A Novel Type of Polyhedral Viruses Infecting Hyperthermophilic Archaea

Ying Liu, Sonoko Ishino, Yoshizumi Ishino, Gerard Pehau-Arnaudet, Mart Krupovic, David Prangishvili

JVI Accepted Manuscript Posted Online 19 April 2017
J. Virol. doi:10.1128/JVI.00589-17

A novel type of polyhedral viruses infecting hyperthermophilic archaea

Running title: New polyhedral virus

Ying Liu,[a] Sonoko Ishino,[b] Yoshizumi Ishino,[b] Gérard Pehau-Arnaudet,[c] Mart Krupovic,[a#] David Prangishvili[a#]

Department of Microbiology, BMGE, Institut Pasteur, 25-28 rue du Dr. Roux, 75015 Paris, France[a] ; Department of Bioscience and Biotechnology, Kyushu University, 6-10-1 Hakozaki, Higashi-ku, Fukuoka, Fukuoka 812-8581, Japan[b]; Ultrapole, UMR 3528, CNRS, Institut Pasteur, 25-28 rue du Dr. Roux, 75015 Paris, France[c]

Address correspondence to: Mart Krupovic, krupovic@pasteur.fr and David Prangishvili, david.prangishvili@pasteur.fr

Word count: for the abstract –171, for the text – 4267.

ABSTRACT Encapsidation of genetic material into polyhedral particles is one of the most common structural solutions employed by viruses infecting hosts in all three domains of life. Here, we describe a new virus of hyperthermophilic archaea, *Sulfolobus* polyhedral virus 1 (SPV1), which condenses its circular double-stranded DNA genome in a manner not previously observed for other known virus. The genome complexed with virion proteins is wound up sinusoidally into a spherical coil which is surrounded by an envelope and further encased by an outer polyhedral capsid apparently composed of the 20 kDa virion protein. Lipids selectively acquired from the pool of host lipids are integral constituents of the virion. None of the major virion proteins of SPV1 show similarity to structural proteins of known viruses. However, minor structural proteins, which are predicted to mediate host recognition, are shared with other hyperthermophilic archaeal viruses infecting members of the order Sulfolobales. The SPV1 genome consists of 20,222 bp and encodes 45 open reading frames, only one fifth of which could be functionally annotated.

IMPORTANCE Viruses infecting hyperthermophilic archaea display a remarkable morphological diversity, often presenting architectural solutions not explored employed by known viruses of bacteria and eukaryotes. Here we present the isolation and characterization of *Sulfolobus* polyhedral virus 1, which condenses its genome into a unique spherical coil. Due to the original genomic and architectural features of SPV1, the virus should be considered as a representative of a new viral family.

KEYWORDS archaea, hyperthermophile, virion structure, viral genome

Journal of Virology

One of the most unexpected results of recent studies on viral diversity is the unveiling of astounding morphological variety of DNA viruses in geothermal environments with temperatures exceeding 80°C (1). Hardly over two dozen viruses isolated from such environments, all parasitizing hyperthermophilic Archaea, display diverse virion shapes many of which were never observed among viruses from the two other domains of life, Bacteria and Eukarya. Due to distinct genome compositions and peculiar morphological properties of these viruses, 11 new virus families were established by ICTV for their classification (1-4).

The remarkable morphological diversity of hyperthermophilic archaeal viruses radically contrasts the limited structural diversity of dsDNA viruses of Bacteria, nearly all of which, with the exception of several pleomorphic species, have icosahedral particles, with or without helical tails (5). The existing picture seems to accurately reflect the actual differences between bacterial and archaeal viruses. The morphological landscape of bacterial viruses is formed by over six thousand species, whereas that of hyperthermophilic archaeal viruses – by only a couple of dozens known species. Moreover, while no new morphotype of bacterial dsDNA virus has been identified since half a century, despite the isolation of thousands of new species (5), the variety of archaeal viral morphotypes continues to expand with the description of viruses from new environments and hosts (6-15). There are reasons to believe that the known variety of archaeal viruses represents no more than the tip of an iceberg and that comprehensive information on them may shed light on the problems of origin and evolution of viruses and virus-host interactions (16-28).

In attempts to further contribute to the knowledge on the diversity of archaeal viruses, we have analyzed virus-host systems in the hot, acidic springs in Beppu, Japan. Here, we report on the isolation and characterization of a polyhedral hyperthermophilic archaeal virus, *Sulfolobus* polyhedral virus 1 (SPV1), with a type of virion organization not previously observed for other known icosahedral viruses.

RESULTS

Virus isolation and virus-host relationships. The aliquots of an environmental sample collected from the hot, acidic spring Umi Jigoku in Beppu, Japan, were used to inoculate the medium favorable for the growth of members of the genus *Sulfolobus*, which are known to dominate in acid thermal environments. After incubation at 75°C for 10 days, cell growth was detected and the presence of polyhedral virus-like-particles (VLPs) was observed in the enrichment culture by transmission electron microscopy (TEM). The particles appeared to be

isometric, likely icosahedral, uniform in overall appearance and size (not shown). From the same culture, 50 isolates, numbered S1 to S50, were colony-purified. None of the isolates was capable of the VLP production. They were further analyzed for the ability to replicate the observed VLPs. Aliquots of the cell-free enrichment culture (5 µl) were applied onto Phytagel plates that contained cells of each of the 50 isolates, prior to the lawn development. Cells of 40 isolates grew as lawns on the plates after incubation at 75°C for three days. Turbid zones of growth inhibition were formed around drops applied on the lawn formed by the isolate S38. The 16S rDNA sequence of the isolate revealed that it represents a species from the genus *Sulfolobus*. From the lawn of *Sulfolobus* sp. S38 the zone of growth inhibition was excised and placed in the growing cell culture of the same species. As a result, the production of polyhedral VLPs was observed in the culture. For the verification of the infectious nature of these particles, they were collected from the cell-free culture supernatant and added to the growing culture of *Sulfolobus* sp. S38. A dramatic increase of the VLP concentration, as observed by TEM, indicated effective replication of the particles and confirmed that they represent infectious virions of a *Sulfolobus* virus.

Sulfolobus sp. S38 was subjected to an additional round of colony purification and the isolated strain, designated *Sulfolobus* sp. S38A, was selected as a standard virus host for all following experiments. The 16S rRNA gene sequence revealed that the isolate represents a strain of *Sulfolobus shibatae*, with the highest identity (99%) to the sequence of *S. shibatae* B12. The polyhedral virus that replicated in *Sulfolobus* sp. S38A was named *Sulfolobus* polyhedral virus 1 (SPV1). The virus SPV1 was purified by 5-20% sucrose rate-zonal centrifugation followed by isopicnic centrifugation in the gradient of CsCl. The morphology and size of SPV1 virions was identical to that of the VLPs observed in the enrichment culture (Fig. 1).

To study the host range of SPV1, purified virus particles were added to exponentially growing cultures of potential hosts and virus propagation was monitored by TEM. Besides *Sulfolobus* sp. S38A, the virus could replicate in *S. islandicus* strain REY15A (29). *S. islandicus* strains HVE10/4, LAL14, *S. solfataricus* strains P1, P2, and 98/2 as well as *S. acidocaldarius* did not serve as hosts for the virus SPV1.

