Comparative genetic and genomic analysis of the novel fusellovirus Sulfolobus spindle-shaped virus 10

David A. Goodman and Kenneth M. Stedman*,†

Biology Department, Center for Life in Extreme Environments, Portland State University, P.O Box 751, Portland, OR 97207-0751, USA

*Corresponding author: E-mail: kstedman@pdx.edu
†http://orcid.org/0000-0001-6086-0238

Abstract

Viruses that infect thermophilic Archaea are unique in both their structure and genetic makeup. The lemon-shaped fuselloviruses—which infect members of the order *Sulfolobales*, growing optimally at 80 °C and pH 3—are some of the most ubiquitous and best studied viruses of the thermoacidophilic Archaea. Nonetheless, much remains to be learned about these viruses. In order to investigate fusellovirus evolution, we have isolated and characterized a novel fusellovirus, *Sulfolobus* spindle-shaped virus 10 (formerly SSV-L1). Comparative genomic analyses highlight significant similarity with both SSV8 and SSV9, as well as conservation of promoter elements within the *Fuselloviridae*. SSV10 encodes five ORFs with no homology within or outside of the *Fuselloviridae*, as well as a putatively functional Cas4-like ORF, which may play a role in evading CRISPR-mediated host defenses. Moreover, we demonstrate the ability of SSV10 to withstand mutation in a fashion consistent with mutagenesis in SSV1.

Key words: Archaea; biogeography; CRISPR/Cas

1. Introduction

In the years since their identification 40 years ago, the Archaea have been separated into four main phyla—*Crenarchaeota*, *Euryarchaeota*, *Thaumarchaeota*, and *Korarchaeota*—with many new phyla being identified using metagenomics. While these diverse archaeal lineages share some traits with both Eukarya and Bacteria, they exhibit many unique phenotypes and processes as well (Torarinsson et al. 2005; Stetter 2006; Bize et al. 2009; Wu et al. 2014; Zhai et al. 2017). Likewise, archaeal viruses—particularly those infecting Crenarchaea—come packaged in strikingly diverse and interesting ways, exhibiting unique morphologies not known to exist in other domains, including spindle-shaped, bottle-shaped, and droplet-shaped virions (Prangishvili et al. 2006; Prangishvili 2013; Rensen et al. 2016). The majority of archaeal viruses harbor either a linear or circular double-stranded DNA (ds-DNA) genome, and many archaeal virus open reading frames (ORFs) encode gene products of unknown function, sharing little or no similarity to known sequences (Krupovic et al. 2018). These Archaea-specific genes may be responsible for archaeal viruses' unique structures, methods of entry and egress from their hosts (Quemin et al. 2015), as well as possibly conferring thermotolerance. Unfortunately, the molecular basis for the morphological novelty of these viruses is still poorly understood, and the mechanisms involved in viral infection of Archaea even less so. Studying these archaeal systems can illuminate protein folding, stability, and protein–protein interactions under extreme conditions, as well as potentially expanding our understanding of viral evolution in general.

Sulfolobus is the model thermoacidophilic crenarchaeal genus, found in acidic hot springs worldwide. Members of the genus are infected by members of the viral family *Fuselloviridae* (Table 1). Major archaeal biological functions including replication, transcription, and translation have been studied using *Sulfolobus* species as model organisms (Chen et al. 2005;

Table 1. Fusellovirus genomes used in this study.

Virus genome (location)[a]	Size (bp)	Annotated ORFs	Accession	Reference
ASV1 (ISL)	24,186	38	NC_013585	(Redder et al. 2009)
SSV1 (JPN)	15,465	35	NC_001338	(Palm et al. 1991)
SSV2 (ISL)	14,795	35	NC_005265	(Stedman et al. 2003)
SSV3 (ISL)	15,230	32	KY579375	(Stedman et al. 2006)
SSV4 (ISL)	15,135	34	NC_009986	(Peng 2008)
SSV5 (ISL)	15,330	34	NC_011217	(Redder et al. 2009)
SSV6 (ISL)	15,684	33	NC_013587	(Redder et al. 2009)
SSV7 (ISL)	17,602	33	NC_013588	(Redder et al. 2009)
SSV8 (USA)	16,473	37	NC_005360	(Wiedenheft et al. 2004)
SSV9 (RUS)	17,385	31	NC_005361	(Wiedenheft et al. 2004)
SSV10 (USA)	16,271	40	KY563228	This work

[a](Location): ISL, Iceland; JPN, Japan; RUS, Russia.

Barry and Bell 2006; Duggin and Bell 2006; Samson et al. 2013). Furthermore, the mechanisms and capacity of the crenarchaeal CRISPR/Cas arrays have been extensively studied in recent years, highlighting the complex nature of these adaptive immune systems. *Sulfolobus* species and their close relatives utilize a combination of CRISPR/Cas type I-A acquisition complex along with type I-A and type III-B interference complexes (Gudbergsdottir et al. 2011; Lemak et al. 2013; He et al. 2016; Liu et al. 2016; Liu et al. 2017). However, very few studies have examined the CRISPR/Cas response in response to direct interaction of *Sulfolobus* with its naturally occurring viruses (Erdmann et al. 2014; Fusco et al. 2015a; Manica et al. 2011, 2013).

