Viruses control natural microbial communities. Identification of virus-host pairs relies either on their cultivation or on metagenomics and tentative assignment based on genomic signatures. Both approaches have severe drawbacks when aiming to target such pairs within the uncultured majority. Here we present an unambiguous way to assign viruses to hosts that does not rely on any previous information about either of them nor requires their cultivation. First, genomic contents of individual cells present in an environmental sample are retrieved by means of single-cell genomic technologies. Then, individual cell genomes are hybridized against a set of individual viral genomes from the same sample, previously immobilized on a microarray. Infected cells will yield positive hybridization as they carry viral genomes, which can be then sequenced and characterized. Using this method, we pinpoint viruses infecting the ubiquitous hyperhalophilic *Nanohaloarchaeota*, included in the so-called ‘microbial dark matter’ (the uncultured fraction of the microbial world).
Microbes and their viruses constitute the most abundant and diverse group within the biosphere. The interactions among viruses and their microbial hosts have a central influence on biogeochemical cycles, on the control of numbers, diversity and evolution of microbes and even on human health. However, owing to limitations in the available techniques, there is a lack of knowledge about the interaction patterns between viruses and hosts in natural communities given that their description relies on the identification of viruses, hosts and viral–host ranges. Although this can be readily accomplished for isolated virus–host pairs, it is not technically feasible for uncultured viruses-hosts, which constitute the majority of microbes on the planet. Metagenomic analyses of cellular and viral fractions have provided valuable information on ecologically relevant virus–host interactions, such as in *Prochlorococcus*, from which previous genomic information was available. However, shotgun metagenomics does not allow for the unambiguous identification of individual virus–host pairs given the limitations of reconstructing individual viral genomes from short read data sets. Cloning of individual viral genomes from environmental samples into fosmids circumvents assembly limitations and allows for the tentative assignment of viruses to their hosts based on guanine/cytosine (GC) content and genomic signature comparisons of viral and host genomes. However, although this approach is very useful, it has technical limitations and, in addition, can only be used for assigning viruses to hosts from which genomic information is previously available. Moreover, in the absence of further proof, the assignment remains partly speculative even if complete genomes are recovered, given that there are well-known virus–host pairs with deviant genomic signatures.

Recently, Allers et al. have developed a PhageFISH method that detects both replicating and encapsidated (intracellular and extracellular) viral DNA, while simultaneously identifying and quantifying host cells during all stages of infection. For this purpose, probes targeting the viral genome and the small subunit (SSU) rRNA of the microbial host are used. This method offers great possibilities to study virus–microbe interactions in nature and can bridge the gap between metagenomics and direct quantification of viral–host pairs in natural samples. However, it still relies on the previous information of such pairs and cannot distinguish between very close viral genomes or microbial hosts with identical SSU rRNAs.

To the best of our knowledge, there are only two previous examples in which viruses have been unambiguously assigned to their uncultured hosts, both using single-cell technologies. In the first case, the sequencing of an individual marine protist revealed the presence of a single-stranded DNA virus infecting the host cell at the time of sampling. This finding illustrates the feasibility of describing virus–host systems by isolation of all individuals present in a sample followed by the sequencing of their genomes. However, this approach would require a considerable sequencing and downstream bioinformatics effort since without any previous information, both infected and uninfected hosts would have to be sequenced and analysed. In the second example, viruses infecting individual cells residing in the termite hindgut were detected by PCR with specific primers for a viral marker gene. However, there are no viral markers present in all viral genomes and thus, either previous information regarding the viruses present in the analysed sample must be available or the search has to be restricted to a group of viruses with known markers.

Here, to circumvent these limitations, we describe a method that unambiguously assigns viruses to uncultured hosts and does not rely on previous information of any of them nor requires their cultivation. This approach takes advantage of two high-throughput techniques that have proven very useful in microbial ecology: single-cell genomics and microarrays. We use this method to detect virus–host pairs in order to investigate virus–microbe infection networks in hypersaline environments. Hypersaline systems harbour the highest densities of viruses reported so far for aquatic samples, as well as a diverse assemblage of Bacteria and Archaea, that is often dominated by the square archaean *Haloquadratum walsbyi* and contains significant numbers of the recently described *Nanohaloarchaeota*. The *Nanohaloarchaeota*, along with four other major uncultured prokaryote groups within the unexplored ‘microbial dark matter’, form a monophyletic superphylum called DPANN, for which cultured representatives are not currently available. Here, we target viruses infecting *Nanohaloarchaeota* after proving the feasibility of our protocol with the appropriate controls.

