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

Xu Peng

Correspondence

Xu Peng

peng@mermaid.molbio.ku.dk

Danish Archaea Centre, Department of Molecular Biology, Biocenter, Copenhagen University, Ole Maaløes Vej 5, DK-2200 Copenhagen N, Denmark

A fusellovirus SSV4 and a pRN-like plasmid pXZ1 were co-isolated from a single strain of *Sulfolobus*. In contrast to the previously characterized virus–plasmid hybrids pSSVx and pSSVi, which can coexist intracellularly with a fusellovirus, pXZ1 is not packaged into viral particles and shows no viral infectivity. The virus and plasmid carry genomes of 15 135 and 6970 bp, respectively. For SSV4, 33 predicted ORFs are compactly organized with a strong preference for UGA stop codons, three-quarters of which overlap with either the Shine–Dalgarno motif or the start codon of the following gene. pXZ1 carries seven ORFs, three of which encode an atypical RepA, a PirA and a CopG protein. A fourth ORF exhibits a high nucleotide sequence identity to the SSV4 integrase gene, which suggests that it has been transferred to the plasmid from SSV4. A single point mutation within an otherwise identical 500 bp region of the integrase gene occurs in the viral attachment site (*attP*), which corresponds to the anticodon region of the targeted tRNA gene in the host chromosome. This point mutation confers on pXZ1 the ability to integrate into the tRNA^{Glu}[CUC] gene, which differs from the integration site of SSV4, tRNA^{Glu}[UUC]. SSV4 and pXZ1 were also shown experimentally to integrate into separate sites on the host chromosome. This is believed to be the first report of a pRN plasmid sharing its natural host with a fusellovirus and carrying a highly similar integrase gene.

Received 5 September 2007

Revised 30 October 2007

Accepted 1 November 2007

INTRODUCTION

Over the past 20 years, many extrachromosomal and mobile genetic elements have been characterized for the Archaea in general, and the crenarchaeal *Sulfolobus* genus in particular. These include many novel viruses, conjugative plasmids and smaller cryptic plasmids, as well as autonomous insertion sequence (IS) elements and non-autonomous miniature inverted repeat transposable elements (MITEs) (reviewed by: Prangishvili *et al.*, 2006; Lipps, 2007; Brügger *et al.*, 2007). These genetic elements play important roles in the evolution of their host organisms and some have provided a basis for studying archaeal cellular processes and for developing archaeal genetic systems.

Abbreviations: cccDNA, covalently closed circular DNA; HTH, helix–turn–helix; RHH, ribbon–helix–helix; SD, Shine–Dalgarno; TEM, transmission electron microscopy.

The GenBank/EMBL/DBJ accession numbers for the sequences of SSV4 and pXZ1 are EU030938 and EU030940, respectively.

Supplementary tables showing primers for amplification of Southern hybridization probes and for the PCR test of SSV4 integration, the properties of SSV4 ORFs and operons, and reannotated ORFs in the genomes of SSV2, SSV RH, SSV K1 and SSV1, are available with the online version of this paper.

The *Fuselloviridae* is one of seven crenarchaeal viral families that have been classified recently. Four members (SSV1, SSV2, SSV K1 and SSV RH), all sharing the same morphology, i.e. a spindle shape with a short tail, have been isolated from widely different geographical locations. SSV1 (also known as SAV1) was identified in *Sulfolobus shibatae* from Beppu, Japan, as a plasmid and shown later on to be the episomal form of a novel UV-inducible virus (Martin *et al.*, 1984). SSV2 was isolated from a *Sulfolobus islandicus* strain sampled from a solfataric hot spring in Reykjanes, Iceland (Stedman *et al.*, 2003), while SSV K1 was from the Kamchatka region of Eastern Russia and SSV RH was isolated from Yellowstone National Park, USA (Wiedenheft *et al.*, 2004). Each of these viruses can infect *Sulfolobus solfataricus* P2, which was isolated from Naples, Italy, indicating that they have a broader host range than other known crenarchaeal viruses (Prangishvili *et al.*, 2006).

Each virus contains a circular dsDNA genome of about 15 kb with about 34 ORFs which show little or no similarity to sequences in public databases. However, comparative genome analyses of the four viruses have revealed that about one half of each genome is highly conserved while the other half is variable (Wiedenheft *et al.*,

2004). Four gene products have been assigned functions: two viral coat proteins VP1 and VP3 which are shared between the viruses, another coat protein VP2 encoded only by SSV1, and a tyrosine integrase. The last facilitates recombination between the viral and archaeal attachment sites, *attP* and *attA*, respectively, producing a partitioned integrase gene upon integration into the host chromosome (Muskhelishvili *et al.*, 1993; Serre *et al.*, 2002).

