



HAL
open science

New virus isolates from Italian hydrothermal environments underscore the biogeographic pattern in archaeal virus communities

Diana P Baquero, Patrizia Contursi, Monica Piochi, Simonetta Bartolucci, Ying Liu, Virginija Cvirkaite-Krupovic, David Prangishvili, M Krupovic

► To cite this version:

Diana P Baquero, Patrizia Contursi, Monica Piochi, Simonetta Bartolucci, Ying Liu, et al.. New virus isolates from Italian hydrothermal environments underscore the biogeographic pattern in archaeal virus communities. *The International Society of Microbiological Ecology Journal*, 2020, [Epub ahead of print], 10.1038/s41396-020-0653-z . pasteur-02557115

HAL Id: pasteur-02557115

<https://pasteur.hal.science/pasteur-02557115>

Submitted on 28 Apr 2020

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.

SUPPLEMENTARY INFORMATION

New virus isolates from Italian hydrothermal environments underscore the biogeographic pattern in archaeal virus communities

Diana P. Baquero^{1,2}, Patrizia Contursi³, Monica Piochi⁴, Simonetta Bartolucci³, Ying Liu¹, Virginija Cvirkaite-Krupovic¹, David Prangishvili^{1,*} and Mart Krupovic^{1,*}

¹ Archaeal Virology Unit, Department of Microbiology, Institut Pasteur, 75015 Paris, France

² Sorbonne Université, Collège Doctoral, 7 Quai Saint-Bernard, 75005 Paris, France

³ University of Naples Federico II, Department of Biology, Naples, Italy

⁴ Istituto Nazionale di Geofisica e Vulcanologia, Sezione Osservatorio Vesuviano, Naples, Italy

* Correspondence to: david.prangishvili@pasteur.fr and mart.krupovic@pasteur.fr

Archaeal Virology Unit, Institut Pasteur, 75015 Paris, France

Tel: 33 (0)1 40 61 37 22

Running title

Biogeographic pattern in archaeal virus communities

SUPPLEMENTARY MATERIALS AND METHODS

Enrichment cultures

Nine samples were collected from hot springs, mud pools and hydrothermally altered terrains at the solfataric field of the Campi Flegrei volcano in Pozzuoli, Italy — the study area known as Pisciarelli — with temperatures ranging from 81 to 96°C and pH between 1 and 7 (Table S1). Each sample was inoculated into medium favoring the growth of *Sulfolobus/Saccharolobus* (basal salt solution supplemented with 0.2% tryptone, 0.1% yeast extract, 0.1% sucrose, pH 3.5), *Acidianus* (basal salt solution supplemented with 0.2% tryptone, 0.1% yeast extract, sulfur, pH 3.5), and *Pyrobaculum* (0.1% yeast extract, 1% tryptone, 0.1-0.3% Na₂S₂O₃ × 5 H₂O, pH 7) species [1, 2]. The enrichment cultures were incubated for 10 days at 75°C, except for *Pyrobaculum* cultures, which were grown at 90°C for 15 days without shaking.

VLP concentration and purification

Following the removal of cells from the enrichment cultures (7,000 rpm, 20 min, Sorvall SLA-1500 rotor), VLPs were collected and concentrated by ultracentrifugation (40,000 rpm, 2h, 10°C, Beckman SW41 rotor). The particles were resuspended in buffer A: 20 mM KH₂PO₄, 250 mM NaCl, 2.14 mM MgCl₂, 0.43 mM Ca(NO₃)₂, <0.001% trace elements of *Sulfolobales* medium, pH 6 [3] and visualized by transmission electron microscopy (TEM) as described below. VLPs were further purified by centrifugation in a CsCl buoyant density gradient (0.45 g ml⁻¹) with a Beckman SW41 rotor at 39,000 rpm for 20 h at 10°C. All opalescent bands were collected with a needle and a syringe, dialyzed against buffer A and analyzed by TEM for the presence of VLPs.

Isolation of virus-host pairs

Strains of cells were colony purified by plating dilutions of the *Sulfolobus/Saccharolobus* enrichment cultures onto Phytigel (0.7% [weight/volume]) plates incubated for 7 days at 75 °C. Brownish colonies were toothpicked, inoculated into 1 mL of growth medium and incubated at 75°C for 3 days. The lawns for isolation of sensitive hosts were prepared as described previously [2]. Inhibition zones were cut out from the phytigel and inoculated into exponentially growing cultures of the corresponding isolates. After incubation at 75°C for 3 days, the virus-like particles (VLPs) were concentrated by ultracentrifugation (40,000 rpm, 2h, 10°C, Beckman SW41 rotor), resuspended in buffer A and observed by TEM (for detailed information, refer to the main text). Pure virus strains were obtained by two rounds of single-plaque purification, as described previously [4].

For *Pyrobaculum* enrichment cultures, a different approach was taken, because it was not possible to obtain single strain isolates. Growing liquid cultures of *P. arsenaticum* 2GA [1], *P. arsenaticum* PZ6 (DSM 13514) [5], *P. calidifontis* VA1 (DSM 21063) [6] and *P. oguniense* TE7 (DSM 13380) [7] were mixed with concentrated VLPs and incubated for 15 days at 90°C. Increase in the number of extracellular virus particles was verified by TEM. The relative number of produced virions was assessed by counting the VLPs in different fields under TEM. A cell-free culture with the same amount of virus particles as used for the infection assay and treated under the same conditions was used as a control. In order to obtain cultures producing just one type of viral particles, eight ten-fold serial dilutions (10 ml) of infected cells were established, incubated at 90°C and observed by TEM. Because a plaque assay could not be established for viruses infecting *Pyrobaculum*, we relied on TEM observations for assessment of virus-host interactions.

Transmission electron microscopy and VLP analysis

For negative-stain TEM, 5 µl of the samples were applied to carbon-coated copper grids, negatively stained with 2% uranyl acetate and imaged with the transmission electron microscope FEI Tecnai Biotwin. The dimensions of the negatively stained virus particles (n=80) were determined using ImageJ [8].

Infection experiments

To test the effect of the viruses on host cell growth, exponentially growing cultures of *S. solfataricus* POZ149, *M. sedula* POZ202, *A. brierleyi* POZ9 and *P. arsenaticum* 2GA were infected with their viruses at a multiplicity of infection (MOI) of about 1. The MOI was established by plaque counting in the case of SSRV1, whereas quantification by qPCR was used for the other viruses. Cultures of *S. solfataricus* POZ149, *M. sedula* POZ202 and *A. brierleyi* POZ9 were incubated at 75°C with shaking, while *P. arsenaticum* 2GA liquid cultures were grown at 90°C without agitation. The cell turbidity (OD600) was measured at appropriate time intervals. The infection experiments were performed in triplicate.

Virus quantification by qPCR

Viruses for which plaque test could not be established were quantified by estimating the number of viral genome copies by quantitative (q)PCR. Primers targeting specific virus sequences and 2 μ L of each sample were mixed with the qPCR kit (Luna Universal qPCR Master Mix, New England Biolabs). A melting curve for each pairs of primers and a calibration curve with decreasing concentration values of viral DNA ranging between 100-0.01 ng/ μ L were performed.