### Results

#### Overview of the method.

In short (Fig. 1), individual cells presented in an environmental sample are separated by means of fluorescence-activated cell-sorting, lyed and their genomes amplified by multiple displacement amplification (MDA). In parallel, the viral fraction of the sample is concentrated and individual viral genomes are purified and cloned in fosmids, which are immobilized on a microarray (‘virochip’). Then, single-amplified genomes (SAGs) from individual cells are hybridized with the ‘virochip’. If a single cell is infected by a virus at the time of sampling, then its SAG would yield a hybridization signal with the ‘virochip’ (provided that the corresponding virus has been cloned). Further sequencing analysis of the SAG and the corresponding cloned viral genome would allow for the identification of both of them and confirm the presence in the sample of such virus–host pair. This approach can be used to look for viruses infecting specific groups of prokaryotes or even to target eukaryotic cells. For these purposes, targeted SAGs could be identified before hybridization by means of, for instance, SSU rRNA gene sequencing.

#### Microarray construction, SAG isolation and hybridization.

Before carrying out the experiments described below, control microarrays were constructed and hybridized as described in the Methods section and in Supplementary Fig. 1. Samples were taken from crystallizer pond CR30 of Bras del Port solar salters (Santa Pola, Spain), which has been extensively studied by a vast array of microbial ecology techniques. A 50-μl sample was used for single-cell sorting, generating a total of 936 SAGs that were screened for the 16S rRNA gene for identification. A total of 52 SAGs corresponded to *Nanohaloarchaeota* (Supplementary Fig. 2) and were used for further analysis. In parallel, individual haloviral genomes present in 21 of the same sample were purified, cloned in fosmids and used for the construction of the ‘virochip’. Fosmids were selected as cloning vectors because they are kept as single copy in the *Escherichia coli* cells, thus minimizing biases against unstable inserts and increasing the cloning efficiency of haloviral genomes. Besides, the optimum insert size for fosmids (that is, between 30 and 45 kb) corresponds to the size of most haloviral genomes detected in CR30 (ref. 12). The ‘virochip’, containing a total of 384 haloviral genomes, was hybridized with the pooled genomes of the 52 nanohaloarchaeal SAGs. As shown in Fig. 1, one of the fosmids (fosmid C23) yielded a strong hybridization signal.

#### Characterization of a nanohaloarchaeon–virus pair.

To ascertain which of the *Nanohaloarchaeota* was infected with the cloned virus, a new microarray (‘Nanohaloarchaeal chip’) was constructed with the 52 individual genomes (Fig. 1) and...
hybridized against the purified fosmid C23. Finally, the tentative virus-containing SAG AB578-D14 (henceforth named as D14) and the fosmid insert were sequenced (Supplementary Table 1). Sequencing indicated that the fosmid insert (of around 30 kb) contained a concatamer of three units of the viral genome. Concatamerization is frequently observed in pulsed field gel electrophoresis preparations of viral genomes with cohesive ends20. The size of the repeated unit, that is, the cloned haloviral genome, was 10,021 bp. As expected, the SAG contained the cloned viral genome (Fig. 2a), indicating that indeed the nanohaloarchaeon D14 was infected at the time of sampling with the cloned virus (we will refer to this virus as ‘nanohaloarchaeal virus 1’, NHV-1). This was further supported by sequencing data that showed that 99.8% of the recovered viral genome from host D14 was identical to the viral genome cloned in the fosmid and immobilized in the ‘virochip’ (Fig. 2a).