Of particular interest is that SSV2 was isolated together with a satellite, pSSVx, which can spread through a *Sulfolobus* culture in the presence of a helper virus such as SSV1 or SSV2 (Arnold *et al.*, 1999; Stedman *et al.*, 2003). The plasmid-virus hybrid pSSVx is a fusion between a pRN-type plasmid and two genes originating from SSVs which were considered to be involved in the packaging of pSSVx (Arnold *et al.*, 1999). Another virus satellite, pSSVi, was recently identified in *S. solfataricus* P2 cells, where it was present only in an integrated form. Upon transformation of the cells by SSV2 DNA prepared from *S. islandicus* REY31A, pSSVi was excised from the host genome and replicated actively (Wang *et al.*, 2007). Like pSSVx, pSSVi can also be packaged into a spindle-like viral particle and spread with the help of SSV1 or SSV2. pSSVi resembles members of the pRN plasmid family in genome organization but encodes an SSV-type integrase (Wang *et al.*, 2007). The conserved genome region of pSSVx and of the other pRN-type plasmids isolated from Iceland includes a replication protein RepA, a putative copy number-control protein CopG and a putative regulatory protein PlrA (Peng *et al.*, 2000; Lipps, 2007). However, three pRN-type plasmids isolated from New Zealand carry similar *copG* and/or *plrA* genes but encode either a much less conserved RepA or a completely different replication protein (Greve *et al.*, 2005). For example, the N-terminal sequence of RepA in pTIK4 shows no sequence similarity to the corresponding ORF regions of other pRN plasmids and the large ORF in pTAU4 encodes a putative minichromosome maintenance protein (MCM). Thus, the pRN plasmids constitute a diverse family. Although a few pRN-type elements have been discovered in an integrated form in the chromosomes of *Sulfolobus* (e.g. Peng *et al.*, 2000), the only free form of a pRN-type plasmid encoding an integrase is the recently described plasmid-virus hybrid pSSVi (Wang *et al.*, 2007).

The work presented here describes the co-isolation of a new member of the *Fuselloviridae*, SSV4, and an integrase-containing pRN-type plasmid, pXZ1, from a pure *Sulfolobus* strain. In contrast to pSSVx and pSSVi, pXZ1 did not spread together with SSV4. Genomic features of the two elements are described and their site-specific integration into host chromosomes is demonstrated.

METHODS

Isolation of virus-containing strains from *Sulfolobus* enrichments. Hot-spring samples collected from various sites in Iceland and enriched in Wolfram Zillig's laboratory were generously made

available to us after his retirement from the Max-Planck-Institute, Munich, Germany. In order to obtain pure virus-containing strains, enrichments were grown in standard *Sulfolobus* medium at 80 °C, as described by Zillig *et al.* (1994). Single-colony clones were then isolated from 0.7% Gelrite plates (Kelco) and inoculated into 50 ml medium. Cultures were then screened for plaque formation using *S. solfataricus* P2 as an indicator strain, as described by Schleper *et al.* (1992). Positive clones were further cultured in 250 ml medium and supernatants were investigated for the presence of virus-like particles by transmission electron microscopy (TEM), as described by Zillig *et al.* (1994). SSV4 was identified from a pure strain isolated from the enrichment *S. islandicus* ARN3 (collected from Arnavatn, Iceland). Since a few pure strains had already been isolated from the same enrichment (Prangishvili *et al.*, 1998), this particular strain was designated *S. islandicus* ARN3/6.

Virus purification and DNA isolation. Cells were isolated from cultures by centrifugation at 7000 g for 15 min at 4 °C, and extrachromosomal covalently closed circular (ccc)DNA was purified using Plasmid Miniprep and Maxiprep Kits (Qiagen). Virus particles were precipitated from culture supernatants by adding 1 M NaCl and 10% (w/v) PEG 6000 (final concentrations) and stirred overnight at 4 °C. The precipitated virus was collected by centrifuging at 15 000 g for 30 min. Virus particles were purified by centrifugation at 69 000 g for 24 h at 20 °C in a CsCl density gradient (0.39 g ml⁻¹). The virus band was removed and dialysed against 10 mM Tris/acetate buffer, pH 6, and stored at 4 °C. Nucleic acid was isolated from the purified virus as described by Zillig *et al.* (1994).

DNA analysis and sequencing. DNAs were digested with the restriction enzyme *Bgl*II, and fragments were separated by agarose gel electrophoresis and visualized by staining with ethidium bromide or SYBR gold (Invitrogen). Shotgun libraries were constructed by inserting sonicated and end-repaired DNA fragments in the size range 1.5–4 kb into a pUC18 vector. Clones were sequenced in MegaBACE sequencers (Amersham Biotech) to yield a fivefold sequence coverage of the genomes. Contig assemblies were accomplished with Sequencher version 4.3 (Gene Codes). Any remaining gaps or ambiguous regions, including the highly similar integrase genes, were PCR-amplified, and the products were sequenced on both strands.

Sequence analyses. ORFs were located and adjusted using putative TATA-like promoter motifs and Shine–Dalgarno (SD) motifs (Torarinsson *et al.*, 2005) in the Artemis V9 program (Sanger Institute). Employing all three start codons (AUG, GUG and UUG), genes were defined by initially selecting for the largest possible ORFs. When the presence of TATA-like and/or SD motifs supported the existence of a shorter coding region, this was selected. BLASTP searches were performed against the EMBL/NCBI database (Altschul *et al.*, 1997) and conserved protein domains were detected by MotifScan in ExPASy Proteomics Tools (<http://www.expasy.org/tools/>). Nucleotide sequence alignments between SSVs were accomplished by BLASTN searches against the *Sulfolobus* Database (Brügger, 2007).

