Genomics and biology of Rudiviruses, a model for the study of virus–host interactions in Archaea

David Prangishvili*[1], Eugene V. Koonin† and Mart Krupovic*

*Institut Pasteur, Department of Microbiology, 25 rue du Dr. Roux, Paris 75015, France, and †National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD 20894, U.S.A.

Abstract

Archaeal viruses, especially viruses that infect hyperthermophilic archaea of the phylum Crenarchaeota, constitute one of the least understood parts of the virosphere. However, owing to recent substantial research efforts by several groups, archaeal viruses are starting to gradually reveal their secrets. In the present review, we summarize the current knowledge on one of the emerging model systems for studies on crenarchaeal viruses, the *Rudiviridae*. We discuss the recent advances towards understanding the function and structure of the proteins encoded by the rudivirus genomes, their role in the virus life cycle, and outline the directions for further research on this model system. In addition, a revised genome annotation of SIRV2 (*Sulfolobus islandicus* rod-shaped virus 2) is presented. Future studies on archaeal viruses, combined with the knowledge on viruses of bacteria and eukaryotes, should lead to a better global understanding of the diversity and evolution of virus–host interactions in the viral world.

Introduction

Studies of viruses of hyperthermophilic archaea resulted in the description of many new, previously unsuspected, virion morphotypes [1]. However, the biology of these viruses remained largely enigmatic. In the last few years, a substantial effort was made to decipher the functions of proteins encoded by archaeal viruses and to characterize different stages of the viral infection cycles. As a result, several virus–host systems, among both the Crenarchaota [2–4] and the Euryarchaeota [5–8], are emerging as promising models to study in more detail archaeal virus–host interactions. In the present review, we summarize the available information on one of such model systems, the hyperthermophilic crenarchaeon *Sulfolobus islandicus* and its rod-shaped virus SIRV (*Sulfolobus islandicus* rod-shaped virus) 2, provide a revised genome annotation of SIRV2 (Supplementary Table S1 at http://www.biochemsoctrans.org/bst/041/bst0410443add.htm), which should aid future functional studies with this virus, and briefly address the comparative genomics of the rudiviruses.

The *Rudiviridae*, comprising linear non-enveloped ds (double-stranded) DNA viruses [9] (Figure 1), is one of the nine currently recognized families of crenarchaeal viruses [1,10]. The rudiviruses appear to share a common ancestry with another family of filamentous crenarchaeal viruses, the *Lipothrixviridae*. The two families are unified in the order *Ligamenvirales* [11]. The family *Rudiviridae* consists of one genus, *Rudivirus*, and four species: SIRV1, SIRV2, ARV1 (*Acidianus* rod-shaped virus 1) and SRV (*Stygiolobus* rod-shaped virus). All of these viruses originate from terrestrial hot acidic springs in Europe: SIRV1 and SIRV2 are from Iceland (Kverkfjöll and Hveragerdi respectively), ARV1 is from Italy (Pozzuoli), and SRV is from Portugal (San Miguel, the Azores), and respectively infect hyperthermophilic species from the genera *Sulfolobus*, *Acidianus* and *Stygiolobus* of the order Sulfolobales [9,12,13]. Recently, an additional rudiviral genome has been sequenced (GenBank® accession number JX944686). According to the GenBank® record, this virus, SMRV1 (Sulfolobales Mexican rudivirus 1), has been recovered from a hot spring located in Los Azufres National Park, Mexico. Of the five rudivirus isolates, SIRV1 and SIRV2 are the most closely related, with 73% identity across the complete genome sequences (Figure 2). Moreover, SIRV1 and SIRV2 infect closely related strains of *S. islandicus*. Most of the current knowledge on the *Rudiviridae* stems from studies carried out with SIRV1 and SIRV2. Given the extent of similarity between these two viruses, results obtained with one virus are likely to be directly transferable to the other.

Virion architecture

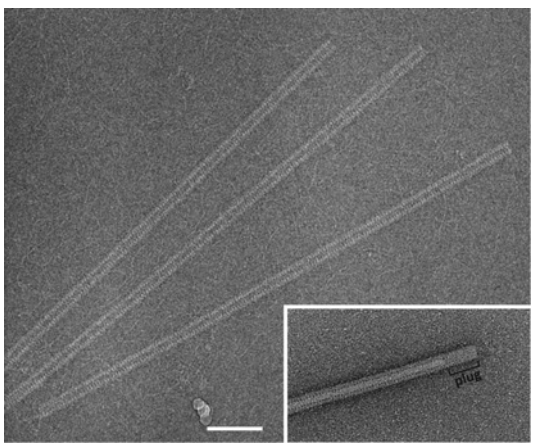
Virion composition and organization of rudiviruses have been investigated biochemically and by electron microscopy. The virion of SIRV2, the type member of the family, is stiff, rod-shaped and measures ~23 nm×900 nm (Figure 1). It contains no envelope and represents a tube-like superhelix formed by dsDNA and multiple copies of the viral protein P134. Negative-contrast electron micrographs suggest that,

Key words: Archaea, filamentous virus, genome replication, hyperthermophile, transcription regulation, virus egress.

Abbreviations used: AFV, *Acidianus* filamentous virus; ARV1, *Acidianus* rod-shaped virus 1; CRISPR, cluster of regularly interspaced palindromic repeats; Cas, CRISPR-associated sequences; ds, double-stranded; GTase, glycosyltransferase; Hjr, Holliday junction resolvase; HTH, helix–turn–helix; ITR, inverted terminal repeat; MCP, major capsid protein; ORF, open reading frame; RHH, ribbon–helix–helix; SIRV, *Sulfolobus islandicus* rod-shaped virus; SRV, *Stygiolobus* rod-shaped virus; ss, single-stranded; STIV, *Sulfolobus* turreted icosahedral virus; STSV1, *Sulfolobus tengchongensis* spindle-shaped virus 1; VAP, virus-associated pyramid.

[1]To whom correspondence should be addressed (email david.prangishvili@pasteur.fr).

Figure 1 | Cryo-electron micrograph of SIRV2 virions
Scale bar, 100 nm. Inset, negative-contrast electron micrograph of a terminal portion of the SIRV2 virion.



at each end, the virion tube carries plugs, ~50 nm×6 nm (Figure 1, inset). However, these plugs are absent from cryo-electron micrographs (Figure 1), and thus may represent an artefact of negative staining. Notably, as in the case of filamentous viruses infecting bacteria {circular ss (single-stranded) DNA genomes, family *Inoviridae* [14]} and plants (linear ssRNA genomes [15]), the virion length in rudiviruses is proportional to the length of the genomic dsDNA [13]. The overall virion organization of rudiviruses also somewhat resembles that of rod-shaped plant-infecting viruses of the family *Virgaviridae*. Indeed, the pitch of the virion helix is about the same (~2.3 nm) for SIRV2 and tobacco mosaic virus [10]. However, the crystal structure of the rudiviral MCP (major capsid protein), P134, revealed a unique four-helix bundle topology [16] (Figure 2), which radically differs from the fold of the tobacco mosaic virus MCP [17]. By contrast, the rudiviral MCP shares a common fold with the MCPs of the enveloped lipothrixvirus AFV (Acidianus filamentous virus) 1 [18], also a member of the order *Ligamenvirales* [11]. The rudiviral MCP can self-assemble to produce filamentous helical structures with diameter and pitch similar to those of the native virions [13].

