non-redundant database (Sayers *et al*., 2016; Coutinho *et al*., 2017).

Cell and virus metagenome annotation and comparison

Functional annotation of predicted genes from cell and virus assembled metagenomes was done by JGI (IMG/M ER: http://img.jgi.doe.gov/mer) (Huntemann *et al*., 2014) and DIAMOND BLASTp (Buchfink *et al*., 2014). For virus metagenomes the NR NCBI database was used (December 2019 update); the Uniprot database (Bateman *et al*., 2015) was used for cell metagenomes. Raw cell and viral metagenomes 'all versus all' comparisons were performed by BLASTn stand-alone using the same parameters described below.

Recovery of draft genomes from cell metagenomes

MaxBin software (Wu *et al*., 2014) was used to bin draft genomes from cell metagenomes. Cell contigs larger than 1000 bp were grouped in bins based on tetranucleotide frequencies and contig sequencing depth

levels. The quality of the recovered bins (completeness and contamination) was analysed with CheckM (Parks *et al*., 2015) and the HMM.essential.rb script from enveomics package (Rodriguez-R and Konstantinidis, 2016), which identified lineage-specific marker genes in each bin. Potential contamination of MAGs was removed following the tutorial available by Anvio's tools v5.5 (Eren *et al*., 2015). Predicted genes from MAGs were annotated against both UniProt databases (SwissProt and TrEMBL) using the DIAMOND BLASTp tool (Buchfink *et al*., 2014). Viral contigs larger than 20 000 bp were extracted and checked by BLASTn against the NCBI Reference Sequence (RefSeq) database (http://www.ncbi.nlm.nih.gov/RefSeq) to be confident that they were viral contigs. MAG quality and taxonomic affiliation were additionally checked by the MiGA platform (Rodriguez-R *et al*., 2018) and Genome taxonomy database (Chaumeil *et al*., 2019).

Abundance of recovered draft viral and cell genomes

The abundances of recovered draft genomes were assessed by fragment recruitment using the trimmed reads of each sample. The reads of cell and viral metagenomes were mapped against the reference microbial (viral or cell) genomes using stand-alone BLASTn with a cutoff of 70% query coverage, *e*-value $\geq 10^{-1}$ and the 'best hit' option. Next, cell and viral abundances were calculated considering only hits with identities ≥98% (cell) or 95% (virus). Fragment recruitment data were plotted by using the enveomics.R package in the R statistical tool (Rodriguez-R and Konstantinidis, 2016). MAG diversities were calculated by ANIr (average nucleotide identity of mapped reads) as previously described (Meziti *et al*., 2019) and SNPs values were calculated using map to reference (≥95% identity, minimum coverage 20× and 0.25 of polymorphism frequency) in the Genious 6.1.8 platform (Kearse *et al*., 2012). The mutation rate was calculated as the number of SNPs/genome size (Mbp).

Tree reconstructions based on 16S, 23S and 5S rRNA genes

Gene fragments from metagenome trimmed reads that contained the 16S rRNA gene were extracted using Parallel-META v 2.4 software. Sequences were grouped in operational taxonomic units (OTUs) at 98.7% similarity using UCLUST (Edgar and Bateman, 2010) implemented in QIIME script *pick_closed_reference_otus.py* (Caporaso *et al*., 2010). For phylogenetic purposes, the longest sequence from each OTU was aligned using the SINA tool (SILVA Incremental Aligner (Pruesse *et al*., 2007) and were added to the reference database SILVA REF123 using the parsimony method implemented in the ARB v6.0.2 software (Ludwig *et al*., 2004). The OTUs were then clustered into OPUs (Mora-Ruiz *et al*., 2016).

The almost-complete 16S rRNA, 23S rRNA and 5S rRNA genes from assembled contigs were extracted using RNAmmer 1.2 Server (Lagesen *et al*., 2007). The 16S and 23S rRNA genes were added by parsimony to the SILVA SSU and LSU REF 123 databases respectively. Sequences were used to reconstruct *de novo* trees using the neighbour-joining algorithm and Jukes-Cantor correction with the close relative sequences selected from the SILVA database.