Currently, *Sulfolobus* spindle-shaped virus 1 (SSV1)—isolated from Beppu, Japan (Martin et al. 1984)—is the best-understood member of the *Fuselloviridae*. The complete sequence of the 15.5 kb circular, ds-DNA genome of SSV1 was determined in 1991 (Palm et al. 1991). SSV1 replicates as an episome inside infected cells as well as a spindle-shaped virion. Virus production does not appear to kill host cells but can be induced by UV irradiation (Martin et al. 1984), which, within the fuselloviruses, seems to be unique to SSV1 (Reiter et al. 1988; Stedman et al. 2003). UV induction was utilized in conjunction with Northern analyses to identify a total of 11 transcripts spanning 34 ORFs (Reiter et al. 1988; Fröls et al. 2007a). Additionally, a newly discovered transcript, T_{lys}, is thought to play a role in maintaining repression of SSV1 replication in the non-induced state (Fusco et al. 2013). Most recently, SSV1 was shown to be surprisingly tolerant to mutation, with ~50 per cent of ORFs able to tolerate insertion or deletion without completely abrogating activity (Iverson et al. 2017a). This is intriguing as presumably some of these novel genes and their unique characteristics at high temperatures are responsible for both infection processes and thermotolerance of SSVs.

Attempting to determine the function of novel fusellovirus protein products is challenging, however, as they lack clear genomic homology outside of the viral family, with only ORF D335—a viral integrase of the tyrosine recombinase family—having unambiguous similarity to previously characterized proteins (Reiter et al. 1989; Muskhelishvili et al. 1993; Eilers et al. 2012). Nevertheless, structural studies and proteomic analyses have illuminated some functional aspects of the life cycle of viruses of Archaea (Kraft et al. 2005; Kraft et al. 2004a,b; Schlenker et al. 2012; Eilers et al. 2012). This work furthers understanding of archaeal fuselloviruses, via a comparative genetic and genomic analysis of the novel SSV10 (formerly SSVL-1) with other members of the *Fuselloviridae*.

SSV10, isolated from a hot spring in Lassen Volcanic National Park, encapsidates a 16.3 kbp circular, double-stranded DNA genome. SSV10 encodes 40 ORFs, of which five have no homology to other ORFs within the *Fuselloviridae*. Moreover, SSV10 has a variable host range, infecting six of thirteen tested *Sulfolobus* isolates, three more than SSV1 (Ceballos et al. 2012). SSV10 also encodes a putatively functional Cas4-like ORF, which may dictate host range and play a role in evasion of host CRISPR/Cas interference. An in-depth analysis of genetic relationships within the *Fuselloviridae* is also reported herein. We also demonstrate the ability of SSV10 to tolerate transposon mutagenesis similarly to SSV1 (Iverson et al. 2017a) and present deletion mutagenesis in SSV10.

2. Methods

2.1 Archaeal and bacterial strains

Sulfolobus solfataricus strain 441 (S441) is a *Sulfolobus* isolate used as the host for the experiments presented in this work (unless otherwise specified). S441 was originally isolated from Devil's Kitchen in Lassen Volcanic National Park (Ceballos et al. 2012) and is susceptible to infection by multiple wild-type SSVs as well as an array of SSV1 deletion and insertion mutants (Iverson et al. 2017a,b). All Sulfolobus strains, both infected and uninfected, were isolated as in Zillig et al. (1993). Briefly, single colonies were isolated from enrichment cultures and screened for virus production by spot on lawn assays of uninfected Sulfolobus strains. *Sulfolobus* strain S355 is also an isolate from Devil's Kitchen, from which SSV10 (formerly SSV-L1) was purified. Transformax EC100D pir+ *Escherichia coli* (Epicentre) was used as host for transformations of transposon-containing constructs.

2.2 Growth media and culturing conditions

All *Sulfolobus* strains, both infected and uninfected, were cultured aerobically in Yeast-sucrose (YS) liquid media or on Gelrite plates at 75°–80 °C as described previously (Iverson and Stedman 2012). Media pH was adjusted to between 3.0 and 3.2 using 50 per cent H_2SO_4 and autoclaved prior to use. *E. coli* strains were grown aerobically in LB media both liquid and on agar plates supplemented with 50 µg/ml Kanamycin when required (Green and Sambrook 2012). Table 2 contains a complete list of the strains used in this study and their genotypes.

Table 2. *Sulfolobus* and *E. coli* Strains.

