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
Institut Pasteur, Department of Microbiology,
75015 Paris, France
Tel: 33 (0)1 40 61 37 22

Running title

Biogeographic pattern in archaeal virus communities

Competing Interests

The authors declare that they have no competing interests.

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 establishment of five 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 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]. Portoglobovirus 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 homologues 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 viruses from different families 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 [2]. However, despite their diversity, distinctiveness and abundance, the number of isolated species of viruses infecting hyperthermophilic archaea remains low compared to 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 [41]. Similarly, using a combination of viral assemblage sequencing and network analysis, it has been estimated that out of 110 identified

115 virus groups, less than 10% represent known archaeal viruses, suggesting that the vast majority
116 of virus clusters represent unknown viruses, likely infecting archaeal hosts [42]. Furthermore, the
117 evolution and structuring of virus communities in terrestrial hydrothermal settings remains
118 poorly understood. Here, to improve understanding on these issues, we explored the diversity of
119 archaeal viruses at the active solfataric field of the Campi Flegrei volcano [43-45] in Pozzuoli,
120 Italy, namely the Pisciarelli hydrothermal area. The field is known for hot acidic environmental
121 conditions [45] and its microbial communities are dominated by extremophilic microbes [34, 46-
122 49]. However, the area we selected is characterized by hydrothermal conditions that are
123 continuously changing in the short term due to the volcano dynamics, which could have an effect
124 on the composition of resident microbial communities and their viruses. We report on the
125 isolation and characterization of five new archaeal viruses belonging to three different families
126 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 enrichments 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, 50]. 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 (Figure 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* (Figure 1a), *Bicaudaviridae* (Figure 1b), *Ampullaviridae* (Figure 1c), *Rudiviridae* (Figure 1d) and *Lipothrixviridae* (Figure 1e). VLPs propagated in the *Pyrobaculum* medium resembled members of the families *Globuloviridae* (Figure 1f), *Tristromaviridae* (Figure 1g) and *Guttaviridae* (Figure 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-

158 producing samples I3 and I9. Concentrated VLPs were first tested against the isolates by spot
159 test. In case of cell growth inhibition, a liquid culture of the isolate was established and infected
160 with the halo zone observed in the spot test. The production of the viral particles was
161 subsequently verified by TEM. As a result, three strains replicating VLPs were identified.
162 Comparison of their 16S rRNA gene sequences showed that the three strains belong to three
163 different genera of the order Sulfolobales, namely, *Saccharolobus* (until recently known as
164 *Sulfolobus*), *Acidianus* and *Metallosphaera*. The 16S rRNA genes of these strains, POZ9,
165 POZ149 and POZ202, respectively, displayed 100%, 99% and 99% identity to the corresponding
166 genes of *Acidianus brierleyi* DSM 1651 (NZ_CP029289), *Saccharolobus solfataricus* Ron 12/III
167 (X90483) and *Metallosphaera sedula* SARC-M1 (CP012176). Rod-shaped particles of different
168 sizes were propagated successfully in the three isolated strains (Figure 2a-c) and, following the
169 nomenclature used for other rudiviruses, were named *Metallosphaera* rod-shaped virus 1
170 (MRV1), *Acidianus* rod-shaped virus 3 (ARV3) and *Saccharolobus solfataricus* rod-shaped virus
171 1 (SSRV1), respectively.
172
173 Negatively stained MRV1, ARV3 and SSRV1 virions are rod-shaped particles measuring 630 ±
174 20 × 25 ± 2 nm, 670 ± 40 × 23 ± 4 nm and 750 ± 30 × 24 ± 3 nm, respectively (Figure 2a-c).
175 Similar to members of the *Rudiviridae* family, the three viral particles have terminal fibers
176 located at each end of the virion, which have been shown to play a role in host recognition in the
177 case of *Sulfolobus islandicus* rod-shaped virus 2 (SIRV2) [51].
178
179 The three rudiviruses displayed different infection dynamics. Infection of *S. solfataricus* POZ149
180 cultures with SSRV1 (MOI=3) resulted in severe growth retardation (Figure S1a). The optical
181 density of infected cultures remained constant for the first 24 hpi, similar to what has been
182 reported previously for the prototypical rudivirus SIRV2 [28]. By contrast, infection with MRV1
183 and ARV3 had no apparent effect on the growth of *M. sedula* POZ202 and *A. brierleyi* POZ9,
184 respectively (Figure S1a). Nevertheless, production of extracellular virions was detected in all
185 three cultures, starting at 8 (for SSRV1) or 12 (for MRV1 and ARV3) hpi and peaking at 24-32
186 hpi (Figure S2).
187

