Isolation and characterization of hyperthermophilic archaeal viruses

Diana Paola Baquero Uriza

Sorbonne Université
École doctorale Complexité du Vivant – ED515
7, Quai Saint-Bernard, case 32
75252 Paris cedex 05, France

Unité Virologie des Archées
Département de Microbiologie – Institut Pasteur
25, rue du Dr. Roux
75015, Paris

THESE DE DOCTORAT DE LA SORBONNE UNIVERSITE

Spécialité : *Microbiologie*

Présentée par

M. Diana Paola Baquero Uriza

Pour obtenir le grade de

DOCTEUR DE LA SORBONNE UNIVERSITE

ISOLATION AND CHARACTERIZATION OF HYPERTHERMOPHILIC ARCHAEAL VIRUSES

Soutenue le 04 Décembre 2020 devant le jury composé de :

Prof. Guennadi Sezonov	*Président du jury*
Dr. Claire Geslin	*Rapporteur de thèse*
Dr. Pascale Boulanger	*Rapporteur de thèse*
Dr. Mart Krupovic	*Directeur de thèse*
Dr. David Prangishvili	*Superviseur de thèse*

Sorbonne Université
École doctorale Complexité du Vivant – ED515
7, Quai Saint-Bernard, case 32
75252 Paris cedex 05, France

Unité Virologie des Archées
Département de Microbiologie – Institut Pasteur
25, rue du Dr. Roux
75015, Paris

THESE DE DOCTORAT DE LA SORBONNE UNIVERSITE

Spécialité : *Microbiologie*

Présentée par

M. Diana Paola Baquero Uriza

Pour obtenir le grade de

DOCTEUR DE LA SORBONNE UNIVERSITE

ISOLATION AND CHARACTERIZATION OF HYPERTHERMOPHILIC ARCHAEAL VIRUSES

Soutenue le 04 Décembre 2020 devant le jury composé de :

Prof. Guennadi Sezonov	*Président du jury*
Dr. Claire Geslin	*Rapperteur de thèse*
Dr. Pascale Boulanger	*Rapporteur de thèse*
Dr. Mart Krupovic	*Directeur de thèse*
Dr. David Prangishvili	*Superviseur de thèse*

Isolation and characterization of hyperthermophilic archaeal viruses

A mis padres,

a todos quienes me

han inspirado en este camino.

Isolation and characterization of hyperthermophilic archaeal viruses

1. RESUME

Les environnements géothermiques extrêmes représentent un habitat pour les virus d'archées qui ont un contenu génomique unique et des morphologies remarquables, dont la plupart n'ont pas été décrites chez les virus infectent les bactéries et les eucaryotes. Cependant, notre connaissance de ces virus reste insuffisante. En effet, le nombre d'espèces de virus connu infectant les archées est faible par rapport aux virus eucaryotes ou bactériens connus. De plus, l'absence de relations avec d'autres virus connus et le caractère distinctif de leurs génomes suggèrent que les mécanismes d'interaction virus-hôte sont également susceptibles d'être nouveaux. Par conséquent, dans le cadre de mes études, je me suis concentré sur deux axes de recherche: (i) l'isolation de nouveaux virus d'archées hyperthermophiles et (ii) la caractérisation des mécanismes d'assemblage et de libération des virions chez les archées.

Nous avons étudié la diversité des virus dans les champs sulfureux du volcan Campi Flegrei à Pozzuoli, en Italie. Cinq nouveaux virus infectant des hyperthermophiles neutrophiles du genre *Pyrobaculum* et des acidophiles hyperthermophiles appartenant à trois genres différents de l'ordre des Sulfolobales, à savoir, *Saccharolobus*, *Acidianus* et *Metallosphaera*, ont été isolés. Les virus caractérisés appartiennent aux familles *Rudiviridae*, *Globuloviridae* et *Tristromaviridae*. Remarquablement, l'analyse phylogénomique des rudivirus nouvellement isolés et précédemment séquencés a révélé un schéma biogéographique clair, dans lequel tous les rudivirus italiens forment un clade monophylétique, suggérant une structuration géographique des communautés virales dans les environnements géothermiques extrêmes.

L'un des virus filamenteux non enveloppés isolés à Pozzuoli, à savoir Saccharolobus solfataricus rod-shaped virus 1 (SSRV1), et le virus enveloppé Sulfolobus islandicus filamentous virus (SIFV) ont été caractérisés structurellement et biochimiquement. L'étude a permis de révéler des caractéristiques structurelles conservées dans les virus d'archées filamenteux et de clarifié la relation évolutive entre les virus filamenteux non enveloppés et enveloppés.

Le deuxième axe de recherche s'est orienté autour de la compréhension des mécanismes d'assemblage et de libération du virus filamenteux enveloppé SIFV. Celui-ci infecte l'archéon hyperthermophile et acidophile *Sulfolobus islandicus*. Nos résultats montrent que SIFV est

lytique et libéré par des portails pyramidaux formés dans la membrane de la cellule hôte, ce qui est inattendu pour des virus enveloppés. La tomographie électronique à deux axes a révélé que les virions de SIFV acquièrent leurs enveloppes à l'intérieur de l'hôte, soit par la formation de membrane *de novo*, soit par le trafic de lipides de la membrane cytoplasmique vers les centres d'assemblage des virions.

Collectivement, nos résultats montrent que les systèmes hydrothermaux continentaux à haute température abritent un virome très diversifié et nous éclairent sur l'évolution des virus d’archées. De plus, nos données contribuent aux connaissances sur les interactions virus-hôte dans les environnements géothermiques extrêmes.

Mots clés

Archées, assemblage, diversité, hyperthermophiles, interactions virus-hôte, libération, structures de virions, structures pyramidales, virus d’archées, virus filamenteux.

2. ABSTRACT

Extreme geothermal environments are inhabited by archaeal viruses with unique genome contents and remarkable morphologies, many of which have not been described among viruses infecting bacteria or eukaryotes. However, the number of known species of viruses infecting archaea remains low compared to the eukaryotic or bacterial viruses. Moreover, the lack of relationships to other known viruses and distinctiveness of their genomes suggest that the mechanisms of virus-host interaction are likely to be also novel. Therefore, in the framework of my studies, I have focused on two major lines of research: (i) isolation and characterization of new hyperthermophilic archaeal viruses and (ii) molecular mechanisms of virus-host interactions in Archaea.

We investigated the virus diversity in the sulfurous fields of the Campi Flegrei volcano in Pozzuoli, Italy. Five new archaeal viruses infecting neutrophilic hyperthermophiles of the genus *Pyrobaculum* and acidophilic hyperthermophiles belonging to three different genera of the order Sulfolobales, namely, *Saccharolobus*, *Acidianus*, and *Metallosphaera* were isolated. The newly characterized viruses belong to the families *Rudiviridae*, *Globuloviridae* and *Tristromaviridae.* Notably, phylogenomic analysis of the newly isolated and previously sequenced rudiviruses revealed a clear biogeographic pattern, with all Italian rudiviruses forming a monophyletic clade, suggesting geographical structuring of virus communities in extreme geothermal environments.

One of the non-enveloped rudiviruses isolated in Pozzuoli, namely, Saccharolobus solfataricus rod-shaped virus 1 (SSRV1), and the enveloped Sulfolobus islandicus filamentous virus (SIFV) were structurally and biochemically characterized. The study has revealed conserved structural features shared by these viruses and clarified the evolutionary relationship between non-enveloped and enveloped filamentous viruses.

The second line of research focused on understanding the mechanisms of virion assembly and release on the example of the enveloped filamentous virus SIFV, which infects the hyperthermophilic and acidophilic archaeon *Sulfolobus islandicus*. Our results showed that SIFV is a lytic virus, which is released through pyramidal portals formed in the host cell membrane, a highly unexpected egress mechanism for an enveloped virus. Interestingly, dual-

axis electron tomography revealed that SIFV virions acquire their lipid envelopes inside the host through an unknown mechanism, involving either de novo membrane formation or trafficking of lipids from the cytoplasmic membrane to virion assembly centers.

Collectively, our results show that high-temperature continental hydrothermal systems harbor a highly diverse virome and shed light on the evolution of archaeal viruses. Moreover, our data contribute to the knowledge on virus-host interplay in extreme geothermal environments.

Keywords

Archaea, archaeal viruses, assembly, diversity, filamentous viruses, hyperthermophiles, release, virion structures, virus-associated pyramids, virus-host interactions.

3. PUBLICATIONS

1. **Baquero, D.P**., Contursi, P., Piochi, M., Bartolucci, S., Liu, Y., Cvirkaite-Krupovic, V., Prangishvili, D., and Krupovic, M. (2020) New virus isolates from Italian hydrothermal environments underscore the biogeographic pattern in archaeal virus communities. The ISME journal *14*, 1821-1833.

2. Wang, F. [#], **Baquero, D.P.[#]**, Beltran, L.C., Su, Z., Osinski, T., Zheng, W., Prangishvili, D., Krupovic, M., and Egelman, E.H. (2020) Structures of filamentous viruses infecting hyperthermophilic archaea explain DNA stabilization in extreme environments. Proceedings of the National Academy of Sciences of the United States of America *117*, 19643-19652.

3. **Baquero, D. P. [#]**, Gazi, A.[#], Sachse, M., Liu, J., Schouten, S., Prangishvili, D., and Krupovic, M. A filamentous archaeal virus is enveloped inside the cell and released through pyramidal portals. *Manuscript in preparation.*

# equal contribution

4. INTRODUCTION

4.1 Archaea: The third domain of life

The current classification of life recognizes three cellular domains, namely Eukarya, Bacteria and Archaea. This classification was proposed by Carl R. Woese and George E. Fox in 1977 based on ribosomal rRNA sequences (Woese and Fox, 1977). The development of comparative genomics and phylogenetic analyses enabled Woese and collaborators to propose that the third domain of life, the Archaea, could be subdivided into two phyla: Euryarchaeota and Crenarchaeota (Woese et al., 1990). Whereas Euryarchaeota is highly diverse in terms of occupied habitats and encompasses methanogens, extreme halophiles, acidophiles, mesophiles and thermophiles, Crenarchaeota is phylogenetically and metabolically less diverse and is mainly composed of thermophiles and hyperthermophiles (Baker et al., 2020; Gupta and Shami, 2011; Woese et al., 1990). Remarkably, during the last decades, advances in cultivation-independent genomic approaches and sequencing technologies have uncovered an unsuspected diversity of new archaeal lineages distinct from Crenarchaeota and Euryarchaeota. These findings have permitted the classification of Archaea into four superphyla: Euryarchaeota, TACK (acronym for Thaumarchaeota, Aigarchaeota, Crenarchaeota and Korarchaeota, the original phyla assigned to this superphylum), DPANN (acronym for Diapherotrites, Parvarchaeota, Aenigmarchaeota, Nanoarchaeota, Nanohaloarchaeota) and Asgard (Baker et al., 2020; Eme et al., 2017; Guy and Ettema, 2011; MacLeod et al., 2019; Rinke et al., 2013; Woese et al., 1990; Zaremba-Niedzwiedzka et al., 2017) (Figure 1).

Euryarchaeota comprises the greatest number of metabolically diverse cultured lineages, including methanogenic, halophilic, acidophilic and thermophilic archaea (Andrei et al., 2012; Biddle et al., 2006; Fox et al., 1977; Larsen, 1986; Reysenbach et al., 2006; Smith and Hungate, 1958; Takai and Horikoshi, 1999; Takai et al., 2001). Originally, it was thought that archaea from this phylum were exclusively involved in methane production (methanogens). However, subsequent studies have shown that euryarchaea have a much broader ecological impact, playing key roles in biogeochemical cycles (Haroon et al., 2013; Orphan et al., 2002; Raghoebarsing et al., 2006). In addition, recent analyses have shown that certain euryarchaeotes are responsible for anaerobic methane oxidation and might degrade other short-chain hydrocarbons, such as butane (Wang et al., 2019c). In contrast, most members of the Crenarchaeota are hyperthermophiles with optimal growth temperatures around 80°C, thriving in sulfur-rich, hot terrestrial and marine environments (Woese et al., 1990). While some

crenarchaea are sulfur-cycling chemolithoautotrophs, many others are anaerobic heterotrophs. The most extensively studied group of crenarchaea corresponds to *Sulfolobus* spp., first isolated from acidic hot springs in Yellowstone National Park (Brock et al., 1972). Notably, development on genetic tools have turned *Sulfolobus* into a well-stablished model for studies on hyperthermophilic archaea and their viruses (Zhang et al., 2018a).

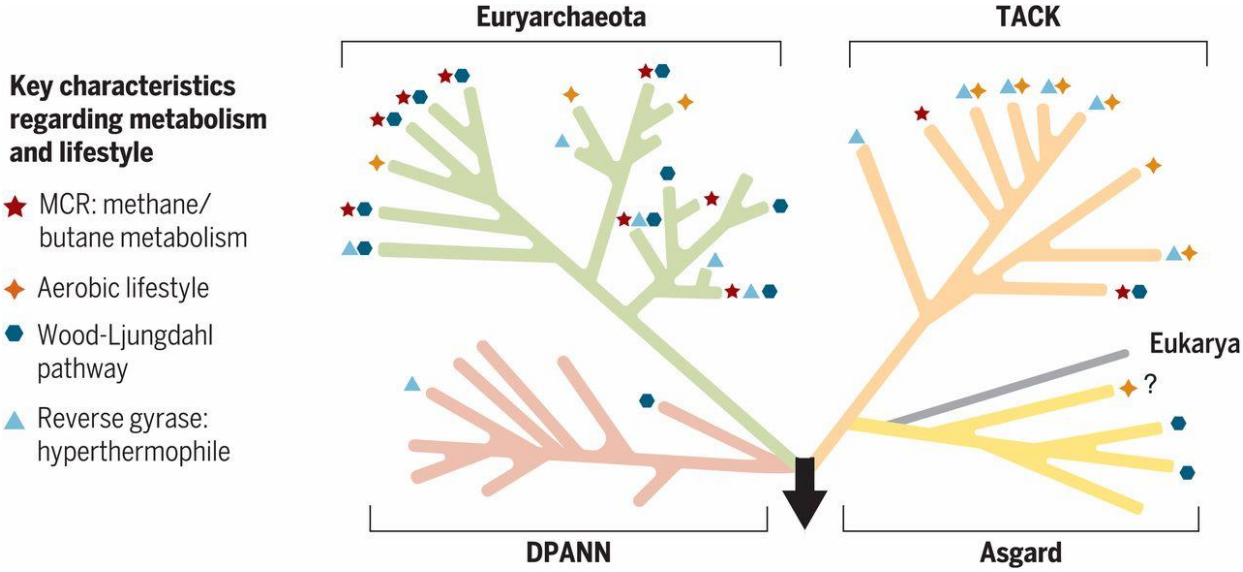


Figure 1 The archaeal three of life. Schematic phylogenetic tree including major archaeal groups for which genomic data are currently available. The currently recognized groups include the Euryarchaeota (green branches) and archaea of the TACK, Asgard, and DPANN superphyla (tan, yellow, and pink branches, respectively). Eukaryotes (gray branch) are suggested to have emerged from the Asgard archaea upon endosymbiosis with an alphaproteobacterial partner (the mitochondrial endosymbiont). For each archaeal lineage, key characteristics regarding metabolism and lifestyle are depicted. The figure is reproduced with permission from Spang, A., Caceres, E.F., and Ettema, T.J.G. (2017). Genomic exploration of the diversity, ecology, and evolution of the archaeal domain of life. Science 357, eaaf3883.

Advances in DNA sequencing, metagenomics and computational approaches during the past few years have prompted the discovery of a vast archaeal diversity in diverse ecosystems, revealing several new taxonomic groups and the description of more than 30 phyla within this domain (Adam et al., 2017; Baker et al., 2020; Spang et al., 2017). The first additional phylum was termed Korarchaeota and includes the largest group of deep-branching unclassified archaea. Representatives of this phylum thrive in a variety of geographical and geochemical conditions, including both terrestrial and marine environments, and remain uncultivable under laboratory conditions (Barns et al., 1996; Barns et al., 1994; Elkins et al., 2008). Recent metagenomic analyses of deep-sea hydrothermal sediments have increased the available number of Korarchaeota genomes. Interestingly, several genomes of Korarchaeota are

predicted to contain *mcr*-like genes, which are proposed to be involved in methane cycling coupled with sulfate reduction (Dombrowski et al., 2018; McKay et al., 2019).

Studies on genetic diversity of natural communities have uncovered several uncultivated lineages related to Crenarchaeota, which were commonly named as "mesophilic Crenarchaeota" due to their ability to grow in moderate-temperature conditions. However, phylogenomic analyses of conserved genes and differences in gene content have located these lineages into a new phylum, namely Thaumarchaeota (Brochier-Armanet et al., 2008). Members of the phylum Thaumarchaeota are abundant in marine and soil ecosystems and play crucial roles in nitrogen cycling by mediating chemolithotrophic ammonia oxidation (Karner et al., 2001; Santoro et al., 2010). In addition, certain members have been reported to obtain ammonia from urea and cyanate, as well as to encode a complete aerobic pathway for highly energy-efficient carbon fixation (Baker et al., 2012; Hallam et al., 2006; Könneke et al., 2014; Palatinszky et al., 2015). The genome analysis of the uncultivated *Candidatus* 'Caldiarchaeum subterraneum' led to the proposal of a sister lineage of Thaumarchaeota, termed Aigarchaeota (Nunoura et al., 2011). Members of the Aigarchaeota are widely distributed in marine hydrothermal environments as well as in terrestrial and subsurface geothermal systems (Beam et al., 2016; Hedlund et al., 2015; Hua et al., 2018; Nunoura et al., 2005). Phylogenetic analysis of single-copy core genes showed that **T**haumarchaeota, **A**igarchaeota, **C**renarchaeota and **K**orarchaeota are closely related and form a monophyletic assemblage, which was called the "TACK" superphylum. Interestingly, members of this superphylum share a common set of genes involved in cytokinesis, protein recycling, membrane remodelling and cell shape determination (Guy and Ettema, 2011; Hua et al., 2018).

An additional phylum, 'Nanoarchaeota', was proposed with the discovery of *Nanoarchaeum equitans*, a nanosized organism (500 nm in diameter) that turned out to be an obligate symbiont of the crenarchaeon *Ignicoccus hospitalis* (Huber et al., 2002). In agreement with its host dependency, genomic analysis of *N. equitans* showed that its genome is one of the smallest known (~490 kb) and lacks essential genes for lipid, amino acid and nucleotide biosynthesis (Waters et al., 2003). Notably, the development of single-cell and metagenomic approaches have unveiled many other nanosized and/or genome-reduced archaeal lineages during the past few years (Castelle and Banfield, 2018; Dombrowski et al., 2019; Probst et al., 2018; Rinke et al., 2013). These genome-reduced Archaea have been proposed to form a monophyletic and

deep-branching archaeal superphylum, named DPANN - referring to the phyla **D**iapherotrites, **P**arvarchaeota, **A**enigmarchaeota, **N**anoarchaeota and **N**anohaloarchaeota. Unlike *N. equitans*, which has been extensively studied, very little is known about the other DPANN lineages and our current knowledge mostly relies on genomic data (Dombrowski et al., 2019; Rinke et al., 2013).

The recent discovery of Lokiarchaeota in 2015, a clade found in deep marine sediments, has provided important insights into the potential archaeal origin of eukaryotes (Spang et al., 2015). Phylogenomic approaches have suggested that Lokiarchaeota forms a sister group to Eukaryotes, suggesting a shared common ancestor. Thus, it has been postulated that the first eukaryotic cells have evolved from an archaeal ancestor (Spang et al., 2015). Genomic analyses have revealed that Lokiarchaeota encode a handful of genes that were previously thought to be specific to eukaryotes. These shared eukaryotic signatures proteins (ESP) comprise cytoskeletal components (actin homologues, gelsolin-domain proteins and actin regulator profilins), constituents of the ubiquitin modifier system and a variety of small GTPases involved in regulatory processes such as cytoskeleton remodeling and signal transduction (Akıl and Robinson, 2018; Spang et al., 2015; Zaremba-Niedzwiedzka et al., 2017). Notably, lokiarchaea encode biosynthetic pathways that might be involved in formation of the bacterial/eukaryotic-type ester-linked lipid membrane, rather than the ether-based membrane as in the case of other archaea (Villanueva et al., 2017), supporting the hypothesis that an archaeal ancestor was at the origin of the first eukaryotic cells (Spang et al., 2015).

Searches of novel archaeal lineages related to Lokiarchaeota, led to the metagenomic discovery of the Asgard superphylum, which includes new uncultivated members spread in a wide variety of environments, including hydrothermal sediments, microbial mats as well as freshwater and marine environments (Dombrowski et al., 2018; MacLeod et al., 2019; Seitz et al., 2016; Spang et al., 2015; Tully et al., 2018; Wong et al., 2018; Zaremba-Niedzwiedzka et al., 2017). Some members of the Asgard superphylum have been reported as mixotrophs, with putatively significant roles in sulfur and nitrogen cycling (Liu et al., 2018b; Seitz et al., 2016). A new class of rhodopsin has also been identified in the genomes of Asgard, indicating a possible phototrophic lifestyle (MacLeod et al., 2019; Pushkarev et al., 2018). Importantly, following a decade of experimentation, an Asgard archaeon related to Lokiarchaeota was successfully isolated from deep-sea marine sediments (Imachi et al., 2020). *Ca.* 'Prometheoarchaeum

syntrophicum' is an anaerobic, extremely slow-growing strain that degrades amino acids through syntrophy. Morphologically, *Ca*. P. syntrophicum possesses unique long, often branched protrusions and has no visible eukaryote-like intracellular complexes, as previously predicted for Asgard archaea (Imachi et al., 2020; Zaremba-Niedzwiedzka et al., 2017).

The proposal of Asgard as the closest prokaryotic relatives of eukaryotes has raised a debate as to whether eukaryotes have evolved as a sister clade of Archaea (three-domain hypothesis) or branched off from within Archaea (two-domain theory) (Da Cunha et al., 2017; Da Cunha et al., 2018; Heinz and Domman, 2017; Lake, 2015; Sun et al., 2020; van der Gulik et al., 2017; Zhou et al., 2018). Notably, recent discovery of six additional Asgard phyla and phylogenetic analysis of universally conserved genes of the expanded Asgard dataset showed no strong support for the origin of eukaryotes from within Asgard but rather suggested a three-domain topology for the tree of life, with eukaryotes branching outside archaea (Liu et al., 2020). Cultivation efforts, coupled with advances in molecular biology, genetics and phylogenetic approaches are expected to shed further light on the evolutionary connections within and between the domains of life.

4.2 Environmental distribution of Archaea

Archaea were long considered to exist solely in environments with harsh conditions, at the extremes of temperature, salinity or pH. Indeed, some archaeal species hold records for growing at the highest temperature (*Methanopyrus kandleri* strain 116, 122°C), lowest pH (*Picrophilus oshimae* and *Picrophilus torridus*, pH ~ 0) and highest salt concentration (*Halobiforma haloterrestris*, ~ 5.2 M NaCl) (Christian and Waltho, 1962; Schleper et al., 1995; Takai et al., 2008). However, cultivation-independent analyses in the last decade revealed that archaea are ubiquitous in almost all environments analyzed until now, including diverse aquatic and soil ecosystems, and are part of the microbiome of various animals, including humans (Bates et al., 2011; Cowie et al., 2011; Karner et al., 2001; Koskinen et al., 2017; Probst et al., 2013; Raymann et al., 2017).

Thermophilic and hyperthermophilic archaea are found in a variety of globally diverse geothermal environments, ranging from deep sea hydrothermal vents to terrestrial hot springs (Ding et al., 2017; Kambura et al., 2016; Pagaling et al., 2012; Satoh et al., 2013; Straub et al.,

2018; Zhang et al., 2016). While thermophiles are defined as growing between 50°C and 70°C, the optimal growth temperature of hyperthermophiles is ~80°C and can be up to 122°C (Stetter, 1996). In general, hyperthermophiles consist of aerobic and anaerobic chemolithoautotrophs and heterotrophs growing at neutral or acidic pH. Acidophilic hyperthermophiles, members of the genera *Sulfolobus*, *Acidianus*, *Metallosphaera* and *Stygiolobus*, are found in terrestrial and marine solfataric fields and grow optimally at pH of around 3 (Ai et al., 2016; Grogan, 1989; Stetter, 1996; Wheaton et al., 2015). Notably, both thermophiles and hyperthermophiles have developed similar adaptations to thrive in such harsh conditions. These specific structural adaptations to avoid irreversible unfolding and undesired aggregation include an increased hydrophobic core of the protein, larger number of disulfide bonds and salt-bridging (Cacciapuoti et al., 2012; Fukuchi and Nishikawa, 2001; Karshikoff and Ladenstein, 2001; Reed et al., 2013; Tomazic and Klibanov, 1988; Vieille and Zeikus, 2001).

Halophilic archaea thrive in environments with extremely high concentrations of salt (salinities ranging from 10% to ~ 36%) and are found in a wide variety of environments that differ in location, climate and chemical composition, such as salt lakes, soda lakes and seawater ponds (Deng et al., 2015; Luk et al., 2014; Sorokin et al., 2014; Tazi et al., 2014). Notably, haloarchaea dominate the microbial communities in ecosystems with NaCl concentrations of around 15-20%, with bacterial halophiles accounting for less than a quarter of the population (Oren, 2002). Other archaeal species have been found in ecological niches with high pH (alkaliphiles), low temperatures (psycrophiles), high pressure (piezophile) or anoxia (anaerobes), highlighting the enormous diversity and ubiquity of archaea (Dalmasso et al., 2016; Duckworth et al., 1996; Franzmann et al., 1997; Franzmann et al., 1992; Mayer and Müller, 2014; Tazi et al., 2014).

Methanogenic archaea are the only known organisms capable of using CO_2 and H_2 as energy and carbon sources to produce methane. Despite their strict anaerobic requirements, methanogenic archaea are globally distributed in different environments, including hydrothermal vents, soils, swamps, freshwater sediments, marine ecosystems and the intestinal tract of certain animals (Hackstein and van Alen, 1996; Jones et al., 1983; Knittel et al., 2005; Valentine, 2002). Remarkably, methanogens have a quantitatively large impact on the global carbon cycle, with estimations of around 1 billion tons of methane per year formed by methanogenic archaea in anoxic environments (Thauer et al., 2008).

Although no pathogenic archaeon has been identified so far, many reports have shown archaea to be widely distributed in the human gut as well as in the vaginal cavity, oral cavity and skin (Bang and Schmitz, 2015; Belay et al., 1990; Borrel et al., 2020; Dridi et al., 2011; Probst et al., 2013). Interestingly, a study has demonstrated that two of the most common archaea in the human intestinal tract, *Methanobrevibacter smithii* and *Methanosphaera stadtmanae*, are recognized by the human innate immunity system, causing the activation of an inflammatory cytokine response to different extents (Bang et al., 2014).

Metagenomic analyses of several environmental samples collected around the globe have shown that, besides being ubiquitous, archaea are abundant in different ecological niches. For instance, archaea represent about 20% of all microbial biomass present in the oceans (DeLong and Pace, 2001). Similarly, a global quantification of archaeal and bacterial 16S rRNA genes in subseafloor sediment samples showed that archaeal cells represents 37.3% of the total microbial cells present in the subseafloor sedimentary biosphere, corresponding to 1.1×10^{29} cells on Earth (Hoshino and Inagaki, 2019). Moreover, ammonia-oxidizing archaea, which belong to the phylum Thaumarchaeota and contribute significatively to the global nitrogen and carbon cycling, have been shown to represent one of the most abundant microbial groups in the ocean. Quantitative analysis demonstrated that Thaumarchaeota represent up to 20% (about 1.0×10^{28} cells) of all microbial prokaryotes present in the oceans, suggesting a key role of ammonia oxidizers in this vast ecosystem (Karner et al., 2001; Pester et al., 2011). Notably, the archaeal gene encoding a subunit of the enzyme ammonia monooxygenase (*amoA*) was found to be 3,000-fold more abundant than its bacterial counterpart in agricultural soils (Leininger et al., 2006), pointing out the important role of ammonia-oxidizing archaea in distinct ecological niches as well as in the global biogeochemistry.

4.3 Characteristics of Archaea

Although the major fraction of archaeal genes are domain-specific, some archaeal genes have homologues in ether bacteria or eukaryotes (Allers and Mevarech, 2005; Leigh et al., 2011; Yutin et al., 2008). Besides being morphologically nearly indistinguishable from bacteria, archaea share with bacteria a similar chromosomal organization and lack of intracellular compartments (Bell and Jackson, 1998). Similarly, archaeal operational genes involved in

central metabolism, energy conversion and biosynthesis are most similar to their bacterial counterparts (Allers and Mevarech, 2005). Archaea and bacteria also share several defense systems against mobile genetic elements (MGE), including restriction-modification, abortive-infection and toxin-antitoxin systems as well as CRISPR-Cas adaptive immunity and Argonaute-based innate immunity (Makarova et al., 2013). Nevertheless, a closer look at Archaea has revealed several genetic and biochemical similarities with Eukarya. These common features include the information processing machineries for DNA replication, transcription and repair, e.g., the similar structures of the DNA-dependent RNA polymerases and occurrence of histones (Allers and Mevarech, 2005; Baker et al., 2020; Henneman et al., 2018; Huet et al., 1983). Likewise, the lack of peptidoglycan in the cell wall is also a shared attribute in Archaea and Eukarya (Kandler and König, 1978). Among the archaea-specific features, the most remarkable are the ether-based lipid membranes and the uniqueness of certain physiologies and metabolic pathways, e.g., the production of methane by methanogenic archaea and the capacity to thrive in harsh environmental conditions such as hypersaline environments and extremely high temperatures (Kates, 1977; Koga et al., 1993; Smith and Hungate, 1958; Woese et al., 1990).

4.3.1 Cell morphology and structure

Similar to bacteria, archaea are single-celled organisms, lacking the nucleus and membrane-bound organelles. Cells are relatively small, most ranging from 0.5 to 4 μm in diameter and display a variety of shapes, with coccoid and rod-shaped being the most common morphologies. Interestingly, unique forms, like triangular discs or square box-shaped morphologies have been described for haloarchaea (Javor et al., 1982; Oren, 1999; Rodrigues-Oliveira et al., 2017). Cell wall-lacking archaea of the order *Thermoplasmatales* are pleomorphic, varying in shape from spherical to filamentous structures. Notably, the presence of a lipopolysaccharide-like material (consisting of a unique tetraether lipid with mannose and glucose units) in the membrane of *Thermoplasma acidophilum* has been suggested to ensure the membrane stability under extremely hot and acidic conditions (Shimada et al., 2002). Archaea also have been reported to form biofilms in a diversity of ecological niches, including Antarctic seawater, acid mine drainages, alkaline lakes and the human gut (Couradeau et al., 2011; Edwards et al., 2000; van Wolferen et al., 2018; Webster and Negri, 2006). Despite the apparent ubiquity of archaeal biofilms, little is known about their formation mechanisms, especially when compared to the wealth of available information for bacterial biofilms. However, some studies have suggested that biofilms might offer certain advantages to archaeal

cells such as protection against harsh external conditions, favoring gene transfer and supporting symbiotic relationship with other microorganisms (Chimileski et al., 2014; Koerdt et al., 2010; Wegener et al., 2015). In addition, electron microscopic observations as well as direct studies on the archaeal cell surfaces have shown that archaea display a variety of surface structures, some of which are unique to archaea (Albers and Meyer, 2011) .

4.3.2 Cell envelope characteristics

The cell envelope of archaea displays some unique features that distinguish archaea from bacteria end eukaryotes. The most notable difference includes the composition of the lipid membrane: while bacterial and eukaryotic membranes are composed of fatty acid chains ester-linked to glycerol moieties, archaeal membrane lipids are characterized by ether linkages between isoprene-based alkyl chains and glycerol moieties (Figure 2) (De Rosa et al., 1986; Gribaldo and Brochier-Armanet, 2006; Jain et al., 2014; Kates, 1977; Koga et al., 1993; Villanueva et al., 2014). Interestingly, while haloarchaea and other mesophilic archaea largely contain diether lipids (archaeol) of C_{20} phytanyl chains that assemble into bilayer membranes, certain acidophilic and hyperthermophilic archaea encompass tetraether lipids (glycerol dibiphytanyl glycerol tetraethers [GDGT]) of C_{40} isoprenoid chains that span the entire thickness of the membrane yielding a unique lipid monolayer (Koga and Morii, 2007; Villanueva et al., 2014). This lipid monolayer has been suggested to decrease the membrane fluidity and, therefore, be an important adaptation of hyperthermophiles and acidophiles to extreme geothermal conditions by reducing the membrane permeability (Jacquemet et al., 2009; Siliakus et al., 2017).

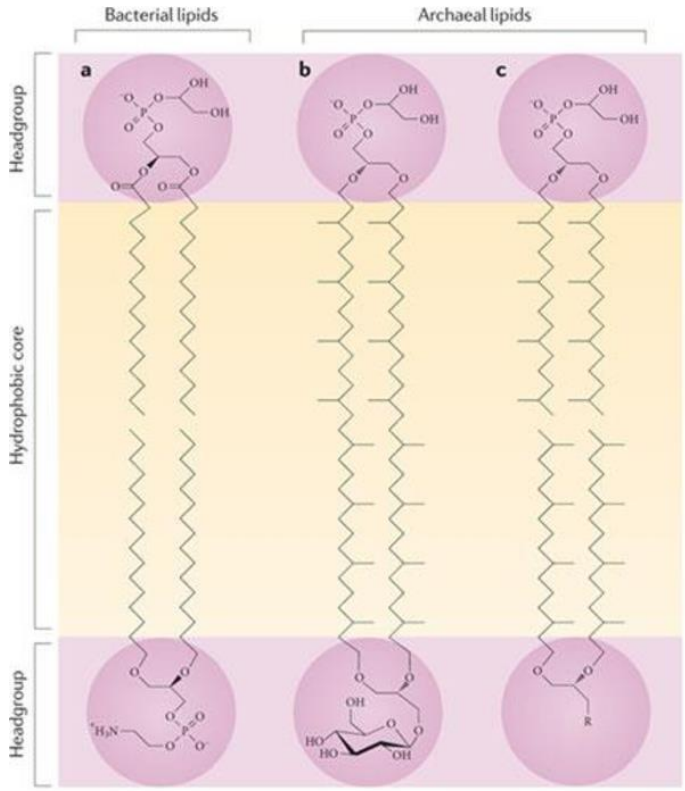


Figure 2. The lipids found in the archaeal membrane are fundamentally different from those found in eukaryotic and bacterial membranes. In eukaryotes and bacteria, the glycerol moiety is ester-linked to an *sn*-glycerol-3-phosphate backbone, whereas in archaea the isoprenoid side chains are ether-linked to an *sn*-glycerol-1-phosphate moiety. The *sn*1 stereochemistry of the glycerol backbone is a truly archaeal feature, as ether lipids occur in minor amounts in eukaryotes and bacteria. The common bilayer-forming lipids in bacteria are phophatidylglycerol (upper lipid) and phosphatidylethanolamine (lower lipid) (see the figure, part a). Part b of the figure shows the structure of monolayer-forming tetraether lipids; for example, the glycophospholipid from the thermoacidophilic archaeon *Thermoplasma acidophilum*, in which the hydrophobic core consists of $C_{40}C_{40}$ caldarchaeol. Part c of the figure shows a bilayer formed of archaeal diether lipids, which can be found, for example, in Halobacteriales. The hydrophobic core consists of $C_{20}C_{20}$ archaeol isoprenoids. The headgroups of phospholipids can be a range of polar compounds — for example, glycerol, serine, inosine, ethanolamine, myo-inositol or aminopentanetetrols. Glycolipids also exhibit a range of sugar residues — for example, glucose, mannose, galactose, gulose, *N*-acetylglucosamine or combinations thereof. The figure is reproduced with permission from Albers, S.V., and Meyer, B.H. (2011). The archaeal cell envelope. Nature reviews Microbiology 9, 414-426.

Most archaea are surrounded by a thin highly-ordered protein layer anchored to the cell membrane, which is knowns as the surface (S-) layer. The S-layer is proposed to maintain an osmotic balance, protect the cell from harsh environmental conditions and contribute to cell shape (Albers and Meyer, 2011; Sára and Sleytr, 2000; Sleytr, 1976; Sleytr et al., 1999). Usually, S-layer proteins carry extensive post-translational modifications, such as N- and O-glycosylation, which are believed to be a common strategy to protect the cells from the extreme conditions of the environment (Eichler, 2003; Palmieri et al., 2013). Remarkably, S-layer proteins are able to assemble into a two-dimensional crystalline protein array with distinct geometries, with the hexagonal lattice being the most common in archaeal cells (Albers and Meyer, 2011; Sleytr et al., 1999). This intrinsic self-assembling property makes S-layer suitable

for a variety of nanotechnology applications (Ilk et al., 2011). Bacterial and archaeal S-layers are usually composed of one (glyco)protein (Rodrigues-Oliveira et al., 2017), although in Sulfolobales, the S-layer is composed of two conserved glycosylated proteins, SlaA and SlaB(Veith et al., 2009). The S-layer model for *Sulfolobus* has been proposed, in which both proteins have recognizable and distinct roles during the S-layer building: the small, membrane-bound protein SlaB acts as a stalk that anchors the large protein SlaA to the cytoplasmic membrane, leading to the formation of a crystal-like matrix that surrounds the cell (Veith et al., 2009; Zhang et al., 2018b).

Few archaeal species possess a cell wall consisting of pseudomurein polymers, which superficially are equivalent, but not evolutionarily related to the bacterial peptidoglycan (Claus and König, 2010; König et al., 1993; Visweswaran et al., 2011). The rod-shaped methanogenic archaea *Methanospirillum hungatei* and *Methanosaeta concilii* possess a more complex cell envelope composed of unique tubular paracrystalline proteinaceous sheaths. Structurally, these sheaths are less porous and more stable than regular S-layer due to the recurrent presence of disulphide bonds between cysteine residues (Beveridge and Graham, 1991). *Methanospirillum* and *Methanosaeta* grow and divide within these proteinaceous sheaths, which enclose an entire chain of cells, rather than single cells (Beveridge et al., 1986; Beveridge et al., 1985; Zeikus and Bowen, 1975).

4.3.3 Cell surface appendages

Similar to bacteria, archaea possess distinct cell surface appendages, including archaeal flagella (archaella) implicated in cell motility and diverse pili involved in intercellular communication, biofilm formation and adhesion to various organic and inorganic substrates (Albers and Meyer, 2011; Pohlschroder and Esquivel, 2015). Interestingly, although archaeal and bacterial flagella are rotating organelles, structural and molecular characterization has revealed that the two structures are **evolutionarily** unrelated (Albers and Jarrell, 2018; Jarrell and Albers, 2012; Streif et al., 2008; Trachtenberg and Cohen-Krausz, 2006). Accordingly, the archaellum acts as a propeller-type driven machine superficially similar to bacterial flagellum (Daum and Gold, 2018; Jarrell and Albers, 2012), but structurally related to archaeal type IV pili (T4P) (Wang et al., 2020b; Wang et al., 2019a). Furthermore, whereas bacterial flagella use the energy in the proton motive force to drive the rotation, archaella employ ATP hydrolysis (Albers and Jarrell, 2018; Macnab, 1999; Streif et al., 2008).

Bioinformatic and structural approaches have shown that many archaeal surface structures are homologous to bacterial T4P (Makarova et al., 2016). Indeed, archaeal T4P are among the most commonly found surface appendages in described archaea and correspond to proteinaceous structures formed from subunits of the type IV pilin proteins (Pohlschroder and Esquivel, 2015). Recent structural studies have elucidated the structure of T4P in hyperthermophilic archaea; surprisingly, *Sulfolobus islandicus* T4P has an extremely stable structure that resists harsh treatments such as digestion by trypsin and pepsin as well as boiling in sodium dodecyl sulfate and 5M guanidinium-HCl. The resilience of the pilus was attributed to extensive surface glycosylation on serine and threonine residues, which are anomalously abundant in the *Sulfolobus* T4P, compared to the bacterial T4P and other *Sulfolobus* proteins (Wang et al., 2019a). Importantly, cryo-EM structures of the T4P of the acidophilic hyperthermophile *Saccharolobus solfataricus* and the neutrophilic hyperthermophile *Pyrobaculum arsenaticum* suggested that the extensive glycosylation previously observed in the *Sulfolobus islandicus* pilus is an adaptation to acidic environments, rather than extreme temperatures. At even higher temperatures but neutral pH, much less glycosylation was observed in *Pyrobaculum* than in *Sulfolobus* and *Saccharolobus* T4P (Figure 3) (Wang et al., 2020b).

Cannulae and hami are less-studied surface appendages that are unique to archaea. Cannulae are found exclusively in the marine hyperthermophilic archaea of the genus *Pyrodictium*. Interestingly, *Pyrodictium* spp. are found as a dense network of cells and tubules, because dividing cells remain interconnected through cannulae. The function of these structures remains unclear, but it has been proposed that they might be involved in exchange of nutrients or cell attachment (Horn et al., 1999; Stetter et al., 1983). The euryarchaeon SM1, which grows in cold sulphidic springs, displays on its surface helical filaments named hami, from which three hooks emanate every 4 nm (Moissl et al., 2005; Rudolph et al., 2001). Hami are proposed to favor cell-cell attachments for the formation of microbial communities (Albers and Meyer, 2011).

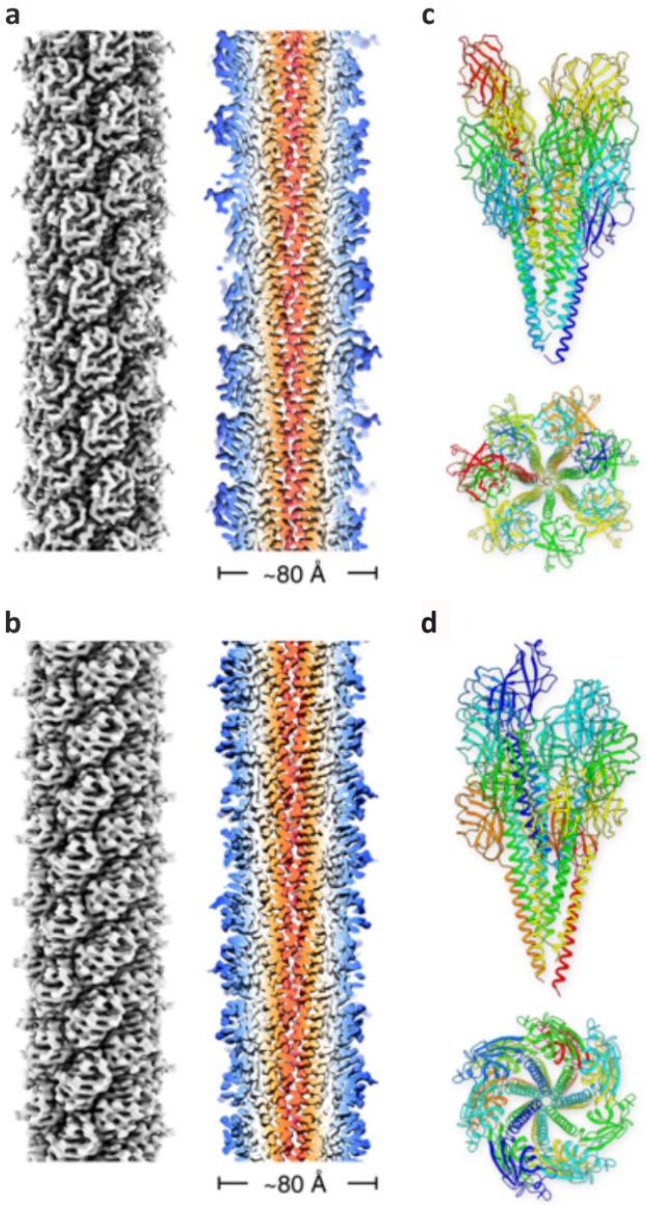


Figure 3. Cryo-EM of the *Pyrobaculum arsenaticum* pilus and the *Saccharolobus solfataricus* pilus. Cryo-EM reconstruction of the *P. arsenaticum* pilus at 3.8 Å resolution (a) and the *S. solfataricus* pilus at 3.4 Å resolution (b). Thin slices parallel to the helical axis of the pilus are shown, colored by the helical radius. (c, d) Side view and top view of the *P. arsenaticum* and *S. solfataricus* pilus atomic models, built into the cryo-EM maps shown in a, b. The model is colored by chain. Image modified and reproduced with permission from Wang, F., Baquero, D.P., Su, Z., Beltran, L.C., Prangishvili, D., Krupovic, M., and Egelman, E.H. (2020). The structures of two archaeal type IV pili illuminate evolutionary relationships. Nature Communications 11, 3424.

4.3.4 Cell division

Similar to bacteria, archaea reproduce asexually through binary fission (Makarova et al., 2010; Samson and Bell, 2011). Although no archaeal species are known to produce spores, some haloarchaea undergo phenotypic switching to thick-walled cells that are resistant to osmotic shock, superficially resembling spores (Kostrikina et al., 1991). Unlike in bacteria and eukaryotes, cell division in archaea remains poorly understood. Yet, archaeal division follows

the four regular conceptual stages of cytokinesis: (i) placement of the early division components at the division site; (ii) recruitment of later division components to the divisional site, (iii) application of a force on the membrane that leads to constriction, and (iv) membrane abscission generating the segregation of the daughter cells (Caspi and Dekker, 2018). Interestingly, whereas Crenarchaeota encompass homologs of the eukaryotic Endosomal Sorting Complex Required for Transport (ESCRT) machinery, which is involved in membrane abscission during cytokinesis, almost all members of the Euryarchaeaota encode homologs of the bacterial tubulin-like protein FtsZ, which forms contracting rings at the septum during cell division (Bernander, 1998; Hobel et al., 2008; Lindås et al., 2008).

4.4 Diversity of Archaeal Viruses

Viruses infecting archaea constitute a distinctive part of the virosphere and exhibit diverse virion morphologies, many of which have never been observed among viruses infecting bacteria or eukaryotes. Diversity of archaeal viruses is also reflected in their genome content, with ~75% of the genes lacking detectable homologs in sequence databases (Krupovic et al., 2018). The vast majority of isolated archaeal viruses infect hyperthermophiles and hyperhalophiles of the phyla Crenarchaeaota and Euryarcheaota, respectively (Luk et al., 2014; Prangishvili et al., 2017). However, a handful of viruses have been recently isolated and described to infect methanogens (Pfister et al., 1998b; Thiroux et al., 2020; Weidenbach et al., 2017; Wolf et al., 2019) and ammonia-oxidizing thaumarchaea (Kim et al., 2019). In addition, putative viruses associated with thaumarchaea (Ahlgren et al., 2019; Chow et al., 2015; Danovaro et al., 2016; Krupovic et al., 2019; Krupovic et al., 2011; Roux et al., 2016), marine thermoplasmata (Vik et al., 2017) and marine group II Euryarchaeota (Nishimura et al., 2017; Philosof et al., 2017) were identified using metagenomics, single cell genomics and mining of the archaeal genome sequences for proviruses, suggesting that archaeal viruses play important ecological roles in different ecosystems. Indeed, it has been demonstrated that in the deep ocean, virus-mediated lysis of archaea is more rapid than that of bacteria, with the average abatement rate of 3.2% per day versus 1.6% per day, respectively. As a result, turnover of archaea in surface deep-sea sediments accounts for up to one-third of the total microbial biomass killed, resulting in the release of ~0.3 to 0.5 gigatons of carbon per year globally (Danovaro et al., 2016).

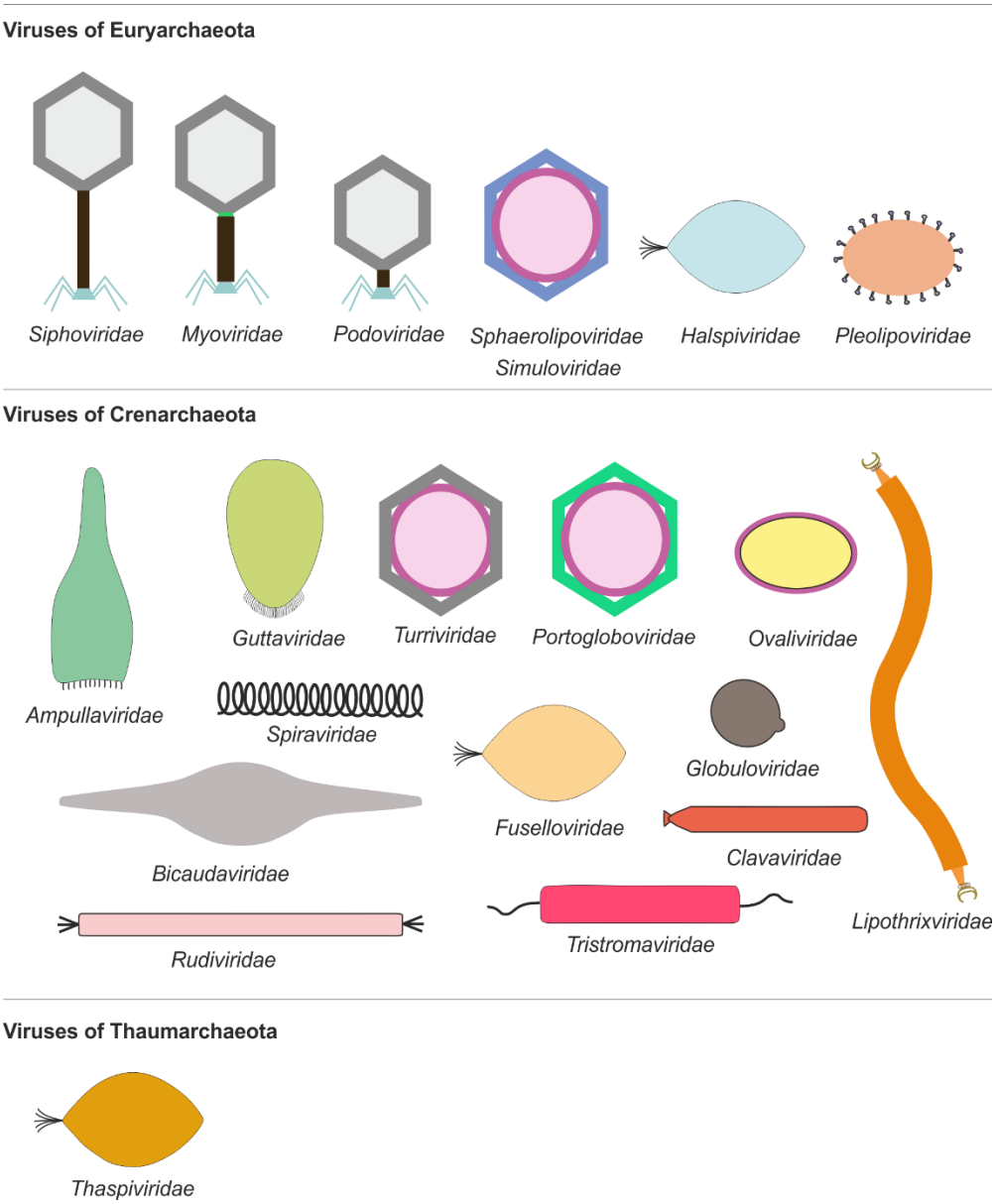


Figure 4. Representation of virion morphotypes of archaeal viruses infecting members of the phyla Euryarchaeota, Crenarchaeota and Thaumarchaeota. Names of viral families are indicated below the schematic virus particles.

Based on their diverse virion morphologies and genomic properties, the characterized archaeal viruses are currently classified into 20 families (Adriaenssens et al., 2020) (Figure 4 and Table 1). Notably, taxonomic classification of a variety of additional groups remains to be defined (Geslin et al., 2007; Gorlas et al., 2012; Liu et al., 2019; Mochizuki et al., 2011; Munson-McGee et al., 2020; Wagner et al., 2017; Weidenbach et al., 2017). Archaeal virome encompasses virus groups which are evolutionarily related to viruses of bacteria and eukaryotes, but most archaeal virus lineages have no counterparts among viruses from the two other domains (Krupovic et al., 2020). Moreover, comparative genomics and bipartite gene-genome network analyses have shown that most of the hyperthermophilic archaeal viruses are

disconnected from the rest of the virosphere and likely display distinct evolutionary origins (Iranzo et al., 2016a; Iranzo et al., 2016b).

Table 1. Representative viruses of the Archaea

Family name	**Virion shape**	**Representative member**	**Host**	**Genome topology and size (bp)**
Ampullaviridae	Bottle-shaped, fibers at the broad end	Acidianus bottle-shaped virus (ABV)	*Acidianus convivator*	Linear dsDNA, 23,900
Bicaudaviridae	Spindle-shaped with one or two appendages	Acidianus two-tailed virus (ATV)	*Acidianus convivator*	Circular dsDNA, 62,730
Clavaviridae	Bacilliform	Aeropyrum pernix bacilliform virus 1 (APBV1)	*Aeropyrum pernix*	Circular dsDNA, 5,278
Fuselloviridae	Spindle-shaped with fibers located at one end	Sulfolobus spindle-shaped virus 1 (SSV1)	*Saccharolobus shibatae*	Circular dsDNA, 15,465
Globuloviridae	Spherical	Pyrobaculum spherical virus (PSV)	*Pyrobaculum* sp. D11	Linear dsDNA, 28,337
Guttaviridae	Droplet/ovoid-shaped	Aeropyrum pernix ovoid virus 1 (APOV1)	*Aeropyrum pernix*	Circular dsDNA, 13,769
Halspiviridae	Spindle-shaped with fibers located at one end	Haloarcula hispanica virus 1 (His1)	*Haloarcula hispanica*	Linear dsDNA, 14,462
Lipothrixviridae	Flexible filament with terminal structures	Acidianus filamentous virus 1 (AFV1)	*Acidianus hospitalis*	Linear dsDNA, 21,080
Myoviridae	Icosahedral head with long contractile tail	Halorubrum sodomense tailed virus 2 (HSTV-2)	*Halorubrum sodomense*	Linear dsDNA, 68,527

Ovaliviridae	Ellipsoid	Sulfolobus ellipsoid virus 1 (SEV1)	*Sulfolobus* sp. A20	Linear dsDNA, 23,219
Pleolipoviridae	Pleomorphic	Halorubrum pleomorphic virus 1 (HRPV-1)	*Halorubrum* spp.	Circular ssDNA, 7,048
Podoviridae	Icosahedral head with short tail	Haloarcula sinaiiensis tailed virus 1 (HSTV-1)	*Haloarcula sinaiiensis*	Linear dsDNA, 32,189
Portogloboviridae	Icosahedral capsid	Sulfolobus polyhedral virus 1 (SPV1)	*Saccharolobus shibatae*	Circular dsDNA, 20,222
Rudiviridae	Rigid rod, with terminal fibers	Sulfolobus islandicus rod-shaped virus 2 (SIRV2)	*Sulfolobus islandicus*	Linear dsDNA, 35,450
Sphaerolipoviridae	Icosahedral capsid	Haloarcula virus SH1	*Haloarcula hispanica*	Linear dsDNA, 30,898
Simuloviridae	Icosahedral capsid	Natrinema virus SNJ1	*Natrinema* sp. J7-1	Circular dsDNA, 16,341
Siphoviridae	Icosahedral head with long non-contractile tail	Haloarcula vallismortis tailed virus 1 (HVTV-1)	*Haloarcula vallismortis*	Linear dsDNA, 10,232
Spiraviridae	Coil-shaped with two terminal appendages	Aeropyrum coil-shaped virus (ACV)	*Aeropyrum pernix*	Circular ssDNA, 24,893
Thaspiviridae	Spindle-shaped with fibers located at one end	Nitrosopumilus spindle-shaped virus 1 (NSV1)	*Nitrosopumilus* sp. SW	Linear dsDNA, 27,548
Tristromaviridae	Filamentous with terminal fibers	Pyrobaculum filamentous virus 1 (PFV1)	*Pyrobaculum arsenaticum*	Linear dsDNA, 17,714
Turriviridae	Icosahedral capsid	Sulfolobus turreted icosahedral virus (STIV)	*Saccharolobus solfataricus*	Circular dsDNA, 17,663

4.4.1 Viruses with unique morphologies

Certain families of hyperthermophilic viruses comprise members with structural features that are not found among viruses infecting bacteria or eukaryotes. Virions of Acidianus bottled-shaped virus (ABV), the representative of the *Ampullaviridae* family, exhibit a distinctive bottle-shaped morphology (Häring et al., 2005a; Prangishvili et al., 2018a). The linear dsDNA viral genome is condensed into a cone-shaped nucleocapsid that is further encased in a lipid-containing envelope. The broad end of the bottle-shaped virions is decorated with short, thick filaments inserted into a structural disc, and the pointed end appears to be involved in attachment to host cells. Ampullaviruses have been detected in terrestrial hot springs in Italy, Iceland and Japan (Gudbergsdóttir et al., 2016; Häring et al., 2005a); however, only one representative of this family has been isolated.

The virions of Aeropyrum coil-shaped virus (ACV), the sole member of the family *Spiraviridae,* infect hyperthermophilic archaea from the genus *Aeropyrum* (order Desulfurococcales, phylum Crenarchaeaota) and are hollow, non-enveloped cylindrical particles with filamentous appendages protruding from both termini of the helical structure (Mochizuki et al., 2012; Prangishvili et al., 2020). The viral genome is a single-stranded, positive-sense DNA molecule. The virions are formed by the condensation of the circular ssDNA genome and capsid proteins into a rope-like structure that is further condensed into a spring-like coil (Mochizuki et al., 2012).

The *Guttaviridae* family includes Sulfolobus neozealandicus droplet-shaped virus (SNDV) and Aeropyrum pernix ovoid virus 1 (APOV1), infecting hyperthermophilic archaea from the crenarchaeal orders Sulfolobales and Desulfurococcales, respectively (Arnold et al., 2000a; Mochizuki et al., 2011; Prangishvili et al., 2018c). Guttaviruses have slightly pleomorphic enveloped virions with circular dsDNA genomes. Virions of Sulfolobus neozealandicus droplet-shaped virus (SNDV) resemble elongated droplets with varying dimensions and a bunch of thick filaments covering the pointed end of the particle (Arnold et al., 2000a). Virions of another member of the family, Aeropyrum pernix ovoid virus 1 (APOV1), are ovoid without exhibiting attached fibers (Mochizuki et al., 2011).

4.4.2 Filamentous viruses

Unlike bacterial and eukaryotic filamentous viruses that have ssDNA and ssRNA genomes, respectively, filamentous archaeal viruses possess dsDNA genomes. They have been grouped

into the families *Rudiviridae* (Prangishvili et al., 1999), *Lipothrixviridae* (Prangishvili and Krupovic, 2012), *Tristromaviridae* (Prangishvili et al., 2019b) and *Clavaviridae* (Prangishvili et al., 2019a), and infect hyperthermophilic hosts from the orders Sulfolobales, Desulfurococcales and Thermoproteales, of the phylum Crenarchaeota. The three former families are evolutionarily related and are unified into a class *Tokiviricetes* (Wang et al., 2020c). Unlike tokiviruses, which have linear dsDNA genomes, clavaviruses have some of the smallest known circular dsDNA genomes (5.2 kb) (Mochizuki et al., 2010). Notably, members of the *Lipothrixviridae and Tristromaviridae* families are enveloped with a lipid membrane (Kasson et al., 2017; Rensen et al., 2016), a feature not observed in filamentous viruses infecting bacteria or eukaryotes.

Members of the *Rudiviridae* infect hyperthermophilic acidophiles of the order Sulfolobales and have non-enveloped rod-shaped virions decorated with three terminal fibers at each end (Prangishvili et al., 1999). Virions of rudiviruses are formed by multiple copies of a homodimeric MCP that wraps around the linear dsDNA, condense it into a hollow helical tube and maintain it in an A-form, as revealed by cryo-electron microscopy reconstruction of the virion of Sulfolobus rod-shaped virus 2 (SIRV2) (DiMaio et al., 2015).

Viruses from the *Lipothrixviridae* family infect hyperthermophilic acidophiles of the order Sulfolobales and possess flexible virions decorated at each terminus with thin filaments, varying in number and arrangement for different species (Arnold et al., 2000b; Bettstetter et al., 2003; Häring et al., 2005b; Vestergaard et al., 2008). Exceptionally, the termini of *Acidianus* filamentous virus 1 (AFV1) are decorated with claw-like structures which clasp upon interaction with the host cell appendages (Bettstetter et al., 2003). The virion nucleoprotein core is built from multiple copies of the two MCP, which are similar in structure to each other as well as to the single MCP of the *Rudiviridae*. Same as in rudiviruses, the structural unit is a dimer of the two MCPs, multiple copies of which bind linear dsDNA, condense it into a helical tube and maintain it in an A-form. However, in this case, the helical nucleocapsid is covered by a membrane containing lipids selectively acquired from the pool of host lipids. Remarkably, viral membranes are twice thinner than the cytoplasmic membrane of the host (Kasson et al., 2017; Liu et al., 2018a; Wang et al., 2020a). Lipid analysis has shown that AFV1 selects for flexible glycerol dibiphytanyl glycerol tetraether (GDGT) lipids, which lack cyclopentane rings and are rather flexible to bent into a unique U-shaped, horseshoe conformation (Kasson et al., 2017). However, the function of the viral envelope as well as the

mechanism of envelopment and lipid selection remain to be understood. Due to the pronounced similarities in their structural and genomic properties, which suggest relatively close evolutionary relationships, the families *Rudiviridae* and *Lipothrixiridae* have been unified in the order *Ligamenvirales* (Prangishvili and Krupovic, 2012).

Viruses belonging to the *Tristromaviridae* family infect neutrophilic hyperthermophiles of the order Thermoproteales and have flexible filamentous virions with terminal filaments attached to one or both ends, which are responsible for host recognition by binding to T4P (Rensen et al., 2016; Wang et al., 2020b). Although initially thought to be evolutionarily unrelated to other known viruses due to lack of any identifiable sequence similarity, cryo-EM structure of the Pyrobaculum filamentous virus 2 (PFV2) has uncovered an unsuspected relationship to rudiviruses and lipothrixviruses (Wang et al., 2020c). Similar to lipothrixviruses, tristromaviruses consist of an envelope and an inner helical core built from linear dsDNA and two MCPs. The genome is also condensed and maintained in an A-form conformation, suggesting that such structure is a general adaptation of crenarchaeal viruses to extreme temperatures rather than low pH, because viruses infecting both neutrophilic (Thermoproteales) and acidophilic (Sulfolobales) hosts accommodate this DNA conformation (DiMaio et al., 2015; Kasson et al., 2017; Liu et al., 2018a; Wang et al., 2020c). Remarkably, unlike in lipothrixviruses, the lipid envelope of tristromaviruses has the same thickness and lipid composition as the host membrane (Rensen et al., 2016), highlighting the variety of mechanisms employed by filamentous enveloped viruses for lipid selection.

The sole known member of the *Clavaviridae* family, Aeropyrum pernix bacilliform virus 1 (APBV1), infects neutrophilic hyperthermophiles of the order Desulfurococcales and has non-enveloped, rigid bacilliform virions with two asymmetric terminal cap structures at each end, one pointed and the other rounded. The virions are built from circular dsDNA, multiple copies of a single MCP and three minor structural proteins (Mochizuki et al., 2010). Cryo-EM structure of the virion has revealed that the DNA is tightly packed within the capsid as a left-handed superhelix and held in place by the interactions with positively charged residues lining the internal surface of the virion tube (Ptchelkine et al., 2017). A model for the assembly of APBV1 has been proposed in which the virion assembly starts when one of the cap structures recognizes and binds to the viral genome at three specific sites, forming three long loops. These gradually intertwine under the guidance of the concomitant capsid assembly to form a left-

handed superhelix. Finally, once the DNA is fully covered with the MCP, the open end of the virion is sealed by a second cap structure (Ptchelkine et al., 2017).

4.4.3 Fusiform viruses

Spindle-shaped viruses are widespread in archaea-dominated terrestrial and marine environments. Tailless spindle-shaped viruses have been isolated from hyperthermophilic crenarchaea (Arnold et al., 1999; Goodman and Stedman, 2018; Redder et al., 2009; Schleper et al., 1992; Stedman, 2011; Wiedenheft et al., 2004; Zhang et al., 2020), hyperhalophilic euryarchaea (Bath et al., 2006; Bath and Dyall-Smith, 1998) and ammonia-oxidizing thaumarchaea (Kim et al., 2019), and are classified into families *Fuselloviridae*, *Halspiviridae* and *Thaspiviridae*, respectively. Tailed spindle-shaped viruses containing one or two tail-like appendages extending from the pointed ends of the virion are classified into the family *Bicaudaviridae* (Prangishvili et al., 2018b; Prangishvili et al., 2006b). Despite the shared morphology, the two families appear to be evolutionarily unrelated and their virions are constructed from unrelated major capsid proteins, suggesting convergent evolution (Krupovic et al., 2014).

Known representatives of the *Fuselloviridae* family are isolated from terrestrial hot springs and infect hosts from the order Sulfolobales, phylum Crenarchaeota (Arnold et al., 1999; Redder et al., 2009; Schleper et al., 1992; Stedman, 2011; Zhang et al., 2020). Fuselloviruses have enveloped lemon-shaped virions decorated with terminal fibers at one of the two pointed ends. However, other family members have been reported to be more pleomorphic and elongated, with three relatively thick filaments at one pointed end (Redder et al., 2009). The virions of SSV1, the type virus of *Fuselloviridae*, consist of four virus-encoded structural proteins, VP1-VP4 (Quemin et al., 2015), with the circular dsDNA genome being packaged within the virion as a nucleoprotein complex.

Similar to fuselloviruses, the sole member of the newly created *Halspiviridae* family, His1, has enveloped spindle-shaped virions with short terminal fibers located at one of the two pointed ends (Bath and Dyall-Smith, 1998). However, despite their morphological similarities and the homology of their major capsid proteins, His1 differs from fuselloviruses by i) infecting halophilic archaea, instead of hyperthermophilic archaea, ii) having linear dsDNA, rather than circular, and iii) encoding a protein-primed family B DNA polymerase (Adriaenssens et al., 2020; Bath et al., 2006). *Thaspiviridae* is other recently created family that groups spindle-shaped viruses infecting marine ammonia-oxidizing thaumarchaea from the phylum

Thaumarchaeota. Similar to His1, thaspiviruses have linear dsDNA genomes and do not display significant sequence similarity to genomes of other archaeal viruses (Kim et al., 2019). Interestingly, two unclassified spindle-shaped viruses infecting members of the order Thermococcales, phylum Euryarchaeota, Pyrococcus abyssi virus 1 (PAV1) and Thermococcus prieurii virus 1 (TPV1), have been isolated from deep-sea hydrothermal vents (Geslin et al., 2007; Geslin et al., 2003; Gorlas et al., 2012). Despite their morphological resemblance to fuselloviruses, the genomes of PAV1 and TPV1 are not recognizably similar to those of fuselloviruses, suggesting that the two viruses are likely to be classified into a new separate family.

Representatives of the *Bicaudaviridae* family infect hosts from the order Sulfolobales (Häring et al., 2005c; Krupovic et al., 2014; Prangishvili et al., 2006b). Remarkably, virions of Acidianus two-tailed virus (ATV), the only currently classified member of the family *Bicaudaviridae*, undergoes a morphological transformation outside and independent of the host cell. ATV particles are released from host cells as tailless spindle-shaped particles, after which they develop tail-like appendages at both ends of the virion (Häring et al., 2005c; Prangishvili et al., 2006b). 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 (Prangishvili et al., 2006b).

