AEM
Journals.ASM.org

Virioplankton Community Structure in Tunisian Solar Salterns

Ines Boujelben,[a,b] Pablo Yarza,[b] Cristina Almansa,[c] Judith Villamor,[b] Sami Maalej,[a] Josefa Antón,[b,d] and Fernando Santos[b]

Département des Sciences de la Vie, Faculté des Sciences de Sfax, Université de Sfax, Sfax, Tunisia[a]; Departamento de Fisiología, Genética y Microbiología, Universidad de Alicante, Alicante, Spain[b]; Servicios Técnicos de Investigación (SSTTI), Unidad de Microscopía, Universidad de Alicante, Alicante, Spain[c]; and Instituto Multidisciplinar para el Estudio del Medio Ramón Margalef, Universidad de Alicante, Alicante, Spain[d]

The microbial community inhabiting Sfax solar salterns on the east coast of Tunisia has been studied by means of different molecular and culture-dependent tools that have unveiled the presence of novel microbial groups as well as a community structure different from that of other coastal hypersaline environments. We have focused on the study of the viral assemblages of these salterns and their changes along the salinity gradient and over time. Viruses from three ponds (C4, M1, and TS) encompassing salinities from moderately hypersaline to saturated (around 14, 19, and 35%, respectively) were sampled in May and October 2009 and analyzed by transmission electron microscopy (TEM) and pulsed-field gel electrophoresis (PFGE). Additionally, for all three October samples and the May TS sample, viral metagenomic DNA was cloned in fosmids, end sequenced, and analyzed. Viral concentration, as well as virus-to-cell ratios, increased along the salinity gradient, with around 10^{10} virus-like particles (VLPs)/ml in close-to-saturation ponds, which represents the highest viral concentration reported so far for aquatic systems. Four distinct morphologies could be observed with TEM (spherical, tailed, spindled, and filamentous) but with various proportions in the different samples. Metagenomic analyses indicated that every pond harbored a distinct viral assemblage whose G+C content could be roughly correlated with that of the active part of the microbial community that may have constituted the putative hosts. As previously reported for hypersaline metaviromes, most sequences did not have matches in the databases, although some were conserved among the Sfax metaviromes. BLASTx, BLASTp, and dinucleotide frequency analyses indicated that (i) factors additional to salinity could be structuring viral communities and (ii) every metavirome had unique gene contents and dinucleotide frequencies. Comparison with hypersaline metaviromes available in the databases indicated that the viral assemblages present in close-to-saturation environments located thousands of kilometers apart presented some common traits among them in spite of their differences regarding the putative hosts. A small core metavirome for close-to-saturation systems was found that contained 7 sequences of around 100 nucleotides (nt) whose function was not hinted at by *in silico* search results, although it most likely represents properties essential for hyperhalophilic viruses.

Haloviruses infect halophilic microorganisms and are found in hypersaline environments. These systems harbor the highest density of viruses reported for aquatic systems, with concentrations reaching up to 10^9 virus-like particles (VLPs) per ml and virus-to-cell ratios up to 100 (15), a number that is considerably higher than ratios in most environments (from 3 to 10 viruses per microbial cell [41]). Furthermore, haloviruses may be the main biological factor controlling prokaryotic populations in hypersaline environments, since there are not usually eukaryotic predators at salinities above 25% (reference 52 and references therein). The study of haloviruses has depended traditionally on the isolation of the virus-host systems. Almost 70 haloviruses have been isolated, mainly from extremely halophilic *Archaea* (5, 27, 46). Most of the virus isolates correspond to *Caudovirales* (head-tailed double-stranded-DNA [dsDNA]-genome viruses), although spherical, "spindle-like," or pleomorphic morphologies have also been obtained by cultivation (5, 13, 14). However, most isolated haloviruses infect hosts that frequently constitute minor components of natural prokaryotic communities, and genomic sequences from isolated haloviruses are only rarely related to viral sequences directly retrieved from the environment. In fact, there have been no published reports on the isolation of viruses infecting two of the most conspicuous groups of extremely halophilic prokaryotes, the archaeon *Haloquadratum walsbyi* (19, 20, 23) and the bacterium *Salinibacter ruber* (3, 4, 34), and closely related microbes.

Viral communities from solar salterns (La Trinitat and Bras del Port in Spain), the Dead Sea, and hypersaline lakes (the Retba, Great Salt, and Mono lakes) have been observed by transmission electron microscopy (TEM) (15, 17, 22, 30, 38, 50), unveiling the presence of four main morphologies in variable proportions: head tailed, icosahedral, spindle like, and filamentous. Haloviral genomic diversity, in terms of genome size distribution, has also been analyzed by pulsed-field gel electrophoresis (PFGE) (25, 31, 47, 48, 50, 51). In some cases, while several morphotypes can be observed by TEM in a given sample, only a single DNA band is observed in PFGE gels, indicating that viral "species" present in the sample harbor a very similar genome size (28, 50, 51). Sequence data from hypersaline systems (viral metagenomes from the Mono and Reba lakes and Bras del Port and San Diego solar salterns) have also shown that the uncultured halovirus community is highly diverse and not related to the culturable fraction (28, 45, 47, 51, 53).

We used culture-independent approaches to analyze the viral assemblage of selected ponds from Sfax multipond solar salterns

Received 7 June 2012 **Accepted** 6 August 2012
Published ahead of print 17 August 2012
Address correspondence to Josefa Antón, anton@ua.es.
Supplemental material for this article may be found at http://aem.asm.org/.

doi:10.1128/AEM.01793-12

in southeast Tunisia. During the past 4 years, this thalassohaline system has been the object of many studies that have allowed the isolation and description of new halophilic bacterial and archaeal strains (7, 56), as well as the description, using the rRNA approach, of the communities inhabiting pond waters (8, 21, 55), sediments (10), magnesium-rich brines (11), and salt crystals (9).

The prokaryotic community in Sfax crystallizers (ponds where sodium chloride precipitates) is dominated by *Archaea*, as revealed by fluorescence *in situ* hybridization (FISH) (21). *Haloquadratum* spp., *Halorubrum* spp., and uncultured *Halobacteriaceae*-related sequences were found to be predominant in both 16S rRNA gene libraries and denaturing gradient gel electrophoresis (DGGE) profiles (8, 21). However, studies carried out by combining cell sorting of Sybr green-stained cells and sequencing of 16S rRNA genes indicated that in the crystallizers nearly 70% of *Haloquadratum* sequences corresponded to the low nucleic acid content (LNA) fraction of cells (55), thus probably indicating that a significant percentage of the archaeal assemblage was not very active. Compared to that of crystallizers, the species composition from medium-salinity ponds (from 10 to 20% salinity) showed more fluctuations and was dominated by bacterial populations (mostly *Bacteroidetes* and *Proteobacteria*), which exceeded archaeal representatives (21). Apart from these 16S rRNA gene sequences, classically retrieved from hypersaline environments, Narasingarao et al. (36) have recently found in Sfax salterns (and other hypersaline settings) sequences from the new haloarchaeal group "Nanohaloarchaea."

All the above-mentioned studies were focused on Sfax prokaryotic populations, while in our study the focus was on the viral assemblage and how it changes with salinity and time. For this purpose, the viral community present in three saltern ponds of increasing salinity was sampled in May and October 2009 and analyzed by transmission electron microscopy, pulsed-field gel electrophoresis, and metagenomic investigation of selected samples.