At each end, the SIRV2 virion carries three terminal fibres with which it attaches to the cellular appendages of the host. The largest viral protein P1070 is a component of these fibres [19]. The fibres appear to be built up of multiple subunits ordered in a linear array [13]. Consistently, analysis of the sequence revealed the presence of a coiled-coil domain, suggesting that each fibre is a homomultimer of intertwining chains of P1070. Besides P134 and P1070, two other viral proteins, encoded by ORF488 and ORF564, were found to be present in the SIRV2 virion, albeit in a very low amount [13]. The four capsid proteins are encoded in all members of the *Rudiviridae* family, and the degree of sequence conservation is very high, especially for the major capsid protein (83–95% identity).

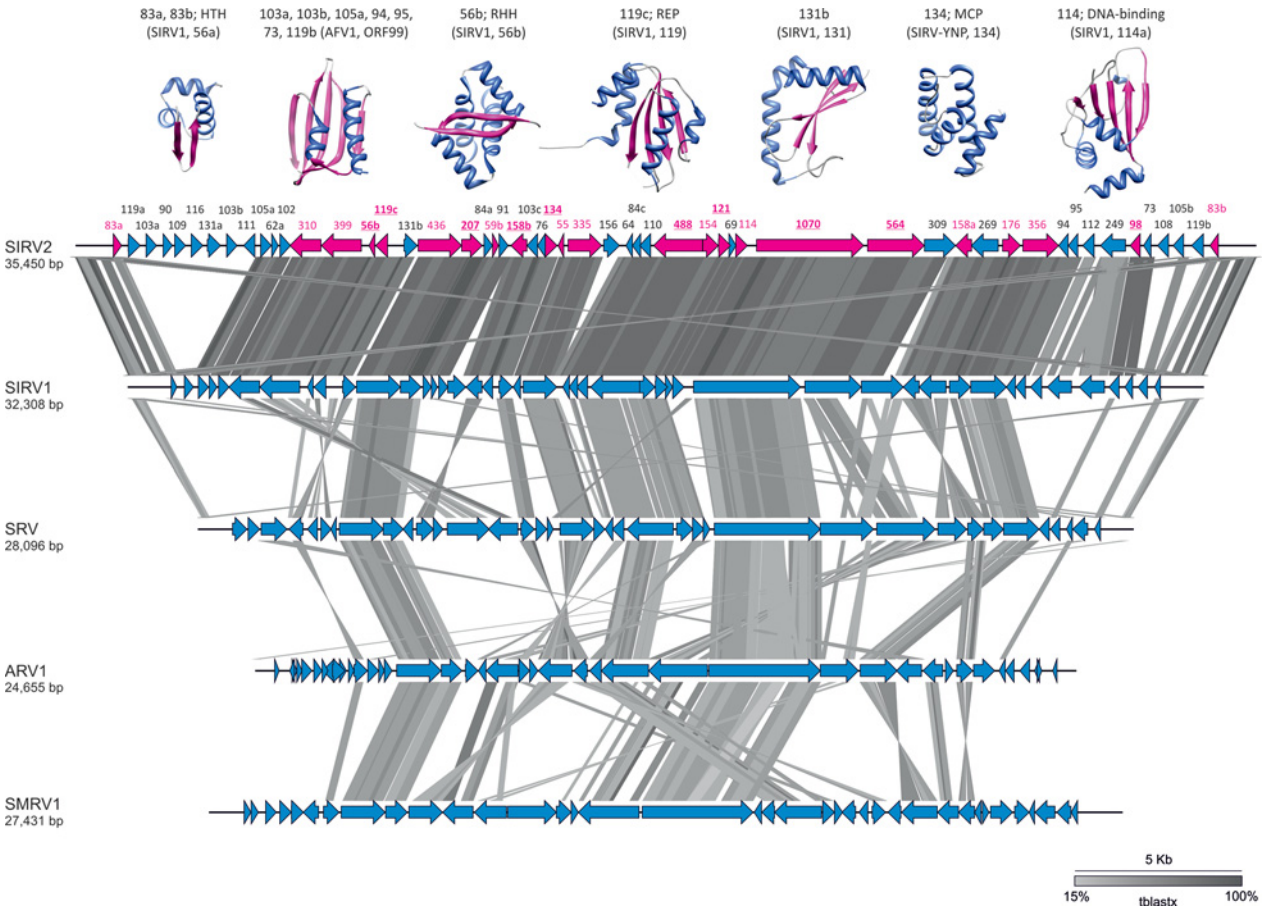
Genome, its replication and nucleic acid metabolism

The linear genome of SIRV2 consists of 35 502 bp and encompasses 1628 bp-long ITRs (inverted terminal repeats). The genomes of other members of the family also contain the ITRs; however, these differ in size and sequence. The two strands of the linear dsDNA of SIRV1 are covalently linked at both ends of the genome, forming a continuous polynucleotide chain [20]. Such structural design of the genome combined with the absence of an identifiable virus-encoded DNA polymerase suggested a unique mechanism of rudiviral genome replication. SIRV1 served as an experimental model for such studies. An insight into the DNA replication mechanism was provided by the observation of single-stranded nicks at a conserved position, 11 nucleotides from the genome terminus, in approximately 5% of DNA molecules isolated from SIRV1 virions [20]. More insights into the replication process were offered by the detection of head-to-head or tail-to-tail linked replicative intermediates in SIRV1-infected cells [21]. On the basis of these results, the self-priming model of SIRV1 genome replication was proposed [22]. According to this model, the replication is initiated by the introduction of single-stranded nick at position 11 from the terminus followed by unfolding of the hairpin loop and reconstruction of the palindrome by elongating the free 3′-hydroxy end. The elongated DNA strand then folds back on itself and replication proceeds by elongation of this structure, resulting in formation of head-to-head and tail-to-tail linked replicative intermediates, which adopt cruciform topology at the borders of genome units by extrusion of the palindromic linkers formed by the ITRs. The resolution of these Holliday junction-like structures gives rise to new copies of the viral genome.

Two SIRV1/2-encoded proteins could be major players in the proposed DNA replication mechanism. The resolution of the crystal structure of P119 of SIRV1 [23,24] enabled recognition of the protein as a member of the Rep superfamily of proteins (Figure 2), which are site-specific endonucleases involved in the initiation of rolling-circle replication of diverse viruses and plasmids [25]. Although P119 and its homologues in other rudiviruses do not show significant sequence similarity to any other protein sequences in database searches performed using PSI-BLAST, HHpred search showed significant similarity ($P = 91.5\%$) to a tyrosine transposase of the TnpA family [26] and limited similarity to other tyrosine transposases and Rep proteins. Moreover, the diagnostic sequence motifs of the Rep superfamily [25] could be identified in the multiple alignment of the rudivirus initiator proteins. The experimentally demonstrated nicking and joining activities of P119, which functions as a dimer, could be involved both in initiation of SIRV1 genome replication by nicking DNA strands close to the terminus and in the formation of a new contiguous DNA strand as a result of the joining activity [23]. Moreover, one of the nicking target sites of the recombinant P119 corresponded to the main nicking site identified in DNA isolated from

Figure 2 | **Genomic relationships between members of the family *Rudiviridae***

SIRV2 ORFs for which a function has been demonstrated or inferred *in silico* are shown in magenta. The names of ORFs, which encode proteins with experimentally verified functions, are underlined. Pairwise tblastx hits between rudiviral genomes are indicated by different shades of grey (the identity scale is included in the Figure). The structures of proteins of the *Ligamenvirales* (rudiviral or lipothrixviral) with orthologues in SIRV2 genome are shown above the SIRV2 genome map, with the source of each structure indicated in parentheses. The protein structures are coloured according to the secondary-structure elements: α-helices, blue; β-strands, magenta; coils, grey. The Figure was prepared using EasyFig [52] and UCSF Chimera [53].