The obtained 5S rRNA genes were compared using the online NCBI BLASTn program and 5SrRNAdb (http://combio.pl/rrna/) to select the closest relative sequences. Sequences were aligned using the MUSCLE v3.8.31 (Edgar, 2004) program and phylogenetic reconstruction was performed using the algorithm neighbour-joining in ARB v6.0.2 (Ludwig *et al*., 2004).

Viral genome diversity analyses

To calculate the percentage of polymorphisms of viral genomes, raw reads from each Añana viral metagenome were mapped against the selected viral genomes using the sensitive local mode of Bowtie2 (Langmead and Salzberg, 2012). Then, in order to obtain counts of synonymous and nonsynonymous mutations in each ORF, the output files were analysed by DiversiTools (http://josephhughes.github.io/DiversiTools/). As recommended by Coutinho *et al*. (2019a) and Coutinho *et al*. (2019), to estimate the percentage of polymorphic sites and pN/pS ratios (Schloissnig *et al*., 2013), only the reads with coverage equal or higher than 5 were considered, and codon mutations must have been detected at least four times in at least 1% of mapped reads. The ORFs included in the analyses must be present in the four Añana viral metagenomes. Only viral genomes with at least eight ORFs that passed the mentioned criteria were included in the diversity analyses. The resulting mutation data were used to estimate the percentage of polymorphic sites and pN/pS ratios, which were calculated as previously described by Schloissnig *et al*. (2013). The pN and pS values equal to zero were included in the study, consequently, we obtained $pN/pS = 0$ (maximum purification) and $pN/pS = \infty$ (diversification), and for this reason the data were transformed by RStudio (http://www.rstudio.com) for the subsequent analyses. The sequencing depth of each viral genome in each viral metagenome was calculated using the BLASTn output filter by coverage (equal or >70%) and identity (equal or >95%) using

BlastTab.seqdetph.pl from enveomic box (Rodriguez-R and Konstantinidis, 2016).

Statistical analysis of viral microdiversity analyses

SNPs and pN/pS ratios were compared between pairs of samples using Wilcoxon rank-sum tests implemented in the coin package (Hothorn *et al*., 2006) of RStudio (http://www.rstudio.com). First, the significant differences (*p*-value equal or >0.05) between samples were tested using the 'two side' option, which was then followed by the 'less and greater' option to ascertain the direction of the change.

Sequence data

Raw sequence reads from each metagenome are publicly available at the repository of Discovery Environment (Cyverse) through the following links: https://de.cyverse.org/dl/d/25B33A44-98A0-4ABA-B706-B71C1FFEBEDB/Ramos-Barbero.zip for cell metagenomes and https://data.cyverse.org/dav-anon/iplant/home/dolores/Ramos_Barbero_et_al_2021_part_1/1_Ramos_et_al_2021_Clean_metaviromes_sequences.rar for viral metagenome. Additionally, the *bona fide* viral genomes and MAGs are available in the link https://data.cyverse.org/dav-anon/iplant/home/dolores/Ramos_Barbero_et_al_2021_part_2/7_Ramos_et_al_2021_bonafide_viral_and_MAGs_sequences.rar.

Acknowledgements

We would like to thank Añana Salt Valley Foundation, and Andoni Erkiaga Agirre, its director at the time of sampling, for their kind help. Thanks to Leire Arana, Edorta Loma and Kika Colom for their help with sampling and to Eduardo González-Pastor for telling us about the Añana Salt Valley. We thank Heather Maughan for the professional English editing and the critical reading of the manuscript and Esther Rubio-Portillo for her help with statistical analyses. This work was funded by the Spanish Ministry of Science, Innovation and Universities grant MICROMATES (PGC2018-096956-B-C41 and C44, to J.A./F.S. and R.R.-M.), which was also supported with European Regional Development Fund (FEDER) funds, and by the Generalitat Valenciana grant PROMETEO/2017/129.