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 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; Figure 2d) and Pyrobaculum spherical virus 2 (PSV2; Figure 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 ± 20 × 34 ± 4 nm in size with terminal filaments of up to 130 nm in length attached to one or both ends of the virions (Figure 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 (Figure 2e), resembling Pyrobaculum spherical virus (PSV) particles [49]. Unfortunately, the hosts for other VLPs shown in Figure 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 (Figure 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 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 Figure S3). Notably, however, adsorption assay (see Supplementary Materials and Methods) did not show appreciable binding of SSRV1 virions to most of the non-host strains (Figure 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 Figure S3). These results suggest that MRV1 is able to inject its DNA into hosts 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, Figure S3), whereas PSV2 genome was detected not only in the host strain but also in *P. oguniense* TE7 and *P. arsenaticum* PZ6, 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, 49]. The coding region of the PSV2 genome is flanked by 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* (Figure 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) [52], have closest homologs in PSV [49] with amino acid sequence identities ranging between 28% and 65% (Figure 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 [53]. 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 *Saccharolobus 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 member in the Tristromaviridae family

The linear genome of the isolated PFV2 carries 17,602 bp 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, 49], 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 [54]. 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 (Figure 4). PFV1 gene 27 encodes a minor virion protein, whereas genes 29 and 30 encode putative lectin-like carbohydrate-binding proteins (Rensen et al., 2016). 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) [55], the only other known member of the *Tristromaviridae* family [54] (Figure 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 (Figure 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 (Figure 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 [48]; 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 [56]. 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 (Figure 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; Figure S2), but no evident pyramids were observed on the surface of the infected cells (Figure 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 [57]. 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 [58]. Our results unequivocally show that the 19 sequenced rudiviruses fall into six clades corresponding to the geographical origins of the virus isolation (Figure 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.

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 [57, 59]. 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 [60-62]. Thus, to investigate which hosts were exposed to Italian rudiviruses, we searched the CRISPRdb database [63] 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 (Figure 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 [64], 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 compared to the other viruses (Figure 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 [57]. 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), "A*zorudivirus*" (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 [65].

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, 66-69], 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 [49, 70-73], 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 [48]. 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 *Sulfolobus islandicus*-infecting rudiviruses isolated from hot springs in Iceland and United States [57]. 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* [73-75]. 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 [76].

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 [73, 77]. 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 [78, 79]. 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 [80].

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 (Figure S3). Given the basal position of *Acidianus* rudiviruses (Figure 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 [81]. 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 [82, 83]. 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 [59], was found to be targeted by 521 unique CRISPR spacers associated with all four principal CRISPR repeat sequences present in Sulfolobales [77], 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 [77]. 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% [61, 62, 84]. 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.

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. 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. 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.

COMPETING INTERESTS

The authors declare that they have no conflict of interest.

SUPPLEMENTARY INFORMATION

Supplementary information is available at ISME J website.