Host range

The following laboratory collection strains were used to test their ability to replicate viruses MRV1, ARV3 and SSRV1: *A. convivator* [9], *A. hospitalis* W1 [10], *Saccharolobus solfataricus* strains P1 (GenBank accession no. NZ_LT549890) and P2 [11], *Sulfolobus islandicus* strains REN2H1 [2], HVE10/4 [12], LAL14/1 [13], REY15A [12], Δ C1C2 [14], and *Sulfolobus acidocaldarius* strain DSM639 [15]. CsCl gradient-purified virus particles were added directly to the growing cultures of *Acidianus* species and virus propagation was evaluated by TEM. For *Sulfolobus* and *Saccharolobus* species, spot tests were performed and the replication of the virus was evidenced by the presence of an inhibition zone on the lawn of the tested strains.

The host ranges of PFV2 and PSV2 were examined by adding purified virions to the growing cultures of *P. arsenaticum* PZ6 (DSM 13514) [5], *P. calidifontis* VA1 (DSM 21063) [6] and *P. oguniense* TE7 (DSM 13380) [7]. In the absence of the plaque test, virus propagation was monitored by TEM, as described above.

Adsorption assay

Different members of the family *Sulfolobaceae* (Table 1) were infected with SSRV1 at a multiplicity of infection (MOI) of 0.05. Infected cells were incubated with agitation for one hour at 75°C. After one hour of incubation, cells were pelleted and the supernatant was kept at 4°C. The percentage of unadsorbed virus particles was determined by plaque assay comparing the viral concentration in the supernatant at 1hpi with the virus titer in the control, a cell-free culture incubated under the same conditions as the treated cultures. *S. solfataricus* POZ149 was used as an indicator strain for titrations. Experiments were conducted in triplicate.

Entry of viral DNA into the cells

PCR reactions targeting viral genome sequences were performed on infected cells as a template to study whether viral DNA has entered into the cells. Each strain of the family *Sulfolobaceae* listed in the Table 1 was infected with MRV1, ARV3 and SSRV3. Fresh growing cultures were infected with an estimated MOI of 1, incubated with agitation for one hour at 75°C, and centrifuged to remove non-bound particles. Three additional washes of the pellet were performed to ensure the removal of the residual free virus particles. The pellet was kept at -20°C until further processing. Cells were resuspended in 40 μ L of the respective media and treated at 95°C for 20 minutes. Detection of viral genomes in the cells was performed by PCR with specific primers for each virus. Strains of the genus *Pyrobaculum* were used to test the DNA entry of PSV2 and PFV2 genomes. The samples were maintained at 90°C without shaking for one hour and treated as described above.

16S rRNA gene sequence determination

DNA was extracted from exponentially growing cell cultures using the Wizard Genomic DNA Purification Kit (Promega). The 16S rRNA genes were amplified by PCR using 519F (5'-CAGCMGCCGCGGTAA) and 1041R (5'-GGCCATGCACCWCCTCTC) primers [16]. The identity of the isolated strains was determined by blastn searches queried with the corresponding 16S rRNA gene sequences against the non-redundant nucleotide sequence database at NCBI.

Viral DNA extraction, sequencing and analysis of viral genomes

Viral DNA was extracted from purified virus particles as described previously [1]. Sequencing libraries were prepared from 100 ng of DNA with the TruSeq DNA PCR-Free library Prep Kit from Illumina and sequenced on Illumina MiSeq platform with 150-bp paired-end read lengths (Institut Pasteur, France). Raw sequence reads were processed with Trimmomatic v.0.3.6 [17] and assembled with SPAdes 3.11.1 [18] with default parameters. Terminal inverted repeats (TIRs) of the viral genomes were identified by aligning the two corresponding genome termini with BLASTN. Open reading frames (ORF) were predicted by GeneMark.hmm [19] and RAST v2.0 [20]. Each predicted ORF was manually inspected for the presence of putative ribosome-binding sites upstream of the start codon. The *in silico*-translated protein sequences were analyzed by BLASTP [21] against the non-redundant protein database at the NCBI with an upper threshold E-

value of $1e-03$. Searches for distant homologs were performed using HHpred [22] against PFAM (Database of Protein Families), PDB (Protein Data Bank) and CDD (Conserved Domains Database) databases. Transmembrane domains were predicted using TMHMM [23]. CRISPR spacers matching the isolated viruses were searched for in the CRISPRdb by blastn with default parameters [24].

Phylogenomic analysis

All pairwise comparisons of the nucleotide sequences of rudivirus genomes were conducted using the Genome-BLAST Distance Phylogeny (GBDP) method implemented in VICTOR, under settings recommended for prokaryotic viruses [25]. The resulting intergenomic distances derived from pairwise matches (local alignments) were used to infer a balanced minimum evolution tree with branch support via FASTME including SPR postprocessing for D6 formula, which corresponds to the sum of all identities found in high-scoring segment pairs divided by total genome length. Branch support was inferred from 100 pseudo-bootstrap replicates each. The tree was rooted with members of the family *Lipothrixviridae*.

Genome sequences of 19 rudiviruses were further compared using the Gegenees, a tool for fragmented alignment of multiple genomes [26]. The comparison was done in the BLASTN mode, with the settings 200/100. The cutoff threshold for non-conserved material was 40%.

