HAL open science

# Diversity of Hyperthermophilic Archaeal Viruses

David Prangishvili, Mart Krupovic, Diana P. Baquero

▶ **To cite this version:**

David Prangishvili, Mart Krupovic, Diana P. Baquero. Diversity of Hyperthermophilic Archaeal Viruses. Bamford DH; Zuckerman M. Encyclopedia of Virology, 4th edition, 4, Elsevier, pp.359-367, 2021, 978-0-12-814516-6. 10.1016/B978-0-12-814515-9.00083-7 . pasteur-03698262

**HAL Id: pasteur-03698262**

**https://pasteur.hal.science/pasteur-03698262v1**

Submitted on 20 Jun 2022

**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 4.0 International License

Prangishvili D, Krupovic M, Baquero DP (2021) Diversity of hyperthermophilic archaeal viruses. In: Bamford DH, Zuckerman M (eds), *Encyclopedia of Virology*, 4th Edition, vol. 4, pp. 359-367. Oxford: Academic Press. doi: 10.1016/B978-0-12-814515-9.00083-7

# Diversity of Hyperthermophilic Archaeal Viruses

**David Prangishvili**

Archaeal Virology Unit, Department of Microbiology, Institut Pasteur, France
Ivane Javakhishvili Tbilisi State University, Tbilisi, Georgia
david.prangishvili@pasteur.fr

**Mart Krupovic**

Archaeal Virology Unit, Department of Microbiology, Institut Pasteur, France
mart.krupovic@pasteur.fr

**Diana Paola Baquero**

Archaeal Virology Unit, Department of Microbiology, Institut Pasteur, France
Sorbonne Universités, Collège Doctoral, Paris, 7505, France
dbaquero@pasteur.fr

## Abstract

Extreme thermal environments represent a habitat for viruses with remarkable morphological diversity and unique genome contents, many of which have not been described in moderate temperature environments. Most of these viruses infect hyperthermophilic archaea of the phylum Crenarchaeota and are currently classified into 13 families**.** Here, we summarize the state-of-the-art knowledge on the diversity and virion structures of crenarchaeal viruses. We then highlight the genomic features of hyperthermophilic archaeal viruses and emphasize how structural and functional genomic approaches have shed light into the characterization of viral proteins. We finally discuss the origins and evolutionary relationships between crenarchaeal viruses, underpinning their distinctiveness in the global virosphere.

## Keywords

Archaea, Crenarchaea, diversity, genomics, hyperthermophilic environments, morphologies, virion structures, viruses.

## Glossary

**A-form** One of the three major forms of double-stranded DNA, with a 23 Å helical diameter and 11 bp per helix turn.
**Acidophilic** Thriving under highly acidic conditions.
**Capsid** Protein shell that encloses the genetic material of the virus.
**Convergent evolution** Independent evolution of similar features in species of different lineages.
**Core genes** One or more genes strongly conserved at the nucleotide sequence level among a related group of genomes.
**Homologous recombination** Recombination between two identical or similar DNA sequences.
**Hyperthermophilic** Requiring extremely high temperatures for optimal growth.

**Inverted terminal repeats** Short, related or identical sequences located in reverse orientation at the ends of the viral genome.
**Jelly-roll fold** A protein fold composed of eight β-strands arranged in two antiparallel four stranded β-sheets.
**Pleomorphic viruses** Viruses with asymmetric or variable virion morphology.
**Protein glycosylation** Post-translational modification where a carbohydrate molecule is covalently bound to a predetermined region of a protein to form a glycoprotein.
**Proviruses** Viral genomes integrated into the host chromosome.
**Structural genomics** Description of the three-dimensional structure of a protein encoded by a given genome.
**Synteny** The shared ordering of genomic segments along a chromosome.
**Thermophilic** Requiring high temperatures for optimal growth.
**Viral envelope** Lipid layer present in many types of viruses that protects their genetic material.
**Virion** Infectious mature virus particle

**Nomenclature**

**Å** Angstrom
**ABV** Acidianus bottle-shaped virus
**ACV** Aeropyrum coil-shaped virus
**AFV1** Acidianus filamentous virus 1
**APBV1** Aeropyrum pernix bacilliform virus 1
**APOV1** Aeropyrum pernix ovoid virus 1
**ATV** Acidianus two-tailed virus
**bp** base pair
**cryo-EM** cryo-electron microscopy
**dsDNA** double-stranded DNA
**GDGT** glycerol dibiphytanyl glycerol tetraether
**ITRs** inverted terminal repeats
**kb** kilobase
**kDa** kilodalton
**MCP** major capsid protein
**nm** nanometer
**NMR** nuclear magnetic resonance
**nt** nucleotide
**ORF** open reading frame
**PCNA** proliferating cell nuclear antigen
**PFV1** Pyrobaculum filamentous virus 1
**PSV** Pyrobaculum spherical virus
**SEV1** Sulfolobus ellipsoid virus 1
**SIRV1** Sulfolobus islandicus rod-shaped virus 1
**SIRV2** Sulfolobus islandicus rod-shaped virus 2
**SNDV** Sulfolobus neozealandicus droplet-shaped virus
**SPV1** Sulfolobus polyhedral virus 1
**ssDNA** single-stranded DNA
**ssRNA** single-stranded RNA
**SSV1** Sulfolobus spindle-shaped virus 1
**STIV** Sulfolobus turreted icosahedral virus
**TSPV1** Thermoproteus spherical piliferous virus 1

## Introduction

One of the most surprising results of the studies on viral diversity on our planet is the observation of an astounding number of different viral morphologies in the habitats where large-scale

biodiversity is least expected - in geothermally heated environments where temperatures exceed 80°C and pH values are often below pH 3. Many virus morphologies observed here have never been detected in environments with less extreme conditions. Besides common filamentous and spherical virions, this diversity includes particles resembling bottles, droplets, coils and spindles, which can be tailless, tailed or two-tailed. Viruses with all these morphologies have been isolated from hot terrestrial springs of Europe, Asia, and North America. The hosts for this collection of viruses are archaea from the phylum Crenarchaeaota – members of the genera *Acidianus, Aeropyrum, Metallosphaera, Pyrobaculum, Stygiolobus, Sulfolobus* and *Thermoproteus*. All these hosts grow optimally at temperatures above 80°C and thus are referred to as hyperthermophiles. Viral infection occurs most efficiently at optimal temperatures of host growth and thus also the viruses are considered to be hyperthermophiles. Hyperthermophilic viruses are extremely thermostable in aggressive environmental conditions of their natural habitats, as well as in the laboratory conditions, e.g. thermal inactivation of some of them requires autoclaving at 121ºC at least for 40 min.

