New insights into the diversity and evolution of the archaeal mobilome from three complete genomes of Saccharolobus shibatae

Sofia Medvedeva, David Brandt, Virginija Cvirkaite-Krupovic, Ying Liu, Konstantin Severinov, Sonoko Ishino, Yoshizumi Ishino, David Prangishvili, Jörn Kalinowski, Mart Krupovic

New insights into the diversity and evolution of the archaeal mobilome from three complete genomes of *Saccharolobus shibatae*

Sofia Medvedeva[1,2], David Brandt[3], Virginija Cvirkaite-Krupovic[1], Ying Liu[1], Konstantin Severinov[2,4,5], Sonoko Ishino[6], Yoshizumi Ishino[6], David Prangishvili[1,7], Jörn Kalinowski[3], Mart Krupovic[1]*

[1] Archaeal Virology Unit, Institut Pasteur, Paris 75015, France
[2] Center of Life Science, Skolkovo Institute of Science and Technology, Moscow 121205, Russia
[3] Center for Biotechnology, Universität Bielefeld, Bielefeld 33615, Germany
[4] Waksman Institute, Rutgers University, Piscataway, NJ, 08854, USA
[5] Institute of Molecular Genetics, Russian Academy of Sciences, Moscow 123182, Russia
[6] Department of Bioscience and Biotechnology, Graduate School of Bioresource and Bioenvironmental Sciences, Kyushu University, 744 Motooka, Nishi-ku, Fukuoka, Fukuoka, 819-0395, Japan
[7] Ivane Javakhishvili Tbilisi State University, Tbilisi 0179, Georgia

* - Correspondence to
E-mail: mart.krupovic@pasteur.fr
Tel: +33 1 40 61 37 22

Running head: Mobilome of *Saccharolobus shibatae*

SUMMARY

Saccharolobus* (formerly *Sulfolobus*) *shibatae* B12, isolated from a hot spring in Beppu, Japan in 1982, is one of the first hyperthermophilic and acidophilic archaeal species to be discovered. It serves as a natural host to the extensively studied spindle-shaped virus SSV1, a prototype of the *Fuselloviridae* family. Two additional *Sa. shibatae* strains, BEU9 and S38A, sensitive to viruses of the families *Lipothrixviridae* and *Portogloboviridae*, respectively, have been isolated more recently. However, none of the strains has been fully sequenced, limiting their utility for studies on archaeal biology and virus-host interactions. Here, we present the complete genome sequences of all three *Sa. shibatae* strains and explore the rich diversity of their integrated mobile genetic elements (MGE), including transposable insertion sequences, integrative and conjugative elements, plasmids, and viruses, some of which were also detected in the extrachromosomal form. Analysis of related MGEs in other Sulfolobales species and patterns of CRISPR spacer targeting revealed a complex network of MGE distributions, involving horizontal spread and relatively frequent host switching by MGEs over large phylogenetic distances, involving species of the genera *Saccharolobus*, *Sulfurisphaera* and *Acidianus*. Furthermore, we characterize a remarkable case of a virus-to-plasmid transition, whereby a fusellovirus has lost the genes encoding for the capsid proteins, while retaining the replication module, effectively becoming a plasmid.

INTRODUCTION

Terrestrial acidic hydrothermal environments are dominated by Archaea and their viruses (Wang et al., 2015; Prangishvili et al., 2017; Munson-McGee et al., 2018a; Lewis et al., 2021). Members of the order Sulfolobales are particularly abundant in such habitats and thrive in sulfuric thermal springs and solfataras around the globe (Kvist et al., 2007; Kozubal et al., 2012; Munson-McGee et al., 2018b; Liu et al., 2019). The order Sulfolobales includes 7 genera (*Acidianus*, *Metallosphaera*, *Saccharolobus*, *Sulfolobus*, *Sulfurisphaera*, *Stygiolobus*, *Sulfodiicoccus* and *Sulfuracidifex*) with more than 30 species in total (Counts et al., 2020). Besides variation in optimal growth temperature (65-88°C) and pH (0.7 - 4.5), Sulfolobales members are characterized by different morphological and genetic features, and metabolic capacities, such as oxidation/reduction of sulfur and iron, oxygen tolerance, sugar catabolism, etc.

A notable characteristic of the Sulfolobales is the remarkable diversity of their mobile genetic elements (MGE), including unique groups of viruses, plasmids and transposons, which collectively play a profound role in the hosts' genome evolution by gene capture and horizontal gene transfer (Zillig et al., 1998; Brügger et al., 2004; Lipps, 2006; Wang et al., 2015; Krupovic et al., 2018). The MGEs provide hotspots for genomic rearrangements and promote species diversification through evolutionary arms race (Redder and Garrett, 2006). To cope with constant onslaught by MGEs, especially viruses, the Sulfolobales customary carry CRISPR-Cas and toxin-antitoxin systems which are concentrated in genomic regions called defense islands (Makarova et al., 2013). The adaptive CRISPR-Cas immunity system consists of CRISPR arrays, which contain spacers, small DNA fragments some of which are derived from MGEs, and CRISPR-associated (*cas*) genes. The *cas* genes form two major, functionally distinct modules: the adaptation module captures new MGE fragments and incorporates them into CRISPR arrays, whereas the interference module recognizes and cleaves the DNA or RNA containing sequences matching acquired fragments. The Sulfolobales species typically encode CRISPR-Cas systems of types I and III, with 1-2 adaptation modules, 1-5 interference modules and 2-6 CRISPR arrays. The toxin-antitoxin systems function in a different fashion. Although in archaea, the role of toxin-antitoxin systems has not been studied in detail, it has been shown in bacteria that upon virus infection, toxins induce cell dormancy or death (i.e., altruistic suicide), thereby blocking the virus spread in the population (Lopatina et al., 2020).

Insertion sequences (IS) are transposable elements capable of relocation from one genomic position to another with the help of transposases encoded within the IS boundaries (for autonomous transposons) or somewhere else in the genome (for non-autonomous transposons). IS elements are particularly abundant in the genomes of members of the genus *Saccharolobus* (Redder et al., 2001; Redder and Garrett, 2006) and are often upregulated in response to various forms of stress, including virus infection (Quax et al., 2013). The plasmids of Sulfolobales are generally divided into larger conjugative and small non-conjugative ones, depending on the presence of conserved genes involved in conjugation, including ATPases VirB4/TrbE and VirD4/TraG (Prangishvili et al., 1998; Stedman et al., 2000; Erauso et al., 2006; Lipps, 2006). Other conserved genes carried by conjugative plasmids encode replication- and transcription regulation-related proteins. There is no consistent classification system for non-conjugative plasmids and they are typically assigned to types based on the major replication protein they encode. Plasmids of the most common type, pRN-like (represented by pRN1 plasmid of *Saccharolobus* [formerly *Sulfolobus*] *islandicus*), encode a large replication protein consisting of the N-terminal primase-polymerase domain and the C-terminal superfamily 3 helicase domain (Lipps, 2006, 2011).

Viruses infecting members of the Sulfolobales are classified into eight different families, which exhibit a broad range of morphologies, gene contents and infection strategies (Prangishvili et al., 2017; Krupovic

et al., 2018; Munson-McGee et al., 2018a). The first Sulfolobales virus, Sulfolobus spindle-shaped virus 1 (SSV1), the prototype member of the family *Fuselloviridae*, was discovered in 1982 together with its host *Saccharolobus* (formerly known as *Sulfolobus*) *shibatae* B12 (Yeats et al., 1982; Grogan et al., 1990). SSV1 is integrated in the tRNA-Arg gene of *Sa. shibatae* B12 (Reiter et al., 1989; Schleper et al., 1992) and its reactivation is induced by UV-irradiation (Schleper et al., 1992; Fusco et al., 2015), which leads to budding of spindle-shaped virus particles covered with a lipid-containing envelope (Quemin et al., 2015; Quemin et al., 2016). Following the discovery of SSV1, numerous related viruses, infecting different species of the *Saccharolobus* and *Acidianus*, have been isolated around the globe, testifying to the ubiquity and importance of the *Fuselloviridae* in extreme geothermal habitats (Wiedenheft et al., 2004; Redder et al., 2009; Krupovic et al., 2014; Goodman and Stedman, 2018; Pauly et al., 2019). Fuselloviruses have been extensively studied from structural, biochemical and genomic standpoints (Lawrence et al., 2009; Contursi et al., 2014; Quemin et al., 2015; Stedman et al., 2015; Pauly et al., 2019; Baquero et al., 2020a), with SSV1 being one of the best characterized archaeal viruses overall. Recently, two new viruses infecting *Sa. shibatae* strains, S38A and BEU9, have been isolated along with their viruses, namely, Sulfolobus polyhedral virus 1 (SPV1) (Liu et al., 2017), and Sulfolobus filamentous virus 1 (SFV1) (Liu et al., 2018). Both virus-host pairs were isolated from the Beppu hot spring complex from which the B12 strain has been isolated more than 30 years prior (Liu et al., 2019). SPV1 and SFV1 have been structurally characterized and classified into families *Portogloboviridae* and *Lipothrixviridae*, respectively (Liu et al., 2018; Wang et al., 2019a). Interestingly, despite unrelated virion architectures (i.e., icosahedral versus enveloped helical capsids), the dsDNA genomes of both viruses adopt the A-form in the virions, suggesting that this unusual DNA conformation is one of the adaptations to high temperature environments (Wang et al., 2020b). Furthermore, it has been shown that SPV1 genome carries mini-CRISPR arrays and apparently deploys them to eliminate other coinfecting MGEs, including closely related viruses (Medvedeva et al., 2019).

The *Sa. shibatae* strains and their viruses, and in particular, B12-SSV1, became important models for studying archaeal transcription (Martin et al., 1984; Reiter et al., 1987; Palm et al., 1991; Qureshi et al., 1997; Qureshi and Jackson, 1998; Wojtas et al., 2012), the mechanisms of viral genome integration (Clore and Stedman, 2007), UV-inducible gene expression and more. The SSV1 genome has been harnessed for developing vectors for protein expression, genetic manipulation and gene silencing in *Saccharolobus* species (Stedman et al., 1999; Jonuscheit et al., 2003; Zebec et al., 2014). *Sa. shibatae* itself also served as a valuable model for investigating different aspects of adaptation to high temperature environments, such as DNA relaxation by topoisomerases (Bergerat et al., 1994; Jaxel et al., 1996), RNA stability (Zhang and Zheng, 2018) and served as a source for discovery of thermostable proteins with potential industrial applications (Van et al., 2007; Boyce and Walsh, 2018). However, the application of these models has been strongly limited by the lack of the complete genome sequence for any of the *Sa. shibatae* strains. Consequently, *Sa. shibatae* was overshadowed by its close relatives with available full genome sequences, namely, *Saccharolobus solfataricus* and *Sa. islandicus* (She et al., 2001b; Guo et al., 2011; Zhang et al., 2018), which are widely used to study the diversity of archaeal MGEs (viruses, plasmids, insertion sequences), virus-host interactions, defense and anti-defense systems.

To expand the number of available models for the studies of archaeal cell biology, metabolism and virus-host interactions, here we present the complete genome sequences of all three isolated *Sa. shibatae* strains, B12, S38A and BEU9. Analysis and comparison of the three genomes coupled with mapping of a large set of CRISPR spacers sequenced directly from the hot spring from which these strains have been isolated revealed a diverse mobilome associated with *Sa. shibatae*, including novel conjugative and cryptic

plasmids, virus-like elements and transposons. Comparative genomics of these elements provided non-trivial insights into the evolution of archaeal MGE, including a remarkable case of virus evolution into a plasmid. Our data suggest a frequent horizontal transfers of MGE over large phylogenetic distances, crossing the genus boundary, which we attribute to the pressure exerted by CRISPR targeting.

RESULTS AND DISCUSSION

Core genome and strain-specific genes

To better understand the relationship between *Sa. shibatae* and other members of the Sulfolobales as well as to gain further insights into virus-host interactions in archaea, we have determined the genome sequences of three *Sa. shibatae* strains, namely, B12 (Yeats et al., 1982), S38A (Liu et al., 2017) and BEU9 (Liu et al., 2018). A combination of short-read and long-read sequencing yielded complete circular chromosomes for all three strains. The genomes were similar in size – 2,879,035 bp for B12, 2,791,785 bp for S38A and 2,740,857 bp for BEU9 – with the variation stemming from differences in the content of mobile genetic elements integrated into various chromosomal loci (see below). The three genomes have a GC% content of 35.4-35.7 and share genome-wide average nucleotide identity (ANI) of 98.8-99.2% (≥ 99.7% identity in their 16S rRNA genes). Phylogenetic analysis based on concatenated alignments of ribosomal proteins and the B subunit of RNA polymerase generally confirmed the established relationships between different genera in the order Sulfolobales and showed that *Sa. shibatae* belongs to a clade containing *Sa. solfataricus* and *Sa. islandicus*, as reported previously (Grogan et al., 1990; Dai et al., 2016). However, unlike in the previous 16S rRNA-based phylogenies, where *Sa. shibatae* typically occupies a basal position relative to *Sa. solfataricus* and *Sa. islandicus* species (Dai et al., 2016), in our present analysis, *Sa. shibatae* formed a sister group to *Sa. islandicus* (Fig. 1A), suggesting a closer evolutionary relationship.

Each *Sa. shibatae* genome encodes ~3200 genes, which can be assigned to ~2000 functional groups using arCOG classification (Makarova et al., 2015). The three *Sa. shibatae* strains share more than 93% of arCOG clusters (Fig. 1B), which represent the core genome. The majority of strain-specific arCOGs are concentrated within the hypervariable regions of the corresponding genomes (Fig. 1C). Similar hypervariable regions have been previously observed in *Sa. islandicus* (Jaubert et al., 2013), where they occupy the same genomic position as in *Sa. shibatae* (Fig. 1C). The core genome of *Sa. shibatae* is also similar to that of *Sa. islandicus*, which has been extensively analyzed elsewhere (Reno et al., 2009; Jaubert et al., 2013; Zhang et al., 2018) and will not be considered here. Instead, we will focus on the hypervariable regions and mobile genetic elements.