SIRV1 virions, suggesting that cleavage of the sequence by P119 observed *in vitro* is relevant *in vivo* [23]. Owing to its catalytic properties, P119 could potentially be involved also in the resolution of replicative intermediates. However, the more likely candidate to perform the resolution of Holliday junctions at the boarders of two genome units is P121. This protein shows high sequence similarity to archaeal Hjrs (Holliday junction resolvases) and has been shown experimentally to possess Hjr activity; it functions as a dimer with two active centres which independently introduce two nicks in the DNA strands of the Holliday junction [27].

Sequence analysis led to the prediction that two other SIRV1/2 proteins, P158b and P207, participate in nucleic acid metabolism (Supplementary Table S1); in both cases, the putative enzymatic activities were confirmed in biochemical assays. The recombinant P158b, as predicted, was shown to represent a dUTPase, which catalyses the hydrolysis of dUTP to dUMP [28]. Among the rudiviruses, only SIRV1 and SIRV2 encode the dUTPase, but a highly conserved homologue is present in an otherwise unrelated *Sulfolobus* virus, STSV1 (*Sulfolobus tengchongensis* spindle-shaped virus 1), and in numerous archaea and bacteria. The SIRV dUTPase might play an important role in adjusting the intracellular concentration of dTTP in infected cells. Notably, the GC content of SIRV is substantially lower than that of its host (25% compared with 38% GC). Thus the pool sizes of dTTP in host cells might not be optimal for supporting rapid growth of the virus. P207, a member of the RecB nuclease superfamily [29], was indeed shown to display a single-strand-specific endonuclease activity with a cleavage mechanism similar to that of the RecB nuclease [30]. Like the dUTPase P158b, the endonuclease P207 is only present in SIRV1 and SIRV2 among the rudiviruses. Notably, among all members of the RecB family, this protein shows the strongest sequence similarity to the Cas4 protein of the CRISPR (cluster of regularly interspaced palindromic repeats)–Cas (CRISPR-associated

sequences) antivirus defence system [31], suggesting the possibility that the common ancestor of SIRV1 and SIRV2 acquired this gene from a CRISPR–Cas locus. The exact role of the ssDNA endonuclease P207 during the infection cycle is unclear. One possibility is that P207 plays a role in host chromosome degradation during SIRV lytic infection.

Transcription regulation and DNA-binding proteins

The analysis of gene expression of the viruses SIRV1 and SIRV2 by Northern blot hybridization, from 30 min to 3 h after infection, revealed that there is little temporal regulation of viral gene expression [32]. Many genes are clustered and appeared to be transcribed as polycistronic messengers. To promote transcription of its genes, SIRV1 was shown to co-opt host-encoded transcription activator Sta1, which displays a canonical winged HTH (helix–turn–helix) fold [33]. Another candidate involved in viral gene expression is P56b (SvtR) that, in experiments *in vitro*, repressed transcription from several viral promoters including the promoters of its own gene and the gene for the largest structural protein, P1070 [34]. The NMR structure of the protein revealed a typical RHH (ribbon–helix–helix) fold; it has been shown that P56b forms a dimer (Figure 2) and binds DNA with its β-sheet face [34].

In addition to P56b, several other putative DNA-binding proteins were predicted by structural (summarized in [35]) and comparative genomics approaches [29] (Supplementary Table S1). These include P83a/83b (HTH motif), P59b (RHH motif), P55 (zinc-binding domain) and P114 (Table 1). Although the exact function of these proteins has yet to be determined, the presence of typical DNA-binding domains suggests that they might be involved in the regulation of viral and/or cellular promoters. The protein P114 is of special interest because it is highly conserved not only in rudiviruses and lipothrixviruses, but also in other diverse viruses of Crenarchaeota. This protein adopts a unique structural fold, and its homologues in lipothrixvirus AFV3 and STIV (*Sulfolobus* turreted icosahedral virus) were shown to bind dsDNA [36,37].

Viral cycle and virion egress

Owing to the absence of visible indications of cell lysis, it was originally presumed that rudiviruses are not lytic [9]. This view has been challenged by the recent in-depth analysis of SIRV2–*Sulfolobus* interactions [38]. Unexpectedly, massive degradation of the host chromosome was detected starting from early stages of infection. This was followed by virion assembly in the cytoplasm, in the form of several (three or four) densely packed bundles of approximately 50 virions, arranged side by side. Parallel to the virion assembly, heptagonal pyramidal formations, termed VAPs (virus-associated pyramids), are formed at the cell surface, rupturing the S-layer and pointing outwards [38]. The virus cycle ends with opening of the VAPs, allowing release of the cytoplasm and mature virions.

The VAPs could be isolated as stable structural units, hollow baseless pyramids with seven faces (isosceles triangles with angles of 74° and 33°) from the membrane fraction of SIRV2-infected cells [39]. These structures have been shown to consist solely of multiple copies of a single SIRV2 protein, P98 [39,40], which is predicted to be a type II membrane protein [40] and is self-sufficient for formation of pyramidal structures with sevenfold symmetry [39]. One member of the *Rudividae*, ARV1 [12], does not encode a P98 homologue and apparently exploits a different, currently unclear, mechanism of virion egress. Surprisingly, a homologue of P98 is present in STIV that has been shown to exploit a virion egress mechanism similar to that of rudiviruses [2,41,42]. Conceivably, these findings reflect independent evolution of virion morphogenesis and egress systems in archaeal viruses.

Covalent modification of various substrates and other functions

DNA viruses often encode proteins that are responsible for covalent modification of various cellular or viral substrates. These functions are likely to play important roles in modulating virus–host interactions at all stages of the infection cycle and are usually derived from the host genome, judging from typically high sequence similarity to and abundance of cellular homologues [29,43]. Among archaeal viruses, perhaps the most prevalent class of modification-conferring enzyme is GTases (glycosyltransferases). These enzymes are encoded by both euryarchaeal (e.g. His1) [44] and various crenarchaeal viruses [29], including the recently described *Aeropyrum* coil-shaped virus with a single-stranded DNA genome [10]. Genomes of members of the *Ligamenvirales* are particularly enriched in GTase-encoding genes, although the exact number varies among these viruses. SIRV2 encodes three GTases (Table 1 and Supplementary Table S1), whereas the lipothrixvirus SIFV (*Sulfolobus islandicus* filamentous virus) encodes five. The three GTases of SIRV2 do not seem to have evolved via recent duplications; rather, each is represented by highly conserved orthologues in all known rudiviruses and, accordingly, all three can be inferred to be ancestral in this virus family. Moreover, the ultimate origins of these GTase might be different, with one of them (P356) being of apparent bacterial provenance. To date, none of the GTases encoded by archaeal viruses has been functionally characterized. However, irrespective of the presence of a GTase gene in a viral genome, it has been demonstrated that structural proteins of archaeal viruses are often glycosylated [45–47]. The same is true for rudivirus ARV1, the major capsid protein of which has been shown to be sugar-modified [12]. Besides virion proteins, GTases might also be responsible for modification of viral genomes, cellular proteins or host cell envelope. However, the latter modifications remain to be investigated.