Based on their diverse morphological and genomic properties, the isolated and characterized viruses of Crenarchaeota are currently classified into 13 families (**Table 1**), 1 order *Ligamenvirales* and 1 proposed class *Tokiviricetes*. The genomes of members of all families, except *Spiraviridae*, have double-stranded (ds) DNA genomes, circular or linear (Table 1), whereas members of the *Spiraviridae* have circular single-stranded (ss) DNA genome.

<Table 1 near here>

## Morphology and structure

The families of crenarchaeal viruses have virions with diverse characteristic morphologies. The bottle-shaped, tailed spindle-shaped, tailless spindle-shaped, bacilliform, coil-shaped, droplet-shaped morphologies of members of the *Ampullaviridae, Bicaudaviridae, Fuselloviridae, Clavaviridae, Spiraviridae,* and *Guttaviridae,* correspondingly, are unprecedented among viruses of Bacteria and Eukarya, and represent Archaea-specific virion morphotypes. Specific for Archaea are also dsDNA viruses with filamentous virions, including members of the *Lipothrixviridae, Rudiviridae* and *Tristromaviridae*, as well as viruses with spherical virions carrying helical nucleoprotein core – members of *Globuloviridae, Ovaliviridae* and *Portogloboviridae*. The only family of crenarchaeal viruses with structural resemblance to viruses from other domains of life is *Turriviridae*. Turriviruses have non-enveloped icosahedral virions, with an internal lipid layer enclosing circular dsDNA genome. The major and minor capsid proteins have the double and single jelly-roll fold, respectively, and are related to those found in a wide range of bacterial and eukaryotic viruses. In other studied cases, capsid proteins of crenarchaeal viruses have unique structural folds.

Structures of virions from the families *Clavaviridae, Lipothrixviridae, Portogloboviridae, Rudiviridae,* and *Tristromaviridae,* were reconstructed at near atomic resolution using cryo-electron microscopy (cryo-EM). The results shed light on the mechanisms of virion morphogenesis and provide information on the molecular basis of high thermostability of the corresponding virions. Remarkably, in the reconstructed virions of members of the four latter families the packed dsDNA could be observed and was found to be in A-form: the phosphate-phosphate distance along DNA backbone is about 5.9 Å, as opposed to 7 Å for common B-form DNA. The common occurrence of A-DNA in hyperthermophilic viruses suggests that it may be the prevalent storage form of DNA in most extreme environments.

Most known viruses of Crenarchaeota are covered with envelopes. In those cases, where this was studied, the viruses form envelopes from lipids selectively acquired from the pool of host lipids.

### *Viruses with particular morphologies*

#### Family *Ampullaviridae* (from Latin *ampulla* for "bottle")

The enveloped virion of Acidianus bottle-shaped virus (ABV), the only isolated member of the family *Ampullaviridae*, resembles in its shape a champagne bottle (**Figure 1a**). It has an overall length

of 230 nm and a width varying from 75 nm at the broad end to 4 nm at the pointed end. The broad end of the virion is decorated with 20 thin rigid filaments, which appear to be inserted into a disc and interconnected at their bases. The cone-shaped inner core is formed by a toroidally supercoiled nucleoprotein filament 7 nm in width. It is presently unclear whether the pointed end or the filaments on the broader end are involved in adsorption to the cell surface and channelling of viral DNA into host cells. The virion contains six major proteins in the size range between 15 to 80 kDa.

<Figure 1 near here>

**Family *Bicaudaviridae* (from Latin *bi*, "two", and *cauda* for "tail")**

The virions of Acidianus two-tailed virus (ATV), the only currently classified member of the family *Bicaudaviridae*, are released from host cells as spindle-shaped particles with overall dimensions of approximately 120 x 300 nm (**Figure 1b**). Upon further incubation at temperatures above 75°C, appendages protrude from both pointed ends of the virion, and the lemon-shaped virion body shrinks to approximately 85 x 150 nm. The tails are heterogeneous in length, reaching 400 nm; the maximum length of the virion including tails reaches about 1000 nm. They have a tube-like structure and terminate with a narrow channel, which is 2 nm in width, and a terminal anchor-like structure formed by two furled filaments, each with a width of 4 nm. The virion contains at least eleven proteins with molecular masses in the range of 12-90 kDa.

Extracellular morphological development of the ATV virion takes place specifically at temperatures above 75 °C, close to that of the natural habitat, and does not require the presence of host cells, an exogenous energy sources or specific co-factors. However, the mechanism of development of the ATV tails remains unknown.

**Family *Clavaviridae* (from Latin *clava* for "club", "stick")**

Aeropyrum pernix bacilliform virus 1 (APBV1), the only known member of the family *Clavaviridae*, has non-enveloped bacilliform, rigid virions with dimensions of about 143 x 16 nm (**Figure 1c**). The terminal cap structures, one of which is pointed and the other rounded, most likely are involved in DNA packaging and recognition of the host cell.

The virions carry multiple copies of a single major capsid protein (MCP) of about 10 kDa and three minor capsid proteins which have molecular masses in the range of 9.5–21.5 kDa. The MCP consists of two α-helices linked with a β-hairpin, a structural fold not seen in the capsid proteins of other known viruses. The structure of APBV1 virion was determined by cryo-electron microscopy at near-atomic resolution. The structure reveals how the MCPs pack together forming a tubular structure: each MCP molecule makes extensive hydrophobic contacts to six other neighbouring subunits, forming very tight hydrophobic interface which apparently contributes to the virion stability at extremely high temperatures. The inner surface of the tubular structure is positively charged, allowing efficient interactions with the circular dsDNA genome and its packaging as a left-handed superhelix. The structure allowed to propose an assembly model where the dsDNA genome and the capsid assembly in a concerted fashion. According to this model, the virion assembly starts by binding of three specific sites of the circular dsDNA with one of the cap structures; this forms three loops, which gradually intertwine directed by protein assembly into the tubular structure; after all DNA has been covered by protein, the open end of the virion is sealed by another cap structure.