4.4.4 Icosahedral viruses

Similar to bacteria and eukaryotes, archaea are also infected with icosahedral viruses. Tailless icosahedral viruses with an internal membrane layer are classified into the families *Turriviridae* (Rice et al., 2004), *Portogloboviridae* (Liu et al., 2017), *Sphaerolipoviridae* (Demina et al., 2017) and *Simuloviridae* (Zhang et al., 2012). The former two families include viruses infecting hyperthermophilic crenarchaea of the genus *Sulfolobus*/*Saccharolobus*, whereas the latter two families, until recently, constituted two genera of the same family, *Sphaerolipoviridae* (Pawlowski et al., 2014), and include evolutionarily related viruses of hyperhalophilic archaea. Overall virion organization is the same for viruses from all four families of tailless archaeal viruses: the icosahedral protein capsid covers an internal membrane layer, which encloses the dsDNA genome. Nevertheless, structural studies have revealed that the structures of the protein capsids are very different in members of the distinct families.

The *Portogloboviridae* family currently includes two closely related viruses, *Sulfolobus* polyhedral virus 1 (SPV1) and SPV2 (Liu et al., 2019), although only SPV1 has been isolated as a pure strain (Liu et al., 2017). SPV1 is unique among bacterial and archaeal icosahedral viruses in that instead of encapsidating a naked dsDNA molecule, its circular genome is condensed and packed in the form of a unique spherical nucleoprotein coil. The latter is surrounded by an internal lipid membrane and an outer icosahedral protein shell (Wang et al., 2019b). Cryo-EM reconstruction at near-atomic resolution revealed that the icosahedral capsid is built from a single jelly-roll (SJR) capsid protein that forms hexameric capsomers, whereas the five-fold vertices are occupied by a penton protein, which also has the SJR fold and forms pentamers. Analysis of the composition of the internal membrane of SPV1 showed that lipids are selectively acquired from the host cell, as is the case for members of the *Lipothrixviridae, Turriviridae* and *Fuselloviridae* families (Kasson et al., 2017; Liu et al., 2018a; Maaty et al., 2006; Quemin et al., 2015) (Kasson et al., 2017; Liu et al., 2017b; Maaty et al., 2006; Quemin et al., 2015), suggesting that this strategy could be a common mechanism among crenarchaeal viruses. Moreover, the virion reconstruction revealed that also in portogloboviruses, the dsDNA is in an A-form (Wang et al., 2019b), which indicates that A-DNA may be the prevalent storage form of DNA in extreme geothermal environments.

STIV and STIV2 are the only isolated members of the *Turriviridae* family. Virions of STIV are similar in their overall design to the virions of the *Portogloboviridae,* i.e. icosahedral particles with spikes protruding from the vertices of the polyhedron and an internal lipid membrane layer (Happonen et al., 2010; Rice et al., 2004; Veesler et al., 2013). However, there is major difference between these two types of virions in the ways of genome packaging. Unlike in portogloboviruses, the DNA is present in a naked state as in icosahedral bacterial viruses with an inner lipid layer from the families *Tectiviridae* and *Corticoviridae*. The structural similarity between these bacterial and archaeal viruses is extended to the structures of the MCPs, which carry the double-jelly roll (DJR) fold, and the minor capsid (penton) proteins, which have the SJR folds (Veesler et al., 2013). In recognition of the common ancestry, all viruses encoding DJR MCPs, including *Turriviridae*, were recently unified by the International Committee on Taxonomy of Viruses (ICTV) into the kingdom *Bamfordvirae* (Koonin et al., 2020).

Until recently, family *Spaherolipoviridae* included three genera, *Alphasphaerolipovirus*, *Betasphaerolipovirus* and *Gammasphaerolipovirus* (Pawlowski et al., 2014). Members of the

Alphasphaerolipovirus and *Betasphaerolipovirus* genera infect halophilic archaea and have linear and circular dsDNA genomes, respectively. By contrast, gammasphaerolipoviruses infect thermophilic bacteria (genus *Thermus*) and encapsidate circular dsDNA genomes (Jaatinen et al., 2008). Viruses from the three genera share overall architecture of the capsid, built using two SJR MCPs, which are oriented vertically with respect to the capsid surface, closely resembling the DJR MCPs of viruses in the *Bamfordvirae* (Gil-Carton et al., 2015; Rissanen et al., 2013; Zhang et al., 2012). However, in terms of gene content and sequence similarity, viruses from the three genera share very little in common (Aiewsakun et al., 2018). Accordingly, it has been recently proposed to move genera *Betasphaerolipovirus* and *Gammasphaerolipovirus* into separate families, *Simuloviridae* and *Matsushitaviridae*, respectively.

4.4.5 Spherical and pleomorphic viruses

Spherical viruses infecting Archaea can be grouped into the families *Globuloviridae* and *Ovaliviridae*. Members of the *Globuloviridae* infect neutrophilic hyperthermophiles of the order Thermoproteales, whereas the sole representative of the family *Ovaliviridae* infects a hyperthermophilic acidophilic host from the order *Sulfolobales*, both in the phylum Crenarchaeota (Ahn et al., 2006; Häring et al., 2004; Hartman et al., 2020; Wang et al., 2018a). By contrast, archaeal pleomorphic viruses with pseudo-spherical shapes of the *Pleolipoviridae* family infect haloarchaea of the phylum Euryarchaeota (Demina and Oksanen, 2020).

The spherical virions of the *Globuloviridae* are covered with a lipid-containing envelope, which encases a tightly-packed superhelical core consisting of linear dsDNA and multiple copies of the MCP. Multiple protrusions, spherical in their appearance, are found on the virion surface; however, their role remains unknown (Ahn et al., 2006; Häring et al., 2004; Hartman et al., 2020). Observation of partially disrupted virions of Pyrobaculum spherical virus (PSV) suggested that there is no regular arrangement of the superhelical nucleoprotein filament inside the virion. Exceptionally, the newly characterized Thermoproteus spherical piliferous virus 1 (TSPV1) has numerous unusual filaments decorating its surface that can extend tens of micrometres from the virion and are presumed to be involved in host interactions (Hartman et al., 2020).

Virions of the sole member of the family *Ovaliviridae*, Sulfolobus ellipsoid virus 1 (SEV1), are ellipsoid, with the nucleoprotein coil being surrounded by a lipid membrane. In contrast to

the globuloviruses, the virions of SEV1 are formed by nucleoprotein filaments that are regularly arranged (Wang et al., 2018a). The linear dsDNA is covered by multiple copies of the MCPs and condensed into a unique spherical nucleoprotein coil, which is enclosed by a membrane decorated with protruding spikes. It has been proposed that the viral DNA wraps around the longitudinal axis of the virion to form a multilayered disk-like structure with a central hole; sixteen of these structures stack together to form an unusual spool-like capsid. Interestingly, electron micrographs of infected cells showed enveloped viral capsids in the cytoplasm of the host cell, indicating that the virus acquires the lipid membrane intracellularly by an unknown mechanism (Wang et al., 2018a).

The *Pleolipoviridae* family comprises globally distributed archaeal viruses with pseudo-spherical, enveloped morphologies carrying either ssDNA or dsDNA in circular or linear topologies. The envelope of pleolipoviruses contains two major membrane proteins, whereas internal nucleoproteins are absent (Pietilä et al., 2016). The currently characterized pleolipoviruses have been isolated on halophilic archaeal strains of the class *Halobacteria,* some of which displaying a wide range of tolerance to NaCl concentrations (Demina et al., 2016)**.** Based on the gene content and whole-genome sequence identity of their members**,** *Pleolipoviridae* is divided into three genera: *Alphapleolipovirus*, *Betapleolipovirus* and *Gammapleolipovirus* (Bamford et al., 2017). Alphapleolipoviruses encode a putative rolling-circle replication initiation protein, whereas betapleolipoviruses genomes contain a protein with a predicted winged-helix DNA-binding domain and gammapleolipoviruses encode a putative protein-primed DNA polymerase (Demina and Oksanen, 2020; Pietilä et al., 2016).

4.4.6 Head-tailed viruses

Archaea are hosts to head-tailed viruses, which closely resemble tailed bacteriophages of the order *Caudovirales* (realm *Duplodnaviria*). Archaeal head-tailed viruses with all three prototypical tail morphologies have been isolated: long, contractile tails (typical of members of the *Myoviridae*), short tails (*Podoviridae*) and long, non-contractile tails (*Siphoviridae*) (Pietilä et al., 2014; Sencilo and Roine, 2014). Similar to their bacterial relatives, virions of archaeal head-tailed viruses comprise an icosahedral capsid (head) and a helical tail containing tail fibers, which play a role in cell recognition. All head-tailed viruses encapsidate linear dsDNA genomes, which in archaeal viruses can vary in length from 26 kb to 144 kb. Archaeal head-tailed viruses have been isolated from extreme halophiles (Atanasova et al., 2016) as well as from mesophilic (Pfister et al., 1998a; Wolf et al., 2019) and marine hyperthermophilic methanogens (Thiroux et al., 2020). However, related proviruses have been detected in a wide

range of archaea belonging to different phyla, suggesting that the last common archaeal ancestor (LACA) has been already infected with head-tailed viruses (Krupovic et al., 2020).

4.5 Features of hyperthermophilic archaeal viruses

All characterized hyperthermophilic archaeal viruses possess DNA genomes, which can be either ssDNA or dsDNA, linear or circular. The average genome sizes are rather small (20-30 kb) with the clavavirus APBV1 having one of the smallest known dsDNA genomes (5.5 kb) (Mochizuki et al., 2010). Notably, however, the sole representative of the *Spiraviridae* family, ACV, possesses one of the largest genomes (24.9 kb) among currently described ssDNA genomes (Mochizuki et al., 2012). Viruses with linear genomes use different strategies to protect and replicate the genome termini, such as covalently closed hairpins, covalently attached terminal proteins and terminal inverted repeats (TIRs) (Blum et al., 2001; Peng et al., 2001; Pina et al., 2014; Prangishvili and Krupovic, 2012). Although no RNA viruses infecting archaea have been isolated, the existence of such viruses has been postulated based on metagenomic sequencing of samples from archaea-dominated acidic hot springs in the Yellowstone National Park, USA (Bolduc et al., 2012; Bolduc et al., 2015).

Along with the extraordinary diversity of morphotypes, another outstanding feature of hyperthermophilic archaeal viruses is the uniqueness of their genome contents, with a very low proportion of genes with recognizable homologues in public databases. Indeed, family-specific comparison of viral proteomes against the available sequences revealed that a fraction of ~85% of the crenarchaeal virus proteins do not have identifiable homologs when an E-value threshold of <1e-5 is used (Krupovic et al., 2018). In line with this observation, the analysis of the evolutionary relationships between all dsDNA viruses using a bipartite network approach, which traces connections between viral genomes through shared gene families, revealed that hyperthermophilic archaeal viruses are largely disconnected from the global dsDNA virosphere (Iranzo et al., 2016b) (Figure 5). Moreover, the families themselves are largely disconnected from each other and share just a small number of common genes, suggesting that they have evolved independently of one another (Iranzo et al., 2016a; Iranzo et al., 2016b). Most of the common genes shared between the families are involved in transcriptional regulation and glycosylation and could be independently acquired by the viruses from their respective hosts (Krupovic et al., 2018).

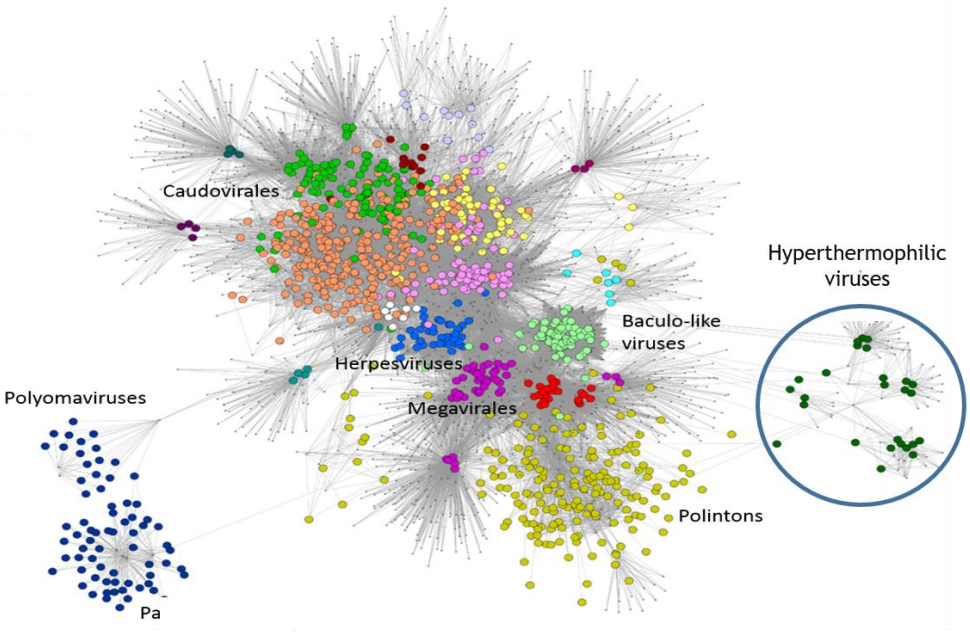


Figure 5. The dsDNA virus world as a bipartite network. Nodes corresponding to genomes are depicted as larger circles, and nodes corresponding to core gene families are depicted as dots. An edge is drawn whenever a genome harbors a representative of a core gene family. The modular structure of the network is highlighted by coloring genome nodes according to the module to which they belong. Image modified and reproduced with permission from Iranzo, J., Krupovic, M., and Koonin, E.V. (2016). The Double-Stranded DNA Virosphere as a Modular Hierarchical Network of Gene Sharing. mBio 7(4) e00978-16.

It is worth noting that reconstruction of the evolutionary history of the individual virus families is hindered by scarcity of known members in the families. Only one or two members have been described in the families *Ampullaviridae*, *Bicaudaviridae*, *Clavaviridae*, *Spiraviridae*, *Guttaviridae*, *Tristomaviridae*, *Portogoboviridae*, and *Turriviridae*. Even in the most populated families, *Fuselloviridae*, *Rudiviridae*, and *Lipothrixviridae* the number of known members is rather limited precluding tracing any general evolutionary trends (Iranzo et al., 2016a; Iranzo et al., 2016b).

4.6 Virus-host interactions

Although the knowledge of virus-host interactions in archaea remain highly fragmented, the increasing number of genetic tools developed for archaea and their viruses as well as application of advanced structural and functional genomics efforts have yielded valuable information on certain aspects of virus-host interactions. Some of these mechanisms are shared with bacterial and/or eukaryotic viruses, whereas others are unique to archaeal viruses.

The virus infection cycle can be subdivided into several stages. The infection starts with the recognition and binding to specific host receptors on the host cell surface which leads to the delivery of the viral genetic material into the cell interior (Hartman et al., 2019; Hong et al., 2015; Quemin et al., 2013). Following the entry, many viruses hijack the host replication, transcription and translation machineries to produce multiple copies of the virus progeny (Gardner et al., 2014; Martínez-Alvarez et al., 2016; Martínez-Alvarez et al., 2017; Pina et al., 2014). For some enveloped viruses, the morphogenesis and egress are concomitant, whereas in the case of non-enveloped viruses the virion assembly typically precedes the release. Thus far, two different egress strategies have been elucidated for archaeal viruses: a budding mechanism similar to that of some eukaryotic enveloped viruses and a unique lytic mechanism employed by certain crenarchaeal viruses which involves the formation of pyramidal structures on the host cell surface (Bize et al., 2009a; Brumfield et al., 2009; Quemin et al., 2016; Wang et al., 2018a). Similar to bacteriophages and certain eukaryotic viruses, some archaeal viruses can undergo lysogenic life cycle, whereby the virus genome integrates into the host chromosome as a provirus or remains as an episomal element until the cell host is exposed to certain stimuli or stress conditions that induce viral replication (Liu et al., 2015; Mochizuki et al., 2011; Muskhelishvili et al., 1993; Prangishvili et al., 2006b; Schleper et al., 1992; Zhan et al., 2015; Zhang et al., 2012). In addition, some viruses can establish a carrier state, whereby the virus is stably maintained in a fraction of the cellular population, without causing cell lysis with the remaining cells being transiently resistant to the viral infection (Bettstetter et al., 2003; Papathanasiou et al., 2019). Archaea and their viruses have evolved defense and counter-defense mechanisms to survive, among which CRISPR-Cas system and anti-CRISPR (Acr) proteins have been the most studied (Athukoralage et al., 2020; Bhoobalan-Chitty et al., 2019; Fusco et al., 2015a; He et al., 2018; Koonin et al., 2017; León-Sobrino et al., 2016; Quax et al., 2013).

4.6.1 Virus entry

The infection cycle begins with the recognition of a suitable host through specific interactions between a receptor-binding protein exposed on the virion and a receptor located on the surface of the host cell. Once the viral protein successfully binds to the host receptor, the virus particle typically undergoes a conformational change that results in the delivery of the genetic material into the cytoplasm of the host cell (Poranen et al., 2002). For the well-studied viruses infecting bacteria, different cell surface structures have been recognized to be targeted by viruses, including pili, flagella, peptidoglycan, lipopolysaccharide and integral membrane proteins

(Davison et al., 2005; Gaidelyte et al., 2006; Guerrero-Ferreira et al., 2011; Marti et al., 2013; Romantschuk and Bamford, 1985; Shin et al., 2012; Xia et al., 2011). Viral attachment and entry have been poorly studied in archaea and only recent efforts have provided the first insights into the entry mechanisms and possible receptors of hyperthermophilic and hyperhalophilic archaeal viruses (Hartman et al., 2019; Kukkaro and Bamford, 2009; Quemin et al., 2013; Rowland et al., 2020; Wang et al., 2019a).

4.6.1.1 Interaction with cellular appendages

Similar to bacteria, electron microscopy observations have showed that many filamentous viruses bind to pili. For instance, termini of the filamentous virions of members of the *Lipothrixviridae* family are decorated with diverse structures, which have been proposed to play a role in the viral attachment to the host cell, specifically to pilus-like appendages (Arnold et al., 2000a; Bettstetter et al., 2003; Häring et al., 2005b).

The interaction of SIRV2, the prototype virus of the family *Rudiviridae*, with cellular appendages has been studied in more detail. Both ends of the rod-shaped SIRV2 virions contain three terminal fibers which specifically interact with the pili abundantly present on the surface of *Sulfolobus islandicus* LAL14/1 cells (Figure 6). Notably, when the pili were detached from the cells, the virus interacted nearly exclusively with the tips of the pili, whereas in the context of metabolizing cells, the virions were observed both at the tips and on the sides of the pili, suggesting that virions move along the pili toward the cell surface. However, the energy source for this movement remains unclear. Once at the cell surface, the SIRV2 virions appear to disassemble, presumably as a consequence of the delivery of the viral DNA into the host cytoplasm (Quemin et al., 2013). Similarly, rudivirus SSRV1 and tristromavirus PFV2 have also been shown to interact with the type IV pili of *Saccharolobus solfataricus* and *Pyrobaculum arsenaticum*, respectively (Wang et al., 2020b). Deletion of the orthologous pilin genes, *pilA1* and *pilA2*, in *S. islandicus* M.16.4 resulted in loss of T4P and resistance to rudivirus SIRV8, a close relative of SIRV2, which could no longer adsorb to the host cell (Rowland et al., 2020). These observations further reinforce the critical role of pili during the early stages of archaeal virus infection.

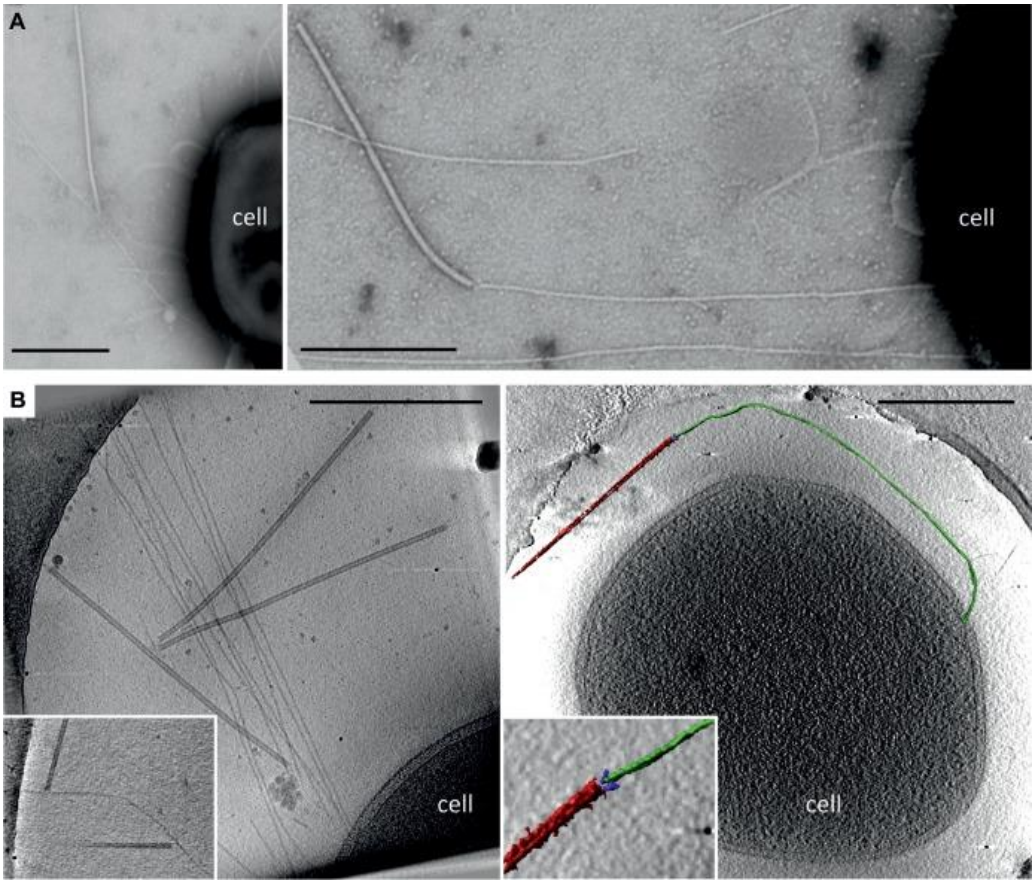


Figure 6. Electron micrographs of SIRV2 interaction with S. islandicus LAL14/1 cells. Samples were collected 1 min postinfection and negatively stained for TEM (A) or plunge-frozen for electron cryotomography (cryo-ET) (B). The virions interact both at the filament tips (right panels) and along the length of the filaments (left panels). The inset in the lower left panel depicts two virions bound to the sides of a single filament. The lower right panel shows a segmented tomographic volume of the SIRV2 virion (red) attached to the tip of an *S. islandicus* filament (green). The three terminal virion fibers that appear to mediate the interaction are shown in blue (the inset depicts a magnified view of the interaction between the virion fibers and the tip of the filament). Scale bars, 500 nm. Image reproduced with permission from Quemin, E.R., Lucas, S., Daum, B., Quax, T.E., Kühlbrandt, W., Forterre, P., Albers, S.V., Prangishvili, D., and Krupovic, M. (2013). First insights into the entry process of hyperthermophilic archaeal viruses. J Virol 87, 13379-13385.

Attachment to pili-like filaments was also confirmed for Sulfolobus turreted icosahedral virus (STIV), the type member of the *Turriviridae* family, which specifically recognizes unidentified pili-like filaments of its host, *Saccharolobus solfataricus* (Hartman et al., 2019). The icosahedral STIV virions are decorated with turret-like protrusions at each of the fivefold vertexes (Veesler et al., 2013). A three-dimensional reconstruction of the STIV-pilus interaction using cryo-electron tomography displayed that the turrets physically interact with the *S. solfataricus* pilus (Hartman et al., 2019). Furthermore, tomographic reconstruction and sub-tomogram averaging unequivocally showed that pilus recognition occurs at the cleft between the second and third jelly-roll domains of the pentameric turret protein C381. Notably, structurally similar jelly-roll domains have been implicated in receptor recognition in highly diverse viruses, including P22-like head-tailed phages (*Caudovirales*), tectiviruses and

adenoviruses (Bewley et al., 1999; Bhardwaj et al., 2011; Hartman et al., 2019; Leavitt et al., 2013; Roelvink et al., 1999). Despite the molecular insights into the receptor binding, as in the case of SIRV2, it remains unclear how (and whether) the virus moves along the filaments and how circular virus genome is delivered into the host cytoplasm. Notably, adsorption to the pili of both SIRV2 and STIV was shown to occur exclusively at physiologically-relevant, high temperatures (Hartman et al., 2019; Quemin et al., 2013).

4.6.1.2 Interaction with cell-surface

The relative simplicity of the archaeal cell envelope likely underlies the mechanisms of virus entry and egress. Spindle-shaped viruses of the *Fuselloviridae* family are often attached to extracellular membrane vesicles or cellular fragments, suggesting that viral receptors are exposed on the host cell surface, with S-layer itself being a prime suspect. The spindle-shaped virion of the model fusellovirus SSV1 at one of the two pointed ends of the virion contains short terminal fibers, likely composed of protein VP4, which are postulated to mediate host recognition (Quemin et al., 2015; Stedman et al., 2015). Despite the observations suggesting possible interactions between SSV1 terminal fibers and the host cell surface, two recent studies on the *Sulfolobus* S-layer produced somewhat contradictory results regarding the role of S-layer in fusellovirus infection. On the one hand, the essential role of the S-layer in SSV1 infection appears to be supported by the finding that *Saccharolobus solfataricus* cells in which *slaB* encoding the membrane-anchoring S-layer protein was downregulated by a CRISPR-based silencing technology are less susceptible to SSV1 infection (Zink et al., 2019). On the other hand, cells in which both genes encoding the S-layer proteins SlaA and SlaB were deleted remained susceptible to infection with SSV9, a close relative of SSV1, suggesting that S-layer is not essential for either adsorption or infection by SSV9 (Rowland et al., 2020). These discrepant results call for additional studies focusing on the entry mechanism of fuselloviruses.

The entry process has also been explored for viruses infecting halophilic archaea. Insights have been obtained into the DNA ejection process in the halophilic spindle-shaped virus His1, the type member of the *Halspiviridae* family. Similar to fuselloviruses, at one of the pointed ends, His1 virion carries a putative receptor-binding module consisting of a central hub and six spikes (Hong et al., 2015). *In vitro* analysis demonstrated that His1 DNA ejection is unidirectional, occurs at a rate comparable to that of bacteriophage λ and is dependent on external osmotic pressure (Hanhijärvi et al., 2013). Notably, *in vitro*, the His1 DNA ejection was only partial,

suggesting that cellular factors are required for completion of the nucleic acid transfer. Interestingly, upon DNA ejection, the lemon-shaped virions transform into empty tubes, indicating that capsid proteins are capable of undergoing substantial quaternary structural changes (Hong et al., 2015).

Members of the family *Pleolipoviridae* have pleomorphic virions, which resemble membrane vesicles decorated with protruding spikes that in all likelihood participate in host attachment and membrane fusion processes (El Omari et al., 2019). It has been suggested that upon binding to the receptor, the spike protein VP5 of Halorubrum pleomorphic virus 6 (HRPV-6) undergoes conformation change and drives fusion of the viral membrane with the host cytoplasmic membrane. Structural analysis of HRPV-6 virion by cryo-electron microscopy and crystallography revealed that HRPV-6 VP5 has a unique V-shaped fold that is unrelated to the previously reported class I–III viral fusion protein (El Omari et al., 2019).

The host recognition by archaeal head-tailed viruses appears to be mediated by tail fiber proteins, resembling the initial interactions of the bacteriophages of the order *Caudovirales*. Interestingly, the head-tailed dsDNA virus φCh1 infecting the haloalkaliphilic archaeon *Natrialba magadii* encodes a phase variation system (Klein et al., 2012). The system consists of an invertible region including a site-specific recombinase of the tyrosine recombinase superfamily interspersed between convergently oriented ORFs 34 and 36, which encode viral tail fiber proteins. Recombination in this region leads to an exchange of the gene fragments encoding the carboxy-termini of the tail fiber proteins, thereby creating a set of heterogeneous gp34 and gp36 proteins with distinct C-termini. Notably, only ORF34 is expressed during the virus infection indicating that fiber proteins of φCh1 are produced exclusively from this ORF. Binding assays showed that only one type of the heterogeneous tail fibers encoded by ORF34 interacts with *N. magadii,* suggesting that other tail fiber variants enable adsorption to other host strains. Galactose moieties present on the cell surface were implicated in host recognition by φCh1 (Klein et al., 2012). Thus, as reported for bacterial viruses, the generation of two types of tail fiber proteins with distinct binding features might expand the φCh1 host range to quickly respond to habitat changes.

4.6.1.3 Kinetics of virus entry

There is a wide variation in terms of adsorption kinetics among archaeal viruses. Generally, hyperthermophilic archaeal viruses tend to display rapid adsorption rates, whereas halophilic viruses are notoriously slow. For instance, it has been shown that SIRV2 adsorbs very rapidly, with ~80% of the virions bound to the host cell within the first 30 seconds of infection (Quemin et al., 2013). Similarly, ~50% of spindle-shaped SMV1 virions were attached to the cells within 1 minute post-infection (p. i.) (Uldahl et al., 2016). By contrast, only 30% of the salterprovirus His1 and siphovirus HHTV-1 virions adsorb to their host in 3 h (Bath and Dyall-Smith, 1998; Kukkaro and Bamford, 2009). Similarly, the adsorption of the tailless icosahedral haloarchaeal virus HCIV-1 is rather efficient but slow, with the 80% of the particles adsorbed to cells after 4-5 hours p.i. While the high adsorption rates of hyperthermophilic viruses are thought to minimize the time they are exposed to harsh extracellular conditions, including acidic pH and high temperatures (Quemin et al., 2013), the low adsorption rates of haloarchaeal viruses are hypothesized to reflect an evolutionary adaptation to the long generation time of the hosts, whereby rapid adsorption might deplete the host population (Kukkaro and Bamford, 2009). Alternatively, it has been proposed that the changing salinity conditions of natural hypersaline environments may have favored the slow-adsorbing viruses that bind efficiently only under a specific range of salt concentrations (Porter et al., 2013).

4.6.2 Genome replication

Very few studies have investigated experimentally the mechanisms of archaeal virus genome replication. In most cases, the mode of genome replication has been inferred from recognizable virus-encoded replication-associated genes, including rolling-circle replication initiation endonucleases (RCRE), DNA polymerases, replicative helicases and other components of the host replisome. Members of the families *Ampullaviridae, Thaspiviridae, Halspiviridae* and *Ovaliviridae* as well as members of the genus *Gammapleolipovirus* (*Pleolipoviridae*) have linear dsDNA genomes and encode protein-primed family B DNA polymerases (Bath et al., 2006; Kim et al., 2019; Peng et al., 2007; Wang et al., 2018a). By contrast, certain head-tailed haloarchaeal viruses encode RNA-primed DNA polymerases closely related to the corresponding proteins of their hosts (Sencilo and Roine, 2014). Generally, there is a correlation between the viral genome size and the completeness of the viral DNA replication machinery. Viruses with small to medium-sized genomes (5-50 kb) commonly encode only essential components of the replication machinery, with the rest of the components being

recruited from the host, whereas viruses with large genomes (>100 kb) appear to depend minimally on the host replisome, encoding nearly complete DNA replication machineries including DNA polymerases, proliferating cell nuclear antigen (PCNA), primases, replicative helicases and homologs of the archaeal Orc1/Cdc6 replication initiators (Kazlauskas et al., 2016). Global analysis of dsDNA viral genomes demonstrated that replicative helicases are the most common replication proteins (75% of the genomes) in the dsDNA virus world. Indeed, some archaeal viruses with medium-sized genomes encode replicative minichromosome maintenance (MCM) helicases. Interestingly, phylogenetic analysis indicates that the *mcm* genes found in haloarchaeal, methanosarcinal, and methanococcal (pro)viruses were acquired from their respective hosts many times independently (Krupovic et al., 2010; Krupovič et al., 2010).

The RCRE are encoded by members of the *Pleolipoviridae* (genus *Alphapleolipovirus*) (Pietilä et al., 2016) and *Sphaerolipoviridae* (genus *Betasphaerolipovirus*) (Wang et al., 2016). However, experimental evidence for rolling circle replication, showing the presence of ssDNA replicative intermediates, has been obtained only in the case of sphaerolipovirus SNJ1 (Wang et al., 2018c). It has been shown that the SNJ1 RepA protein, an RCRE of the HUH superfamily, is indispensable for the genome replication of SNJ1 (Wang et al., 2016). Closest homologs of SNJ1 RepA are encoded by plasmids of halophilic archaea, while more divergent homologues (13-19% identity) are also identified in euryarchaea from the orders *Methanosarcinales* and *Thermoplasmatales*, as well as in ammonia-oxidizing archaea of the phylum *Thaumarchaeota*, underpinning a wide distribution of the SNJ1 RepA-like proteins across mobile genetic elements from diverse archaeal lineages. Interestingly, SNJ1 RepA also shares similarity with bacterial transposases of the IS*91* family, which transpose via a rolling-circle-like mechanism, highlighting evolutionary connections between viruses, plasmids and transposons (Wang et al., 2018c). A divergent RCRE has been also identified in rod-shaped viruses of the *Rudiviridae* family (Oke et al., 2011). The protein has been structurally characterized and studied biochemically. *In vitro*, the Rep protein displays the expected nicking activity. However, its role in the viral genome replication remains unclear, because it does not seem to be expressed during the virus life cycle, at least, under laboratory conditions (Oke et al., 2011; Okutan et al., 2013; Quax et al., 2013).

Although genome replication has been most extensively studied for rudiviruses, the actual mechanism remains enigmatic. Members of the *Rudiviridae*, such as SIRV2, encode several proteins proposed to be involved in DNA replication, recombination and repair. Those proteins include the above mentioned Rep (Oke et al., 2011), a ssDNA-binding protein with a unique fold (Guo et al., 2015), a ssDNA-annealing ATPase (Guo et al., 2015), a Cas4-like ssDNA nuclease (Gardner et al., 2011), a dUTPase (Prangishvili et al., 1998) and a Holliday junction resolvase (Birkenbihl et al., 2001). Furthermore, yeast two-hybrid analysis showed that five SIRV2 proteins interact with the host DNA sliding clamp PCNA, known as a "molecular toolbelt" which interacts with multiple components of the host replisome (Gardner et al., 2014; Pan et al., 2011). Presumably, SIRV2 recruits the host replication machinery for the assembly of the replisome on the viral DNA template. An immunofluorescence study of SIRV2 infected cells demonstrated that viral DNA synthesis is confined to a focus near the periphery of the cell. The study also provided evidence that the viral ssDNA-binding protein (gp17) as well as host PCNA and DNA polymerase I (Dpo1) are recruited to the site of viral DNA synthesis, confirming their essential role in the viral genome replication (Martínez-Alvarez et al., 2017). Given that SIRV2 does not encode an identifiable DNA polymerase, the host polymerase Dpo1 found at the viral DNA synthesis site is likely to be involved in the replication of SIRV2 genome. An exceedingly complex model of SIRV2 replication has been proposed, whereby the virus employs a combination of strand-displacement, rolling-circle and strand-coupled replication mechanisms, which yields highly branched intermediates of about >1200 kb (~34 viral genome units) with unusual 'brush-like' structures (Martínez-Alvarez et al., 2016; Martínez-Alvarez et al., 2017). However, it remains unclear how these three replication mechanisms are coordinated and whether all of them are essential for SIRV2 genome replication. In addition, the particular role of the host and viral proteins involved in the orchestration of viral DNA replication remains obscure.

It should be noted, however, that most archaea-specific viruses do not encode identifiable DNA replication proteins, implying that these viruses either depend on the host DNA replication machinery or employ novel uncharacterized mechanisms for DNA replication. For instance, the genome replication of the lipothixvirus AFV1 has been suggested to start by the formation of a D-loop and progress by the strand displacement replication mechanism, whereas termination relies on recombination events through the formation of terminal-loop-like structures (Pina et al., 2014). However, the genes involved in this unique mechanism of

replication are currently unknown. Similarly, the conspicuous absence of recognizable genes encoding type B DNA polymerases in viral genomes with linear dsDNA and terminal proteins, as in the case of the haloarchaeal sphaerolipoviruses PH1, SH1 and HHIV-2, suggests novel mechanisms of genome replication. A model resembling that of the *Streptomyces* linear plasmids has been proposed, whereby the cellular polymerase is employed for the viral genome replication and the terminal proteins are involved in the end patching of the DNA after replication (Porter et al., 2013; Yang et al., 2006; Yang et al., 2002).

4.6.3 Genome integration

Many archaeal viruses are temperate and can undergo a lysogenic pathway in which they coexist with the host cell as proviruses. In most cases, the temperate viruses integrate their genomes into the host chromosome by the activity of viral site-specific integrases. However, sometimes proviruses can exist as circular extrachromosomal plasmids. For instance, the genomes of the euryarchaeal head-tailed virus φCh1 (family *Myoviridae*), pleomorphic virus SNJ2 (family *Pleolipoviridae*) and the crenarchaeal spindle-shaped SSV1 (family *Fuselloviridae*) integrate into the chromosome of *Natrialba magadii*, *Natrinema* sp. J7-1 and *Sulfolobus shibatae* cells, respectively, whereas the genomes of the sphaerolipovirus SNJ1 and myovirus φH are stably maintained in a non-integrated circular form in *Natrinema* sp. J7-1 and *Halobacterium halobium* cells, respectively (Iro et al., 2007; Ken and Hackett, 1991; Liu et al., 2015; Luk et al., 2014; Wang et al., 2018c; Zhang et al., 2012). The induction of the provirus replication typically occurs as a response to stressful conditions, such as DNA damage (e.g., by UV light or mitomycin C), temperature shock and shift from aerobic to anaerobic conditions. Depending on the virion release mechanism, the provirus induction can lead either to cell lysis, as for the head-tailed archaeal viruses and SNJ1, or to growth retardation, as observed for the fusellovirus SSV1 and pleolipovirus SNJ2 (Chen et al., 2020; Liu et al., 2015; Mei et al., 2015; Pietilä et al., 2013a; Pietilä et al., 2013b; Schleper et al., 1992; Torsvik and Dundas, 1974; Zhang et al., 2012).

Site-specific integrases of archaeal viruses, all belonging to the tyrosine recombinase superfamily, have been experimentally studied for two groups of viruses: fuselloviruses SSV1 (Muskhelishvili et al., 1993) and SSV2 (Zhan et al., 2015), and pleolipovirus SNJ2 (Wang et al., 2018b). The integrase of SNJ2 catalyzes homologous recombination between the viral attachment site located next to the integrase gene and the homologous site located within one

of the tRNA genes on the host chromosome, leading to merger of the two genomes (Wang et al., 2018b). Interestingly, integrases of fuselloviruses are unique in that the attachment site lies within the integrase gene itself and homologous recombination leads to disruption of the integrase gene into two fragments which flank the integrated provirus (Muskhelishvili et al., 1993; Serre et al., 2002; Zhan et al., 2015). Nevertheless, phylogenetic analysis has shown that SSV-like integrases have evolved from the more typical SNJ2-like integrases (Wang et al., 2018b).

4.6.4 Transcription

Genomic analysis of archaeal viruses has shown a lack of genes encoding identifiable RNA polymerases as well as the presence of archaeal promoter elements, indicating a strong dependence on the host transcriptional machinery. However, the existence of putative transcription regulators in viral genomes suggests a viral-driven modulation of the host transcription to redirect it towards the expression of specific viral genes in an efficient and temporal manner (Sheppard and Werner, 2017). Bioinformatics, biochemical, structural and transcriptomic efforts have provided valuable insights into the role of viral transcription regulators as well as into the modulation of the virus/host gene expression during the course of infection.

4.6.4.1 Transcription regulators

A number of archaeal virus ORFs have been predicted to contain DNA-binding motifs typical of cellular transcription factors, with the majority displaying ribbon-helix-helix (RHH), winged helix-turn-helix (wHTH) or zinc (Zn) finger motifs. Notably, RHH and wHTH transcription factor sequences encoded by archaea and their viruses are bacterial-like, whereas archaeal Zn-finger domains resemble those of eukaryotes (Guillière et al., 2013; Guillière et al., 2009; Prangishvili et al., 2006a). The typically small size of such putative transcription factors makes them amenable to structural studies and during the past years some have been experimentally characterized and their structures have been solved. For instance, The p06 protein of lipothrixvirus AFV1 is the first experimentally characterized archaeal Zn-finger protein. The protein preferentially binds to GC-rich DNA regions and displays a classical Zn-finger motif, albeit with an atypical substitution of one of the residues responsible for chelation of the Zn ion (Guillière et al., 2013).

Most of the functionally characterized transcription regulators encoded by archaeal viruses are repressors, often also autoregulating their own expression. For example, the *Sulfolobus* virus transcription regulator (SvtR), encoded by rudivirus SIRV1, contains an RHH fold and forms a homodimer that resembles bacterial repressors CopG, NikR and MetJ. Functional characterization of SvtR showed that its high-affinity binding-site corresponds to the promoter region of the *gp30* gene, which encodes a structural protein involved in the assembly of terminal fibers. Hence, it has been suggested that the primary function of SvtR is to prevent the premature expression of the structural protein encoded by *gp30*. Furthermore, SvtR binds to the promoter of its own gene (*gp08*), downregulating its own expression (Guillière et al., 2009). The *Acidianus* virus transcription regulator (AvtR), encoded by lipothrixvirus AFV6, is highly conserved in the *Betalipothrixvirus* genus and consists of two RHH motifs connected by a linker (Peixeiro et al., 2013). AvtR represses the expression of its own gene (*gp29*), but is able to both activate and repress the expression of the *gp30* gene in a concentration-dependent manner. Although in the case of both genes, AvtR binding sites are distant from the TATA boxes, DNase I footprinting assays showed that AvtR protects a region of approximately 100 nucleotides between the divergently oriented *gp29* and *gp30* promoters. This finding suggests that AvtR regulation depends on protein oligomerization on the DNA template. It has been proposed that AvtR binds to the initial binding site with high affinity and its oligomerization along the viral DNA induces cooperative binding to degenerate secondary sites with weaker affinity (Peixeiro et al., 2013).

The transcriptional regulator F55 encoded by the fusellovirus SSV1 is one of the more extensively studied transcription factors encoded by archaeal viruses. Upon exposure to UV light, F55 plays a crucial role in the transition from the carrier to the induced state of SSV1. F55 is a dimer with an RHH DNA-binding motif. F55 recognizes tandem repeat sequences located within the promoters of the immediate early-induced transcripts T5, T6 and Tind as well as in its own promoter (Fusco et al., 2013). Notably, binding of F55 to the target sequences leads to repression of the gene expression, whereas its dissociation causes activation of the transcription upon exposure of the SSV1 infected cells to UV light (Fusco et al., 2015b). A recent study using a variant of electrophoretic mobility shift assay (EMSA) coupled to mass spectrometry revealed that host RadA recombinase is associated with F55 when bound to the specific promoter sequences, forming a RadA-F55-dsDNA complex (Fusco et al., 2020). The RadA recombinase belongs to the RecA/RadA/Rad51 protein superfamily and promotes DNA

repair and recombination in hyperthermophilic archaea. Therefore, it has been proposed that RadA is a molecular sensor of the SSV1 host DNA damage, analogous to the role of bacterial RecA protein in the life cycle of phage λ. According to the proposed model, the exposure of infected cells to UV light causes massive DNA degradation that leads to accumulation of ssDNA regions. The RadA protein, which is coupled to F55 on the viral dsDNA genome, is recruited to the ssDNA regions and is progressively released from the RadA-F55-dsDNA complex, causing dissociation of F55 from the viral target sequences, thereby leading to activation of transcription of the immediate early transcripts. Notably, however, *in vitro* addition of ssDNA to the stable RadA-F55-dsDNA complex did not result in the release of the transcriptional block, prompting further studies to understand the specific role of RadA in the activation of the SSV1 transcription (Fusco et al., 2020).

The regulation of lysogeny has been also investigated for the haloarchaeal icosahedral virus SNJ1. The product of SNJ1 ORF4 has been recently suggested to controls the lysis-lysogeny switch (Chen et al., 2020). The expression level of ORF4 in the host cell appears to play a crucial role in the repression of genes responsible for the lytic pathway. Consistently, in the absence of ORF4, SNJ1 produces clear plaques, whereas in its presence, the plaques are turbid. In addition, ORF4 has been shown to play a key role in conferring immunity to the host cell against subsequent infections by SNJ1 (homotypic superinfection immunity), presumably by repressing the genome replication of the superinfecting viruses. ORF4 is conserved in other SNJ1-like proviruses, suggesting that the mechanisms behind the lysis-lysogeny switch and superinfection immunity is conserved in SNJ1-like proviruses (Chen et al., 2020).

Bicaudavirus ATV encodes an atypical transcriptional regulator, ORF145, which has apparently evolved from the major capsid protein, with which it shares significant sequence similarity. Similar to the major capsid protein, ORF145 is abundantly present in ATV virions. The protein binds to the host RNA polymerase (RNAP) with nanomolar affinity and inactivates it in a reversible manner via an allosteric mechanism. ORF145, renamed as RNAP inhibitory protein (RIP), binds apically to the DNA-binding channel of the RNAP and locks in a fixed position the RNAP clamp domain that typically switches between open and closed conformations during the transcription cycle. The high-affinity complex formed between RIP and the host RNAP inhibits the formation of transcription pre-initiation complexes and represses abortive and productive initiation as well as transcription elongation. Interestingly,

RIP efficiently hinders the transcription directed from both host and viral promoters, suggesting a global inhibitory activity. Although the biological significance of a global transcriptional shutdown is unclear, it has been proposed that global repression prevents the host defense response. Notably, presence of RIP in the ATV virions suggests that host transcription might be repressed during the very early stages of the viral infection, potentially, to prevent the activation of the host type III-B CRISPR system (Sheppard et al., 2016; Sheppard and Werner, 2017).

4.6.4.2 Transcriptional control

Whole-genome transcriptomic analyses using DNA microarray and RNAseq technologies have provided novel insights into virus-host interactions in crenarchaea. For example, SSV1 infected cells exhibit a tight chronological regulation of viral gene expression upon UV irradiation highly reminiscent of the strategy used by many bacterial and eukaryotic viruses. Thus, SSV1 temporal control leads to three clearly distinguishable sets of genes: immediate early, early and late genes (Fröls et al., 2007). Transcriptomic analysis of the closely related virus SSV2 also displays a temporal regulation of gene expression which occurs in a distributive fashion with expressed genes not being adjacently located (Ren et al., 2013). Contrary to SSV1 and SSV2, analysis of gene expression of the lytic viruses SIRV2 and STIV display little temporal regulation of the viral gene transcription, with SIRV2 starting the transcription at multiple sites in the genome (Kessler et al., 2004; Ortmann et al., 2008). Moreover, an RNAseq analysis on the two-tailed virus STVS2 shows that transcription of the majority of viral genes starts shortly after infection and increases throughout the infection cycle (León-Sobrino et al., 2016).

Studies on the host gene expression during virus infection showed that host response differs among archaeal viruses. While a small proportion of host genes was differentially expressed during the induction of SSV1 lysogens, the expression of several host genes encoding DNA replication, repair and transcription proteins was increased upon the infection with SSV2 (Fröls et al., 2007; Ren et al., 2013). Likewise, the transcription of more than one third of the *Sulfolobus* genes was differentially regulated as consequence of the SIRV2 infection. Notably, transcriptomic analysis of SIRV2 and STSV2 display a strong upregulation of antiviral defense genes including those for CRISPR-Cas and toxin-antitoxin systems. Consistently, infected *Sulfolobus islandicus* cells are actively undergoing CRISPR spacer acquisition from STSV2 (León-Sobrino et al., 2016; Quax et al., 2013). These findings indicate that host response is virus-dependent since some viruses trigger a massive host response, whereas others, such as

SSV1, are nearly unnoticed by the host. Interestingly, the same set of host genes can be up- and down-regulated by two different viruses, which suggests that distinct groups of host functions are required for the propagation of different archaeal viruses. For instance, the crenarchaeal cell division operon (*cdv*), homologous to the eukaryotic ESCRT machinery, is downregulated upon SIRV2 and STSV2 infection but upregulated during the STIV infection (Krupovic et al., 2018; León-Sobrino et al., 2016; Ortmann et al., 2008; Quax et al., 2013).

Analysis of the host gene expression during virus infection also revealed significant upregulation of the host genes implicated in DNA replication and repair. The infection with fuselloviruses SSV1 and SSV2 leads to upregulation of genes encoding reverse gyrases, two subunits of the topoisomerase VI and the replication initiation protein Orc1/Cdc6, whereas the expression of the genes encoding the MCM helicase, PCNA and Dpo1 were boosted only in the case of SSV2 (Fröls et al., 2007; Ren et al., 2013). Similarly, reverse gyrase and Orc1/Cdc6-like genes were upregulated during the STIV infection, whereas genes implicated in energy production and metabolism were downregulated (Ortmann et al., 2008). The STSV2 infection leads to increased transcription of Holliday junction resolvase, DNA topoisomerase I and several DNA repair proteins, albeit the expression of the reverse gyrase, unlike for other crenarchaeal viruses, was downregulated (León-Sobrino et al., 2016). Differential regulation of genes encoding proteins involved in transcription suggests that viral transcription also relies on the host cell machinery. For instance, infection with SSV2, SIRV2 and STIV resulted in the upregulation of transcription initiation factor IIB, different subunits of the DNA-directed RNA polymerase and several transcriptional regulators (Okutan et al., 2013; Ortmann et al., 2008; Ren et al., 2013).

4.6.5 Virion egress

The last stage of the viral infection cycle corresponds to the egress of new virions from the host cell. So far, release mechanisms have been studied in detail only for a handful of archaeal viruses. Generally, there are two types of archaeal viruses – those that disrupt the host cell upon virion egress (i.e., lytic viruses) and those that do not (i.e., non-lytic viruses).

4.6.5.1 Cell membrane disruption

Lytic archaeal viruses have evolved several unrelated mechanisms for disruption of the cell envelope. The most extensively characterized mechanism involves formation of large pyramidal portals, dubbed virus-associated pyramids (VAP) (Figure 7). VAPs develop on the surface of the infected cells, protrude through the S-layer and open outward as flower petals,

generating apertures through which the mature virions exit from the cell. This mechanism is used by viruses from at least three unrelated families, namely, *Rudiviridae*, *Turriviridae* and *Ovaliviridae*. VAPs of the rudivirus SIRV2 and turrivirus STIV exhibit seven-fold symmetry, while those of the ovalivirus SEV1 are six-sided. The VAPs consist of multiple copies of a single 10 kDa viral protein containing a transmembrane domain which promotes its insertion into the cellular membrane (Bize et al., 2009b; Brumfield et al., 2009; Quax et al., 2011; Quax et al., 2010; Snyder et al., 2011; Wang et al., 2018a). Interestingly, heterologous expression of the SIRV2 pyramid protein P98 in archaeal (*Sulfolobus acidocaldarius*), bacterial (*Escherichia coli*) or eukaryotic (*Saccharomyces cerevisiae*) cells resulted in correct protein insertion into the membrane and formation of VAPs. Nevertheless, the signal triggering the opening of the pyramid structures appears to be archaea-specific, because pyramids expressed in bacteria and eukaryotes were never observed in the open conformation (Daum et al., 2014). In addition, pyramidal structures have also been observed on the surface of an unidentified crenarchaeon, probably of the order *Thermoproteales*, infected with an unknown filamentous virus as well as on the surface of *Pyrobaculum oguniense* cells, suggesting that the VAP-based strategy is common among crenarchaeal viruses (Bize et al., 2008; Rensen et al., 2015).

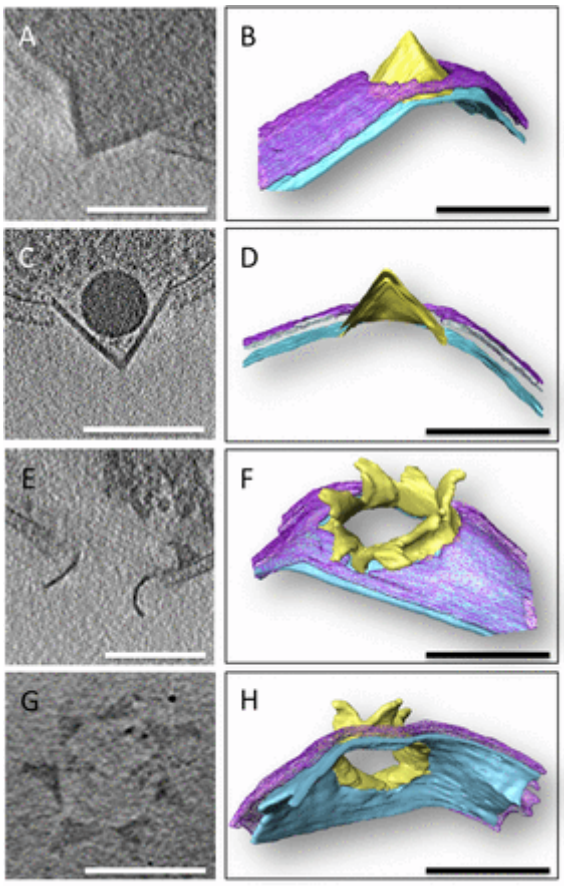


Figure 7. VAPs in closed and open conformation. Tomographic slice (A, C, E, and G) and segmented, surface-rendered volumes (B, D, F, and H) of VAPs in the membrane of SIRV2-infected S. islandicus cells. VAPs are

either closed (A–D) or open (E–H). The S-layer is purple, the cell membrane is blue, and the VAP is yellow. (Scale bars, 200 nm.). Image reproduced with permission from Daum, B., Quax, T.E.F., Sachse, M., Mills, D.J., Reimann, J., Yildiz, Ö., Häder, S., Saveanu, C., Forterre, P., Albers, S.-V., et al. (2014). Self-assembly of the general membrane-remodeling protein PVAP into sevenfold virus-associated pyramids. Proceedings of the National Academy of Sciences 111, 3829-3834.

Head-tailed viruses infecting euryarchaea also lyse their host cells, but the underlying mechanism remains unknown. Bacterial relatives of archaeal head-tailed viruses lyse the cells using the holin–endolysin system, in which holin is a small membrane protein which forms lesions in the membrane, whereas endolysin is a peptidoglycan digesting enzyme. Phages infecting gram-negative bacteria often encode additional lysis proteins responsible for disintegration of the outer membrane (Bernhardt et al., 2001; Catalão et al., 2013; Young, 2013). Thus far, homologs of the components constituting the holin-endolysin systems have not been identified in archaeal viruses. However, several viruses infecting methanogenic archaea, including siphovirus ψM1 infecting *Methanothermobacter marburgensis*, encode pseudomurein endoisopeptidases which degrade the pseudomurein layer. Activity assays confirmed the cell wall-degrading activity of the pseudomurein endoisopeptidases which cleave the *ε*-isopeptide bond between alanine and lysine in the peptide chain of the pseudomurein. However, it remains unclear how these enzymes cross the cell membrane to reach the cell wall since no genes encoding potential holins have been identified in the genome of methanogenic viruses (Luo et al., 2001; Schofield et al., 2015). Furthermore, pseudomurein is not universally present in euryarchaea or even methanogens. How head-tailed viruses infecting other archaea, such as halophiles, are released from the host remains unclear.

Lytic life cycle has been also demonstrated for filamentous enveloped viruses of the *Tristromaviridae* family which infect hyperthermophilic *Pyrobaculum* species (phylum *Crenarchaeota*). Electron microscopy analysis has shown that at the late stages of infection, the cells are packed with virions and their envelope is slashed by long, straight cuts not observed for other archaeal virus-host systems (Rensen et al., 2016). However, the mechanism underlying this lysis mechanism has not been investigated.

4.6.5.2 Viral release without membrane disruption

Many archaeal viruses are released from the host without causing cell lysis. This type of virion egress has been most extensively studied on the example of fusellovirus SSV1 (Quemin et al., 2016). The assembly of SSV1 virions is concomitant with the egress and occurs by a mechanism that resembles the budding of enveloped eukaryotic viruses (Figure 8). Dual-axis

electron tomography analysis has shown that viral nucleoprotein complexes are extruded through the host cytoplasmic membrane in the form of tubular intermediate structures which share a continuous envelope with the host membrane. Subsequently, the SSV1 virions attached to the cell membrane undergo maturation to the characteristic spindle-shaped morphology. Formation of constricted ring-like structures at the trailing end of the virion bud precedes the separation of the SSV1 virion from the cell membrane (Quemin et al., 2016). The ring-like structures observed during the final step of SSV1 budding resembles the budding necks observed prior to the ESCRT machinery-mediated membrane scission during egress of some eukaryotic viruses, including human immunodeficiency virus and Ebola viru. The ESCRT system drives key membrane-remodeling processes in eukaryotes, including cytokinesis, multivesicular body biogenesis and viral budding processes (Hurley and Hanson, 2010). Notably, proteins homologous to eukaryotic ESCRT components are conserved in several members of the Sulfolobales and play a central role in cell division (Makarova et al., 2010; Samson et al., 2008), suggesting that, as for eukaryotic viruses, SSV1 budding may rely on a cellular membrane remodeling machinery (Quemin et al., 2016). Similarly, pleolipoviruses have been suggested to possess a nonlytic life cycle and be released through a budding mechanism (Pietilä et al., 2016; Roine et al., 2010). Notably, however, unlike crenarchaea, halophilic archaea do not encode the ESCRT machinery.

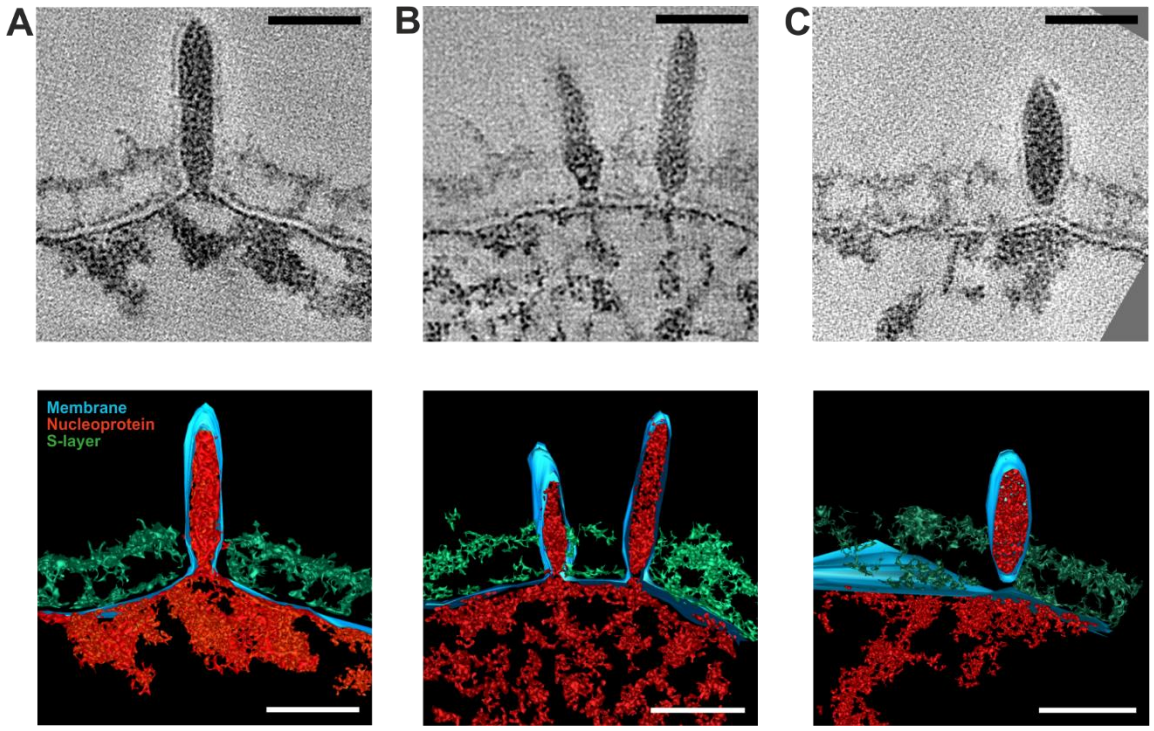


Figure 8. Different stages of SSV1 budding. (A to C) Slices through tomograms (top) and volume segmentations (bottom) showing concomitant assembly and release of SSV1 virions. Scale bars, 50 nm. Image modified and

reproduced with permission from Quemin, E.R.J., Chlanda, P., Sachse, M., Forterre, P., Prangishvili, D., and Krupovic, M. (2016). Eukaryotic-Like Virus Budding in Archaea. mBio 7, e01439-16.

4.6.6 Antiviral defense and viral counterdefense mechanisms

In most environments, viruses and their hosts are engaged in a constant evolutionary arms race. In this context, a broad range of host defense strategies as well as viral counterdefense mechanisms have been described in bacteria (Azam and Tanji, 2019), whereas in archaea such mechanisms remain poorly understood, with the exception of the CRISPR-Cas system.

4.6.6.1 Antiviral defense mechanisms

Mechanistically, the defense systems of archaea and bacteria can be classified into three main groups: (i) variation of virus receptors, (ii) innate and adaptive immunity and (iii) dormancy and programmed cell death. The variation of virus receptors includes programmed changes, such as phase variation and physical masking of the receptor in order to hamper the successful binding of the virus to the host cell. Defense mechanisms that rely on immunity involve the recognition and inactivation of invader genetic material either by nonspecific innate immunity, such as restriction modification modules and Argonaute-based innate immunity, or highly specific adaptive immunity, represented by CRISPR-Cas systems. Finally, strategies based on induction of dormancy or programmed cell death upon viral infection include toxin-antitoxin (TA) systems, in which the infection disrupts the balance of the host toxin-antitoxin complex leading to retardation of the cell growth or death (Koonin et al., 2017). During the past few years the study of antiviral defense mechanisms in archaea has been focused on CRISPR-Cas systems, whereas hints about the role of toxin-antitoxin system in virus-host interactions has been indirectly provided by transcriptomic analysis of infected cells.

CRISPR-Cas systems

The CRISPR-Cas systems represent prokaryotic adaptive immunity mechanism present in about 40% of bacteria and almost all archaea (Makarova et al., 2020). This system protects cells against invasion of mobile genetic elements, such as viruses and plasmids. CRISPR-Cas immunity involves three main stages: (a) adaptation stage, during which new virus/plasmid-derived sequences (spacers) are integrated into the CRISPR array adjacent to the leader sequence of the CRISPR array, (ii) processing stage, during which the CRISPR array is transcribed and processed into separate CRISPR RNAs (crRNAs) containing the spacer sequence with 5′ and 3′ tags from the flanking repeats of the CRISPR array, and (iii) interference stage, whereby the crRNA binds to the assembled CRISPR effector complex

(interference module) to recognize and degrade DNA and/or RNA molecules containing the protospacer sequence. Based on gene synteny and the composition of the effector complexes, CRISPR-Cas systems have been divided into two classes, each subdivided into three types and several subtypes. CRISPR-Cas systems are very abundant among archaea and often several different types of CRISPR systems are encoded on the same archaeal genome (Makarova et al., 2020).

It has been observed that SIRV2 infection leads to massive activation of the CRISPR-Cas systems present in *Sulfolobus islandicus* LAL14/1 (Quax et al., 2013). The genome of *S. islandicus* LAL14/1 contains six CRISPR-*cas* loci encoding complexes of the three subtypes: I-A, I-D and III-B (Jaubert et al., 2013). The SIRV2 infection led to a sharp increase of the transcription levels of all CRISPR-*cas* arrays, except for one incomplete type III-B CRISPR-Cas module lacking the CRISPR array. Moreover, the expression of CRISPR arrays was activated immediately after SIRV2 infection and steadily increased during the course of the infection, with the highest levels of *cas* expression being reached 1 hpi. Notably, the expression levels of different CRISPR-Cas loci varied upon SIRV2 infection, suggesting their specialized roles during viral infection. The ability of SIRV2 to propagate in *S. islandicus* LAL14/1 despite the presence of active CRISPR-Cas systems carrying spacers matching the viral genome suggested the presence of viral anti-CRISPR (Acr) proteins which could counteract the host immune response (Quax et al., 2013). Similarly, a transcriptomic analysis on STSV2 infected cells revealed a differential expression level of the CRISPR-Cas systems upon infection. *S. islandicus* REY15A, the STSV2 host, encodes one type I-A and two type III CRISPR systems. Upon STSV2 infection, the type I-A module was strongly upregulated, whereas the expression level of one type III-B complex was downregulated through the course of STSV2 infection and the second type III-B complex was weakly upregulated (León-Sobrino et al., 2016). The interaction between tailed spindle-shaped virus SMV1 and the CRISPR-Cas systems of *S. islandicus* REY15A was studied in more detail by constructing strains carrying plasmid-borne mini-CRISPR arrays targeting SMV1 genome (Guo et al., 2019). The CRISPR response against SMV1 was type-specific, with the III-B CRISPR complex showing a tight control on the inhibition of the viral replication and proliferation during the infection, whereas the I-A CRISPR complex gradually lost the control of the viral proliferation allowing viral replication and release. The absence of SMV1 escape mutations conferring tolerance to I-A CRISPR

system suggests that the virus has evolved a mechanism, probably an Acr protein, specific against the type I-A CRISPR system (Guo et al., 2019).

Rudivirus SIRV3, a closely related virus to SIRV2, undergoes a host-dependent carrier state infection in *S. islandicus* REY15A whereby the virus is maintained in a small fraction of the population over several days without apparent chromosomal DNA degradation, disturbance of the cell growth or induction of detectable CRISPR-Cas response. Notably, coinfection with the bicaudavirus SMV1 did not affect the SIRV3 carrier state and led to the coexistence of both viruses in the culture over 12 days, despite the induction of CRISPR spacer acquisition from SIRV3 DNA and the increased transcription of the subtype I-A CRISPR-Cas module. A plausible explanation for the maintenance of both viruses in the cell cultures seems to be that host CRISPR-Cas systems are inhibited by the virus-encoded Acr proteins encoded by SIRV3 and SMV1, which most likely target different types of CRISPR-Cas complexes encoded by the host (Papathanasiou et al., 2019). Transcriptomic analysis of *S. solfataricus* P2 cells infected with the fusellovirus SSV2 showed strong upregulation of the six CRISPR loci. By contrast, such effect was not observed when *S. solfataricus* P2 was infected with SSV1, a close relative of SSV2. Surprisingly, the co-infection with both viruses caused the silencing of the host CRISPR response. The major difference between SSV1 and SSV2 lies in the UV-inducible operon present in SSV1 but not in SSV2. Therefore, it has been speculated that SSV1 UV-inducible operon encodes transcription factors that may silence the CRISPR-Cas response in the SSV1-infected strain (Fusco et al., 2015a).

Interestingly, CRISPR-mediated defense systems appear to be employed not only by cells to defend against mobile genetic elements, but also by viruses for interviral conflicts. It was discovered that some archaeal viruses carry mini-CRISPR arrays (with 1-2 repeat-spacer units), which are preceded by promoter-containing leader sequences and genetic determinants required for insertion of new spacers (Medvedeva et al., 2019). Remarkably, most of these virus-borne spacers target closely related viruses present in the same population. For instance, SPV1 and SPV2, two closely related viruses of the family *Portogloboviridae* isolated from a Japanese hot spring, were found to possess mini-CRISPR arrays with spacers reciprocally targeting each other (Medvedeva et al., 2019). The existence of related viruses carrying mini-CRISPR arrays against each other in the same population suggests that viruses have adapted the host defense system for interviral conflicts. Furthermore, virus-borne mini-CRISPR arrays

may represent a mechanism of heterotypic superinfection exclusion, in which a cell infected by one virus becomes resistant to another closely-related virus (Liu et al., 2019; Medvedeva et al., 2019).