MATERIALS AND METHODS

Sampling. Samples from ponds C4, M1, and TS were taken in May 2009 and October 2009 from the Sfax solar salterns located on the central eastern coast of Tunisia (34°39′N, 10°42′E). Salinity of the six samples (named C4May, C4Oct, M1May, M1Oct, TSMay, and TSOct) was measured *in situ* with a hand refractometer (Eclipse). Temperature and pH were also measured *in situ* with a mercury glass thermometer graduated at 0.1°C and a pH meter, respectively.

DAPI, FISH, and Sybr gold counts. Water samples were fixed with formaldehyde (7% final concentration) during 16 h at 4°C and filtered by 0.2-μm GTTP filters (Millipore). Cells were stained with DAPI (4′,6-diamidino-2-phenylindole) after *in situ* hybridizations with EUB338 and Arc915 probes (1) as described by Antón et al. (2). Cells were visualized and counted in an epifluorescence microscope (Leica, type DM4000B; Vashaw Scientifics Inc., Norcross, GA).

For viral counts, 10 to 100 μl from each water sample was fixed with formaldehyde (4% final concentration) for 30 min at room temperature. From 1 to 100 μl of fixed samples was filtered using 0.02-μm Anodisc 25 filters (Whatman). Sybr gold (Sigma) was used for staining viral particles, which were then visualized under an epifluorescence microscope (Leica, type DM4000B; Vashaw Scientifics Inc., Norcross, GA) and counted as described above.

Virus concentration. Cells from 300 to 600 ml of water (depending on the sample) were removed by centrifugation (30,000 × *g*, 1 h, 20°C; Avanti J-30I, Beckman with a JA rotor), and the supernatants were concentrated by tangential flow filtration through a Vivaflow system with a 30,000-molecular-weight-cutoff (MWCO) filter cassette. Water concentrates were filtered using 0.2-μm filters, and cell-free virus suspensions were finally reconcentrated to 250 μl by using 10,000 MWCO Amicon Ultra centrifugal filters (Millipore).

Transmission electron microscopy. A 10-μl sample from each viral suspension was fixed with formaldehyde (4% final concentration) for 30 min at 4°C and stained for 5 min with uranyl acetate (0.5%) on Formvar-coated carbon grids (Electron Microscopy Sciences). Virus-like particles were observed in a Jeol JEM-2010 transmission electron microscope operating at 200 kV. To see the proportion of the different viral morphotypes, 160 TEM images were taken at the same magnification and more than 900 VLPs were counted.

Viral DNA purification. Virus concentrates were mixed with equal volumes of 1.6% low-melting-point agarose (Pronadisa), dispensed into 100-μl molds, and allowed to solidify at 4°C. Agarose plugs were incubated for 1 h with 4 units of Turbo DNase (Ambion) to digest dissolved DNA. The plugs were then incubated overnight at 50°C in ESP (0.5 M EDTA [pH 9.0], 1% *N*-laurylsarcosine, 1 mg/ml proteinase K) for disruption of viral capsids. Some plugs from each sample were then used for pulsed-field gel electrophoresis (see below), and the rest were used for viral DNA extraction. For DNA extraction, plugs were washed with TE-Pefabloc (10 mM Tris-HCl, 1 mM EDTA [pH 8.0], 3 mM Pefabloc) to inactivate proteinase K and incubated for 15 min at 65°C. The mixture of viral DNA and melted agarose was treated with β-agarase (New England BioLabs) for 1.5 h at 42°C (1 enzyme unit per 0.1 g of melted mixture). DNA was then concentrated and purified using Microcon YM-100 centrifugal filter devices (Millipore) and checked for quality on a 1% agarose gel.

PFGE and Southern analysis. Plugs containing viral DNA were subjected to PFGE in a 1% low-electroendosmosis (LE) agarose gel (FMC) in Tris-borate-EDTA (TBE) 0.5× buffer, using a Bio-Rad (Richmond, CA) Chef DR-III system operating at 6 V/cm, with a 1- to 25-min pulse ramp, at 14°C for 24 h. A lambda low-range DNA size ladder (New England BioLabs) served as a molecular weight marker. The gel was visualized after staining with ethidium bromide (1 μg/ml) and washed with distilled water. DNA was then transferred onto a membrane for Southern hybridization against labeled viral DNA from the pond sample TSOct. The DIG High Prime DNA labeling and detection starter kit II (Roche) was used for Southern analysis according to the manufacturer's protocol.

Fosmid library constructions and sequencing. Four viral metagenomic libraries were constructed from samples TSMay, C4Oct, M1Oct, and TSOct. From 0.5 to 1 μg of viral DNA was end-repaired and cloned using the CopyControl HTP Fosmid library production kit with the pCC2FOS vector (Epicentre) according to the manufacturer's recommendations. Fosmid clones were induced, and fosmid ends sequenced by the GATC Biotech service (Konstanz, Germany) using pCC2FOS sequencing primers (Epicentre).

Sequence analyses. (i) Assembly. High-quality and manually revised sequences were assembled to search for groups of similar viruses with the SeqMan application of the Lasergene software (DNASTAR Corporation) and the following assembly parameters: minimum match size of 50 nucleotides, minimum identity of 95%, and minimum alignment length of 100 nucleotides. An assembly with the two sequences of each fosmid (parameters: minimum match size of 50 nucleotides, minimum identity of 100%, and minimum alignment length of 100 nucleotides) was carried out to search for inverted terminal repeats in the viral genomes.

(ii) BLASTn analysis. The 4 metaviromes reported in the present work were complemented with a selection of 24 additional metagenomes found in the literature, and all possible pairwise BLASTn comparisons were carried out. Additional data sets comprised metagenomes from C4Oct, M1Oct, TSOct, and TSMay (present work); 19 high-, medium-, and low-salinity metagenomes from San Diego (26); 3 metagenomes from Santa Pola's CR30 crystallizer pond (28, 32, 51); 1 metagenome from Lake Retba (53); and 1 metagenome from the Dead Sea (18). The BLASTn results were automatically parsed to calculate the amount of nonredundant nucleotides from an Sfax metavirome "A" that were shared with a second selected metagenome "B" by means of meaningful BLAST high-

TABLE 1 Physicochemical characteristics and counts for total cells, *Archaea*, *Bacteria*, and viruses

Sample	Salinity (%)	Temp (°C)	pH	Cells/ml ($\times 10^8$) (mean ± SD)	% *Archaea*/% *Bacteria*	VLPs/ml ($\times 10^8$) (mean ± SD)	Virus/cell ratio
C4May	13.8	24	8.04	1.14 ± 0.12	ND[a]/47	1.92 ± 0.18	1.7
C4Oct	14.7	21	8.4	1.13 ± 0.12	ND/67	5.39 ± 0.48	4.8
M1May	19	27	8.51	4.92 ± 0.38	ND/ND	7.72 ± 0.38	1.6
M1Oct	18.5	23	8.47	1.09 ± 0.10	ND/47	11.7 ± 1.1	10.7
TSMay	34	34	7.60	2.71 ± 0.49	59/9	134 ± 11	49.4
TSOct	36	27	7.46	2.33 ± 0.26	58/20	126 ± 12	54.1

[a] ND, not detected.

scoring segment pairs (HSPs) (i.e., a maximum E-value of 0.001, a minimum of 80% sequence identity, and a minimum alignment length of 100 positions).

(iii) BLASTx searches. In parallel with automatic annotation through the Seed platform (6), we subjected our contigs to BLASTx runs against the nonredundant nucleotide database using the on-line BLAST tool at NCBI (http://blast.ncbi.nlm.nih.gov/Blast.cgi). For each contig, the five best matches with a maximum E-value of 0.001 were automatically retrieved and manually inspected. For a given BLAST-analyzed sequence a functional category was assigned if four of the five best matches were related to the same protein. If not, the BLAST-analyzed sequence was assigned as coding for a "conserved hypothetical protein."