Strain	Description/genotype	Reference
E. coli EC100D *pir*[+]	F[−] mcrA Δ(mrr-hsdRMS-mcrBC) ϕ80dlacZΔM15 ΔlacX74 recA1 endA1 araD139 Δ(ara, leu)7697 galU galK λ[−] rpsL (Str[R]) nupG pir+(DHFR)	Epicentre, Inc
S. solfataricus S441	*S. solfataricus* isolate; SSV host	(Ceballos et al. 2012)
S. solfataricus S355	Original SSV10-infected isolate	This work
S. solfataricus S592	S441 infected with SSV10	This work
S. solfataricus P1	*S. solfataricus* isolate with complete genome sequence	DSM 1616 (Zillig et al. 1980; Liu et al. 2016)
S. solfataricus P2	*S. solfataricus* isolate with complete genome sequence	DSM 1617 (Zillig et al. 1980; She et al. 2001)
S. solfataricus GΘ	*S. solfataricus* MT4 derivative *lacS*	(Cannio et al. 1998)

2.3 Isolation of viral DNA

SSV10 episomal DNA was isolated from *Sulfolobus* strain S355 first by alkaline lysis, then treated with phenol: chloroform: iso-amyl alcohol (25:24:1) (Schleper et al. 1992) and ethanol precipitation. DNA extraction was confirmed via agarose gel electrophoresis after restriction endonuclease digestion with *Eco*RI (New England Biolabs) before being further purified using a GeneJet Plasmid Miniprep Kit (Thermo-Fisher).

2.4 Transposon mutagenesis of the SSV10 genome

The EZ-Tn5™ < R6Kγori/KAN-2 > Tnp Transposome™ Kit (Lucigen) was used to generate insertional mutations in SSV10. The EZ-Tn5™ transposon is 2001 base pairs long and contains an *E. coli* origin of replication as well as a kanamycin resistance gene for selection. As in Iverson et al. (2017a), we replaced the manufacturer's recommended equimolar ratio of target DNA to EZ-Tn5 transposon with a 30:1 molar ratio (Iverson et al. 2017a). SSV10::Tn5 reactions were transformed into Transformax EC100D pir[+] *E. coli* and grown on LB agar plates with kanamycin. Plasmid DNA purified from single colonies were isolated and screened for mutagenized SSV10 DNA via *Eco*RI or *Eco*RV endonuclease digestion.

2.5 Knockout mutants of SSV10

Deletions of SSV10 ORFs were made using long-inverse polymerase chain reaction (LIPCR) as described previously (Clore and Stedman 2007; Iverson and Stedman 2012; Iverson et al. 2017a,b). The T_m calculator software (http://tmcalculator.neb.com/#!/) was used to estimate the annealing temperatures for each primer pair (Supplementary Table S1). Optimal conditions (template concentration, primer melting temperature [T_m], extension time, etc.) were derived experimentally for each set of primers. LIPCR primers were designed to excise as much of the ORF as possible without disrupting flanking coding regions, ideally leaving only the start and stop codons intact. Deletion mutants were generated in the SSV10::Tn5 shuttle vector DAG593. Purified template DNA, ranging between 150 and 250 ng/μl, was initially diluted 1:30 in 30 μl of TE Buffer, then further diluted 10 and 100-fold in either TE Buffer or double-distilled water. LIPCR was performed using Phusion® DNA polymerase at a concentration of 0.005 U as previously described (Iverson and Stedman 2012). The blunt ends of linearized LIPCR products were phosphorylated using T4 polynucleotide kinase (Thermo-Fisher), ligated using T4 ligase (Thermo-Fisher), and transformed into chemically competent Transformax® EC100D pir[+] *E. coli*.

2.6 Testing Tn5 mutants for infectivity

Sulfolobus cultures at mid-logarithmic growth—an OD$_{600}$ of 0.20 to 0.25—were prepared for transformation essentially as in Schleper et al. (1992). Cells (50 ml) were pelleted (15 min.@3,000g)

and washed in decreasing volumes of 20 mM sucrose to a final volume of 400 μl with a final concentration of ~10^{10} cells/ml. One hundred microliters of washed cells were added to a chilled 0.1-cm-gap-length cuvette (VWR), and 2 μl of SSV10 or SSV10::Tn5 (~300–1,000 ng total DNA) was added to the cells. Transformations were performed via electroporation (Gene Pulser II; Bio-Rad) at 1.5 kV, 400 Ω, and 25 μF. Transformed cells were immediately diluted with 1 ml of 75 °C YS media, transferred to a 1.5-ml microcentrifuge tube, and allowed to recover for at least 1 h at 75 °C. Following recovery, cells were transferred to 50 ml of preheated YS in long-neck Erlenmeyer flasks and grown with shaking between 74° and 78 °C.

Spot-on-lawn (halo) assays were performed as in Iverson and Stedman (2012). Briefly, 5 μl of transformed cultures were spotted on a lawn of uninfected *Sulfolobus* at mid-logarithmic growth (OD$_{600}$ = 0.2–0.5) in a 0.2 per cent YS Gelrite soft-layer poured over a 1 per cent YS Gelrite plate. Plates were incubated at 75 °C for 48 to 96 h. Halo assays were prepared in duplicate, typically 72 and 96 h after transformation of *Sulfolobus*.