Toxin-antitoxin system

The toxin-antitoxin (TA) modules consist of a toxin protein capable of inhibiting the cell growth and an antitoxin that neutralizes the action of the toxin. TA modules can be encoded by both cellular organisms and mobile genetic elements, more commonly, plasmids but also viruses. The toxin and its cognate antitoxin form a stable complex preventing the toxin from exerting its toxic effect in normally growing cells. Whereas toxin is a stable protein, often a nuclease, the antitoxin is typically labile, with quick turnaround, and has to be constantly synthesized. Under stressful circumstances, when the production of antitoxin is perturbed, the toxin is unleashed, leading to cell dormancy or death. Given that virus infection can lead to disbalance in toxin-antitoxin equilibrium, TA clusters have been proposed to represent a mechanism of abortive infection, in which an infected cell commits "altruistic suicide" to protect the population (Yamaguchi et al., 2011). Although the role of TA systems in virus-host interactions has been poorly understood in archaea, it has been suggested that *csa5* represents a toxin gene in the type I-A CRISPR system of *Saccharolobus solfataricus* P2, because Csa5 is toxic in the CRISPR-deficient *S. solfataricus* strain. Infection with rudivirus SIRV2 leads to induction of the Csa5 expression and formation of Csa5 oligomers in *Sulfolobus* cells, suggesting that Csa5 may be involved in programmed cell death in response to virus infections (He et al., 2014). Future work is required to gain insights into the mechanism of Csa5 toxicity and its biological relevance.

Transcriptomic analyses have revealed an upregulation of the type II TA systems upon viral infection. The type II TA systems are widespread among members of the order Sulfolobales and consist of an antitoxin protein that directly binds and inhibits the effect of a toxin protein. *S. islandicus* LAL14/1 encodes 16 operons of the family VapBC (virulence associated proteins B and C) and 6 operons of the family HEPN-NT (higher eukaryote and prokaryote nucleotide binding-nucleotidyltransferases), whereby the toxins are proposed to perform RNA cleavage. Notably, the expression of 11 out of 16 VapBC and 3 out of 6 HEPN-NT loci increased after SIRV2 infection at different time points. In most cases, the expression level of both genes coding the toxin and antitoxin was upregulated to similar extents (Quax et al., 2013). Likewise, bicaudavirus STSV2 infection caused a strong upregulation of the host TA gene pairs,

including *vapBC* (León-Sobrino et al., 2016). Even though the role of TA operons remains unclear, the increase in TA gene expression after infection strongly suggests a function in host defense response.

4.6.6.2 Counterdefense mechanisms

Viruses have evolved strategies to evade the targeting by host CRISPR-Cas adaptive immunity. The simplest mechanism involves mutations in the protospacer adjacent motifs (PAM), a short motif necessary for correct CRISPR targeting. A more sophisticated mechanism relies on the dedicated anti-CRISPR proteins (Acrs) which inhibit CRISPR-Cas by a diversity of mechanisms, including binding to specific subunits of the effector complexes. Although Acrs have been studied in many bacteriophages, only three archaeal Acr proteins have been characterized (Athukoralage et al., 2020; Bhoobalan-Chitty et al., 2019; He et al., 2018).

All three archaeal Acrs were discovered in rudiviruses, although some of them display much broader distribution. The AcrID1 protein inhibits subtype I-D CRISPR-Cas system by binding directly to the Cas10d subunit of the I-D CRISPR-Cas effector complex. The AcrID1-Cas10d interaction blocks the interference stage of the I-D CRISPR-Cas response in which the CRISPR RNAs (crRNAs) and Cas proteins form an effector complex that recognizes the viral sequence complementary to that of the crRNA spacer and cleaves it. Notably, AcrID1 is a conserved dimeric αβ-sandwich protein widely distributed in viruses infecting hyperthermophilic crenarchaeota of the order *Sulfolobales,* including rudiviruses, lipothrixviruses, fuselloviruses and monocaudaviruses, with about 50 homologues (He et al., 2018).

The two other Acr proteins disarm the type III CRISPR systems. Type III CRISPR-Cas systems exhibit a high complexity and is classified into four subtypes (A–D), of which subtypes III-A (Csm) and III-B (Cmr) have been the most studied ones. In both cases, the type III effector complex binds to the viral protospacer causing the activation of the Cas10 protein as well as the synthesis of cyclic oligoadenylates (cOAs) from ATP. The presence of cOAs, in turn, triggers the activation of the Csm6 RNase in the III-A system and Csx1 in the III-B system to cleave the viral mRNA. It has been recently demonstrated that SIRV2 gp48, which is conserved in several members of *Rudiviridae* and *Lipothrixviridae,* is an Acr protein that exclusively inhibits the subtype III-B CRISPR-Cas system (AcrIIIB1) of *S. islandicus* LAL14/1. AcrIIIB1 was demonstrated to bind to two distinct effector complexes of the subtype III-B system, Cmr-

α and Cmr-γ, suggesting that the mechanism by which AcrIIIB1 inhibits the subtype III-B response is by interfering with the Csx1 RNase activation process (Bhoobalan-Chitty et al., 2019). The third archaeal Acr belongs to the DUF1874 protein family and was named AcrIII-1 family, because unlike all other known Acr of bacteria or archaea, it is not specific for a particular subtype, but blocks all subtypes of type III which use cyclic oligoadenylate signalling. AcrIII-1 is an enzyme with a ring nuclease activity, rapidly degrading the cyclic tetra-adenylate (cA_4) second messenger into a linear di-adenylate (ApA>P) with a cyclic 2′,3′-phosphate, thereby preventing the activation of the type III-associated RNase and, therefore, blocking the host type III CRISPR defense system (Athukoralage et al., 2020). Given that the target of AcrIII-1 is a signaling molecule (cA_4) with a constant structure, rather than specific CRISPR effector protein, this Acr is likely to be able to inhibit any type III CRISPR subtype using cA_4 as part of its activation. Consistently, the AcrIII-1 family is widely distributed in viruses infecting both archaea and bacteria as well as plasmids and proviruses.

5. AIMS OF THE STUDY

Viruses of hyperthermophilic archaea represent an integral but unique part of the virosphere, with many members displaying virion architectures not observed among viruses infecting bacteria or eukaryotes. However, despite their distinctiveness and importance in the viral world, the number of isolated species of viruses infecting archaea is low compared to the known eukaryotic or bacterial viruses. Moreover, the understanding on virus-host interplay in Archaea remains scarce. To address these outstanding questions, my PhD thesis project had three main aims:

1. to isolate new hyperthermophilic archaeal virus-host systems **(chapter 1)**,
2. to characterize the newly isolated viruses by a combination of various microscopy, genomic and microbiological approaches **(chapters 1 and 2)**, and
3. to study the mechanisms of virion assembly and release employed by an enveloped filamentous archaeal virus **(chapter 3)**.

6.RESULTS

6.1 CHAPTER 1

Isolation and characterization of new archaeal viruses from Italian hydrothermal environments

Contribution: I participated in the collection of the environmental samples and design of the research, performed all the experiments, contributed to the analysis of data and wrote the first draft of the manuscript.

The ISME Journal (2020) 14:1821–1833
https://doi.org/10.1038/s41396-020-0653-z

ISME

ARTICLE

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

Diana P. Baquero[1,2] · Patrizia Contursi[3] · Monica Piochi[4] · Simonetta Bartolucci[3] · Ying Liu[1] · Virginija Cvirkaite-Krupovic[1] · David Prangishvili[1,5] · Mart Krupovic[1]

Received: 13 November 2019 / Revised: 28 March 2020 / Accepted: 31 March 2020 / Published online: 22 April 2020

Abstract

Viruses of hyperthermophilic archaea represent one of the least understood parts of the virosphere, showing little genomic and morphological similarity to viruses of bacteria or eukaryotes. Here, we investigated virus diversity in the active sulfurous fields of the Campi Flegrei volcano in Pozzuoli, Italy. Virus-like particles displaying eight different morphotypes, including lemon-shaped, droplet-shaped and bottle-shaped virions, were observed and five new archaeal viruses proposed to belong to families *Rudiviridae*, *Globuloviridae* and *Tristromaviridae* were isolated and characterized. Two of these viruses infect neutrophilic hyperthermophiles of the genus *Pyrobaculum*, whereas the remaining three have rod-shaped virions typical of the family *Rudiviridae* and infect acidophilic hyperthermophiles belonging to three different genera of the order Sulfolobales, namely, *Saccharolobus*, *Acidianus,* and *Metallosphaera*. Notably, Metallosphaera rod-shaped virus 1 is the first rudivirus isolated on *Metallosphaera* species. Phylogenomic analysis of the newly isolated and previously sequenced rudiviruses revealed a clear biogeographic pattern, with all Italian rudiviruses forming a monophyletic clade, suggesting geographical structuring of virus communities in extreme geothermal environments. Analysis of the CRISPR spacers suggests that isolated rudiviruses have experienced recent host switching across the genus boundary, potentially to escape the targeting by CRISPR-Cas immunity systems. Finally, we propose a revised classification of the *Rudiviridae* family, with the establishment of six new genera. Collectively, our results further show that high-temperature continental hydrothermal systems harbor a highly diverse virome and shed light on the evolution of archaeal viruses.

Introduction

One of the most remarkable features of hyperthermophilic archaea is the diversity and uniqueness of their viruses. Most of these viruses infect members of the phylum Crenarchaeota and are evolutionarily unrelated to viruses infecting bacteria, eukaryotes or even archaea thriving at moderate temperature [1–4]. Thus far, unique to hyperthermophilic archaea are rod-shaped viruses of the families *Rudiviridae* and *Clavaviridae*; filamentous enveloped viruses of the families *Lipothrixviridae* and *Tristromaviridae*; as well as spherical (*Globuloviridae*), ellipsoid (*Ovaliviridae*), droplet-shaped (*Guttaviridae*), coil-shaped (*Spiraviridae*) and bottle-shaped (*Ampullaviridae*) viruses [1, 5]. Hyperthermophilic archaea are also infected by two types of spindle-shaped viruses, belonging to the families *Fuselloviridae* and *Bicaudaviridae* [6, 7]. Whereas bicaudaviruses appear to be restricted to hyperthermophiles, viruses distantly related to fuselloviruses are also known to infect hyperhalophilic archaea [8], marine hyperthermophilic

Supplementary information The online version of this article (https://doi.org/10.1038/s41396-020-0653-z) contains supplementary material, which is available to authorized users.

✉ David Prangishvili
david.prangishvili@pasteur.fr

✉ Mart Krupovic
mart.krupovic@pasteur.fr

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

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

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

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

[5] Ivane Javakhishvili Tbilisi State University, Tbilisi 0179, Georgia

archaea [9, 10] and marine ammonia-oxidizing archaea from the phylum Thaumarchaeota [11]. Finally, hyperthermophilic archaea are also infected by three groups of viruses with icosahedral virions, *Turriviridae* [12], *Portogloboviridae* [13] and two closely related, unclassified viruses infecting *Metallosphaera* species [14]. Portogloboviruses SPV1 is structurally and genomically unrelated to other known viruses [13, 15], whereas viruses structurally similar to turriviruses are widespread in all three domains of life [1, 16]. Structural studies on filamentous and spindle-shaped crenarchaeal viruses have illuminated the molecular details of virion organization and further underscored the lack of relationship to viruses of bacteria and eukaryotes [17–23].

The uniqueness of hyperthermophilic archaeal viruses extends to their genomes, with ~75% of the genes lacking detectable homologs in sequence databases [24]. All characterized archaeal viruses have DNA genomes, which can be single-stranded (ss) or double-stranded (ds), linear or circular. Comparative genomic and bipartite network analyses have shown that viruses of hyperthermophilic archaea share only few genes with the rest of the virosphere [25]. Furthermore, with some exceptions (see below), most of the genes in archaeal viruses are family-specific [26]. These observations, in combination with the structural studies, led to the suggestion that crenarchaeal viruses have originated on multiple independent occasions and constitute a unique part of the virosphere [1].

Although the infection cycles of crenarchaeal viruses have been studied for just a handful of representatives, the available data has already provided valuable insight into the virus–host interaction strategies in archaea. Two different egress strategies have been elucidated. The enveloped virions of fusellovirus SSV1 are assembled at the host cell membrane and are released from the cell by a budding mechanism similar to that of some eukaryotic enveloped viruses [27]. By contrast, lytic crenarchaeal viruses belonging to three unrelated families, *Rudiviridae*, *Turriviridae* and *Ovaliviridae*, employ a unique release mechanism based on the formation of pyramidal protrusions on the host cell surface, leading to perforation of the cell envelope and release of intracellularly assembled mature virions [5, 28, 29]. Remarkably, the pyramids are formed by a single virus-encoded protein of less than 100 amino acids and the corresponding gene has been apparently exchanged horizontally among viruses from different families [30, 31]. Differently from bacteriophages, many hyperthermophilic archaeal viruses encode divergent glycosyltransferases of either GT-A or GT-B superfamily and some carry multiple gene copies, suggesting an important function [2, 24, 32]. Consistently, virions of many crenarchaeal viruses are glycosylated [16, 17, 19, 33, 34], although the exact physiological role of the glycosylation remains unknown. A recent study has suggested that glycosylation confers solubility and stability to macromolecular assemblies, such as type 4 pili and potentially virions, in extreme environments [35]. Another functional group of proteins which is broadly distributed across different families of crenarchaeal viruses includes anti-CRISPR (Acr) proteins. Indeed, CRISPR-Cas systems are prevalent in archaea in general and hyperthermophiles in particular [36]. Recent studies have uncovered three families of Acr proteins widespread in archaeal viruses and plasmids, which block CRISPR-Cas systems of types I and III by different mechanisms [37–39].

Single-cell sequencing combined with environmental metagenomics of hydrothermal microbial community from Yellowstone National Park [40] led to the estimation that >60% of cells contain at least one virus type and a majority of these cells contain two or more virus types [41]. However, despite their diversity, distinctiveness, and abundance, the number of isolated species of viruses infecting hyperthermophilic archaea remains low compared to that of the known eukaryotic or bacterial viruses [2]. Indeed, it has been estimated that only about 0.01–0.1% of viruses present in geothermal acidic environments have been isolated [42]. Similarly, using a combination of viral assemblage sequencing and network analysis, it has been estimated that out of 110 identified virus groups, less than 10% represent known archaeal viruses, suggesting that the vast majority of virus clusters represent unknown viruses, likely infecting archaeal hosts [43]. Furthermore, the evolution and structuring of virus communities in terrestrial hydrothermal settings remain poorly understood. Here, to improve understanding on these issues, we explored the diversity of archaeal viruses at the active solfataric field of the Campi Flegrei volcano [44–46] in Pozzuoli, Italy, namely, the Pisciarelli hydrothermal area. The field is known for hot acidic environmental conditions [46] and its microbial communities are dominated by extremophilic microbes [34, 47–50]. However, the area we selected is characterized by hydrothermal conditions that are continuously changing in the short term due to the volcano dynamics, which could have an effect on the composition of resident microbial communities and their viruses. We report on the isolation and characterization of five new archaeal viruses belonging to three different families and infecting hosts from five different crenarchaeal genera.

Materials and methods

Materials and Methods are available in the Supplementary Information.

Results

Diversity of virus-like particles in enrichment cultures

Nine environmental samples (I1-I9) were collected from hot springs, mud pools and hydrothermally altered terrains of the solfataric fields of the Campi Flegrei volcano in Pozzuoli, Italy, with temperatures ranging from 81 to 96 °C and pH values between 1 and 7 (Supplementary Information, Table S1). The enrichment cultures were obtained by inoculating the samples into different media favoring the growth of hyperthermophilic members of the genera *Sulfolobus/Saccharolobus*, *Acidianus* and *Pyrobaculum* [34, 51]. In particular, samples with acidic pH were inoculated into medium favoring the growth of *Sulfolobus/Saccharolobus* and *Acidianus* species, whereas those with neutral pH were inoculated into the *Pyrobaculum* medium (see Supplementary Materials and Methods for details). Virus-like particles (VLPs) were collected from cell-free culture supernatants and visualized by transmission electron microscopy as described in Supplementary Information.

A variety of VLPs were detected in the *Sulfolobus/Saccharolobus* enrichment cultures of samples I3 and I9 and the *Pyrobaculum* enrichment culture of sample I4 (Fig. 1). Based on virion morphologies, the VLPs detected in the two samples inoculated in *Sulfolobus/Saccharolobus* medium could be assigned to five archaeal virus families: *Fuselloviridae* (Fig. 1a), *Bicaudaviridae* (Fig. 1b), *Ampullaviridae* (Fig. 1c), *Rudiviridae* (Fig. 1d) and *Lipothrixviridae* (Fig. 1e). VLPs propagated in the *Pyrobaculum* medium resembled members of the families *Globuloviridae* (Fig. 1f), *Tristromaviridae* (Fig. 1g) and *Guttaviridae* (Fig. 1h). We next set out to establish pure cultures of these different viruses and to isolate their respective hosts.

Isolation of virus–host pairs

In order to isolate VLP-propagating strains, 215 single-strain isolates were colony purified from the enrichment cultures established in the *Sulfolobus/Saccharolobus* medium of the VLP-producing samples I3 and I9. Concentrated VLPs were first tested against the isolates by spot test. In case of cell growth inhibition, a liquid culture of the isolate was established and infected with the halo zone observed in the spot test. The production of the viral particles was subsequently verified by TEM. As a result, three strains replicating VLPs were identified. Comparison of their 16S rRNA gene sequences showed that the three strains belong to three different genera of the order Sulfolobales, namely, *Saccharolobus* (until recently known as *Sulfolobus*), *Acidianus* and *Metallosphaera*. The 16S rRNA genes of these strains, POZ9, POZ149 and POZ202, respectively, displayed 100, 99, and 99% identity to the corresponding genes of *Acidianus brierleyi* DSM 1651 (NZ_CP029289), *Saccharolobus solfataricus* Ron 12/III (X90483) and *Metallosphaera sedula* SARC-M1 (CP012176). Rod-shaped particles of different lengths were propagated

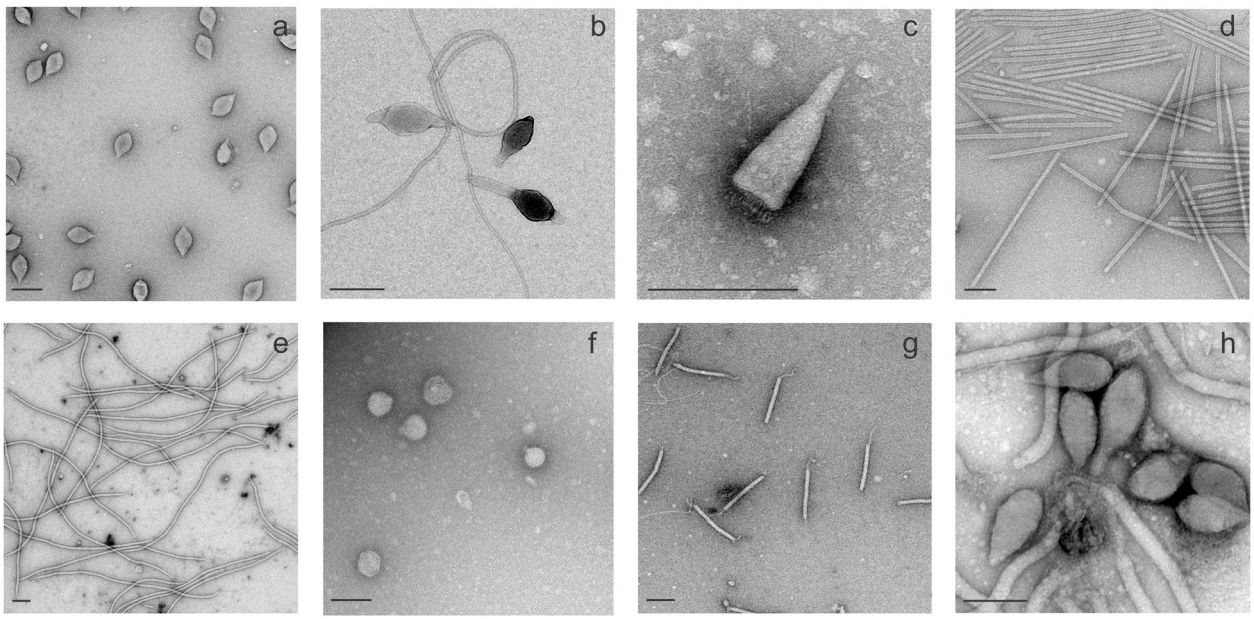


Fig. 1 Electron micrographs of the VLPs observed in enrichment cultures. a Fuselloviruses (tailless lemon-shaped virions). **b** Bicaudaviruses (large, tailed lemon-shaped virions). **c** Ampullaviruses (bottle-shaped virions). **d** Rudiviruses (rod-shaped virions). **e** Lipothrixviruses (filamentous virions). **f** Globuloviruses (spherical enveloped virions). **g** Tristromaviruses (filamentous enveloped virions). **h** Guttaviruses (droplet-shaped virions). Samples were negatively stained with 2% (wt/vol) uranyl acetate. Scale bars: 200 nm.

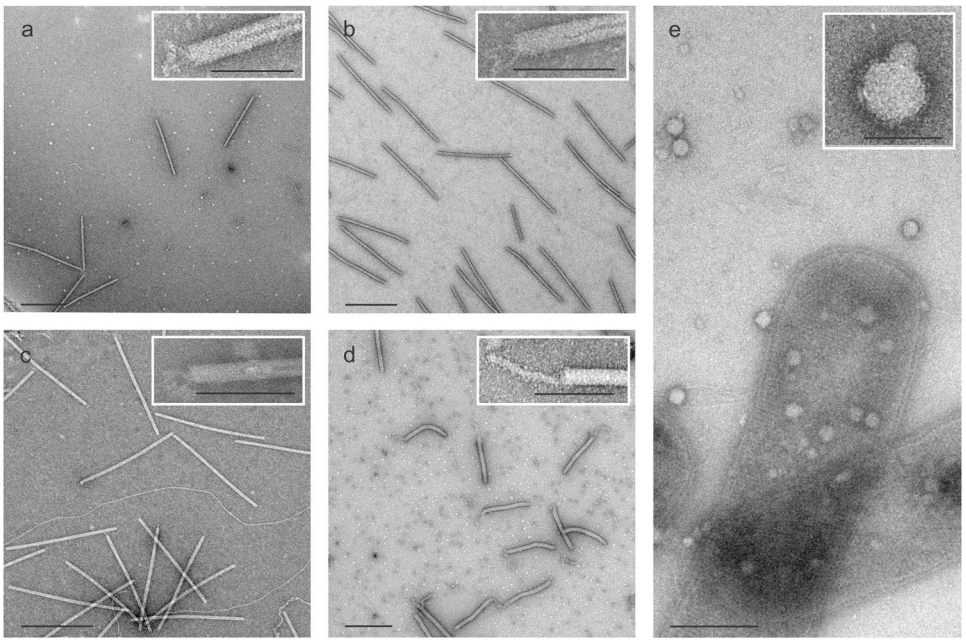


Fig. 2 Electron micrographs of the five isolated viruses. **a** Metallosphaera rod-shaped virus 1. **b** Acidianus rod-shaped virus 3. **c** Saccharolobus solfataricus rod-shaped virus 1. **d** Pyrobaculum filamentous virus 2. **e** Pyrobaculum spherical virus 2. Samples were negatively stained with 2% (wt/vol) uranyl acetate. Scale bars: 500 nm; in insets: 100 nm.

successfully in the three isolated strains (Fig. 2a–c) and, following the nomenclature used for other rudiviruses, were named Metallosphaera rod-shaped virus 1 (MRV1), Acidianus rod-shaped virus 3 (ARV3) and Saccharolobus solfataricus rod-shaped virus 1 (SSRV1), respectively.

Negatively stained MRV1, ARV3, and SSRV1 virions are rod-shaped particles measuring $630 \pm 20 \times 25 \pm 2$ nm, $670 \pm 40 \times 23 \pm 4$ nm and $750 \pm 30 \times 24 \pm 3$ nm, respectively (Fig. 2a–c). Similar to members of the *Rudiviridae* family, the three viral particles have terminal fibers located at each end of the virion, which have been shown to play a role in host recognition in the case of Sulfolobus islandicus rod-shaped virus 2 (SIRV2) [52].

The three rudiviruses displayed different infection dynamics. Infection of *S. solfataricus* POZ149 cultures with SSRV1 (MOI = 3) resulted in severe growth retardation (Fig. S1a). The optical density of infected cultures remained constant for the first 24 hpi, similar to what has been reported previously for the prototypical rudivirus SIRV2 [28]. By contrast, infection with MRV1 and ARV3 had no apparent effect on the growth of *M. sedula* POZ202 and *A. brierleyi* POZ9, respectively (Fig. S1a). Nevertheless, production of extracellular virions was detected in all three cultures, starting at 8 (for SSRV1) or 12 (for MRV1 and ARV3) hpi and peaking at 24–32 hpi (Fig. S2).

A different approach was chosen to identify the hosts of the viral particles detected in the *Pyrobaculum* medium. Exponentially growing liquid cultures of *Pyrobaculum* strains were mixed with concentrated VLPs and incubated for 15 days at 90 °C (see Supplementary Materials and Methods). The replication of the particles was monitored by TEM. *P. arsenaticum* 2GA propagated filamentous and spherical particles, named Pyrobaculum filamentous virus 2 (PFV2; Fig. 2d) and Pyrobaculum spherical virus 2 (PSV2; Fig. 2e), respectively. Because *P. arsenaticum* 2GA could not be grown as a lawn on solid medium, dilutions of infected cells were made in order to establish cultures infected with just one type of viral particles.

Negatively stained virions of PFV2 are filamentous and flexible particles of about $450 \pm 20 \times 34 \pm 4$ nm in size with terminal filaments of up to 130 nm in length attached to one or both ends of the virions (Fig. 2d), similar to what has been reported for PFV1 [34]. PSV2 virions are spherical particles of around 90 ± 20 nm of diameter, with a variable number of bulging protrusions on their surface (Fig. 2e), resembling Pyrobaculum spherical virus (PSV) particles [50]. Unfortunately, the hosts for other VLPs shown in Fig. 1 could not be isolated, either due to unfavorable growth under laboratory conditions or due to virus-mediated extinction of the corresponding host cell populations.

Infection of *P. arsenaticum* 2GA with PSV2 resulted in a slight retardation of the host growth, with no detectable cell debris throughout the incubation, suggesting that the virus is not lytic (Fig. S1b). By contrast, infection with the filamentous virus PFV2 resulted in growth retardation of *P. arsenaticum* 2GA. This observation is consistent with the previous results showing that the closely related virus PFV1 lyses its host through an unknown mechanism [34].

Host range

To test the host ranges of the five isolated viruses, strains of the family *Sulfolobaceae* available in our laboratory

Table 1 Host range of the archaeal viruses isolated in this study.

Archaeal strain	Newly-isolated archaeal virus				
	MRV1	ARV3	SSRV1	PSV2	PFV2
Metallosphaera sedula POZ202	H	–	–	–	–
Acidianus brierleyi POZ9	–*	H	–	–	–
Acidianus convivator	–	–*	–*	–	–
Acidianus hospitalis W1	–	+*	–*	–	–
Saccharolobus solfataricus P1	–	–	–*	–	–
Saccharolobus solfataricus P2	–*	–	–*	–	–
Saccharolobus solfataricus POZ149	–	–	H	–	–
Sulfolobus islandicus LAL14/1	–	–	–*	–	–
Sulfolobus islandicus REN2H1	–	–	–	–	–
Sulfolobus islandicus HVE10/4	–	–	–*	–	–
Sulfolobus islandicus REY15A	–	–	–	–	–
Sulfolobus islandicus ΔC1C2	–	–	–	–	–
Sufolobus acidocaldarius DSM 639	–	–	–	–	–
Pyrobaculum arsenaticum PZ6 (DSM 13514)	–	–	–	–*	–
Pyrobaculum arsenaticum 2GA	–	–	–	H	H
Pyrobaculum calidifontis VA1 (DSM 21063)	–	–	–	–	–
Pyrobaculum oguniense TE7 (DSM 13380)	–	–	–	–*	+*

H isolation host, + supports virus replication, * supports DNA delivery, – no virus production observed.

collection (Table 1) were infected with MRV1, ARV3, and SSRV1, whereas strains of *Pyrobaculum* (Table 1) were infected with PSV2 and PFV2. The production of virions was verified by spot test (for MRV1, ARV3 and SSRV1) and TEM. Only two strains were found to serve as additional hosts for ARV3 and PFV2. *A. hospitalis* W1 supported the propagation of ARV3, whereas *P. oguniense* TE7 served as a host of PFV2 (see Supplementary Materials and Methods for details).

We next investigated whether the infection in most of the tested strains is blocked prior or following the entry into the cell. To this end, the corresponding cells were incubated with the virus for 1 h, excess of the viruses was removed by extensive washes and presence of the viral DNA in the cells was tested by PCR. SSRV1 DNA could be detected in the largest number of strains. In addition to the host *S. solfataricus* POZ149 cells, SSRV1 DNA was present in *A. convivator*, *A. hospitalis* W1, *S. solfataricus* strains P1 and P2 as well as in *S. islandicus* strains HVE10/4 and LAL14/1 (Table 1 and Fig. S3a). Notably, however, adsorption assay (see Supplementary Materials and Methods) did not show appreciable binding of SSRV1 virions to most of the non-host strains (Fig. S4), consistent with the lower sensitivity of the latter assay compared to PCR. Indeed, the signal of SSRV1 DNA amplification was substantially fainter in the non-host cells, compared to the designated host. ARV3 DNA was detected in all three *Acidianus* strains, whereas that of MRV1, in addition to the host *Metallosphaera* strain, was found in *A. brierleyi* POZ9 and *S. solfataricus* P2 (Table 1 and Fig. S3b,c). These results suggest that MRV1 is able to deliver its DNA into cells from three different genera isolated in the same location, but the infection is blocked at a later, post-entry stage of the infection cycle.

PFV2 DNA was detected in *P. arsenaticum* 2GA and *P. oguniense* TE7 cultures, consistent with the observations made by TEM (Table 1, Fig. S3d), whereas PSV2 genome was detected not only in the host strain but also in *P. oguniense* TE7 and *P. arsenaticum* PZ6 (Fig. S3e), although no virions were observed in the latter strains by TEM.

Genome organization

Genomes of the five viruses were isolated from the purified virions and treated with DNase I, type II restriction endonucleases (REases) and RNase A. None of the viral genomes were sensitive to RNase A, but could be digested by DNase I and REases, indicating that all viral genomes consist of dsDNA molecules. The genomes were sequenced on Illumina MiSeq platform, with the assembled contigs corresponding to complete or near-complete virus genomes. The general properties of the virus genomes are summarized in Table 2.

New species in the *Globuloviridae* family

The linear dsDNA genome of PSV2 is 18,212 bp in length and has a GC content of 45%, which is similar to that of other *Pyrobaculum*-infecting viruses (45–48%) [34, 50]. The coding region of the PSV2 genome is flanked by

Table 2 Genomic features of the hyperthermophilic archaeal viruses isolated in this study.

Virus	Host	Size (bp)	TIR	Topology	GC%	ORFs	Accession #
MRV1	*Metallosphaera sedula* POZ202	20269	+	linear	34.12	27	MN876843
ARV3	*Acidianus brierleyi* POZ9	23666	–	linear	32.02	33	MN876842
SSRV1	*Saccharolobus solfataricus* POZ149	26097	–	linear	32.31	37	MN876841
PSV2	*Pyrobaculum arsenaticum* 2GA	18212	+	linear	45.01	32	MN876845
PFV2	*Pyrobaculum arsenaticum* 2GA	17602	+	linear	45.32	39	MN876844

TIR terminal inverted repeats, *ORFs* open reading frames.

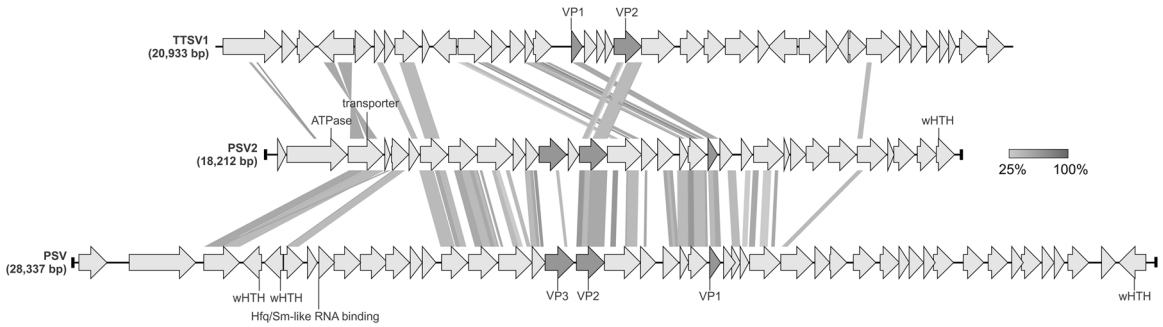


Fig. 3 Genome alignment of the three members of the *Globuloviridae* family. The open reading frames (ORFs) are represented by arrows that indicate the direction of transcription. The terminal inverted repeats (TIRs) are denoted by black bars at the ends of the genomes. Genes encoding the major structural proteins are shown in dark gray. The functional annotations of the predicted ORFs are depicted above/below the corresponding ORF. Homologous ORFs and ORF fragments are connected by shading in grayscale based on the level of amino acid sequence identity between the homologous regions. VP, virion protein; wHTH, winged helix-turn-helix domain.

perfect 55 bp-long terminal inverted repeats (TIR), confirming the linear topology and (near) completeness of the genome. It contains 32 predicted open reading frames (ORFs), all located on the same strand. Thirteen PSV ORFs contain at least one predicted membrane-spanning region. Notably, two of them (ORFs 3 and 9) have nine predicted transmembrane domains (Table S2).

Globuloviruses stand out as some of the most mysterious among archaeal viruses, with 98% of their proteins showing no similarity to sequences in public databases and lacking functional annotation [24]. Homologs of PSV2 proteins were identified exclusively in members of the *Globuloviridae* (Fig. 3), corroborating the initial affiliation of PSV2 into the family *Globuloviridae* based on the morphological features of the virion. Nineteen PSV2 ORFs, including those encoding the three major structural proteins (VP1-VP3) [53], have closest homologs in PSV [50] with amino acid sequence identities ranging between 28% and 65% (Fig. 3); 13 of these ORFs are also shared with Thermoproteus tenax spherical virus 1 (TTSV1; $E < 1e{-}05$), the only other characterized member of the *Globuloviridae* family [54]. The remaining 13 PSV2 ORFs yielded no significant matches to sequences in public databases. The three genomes display no appreciable similarity at the nucleotide sequence level, indicating considerable sequence diversity within the *Globuloviridae* family. Notably, among five PSV proteins for which high-resolution structures are available [32], only one protein with a unique fold, PSV gp11 (ORF239), is conserved in PSV2 ($E = 2e{-}17$).

Sensitive profile-profile comparisons using HHpred allowed functional annotation of only four PSV2 proteins. The PSV2 ORF2 encodes a protein with an AAA+ ATPase domain, which is most closely related to those found in ClpB-like chaperones and heat shock proteins (HHpred probability of 99.6%; Supplementary Information, Table S2). ORF3 encodes one of the two proteins with nine putative transmembrane domains and is predicted to function as a membrane transporter, most closely matching bacterial and archaeal cation exchangers (HHpred probability of 95.5%). Interestingly, the product of ORF4 is predicted to be a circadian clock protein KaiB, albeit with a lower probability (HHpred probability of 93%). Finally, ORF32 shares homology with a putative transcriptional regulator with the winged helix-turn-helix domain (wHTH) of *S. solfataricus* (HHpred probability of 94.4%). In addition, ORF11 encodes a functionally uncharacterized DUF1286-family protein conserved in archaea and several *Saccharolobus*-infecting viruses (HHpred probability of 99.5%; Supplementary Information, Table S2).

New species in the *Tristromaviridae* family

The linear genome of PFV2 is 17,602 bp long and contains 39 ORFs, all except one located on the same strand. The coding region is flanked by 59 bp-long TIRs. The GC

content (45.3%) of the genome is similar to that of PSV2 and other *Pyrobaculum*-infecting viruses [34, 50], but is considerably lower than in *P. arsenaticum* PZ6 (58.3%), and *P. oguniense* TE7 (55.1%). Eleven of the PFV2 ORFs were predicted to encode proteins with one or more membrane-spanning domains (Supplementary Information, Table S3).

The PFV2 genome is 98.9% identical over 70% of its length to that of PFV1, the type species of the *Tristromaviridae* family [55]. PFV1 and PFV2 were isolated ~3 years apart, from the same solfataric field in Pozzuoli [34], suggesting that the population of tristromaviruses is relatively stable over time. Accordingly, 36 of the 39 PFV2 ORFs are nearly identical to those of PFV1, with the ORFs encoding the three major structural proteins (VP1, VP2 and VP3) showing amino acid sequence identities higher than 96.6% (Supplementary Information, Table S3). Two events account for the differences between PFV1 and PFV2: (i) a deletion spanning most of the PFV1 gene 27 (including codons 72–495) as well as the downstream genes 28–30, and (ii) insertion of a four-gene block between PFV1 genes 36 and 37 in PFV2 (Fig. 4). PFV1 gene 27 encodes a minor virion protein, whereas genes 29 and 30 encode putative lectin-like carbohydrate-binding proteins [34]. The absence of the corresponding genes in PFV2 genome suggests that they are dispensable for the PFV1/PFV2 infection cycle. Notably, a homolog of the PFV1 glycoside hydrolase gene 28 is reinserted into PFV2 genome as part of the four-gene block (PFV2 ORF34). However, the two genes do not appear to be orthologous in PFV1 and PFV2 genomes, because they share much lower sequence similarity compared to other orthologous genes (50% versus average 96.5% identity). By contrast, PFV2 ORF35 has no counterpart in PFV1 but is homologous to the glycosyltransferase gene of Thermoproteus tenax virus 1 (TTV1) [56], the only other known member of the *Tristromaviridae* family [55] (Fig. 4). Thus, comparison of the closely related tristromavirus sequences revealed active genome remodeling in this virus group, involving both deletions and horizontal acquisition of new genes.

New rod-shaped viruses

The linear genomes of the isolated rudiviruses have a length ranging from 20,269 to 26,079 bp and a GC content varying between 32.02 and 34.12%. The MRV1 genome contains 80 bp-long TIR, suggesting that the genome is coding-complete (i.e., contains all protein-coding genes). Although no TIRs could be identified for ARV3 and SSRV1, comparison with the genomes of other rudiviruses (Fig. 5) suggests that the two genomes are also nearly complete.

The genomes of MRV1, ARV3, and SSRV1 contain 27, 33, and 37 ORFs, respectively (Supplementary Tables S4–S6), and display high degree of gene synteny (Fig. 5). Comparison of the three genomes showed that they share 26 putative proteins, with amino acid sequence identities ranging between 35.1% and 93.9%. Among the previously reported rudiviruses, the three newly isolated viruses share the highest similarity with ARV2 (ANI of ~78%), which was metagenomically sequenced from samples collected in the same Pozzuoli area [49]; this group of viruses shares 20 genes.

Rudiviruses represent one of the most extensively studied families of archaeal viruses with many of the viral proteins being functionally and structurally characterized [57]. Most of the MRV1, ARV3 and SSRV1 ORFs have orthologs in at least one other member of the *Rudiviridae*. For instance, 96%, 76%, and 84% of the MRV1, ARV3, and SSRV1 ORFs, respectively, have orthologs in ARV2. Genes shared with other rudiviruses include those for the major and minor capsid proteins, several transcription factors with ribbon-helix-helix motifs, three glycosyltransferases, GCN5-family acetyltransferase, Holliday junction resolvase, SAM-dependent methyltransferase and a gene cassette encoding the ssDNA-binding protein, ssDNA annealing ATPase and Cas4-like ssDNA exonuclease

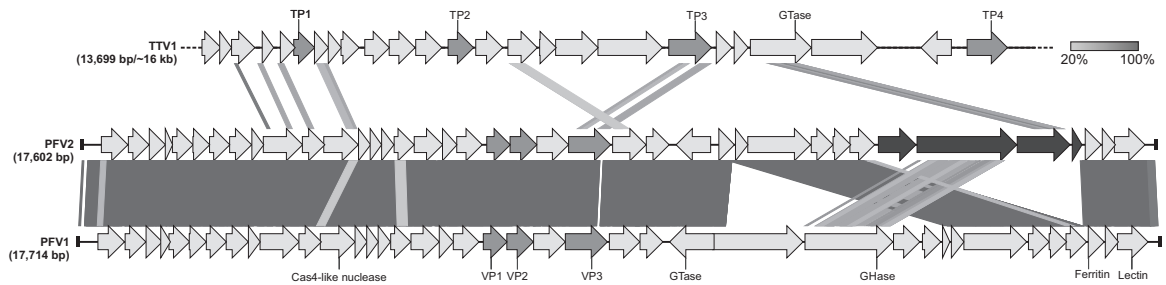


Fig. 4 Genome comparison of the three members of the *Tristromaviridae* family. The open reading frames (ORFs) are represented by arrows that indicate the direction of transcription. The terminal inverted repeats (TIRs) are denoted by black bars at the ends of the genomes. Genes encoding the major structural proteins are shown in dark gray, whereas the four-gene block discussed in the text is shown in black. The functional annotations of the predicted ORFs are depicted above/below the corresponding ORFs. Homologous genes are connected by shading in grayscale based on the level of amino acid sequence identity. The dotted line represents the incompleteness of the TTV1 genome. GHase, glycoside hydrolase; GTase, glycosyltransferase; TP/VP, virion protein.

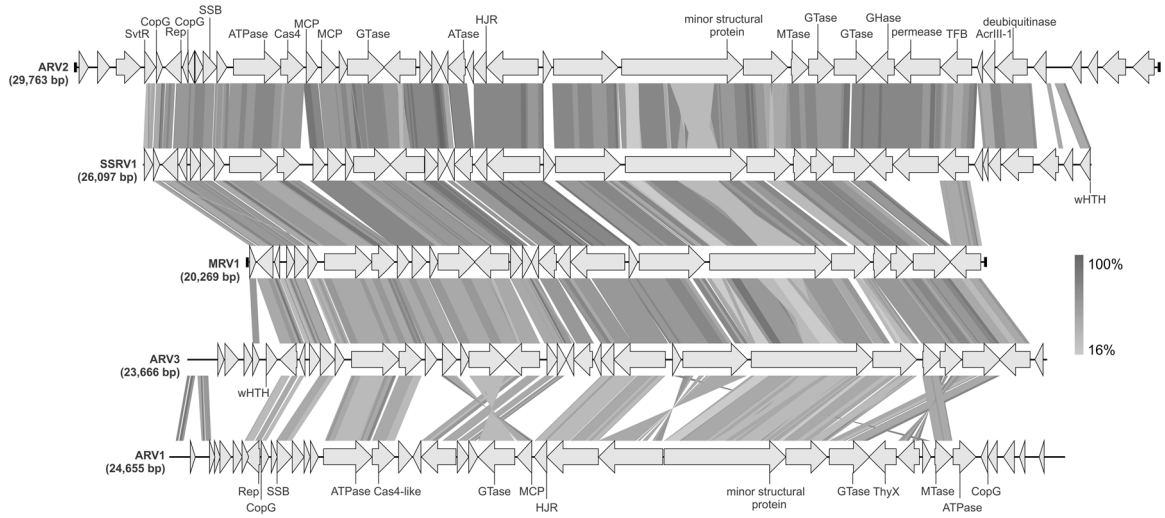


Fig. 5 Genome alignment of the Italian rudiviruses. ARV3, MRV1, and SSRV1 are newly isolated members of the *Rudiviridae* family, whereas ARV1 and ARV2 were reported previously [49, 68]. The open reading frames (ORFs) are represented by arrows that indicate the direction of transcription. The terminal inverted repeats (TIRs) are denoted by black bars at the ends of the genomes. The functional annotations of the predicted ORFs are depicted above/below the corresponding ORFs. Homologous genes are connected by shading in grayscale based on the amino acid sequence identity. AcrIII-1, anti-CRISPR protein blocking type III CRISPR-Cas systems; ATase, acetyltransferase; CopG, ribbon-helix-helix motif-containing transcription regulator; GHase, glycoside hydrolase; GTase, glycosyltransferase; HJR, Holliday junction resolvase; MCP, major capsid protein; MTase, methyltransferase; Rep, replication initiation protein; SSB, ssDNA binding protein; TFB, Transcription factor B; ThyX, thymidylate synthase; wHTH, winged helix-turn-helix domain.

(Fig. 5). The conservation of these core proteins in the expanding collection of rudiviruses underscores their critical role in viral reproduction. Conspicuously missing from the MRV1 and ARV3 are homologs of the SIRV2 P98 protein responsible for the formation of pyramidal structures for virion egress [30]. Cells infected with ARV3 and MRV1 were imaged by transmission electron microscopy at the peak of virion release (24 and 32 hpi; Fig. S2), but no evident pyramids were observed on the surface of the infected cells (Fig. S5). Thus, the mechanisms of MRV1 and ARV3 egress remain enigmatic and deserve further investigation.

Homologs of known Acr proteins encoded by diverse crenarchaeal viruses [37–39] are also missing from MRV1 and ARV3 genomes. Notably, SSRV1 carries a gene for the recently characterized AcrIII-1, which blocks antiviral response of type III CRISPR-Cas systems by cleaving the cyclic oligoadenylate second messenger [37]. Given that CRISPR-Cas systems are highly prevalent in hyperthermophilic archaea [36], the lack of identifiable anti-CRISPR genes in MRV1 and ARV3 is somewhat unexpected, suggesting that the two viruses encode novel Acr proteins. Similarly, lack of the genes encoding recognizable P98-like pyramid proteins in MRV1 and ARV3 (as well as in ARV1 and ARV2) suggests that these viruses have evolved a different solution for virion release.

Besides the core genes, rudiviruses are known to carry a rich complement of variable genes, which typically occupy the termini of linear genomes and are shared with viruses isolated from the same geographical location [58]. For instance, MRV1, ARV3, and SSRV1 carry several genes, which are exclusive to Italian rudiviruses. These include a divergent glycoside hydrolase, putative metal-dependent deubiquitinase, alpha-helical DNA-binding protein, transcription initiation factor and several short hypothetical proteins, which are likely candidates for Acrs (Supplementary Information, Tables S4–S6). Notably, some of these hypothetical proteins are shared with other crenarchaeal viruses isolated from the same location. In particular, ARV3 ORFs 5 and 33 as well as SSRV1 ORF31 are homologous to the putative proteins of the bicaudaviruses Acidianus two-tailed virus (ATV) and ATV2, whereas SSRV1 ORF36 is conserved in lipothrixvirus Acidianus filamentous virus 6.

A biogeographic pattern in the *Rudiviridae* family

To gain insight into the global architecture of the rudivirus populations and the factors that govern it, we performed phylogenomic analysis of all available rudivirus genomes using the Genome-BLAST Distance Phylogeny method implemented in VICTOR [59]. Our results unequivocally show that the 19 sequenced rudiviruses fall into six clades corresponding to the geographical origins of the virus isolation (Fig. 6), suggesting local adaptation of the corresponding viruses. Thus, on the global scale, horizontal spread of rudivirus virions between geographically remote continental hydrothermal systems appears to be restricted.

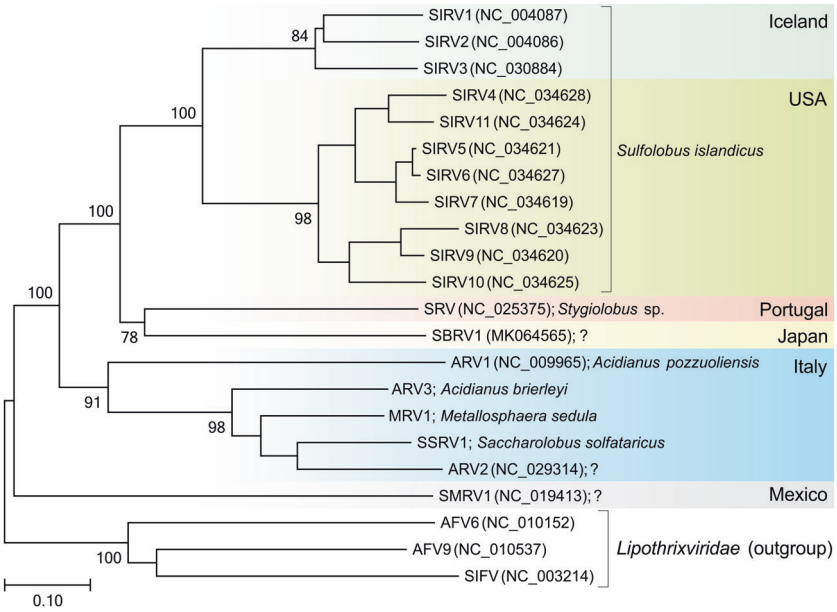


Fig. 6 Inferred phylogenomic tree of all known members of the *Rudiviridae* family based on whole genome VICTOR [59] analysis at the amino acid level. The tree is rooted with lipothrixviruses, and the branch length is scaled in terms of the Genome BLAST Distance Phylogeny (GBDP) distance formula D6. Only branch support values >70% are shown. For each genome, the abbreviated virus name, GenBank accession number and host organism (when known) are indicated. Question marks denote that the host is not known. The tree is divided into colored blocks according to the geographical origin of the compared viruses.

Remarkably, despite forming a monophyletic group, all rudiviruses originating from Italy infect relatively distant hosts, belonging to three different genera of the order Sulfolobales. Such pattern was somewhat unexpected, because for all previously characterized rudiviruses, the genetic divergence of the viruses paralleled that of their respective hosts [58, 60]. Thus, we hypothesized that such pattern of host specificities might signify host switching events in the history of the Italian rudivirus assemblage. CRISPR arrays, which contain spacer sequences derived from mobile genetic elements, keep memory of past infections and are commonly used as indicators for matching the uncultivated viruses to their potential hosts [61–63]. Thus, to investigate which hosts were exposed to Italian rudiviruses, we searched the CRISPRdb database [64] for the presence of spacers matching the corresponding viral genomes. Spacer matches were found for all five virus genomes, albeit with different levels of identity. Matches with 100% identity were obtained for ARV2, ARV3, SSRV1, and MRV1 (Supplementary Information, Table S7). Spacers matching ARV2 and ARV3 were found in *Saccharolobus solfataricus* P1, whereas ARV2 was also targeted by a CRISPR spacer (100% identity) from *Metallosphaera sedula* DSM 5348. Unexpectedly, MRV1, which infects *Metallosphaera* species, was matched by spacers from different strains of *S. solfataricus*. Conversely, SSRV1 infecting *S. solfataricus* was targeted by multiple spacers from CRISPR arrays of *M. sedula* DSM 5348 (Supplementary Information, Table S7). These results are consistent with the possibility that in the recent history of Italian rudiviruses, host switching, even across the genus boundary, has been relatively common.

Revised classification of rudiviruses

Of the 19 rudiviruses for which (near) complete genome sequences are available, only three (SIRV1, SIRV2, and ARV1) are officially classified. All three viruses are included in the same genus, *Rudivirus*. Here, we propose a taxonomic framework for classification of all cultivated and uncultured rudiviruses for which genome sequences are available. As mentioned above, phylogenomic analysis revealed six different clades (Fig. 6), highlighting considerable diversity of the natural rudivirus population, which has stratified into several assemblages, warranting their classification into distinct, genus-level taxonomic units. To this end, we compared the genome sequences using the Gegenees tool [65], which fragments the genomes and calculates symmetrical identity scores for each pairwise comparison based on BLASTn hits and a genome length. The analysis revealed seven clusters of related genomes, which were generally consistent with those obtained in the phylogenomic analysis (Table S8). Notably, due to considerable sequence divergence, ARV1 falls into a separate cluster from other Italian rudiviruses. Consistently, in the phylogenomic tree, ARV1 forms a sister group to other Italian rudiviruses. Furthermore, among the five Italian rudiviruses, the genome of ARV1 is most divergent, displaying multiple gene and genomic segment inversions and relocations compared to the other viruses (Fig. 5). Viruses from the seven clades also differ considerably in terms of the variable gene contents. For instance, viruses SIRV1 and SIRV2 isolated in Iceland share 11 genes that are absent from the USA SIRVs [58]. Thus, to acknowledge the differences between the known rudiviruses, we propose to classify them into seven

genera: "*Icerudivirus*" (former *Rudivirus*, to include SIRV1-SIRV3), "*Mexirudivirus*" (SMRV1), "*Azorudivirus*" (SRV), "*Itarudivirus*" (ARV1), "*Hoswirudivirus*" (ARV2, ARV3, MRV1, SSRV1; *hoswi-*, for host switching), "*Japarudivirus*" (SBRV1) and "*Usarudivirus*" (SIRV4-SIRV11). We note that this classification would be consistent with the current practices of the International Committee on Taxonomy of Viruses (ICTV) to classify viruses based on their protein and genomic sequences [66].

Discussion

Here, we reported the results of our exploration of the diversity of hyperthermophilic archaeal viruses at the solfataric field in Pisciarelli, Pozzuoli. Previous sampling of archaeal viruses in the thermal springs in Pisciarelli led to the isolation of viruses from the families *Ampullaviridae*, *Bicaudaviridae*, *Lipothrixviridae*, *Rudiviridae* and *Tristromaviridae* [34, 67–70], whereas those of the families *Fuselloviridae*, *Globuloviridae* and *Guttaviridae*, which we observed in the initial enrichment cultures, have not been previously reported from the sampled Pisciarelli sites. Nevertheless, similar virion morphologies have been observed in high-temperature continental hydrothermal systems from other geographical locations across the globe [50, 71–74], pointing to global distribution of most groups of hyperthermophilic archaeal viruses. However, how these virus communities are structured and whether geographically remote hydrothermal ecosystems undergo virus immigration is not fully understood. Notably, metagenomics analysis of two hot springs located in Pozzuoli has shown that representatives of the *Lipothrixviridae*, *Bicaudaviridae*, *Ampullaviridae* and *Rudiviridae* families dominated the corresponding virus communities at the time of sampling, collectively amounting to over 90% of the sequencing reads [49]. We have observed virions representing all four virus families by TEM in our enrichment cultures, corroborating the results of the metagenomic sequencing, and could isolate three representatives of the *Rudiviridae* family.

A previous study has revealed a biogeographic pattern among *S. islandicus*-infecting rudiviruses isolated from hot springs in Iceland and United States [58]. That is, viruses from the same geographical location were more closely related to each other than they were to viruses from other locations and the larger the distance the more divergent the virus genomes were. Similar observation has been made for the *Sulfolobus*-infecting spindle-shaped viruses of the family *Fuselloviridae* [74–76]. By contrast, a study focusing on three relatively closely spaced hot springs in Yellowstone National Park concluded that horizontal virus movement, rather than mutation, is the dominant factor controlling the viral community structure [77].

Phylogenomic and comparative genomic analysis of the rudiviruses reported herein and those sequenced previously revealed a strong biogeographic pattern, suggesting that diversification and evolution of rudiviruses is influenced by spatial confinement within discrete high-temperature continental hydrothermal systems, with little horizontal migration of viral particles over large distances. Consistently, analyses of the CRISPR spacers carried by hyperthermophilic archaea predominantly target local viruses, further indicating geographically defined co-evolution of viruses and their hosts [74, 78]. This is in stark contrast with the global architecture of virus communities associated with hyperhalophilic archaea, where genomic similarity between viruses does not correspond to geographical distance [79, 80]. Notably, however, it has been suggested that reversible silicification of virus particles, which is conceivable in hot spring environments, might promote long-distance host-independent virus dispersal [81].

We also show, for the first time, that relatively closely related rudiviruses infect phylogenetically distant hosts, belonging to three different genera. Interestingly, whereas ARV3 could deliver its DNA exclusively into *Acidianus* cells, the genomes of SSRV1 and MRV1 were detected not only in their respective *Saccharolobus* and *Metallosphaera* hosts, but also in the non-host *Acidianus* cells (Fig. S3). Given the basal position of *Acidianus* rudiviruses (Fig. 6), this pattern is best consistent with the host switch events in the history of Italian rudiviruses, whereby the ancestral *Acidianus* virus gained the ability to infect *Saccharolobus* and *Metallosphaera* hosts. At least in the case of rudivirises, relatively few genetic changes appear to be necessary for gaining the ability to infect a new host. In tailed bacteriophages, host range switches are typically associated with mutations in genes encoding the tail fiber proteins responsible for host recognition [82]. Several molecular mechanisms underlying mutability of the tail fiber genes have been described, including genetic drift, diversity generating retroelements and phase variation cassettes. The latter two systems have been demonstrated to function also in viruses infecting anaerobic methane-oxidizing (ANME) and hyperhalophilic archaea, respectively [83, 84]. Both mechanisms depend on specific enzymes, namely, reverse-transcriptase and invertase. However, neither of the two systems is present in rudiviruses, suggesting that genetic drift is the most likely mechanism responsible for generating diversity in the gene(s) encoding receptor-binding protein(s) of rudiviruses.

Analysis of CRISPR spacers from hyperthermophilic archaea confirmed that viruses closely related to those isolated in this study were infecting highly different hosts in

the recent past. Indeed, viruses infecting *Saccharolobus* and *Metallosphaera* species are targeted by spacers from CRISPR arrays of *Metallosphaera* and *Saccharolobus*, respectively (Table S7). This observation further reinforces a relatively recent host switch event. Notably, this phenomenon appears to be also applicable to rudiviruses from other geographical locations. In particular, SBRV1, a rudivirus from a Japanese hot spring [60], was found to be targeted by 521 unique CRISPR spacers associated with all four principal CRISPR repeat sequences present in Sulfolobales [78], suggesting either very broad host range or frequent host switches. Furthermore, we have recently shown that CRISPR targeting is an important factor driving the genome evolution of hyperthermophilic archaeal viruses [78]. Given the presence of multiple CRISPR spacers matching rudivirus genomes, we hypothesize that necessity to switch hosts might be, to a large extent, driven by CRISPR targeting. Notably, matching of CRISPR spacers to protospacers in viral genomes is one of the widely used approaches of host identification for viruses discovered by metagenomics, with the estimated host genus prediction accuracy of 70–90% [62, 63, 85]. Our results suggest that in the case of rudiviruses, spacer matching might not provide accurate predictions beyond the rank of family (i.e., *Sulfolobaceae*).

Collectively, we show that terrestrial hydrothermal systems harbor a highly diverse virome represented by multiple families with unique virus morphologies not described in other environments. Genomes of the newly isolated viruses, especially those infecting *Pyrobaculum* species, remain a rich source of unknown genes, which could be involved in novel mechanisms of virus–host interactions. Furthermore, our results suggest that global rudivirus communities display biogeographic pattern and diversify into distinct lineages confined to discrete geographical locations. This diversification appears to involve relatively frequent host switching, potentially evoked by host CRISPR-Cas immunity systems. Future studies should focus on understanding the molecular changes allowing rudiviruses to efficiently infect and multiply in new hosts, attaining host range expansion and escaping CRISPR targeting.

Data availability

Genome sequences of the isolated viruses have been deposited in GenBank and their accession numbers are listed in Table 2.

Acknowledgements This work was supported by l'Agence Nationale de la Recherche (France) project ENVIRA (to M.K.) and the European Union's Horizon 2020 research and innovation program under grant agreement 685778, project VIRUS X (to D.P.). Y.L. is a recipient of the Pasteur-Roux-Cantarini Fellowship from Institut Pasteur. D.P.B. is part of the Pasteur—Paris University (PPU) International PhD Program, which has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No 665807. We are also grateful to the Ultrastructural BioImaging (UTechS UBI) unit of Institut Pasteur for access to electron microscopes and Marc Monot from the Biomics Platform of Institut Pasteur for helpful discussions on genome assembly.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

1. Prangishvili D, Bamford DH, Forterre P, Iranzo J, Koonin EV, Krupovic M. The enigmatic archaeal virosphere. Nat Rev Microbiol. 2017;15:724–39.
2. Munson-McGee JH, Snyder JC, Young MJ. Archaeal viruses from high-temperature environments. Genes (Basel). 2018;9: E128.
3. Wang H, Peng N, Shah SA, Huang L, She Q. Archaeal extrachromosomal genetic elements. Microbiol Mol Biol Rev. 2015;79:117–52.
4. Dellas N, Snyder JC, Bolduc B, Young MJ. Archaeal Viruses: Diversity, Replication, and Structure. Annu Rev Virol. 2014;1: 399–426.
5. Wang H, Guo Z, Feng H, Chen Y, Chen X, Li Z, et al. Novel Sulfolobus virus with an exceptional capsid architecture. J Virol. 2018;92:e01727–17.
6. Krupovic M, Quemin ER, Bamford DH, Forterre P, Prangishvili D. Unification of the globally distributed spindle-shaped viruses of the Archaea. J Virol. 2014;88:2354–8.
7. Contursi P, Fusco S, Cannio R, She Q. Molecular biology of fuselloviruses and their satellites. Extremophiles. 2014;18: 473–89.
8. Bath C, Dyall-Smith ML. His1, an archaeal virus of the Fuselloviridae family that infects Haloarcula hispanica. J Virol. 1998;72:9392–5.
9. Gorlas A, Koonin EV, Bienvenu N, Prieur D, Geslin C. TPV1, the first virus isolated from the hyperthermophilic genus Thermococcus. Environ Microbiol. 2012;14:503–16.
10. Geslin C, Le Romancer M, Erauso G, Gaillard M, Perrot G, Prieur D. PAV1, the first virus-like particle isolated from a hyperthermophilic euryarchaeote, "Pyrococcus abyssi". J Bacteriol. 2003;185:3888–94.
11. Kim JG, Kim SJ, Cvirkaite-Krupovic V, Yu WJ, Gwak JH, Lopez-Perez M, et al. Spindle-shaped viruses infect marine ammonia-oxidizing thaumarchaea. Proc Natl Acad Sci USA. 2019;116:15645–50.
12. Rice G, Stedman K, Snyder J, Wiedenheft B, Willits D, Brumfield S, et al. Viruses from extreme thermal environments. Proc Natl Acad Sci USA. 2001;98:13341–5.
13. Liu Y, Ishino S, Ishino Y, Pehau-Arnaudet G, Krupovic M, Prangishvili D. A novel type of polyhedral viruses infecting hyperthermophilic archaea. J Virol. 2017;91:e00589–17.
14. Wagner C, Reddy V, Asturias F, Khoshouei M, Johnson JE, Manrique P, et al. Isolation and characterization of Metallosphaera turreted icosahedral virus, a founding member of a new family of archaeal viruses. J Virol. 2017;91:e00925–17.