Similar to *Sa. islandicus*, the hypervariable regions of *Sa. shibatae* contain multiple insertion sequences (IS elements), toxin-antitoxin systems, CRISPR-Cas systems and distinct clusters of metabolic genes, which appear to contribute different adaptive capabilities to the three *Sa. shibatae* strains. Strain B12 contains the largest number of strain-specific genes and encompasses a ~100 kbp region encoding multiple glycosyltransferases and membrane proteins, which could be components of a secretion system. A similar region encompassing numerous glycosyltransferases has been previously described in several *S. acidocaldarius* strains and in *Sa. islandicus* HVE10/4 (Counts et al., 2020). By contrast, BEU9 carries a putative operon (*livKFGH*) constituting the leucine, isoleucine, and valine (LIV) branched-chain amino acid transport system (Ribardo and Hendrixson, 2011), not found in the two other strains.

One of the notable features in the hypervariable region of the S38A genome is the presence of several [NiFe]-hydrogenase clusters, which are not found in other *Sa. shibatae* strains. The [NiFe]-hydrogenase clusters typically consist of large and small hydrogenase subunits, membrane subunits and several maturation factors (Søndergaard et al., 2016). Based on the sequence similarities of the large subunits, hydrogenases are classified into ~30 types (Søndergaard et al., 2016). In *Sa. shibatae* S38A, we identified four different [NiFe]-hydrogenase types, namely, 1g, 1h, 2d and 2e (Fig. 2). Type 1g includes membrane-bound hydrogenases specific of Crenarchaeota, which were shown to participate in anaerobic reduction of elemental sulfur (Laska et al., 2003), whereas type 1h hydrogenases are oxygen-tolerant enzymes widespread in soil bacteria which oxidize atmospheric hydrogen (Greening et al., 2015). Functions of the type 2e and 2d hydrogenases are not known yet. Despite playing an important role in respiration and sulfur metabolism, [NiFe]-hydrogenases display patchy distribution in Sulfolobales genomes (present in 8 out of 21 *Sa. islandicus* genomes, 0/4 *S. acidocaldarius*, 2/2 *Sa. solfataricus*, 4/7 *Metallosphaera*, 8/8 *Acidianus* and 1/2 *Sulfurisphaera*). Notably, the 1h, 2d and 2e cluster of S38A display closest sequence similarity to the corresponding genes of *Sa. islandicus* HVE10/4, whereas the 1g cluster has been apparently lost in the latter strain (Fig. 2). Based on the genome analysis it appears that, unlike B12 and BEU9, S38A is capable of anaerobic growth by reduction of elemental sulfur, emphasizing that closely related strains can display considerable variation in metabolic capacities.

Defense systems

CRISPR-Cas systems

The hypervariable regions of the three *Sa. shibatae* strains encompass overlapping but distinct sets of CRISPR-Cas systems. Each strain carries four CRISPR arrays (39-134 spacers in length), several adaptation modules (I-A and I-D types) as well as complete and incomplete interference modules which can be assigned to I-A, I-D, III-B or III-D types based on the presence of signature genes (Fig. 3A). A type I-A system with adaptation and interference modules, and two cognate divergently oriented CRISPR arrays is present in all three strains as well as in *Sa. islandicus* strains (Fig. 3A). By contrast, all other CRISPR-Cas systems display more patchy distribution. Namely, two III-B type interference modules are shared by BEU9 and S38A, whereas I-A and III-D interference modules are shared by S38A and B12. The latter two strains also share two additional I-A adaptation modules with two additional CRISPR arrays. Notably, although spacer contents in these CRISPR arrays differ between the two strains, S38A and B12 share two adjacent, leader-distal spacers (Fig. 3A), indicative of the common ancestry of the corresponding arrays. In addition, the BEU9 strain carries I-D adaptation and interference modules along with two CRISPR arrays with CRISPR repeat sequences not found in B12 or S38A, and a III-B interference module (Fig. 3A). The provenance of these additional I-D and III-B CRISPR-Cas systems remains unclear. All complete III-B modules belong to the Cmr-β subtype, most closely similar to the corresponding module of *Sa. solfataricus* P2 (Yu et al., 2021). The fact that type III interference modules are not accompanied by cognate adaptation modules suggests that they function using the CRISPR spacers of type I systems in a PAM-independent manner, as demonstrated for other *Saccharolobus* species (Deng et al., 2013; Yu et al., 2021).

CRISPR spacers represent an archive of past virus-host interactions. The CRISPR arrays of the three *Sa. shibatae* strains collectively include 898 spacers, of which 97 (10.8%) could be matched (>85% identity) to known viruses and other MGEs of Sulfolobales (Fig. 3B). The majority of these spacers (n=67) target rod-shaped virus SBRV1 (family *Rudiviridae*), which has been sequenced (but not isolated) from the same enrichment culture from which *Sa. shibatae* BEU9 and S38A were isolated (Liu et al., 2019). Notably, some of the SBRV1-matching spacers are adjacent to each other in the CRISPR arrays (Figure 3B),

indicating that they were acquired consecutively, possibly in the course of primed adaptation. The extensive level of CRISPR targeting suggests that SBRV1 is able to escape CRISPR-Cas immunity by high mutation or replication rate or an anti-CRISPR mechanism. Rudivirus SIRV2 (and many other rudiviruses) is known to encode at least three different anti-CRISPR (Acr) proteins, neutralizing type I-D and type III systems (He et al., 2018; Bhoobalan-Chitty et al., 2019; Athukoralage et al., 2020). None of these Acrs have homologs in SBRV1, suggesting that the virus may encode novel proteins to block the CRISPR-Cas immunity. Notably, orthologs of the proteins constituting the adhesive type IV pili, which are recognized as host receptors by all known rudiviruses (Wang et al., 2019b; Rowland et al., 2020; Wang et al., 2020a), are conserved in all three *Sa. shibatae* strains.

Other viruses targeted by *Sa*. *shibatae* CRISPR spacers were previously identified as infecting one of three sequenced strains: fusellovirus SSV1 (5 spacers in S38A and 1 spacer in BEU9, albeit the latter with 84% identity), lipothrixvirus SFV1 (1 spacer in B12 and 1 spacer in S38A) and portoglobovirus SPV1 (2 spacers in B12 and 2 spacers in BEU9). Notably, S38A, the host of SPV1, contains one spacer which is 95% identical to the protospacer in the SPV1 genome. In addition, rudivirus SIRV10, fusellovirus SSV9 as well as bicaudaviruses SMV1 and SMV3 were targeted by one CRISPR spacer each (Fig. 3B). Overall, CRISPR targeting pattern is consistent with the adaptation to local viral communities, with viruses originating from the same hot spring being targeted more extensively than those isolated from geographically remote hot springs (Table S1), as has been observed previously for rudiviruses isolated from other geographic locations (Bautista et al., 2017). Except for the above mentioned pair of spacers shared by S38A and B12, the three *Sa. shibatae* strains do not contain shared spacers. Notably, we identified a spacer in S38A which targets a gene encoding an ABC-transporter in B12 and BEU9 (100% identity); the ortholog of this gene is not present in S38A, perhaps because it has been lost due to self-targeting by CRISPR.

Toxin-antitoxin systems

Toxin-antitoxin (TA) systems can function in antiviral defense by inducing cell dormancy or death, thereby preventing virus spread in the population (Lopatina et al., 2020). Indeed, it has been shown that both CRISPR-Cas and TA systems are strongly upregulated during infection of *Sa. islandicus* (Quax et al., 2013). The complement of TA systems is similar in all three *Sa. shibatae* strains (Table 1; Table S2). The identified toxins belong to RelE, PIN (PilT N-terminal) and HEPN (Higher Eukaryotes and Prokaryotes Nucleotide-binding) families and all are predicted to function as ribonucleases. These toxins are associated with structurally and functionally distinct antitoxins (Table 1), emphasizing the modular evolution of TA systems (Makarova et al., 2013). Toxins of the PIN and HEPN are conserved in all *Sa. islandicus* strains (Jaubert et al., 2013), whereas RelE family toxins are present in *Sa. tokodaii* and certain *Sa. islandicus* strains (Makarova et al., 2015).

Sa. shibatae **strains are associated with a diverse mobilome**

Transposable elements

As in the case of *Sa. solfataricus* and *Sa. islandicus*, *Sa. shibatae* genomes are heavily infested by IS elements. The IS elements of *Saccharolobus* have been implicated in insertional mutagenesis and genome rearrangements, which are reported to occur at relatively high frequencies (Martusewitsch et al., 2000; Blount and Grogan, 2005; Redder and Garrett, 2006). Using the ISsaga tool (Varani et al., 2011), we identified 10 families of autonomous and 5 families of non-autonomous transposons (Fig. 4; Table S3). The identified autonomous transposons of the IS*1*, IS*5*, IS*6*, IS*256*, IS*630*, IS*H3* families encode DDE-type transposases, whereas IS*110* encode evolutionarily related transposase of the DEDD family. By

contrast, IS*200*/IS*605* and IS*607* encode distinct transposases belonging to the tyrosine and serine recombinase superfamilies, respectively (Siguier et al., 2014). The non-autonomous transposons, namely, TnpB and miniature inverted repeat transposable elements (MITE), lack the transposase genes and are mobilized by the transposases of their autonomous siblings. The overall number and diversity of IS elements is comparable with that reported for *Sa. islandicus* and *Sa. solfataricus* (Brügger et al., 2004; Redder and Garrett, 2006; Guo et al., 2011; Jaubert et al., 2013), although there is notable variation between the three *Sa. shibatae* strains. In particular, IS*110* and IS*256*, and to a lesser extent, IS*1* and IS*5* are considerably more abundant in S38A and BEU9 than in B12, suggesting active proliferation of these IS families following the diversification of the common ancestor of S38A and BEU9 from that of B12. Furthermore, IS*200*/IS*605* and IS*607* family elements were not present in B12. Interestingly, a B12 CRISPR array of type I-A carries a spacer matching the IS*200*/IS*605* found in S38A and BEU9. This observation suggests that CRISPR system might participate in controlling the intragenomic transposon proliferation.

Integrated mobile genetic elements

We have previously shown that MGE integrated in archaeal genomes can be efficiently identified by matching the CRISPR spacers across the corresponding cellular genomes (Medvedeva et al., 2019). Thus, we mapped a large collection of CRISPR spacers sequenced from the uncultivated Sulfolobales community of the Beppu thermal field (~40,000 unique spacer sequences; (Medvedeva et al., 2019)) and spacers extracted from completely sequenced Sulfolobales genomes (~10,000 spacers) onto the *Sa. shibatae* genomes. As expected, CRISPR spacer hits (protospacers) were unevenly distributed across the *Sa. shibatae* genomes, with certain genomic islands being matched by multiple spacers (Fig. 5). The regions displaying elevated density of protospacers were further analyzed for the presence of features characteristic of integrated (i)MGEs, namely, genes encoding transposases, integrases and virus- or plasmid-specific proteins. To define the precise boundaries of the iMGE, the prospective regions were searched for the presence of direct repeats corresponding to the attachment (att) sites. In total, we predicted 14 iMGE (Fig. 5, Table 2, Table S4), among which only fusellovirus SSV1 has been described previously (Palm et al., 1991; Schleper et al., 1992). Somewhat surprisingly, considering the high insertional mutagenesis rate in *Sulfolobus* (Martusewitsch et al., 2000), the genome sequence of SSV1 determined in this study was identical to that obtained 30 years ago (Palm et al., 1991). In addition to SSV1, B12 contains five other iMGE (E1-E5), whereas S38A and BEU9 carry four iMGEs each. The iMGEs vary in size from 6.7 kbp (for BEU9-E2) to 37.8 kbp (for B12-E5). In SSV1-infected cells, the virus genome is present in both extrachromosomal and integrated forms and the former is considered to be essential for the excision of the latter (She et al., 2001a). Mapping of the paired-end reads revealed that not only SSV1 is present in both integrated and extrachromosomal forms, but the same is true for B12-E5 and BEU9-E1 (Fig. S1). No significant increase in sequencing coverage was observed for these iMGEs, suggesting that they are maintained in the cells at low copy number.

Sa. shibatae iMGE target tRNA genes for integration

Two types of recombinases, both belonging to the tyrosine recombinase superfamily, were reported for iMGE of Sulfolobales (She et al., 2004). The pNOB8-like integrase gene remains intact upon iMGE integration, because the att site is located adjacent to the integrase gene. By contrast, in the case of the SSV-like integrases, the att site is located within the integrase gene and upon integration, the gene is split into two separate gene fragments flanking the integrated element. The identified *Sa. shibatae* iMGEs encode both types of integrases, but all have integrated into the genome by recombining with the 3'-distal regions of diverse tRNA genes (Table 2). Although protein-coding genes and intergenic regions are also

known to serve as integration sites for archaeal viruses and plasmids (Krupovic et al., 2010a; Krupovic et al., 2010b; Kazlauskas et al., 2018; Krupovic et al., 2019), the tRNA genes are by far the most common att sites (Badel et al., 2021). Notably, we found that the SSV1 genome occupied not only the previously reported tRNA-Arg (CCG) gene (Reiter et al., 1989; Schleper et al., 1992), but was also integrated into a secondary target site, tRNA-Arg (TCG). Interestingly, in the latter case, SSV1 shared the target site with B12-E4 and the two elements were integrated in tandem. In vitro studies have shown that for efficient recombination, SSV1 integrase does not require nucleotide sequences flanking the att homology region itself (Muskhelishvili et al., 1993). Thus, the same tRNA gene, in principle, can accommodate multiple consecutive integrations by the same or different iMGEs as is indeed observed. Notably, despite recognizing the same tRNA gene on the host chromosome, SSV1 and B12-E4 encode SSV-like and pNOB8-like integrases, respectively, suggesting convergent evolution of the integration specificity in the two elements. Similarly, B12-E1 and B12-E2 also recognize overlapping att sites and are tandemly integrated into the same tRNA-Thr (GGT) gene (Fig. 5A, Table 2). Tandem integrations have been previously described for euryarchaeal and thaumarchaeal iMGEs, including viruses, plasmids and casposons (Krupovic and Bamford, 2008; Krupovic et al., 2019), and it has been suggested that recombination between such tandemly integrated elements promotes their modular genome evolution (Badel et al., 2021).