SUPPLEMENTARY REFERENCES

1. Rensen EI, Mochizuki T, Quemin E, Schouten S, Krupovic M, Prangishvili D. A virus of hyperthermophilic archaea with a unique architecture among DNA viruses. *Proc Natl Acad Sci U S A*. 2016; 113:2478-83.
2. Zillig W, Kletzin A, Schleper C, Holz I, Janekovic D, Hain J *et al*. Screening for Sulfolobales, their Plasmids and their Viruses in Icelandic Solfataras. *Syst Appl Microbiol*. 1993; 16:609-628.
3. Quemin ER, Pietila MK, Oksanen HM, Forterre P, Rijpstra WI, Schouten S *et al*. Sulfolobus spindle-shaped virus 1 contains glycosylated capsid proteins, a cellular chromatin protein, and host-derived lipids. *J Virol*. 2015; 89:11681-91.
4. Schleper C, Kubo K, Zillig W. The particle SSV1 from the extremely thermophilic archaeon Sulfolobus is a virus: demonstration of infectivity and of transfection with viral DNA. *Proc Natl Acad Sci U S A*. 1992; 89:7645-9.
5. Huber R, Sacher M, Vollmann A, Huber H, Rose D. Respiration of arsenate and selenate by hyperthermophilic archaea. *Syst Appl Microbiol*. 2000; 23:305-14.
6. Amo T, Paje ML, Inagaki A, Ezaki S, Atomi H, Imanaka T. *Pyrobaculum calidifontis* sp. nov., a novel hyperthermophilic archaeon that grows in atmospheric air. *Archaea*. 2002; 1:113-21.
7. Sako Y, Nunoura T, Uchida A. *Pyrobaculum oguniense* sp. nov., a novel facultatively aerobic and hyperthermophilic archaeon growing at up to 97 degrees C. *Int J Syst Evol Microbiol*. 2001; 51:303-9.
8. Schneider CA, Rasband WS, Eliceiri KW. NIH Image to ImageJ: 25 years of image analysis. *Nat Methods*. 2012; 9:671-5.
9. Prangishvili D, Vestergaard G, Haring M, Aramayo R, Basta T, Rachel R *et al*. Structural and genomic properties of the hyperthermophilic archaeal virus ATV with an extracellular stage of the reproductive cycle. *J Mol Biol*. 2006; 359:1203-16.
10. Bettstetter M, Peng X, Garrett RA, Prangishvili D. AFV1, a novel virus infecting hyperthermophilic archaea of the genus acidianus. *Virology*. 2003; 315:68-79.
11. She Q, Singh RK, Confalonieri F, Zivanovic Y, Allard G, Awayez MJ *et al*. The complete genome of the crenarchaeon Sulfolobus solfataricus P2. *Proc Natl Acad Sci U S A*. 2001; 98:7835-40.
12. Guo L, Brugger K, Liu C, Shah SA, Zheng H, Zhu Y *et al*. Genome analyses of Icelandic strains of Sulfolobus islandicus, model organisms for genetic and virus-host interaction studies. *J Bacteriol*. 2011; 193:1672-80.
13. Jaubert C, Danioux C, Oberto J, Cortez D, Bize A, Krupovic M *et al*. Genomics and genetics of Sulfolobus islandicus LAL14/1, a model hyperthermophilic archaeon. *Open Biol*. 2013; 3:130010.
14. Gudbergssdóttir S, Deng L, Chen Z, Jensen JV, Jensen LR, She Q *et al*. Dynamic properties of the Sulfolobus CRISPR/Cas and CRISPR/Cmr systems when challenged with vector-borne viral and plasmid genes and protospacers. *Mol Microbiol*. 2011; 79:35-49.
15. Chen L, Brugger K, Skovgaard M, Redder P, She Q, Torarinsson E *et al*. The genome of Sulfolobus acidocaldarius, a model organism of the Crenarchaeota. *J Bacteriol*. 2005; 187:4992-9.
16. Pausan MR, Csorba C, Singer G, Till H, Schopf V, Santigli E *et al*. Exploring the Archaeome: Detection of Archaeal Signatures in the Human Body. *Front Microbiol*. 2019; 10:2796.
17. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 2014; 30:2114-20.
18. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS *et al*. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol*. 2012; 19:455-77.
19. Lukashin AV, Borodovsky M. GeneMark.hmm: new solutions for gene finding. *Nucleic Acids Res*. 1998; 26:1107-15.
20. Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA *et al*. The RAST Server: rapid annotations using subsystems technology. *BMC Genomics*. 2008; 9:75.
21. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W *et al*. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res*. 1997; 25:3389-402.
22. Söding J, Biegert A, Lupas AN. The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res*. 2005; 33:W244-8.
23. Krogh A, Larsson B, von Heijne G, Sonnhammer EL. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol*. 2001; 305:567-80.
24. Pourcel C, Touchon M, Villeriot N, Vernadet JP, Couvin D, Toffano-Nioche C *et al*. CRISPRCasdb a successor of CRISPRdb containing CRISPR arrays and cas genes from complete genome sequences, and tools to download and query lists of repeats and spacers. *Nucleic Acids Res*. 2020; 48:D535-D544.

25. Meier-Kolthoff JP, Goker M. VICTOR: genome-based phylogeny and classification of prokaryotic viruses. *Bioinformatics*. 2017; 33:3396-3404.
26. Ågren J, Sundström A, Håfström T, Segerman B. Gegenees: fragmented alignment of multiple genomes for determining phylogenomic distances and genetic signatures unique for specified target groups. *PLoS One*. 2012; 7:e39107.
27. Piochi M, Mormone A, Strauss H, Balassone G. The acid sulfate zone and the mineral alteration styles of the Roman Puteoli (Neapolitan area, Italy): clues on fluid fracturing progression at the Campi Flegrei volcano. *Solid Earth*. 2019; 10:1809-1831.

SUPPLEMENTARY FIGURES

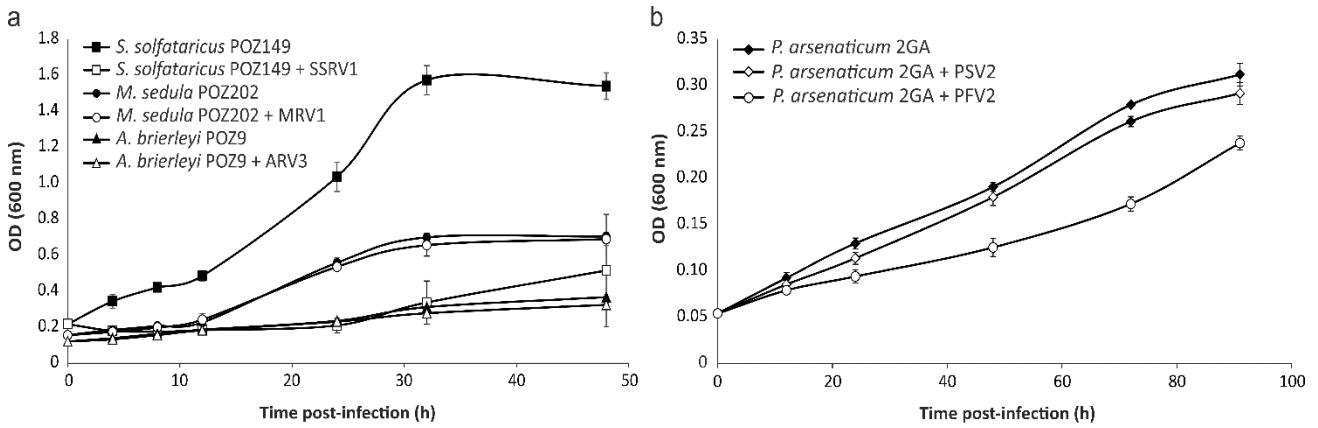


Figure S1. Impact of the isolated viruses on the growth of their respective hosts. a Growth curves of cell cultures infected (open symbols) or non-infected (filled symbols) with the three ruidiviruses. **b** Growth curves of *P. arsenaticum* 2GA cells infected (open symbols) or non-infected (filled symbols) with PFV2 and PSV2. The optical density (OD) at 600 nm was measured over a period of 48 h for the cultures infected with ruidiviruses (**a**) and 91h for *Pyrobaculum* cultures (**b**). Error bars represent standard deviation from three independent measurements.

