



**HAL**  
open science

## A Novel Type of Polyhedral Viruses Infecting Hyperthermophilic Archaea

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

► **To cite this version:**

Ying Liu, Sonoko Ishino, Yoshizumi Ishino, Gerard Pehau-Arnaudet, Mart Krupovic, et al.. A Novel Type of Polyhedral Viruses Infecting Hyperthermophilic Archaea. *Journal of Virology*, 2017, 91 (13), pp.e00589-17. 10.1128/JVI.00589-17. pasteur-01977362

**HAL Id: pasteur-01977362**

**<https://pasteur.hal.science/pasteur-01977362>**

Submitted on 10 Jan 2019

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - ShareAlike 4.0 International License

1 **A novel type of polyhedral viruses infecting hyperthermophilic**  
2 **archaea**

3 Running title: New polyhedral virus

4 Ying Liu,<sup>a</sup> Sonoko Ishino,<sup>b</sup> Yoshizumi Ishino,<sup>b</sup> Gérard Pehau-Arnaudet,<sup>c</sup> Mart  
5 Krupovic,<sup>a#</sup> David Prangishvili<sup>a#</sup>

6 Department of Microbiology, BMGE, Institut Pasteur, 25-28 rue du Dr. Roux, 75015 Paris,  
7 France<sup>a</sup>; Department of Bioscience and Biotechnology, Kyushu University, 6-10-1 Hakozaki,  
8 Higashi-ku, Fukuoka, Fukuoka 812-8581, Japan<sup>b</sup>; Utrapole, UMR 3528, CNRS, Institut Pasteur,  
9 25-28 rue du Dr. Roux, 75015 Paris, France<sup>c</sup>

10

11 Address correspondence to: Mart Krupovic, [krupovic@pasteur.fr](mailto:krupovic@pasteur.fr) and David Prangishvili,  
12 [david.prangishvili@pasteur.fr](mailto:david.prangishvili@pasteur.fr)

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

14

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

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

34 **KEYWORDS** archaea, hyperthermophile, virion structure, viral genome

35

36

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

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

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

62

## 63 RESULTS

64

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

71 isometric, likely icosahedral, uniform in overall appearance and size (not shown). From the  
72 same culture, 50 isolates, numbered S1 to S50, were colony-purified. None of the isolates was  
73 capable of the VLP production. They were further analyzed for the ability to replicate the  
74 observed VLPs. Aliquots of the cell-free enrichment culture (5  $\mu$ l) were applied onto Phytigel  
75 plates that contained cells of each of the 50 isolates, prior to the lawn development. Cells of 40  
76 isolates grew as lawns on the plates after incubation at 75°C for three days. Turbid zones of  
77 growth inhibition were formed around drops applied on the lawn formed by the isolate S38. The  
78 16S rDNA sequence of the isolate revealed that it represents a species from the genus  
79 *Sulfolobus*. From the lawn of *Sulfolobus* sp. S38 the zone of growth inhibition was excised and  
80 placed in the growing cell culture of the same species. As a result, the production of polyhedral  
81 VLPs was observed in the culture. For the verification of the infectious nature of these particles,  
82 they were collected from the cell-free culture supernatant and added to the growing culture of  
83 *Sulfolobus* sp. S38. A dramatic increase of the VLP concentration, as observed by TEM,  
84 indicated effective replication of the particles and confirmed that they represent infectious virions  
85 of a *Sulfolobus* virus.

86 *Sulfolobus* sp. S38 was subjected to an additional round of colony purification and the  
87 isolated strain, designated *Sulfolobus* sp. S38A, was selected as a standard virus host for all  
88 following experiments. The 16S rRNA gene sequence revealed that the isolate represents a  
89 strain of *Sulfolobus shibatae*, with the highest identity (99%) to the sequence of *S. shibatae*  
90 B12. The polyhedral virus that replicated in *Sulfolobus* sp. S38A was named *Sulfolobus*  
91 polyhedral virus 1 (SPV1). The virus SPV1 was purified by 5-20% sucrose rate-zonal  
92 centrifugation followed by isopicnic centrifugation in the gradient of CsCl. The morphology and  
93 size of SPV1 virions was identical to that of the VLPs observed in the enrichment culture (Fig.  
94 1).

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

100 Infection of exponentially growing cells of *Sulfolobus* sp. 38A with SPV1 at an MOI of about  
101 7 caused only slight retardation of host growth and did not lead to host cell death and lysis, as  
102 validated by the OD<sub>600</sub> measurements and enumeration of viable cells in the cell culture (Fig. 2).  
103 In line with these observations, application of 5  $\mu$ l of concentrated virus preparation onto freshly  
104 prepared lawns of *Sulfolobus* sp.38A did not cause, after development of the lawns, appearance

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

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

125 Analysis of the cryo-EM micrographs of intact virions taken in two different projections  
126 provided information on the organization of the nucleoprotein filament in the virion (Fig. 1B). The  
127 patterns projected in the axial views depict 7 concentric rings (Fig. 1B, open arrowhead),  
128 whereas those projected in the side views depict 14 linear striations (Fig. 1B, filled arrowhead).  
129 The observations suggest that the nucleoprotein filament, about 3 nm in width, is wound up  
130 sinusoidally into a spherical coil. This spherical coil is about 60 nm in diameter and comprises  
131 14 loops of the nucleoprotein filament with varying diameters and even spacing between them  
132 (Fig. 1B).

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

139 which is similar to that of various *Sulfolobus* genomes (e.g., 35% for *S. islandicus* (30)). The  
140 SPV1 genome contains 45 open reading frames (ORF), which are tightly arranged and occupy  
141 89.1% of the genome (Fig. 4). The ORFs are generally short, with median length of 103 codons.