Diversity and evolution of *Sa. shibatae* iMGE

Based on the presence of signature genes, *Sa. shibatae* iMGEs can be grouped into three major classes: (i) viruses and virus-derived iMGEs; (ii) conjugative plasmids; (iii) cryptic non-conjugative iMGEs. Below we will describe the three iMGE classes in more detail.

Viruses and virus-derived elements

As mentioned above, the three *Sa. shibatae* strains serve as hosts for three distinct viruses (SSV1, SFV1 and SPV1) representing different families. Among these, only SSV1 was identified as a provirus, even though icosahedral portoglobovirus SPV1 also establishes a chronic infection in S38A and does not appear to lyse the host cell upon virion egress. This is consistent with the lack of the integrase gene in either SPV1 or filamentous lipothrixvirus SFV1 (Liu et al., 2017; Liu et al., 2018).

Strains S38A and BEU9 carry CRISPR spacers targeting SSV1 (5 in S38A and 1 in BEU9; Fig. 3B, Table S1) and are immune to this virus (Fig. S2). However, both strains carry nearly identical iMGEs, S38A-E3 and BEU9-E4, respectively, which exclusively encode genes found in fuselloviruses, including SSV1 (Fig. 6A). Comparative genomic analysis has shown that S38A-E3/BEU9-E4 have evolved from a fusellovirus ancestor through the loss of a block of nine genes, including those encoding the major and minor capsid proteins. Notably, the deletion has resulted in the fusion of the 5'-terminal region of a gene encoding for the putative host recognition protein VP4 (Quemin et al., 2015) and the 3'-terminal fragment of gene *b277*, producing a novel chimeric gene (Fig. 6A). Such recombination-based mechanism could play an important role in the emergence of new archaeal virus genes lacking homologs in sequence databases, which typically constitute ~75% of the archaeal virus genomes (Krupovic et al., 2018). Indeed, it has been previously noted that fusellovirus genomes are prone to recombination (Redder et al., 2009; Pauly et al., 2019; Zhang et al., 2020).

Interestingly, S38A and BEU9 CRISPR spacers target SSV1 in the structural module, suggesting that the deletion of the nine-gene block in S38A-E3/BEU9-E4 has been instigated by CRISPR targeting. We have previously shown that CRISPR targeting drives the evolution of archaeal virus genomes (Medvedeva et

al., 2019) and our present results further support this conclusion. The integrase and helicase genes of S38A-E3/BEU9-E4 are intact, which should allow this element to excise from the host genome, replicate and reintegrate into the tRNA-Arg (GCG) gene, but not to form virions for spread between the cells. Thus, S38A-E3/BEU9-E4 exists like an integrative plasmid, rather than a virus. We note that S38A-E3/BEU9-E4 presents one of the most explicit examples of an evolutionary transition between two types of MGEs, that is, between a virus and a plasmid, which appears to be a recurrent phenomenon underlying the diversification of the viral world (Kazlauskas et al., 2019).

Importantly, S38A-E3/BEU9-E4 emergence might not be an isolated case of virus-to-plasmid transition involving fuselloviruses. Notably, the replication protein (WP_012954641) of pSSVi (Guo and Huang, 2010), a satellite plasmid of *Sa. solfataricus* P2 which, in the presence of fuselloviruses SSV1 or SSV2, is packaged into spindle-shaped virus particles and spreads in the population (Wang et al., 2007), is closely related to the helicase of S38A-E3/BEU9-E4 (66.73% identity over 526 aa alignment). Thus, pSSVi might have also evolved from a fusellovirus by losing the structural module and hence its unexplained ability to hijack the fusellovirus particles might be an ancestral property of this MGE. We predict that upon coinfection with a fusellovirus, S38A-E3/BEU9-E4 might behave as a satellite, akin to pSSVi, and be packed into the viral particles.

Conjugative plasmids

Sa. shibatae B12 and BEU9 carry two novel conjugative plasmids, B12-E5 (pB12E5) and BEU9-E1 (pBEU9E1), both detected in extrachromosomal and integrated forms (Fig. S1). The two plasmids are only distantly related to each other or to the previously reported plasmids (Fig. 6B). Nevertheless, their gene contents are similar to those of other conjugative plasmids of Sulfolobales (Prangishvili et al., 1998; Stedman et al., 2000; Erauso et al., 2006). Both plasmids encode a number of diverse transcription regulators, pNOB8-like integrases, segregation proteins (ParA and/or ParB) as well as the components of conjugation machinery, namely, ATPases VirB4/TrbE and VirD4/TraG, and the transmembrane pore-forming subunit VirB6/TrbL (Fig. 6B). The plasmids also encode several putative toxins, including an mRNA interferase and a GCN5-related N-acetyl-transferase, which might be components of toxin-antitoxin systems and promote plasmid stabilization through the post-segregational killing mechanism. Notably, both plasmids carry a gene encoding a homolog of the fusellovirus SSV1 protein D63, which was proposed to function as an adaptor in macromolecular assembly (Kraft et al., 2004), suggesting a horizontal gene transfer between *Sa. shibatae* viruses and plasmids.

Cryptic non-conjugative plasmids

Most of the identified iMGEs belong to this category and likely represent medium-sized (7-14 kb), non-conjugative plasmids (Table 2, Table S4, Fig. 7A). Based on the encoded DNA replication proteins, these plasmids can be divided into those replicating by the rolling-circle mechanism (B12-E2, S38A-E2 and BEU9-E2) and pRN-like plasmids (B12-E1, B12-E3, B12-E4, S38A-E4), which encode characteristic primases-polymerases of the archaeo-eukaryotic primase superfamily (Lipps, 2009; Kazlauskas et al., 2018), whereas S38A-E1/BEU9-E3 do not encode recognizable replication proteins, except for a superfamily (SF) 1 helicase. The latter element lacks a gene for an identifiable integrase and hence might be inactive.

The plasmids and viruses which use rolling-circle mechanism of DNA replication are widespread in bacteria and archaea (Kazlauskas et al., 2019), but thus far have not been described in Crenarchaeota, including Sulfolobales. The replication mechanism used by these MGEs is initiated by the MGE-encoded

rolling-circle replication endonuclease (RCRE), which nicks the double-stranded replicative intermediate and attaches covalently to the 5' end of the nicked strand through the catalytic tyrosine residue, whereas the free 3' end serves as a primer for the synthesis of the new strand; once the full new circle is synthesized, RCRE catalyzes the ligation (and liberation) of the ssDNA circle. Based on the presence of two or one tyrosine residues in the catalytic motif 3, RCRE are broadly classified into superfamilies 1 and 2 (Koonin and Ilyina, 1993). The RCRE encoded by B12-E2/S38A-E2 (the two elements are 95% identical) and BEU9-E2 contain 1 and 2 tyrosine residues in motif 3 (Fig. 7B). However, closer examination of the B12-E2/S38A-E2 protein showed that the lack of the second tyrosine is due to truncation of the protein, which likely renders it inactive. Consistently, the integrase encoded by this MGE is inactivated by frame-shift mutations (Fig. 7A). By contrast, the RCRE encoded by BEU9-E2 appears to be intact, with all signature motifs being conserved (Fig. 7B). BLASTP searches showed that BEU9-E2 RCRE is most closely related to the previously uncharacterized homologs from iMGE present in the genomes of several crenarchaeal species as well as in Sulfolobus monocaudavirus 2, although the latter appears to be inactivated. More distant homologs, retrieved during the second PSI-BLAST iteration, are encoded by plasmids and viruses of Euryarchaeota, including haloarchaeal pleolipoviruses HRPV-1 and HRPV-2 (Senčilo et al., 2012), provirus of *Methanococcus voltae* A3 (Krupovic and Bamford, 2008), *Thermococcus prieurii* plasmid pTP2 (Gorlas et al., 2013) and diverse integrated elements. Collectively, these results extend the presence of the rolling-circle plasmids to Crenarchaeota.

The pRN-like plasmids, typified by the *Sa. islandicus* plasmid pRN1, encode large fusion proteins consisting of the N-terminal primase-polymerase of the archaeo-eukaryotic primase superfamily (PrimPol) and the C-terminal SF3 helicase domains (Lipps, 2011). Similar PrimPol-domain proteins are widespread in bacterial and archaeal MGEs, and comprise many different families (Kazlauskas et al., 2018). The PrimPol-helicase proteins encoded by the integrative *Sa. shibatae* plasmids display considerable sequence diversity and do not form a monophyletic clade in maximum likelihood phylogenetic analysis (Fig. 7C). Instead, the four plasmids cluster with different groups of pRN-like plasmids from different species of the order Sulfolobales. The B12-E1 plasmid forms a clade with the *Sa. islandicus* elements, including plasmid pXZ1 which, similar to B12-E1, encodes an integrase and has been experimentally demonstrated to integrate into a tRNA gene (Peng, 2008). By contrast, the protein encoded by B12-E4 is most closely related to the homologs of integrative elements from *Sulfurisphaera tokodaii* and *Sulfurisphaera ohwakuensis* TA-1, whereas B12-E3 encodes a divergent protein which forms a separate branch in the phylogenetic tree. Finally, S38A-E4 is at the base of the clade including *S. neozealandicus* plasmid pORA1 (NC_006906) and *Sa. solfataricus* plasmid pIT3 (NC_005907) as well as integrated elements of *Sa. islandicus*. Overall, this analysis shows that pRN-like elements of *Sa. shibatae* are very diverse and their evolution does not follow that of the host species.

Patterns of MGE distribution in Sulfolobales

To understand the patterns of iMGE distribution in Sulfolobales, we systematically assessed the presence/absence of iMGEs related to those of *Sa. shibatae* in other Sulfolobales members and overlaid this information with the CRISPR targeting data (Fig. 8). This analysis revealed a complex network of MGE distributions in different strains, which did not correspond to the phylogenetic or geographical distance between the strains harboring the corresponding MGE. For instance, BEU9-E4/S38A-E3 was not found in B12, but was instead present in *Sa. islandicus* strains Y.G.57.14 and L.S.2.15 isolated in the USA (Reno et al., 2009); BEU9-E2 was not found in either of the two other *Sa. shibatae* strains, but was identified in *Sa. islandicus* L.S.2.15 and *A. brierleyi* strains; B12-E4 was restricted to B12, but its relatives were detected in two species of the genus *Sulfurisphaera* (*S. tokodai* and *S. ohwakuensis*). The

mechanism underlying the transfer of non-conjugative and non-viral MGEs between different species remains unknown. One possibility is that the MGEs are exchanged through extracellular membrane vesicles. Indeed, it has been recently demonstrated that membrane vesicles produced by *Saccharolobus* species mediate intercellular transfer of plasmid DNA in the laboratory and similar vesicles were observed in environmental samples (Liu et al., 2021).

The patchy distribution of the iMGEs in *Sa. shibatae* strains could be to a considerable extent attributed to CRISPR targeting, that is, strains devoid of a particular element harbor CRISPR spacers matching this MGE. Furthermore, CRISPR targeting occurred across the genus boundary, with MGE from *Sa. shibatae* being matched by spacers from *S. tokodai* and *Sa. islandicus*, and vice versa (Fig. 8). Thus, CRISPR targeting might not be sufficient for host assignment of the Sulfolobales MGEs, at least, not when a precision above the family rank is sought. The same conclusion has been reached based on the analysis of the lytic rod-shaped viruses infecting *Saccharolobus*, *Metallosphaera* and *Acidianus* species (Baquero et al., 2020b). Nevertheless, analysis of the CRISPR spacers points to local adaptations of the hosts to their MGEs, with species from the same geographical location having more matching spacers against local viruses than species from remote locations (Table S1). Collectively, these data indicate that host switching by MGEs is relatively frequent in Sulfolobales and, to a large part, is driven by the CRISPR targeting.

EXPERIMENTAL PROCEDURES

Strains and their cultivation

Sa. shibatae strains B12 (Yeats et al., 1982), S38A (Liu et al., 2017) and BEU9 (Liu et al., 2018) were grown aerobically (120 rpm) at 76°C in the *Sulfolobus* growth medium as described previously (Zillig et al., 1993). Plaque assays with SSV1 on the lawns of *Sa. solfataricus* P2 (known host and positive control) and *Sa. shibatae* strains S38A and BEU9 were performed as described previously (Quemin et al., 2015).

Genome purification and sequencing

The DNA was extracted from exponentially growing cultures of *Sa. shibatae* strains B12, S38A and BEU9 as described previously (Zaparty et al., 2010). The DNA was sequenced in the framework of the Virus-X project (Aevarsson et al., 2021), using the Oxford Nanopore GridION Mk1 sequencer with FLO-MIN106D flow cells and SQK-RBK004 libraries, as well as the Illumina MiSeq and Nextera libraries. Both were carried out according to the manufacturer's instructions. Hybrid genome assemblies were done using Unicycler v0.4, which includes short-read assembly, contig-bridging with long reads and a final polishing step (Wick et al., 2017).

Identification of insertion sequences

IS elements were predicted and classified into families using the ISsaga platform (Varani et al., 2011). The 'probable false-positive' predicted by ISsaga were excluded from the final results. Exact coordinates for all identified IS elements are provided in Supporting Information
Table S3.

Mapping of the integration sites

The precise boundaries of integration were defined based on the presence of direct repeats corresponding to attachment sites or target site duplications. The direct and inverted repeats were searched for using Unipro UGENE (Okonechnikov et al., 2012). Whenever possible, additional validation of the MGE integration sites was obtained by comparing sequences of genomes containing the putative iMGEs with those of closely related genomes that do not contain such insertions using blastn algorithm and Mauve full-genome alignments. iMGE related to those of *Sa. shibatae* and integrated in the genomes of other members of the Sulfolobales were searched for using tblastx. The elements were considered related if the length of the alignment generated by tblastx was > 70% of the length of the corresponding *Sa. shibatae* iMGE.