REFERENCES

Amann, R., and Fuchs, B.M. (2008) Single-cell identification in microbial communities by improved fluorescence in situ hybridization techniques. *Nat Rev Microbiol* **6**: 339–348.

Anantharaman, K., Brown, C.T., Hug, L.A., Sharon, I., Castelle, C.J., Probst, A.J., *et al*. (2016) Thousands of microbial genomes shed light on interconnected biogeochemical processes in an aquifer system. *Nat Commun* **7**: 1–11.

Antón, J., Oren, A., Benlloch, S., Rodrı! Guez-Valera, F., Amann, R., and Rosselló-Mora, R. (2002) Salinibacter ruber gen. nov., sp. nov., a novel, extremely halophilic member of the Bacteria from saltern crystallizer ponds. *Int J Syst Evol Microbiol* **52**: 485–491.

Aronesty, E. (2011) *Command-line tools for processing biological sequencing data. ea-utils*. Expression Analysis, Durham, NC.

Baati, H., Guermazi, S., Gharsallah, N., Sghir, A., and Ammar, E. (2010) Microbial community of salt crystals processed from Mediterranean seawater based on 16S rRNA analysis. *Can J Microbiol* **56**: 44–51.

Baker, B.J., Comolli, L.R., Dick, G.J., Hauser, L.J., Hyatt, D., Dill, B.D., *et al*. (2010) Enigmatic, ultrasmall, uncultivated Archaea. *Proc Natl Acad Sci U S A* **107**: 8806–8811.

Bateman, A., Martin, M.J., O'Donovan, C., Magrane, M., Apweiler, R., Alpi, E., *et al*. (2015) UniProt: a hub for protein information. *Nucleic Acids Res* **43**: D204–D212.

Boucher, Y., Cordero, O.X., Takemura, A., Hunt, D.E., Schliep, K., Bapteste, E., *et al*. (2011) Local mobile gene pools rapidly cross species boundaries to create endemicity within global Vibrio cholerae populations. *MBio* **2**: e00335-10.

Boujelben, I., Gomariz, M., Martínez-García, M., Santos, F., Peña, A., López, C., *et al*. (2012a) Spatial and seasonal prokaryotic community dynamics in ponds of increasing salinity of Sfax solar saltern in Tunisia. *Antonie Van Leeuwenhoek* **101**: 845–857.

Boujelben, I., Yarza, P., Almansa, C., Villamor, J., Maalej, S., Antón, J., and Santos, F. (2012b) Virioplankton community structure in Tunisian solar Salterns. *Appl Environ Microbiol* **78**: 7429–7437.

Buchfink, B., Xie, C., and Huson, D.H. (2014) Fast and sensitive protein alignment using DIAMOND. *Nat Methods* **12**: 59–60.

Caporaso, J.G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F.D., Costello, E.K., *et al*. (2010) QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* **7**: 335–336.

Chan, P.P., and Lowe, T.M. (2009) GtRNAdb: a database of transfer RNA genes detected in genomic sequence. *Nucleic Acids Res* **37**: 93–97.

Chaumeil, P.-A., Mussig, A.J., Hugenholtz, P., and Parks, D.H. (2019) GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. *Bioinformatics* **36**: 1925–1927.

Colangelo-Lillis, J., Wing, B.A., and Whyte, L.G. (2016) Low viral predation pressure in cold hypersaline Arctic sediments and limits on lytic replication. *Environ Microbiol Rep* **8**: 250–260.

Coleman, M.L. (2006) Genomic Islands and the ecology and evolution of Prochlorococcus. *Science* **311**: 1768–1770.

Cordero, O.X., and Polz, M.F. (2014) Explaining microbial genomic diversity in light of evolutionary ecology. *Nat Rev Microbiol* **12**: 263–273.

Coutinho, F.H., Edwards, R.A., and Rodríguez-Valera, F. (2019a) Charting the diversity of uncultured viruses of Archaea and bacteria. *BMC Biol* **17**: 1–16.