15. Wang F, Liu Y, Su Z, Osinski T, de Oliveira GAP, Conway JF, et al. A packing for A-form DNA in an icosahedral virus. Proc Natl Acad Sci USA. 2019;116:22591–7.
16. Maaty WS, Ortmann AC, Dlakic M, Schulstad K, Hilmer JK, Liepold L, et al. 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. 2006;80:7625–35.
17. Liu Y, Osinski T, Wang F, Krupovic M, Schouten S, Kasson P, et al. Structural conservation in a membrane-enveloped filamentous virus infecting a hyperthermophilic acidophile. Nat Commun. 2018;9:3360.
18. Hochstein R, Bollschweiler D, Dharmavaram S, Lintner NG, Plitzko JM, Bruinsma R, et al. Structural studies of Acidianus tailed spindle virus reveal a structural paradigm used in the assembly of spindle-shaped viruses. Proc Natl Acad Sci USA. 2018;115:2120–5.
19. Ptchelkine D, Gillum A, Mochizuki T, Lucas-Staat S, Liu Y, Krupovic M, et al. Unique architecture of thermophilic archaeal virus APBV1 and its genome packaging. Nat Commun. 2017; 8:1436.
20. Kasson P, DiMaio F, Yu X, Lucas-Staat S, Krupovic M, Schouten S, et al. Model for a novel membrane envelope in a filamentous hyperthermophilic virus. Elife. 2017;6:e26268.
21. DiMaio F, Yu X, Rensen E, Krupovic M, Prangishvili D, Egelman EH. Virology. A virus that infects a hyperthermophile encapsidates A-form DNA. Science. 2015;348:914–7.
22. Hong C, Pietila MK, Fu CJ, Schmid MF, Bamford DH, Chiu W. Lemon-shaped halo archaeal virus His1 with uniform tail but variable capsid structure. Proc Natl Acad Sci USA. 2015;112: 2449–54.
23. Stedman KM, DeYoung M, Saha M, Sherman MB, Morais MC. Structural insights into the architecture of the hyperthermophilic Fusellovirus SSV1. Virology. 2015;474:105–9.
24. Krupovic M, Cvirkaite-Krupovic V, Iranzo J, Prangishvili D, Koonin EV. Viruses of archaea: Structural, functional, environmental and evolutionary genomics. Virus Res. 2018;244:181–93.
25. Iranzo J, Krupovic M, Koonin EV. The double-stranded DNA virosphere as a modular hierarchical network of gene sharing. MBio. 2016;7:e00978–16.
26. Iranzo J, Koonin EV, Prangishvili D, Krupovic M. Bipartite network analysis of the Archaeal Virosphere: evolutionary connections between viruses and capsidless mobile elements. J Virol. 2016;90:11043–55.
27. Quemin ER, Chlanda P, Sachse M, Forterre P, Prangishvili D, Krupovic M. Eukaryotic-like virus budding in archaea. MBio. 2016;7:e01439–16.
28. Bize A, Karlsson EA, Ekefjard K, Quax TE, Pina M, Prevost MC, et al. A unique virus release mechanism in the Archaea. Proc Natl Acad Sci USA. 2009;106:11306–11.
29. Brumfield SK, Ortmann AC, Ruigrok V, Suci P, Douglas T, Young MJ. Particle assembly and ultrastructural features associated with replication of the lytic archaeal virus sulfolobus turreted icosahedral virus. J Virol. 2009;83:5964–70.
30. Quax TE, Krupovic M, Lucas S, Forterre P, Prangishvili D. The Sulfolobus rod-shaped virus 2 encodes a prominent structural component of the unique virion release system in Archaea. Virology. 2010;404:1–4.
31. Snyder JC, Brumfield SK, Peng N, She Q, Young MJ. Sulfolobus turreted icosahedral virus c92 protein responsible for the formation of pyramid-like cellular lysis structures. J Virol. 2011;85:6287–92.
32. Krupovic M, White MF, Forterre P, Prangishvili D. Postcards from the edge: structural genomics of archaeal viruses. Adv Virus Res. 2012;82:33–62.
33. Quemin ER, Pietila MK, Oksanen HM, Forterre P, Rijpstra WI, Schouten S, et al. Sulfolobus spindle-shaped virus 1 contains glycosylated capsid proteins, a cellular chromatin protein, and host-derived lipids. J Virol. 2015;89:11681–91.
34. Rensen EI, Mochizuki T, Quemin E, Schouten S, Krupovic M, Prangishvili D. A virus of hyperthermophilic archaea with a unique architecture among DNA viruses. Proc Natl Acad Sci USA. 2016;113:2478–83.
35. Wang F, Cvirkaite-Krupovic V, Kreutzberger MAB, Su Z, de Oliveira GAP, Osinski T, et al. An extensively glycosylated archaeal pilus survives extreme conditions. Nat Microbiol. 2019;4:1401–10.
36. Iranzo J, Lobkovsky AE, Wolf YI, Koonin EV. Evolutionary dynamics of the prokaryotic adaptive immunity system CRISPR-Cas in an explicit ecological context. J Bacteriol. 2013;195: 3834–44.
37. Athukoralage JS, McMahon SA, Zhang C, Grüschow S, Graham S, Krupovic M, et al. A viral ring nuclease anti-CRISPR subverts type III CRISPR immunity. Nature. 2020;577:572–5.
38. Bhoobalan-Chitty Y, Johansen TB, Di Cianni N, Peng X. Inhibition of type III CRISPR-Cas immunity by an archaeal virus-encoded anti-CRISPR protein. Cell. 2019;179:448–58 e11.
39. He F, Bhoobalan-Chitty Y, Van LB, Kjeldsen AL, Dedola M, Makarova KS, et al. Anti-CRISPR proteins encoded by archaeal lytic viruses inhibit subtype I-D immunity. Nat Microbiol. 2018;3:461–9.
40. Hurwitz S, Lowenstern JB. Dynamics of the Yellowstone hydrothermal system. Rev Geophys. 2014;52:375–411.
41. Munson-McGee JH, Peng S, Dewerff S, Stepanauskas R, Whitaker RJ, Weitz JS, et al. A virus or more in (nearly) every cell: ubiquitous networks of virus-host interactions in extreme environments. ISME J. 2018;12:1706–14.
42. Snyder JC, Bateson MM, Lavin M, Young MJ. Use of cellular CRISPR (clusters of regularly interspaced short palindromic repeats) spacer-based microarrays for detection of viruses in environmental samples. Appl Environ Microbiol. 2010;76:7251–8.
43. Bolduc B, Wirth JF, Mazurie A, Young MJ. Viral assemblage composition in Yellowstone acidic hot springs assessed by network analysis. ISME J. 2015;9:2162–77.
44. Piochi M, Kilburn CRJ, Di Vito MA, Mormone A, Tramelli A, Troise C, et al. The volcanic and geothermally active Campi Flegrei caldera: an integrated multidisciplinary image of its buried structure. Int Journ Earth Sci. 2014;103:401–21.
45. Piochi M, Mormone A, Balassone G, Strauss H, Troise C, De Natale G. Native sulfur, sulfates and sulfides from the active Campi Flegrei volcano (southern Italy): Genetic environments and degassing dynamics revealed by mineralogy and isotope geochemistry. J Volcano Geotherm Res. 2015;304:180–93.
46. Piochi M, Mormone A, Strauss H, Balassone G. The acid sulfate zone and the mineral alteration styles of the Roman Puteoli (Neapolitan area, Italy): clues on fluid fracturing progression at the Campi Flegrei volcano. Solid Earth. 2019;10:1809–31.
47. Menzel P, Gudbergsdottir SR, Rike AG, Lin L, Zhang Q, Contursi P, et al. Comparative metagenomics of eight geographically remote terrestrial hot springs. Micro Ecol. 2015;70:411–24.
48. Ciniglia C, Yoon HS, Pollio A, Pinto G, Bhattacharya D. Hidden biodiversity of the extremophilic Cyanidiales red algae. Mol Ecol. 2004;13:1827–38.
49. Gudbergsdóttir SR, Menzel P, Krogh A, Young M, Peng X. Novel viral genomes identified from six metagenomes reveal wide distribution of archaeal viruses and high viral diversity in terrestrial hot springs. Environ Microbiol. 2016;18:863–74.
50. Häring M, Peng X, Brugger K, Rachel R, Stetter KO, Garrett RA, et al. Morphology and genome organization of the virus PSV of the hyperthermophilic archaeal genera Pyrobaculum and

Thermoproteus: a novel virus family, the Globuloviridae. Virology. 2004;323:233–42.
51. Zillig W, Kletzin A, Schleper C, Holz I, Janekovic D, Hain J, et al. Screening for Sulfolobales, their Plasmids and their Viruses in Icelandic Solfataras. Syst Appl Microbiol. 1993;16:609–28.
52. Quemin ER, Lucas S, Daum B, Quax TE, Kuhlbrandt W, Forterre P, et al. First insights into the entry process of hyperthermophilic archaeal viruses. J Virol. 2013;87:13379–85.
53. Prangishvili D, Krupovic M, ICTV Report Consortium. ICTV virus taxonomy profile: *Globuloviridae*. J Gen Virol. 2018;99: 1357–8.
54. Ahn DG, Kim SI, Rhee JK, Kim KP, Pan JG, Oh JW. TTSV1, a new virus-like particle isolated from the hyperthermophilic crenarchaeote Thermoproteus tenax. Virology. 2006;351:280–90.
55. Prangishvili D, Rensen E, Mochizuki T, Krupovic M, ICTV Report Consortium. ICTV Virus Taxonomy Profile: Tristromaviridae. J Gen Virol. 2019;100:135–6.
56. Neumann H, Schwass V, Eckerskorn C, Zillig W. Identification and characterization of the genes encoding three structural proteins of the Thermoproteus tenax virus TTV1. Mol Gen Genet. 1989;217:105–10.
57. Prangishvili D, Koonin EV, Krupovic M. Genomics and biology of Rudiviruses, a model for the study of virus-host interactions in Archaea. Biochem Soc Trans. 2013;41:443–50.
58. Bautista MA, Black JA, Youngblut ND, Whitaker RJ. Differentiation and Structure in Sulfolobus islandicus Rod-Shaped Virus Populations. Viruses. 2017;9:E120.
59. Meier-Kolthoff JP, Goker M. VICTOR: genome-based phylogeny and classification of prokaryotic viruses. Bioinformatics. 2017;33: 3396–404.
60. Liu Y, Brandt D, Ishino S, Ishino Y, Koonin EV, Kalinowski J, et al. New archaeal viruses discovered by metagenomic analysis of viral communities in enrichment cultures. Environ Microbiol. 2019;21:2002–14.
61. Roux S, Krupovic M, Daly RA, Borges AL, Nayfach S, Schulz F, et al. Cryptic inoviruses revealed as pervasive in bacteria and archaea across Earth's biomes. Nat Microbiol. 2019;4:1895–906.
62. Roux S, Adriaenssens EM, Dutilh BE, Koonin EV, Kropinski AM, Krupovic M, et al. Minimum Information about an Uncultivated Virus Genome (MIUViG). Nat Biotechnol. 2019;37:29–37.
63. Edwards RA, McNair K, Faust K, Raes J, Dutilh BE. Computational approaches to predict bacteriophage-host relationships. FEMS Microbiol Rev. 2016;40:258–72.
64. Pourcel C, Touchon M, Villeriot N, Vernadet JP, Couvin D, Toffano-Nioche C, et al. CRISPRCasdb a successor of CRISPRdb containing CRISPR arrays and cas genes from complete genome sequences, and tools to download and query lists of repeats and spacers. Nucleic Acids Res. 2020;48:D535–44.
65. Ågren J, Sundström A, Håfström T, Segerman B. Gegenees: fragmented alignment of multiple genomes for determining phylogenomic distances and genetic signatures unique for specified target groups. PLoS ONE. 2012;7:e39107.
66. Simmonds P, Adams MJ, Benko M, Breitbart M, Brister JR, Carstens EB, et al. Consensus statement: Virus taxonomy in the age of metagenomics. Nat Rev Microbiol. 2017;15:161–8.
67. Häring M, Rachel R, Peng X, Garrett RA, Prangishvili D. 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. J Virol. 2005;79:9904–11.
68. Vestergaard G, Haring M, Peng X, Rachel R, Garrett RA, Prangishvili D. A novel rudivirus, ARV1, of the hyperthermophilic archaeal genus Acidianus. Virology. 2005;336:83–92.
69. Prangishvili D, Vestergaard G, Haring M, Aramayo R, Basta T, Rachel R, et al. Structural and genomic properties of the hyperthermophilic archaeal virus ATV with an extracellular stage of the reproductive cycle. J Mol Biol. 2006;359:1203–16.
70. Vestergaard G, Aramayo R, Basta T, Haring M, Peng X, Brugger K, et al. Structure of the acidianus filamentous virus 3 and comparative genomics of related archaeal lipothrixviruses. J Virol. 2008;82:371–81.
71. Arnold HP, Ziese U, Zillig W. SNDV, a novel virus of the extremely thermophilic and acidophilic archaeon Sulfolobus. Virology. 2000;272:409–16.
72. Mochizuki T, Sako Y, Prangishvili D. Provirus induction in hyperthermophilic archaea: characterization of Aeropyrum pernix spindle-shaped virus 1 and Aeropyrum pernix ovoid virus 1. J Bacteriol. 2011;193:5412–9.
73. Wiedenheft B, Stedman K, Roberto F, Willits D, Gleske AK, Zoeller L, et al. Comparative genomic analysis of hyperthermophilic archaeal Fuselloviridae viruses. J Virol. 2004;78:1954–61.
74. Pauly MD, Bautista MA, Black JA, Whitaker RJ. Diversified local CRISPR-Cas immunity to viruses of Sulfolobus islandicus. Philos Trans R Soc Lond B Biol Sci. 2019;374:20180093.
75. Held NL, Whitaker RJ. Viral biogeography revealed by signatures in Sulfolobus islandicus genomes. Environ Microbiol. 2009;11: 457–66.
76. Stedman KM, Clore A, Combet-Blanc Y (2006). Biogeographical diversity of archaeal viruses. In: Logan NA, Lappin-Scott HM, Oyston PCF (eds). *Prokaryotic Diversity: Mechanisms and Significance*. Cambridge University Press: Cambridge. pp 131–43.
77. Snyder JC, Wiedenheft B, Lavin M, Roberto FF, Spuhler J, Ortmann AC, et al. Virus movement maintains local virus population diversity. Proc Natl Acad Sci USA. 2007;104:19102–7.
78. Medvedeva S, Liu Y, Koonin EV, Severinov K, Prangishvili D, Krupovic M. Virus-borne mini-CRISPR arrays are involved in interviral conflicts. Nat Commun. 2019;10:5204.
79. Atanasova NS, Roine E, Oren A, Bamford DH, Oksanen HM. Global network of specific virus-host interactions in hypersaline environments. Environ Microbiol. 2012;14:426–40.
80. Atanasova NS, Bamford DH, Oksanen HM. Virus-host interplay in high salt environments. Environ Microbiol Rep. 2016;8: 431–44.
81. Laidler JR, Shugart JA, Cady SL, Bahjat KS, Stedman KM. Reversible inactivation and desiccation tolerance of silicified viruses. J Virol. 2013;87:13927–9.
82. De Sordi L, Khanna V, Debarbieux L. The gut microbiota facilitates drifts in the genetic diversity and infectivity of bacterial viruses. Cell Host Microbe. 2017;22:801–8, e3.
83. Paul BG, Bagby SC, Czornyj E, Arambula D, Handa S, Sczyrba A, et al. Targeted diversity generation by intraterrestrial archaea and archaeal viruses. Nat Commun. 2015;6:6585.
84. Klein R, Rossler N, Iro M, Scholz H, Witte A. Haloarchaeal myovirus phiCh1 harbours a phase variation system for the production of protein variants with distinct cell surface adhesion specificities. Mol Microbiol. 2012;83:137–50.
85. Roux S, Brum JR, Dutilh BE, Sunagawa S, Duhaime MB, Loy A, et al. Ecogenomics and potential biogeochemical impacts of globally abundant ocean viruses. Nature. 2016;537:689–93.

SUPPLEMENTARY INFORMATION

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

Diana P. Baquero[1,2], Patrizia Contursi[3], Monica Piochi[4], Simonetta Bartolucci[3], Ying Liu[1], Virginija Cvirkaite-Krupovic[1], David Prangishvili[1,*] and Mart Krupovic[1,*]

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

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

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

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

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

Archaeal Virology Unit, Institut Pasteur, 75015 Paris, France

Tel: 33 (0)1 40 61 37 22

Running title

Biogeographic pattern in archaeal virus communities

SUPPLEMENTARY MATERIALS AND METHODS

Enrichment cultures

Nine samples were collected from hot springs, mud pools and hydrothermally altered terrains at the solfataric field of the Campi Flegrei volcano in Pozzuoli, Italy — the study area known as Pisciarelli — with temperatures ranging from 81 to 96°C and pH between 1 and 7 (Table S1). Each sample was inoculated into medium favoring the growth of *Sulfolobus/Saccharolobus* (basal salt solution supplemented with 0.2% tryptone, 0.1% yeast extract, 0.1% sucrose, pH 3.5), *Acidianus* (basal salt solution supplemented with 0.2% tryptone, 0.1% yeast extract, sulfur, pH 3.5), and *Pyrobaculum* (0.1% yeast extract, 1% tryptone, 0.1-0.3% $Na_2S_2O_3 \times 5\ H_2O$, pH 7) species [1, 2]. The enrichment cultures were incubated for 10 days at 75°C, except for *Pyrobaculum* cultures, which were grown at 90°C for 15 days without shaking.

VLP concentration and purification

Following the removal of cells from the enrichment cultures (7,000 rpm, 20 min, Sorvall SLA-1500 rotor), VLPs were collected and concentrated by ultracentrifugation (40,000 rpm, 2h, 10°C, Beckman SW41 rotor). The particles were resuspended in buffer A: 20 mM KH_2PO_4, 250 mM NaCl, 2.14 mM $MgCl_2$, 0.43 mM $Ca(NO_3)_2$, <0.001% trace elements of *Sulfolobales* medium, pH 6 [3] and visualized by transmission electron microscopy (TEM) as described below. VLPs were further purified by centrifugation in a CsCl buoyant density gradient (0.45 g ml^{-1}) with a Beckman SW41 rotor at 39,000 rpm for 20 h at 10°C. All opalescent bands were collected with a needle and a syringe, dialyzed against buffer A and analyzed by TEM for the presence of VLPs.

Isolation of virus-host pairs

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

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

Transmission electron microscopy and VLP analysis

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

Infection experiments

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

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

Host range
The following laboratory collection strains were used to test their ability to replicate viruses MRV1, ARV3 and SSRV1: *A. convivator* [9], *A. hospitalis* W1 [10], *Saccharolobus solfataricus* strains P1 (GenBank accession no. NZ_LT549890) and P2 [11], *Sulfolobus islandicu*s strains REN2H1 [2], HVE10/4 [12], LAL14/1 [13], REY15A [12], ΔC1C2 [14], and *Sulfolobus acidocaldarius* strain DSM639 [15]. CsCl gradient-purified virus particles were added directly to the growing cultures of *Acidianus* species and virus propagation was evaluated by TEM. For *Sulfolobus* and *Saccharolobus* species, spot tests were performed and the replication of the virus was evidenced by the presence of an inhibition zone on the lawn of the tested strains.

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

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

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

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

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

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

Phylogenomic analysis

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

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

SUPPLEMENTARY REFERENCES

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

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

SUPPLEMENTARY FIGURES

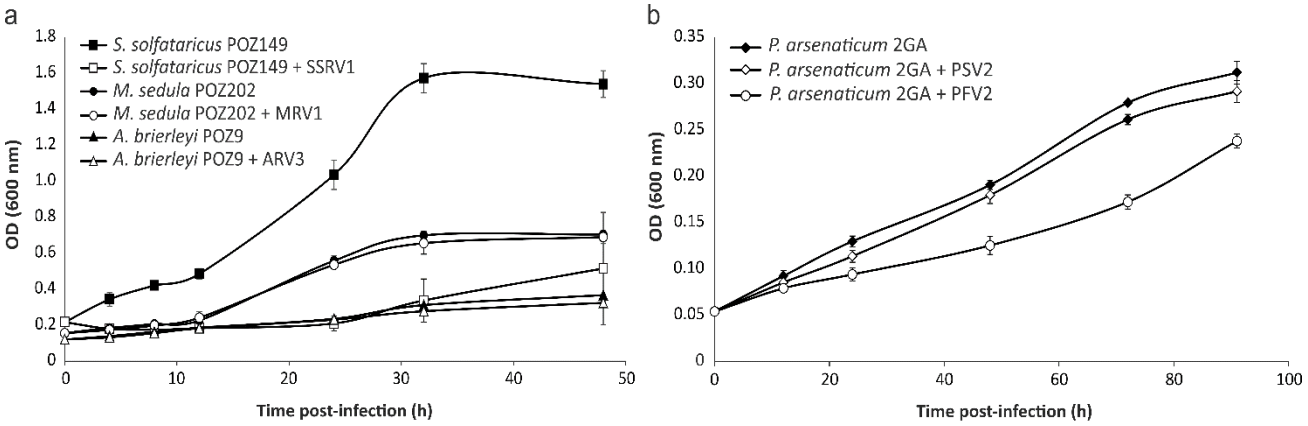


Figure S1. Impact of the isolated viruses on the growth of their respective hosts. a Growth curves of cell cultures infected (open symbols) or non-infected (filled symbols) with the three rudiviruses. **b** Growth curves of *P. arsenaticum* 2GA cells infected (open symbols) or non-infected (filled symbols)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

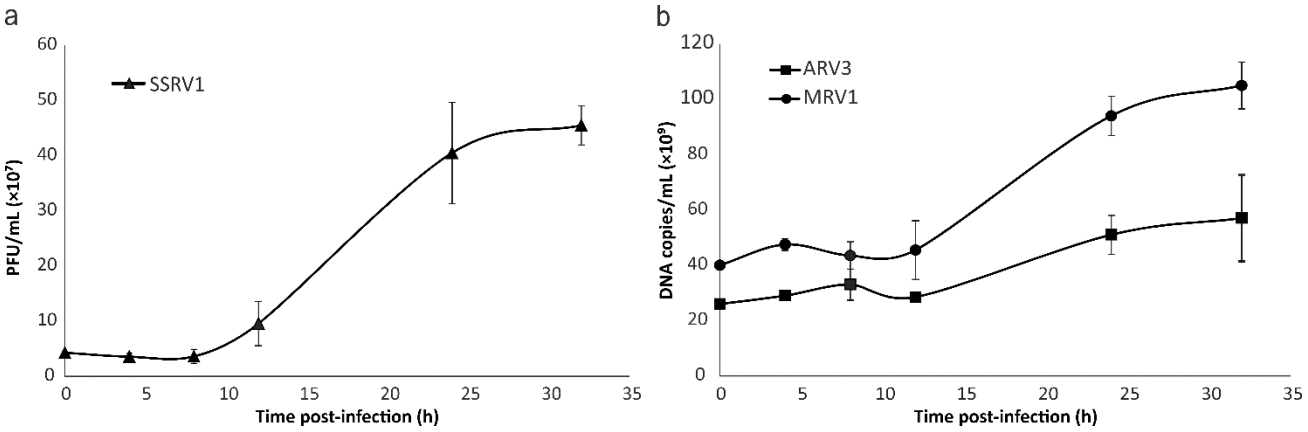


Figure S2. Quantification of virus production. a Production of SSRV1 particles was measured by titration and plaque assay at the indicated time points. **b** Release of ARV3 and MRV1 DNA copies was followed by qPCR over a period of 32 h by using primers targeting specific virus sequences. Error bars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

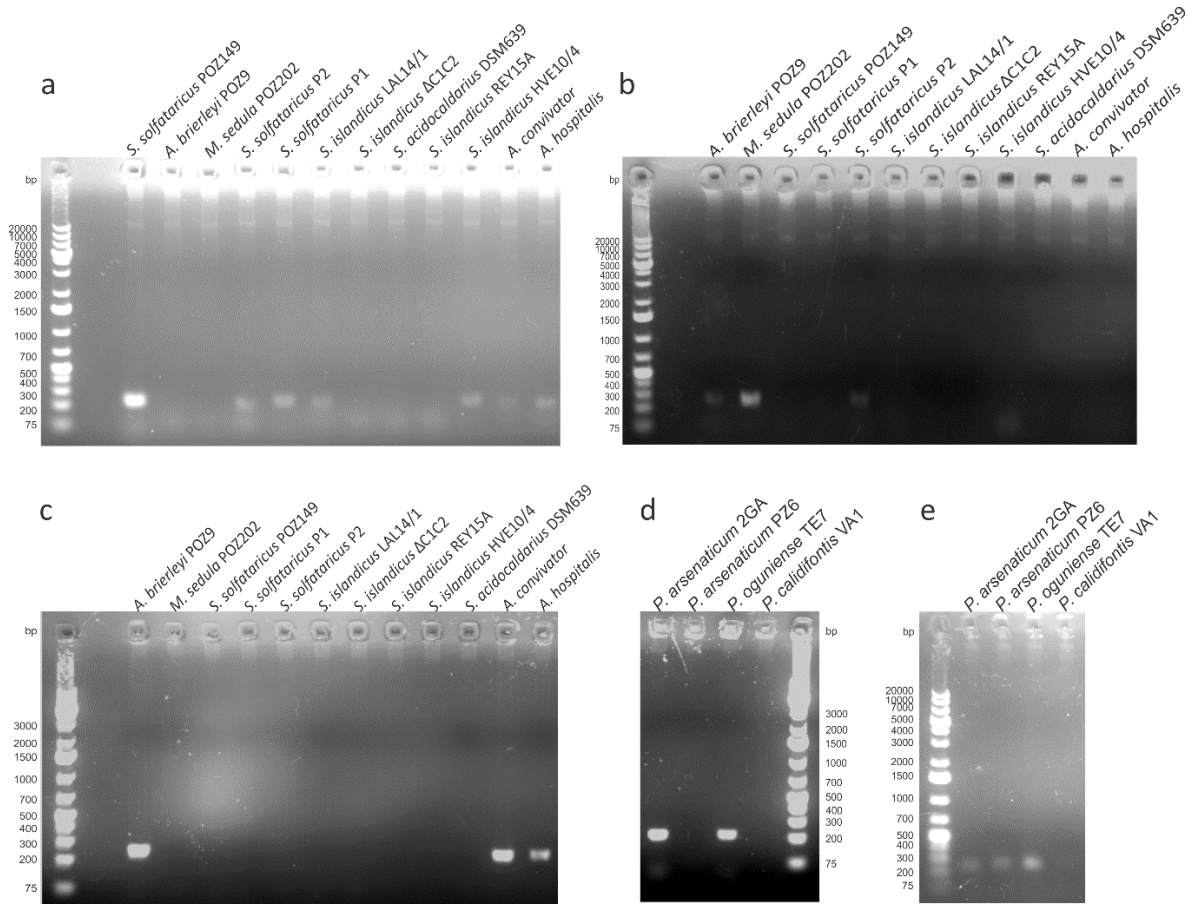


Figure S3. Entry of viral DNA into cells. Detection of SSRV1 (**a**), MRV1 (**b**), ARV3 (**c**), PFV2 (**d**) and PSV2 (**e**) DNA in different archaeal cells after 1 hpi. Members of the family *Sulfolobaceae* were mixed with the rudiviruses SSRV1, MRV1 and ARV3, whereas PFV2 and PSV2 were mixed with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

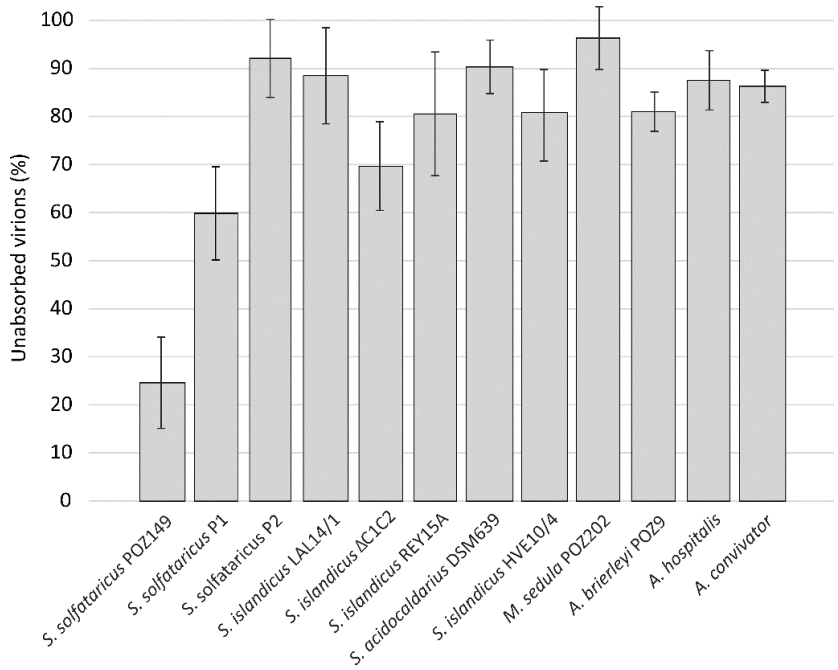


Figure S4. Adsorption of SSRV1 to different strains from the family *Sulfolobaceae*. Cells were mixed with SSRV1 at an MOI of 0.05 and kept at 75°C for 1 h. The number of unbound virions was measured by titration and plaque assay of the cell-free supernatants at 1 hpi. Error bars represent standard deviation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

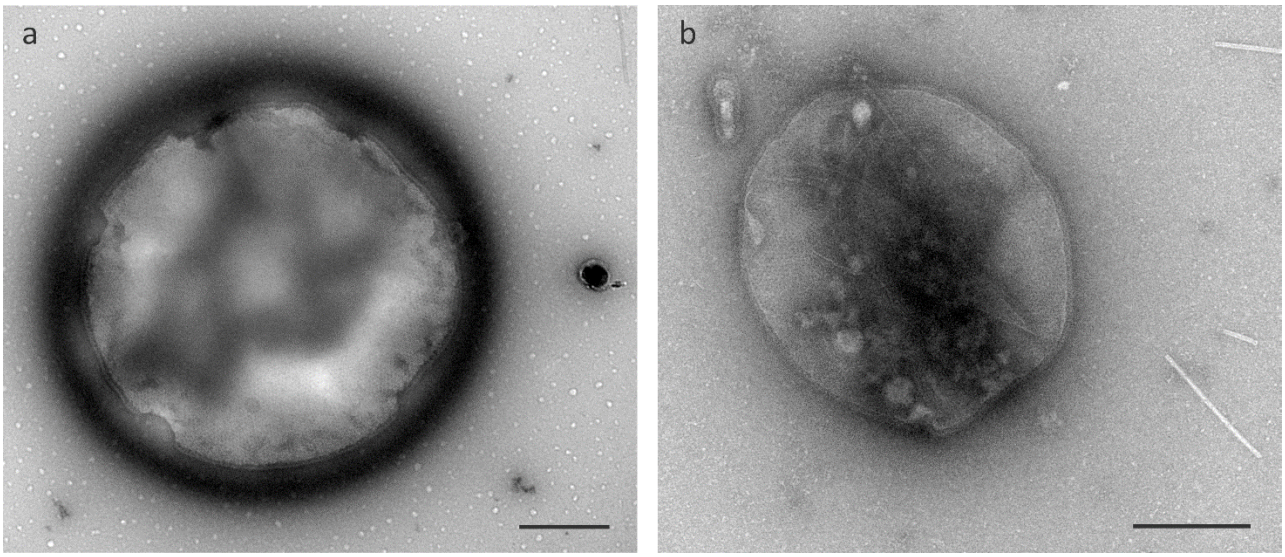


Figure S5. Transmission electron micrographs of infected cells at 32 hpi. a M*etallosphaera sedula* POZ202. **b** *Acidianus brierleyi* POZ9. The samples were negatively stained with 2% uranyl acetatse. Scale bars: 500 nm.

SUPPLEMENTARY TABLES

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

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

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

Table S2. Functional annotation of the PSV2 genome.

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

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

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

Table S3. Functional annotation of the PFV2 genome.

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

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

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

Table S4. Functional annotation of the SSRV1 genome.

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

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

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

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

Table S5. Functional annotation of the MRV1 genome.

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

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

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

Table S6. Functional annotation of the ARV3 genome.

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

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

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

Table S7. CRISPR spacer matches to the genomes of Italian rudiviruses and tristromavirus PFV2 obtained in the CRISPRdb database.

Virus	Alignment	Spacer source	Identity, %	E-value
ARV1	GAAGACAATTCGTTTTTTTTCAATCCAAGAGATGAAAAT ||||| ||||| ||| |||||||||||||||| GAAGAAAATTCATTTATTTTCAATCCAAGAGATGAAAAT	*Saccharolobus solfataricus* 98/2	92	3e-05
ARV2	TCAGATTCTGGCAGTTGATCCACAAATTTCCTAATT ||||||||||||||||||||||||||||| TCAGATTCTGGCAGTTGATCCACAAATTTCCTAATT	*Saccharolobus solfataricus* P1	100	4e-10
	TAAACCGCAACCTTTTATAAAATCTTCAACTGTAA ||||||||||||||||||||||||||||| TAAACCGCAACCTTTTATAAAATCTTCAACTGTAA	*Metallosphaera sedula* DSM 5348	100	1e-09
	CGACATTTTCATTATGGATTGTGAGGGATGTGA |||||||||||||||||||||||||||| CGACATTTTCATTATGGATTGTGAGGGATGTGA	*Saccharolobus solfataricus* 98/2	100	2e-08
	CCAATTCTTGCAAGTGTTGTCCAAATATTTGACCGAAAA |||||||||| |||||||||||||||| |||| CCAATTCTTGTAAGTGTTGTCCAAATATTTGACC-AAAA	*Saccharolobus solfataricus* 98/2	95	6e-08
	AATTTCAAACTTGCAAACGCAACAATTAAATA |||||||||||||||||||||||||||| AATTTCAAACTTGCAAACGCAACAATTAAATA	*Saccharolobus solfataricus* 98/2	100	6e-08
ARV3	TCATCTACGGTTGCAAGTTCCATAGCCAAGTTGTCAAG ||||||||||||||||||||||||||||| TCATCTACGGTTGCAAGTTCCATAGCCAAGTTGTCAAG	*Saccharolobus solfataricus* P1	100	3e-11
	TCAAAAGGTCTTATTTTCGATGAAATTCAAACTTGGAAA ||||| ||||| ||||| |||||||||||||| TCAAATGGTCTAATTTTTGATGAAATTCAAACTTGGAAA	*Sulfolobus islandicus* HVE10/4	92	3e-05
	GAAGAAAACTCATTTTTTTTCAATCCTAGAGATGAAAAT |||||||| |||||| ||||||||| ||||||||||| GAAGAAAATTCATTTATTTTCAATCCAAGAGATGAAAAT	*Saccharolobus solfataricus* 98/2	92	3e-05
	AATTGTAGGTGCAGATTGCGGATCATCAGCTTTGTATTT |||||||| |||||||||| ||||| |||||||||| AATTGTAGGGGCAGATTGCGGTTCATCGGCTTTGTATTT	*Saccharolobus solfataricus* P2	92	3e-05
	GCAGTTTTCGTAACCCTTCTAATAAAATACGATCTGTC |||||||||||| ||||||||||||| |||||||| GCAGTTTTCGTAATTCTTCTAATAAAATAAGATCTGTC	*Sulfolobus islandicus* HVE10/4	92	1e-04
SSRV1	TCTTTAGTCGCTATAATGTCGGGTAATGGTTGCTTACCAG |||||||||||||||||||||||||||||||||| TCTTTAGTCGCTATAATGTCGGGTAATGGTTGCTTACCAG	*Metallosphaera sedula* DSM 5348	100	3e-12
	AATATTTGTAAATTTCATCAAACATTTTCCTAACTTTATC |||||||||||||||||||||||||||||||||| AATATTTGTAAATTTCATCAAACATTTTCCTAACTTTATC	*Metallosphaera sedula* DSM 5348	100	3e-12
	CTATGAAATTATCGCATTCGTATGCAAAATATTTCTGGATCT ||||||||| |||||||||||||||||||||||| CTATGAAATTGTCGCATTCGTATGCAAAATATTTCTGGATCT	*Metallosphaera sedula* DSM 5348	98	3e-11
	CTATGAAATTATCGCATTCGTATGCAAAATATTTCTGGATCT ||||||||| |||||||||||||||||||||||| CTATGAAATTGTCGCATTCGTATGCAAAATATTTCTGGATCT	*Metallosphaera sedula* DSM 5348	98	3e-11
	TGAAAATATTCCAGCGTTATACTTTTGCTTAATGTACTCTTT ||||||||| |||||||||||||||||||||||| TGAAAATATTTCAGCGTTATACTTTTGCTTAATGTACTCTTT	*Metallosphaera sedula* DSM 5348	98	3e-11
MRV1	ATTTTCAGCTGAAAATTTGAAATCTGTAGATTT |||||||||||||||||||||||||||||| ATTTTCAGCTGAAAATTTGAAATCTGTAGATTT	*Saccharolobus solfataricus* P2	100	1e-08
	ATGCCGACATTTTCATTATGGATTGTGAAGGATGTGA ||||||||||||||||||||||||| |||||||| ATGCCGACATTTTCATTATGGATTGTGAGGGATGTGA	*Saccharolobus solfataricus* 98/2	97	1e-08
	AAAAAATATGATATTAAGTTTTTATACAAAAGATTTCTA |||||||||||||| || |||||||||||||||| AAAAAATATGATATAAACTTTTTATACAAAAGATTTCTA	*Sulfolobus* sp. A20	95	2e-07
	AATTTACGATATTATACCAGTTGAACCTGCAAAAAATGT ||||||||||||||||| ||||| ||||||||||| AATTTACGATATTATACCGGTTGAGCCTGCAAAAAATGT	*Saccharolobus solfataricus* P2	95	2e-07
	AAGTTAGGTTGCGAAACTGAAAAAGCAATTGAATATGA |||||||||||||| ||||| |||||||||||||| AAGTTAGGTTGCGACACTGAGAAAGCAATTGAATATGA	*Saccharolobus solfataricus* P1	95	6e-07
PFV2	TGGGTCTGGTCTTCTAACCGGTTTTA-GCGGGTTTTTGTGGG |||||||| ||||||| || |||| | ||||||||||| TGGGTCTGGGCTTCTAACTGGCTTTATG-GGGTTTTTGTGGG	*Pyrobaculum oguniense* TE7	88	9e-04

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

	SMRV1	SRV	ARV1	ARV3	MRV1	ARV2	SSRV1	SBRV1	SIRV1	SIRV2	SIRV3	SIRV10	SIRV8	SIRV9	SIRV4	SIRV5	SIRV6	SIRV7	SIRV11	
SMRV1	100	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	*Mexirudivirus*
SRV	0	100	0	0	0	0	0	0	0	0	0	0	43	0	0	0	0	0	0	*Azorudivirus*
ARV1	0	0	100	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	*Itarudivirus*
ARV3	0	0	0	100	51	52	52	0	0	0	0	0	0	0	0	0	0	0	0	
MRV1	0	0	0	51	100	58	58	0	0	0	0	0	0	0	0	0	0	0	0	*Hoswirudivirus*
ARV2	0	0	0	53	58	100	60	0	0	0	0	0	0	0	0	0	0	0	0	
SSRV1	0	0	0	51	58	60	100	0	0	0	0	0	0	0	0	0	0	0	0	
SBRV1	0	0	0	0	0	0	0	100	41	0	41	0	40	40	0	0	0	0	42	*Japarudivirus*
SIRV1	0	0	0	0	0	0	0	0	100	71	71	49	49	49	47	46	46	46	46	
SIRV2	0	0	0	0	0	0	0	0	71	100	72	48	50	51	44	49	49	49	45	*Icerudivirus*
SIRV3	0	0	0	0	0	0	s0	41	70	72	100	51	52	55	47	47	47	47	43	
SIRV10	0	0	0	0	0	0	0	0	49	47	51	100	77	81	66	70	69	70	66	
SIRV8	0	42	0	0	0	0	0	43	50	49	49	77	100	93	67	68	68	67	65	
SIRV9	0	0	0	0	0	0	0	42	48	51	51	81	93	100	67	70	69	69	66	
SIRV4	0	0	0	0	0	0	0	0	48	43	49	66	67	67	100	83	84	82	85	*Usarudivirus*
SIRV5	0	0	0	0	0	0	0	0	48	49	46	69	68	70	83	100	100	97	83	
SIRV6	0	0	0	0	0	0	0	0	45	49	47	69	67	70	83	100	100	97	83	
SIRV7	0	0	0	0	0	0	0	0	47	47	47	70	67	70	82	97	97	100	82	
SIRV11	0	0	0	0	0	0	0	0	49	42	44	65	66	66	85	84	84	81	100	

6.2 CHAPTER 2

Structural and biochemical characterization of the filamentous archaeal viruses SSRV1 and SIFV

Contribution: I produced highly concentrated samples of viruses SIFV and SSRV1 for cryo-EM analyses. I analysed the protein constituents of the two viruses by SDS-PAGE, prepared the biological materials for mass-spectrometry analysis and analysed the mass-spectrometry data.

Structures of filamentous viruses infecting hyperthermophilic archaea explain DNA stabilization in extreme environments

Fengbin Wang[a,1], Diana P. Baquero[b,c,1], Leticia C. Beltran[a], Zhangli Su[a], Tomasz Osinski[a], Weili Zheng[a], David Prangishvili[b,d], Mart Krupovic[b,2], and Edward H. Egelman[a,2]

[a]Department of Biochemistry and Molecular Genetics, University of Virginia, Charlottesville, VA 22908; [b]Archaeal Virology Unit, Department of Microbiology, Institut Pasteur, 75015 Paris, France; [c]Collège Doctoral, Sorbonne Universités, 75005 Paris, France; and [d]Academia Europaea Tbilisi Regional Knowledge Hub, Ivane Javakhishvili Tbilisi State University, 0179 Tbilisi, Georgia

This contribution is part of the special series of Inaugural Articles by members of the National Academy of Sciences elected in 2019.

Contributed by Edward H. Egelman, June 26, 2020 (sent for review June 1, 2020; reviewed by Tanmay A. M. Bharat and Jack E. Johnson)

Living organisms expend metabolic energy to repair and maintain their genomes, while viruses protect their genetic material by completely passive means. We have used cryo-electron microscopy (cryo-EM) to solve the atomic structures of two filamentous double-stranded DNA viruses that infect archaeal hosts living in nearly boiling acid: *Saccharolobus solfataricus* rod-shaped virus 1 (SSRV1), at 2.8-Å resolution, and *Sulfolobus islandicus* filamentous virus (SIFV), at 4.0-Å resolution. The SIFV nucleocapsid is formed by a heterodimer of two homologous proteins and is membrane enveloped, while SSRV1 has a nucleocapsid formed by a homodimer and is not enveloped. In both, the capsid proteins wrap around the DNA and maintain it in an A-form. We suggest that the A-form is due to both a nonspecific desolvation of the DNA by the protein, and a specific coordination of the DNA phosphate groups by positively charged residues. We extend these observations by comparisons with four other archaeal filamentous viruses whose structures we have previously determined, and show that all 10 capsid proteins (from four heterodimers and two homodimers) have obvious structural homology while sequence similarity can be nonexistent. This arises from most capsid residues not being under any strong selective pressure. The inability to detect homology at the sequence level arises from the sampling of viruses in this part of the biosphere being extremely sparse. Comparative structural and genomic analyses suggest that nonenveloped archaeal viruses have evolved from enveloped viruses by shedding the membrane, indicating that this trait may be relatively easily lost during virus evolution.

cryo-EM | extremophiles | hyperthermophilic archaea | filamentous viruses

Viruses may constitute the largest source of genetic diversity on Earth (1). They greatly impact human health but also have large and direct impacts on the ecosystem of the planet (2). Random sampling of ocean and lake water led to the surprising observation that the concentration of viruses was on the order of 10^8 per milliliter in these aquatic environments (3), and current estimates for the number of viruses in the Earth's oceans are on the order of 10^{30} (2), with estimates of 10^{31} viruses on the Earth in total (4), compared with $\sim 10^{23}$ stars in the universe. Studying viruses has thus been of great interest in many areas of biology and has had an enormous impact on structural biology due to the relative simplicity of virus particles compared to any cellular organisms (5).

Many pathways for DNA repair exist in all living organisms, with some conserved from bacteria to humans (6). These mechanisms are essential to preserve the integrity of genomes due to both exogenous as well as endogenous sources of DNA damage (7). In contrast, when outside the host, viruses must depend upon entirely passive means to protect their genomes using a very limited number of different structural proteins. In some instances, there may only be a single protein (present in many copies) in a virion (8). Viruses that infect archaea have been of particular interest (9–12), as many of these have been found in the most extreme aquatic environments on earth: Temperatures of 80 to 90 °C with pH values of ∼2 to 3. How DNA genomes can be passively maintained in such conditions is of interest not only in terms of evolutionary biology and the origin of life on Earth (1, 13), but has potential consequences for everything from biotechnology (14) to human therapeutics (15).

We have previously used cryo-electron microscopy (cryo-EM) to determine the atomic structures of four archaeal filamentous viruses that infect hyperthermophiles (16–19). One of these, SIRV2, had a protein coat composed of a homodimer of the major capsid protein (MCP) (19), while AFV1 (18), SFV1 (17), and PFV2 (16) had protein coats composed of heterodimers. AFV1, SFV1, and PFV2 are membrane enveloped, while SIRV2 is not. These viruses are classified into three different families, *Rudiviridae* (SIRV2), *Lipothrixviridae* (AFV1 and SFV1), and *Tristromaviridae* (PFV2), with families *Rudiviridae* and *Lipothrixviridae*

Significance

Viruses pose enormous threats to human health around the world but can also be potential tools for everything from gene therapy to medical imaging to drug delivery. We have used cryo-electron microscopy to determine the atomic structure of two filamentous viruses that infect hosts living in some of the most extreme environments on the Earth's surface, springs of nearly boiling acid. We show how the structure of the proteins that protect the viral DNA has been conserved in a family of filamentous viruses found in diverse extreme locations around the globe, suggesting that all had a common origin. These results have implications for understanding the evolution of viruses, as well as for developing new methods for packaging DNA in biotechnology.

Author contributions: F.W., D.P., M.K., and E.H.E. designed research; F.W., D.P.B., L.C.B., W.Z., D.P., and M.K. performed research; Z.S. and T.O. contributed new reagents/analytic tools; F.W., D.P.B., Z.S., M.K., and E.H.E. analyzed data; and F.W., M.K., and E.H.E. wrote the paper.

Reviewers: T.A.M.B., University of Oxford; and J.E.J., The Scripps Research Institute.

The authors declare no competing interest.

[1]F.W. and D.P.B. contributed equally to this work.

[2]To whom correspondence may be addressed. Email: mart.krupovic@pasteur.fr or egelman@virginia.edu.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2011125117/-/DCSupplemental.

First published August 5, 2020.

unified into the order *Ligamenvirales* (20). Furthermore, a class "*Tokiviricetes*" has been proposed for unification of the three families of structurally related viruses (16). Here, we extend these observations by solving the structure of the nonenveloped *Saccharolobus solfataricus* rod-shaped virus 1 (SSRV1) (21) at 2.8-Å resolution, the highest resolution yet achieved for these filamentous viruses, and solving the structure of the membrane-enveloped *Sulfolobus islandicus* filamentous virus (SIFV) (22) at 4.0-Å resolution. With six structures from this group of filamentous viruses, we can make a number of comparisons providing insights into what has been conserved and what has diverged. Although the sequence similarity of the protein subunits within this virus assemblage can be undetectable (17), all share a relatively conserved fold that encapsidates A-form DNA, demonstrating common ancestry. The high resolution of the SSRV1 structure allows us to look with greater confidence at how the A-form is maintained in the virion. While it has been reasonable to assume that enveloped and nonenveloped viruses must have diverged early due to entirely different mechanisms of entry and egress from cells (23), we can show striking similarities between enveloped and nonenveloped archaeal filamentous viruses. It is also clear that other families of archaeal filamentous viruses infecting hyperthermophiles are likely to exist, as the structure of APBV1 (24) shows no apparent homology with the six viruses we now describe.

Results

Two Major Capsid Protein Homologs in SSRV1. All members of the proposed genus "*Hoswirudivirus*," including SSRV1, encode two proteins homologous to the MCP of SIRV2 (21). The two proteins, SSRV1 gp10 (108 aa; QJF12286) and gp11 (134 aa; QJF12287), are encoded next to each other and can be aligned in their C-terminal regions (22% identity over 74 aa; E = 2e-09), whereas the characteristic N-terminal region is missing in gp10. To investigate whether the virions of rudiviruses in the genus "*Hoswirudivirus*" are constructed from a heterodimer, similar to lipothrixviruses and tristromaviruses, or from a homodimer, like in all other known rudiviruses, SSRV1 virions were purified and subjected to SDS/PAGE (*SI Appendix*, Fig. S1) and mass spectrometry analyses, which showed that only gp11 protein is present in the virions. This result indicates that, if expressed, gp10 is not incorporated into the SSRV1 virions at detectable levels.

Cryo-EM of SSRV1. From cryo-EM images of SSRV1 (Fig. 1*A*), 470,216 overlapping segments were extracted and used for a three-dimensional (3D) reconstruction. The possible helical symmetries were determined from an averaged power spectrum (Fig. 1*B*), which showed multiple orders of the 1-start helix (n = 1, 2, 3, 4...) as well as additional layer lines. Indexing the additional layer lines was done by trial and error, and only one symmetry (~14.67 subunits per turn of the 43.1-Å pitch 1-start helix) yielded a reconstruction with interpretable features,

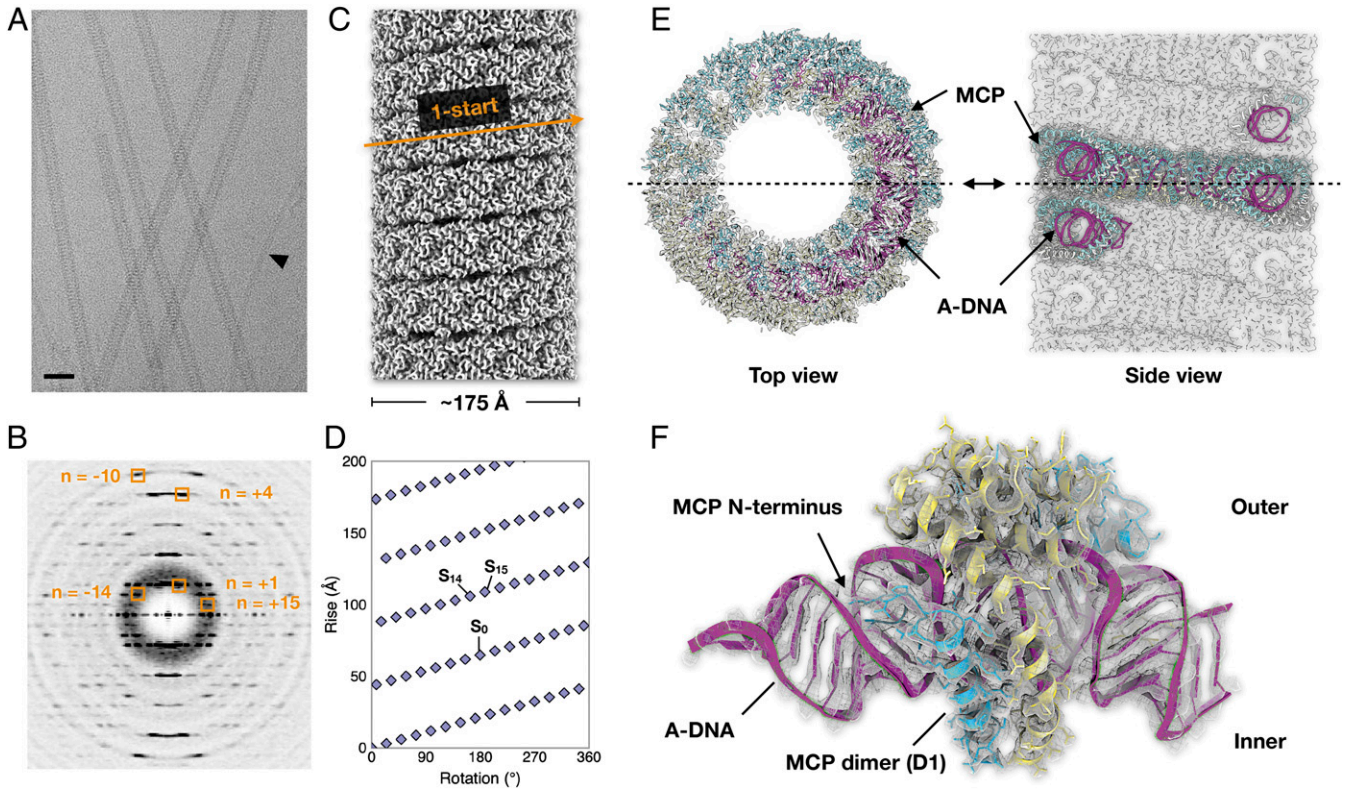


Fig. 1. Cryo-EM of the *S. solfataricus* rod-shaped virus (SSRV1). (*A*) Representative cryo-electron micrograph of SSRV1 virions. A *Saccharolobus solfataricus* Type IV pilus, indicated by black arrowhead, is the cellular receptor for SSRV1. (Scale bar, 50 nm.) (*B*) Averaged power spectrum of the segments used in the IHRSR reconstruction. The layer lines that were used to initially estimate the helical symmetry are indicated, along with the Bessel orders for the correct symmetry found. (*C*) Surface of the SSRV1 cryo-EM reconstruction at 2.8-Å resolution. The right-handed ~43-Å pitch 1-start helix passing through every asymmetric unit is shown. (*D*) Helical net of SSRV1, using the convention that the surface is unrolled and we are viewing it from the outside. Since there are 14.67 subunits per turn of the right-handed 1-start helix, subunits S_{14} and S_{15} will be above the reference subunit S_0 indicated. (*E*) Top and side views of the SSRV1 atomic model fit into the EM map. The side view is from a central slice of the map/model, indicated by the dashed line. A-form DNA is colored magenta, and MCP dimers are cyan and yellow. (*F*) Ribbon models for the SSRV1 protein dimer and 36 bp of A-DNA, fit into the EM map. The clear separation of DNA base pairs is seen in the map, despite averaging over the entire genomic sequence with the imposition of helical symmetry.

including side chains (Fig. 1 *C*–*F*). It was apparent that the asymmetric unit contained a symmetrical dimer of coat proteins wrapped around 12 bp of A-form DNA, so a dihedral symmetry was imposed on the overall reconstruction. This dihedral symmetry relates the 5′-to-3′ strand of the DNA to the 3′-to-5′ strand. Although the helical symmetry imposed will average together 12 bp of DNA throughout the genome, the separation of these base pairs is still preserved very well in the reconstructed volume (Fig. 1*F*) even though all sequence information has been lost. The DNA supercoils in the virion with one right-handed supercoil for every 43.1-Å turn of the protein helix (Fig. 1 *C* and *E*). Since there are 14.67 asymmetric units per turn of this 1-start helix, there are 44 asymmetric units in three turns containing 528 bp. The native twist of the DNA is therefore 528 bp in 47 turns (44 local turns plus three right-handed supercoil turns), which is 11.2 bp/turn, similar to what was found for SIRV2 (19). The helical radius of the DNA solenoidal supercoil is ∼60 Å.

Since the EM reconstruction reached 2.8-Å resolution (*SI Appendix*, Figs. S2 and S3), it was possible to do an unambiguous C_α trace through the density map resulting in a chain containing 131 residues, confirming that the 108-aa-long gp10 does not function as an MCP. The MCP of SSRV1, gp11, shares 80% sequence identity with the MCP of SIRV2. A SIRV2-based homology model was then fit and refined into the map (Table 1 and Fig. 1 *E* and *F*). The overall fold of the SSRV1 model is very similar with that of SIRV2, but with a more accurate model due to the much higher resolution (2.8 versus 4 Å).

Cryo-EM of SIFV. The membrane-enveloped SIFV had previously been demembranated, imaged by negative staining and reconstructed, but no high-resolution information was obtained (22). Here, we reexamined SIFV using cryo-EM, and it was reconstructed by a similar approach as described above. From cryo-electron micrographs (Fig. 2*A*), 167,541 overlapping segments were extracted and processed. The averaged power spectrum (Fig. 2*B*) showed that the pitch of the 1-start helix was 51.3 Å, significantly larger than the 43.1-Å pitch of SSRV1. Again, the symmetry was determined by trial and error, and the correct symmetry was 9.35 subunits per turn of this 51.3-Å pitch helix (Fig. 2 *C*–*E*). In contrast to the symmetrical protein homodimer in SSRV1, there was a pseudosymmetrical heterodimer formed by MCP1 and MCP2 in SIFV (Figs. 1*F*, 2*F*, and *SI Appendix*, Fig. S2), and as a consequence the virions lacked the overall dihedral symmetry in SSRV1 and were polar (Fig. 2*F*). The final EM map had ∼4.0-Å resolution (*SI Appendix*, Figs. S2 and S3), and an unambiguous C_α trace was made through the density map for both MCPs within one asymmetric unit. Each asymmetric unit also contained 12 bp of A-form DNA. The DNA supercoils in SIFV with one right-handed supercoil for every 51.3-Å turn of the protein helix (Fig. 2 *C* and *E*). Since there are 9.35 asymmetric units per turn of this 1-start helix, there are ∼28 asymmetric units in three turns containing 336 bp. The native twist of the DNA is therefore 336 bp in 31 turns (28 local turns plus three right-handed supercoil turns), which is 10.8 bp/turn, similar to what was found for AFV1 (18). The values of 10.8 for SIFV and 11.2 for SSRV1 bracket 11.0 bp per turn, the canonical value for A-form DNA (25). The helical radius of the DNA solenoidal supercoil is ∼40 Å.

The membrane can occasionally be seen to be absent in SIFV (Fig. 2*A*, white arrowhead), and we suggest that this is an artifact of cryo-EM sample preparation, which results from the fluid flow and large shear forces that can be present just prior to vitrification (26). Virions lacking the membrane appear to be more flexible, as was also observed for AFV1 (18). Since the membrane is fluid, we suggest that the greater rigidity of the virions that are enveloped is not due to the rigidity of the membrane, per se, but rather that the presence of the membrane constrains the protein. Although helical symmetry has been imposed on the membrane during the 3D reconstruction, no features appeared in the membrane that might be due to perturbations of the membrane by the protein subunits, or integral membrane proteins that were aligned with the capsid proteins. However, missing residues 122 to 132 appear to form a loop that extends out from the nucleoprotein core to the membrane (Fig. 2*G*), which has not been seen in the three other membrane-enveloped viruses: AFV1, SFV2, and PFV2. This loop does not appear to generate ordered density in the membrane at this location, and the resolution of the density connecting the nucleoprotein capsid to the membrane is too poor to allow for an unambiguous chain trace in this region. We therefore cannot say whether this loop inserts into the membrane, interacts with the membrane headgroups, or is associated with an integral membrane protein that is disordered. The membrane thus appears in all respects as a fluid, and the radial density profile of the membrane (Fig. 2*C*) is shown after cylindrical averaging and low-pass filtering. As with AFV1 (18), the membrane is anomalously thin (∼20 Å).

Protein–DNA Interactions. Similar to what was previously observed for SFV1 (17), it is clear from the EM density that the MCPs in both SSRV1 and SIFV contain N-terminal portions that project into a DNA groove (Fig. 3*A*). While some of the residues in these regions are disordered, we can clearly see the side-chain densities for a few polar N-terminal residues in SSRV1, such as Lys3, Arg5, and Arg8, as they are making contacts with the DNA phosphate backbone (Fig. 3 *A* and *B*). Such side-chain densities are not observed in SIFV, even though both MCPs of SIFV are also arginine and lysine rich in their N terminus (Fig. 3*A*). This is presumably due to greater flexibility and the resulting lower resolution in that region. Similarly, since SIRV2 is structurally very similar to SSRV1 and AFV1 is similar to SIFV, it is very likely their MCPs also insert into a DNA groove, but this was not observed previously due to limited resolution. Interestingly, in the tristromavirus PFV2, a virus evolutionarily divergent from

Table 1. Cryo-EM and refinement statistics of the SSRV1 and SIFV filaments

Parameter	SSRV1	SIFV
Data collection and processing		
Voltage, kV	300	300
Electron exposure, e^{-}·Å^{-2}	50	44
Pixel size, Å	1.08	1.4
Particle images, *n*	470,216	167,541
Helical symmetry		
Point group	D1	C1
Helical rise, Å	2.94	5.48
Helical twist, °	24.53	38.49
Map resolution, Å		
Map:map FSC (0.143)	2.7	3.9
Model:map FSC (0.38)	2.8	4.0
d_{99}	2.8	4.3
Refinement and model validation		
Map-sharpening B-factor, Å^2	−122	−185
Bond lengths rmsd, Å	0.007	0.005
Bond angles rmsd, °	0.692	0.704
Clashscore	3.71	18.8
Poor rotamers, %	0	0
Ramachandran favored, %	98.4	91.9
Ramachandran outlier, %	0	0
MolProbity score	1.16	2.26
Deposition ID		
PDB (model)	6WQ0	6WQ2
EMDB (map)	EMD-21867	EMD-21868

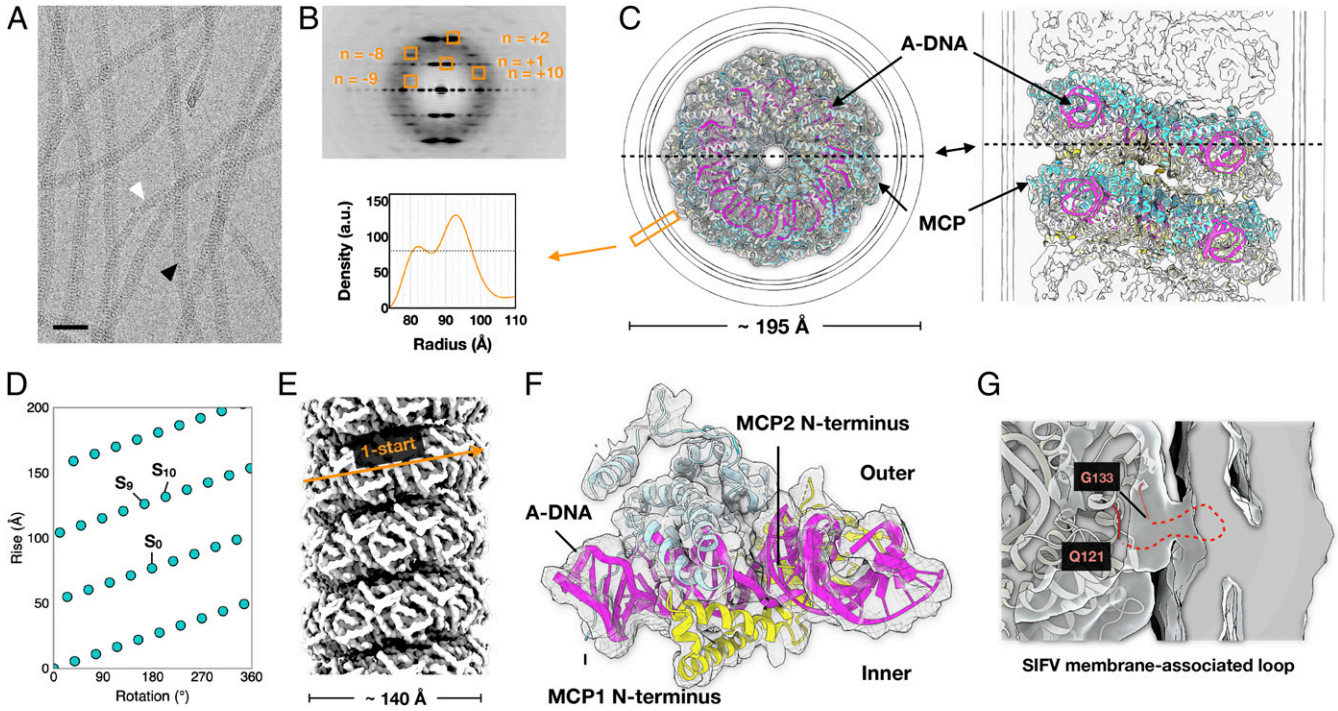


Fig. 2. Cryo-EM of the *S. islandicus* filamentous virus (SIFV). (*A*) Representative cryo-electron micrograph of SIFV. Most of the virions are enveloped by a membrane, indicated by a black arrowhead. Some virions have lost their membrane, and one is indicated by a white arrowhead. This results in a narrower diameter, and these virions appear to have an increased flexibility. (Scale bar, 50 nm.) (*B*) Averaged power spectrum of the segments used in the IHRSR reconstruction. The layer lines that were used to initially estimate the helical symmetry are indicated, along with the Bessel orders for the correct symmetry found. (*C*) Top and side views of the SIFV atomic model fit into the EM map. The side view is from a central slice of the map/model, indicated by the dashed line. A-form DNA is colored magenta, and MCP dimers are in blue and yellow. The membrane has been filtered to 7-Å resolution and cylindrically averaged in this figure. The radial density profile of the cylindrically averaged and filtered membrane is shown. (*D*) Helical net of SIFV, using the convention that the surface is unrolled and we are viewing from the outside. Since there are 9.35 subunits per turn of the right-handed 1-start helix, subunits S_9 and S_{10} will be above the reference subunit S_0 indicated. (*E*) The surface of the 3D reconstruction of SIFV at 4.0-Å resolution, with the membrane removed. (*F*) Ribbon models of the SIFV MCP dimer (cyan and yellow) and 36 bp of A-DNA (magenta), fit into the EM map. (*G*) An 11-aa loop (residues 122 to 132) of SIFV-1 associated with the membrane. An unambiguous backbone trace in this region was not possible due to the low resolution resulting from structural flexibility. The membrane shown here is not cylindrically averaged.

members of the *Rudiviridae* and *Lipothrixviridae* but sharing a similar fold for the MCPs, such N-terminal projections into DNA were not observed at a fairly high resolution of 3.4 Å (16).

A characteristic feature of filamentous archaeal viruses is that their genomic double-stranded DNA (dsDNA) is complexed in A-form (*SI Appendix*, Fig. S4), which is considered to be more resilient to various forms of stress, compared to the physiologically more common B-form (16–19). In both SSRV1 (Fig. 3*B*) and SIFV (Fig. 3*C*), there is an extensive coordination of the phosphate groups in the DNA backbone by positively charged protein sidechains, primarily arginines and lysines. This provides a high degree of specificity in maintaining the DNA in an A-form conformation. In contrast, a nonspecific aspect of the protein–DNA interactions may be a desolvation of the DNA by the extensive wrapping of the protein around the DNA. For SSRV1, we have a resolution (2.8 Å) where one begins to see highly ordered bound water molecules (Fig. 3*D*). There are ~20 ordered water molecules seen around 12-bp A-DNA within the asymmetric unit of SSRV1, with most of them coordinating the phosphate groups. The phosphate groups in the SSRV1 DNA are therefore presumably making hydrogen bonds with either these waters or adjacent protein side chains. In a 0.83-Å resolution crystal structure of A-DNA, most of the bound waters were found surrounding the phosphate groups (27). The water molecules forming a "spine of hydration" in the minor groove of B-DNA (27–29), hypothesized to stabilize B-DNA, do not exist in the SSRV1 structure.

