nature microbiology

Article

https://doi.org/10.1038/s41564-025-02149-7

Viruses and virus satellites of haloarchaea and their nanosized DPANN symbionts reveal intricate nested interactions

Received: 13 February 2025

Accepted: 15 September 2025

Published online: 17 October 2025

Check for updates

Yifan Zhou[1,2], Ana Gutiérrez-Preciado[3], Ying Liu[1], David Moreira[3], Michail M. Yakimov[4], Purificación López-García[3] & Mart Krupovic[1]

Nested symbioses, including hyperparasitism in which parasites exploit other parasites within a host, are common in nature. However, such nested interactions remain poorly studied in archaea. Here we characterize this phenomenon in ultra-small archaea of the candidate phylum Nanohaloarchaeota, members of the DPANN superphylum (named after its first representative phyla: Diapherotrites, Parvarchaeota, Aenigmarchaeota, Nanoarchaeota and Nanohaloarchaeota) that form obligate interactions with halophilic archaea of the class Halobacteria. We reconstructed the viromes from geothermally influenced salt lakes in the Danakil Depression, Ethiopia, and find that nanohaloarchaea and haloarchaea are both associated with head-tailed, tailless icosahedral, pleomorphic and spindle-shaped viruses, representing 16 different families. These viruses exhibit convergent adaptation to hypersaline environments, encode diverse auxiliary metabolic genes and exchange genes horizontally with each other. We further characterize plasmid-derived satellites that independently evolved to parasitize spindle-shaped viruses of haloarchaea and nanohaloarchaea, revealing another layer of nested symbiosis. Collectively, our findings highlight the complexity of virus–host and virus–virus interactions in hypersaline environments.

Outside of laboratory settings, microorganisms rarely, if ever, exist in isolation and instead establish complex networks of symbiotic interactions with other cellular organisms and viruses, which range from mutualism to parasitism. Remarkably, it has been shown that a considerable fraction of the archaeal diversity is represented by ubiquitous, ultra-small cells with highly reduced genomes, namely, members of the kingdom Nanobdellati[1], also known as the DPANN superphylum (named after its first representative phyla: Diapherotrites, Parvarchaeota, Aenigmarchaeota, Nanoarchaeota and Nanohaloarchaeota)[2–4]. Like all cellular organisms, DPANN archaea are infected by viruses[5–7], but the diversity and impact of these viruses on their hosts remain poorly understood. It remains unclear whether viruses of DPANN archaea underwent genome reduction to adapt to their highly reduced hosts; whether the DPANN virome has evolved through spillover from the viromes of their respective hosts; whether horizontal gene exchange is common between viruses infecting symbionts and their hosts; and whether viruses of symbionts can boost the functional potential of their hosts through auxiliary metabolic genes (AMGs). Furthermore, it is becoming increasingly recognized that prokaryotic viruses themselves can be targeted by mobile genetic elements, such as virus satellites[8–10]; however, whether such additional layers of nested symbiosis extend to viruses of symbiotic hosts is unknown. To address some of these

[1]Institut Pasteur, Université Paris Cité, CNRS UMR6047, Cell Biology and Virology of Archaea Unit, Paris, France. [2]Collège Doctoral, Sorbonne Université, Paris, France. [3]Ecologie Systématique Evolution, CNRS, Université Paris-Saclay, AgroParisTech, Gif-sur-Yvette, France. [4]Extreme Microbiology, Biotechnology and Astrobiology Group, Institute of Polar Sciences, ISP-CNR, Messina, Italy. e-mail: mart.krupovic@pasteur.fr

questions, in this study, we focus on one of the lineages of DPANN archaea, Nanohaloarchaeota, associated with halophilic archaea of the class Halobacteria.

Haloarchaea and nanohaloarchaea are usually the dominant archaeal lineages in hypersaline environments[11–14]. Culture-based experiments have shown that nanohaloarchaea possess small cells, approximately 250–500 nm in diameter, and genome sizes of around 1 Mb (refs. 15–17). The genome reduction is typically accompanied by the loss of metabolic and biosynthetic pathways of basic biomolecules, such as nucleotides, amino acids and lipids. Accordingly, nanohaloarchaea develop obligate symbiotic, sometimes mutualistic[17], relationships with autonomously growing haloarchaea[15–18].

Viruses can affect the composition of microbial communities through host cell lysis and rewire the metabolic capacity of their hosts by supplying diverse AMGs[19,20]. Hypersaline ecosystems represent some of the most virus-rich environments on the planet[21,22]. The majority of haloarchaeal viruses (HVs) comprise head-tailed viruses that are classified into 12 families within the class *Caudoviricetes*[23]. However, non-tailed morphotypes are also common in hypersaline environments and include viruses with pleomorphic (*Pleolipoviridae*), tailless icosahedral (*Sphaerolipoviridae* and *Simuloviridae*) and spindle-shaped (*Halspiviridae*) virions[24–26]. Although there is a growing appreciation of the diversity and evolution of HVs, little is known about nanohaloarchaeal viruses (NHVs)[5–7,27–29].

Here, we use metagenomics to explore the diversity of HVs and NHVs in geothermally influenced salt lakes in the Danakil Depression, Ethiopia, some of the most extreme ecosystems known, dominated by haloarchaea and nanohaloarchaea[30,31]. Our results fill the knowledge gap on the diversity and evolution of HVs and NHVs, and illuminate the complexity of nested virus–host and virus–virus interactions.

Results

Virus-like particles in samples from Danakil salt lakes

To explore the viromes of consortia consisting of haloarchaea and symbiotic nanohaloarchaea, we focused on salt lake samples from the Danakil Depression, Ethiopia, a region with multiple hypersaline systems of diverse hydrochemistry[30,31] (Fig. 1a). Haloarchaea and nanohaloarchaea were dominant in samples from Lake Assale or Karum (samples Ass and 9Ass collected during different years), a cave reservoir at the Dallol proto-volcano salt canyons (9Gt) and two of the Western Canyon Lakes (WCL2 and WCL3)[31] (Fig. 1b). Transmission electron microscopy analysis of enrichment cultures from 9Ass, 9Gt and WCL3 samples revealed the presence of morphologically diverse virus-like particles (Fig. 1c and Supplementary Text).

Haloarchaea and nanohaloarchaea are associated with distinct viromes

Both haloarchaeal and nanohaloarchaeal species carry clustered regularly interspaced short palindromic repeats (CRISPR) arrays[18,32]. Thus, we assembled a database of CRISPR sequences specific to either haloarchaea or nanohaloarchaea (Methods). A total of 64 haloarchaea-specific and 11 nanohaloarchaea-specific CRISPR sequences were identified in the Danakil metagenomes, which together with related CRISPR arrays from the Genome Taxonomy Database (GTDB) and Earth's Microbiomes (EM) dataset[33], formed 6 distinct clusters (Fig. 1d and Extended Data Fig. 1). The haloarchaea-specific and nanohaloarchaea-specific CRISPR arrays contained 60,722 and 1,179 spacers, respectively (Fig. 1e).

Next, five Danakil salt lake metagenomes were analysed for the presence of viral contigs using geNomad[34] and VirSorter2[35], which collectively yielded 2,085 non-redundant viral contigs (≥5 kb). The vConTACT gene-sharing network analysis showed that the majority (n = 1,641) of Danakil viral sequences (including 950 singletons) represent novel viruses (Extended Data Fig. 2). A total of 209 and 44 viral contigs (<2% of the whole virome) were targeted (≥30 bp exact match) by haloarchaeal and nanohaloarchaeal spacers, respectively (Fig. 1f), with none being targeted by both nanohaloarchaeal and haloarchaeal spacers. Notably, although some archaea use CRISPR systems to fend against DPANN ectosymbionts[36], none of the retrieved haloarchaeal spacers targeted nanohaloarchaeal genomes.

We assembled 11 and 7 circular (complete) genomes of Danakil HVs and NHVs, respectively (Supplementary Table 1). Based on the presence of signature virion morphogenesis genes, these viruses could be broadly assigned to four groups: (1) head-tailed viruses (class *Caudoviricetes*; eight HVs and three NHVs), (2) tailless icosahedral viruses (realm *Singelaviria*; one HV and one NHV), (3) pleomorphic viruses (family *Pleolipoviridae*; one HV and one NHV) and (4) spindle-shaped viruses (one HV and two NHVs). Hereinafter, we refer to Danakil haloarchaeal and nanohaloarchaeal tailed viruses as DHTVs and DNTVs; to tailless icosahedral viruses as DHIVs and DNIVs; to pleomorphic viruses as DHPVs and DNPVs; and to spindle-shaped viruses as DHSVs and DNSVs, respectively. In addition to viral genomes, we identified nine spacer-targeted small plasmid-like circular elements associated with haloarchaea (see below).

Analysis of the distribution of HVs and NHVs across the five Danakil metagenomes revealed that the viromes varied across the studied ecosystems, with the patterns for the cave reservoir (9Gt), the Lake Assale samples (Ass, 9Ass) and the Western Canyon Lakes (WCL2, WCL3) being clearly distinct (Fig. 1g). Overall, all lakes contained both HVs and NHVs, but certain viruses were exclusive to particular lakes (Supplementary Text).

Deep divergence of haloarchaeal and nanohaloarchaeal viromes

Searches against the IMG/VR and National Center for Biotechnology Information (NCBI) databases for the presence of virus genomes related to NHVs identified in the Danakil metagenomes (Methods) yielded additional complete or near-complete genomes related to DNTVs (n = 7), DNIV-1 (n = 3), DNPV-1 (n = 1) and DNSVs (n = 1; Extended Data Figs. 3–6 and Supplementary Table 2). Host assignments for all complete genomes were done using a combination of CRISPR targeting and blastp searches for proteins matching the reference cellular or viral proteomes (Supplementary Tables 1 and 2 and Methods).

Comparison of the inferred proteomes of HVs and NHVs showed that the corresponding virus groups are only distantly related to each other (Fig. 2), arguing against recent horizontal virus transfer between haloarchaeal and nanohaloarchaeal hosts, instead suggesting deep divergence of the corresponding viromes. Members of the same family exhibited similar genome lengths and GC% content (Fig. 2). Notably, across the discovered virus groups, HVs and NHVs displayed comparable genome lengths, indicating that, unlike their hosts, NHVs did not evolve by genome reduction.

Our results suggest that HVs and NHVs can be classified into 16 families (Supplementary Table 1 and Supplementary Text), 12 of which have not been described previously, emphasizing sparse sampling of this part of the archaeal virome.

Amino acid composition of HVs and NHVs reflects adaptation to hypersaline environments

To thrive under close-to-saturating salt concentrations, nanohaloarchaea and haloarchaea have independently adopted the 'salt-in' strategy[12]. Their intracellular K^+ concentration roughly matches that of salt cations in the surrounding environment, which triggers a massive adaptive acidification of proteomes[27,31,37]. Accordingly, we examined the amino acid usage (AAU) patterns and proteome-wide isoelectric point (pI) distributions of HVs and NHVs, and compared them with those of the respective predicted hosts (Extended Data Fig. 7). Haloarchaeal and nanohaloarchaeal proteomes are enriched in aspartic (D) and glutamate (E) residues, which contribute to the overall acidic pI[31]. Alanine is also specifically enriched in haloarchaeal proteomes, presumably to decrease hydrophobic forces in the high salt environments. The same

https://doi.org/10.1038/s41564-025-02149-7

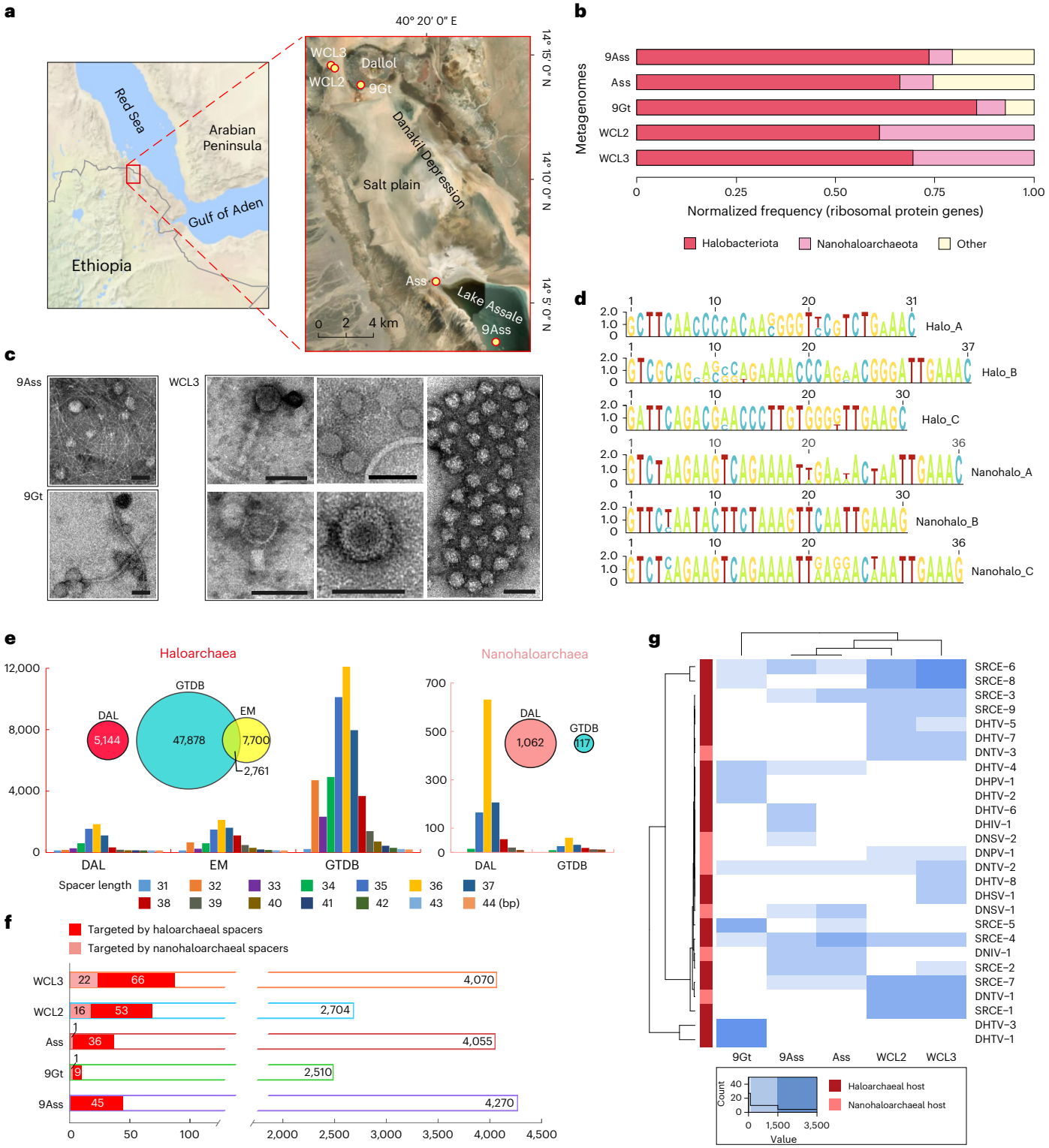


Fig. 1 | Microbial and viral diversity in the salt lakes of the north Danakil Depression, Ethiopia. a, Locations of the five sampling sites in the Danakil Depression. **b**, Microbial community composition in the environmental samples collected from the five salt lakes. The composition is inferred from the previously described normalized frequency of selected ribosomal proteins[31]. **c**, Electron micrographs of virus-like particles observed in the enrichment cultures established from environmental samples collected from the 9Ass, 9Gt and WCL3 salt lakes. The virus-like particles were stained with 2% (w/v) uranyl acetate. Scale bars, 100 nm. **d**, The consensus sequences of the six CRISPR variants specific to Halobacteriota and Nanohaloarchaeota. **e**, The total numbers and length distributions of spacers extracted from the Danakil Depression (DAL) metagenomes, EM Project and GTDB. **f**, The results of CRISPR spacer targeting. The open bars represent the total number of viral sequences identified in each metagenome, whereas the filled bars represent the number of viral sequences targeted by Halo (red) or Nanohalo (pink) spacers. **g**, Heatmap showing the distribution and relative abundance of HVs and NHVs and plasmid-like elements (mean coverage, rows) in Danakil salt lakes (columns). The intensity of the blue colour represents relative abundance (the distribution of abundance values across all virus–site combinations is shown in the inset below the panel). The putative type of archaeal host is indicated in red or pink (left column).

