Rolling-circle replication initiation protein of haloarchaeal sphaerolipovirus SNJ1 is homologous to bacterial transposases of the IS91 family insertion sequences

Yuchen Wang, Beibei Chen, Mengzhuo Cao, Linshan Sima, David Prangishvili, Xiangdong Chen, Mart Krupovic

JOURNAL OF GENERAL VIROLOGY

SHORT COMMUNICATION
Wang *et al.*, *Journal of General Virology* 2018;99:416–421
DOI 10.1099/jgv.0.001009

MICROBIOLOGY SOCIETY

Rolling-circle replication initiation protein of haloarchaeal sphaerolipovirus SNJ1 is homologous to bacterial transposases of the IS*91* family insertion sequences

Yuchen Wang,[1] Beibei Chen,[1] Mengzhuo Cao,[1] Linshan Sima,[1] David Prangishvili,[2] Xiangdong Chen[1,3,*] and Mart Krupovic[2,*]

Abstract

For most archaeal viruses, the mechanisms of genome replication are poorly understood, while the nature and provenance of their replication proteins are usually unknown. Here we show that replication of the circular double-stranded DNA genome of the halophilic *Natrinema* virus SNJ1, a member of the family *Sphaerolipoviridae*, is associated with the accumulation of single-stranded replicative intermediates, which is typical of rolling-circle replication. The homologues of RepA, the only enzyme that is indispensable for SNJ1 genome replication, are widespread in archaea and are most closely related to bacterial transposases of the IS*91* and IS*CR* family insertion sequences, as opposed to other viral rolling-circle replication initiation proteins. Our results provide insights into the replication mechanism of archaeal viruses and emphasize the evolutionary connection between viruses and other types of mobile genetic elements.

The viruses that infect archaea represent one of the most enigmatic components of the virosphere [1, 2]. These viruses display a wide range of virion morphologies, many of which have never been observed among viruses infecting bacteria and eukaryotes. The gene contents of archaeal viruses are also unique, with most genes having no homologues in other known viruses [3, 4]. Interestingly, several groups of archaeal viruses display evolutionary connections to non-viral mobile genetic elements (MGEs) [5], such as various plasmids and casposons, a recently discovered group of transposon-like integrative MGEs encoding Cas1-like endonucleases [6]. In particular, viruses and these non-viral MGEs share the main proteins responsible for genome replication, suggesting that some archaeal viruses could have evolved directly from non-viral MGEs, or that the genome replication modules have been recurrently exchanged across different MGE classes. Nevertheless, genome replication remains poorly understood for the vast majority of archaeal viruses and has only been studied experimentally for a handful of virus–host systems [7–10]. We have recently described the minimal module required for the replication and maintenance of the circular double-stranded (ds) DNA genome (16.3 kb) of halophilic archaeal virus SNJ1 [9, 11], the type species of the genus *Betasphaerolipovirus* within the family *Sphaerolipoviridae* [12]. Unlike alphashaerolipoviruses, which have linear dsDNA genomes and are lytic [12–16], SNJ1 is a temperate virus propagating within its host *Natrinema* sp. J7-1 cells as a circular episome, and it can be induced by mitomycin C (MMC) treatment [11, 17]. The only gene that is indispensable for SNJ1 genome replication encodes for a 481 aa-long protein RepA, a divergent endonuclease of the HUH superfamily [9], a group of enzymes responsible for the initiation of rolling-circle replication in various viruses and plasmids, as well as for the transposition of certain transposons [18–20]. Homologues of the SNJ1 RepA protein are encoded by small euryarchaeal plasmids, *Halorubrum saccharovorum* plasmid pZMX101 (3.9 kb) and *Methanosarcina acetivorans* plasmid pC2A (5.4 kb) [9]. Interestingly, however, whereas pC2A was predicted to replicate via the rolling-circle mechanism [21], it was suggested that pZMX101 employs a theta-like replication strategy [22], raising questions regarding the actual genome replication mechanism utilized by the two plasmids as well as by the virus SNJ1.