Southern hybridization and PCR. Total DNAs from *S. islandicus* ARN3/6 and virus-infected *S. solfataricus* P2 cells were extracted as described by Zillig *et al.* (1994). Total DNAs and extrachromosomal cccDNA isolated from *S. islandicus* ARN3/6 cells were digested with the restriction enzyme *Bgl*II and the resulting fragments were analysed by standard Southern blot hybridization procedures (Sambrook & Russell, 2001). DNA fragments upstream of the integrase gene in SSV4 and pXZ1 were PCR-amplified and labelled by DIG-11-dUTP, alkali-labile, with a Random Primed DNA Labeling Kit (Roche Applied Science). After hybridization, DIG-labelled DNA was detected by chemiluminescent CDP-Star (Roche Applied Science).

PCR was performed to confirm integration of SSV4 into the *S. solfataricus* P2 chromosome. The primers used in the experiment, as well as in generating probes for Southern hybridization analyses, were purchased from TAG Copenhagen. Sequences of PCR primers are available in Supplementary Table S1.

RESULTS AND DISCUSSION

Isolation of SSV4 and pXZ1

Members of the *Fuselloviridae* are widespread in *Sulfolobus* species throughout Iceland (Zillig *et al.*, 1994). Consistent with this, most of the samples from the Zillig Iceland collection produced plaque-like haloes on cell lawns of *S. solfataricus* P2. In order to obtain pure virus-containing strains, single-colony clones were isolated from the enrichments and screened for the presence of viruses (see Methods).

Spindle-shaped virus particles were observed by TEM in the culture supernatant of a pure strain, *S. islandicus* ARN3/6, indicating the possible presence of an SSV-like virus (Fig. 1a). Moreover, the cell-free supernatant of the culture was able to produce clear plaques on cell lawns of *S. solfataricus* P2. In order to demonstrate infectivity of the putative virus, a Gelrite slice containing the plaque zone and the surrounding *S. solfataricus* P2 cells was inoculated into fresh medium. The supernatant of the late-exponential-phase culture was shown by TEM to contain spindle-shaped particles with a morphology identical to that

observed from the original host (Fig. 1a). Moreover, viruses purified from the infected *S. solfataricus* P2 cells could also produce plaques on cell lawns of uninfected *S. solfataricus* P2 cells. This demonstrated that the spindle-shaped particles were indeed infectious virions. Subsequent sequencing showed that virions contained a dsDNA genome similar to those of the previously characterized members of the *Fuselloviridae* (see below). Therefore, the virus was named SSV4.

A *Bgl*II restriction digestion of extrachromosomal cccDNA isolated from *S. islandicus* ARN3/6 cells yielded a substoichiometric band pattern (Fig. 1b). Four large fragments ranging from 2 to about 6.5 kb were produced from SSV4 genomic DNA, which correlated accurately with its genome sequence (see below). Moreover, a 5.5 kb fragment was found to be present at a higher molar ratio than the SSV4 bands (Fig. 1b, lane 1). However, DNA extracted from virions did not contain this fragment, nor did the extrachromosomal cccDNA or viral DNA prepared from infected *S. solfataricus* P2 cultures (Fig. 1b). This suggested that the DNA was not packaged into viral particles. The fragment was sequenced and shown to constitute a pRN-like plasmid which we named pXZ1. In more than 10 individual infectivity assays in which PCR and/or Southern hybridization experiments (see below) were employed to investigate the spread of the two elements, only SSV4 was found to spread in *S. solfataricus* P2. This confirmed that pXZ1 is a plasmid.

SSV4 and pXZ1 were both stably maintained in *S. islandicus* ARN3/6. After more than 10 continuous transfers of the culture into fresh medium, both elements remained detectable. Moreover, each of 20 colonies isolated from the colony-purified *S. islandicus* ARN3/6 was found to contain both SSV4 and pXZ1 (data not shown). Copy-number fluctuation was observed for both SSV4 and pXZ1. In the first few cultures of *S. islandicus* ARN3/6, or when SSV4 was first introduced into *S. solfataricus* P2, their copy numbers were quite high, as judged from a high DNA yield from the plasmid purification and the clear restriction pattern of SSV4 and pXZ1 in the total DNA digest. However, the copy number of both elements fell after continuous transfers of the culture, or after long-term storage of the cell stock at -80°C . The highest copy number observed was about 30 per cell, while the lowest was about two per cell for each element, as revealed by Southern hybridization studies (see below).