https://doi.org/10.1038/s41564-025-02149-7

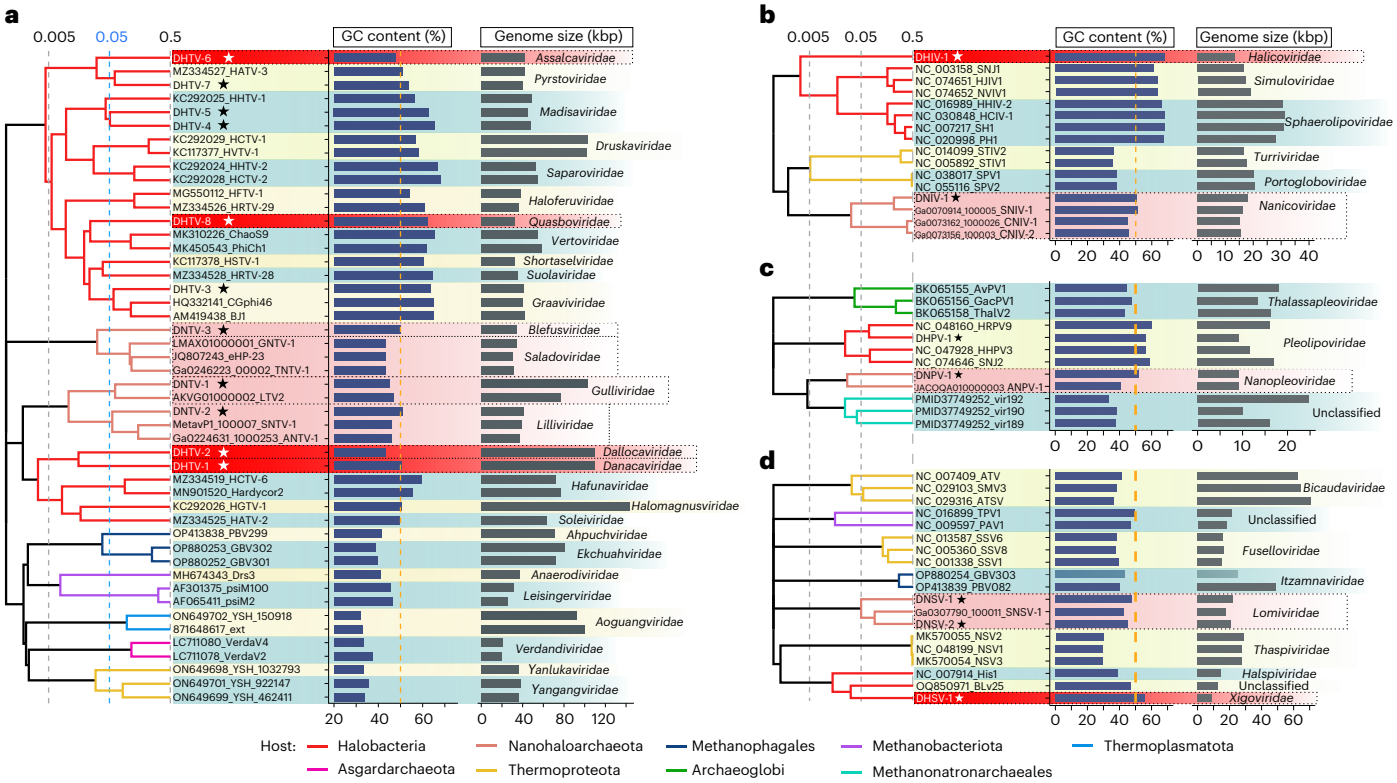


Fig. 2 | Genome-wide proteomic trees of the four virus groups. The bars next to each genome indicate the GC% content and genome length, respectively. **a**, Head-tailed viruses (*Caudoviricetes*). **b**, Tailless icosahedral viruses. **c**, Pleomorphic viruses. **d**, Spindle-shaped viruses. Viruses from the Danakil Depression are indicated with stars. Proposed families are highlighted in red (HVs) or pink (NHVs). The proteomic trees are based on all-versus-all proteomic similarity matrix and are mid-point rooted. Branch lengths are log-scaled. Branches are coloured based on the viral host groups, with the key provided at the bottom of the figure.

AAU and pI patterns were found in the viruses predicted to infect the two groups of halophilic archaea (Extended Data Fig. 7a,b).

Adaptation of HVs and NHVs to their hosts is also evident at the level of nucleotide composition. Analysis of the GC% content showed that nanohaloarchaea (42.9%; $n = 38$) have considerably lower GC% content compared with that of haloarchaea (64.1%; $n = 749$), and the same was true for head-tailed viruses predicted to infect the 2 host groups (58.2% ($n = 62$) versus 46.1% ($n = 9$) for DHTVs and DNTVs, respectively; Extended Data Fig. 7c). This observation is consistent with the results of the previous analyses showing that the GC% content of DNA viruses and their hosts tend to correlate[38].

The observed patterns suggest that HVs and NHVs have adapted their AAU and GC% contents to mirror those of their hosts, presumably to optimize the compatibility with the intracellular pools of amino acids and nucleotides. The distinct GC and AAU patterns further suggest that haloarchaeal and nanohaloarchaeal viromes have diverged early on and coevolved with their hosts for an extended period.

Virus AMGs can enhance the hosts' metabolic potential

Viruses with larger genomes often carry AMGs, which can supplement and enhance the functional potential of the host cells[39,40]. Although nanohaloarchaea have dramatically reduced genomes, their viruses do not appear to follow the same trend. Indeed, HVs and NHVs displayed similar genome sizes, with the largest viral genomes in both host groups exceeding 100 kb (DHTV-1 with 109.7 kb versus DNTV-1 with 103.4 kb).

Most of the functionally annotated genes in the HVs and NHVs with smaller genomes (<50 kb) encode the core functions involved in virion morphogenesis and genome replication, whereas viruses with large genomes encode more diverse functions. For instance, DHTV-1 and DHTV-2 encode nearly complete replisomes and a number of AMGs (Supplementary Text). By contrast, the similarly sized genome of DNTV-1 lacks the expansive set of genome replication proteins and encodes only the small subunit of the sliding clamp loader and MCM helicase (Fig. 3 and Supplementary Table 3). The latter protein is commonly found as a sole replication protein in archaeal viruses with smaller genomes[41]. Consistently, in the proteomic tree (Fig. 2a), DNTV-1 clusters with smaller DNTVs, suggesting that this virus underwent genome expansion independently from tailed viruses of halophilic and marine archaea.

Although nanohaloarchaea largely depend on haloarchaea for the supply of amino acids and nucleotides, similar to DHTV-1 and DHTV-2, DNTV-1 encodes several proteins implicated in nucleotide metabolism, including deoxyuridine triphosphatase (open reading frame (ORF)88), thymidine kinase (ORF91) and the large subunit of the ribonucleoside-diphosphate reductase (RNR; ORF77), which uses thioredoxin (ORF93) as a cofactor to convert ribonucleotides into deoxyribonucleotides (Fig. 3 and Supplementary Table 3). Notably, phage-encoded RNRs have been shown to be expressed upon infection, leading to a dramatic increase in the intracellular concentration of deoxynucleotides[42]. A relative of DNTV-1, LTV2, a virus found in hypersaline Lake Tyrrell, Australia[28], also encodes an RNR (Extended Data Fig. 4a and Supplementary Table 3). Although DNTV-1 and LTV2 are assigned to the same family (Fig. 2a and Extended Data Fig. 5), their RNRs are not orthologous but have been independently acquired horizontally from nanohaloarchaeal hosts (Fig. 4a and Supplementary Text).

In addition, DNTV-1 carries several genes implicated in translation and amino acid metabolism. In particular, it encodes 29 tRNAs for 19 different amino acids (Supplementary Table 4), including 5 tRNAs for the negatively charged Asp and Glu residues that are enriched in the nanohaloarchaeal and NHV proteomes (Extended Data Fig. 7a).

https://doi.org/10.1038/s41564-025-02149-7

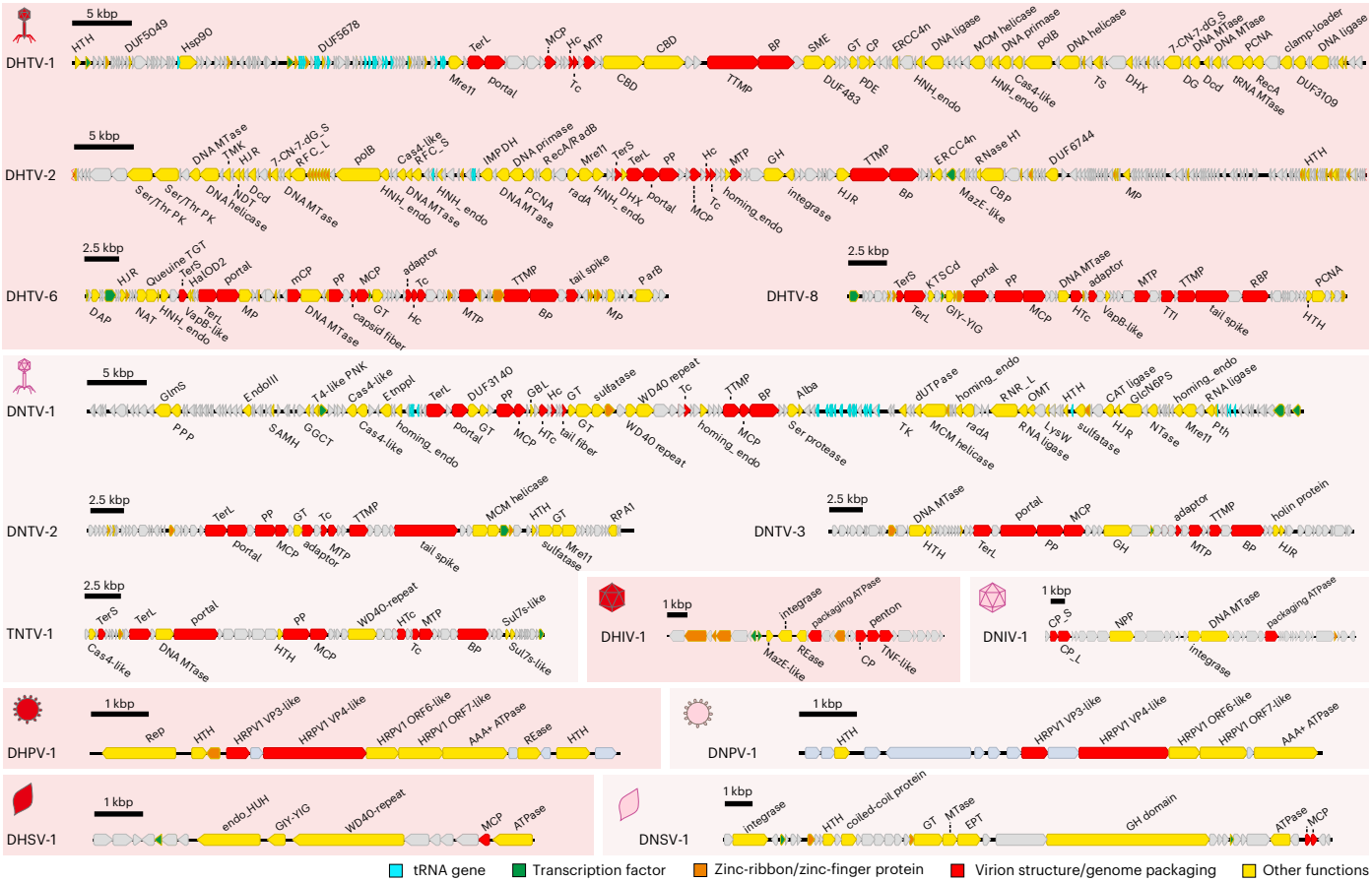


Fig. 3 | Genome maps of HVs and NHVs from the Danakil Depression. For head-tailed viruses, only representatives of the previously undescribed proposed families are shown. Each family is represented by a single genome. HTH, helix-turn-helix; DUF, domain of unknown function; Hsp90, ATP-dependent chaperone Hsp90; Mre11, nuclease Mre11; TerS and TerL, small and large subunits of the terminase, respectively; MCP, major capsid protein; Hc, head-closure protein; Tc, tail-completion protein; MTP, major tail protein; CBD, carbohydrate binding domain; TTMP, tail tape measure protein; BP, baseplate protein; SME, sulfatase-maturating enzyme; GT, glycosyltransferase; PDE, phosphodiesterase; CP, cysteine protease; ERCC4n, ERCC4-type nuclease; HNH_endo, HNH family endonuclease; polB, family B DNA polymerase; DHX, DEAD/DEAH-box helicase; 7-CN-7-dG_S, 7-cyano-7-deazaguanine synthase; DNA MTase, DNA methyltransferase; DG, DNA glycosylase; Dcd, dCTP deaminase; transfer RNA MTase, tRNA methyltransferase; PCNA, DNA polymerase sliding clamp; RecA, RecA ATPase; Ser/Thr PK, serine/threonine protein kinase; TS, thymidylate synthase; TMK, thymidylate kinase; TK, thymidine kinase; HJR, Holliday junction resolvase; NDT, nucleoside 2-deoxyribosyltransferase; RCF_L, replication factor C large subunit; RCF_S, replication factor C small subunit; IMPDH, inosine-5′-monophosphate dehydrogenase; RecA/RadB-like, RecA/RadB-like recombination protein; RadA, DNA repair and recombination protein RadA; PP, prohead protease; homing_endo, homing endonuclease; GH, glycoside hydrolase; RNase H1, ribonuclease HI; MazE-like, MazE-like antitoxin; CBP, cobalamin biosynthesis protein; MP, metalloprotease; DAP, DNA annealing protein; NAT, *N*-acetyltransferase; Queuine TGT, queuine tRNA-guanine transglycosylase; VapB-like, VapB-like antitoxin; HalOD2, haloarchaeal output domain 2; mCP, minor capsid protein; ParB, ParB family DNA-binding protein; KTSCd, lysine (K) tRNA synthetase C-terminal domain; GIY-YIG, GIY-YIG family nuclease; HTc, head-tail connector protein; TTI, tail tube initiator; RBP, receptor binding protein; GlmS, glucosamine 6-phosphate synthase; PPP, phosphotyrosine protein phosphatase; EndoIII, endonuclease III; SAMH, *S*-adenosyl-L-methionine hydrolase; GGCT, gamma-glutamyl cyclotransferase; T4-like PNK, T4-like polynucleotide kinase; Etnppl, ethanolamine phosphate transferase; GBL, galactose-binding lectin; WD40 repeat, WD40 repeat-containing protein; Alba, DNA/RNA-binding protein Alba; RNR_L, ribonucleotide reductase large subunit; OMT, *O*-methyltransferase; LysW, amino group carrier protein LysW; GlcN6PS, glucosamine 6-phosphate synthetase; CAT ligase, carboxylate-amine/thiol ligase; Pth, peptidyl-tRNA hydrolase; PPA1, archaeal replication protein A1; Sul7s-like, Sul7s-like DNA binding protein; CP_S, capsid protein (small); CP_L, capsid protein (large); REase, restriction endonuclease; TNF-like, TNF-like jelly-roll domain protein; NPP, nucleotide pyrophosphatase/phosphodiesterase; Rep, rolling circle replication endonuclease; HRPV1 VP3-like, HRPV1 VP3-like matrix protein; HRPV1 VP4-like, HRPV1 VP4-like membrane fusion protein; endo_HUH, endonuclease of the HUH superfamily; EPT, ethanolamine phosphate transferase.