To gain insights into the replication mechanism of the SNJ1 genome and clarify the discrepancy reported in the

Received 15 November 2017; Accepted 4 January 2018
Author affiliations: [1]State Key Laboratory of Virology, College of Life Sciences, Wuhan University, Wuhan, PR China; [2]Department of Microbiology, Institut Pasteur, Unité Biologie Moléculaire du Gène chez les Extrêmophiles, Paris, France; [3]China Center for Type Culture Collection, Wuhan, PR China.
***Correspondence:** Xiangdong Chen, xdchen@whu.edu.cn; Mart Krupovic, krupovic@pasteur.fr
Keywords: archaeal viruses; halophilic viruses; genome replication; Sphaerolipoviridae.
Abbreviations: IS, insertion sequence; MGE, mobile genetic element; MBN, mung bean nuclease; MMC, mitomycin C; PSI-BLAST, Position-Specific Iterative Basic Local Alignment Search Tool; Rep, replication initiation protein.

literature, we analysed lysogenic *Natrinema* sp. J7-1 cells for the presence of single-stranded replicative intermediates of the SNJ1 genome, a signature of rolling-circle replication [20, 23–25]. To this end, total DNA was purified from the SNJ1 virions as well as from *Natrinema* sp. J7-1 cells with or without MMC induction of the SNJ1 provirus. The DNA preparations were electrophoresed on two 1 % agarose gels and soaked in either denaturing (Fig. 1a) or non-denaturing (Fig. 1b) buffer. The DNA bands were then transferred onto positively charged nylon membranes and hybridized with DIG-labelled probes that were complementary to either strand of the SNJ1 genome (Table 1). As expected, under denaturing conditions the probes hybridized with the SNJ1 genome extracted from both viral particles as well as from the *Natrinema* sp. J7-1 cells with or without MMC induction (Fig. 1a). By contrast, when the gel was soaked in the non-denaturing buffer a positive signal was obtained exclusively with the samples extracted from the *Natrinema* sp. J7-1 cells, and not from the SNJ1 virions (Fig. 1b). Notably, SNJ1-specific ssDNA was detected in both MMC-induced and non-induced samples, which was not unexpected given the stable replication of the SNJ1 provirus in *Natrinema* sp. J7-1 and the detection of RepA in *Natrinema* sp. J7-1 by Western blot, with and without MMC induction [9]. The observed ssDNA bands disappeared when the DNA preparations were treated with the ssDNA-specific mung bean nuclease (Fig. 1b). These results demonstrate the presence of ssDNA replicative intermediates of the SNJ1 genome in *Natrinema* sp. J7-1 cells, strongly suggesting that SNJ1 employs a rolling-circle mechanism of genome replication.

To further validate that RepA is a genuine member of the HUH superfamily, we performed site-directed mutagenesis of the conserved sequence motifs known to be essential for the catalytic activity in other HUH endonucleases. First, we targeted the diagnostic HUH motif (also known as motif 2 [26]), which coordinates a divalent metal ion facilitating DNA cleavage by localizing and polarizing the scissile phosphodiester bond [18]; the two His residues within the 'HVH' motif of SNJ1 RepA were changed to Ala (Fig. 2a). Second, we mutated the catalytic Tyr residue to Phe within motif 3 (Fig. 2a). The mutants were generated by amplification of the NcoI-SmaI restriction fragment of the SNJ1 genome containing the *repA* gene using the overlapping PCR approach with primers carrying the desired mutations (Table 1). After sequencing, the fragments were ligated into NcoI-SmaI predigested shuttle vector pYC-SHS, which relies on the SNJ1 *repA* gene for replication in *Natrinema* cells [9]. The resulting plasmids, pYC-SHS-Ala and pYC-SHS-Phe, carry mutations within motifs 2 and 3, respectively. The two vectors, as well as pYC-SHS, a positive control, were transformed into *Natrinema* sp. CJ7 cells and the transformation efficiencies were measured. Whereas the transformation of pYC-SHS yielded $0.85\pm0.17\times10^3$ c.f.u. per µg of DNA, the transformation efficiencies for pYC-SHS-Ala and pYC-SHS-Phe were both equal to zero in three independent experiments. This result confirms that the HUH motif (motif 2) and the catalytic Tyr residue in motif 3 (Fig. 2a) are both essential for the functionality of RepA.