(iv) BLASTp analysis. All 4 of the metaviromes were self-compared and compared with each other to ascertain the degree of functional conservation within the same sample across the salinity gradient and during different seasons. Open reading frames (ORFs) were automatically extracted from SEED annotation (6). The same procedure and parameters as in reference 51 were used, except that matches against the query sequence itself were preserved (to generate a straight line for a better visualization of Fig. 4A).

(v) Dinucleotide frequencies. The statistical procedure for dinucleotide analysis was carried out as in reference 51. Although three principal components (PCs) were enough to explain 72.58% of the variance, four PCs (82.67% variance explained) provided a better explanation of the data and were selected for subsequent clustering approaches. Seven *k*-mean groups were modeled and proved for significance (at 0.05) with the analysis of similarity (ANOSIM) test (43).

(vi) Hypersaline core metavirome. All known metaviromes obtained from high-salinity ponds (see references above) were classified into four distinct new data sets according to distinct sample sites: (i) the *SanDiego* data set, made up of the 3 metaviromes of high salinity from the San Diego solar salterns in the United States; (ii) the *CR30* data set, containing the 2 metaviromes from the Santa Pola CR30 crystallizer pond in Spain; (iii) the *TSall* data set, formed by joining the two TSMay and TSOct Sfax saltern metaviromes reported in the present work; and (iv) the *LakeRetba* data set, representing the metavirome reported for this lake in Senegal. Pairwise BLASTn comparisons of the four data sets (which were performed following standard procedures cited above) allowed us to recognize the sequences showing meaningful HSPs in all four of the data sets. The size of the calculated "core" is given in terms of the shortest sequences found in a group of matching sequences.

Nucleotide sequence accession numbers. The sequences reported in this article have been deposited in the GenBank database under accession numbers JQ793966 to JQ794444.

RESULTS AND DISCUSSION

Abundance of prokaryotic cells. The three ponds analyzed in this study were distributed along a salinity gradient, from 13.8% to salt saturation (Table 1). Ponds C4 and M1 were considered to be of medium salinity, whereas pond TS was a crystallizer and therefore had salt concentrations close to saturation. The ponds did not undergo drastic changes in salinity before and after summer, as shown by the values obtained in May and October 2009 (Table 1). The numbers of cells per milliliter ranged from 1.09×10^8 (in M1Oct) to 4.92×10^8 (in M1May), which were slightly higher than DAPI counts previously reported for samples taken from the same ponds in 2007 and 2008 (21) and very close to cytometric characterizations carried out by Trigui et al. (55) from M1 and TS samples also taken in October 2009. A temporal variation in the DAPI counts was clearly observed only in the M1 pond, where cell numbers in May were 4.5 times higher than in October. These data are in agreement with the above-mentioned work by Boujelben et al. (21), who also found larger variations in the M1 prokaryotic assemblage, both in numbers and diversity (21).

The proportion of cells detected by FISH was significantly different in TS crystallizer and medium-salinity ponds (Table 1). While *Archaea* were not detected in C4 and M1 samples, they dominated the prokaryotic community in TS (around 60% of the total DAPI counts), as happened with most of the salt-saturated waters studied formerly (2, 33, 35). Previous studies carried out in Sfax salterns (55) demonstrated that most cells in a sample (also taken in October 2009) from the TS crystallizer pond had a high nucleic acid (HNA) content and therefore were highly active. Thus, the high proportion of FISH-detected cells in TS could correspond to the HNA fraction, a result that is in agreement with the idea that FISH detects the active fraction of a community (37). In the same study by Trigui et al. (55), the sample taken in October 2009 from pond M1 harbored a large proportion (around 64%) of low-nucleic-acid-content (LNA) cells. Accordingly, the proportion of cells detected by FISH was considerably lower than in TS (Table 1), with *Bacteria* as the dominant component of the community.

Abundance of virus-like particles. The numbers of VLPs (Table 1) in Sfax ponds were always higher than the number of cells, reaching more than 10^{10} particles per ml in the crystallizer TS. This is the highest number of viruses reported to date for aquatic systems. Virus-to-cell ratios in TS were in the range of values obtained for other salt-saturated environments (between 42 and 100 viruses per host), like the Dead Sea, the Santa Pola crystallizer CR30, the Great Salt Lake, and the San Diego crystallizers (for a discussion of virus-to-cell ratios see reference 52).

Transmission electron microscopy of VLPs. Transmission electron microscopy was used to ascertain the different viral morphotypes (Fig. 1A) and their relative abundance. More than 900 VLPs were analyzed in total, 53% of which were in the range between 50 and 100 nm, and 41.5% had a size smaller than 50 nm; some filamentous viruses above 100 nm were also observed. Spherical viruses (including some head-tailed viruses that could have lost their tails in the TEM preparations) dominated the viral community in Sfax salterns, reaching up to 80% in C4 samples (Fig. 1B). Although most spherical VLPs in Sfax salterns showed a