DATA AVAILABILITY

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

REFERENCES

1. Prangishvili D, Bamford DH, Forterre P, Iranzo J, Koonin EV, Krupovic M. The enigmatic archaeal virosphere. Nat Rev Microbiol. 2017; 15:724-739.
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 U S A. 2019; 116:15645-15650.
12. Rice G, Stedman K, Snyder J, Wiedenheft B, Willits D, Brumfield S *et al*. Viruses from extreme thermal environments. Proc Natl Acad Sci U S A. 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 U S A. 2019; 116:22591-22597.
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. Journal of virology. 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 U S A. 2018; 115:2120-2125.
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 U S A. 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; 244:181-193.
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; 90:11043-11055.
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 U S A. 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. Journal of virology. 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 U S A. 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. Nature microbiology. 2019; 4:1401-1410.
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; 577:572-575.
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; 179:448-458 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; 3:461-469.
40. Hurwitz S, Lowenstern JB. Dynamics of the Yellowstone hydrothermal system. Rev Geophys. 2014; 52:375-411.

630 41. Snyder JC, Bateson MM, Lavin M, Young MJ. Use of cellular CRISPR (clusters of regularly interspaced
631 short palindromic repeats) spacer-based microarrays for detection of viruses in environmental
632 samples. Appl Environ Microbiol. 2010; 76:7251-8.
633 42. Bolduc B, Wirth JF, Mazurie A, Young MJ. Viral assemblage composition in Yellowstone acidic hot
634 springs assessed by network analysis. ISME J. 2015; 9:2162-77.
635 43. Piochi M, Kilburn CRJ, Di Vito MA, Mormone A, Tramelli A, Troise C *et al*. The volcanic and
636 geothermally active Campi Flegrei caldera: an integrated multidisciplinary image of its buried
637 structure. Int Journ Earth Sci. 2014; 103:401-421.
638 44. Piochi M, Mormone A, Balassone G, Strauss H, Troise C, De Natale G. Native sulfur, sulfates and
639 sulfides from the active Campi Flegrei volcano (southern Italy): Genetic environments and
640 degassing dynamics revealed by mineralogy and isotope geochemistry. J Volcanol Geotherm Res.
641 2015; 304:180-193.
642 45. Piochi M, Mormone A, Strauss H, Balassone G. The acid sulfate zone and the mineral alteration styles
643 of the Roman Puteoli (Neapolitan area, Italy): clues on fluid fracturing progression at the Campi
644 Flegrei volcano. Solid Earth. 2019; 10:1809-1831.