Other translation-related genes include archaeal-type peptidyl-tRNA hydrolase (ORF50; Figs. 3 and 4b), two copies of phage T4-like RNA ligase 2 (ORF51 and ORF76; Fig. 3 and Supplementary Table 3), and a GGCT/AIG2 family gamma-glutamyl cyclotransferase (ORF158; Fig. 4b), an enzyme catabolizing L-γ-glutamyl-L-α-amino acid dipeptides formed during the γ-glutamyl cycle to 5-oxo-L-proline and a free amino acid[43].

Finally, DNTV-1 ORF62 encodes a glucosamine 6-phosphate synthase (GlmS), which catalyses conversion of Gln and fructose-6-phosphate to Glu and glucosamine-6-phosphate[44]. The protein contains two domains (Fig. 4b), an N-terminal glutaminase domain, which catalyses the hydrolysis of Gln to Glu and ammonia, and a C-terminal synthase domain, catalysing amination of fructose-6-phosphate. In addition, DNTV-1 ORF23 encodes a stand-alone GlmS glutaminase domain[45] (Fig. 4b), suggesting that it participates in Gln deamination but not carbohydrate amination. This activity could increase the pool of intracellular Glu, one of the most prevalent amino acids in the proteomes of NHVs and nanohaloarchaea (Extended Data Fig. 7a). In a phylogenetic tree, DNTV-1 GlmS branches within a clade containing nanohaloarchaeal homologues and is closest to proteins encoded by members of the family Asbonarchaeaceae (Fig. 4c), a basal lineage of Nanohaloarchaeota recently described in Danakil salt lakes[12]. In addition, four

https://doi.org/10.1038/s41564-025-02149-7

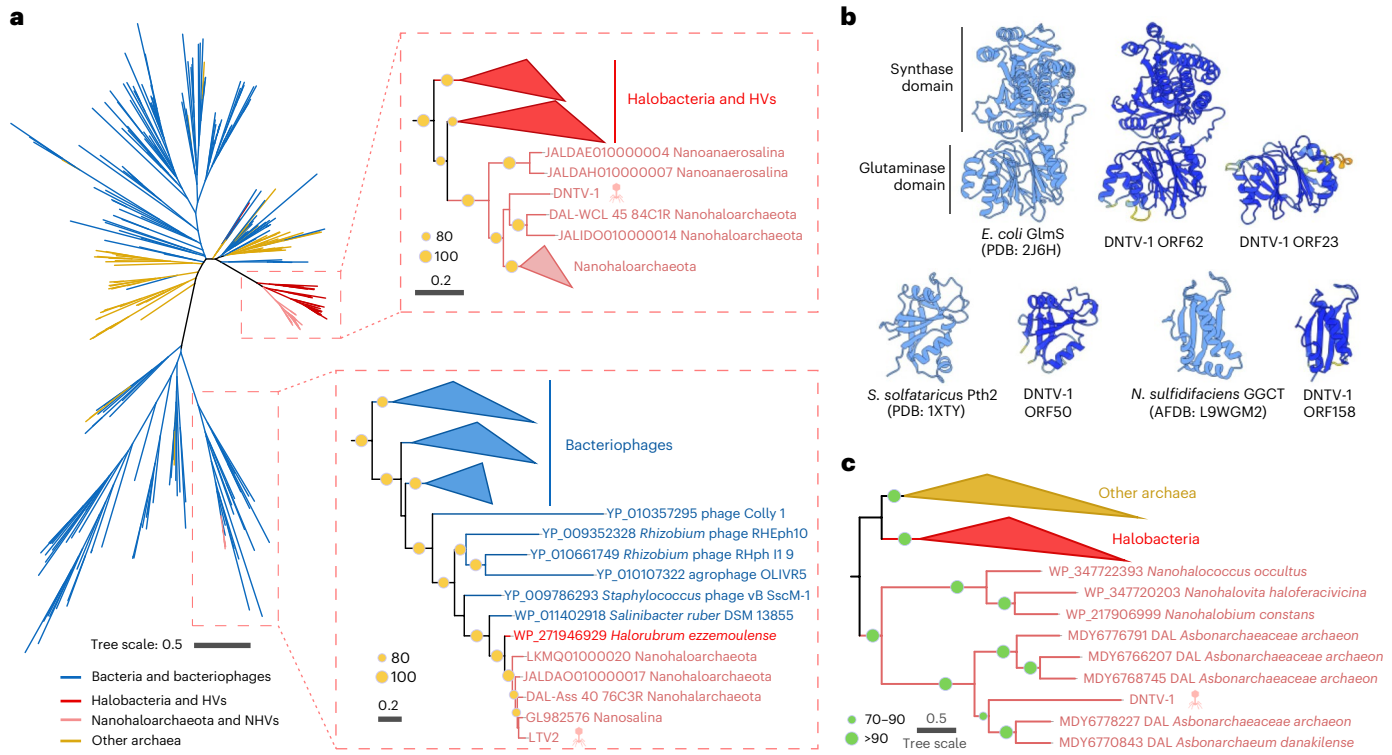


Fig. 4 | Phylogenetic analysis and structural modelling of AMGs encoded by DNTV-1. a, Unrooted maximum likelihood phylogeny of RNR large subunit. Zoom-ins on the branches including DNTV-1 and LTV2 are shown on the right. Branches with bootstrap support values higher than 80% are indicated with yellow circles. **b**, Predicted structural models of DNTV-1 encoded GlmS (ORF62), stand-alone GlmS glutaminase domain (ORF23), archaeal-type peptidyl-tRNA hydrolase (Pth2; ORF50), GGCT/AIG2 family gamma-glutamyl cyclotransferase (ORF158). *E. coli*, *Escherichia coli* (Protein Data Bank (PDB): 2J6H); *S. solfataricus*, *Saccharolobus solfataricus* (PDB: 1XTY); *N. sulfidifaciens*, *Natronorubrum sulfidifaciens* (AlphaFold Protein Structure Database (AFDB): L9WGM2). **c**, Mid-point rooted maximum likelihood phylogeny of GlmS. Branches with bootstrap support values higher than 70% are indicated with green circles.

other DNTV-1 proteins yielded best blastp hits to Asbonarchaeaceae (Supplementary Table 5). Notably, DNTV-2 was binned into the Asbonarchaeaceae metagenome-assembled genome (DAL-9Gt_90_72C1R, 72% completeness and 1% contamination). These results suggest that DNTV-1 and DNTV-2, and potentially other NHVs, infect Asbonarchaeaceae members.

The presence of a greater diversity of AMGs in the DNTV-1 genome compared with DHTVs with larger genomes may reflect the metabolic limitations of the nanohaloarchaeal host, providing a fitness advantage to viruses that encode AMGs. However, inferences based on the gene content analysis described above will have to be experimentally validated using transcriptomics and metabolomics once the corresponding virus–host systems are isolated.

Horizontal gene transfer between HVs and NHVs

Although our data suggest that HVs and NHVs did not switch hosts across the phylum boundary, blastp analysis of the viral proteomes pinpointed 18 NHV genes that appear to have been horizontally transferred between the two virus groups (Supplementary Table 6). Most of these genes encode major morphogenesis proteins, such as major capsid proteins (MCPs) and genome packaging ATPases, and homing endonucleases. To formally assess the potential gene transfers between HVs and NHVs, we focused on the morphogenetic module genes of head-tailed and tailless icosahedral viruses. Phylogenetic analysis suggests that a gene block encoding the MCP and portal protein—but not the large terminase subunit (TerL)—in Halorubrum tailed virus 28 was horizontally acquired from NHVs (Fig. 5a–d, Extended Data Fig. 8 and Supplementary Text). Conversely, the TerL of DNTV-3 appears to have been acquired from HVs, suggesting an HV-to-NHV transfer (Fig. 5c,d).

Exchange of the structural modules appears not to be restricted to head-tailed HVs and NHVs. The four tailless icosahedral NHVs share a set of nine genes and can be classified within the same family (Fig. 2b). However, the two single jelly-roll MCPs of DNIV-1 and its close relative SNIV-1 show high sequence similarity (49.4% and 30.7% identity) with the homologues encoded by icosahedral HVs of the family *Simuloviridae* (Supplementary Table 6). By contrast, the two other DNIV-1-like NHVs, CNIV-1 and CNIV-2, encode distinct MCP variants that are not recognizably similar at the sequence level to the simulovirus or DNIV-1 MCPs (Fig. 5g). Consistently, phylogenetic analysis placed the DNIV-1 and SNIV-1 MCPs deep within the HV clade (Fig. 5e). By contrast, the FtsK-like genome packaging ATPases of the tailless icosahedral HVs and NHVs formed distinct clades corresponding to their host associations (Fig. 5f). These results suggest that the ancestral MCPs, resembling those of CNIV-1 and CNIV-2, were replaced in DNIV-1 and SNIV-1 by homologues from HVs. Collectively, these results suggest common bidirectional horizontal exchange of homologous genes between HVs and NHVs, probably facilitated by cytoplasmic exchange between the haloarchaeal and nanohaloarchaeal hosts[46,47] and driven by the need to escape the host defences (Supplementary Text).

Spindle-shaped viruses of haloarchaea and nanohaloarchaea are associated with circular satellite molecules

The signature morphogenetic module of spindle-shaped viruses includes the hydrophobic α-helical MCP and the AAA+ ATPase[48,49]. Assembly of the Danakil metagenomes yielded nine complete genomes of small circular elements (2.7–9.0 kb, median 3.6 kb) that encode one or both of the two spindle-shaped virus signature proteins. Seven of these elements, spindle-shaped virus-related circular elements (SRCEs), were targeted by haloarchaeal CRISPR spacers (Supplementary Table 7).

https://doi.org/10.1038/s41564-025-02149-7

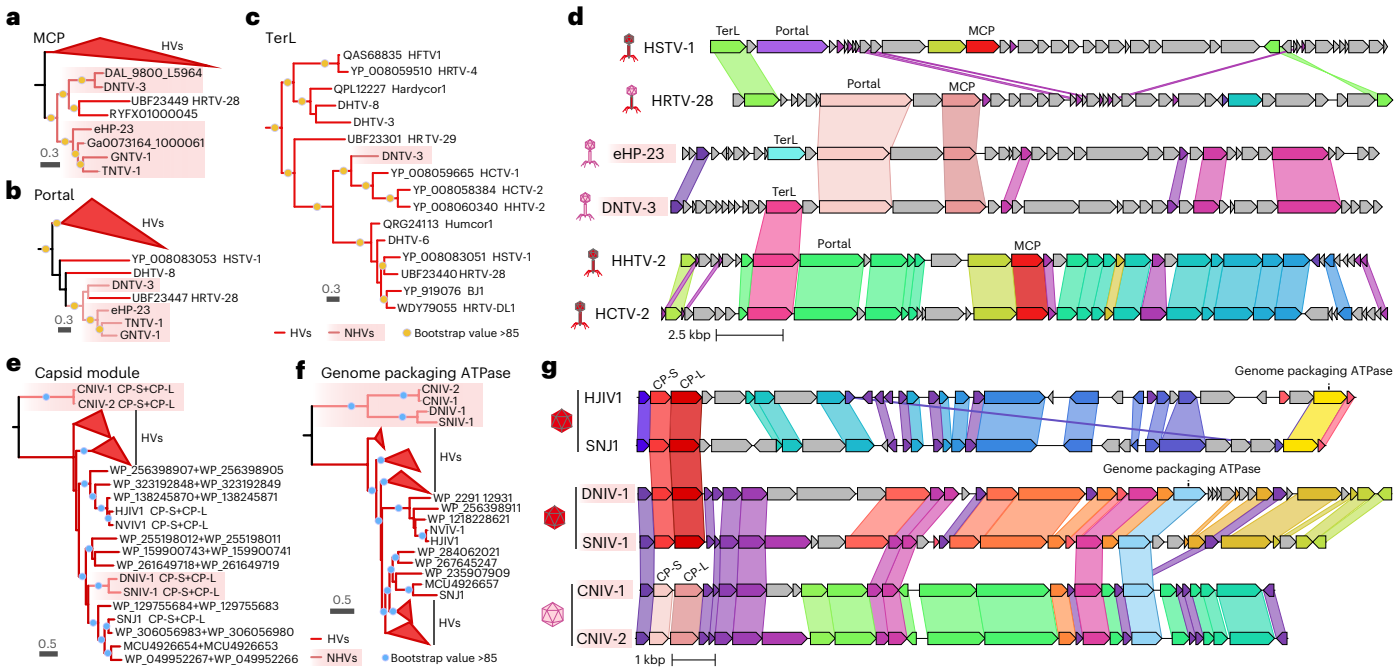


Fig. 5 | Horizontal gene transfer between HVs and NHVs. a–**c**, Maximum likelihood phylogenies of the MCPs (**a**), large subunit of the terminase (TerL, **b**) and portal (**c**) of head-tailed HVs and NHVs. Branches with bootstrap support values higher than 85% are indicated with yellow circles. Expanded trees are shown in Extended Data Fig. 8. **d**, Comparison of the (partial) genome maps of HSTV-1 (KC117378: 1–25,344 bp), HRTV-28 (MZ334528: 1–24,686 bp), ePH23 (JQ807243: 1–13,391 bp, 18,015–30,840 bp), DNTV-3 (7,207–33,827 bp), HHTV-2 (KC292024: 12,973–38,639 bp) and HCTV-2 (KC292028: 6,250–33,408 bp). Homologous genes (>25% identity) are highlighted using the same colour and linked via shadings. **e**,**f**, Maximum likelihood phylogenies of the concatenated sequences of the large and small MCPs (CP_L and CP_S, respectively; **e**) and genome packaging ATPase (**f**) of tailless icosahedral HVs and NHVs. Branches with bootstrap support values higher than 85% are indicated with blue circles. **g**, Comparison of the genome maps of HJIV1, SNJ1, DNIV-1, SNIV-1, CNIV-1 and CNIV-2. Homologous genes (>25% identity) are highlighted using the same colour and linked via shadings. Relevant genes are labelled.

Consistently, phylogenetic analysis of the SRCE AAA+ ATPases showed that they form a monophyletic clade with the homologues encoded by spindle-shaped HVs (Fig. 6a), including DHSV-1, suggesting that these SRCEs replicate in haloarchaea. Further searches against the IMG/VR database queried with the ATPases and MCPs of DNSVs led to the identification of four SRCE-like elements, which in the phylogenetic analysis of the AAA+ ATPases clustered with DNSV-1 and DNSV-2, suggesting nanohaloarchaeal hosts. Thus, the AAA+ ATPase phylogeny supports the early divergence of haloarchaeal and nanohaloarchaeal spindle-shaped viruses and suggests that both groups of viruses are associated with SRCEs.

Besides the two viral proteins, SRCEs encode plasmid-like replication proteins, namely, rolling circle replication initiation endonucleases (Reps) or archaeo-eukaryotic primases-helicases belonging to several different families (Fig. 6b,c). The combination of the spindle-shaped virus signature proteins with plasmid replication proteins found in SRCEs resembles the features of the satellite plasmid pSSVx of the thermoacidophilic archaeon *Saccharolobus islandicus*. Similar to SRCEs, pSSVx combines a replication module closely related to that of *Saccharolobus* plasmids with homologues of two fusellovirus SSV2 proteins, including the AAA+ ATPase[9]. pSSVx is maintained in cells as a plasmid but upon coinfection with SSV2, pSSVx is encapsidated into smaller spindle-shaped particles and spreads horizontally in the population[9]. We hypothesize that SRCEs represent satellites associated with haloarchaeal and nanohaloarchaeal spindle-shaped viruses. Notably, similar SRCE-like agents were previously documented in the hypersaline Lake Reba, Senegal[50], but were not considered to represent virus satellites. Sequence similarity networks and the Rep phylogenetic analysis showed that SRCEs formed five different clusters, including previously described MSBL1, pGRB1-like and pC2A-like Rep families, as well as two additional clusters consisting of haloarchaeal and nanohaloarchaeal SRCEs and plasmids (Fig. 6c).