Infection of exponentially growing cells of *Sulfolobus* sp. 38A with SPV1 at an MOI of about 7 caused only slight retardation of host growth and did not lead to host cell death and lysis, as validated by the OD_{600} measurements and enumeration of viable cells in the cell culture (Fig. 2). In line with these observations, application of 5 µl of concentrated virus preparation onto freshly prepared lawns of *Sulfolobus* sp.38A did not cause, after development of the lawns, appearance

of zones of cell lysis (not shown). SPVConsequently, we failed to establish plaque assays on the lawn of *Sulfolobus* sp. S38A cells. Thus, virus titre was estimated by TEM, based on the comparison of SPV1 particle numbers with the numbers of virions in a standard preparation of *Sulfolobus* rod-shaped virus 2 with the known titre. Significant increase of SPV1 titre in the culture of infected cells could be observed about 24 h post infection and did not cause decrease in the turbidity of the cell culture (Fig. 2). All these observations suggest that SPV1 may be a non-lytic virus.

Virion structure. The virions of SPV1, as observed by TEM and cryo-electron microscopy (cryo-EM), represent polyhedra with the diameter of about 87 nm from vertex to vertex, and 83 nm from facet to facet (Fig.1A,B). They carry well-distinguishable outer shell which can be removed partially (Fig. 3A,B) or completely (Fig. 3C,D) from the intact virions by one cycle of freezing and thawing or prolonged storage. The shell seems to be responsible for the polyhedral shape of the virions. The cores remaining after removal of the shell have slightly pleomorphic, spherical shape (Fig. 3C,D). Protrusions were observed on the surface of the core and can represent either remains of the outer shell or specific structures mediating the contact between the core and the outer shell (Fig. 3C,D filled arrowheads). The cores consist of the envelope (Fig. 3D, black arrow) and viral DNA packaged in the form of condensed nucleoprotein filament (Fig.3D). The nucleoprotein filament remained condensed even after partial detachment of the core envelope (Fig. 3C, inset). In such cases, unwinding and release of the nucleoprotein filament could be observed (Fig. 3C, open arrowheads).

Analysis of the cryo-EM micrographs of intact virions taken in two different projections provided information on the organization of the nucleoprotein filament in the virion (Fig. 1B). The patterns projected in the axial views depict 7 concentric rings (Fig. 1B, open arrowhead), whereas those projected in the side views depict 14 linear striations (Fig. 1B, filled arrowhead). The observations suggest that the nucleoprotein filament, about 3 nm in width, is wound up sinusoidally into a spherical coil. This spherical coil is about 60 nm in diameter and comprises 14 loops of the nucleoprotein filament with varying diameters and even spacing between them (Fig. 1B).

Virus genome. The genome of SPV1 was extracted from highly purified virions and sequenced using the Illumina platform. Assembly of the sequencing reads resulted in a single circular contig of 20,222 bp. The viral genome was sensitive to various type II restriction endonucleases; digestion of the genome with the single-cutters XbaI and BglII resulted in a linearized product, which migrated in the agarose gel as a single sharp band, consistent with the genome being a circular dsDNA molecule. The SPV1 genome has a GC content of 38.3%,

which is similar to that of various *Sulfolobus* genomes (e.g., 35% for *S. isladicus* (30)). The SPV1 genome contains 45 open reading frames (ORF), which are tightly arranged and occupy 89.1% of the genome (Fig. 4). The ORFs are generally short, with median length of 103 codons.

Homology searches using the BLASTP program (31) revealed that less than one third (~27%) of SPV1 gene products are significantly similar (cutoff of E=1e-03) to sequences in the nonredundant protein database (Table 1). Notably, 10 ORFs had homologs in various Sulfolobales viruses, including members of the families *Rudiviridae*, *Lipothrixviridae*, *Fuselloviridae* and *Turriviridae*. Homologs of two additional ORFs were encoded in the genomes of members of the order Sulfolobales, rather than their viruses. A combination of BLASTP and a more sensitive hidden Markov model (HMM)-based HHpred (32) analyses allowed the assignment of putative functions to just one fifth of all SPV1 ORFs (nine ORFs, 20%). Seven ORFs encoded putative proteins containing various DNA-binding domains, including zinc finger (ORF02-234, ORF07-40, ORF15-65), (winged) helix-turn-helix (ORF16-96, ORF22-147, ORF35-111), and ribbon-helix-helix (ORF23-115) domains, which are frequently encoded by crenarchaeal viruses (33-35). In addition, ORF29-310 and ORF32-168 encode glycosyl transferase with the GT-B fold and the S-adenosyl-L-methionine-dependent methyltransferase, respectively. The closest homologs of the two proteins are encoded by *Acidianus*-infecting viruses of the order *Ligamenvirales*. In addition, ORF25-400 and ORF26-357 are homologous to each other and to the minor structural proteins conserved in all members of the family *Rudiviridae* (36) as well as in some members of the families *Lipothrixviridae* and *Bicaudaviridae*, all infecting members of the archaeal order Sulfolobales. Notably, whereas the N-terminal regions of ORF25-400 and ORF26-357 products showed highest similarity to the viral homologs, the central regions of the corresponding proteins were more similar to homologs encoded by *Candidatus* Nanopusillus acidilobi, a symbiotic archaeon of the order Nanoarchaeota inhabiting terrestrial geothermal environments (37).

Virion proteins. Highly purified virions of SPV1 were analyzed by SDS-PAGE. Three prominent bands (bands B1, B2, and B3), and four weaker bands (B4, B5, B6 and B7), were observed on the gel stained with Coomassie Brilliant Blue (Fig. 5A). The protein content in each separate band was analyzed by liquid chromatography and tandem mass spectrometry (LC-MS/MS) (Fig. 5A,C). The results of the analysis have revealed that the protein in band B1 with apparent molecular mass of about 8 kDa is encoded by SPV1 ORF45-78 (Fig. 5C). We denote it as viral protein 1 (VP1). The prominent band B2 contained two proteins — both with apparent molecular mass of 12 kDa — encoded by SPV1 ORF35-111 and ORF44-119, which we denote as VP2 and VP3, correspondingly (Fig. 5A,C). The third prominent band, B3, covers a wide area

corresponding to proteins with the apparent molecular mass in the range of 20 to 32 kDa. It contains a single protein encoded by SPV1 ORF43, denoted as VP4 (Fig. 5A,C). The traces of protein VP4 we detected also across the whole length of the gel above B3. The diffused appearance of the band B3 could be caused by different degrees of putative posttranscriptional modifications of the VP4 as well as formation of protein multimers. Along with the genes of the four major virion proteins, the genes of the minor virion proteins were also identified: ORF42-304 (band B4, VP5), ORF41-305 (band B5, VP6), ORF40-337 (band B6, VP7), ORF25-400 and ORF26-357 (band B7, VP8 and VP9, correspondingly) (Fig. 5A,C).

Sequence analysis of the virion components showed that products of ORF40-337, ORF41-305, ORF42-304 and ORF44-119 are predicted to be integral membrane proteins with 1 to 3 membrane spanning α-helical segments. This result suggests that the four minor structural proteins might be embedded within the envelope surrounding the virion core and located beneath the outer shell. The five remaining virion proteins are predicted to be soluble. Among these, VP1 encoded by ORF45-78 is a highly-basic (pI=9.53), α-helical protein of 78 amino acids, which due to its positive charge is likely to bind to the viral genome. Interestingly, VP2 is also α-helical and is homologous to helix-turn-helix domain-containing transcription factors ubiquitous in archaea (but not viruses), suggesting that it is also a DNA-binding protein. The VP4, the abundant protein constituent of the SPV1 virion, shares no similarity to other cellular or viral proteins. Importantly, partially disassembled virions in which the outer shell was largely removed by freezing and thawing contained a dramatically reduced amount of VP4, whereas the amounts of VP1-3 remained rather constant (Fig. 5B). This result suggests that VP4 is the main components of the external polyhedral protein shell. Analysis of the predicted secondary structure suggests that VP4 is rich in β-strands. Finally, given that homologs of virion proteins encoded by ORF25-400 and ORF26-357 are found in morphologically different viruses infecting archaea of the order Sulfolobales, the two proteins are likely to mediate certain aspects of virus-host interactions. Consistently, both proteins contain the functionally uncharacterized domain DUF2341, which is widespread in bacterial and archaeal proteins and is usually fused to various lectin, immunoglobulin and CARDB-like cell adhesion domains (PF10102), suggesting that products of ORF25-400 and ORF26-357 are also likely to participate in host recognition.