Annotation of the MGE genes

For each analyzed gene, the functional annotations were assigned using the PSI-BLAST program with position specific scoring matrixes derived from arCOG alignments (Altschul et al., 1997; Makarova et al., 2015). To detect remote homology, additional searches were performed using PSI-BLAST (Altschul et al., 1997) against the non-redundant protein database at NCBI and HHpred against the PDB, CDD, SCOPe and Pfam databases available through the MPI Bioinformatics Toolkit (Gabler et al., 2020) (Table S4).

Phylogenetic analysis

For phylogenetic analysis, MCP and ATPase sequences from each (pro)virus were concatenated and aligned using MAFFT (Katoh et al., 2019). Poorly aligned (low information
content) positions were removed using trimal (Capella-Gutierrez et al., 2009) with the gap threshold of 0.2. The final alignment contained 1034 positions. The maximum likelihood phylogenetic tree was constructed using the IQ-tree program with the automatic selection of the best-fit substitution model for a

given alignment (Minh et al., 2020). The best model identified by IQ-tree was VT +F +I +G4. The tree was mid-point rooted for convenient visualization. The branch support was assessed using SH-aLRT implemented in IQ-tree.

Genome comparisons

The genomes of MGEs were compared and visualized using EasyFig v2.1 with tblastx algorithm (Sullivan et al., 2011). The complete genomes of the *Sa. shibatae* strains and *Sa. islandicus* were compared using progressive-Mauve with default parameters (Darling et al., 2010). Average nucleotide identity (ANI) and GC% content were calculated using OrthoANIu at https://www.ezbiocloud.net/tools/ani.

CRISPR-Cas prediction and protospacer search

Genome sequences of *Sa. shibatae* strains were analyzed by CRISPRCasFinder with default parameters (Couvin et al., 2018). Findings with evidence level 4 were considered as CRISPR arrays and corresponding spacer sequences were extracted. Spacers sequences were re-oriented with respect to proposed transcription direction. Protospacer search against the database of all isolated viruses and plasmids of *Sulfolobales* was performed by blastn algorithm with the following parameters: word_size 7, minimal e-value 0.001. The dataset of CRISPR spacers from hot springs has been described previously (Medvedeva et al., 2019).

arCOG assignment

For the arCOG assignment, the arCOG database (ar18) was downloaded from https://ftp.ncbi.nih.gov/pub/wolf/COGs/arCOG/tmp.ar18/. Proteins from *Sa. shibatae* were matched to the proteins from the ar18 database with BLASTP (E-value 1e-05, max_target_seqs 1) and the best hit for each *Sa. shibatae* protein sequence was recovered. All arCOGs annotated in the best hit were assigned to the corresponding *Sa. shibatae* proteins.

Accession numbers

The sequence data have been submitted to the GenBank database under the project number PRJNA716249 with the following accession numbers: *Sa. shibatae* B12 chromosome (CP077717), *Sa. shibatae* B12 plasmid pB12E5 (CP077716), *Sa. shibatae* S38A chromosome (CP077713), *Sa. shibatae* BEU9 chromosome (CP077715), *Sa. shibatae* BEU9 plasmid pBEU9E1 (CP077714).

ACKNOWLEDGMENTS

This work was supported by l'Agence Nationale de la Recherche (Grant ENVIRA, ANR-17-CE15-0005-01) and the Emergence(s) project MEMREMA from Ville de Paris (to M.K.); the European Union's Horizon 2020 research and innovation program under grant agreement 685778, project VIRUS X (to D.P. and J.K.). Y.L. was a recipient of the Pasteur-Roux-Cantarini Fellowship from Institut Pasteur. S.M. was partly supported by the Metchnikov fellowship from Campus France.

References

Aevarsson, A., Kaczorowska, A.K., Adalsteinsson, B.T., Ahlqvist, J., Al-Karadaghi, S., Altenbuchner, J. et al. (2021) Going to extremes - a metagenomic journey into the dark matter of life. *FEMS Microbiol Lett* **doi: 10.1093/femsle/fnab067**.

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**: 3389-3402.

Athukoralage, J.S., McMahon, S.A., Zhang, C., Gruschow, S., Graham, S., Krupovic, M. et al. (2020) An anti-CRISPR viral ring nuclease subverts type III CRISPR immunity. *Nature* **577**: 572-575.

Badel, C., Da Cunha, V., and Oberto, J. (2021) Archaeal tyrosine recombinases. *FEMS Microbiol Rev*.

Baquero, D.P., Liu, Y., Wang, F., Egelman, E.H., Prangishvili, D., and Krupovic, M. (2020a) Structure and assembly of archaeal viruses. *Adv Virus Res* **108**: 127-164.

Baquero, D.P., Contursi, P., Piochi, M., Bartolucci, S., Liu, Y., Cvirkaite-Krupovic, V. et al. (2020b) New virus isolates from Italian hydrothermal environments underscore the biogeographic pattern in archaeal virus communities. *ISME J* **14**: 1821-1833.

Bautista, M.A., Black, J.A., Youngblut, N.D., and Whitaker, R.J. (2017) Differentiation and structure in Sulfolobus islandicus rod-shaped virus populations. *Viruses* **9**: 120.

Bergerat, A., Gadelle, D., and Forterre, P. (1994) Purification of a DNA topoisomerase II from the hyperthermophilic archaeon Sulfolobus shibatae. A thermostable enzyme with both bacterial and eucaryal features. *J Biol Chem* **269**: 27663-27669.

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.

Blount, Z.D., and Grogan, D.W. (2005) New insertion sequences of Sulfolobus: functional properties and implications for genome evolution in hyperthermophilic archaea. *Mol Microbiol* **55**: 312-325.

Boyce, A., and Walsh, G. (2018) Expression and characterisation of a thermophilic endo-1,4-beta-glucanase from Sulfolobus shibatae of potential industrial application. *Mol Biol Rep* **45**: 2201-2211.

Brügger, K., Torarinsson, E., Redder, P., Chen, L., and Garrett, R.A. (2004) Shuffling of Sulfolobus genomes by autonomous and non-autonomous mobile elements. *Biochem Soc Trans* **32**: 179-183.

Capella-Gutierrez, S., Silla-Martinez, J.M., and Gabaldon, T. (2009) trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**: 1972-1973.

Clore, A.J., and Stedman, K.M. (2007) The SSV1 viral integrase is not essential. *Virology* **361**: 103-111.

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

Counts, J.A., Willard, D.J., and Kelly, R.M. (2020) Life in hot acid: a genome-based reassessment of the archaeal order Sulfolobales. *Environ Microbiol*.

Couvin, D., Bernheim, A., Toffano-Nioche, C., Touchon, M., Michalik, J., Neron, B. et al. (2018) CRISPRCasFinder, an update of CRISRFinder, includes a portable version, enhanced performance and integrates search for Cas proteins. *Nucleic Acids Res* **46**: W246-W251.

Dai, X., Wang, H., Zhang, Z., Li, K., Zhang, X., Mora-Lopez, M. et al. (2016) Genome Sequencing of Sulfolobus sp. A20 from Costa Rica and Comparative Analyses of the Putative Pathways of Carbon, Nitrogen, and Sulfur Metabolism in Various Sulfolobus Strains. *Front Microbiol* **7**: 1902.

Darling, A.E., Mau, B., and Perna, N.T. (2010) progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. *PLoS One* **5**: e11147.

Deng, L., Garrett, R.A., Shah, S.A., Peng, X., and She, Q. (2013) A novel interference mechanism by a type IIIB CRISPR-Cmr module in Sulfolobus. *Mol Microbiol* **87**: 1088-1099.

Erauso, G., Stedman, K.M., van de Werken, H.J.G., Zillig, W., and van der Oost, J. (2006) Two novel conjugative plasmids from a single strain of Sulfolobus. *Microbiology (Reading)* **152**: 1951-1968.

Fusco, S., Aulitto, M., Bartolucci, S., and Contursi, P. (2015) A standardized protocol for the UV induction of Sulfolobus spindle-shaped virus 1. *Extremophiles* **19**: 539-546.

Gabler, F., Nam, S.Z., Till, S., Mirdita, M., Steinegger, M., Soding, J. et al. (2020) Protein Sequence Analysis Using the MPI Bioinformatics Toolkit. *Curr Protoc Bioinformatics* **72**: e108.

Goodman, D.A., and Stedman, K.M. (2018) Comparative genetic and genomic analysis of the novel fusellovirus Sulfolobus spindle-shaped virus 10. *Virus Evol* **4**: vey022.

Gorlas, A., Krupovic, M., Forterre, P., and Geslin, C. (2013) Living side by side with a virus: characterization of two novel plasmids from Thermococcus prieurii, a host for the spindle-shaped virus TPV1. *Appl Environ Microbiol* **79**: 3822-3828.

Greening, C., Constant, P., Hards, K., Morales, S.E., Oakeshott, J.G., Russell, R.J. et al. (2015) Atmospheric hydrogen scavenging: from enzymes to ecosystems. *Appl Environ Microbiol* **81**: 1190-1199.

Grogan, D., Palm, P., and Zillig, W. (1990) Isolate B12, which harbours a virus-like element, represents a new species of the archaebacterial genus Sulfolobus, Sulfolobus shibatae, sp. nov. *Arch Microbiol* **154**: 594-599.

Guo, L., Brugger, K., Liu, C., Shah, S.A., Zheng, H., Zhu, Y. et al. (2011) Genome analyses of Icelandic strains of Sulfolobus islandicus, model organisms for genetic and virus-host interaction studies. *J Bacteriol* **193**: 1672-1680.

Guo, X., and Huang, L. (2010) A superfamily 3 DNA helicase encoded by plasmid pSSVi from the hyperthermophilic archaeon Sulfolobus solfataricus unwinds DNA as a higher-order oligomer and interacts with host primase. *J Bacteriol* **192**: 1853-1864.

He, F., Bhoobalan-Chitty, Y., Van, L.B., Kjeldsen, A.L., Dedola, M., Makarova, K.S. et al. (2018) Anti-CRISPR proteins encoded by archaeal lytic viruses inhibit subtype I-D immunity. *Nat Microbiol* **3**: 461-469.

Jaubert, C., Danioux, C., Oberto, J., Cortez, D., Bize, A., Krupovic, M. et al. (2013) Genomics and genetics of Sulfolobus islandicus LAL14/1, a model hyperthermophilic archaeon. *Open Biol* **3**: 130010.

Jaxel, C., Bouthier de la Tour, C., Duguet, M., and Nadal, M. (1996) Reverse gyrase gene from Sulfolobus shibatae B12: gene structure, transcription unit and comparative sequence analysis of the two domains. *Nucleic Acids Res* **24**: 4668-4675.

Jonuscheit, M., Martusewitsch, E., Stedman, K.M., and Schleper, C. (2003) A reporter gene system for the hyperthermophilic archaeon Sulfolobus solfataricus based on a selectable and integrative shuttle vector. *Mol Microbiol* **48**: 1241-1252.

Katoh, K., Rozewicki, J., and Yamada, K.D. (2019) MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Brief Bioinform* **20**: 1160-1166.

Kazlauskas, D., Varsani, A., Koonin, E.V., and Krupovic, M. (2019) Multiple origins of prokaryotic and eukaryotic single-stranded DNA viruses from bacterial and archaeal plasmids. *Nat Commun* **10**: 3425.

Kazlauskas, D., Sezonov, G., Charpin, N., Venclovas, C., Forterre, P., and Krupovic, M. (2018) Novel families of archaeo-eukaryotic primases associated with mobile genetic elements of Bacteria and Archaea. *J Mol Biol* **430**: 737-750.

Koonin, E.V., and Ilyina, T.V. (1993) Computer-assisted dissection of rolling circle DNA replication. *Biosystems* **30**: 241-268.

Kozubal, M.A., Macur, R.E., Jay, Z.J., Beam, J.P., Malfatti, S.A., Tringe, S.G. et al. (2012) Microbial iron cycling in acidic geothermal springs of yellowstone national park: integrating molecular surveys, geochemical processes, and isolation of novel fe-active microorganisms. *Front Microbiol* **3**: 109.

Kraft, P., Kummel, D., Oeckinghaus, A., Gauss, G.H., Wiedenheft, B., Young, M., and Lawrence, C.M. (2004) Structure of D-63 from sulfolobus spindle-shaped virus 1: surface properties of the dimeric four-helix bundle suggest an adaptor protein function. *J Virol* **78**: 7438-7442.

Krupovic, M., and Bamford, D.H. (2008) Archaeal proviruses TKV4 and MVV extend the PRD1-adenovirus lineage to the phylum Euryarchaeota. *Virology* **375**: 292-300.

Krupovic, M., Forterre, P., and Bamford, D.H. (2010a) 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. *J Mol Biol* **397**: 144-160.

Krupovic, M., Gribaldo, S., Bamford, D.H., and Forterre, P. (2010b) The evolutionary history of archaeal MCM helicases: a case study of vertical evolution combined with hitchhiking of mobile genetic elements. *Mol Biol Evol* **27**: 2716-2732.

Krupovic, M., Quemin, E.R., Bamford, D.H., Forterre, P., and Prangishvili, D. (2014) Unification of the globally distributed spindle-shaped viruses of the Archaea. *J Virol* **88**: 2354-2358.