**Family *Fuselloviridae* (from Latin *fusello* for "little spindle")**

Virions of this family are enveloped and many have the shape of a lemon or spindle, with a bunch of thin filaments attached to one of the two pointed ends (**Figure 1d**). Other family members are more pleomorphic and elongated, with three relatively thick filaments at one pointed end. The terminal filaments most likely are involved in adsorption to the host cell surface. Virions have dimensions of approximately 60 x 100 nm. The envelope of Sulfolobus spindle-shaped virus 1 (SSV1) contains lipids and proteins VP1 and VP3. The circular dsDNA of SSV1 is positively supercoiled when isolated from virions.

Examination of the SSV1 virion using cryo-electron microscopy (cryo-EM) have provided only limited information on virion architecture, but the study of SSV1 virion assembly and egress using electron tomography was more informative. Both virion assembly and egress were found to be concomitant and occur at the cellular cytoplasmic membrane via a process highly reminiscent of the budding of enveloped viruses that infect eukaryotes, including human immunodeficiency virus, influenza virus, and Ebola virus.

**Family *Guttaviridae* (from Latin *gutta* for "droplet")**

The guttaviruses have slightly pleomorphic enveloped virions. Virions of Sulfolobus neozealandicus droplet-shaped virus (SNDV) resemble elongated droplets with varying dimensions, 110-185 nm in length and 95-70 nm in width (**Figure 1e**). The pointed end of the virion is covered by a beard of thick filaments. Virions of another member of the family, Aeropyrum pernix ovoid virus 1 (APOV1), are ovoid without detectable filamentous attachments.

**Family *Spiraviridae* (from Latin *spira* for "coil")**

The virions of Aeropyrum coil-shaped virus (ACV), the sole member of the family, are hollow, non-enveloped cylindrical particles, measuring about 230 x 20 nm. A short appendage of about 20 nm protrudes from each end of the cylindrical virion (**Figure 1f**). Exceptionally, the viral genome is a single-stranded, positive-sense DNA molecule. The virion carries two MCPs with molecular masses of about 23 and 18.5 kDa, and a few minor virion proteins. The observation of partially degraded virions enabled to propose the following model of virion morphogenesis: the circular ssDNA is covered by multiple copies of the MCP and the two halves of the circular nucleoprotein filament intertwine to form a rope-like structure, which is further condensed into a helix of higher order.

## *Spherical viruses*

**Family *Globuloviridae* (from Latin *globulus* for "small ball")**

The virions of globuloviruses are enveloped, spherical particles, 70-100 nm in diameter (**Figure 2a**). On the surface of the virion are multiple spherical protrusions about 15 nm in diameter. The viral envelope contains host-derived lipids and encases a tightly-packed superhelical nucleoprotein consisting of linear dsDNA and multiple copies of 33 kDa MCP. The virions carry minor capsid proteins. Virions of Thermoproteus spherical piliferous virus 1 (TSPV1) are decorated with numerous highly unusual filaments, which can extend hundreds of nanometers from the virion.

<Figure 2 near here>

**Family *Ovaliviridae* (from Latin *ovalis* for "oval")**

The sole known member of the *Ovaliviridae* family, Sulfolobus ellipsoid virus 1 (SEV1), has virions of ellipsoidal shape which measure about 115 x 80 nm. The virions are enveloped by a lipid containing membrane. The membrane encases circular nucleoprotein filament - consisting of circular dsDNA and DNA-binding MCPs, wrapped multiple times around the central axis of the virion (**Figure 2b**).

**Family *Portogloboviridae* (from Latin *porto* for "to bear", and *globus* for "ball")**

The virions of Sulfolobus polyhedral virus 1 (SPV1) are icosahedral, about 90 nm in diameter (**Figure 2c**). The cryo-EM reconstruction of the virion revealed structural details at near-atomic resolution. The icosahedral capsid is formed by 2 types of MCPs both of which carry variants of the single jelly-roll fold and are arranged in an atypical manner. The protein capsid encloses an internal lipid-containing membrane, which, in turn, encloses a spherical core formed by a circular nucleoprotein

filament consisting of dsDNA and DNA-binding MCP. In the circular nucleoprotein, the dsDNA is complexed by a dimeric DNA-binding MCP and is in A-form. The condensation of the circular nucleoprotein into a spherical core is suggested to occur in two steps: folding of circular nucleoprotein into a raft which is then spooled into concentric shells.

**Family *Turriviridae* (from Latin *turris* for "tower")**

The overall morphology of non-enveloped icosahedral virions of the *Turriviridae* is highly similar to that of bacterial viruses from the families *Tectiviridae* and *Corticoviridae:* the virions are icosahedral with an inner lipid layer and pack naked dsDNA with the help of virus-encoded ATP-dependent molecular machinery (**Figure 2d**). Moreover, members of all three families share the structure of MCPs which exhibit the double-jelly roll fold. Cryo-EM reconstruction of Sulfolobus turreted icosahedral virus (STIV) revealed unique structural features, including the elaborate turret-like structures at the fivefold vertices and the unusual pseudo-T=31 icosahedral lattice on which the virion is built.