In addition to the GTases, SIRV2 encodes three other predicted enzymes potentially involved in modification

Table 1 | **Functions encoded by the SIRV2 genome**

Role	Protein	GenBank® accession number	Evolutionary conservation*	Function	Evidence
Virion structure	P134	NP_666560	Highly conserved homologues in all rudiviruses; structural homologues in lipothrixviruses	Major capsid protein	Experimental evidence
	P488	NP_666567	Highly conserved homologues in all rudiviruses; no other detectable homologues	Structural protein	Experimental evidence
	P1070	NP_666572	Highly conserved homologues in all rudiviruses; middle coiled-coil domains similar to the coiled-coil domains of Smc protein involved in chromosome segregation	Structural protein; terminal fibres	Experimental evidence
	P564	NP_666573	Highly conserved homologues in all rudiviruses; more distant homologues in lipothrixviruses and several archaea	Structural protein	Experimental evidence
DNA binding and transcriptional control	P56b (SvtR)	NP_666549	Among rudiviruses, clear orthologue only in SIRV1; distant homologues in other archaeal viruses; significant similarity to numerous homologues from bacteria and some archaea; RHH domain	Transcriptional regulator	Experimental evidence
	P83a/ P83b	NP_666535/NP_666588	Homologues in all rudiviruses except ARV1. Homologous with oligomerization domains of carbamoyl phosphate synthases (mostly bacterial); HTH domain	DNA-binding and/or protein oligomerization	Structural similarity
	P59b	NP_666555	Highly conserved homologues in all rudiviruses, other archaeal viruses, archaea and bacteria; RHH domain	DNA-binding protein	*In silico* analysis
	P55	NP_666561	Among rudiviruses, conserved only in SIRV1 and SIRV2; more distant homologues in *Thermococcus prieurii* virus 1 and fusellovirus SSV7 (*Sulfolobus* spindle-shaped virus 7); significant similarity to eukaryotic zinc-finger proteins, particularly transcription factors; C_2H_2 zinc finger	DNA-binding protein	*In silico* analysis
	P114	NP_666571	Highly conserved homologues in all rudiviruses, lipothrixviruses and other archaeal viruses; more distant homologues in diverse bacteria; unique protein fold	DNA-binding protein	Structural similarity
Genome replication	P119c	NP_666550	Conserved in all rudiviruses; distant similarity to tyrosine transposases and Rep proteins; conserved motifs required for endonuclease activity	Replication initiator	Experimental evidence
	P121	NP_666569	Conserved in all rudiviruses; strong similarity to archaeal Hjrs	Hjr	Experimental evidence
Nucleic acid metabolism	P158b	NP_666557	Only in SIRV1 and SIRV2 among the rudiviruses. A highly conserved homologue in STSV1 (*Bicaudaviridae*) and numerous archaea and bacteria	dUTPase	Experimental evidence
	P207	NP_666553	Only in SIRV1 and SIRV2 among the rudiviruses, but also in some other archaeal viruses. Homologous with the Cas4 protein of the CRISPR–Cas antivirus immunity systems	ssDNA-specific endonuclease	Experimental evidence

Table 1 | **Continued**

Role	Protein	GenBank® accession number	Evolutionary conservation*	Function	Evidence
Virion egress	P98	NP_666583	Homologues in STIV and all rudiviruses except for ARV1	Formation of VAPs	Experimental evidence
Covalent modification of various substrates	P356	NP_666578	Highly conserved homologues in all rudiviruses and in diverse bacteria. More distant homologues in other viruses	GTase	*In silico* analysis
	P335	NP_666562	Highly conserved homologues in all rudiviruses and in diverse archaea and bacteria. More distant homologues in other viruses	GTase	*In silico* analysis
	P176	NP_666577	Highly conserved homologues in all rudiviruses and lipothrixviruses; distant similarity to bacterial GTases	GTase	*In silico* analysis
	P154	NP_666568	Highly conserved homologues only in SIRV1, SIRV2 and SRV; limited similarity to numerous acetyltransferases of the GCN5 family	Protein acetyltransferase	*In silico* analysis
	P310	NP_666547	Highly conserved homologues in all rudiviruses (except for ARV1), STSV1 and most archaea	Queuine/archaeosine tRNA-ribosyltransferase	*In silico* analysis
	P158a	NP_666575	Highly conserved homologues in all rudiviruses and lipothrixviruses; more distant homologues in numerous archaea and bacteria	AdoMet (S-adenosylmethionine)-dependent (RNA) methyltransferase	*In silico* analysis
Other	P399	NP_666548	Among rudiviruses, only in SIRV1 and SIRV2; moderately conserved homologues in numerous bacteria	Amino acid transporter	*In silico* analysis
	P436	NP_666552	Homologues in all rudiviruses, some lipothrixviruses and in bacteria (ATPase domains of Lon proteases)	AAA+ (ATPase associated with various cellular activities)	*In silico* analysis

*For details, see Supplementary Table S1 at http://www.biochemsoctrans.org/bst/041/bst0410443add.htm.

of various substrates (Table 1). These include a protein acetyltransferase and two RNA-modifying enzymes, namely tRNA-ribosyltransferase (also a member of the GTase superfamily) and S-adenosylmethionine-dependent methyltransferase. Close homologues of the last two proteins are widespread in archaea and are also conserved in lipothrixviruses. In addition, a homologue of the SIRV2 tRNA-ribosyltransferase is encoded by STSV1 [48], a tentative member of the *Bicaudaviridae* family. An apparent function for these proteins is modification of (certain) cellular tRNAs, with consequent potential role in decoding, translation accuracy and control, as well as structural integrity of tRNAs [49].

Finally, SIRV2 encodes two additional proteins that could play important roles during the infection cycle. Protein P436 possesses an AAA+ (ATPase associated with various cellular activities) domain, most closely related to the ATPase domains of ATP-dependent Lon proteases. However, the function of this protein remains enigmatic. Protein P399 possesses 12 transmembrane domains and is homologous with amino acid transporters (Supplementary Table S1). This protein could be important for securing the supply of amino acids for virus propagation and virion assembly, a process which might be compromised as a result of degradation of the cellular chromosome [38]. Interestingly, beyond the rudiviruses, both of these proteins have primarily bacterial, but not archaeal, homologues, emphasizing the apparent bacterial contribution to the evolution of archaeal viruses.

Conclusions

Properties of approximately half of the proteins encoded by the rudiviruses SIRV1 and SIRV2 have been characterized as a result of analysis of their sequences, structures and biochemical characteristics (Table 1). Such a proportion of recognized gene functions is among the highest for crenarchaeal viruses. The discovery of the unique egress mechanism of rudiviruses suggests that the wealth of information on molecular aspects of virus–host interactions from the two other domains of life, Bacteria and Eukarya, is of limited value for the rudiviruses, and perhaps generally for archaeal viruses. Major questions remain to be answered, particularly the following. (i) How do rudiviruses deliver DNA into the host cell? (ii) How is the host replication machinery recruited for preferential production of viral genomes? (iii) How do viral proteins achieve dramatic changes in the host cell leading to the elimination of the host chromosome and establishment of viral factory? (iv) What are the requirements and molecular mechanisms that allow incorporation of rudiviral gene fragments (protospacers) into CRISPR loci of Sulfolobales [50,51]? (v) What drives the assembly of individual virions as well as virion bundles in the cytoplasm of the host cell? (vi) How do rudiviruses achieve well-orchestrated opening of the virion release structures, VAPs, and extrusion of linear virions through the perforations? Answering these questions and deciphering molecular details of the life cycle of rudiviruses promises to unravel unknown aspects of virus–host interaction and could provide novel insights into the origin and evolution of viruses.

Acknowledgements

Thanks are due to all past and present laboratory members and all collaborators who have contributed to the studies described in the present review.