142 Homology searches using the BLASTP program (31) revealed that less than one third  
143 (~27%) of SPV1 gene products are significantly similar (cutoff of  $E=1e-03$ ) to sequences in the  
144 nonredundant protein database (Table 1). Notably, 10 ORFs had homologs in various  
145 Sulfolobales viruses, including members of the families *Rudiviridae*, *Lipothrixviridae*,  
146 *Fuselloviridae* and *Turriviridae*. Homologs of two additional ORFs were encoded in the genomes  
147 of members of the order Sulfolobales, rather than their viruses. A combination of BLASTP and a  
148 more sensitive hidden Markov model (HMM)-based HHpred (32) analyses allowed the  
149 assignment of putative functions to just one fifth of all SPV1 ORFs (nine ORFs, 20%). Seven  
150 ORFs encoded putative proteins containing various DNA-binding domains, including zinc finger  
151 (ORF02-234, ORF07-40, ORF15-65), (winged) helix-turn-helix (ORF16-96, ORF22-147,  
152 ORF35-111), and ribbon-helix-helix (ORF23-115) domains, which are frequently encoded by  
153 crenarchaeal viruses (33-35). In addition, ORF29-310 and ORF32-168 encode glycosyl  
154 transferase with the GT-B fold and the S-adenosyl-L-methionine-dependent methyltransferase,  
155 respectively. The closest homologs of the two proteins are encoded by *Acidianus*-infecting  
156 viruses of the order *Ligamenvirales*. In addition, ORF25-400 and ORF26-357 are homologous to  
157 each other and to the minor structural proteins conserved in all members of the family  
158 *Rudiviridae* (36) as well as in some members of the families *Lipothrixviridae* and *Bicaudaviridae*,  
159 all infecting members of the archaeal order Sulfolobales. Notably, whereas the N-terminal  
160 regions of ORF25-400 and ORF26-357 products showed highest similarity to the viral  
161 homologs, the central regions of the corresponding proteins were more similar to homologs  
162 encoded by *Candidatus* *Nanopusillus acidilobi*, a symbiotic archaeon of the order  
163 Nanoarchaeota inhabiting terrestrial geothermal environments (37).

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

173 corresponding to proteins with the apparent molecular mass in the range of 20 to 32 kDa. It  
174 contains a single protein encoded by SPV1 ORF43, denoted as VP4 (Fig. 5A,C). The traces of  
175 protein VP4 we detected also across the whole length of the gel above B3. The diffused  
176 appearance of the band B3 could be caused by different degrees of putative posttranscriptional  
177 modifications of the VP4 as well as formation of protein multimers. Along with the genes of the  
178 four major virion proteins, the genes of the minor virion proteins were also identified: ORF42-  
179 304 (band B4, VP5), ORF41-305 (band B5, VP6), ORF40-337 (band B6, VP7), ORF25-400 and  
180 ORF26-357 (band B7, VP8 and VP9, correspondingly) (Fig. 5A,C).

181 Sequence analysis of the virion components showed that products of ORF40-337, ORF41-  
182 305, ORF42-304 and ORF44-119 are predicted to be integral membrane proteins with 1 to 3  
183 membrane spanning  $\alpha$ -helical segments. This result suggests that the four minor structural  
184 proteins might be embedded within the envelope surrounding the virion core and located  
185 beneath the outer shell. The five remaining virion proteins are predicted to be soluble. Among  
186 these, VP1 encoded by ORF45-78 is a highly-basic ( $pI=9.53$ ),  $\alpha$ -helical protein of 78 amino  
187 acids, which due to its positive charge is likely to bind to the viral genome. Interestingly, VP2 is  
188 also  $\alpha$ -helical and is homologous to helix-turn-helix domain-containing transcription factors  
189 ubiquitous in archaea (but not viruses), suggesting that it is also a DNA-binding protein. The  
190 VP4, the abundant protein constituent of the SPV1 virion, shares no similarity to other cellular or  
191 viral proteins. Importantly, partially disassembled virions in which the outer shell was largely  
192 removed by freezing and thawing contained a dramatically reduced amount of VP4, whereas  
193 the amounts of VP1-3 remained rather constant (Fig. 5B). This result suggests that VP4 is the  
194 main components of the external polyhedral protein shell. Analysis of the predicted secondary  
195 structure suggests that VP4 is rich in  $\beta$ -strands. Finally, given that homologs of virion proteins  
196 encoded by ORF25-400 and ORF26-357 are found in morphologically different viruses infecting  
197 archaea of the order Sulfolobales, the two proteins are likely to mediate certain aspects of virus-  
198 host interactions. Consistently, both proteins contain the functionally uncharacterized domain  
199 DUF2341, which is widespread in bacterial and archaeal proteins and is usually fused to various  
200 lectin, immunoglobulin and CARDB-like cell adhesion domains (PF10102), suggesting that  
201 products of ORF25-400 and ORF26-357 are also likely to participate in host recognition.

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

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

209

## 210 **DISCUSSION**

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

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

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

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

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

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

279

## 280 MATERIALS AND METHODS

281 **Enrichment culture, isolation of SPV1 and host strain.** The environmental sample of  
282 translucent liquid mixed with red sand was collected from hot acidic spring Umi Jigoku (80°C,  
283 pH 3.7) in Beppu, Japan. An aliquot (10 ml) of the sample was used to inoculate 40 ml of  
284 *Sulfolobus* growth medium (55), and the culture was incubated for 10 days at 75°C in aerobic  
285 conditions without shaking. Cell-free culture supernatant containing VLPs was collected by  
286 centrifugation, followed by filtration through a filter membrane of 0.22 µm pore size (Merck  
287 Millipore). The single strains were colony-purified from the enrichment culture by plating on  
288 Phytalgel™ (Sigma-Aldrich) plates and incubated for 5 days at 75°C as described previously  
289 (55), except that Gelzan™ CM Gelrite was substituted with Phytalgel™. The oligo primers used  
290 for 16S rRNA gene amplification are A21F: 5'-TTCCGGTTGATCCYGCCGGA-3' and U1525R:  
291 5'-AAGGAGGTGATCCAGCC-3'.