By hijacking the morphogenetic module of the helper viruses, satellite nucleic acids can have adverse effects on the propagation of the coinfecting viruses[51]. We thus assessed the distribution and replicative potential of the SRCEs and spindle-shaped viruses, including those with incomplete genomes, in the Danakil metagenomes by comparing their relative abundance (normalized sequencing depths; Extended Data Fig. 9, Supplementary Table 1 and Supplementary Text). The relative abundance of SRCEs was generally higher than that of the spindle-shaped viruses, with SRCE-6 and SRCE-8 being up to 155-fold and 94-fold more abundant than their putative helper virus DHSV-2 (3,557× and 2,163× versus 23×, respectively; Fig. 1g and Supplementary Table 1). Note that SRCE-6, SRCE-8 and DHSV-2 are targeted by spacers from the same CRISPR arrays and are all found in the Western Canyon Lakes (Supplementary Tables 1, 7 and 8). Furthermore, SRCEs displayed broader distribution across Danakil lakes compared with most DHSVs (Extended Data Fig. 9). Given the dominance of the SRCE genomes, it is possible that, during coinfection, the majority of viral structural proteins could be hijacked by SRCEs, thereby negatively impacting the production of the helper virus. Notably, seven different CRISPR arrays contained spacers against more than one SRCE (Fig. 6b and Supplementary Table 8), suggesting that distinct SRCE variants may coexist in the same host cell. Collectively, our results point to the existence of satellite agents that take advantage of both haloarchaeal and nanohaloarchaeal spindle-shaped viruses, uncovering a hidden layer of symbiosis in hypersaline environments.

Discussion

Here, we applied metagenomics to explore the viromes associated with extreme halophilic archaea in geothermal hypersaline lakes of the Danakil Depression close to life-limiting conditions. We found that both haloarchaea and nanohaloarchaea are associated with the same four groups of viruses, namely, head-tailed viruses, tailless icosahedral

https://doi.org/10.1038/s41564-025-02149-7

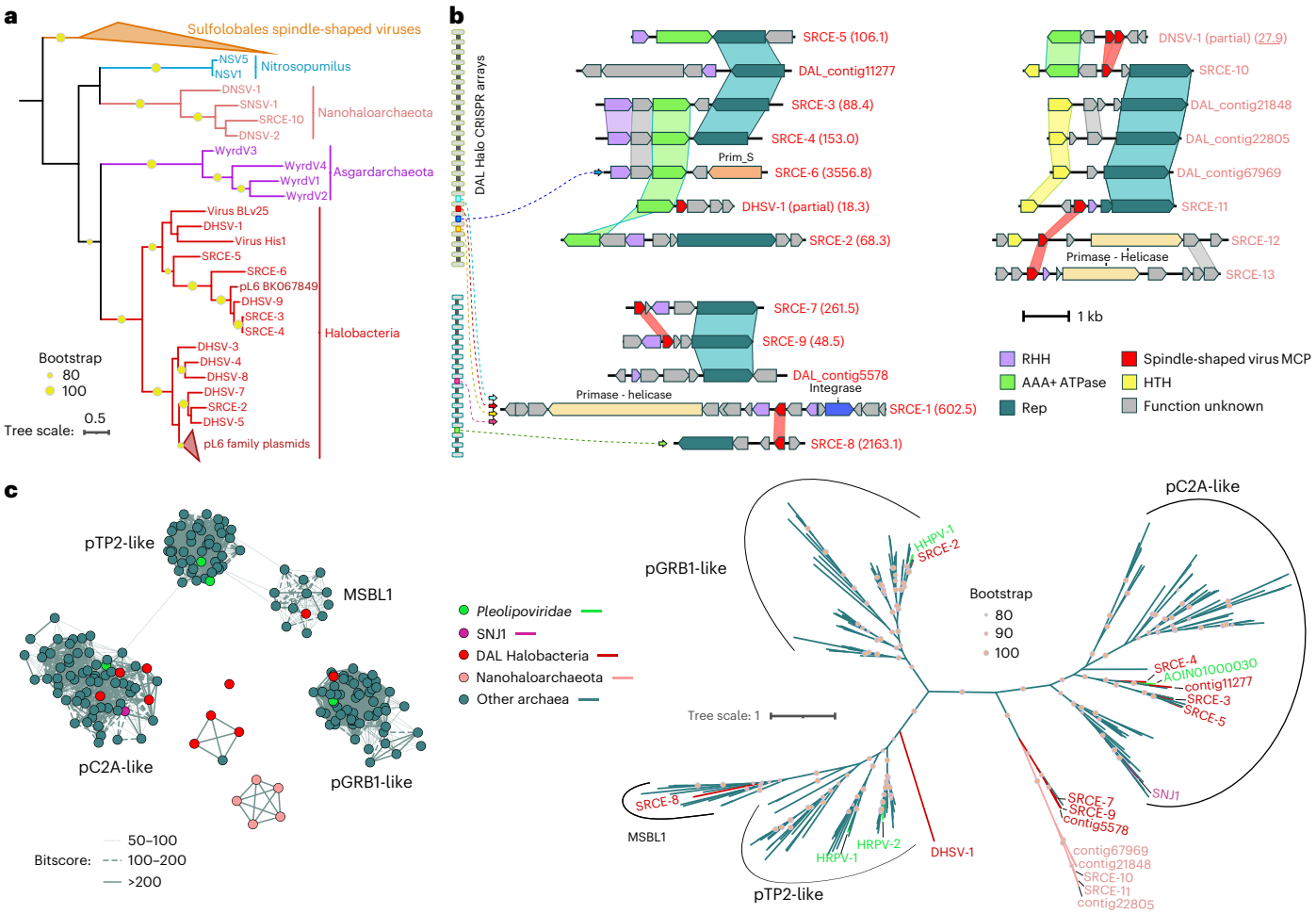


Fig. 6 | Putative satellites associated with spindle-shaped HVs and NHVs. **a**, Maximum likelihood phylogeny of the AAA+ ATPases conserved in spindle-shaped HVs and NHVs and SRCEs. Branches with bootstrap support values higher than 85% are indicated with yellow circles. **b**, Genome maps of SRCEs and related plasmids. Sequencing depth (when available) of each genome is shown in parentheses following the genome name. Functionally equivalent genes are shown in the same colour. Genes with over 20% identity are linked via shadings. Spacers targeting the SRCEs are shown on the left of the figure as coloured bars within the CRISPR arrays, with arrows linking the spacers with the corresponding SCREs. **c**, Left: similarity networks of Rep sequences. Each node represents a Rep sequence, and lines linking the nodes represent the degree of sequence similarity based on BLASTp bitscores. Right: maximum likelihood phylogeny of Rep proteins encoded by diverse archaeal mobile genetic elements. Branches with bootstrap support values higher than 80% are indicated with pink circles.

viruses, pleomorphic viruses and spindle-shaped viruses. These viruses represent 16 families, 12 of which were not described previously, highlighting the sparse sampling of the archaeal virosphere and the distinctiveness of the Danakil viromes. We describe the first spindle-shaped, pleomorphic and tailless icosahedral viruses associated with Nanohaloarchaeota (Supplementary Text). None of the NHVs could be assigned to families containing HVs, and none of the viral genomes was targeted by both haloarchaeal and nanohaloarchaeal spacers, suggesting that the two archaeal lineages are associated with distinct viromes despite living in close contact. Nevertheless, we detected cases of bidirectional horizontal exchange of the morphogenesis modules between HVs and NHVs, probably driven by diverse host defence systems (Supplementary Text).

Some of the large NHVs, such as DNTV-1, carry multiple AMGs predicted to be involved in nucleotide and amino acid metabolism, as well as translation, and may have played an important role in horizontal AMG transfer between viruses and cells, as exemplified by the case of RNR. Nevertheless, it is unlikely that NHVs could complete the replication cycle in nanohaloarchaea that are not attached to the haloarchaeal cells. Thus, the dependence of nanohaloarchaea on haloarchaea for basic biomolecules, such as nucleotides and amino acids, renders NHVs secondarily dependent on haloarchaea besides their direct host nanohaloarchaea. In this context, the discovery of potential satellites, termed SRCEs, associated with haloarchaeal and nanohaloarchaeal spindle-shaped viruses is most striking. NHVs can be considered hyperparasites in their own right, whereas NHV satellites add yet another layer to this nested symbiotic system (Extended Data Fig. 10).

Virus satellites have been known in eukaryotes and bacteria for decades[52–55]. By contrast, in the domain Archaea, only two satellites, pSSVi and pSSVx, had been described previously, both associated with spindle-shaped viruses of the family *Fuselloviridae* infecting thermoacidophilic archaea[9,10]. A common defining property of satellites is that they hijack the virion morphogenesis machinery of helper viruses for horizontal spread in the population. Haloarchaeal and nanohaloarchaeal SRCEs combine plasmid-derived replication modules with incomplete virion morphogenetic modules. In particular, they can encode either a characteristic MCP[49], an AAA+ ATPase or both. Notably, pSSVx associated with fuselloviruses encodes a homologous ATPase, but not the MCP. Thus, in this respect, SRCEs resemble cf-PICI, the recently discovered bacterial satellites that encode the MCPs[54], but appropriate the phage tails for spread.

Although the SRCEs described herein were retrieved from global metagenomes, they are related to enigmatic elements previously detected in the virus-sized fraction of hypersaline samples filtered

through a 0.2-µm filter and pretreated with DNase I to remove free DNA[50], suggesting that SRCEs are bona fide satellites capable of extracellular spread. Consistent with their broad distribution in hypersaline environments, we identified SRCE-like satellites in the IMG/VR database. Furthermore, SRCEs appear to be related to the globally distributed PL6-family plasmids, in which AAA+ ATPase is one of the core genes[56,57]. Indeed, in our phylogenetic analysis, ATPases of the PL6-family plasmids were nested among those encoded by SRCEs (Fig. 6a). Thus, we propose that PL6-family plasmids represent satellites that can hijack the structural module of haloarchaeal spindle-shaped viruses upon coinfection.

We hypothesize that virus satellites could be retained by archaeal cells as a defence strategy against infections by bona fide viruses, as has been proposed for certain eukaryotic satellite viruses[58]. The relative abundance of SRCEs in metagenomes was consistently higher than that of the spindle-shaped viruses, suggesting a negative impact of the satellites on the helper viruses. Upon infection with the cognate helper viruses, the small genomes of SRCEs could replicate faster than the larger genomes of the helper viruses. Alternatively, SRCEs could be present in high copy already in the non-infected cells, in anticipation of virus infection. The SRCE genomes would effectively compete with the helper viruses for the pool of structural proteins required for particle assembly. Such competitions would probably reduce the production of functional virions, limiting the spread of the bona fide viruses, protecting the host population. Therefore, we suggest that SRCEs, or more generally, virus satellites, may act as stabilizers of microbial communities. By curtailing viral outbreaks, they could help preventing the collapse of host populations thus maintaining microbial diversity in ecosystems. Collectively, our results underscore the complexity of natural microbial communities and provide a glimpse at the intricate nested symbiotic interactions between haloarchaea, nanohaloarchaea, their viruses and virus satellites (Extended Data Fig. 10).

Methods

Environmental samples

Brine samples used for this study were collected in January 2019 from different sites in the north Danakil Depression, Ethiopia: a cave reservoir in the Dallol dome salt canyons (9Gt), the Western Canyon Lakes WCL2 and WCL3, and the middle of Lake Karum or Assale close to one of its islands (9Ass)[30]. An additional sample from the northwest rim of Lake Assale was collected in 2016[59]. Their physicochemical parameters and hydrochemistry were characterized, with the geothermally influenced and actively degassing Western Canyon Lakes being the most extreme and chaotropic systems sampled in the region[30]. Microbial biomass with cell diameters ranging from 0.2 to 30 µm was fractionated by filtration, fixed in ethanol in situ and, after DNA purification in the laboratory, used for direct metagenome sequencing[31]. In addition, unfiltered brine samples from 9Ass, 9Gt and WCL3 were collected in sterile flasks and maintained at room temperature under diel conditions for culture enrichment attempts.

Enrichment cultures and observation of virus-like particles

Enrichment cultures of the brine samples were established in CA medium[60], by inoculating a 10 ml aliquot of the brine sample into 40 ml medium. The mixtures were incubated aerobically at 37 °C and 100 rpm agitation for 2 weeks. Cell-free culture supernatants containing virus-like particles were collected by centrifugation (15,303*g*, 30 min, Beckman JLA 16.250 rotor) and concentrated by ultracentrifugation (164,243*g*, 3 h, Beckman rotor 70 Ti). The pellets were suspended in the sample buffer (2.46 M NaCl, 85 mM $MgSO_4 \times 7H_2O$, 88 mM $MgCl_2 \times 6H_2O$, 56 mM KCl, 3 mM $CaCl_2$ and 12 mM Tris–HCl (pH 7.5)). For electron microscopy observations, samples were applied to carbon-coated copper grids, negatively stained with 2% uranyl acetate and observed under the transmission electron microscope FEI Tecnai Biotwin.

Identification of viral sequences

The five Danakil metagenomes analysed in this study were reported recently[31] and are available in GenBank under BioProject PRJNA541281. The accession numbers for individual metagenomes are SAMN37693137 (DAL-Ass), SAMN37693138 (DAL-9Ass), SAMN37693139 (DAL-9Gt), SAMN37693140 (DAL-WCL2) and SAMN37693141 (DAL-WCL3). The microbial community structure was originally inferred from these metagenomes via the identification and relative quantification (reads per kilobase per million mapped reads) of 15 ribosomal proteins phylogenetically assigned to GTDB taxa[31]. The general workflow of identification of archaeal viruses in metagenomes has been recently described in detail elsewhere[61]. In brief, putative viral sequences were sorted out from Danakil metagenome assemblies using two tools, geNomad v1.11.0 (github.com/apcamargo/genomad/)[34] and VirSorter2 v2.2.3 (github.com/jiarong/VirSorter2)[35]. Sequences with virus score over 0.70 (geNomad or VirSorter2) were selected and potential host contaminations were removed using CheckV v1.0.1 (bitbucket.org/berkeleylab/checkv)[62]. The relative abundance of viral genomes was inferred from their normalized coverage (see below). Notably, the metagenomes analysed in this study were produced from samples retained on the 0.2–30-µm filter membranes after filtration, that is, with the majority of extracellular virions being removed. Thus, the reconstructed viromes are likely to be enriched in viruses and satellites actively replicating within their hosts.

Construction of virus gene-sharing networks

Virus gene-sharing networks were built by using vConTACT v.2.0 with default parameters[63]. Input sequences were Danakil viral sequences (≥5 kb) and archaeal viral sequences and bacteriophage sequences downloaded from NCBI Reference Sequence Database (RefSeq release 220). The network file produced by vConTACT v.2.0 was visualized by Cytoscape v3.10.1[64].