In spite of that high level of similarity between both viral genomes (one intracellular, contained within the host D14, and one extracellular, immobilized in the ‘virochip’), a 45-bp region located at the open-reading frame (ORF) 8 (Table 1) displayed 11 single-nucleotide polymorphisms (SNPs) resulting in three non-synonymous substitutions (Fig. 2 and Supplementary Fig. 3). These two genomes could thus correspond to two different virotypes that would be co-occurring in the viral assemblage at the time of sampling. In the case of viruses, a single SNP can impact severely on viral fitness, increasing for instance the adhesion to the host and infection21,22. Genome annotation (Table 1) showed, although NHV-1 lacked definable capsid genes, that most of the viral ORFs coded for hypothetical conserved proteins related to other uncultured haloviruses characterized in previous studies5–7. The lack of integrases (together with the absence of sequence reads overlapping viral and host genomes) suggests a potential lack of lysogenic cycle in NHV-1. In addition, the virus possessed a DNA primase and a viral terminase as well as a putative arsenical resistance repressor-like gene (ORF 7; named as asrR). Interestingly, the catalytic domain of this asrR-like was highly recruited in different geographically distant viral metagenomes (Fig. 2b) and also in a previously described cellular metagenome of the same crystallizer CR30 (ref. 19); identities 77–100%; Fig. 2b). Furthermore, similar asrR-like sequences were also found (Supplementary Table 2) in several prokaryote genome contigs from the hypersaline Lake Tyrrell23, where Nanohaloarchaea were predominant13. Arsenic compounds in hypersaline waters are highly prevalent and toxic for organisms, although prokaryotes have evolved different strategies to detoxify or exploit them24. Whether the viral asrR-like gene is indeed involved is arsenic metabolism or in other transcriptional regulatory pathways requires further attention. However, it resembled other asrR-like genes detected in prokaryote genomes and metagenomes, which suggests a trans-acting regulator with a conserved role in viral fitness. It is also worth noting that the highest recruited NHV-1 genomic region (intergenic space of ORFs 9 and 10) with the cellular metagenome from crystallizer CR30 (Fig. 2) was similar to sequences of the plasmid PL47 of Hqr. walsbyi of viral origin25 and a genomic region between the CRISPR 3 and the insertion element protein (IS2) of that very abundant square archaeon.

Small cell and genome sizes have been predicted as unifying features of the DPANN phyla14. The assembly of the host genome SAG D14 (~1 Mbp; Supplementary Table 3, Supplementary Fig. 4) was similar to that reported for its closest relative Candidatus Nanosalinarum sp. J07AB56 (ref. 13) and in the range of Nanohaloarchaeota group14. Genome comparison showed that although both Nanohaloarchaea shared a high 16S rRNA gene sequence identity (Supplementary Fig. 2), their genomic content was considerably different (Supplementary Fig. 5). However, in both genomes, most genes coded for hypothetical proteins, many of which were shared by both nanohaloarchaeab and present in the corresponding CR30 cellular metagenome19 (Supplementary Figs 6–9 and Supplementary Data 1).

As discussed above, GC content and oligonucleotide frequency signatures have been used to tentatively assign viruses to hosts in
natural assemblages without previous cultivation. In our case, NHV-1 and its nanohaloarchaeon D14 host possessed similar GC content (49% and 51%, respectively). Principal component analysis of dinucleotide frequencies (Fig. 3) revealed that NHV-1 genome grouped with the genomes of the nanohaloarchaeon D14 host and Candidatus Nanosalina (Fig. 3).

Table 1 | Genome annotation of NHV-1, which infects nanohaloarchaeon AB 578-D14.

<table>
<thead>
<tr>
<th>Putative ORFs</th>
<th>Nucleotide position</th>
<th>Predicted function</th>
<th>Best predicted hit (BLASTx)</th>
<th>Closest relative</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>52–342</td>
<td>Hypothetical archaeal protein</td>
<td>1e – 08/56/56/48</td>
<td>Halobacterium halobium J07HQW2</td>
</tr>
<tr>
<td>2</td>
<td>561–4595</td>
<td>DNA primase (Pfam 08275)</td>
<td>0/1134/88/53</td>
<td>Halophilic archaeon DL31</td>
</tr>
<tr>
<td>3</td>
<td>4,939–5,313</td>
<td>Hypothetical viral protein</td>
<td>9e – 65/203/98/82</td>
<td>Uncultured haloarchaeon (contig 157)</td>
</tr>
<tr>
<td>4</td>
<td>5,310–5,675</td>
<td>Putative NAD-dependent epimerase</td>
<td>0.49/38/44/33</td>
<td>Liesonia xyl</td>
</tr>
<tr>
<td>5</td>
<td>5,672–5,818</td>
<td>Unknown</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>6,319–6,546</td>
<td>Hypothetical viral protein</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>6,539–6,922</td>
<td>Hypothetical viral protein. Putative arsenical resistance protein repressor (NCBI-curated domain Cl00090; accession code cl17220)</td>
<td>6e – 36/126/100/96</td>
<td>Uncultured haloarchaeon (contig 152)</td>
</tr>
<tr>
<td>8</td>
<td>7,099–7,650</td>
<td>Hypothetical viral protein</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>7,881–8,174</td>
<td>Hypothetical viral protein</td>
<td>9e – 60/196/93/51</td>
<td>Environmental halophage eHP-16, 6, 12</td>
</tr>
<tr>
<td>10</td>
<td>8,550–10,019</td>
<td>Viral terminase (large subunit; Pfam03237)</td>
<td>1e – 45/153/100/79</td>
<td>Uncultured haloarchaeon (contig 42)</td>
</tr>
</tbody>
</table>