292 The following collection strains were analyzed for the ability to replicate the virus SPV1: *S.*  
293 *islandicus* LAL14/1 (30), REY15A (29) and HVE 10/4 (29), *S. solfataricus* P1 (Genbank:  
294 NZ\_LT549890), P2 (56) and 98/2 (57) as well as *S. acidocaldarius* DSM 639 (58). Aliquots of  
295 virus preparation were added to the exponentially growing cultures of these strains and after  
296 incubation for 2 days at 78°C the cultures were diluted 1:50 and further grown for 2 days. The  
297 presence of virus particles in the cultures was monitored by TEM.

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

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

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

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

309 **SPV1 production and purification.** For virus production, 250 ml of exponentially growing  
310 *Sulfolobus* sp. S38A cell culture (OD<sub>600</sub>: ~0.5) was infected by SPV1 with an MOI of 0.1-0.5. The  
311 infected culture was incubated for about 24 h at 78°C with shaking, and diluted with fresh  
312 medium 1:5 and incubated further for 36-48 h. The cells were removed by centrifugation (Sorvall  
313 SLA 3000, 9 000 rpm, 30 min), the supernatant was collected and filtered through filter  
314 membranes (Merck Millipore) of pore sizes of 1.2 µm, 0.65 µm and 0.45 µm, subsequently, for  
315 complete removal of cells and cell debris. From the cell-free fraction the virus particles were  
316 collected either by precipitation with ammonium sulfate, 60% (wt/vol) saturation, at 4°C (60), or  
317 by flip-flow filtration through Vivaflow-200 (Sartorius Stedim Biotech, France). The concentrated  
318 SPV1 particles were re-suspended in the buffer containing 50 mM citric acid, 100 mM Na<sub>2</sub>HPO<sub>4</sub>,  
319 and 500 mM NaCl, pH3.6.

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

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

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

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

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

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

353 **Analysis of SPV1 lipids.** Lipids were extracted from highly purified SPV1 preparation and  
354 from non-infected cells of *Sulfolobus*. sp S38A as described previously (68). The lipid extracts  
355 were dissolved in chloroform/ methanol/H<sub>2</sub>O (65:24:4 vol:vol:vol) and separated by thin-layer  
356 chromatography on silica gel 60 plates (Merck) using chloroform/methanol/H<sub>2</sub>O (65:24:4  
357 vol:vol:vol) as the solvent. The lipids were visualized by molybdate (68).