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

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. *Environ Microbiol* **21**: 2056-2078.
Kvist, T., Ahring, B.K., and Westermann, P. (2007) Archaeal diversity in Icelandic hot springs. *FEMS Microbiol Ecol* **59**: 71-80.
Laska, S., Lottspeich, F., and Kletzin, A. (2003) Membrane-bound hydrogenase and sulfur reductase of the hyperthermophilic and acidophilic archaeon Acidianus ambivalens. *Microbiology (Reading)* **149**: 2357-2371.
Lawrence, C.M., Menon, S., Eilers, B.J., Bothner, B., Khayat, R., Douglas, T., and Young, M.J. (2009) Structural and functional studies of archaeal viruses. *J Biol Chem* **284**: 12599-12603.
Lewis, A.M., Recalde, A., Brasen, C., Counts, J.A., Nussbaum, P., Bost, J. et al. (2021) The biology of thermoacidophilic archaea from the order Sulfolobales. *FEMS Microbiol Rev*.
Lipps, G. (2006) Plasmids and viruses of the thermoacidophilic crenarchaeote Sulfolobus. *Extremophiles* **10**: 17-28.
Lipps, G. (2009) Molecular biology of the pRN1 plasmid from Sulfolobus islandicus. *Biochem Soc Trans* **37**: 42-45.
Lipps, G. (2011) Structure and function of the primase domain of the replication protein from the archaeal plasmid pRN1. *Biochem Soc Trans* **39**: 104-106.
Liu, J., Cvirkaite-Krupovic, V., Commere, P.H., Yang, Y., Zhou, F., Forterre, P. et al. (2021) Archaeal extracellular vesicles are produced in an ESCRT-dependent manner and promote gene transfer and nutrient cycling in extreme environments. *ISME J* **doi: 10.1038/s41396-021-00984-0**.
Liu, Y., Ishino, S., Ishino, Y., Pehau-Arnaudet, G., Krupovic, M., and Prangishvili, D. (2017) A novel type of polyhedral viruses infecting hyperthermophilic archaea. *J Virol* **91**: e00589-00517.
Liu, Y., Osinski, T., Wang, F., Krupovic, M., Schouten, S., Kasson, P. et al. (2018) Structural conservation in a membrane-enveloped filamentous virus infecting a hyperthermophilic acidophile. *Nat Commun* **9**: 3360.
Liu, Y., Brandt, D., Ishino, S., Ishino, Y., Koonin, E.V., Kalinowski, J. et al. (2019) New archaeal viruses discovered by metagenomic analysis of viral communities in enrichment cultures. *Environ Microbiol* **21**: 2002-2014.
Lopatina, A., Tal, N., and Sorek, R. (2020) Abortive Infection: Bacterial Suicide as an Antiviral Immune Strategy. *Annu Rev Virol* **7**: 371-384.
Makarova, K.S., Wolf, Y.I., and Koonin, E.V. (2013) Comparative genomics of defense systems in archaea and bacteria. *Nucleic Acids Res* **41**: 4360-4377.
Makarova, K.S., Wolf, Y.I., and Koonin, E.V. (2015) Archaeal Clusters of Orthologous Genes (arCOGs): An Update and Application for Analysis of Shared Features between Thermococcales, Methanococcales, and Methanobacteriales. *Life (Basel)* **5**: 818-840.
Martin, A., Yeats, S., Janekovic, D., Reiter, W.D., Aicher, W., and Zillig, W. (1984) SAV 1, a temperate u.v.-inducible DNA virus-like particle from the archaebacterium Sulfolobus acidocaldarius isolate B12. *EMBO J* **3**: 2165-2168.
Martusewitsch, E., Sensen, C.W., and Schleper, C. (2000) High spontaneous mutation rate in the hyperthermophilic archaeon Sulfolobus solfataricus is mediated by transposable elements. *J Bacteriol* **182**: 2574-2581.
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. *Nat Commun* **10**: 5204.
Minh, B.Q., Schmidt, H.A., Chernomor, O., Schrempf, D., Woodhams, M.D., von Haeseler, A., and Lanfear, R. (2020) IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Mol Biol Evol* **37**: 1530-1534.
Munson-McGee, J.H., Snyder, J.C., and Young, M.J. (2018a) Archaeal viruses from high-temperature environments. *Genes (Basel)* **9**: 128.
Munson-McGee, J.H., Peng, S., Dewerff, S., Stepanauskas, R., Whitaker, R.J., Weitz, J.S., and Young, M.J. (2018b) A virus or more in (nearly) every cell: ubiquitous networks of virus-host interactions in extreme environments. *ISME J* **12**: 1706-1714.
Muskhelishvili, G., Palm, P., and Zillig, W. (1993) SSV1-encoded site-specific recombination system in Sulfolobus shibatae. *Mol Gen Genet* **237**: 334-342.

Okonechnikov, K., Golosova, O., Fursov, M., and team, U. (2012) Unipro UGENE: a unified bioinformatics toolkit. *Bioinformatics* **28**: 1166-1167.
Palm, P., Schleper, C., Grampp, B., Yeats, S., McWilliam, P., Reiter, W.D., and Zillig, W. (1991) Complete nucleotide sequence of the virus SSV1 of the archaebacterium Sulfolobus shibatae. *Virology* **185**: 242-250.
Pauly, M.D., Bautista, M.A., Black, J.A., and Whitaker, R.J. (2019) Diversified local CRISPR-Cas immunity to viruses of Sulfolobus islandicus. *Philos Trans R Soc Lond B Biol Sci* **374**: 20180093.
Peng, X. (2008) Evidence for the horizontal transfer of an integrase gene from a fusellovirus to a pRN-like plasmid within a single strain of Sulfolobus and the implications for plasmid survival. *Microbiology (Reading)* **154**: 383-391.
Prangishvili, D., Bamford, D.H., Forterre, P., Iranzo, J., Koonin, E.V., and Krupovic, M. (2017) The enigmatic archaeal virosphere. *Nat Rev Microbiol* **15**: 724-739.
Prangishvili, D., Albers, S.V., Holz, I., Arnold, H.P., Stedman, K., Klein, T. et al. (1998) Conjugation in archaea: frequent occurrence of conjugative plasmids in Sulfolobus. *Plasmid* **40**: 190-202.
Quax, T.E., Voet, M., Sismeiro, O., Dillies, M.A., Jagla, B., Coppee, J.Y. et al. (2013) Massive activation of archaeal defense genes during viral infection. *J Virol* **87**: 8419-8428.
Quemin, E.R., Chlanda, P., Sachse, M., Forterre, P., Prangishvili, D., and Krupovic, M. (2016) Eukaryotic-like virus budding in Archaea. *mBio* **7**: e01439-01416.
Quemin, E.R., Pietila, M.K., Oksanen, H.M., Forterre, P., Rijpstra, W.I., Schouten, S. et al. (2015) Sulfolobus spindle-shaped virus 1 contains glycosylated capsid proteins, a cellular chromatin protein, and host-derived lipids. *J Virol* **89**: 11681-11691.
Qureshi, S.A., and Jackson, S.P. (1998) Sequence-specific DNA binding by the S. shibatae TFIIB homolog, TFB, and its effect on promoter strength. *Mol Cell* **1**: 389-400.
Qureshi, S.A., Bell, S.D., and Jackson, S.P. (1997) Factor requirements for transcription in the Archaeon Sulfolobus shibatae. *EMBO J* **16**: 2927-2936.
Redder, P., and Garrett, R.A. (2006) Mutations and rearrangements in the genome of Sulfolobus solfataricus P2. *J Bacteriol* **188**: 4198-4206.
Redder, P., She, Q., and Garrett, R.A. (2001) Non-autonomous mobile elements in the crenarchaeon Sulfolobus solfataricus. *J Mol Biol* **306**: 1-6.
Redder, P., Peng, X., Brugger, K., Shah, S.A., Roesch, F., Greve, B. et al. (2009) Four newly isolated fuselloviruses from extreme geothermal environments reveal unusual morphologies and a possible interviral recombination mechanism. *Environ Microbiol* **11**: 2849-2862.
Reiter, W.D., Palm, P., and Yeats, S. (1989) Transfer RNA genes frequently serve as integration sites for prokaryotic genetic elements. *Nucleic Acids Res* **17**: 1907-1914.
Reiter, W.D., Palm, P., Yeats, S., and Zillig, W. (1987) Gene expression in archaebacteria: physical mapping of constitutive and UV-inducible transcripts from the Sulfolobus virus-like particle SSV1. *Mol Gen Genet* **209**: 270-275.
Reno, M.L., Held, N.L., Fields, C.J., Burke, P.V., and Whitaker, R.J. (2009) Biogeography of the Sulfolobus islandicus pan-genome. *Proc Natl Acad Sci U S A* **106**: 8605-8610.
Ribardo, D.A., and Hendrixson, D.R. (2011) Analysis of the LIV system of Campylobacter jejuni reveals alternative roles for LivJ and LivK in commensalism beyond branched-chain amino acid transport. *J Bacteriol* **193**: 6233-6243.
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. *Mol Microbiol* **113**: 718-727.
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. *Proc Natl Acad Sci U S A* **89**: 7645-7649.
Senčilo, A., Paulin, L., Kellner, S., Helm, M., and Roine, E. (2012) Related haloarchaeal pleomorphic viruses contain different genome types. *Nucleic Acids Res* **40**: 5523-5534.
She, Q., Shen, B., and Chen, L. (2004) Archaeal integrases and mechanisms of gene capture. *Biochem Soc Trans* **32**: 222-226.
She, Q., Peng, X., Zillig, W., and Garrett, R.A. (2001a) Gene capture in archaeal chromosomes. *Nature* **409**: 478.

She, Q., Singh, R.K., Confalonieri, F., Zivanovic, Y., Allard, G., Awayez, M.J. et al. (2001b) The complete genome of the crenarchaeon Sulfolobus solfataricus P2. *Proc Natl Acad Sci U S A* **98**: 7835-7840.
Siguier, P., Gourbeyre, E., and Chandler, M. (2014) Bacterial insertion sequences: their genomic impact and diversity. *FEMS Microbiol Rev* **38**: 865-891.
Søndergaard, D., Pedersen, C.N., and Greening, C. (2016) HydDB: A web tool for hydrogenase classification and analysis. *Sci Rep* **6**: 34212.
Stedman, K.M., Schleper, C., Rumpf, E., and Zillig, W. (1999) Genetic requirements for the function of the archaeal virus SSV1 in Sulfolobus solfataricus: construction and testing of viral shuttle vectors. *Genetics* **152**: 1397-1405.
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.
Stedman, K.M., She, Q., Phan, H., Holz, I., Singh, H., Prangishvili, D. et al. (2000) pING family of conjugative plasmids from the extremely thermophilic archaeon Sulfolobus islandicus: insights into recombination and conjugation in Crenarchaeota. *J Bacteriol* **182**: 7014-7020.
Sullivan, M.J., Petty, N.K., and Beatson, S.A. (2011) Easyfig: a genome comparison visualizer. *Bioinformatics* **27**: 1009-1010.
Van, T.T., Ryu, S.I., Lee, K.J., Kim, E.J., and Lee, S.B. (2007) Cloning and characterization of glycogen-debranching enzyme from hyperthermophilic archaeon Sulfolobus shibatae. *J Microbiol Biotechnol* **17**: 792-799.
Varani, A.M., Siguier, P., Gourbeyre, E., Charneau, V., and Chandler, M. (2011) ISsaga is an ensemble of web-based methods for high throughput identification and semi-automatic annotation of insertion sequences in prokaryotic genomes. *Genome Biol* **12**: R30.
Wang, F., Baquero, D.P., Su, Z., Beltran, L.C., Prangishvili, D., Krupovic, M., and Egelman, E.H. (2020a) The structures of two archaeal type IV pili illuminate evolutionary relationships. *Nat Commun* **11**: 3424.
Wang, F., Baquero, D.P., Beltran, L.C., Su, Z., Osinski, T., Zheng, W. et al. (2020b) Structures of filamentous viruses infecting hyperthermophilic archaea explain DNA stabilization in extreme environments. *Proc Natl Acad Sci U S A* **117**: 19643-19652.
Wang, F., Liu, Y., Su, Z., Osinski, T., de Oliveira, G.A.P., Conway, J.F. et al. (2019a) A packing for A-form DNA in an icosahedral virus. *Proc Natl Acad Sci U S A* **116**: 22591-22597.
Wang, F., Cvirkaite-Krupovic, V., Kreutzberger, M.A.B., Su, Z., de Oliveira, G.A.P., Osinski, T. et al. (2019b) An extensively glycosylated archaeal pilus survives extreme conditions. *Nat Microbiol* **4**: 1401-1410.
Wang, H., Peng, N., Shah, S.A., Huang, L., and She, Q. (2015) Archaeal extrachromosomal genetic elements. *Microbiol Mol Biol Rev* **79**: 117-152.
Wang, Y., Duan, Z., Zhu, H., Guo, X., Wang, Z., Zhou, J. et al. (2007) A novel Sulfolobus non-conjugative extrachromosomal genetic element capable of integration into the host genome and spreading in the presence of a fusellovirus. *Virology* **363**: 124-133.
Wick, R.R., Judd, L.M., Gorrie, C.L., and Holt, K.E. (2017) Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads. *PLoS Comput Biol* **13**: e1005595.
Wiedenheft, B., Stedman, K., Roberto, F., Willits, D., Gleske, A.K., Zoeller, L. et al. (2004) Comparative genomic analysis of hyperthermophilic archaeal Fuselloviridae viruses. *J Virol* **78**: 1954-1961.
Wojtas, M.N., Mogni, M., Millet, O., Bell, S.D., and Abrescia, N.G. (2012) Structural and functional analyses of the interaction of archaeal RNA polymerase with DNA. *Nucleic Acids Res* **40**: 9941-9952.
Yeats, S., McWilliam, P., and Zillig, W. (1982) A plasmid in the archaebacterium Sulfolobus acidocaldarius. *EMBO J* **1**: 1035-1038.
Yu, Z., Jiang, S., Wang, Y., Tian, X., Zhao, P., Xu, J. et al. (2021) CRISPR-Cas adaptive immune systems in Sulfolobales: genetic studies and molecular mechanisms. *Sci China Life Sci* **64**: 678-696.
Zaparty, M., Esser, D., Gertig, S., Haferkamp, P., Kouril, T., Manica, A. et al. (2010) "Hot standards" for the thermoacidophilic archaeon Sulfolobus solfataricus. *Extremophiles* **14**: 119-142.
Zebec, Z., Manica, A., Zhang, J., White, M.F., and Schleper, C. (2014) CRISPR-mediated targeted mRNA degradation in the archaeon Sulfolobus solfataricus. *Nucleic Acids Res* **42**: 5280-5288.
Zhang, C., Phillips, A.P.R., Wipfler, R.L., Olsen, G.J., and Whitaker, R.J. (2018) The essential genome of the crenarchaeal model Sulfolobus islandicus. *Nat Commun* **9**: 4908.