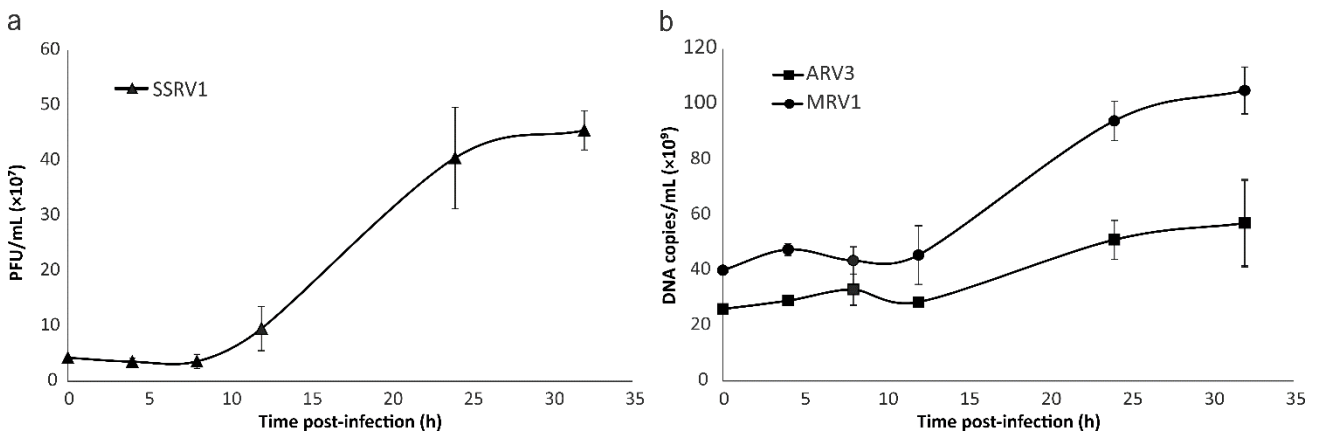


Figure S2. Quantification of virus production. a Production of SSRV1 particles was measured by titration and plaque assay at the indicated time points. **b** Release of ARV3 and MRV1 DNA copies was followed by qPCR over a period of 32 h by using primers targeting specific virus sequences. Error bars represent standard deviation from three independent measurements.

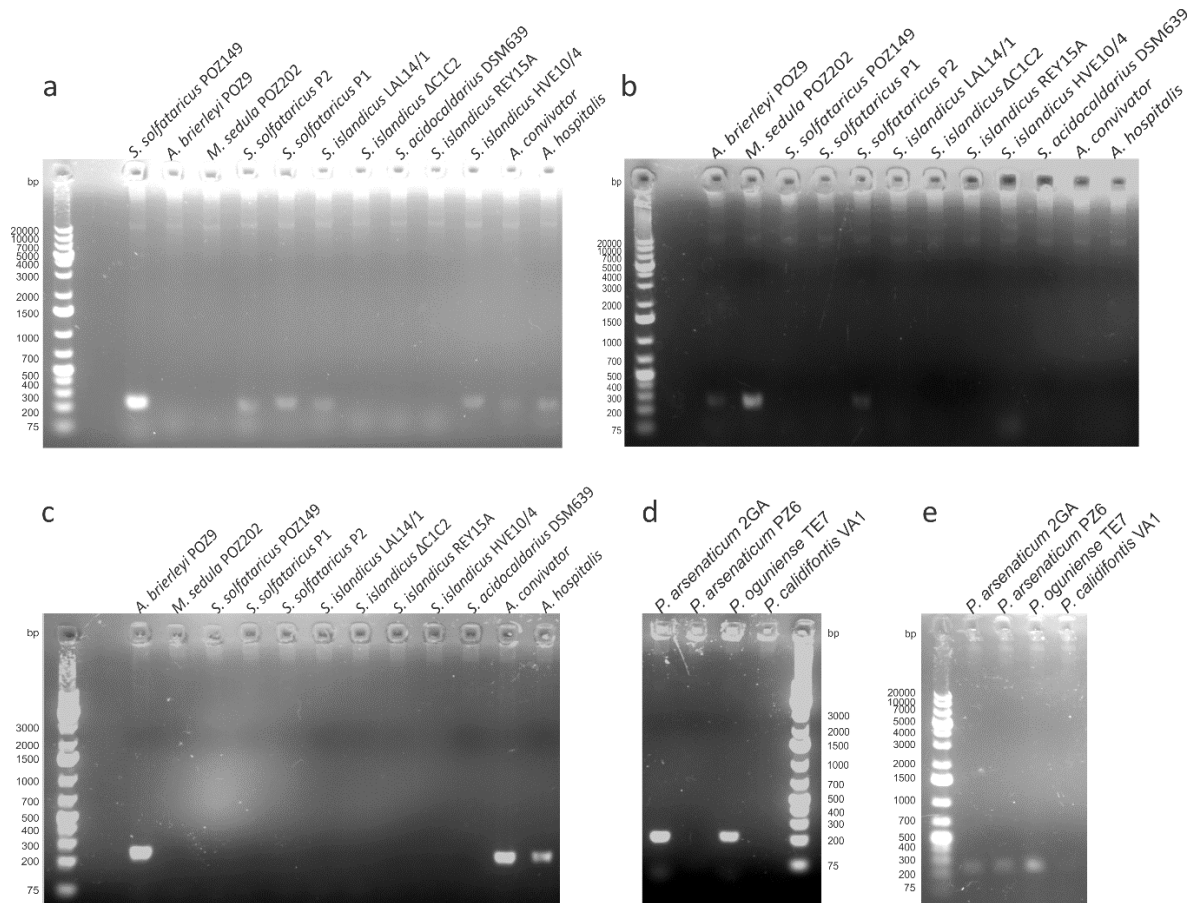


Figure S3. Entry of viral DNA into cells. Detection of SSRV1 (a), MRV1 (b), ARV3 (c), PFV2 (d) and PSV2 (e) DNA in different archaeal cells after 1 hpi. Members of the family *Sulfolobaceae* were mixed with the rudiviruses SSRV1, MRV1 and ARV3, whereas PFV2 and PSV2 were mixed with *Pyrobaculum* strains. PCR reactions targeting viral genomes were performed with extensively washed infected cells as a template.

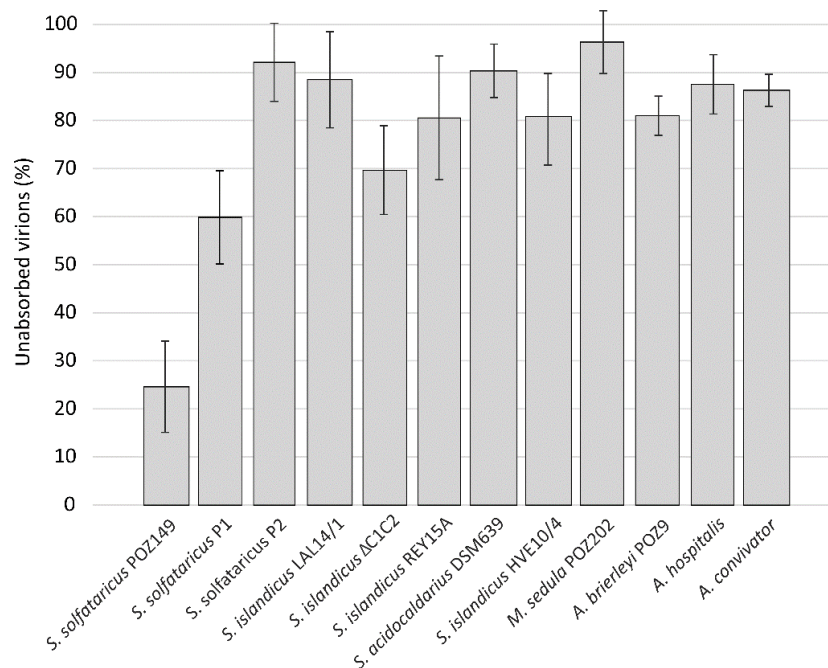


Figure S4. Adsorption of SSRV1 to different strains from the family *Sulfolobaceae*. Cells were mixed with SSRV1 at an MOI of 0.05 and kept at 75°C for 1 h. The number of unbound virions was measured by titration and plaque assay of the cell-free supernatants at 1 hpi. Error bars represent standard deviation from three independent measurements.