ORF, open-reading frame.
*Presence of a conserved domain is indicated in brackets (CDS-BLAST and SWISS-PROT BLAST).
†BLASTx data in non-redundant (nr) Genbank and Uniprot databases. Best hit displayed based on bit-score result.
and also with viral contigs from the above-mentioned CR30 viral metagomic library, which resulted as best BLAST hits for several ORFs of virus NVH-1 (Table 1). Similar results were obtained when tetranucleotide frequency signatures were considered for the analysis (Supplementary Fig. 9). Remarkably, the environmental haloviruses eHP-4 and eHP-25, previously assigned to Nanohaloarchaea hosts according to their codon usage, also clustered with NHV-1. Thus, our data validate the previous assignments of (uncultured) viruses to hosts in hypersaline systems based on genomic signature analyses.

Discussion
Overall, the method presented here can be accommodated within different workflows either to target a specific host group (as done here with the Nanohaloarchaea) following a wider metagenomic approach or as a tool for discovery of novel viral–host pairs without choosing any specific host. The feasibility of any such untargeted approach would depend, however, on the diversity of the system being analysed since, as is the case with metagenomics, more diverse systems would require greater efforts in terms of microarray construction and recovery of SAGs. Furthermore, although our approach has been used to target double-stranded DNA viral assemblages, modifications can be introduced to make it suitable for double-stranded DNA genomes with covalently bound terminal proteins, single-stranded DNA or even RNA viruses. However, our microarray hybridization data does not indicate contamination with free viruses, suggesting an insignificant contribution of ‘swarm’ detection during cell sorting. Nevertheless, a simple pre-enrichment step to remove most free viruses could be routinely implemented before sorting, if needed. In the case of unspecific attached viruses, our data does not support that hypothesis but rather confirms previous assignment data.

Here, we have provided a tool that can be used to analyse virus–microbe infection networks over any range of spatiotemporal scales and to draw valuable information on the evolution of virus genomes and the co-evolution with their hosts. This approach can be used to assign viruses to hosts even without previous information about either of them, which makes it suitable for the exploration of microbial dark matter.

Methods
Sample collection. A 50-μl water sample from CR30, a crystallizer pond of Bras del Port salterns (Santa Pola, Spain, 38°12′N, 0°36′W) taken in June 2011 was used for single-cell sorting. The salinity of the sample was 37.2% and harboured 1.74 × 10^7 cells per ml. Cell counting was performed after 4′,6-diamidino-2-phenylindole-dihydrochloride staining (1 μg ml^-1; Sigma) in an epifluorescence microscope (Leica, type DM4000B; Yasshaw Scientifics Inc.). Two liters of the same sample were used for viral DNA extraction.

Single-cell sorting and analyses. Replicate water samples for single-cell analyses were diluted to 10^4 cells per ml, cryopreserved with 6% glycerol betaine (Sigma-Aldrich) and shipped at −20°C to Single Cell Genomics Center (Maine, USA). For prokaryote detection, diluted subsamples (1 ml) were incubated for 10–120 min with SYTO-9 DNA stain (5 μM final concentration; Invitrogen). The high-nucleic acid cell fraction was targeted for fluorescence-activated cell sorting with a MoFlo (Beckman Coulter) flow cytometer using a 488-nm argon laser for excitation, a 70-μm nozzle orifice and a CyClone robotic arm for droplet deposition into microplates. The cytometer was triggered on side scatter. The single 1 drop mode was used here will sort a drop when it contains only one cell in its centre and no other detectable free particles in the drop, which makes the co-sorting of a cell with a virus very unlikely.