## *Filamentous viruses*

**Proposed class *Tokiviricetes*:**
**order *Ligamenvirales*: family *Rudiviridae* (from the Latin *rudis* for "small rod")**
**order *Ligamenvirales*: family *Lipothrixviridae* (from the Greek *lipos* for "fat" and *thrix* for "hair")**
**family *Tristromaviridae* (from the Greek *tria* for "three" and *stroma* for "layer")**

<Figure 3 near here>

The virions of members of the three families of the proposed class *Tokiviricetes* are filamentous, about 23 nm in width and the lengths in the range of 400-2000 nm, depending on the size of the linear dsDNA genomes and parameters of the nucleoprotein helix (**Figure 3**). At both ends, the virions carry identical terminal structures. In the case of Acidianus filamentous virus 1 (AFV1), a member of the *Lipothrixviridae*, the terminal structure resembles claws and folds upon interaction with the appendages of the host cells (**Figure 3a**). The virions of Sulfolobus islandicus rod-shaped virus 2 (SIRV2), the type member of the *Rudiviridae* family, are decorated with three thin terminal fibers, whereas virions of Pyrobaculum filamentous virus 1 (PFV1), a member of the *Tristromaviridae*, possess bundles of filaments that can reach a length of 80 nm (**Figures 3b and 3c)**.

The virion structures of members of the three families have been reconstructed by cryo-EM at near atomic resolution and revealed similar, previously unknown form of virion organization. In all three families the core of the filamentous virion is a tube-like nucleoprotein formed by condensation of linear dsDNA by dimers of the MCPs – homodimers in case of *Rudiviridae* and heterodimers in the case of the other two families (**Figure 4a**). As a result of binding to DNA, the MCPs form a helix-turn-helix structure and tightly wrap around dsDNA, in a manner not observed in known viruses, and transforms it into A-form (**Figure 4b**). Such arrangement keeps DNA inaccessible for solutes and ensures virion stability in highly aggressive environmental conditions. The MCPs have a molecular mass of about 14.5 kDa. *Rudiviridae* encode a single MCP, whereas each of the two other families encodes two paralogous MCPs. Remarkably, although the MCPs from the three virus families do not share significant sequence similarity, their structures are highly similar – they all carry unique four-helix bundle fold, not seen in the capsid proteins of other known viruses. The structural similarity of the MCPs suggests common origin of the MCPs and provides arguments to postulate evolutionary relationship between the three virus families which employ these MCPs. Thus, it was proposed to unify the families *Rudiviridae*, *Lipothrixviridae*, and *Tristromaviridae* into the new virus class *Tokiviricetes*. The former two families reveal certain similarity also at the genomic level, suggesting closer relationship. Thus, *Rudiviridae* and *Lipothrixviridae* were unified also in a taxon of a lower rank, the order *Ligamenvirales.*

<Figure 4 near here>

The major difference in virion architecture of the three families concerns coating of the nucleoprotein core. In the *Rudiviridae*, it is non-enveloped, in *Lipothrixviridae*, covered with lipid envelope, and in T*ristromaviridae*, covered by a protein matrix which mediates contact with the outer lipid envelope. The matrix layer of the *Tristromaviridae* is formed by an 18 kDa capsid protein, VP3. As in other crenarchaeal viruses, *Lipothrixviridae* and *Tristromaviridae* recruit the lipids for formation of the envelope from the pool of host lipids.

The thickness of the lipid envelope in the reconstructed virions of the lipothrixviruses is about 20–25 Å, only half of the thickness of the cellular membrane. The host membrane, as in all Crenarchaeota, represents a 40 Å monolayer of glycerol dibiphytanyl glycerol tetraether (GDGT) lipids with different numbers of cyclopentane rings. Studies on the composition and structure of the virion envelope showed that the virus selectively acquires more flexible GDGT lipids lacking cyclopentane rings, which can be bent into U-shaped, 'horseshoe' conformation, forming the 20 Å thick virion envelope. Apparently, the membrane with lipids in such conformation, which was previously never observed in the living world, has advantages for survival in aggressive conditions of the natural environment of the hyperthermophiles.

### *General comments*

Summarizing the results on studies on morphology and morphogenesis of viruses of hyperthermophilic Crenarchaeaota, it could be noted that packaging of viral genomes in the form of a nucleocapsid appears to be the major structural characteristic of crenarchaeal viruses. Based on the available information, the assembly of the virion core apparently occurs through concerted coating of the DNA with MCPs and spontaneous self-assembly of such nucleocapsids into minimum free energy structures of different shapes. The self-assembly pathway of virion morphogenesis is common for RNA viruses but not for dsDNA viruses, where capsid assembly and genome packaging are generally separated. dsDNA viruses of Bacteria, herpseviruses of Eukarya, and crenarchaeal *Turriviridae* pack dsDNA in the naked state into the pre-assemble empty capsids. This process requires significant energy expenses — mainly because of the mutual repulsion of the negatively charged phosphate backbone of the DNA — and is facilitated by virus-encoded elaborate molecular machinery. Considering the complexity of such process, it is surprising that the pathways of virion morphogenesis which are common for crenarchaeal viruses are so rare in the world of dsDNA viruses.

In most crenarchaeal virus families, the nucleoprotein core is covered by lipid-containing envelopes. Moreover, the core of the *Portogloboviridae* is encased by an icosahedral protein shell. The mechanisms of coating of the nucleoprotein cores remain unknown. For *Fuselloviridae*, the final step of virion assembly was shown to occur in the course of budding from the host cell in a process, resembling the morphogenesis of enveloped viruses of eukaryotes. For other crenarchaeal viruses, the coating of virions with a lipid envelope takes place apparently intracellularly.

## Genomes

The genomes of crenachaeal viruses are generally small in size (**Table 1**), with the largest ones (60-70 kb) being found in members of the family *Bicaudaviridae*. The 5,278 bp-long genome of the clavavirus APBV1 is among the smallest known dsDNA genomes. The other record holder is the coil-shaped spiravirus ACV – its ssDNA genome with 24.9 kb is the largest among known ssDNA genomes. The genomes can be either circular or linear (**Table 1**). Viruses with linear genomes often carry inverted terminal repeats (ITRs) of different length and exploit different strategies to protect their genome termini, such as covalently closed hairpins or covalently attached terminal proteins.

Along with the extraordinary diversity of morphotypes, another outstanding feature of archaeal viruses is the uniqueness of their genome content, with a very low proportion of genes with recognizable homologues in public databases. Indeed, family-specific comparison of viral proteomes against the sequence databases revealed that ~85% of the crenarchaeal virus proteins do not have identifiable homologs when an E-value threshold of <1e-5 is used.