Virion lipids. Lipids were extracted from purified SPV1 virions and analyzed by thin layer chromatography, as described in Materials and Methods. One major and several minor bands were observed on the chromatogram (Fig. 6). The pattern was dramatically different from that of the host lipids (Fig. 6), in the number of lipid types as well as in their relative abundance, and suggested selective acquisition of viral lipids from the pool of host lipids. Most likely, the lipids

are principal components of the envelope of the virion core. Notably, similar selectivity towards particular lipid species has been observed in several other archaeal viruses (38-40).

DISCUSSION

All viruses with polyhedral capsids known to date obey icosahedral symmetry. Thus, SPV1 is also likely to have icosahedral capsid, but this supposition remains to be verified experimentally. Viruses with icosahedral virions represent more than half of all recognized virus families (41). These viruses display remarkable diversity of virion sizes and complexity, from small capsids composed of a single protein, like in the case of simple ssRNA and ssDNA viruses (42, 43), to extravagantly large, multilayered virions, such as those of mimiviruses (44). There are 3 major architectural classes of viruses with icosahedral capsids and dsDNA genomes, two of which are featured by viruses infecting hosts in all three cellular domains, testifying for their antiquity (41). The first of these architectural classes includes viruses which build their capsids using the MCPs with the so-called HK97-like fold (45). Bacterial and archaeal viruses with the HK97-like MCPs are classified into the order *Caudovirales*, whereas eukaryotic members belong to the order *Herpesvirales*. Besides the MCPs, these viruses employ similar virion assembly and maturation as well as genome packaging mechanisms, which are not found in viruses outside of this virus assemblage (46). The second architectural class encompasses highly diverse dsDNA viruses with the vertical double-jelly roll (DJR) MCPs, including members of the bacterial virus families *Corticoviridae* and *Tectiviridae*, archaeal viruses of the family *Turriviridae*, and eukaryotic viruses of the families *Adenoviridae* and *Lavidaviridae* as well as the proposed order 'Megavirales' that comprises most of the large and giant eukaryotic viruses (47-49). Most viruses within this class contain an internal membrane vesicle located between the protein capsid and the dsDNA genome, and encode homologous A32-like genome packaging ATPases not found in other virus groups. Notably, internal membrane-containing bacterial and archaeal viruses of the family *Sphaerolipoviridae* encode two paralogous MCPs with single jelly-roll folds which form capsomers similar to those formed from the DJR MCPs (50, 51). Thus, the latter group of viruses is considered to be evolutionarily related to, and possibly predate, viruses with the DJR MCPs (52). The third architectural class of viruses includes eukaryotic dsDNA viruses with small genomes (~5–8 kb), namely members of the families *Polyomaviridae* and *Papillomaviridae*. Viruses of the latter group encode single jelly-roll MCPs and are thought to have evolved from eukaryotic ssDNA viruses (42, 53). SPV1 does not fall into any of the three architectural classes and displays several marked differences compared to the members of these classes, as detailed below.

The VP4, which apparently forms the outer shell of the SPV1 virion and is equivalent to the MCPs of other polyhedral viruses, is not recognizably similar to any other viral or cellular protein. The secondary structure prediction suggests that 181 aa-long VP4 contains 10 β-strands and 2 α-helixes. Thus, based on the secondary structure prediction and its sheer size, the protein is highly unlikely to adopt either the DJR or HK97-like fold. Although the predicted secondary structure is consistent with the single jelly-roll fold, SPV1 does not seem to be related to sphaerolipoviruses, all of which employ two MCP for the formation of external icosahedral capsid. Similarly, polyomaviruses and papillomaviruses are architecturally very different from SPV1, in that their genomes are considerably smaller (4.7–8.4 kb) and they do not contain the internal enveloped nucleoprotein core observed in SPV1. Furthermore, unlike dsDNA viruses with the DJR and HK97-like MCPs, which encode distinct types of genome packaging motors, SPV1 lacks recognizable genes for putative ATPases, a class of proteins which is typically among the easiest to recognize by sequence analysis due to the presence of highly conserved Walker A and Walker B motifs. Instead, SPV1 condenses its genome in the form of a unique spherical core, consisting of the highly ordered nucleoprotein complex.

The protein VP2, one of the major structural proteins of SPV1, is homologous to ubiquitous archaeal transcription regulators with the helix-turn-helix motif (Table 1). Notably, some archaeal viruses, such as spindle-shaped virus SSV1, incorporate host-encoded chromatin proteins into their virions (38). It is likely that in the course of evolution, the cellular gene encoding for the DNA-binding protein has been horizontally acquired by the ancestor of SPV1 and adapted to function as a structural component of the virion. Indeed, recruitment of cellular proteins for virion structure appears to be a recurrent theme in the evolution of viruses (41).

Whereas the proteins involved in virion formation, with a notable exception of the putative receptor-binding proteins encoded by ORF25-400 and ORF26-357, are specific to SPV1, a considerable fraction of SPV1 genes are shared with other hyperthermophilic viruses infecting members of the Sulfolobales. The latter category of genes includes several DNA-binding proteins, a glycosyltransferase and a methyltransferase. This observation is in agreement with the recent bipartite network analysis of the known archaeal virosphere which revealed that archaeal virus network consists of 10 modules which are linked via connector genes encoding a small set of widespread proteins, most notably the ribbon-helix-helix domain-containing transcription factors and glycosyltransferases (33), neither of which is a viral hallmark protein (54). However, such a lack of strong connectivity among the modules indicates that most of the viral groups within the archaeal virosphere are evolutionarily distinct. SPV1 integrates into the

global archaeal virus network on the account of the same connector genes but is otherwise unrelated to other archaeal viruses.

Due to the unique genomic and architectural features of SPV1, we propose that the virus be considered as the first representative of a new viral family, which we tentatively name "Portogloboviridae" (from Latin *porto*; to bear, carry, and *globus*; a ball).

MATERIALS AND METHODS

Enrichment culture, isolation of SPV1 and host strain. The environmental sample of translucent liquid mixed with red sand was collected from hot acidic spring Umi Jigoku (80°C, pH 3.7) in Beppu, Japan. An aliquot (10 ml) of the sample was used to inoculate 40 ml of *Sulfolobus* growth medium (55), and the culture was incubated for 10 days at 75°C in aerobic conditions without shaking. Cell-free culture supernatant containing VLPs was collected by centrifugation, followed by filtration through a filter membrane of 0.22 µm pore size (Merck Millipore). The single strains were colony-purified from the enrichment culture by plating on Phytalgel™ (Sigma-Aldrich) plates and incubated for 5 days at 75°C as described previously (55), except that Gelzan™ CM Gelrite was substituted with Phytalgel™. The oligo primers used for 16S rRNA gene amplification are A21F: 5'-TTCCGGTTGATCCYGCCGGA-3' and U1525R: 5'-AAGGAGGTGATCCAGCC-3'.