Flexibility of the Virions. It is clear from looking at cryo-EM images that some of the archaeal filamentous viruses are more rigid than others. A number of formalisms exist for deducing the flexural rigidity, a, of an object from the thermodynamic fluctuations in curvature that occur in solution. These can be described in terms of a persistence length, p, where $p = a/\mathrm{kT}$. The persistence length defines a characteristic distance over which the orientation of a filament persists. It does not mean that a filament can be approximated as a rigid rod over such a length. When the length of a filament, L, is much smaller than p, such a rigid rod approximation is reasonable. When $L >> p$, one has a random coil and there is no correlation between the orientation of segments separated by a distance L. We have used cryo-EM images to estimate p for the six filamentous viruses, but there is an important caveat that must be made. The formalism being used assumes thermodynamic equilibrium, which means in the absence of forces and fluxes. However, we know that filaments being prepared for vitrification experience large fluid flow forces, and compression from the thinning film (26, 30), and thus are being captured far from thermodynamic equilibrium. Nevertheless, we show estimates of p in Fig. 4*A* where the relative (as opposed to absolute) rigidity may be inferred. The strongest correlation is between the nucleocapsid diameter and the persistence length, which is consistent with the fact that for a homogeneous rod the rigidity will scale as the fourth power of the diameter. However, it can be seen that SIFV is significantly more flexible than AFV1, even though both have about the same nucleocapsid diameters. This is due to the significantly greater pitch

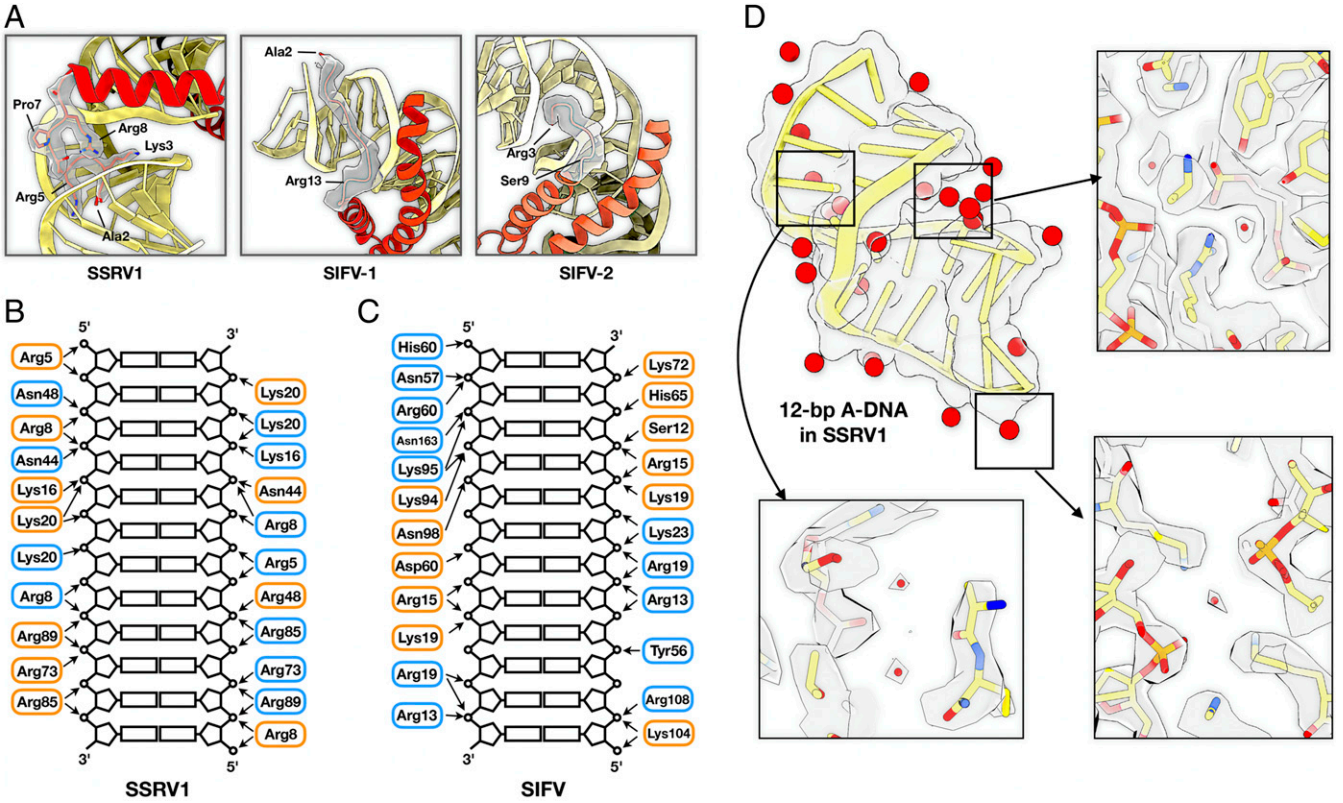


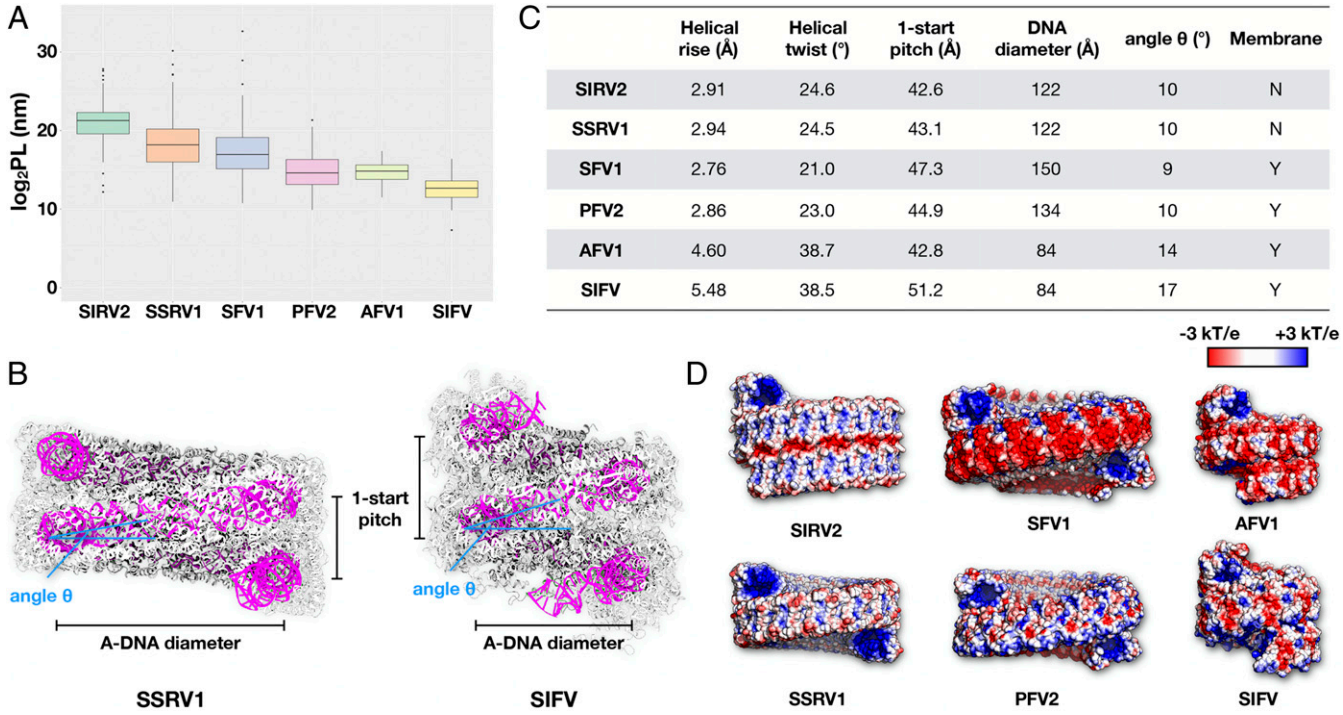
Fig. 3. Extensive MCP–DNA interactions in SSRV1 and SIFV. (*A*) N-terminal residues of SSRV1 MCP, SIFV MCP1 (SIFV-1), and SIFV MCP2 (SIFV-2) inserting into a groove of the A-DNA. The corresponding EM density in this region is shown. For SIFV, these residues are poorly ordered and therefore only a backbone trace is shown. Protein is red, and DNA is yellow. (*B*) Schematic indicating all of the polar protein–DNA contacts in SSRV1. The MCPs form a symmetrical dimer, so residues colored blue are related by the dihedral symmetry to those colored orange. (*C*) Schematic indicating all of the polar protein–DNA contacts in SIFV. Residues from SIFV-1 are colored blue, and those from SIFV-2 are colored orange. (*D*) Water molecules around A-DNA. The 12-bp A-DNA (wrapped by the SSRV1 homodimer) is shown in a yellow ribbon representation, and the water molecules are shown as red spheres. Several close-up views of water and nearby amino acids are shown with the EM density map.

in SIFV (51 vs. 43 Å), which allows more compression of the helical groove during bending (Fig. 4 *B* and *C*). Consistent with this, we expect that the greater flexibility of SFV1 compared to SIRV2 and SSRV1, despite its larger diameter, is due to a larger helical pitch. As with the flexible filamentous plant viruses (31), the SIFV structure is held together by a flexible linker in the protein, which crosses the helical groove and makes contacts with subunits in the next turn of the helix.

Another important factor that contributes to the filament rigidity is the interface between the MCPs. To investigate this, we did PISA (Protein Interfaces, Surfaces, and Assemblies) analysis (32) on all six archaeal virions on three types of interfaces (*SI Appendix*, Fig. S5): 1) the MCP dimer interface; 2) the interface between MCP dimers; and 3) the interface between MCP dimers across the groove of the helix. Within the dimer, as previously reported, protein–protein interfaces are extensive for SIRV2, SSRV1, SFV1, and PFV2 (Fig. 5 and *SI Appendix*, Fig. S5). The interfaces of AFV1 and SIFV are a bit smaller, due to the turn in the N-terminal helix (Fig. 5*C*). The protein–protein interface between adjacent dimers are similar among all six virions, with ∼23% total accessible surface area buried. Interactions between dimers across the groove of the helix are weaker compared to the interface between dimers. When excluding the C-terminal hook, nonenveloped virions SSRV1 and SIRV2 have almost three times larger buried surface area than enveloped virions SFV1, PFV2, AFV1, and SIFV, consistent with the greater rigidity observed in SIRV2 and SSRV1 (Fig. 4*A* and *SI Appendix*, Fig. S5). When including the C-terminal hook in SFV1 and SIFV, the interfaces between MCP dimers across the helical groove are much larger, which probably stabilize the virions.

Next, we looked at the surface electrostatic potential of the six filamentous viruses without A-DNA (Fig. 4*D*). As expected, positively charged tunnels were observed for all six viruses to accommodate their dsDNA genome. The nonenveloped virions SIRV2 and SSRV1 have very similar surface electrostatic potential with a moderate number of charged residues on their surfaces. The electrostatic potential surfaces of the protein cores within enveloped virions are more diverse. For example, SFV1 has a large number of negatively charged residues on its surface. The differences in their surface electrostatic potential may explain differences in their selectivity of lipids acquired from the hosts.

Structural Conservation and Diversity of the Filamentous Virus MCPs. Except for the closely related MCPs of SIRV2 and SSRV1, no other homology is detected at the sequence level. However, a flexible secondary structure alignment indicates obvious homology among the 10 MCPs (Fig. 5*A*). From the structure-based alignment, it is clear that all MCPs have two or three N-terminal α-helices wrapping the DNA in the lumen of the virus (Fig. 5 *A* and *B*), but the details are quite different: SSRV1, SIRV2, and SFV1 have long helices with a kink in the middle to continuously wrap around the DNA; PFV2 is similar to SSRV1, SIRV2, and SFV1, but one of its MCPs, PFV2-2, has domain swapping (Fig. 5*C*); AFV1 and SIFV, on the other hand, form helix–turn–helix motifs, which fold back to cover the DNA on both sides (Fig. 5 *B* and *C*). In the C-terminal domain, all MCPs



C

	Helical rise (Å)	Helical twist (°)	1-start pitch (Å)	DNA diameter (Å)	angle θ (°)	Membrane
SIRV2	2.91	24.6	42.6	122	10	N
SSRV1	2.94	24.5	43.1	122	10	N
SFV1	2.76	21.0	47.3	150	9	Y
PFV2	2.86	23.0	44.9	134	10	Y
AFV1	4.60	38.7	42.8	84	14	Y
SIFV	5.48	38.5	51.2	84	17	Y

Fig. 4. The flexibility of archaeal filamentous virions. (*A*) Estimates of the persistence length of the six archaeal filamentous virions. The measurements for each virus were from 100 filaments randomly selected from cryo-EMs. The measurements are shown in box-and-whisker plots that display five summary statistics (the median, two hinges, and two whiskers), and all "outlying" points individually. (*B*) The filamentous protein–DNA models of SSRV1 and SIFV, with DNA in magenta and protein in gray. (*C*) The parameters of all six archaeal filamentous virus structures, including helical rise and twist, 1-start pitch, A-DNA diameter, pitch angle θ, and the presence/absence of membrane. The DNA diameter is taken as the distance from the axis of the DNA on one side to the other, as this is more precisely defined than something like the outer diameter. (*D*) The electrostatic potential surface of all six archaeal filamentous virus structures, calculated by APBS (63).

have a more conserved four-helix bundle (*SI Appendix*, Fig. S6 *A* and *B*), where the MCPs have three shorter helices (Cα1–3) that combine with the longer helix that wraps the DNA to form the bundle. In the PFV2 MCPs, there is a short additional helix in the C terminus that is not part of the bundle (Fig. 5*C*). Previously, the C-terminal four-helix bundle domain of the major capsid protein from an uncharacterized member of the *Rudiviridae* family (which also includes SIRV1, SIRV2, and SSRV1) was solved by a combination of X-ray crystallography and NMR (33), and its structure is very similar to the C-terminal four-helix bundle domain of SSRV1 and SIRV2 determined by cryo-EM (*SI Appendix*, Fig. S6*C*). Like MCP1 of SFV1, MCP1 of SIFV also has a C-terminal extension reaching out to make extensive contacts with the subunits in the adjacent helical turn (Fig. 5*C*).

A TM-score (34) matrix analysis of the six MCP dimers is shown in Fig. 5*D*. TM score is a metric for measuring the similarity of two protein structures, and it has a value in the range from 0 to 1, where 1 indicates a perfect match between two structures. It has been shown that TM scores below 0.17 correspond to randomly chosen unrelated proteins, while a TM score higher than 0.5 is indicative of the same general protein fold (34). Pairwise comparison of the dimeric MCP structures from the six filamentous viruses followed by single-linkage clustering, which is suited to detect outliers, revealed three major clusters: SIRV2, SSRV1, and SFV1 clustered together with a TM score of >0.75, AFV1 and SIFV dimers formed a smaller cluster with a TM score of 0.72, whereas PFV2 was an outlier (Fig. 5*D*). This clustering is generally consistent with the current classification of the corresponding viruses into three families. Notably, however, SFV1, which is an enveloped virus (family *Lipothrixviridae*), clustered with the nonenveloped rudiviruses rather than with other lipothrixviruses.

To gain further insights into the evolutionary relationship between filamentous archaeal viruses, we performed phylogenomic analysis of all available rudivirus, lipothrixvirus, and tristromavirus genomes using the Genome-BLAST Distance Phylogeny (GBDP) method implemented in VICTOR (35). In the phylogenomic tree based on the shared gene content and rooted with tristromaviruses, all four genera of lipothrixviruses are retrieved with maximal support values. However, family *Lipothrixviridae* is paraphyletic with respect to rudiviruses, which form a sister clade to genus *Gammalipothrixvirus* (Fig. 6). Indeed, 10 of the gammalipothrixvirus AFV1 genes are shared with rudivirus SIRV2 (16). This result suggests that rudiviruses have evolved from within lipothrixvirus diversity.

Discussion

Filamentous helical viruses are common in all three domains of life (36–40), but despite overall similar morphology, virions of bacterial, eukaryotic, and archaeal filamentous viruses are built from unrelated capsid proteins with different structural folds (41). Furthermore, the ways the nucleic acids are protected by the corresponding capsid proteins are radically different in viruses from the different domains of life (19, 31, 42–44). With the two virion structures presented herein, archaeal filamentous viruses of the proposed class "*Tokiviricetes*" (families *Rudiviridae*, *Lipothrixviridae*, and *Tristromaviridae*) emerge as valuable models for understanding the evolution of virus structure and adaptation to extreme environments. High-resolution structures are now available for six evolutionarily related viruses with different degrees of flexibility, providing insights into structural changes underlying the differences in mechanical properties of the viral particles. Archaeal viruses are unique in this respect, because in eukaryotes, rod-shaped viruses, such as tobacco mosaic virus, the

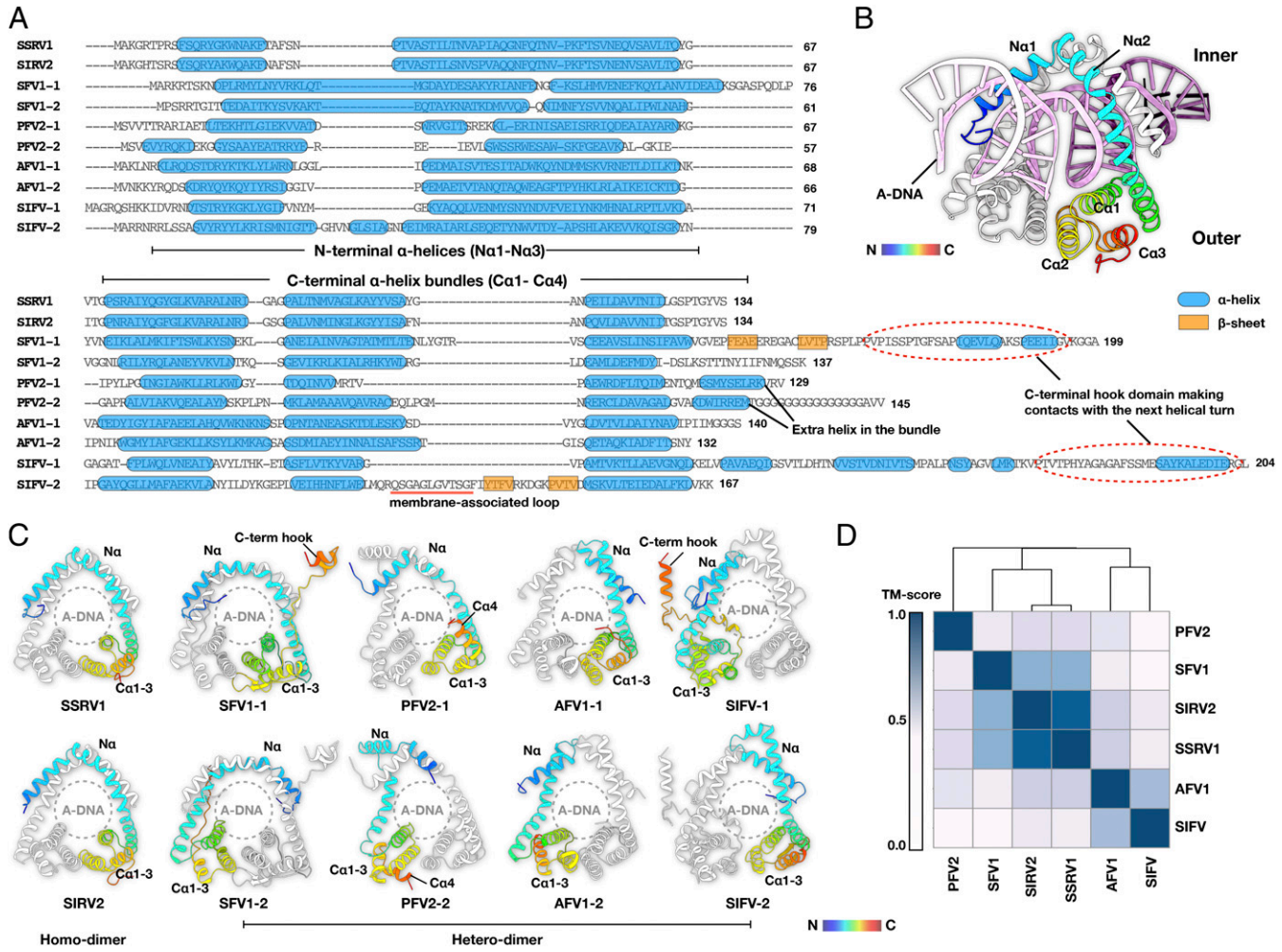


Fig. 5. Structural conservation and diversity of the filamentous virus MCPs. (*A*) Structure-based sequence alignments of 10 MCP sequences from six archaeal filamentous virus structures. The α-helices are indicated by blue rounded rectangles, and β-sheets are indicated by orange rectangles. (*B*) Representative domain architecture of the MCP in known filamentous virus. An SSRV1 dimer with 36-bp A-DNA is shown: One MCP is shown in rainbow coloring and the other in white; DNA is colored magenta. N-terminal α-helices wrap DNA on the luminal side and C-terminal α-helix bundles wrap DNA facing the outer solvent or the membrane. (*C*) MCP comparison of SSRV1, SIRV2, SFV1, PFV2, AFV1, and SIFV. SSRV1 and SIRV2 are homodimers, while the other four MCP dimers are heterodimers. (*D*) All-against-all comparison of the six MCP dimers. The matrix is based on the pairwise TM score calculated from the MM-align server.

first virus ever isolated (42), on the one hand, and flexible viruses, such as potyviruses (45) and alphaflexiviruses (31), on the other hand, are built from nonhomologous capsid proteins. Notably, we find that some of the adaptations in virion structure of archaeal viruses are shared with virions of flexible plant viruses. In particular, the flexibility in SIFV, the most flexible among the compared viruses, can be, in part, attributed to the C-terminal tail, which extends into the adjacent helical turn, where it makes contacts with other capsid protein subunits. As a result, all adjacent helical turns in the virion are tied together, allowing deformations such as compression, extension, and torsion, while still maintaining the structural integrity of the virion. A similar solution has previously been shown for plant alphaflexiviruses (31). More generally, comparison of the persistence lengths of the archaeal viruses showed that nucleocapsid diameter and helical rise are the best predictors for virion rigidity, where the smaller diameter and larger rise are associated with a more flexible virion.

In lipothrixviruses, the nucleocapsid is enveloped by a thin monolayer membrane (17, 18), which is half the thickness of the host membrane from which it is derived, whereas rudiviruses are not enveloped. Due to its importance for various aspects of virus–host interactions, such as during virus entry and egress, the presence or absence of an envelope is generally considered an inherent property of a given virus group, which is not frequently acquired or lost during evolution. Nevertheless, transitions from enveloped to nonenveloped virions have occurred among certain groups of eukaryotic RNA viruses, apparently as an adaptation to plant hosts (46). Comparison of the relatively closely related rudiviruses and lipothrixviruses provides insights into virion envelopment. The DNA is more tightly complexed by the capsid protein in rudiviruses SIRV2 and SSRV1 than in lipothrixviruses SIFV and AFV1, suggesting that the envelope in the latter group of viruses is an adaptation providing an additional layer of protection for the genome against the acidic extracellular environment in which these viruses thrive. Comparison of the (nucleo) capsid surfaces of lipothrixviruses and rudiviruses did not reveal major differences in their hydrophobicity (*SI Appendix*, Fig. S7). This is not entirely surprising, as the protein surface would be facing polar headgroups when a membrane is present. Notably, the major difference in capsid surfaces is actually electrostatic potential, suggesting that acquisition/loss of the membrane has likely been accompanied by extensive sequence changes of the capsid surface charged residues. SIFV contains a unique

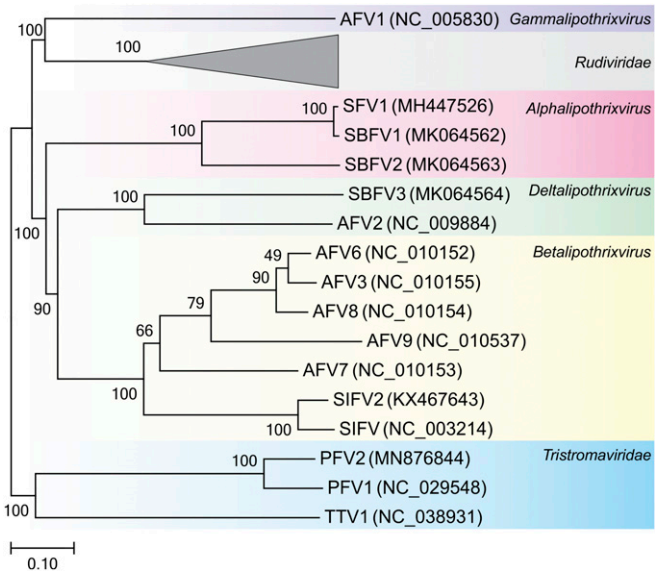


Fig. 6. Inferred phylogenomic tree of archaeal filamentous viruses. The tree is based on whole-genome VICTOR analysis at the amino acid level. The tree is rooted with tristromaviruses, and the branch length is scaled in terms of the Genome BLAST Distance Phylogeny (GBDP) distance formula D6. The numbers above branches are GBDP pseudobootstrap support values from 100 replications. For each genome, the abbreviated virus name and GenBank accession number are indicated. The tree is divided into colored blocks according to the taxonomy of the compared viruses.

surface-exposed insertion in one of the capsid proteins, which might be involved in membrane binding. However, equivalent regions are absent in other lipothrixviruses, suggesting that the determinants underlying envelopment are present in the core regions of the major capsid protein. It remains unclear whether the membrane interacts directly with the MCPs or through an intermediate matrix protein, as in most eukaryotic enveloped viruses (47). Abundant structural proteins other than the MCPs were indeed identified in SFV1 (17) and PFV1 (48), but not in other lipothrixviruses, including SIFV (*SI Appendix*, Fig. S2).

The considerations above beg the question: Was the membrane acquired by the ancestor of rudiviruses to produce the first lipothrixvirus or was it lost in a particular lineage of lipothrixviruses, giving rise to rudiviruses? The existence of a subgroup of rudiviruses encoding relics of a second MCP is consistent with the latter scenario. Notably, the inactivated and functional MCPs are adjacently encoded, as are the two functional MCPs in lipothrixviruses and tristromaviruses. The inactivated MCPs might have been exapted for a different function, a common process in virus evolution (49), thus explaining why they have not been lost. Interestingly, structural comparison of the capsid proteins of filamentous archaeal viruses showed that enveloped filamentous viruses form three clusters, one of which includes SFV1 and nonenveloped rudiviruses. The similarity between the capsid proteins of the three viruses extends also to the overall virion characteristics, including relatively high rigidity and presence of a central cavity, which is not present in SIFV and AFV1. Furthermore, whereas the MCPs of rudiviruses are conserved at the sequence level, those of lipothrixviruses are highly divergent, to the extent that homology between the two MCPs of the same virus or of viruses in the same family but from different genera are not recognizable. This suggests that lipothrixviruses (and their MCPs) have diverged in a more distant past than rudiviruses. Note that there is no reason to assume that evolutionary rate would be different between the two virus families. Based on shared gene content, we have previously proposed that tristromaviruses have diverged from lipothrixviruses and rudiviruses concomitantly with the divergence of their respective hosts, Thermoproteales and Sulfolobales (16), suggesting that the virion organization of the postulated viral ancestor resembled that of lipothrixviruses and tristromaviruses. Finally, phylogenomic analysis of filamentous archaeal viruses also suggests that rudiviruses are a derived rather than ancestral group of viruses. Consequently, currently available data suggest that the most recent common ancestor of extant archaeal filamentous viruses was an enveloped virus with heterodimeric MCPs that encapsidated A-form DNA. Prior to this most recent ancestor, it appears reasonable to imagine a simpler ancestral virus with a homodimeric MCP, resembling the contemporary rudiviruses but not directly related to them. The envelope might have been an ancestral feature, which was shed only in rudiviruses, presumably following the emergence of the MCP capable of efficient DNA protection in harsh environmental conditions. It is almost certain that further exploration of archaeal virus diversity and their structural and biochemical characterization is bound to shed more light on the evolution of this remarkable group of viruses.

The high-resolution structure achieved for SSRV1 allowed us to better understand how B-form DNA is converted into A-form, which is a characteristic feature in this group of viruses (16–19). Notably, storage of DNA in A-form has been convergently evolved by bacterial spores (50, 51) and by an icosahedral hyperthermophilic archaeal virus SPV1 (52), suggesting that A-form is a preferred conformation for DNA storage under extreme environmental conditions. However, A-form DNA might not be the only mechanism of genome protection. For instance, the genome of spindle-shaped virus SSV1 is positively supercoiled in virions (53) and is complexed with the host chromatin protein of the Sul7d family (54). In contrast, the circular genomes of hyperthermophilic bacilliform clavavirus APBV1 (24) and icosahedral turrivirus STIV (55) are encapsidated as naked dsDNA, and it remains unclear whether there are dedicated mechanisms for the DNA protection. Elucidation of these mechanisms should uncover adaptations to life in extreme environments and open doors for biomedical and nanotechnological applications.

Methods

Virus Production and Purification. Exponentially growing cultures of *Saccharolobus solfataricus* POZ149 (21) and *Sulfolobus islandicus* LAL14/1 (56) cells were infected with fresh preparations of SSRV1 and SIFV, respectively. The infected cultures were incubated at 75 °C under agitation for 2 d. After the removal of cells (7,000 rpm, 20 min; Sorvall 1500 rotor), viruses were collected and concentrated by ultracentrifugation (40,000 rpm, 2 h, 10 °C; Beckman 126 SW41 rotor). For cryo-EM analysis, the concentrated particles were resuspended in buffer A (54): 20 mM KH_2PO_4, 250 mM NaCl, 2.14 mM $MgCl_2$, 0.43 mM $Ca(NO_3)_2$, and <0.001% trace elements of Sulfolobales medium, pH 6. For SDS/PAGE and mass spectrometry analyses, virus particles were further purified by ultracentrifugation in a CsCl buoyant density gradient (0.45 g·mL^{-1}) with a Beckman SW41 rotor at 39,000 rpm for 20 h at 10 °C. The opalescent bands were collected with a needle and a syringe, and dialyzed against buffer A.

Analysis of SIFV and SSRV1 Structural Proteins. The purified virions were analyzed by SDS/PAGE, and proteins were stained with InstantBlue (Expedeon). The stained protein bands of SIFV virions were excised from the gel and in-gel digested with trypsin. For SSRV1, the CsCl-purified virions were trypsinized in solution. The generated peptides were separated and identified by nano-LC-MS/MS (Proteomics Platform, Institut Pasteur) using an Ultimate 3000 system (Dionex) coupled to an LTQ-Orbitrap Velos system (Thermo Fisher Scientific). Peptide masses were searched against annotated SSRV1 and SIFV proteomes using Andromeda with MaxQuant software, version 1.3.0.5 (57).

Phylogenomic Analysis. All pairwise comparisons of the amino acid sequences of rudivirus, lipothrixvirus, and tristromavirus genomes were conducted using the GBDP method implemented in VICTOR, under settings recommended for prokaryotic viruses (35). The resulting intergenomic distances derived from

pairwise matches (local alignments) were used to infer a balanced minimum evolution tree with branch support via FASTME including SPR postprocessing for D6 formula, which corresponds to the sum of all identities found in high-scoring segment pairs divided by total genome length. Branch support was inferred from 100 pseudobootstrap replicates each. The tree was rooted with members of the family *Tristromaviridae*.

Cryo-EM Image Analysis. The virus sample (~4.0 μL) was applied to discharged lacey carbon grids and plunge frozen using a Vitrobot Mark IV (FEI). Frozen grids were imaged in a Titan Krios at 300 keV and recorded with a K3 camera at 1.08 Å/pixel (SSRV1) or a Falcon III camera at 1.4 Å/pixel (SIFV). The SIFV data were collected at the University of Virginia core facility, and the SSRV1 data were collected at the National Cryo-EM Facility of the National Cancer Institute. Micrographs were collected using a defocus range of 1 to 2.5 μm, with a total dose of ~50 electrons/Å^2 distributed into ~25 fractions. To get a preliminary helical reconstruction volume in SPIDER, all of the micrographs were first motion corrected and dose weighted by MotionCorr v2, and then CTF-multiplicated by the theoretical CTF. Filament images corresponding to ~20 electrons /Å^2 were extracted using EMAN2 (58). Small subsets containing ~30,000 overlapping 384-pixel-long segments were used to search for the correct helical symmetry. The helical symmetry was determined in SPIDER (59) using IHRSR (60) after searching through all possible symmetries by trial and error, until recognizable secondary structural features were seen. A ~6-Å initial reconstruction was generated from this small subset, and this volume was subsequently filtered to 7 Å as the starting reference used in RELION (61). After using the full dataset in RELION, doing CTF-refinement and Bayesian polishing, the final volume was estimated to have a resolution of 2.8 Å for the SSRV1 and 4.0 Å for SIFV, based on the map:map FSC, model:map FSC and d_{99} (62). The final volumes were then sharpened with a negative B-factor automatically estimated in RELION, and the statistics are listed in Table 1.

Model Building. The density corresponding to a single SSRV1 MCP was segmented from the experimental cryo-EM density using Chimera. Then a starting model was generated by homology modeling using the SIRV2 MCP as the reference, and then docked into the segmented map. Then this model was adjusted manually in Coot and real-space refined in PHENIX. The EM density corresponding to A-DNA was also segmented in Chimera, and the A-DNA was manually put in the map and refined in PHENIX. Finally, the refined single MCP and DNA model were used to generate a filamentous model using the determined helical symmetry, and this filament model was refined against the full cryo-EM map using in PHENIX real-space refinement. MolProbity was used to evaluate the quality of the filament model. The refinement statistics are shown in Table 1.

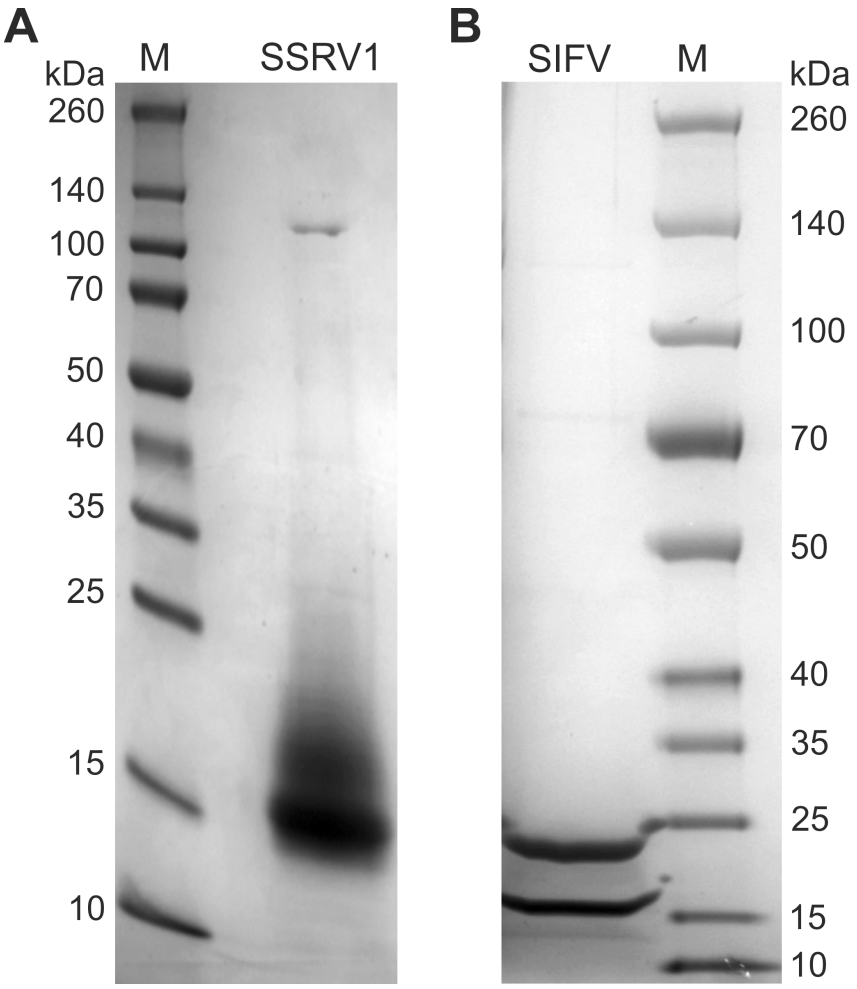
For SIFV, the density corresponding to a single MCP1 or MCP2 was segmented from the experimental filament density using Chimera. The full-length MCP1/MCP2 protein was built de novo into the segmented map using Rosetta-CM, then adjusted manually in Coot and real-space refined in PHENIX. The EM density corresponding to A-DNA was also segmented in Chimera, and the DNA model was manually built and refined in PHENIX. The filament model was refined in the same way as the SSRV1 model described above.

Data Availability. The atomic coordinates for SSRV1 have been deposited at the Protein Data Bank with accession number 6WQ0. The corresponding density map has been deposited at the Electron Microscopy Data Bank with accession number EMD-21867. The atomic coordinates for SIFV have been deposited at the Protein Data Bank with accession number 6WQ2. The corresponding density map has been deposited at the Electron Microscopy Data Bank with accession number EMD-21868.

ACKNOWLEDGMENTS. Imaging of SSRV1 was performed at the National Cancer Institute's National Cryo-EM Facility at the Frederick National Laboratory for Cancer Research under Contract HSSN261200800001E. Imaging of SIFV was done at the Molecular Electron Microscopy Core Facility at the University of Virginia, which is supported by the School of Medicine. This work was supported by NIH Grant R35GM122510 (E.H.E.). M.K. was supported by l'Agence Nationale de la Recherche Grant ANR-17-CE15-0005-01. D.P.B. is part of the Pasteur–Paris University International PhD Program, which has received funding from the European Union's Horizon 2020 Research and Innovation Programme under Marie Sklodowska-Curie Grant Agreement 665807. We thank Thibault Chaze and Mariette Matondo (Pasteur Proteomics Platform) for help with the mass spectrometry analyses.

1. E. V. Koonin, V. V. Dolja, A virocentric perspective on the evolution of life. *Curr. Opin. Virol.* **3**, 546–557 (2013).
2. C. A. Suttle, Viruses in the sea. *Nature* **437**, 356–361 (2005).
3. O. Bergh, K. Y. Børsheim, G. Bratbak, M. Heldal, High abundance of viruses found in aquatic environments. *Nature* **340**, 467–468 (1989).
4. A. R. Mushegian, Are there 10^{31} virus particles on earth, or more, or fewer? *J. Bacteriol.* **202**, e00052-20 (2020).
5. D. L. Caspar, A. Klug, Physical principles in the construction of regular viruses. *Cold Spring Harb. Symp. Quant. Biol.* **27**, 1–24 (1962).
6. Z. Li, A. H. Pearlman, P. Hsieh, DNA mismatch repair and the DNA damage response. *DNA Repair (Amst.)* **38**, 94–101 (2016).
7. J. Xia *et al*., Bacteria-to-human protein networks reveal origins of endogenous DNA damage. *Cell* **176**, 127–143.e24 (2019).
8. R. E. Franklin, K. C. Holmes, The helical arrangement of the protein subunits in tobacco mosaic virus. *Biochim. Biophys. Acta* **21**, 405–406 (1956).
9. M. Krupovic, V. Cvirkaite-Krupovic, J. Iranzo, D. Prangishvili, E. V. Koonin, Viruses of archaea: Structural, functional, environmental and evolutionary genomics. *Virus Res.* **244**, 181–193 (2018).
10. D. Prangishvili *et al*., The enigmatic archaeal virosphere. *Nat. Rev. Microbiol.* **15**, 724–739 (2017).
11. D. Prangishvili, The wonderful world of archaeal viruses. *Annu. Rev. Microbiol.* **67**, 565–585 (2013).
12. J. H. Munson-McGee, J. C. Snyder, M. J. Young, Archaeal viruses from high-temperature environments. *Genes (Basel)* **9**, 128 (2018).
13. A. J. Berliner, T. Mochizuki, K. M. Stedman, Astrovirology: Viruses at large in the universe. *Astrobiology* **18**, 207–223 (2018).
14. F. Pasin, W. Menzel, J. A. Daròs, Harnessed viruses in the age of metagenomics and synthetic biology: An update on infectious clone assembly and biotechnologies of plant viruses. *Plant Biotechnol. J.* **17**, 1010–1026 (2019).
15. M. Y. Chen, S. S. Butler, W. Chen, J. Suh, Physical, chemical, and synthetic virology: Reprogramming viruses as controllable nanodevices. *Wiley Interdiscip. Rev. Nanomed. Nanobiotechnol.* **11**, e1545 (2019).
16. F. Wang, D. P. Baquero, Z. Su, T. Osinski, D. Prangishvili, E. H. Egelman, M. Krupovic, Structure of a filamentous virus uncovers familial ties within the archaeal virosphere. *Virus Evol.* **6**, veaa023 (2020).
17. Y. Liu *et al*., Structural conservation in a membrane-enveloped filamentous virus infecting a hyperthermophilic acidophile. *Nat. Commun.* **9**, 3360 (2018).
18. P. Kasson *et al*., Model for a novel membrane envelope in a filamentous hyperthermophilic virus. *eLife* **6**, e26268 (2017).
19. F. DiMaio *et al*., Virology. A virus that infects a hyperthermophile encapsidates A-form DNA. *Science* **348**, 914–917 (2015).
20. D. Prangishvili, M. Krupovic, A new proposed taxon for double-stranded DNA viruses, the order "Ligamenvirales." *Arch. Virol.* **157**, 791–795 (2012).
21. D. P. Baquero *et al*., New virus isolates from Italian hydrothermal environments underscore the biogeographic pattern in archaeal virus communities. *ISME J.* **14**, 1821–1833 (2020).
22. H. P. Arnold *et al*., A novel lipothrixvirus, SIFV, of the extremely thermophilic crenarchaeon *Sulfolobus*. *Virology* **267**, 252–266 (2000).
23. C. L. Moyer, G. R. Nemerow, Viral weapons of membrane destruction: Variable modes of membrane penetration by non-enveloped viruses. *Curr. Opin. Virol.* **1**, 44–49 (2011).
24. D. Ptchelkine *et al*., Unique architecture of thermophilic archaeal virus APBV1 and its genome packaging. *Nat. Commun.* **8**, 1436 (2017).
25. J. M. Vargason, K. Henderson, P. S. Ho, A crystallographic map of the transition from B-DNA to A-DNA. *Proc. Natl. Acad. Sci. U.S.A.* **98**, 7265–7270 (2001).
26. E. H. Egelman, Cryo-EM: Ice is nice, but good ice can be hard to find. *Biophys. J.* **118**, 1238–1239 (2020).
27. M. Egli *et al*., X-ray crystallographic analysis of the hydration of A- and B-form DNA at atomic resolution. *Biopolymers* **48**, 234–252 (1998).
28. M. L. Kopka, A. V. Fratini, H. R. Drew, R. E. Dickerson, Ordered water structure around a B-DNA dodecamer. A quantitative study. *J. Mol. Biol.* **163**, 129–146 (1983).
29. H. R. Drew, R. E. Dickerson, Structure of a B-DNA dodecamer. III. Geometry of hydration. *J. Mol. Biol.* **151**, 535–556 (1981).
30. V. E. Galkin, A. Orlova, M. R. Vos, G. F. Schröder, E. H. Egelman, Near-atomic resolution for one state of F-actin. *Structure* **23**, 173–182 (2015).
31. F. DiMaio *et al*., The molecular basis for flexibility in the flexible filamentous plant viruses. *Nat. Struct. Mol. Biol.* **22**, 642–644 (2015).
32. E. Krissinel, K. Henrick, Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* **372**, 774–797 (2007).
33. B. R. Szymczyna *et al*., Synergy of NMR, computation, and X-ray crystallography for structural biology. *Structure* **17**, 499–507 (2009).
34. J. Xu, Y. Zhang, How significant is a protein structure similarity with TM-score = 0.5? *Bioinformatics* **26**, 889–895 (2010).
35. J. P. Meier-Kolthoff, M. Göker, VICTOR: Genome-based phylogeny and classification of prokaryotic viruses. *Bioinformatics* **33**, 3396–3404 (2017).
36. S. J. Wylie *et al*.; ICTV Report Consortium, ICTV virus taxonomy profile: *Potyviridae*. *J. Gen. Virol.* **98**, 352–354 (2017).
37. M. Fuchs *et al*., ICTV virus taxonomy profile: *Closteroviridae*. *J. Gen. Virol.* **101**, 364–365 (2020).
38. M. J. Adams *et al*.; Ictv Report Consortium, ICTV virus taxonomy profile: *Virgaviridae*. *J. Gen. Virol.* **98**, 1999–2000 (2017).
39. S. Roux *et al*., Cryptic inoviruses revealed as pervasive in bacteria and archaea across Earth's biomes. *Nat. Microbiol.* **4**, 1895–1906 (2019).

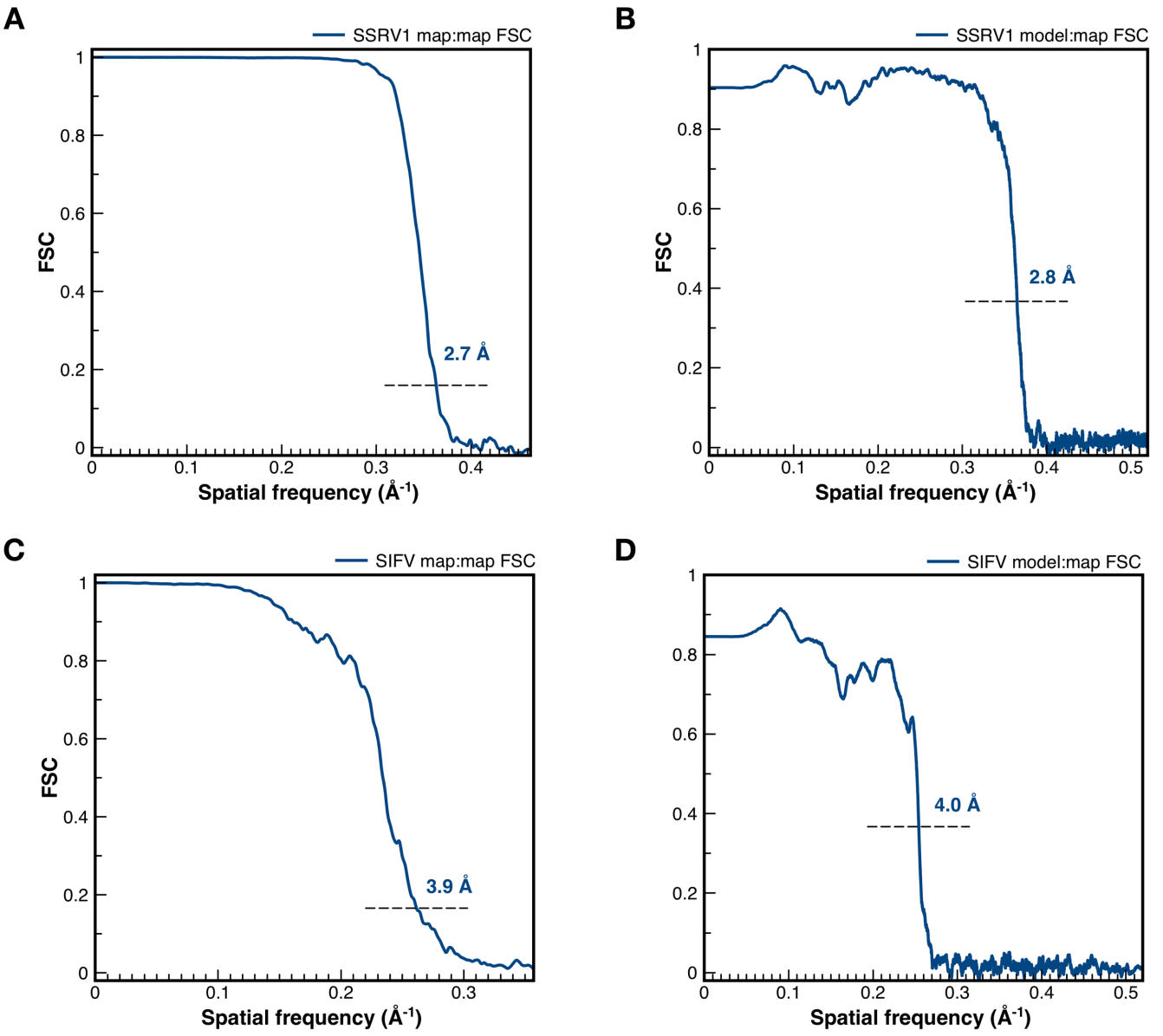
40. G. Stubbs, A. Kendall, Helical viruses. *Adv. Exp. Med. Biol.* **726**, 631–658 (2012).
41. M. Krupovic, E. V. Koonin, Multiple origins of viral capsid proteins from cellular ancestors. *Proc. Natl. Acad. Sci. U.S.A.* **114**, E2401–E2410 (2017).
42. A. Klug, The tobacco mosaic virus particle: Structure and assembly. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **354**, 531–535 (1999).
43. A. K. Tarafder *et al*., Phage liquid crystalline droplets form occlusive sheaths that encapsulate and protect infectious rod-shaped bacteria. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 4724–4731 (2020).
44. A. Grinzato *et al*., Atomic structure of potato virus X, the prototype of the *Alphaflexiviridae* family. *Nat. Chem. Biol.* **16**, 564–569 (2020).
45. M. Zamora, E. Mendez-Lopez, X. Agirrezabala, R. Cuesta, J. L. Lavin, M. A. Sanchez-Pina, M. A. Aranda, M. Valle, Potyvirus virion structure shows conserved protein fold and RNA binding site in ssRNA viruses. *Sci. Adv.* **3**, eaao2182 (2017).
46. R. Kormelink, M. L. Garcia, M. Goodin, T. Sasaya, A. L. Haenni, Negative-strand RNA viruses: The plant-infecting counterparts. *Virus Res.* **162**, 184–202 (2011).
47. L. V. Kordyukova, E. V. Shtykova, L. A. Baratova, D. I. Svergun, O. V. Batishchev, Matrix proteins of enveloped viruses: A case study of influenza A virus M1 protein. *J. Biomol. Struct. Dyn.* **37**, 671–690 (2019).
48. E. I. Rensen *et al*., A virus of hyperthermophilic archaea with a unique architecture among DNA viruses. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 2478–2483 (2016).
49. E. V. Koonin, M. Krupovic, The depths of virus exaptation. *Curr. Opin. Virol.* **31**, 1–8 (2018).
50. K. S. Lee, D. Bumbaca, J. Kosman, P. Setlow, M. J. Jedrzejas, Structure of a protein-DNA complex essential for DNA protection in spores of Bacillus species. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 2806–2811 (2008).
51. P. Setlow, I will survive: DNA protection in bacterial spores. *Trends Microbiol.* **15**, 172–180 (2007).
52. F. Wang *et al*., A packing for A-form DNA in an icosahedral virus. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 22591–22597 (2019).
53. M. Nadal, G. Mirambeau, P. Forterre, W.-D. Reiter, M. Duguet, Positively supercoiled DNA in a virus-like particle of an archaebacterium. *Nature* **321**, 256–258 (1986).
54. E. R. Quemin *et al*., *Sulfolobus* spindle-shaped virus 1 contains glycosylated capsid proteins, a cellular chromatin protein, and host-derived lipids. *J. Virol.* **89**, 11681–11691 (2015).
55. D. Veesler *et al*., Atomic structure of the 75 MDa extremophile *Sulfolobus* turreted icosahedral virus determined by cryoEM and X-ray crystallography. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 5504–5509 (2013).
56. C. Jaubert *et al*., Genomics and genetics of *Sulfolobus islandicus* LAL14/1, a model hyperthermophilic archaeon. *Open Biol.* **3**, 130010 (2013).
57. J. Cox *et al*., Andromeda: A peptide search engine integrated into the MaxQuant environment. *J. Proteome Res.* **10**, 1794–1805 (2011).
58. G. Tang *et al*., EMAN2: An extensible image processing suite for electron microscopy. *J. Struct. Biol.* **157**, 38–46 (2007).
59. J. Frank *et al*., SPIDER and WEB: Processing and visualization of images in 3D electron microscopy and related fields. *J. Struct. Biol.* **116**, 190–199 (1996).
60. E. H. Egelman, A robust algorithm for the reconstruction of helical filaments using single-particle methods. *Ultramicroscopy* **85**, 225–234 (2000).
61. S. He, S. H. W. Scheres, Helical reconstruction in RELION. *J. Struct. Biol.* **198**, 163–176 (2017).
62. P. V. Afonine *et al*., New tools for the analysis and validation of cryo-EM maps and atomic models. *Acta Crystallogr. D Struct. Biol.* **74**, 814–840 (2018).
63. N. A. Baker, D. Sept, S. Joseph, M. J. Holst, J. A. McCammon, Electrostatics of nanosystems: Application to microtubules and the ribosome. *Proc. Natl. Acad. Sci. U.S.A.* **98**, 10037–10041 (2001).



Supplementary Figure 1 Structural proteins of SSRV1 and SIFV virions

A, SDS-PAGE of SSRV1 proteins stained with InstantBlue. Based on the mass spectrometry results, the ~ 120 kDa protein corresponds to SSRV1 gp23 (QJF12299), a minor structural proteins conserved in all rudiviruses and implicated in the formation of terminal fibers responsible for receptor recognition and binding (1).

B, SDS-PAGE of SIFV proteins stained with InstantBlue. M, molecular mass standards.



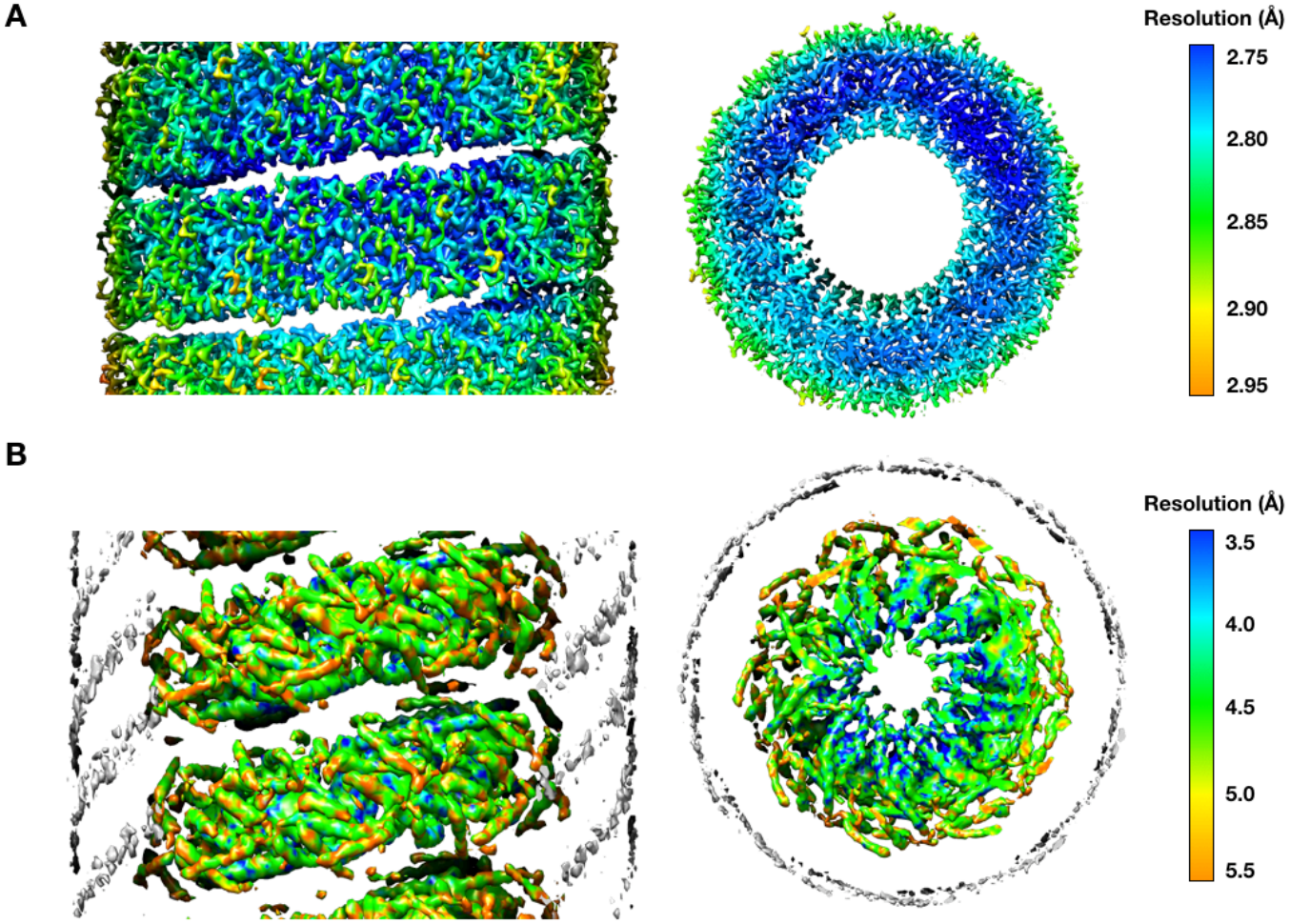
Supplementary Figure 2 Fourier Shell Correlation (FSC) calculations

A, The map:map "gold standard" FSC using the 0.143 criterion estimates the SSRV1 reconstruction to have a resolution of 2.7 Å.

B, The model:map FSC calculation using a 0.38 criterion, which is sqrt (0.143), estimates the SSRV1 reconstruction to have a resolution of 2.8 Å.

C, The map:map "gold standard" FSC using the 0.143 criterion estimates the SIFV reconstruction to have a resolution of 3.9 Å.

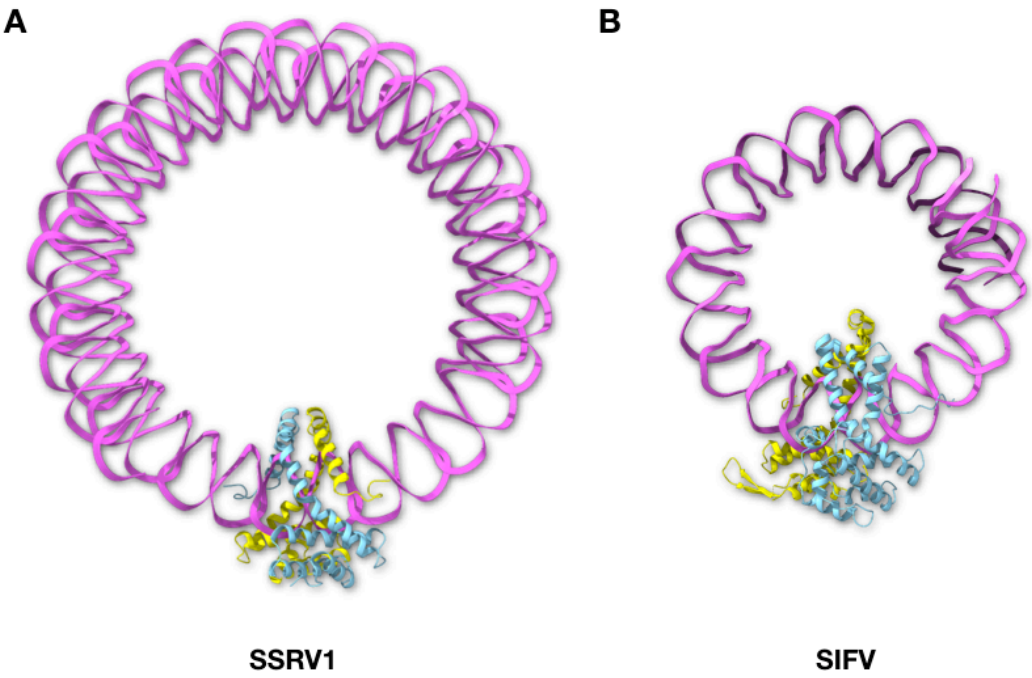
D, The model:map FSC calculation using a 0.38 criterion, which is sqrt (0.143), estimates the SIFV reconstruction to have a resolution of 4.0 Å.



Supplementary Figure 3 Local resolution estimation cryo-EM maps by ResMap

A, SSRV1 side view and top view.

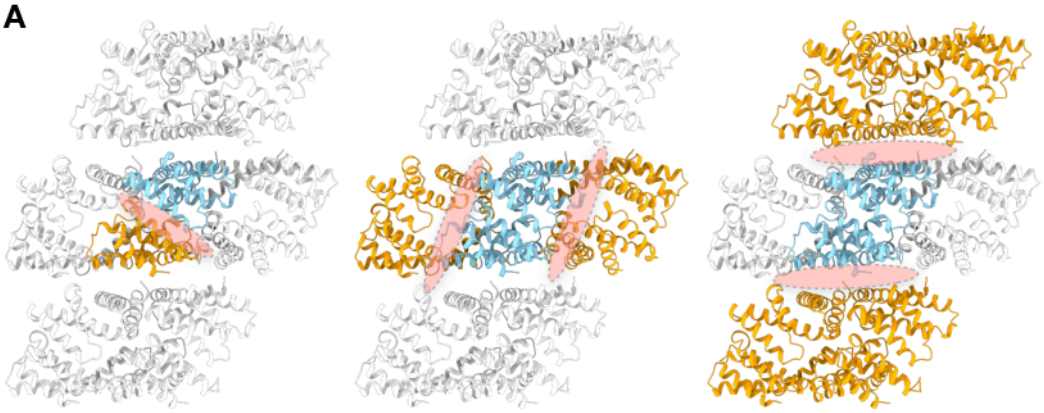
B, SIFV side and top view.



Supplementary Figure 4 Protein-DNA arrangements of SSRV1 and SIFV virions

Top view of SSRV1 (**A**) and SIFV (**B**) virions shows the A-DNA (purple) and one dimer of the capsid protein in ribbon.

A



Dimer interface | Interface between dimers | Interface between dimers and across the groove of the helix

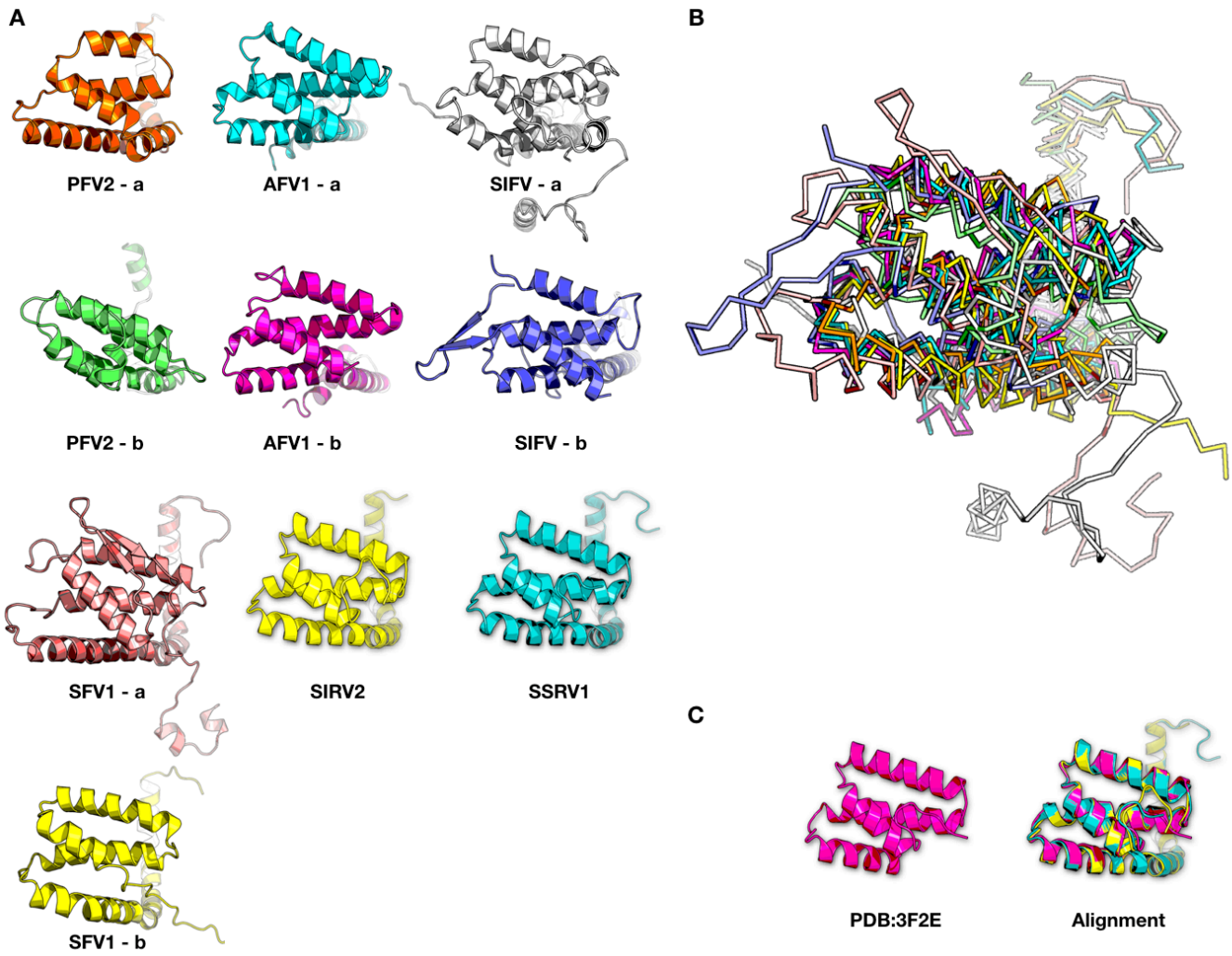
B

(BSA, Å²/%)	Dimer interface	Interface between dimers	Interface between dimers and across the groove of the helix
SIRV2	1491 (17%)	3259 (23%)	1240 (9%)
SSRV1	1536 (17%)	3490 (24%)	1280 (9%)
SFV1	1645 (15%)	4778 (24%)	2275 (11%)
SFV1 - no hook	N.A.	N.A.	410 (2%)
PFV2	1380 (15%)	3511 (25%)	492 (3%)
AFV1	802 (10%)	3137 (21%)	436 (3%)
SIFV	701 (6%)	4863 (22%)	2599 (12%)
SIFV - no hook	N.A.	N.A.	260 (2%)

Supplementary Figure 5 PISA analysis of the six filamentous archaeal virions

A, Illustration of three type of interfaces: dimer interface; interface between dimers; interface between dimers across the groove of the helix.

B, The PISA calculation of the buried surface area (BSA) of the three type of interfaces, and the percentage of the total accessible surface area.

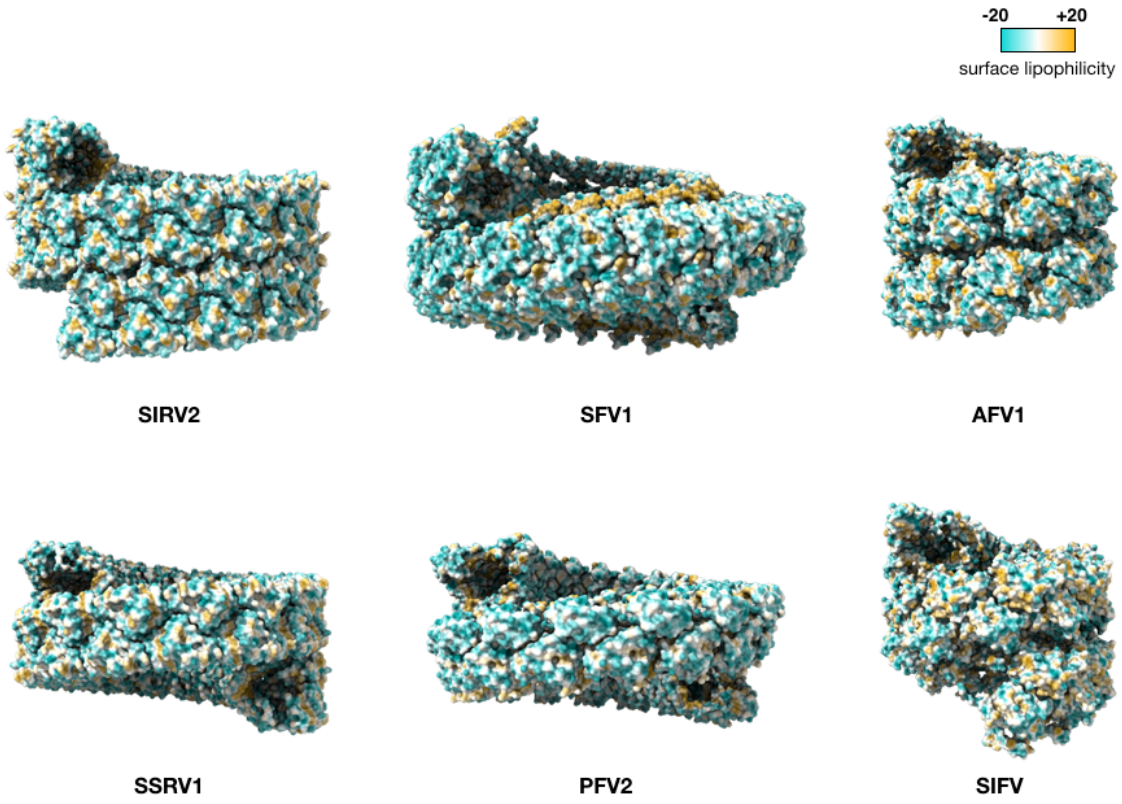


Supplementary Figure 6 C-terminal four-helix bundle domains

A, C-terminal four-helix bundle domains of 10 MCPs of the helical viruses

B, The alignment of the 10 MCPs in (**A**) shows a semi-conserved four-helix bundle domain

C, The crystal structure of Yellowstone SIRV (PDB 3F2E), and its alignment with SIRV2 and SSRV1



Supplementary Figure 7 The surface of the six filamentous archaeal virions colored by lipophilicity

References

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

6.3 CHAPTER 3

The egress of the enveloped filamentous virus SIFV involves intracellular envelopment and formation of virus-associated pyramids

Contribution: I participated in the design of the research, carried out most of the experiments, contributed to the analysis of the data and wrote the first draft of the manuscript.

A filamentous archaeal virus is enveloped inside the cell and released through pyramidal portals

Diana P. Baquero[1,2,#], Anastasia Gazi[3,#], Martin Sachse[3], Junfeng Liu[1], Stefan Schouten[4,5], David Prangishvili[1,6,*] and Mart Krupovic[1,*]

[1] Archaeal Virology Unit, Department of Microbiology, Institut Pasteur, 75015 Paris, France
[2] Collège Doctoral, Sorbonne Universités, 75005 Paris, France
[3] Unité Technologie et service BioImagerie Ultrastructurale, Institut Pasteur, 75015 Paris, France
[4] Department of Marine Microbiology and Biogeochemistry, NIOZ Royal Netherlands Institute for Sea Research, and Utrecht University, PO Box 59, 1790AB, Den Burg (Texel), The Netherlands
[5] Department of Geosciences, Faculty of Earth Sciences, Utrecht University, PO Box 80.021, Utrecht, The Netherlands
[6]Academia Europaea Tbilisi Regional Knowledge Hub, Ivane Javakhishvili Tbilisi State University, 0179 Tbilisi, Georgia

# Contributed equally

* Correspondence to: david.prangishvili@pasteur.fr and mart.krupovic@pasteur.fr
Institut Pasteur, Department of Microbiology,
75015 Paris, France
Tel: 33 (0)1 40 61 37 22

Running title
Intracellular envelopment and release of an archaeal virus

Competing Interests
The authors declare that they have no competing interests.