#### 358 **ACKNOWLEDGEMENTS**

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

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

365

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

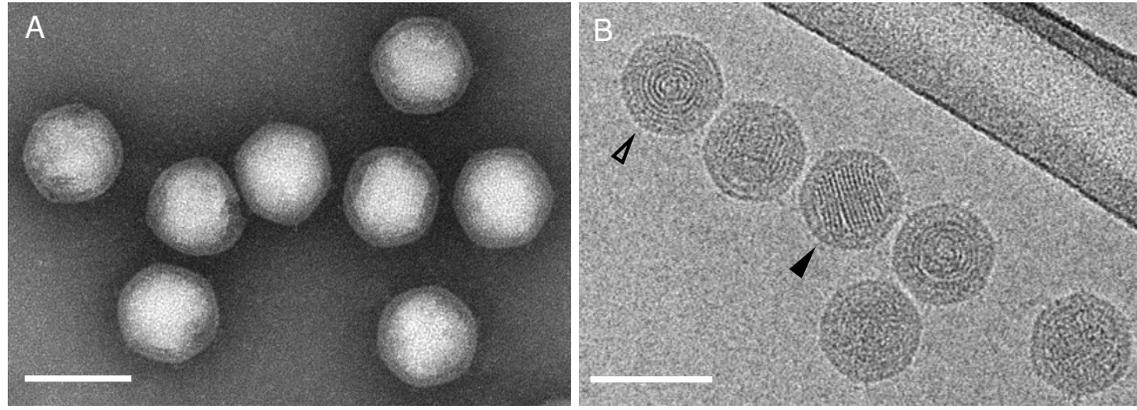
## 399 REFERENCES

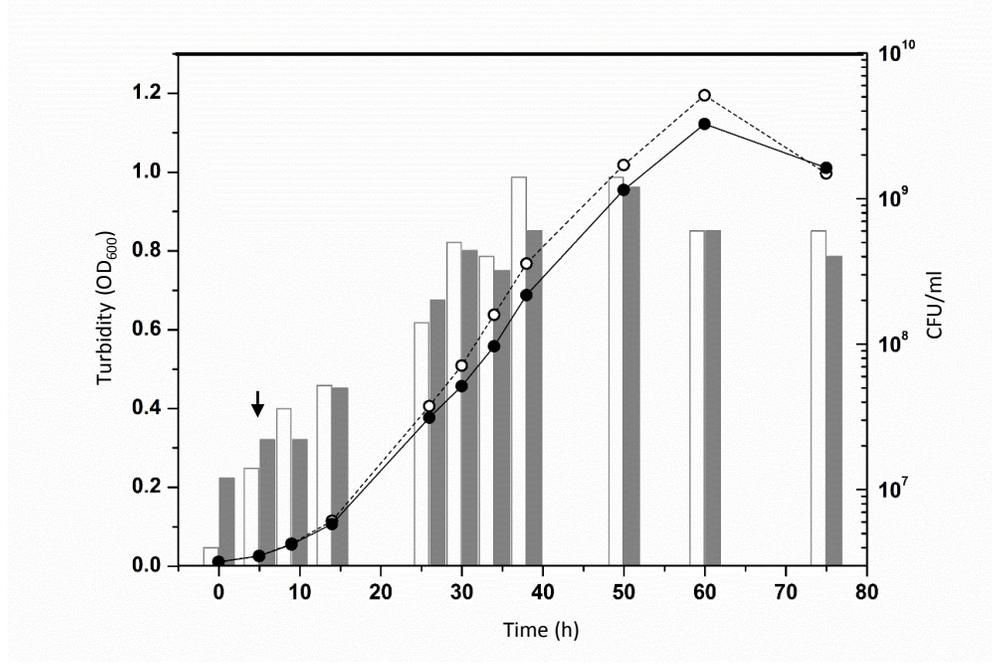
- 400 1. Prangishvili, D. 2013. The wonderful world of archaeal viruses. *Annu Rev Microbiol* **67**:565-85.
- 401 2. Snyder, J. C., B. Bolduc, and M. J. Young. 2015. 40 Years of archaeal virology: Expanding viral
- 402 diversity. *Virology* **479-480**:369-78.
- 403 3. Wang, H., N. Peng, S. A. Shah, L. Huang, and Q. She. 2015. Archaeal extrachromosomal genetic
- 404 elements. *Microbiol Mol Biol Rev* **79**:117-52.
- 405 4. Pietilä, M. K., T. A. Demina, N. S. Atanasova, H. M. Oksanen, and D. H. Bamford. 2014.
- 406 Archaeal viruses and bacteriophages: comparisons and contrasts. *Trends Microbiol* **22**:334-44.
- 407 5. Ackermann, H. W., and D. Prangishvili. 2012. Prokaryote viruses studied by electron
- 408 microscopy. *Arch Virol* **157**:1843-9.
- 409 6. Erdmann, S., B. Chen, X. Huang, L. Deng, C. Liu, S. A. Shah, S. Le Moine Bauer, C. L. Sobrino, H.
- 410 Wang, Y. Wei, Q. She, R. A. Garrett, L. Huang, and L. Lin. 2014. A novel single-tailed fusiform
- 411 *Sulfolobus* virus STSV2 infecting model *Sulfolobus* species. *Extremophiles* **18**:51-60.
- 412 7. Gudbergdottir, S. R., P. Menzel, A. Krogh, M. Young, and X. Peng. 2016. Novel viral genomes
- 413 identified from six metagenomes reveal wide distribution of archaeal viruses and high viral
- 414 diversity in terrestrial hot springs. *Environ Microbiol* **18**:863-74.
- 415 8. Hochstein, R. A., M. J. Amenabar, J. H. Munson-McGee, E. S. Boyd, and M. J. Young. 2016.
- 416 Acidianus Tailed Spindle Virus: a New Archaeal Large Tailed Spindle Virus Discovered by Culture-
- 417 Independent Methods. *J Virol* **90**:3458-68.
- 418 9. Liu, Y., J. Wang, Y. Wang, Z. Zhang, H. M. Oksanen, D. H. Bamford, and X. Chen. 2015.
- 419 Identification and characterization of SNJ2, the first temperate pleolipovirus integrating into the
- 420 genome of the SNJ1-lysogenic archaeal strain. *Mol Microbiol* **98**:1002-20.
- 421 10. Pietilä, M. K., E. Roine, A. Sencilo, D. H. Bamford, and H. M. Oksanen. 2016. Pleolipoviridae, a
- 422 newly proposed family comprising archaeal pleomorphic viruses with single-stranded or double-
- 423 stranded DNA genomes. *Arch Virol* **161**:249-56.
- 424 11. Rensen, E. I., T. Mochizuki, E. Quemin, S. Schouten, M. Krupovic, and D. Prangishvili. 2016. A
- 425 virus of hyperthermophilic archaea with a unique architecture among DNA viruses. *Proc Natl*
- 426 *Acad Sci U S A* **113**:2478-83.
- 427 12. Bath, C., T. Cukalac, K. Porter, and M. L. Dyll-Smith. 2006. His1 and His2 are distantly related,
- 428 spindle-shaped haloviruses belonging to the novel virus group, Salterprovirus. *Virology* **350**:228-
- 429 39.
- 430 13. Mochizuki, T., M. Krupovic, G. Pehau-Arnaudet, Y. Sako, P. Forterre, and D. Prangishvili. 2012.
- 431 Archaeal virus with exceptional virion architecture and the largest single-stranded DNA genome.
- 432 *Proc Natl Acad Sci U S A* **109**:13386-91.
- 433 14. Mochizuki, T., T. Yoshida, R. Tanaka, P. Forterre, Y. Sako, and D. Prangishvili. 2010. Diversity of
- 434 viruses of the hyperthermophilic archaeal genus *Aeropyrum*, and isolation of the *Aeropyrum*
- 435 *pernix* bacilliform virus 1, APBV1, the first representative of the family Clavaviridae. *Virology*
- 436 **402**:347-54.
- 437 15. Gorlas, A., E. V. Koonin, N. Bienvenu, D. Prieur, and C. Geslin. 2012. TPV1, the first virus
- 438 isolated from the hyperthermophilic genus *Thermococcus*. *Environ Microbiol* **14**:503-16.
- 439 16. Contursi, P., S. Fusco, R. Cannio, and Q. She. 2014. Molecular biology of fuselloviruses and their
- 440 satellites. *Extremophiles* **18**:473-89.
- 441 17. Deng, L., F. He, Y. Bhoobalan-Chitty, L. Martinez-Alvarez, Y. Guo, and X. Peng. 2014. Unveiling
- 442 cell surface and type IV secretion proteins responsible for archaeal rudivirus entry. *J Virol*
- 443 **88**:10264-8.
- 444 18. Quemin, E. R., P. Chlanda, M. Sachse, P. Forterre, D. Prangishvili, and M. Krupovic. 2016.
- 445 Eukaryotic-Like Virus Budding in Archaea. *MBio* **7**:e01439-16.