Other potential limitations of the technique are the co-sorting of the host cell with unspecific viruses attached to it or with free viruses placed in its shade, known in flow cytometry as ‘swarm’ detection. However, our microarray hybridization data does not indicate contamination with free viruses, suggesting an insignificant contribution of ‘swarm’ detection during cell sorting. Nevertheless, a simple pre-enrichment step to remove most free viruses could be routinely implemented before sorting, if needed. In the case of unspecific attached viruses, our data does not support that hypothesis but rather confirms previous assignment data.

Here, we have provided a tool that can be used to analyse virus–microbe infection networks over any range of spatiotemporal scales and to draw valuable information on the evolution of virus genomes and the co-evolution with their hosts. This approach can be used to assign viruses to hosts even without previous information about either of them, which makes it suitable for the exploration of microbial dark matter.

Figure 3 | Principal component analyses (PCAs) of the dinucleotide frequency signatures of viruses and hosts. Frequency of dinucleotide signatures was calculated and plotted in a PCA plot for reference halophilic prokaryote genomes, the pair NVH-1-Nanohaloarchaeon host DI4 and the viral contigs from environmental uncultured halophages (‘eHP-number’ and ‘Contig_number’) from García-Heredia et al. and Santos et al., respectively. Viral and host genomes are represented by spheres and stars, respectively. A total of 51 genomes were included in the PCA plot (see Methods), but for convenience, only selected viruses from same group clustering with its putative host were displayed. Axes F1 and F2 explain 76% of the variance between the analysed genomes.
was used for maximal sort purity, which ensures the absence of non-target particles within the target cell drop and the drops immediately surrounding the cell. Single-cell genomic DNA purification and sequencing of PCR products were performed at the Bigelow Laboratory Single Cell Genomics Center (https://scgc.bigelow.org/), as described in detail elsewhere. In brief, individual cells stained with SYTO-9 were sorted using a MoFlo (Beckman Coulter) flow cytometer using a 488-nm argon laser for excitation and a 488-nm argon laser for droplet detection into microplates. Single cells were then lysed using cold KOH and subjected to whole-genome MDA. The MDA products were diluted 50-fold in sterile TE buffer, and 0.5 μl aliquots of the dilute MDA products served as templates in 5 μl real-time PCR screens for 16S rRNA genes. The total 16S rRNA gene sequences obtained from nanohaloarchaea SAGs (~500 bp sequence length) were carefully edited and then aligned using the SILVA aligner (http://www.arb-silva.de/). Only sequences displaying ≥80% of the alignment quality score in the SILVA aligner were considered for the analysis. The alignment was imported into the Geneious R6.1 bioinformatic package (Biomatters Ltd.) and phylogenetic analysis based on neighboring joining and maximum likelihood (1,000 bootstrap replications) was performed.