Wild-type SSV10 DNA and known functional mutant DAG593 were used as positive controls. Negative controls were uninfected *Sulfolobus* cultures. To confirm the identity of the viral DNA in cultures that inhibited host growth, viral DNA was purified from transformed *Sulfolobus* cells and amplified via PCR using specific primers (Supplementary Table S1) flanking the mutated region of the viral DNA. Control PCRs were done using the DNA used for transformation and wild-type SSV10 DNA.

2.7 Sequencing and bioinformatic analysis

Sequencing was performed using transposon-specific primers on 172 SSV10 plasmids containing the EZ-Tn5™ < R6Kγori/ KAN-2 > Tnp transposon. Gaps were closed and ambiguous sequences determined using custom primers (Supplementary Table S1). Sequencing was performed either by Eurofins© Genomics or the DNA Services Core at the Oregon Health & Science University. Geneious™ V. 8.1 (Biomatters, Inc.) was used to identify ORFs in SSV10, and for generating nucleotide and protein alignments. Alignments are available on request. Similar sequences to SSV10 ORFs were identified in the NCBI nr database using the Basic Local Alignment Search Tool (BLASTP) (Altschul et al. 1997). Searching for CRISPR spacers in *Sulfolobus* genomes was done using the BLAT search engine provided in the UCSC Archaeal Genome Browser (Kent 2002; Chan et al. 2012). Specific start codons for each ORF were determined using a combination of the best fit per BLASTP e-value, likely usage rate (Romero and García 1991; Torarinsson et al. 2005) of any particular start codon, and the presence of canonical *Sulfolobus* ribosome-binding sites (Torarinsson et al. 2005). Putative transcripts were identified by the presence of predicted upstream
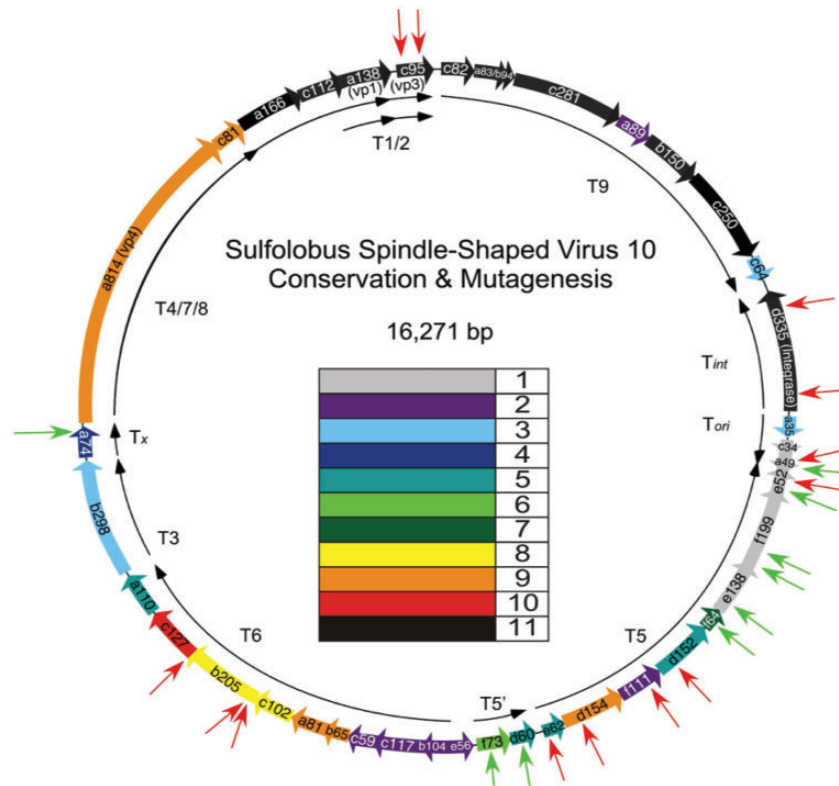
Figure 1. SSV10 Conservation and mutagenesis. Overall conservation (E-value $\leq 1e^{-3}$) of SSV10 ORFs (wide arrows with labels) compared to 10 other members of the Fuselloviridae (Table 1). Completely conserved 'core' ORFs are highlighted in black, unique ORFs are highlighted in grey, intermediately conserved ORFs are color-coded as in the inset ranging from violet if only conserved in 2 Fusellovirus genomes to red if conserved in 10 Fusellovirus genomes. ORFs are labeled as in Genbank Accession KY563228.1. Virus capsid genes and the integrase gene are labeled in parentheses. Putative transcripts are labeled with a capital T. See text for ORF and transcript annotations. Thin arrows outside the genome map represent the location of Tn5 insertion mutants characterized as functional (green) and nonfunctional (red).

promoter elements, identified by their respective TFB-recognition elements (BRE) with adjoining TATA box (Reiter et al. 1987; Kosa et al. 1997; Fröls et al. 2007a; Fusco et al. 2013) and presence of ORFs. Structural models of putative SSV10 proteins were generated using the Protein Homology/analogY Recognition Engine V. 2.0 (Phyre2) web portal for protein modeling, prediction and analysis (Kelley et al. 2015).