Coutinho, F.H., Rosselli, R., and Rodríguez-Valera, F. (2019) Trends of Microdiversity Reveal Depth-Dependent

Evolutionary Strategies of Viruses in the Mediterranean. *mSystem* **4**: e00554–19.

Coutinho, F.H., Silveira, C.B., Gregoracci, G.B., Thompson, C.C., Edwards, R.A., Brussaard, C.P.D., *et al*. (2017) Marine viruses discovered via metagenomics shed light on viral strategies throughout the oceans. *Nat Commun* **8**: 1–12.

Di Meglio, L., Santos, F., Gomariz, M., Almansa, C., López, C., Antón, J., and Nercessian, D. (2016) Seasonal dynamics of extremely halophilic microbial communities in three Argentinian salterns. *FEMS Microbiol Ecol* **92**: fiw184.

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

Edgar, R.C., and Bateman, A. (2010) Search and clustering orders of magnitude faster than BLAST. *Bioinformatices* **26**: 2460–2461.

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

Enault, F., Briet, A., Bouteille, L., Roux, S., Sullivan, M.B., and Petit, M.A. (2017) Phages rarely encode antibiotic resistance genes: a cautionary tale for virome analyses. *ISME J* **11**: 237–247.

Eren, A.M., Esen, O.C., Quince, C., Vineis, J.H., Morrison, H.G., Sogin, M.L., *et al.* (2015) Anvi'o: An advanced analysis and visualization platformfor 'omics data. *PeerJ* **2015**: e1319.

Frankovic, A., Eguiluz, L., and Martínez-Torres, L.M. (2016) Geodynamic evolution of the Salinas de Añana diapir in the Basque-Cantabrian Basin, Western Pyrenees. *J Struct Geol* **83**: 13–27.

Garcia-Doval, C., and van Raaij, M.J. (2012) Structure of the receptor-binding carboxy-terminal domain of bacteriophage T7 tail fibers. *Proc Natl Acad Sci U S A* **109**: 9390–9395.

Ghai, R., Martin-Cuadrado, A.-B., Molto, A.G., Heredia, I.G., Cabrera, R., Martin, J., *et al*. (2010) Metagenome of the Mediterranean deep chlorophyll maximum studied by direct and fosmid library 454 pyrosequencing. *ISME J* **4**: 1154–1166.

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

Guixa-Boixareu, N., Calderón-Paz, J., Heldal, M., Bratbak, G., and Pedrós-Alió, C. (1996) Viral lysis and bacterivory as prokaryotic loss factors along a salinity gradient. *Aquat Microb Ecol* **11**: 215–227.

Hallam, S.J., Putnam, N., Preston, C.M., Detter, J.C., Rokhsar, D., Richardson, P.M., *et al*. (2004) Reverse methanogenesis: testing the hypothesis with environmental genomics. *Science* **305**: 1457–1462.

Hayes, S., Mahony, J., Nauta, A., van Sinderen, D., Hayes, S., Mahony, J., *et al*. (2017) Metagenomic approaches to assess bacteriophages in various environmental niches. *Viruses* **9**: 127.

Hoshino, T., Yilmaz, L.S., Noguera, D.R., Daims, H., and Wagner, M. (2008) Quantification of target molecules needed to detect microorganisms by fluorescence in situ hybridization (FISH) and catalyzed reporter deposition-FISH. *Appl Environ Microbiol* **74**: 5068–5077.

Hothorn, T., Hornik, K., Van De Wiel, M.A., and Zeileis, A. (2006) A lego system for conditional inference. *Am Stat* **60**: 257–263.

Huntemann, M., Mavrommatis, K., Ivanova, N., Mikhailova, N., Ovchinnikova, G., Schaumberg, A., *et al*. (2014) The JGI pipeline for annotation of microbial genomes and metagenomes the JGI pipeline for annotation of microbial genomes and metagenomes. *Calif Digit Libr*: 2–4.

Iranzo, J., Krupovic, M., and Koonin, E.V. (2017) A network perspective on the virus world. *Commun Integr Biol* **10**: e1296614.