Funding

E.V.K. is supported by the intramural funds of the U.S. Department of Health and Human Services (to the National Library of Medicine).

References

1 Pina, M., Bize, A., Forterre, P. and Prangishvili, D. (2011) The archeoviruses. FEMS Microbiol. Rev. **35**, 1035–1054
2 Fu, C.Y. and Johnson, J.E. (2012) Structure and cell biology of archaeal virus STIV. Curr. Opin. Virol. **2**, 122–127
3 Iverson, E. and Stedman, K. (2012) A genetic study of SSV1, the prototypical fusellovirus. Front. Microbiol. **3**, 200
4 Wirth, J.F., Snyder, J.C., Hochstein, R.A., Ortmann, A.C., Willits, D.A., Douglas, T. and Young, M.J. (2011) Development of a genetic system for the archaeal virus *Sulfolobus* turreted icosahedral virus (STIV). Virology **415**, 6–11
5 Klein, R., Rossler, N., Iro, M., Scholz, H. and Witte, A. (2012) Haloarchaeal myovirus phiCh1 harbours a phase variation system for the production of protein variants with distinct cell surface adhesion specificities. Mol. Microbiol. **83**, 137–150
6 Roine, E. and Oksanen, H.M. (2011) Viruses from the hypersaline environment. In Halophiles and Hypersaline Environments: Current Research and Future Trends (Ventosa, A., Oren, A. and Ma, Y., eds), pp. 153–172, Springer, Berlin
7 Senčilo, A., Paulin, L., Kellner, S., Helm, M. and Roine, E. (2012) Related haloarchaeal pleomorphic viruses contain different genome types. Nucleic Acids Res. **40**, 5523–5534
8 Porter, K. and Dyall-Smith, M.L. (2008) Transfection of haloarchaea by the DNAs of spindle and round haloviruses and the use of transposon mutagenesis to identify non-essential regions. Mol. Microbiol. **70**, 1236–1245
9 Prangishvili, D., Arnold, H.P., Gotz, D., Ziese, U., Holz, I., Kristjansson, J.K. and Zillig, W. (1999) A novel virus family, the *Rudiviridae*: structure, virus–host interactions and genome variability of the *Sulfolobus* viruses SIRV1 and SIRV2. Genetics **152**, 1387–1396
10 Mochizuki, T., Krupovic, M., Pehau-Arnaudet, G., Sako, Y., Forterre, P. and Prangishvili, D. (2012) Archaeal virus with exceptional virion architecture and the largest single-stranded DNA genome. Proc. Natl. Acad. Sci. U.S.A. **109**, 13386–13391
11 Prangishvili, D. and Krupovic, M. (2012) A new proposed taxon for double-stranded DNA viruses, the order "Ligamenvirales". Arch. Virol. **157**, 791–795
12 Vestergaard, G., Haring, M., Peng, X., Rachel, R., Garrett, R.A. and Prangishvili, D. (2005) A novel rudivirus, ARV1, of the hyperthermophilic archaeal genus *Acidianus*. Virology **336**, 83–92
13 Vestergaard, G., Shah, S.A., Bize, A., Reitberger, W., Reuter, M., Phan, H., Briegel, A., Rachel, R., Garrett, R.A. and Prangishvili, D. (2008) *Stygiolobus* rod-shaped virus and the interplay of crenarchaeal rudiviruses with the CRISPR antiviral system. J. Bacteriol. **190**, 6837–6845
14 Rakonjac, J., Bennett, N.J., Spagnuolo, J., Gagic, D. and Russel, M. (2011) Filamentous bacteriophage: biology, phage display and nanotechnology applications. Curr. Issues Mol. Biol. **13**, 51–76
15 Kelloniemi, J., Makinen, K. and Valkonen, J.P. (2008) Three heterologous proteins simultaneously expressed from a chimeric potyvirus: infectivity, stability and the correlation of genome and virion lengths. Virus Res. **135**, 282–291
16 Szymczyna, B.R., Taurog, R.E., Young, M.J., Snyder, J.C., Johnson, J.E. and Williamson, J.R. (2009) Synergy of NMR, computation, and X-ray crystallography for structural biology. Structure **17**, 499–507