The following collection strains were analyzed for the ability to replicate the virus SPV1: *S. islandicus* LAL14/1 (30), REY15A (29) and HVE 10/4 (29), *S. solfataricus* P1 (Genbank: NZ_LT549890), P2 (56) and 98/2 (57) as well as *S. acidocaldarius* DSM 639 (58). Aliquots of virus preparation were added to the exponentially growing cultures of these strains and after incubation for 2 days at 78°C the cultures were diluted 1:50 and further grown for 2 days. The presence of virus particles in the cultures was monitored by TEM.

Infection studies. We attempted to establish plaque assay on *Sulfolobus* sp. S38A lawns in different media, as described earlier (55). The plates were incubated for up to 3 days at 75°C.

For testing the effect of SPV1 infection on host cell growth, the cells of *Sulfolobus* sp. S38A at the early logarithmic growth stage were infected with SPV1 with an MOI about 7 and incubated at 78°C with shaking. The cell density (OD_{600}) and the number of viable cells (colony-forming units, CFU) were measured at appropriate time intervals. The CFU counting was performed as described previously (27).

Due to the inability of SPV1 to form plaques on the host lawn, the virus titer was approximately estimated by counting the numbers of virions in meshes on copper grids by TEM.

As a standard, we used the preparation of *Sulfolobus* rod-shaped virus 2 (59) with the known titre.

SPV1 production and purification. For virus production, 250 ml of exponentially growing *Sulfolobus* sp. S38A cell culture (OD_{600}: ~0.5) was infected by SPV1 with an MOI of 0.1-0.5. The infected culture was incubated for about 24 h at 78°C with shaking, and diluted with fresh medium 1:5 and incubated further for 36-48 h. The cells were removed by centrifugation (Sorvall SLA 3000, 9 000 rpm, 30 min), the supernatant was collected and filtered through filter membranes (Merck Millipore) of pore sizes of 1.2 µm, 0.65 µm and 0.45 µm, subsequently, for complete removal of cells and cell debris. From the cell-free fraction the virus particles were collected either by precipitation with ammonium sulfate, 60% (wt/vol) saturation, at 4°C (60), or by flip-flow filtration through Vivaflow-200 (Sartorius Stedim Biotech, France). The concentrated SPV1 particles were re-suspended in the buffer containing 50 mM citric acid, 100 mM Na_2HPO_4, and 500 mM NaCl, pH3.6.

The concentrated SPV1 particles were purified by two rounds of centrifugation. Firstly, they were purified by 5%-20% sucrose rate zonal centrifugation (25 000 rpm, 20 min, 10°C, Beckman rotor SW40Ti). The light-scattering zone was collected and the presence of virus particles was verified by TEM. This fraction was further purified by isopycnic gradient centrifugation in CsCl, as described previously (38).

The outer shell was removed from the SPV1 virions by freezing at -80°C followed by thawing at room temperature. The partially disassembled virions were subjected to CsCl gradient centrifugation, as described above.

Electron microscopy. For TEM and cryo-EM, the samples were prepared and analyzed as described previously (13). The pictures were recorded using a Falcon II (FEI, USA).

Extraction and analysis of SPV1 DNA. Nucleic acid was isolated from purified viral particles as described previously (11). SPV1 genome libraries were prepared with the Nextflex PCRFree kit (Bioo Scientific), and samples were sequenced by Illumina Miseq with paired ended 2 × 250 bp read lengths (Genomics Platform, Institut Pasteur, France). An average coverage of 3000 was obtained. The sequence was assembled using CLC Genomics Workbench software package. ORFs were predicted using GeneMark.hmm v3.25 (61) and RAST v2.0 (62). The *in silico*-translated protein sequences were used as queries to search for sequence homologs in the non-redundant protein database at the National Center for Biotechnology Information using BLASTP (31) with an upper threshold E-value of 1e-3. Searches for distant homologs were performed using HHpred (32) against different protein databases, including PFAM (Database of Protein Families), PDB (Protein Data Bank), CDD

(Conserved Domains Database), and COG (Clusters of Orthologous Groups), which are accessible via the HHpred website. Transmembrane domains were predicted using TMHMM (63), whereas the secondary structure was predicted using Jpred (64) and PsiPred (65). The SPV1 genome sequence has been deposited in the GenBank database (accession no.KY780159).

Analysis of SPV1 structural proteins. The highly purified SPV1 virions were analyzed for protein content by 4-12% gradient NuPAGE Bis-Tris Precast Gel (Thermo Fisher Scientific). Protein bands were stained with Coomassie Blue using InstantBlue (Expedeon). Stained protein bands were excised, and digested 'in gel' with trypsin. The digested peptides were analyzed at the Proteomics Platform of the Institut Pasteur by nano LC-MS/MS using an Ultimate 3000 system (Dionex), as described previously (11). The peptide masses were searched against annotated SPV1 proteins using Andromeda (66) with MaxQuant software, version 1.3.0.5 (67).

Analysis of SPV1 lipids. Lipids were extracted from highly purified SPV1 preparation and from non-infected cells of *Sulfolobus*. sp S38A as described previously (68). The lipid extracts were dissolved in chloroform/ methanol/H_2O (65:24:4 vol:vol:vol) and separated by thin-layer chromatography on silica gel 60 plates (Merck) using chloroform/methanol/H_2O (65:24:4 vol:vol:vol) as the solvent. The lipids were visualized by molybdate (68).

ACKNOWLEDGEMENTS

This work was supported by the European Union's Horizon 2020 research and innovation programme under grant agreement No 685778, project VIRUS-X.

We are grateful to E. V. Koonin and T. G. Senkevich for their help in collecting environmental samples from hot springs in Beppu, Japan, to M. Duchateau (Proteomics Platform, Institut Pasteur) for a help with proteomics analyses, and M. Nilges and the Equipex CACSICE for providing the Falcon II direct detector.

366 **FIGURE LEGENDS**

367

368 FIG1 Electron micrographs of SPV1 virions. (A) Negatively stained with 2% (wt/vol) uranyl
369 acetate. (B) Sample embedded in vitreous ice. The open arrowhead points to the projection in
370 the axial view, the filled arrowhead points to the projection in the side view. Scale bars, 100 nm.

371

372 FIG 2 Impact of SPV1 infection on the growth kinetics of *Sulfolobus sp.* S38A culture. Growth
373 curves (OD_{600}) of non-infected and infected cell cultures are shown, correspondingly, by dotted
374 line (empty circles) and continuous line (filled circles). The bars show numbers of CFU at
375 indicated time intervals, in non-infected (open bars) and infected (filled bars) cultures. The
376 vertical arrow corresponds to the time of infection (5 h).

377

378 FIG 3 Electron micrographs of partially disrupted virions of SPV1. (A, B) Particles partially
379 devoid of outer shell. (C, D) Particles completely devoid of the outer shell. The filled arrowheads
380 point to the protrusions on the surface of the inner core; the black arrow indicates the envelope
381 of the inner core; open arrowheads point to the nucleoprotein filament released from the core.
382 (A, C*)* Negative staining with 2% (wt/vol) uranyl acetate. (B, D) Samples embedded in vitreous
383 ice. Scale bars, 100 nm.

384

385 FIG 4 Genome map of SPV1. The ORFs are represented with arrows that indicate the direction
386 of transcription. Genes identified as structural proteins are shown in grey. Homologs to other
387 ORFs of Sulfolobales viruses are shaded by lines. ORFs encoding predicted membrane
388 proteins are indicated by asterisks.