2.8 Transmission electron microscopy

For transmission electron microscopy, samples were prepared on 400-mesh carbon-Formvar-coated copper grids (Ted Pella, Redding, CA, USA). Culture supernatants were generated by centrifugation at 6,000*g* for 5 minutes. Grids were placed, carbon-Formvar down, on a 5-μl droplet of culture supernatant for 2 minutes. Samples were removed from the grid by wicking. Grids were then stained for 15–60 seconds on 5 μl of 2 per cent uranyl acetate stain (pH 3). Grids were allowed to dry in air overnight and were examined within 48 h of staining. Images were obtained at magnifications ranging from 8,500 to 34,000 on an FEI Tecnai F20 transmission electron microscope (TEM) (FEI Inc. Hillsboro, OR, USA).

3. Results and discussion

3.1 SSV10 is a novel member of the *Fuselloviridae*

An SSV-producing *Sulfolobus* strain, S355 (Table 2), was isolated from Devil's Kitchen in Lassen Volcanic National Park (Clore 2008; Ceballos et al. 2012). Episomal DNA, originally SSV-

L1, now SSV10 (Iverson et al. 2017a), purified from S355 was transformed into the laboratory *Sulfolobus* host S441 via electroporation (Schleper et al. 1992), generating *Sulfolobus* strain S592 (Table 2). Strain S592 produced infectious virus as assessed by halo assay, transmission electron microscopy, and subsequent recovery of SSV10 DNA from transformed cultures (Iverson and Stedman 2012). The SSV10 genome was sequenced from transposon insertions. The GC content of SSV10 is 37.1 per cent, consistent with the 38.5 per cent median of fusellovirus GC content. The SSV10 genome (Accession number KY563228) was annotated to have 40 ORFs using comparisons to previously annotated SSV genomes, a minimum ORF length of 34 codons, usage of start codons AUG, UUG, and GUG, and the presence of a ribosome-binding site (Fig. 1). BRE and corresponding TATA boxes found in SSV1 (Reiter et al. 1987; Fröls et al. 2007a; Fusco et al. 2013) were used to annotate 10 putative transcripts in SSV10 (Fig. 1).

The SSV10 genome and its ORFs were compared to 10 other members of the *Fuselloviridae*. SSV10 shares a unique admixture of features from different SSVs with a mixture of 'core' genes and variable genes, with most of the variable genes in the T5 and T6 'early' transcripts (Reiter et al. 1987; Fröls et al. 2007a), similar to other SSVs (Fig. 1). SSV10 has an average sequence identity of ∼41 per cent within the family and is most similar to Yellowstone National Park's SSV8 (formerly SSVRH), the only other published SSV isolated from North America (Wiedenheft et al. 2004), at 54.7 per cent.

Conservation of ORFs ranges from 12 completely conserved 'core' ORFs to five unique to SSV10 (Fig. 1). About 68 per cent

Table 3. SSV10 mutagenesis.

Plasmid	Position[a]	ORF	Transcript	Positive[b]	Negative	Functional
DAG593	5148	F199	T5	20	2	Y
DAG594	9607	B205	T6	0	3	N
DAG595	7302	E62	T5	0	3	N
DAG597	5673	E138	T5	3	2	Y
DAG615	6453	F111	T5	0	3	N
DAG616	7508	D60	T5'	2	1	Y
DAG627	4699	E52	T5	5	4	Y
DAG637	7743	F73	T5'	2	1	Y
DAG638	5820	F64	T5	3	0	Y
DAG660	4013	INT (D335)	T*int*	0	4	N
DAG674	6261	D152	T5	0	3	N
DAG676	12003	A74	Tx	3	3	Y
DAG681	15995	VP3	T1/T2	0	3	N
DAG685	10144	B205/C127	T6	0	5	N
DAG693	7085	D154	T5	0	3	N
DAG698	5279	F199	T5	4	1	Y
DAG699	4545	A49	T*ori*	2	1	Y
DAG702	4468	A49	T*ori*	0	3	N
DAG719	3203	INT (D335)	T*int*	0	4	N
DAG720	4619	E52	T5	0	3	N
DAG787	16182	VP3	T1/T2	0	3	N
DAG788	9619	B205	T6	0	3	N
DAG821	9530–10147	ΔB205	T6	3	5	Y
DAG825	15984–16271	ΔVP3	T1/T2	3	1	Y

[a]All are Tn5 insertions except DAG821 and DAG825 which are deletions generated in DAG593.
[b]Number of independent transformations that generated cultures that inhibit growth of uninfected *Sulfolobus*.

of ORFs encoded by SSV10 are most similar in amino acid sequence to those in either SSV8 or SSV9 (the latter from Kamchatka, Russia). Although the arrangement and conservation of 'core' ORFs within the genome of SSV10 is similar to that of other SSVs, two ORFs in the T9 transcript are of note. SSV10 ORF C64, homologues of which are only found in ASV1 and SSV8, is located downstream of the highly conserved ORF C250. Additionally, SSV10 ORF A89 is the first known homologue of SSV1 ORF C102a. The region of SSV10 upstream of the integrase gene encodes five novel ORFs, three of which are located on the opposite strand and appear to have their own transcript (T*ori*). Thus, SSV10 is a genetically distinct addition to the *Fuselloviridae*, the first isolated fusellovirus from Lassen Volcanic National Park in the USA, and only the second North American SSV published to date.