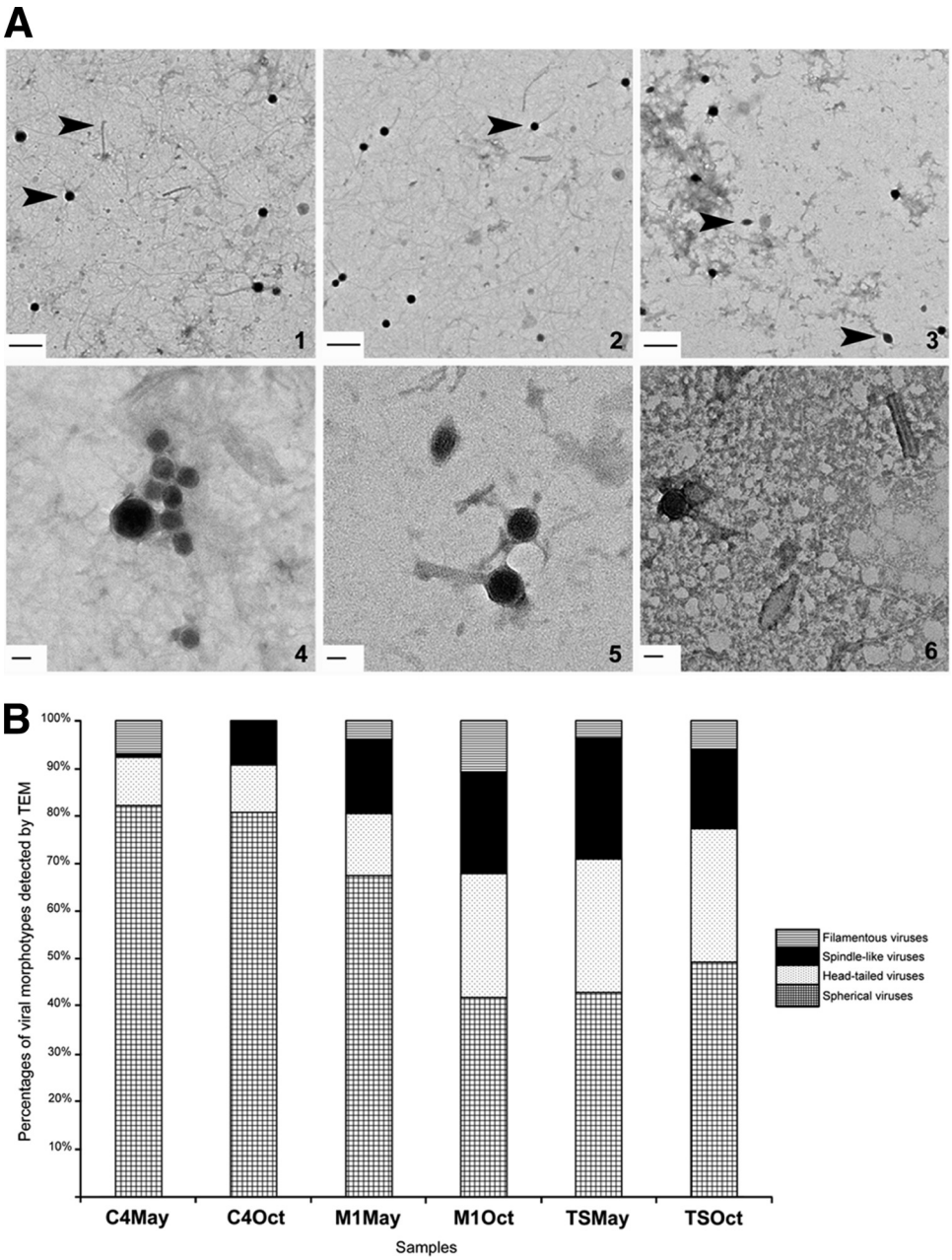

FIG 1 (A) Transmission electron micrographs showing different viral morphotypes found in Sfax salterns. Examples of micrographs used for counting (scale bar in images 1 to 3, 200 nm; viral particles are indicated by arrowheads) are shown, as are details of spherical (scale bar in image 4, 50 nm), spindle-like, and head-tailed (scale bar in image 5, 50 nm) and filamentous (scale bar in image 6, 20 nm) viruses. (B) Distribution of morphotypes (in percentages) detected on every analyzed sample by TEM.

clear virus-like morphology, we cannot rule out the possibility that some of these particles corresponded to DNA-containing membrane vesicles, such as the ones described as hyperthermophilic *Archaea* of the order *Thermococcales* (54).

The number of spherical viruses decreased with salinity, while the proportion of tailed and "spindle-like" (also known as "lemon-shaped") viruses was higher in the M1 and TS ponds, which may indicate that this morphotype is better adapted to higher salinities (Fig. 1B). Indeed, lemon-shaped viruses can reach up to 25% in Mediterranean crystallizers (30) or even dominate the community in the close-to-salt-saturation Lake Retba (17). As the increase in salinity is normally accompanied by an increase of archaeal members, it has been suggested that lemon viruses would infect haloarchaea (30). In fact, His1 and His2 phages, isolated from *Haloarcula hispanica*, belong to this morphotype (13, 14).

Pulsed-field gel electrophoresis of viral assemblages. PFGE was used to evaluate the genomic size pattern of the most abundant members of Sfax virioplankton (viral genomes below a certain concentration cannot be detected by this technique). Genomic patterns included between 2 and 5 different bands with

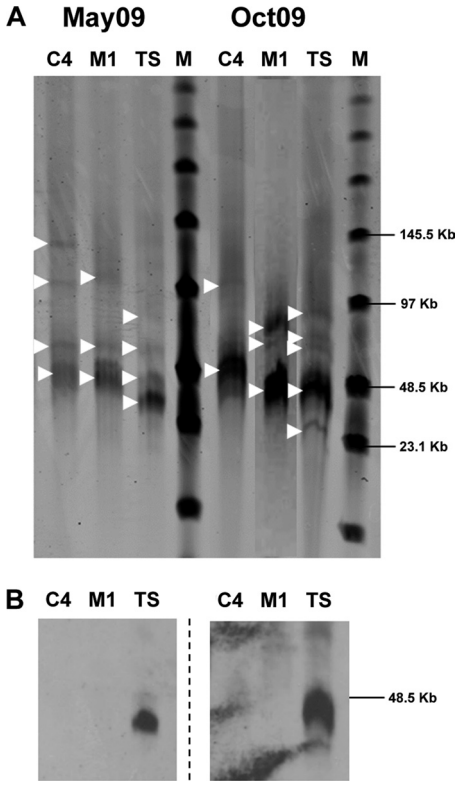

FIG 2 (A) Pulsed-field gel electrophoresis of Sfax saltern viral assemblages. DNA bands are indicated by white triangles. M, lambda PFG marker; sizes of the bands are indicated in kilobases (kb). (B) Southern hybridization between the PFGE viral assemblages and labeled viral DNA from TSOct (the left image corresponds to transferred DNA from the May samples; the right image corresponds to transferred DNA from October samples).

a size range from around 23 kb up to 96 kb. Viral genomes over 96 kb were detected only in C4 and M1 samples, while TS samples showed a narrower range of genome sizes (Fig. 2A). All samples showed a predominant band in the region from around 40 to 50 kb (Fig. 2A) that, as shown by previous studies (50, 51), could comprise an undetermined number of different viral genomes with very similar sizes. PFGE has been previously used to study the viral assemblages in Mediterranean salterns and the Mono Lake, in which the most abundant genomes also ranged from 30 to 60 kb (25, 47, 48), or even in a very narrow region of around 37 kb (50).

Southern hybridization was carried out to evaluate whether the viral genetic diversity was specific for every pond (Fig. 2B). For this purpose, the PFGE gel shown in Fig. 2A was transferred onto a membrane and hybridized against labeled viral DNA from TSOct. There was a hybridization signal only with the lanes corresponding to TSOct (positive control) and TSMay samples. These results suggested that viral populations were specific for a given pond, although some viral genomes (or certain genomic regions) could remain in the crystallizer, at least for a period of several months. Similar hybridization experiments with viral communities have also been carried out with samples from a Mediterranean crystallizer pond and Mono Lake (31, 50). In the first case, the results suggested that a viral genome (or certain genomic regions) directly retrieved from a crystallizer remained in the environment for at least a couple of years (50). In the Mono Lake study (31), results indicated a faster recycling of viral genotypes (or viral species). Fast viral recycling was also observed by metagenomic analyses in the San Diego, CA, salterns (45).

Viral metagenomics from Sfax salterns. One of the best ways to characterize a viral community is the construction of a viral metagenome or metavirome (49). Viral metagenomes have been constructed from different hypersaline environments such as Mono Lake, the San Diego and Santa Pola salterns, and Lake Retba by using different methodologies that include cloning in plasmids and fosmids, pyrosequencing, and combinations of both approaches (references 28 and 52 and references therein).

Here, a total of 4 metaviromes (Table 2) were constructed to (i) describe the viral diversity in the different ponds, (ii) compare the viral populations found along the salinity gradient (samples C4Oct, M1Oct, and TSOct), and (iii) study the viral recycling in the crystallizer after a 5-month period (samples TSMay and TSOct). The metagenomic libraries were constructed with the viral DNA corresponding to the fraction between 23 and 50 kb observed in the PFGE analysis, purified, and directly cloned into fosmids. As explained above, most of the DNA that could be detected by PFGE was included within this size range. The 4 metaviromes constructed from Sfax samples contained a total of around 12 M bp of cloned DNA (Table 2).