Functional annotation of crenarchaeal proteins shows that very few of these proteins are homologous to any sequences in the public databases, be it proteins of other viruses or those of cellular organisms. Consequently, archaeal virus genomes remain a rich source of unknown genes, many of which could be responsible for unique mechanisms of virus-host interactions or possess unexpected properties.

Structural genomics have been performed to promote the functional characterization of *Fuselloviridae*, *Bicaudaviridae*, *Rudiviridae*, *Lipothrixviridae*, *Globuloviridae* and *Turriviridae* proteins. Protein structures were determined by X-ray crystallography, and nuclear magnetic resonance (NMR) spectroscopy. For some viral proteins, the structures displayed unique folds with no homologs in databases, hindering the assignment of putative functions. In few cases, the structural information suggested a function which was verified biochemically. This was the case, e.g. for a novel type of nuclease encoded by the lipothrixvirus AFV1 and replication initiator protein encoded by the rudivirus SIRV2.

Notably, none of the viruses encodes a recognizable DNA-dependent RNA polymerase and only members of two families – *Ampullaviridae* and *Ovaliviridae*, encode recognizable DNA-dependent DNA polymerases. However, crenarchaeal viruses commonly encode multiple transcription factors which may recruit and redirect the host transcription machinery to preferential expression of viral genes.

Comparative genomic analyses have revealed that hyperthermophilic viruses from different families share just a small group of common genes, suggesting independent origins of the distinct groups of archaeal viruses. Ten virus families each have only one or two members with the sequenced genome. Exceptions are *Fuselloviridae* and two families of the order *Ligamenvirales* with higher number of members with sequenced genomes. Comparison of these genomes facilitated functional predictions and contributed to understanding the evolutionary history of the corresponding virus families, and will be briefly detailed below.

Family *Fuselloviridae*

The genomes of all members of *Fuselloviridae* are highly similar at the nucleotide sequence level and display overall gene synteny. A high frequency of homologous recombination has been reported between family members. The majority of predicted ORFs of the fuselloviruses cannot be assigned a function based on homology with sequences in public databases. Comparison of genome sequences has revealed a set of 12 not contiguous 'core' genes shared by all family members, suggesting a common evolutionary history despite differences in the geographical context. The 'core' proteome includes structural proteins, predicted DnaA-like AAA+ ATPase, transcriptional regulators and a tyrosine superfamily integrase, which is involved in the integration of the viral genome into the host tRNA gene. Accordingly, the infected cell cultures contain both the covalently closed circular form of the viral genome and the integrated provirus. Unlike in the case of temperate bacteriophages, the integrase gene of the fuselloviruses is partitioned into two fragments upon the viral genome integration. Notably, studies on the prototypical fusellovirus SSV1 have shown that SSV1 integrase is not essential for the viral cycle. Moreover, nearly half of the predicted SSV1 ORFs were shown to tolerate insertions or deletions without affecting the viral infectivity.

Order *Ligamenvirales*

All members of the order have linear dsDNA genomes. On the example of the genome of the rudivirus SIRV1 it was shown that the two strands of the linear dsDNA genomes are covalently linked at both ends, forming a continuous polynucleotide chain and producing terminal hairpin structures. For other ligamenviruses such detailed analysis of the terminal regions has not been performed. Rudiviruses carry terminal inverted repeats which differ in size and sequence. The longest with 2 kb is found in the rudivirus SIRV1. In both rudiviruses and lipothrixviruses active genome remodelling, involving both deletions and horizontal acquisition of new genes, has been documented. The comparison of the genomes of different members of the same family and different strains of the same species revealed

multiple examples of genomic rearrangements caused by insertion/deletion of sequences that do not disrupt the ORFs. In *Rudiviridae* and *Lipothrixviridae* such sequences were shown to carry 12 bp or multiples thereof. It was suggested that the latter might constitute genetic elements mobilized by archaeal intron splicing.

Members of the *Rudiviridae* share significant similarity in the genome sequence and synteny, and carry a set of conserved core genes, most of which are localized in the middle region of the linear genome. By contrast, the genomes of *Lipothrixviridae* display a considerable variation in gene order and content, suggesting a longer evolutionary history. Based on the sequence similarity, the genomes of *Lipothrixviridae* group into four clusters of related species which form the four genera in the family. The examples of intergenomic recombination have been observed between members of different genera.

Properties of about half of the proteins encoded by the rudiviruses SIRV1 and SIRV2 have been predicted as a result of analysis of their sequences, structures and biochemical characteristics. Such proportion of recognized gene functions is among the highest for crenarchaeal viruses. The proteins with predicted functions include transcription regulators with ribbon-helix-helix and helix-turn-helix motifs, glycosyltransferases, acetyltransferase, Holliday junction resolvase, methyltransferase, dUTPase, ssDNA-binding protein, ssDNA annealing ATPase, Cas4-like exonuclease, as well as a protein specifically interacting with the host proliferating cell nuclear antigen (PCNA) and presumably recruiting it for the replication of the viral genome. Moreover, three proteins of the virus SIRV2 have been identified as anti-CRISPR proteins, which inhibit type I and type III CRISPR-Cas immunity systems of the *Sulfolobus* host. Detailed studies on the life cycle of the rudivirus SIRV2 enabled identification of the viral protein involved in the unique mechanism of virion egress. The function of this protein could not be predicted based on its sequence and was verified only as a results of detailed studies on the virus life cycle, supporting the possibility that many genes of archaeal viruses are implicated in specific aspects of virus-host interactions.

## Evolutionary relationships

The unique morphological and genomic features of crenarchaeal DNA viruses raise important questions regarding their origins. The analysis of the evolutionary relationships between all dsDNA viruses using the bipartite network approach, which traces connections between viral genomes through shared gene families, revealed that crenarchaeal viruses are largely disconnected from the global dsDNA virosphere. The families of crenarchaeal viruses are themselves largely disconnected from each other and share just a small group of common genes, suggesting that most families of crenarchaeal viruses have evolved independently of one another.