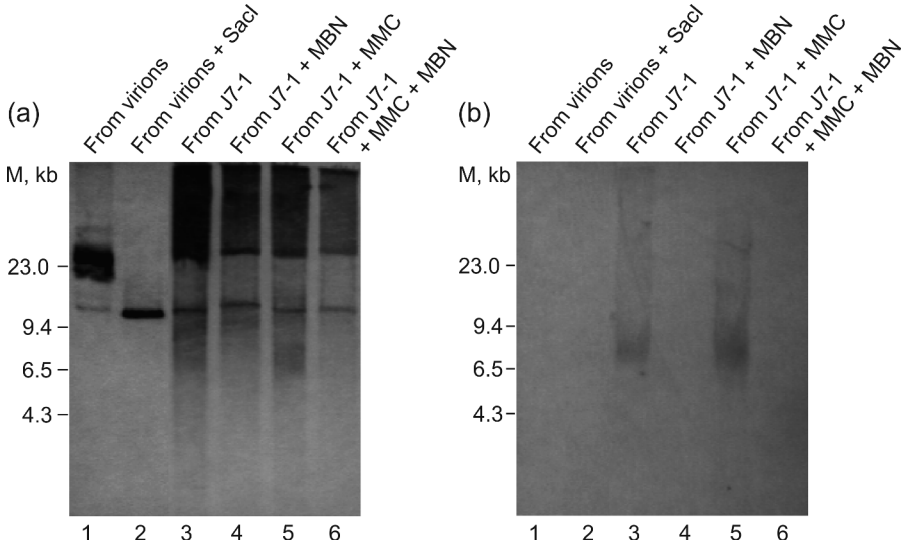

Fig. 1. Southern blot analysis of replicative SNJ1 DNA intermediates. The DNA preparations were electrophoresed on agarose gels at 30 V for 3 h and soaked in either denaturing (a) or non-denaturing (b) buffer. The DNA bands were then transferred onto positively charged nylon membranes and hybridized with DIG-labelled SNJ1-specific probes (Table 1). Lane 1, DNA extracted from SNJ1 virions; lane 2, DNA extracted from virions and linearized with SacI; lane 3, DNA extracted from *Natrinema* sp. J7-1 cells; lane 4, as in lane 3, but treated with ssDNA-specific mung bean nuclease (MBN); lane 5, DNA extracted from *Natrinema* sp. J7-1 cells after mitomycin C (MMC) induction; lane 6, as in lane 5, but treated with MBN. Denaturing buffer: 0.4 M NaOH and 1 M NaCl; non-denaturing buffer: 3 M NaCl and 0.3 M trisodium citrate dihydrate (pH=7, adjusted using NaOH).

Table 1. Primers and probes

Primers and probes	Sequences
Primers for mutagenesis	
SHS-3312-Phe-F	5′-CGAGCCGTCGTCTTCGCGCTCAGTCACTGCG-3′
SHS-3312-Phe-R	5′-CGCAGTGACTGAGCGCGAAGACGACGGCTCG-3′
SHS-3149-Ala-F	5′-AATACCGGCCGGCCGTTGCCCTGATCGGTG-3′
SHS-3149-Ala-R	5′-ATTAGGCACCGATCAGGGCAACGGCCGGCCGGT-3′
SHS-NcoI-1555-F	5′-AATCCATGGGAAACCTCCTAGCTAGCGCTC-3′
SHS-SmaI-6014-R	5′-TAACCCGGGGAGGGGTCACAGTTCCCTC-3′
SNJ1-specific probes	
ORF7-F	5′-CGTTGAGGCGACGAAGCTCGGTGA-3′
ORF11-R	5′-CCTCTTTCGTCAGCGTGTCCAAGCTCT-3′

In our previous work we reported that the closest homologues of the SNJ1 RepA are encoded in halophilic and methanogenic archaea from the phylum *Euryarchaeota* [9]. Given the substantial expansion of the sequence databases during the past 2 years, we decided to perform a more exhaustive sequence analysis of the SNJ1 RepA-like proteins to investigate the extent of their distribution and explore their relationship to other HUH endonucleases. Position-Specific Iterative Basic Local Alignment Search Tool (PSI-BLAST) searches [27] initiated with the SNJ1 RepA sequence resulted in the identification of homologues encoded in highly diverse archaea (Fig. 2b). Consistent with the previous analysis, the closest relatives, retrieved during the first iteration (E value cutoff 5e-03), were encoded in haloarhaea (class *Halobacteria*), but following additional iterations, more divergent (13–19 % identity) homologues were also identified in euryarchaea of the orders *Methanosarcinales* and *Thermoplasmatales*, as well as in ammonia-oxidizing archaea of the phylum *Thaumarchaeota*. Furthermore, we also identified a RepA homologue in an uncultivated archaeon belonging to the candidate MSBL1 division [28]. The wide spread across diverse archaeal lineages notwithstanding, the distribution of RepA-like proteins within the respective groups was highly sporadic, consistent with them being encoded by MGEs. Notably, besides the SNJ1 RepA, only three genes were carried by plasmids, whereas the rest of the homologues were encoded on cellular chromosomes, within the so-called archaeal 'dark matter' islands [29], most likely corresponding to integrated MGEs. Maximum-likelihood phylogenetic analysis has shown that RepA-like sequences form clades that are consistent with the host phylogeny (Fig. 2b), pointing towards coevolution of the corresponding MGEs with their respective hosts, with limited (if any) horizontal exchange between distantly related archaea.