Fig. 1. (a) Electron micrograph of SSV4 (bar, 100 nm). Three virions attached together by their ends are denoted with an arrow. (b) *Bgl*II digests of total extrachromosomal cccDNAs isolated from *S. islandicus* ARN3/6 cells (lane 1), DNA extracted from virions purified from the supernatant of the *S. islandicus* ARN3/6 culture (lane 2), total extrachromosomal cccDNAs isolated from virus-infected *S. solfataricus* P2 cells (lane 3) and DNA extracted from virions purified from the supernatant of the virus-infected *S. solfataricus* P2 culture (lane 4). M, DNA ladder with sizes shown on the left-hand side. The 5.5 kb fragment, of higher molar concentration, is indicated by an arrow.

Genome organization and gene content of SSV4

In order to further characterize the two elements, their complete genomes were sequenced (see Methods). SSV4 virions contain a circular dsDNA genome of 15 135 bp with a G+C content of 38.5%, similar to that of the previously characterized SSVs (Wiedenheft *et al.*, 2004). Thirty-three ORFs were identified, ranging in size from 45 to 808 aa (Fig. 2). About 25 of the ORFs were preceded by putative SD motifs, while only 10 were preceded by

Fig. 2. Genome map of SSV4 aligned with the other SSVs. The left-hand halves of the linearized genomes contain the highly conserved region, while the right-hand halves are more divergent (Wiedenheft *et al.*, 2004). ORFs found in all SSVs are shown by black arrows, whereas those exclusive to a single SSV are indicated by white arrows. ORFs conserved among two to four of the SSVs can be identified by their colour and filling pattern. The names of the reannotated ORFs are positioned above their genome maps. Arrowed lines denote predicted SSV4 operons (top) and experimentally determined SSV1 operons (bottom). For SSV4 alone, putative transcriptional regulator genes are indicated by pink arrow heads.

putative TATA-like sequences, and eight exhibited downstream T-rich sequences which are likely to be transcriptional terminators (Reiter *et al.*, 1988). More than half of the ORFs overlapped with adjacent genes, and 11 were located less than 20 bp from the next gene and were therefore considered to form part of an operon. The information on putative promoter and terminator sequences, as well as the distance between genes, is summarized in Supplementary Table S2. In total, four putative operons were identified that comprised 29 genes. All genes, except the first ones in the putative operons, exhibited SD motifs, whereas in general the single genes and the first genes of operons did not, as has been observed earlier for *Sulfolobus* and other archaeal genomes (Tolstrup *et al.*, 2000; Torarinsson *et al.*, 2005). The annotation of SSV4 operons is supported by the similar organization of the experimentally detected SSV1 co-transcripts (Fig. 2) (Reiter *et al.*, 1987; Fröls *et al.*, 2007).

Sequence comparison revealed that some ORFs in the four SSV genomes characterized earlier are much larger than their predicted SSV4 homologues (Fig. 2). For example, ORF A168 of SSV2 is 44 aa longer than the homologous ORF124 in SSV4 and 39 aa longer than ORF B129 of SSV1. ORF A168, which lacks an SD motif and was assigned a CUG start codon, was reannotated to ORF A168 (124)

encoding 124 aa and carrying an AUG start codon and GGAGG SD motif (Fig. 2, Supplementary Table S3). Other ORFs were also re-examined for the presence of a putative TATA-like motif, an SD motif and/or terminator sequences in the earlier-characterized SSV genomes. This resulted in two to nine genes being reannotated in each viral genome (Fig. 2, Supplementary Table S3). Moreover, similar operon structures were predicted for the genomes of SSV2, SSV RH and SSV K1 (data not shown).

About 90 % of the start codons of SSV4 were AUG, and a few started at GUG or UUG, as found in the genomes of *S. solfataricus* and *Sulfolobus tokodaii* (Garcia-Vallve *et al.*, 2003). However, in contrast to the *Sulfolobus* chromosomal genes that use the UAA stop codon in preference to UGA and UAG, more than 60 % of SSV4 ORFs terminated at UGA (Table 1). This is also true for the other SSVs, which average 50–60 % UGA, but is not true for the other crenarchaeal viruses, which average 20–30 % UGA (Table 1). Further analysis revealed that about three-quarters of the UGA stop codons in SSVs were part of the SD motif of a next gene (GGTGA), or overlapped by 1 (TGATG) or 4 bp (ATGA) with the start codon of the following ORF. This demonstrated that SSV genomes are organized in a highly compact fashion, whereby the TGA sequence can be multifunctional.

Table 1. Percentage stop codon use in genes of fuselloviruses and other crenarchaeal viruses

Percentages of preferred stop codons are highlighted in bold type.

Stop codon	SSV					SIFV	SIRV1	STSV1	STIV	ABV	ATV
	4	2	RH	K1	1						
UGA	66	67	57	44	56	34	27	27	37	38	40
UAA	16	24	37	35	25	57	61	54	46	51	40
UAG	18	9	6	21	19	9	12	19	17	11	20

Besides genes that encode the structural proteins VP1 and VP3 and the integrase, a MarR family transcription regulator (ORF 143), a ribbon-helix-helix (RHH) protein (ORF79), two C2H2 zinc finger proteins (ORFs 80 and 124) and one helix-turn-helix (HTH)-containing protein (ORF 107a) were identified (Fig. 2, Table 2). It is noteworthy that the putative transcription regulators are concentrated in the genomic region corresponding to the two early transcription units T5 and T6 identified for SSV1, which are considered to be responsible for genome replication (Fröls *et al.*, 2007; Fig. 2). This suggests that the putative transcription regulators are involved in the regulation of genome replication through protein-protein and/or protein-DNA interactions.