ABSTRACT

Viruses of Archaea represent a distinctive and enigmatic part of the virosphere, displaying unique virion architectures and little genomic similarity to viruses infecting bacteria or eukaryotes. The lack of relationships to other known viruses suggests that the mechanisms of virus-host interaction are also likely to be novel. To gain new insights into archaeal virus-host interactions, we studied the life cycle of the enveloped, ~2 μm-long Sulfolobus islandicus filamentous virus (SIFV), a member of the family *Lipothrixviridae*, infecting a hyperthermophilic and acidophilic archaeon *Sulfolobus islandicus* LAL14/1. Dual-axis electron tomography, in combination with other microscopy and biochemical techniques, showed that filamentous SIFV virions are assembled in the cell cytoplasm, forming twisted virion bundles organized on a non-perfect hexagonal lattice. Remarkably, our results indicate that envelopment of the helical nucleocapsids takes place inside the cell, rather than by budding as in the case of other known enveloped viruses. The mature virions are released from the cell through large, six-sided pyramidal portals in the cell envelope. The viral protein gp43, exclusive to members of the *Deltalipothrixvirus* and *Betalipothrixvirus genera*, was identified as the protein responsible for the formation of virus-associated pyramids and its heterologous expression in *Escherichia coli* led to the formation of pyramidal structures in the bacterial membrane. Collectively, our results provide insights on the assembly and release of lipothrixviruses, and suggest a novel biological phenomenon for viral envelopment that requires further investigation.

INTRODUCTION

Hyperthermophilic archaeal viruses are among the most enigmatic members of the global virosphere, with many of them displaying unique virion architectures and genomic contents (Dellas et al., 2014; Krupovic et al., 2018; Munson-McGee et al., 2018; Prangishvili et al., 2017). The understanding on virus-host interactions in Archaea remains scarce when compared to bacterial or eukaryotic viruses. However, recent studies have provided first insights into different steps of the infection cycle for several model archaeal viruses, showing that some of the mechanism used by archaeal viruses to interact with the hosts are similar to those of eukaryotic and/or bacterial viruses, whereas others are unique to archaeal viruses (Bize et al., 2009; Martínez-Alvarez et al., 2016; Pina et al., 2014; Quemin et al., 2013; Quemin et al., 2016; Uldahl et al., 2016).

Two major strategies of virion assembly and release have been described for hyperthermophilic archaeal viruses. One strategy is exemplified by the Sulfolobus spindle-shaped virus 1 (SSV1), the prototypic member of the *Fuselloviridae* family, whereby virion assembly is concomitant with its release via budding through the host cell envelope, closely resembling the release of many eukaryotic enveloped viruses, such as HIV-1 and influenza (Quemin et al., 2016). This release strategy typically does not result in the lysis of the infected cell and, following the eukaryotic virus paradigm, is expected to be common to other enveloped archaeal viruses. By contrast, viruses which assemble virions intracellularly employ egress strategy involving the disruption and death of the host cell. Archaeal viruses have evolved a unique cell lysis mechanism based on the formation of large pyramidal structures, dubbed virus-associated pyramids (VAPs), on the host cell surface (Prangishvili and Quax, 2011). The VAPs protrude through the surface protein (S-) layer, the only component of the archaeal cell envelope besides the cytoplasmic membrane (Albers and Meyer, 2011), and at the end of the infection cycle, the triangular facets of the VAP come apart as flower petals, producing apertures through which the mature virions exit the host cell (Bize et al., 2009; Brumfield et al., 2009; Daum et al., 2014; Quax et al., 2011). Thus far, the VAP-based egress mechanism has been shown to be used by viruses belonging to three unrelated families, namely, *Rudiviridae*, *Turriviridae* and *Ovaliviridae*, all infecting hyperthermophilic and acidophilic archaea of the order Sulfolobales. The VAPs formed by non-enveloped rod-shaped rudiviruses and icosahedral turriviruses are seven-sided (i.e., the VAP has seven triangular facets) (Brumfield et al., 2009; Daum et al., 2014; Fu et al., 2010) and are built from homologous proteins which, in all likelihood, have been exchanged between viruses from the two families by horizontal gene transfer (Quax et al., 2010; Snyder et al., 2011). By contrast, the VAPs built by the ovalivirus SEV1 are six-sided, but the protein responsible for the VAP formation has not been identified (Wang et al., 2018). Notably, similar six-sided pyramids have been also observed on the surface of hyperthermophilic neutrophiles of the order Thermoproteales (Bize et al., 2008; Rensen et al., 2015), suggesting that VAP-based egress strategy is widespread among hyperthermophilic archaeal viruses.

Filamentous viruses of the family *Lipothrixviridae* are among the most broadly distributed archaeal viruses, with representatives being isolated from hot springs in Iceland, Italy, Russia, USA and Japan (Arnold et al., 2000; Bettstetter et al., 2003; Bize et al., 2008; Häring et al., 2005; Liu et al., 2019; Liu et al., 2018; Vestergaard et al., 2008). Lipothrixviruses have linear double-stranded (ds) DNA genomes and based on genomic similarities are divided into four genera, *Alphalipothrixvirus*, *Betalipothrixvirus*, *Gammalipothrixvirus* and *Deltalipothrixvirus*. Structural studies have shown that all lipothrixviruses share the same virion organization; namely, linear dsDNA is complexed and condensed by two paralogous major capsid proteins (MCPs) into a helical nucleocapsid, which is further enveloped with a lipid membrane (Kasson et al., 2017; Liu et al., 2018; Wang et al., 2020a). Both ends of the virion are capped with terminal structures responsible for host recognition and binding (Pina et al., 2014). Similar to several other hyperthermophilic archaeal viruses (DiMaio et al., 2015; Wang et al., 2020c; Wang et al., 2019b), the dsDNA in the nucleocapsid of lipothrixviruses is stored in the A-form (Kasson et al., 2017; Liu et al., 2018; Wang et al., 2020a), which is believed to be one of the adaptations to high temperature environments. Remarkably, the lipid envelope surrounding the nucleocapsid of lipothrixviruses is twice thinner than the cytoplasmic membrane of the host cell. It has been shown that gammalipothrixvirus AFV1 selectively recruits from the host those tetraether lipids species, which can

be bent into a U-shaped horseshoe conformation, and molecular dynamics simulation has further suggested that these lipids form a thin monolayer membrane around the nucleocapsid (Kasson et al., 2017). By contrast, the envelope of alphalipothrixvirus SFV1 is strongly enriched in archaeol, a short lipid molecule corresponding to ~1% of lipids in the host membrane (Liu et al., 2018). However, whether the viral envelope is acquired during the budding process, as in the case of the majority of other enveloped viruses (Rheinemann and Sundquist, 2020), remains unknown. Notably, previous studies have suggested that lipothrixviruses are released without causing host cell lysis (Arnold et al., 2000; Bettstetter et al., 2003; Häring et al., 2005), but the exact mechanism has not been investigated.

Here, we have investigated the assembly and release of Sulfolobus islandicus filamentous virus (SIFV), the type member of the *Betalipothrixvirus* genus. The SIFV virions are enveloped, flexible, filamentous particles measuring ~2 μm in length and 24 nm in width (Figure 1A). At each end of the filament, the SIFV virions are decorated with terminal mop-like structures which presumably play a role in host recognition (Arnold et al., 2000; Bettstetter et al., 2003; Häring et al., 2005). Using dual-axis electron tomography, we show that the ~2 μm-long SIFV virions are assembled in the cytoplasm of the infected cells which have the diameter of 1-1.2 μm, and are released at the end of the infection cycle through six-sided VAPs. The 89 aa-long SIFV protein gp43 is sufficient for VAP formation and its heterologous expression in *Escherichia coli* leads to formation of similar structures in the bacterial membrane. The VAP protein is conserved in all members of the *Betalipothrixvirus* and *Deltalipothrixvirus* genera, but is unrelated to any of the previously characterized VAP proteins from other viruses. Unexpectedly, our results show that, differently from other characterized enveloped viruses, SIFV nucleocapsids are enveloped with a lipid membrane inside the host cell by an unknown mechanism.

RESULTS

Efficient SIFV propagation in *Sulfolobus islandicus* LAL14/1

To obtain insights into the life cycle of liprothixviruses and to investigate the mechanism by which their virions are released, we focused on SIFV (Arnold et al., 2000). Although *Sulfolobus islandicus* HVE10/4 has been characterized as the only susceptible host of the virus (Arnold et al., 2000), we found that the virus propagates more efficiently in *Sulfolobus islandicus* LAL14/1 (Zillig et al., 1998), a closely related strain (Jaubert et al., 2013), yielding higher titres of infectious SIFV particles (Figure S1). Consequently, *S. islandicus* LAL14/1 was used as the host in all subsequent experiments. The *S. islandicus* LAL14/1 cells display an irregular coccoid morphology typical of *Sulfolobus* species and have a diameter of ~1 µm (Albers and Siebers, 2014). Thus, SIFV virions are twice as long as the diameter of the host cell (Figure 1B), indicating that internal volume optimization might be required for efficient virion morphogenesis inside the infected cell.

To choose the optimal infection parameters for the study of SIFV morphogenesis and egress, we first determined the length of the infection cycle by performing a one-step growth experiment using a multiplicity of infection (MOI) of 0.01. A sharp increase in the extracellular virus titer at 11 hpi, signified the length of the latent period (Figure 1C). With 26 ± 7 virions being produced per cell, the burst size is comparable to that determined for the filamentous non-enveloped rudivirus SIRV2 (Bize et al., 2009). The adsorption assay showed that SIFV binding to the host cells is highly efficient, with nearly 70% of the virions being attached to the host cells within the first 2 minutes post-infection (Figure 1C), ensuring nearly synchronous infection of the *S. islandicus* population. The adsorption rate constant calculated at 2 minutes post-infection was 5.8×10^{-9} ml min^{-1}, which is similar to those reported for the turrivirus STIV, rudivirus SIRV2 and bicaudavirus SMV1 (Hartman et al., 2019; Quemin et al., 2013; Uldahl et al., 2016). A cell-free control in which SIFV was incubated at 75°C in the *S. islandicus* LAL/1 growth medium was performed to ensure that the observed decrease in the virus titer is a result of virus adsorption rather than virion inactivation in the harsh conditions, i.e., high temperatures and acidic pH. The virus titer of the cell-free control did not change over the 30 min of incubation.

Infection of *S. islandicus* cells using an MOI as low as 0.01 resulted in obvious growth retardation of the culture (Figure 1D), whereas at MOIs > 1 there was a significant decrease in the number of colony forming units (Figure 1E). The effects were more pronounced upon infection with higher MOIs, suggesting that SIFV infection leads to cell death in an MOI-dependent manner. The impact of SIFV on the growth dynamics and viability of the infected cells is reminiscent of those reported for the rod-shaped lytic virus SIRV2 (Bize et al., 2009), which is structurally and evolutionarily related to SIFV, but is not enveloped (DiMaio et al., 2015; Wang et al., 2020a). Notably, SIRV2 causes massive degradation of the host chromosome upon infection (Bize et al., 2009). To determine whether this is also the case during SIFV infection, the intracellular DNA content of non-infected and SIFV-infected cultures was monitored over time by flow cytometry. Unlike for SIRV2, there was no apparent host DNA degradation in the case of SIFV-infected cells (Figure S2).

Envelopment of SIFV virions occurs in the cytoplasm

To gain insights into SIFV virion assembly and envelopment, the infected cells were analyzed using dual-axis electron tomography at 10 and 12 hpi. The reconstructed tomographic volumes were analyzed using a convoluted neural network (CNN) algorithm to recognize virions, ribosomes and S-layer. At 12 hpi, bundles of filamentous particles resembling SIFV virions were observed (Figure 2A-F), whereas at 10 hpi similar structures were hardly detectable (Figure S3A-B), consistent with the results of the one-step growth experiment showing that SIFV virions start to be released 11 hpi (Figure 1C). Reconstructed electron tomograms showed that virions are organized in clusters along the membrane plane (Figure 2A). In cross-sections of the infected cells we could trace up to 70% of the total virion length (i.e., 1.4 µm out of 2 µm), indicating that virion bundles are folded to fit the volume of the cell. Interestingly, ribosome-like structures were occasionally ordered along the viral particles (Figure S4), suggesting an active role during virion morphogenesis.

Analysis of the transversal tomographic sections showed that filamentous virions are present in two forms (Figure 2D, Figure S3B-C): (i) nucleocapsids surrounded by lower density rings that presumably represent viral envelopes, and (ii) nucleocapsids devoid of any visible envelope (Figures 2D-2F, Figure S3C). Linear density profiles measured across four non-enveloped and enveloped nucleocapsids located adjacent to each other showed that the pixel intensities were the same in the region corresponding to the nucleocapsid, whereas profiles of enveloped virions have additional intensities at each side of the cross-section profile, representing the viral envelope (Figures 2G-H). Consistent with this interpretation, no significant difference was found between the diameters of the enveloped and non-enveloped nucleocapsid cores (Figure 2I). However, as expected, the center-to-center distances measured between the enveloped nucleocapsids were significantly larger than those between the non-enveloped nucleocapsids, consistent with the additional spacing contributed by the envelopes (Figure 2J). Collectively, these findings suggest that virion assembly and maturation take place in the cytoplasm of infected cells and proceed through the initial formation of naked nucleocapsids, which are subsequently enveloped. Intriguingly, this means that envelopment itself occurs in the cytoplasm rather than by extrusion of the naked nucleocapsids through the cellular membrane.

To further characterize virion bundles inside the cells, we performed three-dimensional reconstruction of the tomographic data (Figure 3A). Top and lateral projections showed that the virion bundles are organized on a non-perfect hexagonal lattice (Figures 3B, 3E) and are slightly twisted (Figures 3B-D). In particular, virions in the periphery of the bundle twist around the virions in the center at 8° angle (Figure 3C, D).

Characterization of the SIFV envelope

Recent structural characterization of the SIFV virions has revealed that the envelope surrounding the nucleocapsid is twice as thin as the cytoplasmic membrane of the host, as observed for lipothrixviruses from other genera (Kasson et al., 2017; Liu et al., 2018; Wang et al., 2020a). To determine the lipid composition of the viral envelope and host cell membrane, we performed liquid chromatography with tandem mass spectrometry (LC-MS/MS) analysis of the mature SIFV virions and *S. islandicus* LAL14/1 cells. It turned out the lipid composition of the viral envelope is quantitatively very different from that of the host membrane (Figure 2K). The *S. islandicus* LAL14/1 membrane nearly exclusively contains C_{40} glycerol dibiphytanyl glycerol tetraether (GDGT) species, long lipid molecules spanning the entire thickness of the membrane, which is effectively a monolayer of GDGT lipids (Villanueva et al., 2014). The dominant lipid species (~60% of all lipids) is GDGT-4 carrying four cyclopentane rings (Figure S5). By contrast, the envelope of SIFV is strongly enriched in C_{20} *sn*-2,3-glycerol diphytanyl ether lipid (Figure S5), known as archaeol, and GDGT-0, two lipid species that together account for less than 1% of the host lipids (Figure 2K). Thus, similar to some other archaeal viruses (Bamford et al., 2005; Kasson et al., 2017; Liu et al., 2018; Quemin et al., 2015), the lipids are incorporated into the SIFV envelope in a highly selective manner, in line with the observation that the envelope is not acquired through the budding process.

Cryo-EM reconstruction of the mature SIFV virions has suggested that one of the two SIFV MCPs interacts with the lipid envelope through an unstructured loop (Wang et al., 2020a). To verify whether either of the two SIFV MCPs interacts with the membrane within infected cells, *S. islandicus* LAL14/1 cells were sampled at different time points post infection, fractionated into soluble and membrane fractions and analyzed by SDS-PAGE and LC-MS/MS. Interestingly, we found that at 12 hpi both MCPs were largely found in the membrane fraction (see also below). This subcellular localization of the MCPs was obtained repeatedly and was independent of the method used to break the cells.

SIFV is released from the cell through hexagonal VAPs

The intracellular envelopment of SIFV virions raises questions regarding the mechanism of their egress from the host cell. Thus, to better understand this last stage of the SIFV life cycle, infected cells were monitored by transmission electron microscopy (TEM) at different time points after infection. The TEM analysis at 12 and 24 hpi revealed the presence of 6-sided (hexagonal) apertures on the cell surface

(Figure 4A). No such structures were observed on the surface of non-infected cells. Scanning electron microscopy (SEM) analysis confirmed the presence of perforations in the envelope of infected cells at 12 hpi (Figure 4B). The hexagonal apertures closely resemble the opened VAPs previously observed in *Pyrobaculum oguniense* cells (Rensen et al., 2015). The analysis of thin sections of infected cells demonstrated the presence of pyramidal structures in SIFV-infected cells (Figure 4C), similar to those previously described for lytic viruses of the families *Rudiviridae*, *Turriviridae* and *Ovaliviridae* (Bize et al., 2009; Brumfield et al., 2009; Wang et al., 2018).

To obtain further details about the SIFV-induced VAPs, we used dual-axis electron tomography. The VAPs displayed considerable variation in size: the height (measured from the base to the tip of the VAP) ranged from ~38 to 124 nm and the diameter (measured between the opposite sides of the VAPs' base) varied from 48 to 220 nm (n= 22). The SIFV VAPs grow outwards, penetrating and disrupting the S-layer of the host cell (Figures 5A and S6). At 12 hpi, when the virions are being released, open VAPs were also detected; VAP opening leads to the loss of the intracellular content, including the virions, leaving behind the empty "ghost cells" (Figure S6). Occasionally, SIFV VAPs were associated with dense spherical bodies (Figure 5A), also observed in the case of SIRV2 and STIV VAPs (Daum et al., 2014; Fu et al., 2010). The relevance of these structures for VAP formation and/or virus release remains unknown. Notably, unlike in the case of VAPs produced by other viruses (Brumfield et al., 2009; Quax et al., 2011; Wang et al., 2018), where multiple VAPs per cell were observed, in the case of SIFV, no more than three VAPs were detected per section of SIFV-infected cells.

Orthogonal views of SIFV VAPs clearly reveal their hexagonal base (Figure 5B), consistent with the six-sided apertures observed in the cell envelope by negative stain electron microscopy (Figures 4A-B). A three-dimensional (3D) map of a SIFV VAP in closed conformation was obtained by manually segmentation (Figures 5C-D). The reconstruction shows that SIFV pyramids are baseless hollow structures consisting of six triangular sides. Presence of VAPs of highly variable sizes, including small VAPs located beneath the S-layer, it is likely that VAP formation is nucleated by a small hexameric assembly which develops into a six-sided VAP by gradual growth of the triangular facets.

SIFV gp43 is sufficient for VAP formation

SIFV does not encode identifiable homologs of the previously reported VAP proteins P98 and C92 of rudivirus SIRV2 and turrivirus STIV, respectively (Quax et al., 2010; Snyder et al., 2011). Given that the VAPs of rudiviruses and turrivirus are seven-sided (Daum et al., 2014; Fu et al., 2010; Quax et al., 2010), whereas those of SIFV are six-sided, it is conceivable that the proteins forming the two types of VAPs might be unrelated. Thus, to identify the protein responsible for formation of the SIFV VAPs, the proteins enriched in the membrane fraction of SIFV-infected cells at 12 hpi were analyzed by SDS-PAGE and LC-MS/MS. Five protein bands (B1-B5) appeared or grew in intensity in the membrane fraction of infected cells at 12 hpi, compared to earlier time points post infection or the non-infected control (Figure 6A). The upper bands B1 and B2 (with molecular masses of ~20 and ~24 kDa, respectively), also visible in the membrane fraction of infected cells at 10 hpi, were identified as the two major capsid proteins of SIFV, as discussed above. The bands B3 (~15 kDa), B4 (~12 kDa) and B5 (~9 kDa) were detected exclusively in the membrane fraction of infected cells at 12 hpi (Figure 6A). Whereas B3 contained no identifiable virus proteins, LC-MS/MS analysis has shown that bands B4 and B5 contain several viral proteins of unknown functions, namely, SIFV gp15, gp43, gp20 and gp71 (Table S1).

The four proteins were analyzed for the presence of predicted N-terminal transmembrane domain (TMD), a feature found in other VAP proteins (Quax et al., 2010; Snyder et al., 2011). Only gp43 fulfilled this requirement. To investigate if SIFV gp43 is involved in VAP formation, the corresponding gene was cloned and expressed in *Escherichia coli* Rosetta (DE3) pLys. Protein expression was confirmed by western-blot analysis with anti-6XHis antibodies (Figure S7). Electron microscopy analysis of thin sections of gp43-expressing cells after 4 hours after induction showed the presence of multiple VAP-like structures on the cytoplasmic membrane of the bacterial cells protruding towards the

periplasmatic space (Figure 6B). The pyramidal structures were always found in the closed conformation, suggesting that the signal triggering the opening of the pyramid structures is archaea-specific. Nevertheless, at 4 h post induction, the optical density of induced cell culture was significantly lower compared to the non-induced control, suggesting that the protein expression and VAP formation are toxic to bacteria. Collectively, these results indicate that gp43 is the only essential component of VAPs.

The gp43 of SIFV is 89aa-long and is the shortest VAP protein identified to date (Figure S8). To study the distribution of SIFV gp43 homologs, we performed PSI-BLAST searches against the viral non-redundant protein database at NCBI. SIFV gp43 homologs were found to be conserved in all characterized members of the *Betalipothrixvirus* and *Deltalipothrixvirus* genera of the *Lipothrixviridae* family, but have no identifiable homologs in viruses from other families. Thus, SIFV gp43-like proteins form a new family of VAP proteins, distinct from that including other known VAP proteins from rudiviruses and turriviruses (Quax et al., 2010; Snyder et al., 2011). The two protein families display similar features, including the N-terminal TMDs and extensive α-helical content (Figure S8). However, the pattern of amino acid conservation is distinct in the two families. Regardless, it is clear that superficially similar complex structures, such as VAPs, can be built from proteins with highly different sequences.

DISCUSSION

Strategies of virus-host interactions in archaea remain largely unexplored. In this study, we attempted to rectify this situation on the example of SIFV, a member of the family *Lipothrixviridae.* Structural studies have shown that lipothrixvirus virions consist of a helical nucleocapsid enveloped with a thin lipid membrane (Kasson et al., 2017; Liu et al., 2018; Wang et al., 2020a). The nucleocapsid of lipothrixviruses is homologous to the helical capsid of non-enveloped viruses of the *Rudiviridae* family (Wang et al., 2020a), indicating that viruses from the two families have evolved from a common ancestor. Based on phylogenomic and structural studies, it has been suggested that this ancestor was an enveloped virus, resembling lipothrixviruses, and that rudiviruses have emerged by shedding the lipid membrane (Wang et al., 2020a). Given that in most viruses, virion envelope plays key roles during different stages of virus-host interaction, such as genome delivery or virion egress, functional comparison between rudiviruses and lipothrixviruses offers a unique opportunity to study the evolution of virus host-interactions. Furthermore, the fact SIFV virions are twice longer than the diameter of the host cell poses an intriguing question of how the assembly of such virions is orchestrated inside the host cell.

In many ways, the SIFV infection process resembles that of rudiviruses and some other archaeal viruses, although there are also considerable differences. The SIFV infection cycle starts with rapid virion adsorption to the host cell surface. The high rate of adsorption, similar to that documented for other hyperthermophilic archaeal viruses (Hartman et al., 2019; Papathanasiou et al., 2019; Quemin et al., 2013; Uldahl et al., 2016), is likely to be important for limiting the exposure of the viral particles to extreme environmental conditions. Unlike many other archaeal viruses which recognize their hosts through pili (Hartman et al., 2019; Rowland et al., 2020; Wang et al., 2020b; Wang et al., 2019a), SIFV has been suggested to bind the receptor located directly within the cellular membrane (Arnold et al., 2000). The latent period of SIFV is ~11 h, which is similar to that typical of many other archaeal viruses and might also signify the general preference of hyperthermophilic archaeal viruses to spend more time within, rather than outside of the cell.

Electron tomography analysis has provided the first insights into the intracellular assembly of the SIFV virions. The formation of the SIFV nucleocapsids is highly reminiscent of the assembly of the mature rudivirus virions (Bize et al., 2009; Daum et al., 2014). In the case of both viruses, filamentous (nucleo)capsids are assembled in the cell interior within bundles containing multiple virions. The

virions in the bundles are arranged not randomly but on a hexagonal lattice, resembling the property of many icosahedral virions to form crystalline-line arrays within the cell cytoplasm (Duyvesteyn et al., 2018). The three-dimensional reconstruction has shown that the bundles are twisted at an 8° angle. Although the biological relevance of the SIFV bundle twisting is unclear, a similar behavior has been characterized for many biological filaments and artificial materials, such as carbon nanotube ropes and micropatterned filament arrays (Bruss and Grason, 2013). Interestingly, it has been concluded that the lowest energy state for a bundle of sufficiently flexible and long filaments is generically twisted (Bruss and Grason, 2013). Accordingly, twisting of the SIFV bundles might derive from the geometric frustration of the bulk virion packing and surface energy of non-contacting virions at the boundary of the bundle. In addition to twisting, the bundles undergo more pronounced bending to fit within the cell.

Whereas the non-enveloped capsids represent mature virions primed for egress in the case of rudiviruses, the SIFV nucleocapsids have to be further enveloped. Most of the studied enveloped viruses, including filamentous Ebola viruses, escape from their host cells by budding (Noda et al., 2006; Rheinemann and Sundquist, 2020; Votteler and Sundquist, 2013). Thus, budding is often considered, by default, to be the mechanism of envelope acquisition in enveloped viruses. Indeed, virion morphogenesis and egress of the archaeal fusellovirus SSV1 are concomitant and occur at the cellular cytoplasmic membrane via a mechanism highly reminiscent of the budding of enveloped eukaryotic viruses (Quemin et al., 2016). Similarly, pleolipoviruses and bicaudaviruses have been also proposed to use budding as an exit mechanism (Demina and Oksanen, 2020; Liu et al., 2017). Hence, the finding that SIFV virions are enveloped inside the cell cytoplasm was unexpected. Using electron tomography, we observed both non-enveloped and enveloped SIFV virions within the same cell, suggesting an order of events from non-enveloped nucleocapsids, resembling mature rudivirus virions, to mature, enveloped SIFV virions (Figure 7). In eukaryotes, some viruses acquire envelopes inside the cell by budding through organelles, such as endoplasmic reticulum or Golgi complex (Risco et al., 2003; Roingeard et al., 2004; Shi et al., 2010). However, internal, membrane-bound compartments have never been observed in *Sulfolobus* cells, excluding the possibility that SIFV virions are enveloped by budding through intracellular membranes. Consequently, the envelopment of SIFV might occur by a completely novel mechanism, involving either de novo membrane formation or trafficking of lipids from the cytoplasmic membrane to the virion assembly centers – neither mechanism has been demonstrated for other prokaryotic viruses. Recently, it has been suggested that ovoid-shaped archaeal virus SEV1 also acquires its envelope intracellularly (Wang et al., 2018). Thus, the mechanism of membrane remodeling and envelopment employed by lipothrixviruses might be widespread among evolutionarily unrelated archaeal viruses.

Contrary to our initial expectation based on the egress mechanisms employed by other enveloped viruses, SIFV is a lytic virus. Electron microscopy analysis showed that SIFV induces formation of VAPs on the surface of infected cells, which gradually grow in size (Figure 7). A similar mechanism of virion release has been previously described for viruses from families *Rudiviridae*, *Turriviridae* and *Ovaliviridae* (Bize et al., 2009; Brumfield et al., 2009; Wang et al., 2018). Among these, SIFV VAPs more closely resemble VAPs formed by ovalivirus SEV1 (Wang et al., 2018): VAPs of both viruses are six-sided, rather than seven-sided as observed for rudiviruses and turriviruses (Bize et al., 2009; Brumfield et al., 2009). Notably, whereas VAP proteins of rudiviruses and turriviruses share relatively high sequence similarity (Quax et al., 2010; Snyder et al., 2011), gp43 of SIFV has no homologs in ovalivirus SEV1 and appears to be unrelated to the VAP proteins of rudiviruses and turriviruses (Figure S8). The protein responsible for VAP formation during SEV1 infection remains unknown, but it is likely to represent a separate protein family. Furthermore, gp43 homologs could not be identified in lipothrixviruses of the *Alphalipothrixvirus* and *Gammalipothrixvirus* genera, suggesting that a considerable diversity of protein families capable of VAP formation remains to be discovered in the archaeal virosphere.

Our results show that expression of gp43 in *E. coli* leads to VAP formation in the bacterial membrane. The same result was obtained with the VAP protein of rudivirus SIRV2 (Quax et al., 2011). It should be noted that bacterial and archaeal membranes consist of unrelated lipids: whereas bacterial

membranes are bilayers containing phospholipids (fatty acids linked to glycerol moieties by ester linkages), the membrane of *Sulfolobus islandicus* is largely a monolayer of tetraether lipids (long isoprenoid chains capped on both ends by glycerol moieties through ether linkages). Thus, the inherent ability of the two proteins, lacking any recognizable sequence similarity, to form VAPs in both bacterial and archaeal membranes is remarkable. Whether proteins from the two families have diverged from a common ancestor or have originated independently remains unclear. Regardless, the general replication cycle of enveloped lipothrixviruses and rudiviruses appear to be closely similar, involving formation of helical nucleocapsids which are released through VAPs, suggesting that evolutionary transition from a postulated enveloped lipothrixvirus-like ancestor to the non-enveloped rudivirus-like ancestor did not entail any major adaptations in the mechanisms underlying the virus-host interactions. This finding raises the questions regarding the function of the membrane in lipothrixviruses. We hypothesize that the primary role of the lipothrixvirus envelope is protection of the viral genome in hot and acidic environment. Indeed, structural studies have shown that MCP packing in rigid rod-shaped rudiviruses is tighter than in flexible lipothrixviruses (Wang et al., 2020a). Thus, once the ancestral MCP has evolved towards forming a more robust virus particle, the membrane layer might have become dispensable and was shed.

ACKNOWLEDGEMENTS

This work was supported by l'Agence Nationale de la Recherche (France) project ENVIRA and Emergence(s) project MEMREMA from Ville de Paris (to M.K.). D.P.B. is part of the Pasteur – Paris University (PPU) International PhD Program. This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No 665807.

MATERIAL AND METHODS

Optimal host

To determine the optimal host for SIFV, *Sulfolobus islandicus* LAL14/1 and *Sulfolobus islandicus HVE10/4* (Zillig et al., 1998; Zillig et al., 1993) were grown aerobically at 75°C, pH 3.5 in rich medium containing 0.2% (wt/vol) tryptone, 0.1% (wt/vol) sucrose, 0.1% (wt/vol) yeast extract and mineral salt solution, as described previously (Zillig et al., 1993). Exponentially growing cultures of *S. islandicus* HVE10/4 and LAL14/1 were infected with the same SIFV preparation and incubated at 75°C under agitation. After two days, cells were removed by centrifugation (7,000 rpm, 20 min, Sorvall 1500 rotor) and viruses were collected and concentrated by ultracentrifugation (40,000 rpm, 2 h, 10 °C, Beckman 126 SW41 rotor). The virus titer was measured by plaque assay for both cultures.

Plaque assays

Serial dilutions of the viral preparations were mixed with preheated fresh *S. islandicus* LAL14/1 cells. Subsequently, 5 mL of pre-heated rich medium containing 0.3% Phytagel™ (Sigma-Aldrich, USA) were added to the prepared mixtures, vortexed and poured into plates containing 0.1% yeast extract, 0.2% sucrose (w/v) and 0.7% Phytagel™ (Sigma-Aldrich, USA) (Arnold et al., 2000). After three days of incubation at 75°C, visible plaques of 0.5-3 mm appeared on the plates.

Virus production and purification.

Sulfolobus islandicus LAL14/1 was grown aerobically at 75°C, pH 3.5 in rich medium (Zillig et al., 1993). Early exponentially growing cultures (150 mL) were infected with fresh preparations of SIFV and incubated at 75°C under agitation. After two days, the infected cell cultures were transferred into fresh cultures (250 mL) of *S. islandicus* LAL14/1 cells and incubated for 2 days. Cells were removed by centrifugation (7,000 rpm, 20 min, Sorvall 1500 rotor) and viruses were collected and concentrated by ultracentrifugation (40,000 rpm, 2 h, 10 °C, Beckman 126 SW41 rotor). The concentrated particles were resuspended in buffer A: 20 mM KH_2PO_4, 250 mM NaCl, 2.14 mM $MgCl_2$, 0.43 mM $Ca(NO_3)_2$, and <0.001% trace elements of Sulfolobales medium, pH 6 (Quemin et al., 2015), and stocked at 4°C until used. For SDS/PAGE and mass spectrometry analyses, virus particles were further purified by ultracentrifugation in a CsCl buoyant density gradient (0.45 $g \cdot mL^{-1}$) with a Beckman SW41 rotor at 39,000 rpm for 20 h at 10 °C. The opalescent bands were collected with a needle and a syringe and dialyzed against buffer A for 2 hours.

Adsorption assays

Exponentially growing cultures of *S. islandicus* LAL14/1 cells were infected with SIFV at a MOI of 0.05. Infected cells were incubated under agitation for 30 min at 75°C. Samples of 1 mL were taken at defined time intervals, cells were pelleted at low speed centrifugation, and supernatants were kept at 4°C. The percentage of unadsorbed virus particles was determined by plaque assay comparing the viral concentration in the supernatants with the virus titer in the control, a cell-free culture incubated under the same conditions as the treated cultures. Experiments were conducted in triplicate. The adsorption rate constant (k) was calculated using the following formula, as described previously (Uldahl et al., 2016): $k = 2.3/B_t \times \log_{10}(P_0/P_t)$, where B_t= concentration of cells at a specific time t (cell/mL), P_0 = concentration of the virus at zero time (PFU/mL) and P_t = concentration of not adsorbed viruses at a specific time t (PFU/mL).

One-step growth curve

The virus was added to early exponentially growing cultures of *S. islandicus* LAL14/1 cells at an MOI of 0.01. After 30 min of incubation at 75°C, the cultures were diluted in prewarmed medium to prevent the occurrence of new events of adsorption during the experiment. The diluted cultures were incubated at 75°C under agitation. Samples of 1 mL were collected at defined time points and immediately centrifuged at low speed to separate the free viruses (supernatant) from the cells (pellet). The PFU titers at different time points were determined by plaque assay. The burst size was estimated by dividing the average amount of viruses present in the supernatants after virus release (11-12 hpi) by the average

amount of viruses present in the supernatants before release (0-10 hpi) (Bize et al., 2009). Experiments were conducted in triplicate.

Infection studies

Exponentially growing cultures of *S. islandicus* LAL14/1 were infected with SIFV using an MOI range of 0.01 to 10 and incubated at 75°C with shaking. Samples of 1 mL were collected from each culture at defined time points, and the cell density (OD_{600}) and the number of viable cells (CFU) were measured. The CFU counting was carried out as described previously (Bize et al., 2009). Non-infected cultures were used as controls. Experiments were conducted in triplicate.

Flow cytometry

Non-infected and SIFV infected *S. islandicus* LAL14/1 cells were collected at defined time intervals and fixed with 70% cold ethanol overnight. The fixed cells were pelleted by centrifugation at low speed and resuspended in 1 ml of PBS buffer with 0.05% Tween-20. Cells were pelleted a second time and resuspended in 100 µL of staining buffer containing 40 µg/ml propidium iodide (PI). After staining (>30 min), the samples were analyzed for DNA content using an ImageStreamX MarkII Quantitative imaging analysis flow cytometry (Merck Millipore, Germany). The data of 100,000 imaged cells or particles were collected from each sample and analyzed for DNA content by IDEAS data analysis software.

Transmission electron microscopy

For negative-staining TEM analysis, 5 µl of the samples were applied to carbon-coated copper grids, negatively stained with 2% uranyl acetate (w/v) and imaged with the transmission electron microscope FEI Spirit Tecnai Biotwin operated at 120 kV.

Scanning electron microscopy

Liquid cultures of infected *S. islandicus* LAL14/1 cells were fixed with 2.5% glutaraldehyde for 2 hours at room temperature. Afterwards, fixed cells were washed in 0.1 M HEPES buffer pH 7.2, post fixed for 1 h 30 in 1% osmium tetroxide in 0.1 M HEPES buffer pH 7.2 and rinsed with distilled water. Samples were dehydrated through a graded series of ethanol (25, 50, 75, 95 and 100%), followed by critical point drying with CO2. Dried specimens were sputtered with 20 nm gold palladium using a GATAN Ion Beam Coater and examined and photographed with a JEOL JSM 6700F field emission scanning electron microscope operating at 7 Kv. Images were acquired from the upper SE detector (SEI).

Preparation of thin sections using chemical fixation

Samples at determined time points were fixed by adding glutaraldehyde to the growth medium to a final concentration of 1% during 2 h at room temperature. Cells were pelleted down and resuspended in 20 µL in 0.1M Hepes, pH 7. The suspensions were mixed with 20 µL of low melting point agar (type VII) and solidified on ice. After solidification, the samples were cut into small pieces, post fixed in 1% osmium tetroxide in 0.1M cacodylate for 1 h on ice, dehydrated in graded ethanol series, infiltrated in propylenoxide:Epon (1:1) and final infiltrated in pure Epon. Samples were transferred into embedding molds and polymerized for 48h at 60°C. Embedded cells were cut into 70-nm thin sections with an Ultracut R microtome (Leica) and collected on Formvar-coated copper grids. Thin sections (70 nm) were stained and imaged with the transmission electron microscope FEI Spirit Tecnai Biotwin operated at 120 kV.

Fractioning of SIFV-infected and non-infected S. islandicus LAL14/1 cells

The protein content of SIFV-infected cells was stablished as described previously with small modifications (Quax et al., 2010). Briefly, an exponentially growing culture of *S. islandicus* LAL14/1 was infected with SIFV at a MOI of 3. Liquid samples were collected at determined time points and cells were pelleted using low speed centrifugation. Concentrated cells were resuspended in 20 mM Tris-HCl pH 7 and disrupted by sonication (6 cycles with 20 s of sonication and 40 s of pause). Unbroken cells were removed from the total cell lysate by low-speed centrifugation. Membrane and cytosol

fractions were separated by high-speed centrifugation at 100,000 × g for 40 min at 4°C. The membrane fraction was resuspended in 20 mMTris-HCl pH 7. The proteins in each fraction were solubilized in 1% n-Dodecyl β-d-maltoside and incubated at 1 hour at 37°C. Samples were heat-denaturated in presence of SDS sample loading buffer and 1.25% β-mercaptoethanol. Proteins were separated by electrophoresis on a precast NuPAGE gel 4-12% Bis-Tris Bolt (ThermoFisher) and visualized with Instant Blue™ staining (Expedeon).

Mass spectrometry

The stained protein bands were excised from the gel and in-gel digested with trypsin. The generated peptides were separated and identified by nano-LC-MS/MS (Proteomics Platform, Institut Pasteur) using an Ultimate 3000 system (Dionex) coupled to an LTQ-Orbitrap Velos system (Thermo Fisher Scientific). Peptide masses were searched against annotated SIFV and *S. islandicus* LAL14/1 proteomes using Andromeda with MaxQuant software, version 1.3.0.5

Expression of SIFV gp43 in E. coli

The SIFV ORF43 was amplified from a pure SIFV stock and cloned into the pUC19 plasmid. The vector contains an isopropyl β-D-1-thiogalactopyranoside–inducible promoter that was used for the expression of the His-tagged protein. *E. coli* Rosetta(DE3)pLys (Novagen Merck) cells were transformed with the construct, liquid cultures were grown in 2YS medium and induced with 1 mM IPTG at ODs of 0.4-0.6 for 4 hours. A non-induced cell culture was used as control. Thin sections of *E. coli* cells were prepared as described above.

Western blot

Induced and non-induced *E. coli* cells were harvested by centrifugation at low speed and resuspended in 20 mM Tris-HCl pH 7. Samples were heat-denaturated in presence of SDS sample loading buffer and, optionally, 10 mM DTT as redactor agent, and loaded onto a 4-12% polyacrylamide Bis-Tris gradient gel. Proteins were transferred onto a PVDF membrane. The presence of the 6×His-tagged SIFV gp43 was visualized using a 1:10,000 dilution of the rabbit polyclonal Anti-6X His tag® antibody conjugated with HRP (Abcam).

Sample preparation for electron tomography

Samples for electron microscopy and electron tomography were treated as described previously (Quemin et al., 2015). Cultures at 10 and 12 hours post infection were pelleted by low-speed centrifugation and resuspended in a minimal volume of rich medium. Samples were processed into capillaries (Leica), transferred into a lecithin-coated sample holder and frozen with a high-pressure freezing machine (HPM100, Leica). The samples were subsequently freeze-substituted with 0.5% glutaraldehyde, 1% OsO4, 0.2% uranyl acetate, 2% H2O, and 4% methanol in acetone according to the following schedule: −90°C for 40 h, 5°C/h for 6 h, −60°C for 8 h, 5°C/h for 6 h, and −30°C for 8 h (Leica Microsystems). Cells were rinsed in acetone and slowly infiltrated with Agar 100 epoxy resin (Agar Scientific, United Kingdom). After heat polymerization, embedded cells were cut into 70-nm thin sections with an Ultracut R microtome (Leica) and collected on Formvar-coated copper grids. Thin sections (70 nm) were poststained with 4% uranyl acetate for 45 min and lead citrate staining during 5 min. Samples were imaged with the transmission electron microscope FEI Spirit Tecnai Biotwin operated at 120 kV. For electron tomography, embedded cells were cut into 200-nm thick sections with an Ultracut R microtome (Leica) and collected on Formvar-coated copper grids. Protein A-gold particles of 10 nm were added on both sides of the sections and stained with 2% lead citrate in water.

Dual-axis electron tomography

Grids were loaded on a dual-axis tomography holder and observed with a TECNAI F20 Transmission Electron Microscope (FEI) operating at 200kV and equipped with a 4k x 4k CCD camera (Ultrascan 4000, Gatan). Micrographs, tilt series and maps, in low and middle magnifications, were acquired using SerialEM (Mastronarde, 2005; Schorb et al., 2019). After identifying areas of interest on middle

magnification maps, the areas were baked using a total dose of 1,500 e^-/A^2. The continuous tilt scheme was used for the automatic acquisition of micrographs every 1° over a ±55° range at higher magnification (usually 29K or 50K). After the acquisition of tilt series in all areas of interest, grid was manually rotated by 90° to acquire the second orthogonal tilt axis series in the same areas of interest.

Initial image shifts of the tilt series were estimated using IMOD's function tiltxcorr (Mastronarde and Held, 2017). Alignments were further optimized in IMOD using the tracing of gold fiducials across the tilt series. Three-dimensional reconstructions were calculated in IMOD by weighted back projection using the SIRT-like radial filter to enhance contrast and facilitate subsequent segmentation analysis. The volumes from the two tilt axes were combined to one using fiducials present in IMOD (Mastronarde, 1997).

Segmentation and analysis of tomographic data

Tomograms were displayed and analyzed using the 3dmod interface of IMOD (Kremer et al., 1996). Archaeal cellular membranes were modeled with manual tracing every 40 slices and the use of IMOD's interpolator. Closed or open contours were used, depending if a full archaeon was included in the field of view or not, respectively. Pyramids of interest were manually traced using IMOD's Slicer to orient the planes for tracing parallel to the pyramid's base so the hexagonal shape was easily identifiable. Traced models were meshed to surfaces using the imodmesh function of IMOD. Meshed models were finally used to calculate density maps from them in eman2 (Chen et al., 2017). For the rest of the features that were easily identified: S-layer, ribosomes, virion main bodies and envelope of virions; training sets were prepared based on positive and negative segmentation examples in order to train the convoluted neural network (CNN) algorithms implemented in eman2.3 (Chen et al., 2017). Final segmentation results were visualized as iso-surfaces with UCSF Chimera (Pettersen et al., 2004). Many false positives were removed by thresholding out smaller sized particles for each feature using the 'Hide Dust' tool of UCSF Chimera. For the representation of the S-layer, the CNN segmentation result was used as a mask and the S-layer iso-surfaces were produced based on the original density map of the volume data.

REFERENCES

Albers, S.-V., and Siebers, B. (2014). The Family Sulfolobaceae. In The Prokaryotes: Other Major Lineages of Bacteria and The Archaea, E. Rosenberg, E.F. DeLong, S. Lory, E. Stackebrandt, and F. Thompson, eds. (Berlin, Heidelberg: Springer Berlin Heidelberg), pp. 323-346.

Albers, S.V., and Meyer, B.H. (2011). The archaeal cell envelope. Nature reviews Microbiology *9*, 414-426.

Arnold, H.P., Zillig, W., Ziese, U., Holz, I., Crosby, M., Utterback, T., Weidmann, J.F., Kristjanson, J.K., Klenk, H.P., Nelson, K.E., *et al.* (2000). A novel lipothrixvirus, SIFV, of the extremely thermophilic crenarchaeon Sulfolobus. Virology *267*, 252-266.

Bamford, D.H., Ravantti, J.J., Rönnholm, G., Laurinavicius, S., Kukkaro, P., Dyall-Smith, M., Somerharju, P., Kalkkinen, N., and Bamford, J.K. (2005). Constituents of SH1, a novel lipid-containing virus infecting the halophilic euryarchaeon Haloarcula hispanica. Journal of virology *79*, 9097-9107.

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.

Bize, A., Karlsson, E.A., Ekefjärd, 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. Proceedings of the National Academy of Sciences of the United States of America *106*, 11306-11311.

Bize, A., Peng, X., Prokofeva, M., Maclellan, K., Lucas, S., Forterre, P., Garrett, R.A., Bonch-Osmolovskaya, E.A., and Prangishvili, D. (2008). Viruses in acidic geothermal environments of the Kamchatka Peninsula. Research in microbiology *159*, 358-366.

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. Journal of virology *83*, 5964-5970.

Bruss, I.R., and Grason, G.M. (2013). Topological defects, surface geometry and cohesive energy of twisted filament bundles. Soft Matter *9*, 8327-8345.

Chen, M., Dai, W., Sun, S.Y., Jonasch, D., He, C.Y., Schmid, M.F., Chiu, W., and Ludtke, S.J. (2017). Convolutional neural networks for automated annotation of cellular cryo-electron tomograms. Nature methods *14*, 983-985.

Daum, B., Quax, T.E., Sachse, M., Mills, D.J., Reimann, J., Yildiz, Ö., Häder, S., Saveanu, C., Forterre, P., Albers, S.V., *et al.* (2014). Self-assembly of the general membrane-remodeling protein PVAP into sevenfold virus-associated pyramids. Proceedings of the National Academy of Sciences of the United States of America *111*, 3829-3834.

Dellas, N., Snyder, J.C., Bolduc, B., and Young, M.J. (2014). Archaeal Viruses: Diversity, Replication, and Structure. Annu Rev Virol *1*, 399-426.

Demina, T.A., and Oksanen, H.M. (2020). Pleomorphic archaeal viruses: the family Pleolipoviridae is expanding by seven new species. Archives of virology.

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.

Duyvesteyn, H.M.E., Ginn, H.M., Pietila, M.K., Wagner, A., Hattne, J., Grimes, J.M., Hirvonen, E., Evans, G., Parsy, M.L., Sauter, N.K., *et al.* (2018). Towards in cellulo virus crystallography. Sci Rep *8*, 3771.

Fu, C.Y., Wang, K., Gan, L., Lanman, J., Khayat, R., Young, M.J., Jensen, G.J., Doerschuk, P.C., and Johnson, J.E. (2010). In vivo assembly of an archaeal virus studied with whole-cell electron cryotomography. Structure (London, England : 1993) *18*, 1579-1586.

Häring, M., Vestergaard, G., Brügger, K., Rachel, R., Garrett, R.A., and Prangishvili, D. (2005). Structure and genome organization of AFV2, a novel archaeal lipothrixvirus with unusual terminal and core structures. Journal of bacteriology *187*, 3855-3858.

Hartman, R., Eilers, B.J., Bollschweiler, D., Munson-McGee, J.H., Engelhardt, H., Young, M.J., and Lawrence, C.M. (2019). The Molecular Mechanism of Cellular Attachment for an Archaeal Virus. Structure (London, England : 1993) *27*, 1634-1646.e1633.

Jaubert, C., Danioux, C., Oberto, J., Cortez, D., Bize, A., Krupovic, M., She, Q., Forterre, P., Prangishvili, D., and Sezonov, G. (2013). Genomics and genetics of Sulfolobus islandicus LAL14/1, a model hyperthermophilic archaeon. Open Biol *3*, 130010.

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*.

Kremer, J.R., Mastronarde, D.N., and McIntosh, J.R. (1996). Computer Visualization of Three-Dimensional Image Data Using IMOD. Journal of Structural Biology *116*, 71-76.

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, J., Gao, R., Li, C., Ni, J., Yang, Z., Zhang, Q., Chen, H., and Shen, Y. (2017). Functional assignment of multiple ESCRT-III homologs in cell division and budding in Sulfolobus islandicus. Molecular microbiology *105*, 540-553.

Liu, Y., Brandt, D., Ishino, S., Ishino, Y., Koonin, E.V., Kalinowski, J., Krupovic, M., and Prangishvili, D. (2019). New archaeal viruses discovered by metagenomic analysis of viral communities in enrichment cultures. Environmental Microbiology *21*, 2002-2014.

Liu, Y., Osinski, T., Wang, F., Krupovic, M., Schouten, S., Kasson, P., Prangishvili, D., and Egelman, E.H. (2018). Structural conservation in a membrane-enveloped filamentous virus infecting a hyperthermophilic acidophile. Nature communications *9*, 3360.

Martínez-Alvarez, L., Bell, S.D., and Peng, X. (2016). Multiple consecutive initiation of replication producing novel brush-like intermediates at the termini of linear viral dsDNA genomes with hairpin ends. Nucleic acids research *44*, 8799-8809.

Mastronarde, D.N. (1997). Dual-Axis Tomography: An Approach with Alignment Methods That Preserve Resolution. Journal of Structural Biology *120*, 343-352.

Mastronarde, D.N. (2005). Automated electron microscope tomography using robust prediction of specimen movements. Journal of Structural Biology *152*, 36-51.

Mastronarde, D.N., and Held, S.R. (2017). Automated tilt series alignment and tomographic reconstruction in IMOD. J Struct Biol *197*, 102-113.

Munson-McGee, J.H., Snyder, J.C., and Young, M.J. (2018). Archaeal Viruses from High-Temperature Environments. Genes (Basel) *9*.

Noda, T., Ebihara, H., Muramoto, Y., Fujii, K., Takada, A., Sagara, H., Kim, J.H., Kida, H., Feldmann, H., and Kawaoka, Y. (2006). Assembly and budding of Ebolavirus. PLoS Pathog *2*, e99.

Papathanasiou, P., Erdmann, S., Leon-Sobrino, C., Sharma, K., Urlaub, H., Garrett, R.A., and Peng, X. (2019). Stable maintenance of the rudivirus SIRV3 in a carrier state in Sulfolobus islandicus despite activation of the CRISPR-Cas immune response by a second virus SMV1. RNA biology *16*, 557-565.

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. Journal of computational chemistry *25*, 1605-1612.

Pina, M., Basta, T., Quax, T.E., Joubert, A., Baconnais, S., Cortez, D., Lambert, S., Le Cam, E., Bell, S.D., Forterre, P., *et al.* (2014). Unique genome replication mechanism of the archaeal virus AFV1. Molecular microbiology *92*, 1313-1325.

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., and Quax, T.E. (2011). Exceptional virion release mechanism: one more surprise from archaeal viruses. Curr Opin Microbiol *14*, 315-320.

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.

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. Proceedings of the National Academy of Sciences of the United States of America *108*, 3354-3359.

Quemin, E.R., Lucas, S., Daum, B., Quax, T.E., Kühlbrandt, W., Forterre, P., Albers, S.V., Prangishvili, D., and Krupovic, M. (2013). First insights into the entry process of hyperthermophilic archaeal viruses. Journal of virology *87*, 13379-13385.

Quemin, E.R., Pietilä, M.K., Oksanen, H.M., Forterre, P., Rijpstra, W.I., Schouten, S., Bamford, D.H., Prangishvili, D., and Krupovic, M. (2015). Sulfolobus Spindle-Shaped Virus 1 Contains Glycosylated Capsid Proteins, a Cellular Chromatin Protein, and Host-Derived Lipids. Journal of virology *89*, 11681-11691.

Quemin, E.R.J., Chlanda, P., Sachse, M., Forterre, P., Prangishvili, D., and Krupovic, M. (2016). Eukaryotic-Like Virus Budding in Archaea. mBio *7*, e01439-01416.

Rensen, E., Krupovic, M., and Prangishvili, D. (2015). Mysterious hexagonal pyramids on the surface of Pyrobaculum cells. Biochimie *118*, 365-367.

Rheinemann, L., and Sundquist, W.I. (2020). Virus Budding. In Reference Module in Life Sciences (Elsevier).

Risco, C., Carrascosa, J.L., and Frey, T.K. (2003). Structural maturation of rubella virus in the Golgi complex. Virology *312*, 261-269.

Roingeard, P., Hourioux, C., Blanchard, E., Brand, D., and Ait-Goughoulte, M. (2004). Hepatitis C virus ultrastructure and morphogenesis. Biology of the cell *96*, 103-108.

Rowland, E.F., Bautista, M.A., Zhang, C., and Whitaker, R.J. (2020). Surface resistance to SSVs and SIRVs in pilin deletions of Sulfolobus islandicus. Molecular microbiology *113*, 718-727.

Schorb, M., Haberbosch, I., Hagen, W.J.H., Schwab, Y., and Mastronarde, D.N. (2019). Software tools for automated transmission electron microscopy. Nature methods *16*, 471-477.

Shi, X., van Mierlo, J.T., French, A., and Elliott, R.M. (2010). Visualizing the replication cycle of bunyamwera orthobunyavirus expressing fluorescent protein-tagged Gc glycoprotein. Journal of virology *84*, 8460-8469.

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. Journal of virology *85*, 6287-6292.

Uldahl, K.B., Jensen, S.B., Bhoobalan-Chitty, Y., Martínez-Álvarez, L., Papathanasiou, P., and Peng, X. (2016). Life Cycle Characterization of Sulfolobus Monocaudavirus 1, an Extremophilic Spindle-Shaped Virus with Extracellular Tail Development. Journal of virology *90*, 5693-5699.

Vestergaard, G., Aramayo, R., Basta, T., Häring, M., Peng, X., Brügger, K., Chen, L., Rachel, R., Boisset, N., Garrett, R.A., *et al.* (2008). Structure of the acidianus filamentous virus 3 and comparative genomics of related archaeal lipothrixviruses. Journal of virology *82*, 371-381.

Villanueva, L., Damsté, J.S.S., and Schouten, S. (2014). A re-evaluation of the archaeal membrane lipid biosynthetic pathway. Nature Reviews Microbiology *12*, 438-448.

Votteler, J., and Sundquist, W.I. (2013). Virus budding and the ESCRT pathway. Cell host & microbe *14*, 232-241.

Wang, F., Baquero, D.P., Beltran, L.C., Su, Z., Osinski, T., Zheng, W., Prangishvili, D., Krupovic, M., and Egelman, E.H. (2020a). Structures of filamentous viruses infecting hyperthermophilic archaea explain DNA stabilization in extreme environments. Proceedings of the National Academy of Sciences of the United States of America *117*, 19643-19652.

Wang, F., Baquero, D.P., Su, Z., Beltran, L.C., Prangishvili, D., Krupovic, M., and Egelman, E.H. (2020b). The structures of two archaeal type IV pili illuminate evolutionary relationships. Nature communications *11*, 3424.

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

Wang, F., Cvirkaite-Krupovic, V., Kreutzberger, M.A.B., Su, Z., de Oliveira, G.A.P., Osinski, T., Sherman, N., DiMaio, F., Wall, J.S., Prangishvili, D., *et al.* (2019a). An extensively glycosylated archaeal pilus survives extreme conditions. Nature microbiology *4*, 1401-1410.

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. (2019b). 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*.

Zillig, W., Arnold, H.P., Holz, I., Prangishvili, D., Schweier, A., Stedman, K., She, Q., Phan, H., Garrett, R., and Kristjansson, J.K. (1998). Genetic elements in the extremely thermophilic archaeon Sulfolobus. Extremophiles : life under extreme conditions *2*, 131-140.

Zillig, W., Kletzin, A., Schleper, C., Holz, I., Janekovic, D., Hain, J., Lanzendörfer, M., and Kristjansson, J.K. (1993). Screening for Sulfolobales, their Plasmids and their Viruses in Icelandic Solfataras. Systematic and Applied Microbiology *16*, 609-628.

FIGURE LEGENDS

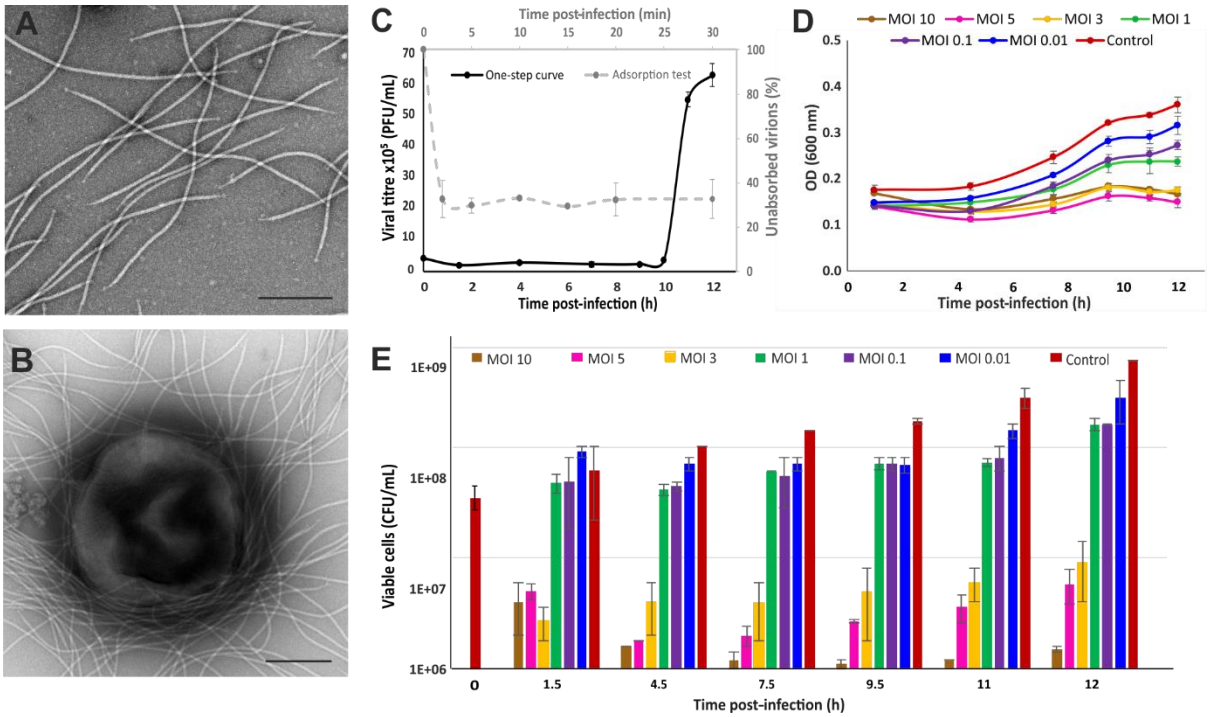


Figure 1. Characterization of the SIFV infection cycle in *S. islandicus* LAL14/1 cells. (A) Electron micrograph of purified SIFV particles negatively stained with 2% uranyl acetate. Scale bar, 500 nm. (B) Electron micrograph of *S. islandicus* LAL14/1 cells infected with SIFV at a MOI of 10. Sample was collected 2 min post-infection and negatively stained with 2% uranyl acetate. Scale bar, 500 nm. (C) One-step growth curve (black) and adsorption kinetics (grey) of SIFV using as host *S. islandicus* LAL14/1. For the one-step growth curve, cells were infected with an MOI of 0.01 and samples of unadsorbed viruses were collected as described in Materials and Methods. For the adsorption test, cells were infected with an MOI of 0.05 and the percentage of unadsorbed virions was determined as described in Materials and Methods. (D) Optical density (OD) of *S. islandicus* LAL14/1 liquid cultures infected at MOIs ranging from 0.01 to 10. (E) Number of viable cells (CFU/mL) of infected *S. islandicus* LAL14/1 liquid cultures at different MOIs (0.01-10).

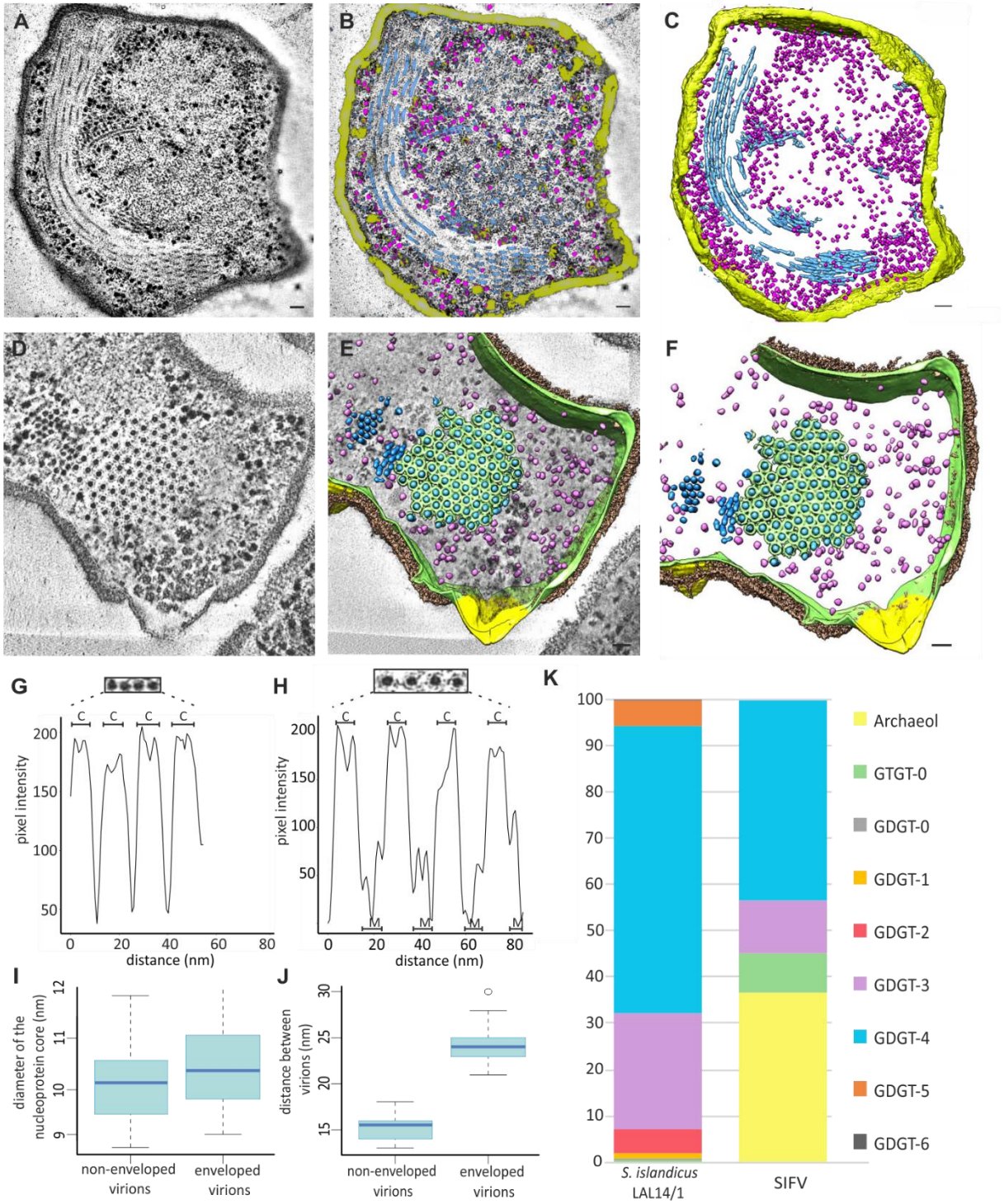


Figure 2. Assembly of SIFV virions in the cytoplasm of the host cell. (A) A slice through a reconstructed tomogram of a sectioned sample of SIFV-infected cells at 12 hpi. (B-C) A segmented and surface-rendering displays of the tomogram in panel A, including various viral and cellular components: S-layer (yellow), virions (blue), ribosomes (purple). Bars, 40 nm. (D) A slice through a reconstructed tomogram of a sectioned sample of SIFV-infected cells at 12 hpi, displaying a transversal view of the virions assembled in the host cytoplasm. (E-F) A segmented and surface-rendering displays of the tomogram in panel D, including various viral and cellular components: S-layer (pink), pyramid in a closed-conformation (yellow), cellular and viral membranes (green), nucleoprotein cores (blue), ribosomes (purple). Bars, 40 nm. (G-H) Linear density profiles of four non enveloped (G) and enveloped virions (H) located adjacent to each other, respectively. C, nucleoprotein core, M, membrane.

(I) Measurement of the diameter (nm) of the nucleoprotein cores of enveloped and non-enveloped virions. (J) Measurement of the distances (nm) between contiguous virions in clusters of enveloped and non-enveloped virions. The distance was measured between the centers of adjacent nucleoprotein cores. (K) Distribution of lipid species identified in S. islandicus LAL14/1 cells and highly purified SIFV virions.