Unexpectedly, following three PSI-BLAST iterations against the non-redundant NCBI database of prokaryotic and viral sequences, numerous (n>50) significant matches were obtained to transposases of the IS*91* family, rather than to rolling-circle replication initiation proteins encoded by other archaeal plasmids and viruses [23, 30–34]. The insertion sequences of the IS*91* family are predominantly found in gammaproteobacteria [35] and are usually located adjacent to pathogenicity- and virulence-related genes [36]. Unlike for the majority of bacterial IS elements, IS*91* transpose via a rolling-circle-like mechanism, which leads to the accumulation of ssDNA transposition intermediates [37] that are similar to those observed in this study for SNJ1 (Fig. 1). Consistently, IS*91* family transposases are related to the replication-initiation HUH endonucleases encoded by the pUB110 family plasmids [38]. A characteristic feature of IS*91* transposases is the presence of a unique N-terminal Zn-binding domain carrying six Cys residues, which precedes the HUH endonuclease domain (Fig. 2a) [36]. The homologous region between the SNJ1 RepA-like and IS*91*-like sequences detected by PSI-BLAST encompassed both the Zn-binding and the HUH endonuclease domains. Detailed inspection showed that all six Cys residues of the Zn-binding domain are conserved in SNJ1 RepA, although in some archaeal RepA-like homologues some of the Cys residues were substituted, occasionally to functionally equivalent His residues. Unlike bona fide IS*91*-like transposases, which have two active site Tyr residues (Y2 transposases) [18], all archaeal RepA-like proteins contain a single catalytic Tyr (Fig. 2a). Notably, the IS*91* family includes a subgroup of transposases encoded by IS*CR* elements, which are often associated with multiple antibiotic-resistance regions [39, 40] and contain only one Tyr residue in the active site (Y1 transposases) [18]. Y1 transposases are also encoded by bacterial IS*200*/IS*605* elements [18, 41]. However, the latter transposases do not contain the Zn-binding domain, display unique signatures within the endonuclease domain (Fig. 2a) and could not be retrieved by PSI-BLAST analysis, suggesting that they are not closely related to SNJ1 RepA-like sequences. To gain further insight into the evolution of SNJ1 RepA-like proteins, we collected a non-redundant dataset of IS*91*-like and IS*CR*-like transposases and performed maximum-likelihood phylogenetic (Fig. 2b) and sequence clustering (Fig. 2c) analyses.

Phylogenetic analysis has shown that archaeal SNJ1 RepA homologues form a well-supported monophyletic clade separated from the IS*91*-like and IS*CR*-like transposases