Genome of pXZ1

The pXZ1 genome contains 6970 bp with a G+C content of 39.4%. A total of seven ORFs were identified in the genome, five of which were assigned functions based on their similarities to sequences in public databases (Fig. 3, Table 3). ORFs 82 and 52 show high similarities to pRN putative regulatory protein (PlrA) and copy number control protein (CopG), respectively, while ORF75 shows a similarity to pDL ORF72, which is located immediately upstream of repA and has been designated CopG-like (Kletzín *et al.*, 1999). Homologues of the latter have also been found at the same genomic region of pSSVx and of two integrated pRN elements in *Sulfolobus* chromosomes,

pST1 and pXQ1 (She *et al.*, 2007). Although ORF1077 showed limited similarity to the putative RepA protein of pTIK4 and a bacterial DNA primase-like protein (Table 3), no recognizable protein domain was detected. This suggests that ORF1077 constitutes a novel type of replication protein. Previous sequence comparison has shown that pRN plasmids isolated from Iceland have very similar replication proteins, whereas those from Italy and New Zealand encode quite different RepAs (reviewed by Lipps, 2007). The presence of a distantly related putative replication protein in the novel pRN plasmid isolated from Iceland, pXZ1, reinforces the diversity of the replication proteins of pRN plasmids.

Two ORFs of unknown function, ORF76 and ORF129, are exclusive to pXZ1 and are located in its variable region. With respect to ORFs in this region of other pRN plasmids (Peng *et al.*, 2000; Wang *et al.*, 2007), ORF129 yields no sequence matches in public databases and ORF76 shows similarity only in a 30 aa region to ORF90a of pRN1.

Based on high sequence similarities to integrase genes encoded in the SSV genomes (Table 3), ORF 331 from pXZ1 was annotated as an integrase-encoding gene. The amino acid sequence identity/similarity between ORF331 and its closest homologue, the SSV4 integrase gene, is 86/94%. Given that pXZ1 was co-isolated with SSV4 from a single-colony culture and that the two elements are stably maintained in the same host, the high similarity strongly suggests that ORF331 originated from SSV4.

Table 2. Functions assigned to SSV4 ORFs

ORF	Homologues in:	Predicted function	Pfam expectation value*
330†	All SSVs	Integrase	—
VP1†	All SSVs	Structural protein VP1	—
VP3†	All SSVs	Structural protein VP3	—
143	SSV2, SSV RH	Transcriptional regulator MarR family	6.5e-05
79	All SSVs	RHH protein, CopG family	9.1e-12
80	All SSVs	Zinc finger, C2H2 type	2.9e-05
124	All SSVs	Zinc finger, C2H2 type	8e-05
107a	All SSVs	HTH-5-containing	0.0001

*Sequences of all SSV homologues were used to search the Pfam database and only the best e-value is given.

†Functional assignments for the three genes are based on their similarity to the experimentally characterized SSV1 homologues (Palm *et al.*, 1991).

Fig. 3. Linearized genome map of pXZ1 aligned with those of pSSVi, pSSVx and pRN1. ORFs constituting the conserved core of pRN plasmids are shown by green arrows, while those of SSV origin are in red. Homologous ORFs are defined by arrows with same colour and filling pattern. ORFs represented by white arrows have no homologues in the other pRN plasmids presented here.

Although a few integrase-containing pRN-type elements have been found in *Sulfolobus* chromosomes (Peng *et al.*, 2000; She *et al.*, 2007), none of the eight previously characterized extrachromosomal pRN plasmids contains an integrase gene (reviewed by Lipps, 2007). Only the recently identified virus satellite pSSVi, originally present in an integrated form in an *S. solfataricus* strain and excised upon introduction of SSV2 DNA into the host cells, is similar to pRN plasmids in genome organization; it encodes an integrase which shows only 38/57 % amino acid identity/similarity to that of SSV2 (Wang *et al.*, 2007). This raises the question of the origin of the integrase genes present in the integrated pRN-like elements. The identification of the pXZ1 integrase gene provides the first strong evidence that it was probably transferred from an SSV.

The genome maps of pXZ1, pSSVi and pSSVx, together with that of pRN1 (Keeling *et al.*, 1998), are aligned in Fig. 3. A common feature shared between the first three is that they coexist with an SSV and each shares homologous genes with SSVs which may be of viral origin. While pXZ1 and pSSVi encode SSV-type integrases, pSSVx contains two genes that are conserved in SSVs (Fig. 3). Thus, co-existing

with an SSV in the same host apparently facilitates the uptake of viral genes into the pRN genomes, which may confer a survival advantage upon the latter by enabling integration into host chromosomes or packaging into viral particles and hence spreading together with the virus. Unlike pSSVi and pSSVx, however, pXZ1 remains as a plasmid and does not spread with SSV4 to other host cells. This could be due to the absence of a packaging signal in the genome of pXZ1.