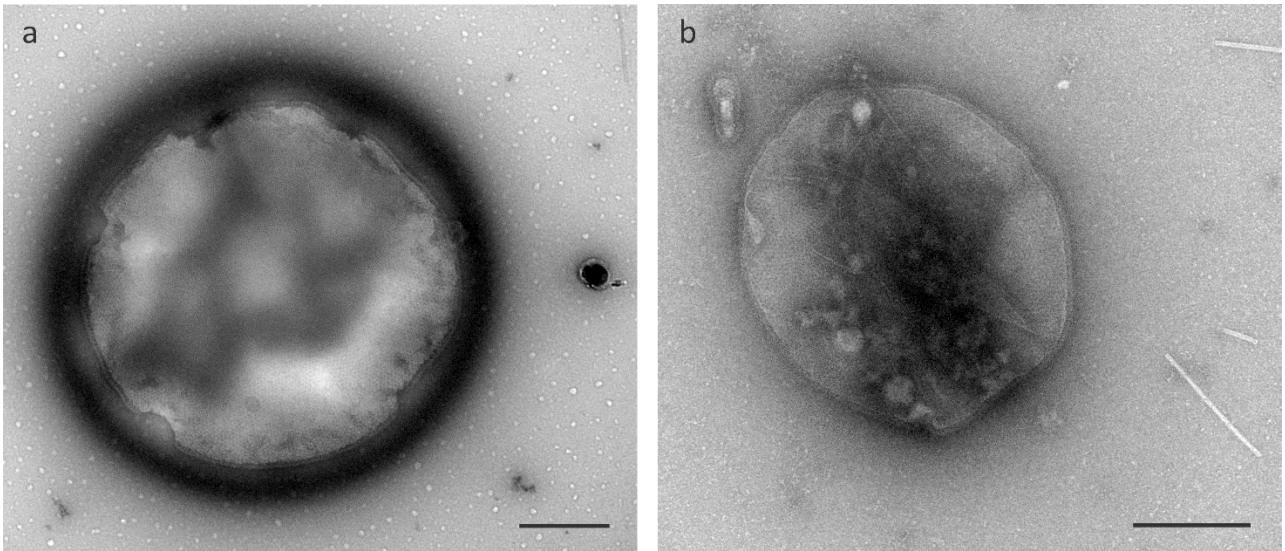


Figure S5. Transmission electron micrographs of infected cells at 32 hpi. a *Metallosphaera sedula* POZ202. **b** *Acidianus brierleyi* POZ9. The samples were negatively stained with 2% uranyl acetate. Scale bars: 500 nm.

SUPPLEMENTARY TABLES

Table S1. Sampling conditions and geological features of the nine samples collected in the solfataric field of Pisciarelli in Pozzuoli, Italy. Coordinates are displayed in the Decimal Degrees format.

Sample	Coordinates	pH	Temperature (°C)	Sample type	Geological features of the sampled sites*
I1	40° 49.749'N 14° 08.823'E	1-2	94	Milky water	Persistent vigorous bubbling water pool, 1-2 m large and several tens of cm deep, located at the base of the southern scarp (L60b)
I2	40° 49.748'N 14° 08.822'E	1	91.8	Milky water	Bubbling water within a hole which is 50 cm wide and 1 m deep, found at the base of the southern scarp (L60)
I3	40° 49.755'N 14° 08.824'E	4-5	81.9-91.2	Grey mud-bearing water	Main mud pool with bubbling zones
I4	40° 49.758'N 14° 08.821'E	7	95.6	Water spray	Small-sized fracture at the soil surface nearby the main mud pool (Fratturina)
I5	40° 49.763'N 14° 08.825'E	N.D.	95	Blackish loose sand (occasional substrate for native sulfur crystallization)	Vapour-emitting fault of the scarp delimiting the mud pool area to the north (L1vent)
I6	40° 49.748'N 14° 08.832'E	N.D.	91.8	Grey mud with whitish sand	Vapour chimney above the scarp, which borders the mud pool area to the south (L20camino)
I7	40° 49.750'N 14° 08.835'E	N.D.	94.8	Grey to black mud	Chimney with low intensity, geyser-like emissions, located above the scarp bordering the mud pool to the south (L19)
I8	40° 49.747'N 14° 08.823'E	N.D.	93.3	Blackish sand (substratum of native sulfur crystallization)	Wall rock with feeble vapour emissions above the scarp proxy to the water pools (L70)
I9	40° 49.745'N 14° 08.821'E	N.D.	92.6	Blackish to whitish sand (substratum of native sulfur crystallization)	Wall rock with feeble vapour emissions found at the base of the western scarp (L100)

* - the information according to [27]. N.D., not determined. Sampling conducted on October 30, 2018.

Table S2. Functional annotation of the PSV2 genome.

ORF	Position	Length, aa	TMD	Annotation	HHpred hit	Probability (%)	Blast hit	E-value	Identity (%)
1	276-491	71	-						
2	515-2092	525	-	AAA+ domain found in chaperone proteins ClpB and heat shock proteins	6AZY_A	99.6	PSV [YP_015523]	4E-09	27.44
3	2122-3045	307	+ (9)	Cation exchanger, membrane transport protein	4KJR_A	95.5	PSV [YP_015524]	3E-94	51.03
4	3085-3225	46	+ (1)	Circadian clock protein KaiB	5JWO_B	93			
5	3261-3710	149	-				PSV [YP_015528]	4E-09	32.46
6	3720-3983	87	-				PSV [YP_015529]	4E-08	36.71
7	4022-4726	234	-				PSV [YP_015536]	5E-42	38.71
8	4755-5486	243	+ (1)				PSV [YP_015537]	2E-64	52.58
9	5516-6424	302	+ (9)				PSV [YP_015538]	4E-42	35.05
10	6448-6765	105	-						
11	6785-7114	109	+ (2)	DUF1286-family protein	PF06939.11	99.5	PSV [YP_015539]	2E-19	47.86
12	7133-7849	238	+ (3)	Structural protein VP3			PSV [YP_015540]	3E-12	29.66
13	7895-8167	90	-						
14	8192-8917	241	+ (6)	Structural protein VP2			PSV [YP_015541]	1E-77	64.62
15	8919-9821	300	+ (6)				PSV [YP_015542]	3E-50	40.22
16	9818-10213	131	+ (1)						
17	10242-10661	139	+ (1)				PSV [YP_015544]	9E-17	34.59
18	10821-11042	73	-				PSV [YP_015545]	7E-22	54.05
19	11042-11509	155	+ (4)				PSV [YP_015546]	2E-65	63.92
20	11562-11819	85	-	Structural protein VP1			PSV [YP_015547]	2E-25	49.41
21	11872-12207	111	-						
22	12437-12727	96	-						
23	12756-13559	267	+ (2)				PSV [YP_015551]	1E-29	34.07
24	13556-13699	47	-						
25	13731-14138	135	-						
26	14135-14740	201	-				PSV [YP_015531]	1E-06	32.09
27	14727-15446	239	-	PSV protein ORF239	2X3M_A	100	PSV [YP_015532]	2E-17	33.52

28	15479-16249	256	+ (4)				PSV [YP_015552]	3E-09	28.35
29	16279-16446	55	-						
30	16439-16990	183	-						
31	17053-17562	169	-						
32	17555-18034	159	-	Hypothetical transcription regulator, WHH	2X4H_B, 2GXG_A	94.4			

TMD, transmembrane domains. Blast hits were obtained with BLASTP, using the nr protein database (E-value of 1e-03). Distant homologs were identified by HHpred.