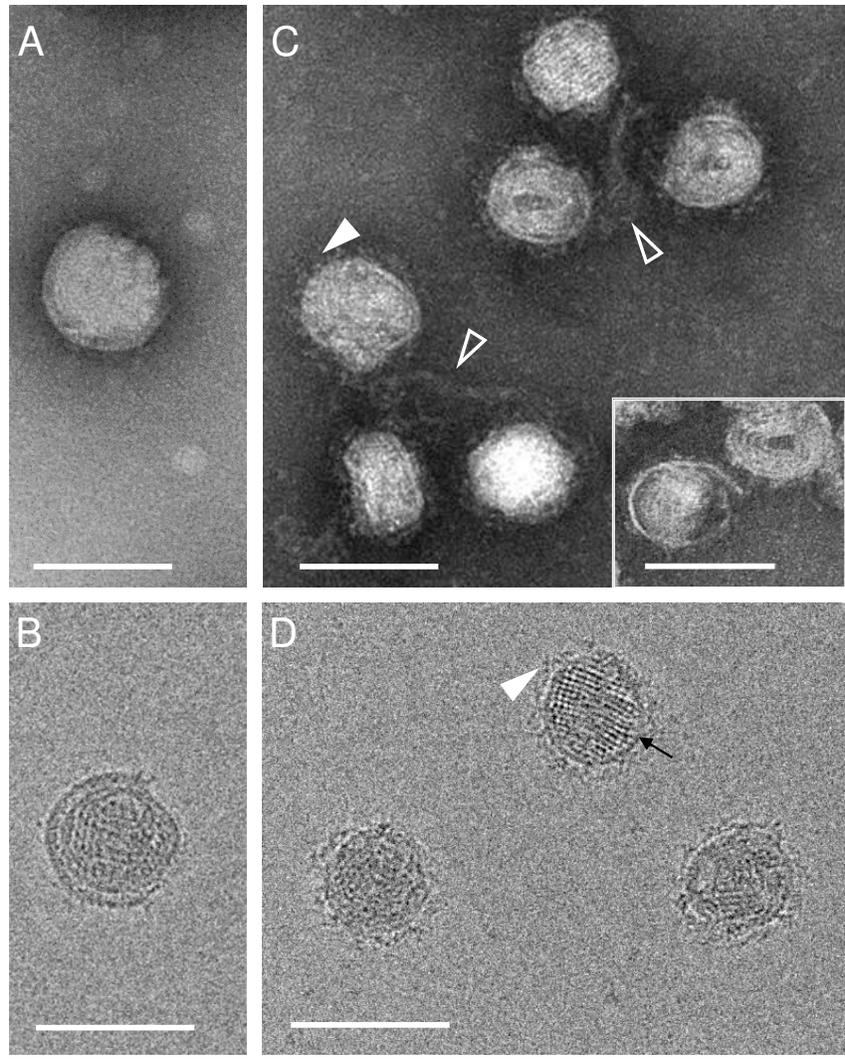
- 446 19. **Quemin, E. R., S. Lucas, B. Daum, T. E. Quax, W. Kuhlbrandt, P. Forterre, S. V. Albers, D.**  
447 **Prangishvili, and M. Krupovic.** 2013. First insights into the entry process of hyperthermophilic  
448 archaeal viruses. *J Virol* **87**:13379-85.
- 449 20. **Snyder, J. C., R. Y. Samson, S. K. Brumfield, S. D. Bell, and M. J. Young.** 2013. Functional  
450 interplay between a virus and the ESCRT machinery in archaea. *Proc Natl Acad Sci U S A*  
451 **110**:10783-7.
- 452 21. **Uldahl, K. B., S. B. Jensen, Y. Bhoobalan-Chitty, L. Martinez-Alvarez, P. Papathanasiou, and X.**  
453 **Peng.** 2016. Life Cycle Characterization of *Sulfolobus* Monocaudavirus 1, an Extremophilic  
454 Spindle-Shaped Virus with Extracellular Tail Development. *J Virol* **90**:5693-9.
- 455 22. **Atanasova, N. S., D. H. Bamford, and H. M. Oksanen.** 2016. Virus-host interplay in high salt  
456 environments. *Environ Microbiol Rep* **8**:431-44.
- 457 23. **Stedman, K. M., M. DeYoung, M. Saha, M. B. Sherman, and M. C. Morais.** 2015. Structural  
458 insights into the architecture of the hyperthermophilic Fusellovirus SSV1. *Virology* **474**:105-9.
- 459 24. **Pietilä, M. K., P. Laurinmaki, D. A. Russell, C. C. Ko, D. Jacobs-Sera, R. W. Hendrix, D. H.**  
460 **Bamford, and S. J. Butcher.** 2013. Structure of the archaeal head-tailed virus HSTV-1 completes  
461 the HK97 fold story. *Proc Natl Acad Sci U S A* **110**:10604-9.
- 462 25. **Veesler, D., T. S. Ng, A. K. Sendamarai, B. J. Eilers, C. M. Lawrence, S. M. Lok, M. J. Young, J. E.**  
463 **Johnson, and C. Y. Fu.** 2013. Atomic structure of the 75 MDa extremophile *Sulfolobus* turreted  
464 icosahedral virus determined by CryoEM and X-ray crystallography. *Proc Natl Acad Sci U S A*  
465 **110**:5504-9.
- 466 26. **Erdmann, S., S. Le Moine Bauer, and R. A. Garrett.** 2014. Inter-viral conflicts that exploit host  
467 CRISPR immune systems of *Sulfolobus*. *Mol Microbiol* **91**:900-17.
- 468 27. **Bize, A., E. A. Karlsson, K. Ekefjard, T. E. Quax, M. Pina, M. C. Prevost, P. Forterre, O. Tenaillon,**  
469 **R. Bernander, and D. Prangishvili.** 2009. A unique virus release mechanism in the Archaea. *Proc*  
470 *Natl Acad Sci U S A* **106**:11306-11.
- 471 28. **Ceballos, R. M., C. D. Marceau, J. O. Marceau, S. Morris, A. J. Clore, and K. M. Stedman.** 2012.  
472 Differential virus host-ranges of the Fuselloviridae of hyperthermophilic Archaea: implications  
473 for evolution in extreme environments. *Front Microbiol* **3**:295.
- 474 29. **Guo, L., K. Brugger, C. Liu, S. A. Shah, H. Zheng, Y. Zhu, S. Wang, R. K. Liljestol, L. Chen, J. Frank,**  
475 **D. Prangishvili, L. Paulin, Q. She, L. Huang, and R. A. Garrett.** 2011. Genome analyses of  
476 Icelandic strains of *Sulfolobus islandicus*, model organisms for genetic and virus-host interaction  
477 studies. *J Bacteriol* **193**:1672-80.
- 478 30. **Jaubert, C., C. Danioux, J. Oberto, D. Cortez, A. Bize, M. Krupovic, Q. She, P. Forterre, D.**  
479 **Prangishvili, and G. Sezonov.** 2013. Genomics and genetics of *Sulfolobus islandicus* LAL14/1, a  
480 model hyperthermophilic archaeon. *Open Biol* **3**:130010.
- 481 31. **Altschul, S. F., T. L. Madden, A. A. Schaffer, J. Zhang, Z. Zhang, W. Miller, and D. J. Lipman.**  
482 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs.  
483 *Nucleic Acids Res* **25**:3389-402.
- 484 32. **Soding, J., A. Biegert, and A. N. Lupas.** 2005. The HHpred interactive server for protein  
485 homology detection and structure prediction. *Nucleic Acids Res* **33**:W244-8.
- 486 33. **Iranzo, J., E. V. Koonin, D. Prangishvili, and M. Krupovic.** 2016. Bipartite Network Analysis of the  
487 Archaeal Virosphere: Evolutionary Connections between Viruses and Capsidless Mobile  
488 Elements. *J Virol* **90**:11043-11055.
- 489 34. **Prangishvili, D., R. A. Garrett, and E. V. Koonin.** 2006. Evolutionary genomics of archaeal  
490 viruses: unique viral genomes in the third domain of life. *Virus Res* **117**:52-67.
- 491 35. **Krupovic, M., M. F. White, P. Forterre, and D. Prangishvili.** 2012. Postcards from the edge:  
492 structural genomics of archaeal viruses. *Adv Virus Res* **82**:33-62.