645 46. Menzel P, Gudbergsdottir SR, Rike AG, Lin L, Zhang Q, Contursi P *et al*. Comparative Metagenomics
646 of Eight Geographically Remote Terrestrial Hot Springs. Microb Ecol. 2015; 70:411-24.
647 47. Ciniglia C, Yoon HS, Pollio A, Pinto G, Bhattacharya D. Hidden biodiversity of the extremophilic
648 Cyanidiales red algae. Mol Ecol. 2004; 13:1827-38.
649 48. Gudbergsdóttir SR, Menzel P, Krogh A, Young M, Peng X. Novel viral genomes identified from six
650 metagenomes reveal wide distribution of archaeal viruses and high viral diversity in terrestrial hot
651 springs. Environ Microbiol. 2016; 18:863-74.
652 49. Häring M, Peng X, Brugger K, Rachel R, Stetter KO, Garrett RA *et al*. Morphology and genome
653 organization of the virus PSV of the hyperthermophilic archaeal genera Pyrobaculum and
654 Thermoproteus: a novel virus family, the Globuloviridae. Virology. 2004; 323:233-42.
655 50. Zillig W, Kletzin A, Schleper C, Holz I, Janekovic D, Hain J *et al*. Screening for Sulfolobales, their
656 Plasmids and their Viruses in Icelandic Solfataras. Syst Appl Microbiol. 1993; 16:609-628.
657 51. Quemin ER, Lucas S, Daum B, Quax TE, Kuhlbrandt W, Forterre P *et al*. First insights into the entry
658 process of hyperthermophilic archaeal viruses. J Virol. 2013; 87:13379-85.
659 52. Prangishvili D, Krupovic M, ICTV Report Consortium. ICTV Virus Taxonomy Profile: *Globuloviridae*. J
660 Gen Virol. 2018; 99:1357-1358.
661 53. Ahn DG, Kim SI, Rhee JK, Kim KP, Pan JG, Oh JW. TTSV1, a new virus-like particle isolated from the
662 hyperthermophilic crenarchaeote Thermoproteus tenax. Virology. 2006; 351:280-90.
663 54. Prangishvili D, Rensen E, Mochizuki T, Krupovic M, ICTV Report Consortium. ICTV Virus Taxonomy
664 Profile: Tristromaviridae. J Gen Virol. 2019; 100:135-136.
665 55. Neumann H, Schwass V, Eckerskorn C, Zillig W. Identification and characterization of the genes
666 encoding three structural proteins of the Thermoproteus tenax virus TTV1. Mol Gen Genet. 1989;
667 217:105-10.
668 56. Prangishvili D, Koonin EV, Krupovic M. Genomics and biology of Rudiviruses, a model for the study of
669 virus-host interactions in Archaea. Biochem Soc Trans. 2013; 41:443-50.
670 57. Bautista MA, Black JA, Youngblut ND, Whitaker RJ. Differentiation and Structure in Sulfolobus
671 islandicus Rod-Shaped Virus Populations. Viruses. 2017; 9:E120.
672 58. Meier-Kolthoff JP, Goker M. VICTOR: genome-based phylogeny and classification of prokaryotic
673 viruses. Bioinformatics. 2017; 33:3396-3404.
674 59. Liu Y, Brandt D, Ishino S, Ishino Y, Koonin EV, Kalinowski J *et al*. New archaeal viruses discovered by
675 metagenomic analysis of viral communities in enrichment cultures. Environ Microbiol. 2019;
676 21:2002-2014.