Iranzo, J., Wolf, Y.I., Koonin, E.V., and Sela, I. (2019) Gene gain and loss push prokaryotes beyond the homologous recombination barrier and accelerate genome sequence divergence. *Nat Commun* **10**: 1–10.

Iribar, V., and Ábalos, B. (2011) The geochemical and isotopic record of evaporite recycling in spas and salterns of the Basque Cantabrian basin, Spain. *Appl Geochem* **26**: 1315–1329.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., *et al*. (2012) Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28**: 1647–1649.

Keerthi, S., Koduru, U.D., Nittala, S.S., and Parine, N.R. (2018) The heterotrophic eubacterial and archaeal coinhabitants of the halophilic *Dunaliella salina* in solar salterns fed by bay of Bengal along south eastern coast of India. *Saudi J Biol Sci* **25**: 1411–1419.

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

Lagesen, K., Hallin, P., Rødland, A., Staerfeldt, H.-H., Rognes, T., and Ussery, D.W. (2007) RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res* **35**: 3100–3108.

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

Li, S.-J., Hua, Z.-S., Huang, L.-N., Li, J., Shi, S.-H., Chen, L.-X., *et al*. (2014) Microbial communities evolve faster in extreme environments. *Sci Rep* **4**: 6205.

Li, W., and Godzik, A. (2006) Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **22**: 1658–1659.

López-Pérez, M., Haro-Moreno, J.M., Gonzalez-Serrano, R., Parras-Moltó, M., and Rodriguez-Valera, F. (2017) Genome diversity of marine phages recovered from Mediterranean metagenomes: size matters. *PLoS Genet* **13**: e1007018.

Ludwig, W., Strunk, O., Westram, R., Richter, L., Meier, H., Yadhukumar, A., *et al*. (2004) ARB: a software environment for sequence data. *Nucleic Acids Res* **32**: 1363–1371.

Mahmoudabadi, G., Milo, R., and Phillips, R. (2017) Energetic cost of building a virus. *Proc Natl Acad Sci U S A* **114**: E4324–E4333.

Martinez-Hernandez, F., Fornas, O., Lluesma Gomez, M., Bolduc, B., de la Cruz Peña, M.J., Martínez, J.M., *et al*.

(2017) Single-virus genomics reveals hidden cosmopolitan and abundant viruses. *Nat Commun* **8**: 15892.

Maturrano, L., Santos, F., Rosselló-Mora, R., and Antón, J. (2006) Microbial diversity in Maras salterns, a hypersaline environment in the Peruvian Andes. *Appl Environ Microbiol* **72**: 3887–3895.

Menzel, P., Ng, K.L., and Krogh, A. (2016) Fast and sensitive taxonomic classification for metagenomics with Kaiju. *Nat Commun* **7**: 1–9.

Meziti, A., Tsementzi, D., Rodriguez-R, L.M., Hatt, J.K., Karayanni, H., Kormas, K.A., and Konstantinidis, K.T. (2019) Quantifying the changes in genetic diversity within sequence-discrete bacterial populations across a spatial and temporal riverine gradient. *ISME J* **13**: 767–779.

Mizuno, C.M., Ghai, R., and Rodriguez-Valera, F. (2014) Evidence for metaviromic islands in marine phages. *Front Microbiol* **5**: 27.

Mizuno, C.M., Rodriguez-Valera, F., Kimes, N.E., and Ghai, R. (2013) Expanding the marine virosphere using metagenomics. *PLoS Genet* **9**: e1003987.

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

Mora-Ruiz, M.d.R., Font-Verdera, F., Orfila, A., Rita, J., and Rosselló-Móra, R. (2016) Endophytic microbial diversity of the halophyte *Arthrocnemum macrostachyum* across plant compartments. *FEMS Microbiol Ecol* **92**: fiw145.

Muniozguren, M.A., Ballesteros, I.M., Gamboa, J., Seoane, S., Alonso, R., Laorden, L., *et al*. (2021) Altererythrobacter muriae sp. nov., isolated from hypersaline Añana Salt Valley spring water, a continental thalassohaline type solar saltern. *Int J Syst Evol Microbiol* **71**: 1–9.