Taxon-specific CRISPR identification and spacer extraction

Genomes of representative species of Halobacteriota and Nanohaloarchaeota were downloaded from GTDB (release 08-RS214), and CRISPR sequences were extracted from the genomes using CRISPRCasFinder v4.2.20[65]. Redundancy removal of CRISPR sequences was done by CD-HIT v4.8.1 clustering[66] with a 100% sequence identity cut-off. Next, to obtain taxon-specific CRISPRs, for example, haloarchaea-specific CRISPRs, the following procedure was conducted: all CRISPR sequences extracted from haloarchaea were used as queries to perform BLASTn v2.16.0 searches against genomes of all other archaeal phyla within the GTDB. CRISPR sequences were removed if they showed matches (>90% identity and >90% coverage) to genomes from other archaeal phyla, thus retaining only those that are unique to haloarchaea (that is, haloarchaea-specific CRISPRs). The same process was applied to obtain nanohaloarchaea-specific CRISPRs. In total, 988 non-redundant haloarchaea-specific and 4 non-redundant nanohaloarchaea-specific CRISPR sequences were obtained from the GTDB database. Using these taxon-specific CRISPRs as references, the presence of haloarchaeal and nanohaloarchaeal CRISPR arrays within Danakil metagenomes were identified using BLASTn v2.16.0 search with the parameters: -word_size 7, -dust no, -perc_identity 90, -qcov_hsp_perc 90. Next, spacers were retrieved from these taxon-linked CRISPR arrays using CRISPRCasFinder v4.2.20. In addition, to augment the spacer dataset for the subsequent establishment of virus–host connections, we extracted haloarchaeal and nanohaloarchaeal CRISPR spacers also from the EM dataset using the same approach. The haloarchaea-specific CRISPR arrays from the Danakil metagenomes, GTDB and EM contained 5,144, 47,878 and 7,700 spacers, respectively. In the case of nanohaloarchaea, 1,062 and 117 spacers were extracted from Danakil metagenomes and GTDB, respectively, whereas EM did not contain identifiable nanohaloarchaeal spacers. Spacers extracted from different databases displayed

similar length distributions, centring at ~36 bp (Fig. 1e). The Danakil haloarchaeal and nanohaloarchaeal spacers showed no overlap with those from the two other databases, pointing to the novelty of the Danakil viral assemblages.

Spacer targeting analysis

To make linkages between Danakil viruses and archaeal host groups, haloarchaeal and nanohaloarchaeal CRISPR spacers were used as queries to search against viral sequences using BLASTn v2.16.0 with the parameters: -word_size 7, -dust no[67,68]. Consequently, a viral sequence hit showing ≥30 bp exact match with a certain spacer was considered the spacer target (that is, protospacer), and the spacer-related archaeal taxon was then assigned as the host group to the corresponding protospacer-carrying viral sequence. If some members of a group of closely related viruses were targeted by CRISPR spacers, all members of that group were considered to infect the same hosts. For instance, DNSV-1 is not targeted by nanohaloarchaeal CRISPR spacers, but its two close relatives, DNSV-2 (near-complete genome) and SNSV-1 (complete genome) are both matched by nanohaloarchaeal spacers (Extended Data Fig. 3a). Thus, all members of this group were considered to infect nanohaloarchaeal hosts. Notably, DNTV-1 was found to be closely related to a viral fragment found in a nanohaloarchaeal single-cell amplified genome[6] (Extended Data Fig. 4a).

Viral sequence extension

The de novo assembled viral sequences (≥3 kb) were pooled together and used as seed sequences for reference assembly as previously described[69], in an attempt to get complete genomes. In brief, all Danakil metagenomic sequencing reads were mapped to seed sequences using Geneious Prime with parameters of 35 bp overlap and 97% identity. The final extended sequences with ≥50 bp direct or inverted terminal repeats were considered complete genomes.

Virus genome annotation and phylogenomic analysis

ORFs of viral sequences were predicted using Prokka v1.14.5[70]. Functional annotation was performed using HHsearch[71] against Pfam, Protein Data Bank (PDB), Conserved Domain Database (CDD) and viral protein sequence (uniprot_sprot_vir70) databases available from the MPI Bioinformatics Toolkit[72]. The proteomic trees were generated using the ViPTree server version 4.0[73].

Identification of relatives of the Danakil NHVs

Sequences of MCPs of Danakil nanohaloarchaeal head-tailed viruses, MCPs and genome-packaging ATPases of Danakil nanohaloarchaeal tailless icosahedral viruses, pleolipovirus VP4-like proteins and AAA+ ATPases of Danakil nanohaloarchaeal pleomorphic viruses, and MCPs and AAA+ ATPases of Danakil nanohaloarchaeal spindle-shaped viruses were used as queries to perform BLASTp searches in NCBI nr database and IMG/VR database. Hits with ≥30% identity and ≥50% coverage were retrieved, and the corresponding viral sequences were designated as relatives of Danakil NHVs and downloaded for subsequent analyses.

Orthologous fraction calculation of the viral genomes

The proportion of orthologous fraction between the viral genomes (same inputs as for ViPTree) was estimated as previously described[61].

Distribution of HVs and NHVs across Danakil salt lakes

The abundance of the viral and SRCE genomes across the different metagenomes was calculated with CoverM v0.6.1[74] by mapping the reads from the five Danakil metagenomes to each of the genome contigs. The trimmed mean was selected for visualization in a heatmap via an ad hoc R script using the gplots package. The dendrograms were computed and reordered based on the means. In Fig. 1g, manual breaks of the data were set for min (0), median (0.6043255), mean (109.1317) and max (3,369.535) for better visualization, whereas in Extended Data Fig. 9, manual breaks of the data were set for min (0), median (1.845191), mean (85.48989) and max (2,669.926).

Phylogenetic analyses

Head-tailed HVs. The MCP sequences of Danakil head-tailed NHVs and HVs were pooled with their homologues from all cultured HVs. Protein sequences were first clustered using CD-HIT v4.8.1 with a 75% identity threshold (option '-c 0.75')[66]. Next, the non-redundant sequences were aligned using Muscle5 with default parameters[75], and non-informative columns were removed from the alignment using trimAl v1.2 with option -gt 0.2[76]. Next, a phylogenetic tree was constructed based on the trimmed alignment using IQ-TREE v2.2.2.2 with the following parameters: -m MFP, -alrt 1000[77]. The resulting phylogeny was visualized using the iTOL v7 online tool[78]. The same process was applied to generate phylogenetic trees for portal proteins and terminase large subunits of head-tailed HVs. The trimmed alignments for the MCP, portal and TerL sequences included 452, 889 and 537 positions, and the best fitting models for phylogenetic reconstructions were Blosum62+F+I+R3, Q.pfam+F+I+G4 and Q.pfam+F+R5, respectively.

Tailless icosahedral NHVs. Phylogenetic trees were constructed for capsid proteins and genome packaging ATPases. The homologues of NHVs' small capsid proteins, large capsid proteins and genome packaging ATPases were collected from the NCBI nr database through BLASTp searches (>30% identity and >50% coverage). The methods for sequence redundancy removal, sequence alignment and trimming, and tree construction and visualization were the same as used for generating the MCP tree of head-tailed HVs. Before tree construction, the trimmed sequence alignments of small and large capsid proteins were concatenated. The trimmed alignments of the CP_S-CP_L and genome packaging ATPase included 423 and 269 positions, respectively. The best-fitting model for the CP_S-CP_L and ATPase phylogenies was Q.pfam+F+I+R4 and Q.pfam+I+G4, respectively.

Spindle-shaped viruses and their satellites. Reference sequences of AAA+ ATPases were extracted from genomes of spindle-shaped viruses of Sulfolobales ($n = 57$), Asgardarchaeota ($n = 4$), *Nitrosopumilus* ($n = 2$) and Halobacteria ($n = 2$). The phylogenetic tree was generated using the same methods as detailed above for head-tailed HVs. The trimmed alignment included 198 positions, and the best-fitting model was LG+I+R4.

RNR large subunits of large head-tailed NHVs. The homologues of NHV RNR large subunits were collected from NCBI Protein Reference Sequences database (Archaea, Bacteria and Viruses) and Danakil metagenomes through BLASTp searches (>50% identity and >75% coverage). Protein sequences were initially clustered using CD-HIT v4.8.1 with a 75% identity threshold[66]. Subsequently, the RNR large subunits of NHVs were aligned with their non-redundant homologues using Muscle5 with default parameters[75]. Non-informative columns were eliminated from the alignment using trimAl v1.2 with the option '-gappyout'[76]. Finally, a maximum likelihood phylogenetic tree was calculated with IQ-TREE v2.2.2.2 with the parameters: -m MFP, -alrt 1000. The trimmed alignment included 477 positions, and the best-fitting model was Q.pfam+I+R10.

GlmS of DNTV-1. The homologues of DNTV-1 GlmS were collected from NCBI Protein Reference Sequences database (Archaea, Bacteria and Viruses) and Danakil Depression metagenomes through three iterations of PSI-BLASTp searches. The phylogenetic tree was generated using the same methods as detailed above for head-tailed HVs. The trimmed alignment included 605 positions, and the best-fitting model was LG+F+R10.

Reps of SRCEs. The homologues of SRCE Reps were collected from ref. 79 (only archaeal and archaeal virus homologues were kept). The phylogenetic tree was generated using the same methods as detailed above for head-tailed HVs. The trimmed alignment included 628 positions, and the best-fitting model was VT+F+R8.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

All assembled genomes were deposited to GenBank (viruses: PQ827550–PQ827567; SRCEs and plasmids: PQ766422–PQ766435). Metagenome-assembled genomes are accessible on GenBank through BioProject PRJNA541281. All identified haloarchaeal and nanohaloarchaeal CRISPRs and spacers are available via GitHub at https://github.com/IfanZHOU/DAL-virome. Source data are provided with this paper.

Code availability

All scripts used in this work are available via GitHub at https://github.com/IfanZHOU/DAL-virome.

References

1. Goker, M. & Oren, A. Valid publication of names of two domains and seven kingdoms of prokaryotes. *Int. J. Syst. Evol. Microbiol.* https://doi.org/10.1099/ijsem.0.006242 (2024).
2. Baker, B. J. et al. Diversity, ecology and evolution of Archaea. *Nat. Microbiol* **5**, 887–900 (2020).
3. Dombrowski, N., Lee, J. H., Williams, T. A., Offre, P. & Spang, A. Genomic diversity, lifestyles and evolutionary origins of DPANN archaea. *FEMS Microbiol. Lett.* https://doi.org/10.1093/femsle/fnz008 (2019).
4. Rinke, C. et al. Insights into the phylogeny and coding potential of microbial dark matter. *Nature* **499**, 431–437 (2013).
5. Emerson, J. B. et al. Virus–host and CRISPR dynamics in Archaea-dominated hypersaline Lake Tyrrell, Victoria, Australia. *Archaea* **2013**, 370871 (2013).
6. Martínez-García, M., Santos, F., Moreno-Paz, M., Parro, V. & Anton, J. Unveiling viral–host interactions within the 'microbial dark matter. *Nat. Commun.* **5**, 4542 (2014).
7. Wu, Z., Liu, S. & Ni, J. Metagenomic characterization of viruses and mobile genetic elements associated with the DPANN archaeal superphylum. *Nat. Microbiol.* **9**, 3362–3375 (2024).
8. Penades, J. R., Seed, K. D., Chen, J., Bikard, D. & Rocha, E. P. C. Genetics, ecology and evolution of phage satellites. *Nat. Rev. Microbiol.* https://doi.org/10.1038/s41579-025-01156-z (2025).
9. Arnold, H. P. et al. The genetic element pSSVx of the extremely thermophilic crenarchaeon *Sulfolobus* is a hybrid between a plasmid and a virus. *Mol. Microbiol.* **34**, 217–226 (1999).
10. Wang, Y. et al. A novel Sulfolobus non-conjugative extrachromosomal genetic element capable of integration into the host genome and spreading in the presence of a fusellovirus. *Virology* **363**, 124–133 (2007).
11. Andrade, K. et al. Metagenomic and lipid analyses reveal a diel cycle in a hypersaline microbial ecosystem. *ISME J.* **9**, 2697–2711 (2015).
12. Baker, B. A. et al. Expanded phylogeny of extremely halophilic archaea shows multiple independent adaptations to hypersaline environments. *Nat. Microbiol* **9**, 964–975 (2024).
13. Feng, Y. et al. The evolutionary origins of extreme halophilic archaeal lineages. *Genome Biol. Evol.* https://doi.org/10.1093/gbe/evab166 (2021).
14. Zhao, D. et al. Comparative genomic insights into the evolution of Halobacteria-associated "Candidatus Nanohaloarchaeota". *mSystems* **7**, e0066922 (2022).
15. Hamm, J. N. et al. Unexpected host dependency of Antarctic Nanohaloarchaeota. *Proc. Natl Acad. Sci. USA* **116**, 14661–14670 (2019).
16. La Cono, V. et al. Nanohaloarchaea as beneficiaries of xylan degradation by haloarchaea. *Microb. Biotechnol.* **16**, 1803–1822 (2023).
17. La Cono, V. et al. Symbiosis between nanohaloarchaeon and haloarchaeon is based on utilization of different polysaccharides. *Proc. Natl Acad. Sci. USA* **117**, 20223–20234 (2020).
18. Reva, O. et al. Functional diversity of nanohaloarchaea within xylan-degrading consortia. *Front. Microbiol.* **14**, 1182464 (2023).
19. Puxty, R. J. & Millard, A. D. Functional ecology of bacteriophages in the environment. *Curr. Opin. Microbiol.* **71**, 102245 (2023).
20. Breitbart, M., Bonnain, C., Malki, K. & Sawaya, N. A. Phage puppet masters of the marine microbial realm. *Nat. Microbiol.* **3**, 754–766 (2018).
21. López-García, P. et al. Metagenome-derived virus–microbe ratios across ecosystems. *ISME J.* **17**, 1552–1563 (2023).
22. Santos, F. et al. Culture-independent approaches for studying viruses from hypersaline environments. *Appl. Environ. Microbiol.* **78**, 1635–1643 (2012).
23. Liu, Y. et al. Diversity, taxonomy, and evolution of archaeal viruses of the class *Caudoviricetes*. *PLoS Biol.* **19**, e3001442 (2021).
24. Atanasova, N. S., Bamford, D. H. & Oksanen, H. M. Haloarchaeal virus morphotypes. *Biochimie* **118**, 333–343 (2015).
25. Luk, A. W., Williams, T. J., Erdmann, S., Papke, R. T. & Cavicchioli, R. Viruses of haloarchaea. *Life* **4**, 681–715 (2014).
26. Dyall-Smith, M., Tang, S. L. & Bath, C. Haloarchaeal viruses: how diverse are they?. *Res. Microbiol.* **154**, 309–313 (2003).
27. Crits-Christoph, A. et al. Functional interactions of archaea, bacteria and viruses in a hypersaline endolithic community. *Environ. Microbiol.* **18**, 2064–2077 (2016).
28. Emerson, J. B. et al. Dynamic viral populations in hypersaline systems as revealed by metagenomic assembly. *Appl. Environ. Microbiol.* **78**, 6309–6320 (2012).
29. Garcia-Heredia, I. et al. Reconstructing viral genomes from the environment using fosmid clones: the case of haloviruses. *PLoS ONE* **7**, e33802 (2012).
30. Belilla, J. et al. Archaeal overdominance close to life-limiting conditions in geothermally influenced hypersaline lakes at the Danakil Depression, Ethiopia. *Environ. Microbiol.* **23**, 7168–7182 (2021).
31. Gutiérrez-Preciado, A. et al. Extremely acidic proteomes and metabolic flexibility in bacteria and highly diversified archaea thriving in geothermal chaotropic brines. *Nat. Ecol. Evol.* **8**, 1856–1869 (2024).
32. Maier, L. K. et al. The nuts and bolts of the *Haloferax* CRISPR–Cas system I-B. *RNA Biol.* **16**, 469–480 (2019).
33. Nayfach, S. et al. A genomic catalog of Earth's microbiomes. *Nat. Biotechnol.* **39**, 499–509 (2021).
34. Camargo, A. P. et al. Identification of mobile genetic elements with geNomad. *Nat. Biotechnol.* **42**, 1303–1312 (2024).
35. Guo, J. et al. VirSorter2: a multi-classifier, expert-guided approach to detect diverse DNA and RNA viruses. *Microbiome* **9**, 37 (2021).
36. Esser, S. P. et al. A predicted CRISPR-mediated symbiosis between uncultivated archaea. *Nat. Microbiol.* **8**, 1619–1633 (2023).
37. Reed, C. J., Lewis, H., Trejo, E., Winston, V. & Evilia, C. Protein adaptations in archaeal extremophiles. *Archaea* **2013**, 373275 (2013).
38. Simon, D., Cristina, J. & Musto, H. Nucleotide composition and codon usage across viruses and their respective hosts. *Front Microbiol.* **12**, 646300 (2021).
39. Hurwitz, B. L. & U'Ren, J. M. Viral metabolic reprogramming in marine ecosystems. *Curr. Opin. Microbiol.* **31**, 161–168 (2016).