677 60. Roux S, Krupovic M, Daly RA, Borges AL, Nayfach S, Schulz F *et al*. Cryptic inoviruses revealed as
678 pervasive in bacteria and archaea across Earth's biomes. Nat Microbiol. 2019; 4:1895-1906.
679 61. Roux S, Adriaenssens EM, Dutilh BE, Koonin EV, Kropinski AM, Krupovic M *et al*. Minimum
680 Information about an Uncultivated Virus Genome (MIUViG). Nat Biotechnol. 2019; 37:29-37.
681 62. Edwards RA, McNair K, Faust K, Raes J, Dutilh BE. Computational approaches to predict
682 bacteriophage-host relationships. FEMS Microbiol Rev. 2016; 40:258-72.
683 63. Pourcel C, Touchon M, Villeriot N, Vernadet JP, Couvin D, Toffano-Nioche C *et al*. CRISPRCasdb a
684 successor of CRISPRdb containing CRISPR arrays and cas genes from complete genome sequences,
685 and tools to download and query lists of repeats and spacers. Nucleic Acids Res. 2020; 48:D535-
686 D544.
687 64. Ågren J, Sundström A, Håfström T, Segerman B. Gegenees: fragmented alignment of multiple
688 genomes for determining phylogenomic distances and genetic signatures unique for specified
689 target groups. PLoS One. 2012; 7:e39107.
690 65. Simmonds P, Adams MJ, Benko M, Breitbart M, Brister JR, Carstens EB *et al*. Consensus statement:
691 Virus taxonomy in the age of metagenomics. Nat Rev Microbiol. 2017; 15:161-168.
692 66. Häring M, Rachel R, Peng X, Garrett RA, Prangishvili D. Viral diversity in hot springs of Pozzuoli, Italy,
693 and characterization of a unique archaeal virus, Acidianus bottle-shaped virus, from a new family,
694 the Ampullaviridae. J Virol. 2005; 79:9904-11.
695 67. Vestergaard G, Haring M, Peng X, Rachel R, Garrett RA, Prangishvili D. A novel rudivirus, ARV1, of the
696 hyperthermophilic archaeal genus Acidianus. Virology. 2005; 336:83-92.
697 68. Prangishvili D, Vestergaard G, Haring M, Aramayo R, Basta T, Rachel R *et al*. Structural and genomic
698 properties of the hyperthermophilic archaeal virus ATV with an extracellular stage of the
699 reproductive cycle. J Mol Biol. 2006; 359:1203-16.
700 69. Vestergaard G, Aramayo R, Basta T, Haring M, Peng X, Brugger K *et al*. Structure of the acidianus
701 filamentous virus 3 and comparative genomics of related archaeal lipothrixviruses. Journal of
702 virology. 2008; 82:371-81.
703 70. Arnold HP, Ziese U, Zillig W. SNDV, a novel virus of the extremely thermophilic and acidophilic
704 archaeon Sulfolobus. Virology. 2000; 272:409-16.
705 71. Mochizuki T, Sako Y, Prangishvili D. Provirus induction in hyperthermophilic archaea:
706 characterization of Aeropyrum pernix spindle-shaped virus 1 and Aeropyrum pernix ovoid virus 1.
707 J Bacteriol. 2011; 193:5412-9.
708 72. Wiedenheft B, Stedman K, Roberto F, Willits D, Gleske AK, Zoeller L *et al*. Comparative genomic
709 analysis of hyperthermophilic archaeal Fuselloviridae viruses. J Virol. 2004; 78:1954-61.
710 73. Pauly MD, Bautista MA, Black JA, Whitaker RJ. Diversified local CRISPR-Cas immunity to viruses of
711 Sulfolobus islandicus. Philos Trans R Soc Lond B Biol Sci. 2019; 374:20180093.
712 74. Held NL, Whitaker RJ. Viral biogeography revealed by signatures in Sulfolobus islandicus genomes.
713 Environ Microbiol. 2009; 11:457-66.
714 75. Stedman KM, Clore A, Combet-Blanc Y (2006). Biogeographical diversity of archaeal viruses. In:
715 Logan NA, Lappin-Scott HM, Oyston PCF (eds). *Prokaryotic Diversity: Mechanisms and*
716 *Significance*. Cambridge University Press: Cambridge. pp 131-143.
717 76. Snyder JC, Wiedenheft B, Lavin M, Roberto FF, Spuhler J, Ortmann AC *et al*. Virus movement
718 maintains local virus population diversity. Proc Natl Acad Sci U S A. 2007; 104:19102-7.
719 77. Medvedeva S, Liu Y, Koonin EV, Severinov K, Prangishvili D, Krupovic M. Virus-borne mini-CRISPR
720 arrays are involved in interviral conflicts. Nat Commun. 2019; 10:5204.
721 78. Atanasova NS, Roine E, Oren A, Bamford DH, Oksanen HM. Global network of specific virus-host
722 interactions in hypersaline environments. Environ Microbiol. 2012; 14:426-40.
723 79. Atanasova NS, Bamford DH, Oksanen HM. Virus-host interplay in high salt environments. Environ
724 Microbiol Rep. 2016; 8:431-44.