389

390 FIG 5 Characterization of SPV1 proteins. (A) SDS-PAGE of proteins in intact virions; (B) SDS-
391 PAGE of proteins in virions with detached outer shell; (C) Identification of the genes encoding
392 proteins in intact virions by LC-MS/MS. B1-B7, protein bands of proteins with identified genes.
393 M, molecular mass standards.

394

395 FIG 6 Thin-layer chromatography of lipids extracted from purified SPV1 virions and uninfected
396 cells of *Sulfolobus* sp. S38A. The filled arrowhead points the main lipid type of the SPV1 virion,
397 and the open arrowheads point to the main lipid types of the host cell.

398

REFERENCES

1. **Prangishvili, D.** 2013. The wonderful world of archaeal viruses. Annu Rev Microbiol **67:**565-85.
2. **Snyder, J. C., B. Bolduc, and M. J. Young.** 2015. 40 Years of archaeal virology: Expanding viral diversity. Virology **479-480:**369-78.
3. **Wang, H., N. Peng, S. A. Shah, L. Huang, and Q. She.** 2015. Archaeal extrachromosomal genetic elements. Microbiol Mol Biol Rev **79:**117-52.
4. **Pietilä, M. K., T. A. Demina, N. S. Atanasova, H. M. Oksanen, and D. H. Bamford.** 2014. Archaeal viruses and bacteriophages: comparisons and contrasts. Trends Microbiol **22:**334-44.
5. **Ackermann, H. W., and D. Prangishvili.** 2012. Prokaryote viruses studied by electron microscopy. Arch Virol **157:**1843-9.
6. **Erdmann, S., B. Chen, X. Huang, L. Deng, C. Liu, S. A. Shah, S. Le Moine Bauer, C. L. Sobrino, H. Wang, Y. Wei, Q. She, R. A. Garrett, L. Huang, and L. Lin.** 2014. A novel single-tailed fusiform Sulfolobus virus STSV2 infecting model Sulfolobus species. Extremophiles **18:**51-60.
7. **Gudbergsdottir, S. R., P. Menzel, A. Krogh, M. Young, and X. Peng.** 2016. Novel viral genomes identified from six metagenomes reveal wide distribution of archaeal viruses and high viral diversity in terrestrial hot springs. Environ Microbiol **18:**863-74.
8. **Hochstein, R. A., M. J. Amenabar, J. H. Munson-McGee, E. S. Boyd, and M. J. Young.** 2016. Acidianus Tailed Spindle Virus: a New Archaeal Large Tailed Spindle Virus Discovered by Culture-Independent Methods. J Virol **90:**3458-68.
9. **Liu, Y., J. Wang, Y. Wang, Z. Zhang, H. M. Oksanen, D. H. Bamford, and X. Chen.** 2015. Identification and characterization of SNJ2, the first temperate pleolipovirus integrating into the genome of the SNJ1-lysogenic archaeal strain. Mol Microbiol **98:**1002-20.
10. **Pietilä, M. K., E. Roine, A. Sencilo, D. H. Bamford, and H. M. Oksanen.** 2016. Pleolipoviridae, a newly proposed family comprising archaeal pleomorphic viruses with single-stranded or double-stranded DNA genomes. Arch Virol **161:**249-56.
11. **Rensen, E. I., T. Mochizuki, E. Quemin, S. Schouten, M. Krupovic, and D. Prangishvili.** 2016. A virus of hyperthermophilic archaea with a unique architecture among DNA viruses. Proc Natl Acad Sci U S A **113:**2478-83.
12. **Bath, C., T. Cukalac, K. Porter, and M. L. Dyall-Smith.** 2006. His1 and His2 are distantly related, spindle-shaped haloviruses belonging to the novel virus group, Salterprovirus. Virology **350:**228-39.
13. **Mochizuki, T., M. Krupovic, G. Pehau-Arnaudet, Y. Sako, P. Forterre, and D. Prangishvili.** 2012. Archaeal virus with exceptional virion architecture and the largest single-stranded DNA genome. Proc Natl Acad Sci U S A **109:**13386-91.
14. **Mochizuki, T., T. Yoshida, R. Tanaka, P. Forterre, Y. Sako, and D. Prangishvili.** 2010. Diversity of viruses of the hyperthermophilic archaeal genus Aeropyrum, and isolation of the Aeropyrum pernix bacilliform virus 1, APBV1, the first representative of the family Clavaviridae. Virology **402:**347-54.
15. **Gorlas, A., E. V. Koonin, N. Bienvenu, D. Prieur, and C. Geslin.** 2012. TPV1, the first virus isolated from the hyperthermophilic genus Thermococcus. Environ Microbiol **14:**503-16.
16. **Contursi, P., S. Fusco, R. Cannio, and Q. She.** 2014. Molecular biology of fuselloviruses and their satellites. Extremophiles **18:**473-89.
17. **Deng, L., F. He, Y. Bhoobalan-Chitty, L. Martinez-Alvarez, Y. Guo, and X. Peng.** 2014. Unveiling cell surface and type IV secretion proteins responsible for archaeal rudivirus entry. J Virol **88:**10264-8.
18. **Quemin, E. R., P. Chlanda, M. Sachse, P. Forterre, D. Prangishvili, and M. Krupovic.** 2016. Eukaryotic-Like Virus Budding in Archaea. MBio **7:**e01439-16.