Primary sequence data, contig assembly, and statistics. Between 76 and 92% of the total available clones yielded good quality end sequences with both forward and reverse vector primers, while the remaining clones were successfully sequenced only with either one of the primers. An assembling step was carried out in order to avoid redundancy in subsequent analyses. The lowest percentage of unique sequences (corresponding to different viral genotypes) was found in the C4Oct sample (45%), suggesting that the increase in salinity is also accompanied by an increase in viral genotype diversity

TABLE 2 Statistics of fosmid libraries

Features	Samples			
	TSMay09	C4Oct09	M1Oct09	TSOct09
Total cloned DNA (Mb)[a]	3.6	3.8	3.5	2.5
No. of sequences	159	181	153	116
Avg length (bp)	592	653	630	513
Total sequenced (kb)	94	118	96	59.5
Total assembled (kb)	73.5	84.3	85	57
G+C content (avg) (%)	61	56	59	58
No. of sequenced fosmids	90	93	87	63
Fosmids with unique sequences (%)	56.6	45	63.2	67.7
Most abundant fosmid (%)[b]	6.6	8.5	7	6.5
BLASTx hits (% of contigs)				
No hits	41.94	40.31	49.64	55.17
Conserved hypothetical proteins	54.61	31.78	30.66	33.62
Terminases	3.23	11.63	11.68	3.45
Nucleases	4.03	5.43	3.65	3.45
Methylases	1.61	0.00	0.73	0.86
Integrases	0.00	1.55	1.46	0.00
Others	0.00	7.75	2.92	2.59

[a] Value calculated assuming 40 kb as the average insert size.
[b] Assembling parameters: match size of 50 nucleotides in a minimal alignment of 100 nucleotides, with a minimum identity of 95%.

(Table 2). The different viral genotypes can infect not only hosts from different prokaryotic species but also hosts from different strains in a given species. In fact, crystallizers are hypersaline systems with a low diversity and a very high microdiversity (24, 32, 39, 40, 42).

Both forward and reverse sequences from a given insert were also analyzed to search for inverted terminal repeats (ITRs) in the viral genomes. These repeats have been found in several linear and double-stranded DNA viruses that contain 5′-covalently linked terminal proteins and whose replication takes place through a protein-primed mechanism (12). In Sfax salterns, ITRs were found in 16% (TSMay), 12% (C4Oct), 14% (M1Oct), and 11.3% (TSOct) of the analyzed fosmids.

G+C content and tentative hosts. Since the G+C content of a given virus is normally very close to that of its host (44), the G+C content of the Sfax cloned viral sequences was determined in order to ascertain their likely hosts among the microbial populations present in the system (8, 21, 55). Surprisingly, most of the viral sequences from the whole Sfax metavirome presented a high G+C content (Table 2). In previous work carried out with the metavirome of the CR30 crystallizer from the Santa Pola salterns (51), half of the viral metagenomic sequences (from both fosmid and plasmid libraries) had a low G+C content while the rest of the sequences showed G+C values above 50%. Dinucleotide frequency values indicated that the high-G+C-content sequences most likely included viruses infecting *Salinibacter ruber* (with a G+C content of around 70%) and high-G+C-content haloarchaea. On the other side, low-G+C-content viral sequences could belong to viruses infecting the low-G+C-content *Hqr. walsbyi,* the dominant microbe in the CR30 crystallizer (2, 16, 29). In Sfax salterns, medium-salinity ponds are dominated by high-G+C-content prokaryotes (21), and accordingly, 70% of the viral sequences analyzed from C4Oct and up to 90% of sequences from M1Oct showed a high G+C content. However, the metaviromes from TS crystallizer presented a low proportion (around 10%) of low-G+C-content viral sequences, an unexpected finding for a system that seems to be dominated by the "square" archaeon (8). A possible explanation for this result could be attributed to a very high proportion of inactive squares in the TS crystallizer. As discussed by Trigui et al. (55), 40% of the cells in the TS crystallizer had low nucleic acid content and could be considered the inactive group, 60% of which were assigned to *Hqr. walsbyi*. In the active (or HNA) cluster, up to 70% of the cells were identified as *Halorubrum* spp. and uncultured *Halobacteriaceae*. Viruses need to infect active cells in order to complete their life cycles, which makes active cells more susceptible to infection than inactive or dormant ones. Therefore, the unexpectedly high proportion of high-G+C-content viruses in TS could be the direct consequence of the dominance of high-G+C-content hosts within the active part of the community.

Comparison of Sfax metaviromes with themselves and with other hypersaline systems by BLASTn. BLASTn comparisons among the 4 metaviromes under study revealed that viral diversity was significantly different in each pond, although medium-salinity sequences were more similar to each other than those from crystallizer samples themselves (see Table S1 in the supplemental material). Comparisons between Sfax sequences and additional metagenomes (see Materials and Methods) using BLASTn showed that most of the C4Oct hits occurred when BLAST analysis was performed against medium-salinity metaviromes from San Diego salterns, while samples M1 and TS more frequently matched high-salinity viral sequences from CR30, San Diego, and Lake Retba sequences. Although hits between M1Oct (a medium-salinity pond) and high-salinity samples from San Diego salterns were found, the percentage was lower than in the case of the TS crystallizer. The number of hits between the TS crystallizer samples and the Lake Retba and CR30 crystallizer samples from the Santa Pola salterns (close-to-salt-saturation environments) were also higher than in the case of sample M1 (for the C4Oct sample hits were not found).

In an attempt to identify the "viral core metagenome" of all the analyzed close-to-saturation environments, we looked for the sequences common to the high-salinity ponds from the San Diego salterns, Lake Retba, and the Santa Pola crystallizer CR30 and TS samples from the Sfax salterns. The search yielded a total of 7 sequences of around 100 nucleotides each that were present in all these samples and absent from the rest of metaviromes in the databases. BLASTx analyses indicated that these sequences coded for hypothetical proteins, which would be conserved among the hypersaline metaviromes, while no motifs suggesting their function or structure could be found. The small size of the core and the small amount of information it provides were due in part to the small size of some of the metaviromes analyzed and the short length of the 454 sequences that constitute some them, which hamper ORF prediction and protein-based comparison. In any case, the presence of identical sequences in all the analyzed data sets is remarkable and suggests that this core metagenome codes for functions that are ecologically relevant in hypersaline systems.

Annotation of Sfax metaviromes. BLASTx analysis showed that between 40 and 55% of the contigs did not have matches in the databases. Most C4 and M1 sequences produced bacterial hits, while a high proportion of both TSMay and TSOct sequences matched viral sequences (up to 22% of total searches in TSMay; Fig. 3A). This might be biased due to the previously sequenced CR30 viral metagenomes (28, 51), as ORFs annotated as "hypothetical proteins" were turned into "conserved." Bacterial hits decreased along the salinity gradient while archaeal hits showed the opposite trend (Fig. 3B), indicating that most of the hosts in salt-saturated environments are *Archaea*.

Terminases and nucleases were the most abundant functions found in the 4 metaviromes (Table 2). Terminases are components of the molecular motor that translocates genomic DNA into empty capsids during DNA packaging in the head-tailed viruses, order *Caudovirales* (dsDNA viruses with head-tail morphology).