Integration of SSV4 and pXZ1

The presence of a putative integrase gene in the genomes of SSV4 and pXZ1 suggests that they are capable of integrating into host chromosomes. To test this possibility, Southern hybridization experiments were performed. Firstly, their putative attachment sites (*attP*) were identified and aligned with the putative *attA* sites in the genome of *S. solfataricus* P2 (Sso) (She *et al.*, 2001). As shown in Fig. 4(a), the *attP* sequences of SSV4 and pXZ1 match perfectly with the 3' halves of the genes for tRNA^{Glu}[UUC] and tRNA^{Glu}[CUC], respectively. Except for one mismatch in the anticodon (UUC vs CUC), the two tRNA genes are identical in the genome of *S. solfataricus* P2, including

Table 3. ORF analysis of pXZ1

ORF	Annotation	Best match (percentage identity/similarity)
331	Integrase	SSV4 integrase (86/94)
82	PlrA	pRN1 PlrA (86/93)
76	Hypothetical	30 aa aligned with pRN1 ORF 90a (66/81)
129	Hypothetical	—
52	CopG	pHEN7 CopG (73/92)
75	CopG-like	pST1 CopG-like (66/82)
1077	Replication protein	pTIK4 putative RepA N-terminal 200 aa (38/53); <i>Syntrophomonas wolfei</i> DNA primase-like protein at positions 432–700 (23/47)

their intron sequences (She *et al.*, 2001). tRNA genes with sequences identical to those of *S. solfataricus* tRNA^{Glu}[UUC] and tRNA^{Glu}[CUC] also occur in the genome of *S. islandicus* HVE10/4 (K. Brügger and others, unpublished results), which was isolated from Iceland and is closely related to *S. islandicus* ARN3/6 (Zillig *et al.*, 1994). Given the close relationship between the two *S. islandicus* strains, the same two tRNA genes are expected to be present in the natural host strain of SSV4 and pXZ1.

Using an SSV4-specific probe (see Methods), integration of the virus was detected by Southern hybridization experiments (Fig. 4b). A fragment of 1.6 kb was detected in addition to the two episomal fragments in total DNA from the virus-infected *S. solfataricus* P2, consistent with the expected 1635 bp host–SSV4 hybrid band assuming tRNA^{Glu}[UUC] to be the attachment site (Fig. 4b, lanes 2 and 2'). An extra band of 4.2 kb was also detected from the

total DNA of *S. islandicus* ARN3/6 (Fig. 4b, lane 1). Although the complete genome sequence is unavailable, the size matches closely with the predicted 4252 bp assuming that tRNA^{Glu}[UUC] in the chromosome of *S. islandicus* HVE 10/4 is the attachment site (K. Brügger and others, unpublished results). This suggests a high sequence similarity between the two *S. islandicus* strains. In order to verify the integration, PCR experiments were employed to amplify the border of the virus–host hybrid from total DNA of virus-infected *S. solfataricus* P2 (Fig. 4d). Subsequent sequencing of the PCR products confirmed the integration of SSV4 into tRNA^{Glu}[UUC].

Hybridization with a pXZ1-specific probe revealed one band in addition to the episomal fragment from *S. islandicus* ARN3/6 total DNA (Fig. 4c), indicating that pXZ1 integrated into one site in the host chromosome. Estimating from the genome sequence of *S. islandicus* HVE

Fig. 4. Integration of SSV4 and pXZ1. (a) Sequence alignment of *attP* sites of SSV4 and pXZ1 with the 3' halves of *S. solfataricus* P2 tRNA^{Glu}[UUC] and tRNA^{Glu}[CUC] genes. The complementary sequences of the tRNA genes are shown in lower-case type and the three nucleotides corresponding to the anticodon are indicated by asterisks. (b, c) Detection of SSV4 (b) and pXZ1 (c) integration into host genomes by Southern hybridization. A diagram showing the putative integration mechanisms is presented on the left, where the *Bgl*II sites and genomic position of the DIG-labelled probe (grey bar on the left of *attP*) are indicated. The empty triangles denote the positions of the first nucleotide of SSV4 and pXZ1 sequences in GenBank. Lane 1, *S. islandicus* ARN3/6 total DNA; lane 2, virus-infected *S. solfataricus* P2 total DNA; lane 3, extrachromosomal cccDNA from *S. islandicus* ARN3/6 cells. Lane 2' in (b) is another version of lane 2 from a much less exposed film to better show the integrated band. M, DNA ladder with sizes shown on the left. DNA bands corresponding to SSV4- or pXZ1–host hybrids are denoted by arrowheads on the right. (d) PCR detection of SSV4 integrated into the *S. solfataricus* P2 chromosome. Genomic positions of primers are indicated on the diagram of the Sso–SSV4 hybrid. Total DNAs used as PCR templates were extracted from non-infected (lanes 1 and 2) and SSV4-infected (lanes 3 and 4) *S. solfataricus* P2 cells. Primer set P1/P2 (lanes 1 and 3) was used to test for the presence of SSV4 DNA in the cells and P3/P4 (lanes 2 and 4) was used to detect integration.