Mutlu, M.B., Martínez-García, M., Santos, F., Peña, A., Guven, K., and Antón, J. (2008) Prokaryotic diversity in Tuz Lake, a hypersaline environment in inland Turkey. *FEMS Microbiol Ecol* **65**: 474–483.

Muyzer, G., De Waal, E.C., and Uitterlinden, A.G. (1993) Profiling of complex microbial populations by denaturing gradient gel electrophoresis analysis of polymerase chain reaction-amplified genes coding for 16S rRNA. *Appl Environ Microbiol* **59**: 695–700.

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

Oren, A. (2014) The ecology of Dunaliella in high-salt environments. *J Biol Res* **21**: 23–23.

Parikka, K.J., Le Romancer, M., Wauters, N., and Jacquet, S. (2017) Deciphering the virus-to-prokaryote ratio (VPR): insights into virus-host relationships in a variety of ecosystems. *Biol Rev* **92**: 1081–1100.

Parks, D.H., Imelfort, M., Skennerton, C.T., Hugenholtz, P., and Tyson, G.W. (2015) CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res* **25**: 1043–1055.

Plata, A., and Erkiaga, A. (2018) *El sistema de producción de sal de Añana. Valle Salado (Araba, País Vasco)*. Universidad del País Vasco, Bilbao, España pp. 170.

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

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

Puente-Sánchez, F., Arce-Rodríguez, A., Oggerin, M., García-Villadangos, M., Moreno-Paz, M., Blanco, Y., *et al*. (2018) Viable cyanobacteria in the deep continental subsurface. *Proc Natl Acad Sci U S A* **115**: 10702–10707.

Ramos-Barbero, M.D., Martin-Cuadrado, A.-B., Viver, T., Santos, F., Martinez-Garcia, M., and Antón, J. (2019) Recovering microbial genomes from metagenomes in hypersaline environments: the good, the bad and the ugly. *Syst Appl Microbiol* **42**: 30–40.

Reisser, W. (2007) *The Hidden Life of Algae Underground*. Dordrecht: Springer, pp. 47–58.

Rodriguez-R, L.M., Gunturu, S., Harvey, W.T., Rosselló-Mora, R., Tiedje, J.M., Cole, J.R., and Konstantinidis, K.T. (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.

Rodriguez-R, L.M., and Konstantinidis, K.T. (2014) Nonpareil: a redundancy-based approach to assess the level of coverage in metagenomic datasets. *Bioinformatics* **30**: 629–635.

Rodriguez-R, L.M., and Konstantinidis, K.T. (2016) The Enveomics Collection: A Toolbox for Specialized Analyses of Microbial Genomes and Metagenomes. *PeerJ Prepr*.

Rodriguez-Valera, F., Mizuno, C.M., and Ghai, R. (2014) Tales from a thousand and one phages. *Bacteriophage* **4**: e28265.

Roudnew, B., Lavery, T., Seymour, J., Smith, R., and Mitchell, J. (2013) Spatially varying complexity of bacterial and virus-like particle communities within an aquifer system. *Aquat Microb Ecol* **68**: 259–266.

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

Roux, S., Enault, F., Ravet, V., Colombet, J., Bettarel, Y., Auguet, J.-C., *et al*. (2016) Analysis of metagenomic data reveals common features of halophilic viral communities across continents. *Environ Microbiol* **18**: 889–903.

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

Roux, S., Páez-Espino, D., Chen, I.-M.A., Palaniappan, K., Ratner, A., Chu, K., *et al*. (2020) IMG/VR v3: an integrated ecological and evolutionary framework for interrogating genomes of uncultivated viruses. *Nucleic Acids Res* **49**: D764–D775.

Rubino, F., Carberry, C., Waters, M., Kenny, D., McCabe, M. S., and Creevey, C.J. (2017) Divergent functional isoforms drive niche specialisation for nutrient acquisition and use in rumen microbiome. *ISME J* **11**: 932–944.