19. **Quemin, E. R., S. Lucas, B. Daum, T. E. Quax, W. Kuhlbrandt, P. Forterre, S. V. Albers, D. Prangishvili, and M. Krupovic.** 2013. First insights into the entry process of hyperthermophilic archaeal viruses. J Virol **87:**13379-85.
20. **Snyder, J. C., R. Y. Samson, S. K. Brumfield, S. D. Bell, and M. J. Young.** 2013. Functional interplay between a virus and the ESCRT machinery in archaea. Proc Natl Acad Sci U S A **110:**10783-7.
21. **Uldahl, K. B., S. B. Jensen, Y. Bhoobalan-Chitty, L. Martinez-Alvarez, P. Papathanasiou, and X. Peng.** 2016. Life Cycle Characterization of Sulfolobus Monocaudavirus 1, an Extremophilic Spindle-Shaped Virus with Extracellular Tail Development. J Virol **90:**5693-9.
22. **Atanasova, N. S., D. H. Bamford, and H. M. Oksanen.** 2016. Virus-host interplay in high salt environments. Environ Microbiol Rep **8:**431-44.
23. **Stedman, K. M., M. DeYoung, M. Saha, M. B. Sherman, and M. C. Morais.** 2015. Structural insights into the architecture of the hyperthermophilic Fusellovirus SSV1. Virology **474:**105-9.
24. **Pietilä, M. K., P. Laurinmaki, D. A. Russell, C. C. Ko, D. Jacobs-Sera, R. W. Hendrix, D. H. Bamford, and S. J. Butcher.** 2013. Structure of the archaeal head-tailed virus HSTV-1 completes the HK97 fold story. Proc Natl Acad Sci U S A **110:**10604-9.
25. **Veesler, D., T. S. Ng, A. K. Sendamarai, B. J. Eilers, C. M. Lawrence, S. M. Lok, M. J. Young, J. E. Johnson, and C. Y. Fu.** 2013. Atomic structure of the 75 MDa extremophile Sulfolobus turreted icosahedral virus determined by CryoEM and X-ray crystallography. Proc Natl Acad Sci U S A **110:**5504-9.
26. **Erdmann, S., S. Le Moine Bauer, and R. A. Garrett.** 2014. Inter-viral conflicts that exploit host CRISPR immune systems of Sulfolobus. Mol Microbiol **91:**900-17.
27. **Bize, A., E. A. Karlsson, K. Ekefjard, T. E. Quax, M. Pina, M. C. Prevost, P. Forterre, O. Tenaillon, R. Bernander, and D. Prangishvili.** 2009. A unique virus release mechanism in the Archaea. Proc Natl Acad Sci U S A **106:**11306-11.
28. **Ceballos, R. M., C. D. Marceau, J. O. Marceau, S. Morris, A. J. Clore, and K. M. Stedman.** 2012. Differential virus host-ranges of the Fuselloviridae of hyperthermophilic Archaea: implications for evolution in extreme environments. Front Microbiol **3:**295.
29. **Guo, L., K. Brugger, C. Liu, S. A. Shah, H. Zheng, Y. Zhu, S. Wang, R. K. Lillestol, L. Chen, J. Frank, D. Prangishvili, L. Paulin, Q. She, L. Huang, and R. A. Garrett.** 2011. Genome analyses of Icelandic strains of Sulfolobus islandicus, model organisms for genetic and virus-host interaction studies. J Bacteriol **193:**1672-80.
30. **Jaubert, C., C. Danioux, J. Oberto, D. Cortez, A. Bize, M. Krupovic, Q. She, P. Forterre, D. Prangishvili, and G. Sezonov.** 2013. Genomics and genetics of Sulfolobus islandicus LAL14/1, a model hyperthermophilic archaeon. Open Biol **3:**130010.
31. **Altschul, S. F., T. L. Madden, A. A. Schaffer, J. Zhang, Z. Zhang, W. Miller, and D. J. Lipman.** 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res **25:**3389-402.
32. **Soding, J., A. Biegert, and A. N. Lupas.** 2005. The HHpred interactive server for protein homology detection and structure prediction. Nucleic Acids Res **33:**W244-8.
33. **Iranzo, J., E. V. Koonin, D. Prangishvili, and M. Krupovic.** 2016. Bipartite Network Analysis of the Archaeal Virosphere: Evolutionary Connections between Viruses and Capsidless Mobile Elements. J Virol **90:**11043-11055.
34. **Prangishvili, D., R. A. Garrett, and E. V. Koonin.** 2006. Evolutionary genomics of archaeal viruses: unique viral genomes in the third domain of life. Virus Res **117:**52-67.
35. **Krupovic, M., M. F. White, P. Forterre, and D. Prangishvili.** 2012. Postcards from the edge: structural genomics of archaeal viruses. Adv Virus Res **82:**33-62.

36. **Prangishvili, D., E. V. Koonin, and M. Krupovic.** 2013. Genomics and biology of Rudiviruses, a model for the study of virus-host interactions in Archaea. Biochem Soc Trans **41:**443-50.
37. **Wurch, L., R. J. Giannone, B. S. Belisle, C. Swift, S. Utturkar, R. L. Hettich, A. L. Reysenbach, and M. Podar.** 2016. Genomics-informed isolation and characterization of a symbiotic Nanoarchaeota system from a terrestrial geothermal environment. Nat Commun **7:**12115.
38. **Quemin, E. R., M. K. Pietila, H. M. Oksanen, P. Forterre, W. I. Rijpstra, S. Schouten, D. H. Bamford, D. Prangishvili, and M. Krupovic.** 2015. Sulfolobus Spindle-Shaped Virus 1 Contains Glycosylated Capsid Proteins, a Cellular Chromatin Protein, and Host-Derived Lipids. J Virol **89:**11681-91.
39. **Bettstetter, M., X. Peng, R. A. Garrett, and D. Prangishvili.** 2003. AFV1, a novel virus infecting hyperthermophilic archaea of the genus acidianus. Virology **315:**68-79.
40. **Roine, E., and D. H. Bamford.** 2012. Lipids of archaeal viruses. Archaea **2012:**384919.
41. **Krupovic, M., and E. V. Koonin.** 2017. Multiple origins of viral capsid proteins from cellular ancestors. Proc Natl Acad Sci U S A **114:**E2401-E2410.
42. **Krupovic, M.** 2013. Networks of evolutionary interactions underlying the polyphyletic origin of ssDNA viruses. Curr Opin Virol **3:**578-86.
43. **Ban, N., S. B. Larson, and A. McPherson.** 1995. Structural comparison of the plant satellite viruses. Virology **214:**571-83.
44. **Klose, T., and M. G. Rossmann.** 2014. Structure of large dsDNA viruses. Biol Chem **395:**711-9.
45. **Suhanovsky, M. M., and C. M. Teschke.** 2015. Nature's favorite building block: Deciphering folding and capsid assembly of proteins with the HK97-fold. Virology **479-480:**487-97.
46. **Rixon, F. J., and M. F. Schmid.** 2014. Structural similarities in DNA packaging and delivery apparatuses in Herpesvirus and dsDNA bacteriophages. Curr Opin Virol **5:**105-10.
47. **Abrescia, N. G., D. H. Bamford, J. M. Grimes, and D. I. Stuart.** 2012. Structure unifies the viral universe. Annu Rev Biochem **81:**795-822.
48. **Krupovic, M., and E. V. Koonin.** 2015. Polintons: a hotbed of eukaryotic virus, transposon and plasmid evolution. Nat Rev Microbiol **13:**105-15.
49. **Sinclair, R. M., J. J. Ravantti, and D. H. Bamford.** 2017. Nucleic and amino acid sequences support structure-based viral classification. J Virol.
50. **Gil-Carton, D., S. T. Jaakkola, D. Charro, B. Peralta, D. Castano-Diez, H. M. Oksanen, D. H. Bamford, and N. G. Abrescia.** 2015. Insight into the Assembly of Viruses with Vertical Single beta-barrel Major Capsid Proteins. Structure **23:**1866-77.
51. **Rissanen, I., J. M. Grimes, A. Pawlowski, S. Mantynen, K. Harlos, J. K. Bamford, and D. I. Stuart.** 2013. Bacteriophage P23-77 capsid protein structures reveal the archetype of an ancient branch from a major virus lineage. Structure **21:**718-26.
52. **Krupovic, M., and D. H. Bamford.** 2008. Virus evolution: how far does the double beta-barrel viral lineage extend? Nat Rev Microbiol **6:**941-8.
53. **Koonin, E. V., V. V. Dolja, and M. Krupovic.** 2015. Origins and evolution of viruses of eukaryotes: The ultimate modularity. Virology **479-480:**2-25.
54. **Iranzo, J., M. Krupovic, and E. V. Koonin.** 2016. The Double-Stranded DNA Virosphere as a Modular Hierarchical Network of Gene Sharing. MBio **7:**e00978-16.
55. **Zillig, W., A. Kletzin, C. Schleper, I. Holz, D. Janekovic, J. Hain, M. Lanzendorfer, and J. K. Kristjansson.** 1994. Screening for Sulfolobales, Their Plasmids and Their Viruses in Icelandic Solfataras. Syst Appl Microbiol **16:**609-628.
56. **She, Q., R. K. Singh, F. Confalonieri, Y. Zivanovic, G. Allard, M. J. Awayez, C. C. Chan-Weiher, I. G. Clausen, B. A. Curtis, A. De Moors, G. Erauso, C. Fletcher, P. M. Gordon, I. Heikamp-de Jong, A. C. Jeffries, C. J. Kozera, N. Medina, X. Peng, H. P. Thi-Ngoc, P. Redder, M. E. Schenk, C. Theriault, N. Tolstrup, R. L. Charlebois, W. F. Doolittle, M. Duguet, T. Gaasterland, R. A.**

Garrett, M. A. Ragan, C. W. Sensen, and J. Van der Oost. 2001. The complete genome of the crenarchaeon Sulfolobus solfataricus P2. Proc Natl Acad Sci U S A **98:**7835-40.