10/4 that is available in our laboratory (K. Brügger and others, unpublished results), integration into tRNA^{Glu}[CUC] would produce a host-pXZ1 hybrid fragment of 3388 bp after *Bgl*III digestion, whereas a 5101 bp fragment would be expected after integration into tRNA^{Glu}[UUC]. Therefore, the presence of a 3.4 kb hybrid fragment strongly indicates a site-specific integration of pXZ1 into the tRNA^{Glu}[CUC] gene (Fig. 4c). As expected, neither the episomal nor the integrated form of pXZ1 was detected in the virus-infected *S. solfataricus* P2 total DNA (Fig. 4c, lane 2). This confirms that pXZ1 is a plasmid and does not spread together with SSV4 in a foreign host.

As mentioned above, the integrase gene encoded in pXZ1 probably derived from SSV4, and this is consistent with the 88% overall nucleotide identity between the two genes. About 500 bp in the centre, including the *attP* site, are almost identical. The only mismatch within this region is found at a position corresponding to the mutated nucleotide in the anticodon of the two tRNA genes (Fig. 4a). Thus, this single point mutation seems to have been positively selected and has conferred on pXZ1 the ability to integrate into a different tRNA gene from that targeted by SSV4, thereby avoiding any competition for integration sites.

CONCLUDING REMARKS

Site-specific integration has been recognized in some bacteriophages (e.g. phage lambda) as a strategy to cope with unfavourable conditions for replication, whereby the integrated phage genome can be passively maintained and replicated in the host chromosome. Since most crenarchaeal viruses do not lyse host cells and possess no lytic/lysogenic switches (Prangishvili *et al.*, 2006), the role of their site-specific integration has rarely been interpreted in terms of their own requirements but rather in terms of the advantages for host genome evolution (She *et al.*, 2007). Given the harsh and destructive environment that most known crenarchaea and their viruses encounter, site-specific integration, at least for viruses with circular genomes, is probably advantageous for virus survival. Thus, taking over and modifying the viral integrase to create a new plasmid integration site in the host chromosome provides an alternative strategy for enhancing plasmid survival. This suggests that plasmids use different strategies to survive in *Sulfolobus* cells growing under extreme conditions. When coexisting with a fusellovirus either they exploit the viral integrase, as described here, to integrate into a different chromosomal site or they exploit the viral packaging system and spread as a virus satellite, as found for pSSVx and pSSVi.

ACKNOWLEDGEMENTS

I am grateful to Professor Roger A. Garrett for his support and encouragement. Dr David Prangishvili is acknowledged for discussions. Hien Phan is thanked for helping with the sequencing of SSV4

and pXZ1. This work was supported by a Steno Stipend from the Danish Research Council for Natural Science. This paper is dedicated to Wolfram Zillig who provided the enrichment cultures examined here.

REFERENCES

- Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**, 3389–3402.
- Arnold, H. P., She, Q., Phan, H., Stedman, K., Prangishvili, D., Holz, I., Kristjansson, J. K., Garrett, R. & Zillig, W. (1999). The genetic element pSSVx of the extremely thermophilic crenarchaeon *Sulfolobus* is a hybrid between a plasmid and a virus. *Mol Microbiol* **34**, 217–226.
- Brügger, K. (2007). The *Sulfolobus* database. *Nucleic Acids Res* **35** (Database issue), D413–D415.
- Brügger, K., Peng, X. & Garrett, R. A. (2007). *Sulfolobus* genomes: mechanisms of rearrangement and change. In *Archaea: Evolution, Physiology and Molecular Biology*, pp. 95–104. Edited by R. A. Garrett & H. P. Klenk. Oxford: Blackwell.
- Fröls, S., Gordon, P. M., Panlilio, M. A., Schleper, C. & Sensen, C. W. (2007). Elucidating the transcription cycle of the UV-inducible hyperthermophilic archaeal virus SSV1 by DNA microarrays. *Virology* **365**, 48–59.
- Garcia-Vallve, S., Guzman, E., Montero, M. A. & Romeu, A. (2003). HGT-DB: a database of putative horizontally transferred genes in prokaryotic complete genomes. *Nucleic Acids Res* **31**, 187–189.
- Greve, B., Jensen, S., Phan, H., Brügger, K., Zillig, W., She, Q. & Garrett, R. A. (2005). Novel RepA-MCM proteins encoded in plasmids pTAU4, pORA1 and pTIK4 from *Sulfolobus neozealandicus*. *Archaea* **1**, 319–325.
- Keeling, P. J., Klenk, H. P., Singh, R. K., Schenk, M. E., Sensen, C. W., Zillig, W. & Doolittle, W. F. (1998). *Sulfolobus islandicus* plasmids pRN1 and pRN2 share distant but common evolutionary ancestry. *Extremophiles* **2**, 391–393.
- Kletzin, A., Lieke, A., Ulrich, T., Charlebois, R. L. & Sensen, C. W. (1999). Molecular analysis of pDL10 from *Acidianus ambivalens* reveals a family of related plasmids from extremely thermophilic and acidophilic archaea. *Genetics* **152**, 1307–1314.
- Lipps, G. (2007). Plasmids. In *Archaea: Evolution, Physiology and Molecular Biology*, pp. 105–112. Edited by R. A. Garrett & H. P. Klenk. Oxford: Blackwell.
- Martin, A., Yeats, S., Janekovic, D., Reiter, W. D., Aicher, W. & Zillig, W. (1984). SAV 1, a temperate u.v.-inducible DNA virus-like particle from the archaeobacterium *Sulfolobus acidocaldarius* isolate B12. *EMBO J* **3**, 2165–2168.
- Muskhelishvili, G., Palm, P. & Zillig, W. (1993). SSV1-encoded site-specific recombination system in *Sulfolobus shibatae*. *Mol Gen Genet* **237**, 334–342.
- Palm, P., Schleper, C., Grampp, B., Yeats, S., McWilliam, P., Reiter, W. D. & Zillig, W. (1991). Complete nucleotide sequence of the virus SSV1 of the archaeobacterium *Sulfolobus shibatae*. *Virology* **185**, 242–250.
- Peng, X., Holz, I., Zillig, W., Garrett, R. A. & She, Q. (2000). Evolution of the family of pRN plasmids and their integrase-mediated insertion into the chromosome of the crenarchaeon *Sulfolobus solfataricus*. *J Mol Biol* **303**, 449–454.
- Prangishvili, D., Albers, S. V., Holz, I., Arnold, H. P., Stedman, K., Klein, T., Singh, H., Hiort, J., Schweier, A. & other authors (1998). Conjugation in Archaea: frequent occurrence of conjugative plasmids in *Sulfolobus*. *Plasmid* **40**, 190–202.