How and when did archaea-specific viruses originate? Why do they only infect archaea? There are at least two non-mutually exclusive explanations. Some of the archaea-specific virus groups could have emerged during the early stages of cellular evolution and been retained in the Archaea but lost in the domains Bacteria and Eukarya. Other archaeal virus groups could have evolved concomitantly with the Archaea or, even more recently, within specific archaeal lineages. The observed limited gene sharing between different groups of archaea-specific viruses seems to make the latter possibility particularly plausible.

## Further Reading

Arnold, H.P., Ziese, U., and Zillig, W. (2000a). SNDV, a novel virus of the extremely thermophilic and acidophilic archaeon Sulfolobus. Virology *272*, 409-416.

Bettstetter, M., Peng, X., Garrett, R.A., and Prangishvili, D. (2003). AFV1, a novel virus infecting hyperthermophilic archaea of the genus acidianus. Virology *315*, 68-79.

Contursi, P., Fusco, S., Cannio, R., and She, Q. (2014). Molecular biology of fuselloviruses and their satellites. Extremophiles : life under extreme conditions *18*, 473-489.

DiMaio, F., Yu, X., Rensen, E., Krupovic, M., Prangishvili, D., and Egelman, E.H. (2015). Virology. A virus that infects a hyperthermophile encapsidates A-form DNA. Science (New York, NY) *348*, 914-917.

Häring, M., Peng, X., Brügger, K., Rachel, R., Stetter, K.O., Garrett, R.A., and Prangishvili, D. (2004). Morphology and genome organization of the virus PSV of the hyperthermophilic archaeal genera Pyrobaculum and Thermoproteus: a novel virus family, the Globuloviridae. Virology *323*, 233-242.

Häring, M., Rachel, R., Peng, X., Garrett, R.A., and Prangishvili, D. (2005a). Viral diversity in hot springs of Pozzuoli, Italy, and characterization of a unique archaeal virus, Acidianus bottle-shaped virus, from a new family, the Ampullaviridae. Journal of virology *79*, 9904-9911.

Iranzo, J., Koonin, E.V., Prangishvili, D., and Krupovic, M. (2016a). Bipartite Network Analysis of the Archaeal Virosphere: Evolutionary Connections between Viruses and Capsidless Mobile Elements. Journal of virology *90*, 11043-11055.

Kasson, P., DiMaio, F., Yu, X., Lucas-Staat, S., Krupovic, M., Schouten, S., Prangishvili, D., and Egelman, E.H. (2017). Model for a novel membrane envelope in a filamentous hyperthermophilic virus. eLife *6*.

Krupovic, M., Cvirkaite-Krupovic, V., Iranzo, J., Prangishvili, D., and Koonin, E.V. (2018). Viruses of archaea: Structural, functional, environmental and evolutionary genomics. Virus research *244*, 181-193.

Liu, Y., Ishino, S., Ishino, Y., Pehau-Arnaudet, G., Krupovic, M., and Prangishvili, D. (2017). A Novel Type of Polyhedral Viruses Infecting Hyperthermophilic Archaea. Journal of virology *91*.

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. Proceedings of the National Academy of Sciences of the United States of America *109*, 13386-13391.

Mochizuki, T., Sako, Y., and Prangishvili, D. (2011). Provirus induction in hyperthermophilic archaea: characterization of Aeropyrum pernix spindle-shaped virus 1 and Aeropyrum pernix ovoid virus 1. Journal of bacteriology *193*, 5412-5419.

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.

Prangishvili, D., Bamford, D.H., Forterre, P., Iranzo, J., Koonin, E.V., and Krupovic, M. (2017). The enigmatic archaeal virosphere. Nature reviews Microbiology *15*, 724-739.

Prangishvili, D., Koonin, E.V., and Krupovic, M. (2013). Genomics and biology of Rudiviruses, a model for the study of virus-host interactions in Archaea. Biochemical Society transactions *41*, 443-450.

Prangishvili, D., and Krupovic, M. (2012). A new proposed taxon for double-stranded DNA viruses, the order "Ligamenvirales". Archives of virology *157*, 791-795.

Prangishvili, D., Vestergaard, G., Häring, M., Aramayo, R., Basta, T., Rachel, R., and Garrett, R.A. (2006). Structural and genomic properties of the hyperthermophilic archaeal virus ATV with an extracellular stage of the reproductive cycle. Journal of molecular biology *359*, 1203-1216.

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.

Quemin, E.R., Chlanda, P., Sachse, M., Forterre, P., Prangishvili, D., and Krupovic, M. (2016). Eukaryotic-Like Virus Budding in Archaea. mBio *7*.

Rensen, E.I., Mochizuki, T., Quemin, E., Schouten, S., Krupovic, M., and Prangishvili, D. (2016). A virus of hyperthermophilic archaea with a unique architecture among DNA viruses. Proceedings of the National Academy of Sciences of the United States of America *113*, 2478-2483.

Wang, F., Baquero, D.P., Su, Z., Osinski, T., Prangishvili, D., Egelman, E.H., and Krupovic, M. (2020). Structure of a filamentous virus uncovers familial ties within the archaeal virosphere. Virus evolution *6*, veaa023.

Wang, F., Liu, Y., Su, Z., Osinski, T., de Oliveira, G.A.P., Conway, J.F., Schouten, S., Krupovic, M., Prangishvili, D., and Egelman, E.H. (2019). A packing for A-form DNA in an icosahedral virus. Proceedings of the National Academy of Sciences of the United States of America *116*, 22591-22597.

Wang, H., Guo, Z., Feng, H., Chen, Y., Chen, X., Li, Z., Hernández-Ascencio, W., Dai, X., Zhang, Z., Zheng, X., *et al.* (2018). Novel Sulfolobus Virus with an Exceptional Capsid Architecture. Journal of virology *92*.