- 493 36. **Prangishvili, D., E. V. Koonin, and M. Krupovic.** 2013. Genomics and biology of Rudiviruses, a  
494 model for the study of virus-host interactions in Archaea. *Biochem Soc Trans* **41**:443-50.
- 495 37. **Wurch, L., R. J. Giannone, B. S. Belisle, C. Swift, S. Utturkar, R. L. Hettich, A. L. Reysenbach, and**  
496 **M. Podar.** 2016. Genomics-informed isolation and characterization of a symbiotic  
497 Nanoarchaeota system from a terrestrial geothermal environment. *Nat Commun* **7**:12115.
- 498 38. **Quemin, E. R., M. K. Pietila, H. M. Oksanen, P. Forterre, W. I. Rijpstra, S. Schouten, D. H.**  
499 **Bamford, D. Prangishvili, and M. Krupovic.** 2015. *Sulfolobus* Spindle-Shaped Virus 1 Contains  
500 Glycosylated Capsid Proteins, a Cellular Chromatin Protein, and Host-Derived Lipids. *J Virol*  
501 **89**:11681-91.
- 502 39. **Bettstetter, M., X. Peng, R. A. Garrett, and D. Prangishvili.** 2003. AFV1, a novel virus infecting  
503 hyperthermophilic archaea of the genus *acidianus*. *Virology* **315**:68-79.
- 504 40. **Roine, E., and D. H. Bamford.** 2012. Lipids of archaeal viruses. *Archaea* **2012**:384919.
- 505 41. **Krupovic, M., and E. V. Koonin.** 2017. Multiple origins of viral capsid proteins from cellular  
506 ancestors. *Proc Natl Acad Sci U S A* **114**:E2401-E2410.
- 507 42. **Krupovic, M.** 2013. Networks of evolutionary interactions underlying the polyphyletic origin of  
508 ssDNA viruses. *Curr Opin Virol* **3**:578-86.
- 509 43. **Ban, N., S. B. Larson, and A. McPherson.** 1995. Structural comparison of the plant satellite  
510 viruses. *Virology* **214**:571-83.
- 511 44. **Klose, T., and M. G. Rossmann.** 2014. Structure of large dsDNA viruses. *Biol Chem* **395**:711-9.
- 512 45. **Suhanovsky, M. M., and C. M. Teschke.** 2015. Nature's favorite building block: Deciphering  
513 folding and capsid assembly of proteins with the HK97-fold. *Virology* **479-480**:487-97.
- 514 46. **Rixon, F. J., and M. F. Schmid.** 2014. Structural similarities in DNA packaging and delivery  
515 apparatuses in Herpesvirus and dsDNA bacteriophages. *Curr Opin Virol* **5**:105-10.
- 516 47. **Abrescia, N. G., D. H. Bamford, J. M. Grimes, and D. I. Stuart.** 2012. Structure unifies the viral  
517 universe. *Annu Rev Biochem* **81**:795-822.
- 518 48. **Krupovic, M., and E. V. Koonin.** 2015. Polintons: a hotbed of eukaryotic virus, transposon and  
519 plasmid evolution. *Nat Rev Microbiol* **13**:105-15.
- 520 49. **Sinclair, R. M., J. J. Ravantti, and D. H. Bamford.** 2017. Nucleic and amino acid sequences  
521 support structure-based viral classification. *J Virol*.
- 522 50. **Gil-Carton, D., S. T. Jaakkola, D. Charro, B. Peralta, D. Castano-Diez, H. M. Oksanen, D. H.**  
523 **Bamford, and N. G. Abrescia.** 2015. Insight into the Assembly of Viruses with Vertical Single  
524 beta-barrel Major Capsid Proteins. *Structure* **23**:1866-77.
- 525 51. **Rissanen, I., J. M. Grimes, A. Pawlowski, S. Mantynen, K. Harlos, J. K. Bamford, and D. I. Stuart.**  
526 2013. Bacteriophage P23-77 capsid protein structures reveal the archetype of an ancient branch  
527 from a major virus lineage. *Structure* **21**:718-26.
- 528 52. **Krupovic, M., and D. H. Bamford.** 2008. Virus evolution: how far does the double beta-barrel  
529 viral lineage extend? *Nat Rev Microbiol* **6**:941-8.
- 530 53. **Koonin, E. V., V. V. Dolja, and M. Krupovic.** 2015. Origins and evolution of viruses of eukaryotes:  
531 The ultimate modularity. *Virology* **479-480**:2-25.
- 532 54. **Iranzo, J., M. Krupovic, and E. V. Koonin.** 2016. The Double-Stranded DNA Virosphere as a  
533 Modular Hierarchical Network of Gene Sharing. *MBio* **7**:e00978-16.
- 534 55. **Zillig, W., A. Kletzin, C. Schleper, I. Holz, D. Janekovic, J. Hain, M. Lanzendorfer, and J. K.**  
535 **Kristjansson.** 1994. Screening for *Sulfolobales*, Their Plasmids and Their Viruses in Icelandic  
536 Solfataras. *Syst Appl Microbiol* **16**:609-628.
- 537 56. **She, Q., R. K. Singh, F. Confalonieri, Y. Zivanovic, G. Allard, M. J. Awayez, C. C. Chan-Weiher, I.**  
538 **G. Clausen, B. A. Curtis, A. De Moors, G. Erauso, C. Fletcher, P. M. Gordon, I. Heikamp-de Jong,**  
539 **A. C. Jeffries, C. J. Kozera, N. Medina, X. Peng, H. P. Thi-Ngoc, P. Redder, M. E. Schenk, C.**  
540 **Theriault, N. Tolstrup, R. L. Charlebois, W. F. Doolittle, M. Duguet, T. Gaasterland, R. A.**