- Prangishvili, D., Forterre, P. & Garrett, R. A. (2006).** Viruses of the Archaea: a unifying view. *Nat Rev Microbiol* **4**, 837–848.
- Reiter, W. D., Palm, P. & Zillig, W. (1987).** Gene expression in archaeobacteria: physical mapping of constitutive and UV-inducible transcripts from the *Sulfolobus* virus-like particle SSV1. *Mol Gen Genet* **209**, 270–275.
- Reiter, W. D., Palm, P. & Zillig, W. (1988).** Transcription termination in the archaeobacterium *Sulfolobus*: signal structures and linkage to transcription initiation. *Nucleic Acids Res* **16**, 2445–2459.
- Sambrook, J. & Russell, D. J. (2001).** *Molecular Cloning, a Laboratory Manual*, 3rd edn. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.
- Schleper, C., Kubo, K. & 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.
- Serre, M. C., Letzelter, C., Garel, J. R. & Duguet, M. (2002).** Cleavage properties of an archaeal site-specific recombinase, the SSV1 integrase. *J Biol Chem* **277**, 16758–16767.
- She, Q., Singh, R. K., Confalonieri, F., Zivanovic, Y., Allard, G., Awayez, M. J., Chan-Weiher, C. C., Clausen, I. G., Curtis, B. A. & other authors (2001).** The complete genome of the crenarchaeon *Sulfolobus solfataricus* P2. *Proc Natl Acad Sci U S A* **98**, 7835–7840.
- She, Q., Zhu, H. & Xiang, X. (2007).** Integration mechanisms: possible role in genome evolution. In *Archaea: Evolution, Physiology and Molecular Biology*, pp. 113–124. Edited by R. A. Garrett & H. P. Klenk. Oxford: Blackwell.
- Stedman, K. M., She, Q., Phan, H., Arnold, H. P., Holz, I., Garrett, R. A. & Zillig, W. (2003).** Relationships between fuselloviruses infecting the extremely thermophilic archaeon *Sulfolobus*: SSV1 and SSV2. *Res Microbiol* **154**, 295–302.
- Tolstrup, N., Sensen, C. W., Garrett, R. A. & Clausen, I. G. (2000).** Two different and highly organized mechanisms of translation initiation in the archaeon *Sulfolobus solfataricus*. *Extremophiles* **4**, 175–179.
- Torarinsson, E., Klenk, H.-P. & Garrett, R. A. (2005).** Divergent transcriptional and translational signals in Archaea. *Environ Microbiol* **7**, 47–54.
- Wang, Y., Duan, Z., Zhu, H., Guo, X., Wang, Z., Zhou, J., She, Q. & Huang, L. (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.
- Wiedenheft, B., Stedman, K., Roberto, F., Willits, D., Gleske, A.-K., Zoeller, L., Snyder, J., Douglas, T. & Young, M. (2004).** Comparative genomic analysis of hyperthermophilic archaeal fuselloviridae viruses. *J Virol* **78**, 1954–1961.
- Zillig, W., Kletzin, A., Schleper, C., Holz, I., Janecovik, D., Hain, J., Lanzendorfer, M. & Kristjansson, J. (1994).** Screening for *Sulfolobales*, their plasmids and viruses in Icelandic solfataras. *Syst Appl Microbiol* **16**, 609–628.

Edited by: J. van der Oost