Zhang, J., Zheng, X., Wang, H., Jiang, H., Dong, H., and Huang, L. (2020) Novel Sulfolobus fuselloviruses with extensive genomic variations. *J Virol* **94**: e01624-01619.
Zhang, X., and Zheng, Q.C. (2018) Exploring the influence of hyperthermophilic protein Ssh10b on the stability and conformation of RNA by molecular dynamics simulation. *Biopolymers* **109**: e23068.
Zillig, W., Kletzin, A., Schleper, C., Holz, I., Janekovic, D., Hain, J. et al. (1993) Screening for Sulfolobales, their plasmids and their viruses in Icelandic solfataras. *Syst Appl Microbiol* **16**: 609-628.
Zillig, W., Arnold, H.P., Holz, I., Prangishvili, D., Schweier, A., Stedman, K. et al. (1998) Genetic elements in the extremely thermophilic archaeon Sulfolobus. *Extremophiles* **2**: 131-140.

Table 1. Toxin-antitoxin systems of *Sa. shibatae*.

Toxin	Antitoxin	B12	S38A	BEU9
RelE (2 arCOGs)	?	4	5	5
PIN domain (7 arCOGs)	?	6	8	9
PIN domain (8 arCOGs)	AbrB (8 arCOGs)	8	8	9
PIN domain (5 arCOGs)	RHH/CopG (7 arCOGs)	7	8	8
HEPN (6 arCOGs)	MNT (7 arCOGs)	16	17	13

Table 2. Integrated mobile genetic elements of *Sa. shibatae*.

Element name	Nucleotide coordinates	Length (bp)	Integration site	att length	MGE type	Replication protein
B12-E1	635422..649821	14400	tRNA-Thr (GGT)	48 bp	Cryptic plasmid	PrimPol-SF3 helicase
B12-E2	649777..660711	10935	tRNA-Thr (GGT)	43 bp	Cryptic plasmid	RCRE
B12-E3	2221911..2229307	7397	tRNA-Ala (CGC)	46 bp	Cryptic plasmid	PrimPol-SF3 helicase
B12-E4	2405111..2413149	8039	tRNA-Arg (TCG)	43 bp	Cryptic plasmid	PrimPol-SF3 helicase
B12-E5* (pB12E5)	2612035..2649874	37840	tRNA-Glu (TTC)	67 bp	Conjugative plasmid	Replication protein A
SSV1*	2839680..2855188	15508	tRNA-Arg (CCG/TCG)	44 bp	Fusellovirus	DnaA-like ATPase
BEU9-E1* (pBEU9E1)	45765..83514	37749	tRNA-His (GTG)	49 bp	Conjugative plasmid	Replication protein A
BEU9-E2	141911..148661	6750	tRNA-Val (CAC)	43 bp	pNOB8-like	RCRE
BEU9-E3	186480..200539	14059	tRNA-Trp (CCA)	30 bp	incomplete	SF1 helicase?
BEU9-E4	2499412..2506668	7256	tRNA-Arg (GCG)	45 bp	virus-derived	helicase
S38A-E1	142104..154888	12784	tRNA-Trp (CCA)	30 bp	incomplete	SF1 helicase?
S38A-E2	648710..659573	10863	tRNA-Thr (GGT)	51 bp	incomplete	RCRE
S38A-E3	2539711..2546966	7255	tRNA-Arg (GCG)	45 bp	virus-derived	helicase
S38A-E4	2775115..2785077	9962	tRNA-Val (GAC)	52 bp	pNOB8-like	PrimPol-SF3 helicase

* – detected as extrachromosomal elements. RCRE, rolling circle replication endonuclease; SF, superfamily

FIGURES AND LEGENDS

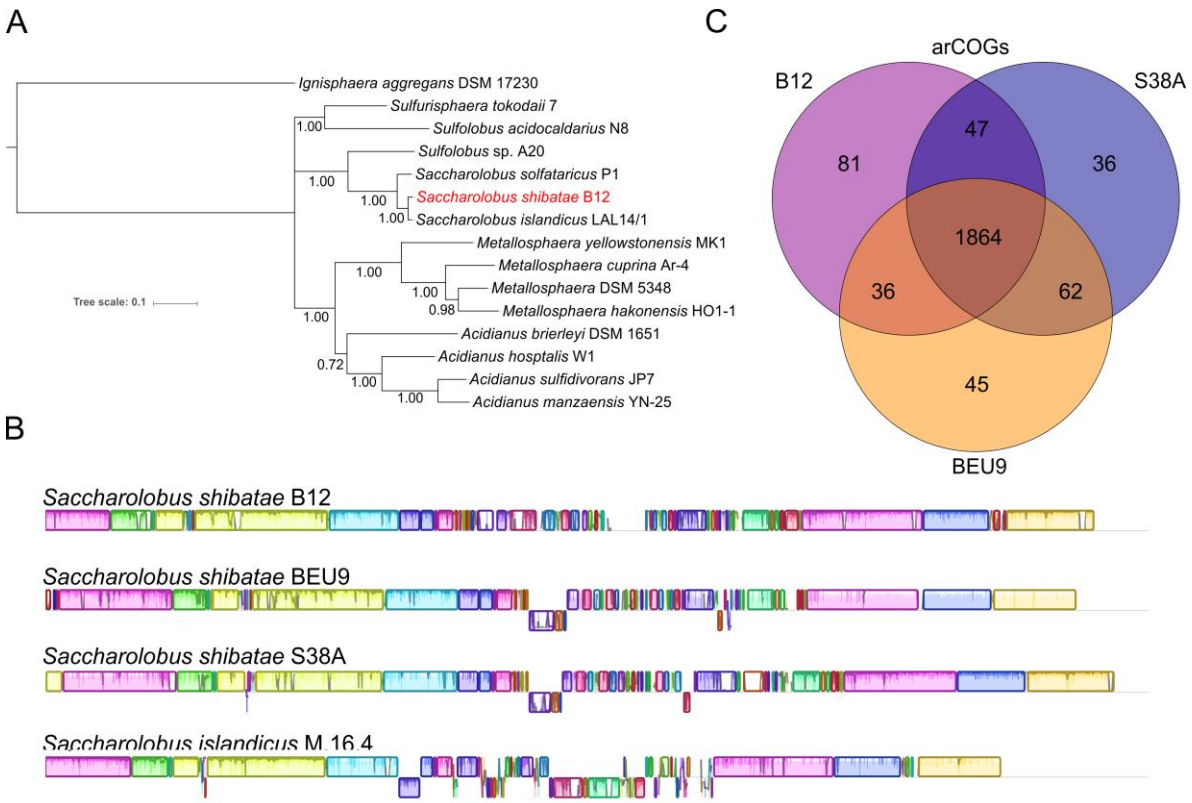

Figure 1. Genomic characteristics of *Sa. shibatae*. A. Phylogenetic relationships between different members of the order Sulfolobales. Maximum-likelihood phylogenetic tree is based on the concatenated alignment of ribosomal proteins and the B subunit of RNA polymerase. Numbers at the nodes represent SH-aLRT support values. Branches with support values below 0.7 were collapsed. The scale bar represents the number of substitutions per site. B. Venn diagram showing the comparison of the arCOG contents of the three *Sa. shibatae* genomes. C. Genomic maps of the three *Sa. shibatae* strains and *Sa. islandicus* M.16.4. The comparison was generated using Mauve (Darling et al., 2010). Homologous regions in the compared genomes are depicted as blocks of the same color.

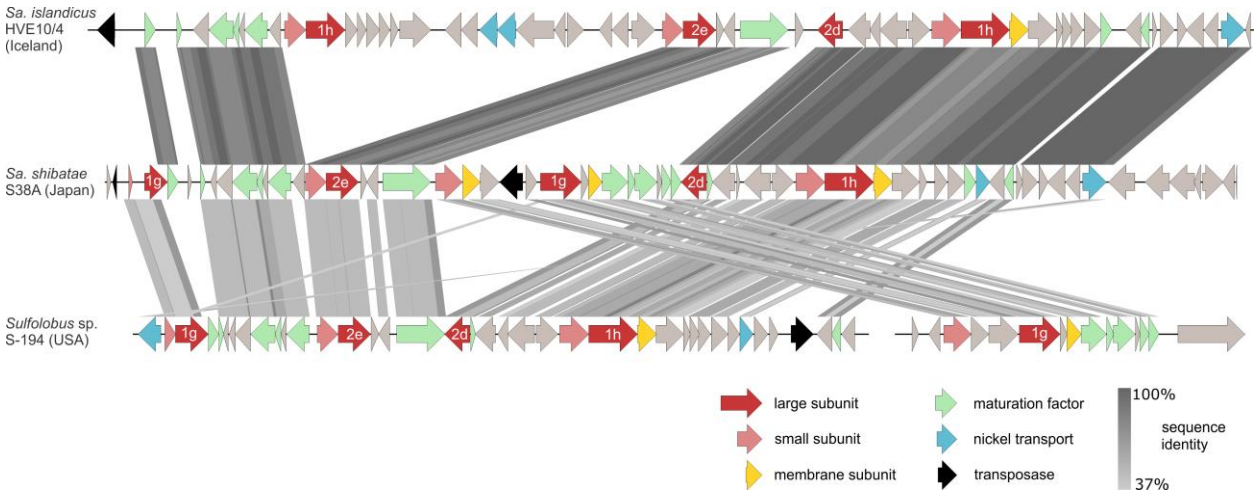

Figure 2. A locus of *Sa. shibatae* S38A genome containing the [NiFe]-hydrogenase clusters. The S38A locus is aligned with the corresponding loci in *Sa. islandicus* HVE10/4 and *Sulfolobus* sp. S-194. Genes constituting the [NiFe]-hydrogenase clusters are color coded with the key provided on the right.

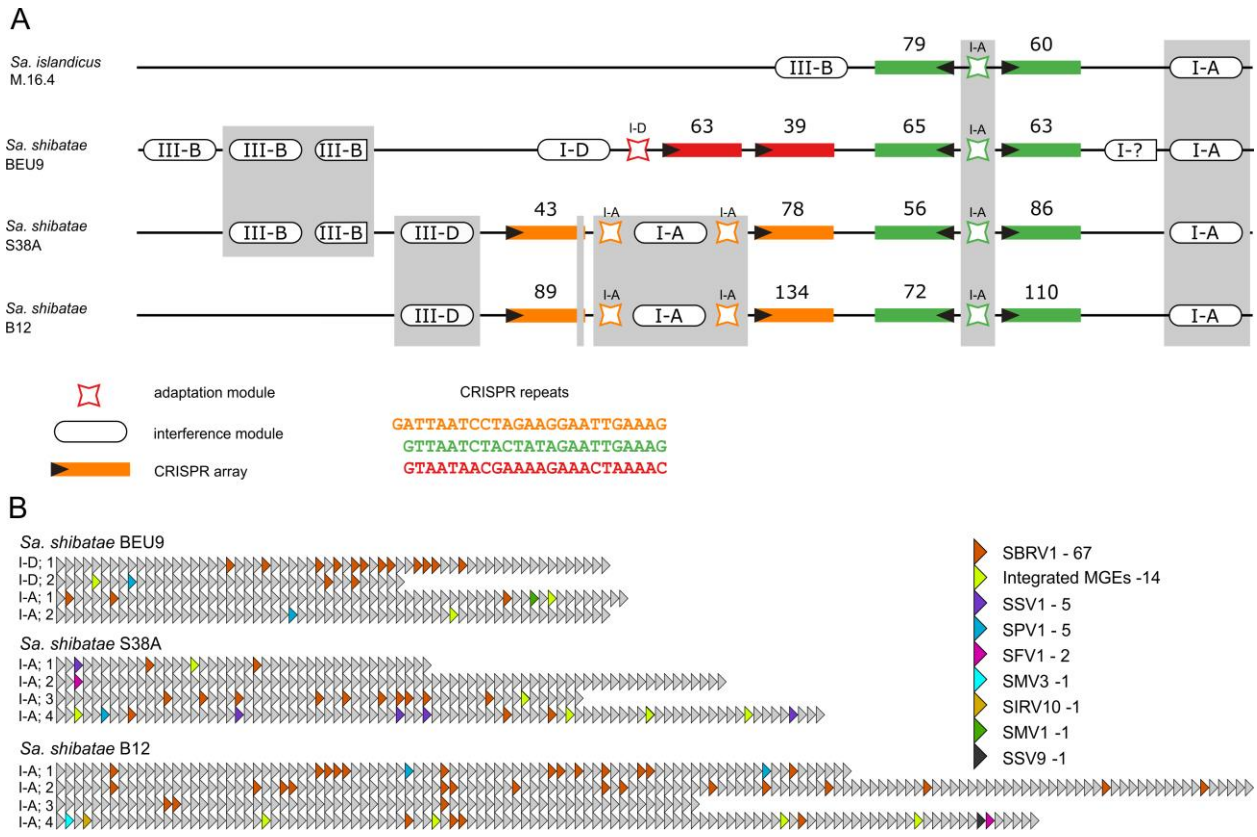

Figure 3. The diversity of CRISPR-Cas systems in *Sa. shibatae* strains. A. Schematic representation of the composition and arrangement of the CRISPR-Cas systems in the three *Sa. shibatae* strains. *Sa. islandicus* M.16.4 is shown for comparison. Numbers above the colored boxes indicate the number of spacers in the corresponding CRISPR arrays, whereas black triangles denote the leader-proximal end of the array. The sequences of the CRISPR repeats characteristic of the three types of arrays are colored similarly and are shown at the bottom of the panel. B. CRISPR arrays. Triangles correspond to individual spacers, with those colored matching diverse viruses and integrated elements (listed on the right).