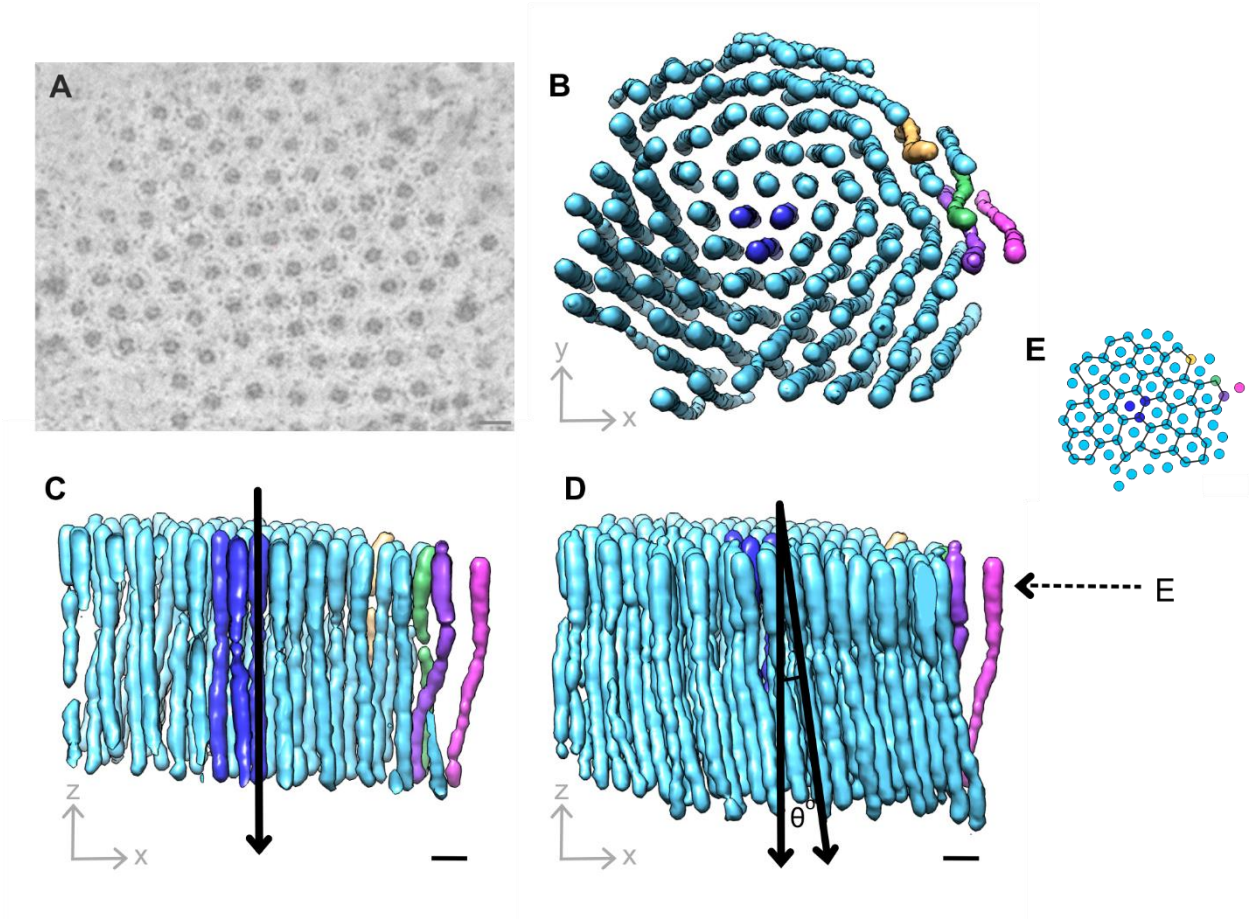


Figure 3. SIFV virions organized into twisted filament bundles. (A) A slice through a reconstructed tomogram of a cluster of enveloped virions observed at 12 hpi. (B) Top view of the array of enveloped virions. Virions located at the center and periphery of the bundle were colored to have them as positional references. (C) Cross-section through the middle of the array displayed in (B) brought to a vertical orientation. (D) Visualization of the total array displayed in (B). Virions located at the periphery were used to calculate the twist angle θ (θ=8°). (E) Black lines trace the non-perfect hexagonal lattice on which the SIFV virions are organized within the bundle.

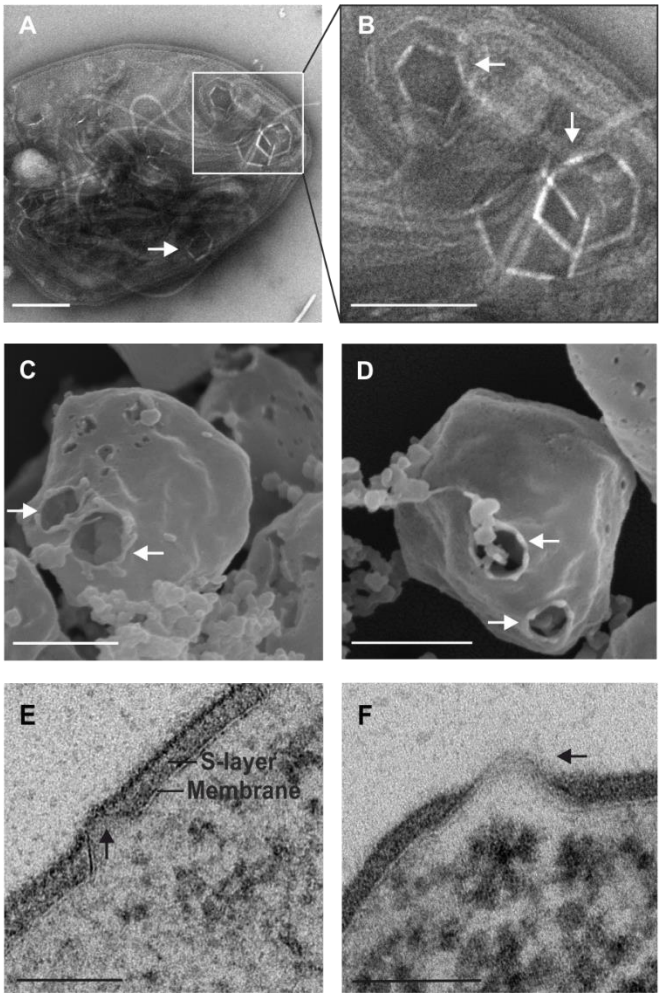
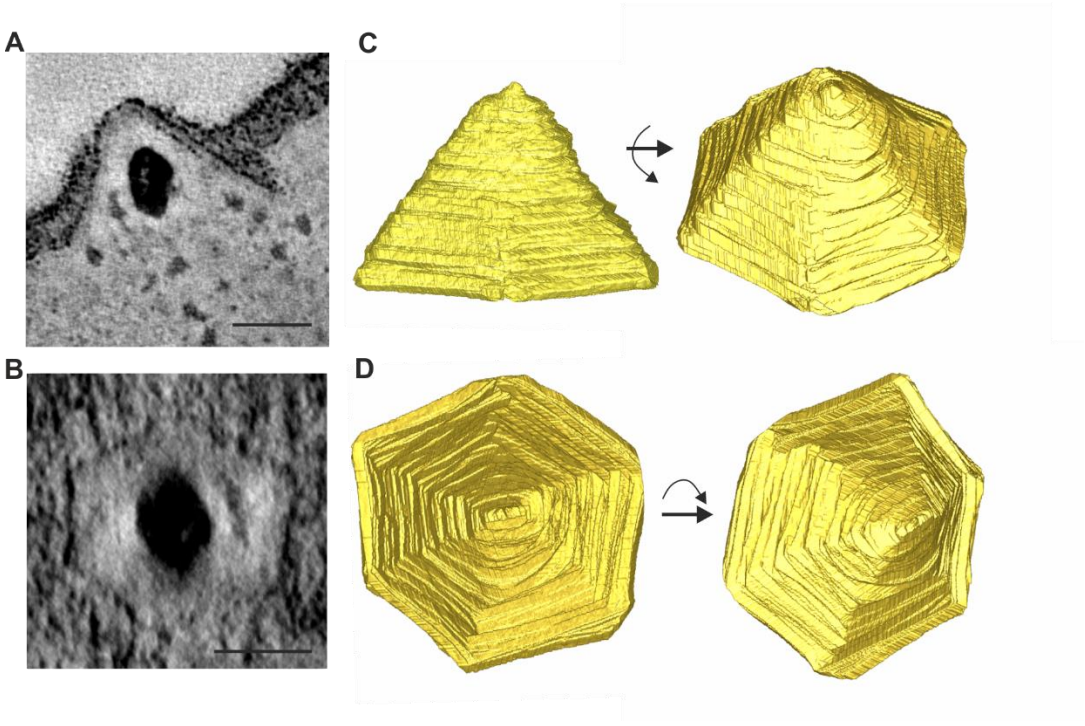


Figure 4. Visualization of infected *S. islandicus* LAL14/1 cells by TEM and SEM. (A-B) Infected *S. islandicus* LAL/14 cells collected at 12 hpi, negatively stained with 2% uranyl acetate and visualized under transmission electron microscopy (TEM). (C-D) Infected *S. islandicus* LAL/14 cells collected at 24 hpi and visualized by scanning electron microscopy (SEM). (E-F) Thin sections (70 nm) of infected *S. islandicus* LAL/14 cells collected at10 hpi and visualized by TEM. Arrows indicate VAPs at different stages. Scale bars, A, 200 nm; B, 100 nm; C, 200 nm; D, 200 nm; E, 200 nm; F, 100 nm.



828
829 **Figure 5. Visualization of the six-sided pyramidal structure formed by SIFV.** (A) A slice through
830 a reconstructed tomogram showing the closed conformation of a SIFV VAP. Scale bar, 50 nm. (B)
831 Orthogonal view of the SIFV VAP shown in panel A, displaying the hexagonal shape of the base. Scale
832 bar, 50 nm. (C) Lateral view of a 3D map in solid representation of a SIFV VAP. (D) Bottom view of
833 a 3D map in solid representation of a SIFV VAP.
834

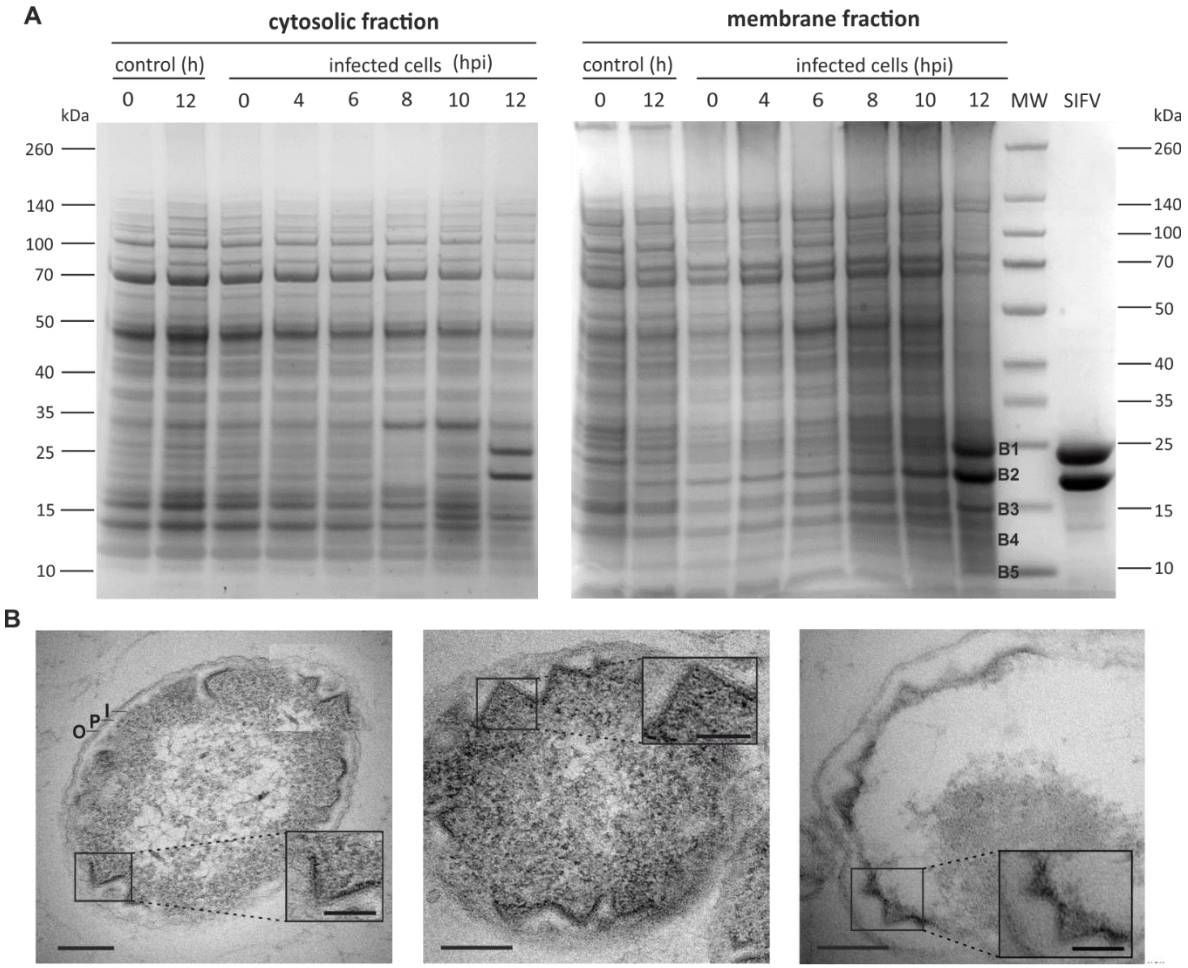


Figure 6. Identification of the viral protein involved in VAP formation. (A) Stained SDS-PAGE gels of the cytosolic and membrane fractions of *S. islandicus* LAL14/1 cells infected with SIFV. Cellular fractions were prepared as described in Materials and Methods. B1-B5 labelling denotes the protein bands that appeared in the membrane fraction as a result of the infection. MW, molecular weight marker. (B) Thin section electron micrographs of *E. coli* cells expressing SIFV gp43. I, inner membrane, P, periplasmic space, O, outer membrane. Scale bars, 200 nm; in insets: 100 nm.

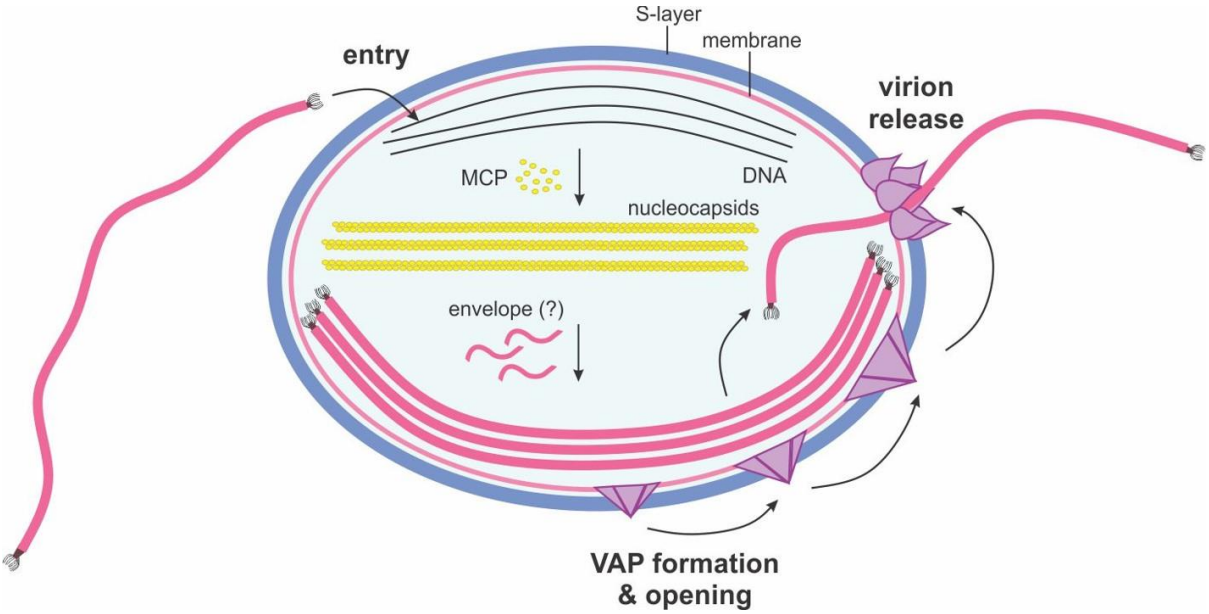


Figure 7. A schematic representation of the SIFV life cycle. The infection starts when the virus binds to its specific host receptor; subsequently, the virus delivers its genetic material into the cell through an unknown mechanism. The viral genome is replicated, and nucleoprotein cores are formed in the cytoplasm by binding of heterodimers of the two MCPs to the linear DNA. Concomitantly with the virion assembly, hexagonal virus-associated pyramids (VAPs) start growing on the surface of infected cells. The formation of mature virions is accomplished when the nucleoprotein cores are enveloped in the cytoplasm through an unclear mechanism. Subsequently, the pyramidal structures disrupt the S-layer and open to produce apertures through which the mature virions exit.

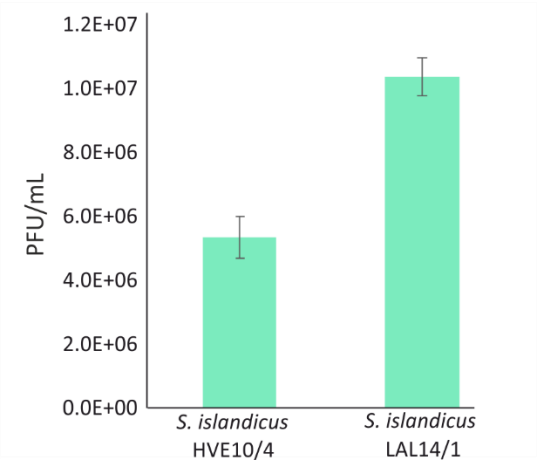


Figure S1. Comparison of the viral titre obtained upon the infection of two close-related *S. islandicus* strains HVE10/4 and LAL14/1 with SIFV. After infection of both strains using the same conditions, the viral titer (PFU/mL) was determined by plaque assay, as described in Material and Methods.

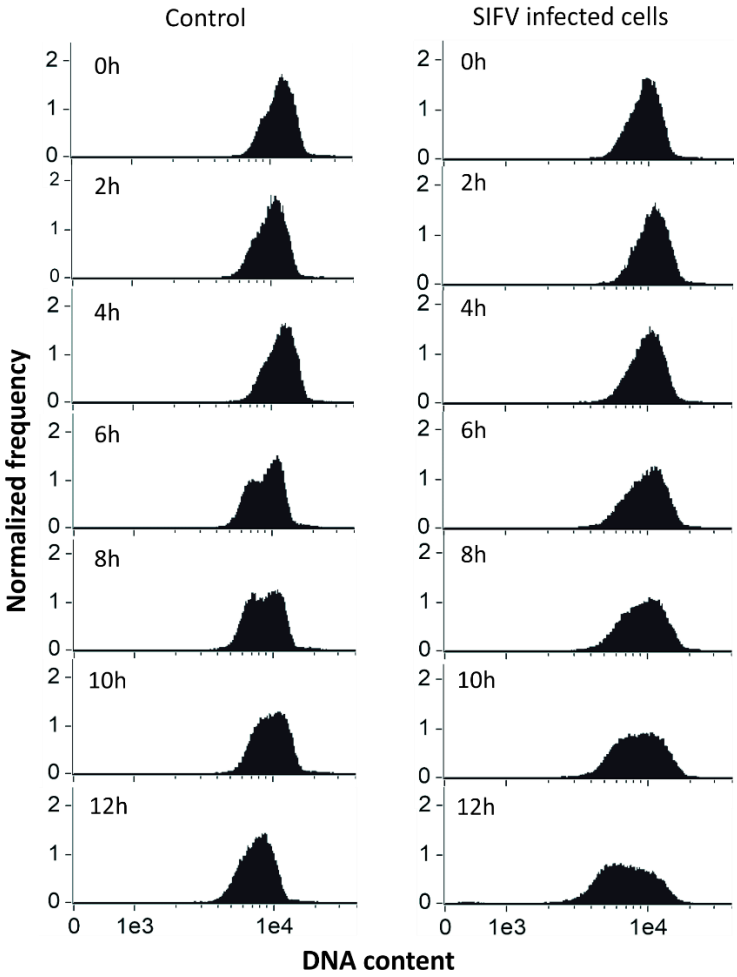


Figure S2. Flow cytometry analysis of infected *S. islandicus* LAL14/1 cells. Representative DNA content distribution of uninfected and SIFV-infected cells at determined time points (MOI ≈ 3). The content distribution was measured as described in Material and Methods.

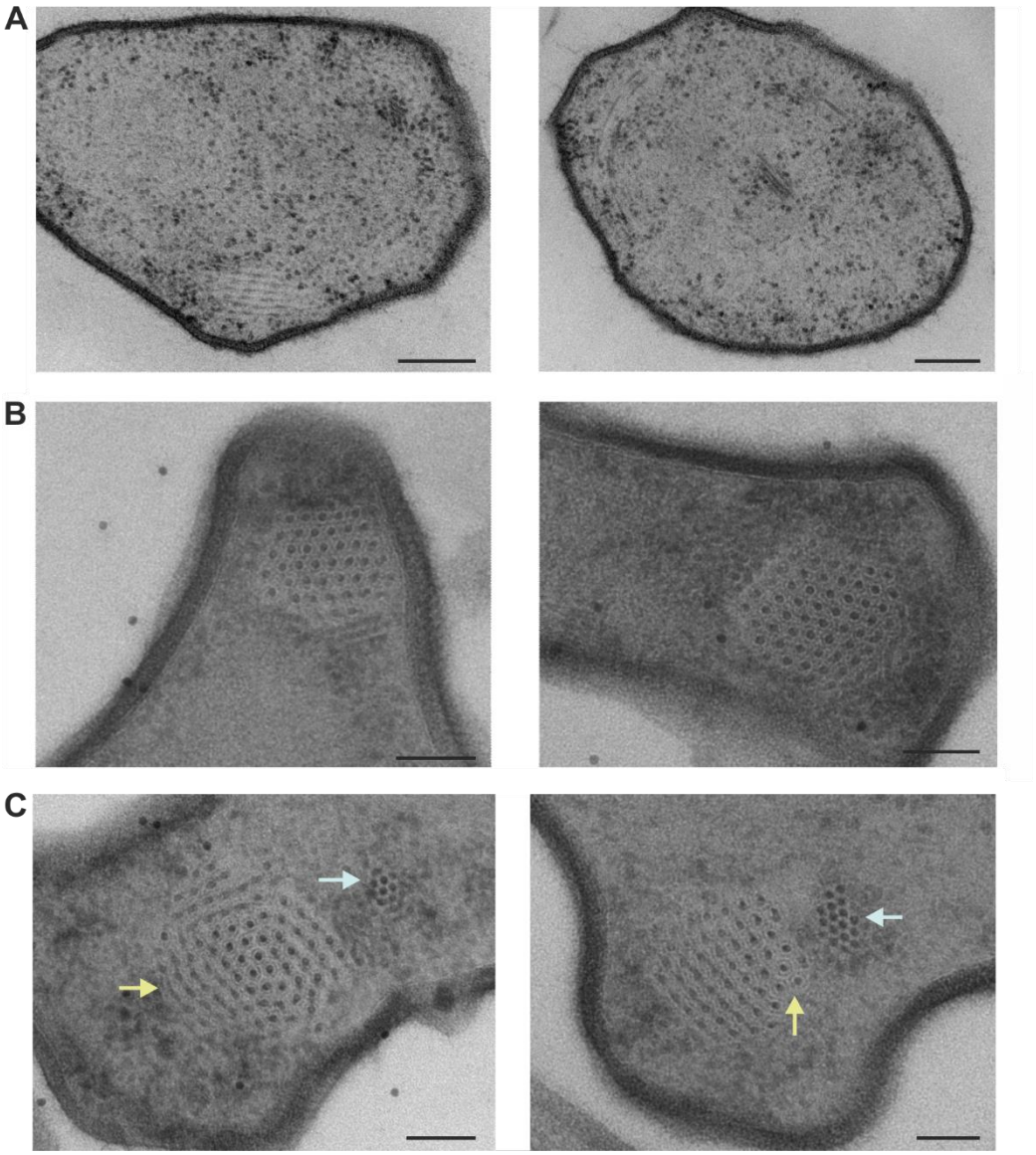


Figure S3. Slices through reconstructed tomograms of *Sulfolobus islandicus* LAL14/1 infected cells at 10 and 12 hpi. (A) Slices displaying SIFV infected cells at 10 hpi. (B) Slices displaying SIFV infected cells at 12 hpi. (C) Slices displaying SIFV infected cells at 12 hpi, wherein clusters of enveloped and non-enveloped nucleocapsids are highlighted with yellow and blue arrows, respectively. Scale bars, 100 nm.

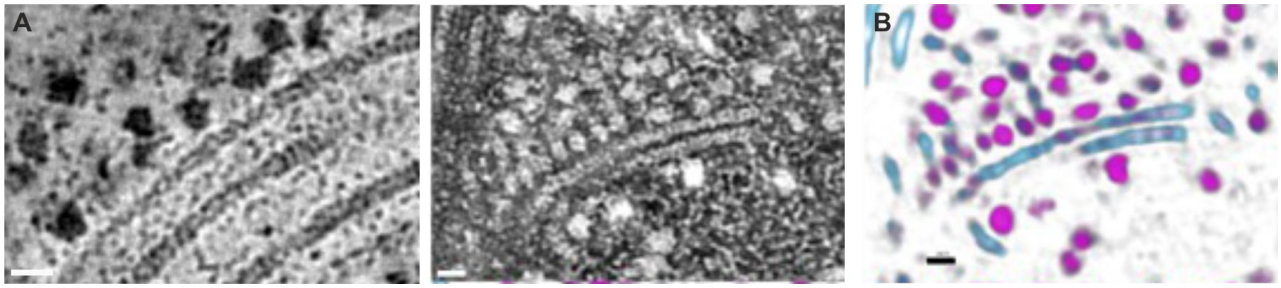


Figure S4. Visualization of ribosomes-like structures localized around the viral particles. (A) Slices through reconstructed tomograms displaying ribosomes-like structures ordered along the SIFV virions. (B) Ribosomes identified by a convoluted neural network (CNN) algorithm (purple) around virion particles (blue). Scale bar, 20 nm.

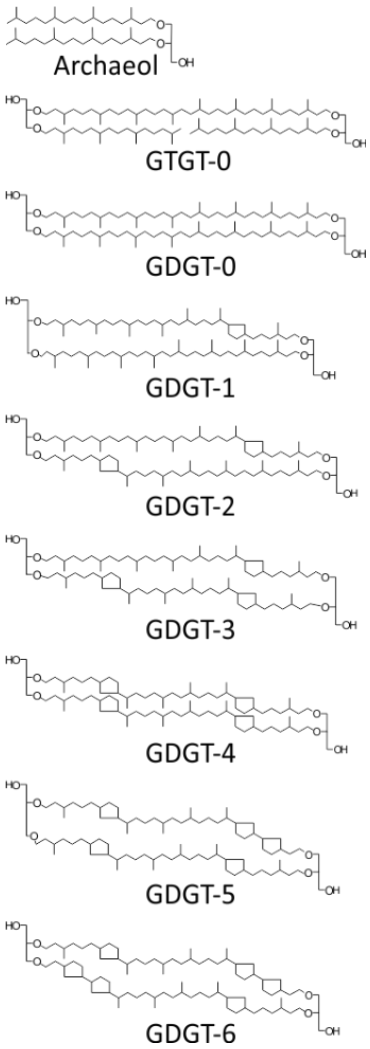


Figure S5. Chemical structures of the lipid species identified in SIFV virions and *S. islandicus* LAL14/1 cells.

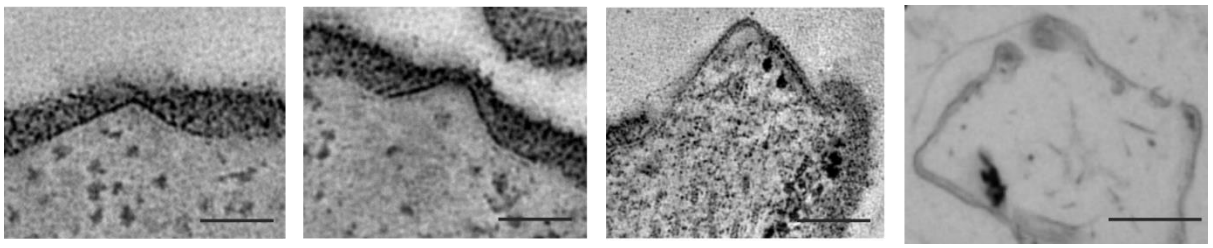


Figure S6. Visualization of the different stages of SIFV VAPs. The first three panels correspond to slices through reconstructed tomograms, whereas the last panel was observed from a thin section electron micrograph. Scale bar, first three panels 100 nm, last panel 400 nm.

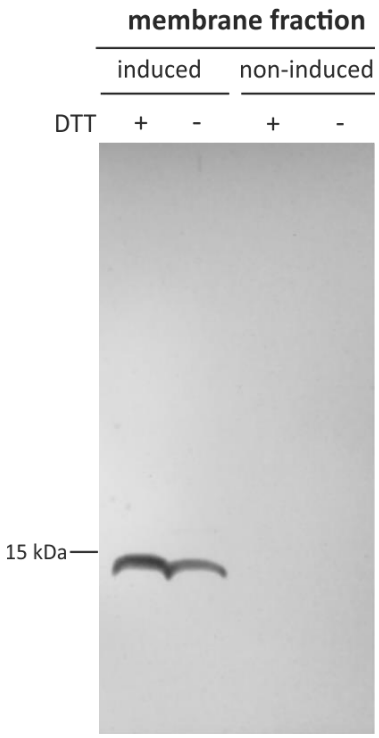


Figure S7. Detection of the His6-tagged SIFV gp43 in induced *E. coli* cells by western blot. Membrane fractions of induced and non-induced samples were collected at 4 hours post-induction. The effect of DTT on the collected fractions was tested.

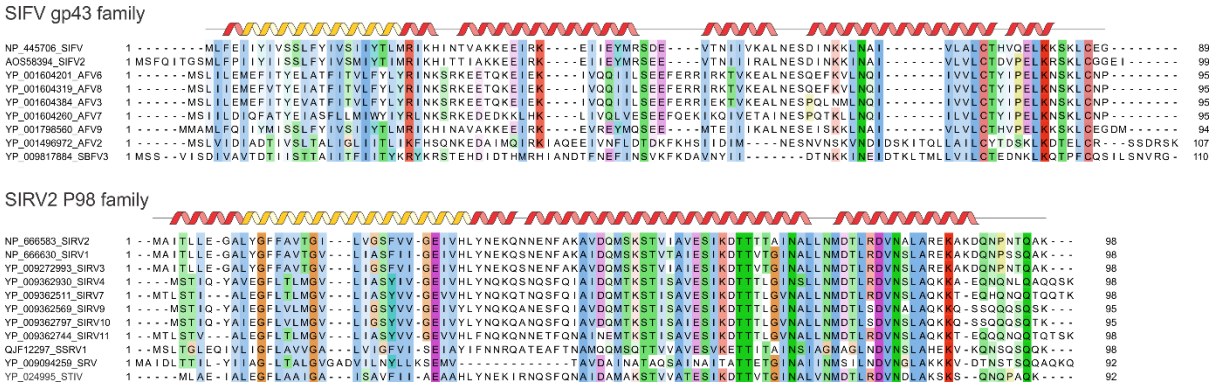


Figure S8. Multiple sequence alignment of the VAP families represented by SIFV gp43 (top) and SIRV2 P98 (bottom). Secondary structure was predicted using PSI-Pred and is shown above the corresponding alignments. Predicted α-helices are shown as ribbons. Yellow ribbons represent the segments predicted to form α-helical transmembrane domains. Each protein is labelled with the corresponding accession number.

Table S1. Proteins detected in the five bands (B1-B5) present in the membrane fraction of infected cells and analyzed by LC-MS/MS. Peptide masses were searched against the annotated SIFV proteome.

Band	Gene	Function	Molecular weight (kDa)	Organism	Sum of IBAQ	Protein ID
B1	SIFV0036	Major capsid protein	22.5	SIFV	6.41E+08	Y036_SIFVH
B2	SIFV0035	Major capsid protein	18.8	SIFV	4.33E+08	Y035_SIFVH
B4	SIFV0015	Hypothetical protein	11.7	SIFV	1.32E+08	Y015_SIFVH
	SIFV0043	Hypothetical protein	10.4	SIFV	6.80E+07	Y043_SIFVH
	SIFV0020	Hypothetical protein	10.8	SIFV	5.31E+07	Y020_SIFVH
B5	SIFV0071	Hypothetical protein	8.5	SIFV	6.09E+06	Y071_SIFVH

7.CONCLUSIONS AND FUTURE PERSPECTIVES

7.1 New archaeal viruses

It has been estimated that only about 0.01–0.1% of viruses present in geothermal acidic environments have been isolated (Snyder et al., 2010). To improve this situation and to expand our knowledge on the diversity of viruses thriving in extreme thermal environments, we explored the virus diversity in Italian hydrothermal environments. As a result, we successfully isolated and characterized five new archaeal viruses (**chapter 1**), which are valuable models for understanding the structure and evolution of archaeal viruses (**chapter 2**) and the mechanisms of their interaction with the hosts.

Globuloviruses and tristromaviruses remain among the most enigmatic archaeal viruses

Pyrobaculum spherical virus 2 (PSV2) and Pyrobaculum filamentous virus 2 (PFV2) were isolated from enrichment cultures grown at 90°C and neutral pH. Based on their morphologies and genome contents, the two viruses were classified into the families *Globuloviridae* and *Tristromaviridae*, respectively (**chapter 1**). Sequence similarity searches showed that both viruses do not share genes with members of other archaeal virus families. In addition, the function for the majority of predicted genes remains unknown, supporting the notion that globuloviruses and tristromaviruses stand out as some of the most mysterious among archaeal viruses.

Remarkably, despite the fact that tristromaviruses do not share genes with other filamentous archaeal viruses, the availability of the pure strain of PFV2 enabled its structural characterization by cryo-EM, which unexpectedly revealed deep evolutionary relationship of tristromaviruses with rudiviruses and lipothrixviruses (Wang et al., 2020c) (Figure 9). The reconstruction of PFV2 at 3.4 Å resolution showed that the nucleocapsid is constructed from two paralogous major capsid proteins, which are homologous to the capsid proteins of members of the *Rudiviridae* and *Lipothrixviridae*. Notably, we also found that the linear dsDNA in tristromavirus virions is also stored in the A-form. Due to the shared virion organization of viruses in the three families (Figure 9), the first class-rank taxon for archaeal viruses, ‘Tokiviricetes’, was proposed to unify them.

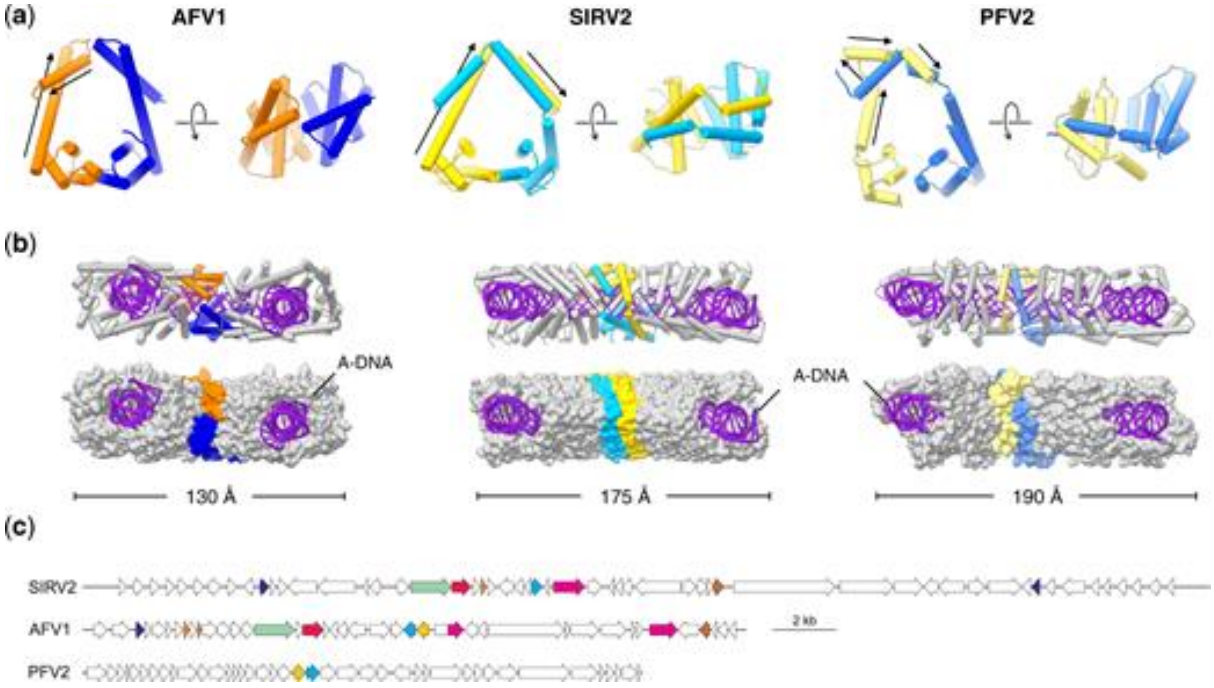


Figure 9. Comparison of the filamentous viruses AFV1, SIRV2 and PFV2. (a) MCP dimer (asymmetric unit) comparison of AFV1, SIRV2 and PFV2. The MCP1 of AFV1, SIRV2 and PFV2 are colored in orange, gold and yellow, respectively. The MCP2 of AFV1, SIRV2 and PFV2 are colored in blue, cyan and light blue, respectively. The N-terminal helices of MCP1 in AFV1, SIRV2 and PFV2 are marked with black arrows. (b) Wrapping of A-DNA in AFV1, SIRV2 and PFV2. Five MCP dimers are displayed: one MCP dimer is colored as in (a); the other four colored in gray. Proteins are shown in ribbon representation (top) and as surfaces (bottom). (c) Comparison of genome maps of rudivirus SIRV2 (NC_004086), lipothrixvirus AFV1 (NC_005830) and tristromavirus PFV2 (MN876844). Homologous genes (E < 1e–04) are indicated with the same colors. The homology between the MCPs of PFV2 (yellow and cyan ORFs) and those of the other two viruses is not recognizable by sequence similarity searches. Figure reproduced with permission 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.

A deeper characterization of tristromaviruses and globulovirues needs to be carried out in the future to increase our knowledge on these mysterious members of the archaeal virosphere. Structural studies on globuloviruses are expected to provide insights into the virion architecture and might reveal unanticipated relationships with other spherical viruses. It is noteworthy that the unexplored genomes of these two families might be a source of genes encoding new functions or involved in novel mechanisms of interaction with their hosts.

Biogeographical pattern and extensive host switching in the Rudiviridae

Three viruses (MRV1, ARV3 and SSRV1) with rod-shaped filamentous virions were isolated from enrichment cultures growing in acidic conditions (pH 3.5) at 75°C and classified into the

Rudiviridae family. Although their genomes display high degree of gene synteny and share 26 putative proteins, the host-range experiments showed that the three viruses infect acidophilic hyperthermophiles belonging to three different genera of the order Sulfolobales, namely, *Saccharolobus*, *Acidianus* and *Metallosphaera*. This discovery would not have been possible without isolation of the pure virus strains.

Phylogenomic analysis of all available rudivirus genomes displayed that all sequenced rudiviruses fall into six clades corresponding to the geographical origins of the virus isolation, suggesting local adaptation of the viruses. This finding contrasts the broadly applicable paradigm that "everything is everywhere", with most viruses moving between different biomes (Breitbart and Rohwer, 2005). Despite forming a monophyletic group, all rudiviruses originating from Italy infect relatively distant hosts, implying possible host switching events. Our hypothesis was further supported by the presence of spacers against the genomes of the five described Italian rudiviruses within the CRISPR arrays of different strains of the order Sulfolobales. Notably, this phenomenon appears to be also applicable to rudiviruses from other geographical locations (**chapter 1**). The three rudiviruses infecting *Saccharolobus*, *Acidianus* and *Metallosphaera* isolated in the course of my PhD thesis will serve as valuable models to study the host range evolution in archaeal viruses.

Although rudiviruses represent one of the most extensively studied families of archaeal viruses with many of the viral proteins being characterized, certain unusual features found in the Italian rudiviruses ARV3 and MRV1 deserve further investigation in the future. The absence of a homologue of the SIRV2 P98, a pyramid-forming protein conserved in other rudiviruses, suggests that MRV1 and ARV3 might have evolved a different strategy of egress. Similarly, homologs of known Acr proteins are also missing from MRV1 and ARV3 genomes. Given that CRISPR-Cas systems are highly prevalent in crenarchaea, the lack of identifiable anti-CRISPR genes in their genomes indicate that these two viruses could encode novel Acr proteins (Athukoralage et al., 2020; Bhoobalan-Chitty et al., 2019; He et al., 2018) (**chapter 1**).

7.2 Cryo-EM structures of the filamentous viruses SSRV1 and SIFV

Structural studies on archaeal viruses help to uncover the molecular adaptations to life under extreme environmental conditions and provide insights into the evolution of archaeal viruses. Unlike other rudiviruses, but similar to lipothrixviruses and tristromaviruses, members of the

proposed genus "*Hoswirudivirus*," carry adjacent genes encoding two homologous major capsid proteins of about 108 and 134-aa long (**chapter 1**), suggesting that viruses of this genus might represent a missing evolutionary link between the families *Rudiviridae* and *Lipothrixviridae*. Our mass spectrometry and structural studies have shown that only one of the two MCPs encoded by SSRV1 is incorporated into the virion (**chapter 2**). This finding, combined with comparative genomics analysis, is most compatible with the evolution of rudiviruses from a lipothrixvirus-like ancestor by shedding of the lipid envelope. This study has provided valuable insights into the long-standing question on the evolutionary relationships among filamentous archaeal viruses. A future study should investigate whether the inactivated MCP of SSRV1 was exapted for a different function during the evolution of this virus lineage.

Determination of the near-atomic structures of lipothrixvirus SIFV and rudivirus SSRV1 enabled a more detailed comparison with other archaeal filamentous viruses to be performed. The finding that DNA genomes in all filamentous viruses are stored as A-form DNA strongly suggests that A-DNA is the preferred form of DNA in viruses from extreme geothermal environments. In addition, the fact that A-form DNA is found in hyperthermophilic viruses infecting both neutrophilic (Thermoproteales) and acidophilic (Sulfolobales) hosts suggests that A-form DNA is a general adaptation of crenarchaeal viruses to extreme temperatures rather than low pH (Wang et al., 2020b). Furthermore, our study provided insights into structural changes underlying the differences in mechanical properties of the viral particles. In particular, comparison of the persistence lengths of the archaeal viruses showed that nucleocapsid diameter and helical rise are the best predictors for virion rigidity, where the smaller diameter and larger rise are associated with a more flexible virion (**chapter 2**). Understanding the exact molecular bases underlying the different mechanical properties of archaeal virus particles might help to harness them as biomaterials for developing new nanotechnological and biomedical applications.

Our results provided near-atomic understanding on how the filamentous virions are constructed and suggested that virion assembly and genome packaging in filamentous hyperthermophilic viruses occur in a coordinated fashion. However, the cellular context of the filamentous virion assembly and in particular, the envelopment of the filamentous nucleocapsids remained unclear. I set out to gain insights into these processes in **chapter 3** using electron tomography.

7.3 SIFV as model to study virus-host interactions in the family *Lipothrixviridae*

The understanding on virus-host interactions in archaea remains highly fragmented. To gain insights into different aspects of the viral life cycle in archaea, I focused on the *Lipothrixviridae* family and selected the Sulfolobus islandicus filamentous virus (SIFV) as a model for these studies.

SIFV is lytic

Previous studies had suggested that members of *Lipothrixviridae* do not cause cell lysis based on the absence of cell debris and the increase of the optical density of infected cultures (Arnold et al., 2000b; Bettstetter et al., 2003; Häring et al., 2005b; Vestergaard et al., 2008). However, we showed for the first time that SIFV affects the cellular growth of its host and, more importantly, employs a lysis mechanism that relies on the formation of pyramidal structures in the membrane of the infected cells (**chapter 3**).

The observation of VAPs in members of the non-related families *Rudiviridae*, *Turriviridae*, *Ovaliviridae* and *Lipothrixviridae* indicates that this strategy is widespread among archaeal viruses infecting *Sulfolobus*, which otherwise share little similarity in virion morphology and gene content (Bize et al., 2009b; Brumfield et al., 2009; Wang et al., 2018a). This observation further reinforces the notion that virion morphogenesis and release mechanisms are uncoupled and evolve independently of each other. Importantly, SIFV protein responsible for the VAP formation, gp43, is not related to the two VAP proteins characterized for rudiviruses and turriviruses, suggesting that the diversity of VAP proteins is much greater that currently realized. The fact that such small proteins (<100 aa) can assemble into large pyramidal structures which can be formed in membranes of organisms from different domains of life is remarkable. Future studies should uncover the intermolecular interactions underlying the VAP formation and reveal the signal and mechanism of VAP opening.

SIFV is enveloped intracellularly

Unexpectedly, our results show that enveloped SIFV virions assemble in the cytoplasm and are released from the cell through pyramidal openings in the cell envelope. Our results suggest that assembly of virions proceeds through formation of naked nucleocapsids and their subsequent envelopment (c**hapter 3**). We hypothesize that envelopment of SIFV nucleocapsids occurs by

a completely novel mechanism, involving either de novo membrane formation or trafficking of lipids from the cytoplasmic membrane to the virion assembly centers – neither mechanism has been demonstrated for other viruses. Further studies using a combination of cryo-EM and cryo-electron tomography approaches are expected to help to understand how the lipid membrane is acquired inside the host cell. Given that proteins homologous to eukaryotic ESCRT components are conserved in several members of the Sulfolobales (Lindås et al., 2008), the infection of *Sulfolobus islandicus* strains in which different ESCRT components have been knocked down would provide insights into the role, if any, of this membrane-remodeling machinery in the release of archaeal filamentous viruses.

The results obtained during my PhD thesis project shed light on the diversity of crenarchaeal viruses and uncovered molecular mechanisms employed by archaeal viruses to interact with their hosts. Further exploration of the archaeal virus diversity and development of new virus-host model systems, coupled with advances in molecular biology, genetics and microscopy tools will continue uncovering the mysteries of the archaeal virosphere.

8.REFERENCES

Adam, P.S., Borrel, G., Brochier-Armanet, C., and Gribaldo, S. (2017). The growing tree of Archaea: new perspectives on their diversity, evolution and ecology. The ISME journal *11*, 2407-2425.

Adriaenssens, E.M., Sullivan, M.B., Knezevic, P., van Zyl, L.J., Sarkar, B.L., Dutilh, B.E., Alfenas-Zerbini, P., Łobocka, M., Tong, Y., Brister, J.R., *et al.* (2020). Taxonomy of prokaryotic viruses: 2018-2019 update from the ICTV Bacterial and Archaeal Viruses Subcommittee. Archives of virology *165*, 1253-1260.

Ahlgren, N.A., Fuchsman, C.A., Rocap, G., and Fuhrman, J.A. (2019). Discovery of several novel, widespread, and ecologically distinct marine Thaumarchaeota viruses that encode amoC nitrification genes. The ISME journal *13*, 618-631.

Ahn, D.-G., Kim, S.-I., Rhee, J.-K., Pyo Kim, K., Pan, J.-G., and Oh, J.-W. (2006). TTSV1, a new virus-like particle isolated from the hyperthermophilic crenarchaeote Thermoproteus tenax. Virology *351*, 280-290.

Ai, C., McCarthy, S., Eckrich, V., Rudrappa, D., Qiu, G., and Blum, P. (2016). Increased acid resistance of the archaeon, Metallosphaera sedula by adaptive laboratory evolution. Journal of industrial microbiology & biotechnology *43*, 1455-1465.

Aiewsakun, P., Adriaenssens, E.M., Lavigne, R., Kropinski, A.M., and Simmonds, P. (2018). Evaluation of the genomic diversity of viruses infecting bacteria, archaea and eukaryotes using a common bioinformatic platform: steps towards a unified taxonomy. The Journal of general virology *99*, 1331-1343.

Akıl, C., and Robinson, R.C. (2018). Genomes of Asgard archaea encode profilins that regulate actin. Nature *562*, 439-443.

Albers, S.-V., and Jarrell, K.F. (2018). The Archaellum: An Update on the Unique Archaeal Motility Structure. Trends in Microbiology *26*, 351-362.

Albers, S.V., and Meyer, B.H. (2011). The archaeal cell envelope. Nature reviews Microbiology *9*, 414-426.

Allers, T., and Mevarech, M. (2005). Archaeal genetics — the third way. Nature Reviews Genetics *6*, 58-73.

Andrei, A.-Ş., Banciu, H.L., and Oren, A. (2012). Living with salt: metabolic and phylogenetic diversity of archaea inhabiting saline ecosystems. FEMS Microbiology Letters *330*, 1-9.

Arnold, H.P., She, Q., Phan, H., Stedman, K., Prangishvili, D., Holz, I., Kristjansson, J.K., Garrett, R., and Zillig, W. (1999). The genetic element pSSVx of the extremely thermophilic crenarchaeon Sulfolobus is a hybrid between a plasmid and a virus. Mol Microbiol *34*, 217-226.

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.

Arnold, H.P., Zillig, W., Ziese, U., Holz, I., Crosby, M., Utterback, T., Weidmann, J.F., Kristjanson, J.K., Klenk, H.P., Nelson, K.E., *et al.* (2000b). A Novel Lipothrixvirus, SIFV, of the Extremely Thermophilic Crenarchaeon Sulfolobus. Virology *267*, 252-266.

Atanasova, N.S., Bamford, D.H., and Oksanen, H.M. (2016). Virus-host interplay in high salt environments. Environmental Microbiology Reports *8*, 431-444.

Athukoralage, J.S., McMahon, S.A., Zhang, C., Grüschow, S., Graham, S., Krupovic, M., Whitaker, R.J., Gloster, T.M., and White, M.F. (2020). An anti-CRISPR viral ring nuclease subverts type III CRISPR immunity. Nature *577*, 572-575.

Azam, A.H., and Tanji, Y. (2019). Bacteriophage-host arm race: an update on the mechanism of phage resistance in bacteria and revenge of the phage with the perspective for phage therapy. Appl Microbiol Biotechnol *103*, 2121-2131.

Baker, B.J., De Anda, V., Seitz, K.W., Dombrowski, N., Santoro, A.E., and Lloyd, K.G. (2020). Diversity, ecology and evolution of Archaea. Nature Microbiology *5*, 887-900.

Baker, B.J., Lesniewski, R.A., and Dick, G.J. (2012). Genome-enabled transcriptomics reveals archaeal populations that drive nitrification in a deep-sea hydrothermal plume. The ISME journal *6*, 2269-2279.

Bamford, D.H., Pietilä, M.K., Roine, E., Atanasova, N.S., Dienstbier, A., Oksanen, H.M., and Ictv Report, C. (2017). ICTV Virus Taxonomy Profile: Pleolipoviridae. The Journal of general virology *98*, 2916-2917.

Bang, C., and Schmitz, R.A. (2015). Archaea associated with human surfaces: not to be underestimated. FEMS Microbiology Reviews *39*, 631-648.

Bang, C., Weidenbach, K., Gutsmann, T., Heine, H., and Schmitz, R.A. (2014). The Intestinal Archaea Methanosphaera stadtmanae and Methanobrevibacter smithii Activate Human Dendritic Cells. PLOS ONE *9*, e99411.

Barns, S.M., Delwiche, C.F., Palmer, J.D., and Pace, N.R. (1996). Perspectives on archaeal diversity, thermophily and monophyly from environmental rRNA sequences. Proceedings of the National Academy of Sciences of the United States of America *93*, 9188-9193.

Barns, S.M., Fundyga, R.E., Jeffries, M.W., and Pace, N.R. (1994). Remarkable archaeal diversity detected in a Yellowstone National Park hot spring environment. Proceedings of the National Academy of Sciences of the United States of America *91*, 1609-1613.

Bates, S.T., Berg-Lyons, D., Caporaso, J.G., Walters, W.A., Knight, R., and Fierer, N. (2011). Examining the global distribution of dominant archaeal populations in soil. The ISME journal *5*, 908-917.

Bath, C., Cukalac, T., Porter, K., and Dyall-Smith, M.L. (2006). His1 and His2 are distantly related, spindle-shaped haloviruses belonging to the novel virus group, Salterprovirus. Virology *350*, 228-239.

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.

Beam, J.P., Jay, Z.J., Schmid, M.C., Rusch, D.B., Romine, M.F., M Jennings, R.d., Kozubal, M.A., Tringe, S.G., Wagner, M., and Inskeep, W.P. (2016). Ecophysiology of an uncultivated lineage of Aigarchaeota from an oxic, hot spring filamentous 'streamer' community. The ISME journal *10*, 210-224.

Belay, N., Mukhopadhyay, B., Conway de Macario, E., Galask, R., and Daniels, L. (1990). Methanogenic bacteria in human vaginal samples. Journal of clinical microbiology *28*, 1666-1668.

Bell, S.D., and Jackson, S.P. (1998). Transcription and translation in Archaea: a mosaic of eukaryal and bacterial features. Trends Microbiol *6*, 222-228.

Bernander, R. (1998). Archaea and the cell cycle. Molecular Microbiology *29*, 955-961.

Bernhardt, T.G., Struck, D.K., and Young, R. (2001). The lysis protein E of phi X174 is a specific inhibitor of the MraY-catalyzed step in peptidoglycan synthesis. The Journal of biological chemistry *276*, 6093-6097.

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.

Beveridge, T.J., and Graham, L.L. (1991). Surface layers of bacteria. Microbiological reviews *55*, 684-705.

Beveridge, T.J., Patel, G.B., Harris, B.J., and Sprott, G.D. (1986). The ultrastructure of Methanothrix concilii, a mesophilic aceticlastic methanogen. Canadian Journal of Microbiology *32*, 703-710.

Beveridge, T.J., Stewart, M., Doyle, R.J., and Sprott, G.D. (1985). Unusual stability of the Methanospirillum hungatei sheath. Journal of bacteriology *162*, 728-737.

Bewley, M.C., Springer, K., Zhang, Y.-B., Freimuth, P., and Flanagan, J.M. (1999). Structural Analysis of the Mechanism of Adenovirus Binding to Its Human Cellular Receptor, CAR. Science *286*, 1579-1583.

Bhardwaj, A., Molineux, I.J., Casjens, S.R., and Cingolani, G. (2011). Atomic structure of bacteriophage Sf6 tail needle knob. The Journal of biological chemistry *286*, 30867-30877.

Bhoobalan-Chitty, Y., Johansen, T.B., Di Cianni, N., and Peng, X. (2019). Inhibition of Type III CRISPR-Cas Immunity by an Archaeal Virus-Encoded Anti-CRISPR Protein. Cell *179*, 448-458.e411.

Biddle, J.F., Lipp, J.S., Lever, M.A., Lloyd, K.G., Sørensen, K.B., Anderson, R., Fredricks, H.F., Elvert, M., Kelly, T.J., Schrag, D.P., *et al.* (2006). Heterotrophic Archaea dominate sedimentary subsurface ecosystems off Peru. Proceedings of the National Academy of Sciences of the United States of America *103*, 3846-3851.

Birkenbihl, R.P., Neef, K., Prangishvili, D., and Kemper, B. (2001). Holliday junction resolving enzymes of archaeal viruses SIRV1 and SIRV21 1Edited by J. Karn. Journal of molecular biology *309*, 1067-1076.

Bize, A., Karlsson, E.A., Ekefjärd, K., Quax, T.E., Pina, M., Prevost, M.C., Forterre, P., Tenaillon, O., Bernander, R., and Prangishvili, D. (2009a). A unique virus release mechanism in the Archaea. Proceedings of the National Academy of Sciences of the United States of America *106*, 11306-11311.

Bize, A., Karlsson, E.A., Ekefjärd, K., Quax, T.E.F., Pina, M., Prevost, M.-C., Forterre, P., Tenaillon, O., Bernander, R., and Prangishvili, D. (2009b). A unique virus release mechanism in the Archaea. Proceedings of the National Academy of Sciences *106*, 11306-11311.

Bize, A., Peng, X., Prokofeva, M., Maclellan, K., Lucas, S., Forterre, P., Garrett, R.A., Bonch-Osmolovskaya, E.A., and Prangishvili, D. (2008). Viruses in acidic geothermal environments of the Kamchatka Peninsula. Research in microbiology *159*, 358-366.

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.

Bolduc, B., Shaughnessy, D.P., Wolf, Y.I., Koonin, E.V., Roberto, F.F., and Young, M. (2012). Identification of Novel Positive-Strand RNA Viruses by Metagenomic Analysis of Archaea-Dominated Yellowstone Hot Springs. Journal of Virology *86*, 5562-5573.

Bolduc, B., Wirth, J.F., Mazurie, A., and Young, M.J. (2015). Viral assemblage composition in Yellowstone acidic hot springs assessed by network analysis. The ISME journal *9*, 2162-2177.

Borrel, G., Brugère, J.-F., Gribaldo, S., Schmitz, R.A., and Moissl-Eichinger, C. (2020). The host-associated archaeome. Nature Reviews Microbiology.

Breitbart, M., and Rohwer, F. (2005). Here a virus, there a virus, everywhere the same virus? Trends Microbiol *13*, 278-284.

Brochier-Armanet, C., Boussau, B., Gribaldo, S., and Forterre, P. (2008). Mesophilic crenarchaeota: proposal for a third archaeal phylum, the Thaumarchaeota. Nature Reviews Microbiology *6*, 245-252.

Brock, T.D., Brock, K.M., Belly, R.T., and Weiss, R.L. (1972). Sulfolobus: a new genus of sulfur-oxidizing bacteria living at low pH and high temperature. Archiv fur Mikrobiologie *84*, 54-68.

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. Journal of Virology *83*, 5964-5970.

Cacciapuoti, G., Fuccio, F., Petraccone, L., Del Vecchio, P., and Porcelli, M. (2012). Role of disulfide bonds in conformational stability and folding of 5'-deoxy-5'-methylthioadenosine phosphorylase II from the hyperthermophilic archaeon Sulfolobus solfataricus. Biochimica et biophysica acta *1824*, 1136-1143.

Caspi, Y., and Dekker, C. (2018). Dividing the Archaeal Way: The Ancient Cdv Cell-Division Machinery. Frontiers in Microbiology *9*.

Castelle, C.J., and Banfield, J.F. (2018). Major New Microbial Groups Expand Diversity and Alter our Understanding of the Tree of Life. Cell *172*, 1181-1197.

Catalão, M.J., Gil, F., Moniz-Pereira, J., São-José, C., and Pimentel, M. (2013). Diversity in bacterial lysis systems: bacteriophages show the way. FEMS Microbiology Reviews *37*, 554-571.

Chen, B., Chen, Z., Wang, Y., Gong, H., Sima, L., Wang, J., Ouyang, S., Gan, W., Krupovic, M., Chen, X., *et al.* (2020). ORF4 of the Temperate Archaeal Virus SNJ1 Governs the Lysis-Lysogeny Switch and Superinfection Immunity. J Virol *94*.

Chimileski, S., Franklin, M.J., and Papke, R.T. (2014). Biofilms formed by the archaeon Haloferax volcaniiexhibit cellular differentiation and social motility, and facilitate horizontal gene transfer. BMC Biology *12*, 65.

Chow, C.E., Winget, D.M., White, R.A., 3rd, Hallam, S.J., and Suttle, C.A. (2015). Combining genomic sequencing methods to explore viral diversity and reveal potential virus-host interactions. Front Microbiol *6*, 265.

Christian, J.H., and Waltho, J.A. (1962). Solute concentrations within cells of halophilic and non-halophilic bacteria. Biochimica et biophysica acta *65*, 506-508.

Claus, H., and König, H. (2010). Cell Envelopes of Methanogens. In Prokaryotic Cell Wall Compounds: Structure and Biochemistry, H. König, H. Claus, and A. Varma, eds. (Berlin, Heidelberg: Springer Berlin Heidelberg), pp. 231-251.

Couradeau, E., Benzerara, K., Moreira, D., Gérard, E., Kaźmierczak, J., Tavera, R., and López-García, P. (2011). Prokaryotic and Eukaryotic Community Structure in Field and Cultured Microbialites from the Alkaline Lake Alchichica (Mexico). PLOS ONE *6*, e28767.

Cowie, R.O.M., Maas, E.W., and Ryan, K.G. (2011). Archaeal diversity revealed in Antarctic sea ice. Antarctic Science *23*, 531-536.

Da Cunha, V., Gaia, M., Gadelle, D., Nasir, A., and Forterre, P. (2017). Lokiarchaea are close relatives of Euryarchaeota, not bridging the gap between prokaryotes and eukaryotes. PLoS genetics *13*, e1006810.

Da Cunha, V., Gaia, M., Nasir, A., and Forterre, P. (2018). Asgard archaea do not close the debate about the universal tree of life topology. PLoS genetics *14*, e1007215.

Dalmasso, C., Oger, P., Selva, G., Courtine, D., L'Haridon, S., Garlaschelli, A., Roussel, E., Miyazaki, J., Reveillaud, J., Jebbar, M., *et al.* (2016). Thermococcus piezophilus sp. nov., a novel hyperthermophilic and piezophilic archaeon with a broad pressure range for growth, isolated from a deepest hydrothermal vent at the Mid-Cayman Rise. Syst Appl Microbiol *39*, 440-444.

Danovaro, R., Dell'Anno, A., Corinaldesi, C., Rastelli, E., Cavicchioli, R., Krupovic, M., Noble, R.T., Nunoura, T., and Prangishvili, D. (2016). Virus-mediated archaeal hecatomb in the deep seafloor. Science Advances *2*, e1600492.

Daum, B., and Gold, V. (2018). Twitch or swim: towards the understanding of prokaryotic motion based on the type IV pilus blueprint. Biological chemistry *399*, 799-808.

Daum, B., Quax, T.E.F., Sachse, M., Mills, D.J., Reimann, J., Yildiz, Ö., Häder, S., Saveanu, C., Forterre, P., Albers, S.-V., *et al.* (2014). Self-assembly of the general membrane-remodeling protein PVAP into sevenfold virus-associated pyramids. Proceedings of the National Academy of Sciences *111*, 3829-3834.

Davison, S., Couture-Tosi, E., Candela, T., Mock, M., and Fouet, A. (2005). Identification of the Bacillus anthracis (gamma) phage receptor. Journal of bacteriology *187*, 6742-6749.

De Rosa, M., Gambacorta, A., and Gliozzi, A. (1986). Structure, biosynthesis, and physicochemical properties of archaebacterial lipids. Microbiological reviews *50*, 70-80.

DeLong, E.F., and Pace, N.R. (2001). Environmental diversity of bacteria and archaea. Systematic biology *50*, 470-478.

Demina, T.A., Atanasova, N.S., Pietilä, M.K., Oksanen, H.M., and Bamford, D.H. (2016). Vesicle-like virion of Haloarcula hispanica pleomorphic virus 3 preserves high infectivity in saturated salt. Virology *499*, 40-51.

Demina, T.A., and Oksanen, H.M. (2020). Pleomorphic archaeal viruses: the family Pleolipoviridae is expanding by seven new species. Archives of virology *165*, 2723-2731.

Demina, T.A., Pietilä, M.K., Svirskaitė, J., Ravantti, J.J., Atanasova, N.S., Bamford, D.H., and Oksanen, H.M. (2017). HCIV-1 and Other Tailless Icosahedral Internal Membrane-Containing Viruses of the Family Sphaerolipoviridae. Viruses *9*.

Deng, Y., Xu, G., and Sui, L. (2015). Isolation and characterization of halophilic bacteria and archaea from salt ponds in Hangu Saltworks, Tianjin, China. Chinese Journal of Oceanology and Limnology *33*, 862-868.

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 *348*, 914-917.

Ding, J., Zhang, Y., Wang, H., Jian, H., Leng, H., and Xiao, X. (2017). Microbial Community Structure of Deep-sea Hydrothermal Vents on the Ultraslow Spreading Southwest Indian Ridge. Frontiers in Microbiology *8*.

Dombrowski, N., Lee, J.-H., Williams, T.A., Offre, P., and Spang, A. (2019). Genomic diversity, lifestyles and evolutionary origins of DPANN archaea. FEMS Microbiology Letters *366*.

Dombrowski, N., Teske, A.P., and Baker, B.J. (2018). Expansive microbial metabolic versatility and biodiversity in dynamic Guaymas Basin hydrothermal sediments. Nature Communications *9*, 4999.

Dridi, B., Raoult, D., and Drancourt, M. (2011). Archaea as emerging organisms in complex human microbiomes. Anaerobe *17*, 56-63.

Duckworth, A.W., Grant, W.D., Jones, B.E., and van Steenbergen, R. (1996). Phylogenetic diversity of soda lake alkaliphiles. FEMS Microbiology Ecology *19*, 181-191.

Edwards, K.J., Bond, P.L., Gihring, T.M., and Banfield, J.F. (2000). An archaeal iron-oxidizing extreme acidophile important in acid mine drainage. Science *287*, 1796-1799.

Eichler, J. (2003). Facing extremes: archaeal surface-layer (glyco)proteins. Microbiology (Reading, England) *149*, 3347-3351.

El Omari, K., Li, S., Kotecha, A., Walter, T.S., Bignon, E.A., Harlos, K., Somerharju, P., De Haas, F., Clare, D.K., Molin, M., *et al.* (2019). The structure of a prokaryotic viral envelope protein expands the landscape of membrane fusion proteins. Nature Communications *10*, 846.

Elkins, J.G., Podar, M., Graham, D.E., Makarova, K.S., Wolf, Y., Randau, L., Hedlund, B.P., Brochier-Armanet, C., Kunin, V., Anderson, I., *et al.* (2008). A korarchaeal genome reveals insights into the evolution of the Archaea. Proceedings of the National Academy of Sciences *105*, 8102-8107.

Eme, L., Spang, A., Lombard, J., Stairs, C.W., and Ettema, T.J.G. (2017). Archaea and the origin of eukaryotes. Nature Reviews Microbiology *15*, 711-723.

Fox, G.E., Magrum, L.J., Balch, W.E., Wolfe, R.S., and Woese, C.R. (1977). Classification of methanogenic bacteria by 16S ribosomal RNA characterization. Proceedings of the National Academy of Sciences of the United States of America *74*, 4537-4541.

Franzmann, P.D., Liu, Y., Balkwill, D.L., Aldrich, H.C., Conway De Macario, E., and Boone, D.R. (1997). Methanogenium frigidum sp. nov., a Psychrophilic, H2-Using Methanogen from Ace Lake, Antarctica. International Journal of Systematic and Evolutionary Microbiology *47*, 1068-1072.

Franzmann, P.D., Springer, N., Ludwig, W., Conway De Macario, E., and Rohde, M. (1992). A Methanogenic Archaeon from Ace Lake, Antarctica: Methanococcoides burtonii sp. nov. Systematic and Applied Microbiology *15*, 573-581.

Fröls, S., Gordon, P.M.K., Panlilio, M.A., Schleper, C., and Sensen, C.W. (2007). Elucidating the transcription cycle of the UV-inducible hyperthermophilic archaeal virus SSV1 by DNA microarrays. Virology *365*, 48-59.

Fukuchi, S., and Nishikawa, K. (2001). Protein surface amino acid compositions distinctively differ between thermophilic and mesophilic bacteria. Journal of molecular biology *309*, 835-843.

Fusco, S., Aulitto, M., Iacobucci, I., Crocamo, G., Pucci, P., Bartolucci, S., Monti, M., and Contursi, P. (2020). The interaction between the F55 virus-encoded transcription regulator and the RadA host recombinase reveals a common strategy in Archaea and Bacteria to sense the UV-induced damage to the host DNA. Biochimica et biophysica acta Gene regulatory mechanisms *1863*, 194493.

Fusco, S., Liguori, R., Limauro, D., Bartolucci, S., She, Q., and Contursi, P. (2015a). Transcriptome analysis of Sulfolobus solfataricus infected with two related fuselloviruses reveals novel insights into the regulation of CRISPR-Cas system. Biochimie *118*, 322-332.

Fusco, S., She, Q., Bartolucci, S., and Contursi, P. (2013). T(lys), a newly identified Sulfolobus spindle-shaped virus 1 transcript expressed in the lysogenic state, encodes a DNA-binding protein interacting at the promoters of the early genes. J Virol *87*, 5926-5936.

Fusco, S., She, Q., Fiorentino, G., Bartolucci, S., and Contursi, P. (2015b). Unravelling the Role of the F55 Regulator in the Transition from Lysogeny to UV Induction of Sulfolobus Spindle-Shaped Virus 1. Journal of Virology *89*, 6453-6461.

Gaidelyte, A., Cvirkaite-Krupovic, V., Daugelavicius, R., Bamford, J.K., and Bamford, D.H. (2006). The entry mechanism of membrane-containing phage Bam35 infecting Bacillus thuringiensis. Journal of bacteriology *188*, 5925-5934.