17 Krupovic, M. and Bamford, D.H. (2011) Double-stranded DNA viruses: 20 families and only five different architectural principles for virion assembly. Curr. Opin. Virol. **1**, 118–124
18 Goulet, A., Blangy, S., Redder, P., Prangishvili, D., Felisberto-Rodrigues, C., Forterre, P., Campanacci, V. and Cambillau, C. (2009) *Acidianus* filamentous virus 1 coat proteins display a helical fold spanning the filamentous archaeal viruses lineage. Proc. Natl. Acad. Sci. U.S.A. **106**, 21155–21160
19 Steinmetz, N.F., Bize, A., Findlay, K.C., Lomonossoff, G.P., Manchester, M., Evans, D.J. and Prangishvili, D. (2008) Site-specific and spatially controlled addressability of a new viral nanobuilding block: *Sulfolobus islandicus* rod-shaped virus 2. Adv. Funct. Mater. **18**, 2478–3486
20 Blum, H., Zillig, W., Mallok, S., Domdey, H. and Prangishvili, D. (2001) The genome of the archaeal virus SIRV1 has features in common with genomes of eukaryal viruses. Virology **281**, 6–9
21 Peng, X., Blum, H., She, Q., Mallok, S., Brugger, K., Garrett, R.A., Zillig, W. and Prangishvili, D. (2001) Sequences and replication of genomes of the archaeal rudiviruses SIRV1 and SIRV2: relationships to the archaeal lipothrixvirus SIFV and some eukaryal viruses. Virology **291**, 226–234
22 Prangishvili, D. (2003) Evolutionary insights from studies on viruses of hyperthermophilic archaea. Res. Microbiol. **154**, 289–294
23 Oke, M., Kerou, M., Liu, H., Peng, X., Garrett, R.A., Prangishvili, D., Naismith, J.H. and White, M.F. (2011) A dimeric Rep protein initiates replication of a linear archaeal virus genome: implications for the Rep mechanism and viral replication. J. Virol. **85**, 925–931
24 Oke, M., Carter, L.G., Johnson, K.A., Liu, H., McMahon, S.A., Yan, X., Kerou, M., Weikart, N.D., Kadi, N., Sheikh, M.A. et al. (2010) The Scottish Structural Proteomics Facility: targets, methods and outputs. J. Struct. Funct. Genomics **11**, 167–180
25 Ilyina, T.V. and Koonin, E.V. (1992) Conserved sequence motifs in the initiator proteins for rolling circle DNA replication encoded by diverse replicons from eubacteria, eucaryotes and archaebacteria. Nucleic Acids Res. **20**, 3279–3285
26 Messing, S.A., Ton-Hoang, B., Hickman, A.B., McCubbin, A.J., Peaslee, G.F., Ghirlando, R., Chandler, M. and Dyda, F. (2012) The processing of repetitive extragenic palindromes: the structure of a repetitive extragenic palindrome bound to its associated nuclease. Nucleic Acids Res. **40**, 9964–9979
27 Birkenbihl, R.P., Neef, K., Prangishvili, D. and Kemper, B. (2001) Holliday junction resolving enzymes of archaeal viruses SIRV1 and SIRV2. J. Mol. Biol. **309**, 1067–1076
28 Prangishvili, D., Klenk, H.P., Jakobs, G., Schmiechen, A., Hanselmann, C., Holz, I. and Zillig, W. (1998) Biochemical and phylogenetic characterization of the dUTPase from the archaeal virus SIRV. J. Biol. Chem. **273**, 6024–6029
29 Prangishvili, D., Garrett, R.A. and Koonin, E.V. (2006) Evolutionary genomics of archaeal viruses: unique viral genomes in the third domain of life. Virus Res. **117**, 52–67
30 Gardner, A.F., Prangishvili, D. and Jack, W.E. (2011) Characterization of *Sulfolobus islandicus* rod-shaped virus 2 gp19, a single-strand specific endonuclease. Extremophiles **15**, 619–624
31 Makarova, K.S., Haft, D.H., Barrangou, R., Brouns, S.J., Charpentier, E., Horvath, P., Moineau, S., Mojica, F.J., Wolf, Y.I., Yakunin, A.F. et al. (2011) Evolution and classification of the CRISPR–Cas systems. Nat. Rev. Microbiol. **9**, 467–477
32 Kessler, A., Brinkman, A.B., van der Oost, J. and Prangishvili, D. (2004) Transcription of the rod-shaped viruses SIRV1 and SIRV2 of the hyperthermophilic archaeon *Sulfolobus*. J. Bacteriol. **186**, 7745–7753
33 Kessler, A., Sezonov, G., Guijarro, J.I., Desnoues, N., Rose, T., Delepierre, M., Bell, S.D. and Prangishvili, D. (2006) A novel archaeal regulatory protein, Sta1, activates transcription from viral promoters. Nucleic Acids Res. **34**, 4837–4845
34 Guilliere, F., Peixeiro, N., Kessler, A., Raynal, B., Desnoues, N., Keller, J., Delepierre, M., Prangishvili, D., Sezonov, G. and Guijarro, J.I. (2009) Structure, function, and targets of the transcriptional regulator SvtR from the hyperthermophilic archaeal virus SIRV1. J. Biol. Chem. **284**, 22222–22237
35 Krupovic, M., White, M.F., Forterre, P. and Prangishvili, D. (2012) Postcards from the edge: structural genomics of archaeal viruses. Adv. Virus Res. **82**, 33–62
36 Keller, J., Leulliot, N., Cambillau, C., Campanacci, V., Porciero, S., Prangishvili, D., Forterre, P., Cortez, D., Quevillon-Cheruel, S. and van Tilbeurgh, H. (2007) Crystal structure of AFV3-109, a highly conserved protein from crenarchaeal viruses. Virol. J. **4**, 12
37 Larson, E.T., Eilers, B.J., Reiter, D., Ortmann, A.C., Young, M.J. and Lawrence, C.M. (2007) A new DNA binding protein highly conserved in diverse crenarchaeal viruses. Virology **363**, 387–396
38 Bize, A., Karlsson, E.A., Ekefjard, K., Quax, T.E., Pina, M., Prevost, M.C., Forterre, P., Tenaillon, O., Bernander, R. and Prangishvili, D. (2009) A unique virus release mechanism in the Archaea. Proc. Natl. Acad. Sci. U.S.A. **106**, 11306–11311
39 Quax, T.E., Lucas, S., Reimann, J., Pehau-Arnaudet, G., Prevost, M.C., Forterre, P., Albers, S.V. and Prangishvili, D. (2011) Simple and elegant design of a virion egress structure in Archaea. Proc. Natl. Acad. Sci. U.S.A. **108**, 3354–3359
40 Quax, T.E., Krupovic, M., Lucas, S., Forterre, P. and Prangishvili, D. (2010) The *Sulfolobus* rod-shaped virus 2 encodes a prominent structural component of the unique virion release system in Archaea. Virology **404**, 1–4
41 Brumfield, S.K., Ortmann, A.C., Ruigrok, V., Suci, P., Douglas, T. and Young, M.J. (2009) Particle assembly and ultrastructural features associated with replication of the lytic archaeal virus *Sulfolobus* turreted icosahedral virus. J. Virol. **83**, 5964–5970
42 Snyder, J.C., Brumfield, S.K., Peng, N., She, Q. and Young, M.J. (2011) *Sulfolobus* turreted icosahedral virus c92 protein responsible for the formation of pyramid-like cellular lysis structures. J. Virol. **85**, 6287–6292
43 Krupovic, M., Prangishvili, D., Hendrix, R.W. and Bamford, D.H. (2011) Genomics of bacterial and archaeal viruses: dynamics within the prokaryotic virosphere. Microbiol. Mol. Biol. Rev. **75**, 610–635
44 Bath, C. and Dyall-Smith, M.L. (1998) His1, an archaeal virus of the *Fuselloviridae* family that infects *Haloarcula hispanica*. J. Virol. **72**, 9392–9395
45 Kandiba, L., Aitio, O., Helin, J., Guan, Z., Permi, P., Bamford, D.H., Eichler, J. and Roine, E. (2012) Diversity in prokaryotic glycosylation: an archaeal-derived N-linked glycan contains legionaminic acid. Mol. Microbiol. **84**, 578–593
46 Maaty, W.S., Ortmann, A.C., Dlakic, M., Schulstad, K., Hilmer, J.K., Liepold, L., Weidenheft, B., Khayat, R., Douglas, T., Young, M.J. and Bothner, B. (2006) Characterization of the archaeal thermophile *Sulfolobus* turreted icosahedral virus validates an evolutionary link among double-stranded DNA viruses from all domains of life. J. Virol. **80**, 7625–7635
47 Mochizuki, T., Yoshida, T., Tanaka, R., Forterre, P., Sako, Y. and Prangishvili, D. (2010) Diversity of viruses of the hyperthermophilic archaeal genus *Aeropyrum*, and isolation of the *Aeropyrum pernix* bacilliform virus 1, APBV1, the first representative of the family *Clavaviridae*. Virology **402**, 347–354
48 Xiang, X., Chen, L., Huang, X., Luo, Y., She, Q. and Huang, L. (2005) *Sulfolobus tengchongensis* spindle-shaped virus STSV1: virus–host interactions and genomic features. J. Virol. **79**, 8677–8686
49 Phillips, G. and de Crécy-Lagard, V. (2011) Biosynthesis and function of tRNA modifications in Archaea. Curr. Opin. Microbiol. **14**, 335–341
50 Guo, L., Brugger, K., Liu, C., Shah, S.A., Zheng, H., Zhu, Y., Wang, S., Lillestol, R.K., Chen, L., Frank, J. et al. (2011) Genome analyses of Icelandic strains of *Sulfolobus islandicus*, model organisms for genetic and virus–host interaction studies. J. Bacteriol. **193**, 1672–1680
51 You, X.Y., Liu, C., Wang, S.Y., Jiang, C.Y., Shah, S.A., Prangishvili, D., She, Q., Liu, S.J. and Garrett, R.A. (2011) Genomic analysis of *Acidianus hospitalis* W1 a host for studying crenarchaeal virus and plasmid life cycles. Extremophiles **15**, 487–497
52 Sullivan, M.J., Petty, N.K. and Beatson, S.A. (2011) Easyfig: a genome comparison visualizer. Bioinformatics **27**, 1009–1010
53 Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C. and Ferrin, T.E. (2004) UCSF Chimera: a visualization system for exploratory research and analysis. J. Comput. Chem. **25**, 1605–1612

Received 13 November 2012
doi:10.1042/BST20120313

SUPPLEMENTARY ONLINE DATA

Genomics and biology of Rudiviruses, a model for the study of virus–host interactions in Archaea

David Prangishvili*[1], Eugene V. Koonin† and Mart Krupovic*

*Institut Pasteur, Department of Microbiology, 25 rue du Dr. Roux, Paris 75015, France, and †National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD 20894, U.S.A.