**Table 1.** Viruses of Crenarchaeota

| **Family name** | **Virion shape** | **Example of species** | **Envelope** | **Genome type and size (bp)** | **Accesion number** |
|---|---|---|---|---|---|
| Viruses with particular shapes | | | | | |
| *Ampullaviridae* | Bottle-shaped, with fibers at the blunt end | Acidianus bottle-shaped virus (ABV) | External | Linear dsDNA, 23,900 | EF432053 |
| *Bicaudaviridae* | Spindle-shaped with two appendages | Acidianus two-tailed virus (ATV) | None | Circular dsDNA, 62,730 | AJ888457 |
| *Clavaviridae* | Bacilliform | Aeropyrum pernix bacilliform (APBV1) | None | Circular dsDNA, 5,278 | AB537968 |
| *Fuselloviridae* | Spindle-shaped with fibers at one end | Sulfolobus spindle-shaped virus 1 (SSV1) | External | Circular dsDNA, 15,465 | XO7234 |
| *Guttaviridae* | Droplet/ovoid-shaped | Aeropyrum pernix ovoid virus 1 (APOV1) | External | Circular dsDNA, 13,769 | NC_028256 |
| *Spiraviridae* | Coils with two identical terminal appendages | Aeropyrum coil-shaped virus (ACV) | None | Linear dsDNA, 24,893 | HE681887 |
| Spherical viruses | | | | | |
| *Globuloviridae* | Spherical | Pyrobaculum spherical virus (PSV) | External | Linear dsDNA, 28,337 | AJ635162 |
| *Ovaliviridae* | Ellipsoid | Sulfolobus ellipsoid virus 1 (SEV1) | External | Circular dsDNA, 23,219 | MF144115 |
| *Portogloboviridae* | Icosahedron | Sulfolobus polyhedral virus 1 (SPV1) | Internal | Circular dsDNA, 20,222 | KY927925 |
| *Turriviridae* | Icosahedron with 12 turrets | Sulfolobus turreted icosahedral virus (STIV) | Internal | Circular dsDNA, 17,663 | AY569307 |
| Filamentous viruses | | | | | |
| *Lipothrixviridae* | Flexible filament with terminal structures | Acidianus filamentous virus 1 (AFV1) | External | Linear dsDNA, 21,080 | AJ567472 |
| *Rudiviridae* | Rigid rod, with terminal fibers | Sulfolobus islandicus rod-shaped virus 2 (SIRV2) | None | Linear dsDNA, 35,450 | AJ344259 |
| *Tristromaviridae* | Filamentous, with terminal fibers | Pyrobaculum filamentous virus 1 (PFV1) | External | Linear dsDNA, 17,714 | KU307456 |

## Figures

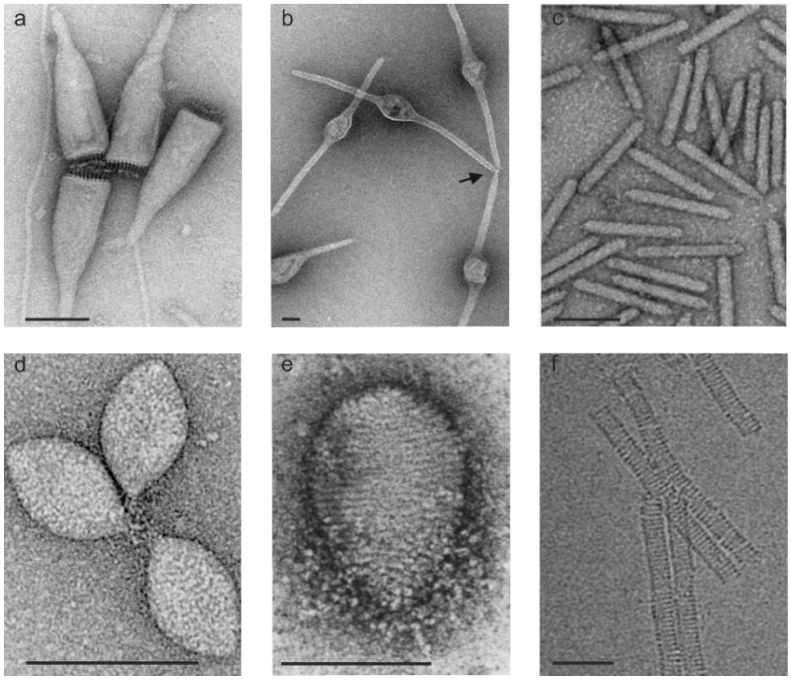


**Figure 1.** Electron micrographs of archaeal viruses with particular morphologies. (a) ABV, Acidianus bottle-shaped virus; (b) ATV, Acidianus two-tailed virus, the arrow indicates virion tails which underwent extracellular development; (c) APBV1, Aeropyrum pernix bacilliform virus 1; (d) SSV1, Sulfolobus spindle-shaped virus 1; (e) SNDV, Sulfolobus neozealandicus droplet-shaped virus, and (f) ACV, Aeropyrum coil-shaped virus. Negative stain with uranyl acetate, except for ACV, which is in vitreous ice. Bars, 100 nm. Image modified from Prangishvili, D., Bamford, D.H., Forterre, P., Iranzo, J., Koonin, E.V., and Krupovic, M. (2017). The enigmatic archaeal virosphere. Nature reviews Microbiology *15*, 724-739.

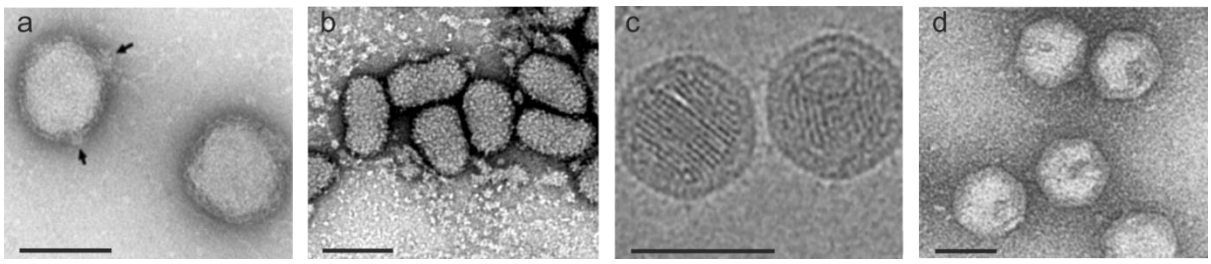


**Figure 2.** Electron micrographs of archaeal viruses with spherical morphologies. (a) PSV1, Pyrobaculum spherical virus 1, the arrows indicate spherical protrusions; (b) SEV1, Sulfolobus ellipsoid virus 1; (c) SPV1, Sulfolobus polyhedral virus 1; and (d) STIV, Sulfolobus turreted icosahedral virus. Negative stain with uranyl acetate. Bars, 100 nm. Image modified from Prangishvili, D., Bamford, D.H., Forterre, P., Iranzo, J., Koonin, E.V., and Krupovic, M. (2017). The enigmatic archaeal virosphere. Nature reviews Microbiology 15, 724-739. SEV1 image is courtesy of L. Huang.