Since it has been shown that archaeal transcription can result in both leaderless mRNAs or mRNAs containing 5'-untranslated regions (Nakagawa et al. 2017), translational start sites may or may not correspond to the first start (AUG, GUG, and UUG) codon of the transcript. SSV10 ORFs C34, A49, and D60 are annotated such that they would require the use of noncanonical start codons CUG, AUU, and AUA, respectively (Table 1).

3.2 SSV10 mutagenesis

To date, SSV1 is the only member of the *Fuselloviridae* whose genetic requirements have been analyzed via mutagenesis. In an effort to further understand the broader genetic requirements within the family, we screened our SSV10 transposon mutants for function via growth inhibition or halo assays as in Iverson et al. (2017a) (Fig. 1; Table 3). SSV10 is genetically tractable,

tolerating both insertion and deletion mutations in a manner similar to SSV1 (Clore and Stedman 2007; Stedman et al. 1999; Iverson and Stedman 2012; Iverson et al. 2017a). A mutant with an insertion in the putative viral nuclease ORF F199 was shown to be functional in the first subset of trials and was subsequently used as the positive control for transformation. *Sulfolobus* transformed with this mutant, DAG593, generated halos in ∼90 per cent of trials (Table 3). Negative results for transformations of mutant DNAs that in other cases generate virus are likely to be false negatives, as transformations are not 100 per cent efficient (Iverson et al. 2017a).

Unique insertion mutations were obtained in thirty-five out of forty SSV10 ORFs, including all ORFs in the putative T5 and T5′ transcripts. Sixteen different mutants in fourteen different ORFs in this coding region were analyzed, as well as insertion mutants in the T6 and T3 transcripts, and the minor capsid gene VP3, all comparable to known functional mutants in SSV1 (Table 4). Two deletion mutants, DAG593_ΔB205 (DAG821) and DAG593_ΔVP3 (DAG825), were also generated using LIPCR (Clore and Stedman 2007; Iverson and Stedman 2012). All functional SSV10::Tn5 insertion and deletion mutants were confirmed via DNA extraction from transformed *Sulfolobus* and subsequent PCR amplification of the mutagenized region (Supplementary Table S1).

Consistent with mutagenesis in SSV1, SSV10 ORFs encoded on the T5 and T5′ transcripts generally tolerate transposon insertionwithout abrogating function. In total, eight of sixteen transposon mutants in the T5 transcripts were functional. Most of these insertions were in ORFs that contain putative DNA binding domains characteristic of transcription regulatory proteins. Furthermore, all eight of these functional mutants occurred in the 3′-end of the T5 transcript, or in the two ORFs in

Table 4. SSV10 Putative Promoters.

Transcript	Promoter sequence[a]	SSV1 Match (BRE/TATA)	Features
T1/2	TTCTGAATTCAGAACTAGGGGGGTTTAAAAAGCTTAGTGATAAGCCCCTA *TTGACCAAGGATG*	Y/Y	One mismatch in BRE
T3	ATTTCGTAATGCATCTTTTTT**AGGCCC**TTTATAAAGTTACACTTTCCTTTTT *CGCTTACAATG*	N/N	Matches T9 promoter
T4/7/8	TTCTTCGTAAGACGAAAATAGATT**AAGCCC**TTTATAAAGTCACATAATTTTTATCGCTTAATG	Y/Y	One mismatch to T9 BRE
T5	AGAAAGAGAGATAGATGAACA**GAAAGA**TTTATATAGTCAGATAGATAGAT *AGATAAATTG*	Y/N	One mismatch in BRE; no direct repeats
T6	TCttgatagattgatagata**GAAAAA**<u>TTTATATA</u>CTCAGAttgatagattgataaata GAGGG *TCAAAAAATG*	Y/N	One mismatch in repeats, BRE
T9	GTATAAAATCTACAAAGACTGAGT**AGGCCC**TTTATAAAGTCATTTTCTTTTTTTCATTCAATG	Y/Y	Most highly conserved
Tx	TACCACATATGCACTCTAAGGC**AAAATT**TTTTATCCTTTCTTTTTTATATGTTTAATCAAATG	N/N	–
Tint	GATAAGATAATTA**TCATCC**TTTTTAAATGCTTACGTGATAAATATAAATGGGCTGAAGGTATG	–	Novel transcript
Tori	AAAACTTTATTACCCATACCTTC**AGCCCA**TTTATATTTATCACGTAAGCATTTAAAAAGGATG	–	Novel transcript
T5′	TTCAGTCCCTTCGTTTTTC**ATGTCC**TTTATTTTTGCATATAACTTGTGATATGAGAAGGATTG	–	Novel transcript