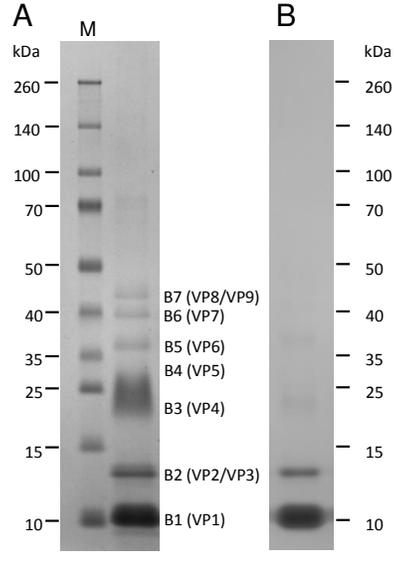
- 541 **Garrett, M. A. Ragan, C. W. Sensen, and J. Van der Oost.** 2001. The complete genome of the  
542 crenarchaeon *Sulfolobus solfataricus* P2. *Proc Natl Acad Sci U S A* **98**:7835-40.
- 543 57. **McCarthy, S., J. Gradnigo, T. Johnson, S. Payne, A. Lipzen, J. Martin, W. Schackwitz, E.**  
544 **Moriyama, and P. Blum.** 2015. Complete Genome Sequence of *Sulfolobus solfataricus* Strain  
545 98/2 and Evolved Derivatives. *Genome Announc* **3**.
- 546 58. **Chen, L., K. Brugger, M. Skovgaard, P. Redder, Q. She, E. Torarinsson, B. Greve, M. Awayez, A.**  
547 **Zibat, H. P. Klenk, and R. A. Garrett.** 2005. The genome of *Sulfolobus acidocaldarius*, a model  
548 organism of the Crenarchaeota. *J Bacteriol* **187**:4992-9.
- 549 59. **Prangishvili, D., H. P. Arnold, D. Gotz, U. Ziese, I. Holz, J. K. Kristjansson, and W. Zillig.** 1999. A  
550 novel virus family, the Rudiviridae: Structure, virus-host interactions and genome variability of  
551 the *sulfolobus* viruses SIRV1 and SIRV2. *Genetics* **152**:1387-96.
- 552 60. **Martin, A., S. Yeats, D. Janekovic, W. D. Reiter, W. Aicher, and W. Zillig.** 1984. SAV 1, a  
553 temperate u.v.-inducible DNA virus-like particle from the archaeobacterium *Sulfolobus*  
554 *acidocaldarius* isolate B12. *Embo J* **3**:2165-8.
- 555 61. **Lukashin, A. V., and M. Borodovsky.** 1998. GeneMark.hmm: new solutions for gene finding.  
556 *Nucleic Acids Res* **26**:1107-15.
- 557 62. **Overbeek, R., R. Olson, G. D. Pusch, G. J. Olsen, J. J. Davis, T. Disz, R. A. Edwards, S. Gerdes, B.**  
558 **Parrello, M. Shukla, V. Vonstein, A. R. Wattam, F. Xia, and R. Stevens.** 2014. The SEED and the  
559 Rapid Annotation of microbial genomes using Subsystems Technology (RAST). *Nucleic Acids Res*  
560 **42**:D206-14.
- 561 63. **Krogh, A., B. Larsson, G. von Heijne, and E. L. Sonnhammer.** 2001. Predicting transmembrane  
562 protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol*  
563 **305**:567-80.
- 564 64. **Drozdetskiy, A., C. Cole, J. Procter, and G. J. Barton.** 2015. JPred4: a protein secondary structure  
565 prediction server. *Nucleic Acids Res* **43**:W389-94.
- 566 65. **Jones, D. T.** 1999. Protein secondary structure prediction based on position-specific scoring  
567 matrices. *J Mol Biol* **292**:195-202.
- 568 66. **Cox, J., N. Neuhauser, A. Michalski, R. A. Scheltema, J. V. Olsen, and M. Mann.** 2011.  
569 Andromeda: a peptide search engine integrated into the MaxQuant environment. *J Proteome*  
570 *Res* **10**:1794-805.
- 571 67. **Cox, J., and M. Mann.** 2008. MaxQuant enables high peptide identification rates, individualized  
572 p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol*  
573 **26**:1367-72.
- 574 68. **Arnold, H. P., W. Zillig, U. Ziese, I. Holz, M. Crosby, T. Utterback, J. F. Weidmann, J. K.**  
575 **Kristjanson, H. P. Klenk, K. E. Nelson, and C. M. Fraser.** 2000. A novel lipothrixvirus, SIFV, of the  
576 extremely thermophilic crenarchaeon *Sulfolobus*. *Virology* **267**:252-66.  
577  
578