80. Laidler JR, Shugart JA, Cady SL, Bahjat KS, Stedman KM. Reversible inactivation and desiccation tolerance of silicified viruses. J Virol. 2013; 87**:**13927-9.
81. 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**:**801-808 e3.
82. 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.
83. 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.
84. 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**:**689-693.

FIGURE LEGENDS

Figure 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. Bars: 200 nm.

Figure 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.

Figure 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 grey. 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.

Figure 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 grey, 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.

Figure 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 [48, 67]. 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. ATase, acetyltransferase; GHase, glycoside hydrolase; GTase, glycosyltransferase; HJR, Holliday junction resolvase; MCP, major capsid protein; MTase, methyltransferase; SSB, ssDNA binding protein; TFB, Transcription factor B; ThyX, thymidylate synthase; wHTH, winged helix-turn-helix domain.

Figure 6. Inferred phylogenomic tree of all known members of the *Rudiviridae* family based on whole genome VICTOR [58] 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.

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

	Newly-isolated archaeal virus				
Archaeal strain	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 DSM639	-	-	-	-	-
Pyrobaculum arsenaticum PZ6 (DSM 13514)	-	-	-	-*	-
Pyrobaculum arsenaticum 2GA	-	-	-	H	H
Pyrobaculum calidifontis VA1 (DSM 21063)	-	-	-	-	-
Pyrobaculum oguniense TE7 (DSM 13380)	-	-	-	-*	+*

789 H, isolation host; +, supports virus replication; *, supports DNA delivery; -, no virus production observed

790

791 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

792 TIR, terminal inverted repeats; ORFs, open reading frames.

793

794

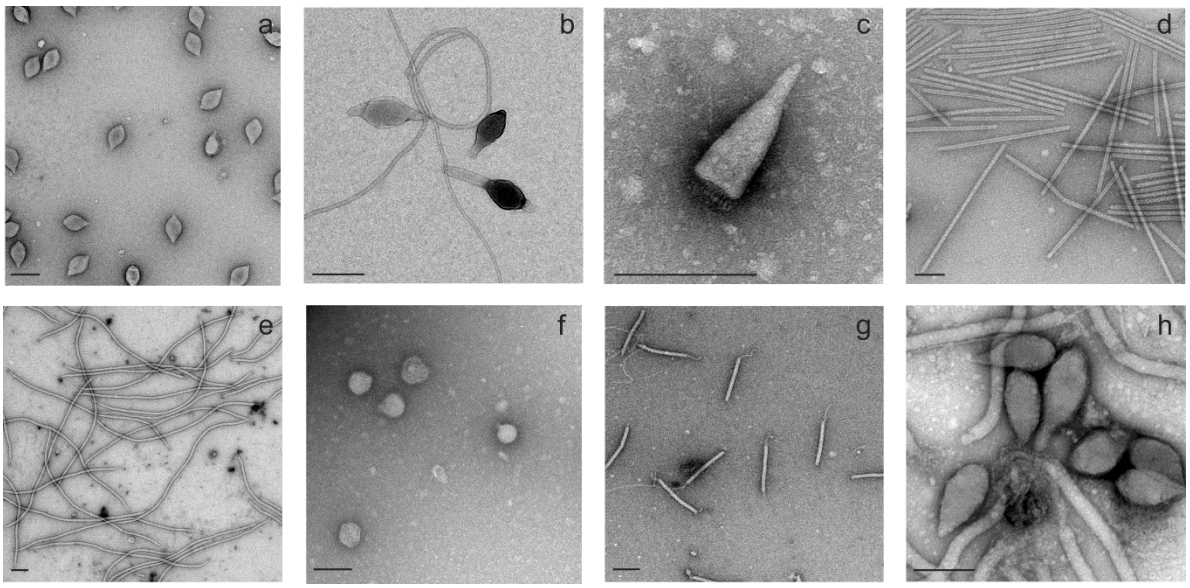

Figure 1.