Table S3. Functional annotation of the PFV2 genome.

ORF	Position	Length. aa	TDM	Annotation	HHpred hit	Probability (%)	Blast hit	E value	Identity (%)
1	316-753	145	-				PFV1 [YP_009237217]	3E-99	97.93
2	761-1102	113	-				PFV1 [YP_009237217]	3E-78	100.00
3	1105-1371	88	-				PFV1 [YP_009237219]	2E-53	95.45
4	1368-1487	39	-				PFV1 [YP_009237220]	1E-07	95.83
5	1484-1822	112	-				PFV1 [YP_009237221]	1E-74	94.64
6	1819-2097	92	-				PFV1 [YP_009237222]	2E-57	97.83
7	2094-2426	110	-				PFV1 [YP_009237223]	3E-75	99.09
8	2423-2785	120	-				PFV1 [YP_009237224]	2E-84	99.17
9	2786-2983	65	-				PFV1 [YP_009237225]	1E-37	100.00
10	2976-3620	214	+(1)				PFV1 [YP_009237226]	2E-156	98.60
11	3617-3985	122	-				PFV1 [YP_009237227]	5E-85	100.00
12	3966-4535	189	-	Cas_Cas4 ; DUF83	PF01930.17	97.9	PFV1 [YP_009237228]	1E-136	99.47
13	4537-4734	65	-				PFV1 [YP_009237229]	4E-39	100.00
14	4727-4918	63	+(2)				PFV1 [YP_009237230]	2E-38	96.83
15	4915-5118	67	+(3)				PFV1 [YP_009237231]	1E-40	100.00
16	5122-5451	109	-				PFV1 [YP_009237232]	1E-71	98.17
17	5453-5920	155	-				PFV1 [YP_009237233]	4E-108	99.35
18	5917-6153	78	-				PFV1 [YP_009237234]	1E-52	100.00
19	6150-6569	139	-				PFV1 [YP_009237235]	5E-95	98.56
20	6640-7029	129	-	PFV1 structural protein VP1			PFV1 [YP_009237236]	8E-88	97.67
21	7026-7463	145	-	PFV1 structural protein VP2			PFV1 [YP_009237237]	4E-97	99.31%

22	7465-7986	173	-				PFV1 [YP_009237238]	5E-119	98.85
23	7983-8675	230	-	Chymotrypsin-like peptidase; Peptidase family C4; PFV1 structural protein VP3	PF00863.19	95.6	PFV1 [YP_009237239]	7E-163	96.96
24	8706-9257	183	+(2)				PFV1 [YP_009237240]	9E-106	94.48
25	9260-9637	125	-				PFV1 [YP_009237241]	4E-83	96.80
26	10314-9748	188	-	STIV glycosyltransferase A197	2C0N_A	90.8	PFV1 [YP_009237242]	1E-141	100.00
27	10450-10704	84	-				PFV1 [YP_009237243]	1E-35	88.73
28	10722-10928	68	+(2)				PFV1 [YP_009237248]	3E-38	95.52
29	10925-11956	343	+(2)				PFV1 [YP_009237249]	0.0	99.13
30	11965-12324	119	+(3)				PFV1 [YP_009237250]	5E-68	98.21
31	12324-12614	96	+(3)				PFV1 [YP_009237251]	3E-61	98.96
32	12602-13009	135	+(1)				PFV1 [YP_009237252]	1E-55	98.82
33	13067-13684	205	+(1)						
34	13702-15339	545	+(1)	Glycoside hydrolase, family 43	cd08984	97.2	PFV1 [YP_009237244]	5E-140	50.21
35	15345-16214	289	-	Glycosyltransferase-like family 2 (Glyco_tranf_2_2)		94.8	<i>Thermoproteus tenax</i> virus 1 [YP_009509148]	2E-50	37.05
36	16248-16409	53	-						
37	16463-16726	87	-	ferritin (<i>Pyrococcus furiosus</i> COM1)	5N5E_C	99.5	PFV1 [YP_009237253]	2E-54	97.70
38	16743-16958	71	-				PFV1 [YP_009237254]	7E-43	97.18
39	16942-17445	167	-	Ricin-type beta-trefoil lectin C-terminal domain (Lectin_C_term)	PF18022.1	99.9	PFV1 [YP_009237255]	3E-118	97.60

TMD, transmembrane domains. Blast hits were obtained with BLASTP, using the nr protein database (E-value of 1e-03). Distant homologs were identified by HHpred.

Table S4. Functional annotation of the SSRV1 genome.

ORF	Position	Length, aa	TMD	Annotation	HHpred hit	Probability (%)	Blast hit	E-value	Identity (%)
1	35-271	78	-				ARV2 [YP_009230213]	1E-19	46.67
2	284-460	58	-	CopG DNA binding proteins, CopG-like RHH motif	1IRQ_A	98.7	SRV [YP_009094260]	3E-23	74.14
3	955-521	144	-	HUH superfamily endonuclease, Rep	2X3G_A	100	ARV2 [YP_009230215]	7E-61	73.24
4	1207-992	71	-	SIRV1 ORF56B, transcription factor	2KEL_B	98.7	ARV2 [YP_009230216]	6E-20	75.51
5	1299-1580	93	-	ParG-like RHH protein with 4 α -helices	PF09274.10	94	ARV2 [YP_009230218]	7E-38	83.33
6	1582-1962	126	-	SIRV1 ORF131	2X5H_D	100	ARV2 [YP_009230219]	1E-62	80.16