40. Enav, H., Mandel-Gutfreund, Y. & Beja, O. Comparative metagenomic analyses reveal viral-induced shifts of host metabolism towards nucleotide biosynthesis. *Microbiome* **2**, 9 (2014).
41. Krupovic, M., Gribaldo, S., Bamford, D. H. & Forterre, P. The evolutionary history of archaeal MCM helicases: a case study of vertical evolution combined with hitchhiking of mobile genetic elements. *Mol. Biol. Evol.* **27**, 2716–2732 (2010).
42. Chevallereau, A. et al. Next-generation "-omics" approaches reveal a massive alteration of host RNA metabolism during bacteriophage infection of *Pseudomonas aeruginosa*. *PLoS Genet.* **12**, e1006134 (2016).
43. Oakley, A. J., Coggan, M. & Board, P. G. Identification and characterization of gamma-glutamylamine cyclotransferase, an enzyme responsible for gamma-glutamyl-epsilon-lysine catabolism. *J. Biol. Chem.* **285**, 9642–9648 (2010).
44. Milewski, S. Glucosamine-6-phosphate synthase—the multi-facets enzyme. *Biochim. Biophys. Acta* **1597**, 173–192 (2002).
45. Isupov, M. N. et al. Substrate binding is required for assembly of the active conformation of the catalytic site in Ntn amidotransferases: evidence from the 1.8 A crystal structure of the glutaminase domain of glucosamine 6-phosphate synthase. *Structure* **4**, 801–810 (1996).
46. Reva, O. N. et al. DPANN symbiont of *Haloferax volcanii* accelerates xylan degradation by the non-host haloarchaeon *Halorhabdus* sp. *iScience* **28**, 111749 (2025).
47. Reva, O. N. et al. Interplay of intracellular and trans-cellular DNA methylation in natural archaeal consortia. *Environ. Microbiol. Rep.* **16**, e13258 (2024).
48. Krupovic, M., Quemin, E. R., Bamford, D. H., Forterre, P. & Prangishvili, D. Unification of the globally distributed spindle-shaped viruses of the Archaea. *J. Virol.* **88**, 2354–2358 (2014).
49. Wang, F. et al. Spindle-shaped archaeal viruses evolved from rod-shaped ancestors to package a larger genome. *Cell* **185**, 1297–1307 (2022).
50. Roux, S. et al. Analysis of metagenomic data reveals common features of halophilic viral communities across continents. *Environ. Microbiol.* **18**, 889–903 (2016).
51. Guyot, V. et al. A newly emerging alphasatellite affects banana bunchy top virus replication, transcription, siRNA production and transmission by aphids. *PLoS Pathog.* **18**, e1010448 (2022).
52. Boyd, C. M. & Seed, K. D. A phage satellite manipulates the viral DNA packaging motor to inhibit phage and promote satellite spread. *Nucleic Acids Res.* **52**, 10431–10446 (2024).
53. de Sousa, J. A. M., Fillol-Salom, A., Penades, J. R. & Rocha, E. P. C. Identification and characterization of thousands of bacteriophage satellites across bacteria. *Nucleic Acids Res.* **51**, 2759–2777 (2023).
54. Alqurainy, N. et al. A widespread family of phage-inducible chromosomal islands only steals bacteriophage tails to spread in nature. *Cell Host Microbe* **31**, 69–82 (2023).
55. Hadidi, A., Czosnek, H. H., Kalantidis, K. & Palukaitis, P. Viroids and satellites and their vector interactions. *Viruses* https://doi.org/10.3390/v16101598 (2024).
56. Dyall-Smith, M. & Pfeiffer, F. Global distribution and diversity of haloarchaeal pL6-family plasmids. *Genes* https://doi.org/10.3390/genes15091123 (2024).
57. Dyall-Smith, M. & Pfeiffer, F. The PL6-family plasmids of haloquadratum are virus-related. *Front. Microbiol.* **9**, 1070 (2018).
58. Koonin, E. V. & Krupovic, M. Polintons, virophages and transpovirons: a tangled web linking viruses, transposons and immunity. *Curr. Opin. Virol.* **25**, 7–15 (2017).
59. Belilla, J. et al. Hyperdiverse archaea near life limits at the polyextreme geothermal Dallol area. *Nat. Ecol. Evol.* **3**, 1552–1561 (2019).
60. Allers, T., Ngo, H. P., Mevarech, M. & Lloyd, R. G. Development of additional selectable markers for the halophilic archaeon *Haloferax volcanii* based on the leuB and trpA genes. *Appl. Environ. Microbiol.* **70**, 943–953 (2004).
61. Zhou, Y., Wang, Y., Prangishvili, D. & Krupovic, M. Exploring the archaeal virosphere by metagenomics. *Methods Mol. Biol.* **2732**, 1–22 (2024).
62. Nayfach, S. et al. CheckV assesses the quality and completeness of metagenome-assembled viral genomes. *Nat. Biotechnol.* **39**, 578–585 (2021).
63. Bin Jang, H. et al. Taxonomic assignment of uncultivated prokaryotic virus genomes is enabled by gene-sharing networks. *Nat. Biotechnol.* **37**, 632–639 (2019).
64. Shannon, P. et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504 (2003).
65. Couvin, D. et al. CRISPRCasFinder, an update of CRISRFinder, includes a portable version, enhanced performance and integrates search for Cas proteins. *Nucleic Acids Res.* **46**, W246–W251 (2018).
66. Li, W. & Godzik, A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **22**, 1658–1659 (2006).
67. Zhou, Y. et al. Diverse viruses of marine archaea discovered using metagenomics. *Environ. Microbiol.* **25**, 367–382 (2023).
68. Edwards, R. A., McNair, K., Faust, K., Raes, J. & Dutilh, B. E. Computational approaches to predict bacteriophage-host relationships. *FEMS Microbiol. Rev.* **40**, 258–272 (2016).
69. Gong, C. et al. Novel virophages discovered in a freshwater lake in China. *Front. Microbiol* **7**, 5 (2016).
70. Seemann, T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30**, 2068–2069 (2014).
71. Steinegger, M. et al. HH-suite3 for fast remote homology detection and deep protein annotation. *BMC Bioinf.* **20**, 473 (2019).
72. Gabler, F. et al. Protein sequence analysis using the MPI Bioinformatics Toolkit. *Curr. Protoc. Bioinform.* **72**, e108 (2020).
73. Nishimura, Y. et al. ViPTree: the viral proteomic tree server. *Bioinformatics* **33**, 2379–2380 (2017).
74. Aroney, S. T. et al. CoverM: read alignment statistics for metagenomics. *Bioinformatics* **41**, btaf147 (2025).
75. Edgar, R. C. Muscle5: high-accuracy alignment ensembles enable unbiased assessments of sequence homology and phylogeny. *Nat. Commun.* **13**, 6968 (2022).
76. Capella-Gutiérrez, S., Silla-Martínez, J. M. & Gabaldon, T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**, 1972–1973 (2009).
77. Minh, B. Q. et al. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol.* **37**, 1530–1534 (2020).
78. Letunic, I. & Bork, P. Interactive Tree of Life (iTOL) v6: recent updates to the phylogenetic tree display and annotation tool. *Nucleic Acids Res.* **52**, W78–W82 (2024).
79. Kazlauskas, D., Varsani, A., Koonin, E. V. & Krupovic, M. Multiple origins of prokaryotic and eukaryotic single-stranded DNA viruses from bacterial and archaeal plasmids. *Nat. Commun.* **10**, 3425 (2019).

Acknowledgements

This work was supported by grants from Ville de Paris (Emergence(s) project MEMREMA) and Agence Nationale de la Recherche (ANR-23-CE13-022 and ANR-21-CE11-0001) to M.K., and the Moore Foundation (https://doi.org/10.37807/GBMF9739), the ANR (ANR-23-CE02-0016-01) and the European Research Council (ERC-2023-AdG 101141745) to P.L.-G. We thank X. Wang, T. Xu and A. Zhou for their help with configuration of software and scripts.

Author contributions

M.K. and Y.Z. conceived the study; Y.Z. assembled and analysed the viral genomes; M.K. annotated the viral genomes and performed structural modelling; Y.L. established enrichment cultures and performed electron microscopy; A.G.-P. analysed the distribution and abundance of viruses and satellites in salt lake metagenomes; D.M. and P.L.-G. collected the environmental samples and provided access to metagenomes; M.M.Y. advised on cultivation of haloarchaeal and nanohaloarchaeal communities. All authors contributed to the conceptualization of the results; Y.Z. and M.K. wrote the paper, which was revised and approved by all coauthors.

Competing interests

The authors declare no competing interests.

Additional information

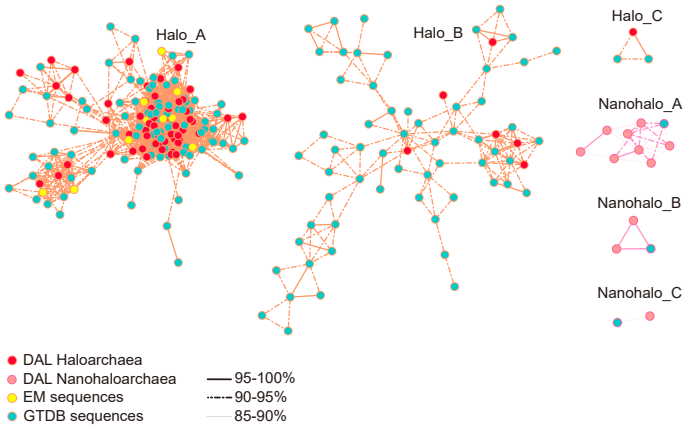
Extended data is available for this paper at https://doi.org/10.1038/s41564-025-02149-7.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41564-025-02149-7.

Correspondence and requests for materials should be addressed to Mart Krupovic.

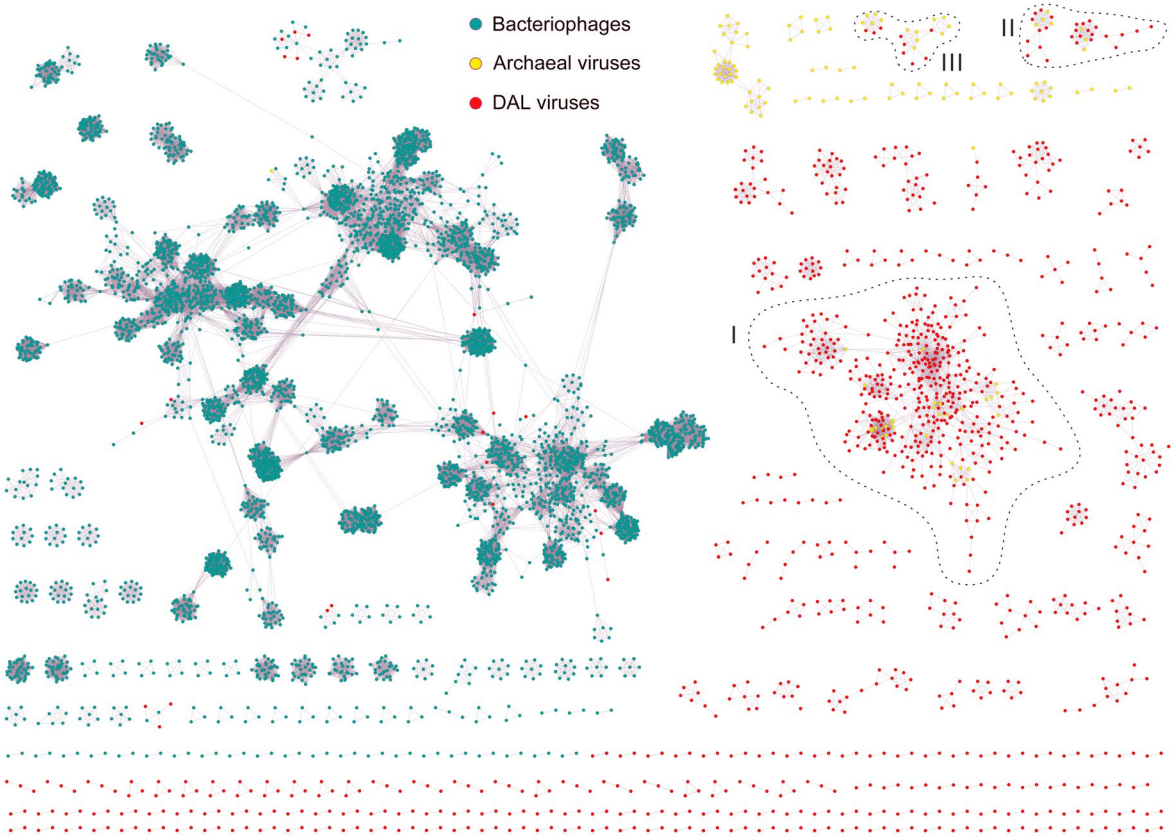
Peer review information *Nature Microbiology* thanks Shingo Kato and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

https://doi.org/10.1038/s41564-025-02149-7



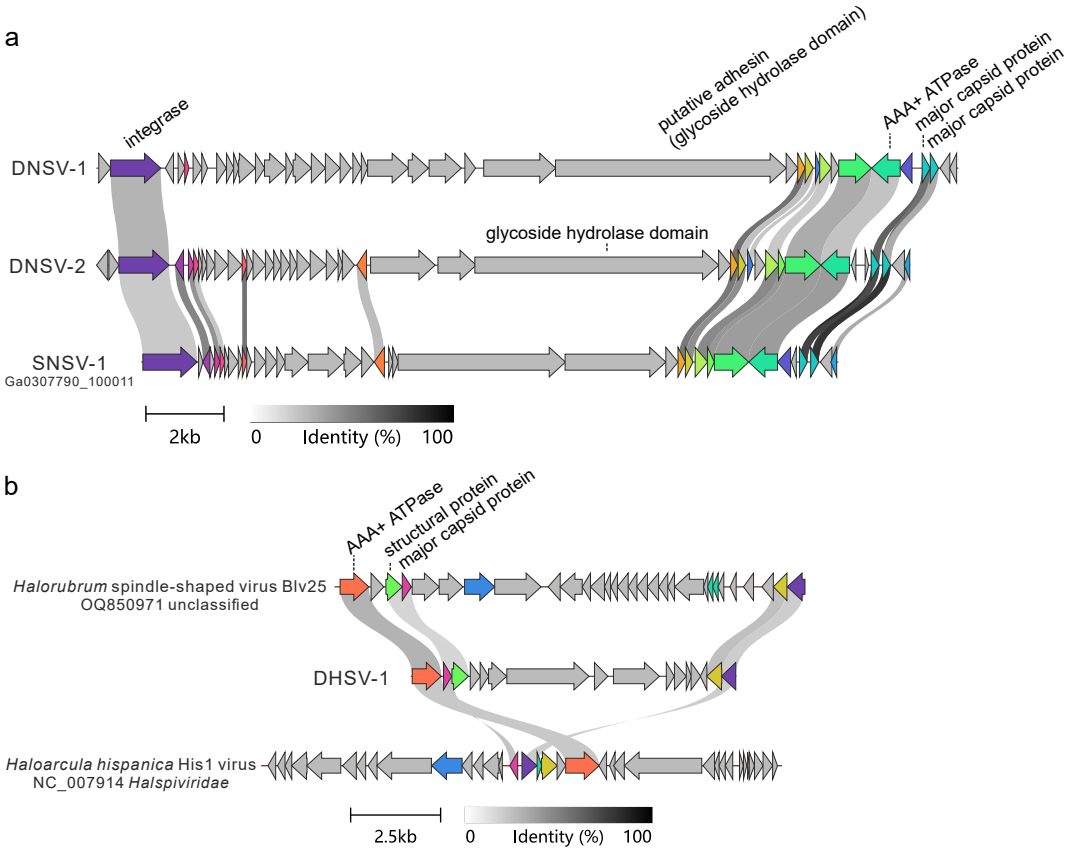
Extended Data Fig. 1 | CRISPR diversity in the genomes of Halobacteriota (Halo) and Nanohaloarchaeota (Nanohalo). Similarity networks of CRISPR sequences. Each node represents a CRISPR sequence and the links between nodes represent the degree of sequence similarity between CRISPR sequences.