**Viral DNA purification and fosmid library.** Two liters of the CR30 water sample were centrifuged at 30,000 g (30 min, 20°C, Avanti J-30, Beckman). The supernatant was then tangentially filtered through a 30,000 molecular-weight-cutoff Vivaflo filter cassette (Sartorius Stedim Biotech) and concentrated to 20 ml. Water was extruded without removal of the sheared DNA from the extruded cells on nylon filters (GV Dupareo, Millipore) and viruses were then ultrafiltrated at 186,000 g for 2 h at 20°C in an Optima MAX-XP Ultrafiltron with the TLA-55 rotor (Beckman Coulter) and re-suspended in 0.5 ml of 25% sea water (SW), containing (in grams per liter): NaBr, 0.65; NaHCO3, 0.17; KCl, 5; CaCl2, 0.72; MgSO4, 2H2O, 49.49; MgCl2, 6H2O, 34.57; NaCl, 195. Halovirus concentrates were mixed with equal volumes of 1.6% low-melting-point agarose (Pronadisa), dispensed into 100-μl moulds, and allowed to solidify at 4°C. Agarose plugs were incubated for 90 min with 5 μl of Turbo DNA-free kit (Ambion) to digest dissolved DNA (according to the manufacturer’s protocol, 2–3 μl of Turbo DNase digest up to 500 μg ml−1 of DNA in a final reaction volume of 50 μl, using 1 μl of Turbo DNase). Agarose plugs were then incubated overnight at 50°C in TB (10 mM Tris–HCl pH 8.0, 1 mM EDTA pH 8.0) and treated by sonication with a 1.5 inch enzyme per 0.1 g of melted mixture) and DNA was then purified using MicroconYM-100 centrifugal filters (Millipore). DNA quality was checked by electrophoresis and its concentration determined by Nanodrop (Thermo Fisher Scientific Inc.). Around 1.5 μg of viral DNA were end-repaired and cloned into pCZ2FOS vector using the CopyControl HTP Fosmid Library Production Kit (Epicentre) according to the manufacturer’s recommendations. The EPI300-T1R strain of *E. coli* (Epicentre) was used as plating strain. The Fosmid Library Production Kit packages optimally into lambda phage heads giving yields with sizes ranging from 30 and 45 kb. All the clones were transferred to four 96-well plates with 500 μl of LB medium containing 0.2% maltose, 12.5 μg ml−1 of chloramphenicol and the inducer CopyControl Induction Solution (Epicentre). For the purification of the fosmid C23, the corresponding clone was grown in 5 ml LB medium containing 0.2% maltose, 12.5 μg ml−1 of chloramphenicol and 0.5% glycerol, and stored at −80°C until use.

**Construction of the ‘control microarray’ and hybridization controls.** For the experimental controls, genomic DNA from the strain M8 of the extremely halophilic bacterium *Salinibacter ruber* and a fosmid containing the virus *PhiM8-CR4* (which infects *M8*, Villamor et al., unpublished) as the insert, were used as the ‘probes’ to be spotted in the ‘control microarray’ (Supplementary Fig. 1a). Probe DNAs were dried using a centrifugal evaporator and re-suspended in microspotting Solution Plus 1X (Arrayit Corp.) to yield five different concentrations: 10, 50, 100, 250 and 500 ng μl−1. Spotting was performed with the MicroGrid-TAS II Arrayer (Genomics Solutions) at 22°C and 50–50% relative humidity on epoxy-substrates (Microarray slides) according to the manufacturer’s protocol. The 25 ml of the DNAs used as probes was spotted five times. On the other hand, total DNA samples from cultures of: (i) non-infected M8 strain and (ii) strain M8 infected with virus *PhiM8-CR4* were used as the ‘targets’. DNA from *S. ruber* cultures was extracted with the DNeasy Blood and Tissue Kit (QIAGEN) according to the manufacturer’s recommendations. Hybridization of the total DNA with the virus *PhiM8-CR4* DNA corresponded to the control clone was grown in 5 ml of terrific broth (TB) medium containing 0.2% maltose, 12.5 μg ml−1 of chloramphenicol and the CopyControl Induction Solution (Epicentre) and the fosmid was extracted using the FosmidMAX DNA Purification Kit (Epicentre). For the labelling, 2.8 μg were treated as described above. Hybridization (see below) yielded a unique signal with the spots corresponding to the nanohaloarchaea SAGs identified as *Nanohaloarchaea*. For the labelling, 2.8 μg were treated as described above. Hybridization (see below) yielded a unique signal with the spots corresponding to the nanohaloarchaea SAGs identified as *Nanohaloarchaea*. For the labelling, 2.8 μg were treated as described above. Hybridization (see below) yielded a unique signal with the spots corresponding to the nanohaloarchaea SAGs identified as *Nanohaloarchaea*. For the labelling, 2.8 μg were treated as described above. Hybridization (see below) yielded a unique signal with the spots corresponding to the nanohaloarchaea SAGs identified as *Nanohaloarchaea*.
technology at the Genomic and Bioinformatic Services of the Autonomous University of Barcelona. Paired-end read libraries for Illumina sequencing were prepared using the Nextera Extender Sample Preparation protocol, in which 50 ng of DNA template was simultaneously fragmented and tagged with sequencing adapters in a single step. Then, sequencing was performed in a MiSeq Benchtop Sequencer (500 cycles run) generating 21.3 and 1.58 Gb for the fosmid and the SAG D14, respectively (Supplementary Table 2). Assembly of viral reads was performed with the VelvetOptimiser (http://bioinformatics.net.au/software.velvetoptimiser.shtml) was initially used to test the optimal parameters for our data that were further used for the assembly. Finally, the new assembler SPAdes33 was specifically designed for sequencing data from SAGs, which outperformed the assembler VELVET-SC for single-cell genomes35, was used with the recommended parameters (‘spades.py –sc –k 21,33,55,77,99,127’). Assembly data clearly demonstrate that SPAdes was the best strategy for the genome assembly of the nanohaloarchaeon D14 single cell (Supplementary Table 3).