[a]TFB-Recognition Element in bold; TATA-box underlined; Start/RBS in italics; T6 direct repeats in lowercase.

the T5′ transcript (Fig. 1). Mutants in both ORFs of the novel T5′ transcript were functional, which is not surprising considering that they are relatively poorly conserved. ORF F73 in the T5′ transcript is a predicted copG-like protein (Table 5) and may either be self-regulating or a regulator of transcription in general, similar to ORF F55 of SSV1, which was also shown to tolerate mutation (Iverson et al. 2017a). However, starkly different from SSV1, homologues of these ORFs in SSV2 were the last to be upregulated (Ren et al. 2013), which may implicate them in establishing the so-called 'carrier state' once the infection cycle has completed. The same may be true of SSV10 ORFs E62 and B65 and their homologues, both which also encode RHH copG-like products in the T5 and T6 transcripts respectively and are expressed late (∼4.5—6 h.p.i.) in SSV2 (Ren et al. 2013).

The four SSV10 ORFs starting from the 5′-end of the T5 transcript—E62, D154, F111, and D152—did not tolerate transposon insertion. SSV10 ORF D154 is predicted to be a MarR-like transcriptional repressor (Fig. 1; Table 5), and if induced early in the infection process, is likely involved in repression of host response to infection. Interestingly, no homologues of these four ORFs are found in SSV1. The essential nature of these ORFs in SSV10 and lack of conservation in SSV1 indicates that SSV1, or SSVs in general, encode functionally similar proteins with different sequences. Conversely, these functions may only be required in SSV10.

Insertions in the viral nuclease ORF F199 are functional, although it is not known whether or not they generate a more virulent phenotype, as was seen with mutants in ORF D244 in SSV1 (Iverson and Stedman 2012). Halos generated by SSV10 ORF F199 insertion mutants were similar to those generated by wild-type SSV10 (Fig. 3). Two mutants in ORF E52 were characterized; one was functional while the other was not. The insertion in the functional E52 mutant is located just one residue downstream of the annotated ATG start, yet there are two more ATG start codons 3 and 7 in-frame residues downstream, which may compensate for the disruption. Conversely, the insertion in the non-functional ORF E52 mutant is located about halfway into the ORF, indicating that the gene product of E52 may be critical for virus function.

Insertions in the SSV10 integrase gene, including one nearly identical to a functional mutant of SSV1 (Iverson et al. 2017a), were all found to be non-functional when transformed into

Sulfolobus strain S441. Furthermore, unlike integrase mutants in SSV1, SSV10 integrase mutants DAG660 and DAG719 (Table 3) were nonfunctional when transformed into *S. solfataricus* strain GΘ. Mutant DAG719 was also nonfunctional when transformed into both *S. solfataricus* strains P1 and P2. SSV integration into the host genome has recently been implicated as a sort of 'mutually assured destruction' method of survival by which virus integration forces the host CRISPR/Cas response to either disable its own CRISPR array or risk self-harm by cleavage of the integrated provirus (Fusco et al. 2015a). Differences between SSV1 and SSV10 as it pertains to integrase mutants may be indicative of the variance in life cycles between members of the *Fuselloviridae* or that there is an integrase gene in *Sulfolobus* strains GΘ and P2 that can complement SSV1 integrase mutants, but not SSV10 integrase mutants.

3.3 The SSV10 putative Cas4 homolog ORF B205

Sequence similarity searches done on the SSV10 ORF B205 using BLASTP (Altschul et al. 1997) (NCBI) predict a putative CRISPR associated Cas4-like protein product (E-value = $4e^{-10}$) that is homologous to ORFs present in eight out of eleven members of the *Fuselloviridae* but is not present in SSV1 (Figs. 1 and 5). *Sulfolobus* genomes encode CRISPR/Cas arrays combining the well conserved type I-A acquisition complex involving Cas4 proteins, along with type I-A and type III-B interference complexes which have been shown to be able to target and neutralize foreign extrachromosomal plasmids *in vivo* (Gudbergsdottir et al. 2011; Manica et al. 2011, 2013; Plagens et al. 2012). For some type I-A systems, Cas4 has been shown form a complex with Cas1 and Cas2, or in some cases, fused to Cas1 directly (Plagens et al. 2012). In *S. solfataricus*, monomers of the Cas4 protein SSO0001 come together to form a decameric toroidal quaternary structure thought to be part of the CRISPR/Cas spacer acquisition complex along with Cas1 and Cas2 (Zhang et al. 2012; Lemak et al. 2013).