https://doi.org/10.1038/s41564-025-02149-7



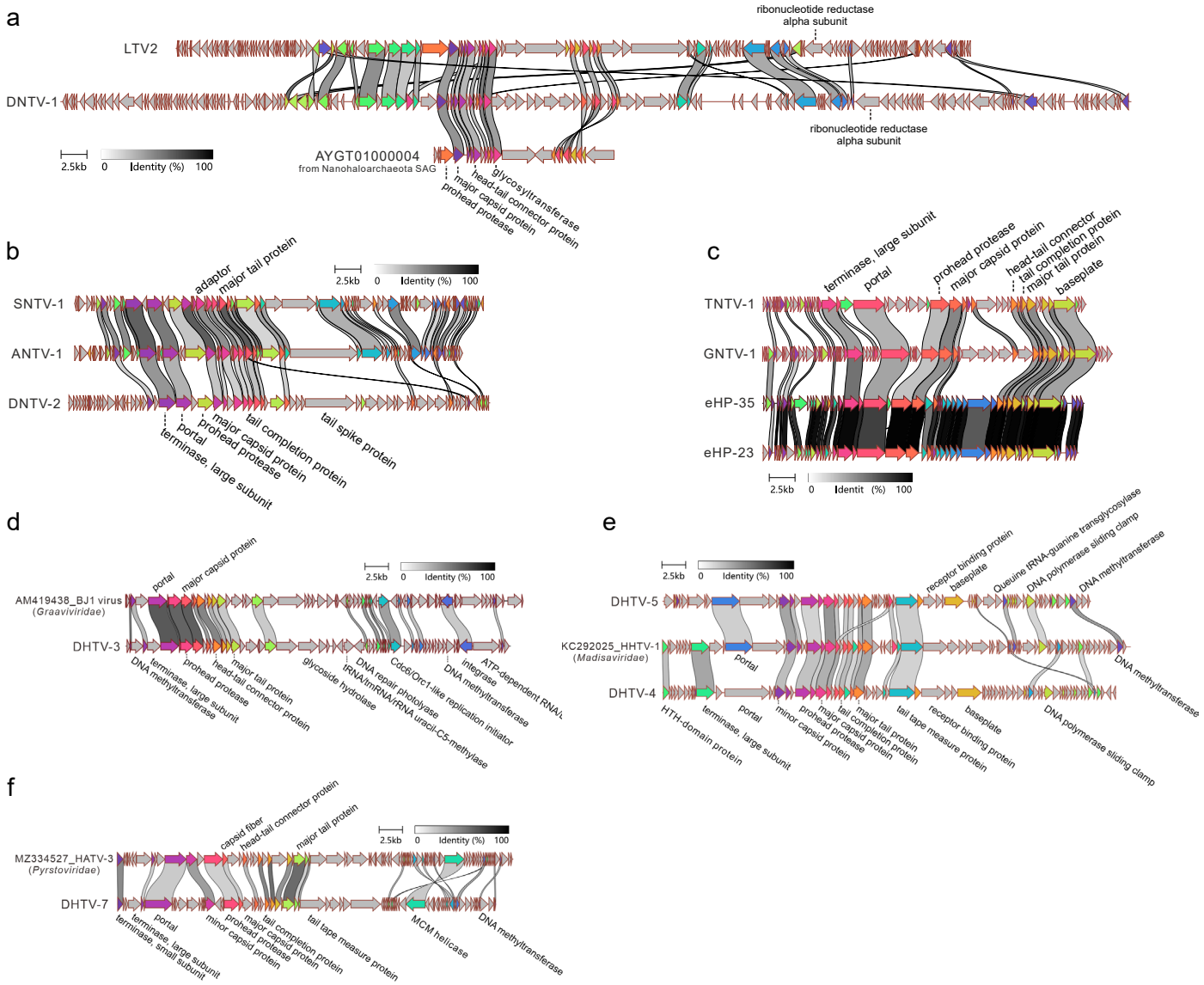
Extended Data Fig. 2 | The vConTACT2 gene-sharing networks of viruses from Danakil Depression and reference prokaryotic DNA viruses. Each node represents a viral sequence and the edges between nodes represent the degree of connectivity based on the fraction of shared proteins. Nodes for reference bacteriophages are colored cyan, reference archaeal viruses are in yellow, and Danakil viruses are in red. The Danakil viral contigs formed three clusters (outlined) with previously described archaeal viruses: I, head-tailed HVs (class *Caudoviricetes*, n = 382); II, tailless icosahedral HVs (families *Simuloviridae* and *Sphaerolipoviridae*, n = 26); III, pleomorphic HVs (family *Pleolipoviridae*, n = 9).

https://doi.org/10.1038/s41564-025-02149-7



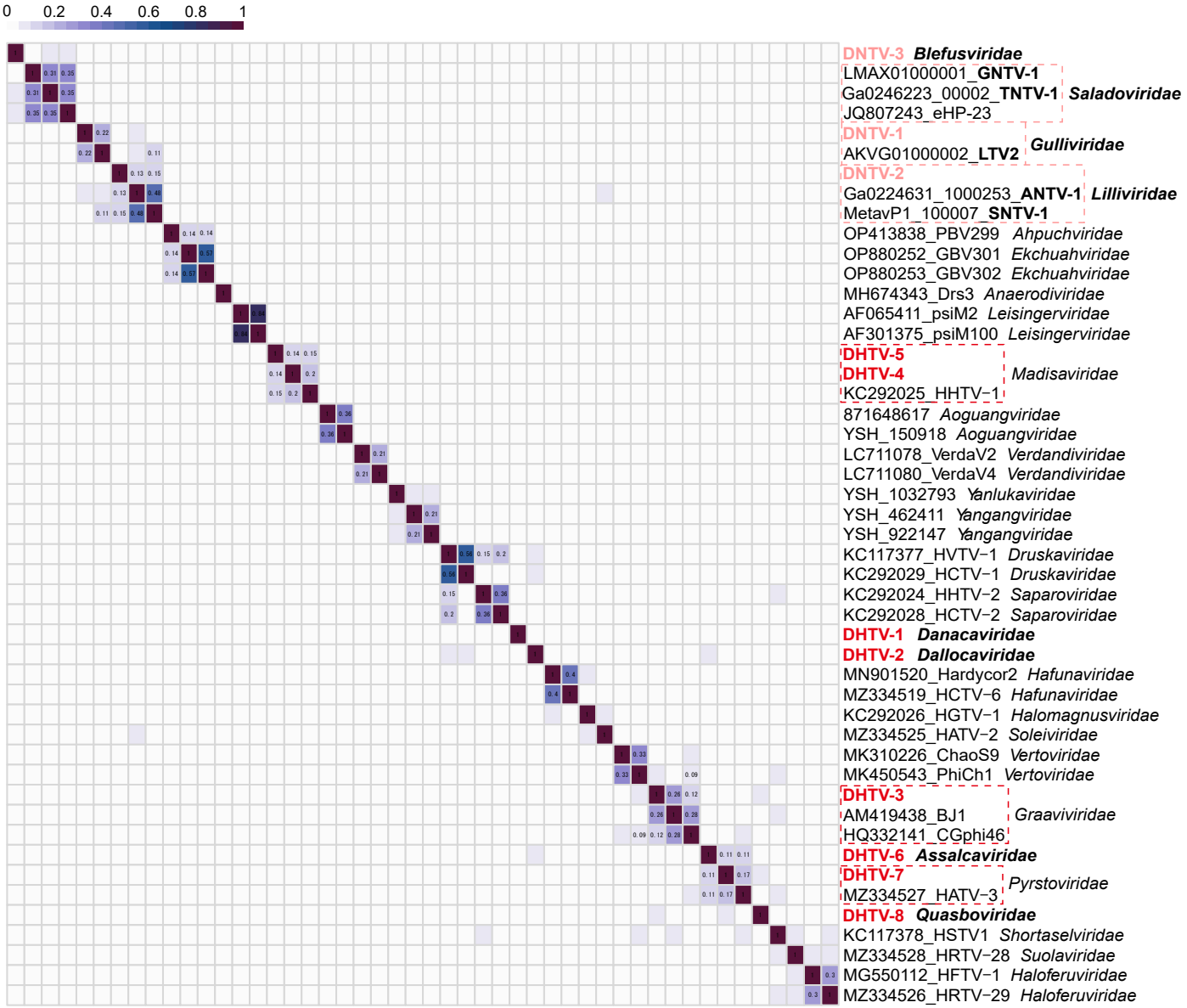
Extended Data Fig. 3 | Genome maps showing the relationship among spindle-shaped viruses. a. Spindle-shaped NHVs. **b**. Spindle-shaped HVs.

https://doi.org/10.1038/s41564-025-02149-7



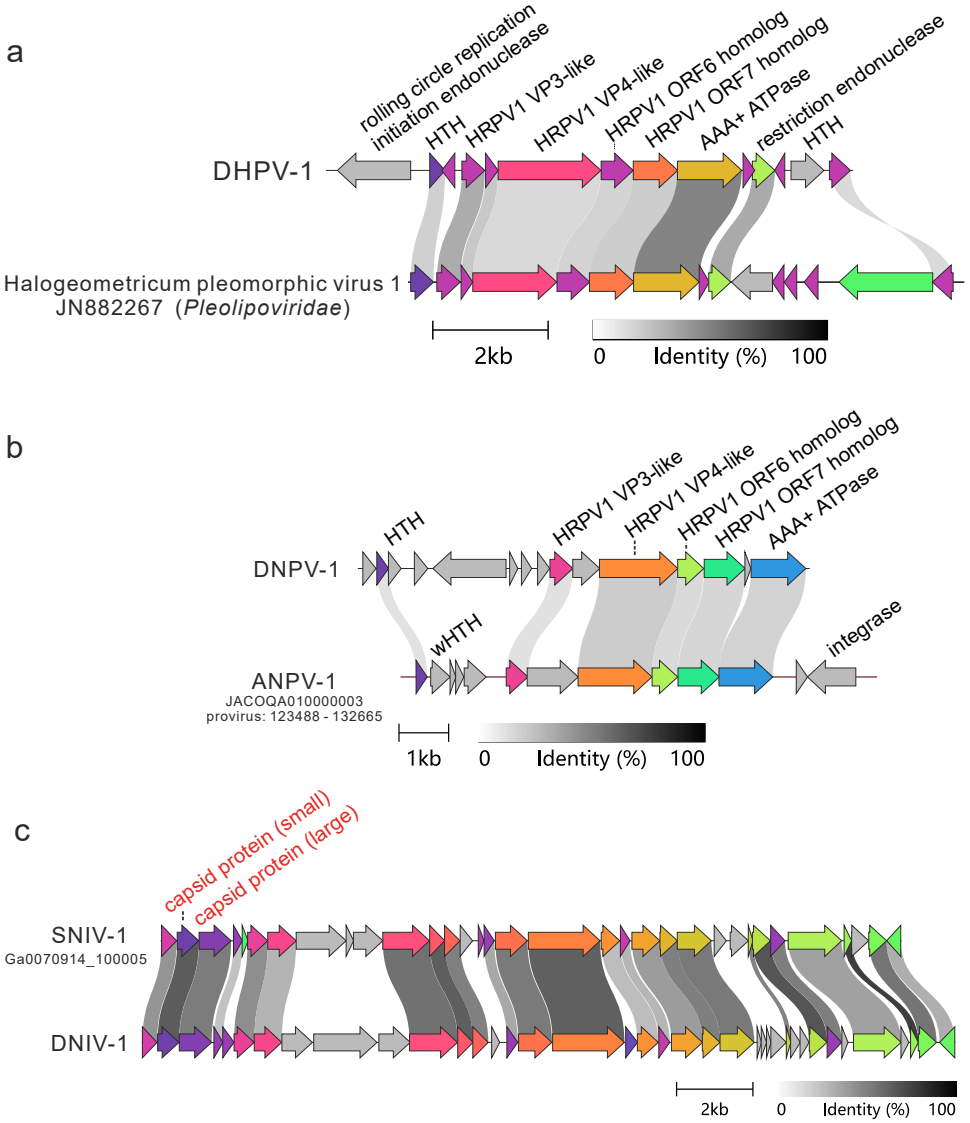
Extended Data Fig. 4 | Genome maps showing the relationships among head-tailed NHVs. a. *Gulliviridae*. **b**. *Lilliviridae*. **c**. *Saladoviridae*. **d**. *Graaviviridae*; **e**. *Madisaviridae*; **f**. *Pyrstoviridae*.

https://doi.org/10.1038/s41564-025-02149-7



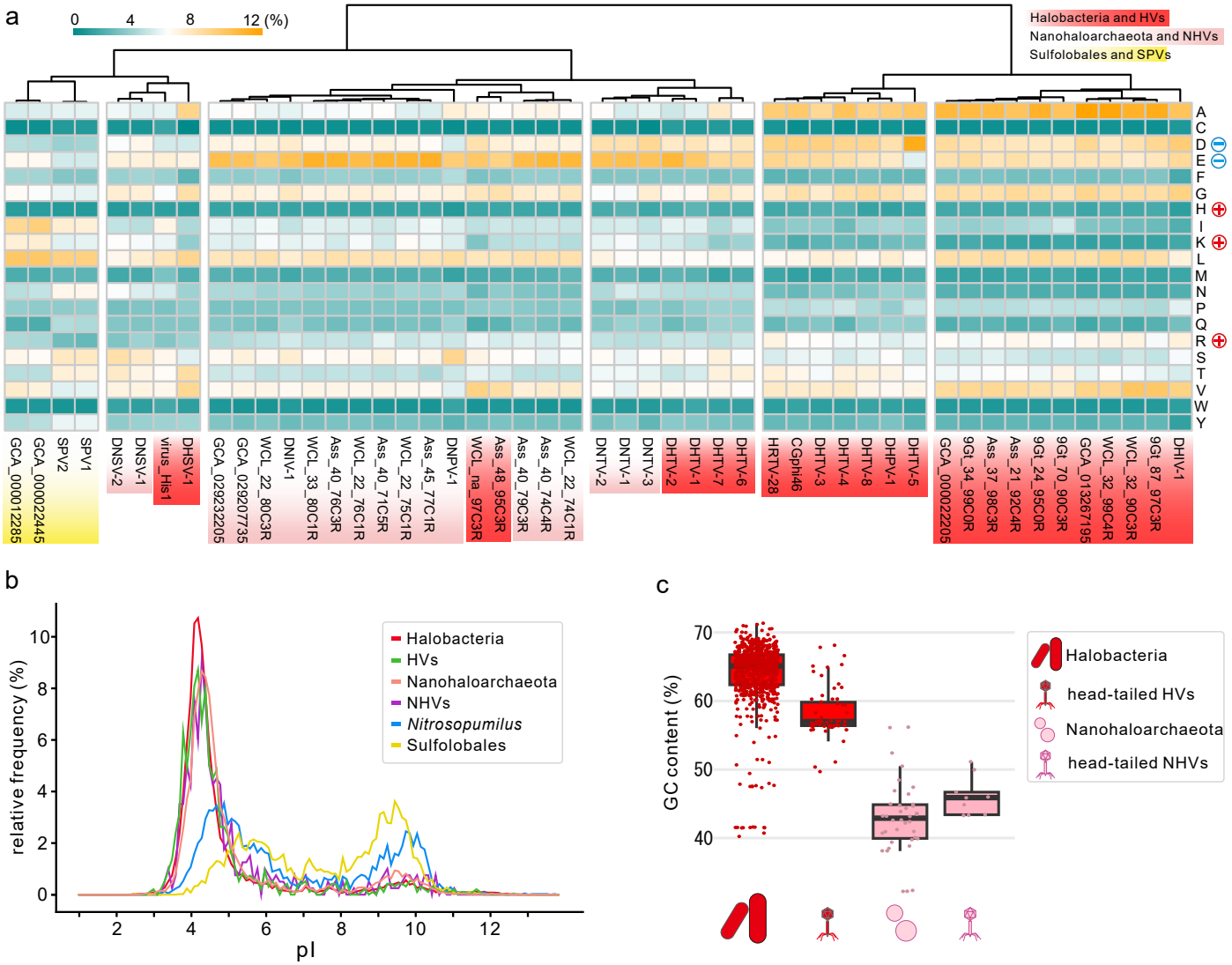
Extended Data Fig. 5 | The heatmap of orthologous fraction among archaeal tailed viruses. HVs and NHVs described in this work are shown in red. Family-level groups of viruses including representatives from the Danakil Depression are boxed. Orthologous fraction values > 0.08 are shown.