Self-to-self comparisons of predicted proteins. We performed BLAST analysis against themselves of all 520 translated ORFs from the 4 metaviromes. The outcome has been summarized in Fig. 4A. First, certain proteins (presumably with relevant functions for haloviruses) appeared to be conserved in several contigs from the same pond. For example, locator 283 from the TS pond, which encodes a conserved hypothetical protein, was repeated nine times in other contigs from TSMay and TSOct. Proteins of unknown function conserved among different viruses inhabiting the same pond were also observed when we characterized the CR30 crystallizer metavirome from the Santa Pola salterns (51). In the Sfax salterns, each sample showed a particular set of ORFs that may have been relevant (repeated in at least 2 distinct contigs) and exclusive (not represented in any other sample). Accordingly, 18 ORFs were exclusively retrieved from the C4Oct sample, 19 ORFs from M1Oct, 6 ORFs from TSOct, and 23 ORFs from TSMay, and 46 ORFs were exclusively retrieved from the TS pond (from a mixture of both May and October samples). On the other hand, many ORFs were conserved between ponds C4 and M1 and between M1 and TS, while C4 and TS did not share common ORFs. These results are in good agreement with the

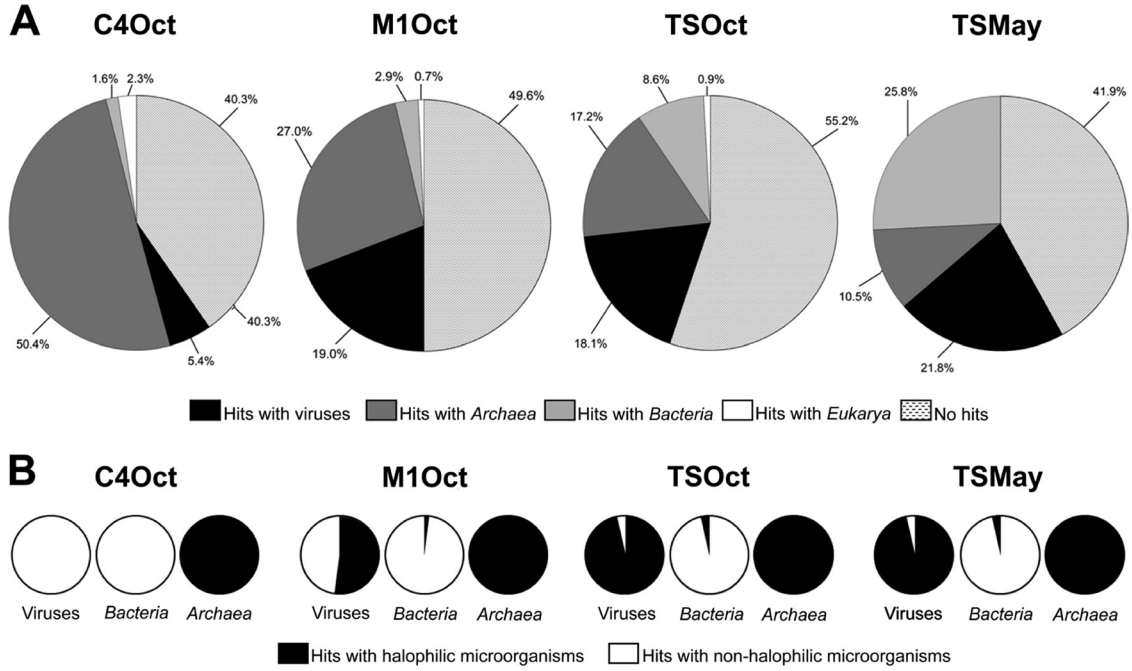

FIG 3 (A) Domain-level classification of the best BLASTx hits produced by the four analyzed metaviromes. (B) Percentages of sequences producing BLASTx matches with halophilic or nonhalophilic *Archaea*, *Bacteria*, or viruses.

BLASTn results that did not show any similarity at the nucleotide level between C4 and TS.

Although most of these genes have unknown functions, the genes are conserved. This most likely indicates that they code for functions relevant for haloviruses within the same pond (e.g., locator 438, a conserved hypothetical protein, is shared by 6 distinct contigs that are all from TSMay), and even between distinct ponds (e.g., locator 188, which encodes a terminase, is shared by 3 M1 contigs and 4 TS contigs). These findings might reflect the different life strategies and adaptations to certain salinity conditions, as indicated by the overlap between the M1 and C4 ponds (e.g., locator 3, a conserved hypothetical protein, and locator 82, a terminase), or the absence of genes conserved between ponds C4 and TS. Additionally, seasonal variation between TSMay and TSOct is supported by different sets of conserved genes (e.g., TSOct-exclusive locators 294 and 364; TSMay-exclusive locators 438 and 316), which indicate viral recycling phenomena.

Dinucleotide frequency analyses. Principal component (PC) analysis based on the dinucleotide abundances (see Materials and Methods) indicated that 72.58% of the global variance could be explained by the three principal components. This value was comparable to others reported for data sets based on multiple environments (57). The three PCs have been used to draw the scatter plots shown in Fig. 4B and C. By using four PCs (82.67% variance explained) we were able to discriminate seven groups of dinucleotide usage (*k*1 to *k*7) in the data set delivered by *k* means and statistically supported by an ANOSIM test. These groups have been plotted with different symbols in Fig. 4B. For comparison, all sequences have also been plotted separately in Fig. 4C and labeled according to their sample of origin. The high overlap between samples might be caused by the low number of sequences analyzed from each pond and/or by the overrepresentation of certain viral genotypes that are common to the three ponds. However, there is an incipient separation of samples from low-to-medium salinity ponds (C4 and M1) from those from high-salinity ponds (TS). Indeed, *k*-means analysis supported this observation and allowed the following deductions: (i) groups *k*1, *k*5, *k*6, and *k*7 show the predominance of ponds C4 and M1; (ii) groups *k*3 and *k*4 show the predominance of pond TS; (iii) group *k*2 is a mixture of all four ponds; and (iv) group *k*3 shows a clear predominance of TSMay over TSOct.

When we compared dinucleotide usage with functional divergences (BLASTp analyses) among samples, we observed that sets of conserved ORFs typically belonged to sequences associated with certain dinucleotide usage groups, indicating a good correlation between the two methods. Thus, ORFs specific for sample. C4Oct were associated with dinucleotide groups *k*1, *k*5, and *k*7, while ORFs specific for M1Oct were mostly associated with groups *k*6, *k*1, and *k*2. The ORFs found only in crystallizer TS (including both May and October samples in combination) were mostly associated with *k*3, *k*4, and *k*6. Within these samples a slight temporal variation was also unveiled by dinucleotide analysis since conserved ORFs from TSOct were associated with *k*4 and *k*6, whereas conserved ORFs from TSMay were mostly associated with *k*3 and *k*4.

Thus, the different tools used to analyze the 4 Sfax metaviromes discussed here have provided similar results: all of them unveil changes in the viral assemblage along the salinity gradient and reveal a closer relationship between the viral communities in ponds M1 and TS than between those in ponds C4 and TS. Since the salinities of C4 and M1 are very close to each other, and much lower than that of TS, these results indicate that other factors besides the total salt concentration are structuring viral communities along salt gradients. In spite of the differences among all the hypersaline systems whose metaviromes have been analyzed to date, it is still possible to find some common traits among them, most likely reflecting the most important strategies for virus survival in close-to-saturation systems. However, the results pre-