57. **McCarthy, S., J. Gradnigo, T. Johnson, S. Payne, A. Lipzen, J. Martin, W. Schackwitz, E. Moriyama, and P. Blum.** 2015. Complete Genome Sequence of Sulfolobus solfataricus Strain 98/2 and Evolved Derivatives. Genome Announc **3**.

58. **Chen, L., K. Brugger, M. Skovgaard, P. Redder, Q. She, E. Torarinsson, B. Greve, M. Awayez, A. Zibat, H. P. Klenk, and R. A. Garrett.** 2005. The genome of Sulfolobus acidocaldarius, a model organism of the Crenarchaeota. J Bacteriol **187:**4992-9.

59. **Prangishvili, D., H. P. Arnold, D. Gotz, U. Ziese, I. Holz, J. K. Kristjansson, and W. Zillig.** 1999. A novel virus family, the Rudiviridae: Structure, virus-host interactions and genome variability of the sulfolobus viruses SIRV1 and SIRV2. Genetics **152:**1387-96.

60. **Martin, A., S. Yeats, D. Janekovic, W. D. Reiter, W. Aicher, and W. Zillig.** 1984. SAV 1, a temperate u.v.-inducible DNA virus-like particle from the archaebacterium Sulfolobus acidocaldarius isolate B12. Embo J **3:**2165-8.

61. **Lukashin, A. V., and M. Borodovsky.** 1998. GeneMark.hmm: new solutions for gene finding. Nucleic Acids Res **26:**1107-15.

62. **Overbeek, R., R. Olson, G. D. Pusch, G. J. Olsen, J. J. Davis, T. Disz, R. A. Edwards, S. Gerdes, B. Parrello, M. Shukla, V. Vonstein, A. R. Wattam, F. Xia, and R. Stevens.** 2014. The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). Nucleic Acids Res **42:**D206-14.

63. **Krogh, A., B. Larsson, G. von Heijne, and E. L. Sonnhammer.** 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J Mol Biol **305:**567-80.

64. **Drozdetskiy, A., C. Cole, J. Procter, and G. J. Barton.** 2015. JPred4: a protein secondary structure prediction server. Nucleic Acids Res **43:**W389-94.

65. **Jones, D. T.** 1999. Protein secondary structure prediction based on position-specific scoring matrices. J Mol Biol **292:**195-202.

66. **Cox, J., N. Neuhauser, A. Michalski, R. A. Scheltema, J. V. Olsen, and M. Mann.** 2011. Andromeda: a peptide search engine integrated into the MaxQuant environment. J Proteome Res **10:**1794-805.

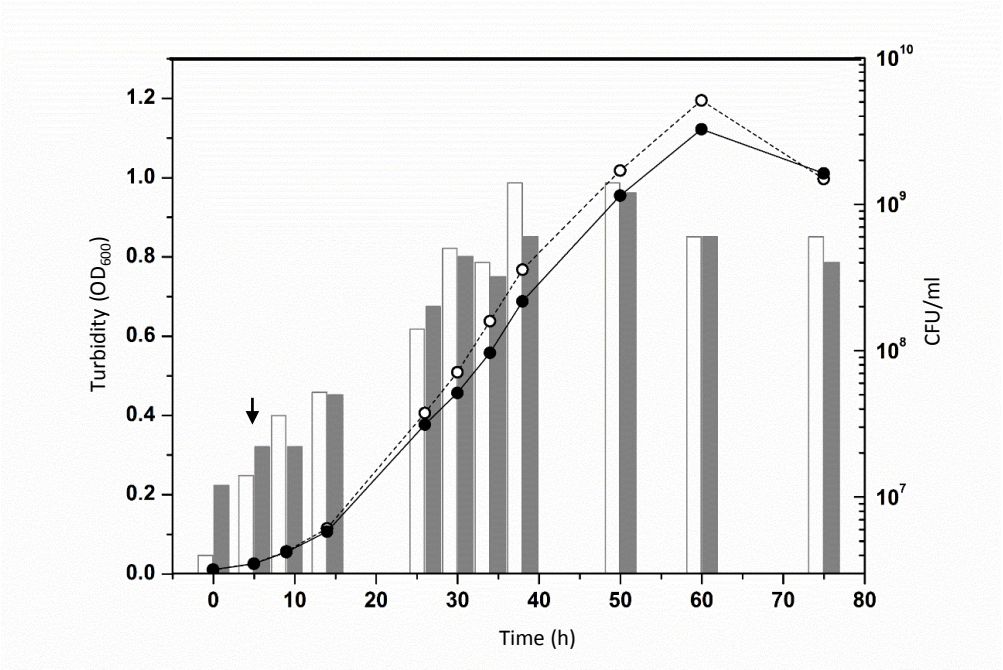
67. **Cox, J., and M. Mann.** 2008. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat Biotechnol **26:**1367-72.

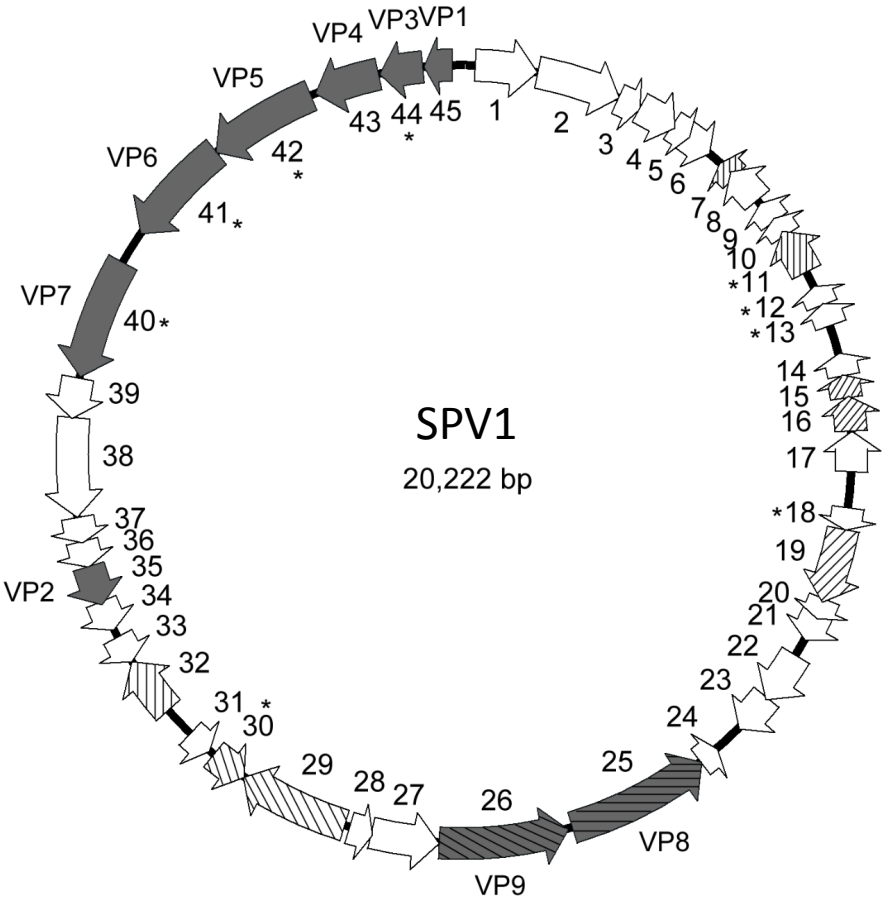
68. **Arnold, H. P., W. Zillig, U. Ziese, I. Holz, M. Crosby, T. Utterback, J. F. Weidmann, J. K. Kristjanson, H. P. Klenk, K. E. Nelson, and C. M. Fraser.** 2000. A novel lipothrixvirus, SIFV, of the extremely thermophilic crenarchaeon Sulfolobus. Virology **267:**252-66.