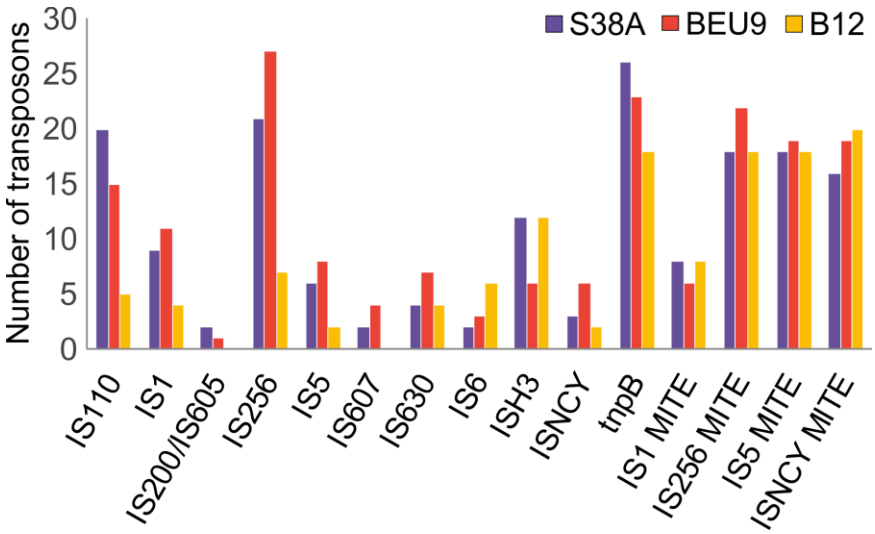

Figure 4. Diversity and distribution of insertion sequences in *Sa. shibatae* strains. MITE, miniature inverted repeat transposable elements.

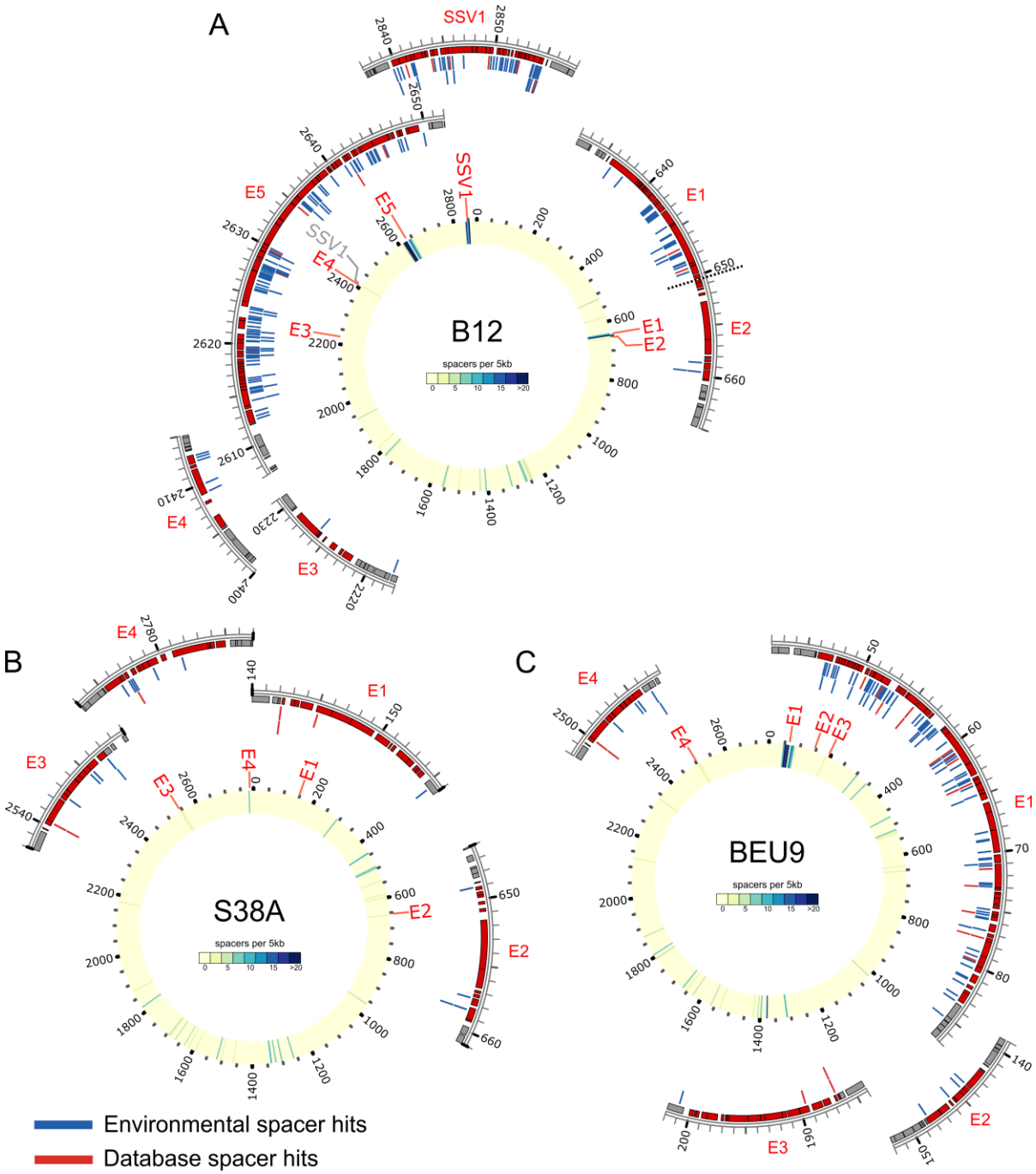

Figure 5. Integrated mobile genetic elements of *Sa. shibatae*. The circular genomes of the three *Sa. shibatae* strains are depicted as yellow rings with the distribution of CRISPR spacers depicted as vertical bars colored according to their density in the genome (scale is provided in the center of each genome map). The location of integrated mobile genetic elements is indicated with red bars and the zoom-in on the distribution of CRISPR spacers targeting the corresponding elements is shown in the outer circle, with the spacers originating from the environmental dataset (Medvedeva et al., 2019) and those extracted from genome sequences shown in blue and red, respectively. The secondary integration site of SSV1 is shown in grey.

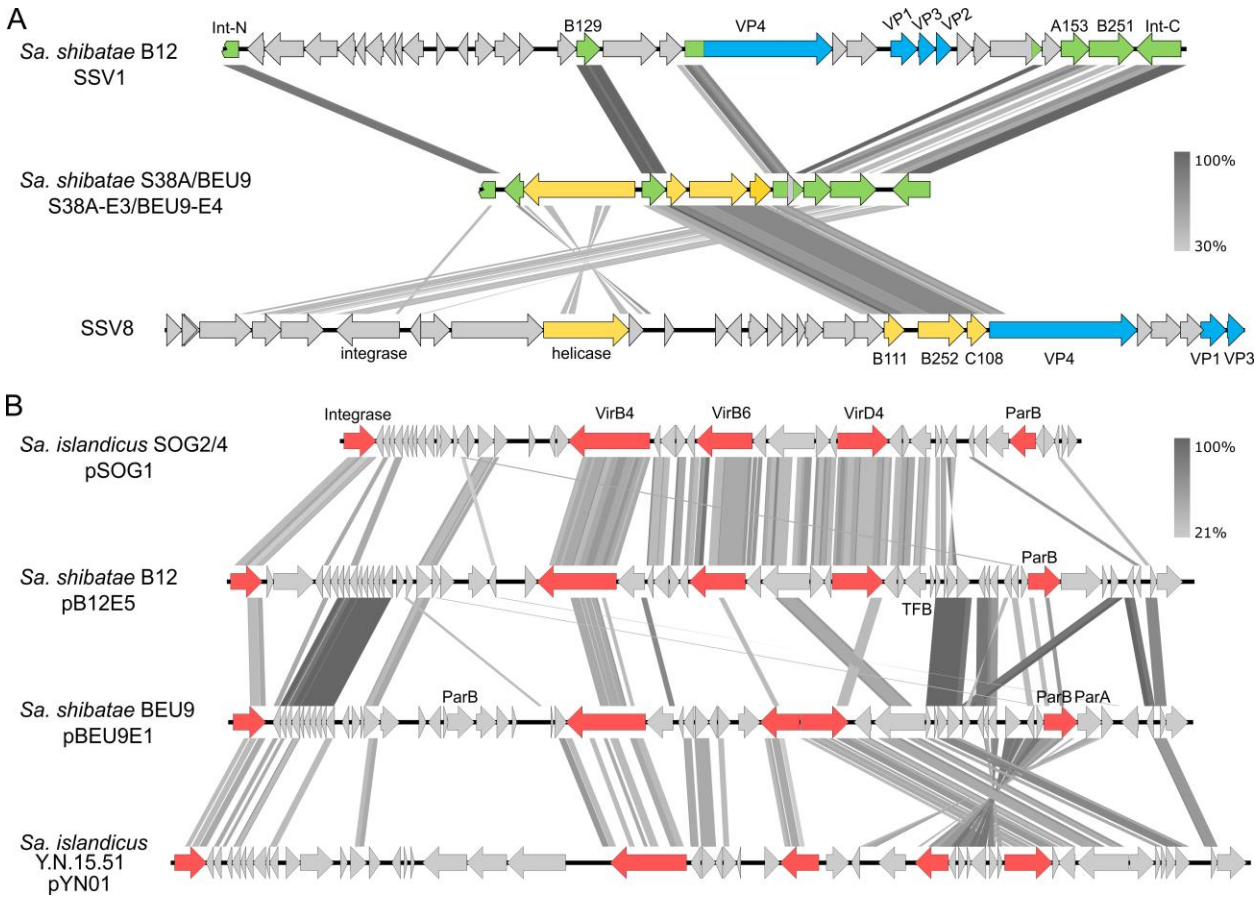

Figure 6. Viral and conjugative elements integrated in *Sa. shibatae* genomes. A. Genome maps of fuselloviruses SSV1 and SSV8, and a derived integrative element S38A-E3/BEU9-E4. Genes shared between SSV1 and S38A-E3/BEU9-E4 are shown in green, whereas those shared with SSV8 are in yellow. Genes encoding structural proteins of fuselloviruses (VP1-VP4) are shown in blue. FuselloGrey shading connects genes displaying sequence similarity at the protein level, with the percent of sequence identity depicted with different shades of grey (see scale on the right). Int-N and Int-C, N- and C-terminal fragments of the integrase. B. Integrative and conjugative plasmids of *Sa. shibatae*. Genes encoding proteins conserved in archaeal conjugative elements are shown in red. Grey shading connects genes displaying sequence similarity at the protein level, with the percent of sequence identity depicted with different shades of grey (see scale on the right).

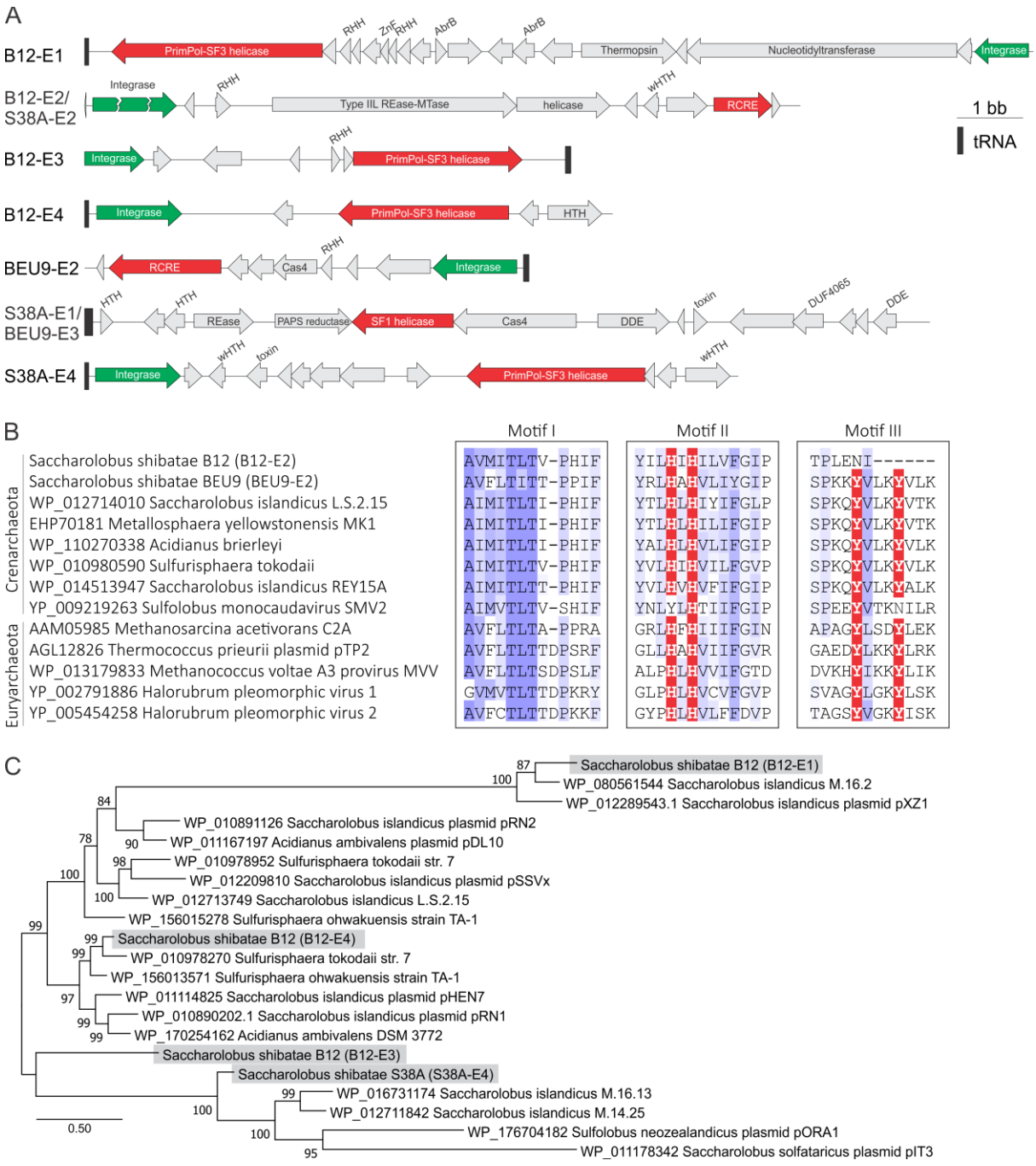

Figure 7. Cryptic plasmids of *Sa. shibatae*. A. Genome maps of cryptic plasmids. Gene encoding genome replication proteins and integrases are shown in red and green, respectively. Abbreviations: RHH, ribbon-helix-helix; PrimPol, primase-polymerase; SF1 and SF3, superfamily 1 and 3, respectively; (w)HTH, (winged) helix-turn-helix; REase, restriction endonuclease; MTase, methyltransferase; PAPS, phosphoadenosine phosphosulfate; DDE, DDE superfamily transposase. B. Alignment of the three conserved motifs (I–III) of superfamily I rolling circle replication initiation proteins encoded by euryarchaeal plasmids and viruses with the corresponding motifs from the putative replication protein of crenarchaeal integrated elements, including those from *Sa. shibatae*. Note the mutations in the proteins encoded by the virus SMV2 and *Sa. shibatae* B12 element E2. C. Maximum likelihood phylogenetic tree of PrimPol-SF3 helicase proteins encoded by diverse plasmids and integrated mobile genetic elements in the order Sulfolobales. Elements of *Sa. shibatae* are shown on grey background. The tree was constructed using the automatic optimal model selection (VT +F +I +G4) and is rooted at the midpoint. Numbers at the nodes represent represent SH-aLRT support values.