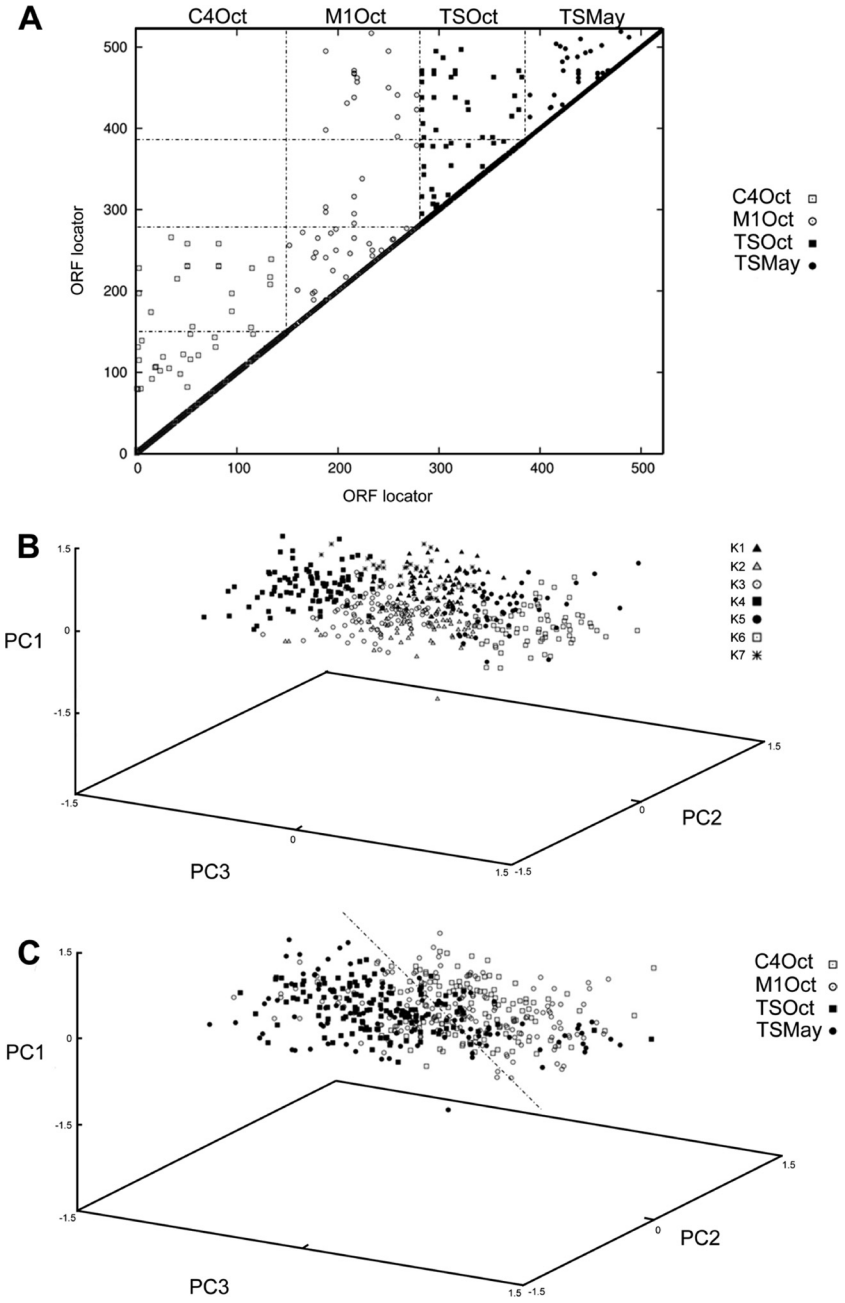

FIG 4 (A) All 520 ORFs from the distinct Sfax samples were assigned to a distinct numeric locator (from 1 to 520) and compared against themselves by using BLASTp. A diagonal straight line is given by self-matches. (B) Scatter plot of the three principal components based on the dinucleotide usage of Sfax salterns' complete sequence collection. (C) Symbols have been assigned according to a delivery into seven meaningful *k*-means groups and four distinct source samples.

sented here also point to a rather wide diversity among both the microbial and viral assemblages in hypersaline systems, reflecting the fact that every individual system has its particularities, and therefore caution should be exerted when general conclusions are being drawn from data retrieved from individual settings.

ACKNOWLEDGMENTS

This work was funded by projects A/018342/08 from the Spanish Agency for International Cooperation (AECI) and CGL2009-12651-C02-01 from the Spanish Ministry of Science and Innovation (MICINN), including FEDER funds.

We thank Inmaculada Meseguer and Manuel Martínez-García for technical assistance and encouraging discussions.

REFERENCES

1. **Amann RI, et al.** 1990. Combination of 16S rRNA-targeted oligonucleotide probes with flow cytometry for analyzing mixed microbial populations. Appl. Environ. Microbiol. **56**:1919–1925.
2. **Antón J, Llobet-Brossa E, Rodríguez-Valera F, Amann R.** 1999. Fluorescence *in situ* hybridization analysis of the prokaryotic community inhabiting crystallizer ponds. Environ. Microbiol. **1**:517–523.
3. **Antón J, et al.** 2008. Distribution, abundance and diversity of the extremely halophilic bacterium *Salinibacter ruber*. Saline Systems **4**:15–25.