Genome annotation. ORFs of viral genome were detected by using the heuristic approach with GenMark Hidden Markov model 36 and annotated by using a combination of BLAST, UNI-PROT BLAST and Conserved-Domain BLAST for the search of functional domains in viral proteins 37. SAG D14 genome was annotated by using the SEED subsystem publicly available at RAST server 38 and the bioinformatics resources of the US Department of Energy Joint Genome Institute (http://www.jgi.doe.gov/) with the pipeline annotation Prodigal 39.

Metagenome recruitment and genome analysis. The basic approach of Rusch et al. 40 was used to estimate recruitment of viral genome and nanohaloarchaeon D14 in previously published prokaryote and viral metagenomes from the same crystalizer pond (CR30) and other similar hypersaline systems. Metagenomic data from crystalizer CR30 studies here were obtained from Santos et al. and Ghai et al.32 Deep-Illumina sequencing data from Lake Tyrrell was from Emerson et al. and Podell et al.23, whereas genomic data of the previously characterized Nanohaloarchaea were from Narasingarao et al. BLAST + v2.2.22 was used to recruit metagenome sequences to reference genome using the following parameters: -evalue = 0.0001 -perc_identity 60 -outfmt 6. Then, BLAST hit output was parsed according to the cut-off identities and nucleotide position of the viral genome. Normalization of the recruited metagenome fraction was performed according to metagenome sizes. Genome comparison of nanohaloarchaeon D14 to Candidatus Nanosalinarum J07AB56 (ref. 13) was performed with stand-alone BLAST version 2.2.22 + in a similar manner but using a threshold coverage and identity values of 80% in BLASTn searches in order to consider reciprocal hits. Nucleotide alignments and whole-genome alignment were performed with ClustalW and Mauve aligner42 implemented in Geneious bioinformatics package (http://www.geneious.com) with the pipeline annotation Prodigal39.

References

Acknowledgements
This work was supported by the projects CGL2012-39627-C03-01 (to J.A.) and AYA2011-24803 (to V.P.) of the Spanish Ministry of Science and Innovation, which are co-financed with FEDER support from the European Union. We thank the staff of the Bras del Port salterns for their help with sampling and also Judith Villamor for providing us with the virus ΦM8-CR4. Finally, we thank Anton Korobeynikov for his technical support with the SPAdes programme.

Author contributions
M.M.-G., F.S. and J.A. designed and performed the experiments, analysed data and wrote the paper. M.M.-P. performed experiments and V.P. wrote the paper.

Additional information
Accession codes: 16S rRNA gene sequences from SAGs are deposited in the Genbank Nucleotide database with accession codes KF771589 to KF771641. Viral and SAG D14 genome sequences are deposited in the Genbank Nucleotide database with accession codes SAMN02369554 and SAMN02369553. The complete archael genome shotgun project is deposited in the GenBank Nucleotide database with accession code AYG10000000 (the version described in this paper is version AYG10000000). SAG D14 genome annotation is deposited in the RAST (Rapid Annotation using Subsystem Technology) server with accession code 6666666.48919. Raw Illumina data for single-cell and fosmid sequencing are deposited in the NCBI Bioproject database with accession code PRJNA222265.

Supplementary Information accompanies this paper at http://www.nature.com/naturecommunications

Competing financial interests: The authors declare no competing financial interests.

Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/

How to cite this article: Martinez-García, M. et al. Unveiling viral–host interactions within the ‘microbial dark matter’. Nat. Commun. 5:4542 doi: 10.1038/ncomms5542 (2014).