7	1963-2235	90	+ (1)				ARV2 [YP_009230220]	5E-37	62.22
8	2379-3686	435	-	AAA+ ATPase domain	PF13337.6	100	ARV2 [YP_009230221]	0.0	89.66
9	3692-4318	208	-	Cas4, CRISPR, MCSG, Exonuclease	4IC1_H	99.8	ARV2 [YP_009230222]	3E-130	84.54
10	4675-5001	1008	-	SIRV2 major capsid protein	3J9X	99.9	ARV2 [YP_009230223]	6E-67	93.46
11	5083-5487	134	-	SIRV2 major capsid protein	3J9X	100	ARV2 [YP_009230224]	1E-83	91.04
12	5568-5792	74	-	DNA-binding domain	2XIG_A	92	ARV2 [YP_009230225]	2E-39	83.78
13	5793-6812	339	-	N-acetyl-alpha-D-glucosaminyl L-malate synthase	5D01_A	99.9	ARV2 [YP_009230226]	0.0	94.10
14	7739-6795	314	-	-	-		ARV2 [YP_009230227]	0.0	88.70
15	7753-8124	123	-				ARV2 [YP_009230228]	4E-74	94.64
16	8117-8365	82	+ (1)				ARV2 [YP_009230229]	9E-37	86.59
17	8572-8366	68	-	-	-		ARV2 [YP_009230230]	1E-28	85.94
18	9062-8565	165	-	GCN5-related N-acetyltransferase; transcriptional regulator (?)	1Z4E_B	99.5	ARV2 [YP_009230231]	4E-109	92.12
19	9461-9114	115	-	Holliday-junction resolvase (<i>S. solfataricus</i>)	1OB8_A	99.6	ARV2 [YP_009230233]	5E-67	93.04
20	10918-9470	482	-				ARV2 [YP_009230234]	0.0	86.84
21	11024-11320	98	+ (1)	SIRV3 pyramid forming protein, c92 STIV			SIRV1 [NP_666630]	5E-26	50.51
22	11351-13138	595	-	Domain of unknown function (DUF2341)	PF10102.9	99.2	ARV2 [YP_009230236]	0.0	82.60
23	13271-16618	1115	-	C-terminal Tale-like DNA-binding domain	4HPZ_A	98.6	ARV2 [YP_009230237]	0.0	86.37
24	16620-17819	399	-				ARV2 [YP_009230238]	5E-172	64.13
25	17899-18369	156	-	FkbM domain methyltransferase	2PY6_A	99	ARV2 [YP_009230239]	7E-100	86.54
26	18376-18966	196	-	dTDP-rhamnosyl transferase, GT2_RfbF_like	cd02526	90.2	ARV2 [YP_009230240]	5E-110	70.26
27	18992-20023	343	-	GDP-mannose-dependent alpha-(1-6)-phosphatidylinositol monomannoside mannosyltransferase	3OKP_A	99.9	ARV2 [YP_009230241]	0.0	92.11
28	20623-20000	207	-	Endoglycosidase; glycoside hydrolase family 18	6Q64_A	88.7	ARV2 [YP_009230242]	7E-134	87.92
29	21886-20633	417	+ (12)	Amino acid transporters	PF00324.21	99.9	ARV2 [YP_009230243]	0.0	87.29
30	22720-21908	270	-	Transcription initiation factor IIB	5IYD_M	100	ARV2 [YP_009230244]	2E-125	64.44
31	23103-22879	74	-				<i>Acidianus</i> two-tailed virus [YP_319834]	8E-24	56.76
32	23256-23113	47	-				ARV2 [YP_009230245]	2E-16	72.34
33	23599-23264	111	-	AcrIII-1 (DUF1874)	PF08960.10	100	ARV2 [YP_009230246]	1E-51	68.47
34	24491-23649	280	-	MPN_archaeal; Mov34/MPN/PAD-1 family: archaeal JAB1/MPN/Mov34 metalloenzyme. This	cd08059	95.5	ARV2 [YP_009230247]	4E-171	81.79

				family contains only archaeal MPN-like domains.					
35	25188-24664	174	-						
36	25582-25331	83	-	-			<i>Acidianus filamentous virus 6</i> [YP_001604177]	4E-11	40.70%
37	26062-25787	91	-	PadR-family transcriptional regulator; PadR, wHTH DNA binding domain	5DYM_A	94.5	ARV1 [YP_001542655]	0.21	34.88%

TMD, transmembrane domains. Blast hits were obtained with BLASTP, using the nr protein database (E-value of 1e-03). Distant homologs were identified by HHpred.

Table S5. Functional annotation of the MRV1 genome.

ORF	Position	Length, aa	TMD	Annotation	HHpred hit	Probability (%)	Blast hit	E-value	Identity (%)
1	16-192	58	-	CopG DNA binding proteins, CopG-like RHH motif	1IRQ_A	98.8	SIRV3 [YP_009272968]	9E-17	62.07
2	663-217	148	-	HUH superfamily endonuclease	2X3G_A	100	ARV2 [YP_009230215]	7E-65	72.11
3	846-700	48	-	Transcription repressor/SIRV1/56B; RHH motif	2KEL_B	98.9	ARV2 [YP_009230216]	7E-20	78.72
4	1034-1252	72	-	ParG-like RHH protein with 4 α -helices	PF09274.10	95.8	ARV2 [YP_009230218]	3E-39	84.72
5	1254-1622	122	-	SIRV1 ORF131	2X5H_D	100	ARV2 [YP_009230219]	1E-59	76.98
6	1623-1895	90	-				ARV2 [YP_009230220]	4E-50	81.11
7	2075-3382	435	-	AAA+ ATPase domain	PF13337.6	100	ARV2 [YP_009230221]	0.0	92.64
8	3379-4014	211	-	Cas4, CRISPR, MCSG, Exonuclease	4IC1_H	99.9	ARV2 [YP_009230222]	2E-133	86.73
9	4083-4409	108	-	SIRV2 major capsid protein	3J9X	99.9	ARV2 [YP_009230223]	4E-60	85.05
10	4491-4895	134	-	SIRV2 major capsid protein	3J9X	100	ARV2 [YP_009230224]	7E-85	91.04
11	4976-5200	74	-	Phage integrase, N-terminal SAM-like domain (Phage_int_SAM_4)	PF13495.6	93.5	ARV2 [YP_009230225]	1E-38	79.73
12	5201-6220	339	-	GDP-mannose-dependent alpha-(1-6)-phosphatidyl inositol monomannoside mannosyltransferase	3OKP_A	99.9	ARV2 [YP_009230226]	0.0	88.20
13	7147-6203	314	-				ARV2 [YP_009230227]	1E-178	84.93
14	7194-7532	112	-				ARV2 [YP_009230228]	1E-66	85.71
15	7525-7773	82	+ (1)				ARV2 [YP_009230229]	2E-36	85.37
16	7968-7774	64	-				ARV2 [YP_009230230]	5E-34	92.19
17	8458-7961	165	-	GCN5-related N-acetyltransferase	2OH1_C	96.8	ARV2 [YP_009230231]	2E-105	87.58
18	8845-8498	115	-	GCN5-related N-acetyltransferase	1OB8_A	99.6	ARV2 [YP_009230231]	2E-105	87.58
19	10341-8854	495	-				ARV2 [YP_009230233]	1E-64	89.57

20	10456-10695	79	+	(2)				ARV2 [YP_009230235]	2E-46	93.67
21	10741-12546	601	-		Domain of unknown function (DUF2341)	PF10102.9	99	ARV2 [YP_009230236]	0.0	78.33
22	12679-16026	1115	-		C-terminal Tale-like DNA-binding domain	4HPZ_A	98.2	ARV2 [YP_009230237]	0.0	84.64
23	16028-17107	359	-					ARV2 [YP_009230238]	5E-113	49.06
24	17187-17657	156	-		FkbM domain methyltransferase	PF05050.12	99.1	ARV2 [YP_009230239]	2E-103	87.82
25	17664-18254	196	-		Glycosyltransferase like family 2 (Glyco_tranf_2_2)	PF10111.9	83.9	ARV2 [YP_009230240]	1E-120	76.29
26	18280-19311	343	-		GDP-mannose-dependent alpha-(1-6)-phosphatidyl inositol monomannoside mannosyltransferase	3OKP_A	99.9	ARV2 [YP_009230241]	0.0	87.72
27	20136-19288	282	-		Transcription initiation factor IIB	5IYD_M	100	ARV2 [YP_009230244]	2E-162	77.09

TMD, transmembrane domains. Blast hits were obtained with BLASTP, using the nr protein database (E-value of 1e-03). Distant homologs were identified by HHpred.

Table S6. Functional annotation of the ARV3 genome.