4. **Antón J, Rosselló-Mora R, Rodríguez-Valera F, Amann R.** 2000. Extremely halophilic *Bacteria* in crystallizer ponds from solar salterns. Appl. Environ. Microbiol. **66**:3052–3057.
5. **Atanasova NS, Roine E, Oren A, Bamford DH, Oksanen HM.** 2011. Global network of specific virus-host interactions in hypersaline environments. Environ. Microbiol. **14**:426–440.
6. **Aziz RK, et al.** 2008. The RAST server rapid annotations using subsystems technology. BMC Genomics **9**:75.
7. **Baati H, Amdouni R, Gharsallah N, Sghir A, Ammar E.** 2010. Isolation and characterization of moderately halophilic bacteria from Tunisian solar saltern. Curr. Microbiol. **60**:157–161.
8. **Baati H, et al.** 2008. Prokaryotic diversity of a Tunisian multipond solar saltern. Extremophiles **12**:505–518.
9. **Baati H, Guermazi S, Gharsallah N, Sghir A, Ammar E.** 2010. Microbial community of salt crystals processed from Mediterranean seawater based on 16S rRNA analysis. Can. J. Microbiol. **56**:44–51.
10. **Baati H, Guermazi S, Gharsallah N, Sghir A, Ammar E.** 2010. Novel prokaryotic diversity in sediments of Tunisian multipond solar saltern. Res. Microbiol. **161**:573–582.
11. **Baati H, Jarboui R, Gharsallah N, Sghir A, Ammar E.** 2011. Molecular community analysis of magnesium-rich bittern brine recovered from a Tunisian solar saltern. Can. J. Microbiol. **57**:975–981.
12. **Bamford DH, Mindich L.** 1984. Characterization of the DNA-protein complex at the termini of the bacteriophage PRD1 genome. J. Virol. **50**: 309–315.
13. **Bath C, Cukalac T, Porter K, Dyall-Smith ML.** 2006. His1 and His2 are distantly related, spindle-shaped haloviruses belonging to the novel group, *Salterprovirus.* Virology **350**:228–239.
14. **Bath C, Dyall-Smith ML.** 1998. His1, an archaeal virus of the *Fuselloviridae* family that infects *Haloarcula hispanica.* J. Virol. **72**:9392–9395.
15. **Baxter BK, et al.** 2011. Haloviruses of Great Salt Lake: a model for understanding viral diversity, p 173–190. *In* Ventosa A, Oren A, Ma Y (ed), Halophiles and hypersaline environments. Springer, Berlin, Germany.
16. **Benlloch S, et al.** 2002. Prokaryotic genetic diversity throughout the salinity gradient of a coastal solar saltern. Environ. Microbiol. **4**:349–360.
17. **Bettarel Y, et al.** 2011. Ecological traits of planktonic viruses and prokaryotes along a full-salinity gradient. FEMS Microbiol. Ecol. **76**:360–372.
18. **Bodaker I, et al.** 2010. Comparative community genomics in the Dead Sea: an increasingly extreme environment. ISME J. **4**:399–407.
19. **Bolhuis H, et al.** 2006. The genome of the square archaeon *Haloquadratum walsbyi*: life at the limits of water activity. BMC Genomics **4**:169.
20. **Bolhuis H, Poele EM, Rodríguez-Valera F.** 2004. Isolation and cultivation of Walsby's square archaeon. Environ. Microbiol. **6**:349–360.
21. **Boujelben I, et al.** 2012. Spatial and seasonal prokaryotic community dynamics in ponds of increasing salinity of Sfax solar saltern in Tunisia. Antonie Van Leeuwenhoek **101**:845–857.
22. **Brum JF, Steward GF.** 2010. Morphological characterization of viruses in the stratified water column of alkaline, hypersaline Mono Lake. Microb. Ecol. **3**:636–643.
23. **Burns DG, Camakaris HM, Janssen PH, Dyall-Smith ML.** 2004. Cultivation of Walsby's square haloarchaeaon. FEMS Microbiol. Lett. **238**: 469–473.
24. **Cuadros-Orellana S, et al.** 2007. Genomic plasticity in prokaryotes: the case of the square haloarchaeon. ISME J. **1**:235–245.
25. **Díez B, Antón J, Guixa-Boixareu N, Pedrós-Alió C, Rodríguez-Valera F.** 2000. Pulsed-field gel electrophoresis analysis of virus assemblages present in a hypersaline environment. Int. Microbiol. **3**:159–164.
26. **Dinsdale EA, et al.** 2008. Functional metagenomic profiling of nine biomes. Nature **452**:629–632.
27. **Dyall-Smith M, Tang SL, Bath C.** 2003. Haloarchaeal viruses: how diverse are they? Res. Microbiol. **154**:309–313.
28. **García-Heredia I, et al.** 2012. Reconstructing viral genomes from the environment using fosmid clones: the case of haloviruses. PLoS One **7**:e33802. doi:10.1371/journal.pone.0038802.
29. **Ghai R, et al.** 2011. New abundant microbial groups in aquatic hypersaline environments. Sci. Rep. **1**:135. doi:10.1038/srep00135.
30. **Guixa-Boixareu N, Calderón-Paz JI, Heldal M, Bratbak G, Pedrós-Alió C.** 1996. Viral lysis and bacterivory as prokaryotic loss factors along a salinity gradient. Aquat. Microb. Ecol. **11**:215–227.
31. **Jiang S, Steward G, Jellison R, Chu W, Choi S.** 2004. Abundance, distribution and diversity of viruses in alkaline, hypersaline Mono Lake, California. Microb. Ecol. **47**:9–17.
32. **Legault BA, et al.** 2006. Environmental genomics of *Haloquadratum walsbyi* in a saltern crystallizer indicates a large pool of accessory genes in an otherwise coherent species. BMC Genomics **4**:171.
33. **Maturrano L, Santos F, Rosselló-Mora R, Antón J.** 2006. Microbial diversity in Maras salterns, a hypersaline environment in the Peruvian Andes. Appl. Environ. Microbiol. **72**:3887–3895.
34. **Mongodin EF, et al.** 2005. The genome of *Salinibacter ruber*: convergence and gene exchange among hyperhalophilic bacteria and archaea. Proc. Natl. Acad. Sci. U. S. A. **102**:18147–18152.
35. **Mutlu MB, et al.** 2008. Prokaryotic diversity in Tuz Lake, a hypersaline environment in inland Turkey. FEMS Microbiol. Ecol. **65**:474–483.
36. **Narasingarao P, et al.** 2012. *De novo* metagenomic assembly reveals abundant novel major lineage of *Archaea* in hypersaline microbial communities. ISME J. **6**:81–93.
37. **Nogales B, et al.** 2001. Combined use of 16S ribosomal DNA and 16S rRNA to study the bacterial community of polychlorinated biphenyl-polluted soil. Appl. Environ. Microbiol. **67**:1874–1884.
38. **Oren A, Bratbak G, Heldal M.** 1997. Occurrence of virus-like particles in the Dead Sea. Extremophiles **1**:143–149.
39. **Papke RT, Koening JE, Rodríguez-Valera F, Doolittle WF.** 2004. Frequent recombination in saltern population of *Halorubrum.* Science **306**: 1928–1029.
40. **Pasić L, et al.** 2009. Metagenomic islands of hyperhalophiles: the case of *Salinibacter ruber.* BMC Genomics **10**:570.
41. **Paul JH, Sullivan MB, Segal AM, Rohwer F.** 2002. Marine phage genomics. Comp. Biochem. Phys. **133**:463–476.
42. **Peña A, et al.** 2010. Fine-scale evolution: genomic, phenotypic and ecological differentiation in two coexisting *Salinibacter ruber* strains. ISME J. **4**:882–895.
43. **Ramette A.** 2007. Multivariate analysis in microbial ecology. FEMS Microbiol. Ecol. **62**:142–160.
44. **Rocha EP, Danchin A.** 2002. Base competition bias might result from competition for metabolic resources. Trends Genet. **16**:276–277.
45. **Rodríguez-Brito B, et al.** 2010. Viral and microbial community dynamics in four aquatic environments. ISME J. **4**:739–751.
46. **Roine E, Oksanen HM.** 2011. Viruses from the hypersaline environment, p 153–172. *In* Ventosa A, Oren A, Ma Y (ed), Halophiles and hypersaline environments. Springer, Berlin, Germany.
47. **Sabet S, Chu W, Jiang SC.** 2006. Isolation and genetic analysis of haloalkaliphilic bacteriophages in a North American soda lake. Microb. Ecol. **51**:543–554.
48. **Sandaa RA, Skjoldal EF, Bratbak G.** 2003. Virioplankton community structure along a salinity gradient in a solar saltern. Extremophiles **7**:347–351.
49. **Santos F, Antón J.** 2011. Viral metagenomics and the regulation of prokaryotic communities, p 33–44. *In* Marco D (ed), Metagenomics: current innovations and future trends. Caister Academic Press, Norfolk, VA.
50. **Santos F, et al.** 2007. Metagenomic approach to the study of halophages: the environmental halophage 1. Environ. Microbiol. **9**:1711–1723.
51. **Santos F, Yarza P, Briones C, Parro V, Antón J.** 2010. The metavirome of a hypersaline environment. Environ. Microbiol. **12**:2965–2976.
52. **Santos F, et al.** 2012. Culture-independent approaches for studying viruses from hypersaline environments. Appl. Environ. Microbiol. **78**:1635–1643.
53. **Sime-Ngando T, et al.** 2011. Diversity of virus-host systems in hypersaline Lake Retba, Senegal. Environ. Microbiol. **13**:1956–1972.
54. **Soler N, Marguet E, Verbavatz JM, Forterre P.** 2008. Virus-like vesicles and extracellular DNA produced by hyperthermophilic archaea of the order Thermococcales. Res. Microbiol. **159**:390–399.
55. **Trigui H, et al.** 2011. Characterization of heterotrophic prokaryote subgroups in the Sfax coastal solar salterns by combining flow cytometry cell sorting and phylogenetic analysis. Extremophiles **15**:347–358.
56. **Trigui H, Masmoudi S, Brochier-Armanet C, Maalej S, Dukan S.** 2 Jun 2011. Characterization of *Halorubrum sfaxense* sp. nov., a new halophilic archaeon isolated from the solar saltern of Sfax in Tunisia. Int. J. Microbiol. doi:10.1155/2011/240191.
57. **Willner D, Vega Thurber R, Rohwer F.** 2009. Metagenomic signatures of 86 microbial and viral metagenomes. Environ. Microbiol. **11**:1752–1766.