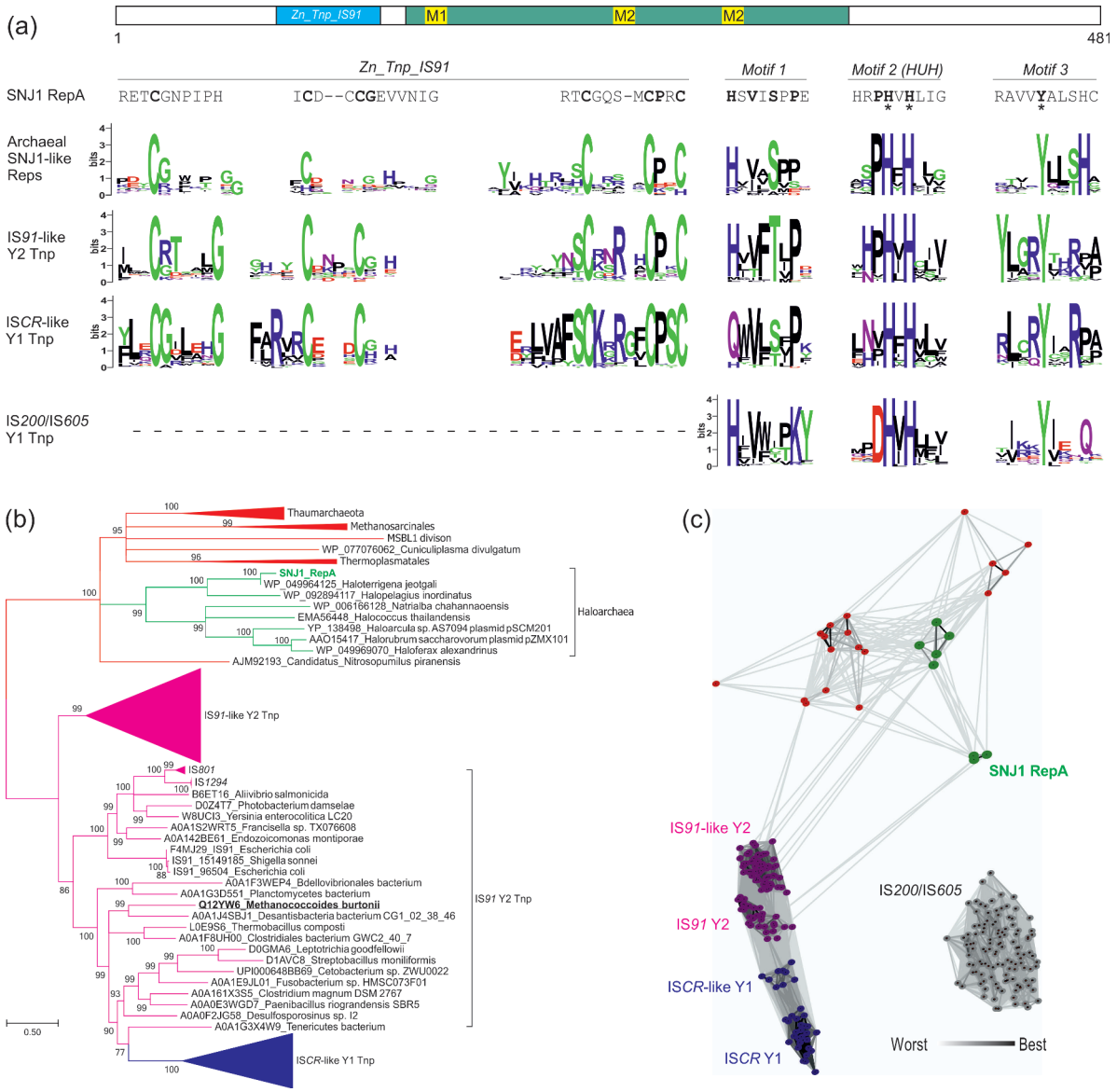

Fig. 2. Evolutionary relationship between SNJ1 RepA and other HUH endonucleases. (a) Top: domain organization of SNJ1 RepA and location of the conserved sequence motifs (M1-M3). Bottom: sequence conservation of the Zn-binding and HUH endonuclease domains in archaeal RepA-like proteins, IS*91*-like Y2 transposases (Tnp), IS*CR*-like Y1 Tnp and IS*200*/IS*605* Tnp. Sequence logos were generated using WebLogo [47]. Residues of the SNJ1 RepA subjected to site-directed mutagenesis are indicated by asterisks. (b) Maximum-likelihood phylogeny of SNJ1 RepA-like and IS*91*-like HUH endonucleases. Haloarchaea, green; other archaea, red; IS*91*-like Y2 transposases, magenta; IS*CR*-like Y1 transposases, blue. The IS*91*-like homologue from *Methanococcoides burtonii* is underlined. The sequences were aligned using MAFFT [48] and manually corrected, and the uninformative positions were removed using the *gappyout* function of the trimAL program [49]. Phylogenetic trees were constructed using the PhyML [50] with an automatic selection of the best-fit substitution model for a given alignment (WAG +G+I+F). A Bayesian-like transformation of aLRT (aBayes), as implemented in PhyML [50], was used to estimate branch support. (c) Clustering of archaeal SNJ1 RepA-like proteins with HUH transposases of the IS*91* and IS*200*/IS*605* families. The colour scheme is the same as in (b).