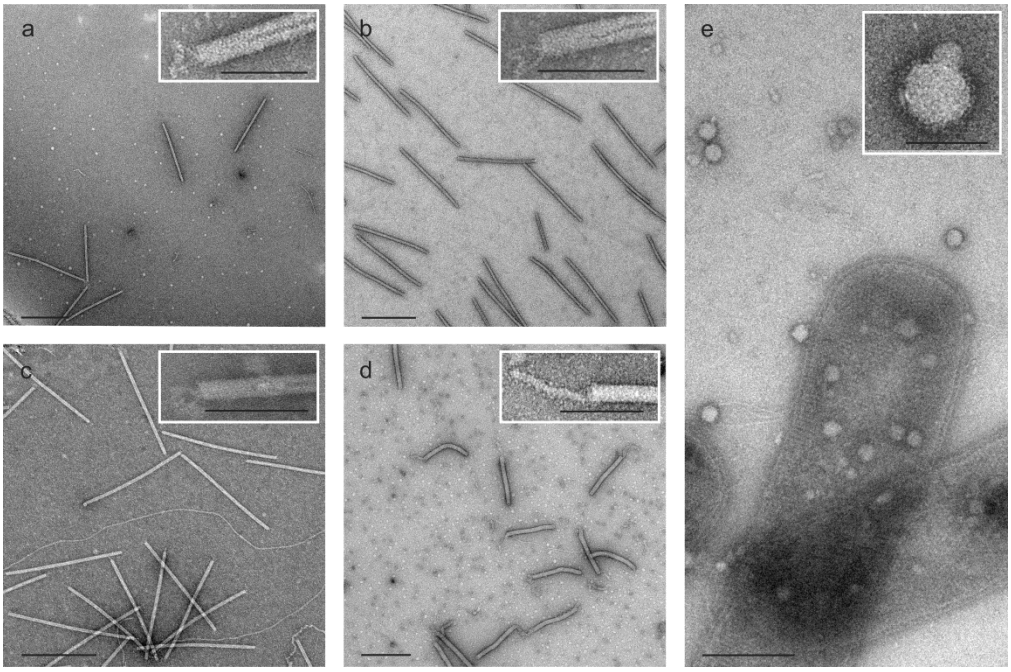

Figure 2

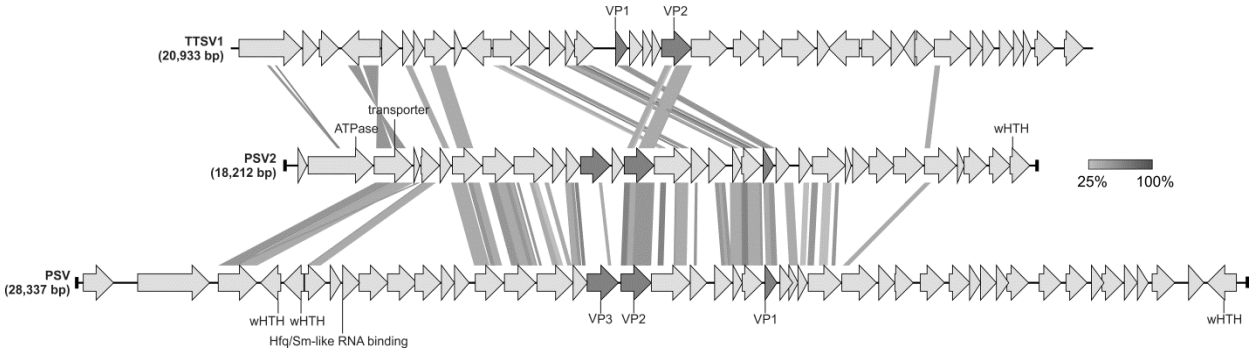

Figure 3

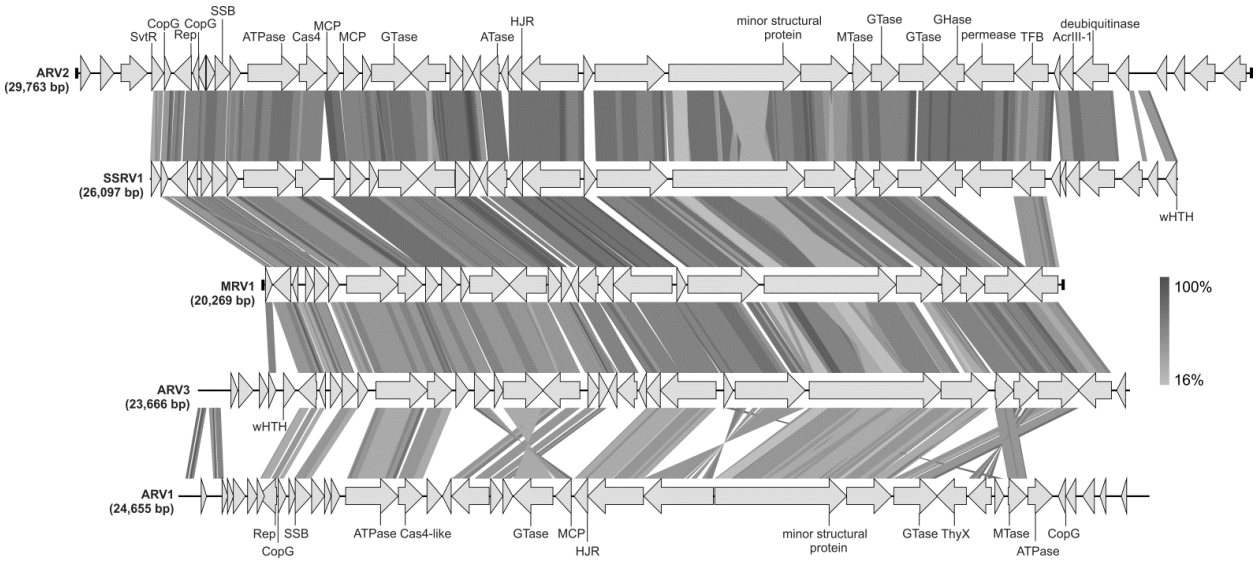

Figure 4

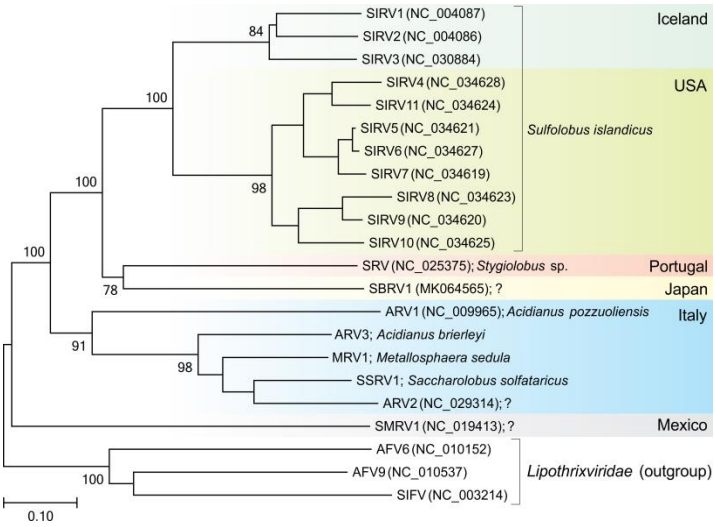
SIRV1 (NC_004087)
SIRV2 (NC_004086)
SIRV3 (NC_030884)
Iceland
SIRV4 (NC_034628)
SIRV11 (NC_034624)
SIRV5 (NC_034621)
SIRV6 (NC_034627)
SIRV7 (NC_034619)
SIRV8 (NC_034623)
SIRV9 (NC_034620)
SIRV10 (NC_034625)
USA
Sulfolobus islandicus
SRV (NC_025375); Stygiolobus sp.
Portugal
SBRV1 (MK064565); ?
Japan
ARV1 (NC_009965); Acidianus pozzuoliensis
ARV3; Acidianus brierleyi
MRV1; Metallosphaera sedula
SSRV1; Saccharolobus solfataricus
ARV2 (NC_029314); ?
Italy
SMRV1 (NC_019413); ?
Mexico
AFV6 (NC_010152)
AFV9 (NC_010537)
SIFV (NC_003214)
Lipothrixviridae (outgroup)
84
100
100
98
100
78
91
98
100
0.10

Figure 6