Gardner, A.F., Bell, S.D., White, M.F., Prangishvili, D., and Krupovic, M. (2014). Protein-protein interactions leading to recruitment of the host DNA sliding clamp by the hyperthermophilic Sulfolobus islandicus rod-shaped virus 2. J Virol *88*, 7105-7108.

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 : life under extreme conditions *15*, 619-624.

Geslin, C., Gaillard, M., Flament, D., Rouault, K., Le Romancer, M., Prieur, D., and Erauso, G. (2007). Analysis of the First Genome of a Hyperthermophilic Marine Virus-Like Particle, PAV1, Isolated from Pyrococcus abyssi. Journal of bacteriology *189*, 4510-4519.

Geslin, C., Le Romancer, M., Erauso, G., Gaillard, M., Perrot, G., and Prieur, D. (2003). PAV1, the first virus-like particle isolated from a hyperthermophilic euryarchaeote, "Pyrococcus abyssi". Journal of bacteriology *185*, 3888-3894.

Gil-Carton, D., Jaakkola, S.T., Charro, D., Peralta, B., Castaño-Díez, D., Oksanen, H.M., Bamford, D.H., and Abrescia, N.G.A. (2015). Insight into the Assembly of Viruses with Vertical Single β-barrel Major Capsid Proteins. Structure (London, England : 1993) *23*, 1866-1877.

Goodman, D.A., and Stedman, K.M. (2018). Comparative genetic and genomic analysis of the novel fusellovirus Sulfolobus spindle-shaped virus 10. Virus evolution *4*.

Gorlas, A., Koonin, E.V., Bienvenu, N., Prieur, D., and Geslin, C. (2012). TPV1, the first virus isolated from the hyperthermophilic genus Thermococcus. Environmental microbiology *14*, 503-516.

Gribaldo, S., and Brochier-Armanet, C. (2006). The origin and evolution of Archaea: a state of the art. Philosophical transactions of the Royal Society of London Series B, Biological sciences *361*, 1007-1022.

Grogan, D.W. (1989). Phenotypic characterization of the archaebacterial genus Sulfolobus: comparison of five wild-type strains. Journal of bacteriology *171*, 6710-6719.

Gudbergsdóttir, S.R., Menzel, P., Krogh, A., Young, M., and Peng, X. (2016). Novel viral genomes identified from six metagenomes reveal wide distribution of archaeal viruses and high viral diversity in terrestrial hot springs. Environmental microbiology *18*, 863-874.

Guerrero-Ferreira, R.C., Viollier, P.H., Ely, B., Poindexter, J.S., Georgieva, M., Jensen, G.J., and Wright, E.R. (2011). Alternative mechanism for bacteriophage adsorption to the motile bacterium Caulobacter crescentus. Proceedings of the National Academy of Sciences *108*, 9963-9968.

Guillière, F., Danioux, C., Jaubert, C., Desnoues, N., Delepierre, M., Prangishvili, D., Sezonov, G., and Guijarro, J.I. (2013). Solution Structure of an Archaeal DNA Binding Protein with an Eukaryotic Zinc Finger Fold. PLOS ONE *8*, e52908.

Guillière, 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. The Journal of biological chemistry *284*, 22222-22237.

Guo, T., Han, W., and She, Q. (2019). Tolerance of Sulfolobus SMV1 virus to the immunity of I-A and III-B CRISPR-Cas systems in Sulfolobus islandicus. RNA biology *16*, 549-556.

Guo, Y., Kragelund, B.B., White, M.F., and Peng, X. (2015). Functional characterization of a conserved archaeal viral operon revealing single-stranded DNA binding, annealing and nuclease activities. Journal of molecular biology *427*, 2179-2191.

Gupta, R.S., and Shami, A. (2011). Molecular signatures for the Crenarchaeota and the Thaumarchaeota. Antonie van Leeuwenhoek *99*, 133-157.

Guy, L., and Ettema, T.J.G. (2011). The archaeal 'TACK' superphylum and the origin of eukaryotes. Trends in Microbiology *19*, 580-587.

Hackstein, J.H.P., and van Alen, T.A. (1996). FECAL METHANOGENS AND VERTEBRATE EVOLUTION. Evolution; international journal of organic evolution *50*, 559-572.

Hallam, S.J., Konstantinidis, K.T., Putnam, N., Schleper, C., Watanabe, Y., Sugahara, J., Preston, C., de la Torre, J., Richardson, P.M., and DeLong, E.F. (2006). Genomic analysis of the uncultivated marine crenarchaeote Cenarchaeum symbiosum. Proceedings of the National Academy of Sciences of the United States of America *103*, 18296-18301.

Hanhijärvi, K.J., Ziedaite, G., Pietilä, M.K., Hæggström, E., and Bamford, D.H. (2013). DNA ejection from an archaeal virus--a single-molecule approach. Biophysical journal *104*, 2264-2272.

Happonen, L.J., Redder, P., Peng, X., Reigstad, L.J., Prangishvili, D., and Butcher, S.J. (2010). Familial relationships in hyperthermo- and acidophilic archaeal viruses. J Virol *84*, 4747-4754.

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. J Virol *79*, 9904-9911.

Häring, M., Vestergaard, G., Brügger, K., Rachel, R., Garrett, R.A., and Prangishvili, D. (2005b). Structure and genome organization of AFV2, a novel archaeal lipothrixvirus with unusual terminal and core structures. Journal of bacteriology *187*, 3855-3858.

Häring, M., Vestergaard, G., Rachel, R., Chen, L., Garrett, R.A., and Prangishvili, D. (2005c). Virology: independent virus development outside a host. Nature *436*, 1101-1102.

Haroon, M.F., Hu, S., Shi, Y., Imelfort, M., Keller, J., Hugenholtz, P., Yuan, Z., and Tyson, G.W. (2013). Anaerobic oxidation of methane coupled to nitrate reduction in a novel archaeal lineage. Nature *500*, 567-570.

Hartman, R., Biewenga, L., Munson-McGee, J., Refai, M., Boyd, E.S., Bothner, B., Lawrence, C.M., and Young, M. (2020). Discovery and Characterization of Thermoproteus Spherical Piliferous Virus 1: a Spherical Archaeal Virus Decorated with Unusual Filaments. Journal of Virology *94*, e00036-00020.

Hartman, R., Eilers, B.J., Bollschweiler, D., Munson-McGee, J.H., Engelhardt, H., Young, M.J., and Lawrence, C.M. (2019). The Molecular Mechanism of Cellular Attachment for an Archaeal Virus. Structure (London, England : 1993) *27*, 1634-1646.e1633.

He, F., Bhoobalan-Chitty, Y., Van, L.B., Kjeldsen, A.L., Dedola, M., Makarova, K.S., Koonin, E.V., Brodersen, D.E., and Peng, X. (2018). Anti-CRISPR proteins encoded by archaeal lytic viruses inhibit subtype I-D immunity. Nature Microbiology *3*, 461-469.

He, F., Chen, L., and Peng, X. (2014). First experimental evidence for the presence of a CRISPR toxin in sulfolobus. Journal of molecular biology *426*, 3683-3688.

Hedlund, B.P., Murugapiran, S.K., Alba, T.W., Levy, A., Dodsworth, J.A., Goertz, G.B., Ivanova, N., and Woyke, T. (2015). Uncultivated thermophiles: current status and spotlight on 'Aigarchaeota'. Current Opinion in Microbiology *25*, 136-145.

Heinz, E., and Domman, D. (2017). Reshaping the tree of life. Nature Reviews Microbiology *15*, 322-322.

Henneman, B., van Emmerik, C., van Ingen, H., and Dame, R.T. (2018). Structure and function of archaeal histones. PLoS genetics *14*, e1007582.

Hobel, C.F., Albers, S.V., Driessen, A.J., and Lupas, A.N. (2008). The Sulfolobus solfataricus AAA protein Sso0909, a homologue of the eukaryotic ESCRT Vps4 ATPase. Biochemical Society transactions *36*, 94-98.

Hong, C., Pietilä, M.K., Fu, C.J., Schmid, M.F., Bamford, D.H., and Chiu, W. (2015). Lemon-shaped halo archaeal virus His1 with uniform tail but variable capsid structure. Proceedings of the National Academy of Sciences *112*, 2449-2454.

Horn, C., Paulmann, B., Kerlen, G., Junker, N., and Huber, H. (1999). In vivo observation of cell division of anaerobic hyperthermophiles by using a high-intensity dark-field microscope. Journal of bacteriology *181*, 5114-5118.

Hoshino, T., and Inagaki, F. (2019). Abundance and distribution of Archaea in the subseafloor sedimentary biosphere. The ISME journal *13*, 227-231.

Hua, Z.-S., Qu, Y.-N., Zhu, Q., Zhou, E.-M., Qi, Y.-L., Yin, Y.-R., Rao, Y.-Z., Tian, Y., Li, Y.-X., Liu, L., *et al.* (2018). Genomic inference of the metabolism and evolution of the archaeal phylum Aigarchaeota. Nature Communications *9*, 2832.

Huber, H., Hohn, M.J., Rachel, R., Fuchs, T., Wimmer, V.C., and Stetter, K.O. (2002). A new phylum of Archaea represented by a nanosized hyperthermophilic symbiont. Nature *417*, 63-67.

Huet, J., Schnabel, R., Sentenac, A., and Zillig, W. (1983). Archaebacteria and eukaryotes possess DNA-dependent RNA polymerases of a common type. The EMBO journal *2*, 1291-1294.

Hurley, J.H., and Hanson, P.I. (2010). Membrane budding and scission by the ESCRT machinery: it's all in the neck. Nature reviews Molecular cell biology *11*, 556-566.

Ilk, N., Egelseer, E.M., and Sleytr, U.B. (2011). S-layer fusion proteins—construction principles and applications. Current Opinion in Biotechnology *22*, 824-831.

Imachi, H., Nobu, M.K., Nakahara, N., Morono, Y., Ogawara, M., Takaki, Y., Takano, Y., Uematsu, K., Ikuta, T., Ito, M., *et al.* (2020). Isolation of an archaeon at the prokaryote–eukaryote interface. Nature *577*, 519-525.

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. J Virol *90*, 11043-11055.

Iranzo, J., Krupovic, M., and Koonin, E.V. (2016b). The Double-Stranded DNA Virosphere as a Modular Hierarchical Network of Gene Sharing. mBio *7*.

Iro, M., Klein, R., Gálos, B., Baranyi, U., Rössler, N., and Witte, A. (2007). The lysogenic region of virus phiCh1: identification of a repressor-operator system and determination of its activity in halophilic Archaea. Extremophiles : life under extreme conditions *11*, 383-396.

Jaatinen, S.T., Happonen, L.J., Laurinmäki, P., Butcher, S.J., and Bamford, D.H. (2008). Biochemical and structural characterisation of membrane-containing icosahedral dsDNA bacteriophages infecting thermophilic Thermus thermophilus. Virology *379*, 10-19.

Jacquemet, A., Barbeau, J., Lemiègre, L., and Benvegnu, T. (2009). Archaeal tetraether bipolar lipids: Structures, functions and applications. Biochimie *91*, 711-717.

Jain, S., Caforio, A., and Driessen, A.J. (2014). Biosynthesis of archaeal membrane ether lipids. Front Microbiol *5*, 641.

Jarrell, K.F., and Albers, S.V. (2012). The archaellum: an old motility structure with a new name. Trends Microbiol *20*, 307-312.

Jaubert, C., Danioux, C., Oberto, J., Cortez, D., Bize, A., Krupovic, M., She, Q., Forterre, P., Prangishvili, D., and Sezonov, G. (2013). Genomics and genetics of Sulfolobus islandicus LAL14/1, a model hyperthermophilic archaeon. Open Biol *3*, 130010.

Javor, B., Requadt, C., and Stoeckenius, W. (1982). Box-shaped halophilic bacteria. Journal of bacteriology *151*, 1532-1542.

Jones, K.L., Rees, J.F., and Grainger, J.M. (1983). Methane generation and microbial activity in a domestic refuse landfill site. European journal of applied microbiology and biotechnology *18*, 242-245.

Kambura, A.K., Mwirichia, R.K., Kasili, R.W., Karanja, E.N., Makonde, H.M., and Boga, H.I. (2016). Bacteria and Archaea diversity within the hot springs of Lake Magadi and Little Magadi in Kenya. BMC microbiology *16*, 136.

Kandler, O., and König, H. (1978). Chemical composition of the peptidoglycan-free cell walls of methanogenic bacteria. Archives of Microbiology *118*, 141-152.

Karner, M.B., DeLong, E.F., and Karl, D.M. (2001). Archaeal dominance in the mesopelagic zone of the Pacific Ocean. Nature *409*, 507-510.

Karshikoff, A., and Ladenstein, R. (2001). Ion pairs and the thermotolerance of proteins from hyperthermophiles: a 'traffic rule' for hot roads. Trends in Biochemical Sciences *26*, 550-557.

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*.

Kates, M. (1977). The phytanyl ether-linked polar lipids and isoprenoid neutral lipids of extremely halophilic bacteria. Progress in the Chemistry of Fats and other Lipids *15*, 301-342.

Kazlauskas, D., Krupovic, M., and Venclovas, Č. (2016). The logic of DNA replication in double-stranded DNA viruses: insights from global analysis of viral genomes. Nucleic acids research *44*, 4551-4564.

Ken, R., and Hackett, N.R. (1991). Halobacterium halobium strains lysogenic for phage phi H contain a protein resembling coliphage repressors. Journal of bacteriology *173*, 955-960.

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. Journal of bacteriology *186*, 7745-7753.

Kim, J.-G., Kim, S.-J., Cvirkaite-Krupovic, V., Yu, W.-J., Gwak, J.-H., López-Pérez, M., Rodriguez-Valera, F., Krupovic, M., Cho, J.-C., and Rhee, S.-K. (2019). Spindle-shaped viruses infect marine ammonia-oxidizing thaumarchaea. Proceedings of the National Academy of Sciences *116*, 15645-15650.

Klein, R., Rössler, N., Iro, M., Scholz, H., and Witte, A. (2012). Haloarchaeal myovirus φCh1 harbours a phase variation system for the production of protein variants with distinct cell surface adhesion specificities. Mol Microbiol *83*, 137-150.

Knittel, K., Lösekann, T., Boetius, A., Kort, R., and Amann, R. (2005). Diversity and distribution of methanotrophic archaea at cold seeps. Applied and environmental microbiology *71*, 467-479.

Koerdt, A., Gödeke, J., Berger, J., Thormann, K.M., and Albers, S.-V. (2010). Crenarchaeal Biofilm Formation under Extreme Conditions. PLOS ONE *5*, e14104.

Koga, Y., and Morii, H. (2007). Biosynthesis of ether-type polar lipids in archaea and evolutionary considerations. Microbiology and molecular biology reviews : MMBR *71*, 97-120.

Koga, Y., Nishihara, M., Morii, H., and Akagawa-Matsushita, M. (1993). Ether polar lipids of methanogenic bacteria: structures, comparative aspects, and biosyntheses. Microbiological reviews *57*, 164-182.

König, H., Hartmann, E., and Kärcher, U. (1993). Pathways and Principles of the Biosynthesis of Methanobacterial Cell Wall Polymers. Systematic and Applied Microbiology *16*, 510-517.

Könneke, M., Schubert, D.M., Brown, P.C., Hügler, M., Standfest, S., Schwander, T., Schada von Borzyskowski, L., Erb, T.J., Stahl, D.A., and Berg, I.A. (2014). Ammonia-oxidizing archaea use the most energy-efficient aerobic pathway for CO₂ fixation. Proceedings of the National Academy of Sciences *111*, 8239-8244.

Koonin, E.V., Dolja, V.V., Krupovic, M., Varsani, A., Wolf, Y.I., Yutin, N., Zerbini, F.M., and Kuhn, J.H. (2020). Global Organization and Proposed Megataxonomy of the Virus World. Microbiology and molecular biology reviews : MMBR *84*.

Koonin, E.V., Makarova, K.S., and Wolf, Y.I. (2017). Evolutionary Genomics of Defense Systems in Archaea and Bacteria. Annual review of microbiology *71*, 233-261.

Koskinen, K., Pausan, M.R., Perras, A.K., Beck, M., Bang, C., Mora, M., Schilhabel, A., Schmitz, R., and Moissl-Eichinger, C. (2017). First Insights into the Diverse Human Archaeome: Specific Detection of Archaea in the Gastrointestinal Tract, Lung, and Nose and on Skin. mBio *8*, e00824-00817.

Kostrikina, N.A., Zvyagintseva, I.S., and Duda, V.I. (1991). Cytological peculiarities of some extremely halophilic soil archaeobacteria. Archives of Microbiology *156*, 344-349.

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.

Krupovic, M., Dolja, V.V., and Koonin, E.V. (2020). The LUCA and its complex virome. Nature Reviews Microbiology *18*, 661-670.

Krupovic, M., Forterre, P., and Bamford, D.H. (2010). Comparative analysis of the mosaic genomes of tailed archaeal viruses and proviruses suggests common themes for virion architecture and assembly with tailed viruses of bacteria. Journal of molecular biology *397*, 144-160.

Krupovič, M., Gribaldo, S., Bamford, D.H., and Forterre, P. (2010). The Evolutionary History of Archaeal MCM Helicases: A Case Study of Vertical Evolution Combined with Hitchhiking of Mobile Genetic Elements. Molecular biology and evolution *27*, 2716-2732.

Krupovic, M., Makarova, K.S., Wolf, Y.I., Medvedeva, S., Prangishvili, D., Forterre, P., and Koonin, E.V. (2019). Integrated mobile genetic elements in Thaumarchaeota. Environmental microbiology *21*, 2056-2078.

Krupovic, M., Quemin, E.R.J., Bamford, D.H., Forterre, P., and Prangishvili, D. (2014). Unification of the Globally Distributed Spindle-Shaped Viruses of the Archaea. Journal of Virology *88*, 2354-2358.

Krupovic, M., Spang, A., Gribaldo, S., Forterre, P., and Schleper, C. (2011). A thaumarchaeal provirus testifies for an ancient association of tailed viruses with archaea. Biochemical Society transactions *39*, 82-88.

Kukkaro, P., and Bamford, D.H. (2009). Virus-host interactions in environments with a wide range of ionic strengths. Environ Microbiol Rep *1*, 71-77.

Lake, J.A. (2015). Eukaryotic origins. Philosophical transactions of the Royal Society of London Series B, Biological sciences *370*, 20140321.

Larsen, H. (1986). Halophilic and halotolerant microorganisms—an overview and historical perspective. FEMS Microbiology Reviews *2*, 3-7.

Leavitt, J.C., Gogokhia, L., Gilcrease, E.B., Bhardwaj, A., Cingolani, G., and Casjens, S.R. (2013). The Tip of the Tail Needle Affects the Rate of DNA Delivery by Bacteriophage P22. PLOS ONE *8*, e70936.

Leigh, J.A., Albers, S.-V., Atomi, H., and Allers, T. (2011). Model organisms for genetics in the domain Archaea: methanogens, halophiles, Thermococcales and Sulfolobales. FEMS Microbiology Reviews *35*, 577-608.

Leininger, S., Urich, T., Schloter, M., Schwark, L., Qi, J., Nicol, G.W., Prosser, J.I., Schuster, S.C., and Schleper, C. (2006). Archaea predominate among ammonia-oxidizing prokaryotes in soils. Nature *442*, 806-809.

León-Sobrino, C., Kot, W.P., and Garrett, R.A. (2016). Transcriptome changes in STSV2-infected Sulfolobus islandicus REY15A undergoing continuous CRISPR spacer acquisition. Molecular Microbiology *99*, 719-728.

Lindås, A.C., Karlsson, E.A., Lindgren, M.T., Ettema, T.J., and Bernander, R. (2008). A unique cell division machinery in the Archaea. Proceedings of the National Academy of Sciences of the United States of America *105*, 18942-18946.

Liu, Y., Brandt, D., Ishino, S., Ishino, Y., Koonin, E.V., Kalinowski, J., Krupovic, M., and Prangishvili, D. (2019). New archaeal viruses discovered by metagenomic analysis of viral communities in enrichment cultures. Environmental microbiology *21*, 2002-2014.

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*, e00589-00517.

Liu, Y., Makarova, K.S., Huang, W.-C., Wolf, Y.I., Nikolskaya, A., Zhang, X., Cai, M., Zhang, C.-J., Xu, W., Luo, Z., *et al.* (2020). Expanding diversity of Asgard archaea and the elusive ancestry of eukaryotes. bioRxiv, 2020.2010.2019.343400.

Liu, Y., Osinski, T., Wang, F., Krupovic, M., Schouten, S., Kasson, P., Prangishvili, D., and Egelman, E.H. (2018a). Structural conservation in a membrane-enveloped filamentous virus infecting a hyperthermophilic acidophile. Nature Communications *9*, 3360.

Liu, Y., Wang, J., Liu, Y., Wang, Y., Zhang, Z., Oksanen, H.M., Bamford, D.H., and Chen, X. (2015). Identification and characterization of SNJ2, the first temperate pleolipovirus integrating into the genome of the SNJ1-lysogenic archaeal strain. Mol Microbiol *98*, 1002-1020.

Liu, Y., Zhou, Z., Pan, J., Baker, B.J., Gu, J.-D., and Li, M. (2018b). Comparative genomic inference suggests mixotrophic lifestyle for Thorarchaeota. The ISME journal *12*, 1021-1031.

Luk, A.W., Williams, T.J., Erdmann, S., Papke, R.T., and Cavicchioli, R. (2014). Viruses of haloarchaea. Life (Basel, Switzerland) *4*, 681-715.

Luo, Y., Pfister, P., Leisinger, T., and Wasserfallen, A. (2001). The genome of archaeal prophage PsiM100 encodes the lytic enzyme responsible for autolysis of Methanothermobacter wolfeii. Journal of bacteriology *183*, 5788-5792.

Maaty, W.S., Ortmann, A.C., Dlakić, M., Schulstad, K., Hilmer, J.K., Liepold, L., Weidenheft, B., Khayat, R., Douglas, T., Young, M.J., *et al.* (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.

MacLeod, F., Kindler, G.S., Wong, H.L., Chen, R., and Burns, B.P. (2019). Asgard archaea: Diversity, function, and evolutionary implications in a range of microbiomes. AIMS microbiology *5*, 48-61.

Macnab, R.M. (1999). The Bacterial Flagellum: Reversible Rotary Propellor and Type III Export Apparatus. Journal of bacteriology *181*, 7149-7153.

Makarova, K.S., Koonin, E.V., and Albers, S.-V. (2016). Diversity and Evolution of Type IV pili Systems in Archaea. Frontiers in Microbiology *7*.

Makarova, K.S., Wolf, Y.I., Iranzo, J., Shmakov, S.A., Alkhnbashi, O.S., Brouns, S.J.J., Charpentier, E., Cheng, D., Haft, D.H., Horvath, P., *et al.* (2020). Evolutionary classification of CRISPR–Cas systems: a burst of class 2 and derived variants. Nature Reviews Microbiology *18*, 67-83.

Makarova, K.S., Wolf, Y.I., and Koonin, E.V. (2013). Comparative genomics of defense systems in archaea and bacteria. Nucleic acids research *41*, 4360-4377.

Makarova, K.S., Yutin, N., Bell, S.D., and Koonin, E.V. (2010). Evolution of diverse cell division and vesicle formation systems in Archaea. Nature reviews Microbiology *8*, 731-741.

Marti, R., Zurfluh, K., Hagens, S., Pianezzi, J., Klumpp, J., and Loessner, M.J. (2013). Long tail fibres of the novel broad-host-range T-even bacteriophage S16 specifically recognize Salmonella OmpC. Mol Microbiol *87*, 818-834.

Martínez-Alvarez, L., Bell, S.D., and Peng, X. (2016). Multiple consecutive initiation of replication producing novel brush-like intermediates at the termini of linear viral dsDNA genomes with hairpin ends. Nucleic acids research *44*, 8799-8809.

Martínez-Alvarez, L., Deng, L., and Peng, X. (2017). Formation of a Viral Replication Focus in Sulfolobus Cells Infected by the Rudivirus Sulfolobus islandicus Rod-Shaped Virus 2. Journal of Virology *91*, e00486-00417.

Mayer, F., and Müller, V. (2014). Adaptations of anaerobic archaea to life under extreme energy limitation. FEMS Microbiology Reviews *38*, 449-472.

McKay, L.J., Dlakić, M., Fields, M.W., Delmont, T.O., Eren, A.M., Jay, Z.J., Klingelsmith, K.B., Rusch, D.B., and Inskeep, W.P. (2019). Co-occurring genomic capacity for anaerobic methane and dissimilatory sulfur metabolisms discovered in the Korarchaeota. Nature Microbiology *4*, 614-622.

Medvedeva, S., Liu, Y., Koonin, E.V., Severinov, K., Prangishvili, D., and Krupovic, M. (2019). Virus-borne mini-CRISPR arrays are involved in interviral conflicts. Nature Communications *10*, 5204.

Mei, Y., He, C., Huang, Y., Liu, Y., Zhang, Z., Chen, X., and Shen, P. (2015). Salinity Regulation of the Interaction of Halovirus SNJ1 with Its Host and Alteration of the Halovirus Replication Strategy to Adapt to the Variable Ecosystem. PLOS ONE *10*, e0123874.

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.

Moissl, C., Rachel, R., Briegel, A., Engelhardt, H., and Huber, R. (2005). The unique structure of archaeal 'hami', highly complex cell appendages with nano-grappling hooks. Mol Microbiol *56*, 361-370.

Munson-McGee, J.H., Rooney, C., and Young, M.J. (2020). An Uncultivated Virus Infecting a Nanoarchaeal Parasite in the Hot Springs of Yellowstone National Park. Journal of Virology *94*, e01213-01219.

Muskhelishvili, G., Palm, P., and Zillig, W. (1993). SSV1-encoded site-specific recombination system in Sulfolobus shibatae. Molecular & general genetics : MGG *237*, 334-342.

Nishimura, Y., Watai, H., Honda, T., Mihara, T., Omae, K., Roux, S., Blanc-Mathieu, R., Yamamoto, K., Hingamp, P., Sako, Y., *et al.* (2017). Environmental Viral Genomes Shed New Light on Virus-Host Interactions in the Ocean. mSphere *2*.

Nunoura, T., Hirayama, H., Takami, H., Oida, H., Nishi, S., Shimamura, S., Suzuki, Y., Inagaki, F., Takai, K., Nealson, K.H., *et al.* (2005). Genetic and functional properties of uncultivated thermophilic crenarchaeotes from a subsurface gold mine as revealed by analysis of genome fragments. Environmental microbiology *7*, 1967-1984.

Nunoura, T., Takaki, Y., Kakuta, J., Nishi, S., Sugahara, J., Kazama, H., Chee, G.J., Hattori, M., Kanai, A., Atomi, H., *et al.* (2011). Insights into the evolution of Archaea and eukaryotic protein modifier systems revealed by the genome of a novel archaeal group. Nucleic acids research *39*, 3204-3223.

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. Journal of Virology *85*, 925-931.

Okutan, E., Deng, L., Mirlashari, S., Uldahl, K., Halim, M., Liu, C., Garrett, R.A., She, Q., and Peng, X. (2013). Novel insights into gene regulation of the rudivirus SIRV2 infecting Sulfolobus cells. RNA biology *10*, 875-885.

Oren, A. (1999). The Enigma of Square and Triangular Halophilic Archaea. In Enigmatic Microorganisms and Life in Extreme Environments, J. Seckbach, ed. (Dordrecht: Springer Netherlands), pp. 337-355.

Oren, A. (2002). Diversity of halophilic microorganisms: environments, phylogeny, physiology, and applications. Journal of industrial microbiology & biotechnology *28*, 56-63.

Orphan, V.J., House, C.H., Hinrichs, K.-U., McKeegan, K.D., and DeLong, E.F. (2002). Multiple archaeal groups mediate methane oxidation in anoxic cold seep sediments. Proceedings of the National Academy of Sciences *99*, 7663-7668.

Ortmann, A.C., Brumfield, S.K., Walther, J., McInnerney, K., Brouns, S.J.J., van de Werken, H.J.G., Bothner, B., Douglas, T., van de Oost, J., and Young, M.J. (2008). Transcriptome Analysis of Infection of the Archaeon Sulfolobus solfataricus with Sulfolobus Turreted Icosahedral Virus. Journal of Virology *82*, 4874-4883.

Pagaling, E., Grant, W.D., Cowan, D.A., Jones, B.E., Ma, Y., Ventosa, A., and Heaphy, S. (2012). Bacterial and archaeal diversity in two hot spring microbial mats from the geothermal region of Tengchong, China. Extremophiles : life under extreme conditions *16*, 607-618.

Palatinszky, M., Herbold, C., Jehmlich, N., Pogoda, M., Han, P., von Bergen, M., Lagkouvardos, I., Karst, S.M., Galushko, A., Koch, H., *et al.* (2015). Cyanate as an energy source for nitrifiers. Nature *524*, 105-108.

Palmieri, G., Balestrieri, M., Peter-Katalinić, J., Pohlentz, G., Rossi, M., Fiume, I., and Pocsfalvi, G. (2013). Surface-exposed Glycoproteins of Hyperthermophilic Sulfolobus solfataricus P2 Show a Common N-Glycosylation Profile. Journal of Proteome Research *12*, 2779-2790.

Pan, M., Kelman, L.M., and Kelman, Z. (2011). The archaeal PCNA proteins. Biochemical Society transactions *39*, 20-24.

Papathanasiou, P., Erdmann, S., Leon-Sobrino, C., Sharma, K., Urlaub, H., Garrett, R.A., and Peng, X. (2019). Stable maintenance of the rudivirus SIRV3 in a carrier state in Sulfolobus islandicus despite activation of the CRISPR-Cas immune response by a second virus SMV1. RNA biology *16*, 557-565.

Pawlowski, A., Rissanen, I., Bamford, J.K., Krupovic, M., and Jalasvuori, M. (2014). Gammasphaerolipovirus, a newly proposed bacteriophage genus, unifies viruses of halophilic archaea and thermophilic bacteria within the novel family Sphaerolipoviridae. Archives of virology *159*, 1541-1554.

Peixeiro, N., Keller, J., Collinet, B., Leulliot, N., Campanacci, V., Cortez, D., Cambillau, C., Nitta, K.R., Vincentelli, R., Forterre, P., *et al.* (2013). Structure and function of AvtR, a novel transcriptional regulator from a hyperthermophilic archaeal lipothrixvirus. J Virol *87*, 124-136.

Peng, X., Basta, T., Häring, M., Garrett, R.A., and Prangishvili, D. (2007). Genome of the Acidianus bottle-shaped virus and insights into the replication and packaging mechanisms. Virology *364*, 237-243.

Peng, X., Blum, H., She, Q., Mallok, S., Brügger, 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.

Pester, M., Schleper, C., and Wagner, M. (2011). The Thaumarchaeota: an emerging view of their phylogeny and ecophysiology. Current Opinion in Microbiology *14*, 300-306.

Pfister, P., Wasserfallen, A., Stettler, R., and Leisinger, T. (1998a). Molecular analysis of Methanobacterium phage psiM2. Mol Microbiol *30*, 233-244.

Pfister, P., Wasserfallen, A., Stettler, R., and Leisinger, T. (1998b). Molecular analysis of Methanobacterium phage ΨM2. Molecular Microbiology *30*, 233-244.

Philosof, A., Yutin, N., Flores-Uribe, J., Sharon, I., Koonin, E.V., and Béjà, O. (2017). Novel Abundant Oceanic Viruses of Uncultured Marine Group II Euryarchaeota. Current biology : CB *27*, 1362-1368.

Pietilä, M.K., Demina, T.A., Atanasova, N.S., Oksanen, H.M., and Bamford, D.H. (2014). Archaeal viruses and bacteriophages: comparisons and contrasts. Trends Microbiol *22*, 334-344.

Pietilä, M.K., Laurinmäki, P., Russell, D.A., Ko, C.C., Jacobs-Sera, D., Butcher, S.J., Bamford, D.H., and Hendrix, R.W. (2013a). Insights into head-tailed viruses infecting extremely halophilic archaea. J Virol *87*, 3248-3260.

Pietilä, M.K., Laurinmäki, P., Russell, D.A., Ko, C.C., Jacobs-Sera, D., Hendrix, R.W., Bamford, D.H., and Butcher, S.J. (2013b). Structure of the archaeal head-tailed virus HSTV-1 completes the HK97 fold story. Proceedings of the National Academy of Sciences of the United States of America *110*, 10604-10609.

Pietilä, M.K., Roine, E., Sencilo, A., Bamford, D.H., and Oksanen, H.M. (2016). Pleolipoviridae, a newly proposed family comprising archaeal pleomorphic viruses with single-stranded or double-stranded DNA genomes. Archives of virology *161*, 249-256.

Pina, M., Basta, T., Quax, T.E., Joubert, A., Baconnais, S., Cortez, D., Lambert, S., Le Cam, E., Bell, S.D., Forterre, P., *et al.* (2014). Unique genome replication mechanism of the archaeal virus AFV1. Mol Microbiol *92*, 1313-1325.

Pohlschroder, M., and Esquivel, R.N. (2015). Archaeal type IV pili and their involvement in biofilm formation. Front Microbiol *6*, 190.

Poranen, M.M., Daugelavicius, R., and Bamford, D.H. (2002). Common principles in viral entry. Annual review of microbiology *56*, 521-538.

Porter, K., Tang, S.L., Chen, C.P., Chiang, P.W., Hong, M.J., and Dyall-Smith, M. (2013). PH1: an archaeovirus of Haloarcula hispanica related to SH1 and HHIV-2. Archaea *2013*, 456318.

Prangishvili, D., Arnold, H.P., Götz, 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.

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., Garrett, R.A., and Koonin, E.V. (2006a). Evolutionary genomics of archaeal viruses: unique viral genomes in the third domain of life. Virus research *117*, 52-67.

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. The Journal of biological chemistry *273*, 6024-6029.

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., Krupovic, M., and Ictv Report, C. (2018a). ICTV Virus Taxonomy Profile: Ampullaviridae. The Journal of general virology *99*, 288-289.

Prangishvili, D., Krupovic, M., and Ictv Report, C. (2018b). ICTV Virus Taxonomy Profile: Bicaudaviridae. The Journal of general virology *99*, 864-865.

Prangishvili, D., Mochizuki, T., Krupovic, M., and Ictv Report, C. (2018c). ICTV Virus Taxonomy Profile: Guttaviridae. The Journal of general virology *99*, 290-291.

Prangishvili, D., Mochizuki, T., Krupovic, M., and Ictv Report, C. (2020). ICTV Virus Taxonomy Profile: Spiraviridae. The Journal of general virology *101*, 240-241.

Prangishvili, D., Mochizuki, T., Liu, Y., Krupovic, M., and Ictv Report, C. (2019a). ICTV Virus Taxonomy Profile: Clavaviridae. The Journal of general virology *100*, 1267-1268.

Prangishvili, D., Rensen, E., Mochizuki, T., Krupovic, M., and Ictv Report, C. (2019b). ICTV Virus Taxonomy Profile: Tristromaviridae. The Journal of general virology *100*, 135-136.

Prangishvili, D., Vestergaard, G., Häring, M., Aramayo, R., Basta, T., Rachel, R., and Garrett, R.A. (2006b). 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.

Probst, A.J., Auerbach, A.K., and Moissl-Eichinger, C. (2013). Archaea on Human Skin. PLOS ONE *8*, e65388.

Probst, A.J., Ladd, B., Jarett, J.K., Geller-McGrath, D.E., Sieber, C.M.K., Emerson, J.B., Anantharaman, K., Thomas, B.C., Malmstrom, R.R., Stieglmeier, M., *et al.* (2018). Differential depth distribution of microbial function and putative symbionts through sediment-hosted aquifers in the deep terrestrial subsurface. Nat Microbiol *3*, 328-336.

Ptchelkine, D., Gillum, A., Mochizuki, T., Lucas-Staat, S., Liu, Y., Krupovic, M., Phillips, S.E.V., Prangishvili, D., and Huiskonen, J.T. (2017). Unique architecture of thermophilic archaeal virus APBV1 and its genome packaging. Nature Communications *8*, 1436.

Pushkarev, A., Inoue, K., Larom, S., Flores-Uribe, J., Singh, M., Konno, M., Tomida, S., Ito, S., Nakamura, R., Tsunoda, S.P., *et al.* (2018). A distinct abundant group of microbial rhodopsins discovered using functional metagenomics. Nature *558*, 595-599.

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. Proceedings of the National Academy of Sciences of the United States of America *108*, 3354-3359.

Quax, T.E., Voet, M., Sismeiro, O., Dillies, M.A., Jagla, B., Coppée, J.Y., Sezonov, G., Forterre, P., van der Oost, J., Lavigne, R., *et al.* (2013). Massive activation of archaeal defense genes during viral infection. J Virol *87*, 8419-8428.

Quax, T.E.F., Krupovič, 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., Lucas, S., Daum, B., Quax, T.E., Kühlbrandt, W., Forterre, P., Albers, S.V., Prangishvili, D., and Krupovic, M. (2013). First insights into the entry process of hyperthermophilic archaeal viruses. J Virol *87*, 13379-13385.

Quemin, E.R., Pietilä, M.K., Oksanen, H.M., Forterre, P., Rijpstra, W.I., Schouten, S., Bamford, D.H., Prangishvili, D., and Krupovic, M. (2015). Sulfolobus Spindle-Shaped Virus 1 Contains Glycosylated Capsid Proteins, a Cellular Chromatin Protein, and Host-Derived Lipids. J Virol *89*, 11681-11691.

Quemin, E.R.J., Chlanda, P., Sachse, M., Forterre, P., Prangishvili, D., and Krupovic, M. (2016). Eukaryotic-Like Virus Budding in Archaea. mBio *7*, e01439-01416.

Raghoebarsing, A.A., Pol, A., van de Pas-Schoonen, K.T., Smolders, A.J.P., Ettwig, K.F., Rijpstra, W.I.C., Schouten, S., Damsté, J.S.S., Op den Camp, H.J.M., Jetten, M.S.M., *et al.* (2006). A microbial consortium couples anaerobic methane oxidation to denitrification. Nature *440*, 918-921.

Raymann, K., Moeller, A.H., Goodman, A.L., and Ochman, H. (2017). Unexplored Archaeal Diversity in the Great Ape Gut Microbiome. mSphere *2*.

Redder, P., Peng, X., Brügger, K., Shah, S.A., Roesch, F., Greve, B., She, Q., Schleper, C., Forterre, P., Garrett, R.A., *et al.* (2009). Four newly isolated fuselloviruses from extreme geothermal environments reveal unusual morphologies and a possible interviral recombination mechanism. Environmental microbiology *11*, 2849-2862.

Reed, C.J., Lewis, H., Trejo, E., Winston, V., and Evilia, C. (2013). Protein Adaptations in Archaeal Extremophiles. Archaea *2013*, 373275.

Ren, Y., She, Q., and Huang, L. (2013). Transcriptomic analysis of the SSV2 infection of Sulfolobus solfataricus with and without the integrative plasmid pSSVi. Virology *441*, 126-134.

Rensen, E., Krupovic, M., and Prangishvili, D. (2015). Mysterious hexagonal pyramids on the surface of Pyrobaculum cells. Biochimie *118*, 365-367.

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.

Reysenbach, A.-L., Liu, Y., Banta, A.B., Beveridge, T.J., Kirshtein, J.D., Schouten, S., Tivey, M.K., Von Damm, K.L., and Voytek, M.A. (2006). A ubiquitous thermoacidophilic archaeon from deep-sea hydrothermal vents. Nature *442*, 444-447.

Rice, G., Tang, L., Stedman, K., Roberto, F., Spuhler, J., Gillitzer, E., Johnson, J.E., Douglas, T., and Young, M. (2004). The structure of a thermophilic archaeal virus shows a double-stranded DNA viral capsid type that spans all domains of life. Proceedings of the National Academy of Sciences of the United States of America *101*, 7716-7720.

Rinke, C., Schwientek, P., Sczyrba, A., Ivanova, N.N., Anderson, I.J., Cheng, J.-F., Darling, A., Malfatti, S., Swan, B.K., Gies, E.A., *et al.* (2013). Insights into the phylogeny and coding potential of microbial dark matter. Nature *499*, 431-437.

Rissanen, I., Grimes, J.M., Pawlowski, A., Mäntynen, S., Harlos, K., Bamford, J.K., and Stuart, D.I. (2013). Bacteriophage P23-77 capsid protein structures reveal the archetype of an ancient branch from a major virus lineage. Structure (London, England : 1993) *21*, 718-726.

Rodrigues-Oliveira, T., Belmok, A., Vasconcellos, D., Schuster, B., and Kyaw, C.M. (2017). Archaeal S-Layers: Overview and Current State of the Art. Frontiers in Microbiology *8*.

Roelvink, P.W., Mi Lee, G., Einfeld, D.A., Kovesdi, I., and Wickham, T.J. (1999). Identification of a conserved receptor-binding site on the fiber proteins of CAR-recognizing adenoviridae. Science *286*, 1568-1571.

Roine, E., Kukkaro, P., Paulin, L., Laurinavicius, S., Domanska, A., Somerharju, P., and Bamford, D.H. (2010). New, closely related haloarchaeal viral elements with different nucleic Acid types. J Virol *84*, 3682-3689.

Romantschuk, M., and Bamford, D.H. (1985). Function of pili in bacteriophage phi 6 penetration. The Journal of general virology *66 (Pt 11)*, 2461-2469.

Roux, S., Brum, J.R., Dutilh, B.E., Sunagawa, S., Duhaime, M.B., Loy, A., Poulos, B.T., Solonenko, N., Lara, E., Poulain, J., *et al.* (2016). Ecogenomics and potential biogeochemical impacts of globally abundant ocean viruses. Nature *537*, 689-693.

Rowland, E.F., Bautista, M.A., Zhang, C., and Whitaker, R.J. (2020). Surface resistance to SSVs and SIRVs in pilin deletions of Sulfolobus islandicus. Molecular Microbiology *113*, 718-727.

Rudolph, C., Wanner, G., and Huber, R. (2001). Natural Communities of Novel Archaea and Bacteria Growing in Cold Sulfurous Springs with a String-of-Pearls-Like Morphology. Applied and environmental microbiology *67*, 2336-2344.

Samson, R.Y., and Bell, S.D. (2011). Cell cycles and cell division in the archaea. Curr Opin Microbiol *14*, 350-356.

Samson, R.Y., Obita, T., Freund, S.M., Williams, R.L., and Bell, S.D. (2008). A role for the ESCRT system in cell division in archaea. Science *322*, 1710-1713.

Santoro, A.E., Casciotti, K.L., and Francis, C.A. (2010). Activity, abundance and diversity of nitrifying archaea and bacteria in the central California Current. Environmental microbiology *12*, 1989-2006.

Sára, M., and Sleytr, U.B. (2000). S-Layer proteins. Journal of bacteriology *182*, 859-868.

Satoh, T., Watanabe, K., Yamamoto, H., Yamamoto, S., and Kurosawa, N. (2013). Archaeal Community Structures in the Solfataric Acidic Hot Springs with Different Temperatures and Elemental Compositions. Archaea *2013*, 723871.

Schleper, C., Kubo, K., and Zillig, W. (1992). The particle SSV1 from the extremely thermophilic archaeon Sulfolobus is a virus: demonstration of infectivity and of transfection with viral DNA. Proceedings of the National Academy of Sciences *89*, 7645-7649.

Schleper, C., Puehler, G., Holz, I., Gambacorta, A., Janekovic, D., Santarius, U., Klenk, H.P., and Zillig, W. (1995). Picrophilus gen. nov., fam. nov.: a novel aerobic, heterotrophic, thermoacidophilic genus and family comprising archaea capable of growth around pH 0. Journal of bacteriology *177*, 7050-7059.

Schofield, L.R., Beattie, A.K., Tootill, C.M., Dey, D., and Ronimus, R.S. (2015). Biochemical Characterisation of Phage Pseudomurein Endoisopeptidases PeiW and PeiP Using Synthetic Peptides. Archaea *2015*, 828693.

Seitz, K.W., Lazar, C.S., Hinrichs, K.-U., Teske, A.P., and Baker, B.J. (2016). Genomic reconstruction of a novel, deeply branched sediment archaeal phylum with pathways for acetogenesis and sulfur reduction. The ISME journal *10*, 1696-1705.

Sencilo, A., and Roine, E. (2014). A Glimpse of the genomic diversity of haloarchaeal tailed viruses. Frontiers in Microbiology *5*.

Serre, M.C., Letzelter, C., Garel, J.R., and Duguet, M. (2002). Cleavage properties of an archaeal site-specific recombinase, the SSV1 integrase. The Journal of biological chemistry *277*, 16758-16767.

Sheppard, C., Blombach, F., Belsom, A., Schulz, S., Daviter, T., Smollett, K., Mahieu, E., Erdmann, S., Tinnefeld, P., Garrett, R., *et al.* (2016). Repression of RNA polymerase by the archaeo-viral regulator ORF145/RIP. Nature Communications *7*, 13595.

Sheppard, C., and Werner, F. (2017). Structure and mechanisms of viral transcription factors in archaea. Extremophiles : life under extreme conditions *21*, 829-838.

Shimada, H., Nemoto, N., Shida, Y., Oshima, T., and Yamagishi, A. (2002). Complete polar lipid composition of Thermoplasma acidophilum HO-62 determined by high-performance liquid chromatography with evaporative light-scattering detection. Journal of bacteriology *184*, 556-563.

Shin, H., Lee, J.-H., Kim, H., Choi, Y., Heu, S., and Ryu, S. (2012). Receptor Diversity and Host Interaction of Bacteriophages Infecting Salmonella enterica Serovar Typhimurium. PLOS ONE *7*, e43392.

Siliakus, M.F., van der Oost, J., and Kengen, S.W.M. (2017). Adaptations of archaeal and bacterial membranes to variations in temperature, pH and pressure. Extremophiles : life under extreme conditions *21*, 651-670.

Sleytr, U.B. (1976). Self-assembly of the hexagonally and tetragonally arranged subunits of bacterial surface layers and their reattachment to cell walls. Journal of Ultrastructure Research *55*, 360-377.

Sleytr, U.B., Messner, P., Pum, D., and Sára, M. (1999). Crystalline Bacterial Cell Surface Layers (S Layers): From Supramolecular Cell Structure to Biomimetics and Nanotechnology. Angewandte Chemie International Edition *38*, 1034-1054.

Smith, P.H., and Hungate, R.E. (1958). Isolation and characterization of Methanobacterium ruminantium n. sp. Journal of bacteriology *75*, 713-718.

Snyder, J.C., Bateson, M.M., Lavin, M., and Young, M.J. (2010). Use of Cellular CRISPR (Clusters of Regularly Interspaced Short Palindromic Repeats) Spacer-Based Microarrays for Detection of Viruses in Environmental Samples. Applied and environmental microbiology *76*, 7251-7258.

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.

Sorokin, D.Y., Berben, T., Melton, E.D., Overmars, L., Vavourakis, C.D., and Muyzer, G. (2014). Microbial diversity and biogeochemical cycling in soda lakes. Extremophiles : life under extreme conditions *18*, 791-809.

Spang, A., Caceres, E.F., and Ettema, T.J.G. (2017). Genomic exploration of the diversity, ecology, and evolution of the archaeal domain of life. Science *357*, eaaf3883.

Spang, A., Saw, J.H., Jørgensen, S.L., Zaremba-Niedzwiedzka, K., Martijn, J., Lind, A.E., van Eijk, R., Schleper, C., Guy, L., and Ettema, T.J.G. (2015). Complex archaea that bridge the gap between prokaryotes and eukaryotes. Nature *521*, 173-179.

Stedman, K.M. (2011). Fusellovirus. In The Springer Index of Viruses, C. Tidona, and G. Darai, eds. (New York, NY: Springer New York), pp. 561-566.

Stedman, K.M., DeYoung, M., Saha, M., Sherman, M.B., and Morais, M.C. (2015). Structural insights into the architecture of the hyperthermophilic Fusellovirus SSV1. Virology *474*, 105-109.

Stetter, K.O. (1996). Hyperthermophilic procaryotes. FEMS Microbiology Reviews *18*, 149-158.

Stetter, K.O., König, H., and Stackebrandt, E. (1983). Pyrodictium gen. nov., a New Genus of Submarine Disc-Shaped Sulphur Reducing Archaebacteria Growing Optimally at 105°C. Systematic and Applied Microbiology *4*, 535-551.

Straub, C.T., Counts, J.A., Nguyen, D.M.N., Wu, C.-H., Zeldes, B.M., Crosby, J.R., Conway, J.M., Otten, J.K., Lipscomb, G.L., Schut, G.J., *et al.* (2018). Biotechnology of extremely thermophilic archaea. FEMS Microbiology Reviews *42*, 543-578.

Streif, S., Staudinger, W.F., Marwan, W., and Oesterhelt, D. (2008). Flagellar Rotation in the Archaeon Halobacterium salinarum Depends on ATP. Journal of molecular biology *384*, 1-8.

Sun, Y., Liu, Y., Pan, J., Wang, F., and Li, M. (2020). Perspectives on Cultivation Strategies of Archaea. Microbial Ecology *79*, 770-784.

Takai, K., and Horikoshi, K. (1999). Genetic diversity of archaea in deep-sea hydrothermal vent environments. Genetics *152*, 1285-1297.

Takai, K., Moser, D.P., DeFlaun, M., Onstott, T.C., and Fredrickson, J.K. (2001). Archaeal diversity in waters from deep South African gold mines. Applied and environmental microbiology *67*, 5750-5760.

Takai, K., Nakamura, K., Toki, T., Tsunogai, U., Miyazaki, M., Miyazaki, J., Hirayama, H., Nakagawa, S., Nunoura, T., and Horikoshi, K. (2008). Cell proliferation at 122 degrees C and isotopically heavy CH4 production by a hyperthermophilic methanogen under high-pressure cultivation. Proceedings of the National Academy of Sciences of the United States of America *105*, 10949-10954.

Tazi, L., Breakwell, D.P., Harker, A.R., and Crandall, K.A. (2014). Life in extreme environments: microbial diversity in Great Salt Lake, Utah. Extremophiles : life under extreme conditions *18*, 525-535.

Thauer, R.K., Kaster, A.-K., Seedorf, H., Buckel, W., and Hedderich, R. (2008). Methanogenic archaea: ecologically relevant differences in energy conservation. Nature Reviews Microbiology *6*, 579-591.

Thiroux, S., Dupont, S., Nesbø, C.L., Bienvenu, N., Krupovic, M., L'Haridon, S., Marie, D., Forterre, P., Godfroy, A., and Geslin, C. (2020). The first head-tailed virus, MFTV1, infecting hyperthermophilic methanogenic deep-sea archaea. Environmental microbiology.

Tomazic, S.J., and Klibanov, A.M. (1988). Mechanisms of irreversible thermal inactivation of Bacillus alpha-amylases. The Journal of biological chemistry *263*, 3086-3091.

Torsvik, T., and Dundas, I.D. (1974). Bacteriophage of Halobacterium salinarium. Nature *248*, 680-681.

Trachtenberg, S., and Cohen-Krausz, S. (2006). The archaeabacterial flagellar filament: a bacterial propeller with a pilus-like structure. Journal of molecular microbiology and biotechnology *11*, 208-220.

Tully, B.J., Graham, E.D., and Heidelberg, J.F. (2018). The reconstruction of 2,631 draft metagenome-assembled genomes from the global oceans. Scientific Data *5*, 170203.

Uldahl, K.B., Jensen, S.B., Bhoobalan-Chitty, Y., Martínez-Álvarez, L., Papathanasiou, P., and Peng, X. (2016). Life Cycle Characterization of Sulfolobus Monocaudavirus 1, an

Extremophilic Spindle-Shaped Virus with Extracellular Tail Development. Journal of Virology *90*, 5693-5699.

Valentine, D.L. (2002). Biogeochemistry and microbial ecology of methane oxidation in anoxic environments: a review. Antonie Van Leeuwenhoek *81*, 271-282.

van der Gulik, P.T.S., Hoff, W.D., and Speijer, D. (2017). In defence of the three-domains of life paradigm. BMC Evolutionary Biology *17*, 218.

van Wolferen, M., Orell, A., and Albers, S.-V. (2018). Archaeal biofilm formation. Nature Reviews Microbiology *16*, 699-713.

Veesler, D., Ng, T.-S., Sendamarai, A.K., Eilers, B.J., Lawrence, C.M., Lok, S.-M., Young, M.J., Johnson, J.E., and Fu, C.-y. (2013). Atomic structure of the 75 MDa extremophile Sulfolobus turreted icosahedral virus determined by CryoEM and X-ray crystallography. Proceedings of the National Academy of Sciences *110*, 5504-5509.

Veith, A., Klingl, A., Zolghadr, B., Lauber, K., Mentele, R., Lottspeich, F., Rachel, R., Albers, S.-V., and Kletzin, A. (2009). Acidianus, Sulfolobus and Metallosphaera surface layers: structure, composition and gene expression. Molecular Microbiology *73*, 58-72.

Vestergaard, G., Aramayo, R., Basta, T., Häring, M., Peng, X., Brügger, K., Chen, L., Rachel, R., Boisset, N., Garrett, R.A., *et al.* (2008). Structure of the acidianus filamentous virus 3 and comparative genomics of related archaeal lipothrixviruses. J Virol *82*, 371-381.

Vieille, C., and Zeikus, G.J. (2001). Hyperthermophilic Enzymes: Sources, Uses, and Molecular Mechanisms for Thermostability. Microbiology and Molecular Biology Reviews *65*, 1-43.

Vik, D.R., Roux, S., Brum, J.R., Bolduc, B., Emerson, J.B., Padilla, C.C., Stewart, F.J., and Sullivan, M.B. (2017). Putative archaeal viruses from the mesopelagic ocean. PeerJ *5*, e3428.

Villanueva, L., Damsté, J.S.S., and Schouten, S. (2014). A re-evaluation of the archaeal membrane lipid biosynthetic pathway. Nature Reviews Microbiology *12*, 438-448.

Villanueva, L., Schouten, S., and Damsté, J.S. (2017). Phylogenomic analysis of lipid biosynthetic genes of Archaea shed light on the 'lipid divide'. Environmental microbiology *19*, 54-69.

Visweswaran, G.R., Dijkstra, B.W., and Kok, J. (2011). Murein and pseudomurein cell wall binding domains of bacteria and archaea--a comparative view. Appl Microbiol Biotechnol *92*, 921-928.

Wagner, C., Reddy, V., Asturias, F., Khoshouei, M., Johnson, J.E., Manrique, P., Munson-McGee, J., Baumeister, W., Lawrence, C.M., and Young, M.J. (2017). Isolation and Characterization of Metallosphaera Turreted Icosahedral Virus, a Founding Member of a New Family of Archaeal Viruses. J Virol *91*.

Wang, F., Baquero, D.P., Beltran, L.C., Su, Z., Osinski, T., Zheng, W., Prangishvili, D., Krupovic, M., and Egelman, E.H. (2020a). Structures of filamentous viruses infecting hyperthermophilic archaea explain DNA stabilization in extreme environments. Proceedings of the National Academy of Sciences *117*, 19643-19652.

Wang, F., Baquero, D.P., Su, Z., Beltran, L.C., Prangishvili, D., Krupovic, M., and Egelman, E.H. (2020b). The structures of two archaeal type IV pili illuminate evolutionary relationships. Nature Communications *11*, 3424.

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

Wang, F., Cvirkaite-Krupovic, V., Kreutzberger, M.A.B., Su, Z., de Oliveira, G.A.P., Osinski, T., Sherman, N., DiMaio, F., Wall, J.S., Prangishvili, D., *et al.* (2019a). An extensively glycosylated archaeal pilus survives extreme conditions. Nat Microbiol *4*, 1401-1410.

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. (2019b). A packing for A-form DNA in an icosahedral virus. Proceedings of the National Academy of Sciences *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.* (2018a). Novel Sulfolobus Virus with an Exceptional Capsid Architecture. J Virol *92*.

Wang, J., Liu, Y., Liu, Y., Du, K., Xu, S., Wang, Y., Krupovic, M., and Chen, X. (2018b). A novel family of tyrosine integrases encoded by the temperate pleolipovirus SNJ2. Nucleic acids research *46*, 2521-2536.

Wang, Y., Chen, B., Cao, M., Sima, L., Prangishvili, D., Chen, X., and Krupovic, M. (2018c). Rolling-circle replication initiation protein of haloarchaeal sphaerolipovirus SNJ1 is homologous to bacterial transposases of the IS91 family insertion sequences. Journal of General Virology *99*, 416-421.

Wang, Y., Sima, L., Lv, J., Huang, S., Liu, Y., Wang, J., Krupovic, M., and Chen, X. (2016). Identification, Characterization, and Application of the Replicon Region of the Halophilic Temperate Sphaerolipovirus SNJ1. Journal of bacteriology *198*, 1952-1964.

Wang, Y., Wegener, G., Hou, J., Wang, F., and Xiao, X. (2019c). Expanding anaerobic alkane metabolism in the domain of Archaea. Nature Microbiology *4*, 595-602.

Waters, E., Hohn, M.J., Ahel, I., Graham, D.E., Adams, M.D., Barnstead, M., Beeson, K.Y., Bibbs, L., Bolanos, R., Keller, M., *et al.* (2003). The genome of Nanoarchaeum equitans: Insights into early archaeal evolution and derived parasitism. Proceedings of the National Academy of Sciences *100*, 12984-12988.

Webster, N.S., and Negri, A.P. (2006). Site-specific variation in Antarctic marine biofilms established on artificial surfaces. Environmental microbiology *8*, 1177-1190.

Wegener, G., Krukenberg, V., Riedel, D., Tegetmeyer, H.E., and Boetius, A. (2015). Intercellular wiring enables electron transfer between methanotrophic archaea and bacteria. Nature *526*, 587-590.

Weidenbach, K., Nickel, L., Neve, H., Alkhnbashi, O.S., Künzel, S., Kupczok, A., Bauersachs, T., Cassidy, L., Tholey, A., Backofen, R., *et al.* (2017). Methanosarcina Spherical Virus, a Novel Archaeal Lytic Virus Targeting Methanosarcina Strains. Journal of Virology *91*, e00955-00917.

Wheaton, G., Counts, J., Mukherjee, A., Kruh, J., and Kelly, R. (2015). The Confluence of Heavy Metal Biooxidation and Heavy Metal Resistance: Implications for Bioleaching by Extreme Thermoacidophiles. Minerals *5*, 397-451.

Wiedenheft, B., Stedman, K., Roberto, F., Willits, D., Gleske, A.K., Zoeller, L., Snyder, J., Douglas, T., and Young, M. (2004). Comparative genomic analysis of hyperthermophilic archaeal Fuselloviridae viruses. J Virol *78*, 1954-1961.

Woese, C.R., and Fox, G.E. (1977). Phylogenetic structure of the prokaryotic domain: The primary kingdoms. Proceedings of the National Academy of Sciences *74*, 5088-5090.

Woese, C.R., Kandler, O., and Wheelis, M.L. (1990). Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. Proceedings of the National Academy of Sciences of the United States of America *87*, 4576-4579.

Wolf, S., Fischer, M.A., Kupczok, A., Reetz, J., Kern, T., Schmitz, R.A., and Rother, M. (2019). Characterization of the lytic archaeal virus Drs3 infecting Methanobacterium formicicum. Archives of virology *164*, 667-674.

Wong, H.L., White, R.A., Visscher, P.T., Charlesworth, J.C., Vázquez-Campos, X., and Burns, B.P. (2018). Disentangling the drivers of functional complexity at the metagenomic level in Shark Bay microbial mat microbiomes. The ISME journal *12*, 2619-2639.

Xia, G., Corrigan, R.M., Winstel, V., Goerke, C., Gründling, A., and Peschel, A. (2011). Wall Teichoic Acid-Dependent Adsorption of Staphylococcal Siphovirus and Myovirus. Journal of bacteriology *193*, 4006-4009.

Yamaguchi, Y., Park, J.H., and Inouye, M. (2011). Toxin-antitoxin systems in bacteria and archaea. Annual review of genetics *45*, 61-79.

Yang, C.-C., Chen, Y.-H., Tsai, H.-H., Huang, C.-H., Huang, T.-W., and Chen, C.W. (2006). In Vitro Deoxynucleotidylation of the Terminal Protein of Streptomyces Linear Chromosomes. Applied and environmental microbiology *72*, 7959-7961.

Yang, C.C., Huang, C.H., Li, C.Y., Tsay, Y.G., Lee, S.C., and Chen, C.W. (2002). The terminal proteins of linear Streptomyces chromosomes and plasmids: a novel class of replication priming proteins. Mol Microbiol *43*, 297-305.

Young, R. (2013). Phage lysis: do we have the hole story yet? Curr Opin Microbiol *16*, 790-797.

Yutin, N., Makarova, K.S., Mekhedov, S.L., Wolf, Y.I., and Koonin, E.V. (2008). The deep archaeal roots of eukaryotes. Molecular biology and evolution *25*, 1619-1630.

Zaremba-Niedzwiedzka, K., Caceres, E.F., Saw, J.H., Bäckström, D., Juzokaite, L., Vancaester, E., Seitz, K.W., Anantharaman, K., Starnawski, P., Kjeldsen, K.U., *et al.* (2017). Asgard archaea illuminate the origin of eukaryotic cellular complexity. Nature *541*, 353-358.

Zeikus, J.G., and Bowen, V.G. (1975). Fine structure of Methanospirillum hungatii. Journal of bacteriology *121*, 373-380.

Zhan, Z., Zhou, J., and Huang, L. (2015). Site-Specific Recombination by SSV2 Integrase: Substrate Requirement and Domain Functions. J Virol *89*, 10934-10944.

Zhang, C., Phillips, A.P.R., Wipfler, R.L., Olsen, G.J., and Whitaker, R.J. (2018a). The essential genome of the crenarchaeal model Sulfolobus islandicus. Nature Communications *9*, 4908.

Zhang, C., Wipfler, R., Li, Y., Wang, Z., Hallett, E., and Whitaker, R. (2018b). Revealing S-layer Functions in the Hyperthermophilic Crenarchaeon Sulfolobus islandicus (bioRxiv).

Zhang, J., Zheng, X., Wang, H., Jiang, H., Dong, H., and Huang, L. (2020). Novel Sulfolobus Fuselloviruses with Extensive Genomic Variations. Journal of Virology *94*, e01624-01619.

Zhang, L., Kang, M., Xu, J., Xu, J., Shuai, Y., Zhou, X., Yang, Z., and Ma, K. (2016). Bacterial and archaeal communities in the deep-sea sediments of inactive hydrothermal vents in the Southwest India Ridge. Scientific Reports *6*, 25982.

Zhang, Z., Liu, Y., Wang, S., Yang, D., Cheng, Y., Hu, J., Chen, J., Mei, Y., Shen, P., Bamford, D.H., *et al.* (2012). Temperate membrane-containing halophilic archaeal virus SNJ1 has a circular dsDNA genome identical to that of plasmid pHH205. Virology *434*, 233-241.

Zhou, Z., Liu, Y., Li, M., and Gu, J.-D. (2018). Two or three domains: a new view of tree of life in the genomics era. Applied Microbiology and Biotechnology *102*, 3049-3058.

Zink, I.A., Pfeifer, K., Wimmer, E., Sleytr, U.B., Schuster, B., and Schleper, C. (2019). CRISPR-mediated gene silencing reveals involvement of the archaeal S-layer in cell division and virus infection. Nature Communications *10*, 4797.

9. AKNOWLEDGEMENTS

I would like to acknowledge all members of the jury for their time and consideration. I thank Dr. Pascale Boulanger and Dr. Claire Geslin for kindly agreeing to review my thesis, and Professor Guennadi Senozov for being the president of the jury. Furthermore, I would like to thank the members of my thesis advisory committee, Dr. Jacomine Krijnse-Locker, Dr. Paulo Tavares and Professor Guennadi Senozov, for all their valuable comments which enriched my PhD project.

I would like to express my deepest gratitude to my PhD directors, Mart and David, for all their guidance, direction and supervision during these three exciting years. Thanks for your patience, confidence, and most of all, for encouraging me and giving me confidence in my own work. I feel very lucky to have counted with your experience and support during all the stages of my PhD. David, thanks for sharing your knowledge and passion for archaeal viruses, I learned to love them throughout these years. Mart, many thanks for all what you have taught me about science, I enjoyed very much our discussions, always spiced with your characteristic sense of humour.

Big thanks to all past members of the former BMGE and current ARVIR unit. I acknowledge Patrick Forterre for giving me the opportunity to perform my PhD in the former BMGE unit.
I am grateful to Ana for all her patience and help from the very beginning of this adventure. Special thanks go to Virginija and Ying for having welcomed me to the lab with such warmth and sparing their time teaching me all the tricks; more importantly, many thanks for their friendship and support. Thanks to Pierre for his scientific advices and interesting and enjoyable stories during lunch time. Thanks to Karyna and Sonia for the nice moments we shared. I would also like to thank Soizick for her help at the beginning of my PhD and for taking care of the former unit.

I would like to thank members of the *Ultrapole*, in particular Anastasia Gazi, Martin Sasche, Christine Schmitt and Gerard Pehaut-Arnaudet, for all their support with the electron microscopy approaches. Thanks to all our collaborators for such wonderful works: Edward Egelman and Fengbin Wang from University of Virginia, Patrizia Contursi and Simonetta

Bartolucci from University of Naples and Monica Piocchi from the Istituto Nazionale di Geofisica e Vulcanologia; it has been a real pleasure to work with all of you.

Thanks to the PPU program for supporting me and giving me one of the most challenging and exciting opportunities I have ever had. Special thanks to Susanna Celli and the PPU steering committee; your kindness and dedication made us feel welcomed and supported during this journey.

Thanks to all my new friends, Rosaria, Pravali, Adrian, Cesar, Mery, Daniela, Camilla; my PhD wouldn't have been so pleasant without you. Gracias a mis amigos de siempre por tener las palabras correctas en el momento adecuado; más de una vez, e incluso a distancia, sus palabras me recargaron de energía. Merci à Quentin pour son patiente et soutien inconditionnel pendent ces années, tu as rendu ma vie en France beaucoup plus agréable. Finalmente, gracias infinitas a mis padres, hermana y tío por su amor incondicional; sin el apoyo constante de ustedes hubiera sido imposible lograr todo lo que he conseguido. A pesar de la distancia, siempre sentí sus palabras de apoyo y cariño muy próximas a mi corazón.

10. MEMBERS OF THE JURY

PhD supervised by:

Mart Krupovic, PhD
Unité Virologie des Archées - Département de Microbiologie
Institut Pasteur
25, rue du Dr. Roux
75015, Paris
France
mart.krupovic@pasteur.fr

David Prangishvili, PhD
Honorary Professor
Institut Pasteur
25, rue du Dr. Roux
75015, Paris
France
david.prangishvili@pasteur.fr

PhD thesis reviewed by:

Claire Geslin, PhD
Laboratoire de Microbiologie des Environnements Extrêmes
Université de Brest
29280 Plouzané
France
claire.geslin@univ-brest.fr

Pascale Boulanger, PhD
Bactériophage T5 – Institut de Biologie Intégrative de la Cellule
Université Paris Sud - Orsay
15 rue G. Clémenceau
France
pascale.boulanger@i2bc.paris-saclay.fr

President of the PhD committee:

Guennadi Sezonov, Prof.
Centre de Recherche des Cordeliers, Equipe 12
Sorbonne Université
15 Rue de l'École de Médecine
75006, Paris
France
guennadi.sezonov@sorbonne-universite.fr