(Fig. 2b). Notably, the archaeal homologue of the IS*91* family Y2 transposases is found exclusively in the genome of a methanogen *Methanococcoides burtonii* [42], which in our phylogeny is nested among bacterial transposases, suggesting that IS*91* transposase was introduced into this archaeon relatively recently by horizontal transfer. IS*CR*-like Y1 transposases emerged from within one of the two major clades of bacterial IS*91*-like Y2 enzymes, suggesting their evolution from IS*91*-like Y2 transposases, independently from the Y1 SNJ1 RepA-like proteins. The results of the

sequence clustering analysis using CLANS [43] (Fig. 2c) were generally consistent with those of the phylogenetic analysis (Fig. 2b). Y2 and Y1 transposases each formed two discrete, tightly interconnected subclusters. The SNJ1 RepA-like sequences were considerably more loosely connected to each other and showed closest similarity to Y2 transposases rather than IS*C*R-like Y1 enzymes. By contrast, the IS*200*/IS*605* family transposases remained disconnected from either IS*91*-like or SNJ1 RepA-like HUH endonucleases (Fig. 2b), emphasizing the close evolutionary relationship between the latter two groups. The long branches in the phylogeny, as well as the results of clustering analysis, both suggest that the sequences of archaeal RepA homologues are more divergent than bacterial IS*91* and IS*CR* family transposases, which is indicative of an ancient association and diversification of RepA-like endonucleases in archaea.

Collectively, our results indicate that IS*91*-like transposases are the closest known relatives of the archaeal SNJ1 RepA-like rolling-circle replication initiation proteins, and that the two groups have shared the most recent common ancestor, which displayed the characteristic domain architecture with the N-terminal Zn-binding and C-terminal HUH endonuclease domains. Although it is not possible at this point to confidently conclude whether the ancestral enzyme of this group functioned as a transposase or a replication initiator, it is clear that in the evolution of HUH endonucleases alternations between replication and transposition have occurred on multiple independent occasions. For instance, the HUH family endonuclease responsible for the rolling-hairpin replication of the linear ssDNA genomes of parvoviruses is known to mediate site-specific integration of the viral genome into the host chromosome [44] and some parvoviruses have apparently switched to exclusively transposon-like propagation [45]. It has been further suggested that many ssDNA viruses, particularly in eukaryotes, occasionally integrate into the cellular genomes with the aid of the rolling-circle replication initiation proteins (reviewed in [46]). It is thus conceivable that the integration of an ancestral plasmid or viral genome into the cellular chromosome, as observed in the archaeal genomes described in this study, and subsequent fixation of the corresponding MGEs could promote the refunctionalization of their Rep proteins into transposases. More generally, transitions between different classes of MGEs appear to be a pervasive and common phenomenon in the evolution of viruses, plasmids and transposons, which amalgamates them into one global genetic network.

Funding information
This work was supported by grants from the National Natural Science Foundation of China (no. 31570174), the National Fund for Fostering Talents of Basic Sciences (J1103513) and the Research (Innovative) Fund of Laboratory of Wuhan University. M. K is supported by l'Agence Nationale de la Recherche (France) project ENVIRA. D. P. is supported by the European Union's Horizon 2020 research and innovation programme under grant agreement 685778, project VIRUS-X.

Conflicts of interest
The authors declare that that there are no conflicts of interest.

Ethical statement
This article does not describe experimental work with humans.