Table S1 | **Genome annotation of SIRV2**

SIFV, *Sulfolobus islandicus* filamentous virus; SSV, *Sulfolobus* spindle-shaped virus; TMD, transmembrane domain.

ORF	Name	Position (orientation)	*Ligamenvirales* structures	Predicted function	HHpred/FFAS03/CD hit	Homologues in other viruses and cellular organisms	BLAST hit	Identity; *E* value
gp01	ORF83a	1138–1389	SIRV1, ORF56a	DNA-binding, HTH				
gp02	ORF119a	1584–1943						
gp03	ORF103a	2121–2432	SIFV, ORF14; AFV1, ORF99			*Lipothrixviridae*, *Fuselloviridae* (SSV6, gp21, YP_003331471)		
gp04	ORF90	2635–2907		TMD				
gp05	ORF109	2982–3311		Coiled-coil domain				
gp06	ORF116	3477–3827						
gp07	ORF131a	3959–4354						
gp08	ORF103b	4542–4853	SIFV, ORF14; AFV1, ORF99			*Lipothrixviridae*, *Fuselloviridae* (SSV2, ORF88a, NP_944469; SSV6, gp20, YP_003331470)		
gp09	ORF111	5053–5388 (−)						
gp10	ORF105a	5562–5879	SIFV, ORF14; AFV1, ORF99			*Lipothrixviridae*, *Fuselloviridae* (SSV2, ORF88a, NP_944469; SSV6, gp20, YP_003331470)		
gp11	ORF62a	5897–6085				Multiple homologues in Crenarchaea	*Sulfolobus islandicus* HVE10/4 (YP_005645095) *Acidianus hospitalis* W1 (YP_004458796)	54/60 (90 %); 3×10^{-30} 52/60 (87 %); 5×10^{-29}
gp12	ORF102	6144–6452			HHpred hit to 1te7, $P=$ 97.9	*Fuselloviridae* (SSV6, gp11, YP_003331461); homologues in archaea and their mobile elements	*Ignisphaera aggregans* DSM 17230 (YP_003860652) *Acidianus hospitalis* plasmid pAH1 (YP_002261293)	32/82 (39 %); 2×10^{-8} 34/87 (39 %); 6×10^{-5}
gp13	ORF310	6441–7373 (−)		Queuine/archaeosine tRNA-ribosyltransferase	HHpred to 2ash (tRNA-guanine transglycosylase), $P=$ 100%	*Bicaudaviridae* (STSV1, ORF18, YP_077211); multiple homologues in archaea	*Acidianus hospitalis* W1 (YP_004457216) *Sulfolobus tokodaii* strain 7 (NP_377656)	122/289 (42 %); 2×10^{-62} 118/301 (39 %); 1×10^{-56}
gp14	ORF399	7382–8581 (−)		Amino acid transporter; 12 TMD	HHpred hits to 3gia and 3l1l, $P=$ 100 %; FFAS03 hit to Pfam PF00324 (amino acid permease), score −53.0	Among rudiviruses, only in SIRV1 and SIRV2; moderately conserved homologues in numerous bacteria	*Acidianus hospitalis* W1 (YP_004457471) *Sulfolobus tokodaii* strain 7 (NP_377632)	140/467 (30 %); 6×10^{-51} 132/460 (29 %); 1×10^{-37}
gp15	ORF56b	8818–8988 (−)	SIRV1, ORF56b		Transcriptional regulator, RHH	Among rudiviruses, clear orthologue only in SIRV1; distant homologues in other archaeal viruses, e.g. *Lipothrixviridae* (SIFV, ORF13, NP_445678); significant similarity to numerous homologues from bacteria and some archaea	*Natranaerobius thermophilus* JW/NM-WN-LF plasmid pNTHE01 (YP_001911143)	19/42 (45 %); 1×10^{-3}

[1]To whom correspondence should be addressed (email david.prangishvili@pasteur.fr).

Table S1 | **Continued**

ORF	Name	Position (orientation)	*Ligamenvirales* structures	Predicted function	HHpred/FFAS03/CD hit	Homologues in other viruses and cellular organisms	BLAST hit	Identity; *E* value
gp16	ORF119c	9013–9372 (−)	SIRV1, ORF119	Genome replication initiation protein		Conserved in all rudiviruses; distant similarity to tyrosine transposases and Rep proteins; conserved motifs required for endonuclease activity		
gp17	ORF131b	9868–10263	SIRV1, ORF131	Coiled-coil domain				
gp18	ORF436	10289–11599		AAA+ (ATPase associated with various cellular activities) domain protein (similar to ATPase domain of Lon-like proteases)	FFAS03 hit to Pfam PF13337, score −81.8; CD hit to Lon_2 (pfam13337; putative ATP-dependent Lon protease), $E = 6.3\times10^{-21}$	*Lipothrixviridae* (AFV1, ORF426, YP_003740; AFV2, YP_001496940)	*Archaeoglobus profundus* DSM 5631 (YP_003400858) *Geobacter sulfurreducens* PCA (NP_953154)	84/299 (28%); 3×10^{-15} 49/193 (25%); 8×10^{-3}
gp19	ORF207	11593–12216		CRISPR-associated Cas4-like protein; RecB-like nuclease	CD hit to PHA01622, $E = 8.4\times10^{-117}$; HHpred hit to 3u4q, $P = 99.6\%$; FFAS03 hit to 3l0a, score −37.6	Only in SIRV1 and SIRV2 among the rudiviruses. *Lipothrixviridae* (AFV1, ORF223, YP_003742; SIFV, ORF06, NP_445671), *Fuselloviridae* (SSV7, B206, NP_445671)	Halophilic archaeon DL31 (YP_004810090) *Sulfolobus islandicus* REY15A (YP_005648526)	52/204 (25%); 5×10^{-8} 45/153 (29%); 1×10^{-5}
gp20	ORF84a	12271–12525						
gp21	ORF59b	12522–12701		Transcriptional regulator, RHH motif;	CD hit to RHH_1 (Pfam01402), $E = 4.4\times10^{-4}$; FFAS03 hit to Pfam PF07878, score −14.6; HHpred hit to 2cpg, $P = 99.5\%$	Homologues in multiple archaeal viruses, including *Lipothrixviridae*, *Bicaudaviridae*, *Fuselloviridae*, STIV1/2; numerous cellular homologues	*Sulfolobus islandicus* M.14.25 (YP_002828386) *Sulfolobus acidocaldarius* DSM 639 (YP_255597)	33/58 (57%); 4×10^{-15} 33/58 (57%); 7×10^{-15}
gp22	ORF91	12720–12995				*Fuselloviridae* (SSV1, C102, NP_039806)	*Sulfolobus islandicus* L.S.2.15 (YP_002830940)	21/44 (48%); 1×10^{-4}
gp23	ORF158b	13075–13551 (−)		dUTPase	CD hit to trimeric dUTPase (cd07557), $E = 8.3\times10^{-22}$; HHpred hit to 2qxx and 1pkh, $P = 100\%$; FFAS03 hit to 1ogh, score −82.6	Only in SIRV1 and SIRV2 among the rudiviruses. *Bicaudaviridae* (STSV1, ORF33, YP_077226), bacterial and archaeal (HF1/2, NP_542557) caudovirales; numerous cellular homologues in archaea and bacteria	*Acidianus hospitalis* W1 (YP_004457798) *Sulfolobus solfataricus* 98/2 (YP_005643010)	109/169 (64%); 1×10^{-66} 112/169 (66%); 5×10^{-65}
gp24	ORF103c	13558–13869 (−)						
gp25	ORF76	13873–14103 (−)						
gp26	ORF134	14058–14462	SIRV-YNP	Major capsid protein		*Lipothrixviridae*		
gp27	ORF55	14493–14660 (−)		Zinc-binding domain (C_2H_2), DNA-binding protein	CD hit to ZnF_C2H2 (smart00355), $E = 5\times10^{-3}$; HHpred hit to 2lt7, $P = 99.6\%$	Among rudiviruses, conserved only in SIRV1 and SIRV2; more distant homologue in *Thermococcus prieurii* virus 1 (gp27, YP_005271249) and fusellovirus SSV7 (B82, YP_003331504); significant similarity to eukaryotic zinc-finger proteins, particularly transcription factors; C_2H_2 zinc-finger		
gp28	ORF335	14792–15799		GT1 family of GTases; C-terminal coiled-coil domain	CD hit to GT1_wlbH_like (cd03798), $E = 8\times10^{-17}$; HHpred hit to 3c48, 3okp, 3fro, 3vue, 3oy2, etc. $P = 100\%$; FFAS03 hit to 3oy2, score −54.7	*Aeropyrum* coil-shaped virus (gp38, CCG27851); *Lipothrixviridae*, *Phycodnaviridae*	*Staphylothermus hellenicus* DSM 12710 (YP_003668896) *Gluconobacter oxydans* 621H (YP_191881)	62/237 (26%); 2×10^{-8} 52/207 (25%); 4×10^{-7}
gp29	ORF156	15862–16332				*Lipothrixviridae* (SIFV, ORF44, NP_445707)		