ORF	Position	Length, aa	TMD	Annotation	HHpred hit	Probability (%)	Blast hit	E-value	Identity (%)
1	839-1012	57	-						
2	1039-1410	123	-						
3	1562-1792	76	-				ARV1 [YP_001542654]	2E-20	57.14%
4	1802-1987	61	-	Transcriptional regulator, CopG-like RHH_7 motif	2K29_B	98.8	<i>Acidianus</i> spindle-shaped virus 1 [YP_003331424]	3E-24	78.18%
5	2173-2478	101	-	Probable transcriptional regulator PA3067	d2hr3a1	87.4	ATV2 [AON96442]	3E-52	91.11%
6	3022-2552	156	-	HUH superfamily endonuclease, Rep	2X3G_A	100	ARV2 [YP_009230215]	3E-50	55.77%
7	3241-3086	51	-	Transcription repressor/SIRV1/56B; RHH motif	2KEL_B	98.9	ARV2 [YP_009230216]	5E-16	70.45%
8	3366-3659	97	-	Transcription repressor/SIRV1/56B; RHH motif	2KEL_B	96.4	ARV2 [YP_009230218]	8E-28	68.06%
9	3661-4053	130	-	SIRV1 ORF131	2X5H_D	100	ARV2 [YP_009230219]	2E-47	60.00%
10	4069-4323	84	-				ARV2 [YP_009230220]	3E-11	35.80%
11	4518-5828	436	-	AAA+ ATPase domain	PF13337.6	100	ARV2 [YP_009230221]	0.0	82.30%
12	5831-6457	208	-	Cas4, CRISPR, MCSG, Exonuclease	4IC1_H	99.8	ARV2 [YP_009230222]	9,0E-115	74.40%
13	6550-6879	109	-	SIRV2 major capsid protein	3J9X	97.2	ARV2 [YP_009230223]	6E-43	60.75%
14	7018-7422	134	-	SIRV2 major capsid protein	3J9X	100	ARV1 [YP_001542641]	7E-84	87.31%

15	7527-7751	74	-	Phage integrase, N-terminal SAM-like domain (Phage_int_SAM_4)	PF13495.6	90.7	ARV2 [YP_009230225]	1E-38	82.43%
16	7752-8771	339	-	GDP-mannose-dependent alpha-(1-6)-phosphatidyl inositol monomannoside mannosyltransferase	3OKP_A	99.9	ARV2 [YP_009230226]	0.0	85.84%
17	9698-8754	314	-				ARV2 [YP_009230227]	2E-177	82.53%
18	9904-10191	95	-				ARV2 [YP_009230228]	3E-47	78.95%
19	10184-10432	82	+ (1)				ARV2 [YP_009230229]	2E-34	84.15%
20	10632-10427	70	-				ARV2 [YP_009230230]	3E-26	78.12%
21	11153-10632	173	-	GCN5-related N-acetyltransferase	2OH1_C	96.8	ARV2 [YP_009230231]	6E-100	90.67%
22	11392-11195	65	+ (2)				ARV2 [YP_009230232]	2E-17	83.33%
23	11741-11385	118	-	Holliday-junction resolvase (<i>S. solfataricus</i>)	1OB8_A	99.7	ARV2 [YP_009230233]	3E-58	77.19%
24	13156-11741	471	-				ARV2 [YP_009230234]	0.0	79.62%
25	13355-13594	79	+ (2)				ARV2 [YP_009230235]	1E-44	88.61%
26	13640-15421	593	-	Domain of unknown function (DUF2341)	PF10102.9	99.3	ARV2 [YP_009230236]	0.0	77.53%
27	15518-18865	1115	-	C-terminal Tale-like DNA-binding domain	4HPZ_A	97.7	ARV2 [YP_009230237]	0.0	82.78%
28	18867-20078	403	-				ARV1 [YP_001542646]	5E-91	39.80%
29	20242-20712	156	-	FkbM domain methyltransferase	PF05050.12	99.1	ARV2 [YP_009230239]	1E-92	80.77%
30	20719-21309	196	-	Glycosyltransferase like family 2 (Glyco_tranf_2_2)	PF10111.9	89.4	ARV2 [YP_009230240]	4E-88	59.48%
31	21335-22372	345	-	GDP-mannose-dependent alpha-(1-6)-phosphatidyl inositol monomannoside mannosyltransferase	3OKP_A	99.9	ARV2 [YP_009230241]	0.0	87.68%
32	23191-22346	281	-	Transcription initiation factor IIB	5IYD_M	100	ARV2 [YP_009230244]	4E-82	44.57%
33	23550-23326	74	-				ATV [YP_319834]	3E-22	50.00%

TMD, transmembrane domains. Blast hits were obtained with BLASTP, using the nr protein database (E-value of 1e-03). Distant homologs were identified by HHpred.

Table S8. Comparison of the rudivirus genome sequences using the Gegenees tool (BLASTN mode and cutoff threshold for non-conserved material of 40%).

	SMRV1	SRV	ARV1	ARV3	MRV1	ARV2	SSRV1	SBRV1	SIRV1	SIRV2	SIRV3	SIRV10	SIRV8	SIRV9	SIRV4	SIRV5	SIRV6	SIRV7	SIRV11		
SMRV1	100	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	<i>Mexirudivirus</i>
SRV	0	100	0	0	0	0	0	0	0	0	0	0	43	0	0	0	0	0	0	0	<i>Azorudivirus</i>
ARV1	0	0	100	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	<i>Itarudivirus</i>
ARV3	0	0	0	100	51	52	52	0	0	0	0	0	0	0	0	0	0	0	0	0	
MRV1	0	0	0	51	100	58	58	0	0	0	0	0	0	0	0	0	0	0	0	0	<i>Hoswirudivirus</i>
ARV2	0	0	0	53	58	100	60	0	0	0	0	0	0	0	0	0	0	0	0	0	
SSRV1	0	0	0	51	58	60	100	0	0	0	0	0	0	0	0	0	0	0	0	0	
SBRV1	0	0	0	0	0	0	0	100	41	0	41	0	40	40	0	0	0	0	0	42	<i>Japarudivirus</i>
SIRV1	0	0	0	0	0	0	0	0	100	71	71	49	49	49	47	46	46	46	46	46	
SIRV2	0	0	0	0	0	0	0	0	71	100	72	48	50	51	44	49	49	49	49	45	<i>Icerudivirus</i>
SIRV3	0	0	0	0	0	0	50	41	70	72	100	51	52	55	47	47	47	47	47	43	
SIRV10	0	0	0	0	0	0	0	0	49	47	51	100	77	81	66	70	69	70	66	66	
SIRV8	0	42	0	0	0	0	0	43	50	49	49	77	100	93	67	68	68	67	67	65	
SIRV9	0	0	0	0	0	0	0	42	48	51	51	81	93	100	67	70	69	69	66	66	
SIRV4	0	0	0	0	0	0	0	0	48	43	49	66	67	67	100	83	84	82	85	85	<i>Usarudivirus</i>
SIRV5	0	0	0	0	0	0	0	0	48	49	46	69	68	70	83	100	100	97	83	83	
SIRV6	0	0	0	0	0	0	0	0	45	49	47	69	67	70	83	100	100	97	83	83	
SIRV7	0	0	0	0	0	0	0	0	47	47	47	70	67	70	82	97	97	100	82	82	
SIRV11	0	0	0	0	0	0	0	0	49	42	44	65	66	66	85	84	84	81	100	100	