References

1. Prangishvili D, Bamford DH, Forterre P, Iranzo J, Koonin EV *et al.* The enigmatic archaeal virosphere. *Nat Rev Microbiol* 2017;15:724–739.
2. Snyder JC, Bolduc B, Young MJ. 40 Years of archaeal virology: expanding viral diversity. *Virology* 2015;479-480:369–378.
3. Iranzo J, Krupovic M, Koonin EV. The double-stranded DNA virosphere as a modular hierarchical network of gene sharing. *MBio* 2016;7:e00978-16.
4. Krupovic M, Cvirkaite-Krupovic V, Iranzo J, Prangishvili D, Koonin EV. Viruses of archaea: Structural, functional, environmental and evolutionary genomics. *Virus Res* 2018;244:181–193.
5. Iranzo J, Koonin EV, Prangishvili D, Krupovic M. Bipartite network analysis of the archaeal virosphere: evolutionary connections between viruses and capsidless mobile elements. *J Virol* 2016;90:11043–11055.
6. Krupovic M, Béguin P, Koonin EV. Casposons: mobile genetic elements that gave rise to the CRISPR-Cas adaptation machinery. *Curr Opin Microbiol* 2017;38:36–43.
7. Martínez-Alvarez L, Bell SD, Peng X. Multiple consecutive initiation of replication producing novel brush-like intermediates at the termini of linear viral dsDNA genomes with hairpin ends. *Nucleic Acids Res* 2016;44:8799–8809.
8. Pina M, Basta T, Quax TE, Joubert A, Baconnais S *et al.* Unique genome replication mechanism of the archaeal virus AFV1. *Mol Microbiol* 2014;92:1313–1325.
9. Wang Y, Sima L, Lv J, Huang S, Liu Y *et al.* Identification, characterization, and application of the replicon region of the halophilic temperate sphaerolipovirus SNJ1. *J Bacteriol* 2016;198:1952–1964.
10. Gardner AF, Bell SD, White MF, Prangishvili D, Krupovic M. Protein-protein interactions leading to recruitment of the host DNA sliding clamp by the hyperthermophilic *Sulfolobus islandicus* rod-shaped virus 2. *J Virol* 2014;88:7105–7108.
11. Zhang Z, Liu Y, Wang S, Yang D, Cheng Y *et al.* Temperate membrane-containing halophilic archaeal virus SNJ1 has a circular dsDNA genome identical to that of plasmid pHH205. *Virology* 2012;434:233–241.
12. Pawlowski A, Rissanen I, Bamford JK, Krupovic M, Jalasvuori M. Gammasphaerolipovirus, a newly proposed bacteriophage genus, unifies viruses of halophilic archaea and thermophilic bacteria within the novel family *Sphaerolipoviridae*. *Arch Virol* 2014;159:1541–1554.
13. Demina TA, Pietilä MK, Svirskaitė J, Ravantti JJ, Atanasova NS *et al.* Archaeal *Haloarcula californiae* Icosahedral virus 1 highlights conserved elements in icosahedral membrane-containing DNA viruses from extreme environments. *MBio* 2016;7:e00699-16.
14. Porter K, Kukkaro P, Bamford JK, Bath C, Kivelä HM *et al.* SH1: a novel, spherical halovirus isolated from an Australian hypersaline lake. *Virology* 2005;335:22–33.
15. Porter K, Tang SL, Chen CP, Chiang PW, Hong MJ *et al.* PH1: an archaeovirus of *Haloarcula hispanica* related to SH1 and HHIV-2. *Archaea* 2013;2013:1–17.
16. Jaakkola ST, Penttinen RK, Vilén ST, Jalasvuori M, Rönnholm G *et al.* Closely related archaeal Haloarcula hispanica icosahedral viruses HHIV-2 and SH1 have nonhomologous genes encoding host recognition functions. *J Virol* 2012;86:4734–4742.
17. Mei Y, Chen J, Sun D, Chen D, Yang Y *et al.* Induction and preliminary characterization of a novel halophage SNJ1 from lysogenic *Natrinema* sp. F5. *Can J Microbiol* 2007;53:1106–1110.