Salter, S.J., Cox, M.J., Turek, E.M., Calus, S.T., Cookson, W.O., Moffatt, M.F., *et al*. (2014) Reagent and

laboratory contamination can critically impact sequence-based microbiome analyses. *BMC Biol* **12**: 1–12.

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

Sayers, E.W., Agarwala, R., Bolton, E.E., Brister, J.R., Canese, K., Clark, K., *et al*. (2016) Database resources of the National Center for biotechnology information. *Nucleic Acids Res* **44**: D7–D19.

Schäfer, H., Bernard, L., Courties, C., Lebaron, P., Servais, P., Pukall, R., *et al*. (2001) Microbial community dynamics in Mediterranean nutrient-enriched seawater mesocosms: changes in the genetic diversity of bacterial populations. *FEMS Microbiol Ecol* **34**: 243–253.

Schloissnig, S., Arumugam, M., Sunagawa, S., Mitreva, M., Tap, J., Zhu, A., *et al*. (2013) Genomic variation landscape of the human gut microbiome. *Nature* **493**: 45–50.

Schmieder, R., and Edwards, R. (2011) Quality control and preprocessing of metagenomic datasets. *Bioinformatics* **27**: 863–864.

Silveira, C.B., Coutinho, F.H., Cavalcanti, G.S., Benler, S., Doane, M.P., Dinsdale, E.A., *et al*. (2020) Genomic and ecological attributes of marine bacteriophages encoding bacterial virulence genes. *BMC Genomics* **21**: 126.

Smillie, C.S., Smith, M.B., Friedman, J., Cordero, O.X., David, L.A., and Alm, E.J. (2011) Ecology drives a global network of gene exchange connecting the human microbiome. *Nature* **480**: 241–244.

Starnawski, P., Bataillon, T., Ettema, T.J.G., Jochum, L.M., Schreiber, L., Chen, X., *et al*. (2017) Microbial community assembly and evolution in subseafloor sediment. *Proc Natl Acad Sci U S A* **114**: 2940–2945.

Ventosa, A., de la Haba, R.R., Sánchez-Porro, C., and Papke, R.T. (2015) Microbial diversity of hypersaline environments: a metagenomic approach. *Curr Opin Microbiol* **25**: 80–87.

Villamor, J., Ramos-Barbero, M.D., González-Torres, P., Gabaldón, T., Rosselló-Móra, R., Meseguer, I., *et al*. (2017) Characterization of ecologically diverse viruses infecting co-occurring strains of cosmopolitan hyperhalophilic Bacteroidetes. *ISME J* **12**: 424–437.

Vreeland, R.H., Rosenzweig, W.D., and Powers, D.W. (2000) Isolation of a 250 million-year-old halotolerant bacterium from a primary salt crystal. *Nature* **407**: 897–900.

Wilkins, L.G.E., Ettinger, C.L., Jospin, G., and Eisen, J.A. (2019) Metagenome-assembled genomes provide new insight into the microbial diversity of two thermal pools in Kamchatka, Russia. *Sci Rep* **9**: 3059.

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

Zhang, F., Zhou, F., Gan, R., Ren, C., Jia, Y., Yu, L., and Huang, Z. (2019) PHISDetector: a tool to detect diverse in silico phage-host interaction signals for virome studies. *bioRXiv* 661074. https://doi.org/10.1101/661074.

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

Supplementary Dataset 1.

Supplementary Dataset 2.

Supplementary Dataset 3.

Supplementary Dataset 4.

Supplementary Dataset 5.

Supplementary Dataset 6.

Supplementary Fig. S1. (A) Abundances of cells and viruses in Añana samples. (B) Examples of photomicrographs used for counting (from top to bottom): SE DAPI stained cells, SE CARD-FISH with bacterial probe, HU CARD-FISH with archaeal probe, Sybr-Gold stained viruses from EC. Bar: 5 μm.

Supplementary Fig. S2. Phylogenetic tree reconstruction using the neighbour-joining algorithm on 16S rRNA genes recovered from assembled metagenomes.