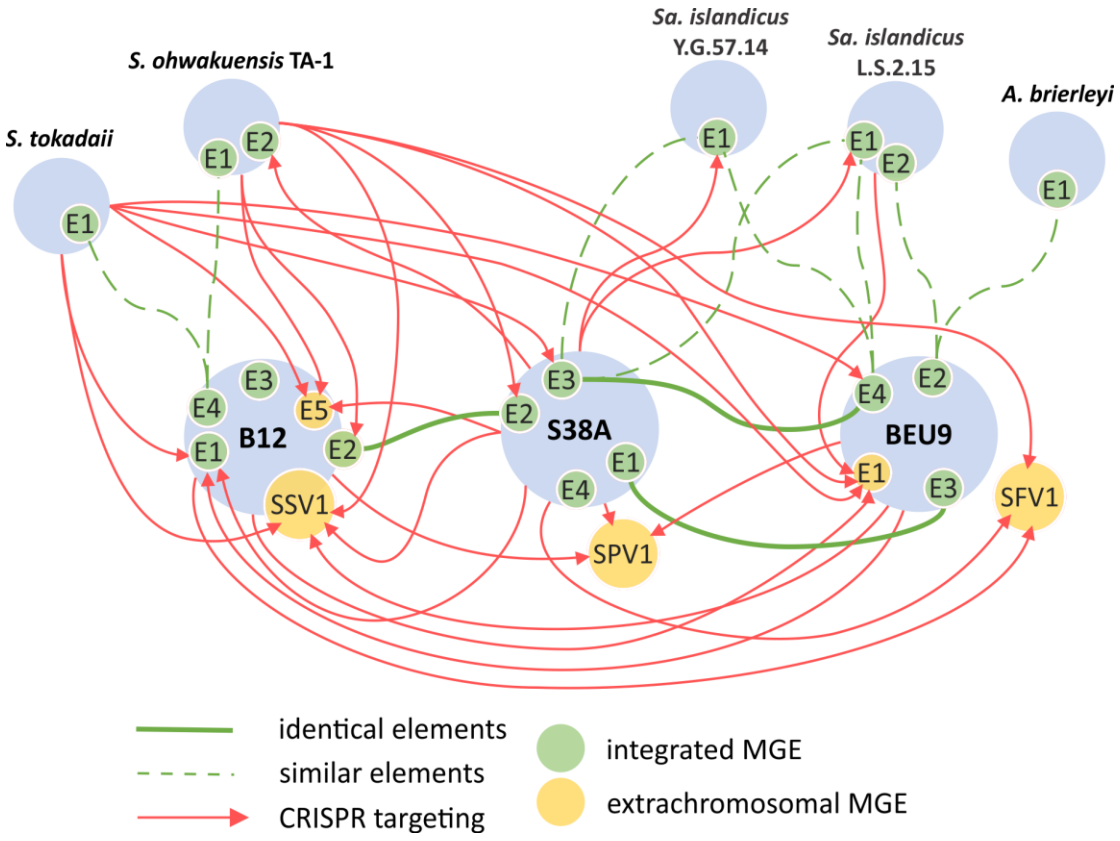

Figure 8. Relationships between integrated (green circles) and extrachromosomal (yellow circles) mobile genetic elements of *Sa. shibatae* and other members of the Sulfolobales. Identical and similar elements are connected with solid and broken green lines, respectively. Red arrows indicate targeting of the corresponding elements by CRISPR spacers carried by particular strains. SFV1 and SPV1 do not integrate into the cellular genome and are thus depicted adjacent to their respective hosts.

SUPPLEMENTARY INFORMATION

Supplementary figures

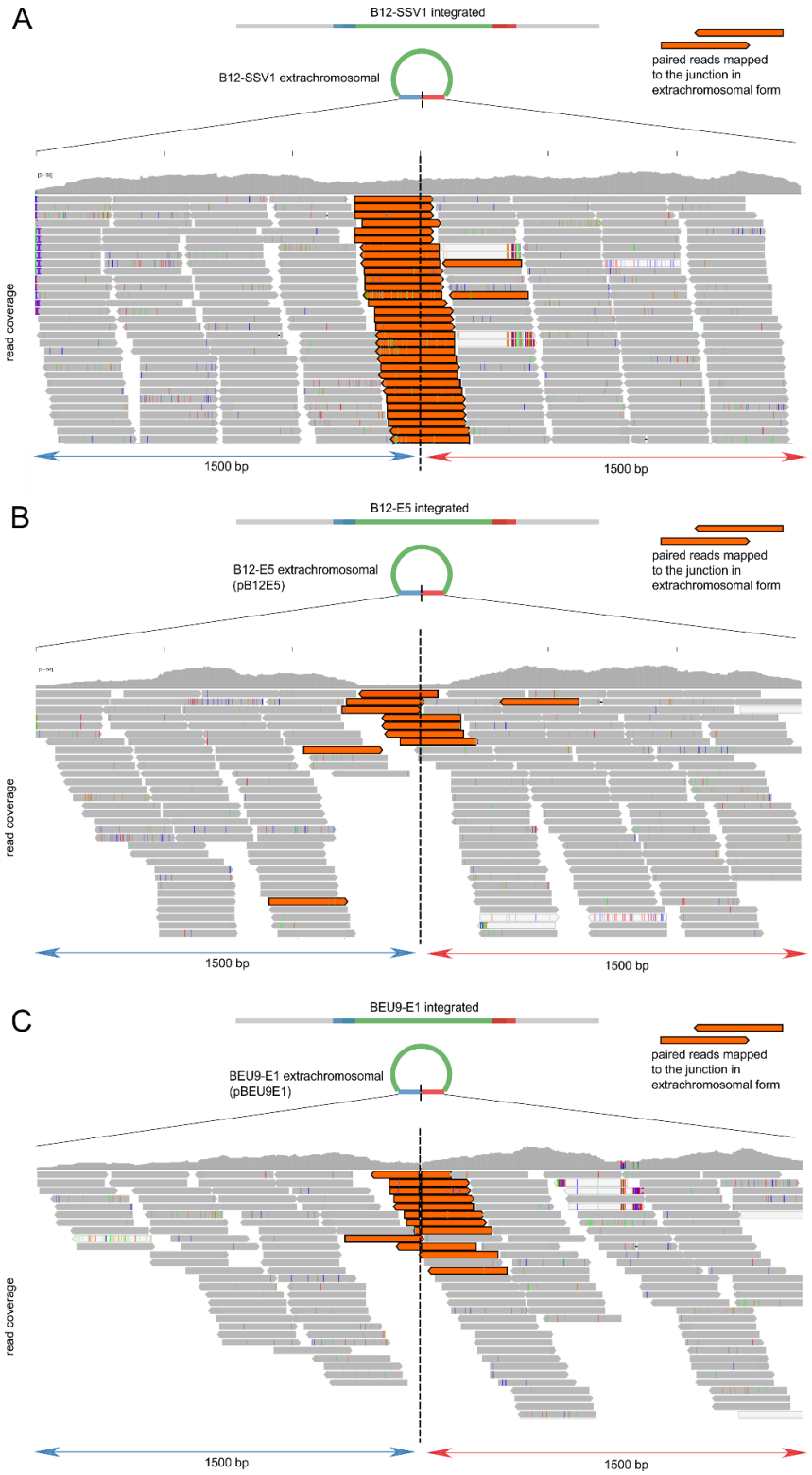

Figure S1. Presence of circular and extrachromosomal forms of the virus SSV1 (A) and plasmids pB12E5 (B) and pBEU9E1 (C). Paired-end sequencing reads are shown as arrows, which are aligned to the corresponding regions of the genomes. Termini of the integrated mobile genetic elements (MGE) are highlighted in blue and red. Upon excision, the MGE is circularized with the termini being joined together. Reads spanning the junction (indicated with a broken line) between the rejoined termini are shown in orange.

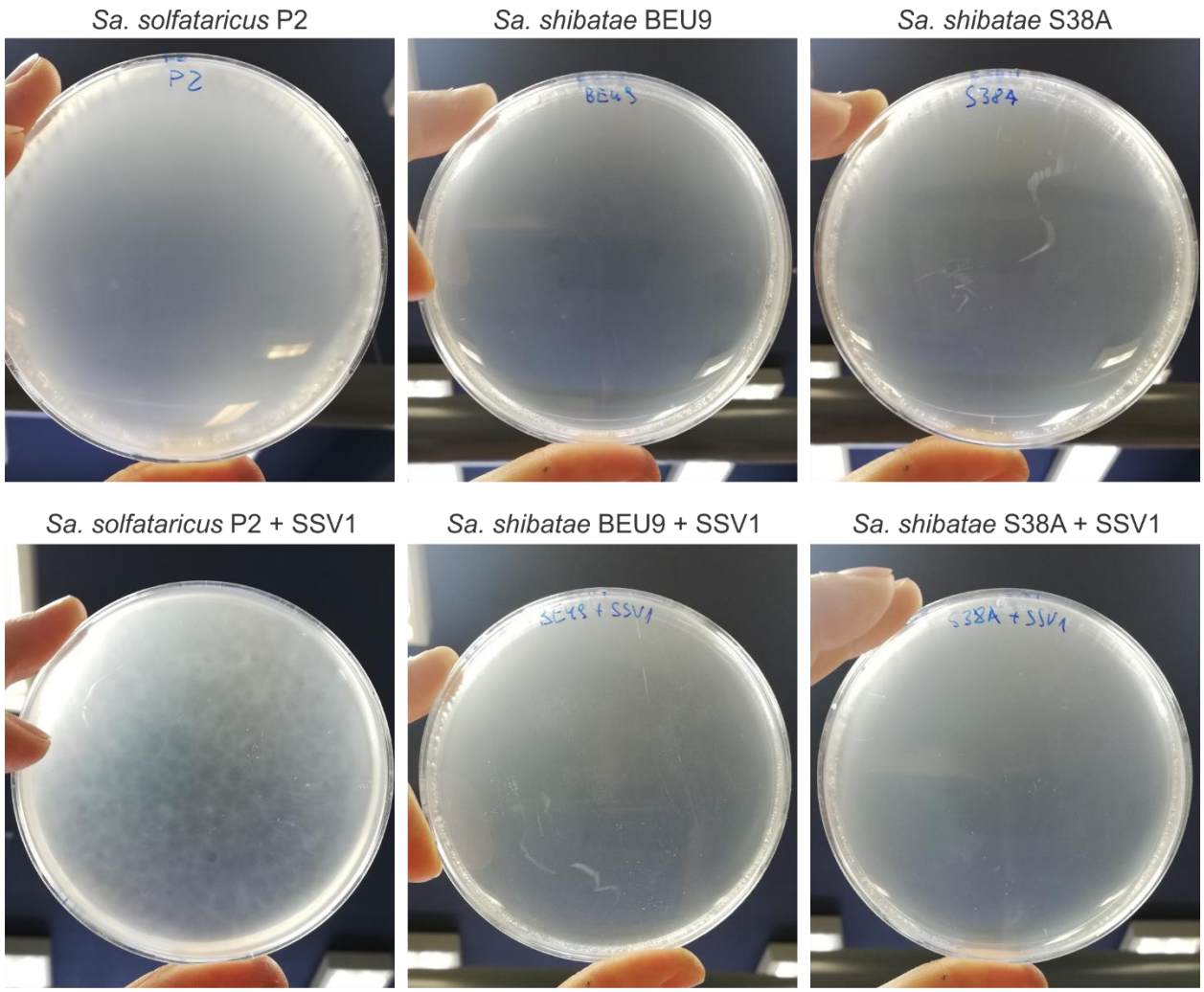

Figure S2. Sensitivity of *Sa. shibatae* strains BEU9 and S38A to SSV1. Top and bottom panels show the lawns of noninfected and infected cells, respectively. *Sa. solfataricus* P2, which is a susceptible host for SSV1, was used as a positive control.

Supplementary tables

Table S1. CRISPR targeting of *Sa. shibatae* viruses.

Origin	Strain	SSV1	SPV1	SFV1
Japan	*Sa. shibatae* B12	host	+	+
	Sa. shibatae S38A	+	host	+
	Sa. shibatae BEU9	+	+	host
	S. tokadaii strain 7	+	-	-
	S. ohwakuensis TA-1	+	-	+
USA	*Sa. isl.* Y.G.57.14	-	-	-
	Sa. isl. L.S.2.15	-	-	-
	A. brierleyi DSM 1651	-	-	-

Table S2. Toxin-antitoxin systems identified in the three *Sa. shibatae* strains.

Table S3. Insertion sequences identified in the three *Sa. shibatae* strains. Nucleotide coordinates for each IS element are provided.

Table S4. Functional annotation of the genes carried by the identified mobile genetic elements integrated in *Sa. shibatae* genomes.