18. Chandler M, de La Cruz F, Dyda F, Hickman AB, Moncalian G *et al.* Breaking and joining single-stranded DNA: the HUH endonuclease superfamily. *Nat Rev Microbiol* 2013;11:525–538.
19. Krupovic M. Networks of evolutionary interactions underlying the polyphyletic origin of ssDNA viruses. *Curr Opin Virol* 2013;3:578–586.
20. Khan SA. Plasmid rolling-circle replication: highlights of two decades of research. *Plasmid* 2005;53:126–136.
21. Metcalf WW, Zhang JK, Apolinario E, Sowers KR, Wolfe RS. A genetic system for Archaea of the genus *Methanosarcina*: liposome-mediated transformation and construction of shuttle vectors. *Proc Natl Acad Sci USA* 1997;94:2626–2631.
22. Zhou L, Zhou M, Sun C, Xiang H, Tan H. Genetic analysis of a novel plasmid pZMX101 from *Halorubrum saccharovorum*: determination of the minimal replicon and comparison with the related haloarchaeal plasmid pSCM201. *FEMS Microbiol Lett* 2007;270:104–108.
23. Forterre P, Krupovic M, Raymann K, Soler N. Plasmids from *Euryarchaeota*. *Microbiol Spectr* 2014;2:PLAS-0027-2014.
24. Te Riele H, Michel B, Ehrlich SD. Single-stranded plasmid DNA in *Bacillus subtilis* and *Staphylococcus aureus*. *Proc Natl Acad Sci USA* 1986;83:2541–2545.
25. Zhou M, Xiang H, Sun C, Li Y, Liu J *et al.* Complete sequence and molecular characterization of pNB101, a rolling-circle replicating plasmid from the haloalkaliphilic archaeon *Natronobacterium* sp. strain AS7091. *Extremophiles* 2004;8:91–98.
26. Ilyina TV, Koonin EV. Conserved sequence motifs in the initiator proteins for rolling circle DNA replication encoded by diverse replicons from eubacteria, eucaryotes and archaebacteria. *Nucleic Acids Res* 1992;20:3279–3285.
27. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 1997;25:3389–3402.
28. van der Wielen PW, Bolhuis H, Borin S, Daffonchio D, Corselli C *et al.* The enigma of prokaryotic life in deep hypersaline anoxic basins. *Science* 2005;307:121–123.
29. Makarova KS, Wolf YI, Forterre P, Prangishvili D, Krupovic M *et al.* Dark matter in archaeal genomes: a rich source of novel mobile elements, defense systems and secretory complexes. *Extremophiles* 2014;18:877–893.
30. Gorlas A, Krupovic M, Forterre P, Geslin C. Living side by side with a virus: characterization of two novel plasmids from *Thermococcus prieurii*, a host for the spindle-shaped virus TPV1. *Appl Environ Microbiol* 2013;79:3822–3828.
31. Pietilä MK, Roine E, Sencilo A, Bamford DH, Oksanen HM. *Pleolipoviridae*, a newly proposed family comprising archaeal pleomorphic viruses with single-stranded or double-stranded DNA genomes. *Arch Virol* 2016;161:249–256.
32. Erauso G, Marsin S, Benbouzid-Rollet N, Baucher MF, Barbeyron T *et al.* Sequence of plasmid pGT5 from the archaeon *Pyrococcus abyssi*: evidence for rolling-circle replication in a hyperthermophile. *J Bacteriol* 1996;178:3232–3237.
33. Holmes ML, Pfeifer F, Dyall-Smith ML. Analysis of the halobacterial plasmid pHK2 minimal replicon. *Gene* 1995;153:117–121.
34. Sencilo A, Paulin L, Kellner S, Helm M, Roine E. Related haloarchaeal pleomorphic viruses contain different genome types. *Nucleic Acids Res* 2012;40:5523–5534.
35. Siguier P, Filée J, Chandler M. Insertion sequences in prokaryotic genomes. *Curr Opin Microbiol* 2006;9:526–531.
36. Garcillán-Barcia MP, de La Cruz F. Distribution of IS91 family insertion sequences in bacterial genomes: evolutionary implications. *FEMS Microbiol Ecol* 2002;42:303–313.
37. del Pilar Garcillán-Barcia M, Bernales I, Mendiola MV, de La Cruz F. Single-stranded DNA intermediates in IS91 rolling-circle transposition. *Mol Microbiol* 2001;39:494–502.
38. Mendiola MV, de La Cruz F. IS91 transposase is related to the rolling-circle-type replication proteins of the pUB110 family of plasmids. *Nucleic Acids Res* 1992;20:3521.
39. Ilyina TS. Mobile ISCR elements: structure, functions, and role in the emergence, increasing and spreading of blocks of bacterial genes of multiple antibiotic resistance. *Mol Gen Mikrobiol Virusol* 2012;4:135–146.
40. Toleman MA, Bennett PM, Walsh TR. ISCR elements: novel gene-capturing systems of the 21st century? *Microbiol Mol Biol Rev* 2006;70:296–316.
41. He S, Corneloup A, Guynet C, Lavatine L, Caumont-Sarcos A *et al.* The IS200/IS605 family and "Peel and Paste" single-strand transposition mechanism. *Microbiol Spectr* 2015;3:MDNA3-0039-2014.
42. Filée J, Siguier P, Chandler M. Insertion sequence diversity in archaea. *Microbiol Mol Biol Rev* 2007;71:121–157.
43. Frickey T, Lupas A. CLANS: a Java application for visualizing protein families based on pairwise similarity. *Bioinformatics* 2004;20:3702–3704.
44. Henckaerts E, Dutheil N, Zeltner N, Kattman S, Kohlbrenner E *et al.* Site-specific integration of adeno-associated virus involves partial duplication of the target locus. *Proc Natl Acad Sci USA* 2009;106:7571–7576.
45. Liu H, Fu Y, Li B, Yu X, Xie J *et al.* Widespread horizontal gene transfer from circular single-stranded DNA viruses to eukaryotic genomes. *BMC Evol Biol* 2011;11:276.
46. Krupovic M, Forterre P. Single-stranded DNA viruses employ a variety of mechanisms for integration into host genomes. *Ann N Y Acad Sci* 2015;1341:41–53.
47. Crooks GE, Hon G, Chandonia JM, Brenner SE. WebLogo: a sequence logo generator. *Genome Res* 2004;14:1188–1190.
48. Katoh K, Rozewicki J, Yamada KD. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Brief Bioinform* 2017.
49. Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 2009;25:1972–1973.
50. Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W *et al.* New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol* 2010;59:307–321.