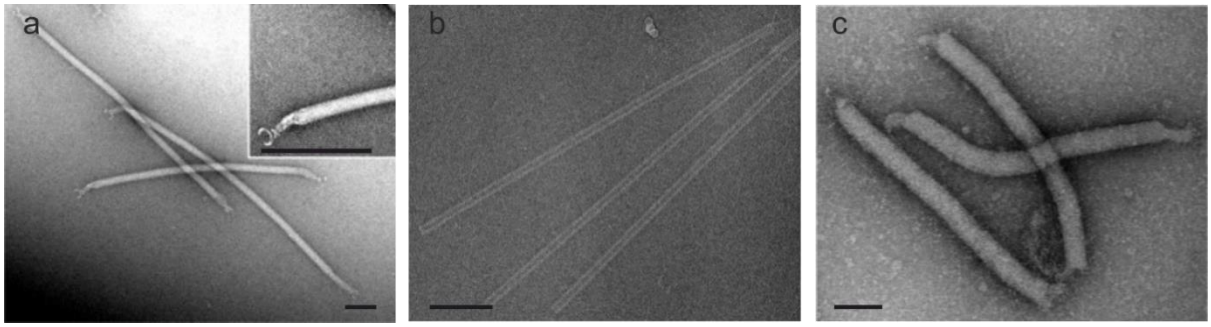


**Figure 3.** Electron micrographs of archaeal viruses with filamentous morphologies. (a) AFV1, Acidianus filamentous virus 1, the inset displays the terminal structure; (b) SIRV2, Sulfolobus islandicus rod-shaped virus 2; and (c) PFV1, Pyrobaculum filamentous virus 1. Negative stain with uranyl acetate, except for SIRV2, which is in vitreous ice. Bars, 100 nm. Image modified from Prangishvili, D., Bamford, D.H., Forterre, P., Iranzo, J., Koonin, E.V., and Krupovic, M. (2017). The enigmatic archaeal virosphere. Nature reviews Microbiology *15*, 724-739.

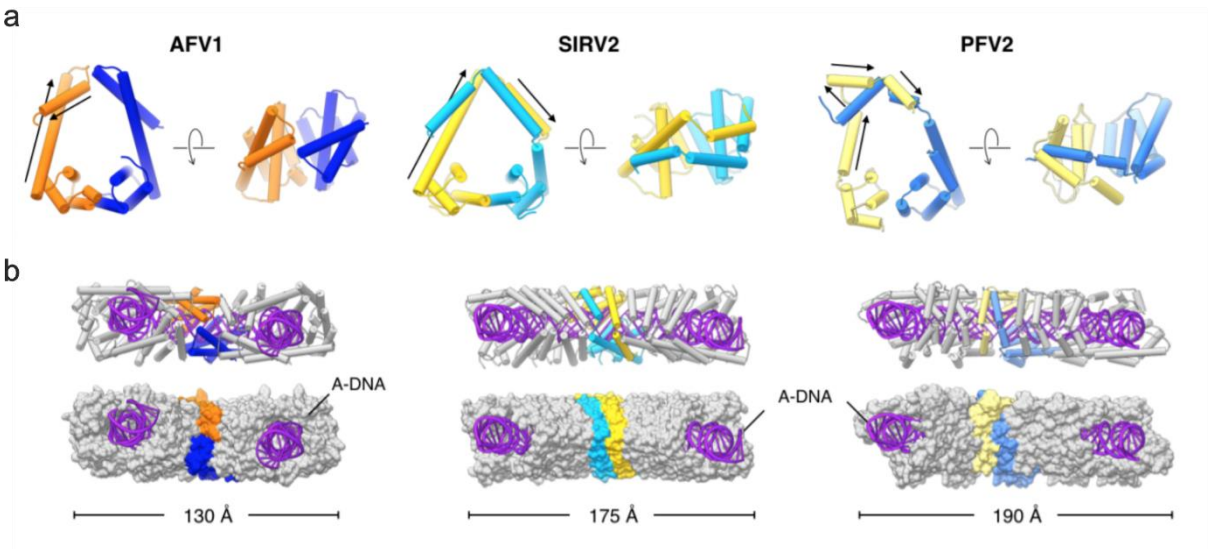


**Figure 4.** Virion organization of the filamentous viruses AFV1, SIRV2 and PFV2. (a) Comparison of the MCP dimer (asymmetric unit) of AFV1 (left), SIRV2 (centre) and PFV2 (right). The MCP1 of AFV1, SIRV2 and PFV2 are colored in orange, gold and yellow, respectively. The MCP2 of AFV1, SIRV2 and PFV2 are coloured in blue, cyan and light blue, respectively. The N-terminal helices of MCP1 in AFV1, SIRV2 and PFV2 are marked with black arrows. (b) Wrapping of A-form DNA in AFV1, SIRV2 and PFV2. One MCP dimer is colored as in (a), other four dimers are coloured in gray. Proteins are shown in ribbon representation (top) and as surfaces (bottom). Image reproduced from Wang, F., Baquero, D.P., Su, Z., Osinski, T., Prangishvili, D., Egelman, E.H., and Krupovic, M. (2020). Structure of a filamentous virus uncovers familial ties within the archaeal virosphere. Virus evolution *6*, veaa023.