Journal of Virology

JVI

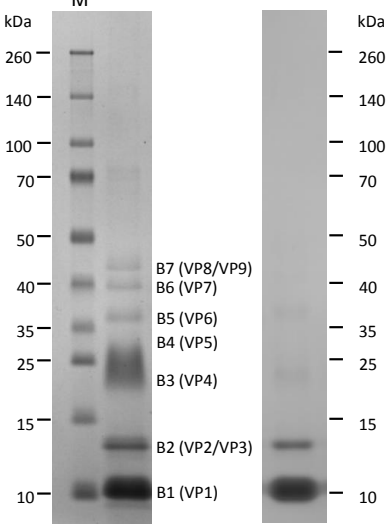
Journal of Virology





Journal of Virology

A



B

C

Protein band	SPV1 protein (ORF)	Number of identified unique peptides	Coverage (%)[a]
B1	VP1 (ORF45-78)	13	100
B2	VP2 (ORF35-111)	24	100
	VP3 (ORF44-119)	10	74.8
B3	VP4 (ORF43-181)	11	54.7
B4	VP5 (ORF42-304)	3	7.9
B5	VP6 (ORF41-305)	2	14.4
B6	VP7 (ORF40-337)	4	17.5
B7	VP8 (ORF25-400)	5	19
	VP9 (ORF26-357)	3	9.2

a. The percentage of the protein sequence covered by identified peptides.

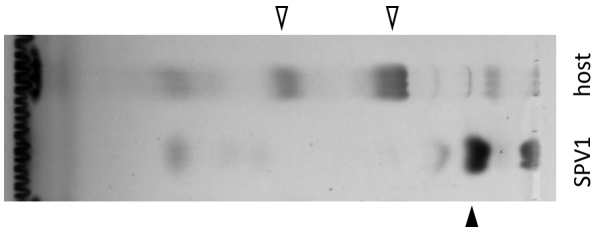


Table 1. Summary of predicted ORFs in the SPV1 genome

ORF	Coordinates	Length, aa	TMD	Annotation	HHpred hit	Probability	BLAST hit	Identity (%), E-value
ORF01-171	77..592	171						
ORF02-234	599..1303	234		Zn-finger DNA-binding protein	2n25	P=99.2		
ORF03-65	1300..1497	65						
ORF04-110	1494..1826	110						
ORF05-39	1813..1932	39						
ORF06-83	1929..2180	83						
ORF07-40	c(2305..2427)	40		C2H2 Zn-binding protein	2dmd	P=99.4	Sulfolobus islandicus rudivirus 3 (YP_009272958)	14/39(36%), 1e-03
ORF08-107	c(2414..2737)	107						
ORF09-63	c(2771..2962)	63						
ORF10-36	c(2946..3056)	36						
ORF11-131	c(3064..3459)	131	1	GepA-like uncharacterized conserved protein	COG3600	P=99.5	Acidianus hospitalis (WP_048054695) Acidianus filamentous virus 3 (YP_001604369)	37/120(31%), 2e-06 33/113(29%), 1.7
ORF12-57	c(3590..3763)	57	2					
ORF13-64	c(3750..3944)	64	1					
ORF14-62	c(4186..4374)	62						
ORF15-65	c(4364..4561)	65		SWIM Zn-finger protein	PF04434	P=97.1	Sulfolobus islandicus (WP_012953064) Sulfolobus monocaudavirus SMV1 (YP_009008084)	28/66(42%), 5e-10 24/61(39%), 4e-09
ORF16-96	c(4548..4838)	96		DNA binding protein, wHTH	PF14947	P=97.6	Sulfolobus spindle-shaped virus 1 (NP_039783)	31/90(34%), 4e-10
ORF17-112	c(4838..5176)	112						
ORF18-62	5475..5663	62	2					
ORF19-208	5660..6286	208			VOG3904	P=100	Sulfolobus spindle-shaped virus 6 (YP_003331465)	62/180(34%), 1e-19
ORF20-34	6283..6387	34						
ORF21-84	6380..6634	84						
ORF22-147	6781..7224	147		DNA binding protein, HTH; conserved in Sulfolobales	3m8j	P=97.6	Acidianus hospitalis (WP_048054614)	43/138(31%); 4e-06
ORF23-115	7205..7552	115		transcriptional regulator, CopG-like RHH	PF12441	P=96.7		
ORF24-43	c(7810..7941)	43						
ORF25-400	c(7967..9169)	400		Structural protein	PF10102	P=98.7	Acidianus rod-shaped virus 1 (YP_001542644) Candidatus Nanopusillus acidilobi (AMD30014)	47/130(36%), 4e-10 72/259(28%), 3e-06
ORF26-357	c(9211..10284)	357		Structural protein	PF10102	P=98.7	Sulfolobus islandicus rudivirus 3 (YP_009272982) Candidatus Nanopusillus acidilobi (AMD29619)	38/106(36%), 1e-08 66/193(34%), 9e-03
ORF27-187	c(10289..10852)	187						
ORF28-64	c(10846..11040)	64		Leucine-rich repeat protein (cell adhesion/invasion)	PF14580	P=98.7		
ORF29-310	11083..12015	310		Glycosyl transferase, GT-B	1rzu	P=99.9	Acidianus filamentous virus 2 (YP_001496947)	93/354(26%), 9e-29

				fold				
ORF30-103	c(12010..12321)	103	2				Sulfolobus turreted icosahedral virus 2 (YP_003591085)	59/103(57%), 3e-25
ORF31-70	c(12371..12583)	70						
ORF32-168	12849..13355	168		SAM-dependent methyltransferase, FkbM-like	2py6	P=99.5	Acidianus rod-shaped virus 2 (YP_009230239)	74/144(51%), 3e-45
ORF33-75	c(13357..13584)	75						
ORF34-84	c(13655..13909)	84						
ORF35-111	c(13884..14219)	111		α-helical protein, transcriptional regulator, HTH	PF04297	P=98.2	Sulfolobales archaeon AZ1 (EWG08170)	25/78(32%), 1e-03
ORF36-72	c(14216..14434)	72						
ORF37-69	c(14424..14633)	69						
ORF38-274	c(14639..15463)	274						
ORF39-112	c(15456..15794)	112						
ORF40-337	c(15807..16820)	337	1	Mainly β-stranded				
ORF41-305	c(17106..18023)	305	1	Mainly β-stranded				
ORF42-304	c(18023..18937)	304	3	Mainly α-helical				
ORF43-181	c(18979..19524)	181		Mainly β-stranded				
ORF44-119	c(19536..19895)	119	2	Mainly α-helical				
ORF45-78	c(19906..20142)	78		α-helical protein				

c, complementary strand; TMD, transmembrane domain. Genes encoding structural proteins are indicated by boldface type.