https://doi.org/10.1038/s41564-025-02149-7



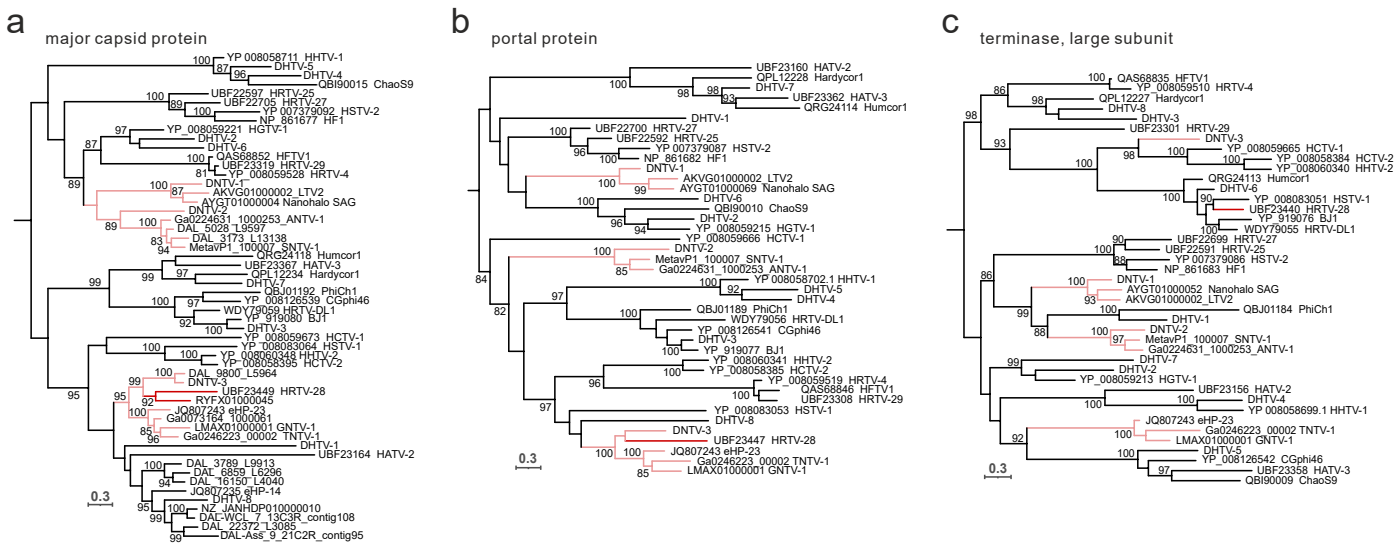
Extended Data Fig. 6 | Genome maps showing the relationships among pleomorphic and tailless icosahedral viruses. a. *Pleolipoviridae* HVs. **b**. *Nanopleoviridae* NHVs. **c**. *Nanicoviridae* NHVs.

https://doi.org/10.1038/s41564-025-02149-7



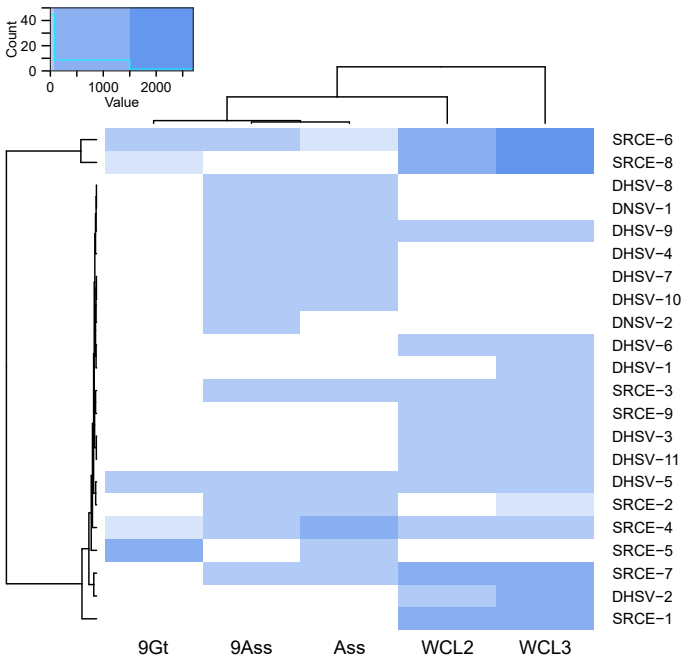
Extended Data Fig. 7 | Adaptation of HVs and NHVs to hypersaline environments. a. Heatmap of amino acid usage patterns. Amino acid frequencies were calculated using proteins encoded by haloarchaea (n = 10), nanohaloarchaea (n = 10) and their viruses (HVs and NHVs) from the Danakil Depression. Genomes of two head-tailed HVs (CGphi46 and HRTV-28) and their hosts as well as two complete genomes of nanohaloarchaea were used for comparison. Genomes of two Sulfolobus species and their viruses SPV1 and SPV2 were also used for comparison. **b**. Distribution of isoelectric point (pI) values inferred for proteins encoded by the Danakil haloarchaea, nanohaloarchaea and their viruses (HVs and NHVs) in comparison with representative archaeal genomes from seawater (*Nitrosopumilus*, n = 2) and hot springs (*Sulfolobus*, n = 2). **c**. Box plots show GC content (%) of genomes of Halobacteriales (n = 749), Nanohaloarchaeota (n = 49) and head-tailed viruses (HVs, n = 62 and NHVs, n = 9). The center line represents the median; the box limits, the first and third quartiles; whiskers extend 1.5 times the interquartile range; data beyond the whiskers are outliers represented as points.

https://doi.org/10.1038/s41564-025-02149-7



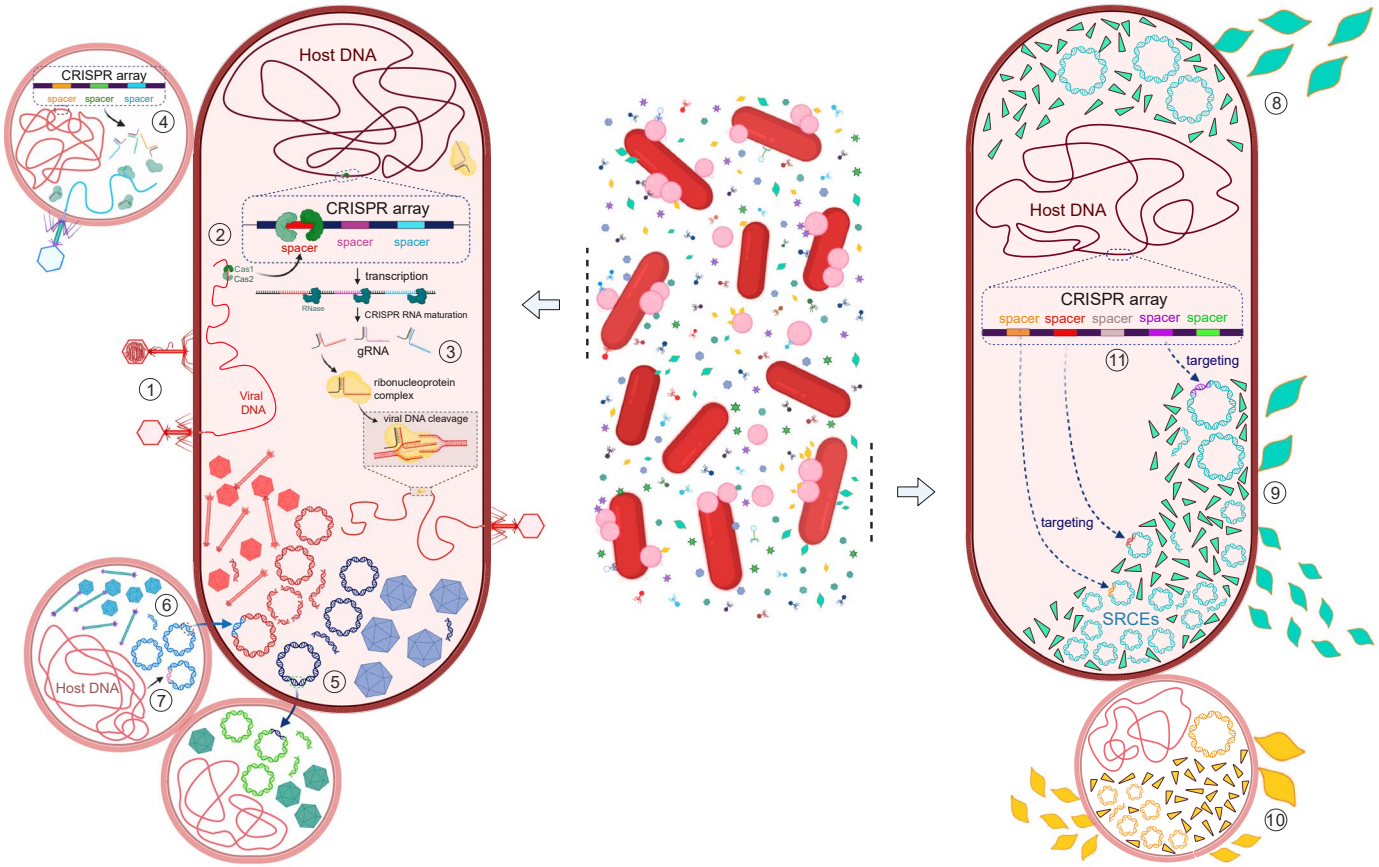
Extended Data Fig. 8 | Maximum likelihood phylogenies of the hallmark proteins of head-tailed viruses. a. Major capsid protein (MCP) of head-tailed HVs and NHVs. **b**. Portal protein of head-tailed HVs and NHVs. **c**. Terminase large subunit (TerL) of head-tailed HVs and NHVs. NHVs are indicated with pink branches.

https://doi.org/10.1038/s41564-025-02149-7



Extended Data Fig. 9 | Heatmap showing the distribution and abundance of haloarchaeal and nanohaloarchaeal spindle-shaped viruses and SRCEs (mean coverage, rows) in Danakil salt lakes (columns). Intensity of the blue color represents relative abundance. The accompanying histogram (upper left) displays the distribution of abundance values across all virus–site combinations.

https://doi.org/10.1038/s41564-025-02149-7



Extended Data Fig. 10 | The complexity of interactions between haloarchaea, nanohaloarchaea, their respective viruses and virus satellites. 1: Viruses infect a haloarchaeal host; 2: A spacer is acquired from viral DNA by the haloarchaeal CRISPR-Cas system; 3: CRISPR spacers are transcribed, matured, and matched with target viral DNA with the help of Cas proteins, leading to the cleavage of the invading viral DNA; 4: Similar CRISPR-Cas immunity processes (1, 2, 3) are happening in nanohaloarchaea; 5: A gene transfer from an icosahedral HV to an icosahedral NHV; 6: A gene transfer from a head-tailed NHV to a head-tailed HV; 7: A gene transfer from a nanohaloarchaeal host to a head-tailed virus; 8: Spindle-shaped HVs infect a haloarchaeal host without interference of virus satellites (SRCEs); 9: Spindle-shaped HVs and SRCEs co-infect a haloarchaeal host. SRCEs replicate and consume the virion components of HVs, which lead to a decrease in HV production; 10: Similar phenomenon (9) is also happening in nanohaloarchaea; 11: the archaeal host carries a CRISPR array with spacers targeting both viruses and virus satellites. Created in BioRender. Krupovic, M. (2025) https://BioRender.com/ptoya59.

nature portfolio

Corresponding author(s): Mart Krupovic

Last updated by author(s): Sep 3, 2025

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☒	☐ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☒	☐ The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐ A description of all covariates tested
☒	☐ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☒	☐ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☒	☐ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Previously published, publicly available metagenomics data was used in this study. The corresponding BioProject accession number is provided in the Data availability section.
Data analysis	The data was analyzed using standard bioinformatics tools, including blastn v2.16.0, CheckV v1.0.1, geNomad v1.11.0, VirSorter v2.2.3, vConTACT v.2.0, Cytoscape v3.10.1, CRISPRCasFinder v4.2.20, CD-HIT v4.8.1, Prokka v1.14.5, HHsearch v3, ViPTree v4.0, CoverM v0.6.1, trimAl v1.2, IQ-TREE v2.2.2.2, iTOL v7. All scripts used in this work are available at github.com/IfanZHOU/DAL-virome

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

All assembled genomes were deposited to GenBank (viruses: PQ827550-PQ827567; SRCEs and plasmids: PQ766422-PQ766435). Metagenome assembled genomes

are accessible on GenBank through BioProject PRJNA541281. All identified haloarchaeal and nanohaloarchaeal CRISPRs and spacers are available at github.com/IfanZHOU/DAL-virome.

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender	*Use the terms sex (biological attribute) and gender (shaped by social and cultural circumstances) carefully in order to avoid confusing both terms. Indicate if findings apply to only one sex or gender; describe whether sex and gender were considered in study design; whether sex and/or gender was determined based on self-reporting or assigned and methods used.* *Provide in the source data disaggregated sex and gender data, where this information has been collected, and if consent has been obtained for sharing of individual-level data; provide overall numbers in this Reporting Summary. Please state if this information has not been collected.* *Report sex- and gender-based analyses where performed, justify reasons for lack of sex- and gender-based analysis.*
Reporting on race, ethnicity, or other socially relevant groupings	*Please specify the socially constructed or socially relevant categorization variable(s) used in your manuscript and explain why they were used. Please note that such variables should not be used as proxies for other socially constructed/relevant variables (for example, race or ethnicity should not be used as a proxy for socioeconomic status).* *Provide clear definitions of the relevant terms used, how they were provided (by the participants/respondents, the researchers, or third parties), and the method(s) used to classify people into the different categories (e.g. self-report, census or administrative data, social media data, etc.)* *Please provide details about how you controlled for confounding variables in your analyses.*
Population characteristics	*Describe the covariate-relevant population characteristics of the human research participants (e.g. age, genotypic information, past and current diagnosis and treatment categories). If you filled out the behavioural & social sciences study design questions and have nothing to add here, write "See above."*
Recruitment	*Describe how participants were recruited. Outline any potential self-selection bias or other biases that may be present and how these are likely to impact results.*
Ethics oversight	*Identify the organization(s) that approved the study protocol.*

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Not applicable.
Data exclusions	No data was excluded.
Replication	Not applicable.
Randomization	Not applicable.
Blinding	Not applicable.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern
☒	☐ Plants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Plants

Seed stocks	Not applicable.
Novel plant genotypes	Not applicable.
Authentication	Not applicable.