Supplementary Fig. S3. A. Metagenome annotation. 1) glycosyltransferase involved in cell wall biosynthesis, 2) uncharacterized membrane protein, 3) signal transduction histidine kinase, 4) putative transposase, 5) nucleotide-binding universal stress protein, UspA family, 6) putative membrane protein, 7) cell division control protein 6, 8) transposase, 9) transcription initiation factor TFIIB, 10) predicted transcriptional regulator, 11) methyl-accepting chemotaxis protein, 12) DNA polymerase I, 13) PAS domain-containing protein, 14) DNA-binding transcriptional regulator, Lrp family, 15) branched-chain amino acid transport system permease protein, 16) ABC-2 type transport system ATP-binding protein, 17) small-conductance mechanosensitive channel, 18) transitional endoplasmic reticulum ATPase, 19) 23S rRNA. Archaeal LSU, 20) UDP-glucose 4-epimerase, 21) uncharacterized conserved protein, 22) transposase and inactivated derivatives, 23) uncharacterized protein conserved in archaea, 24) uncharacterized protein conserved in bacteria, 25) universal stress protein UspA and related nucleotide-binding proteins, 26) histidine kinase-, DNA gyrase B-, and HSP90-like ATPase, 27) predicted membrane protein, 28) PAS fold, 29) Site-specific recombinase XerD, 30) Response regulator containing CheY-like receiver, AAA-type ATPase, and DNA-binding domains, 31) glycosyltransferase, 32) Transcriptional regulators 33) K^+ transport systems, NAD-binding component, 34) pyridine nucleotide-disulphide oxidoreductase, 35) predicted DNA binding protein, 36) cation transport ATPase. B. Vial metagenome annotation. a) DNA polymerase, b) phage tail, c) HNH endonuclease, d) MULTISPECIES: hypothetical, e) terminase large, f) site-specific DNA-methyltransferase, g) phage portal, h) capsid protein, i) PBSX family, j) portal Halovirus, k) site-specific integrase, l) putative DNA, m) AAA family, n) ABC transporter, o) response regulator, p) DEAD/DEAH box, q) phage major, r) ATP-binding protein, s) tyrosine-type recombinase/integrase, t) tape measure, u) fibronectin type, v) putative ATPase, w) adenine methyltransferase, x) IS200/IS605 family, y) phage terminase.

Supplementary Fig. S4. Phylogenetic tree reconstruction using the neighbour-joining algorithm with the retinal binding proteins (RBP, also known as rhodopsins) recovered from assembled contigs of each cellular metagenome. Bar plots indicate the counts per million (CPM) reads mapping to each type of RBP.

Supplementary Fig. S5. MAGs from extremely halophilic microorganisms obtained from Añana metagenomes. For each MAG, bin metrics (number of contigs within the bin, N50 for the obtained bins, bin completeness and degree of contamination, GC content) and My Taxa evaluates MAGs in windows of ten genes and determines the taxonomic affiliation of each one. Additionally, My Taxa provides information about possible cases of horizontal gene transfer or contamination in studied genome, these regions are marked as red circles in the figure.

Supplementary Fig. S6. Añana MAGs recruitment plot in all cell metagenomes.

Supplementary Fig. S7. Taxonomic composition of viral metagenomic reads.

Supplementary Fig. S8. Taxonomic composition of viral contigs from the HU viral metagenome.

Supplementary Fig. S9. Taxonomic composition of viral contigs from the HP viral

metagenome.

Supplementary Fig. S10. Taxonomic composition of viral contigs from the SE viral metagenome.

Supplementary Fig. S11. Taxonomic composition of viral contigs from the EC viral metagenome.

Supplementary Fig. S12. Annotation of cosmopolitan viral genomes.

Supplementary Table S1. Characteristics of the cellular and viral metagenomes analysed.

Supplementary Table S2. Viral abundances and ANIr of viral contigs in cell and virus metagenomes.

Supplementary Table S3. General characteristics and abundance of HU virus subset in Añana viromes and cell metagenomes

Supplementary Table S4.Distribution of pN/pS calculated for viral genomes from all Añana metaviromes