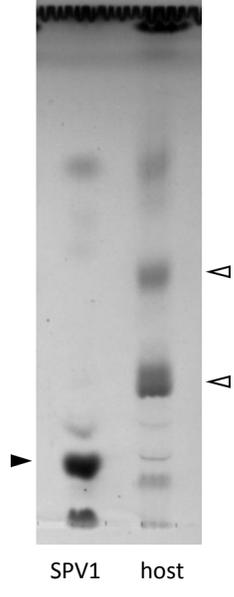




**C**

Protein band	SPV1 protein (ORF)	Number of identified unique peptides	Coverage (%) <sup>a</sup>
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	<i>Sulfolobus islandicus</i> 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	<i>Acidianus hospitalis</i> (WP_048054695) <i>Acidianus filamentous virus 3</i> (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	<i>Sulfolobus islandicus</i> (WP_012953064) <i>Sulfolobus monocaudavirus SMV1</i> (YP_009008084) <i>Sulfolobus spindle-shaped virus 1</i> (NP_039783)	28/66(42%), 5e-10 24/61(39%), 4e-09 31/90(34%), 4e-10
ORF16-96	c(4548..4838)	96		DNA binding protein, wHTH	PF14947	P=97.6		
ORF17-112	c(4838..5176)	112						
ORF18-62	5475..5663	62	2					
ORF19-208	5660..6286	208			VOG3904	P=100	<i>Sulfolobus spindle-shaped virus 6</i> (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 <i>Sulfolobales</i> transcriptional regulator, CopG-like RHH	3m8j	P=97.6	<i>Acidianus hospitalis</i> (WP_048054614)	43/138(31%), 4e-06
ORF23-115	7205..7552	115			PF12441	P=96.7		
ORF24-43	c(7810..7941)	43						
<b>ORF25-400</b>	c(7967..9169)	400		Structural protein	PF10102	P=98.7	<i>Acidianus rod-shaped virus 1</i> (YP_001542644) <i>Candidatus Nanopusillus acidilobi</i> (AMD30014) <i>Sulfolobus islandicus rudivirus 3</i> (YP_009272982) <i>Candidatus Nanopusillus acidilobi</i> (AMD29619)	47/130(36%), 4e-10 72/259(28%), 3e-06 38/106(36%), 1e-08 66/193(34%), 9e-03
<b>ORF26-357</b>	c(9211..10284)	357		Structural protein	PF10102	P=98.7		
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	<i>Acidianus filamentous virus 2</i> (YP_001496947)	93/354(26%), 9e-29

ORF30-103	c(12010..12321)	103	2	fold			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						
<b>ORF35-111</b>	c(13884..14219)	111		$\alpha$ -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						
<b>ORF40-337</b>	c(15807..16820)	337	1	Mainly $\beta$ -stranded				
<b>ORF41-305</b>	c(17106..18023)	305	1	Mainly $\beta$ -stranded				
<b>ORF42-304</b>	c(18023..18937)	304	3	Mainly $\alpha$ -helical				
<b>ORF43-181</b>	c(18979..19524)	181		Mainly $\beta$ -stranded				
<b>ORF44-119</b>	c(19536..19895)	119	2	Mainly $\alpha$ -helical				
<b>ORF45-78</b>	c(19906..20142)	78		$\alpha$ -helical protein				

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