Table S1 | **Continued**

ORF	Name	Position (orientation)	*Ligamenvirales* structures	Predicted function	HHpred/FFAS03/CD hit	Homologues in other viruses and cellular organisms	BLAST hit	Identity; *E* value
gp30	ORF64	16528–16722 (−)				*Lipothrixviridae* (SIFV, ORF62, NP_445725; AFV9, gp04/57, YP_001798522)		
gp31	ORF84c	16719–16973 (−)		Coiled-coil domain, TMD	HHpred hit to 3hnw, $P=$ 91.5%			
gp32	ORF110	16939–17271 (−)						
gp33	ORF488	17361–18827 (−)		Structural protein		Homologues in all rudiviruses; no other detectable homologues		
gp34	ORF154	18836–19300		Protein acetyltransferase; GCN5 family	HHpred hit to 3f8k (GCN5-related N-acetyltransferase from *S. solfataricus*), $P=$ 96.2%			
gp35	ORF121	19307–19672		Archaeal Hjr	CD hit to archeal Hjr (cd00523), $E = 7\times10^{-36}$; HHpred hit to 1ob8 (Hjr from *S. solfataricus*), $P=$ 99.9%	Conserved in all rudiviruses; strong similarity to archaeal Hjrs; homologue in *Thermus* phage P74-26 (YP_001468012	*Sulfolobus islandicus* M.16.27 (YP_002843059) *Metallosphaera sedula* DSM 5348 (YP_001191063)	42/102 (41%); 9×10^{-16} 35/103 (34); 3×10^{-13}
gp36	ORF69	19623–19832						
gp37	ORF114	19829–20173	SIRV1, ORF114a; AFV3, ORF109	Putative DNA-binding protein; unique protein fold		*Lipothrixviridae*, STIV1/2, *Bicaudaviridae*, *Caudovirales*; numerous homologues in bacteria and archaea	*Sulfolobus islandicus* M.14.25 (YP_002829883) *Beggiatoa* sp. PS (ZP_02002472)	48/108 (44%); 8×10^{-22} 27/67 (40%); 4×10^{-9}
gp38	ORF1070	20451–23663		Structural protein, terminal fibres; coiled-coil domains		Homologues in all rudiviruses; middle coiled-coil domains similar to the coiled-coil domains of Smc protein involved in chromosome segregation	*Streptococcus* phage 8140 (CBW39183) *Lactobacillus* phage c5 (ACA63308)	38/139 (27%); 8×10^{-6} 67/280 (24%); 9×10^{-6}
gp39	ORF564	23785–25488		Structural protein		Homologues in all rudiviruses; *Lipothrixviridae* (AFV9, gp54, YP_001798572; also AFV6, AFV3 and AFV8); more distant homologues in several archaea	'*Candidatus* Parvarchaeum acidiphilum' ARMAN-4 (EEZ92844)	111/287 (39%); 3×10^{-34}
gp40	ORF309	25495–26424						
gp41	ORF158a	26425–26901 (−)		SAM-dependent (RNA) methyltransferase	HHpred hit to 3k6r and 3kr9, $P=$ 99.6%	Highly conserved homologues in all rudiviruses and lipothrixviruses (e.g. SIFV, ORF65, NP_445728); more distant homologues in numerous archaea and bacteria	*Picrophilus torridus* DSM 9790 (YP_023074) *Sulfolobus solfataricus* P2 (NP_342327)	41/133 (31%); 4×10^{-5} 41/144 (28%); 8×10^{-5}
gp42	ORF269	26898–27707 (−)		TMD				
gp43	ORF176	27839–28369		GTase	HHpred hit to 1qg8 (spore coat polysaccharide biosynthesis), $P=$ 93.5%	*Lipothrixviridae* (e.g. AFV7, gp39, YP_001604263)		
gp44	ORF356	28451–29521		GTase, Group 1; C-terminal coiled-coil domain	CD hit to PHA01630 (PHA01630), $E = 1.7\times10^{-171}$; HHpred hits to 3fro, 3okp, 2gek, 3oy2, etc, $P=$ 100%; FFAS03 hit to 3oy2, score −56.8	*Lipothrixviridae* (e.g. SIFV ORF45), *Bicaudaviridae* (STSV1, ORF63), *Aeropyrum* coil-shaped virus (gp38), *Phycodnaviridae*	*Brevibacillus* sp. BC25 (ZP_10575973) *Aquifex aeolicus* VF5 (NP_213736)	54/208 (26%); 3×10^{-6} 72/326 (22%); 4×10^{-5}
gp45	ORF94	29522–29806 (−)	SIFV, ORF14, AFV1, ORF99			*Lipothrixviridae*, *Fuselloviridae*		
gp46	ORF95	29807–30094 (−)	SIFV, ORF14, AFV1, ORF99			*Lipothrixviridae*, *Fuselloviridae*		
gp47	ORF112	30244–30582 (−)						
gp48	ORF249	30775–31524 (−)				*Lipothrixviridae* (e.g. SIFV, ORF16, NP_445681)		
gp49	ORF98	31671–31967 (−)		Virion egress, VAP formation, TMD		STIV (C98, YP_024995)		

Table S1 | **Continued**

ORF	Name	Position (orientation)	*Ligamenvirales* structures	Predicted function	HHpred/FFAS03/CD hit	Homologues in other viruses and cellular organisms	BLAST hit	Identity; *E* value
gp50	ORF73	32071–32292 (–)	SIFV, ORF14, AFV1, ORF99			*Lipothrixviridae*, *Fuselloviridae*		
gp51	ORF108	32493–32819 (–)						
gp52	ORF105b	32956–33273 (–)		Coiled-coil domains				
gp53	ORF119b	33508–33867 (–)	SIFV, ORF14, AFV1, ORF99			*Lipothrixviridae*, *Fuselloviridae*		
gp54	ORF83b	34062–34313 (–)	SIRV1, ORF56a	DNA-binding, HTH				

Received 13 November 2012
doi:10.1042/BST20120313