Structural predictions of SSV10 ORF B205 generated using Phyre2 (Kelley et al. 2015) show high-confidence (>90%) matches with various known archaeal and bacterial Cas4 and RecB protein structures, most notably with the Cas4 SSO0001 encoded by *S. solfataricus* (Fig. 2a; Table 5). Moreover, SSV10 ORF B205 and all other SSV-encoded Cas4-like ORFs contain the

Table 5. Phyre2 protein structural predictions for SSV10 ORFs [TS]

ORF[a] (Transcript)	Phyre2 prediction (PDB)	Coverage/confidence (Phyre2)	Predicted function
C82 (T9)	Transmembrane protein	—[b]	Replication complex (2 TMH[c])
A83 (T9)	—	—	Replication complex
B94 (T9)	Transmembrane protein	—	Replication complex (3 TMH)
C281 (T9)	Transmembrane protein	—	Replication complex (1 TMH)
A89 (T9)	—	—	Replication complex
B150 (T9)	—	—	Replication complex
C250 (T9)	DnaA-like AAA+ ATPase	75%@>90	Replication initiation
C64 (T9)	—	—	—
D335 (Tint)	Tyrosine recombinase (PDB: 3UXU)	95%@100	Viral integration
A35 (Tori)	—	—	—
C34 (Tori)	—	—	—
A49 (Tori)	Transmembrane protein	—	Replication complex (1 TMH)
E52 (T5)	—	—	Transcriptional Regulator
F199 (T5)	SSV8 ORF D212 (PDB: 2W8M)	49%@>90	Viral nuclease
E138 (T5)	—	—	—
F64 (T5)	SSV1 ORF D63 (PDB: 1SKV)	92%@100	ROP-like regulator
D152 (T5)	Polymerase II elongation factor ell2	33%@53	Transcriptional regulator
F111 (T5)	Transmembrane protein	—	Replication complex (2 TMH)
D154 (T5)	MarR-like transcriptional regulator (PDB: 3F3X)	72%@>90	Host Interference
E62 (T5)	SSV8 ORF E73/transcriptional repressor (PDBe: 4aai)	77%@60	RHH CopG-like regulator
D60 (T5')	—	—	—
F73 (T5')	SSV8 ORF/transcriptional repressor (PDBe: 4aai)	98%@100	RHH CopG-like regulator
E56	—	—	—
B104 (T6)	SSV1 ORF A100 (PDBe: 4lid)	61%@87.4	DNA binding scaffold
C117 (T6)	—	—	—
C59 (T6)	—	—	—
B65 (T6)	RHH protein	66%@>90	RHH CopG-like regulator
A81 (T6)	Eukaryal C2H2 zinc finger	94%@>90	Transcriptional regulator
C102 (T6)	Transmembrane protein	—	Replication complex (2 TMH)
B205 (T6)	*S. solfataricus* Cas4 SSO0001 (PDB: 4IC1)	95%@100	Anti-CRISPR/DNA repair
C127 (T6)	SSV1 ORF B129 C2H2 Zinc Finger (PDB: 2WBT)	99%@100	Transcriptional regulator
A110 (T6)	Transmembrane protein	—	Virion assembly/egress (3 TMH)
B298 (T3)	Transmembrane protein	—	Virion assembly/egress (5 TMH)
A74 (Tx)	Transmembrane protein	—	Virion assembly/egress (1 TMH)
A814 (VP4) (T4/7/8)	—	—	Terminal fiber
C81 (T4/7/8)	Transmembrane protein	—	Virion assembly/egress (2 TMH)
A166 (T4/7/8)	Transmembrane protein	—	Virion assembly/egress (4 TMH)
C112 (T4/7/8)	ArsR-like transcriptional regulator (PDB: 3F6V)	78%@>90	Host interference (1 TMH)
A138 (VP1) (T1/2)	—	—	Major capsid protein
C95 (VP3) (T1/2)	—	—	Minor capsid protein

[a]Colors correlate to Fig. 1a.
[b]No significant prediction.
[c]Transmembrane helices.

requisite quad-cysteine iron-sulfur [4Fe-4S] binding site as well as conserved motifs required for manganese (Mn^{2+}) binding at the RecB-like active site that allow ATP-independent unwinding of dsDNA and $5' \rightarrow 3'$ exonuclease activity (Lemak et al. 2013) (Fig. 2b). However, it is currently unknown whether these virally encoded Cas4 genes are capable of forming higher order structures leading to unwinding and exonuclease activities similar to their host-encoded counterparts (Lemak et al. 2013).

Interestingly, Cas4 homologs are seemingly missing from the SSV1 and ASV1 genomes, leading to the assumption that SSV1 and ASV1 have developed alternative ways of evading host defenses. The *Thermoproteus tenax* virus 1 encodes a deconstructed version of a Cas4 nuclease, which seems to have split into two parts, losing its canonical function and repurposed as the nucleocapsid protein TP1 (Krupovic et al. 2015). While this possibility cannot be ruled out in the case of SSV10 and others at this time, it is highly unlikely given the presence of a full-length protein product complete with necessary functional motifs. Furthermore, upregulation of Cas4 has been observed in multiple CRISPR/Cas type I-A systems when abiotic stress is induced, leading to the idea that it may play multifunctional roles apart from the canonical CRISPR/Cas response (Fröls et al. 2007b; Plagens et al. 2012). Since Cas4 in particular has been characterized as a 'first responder' in the event of abiotic stressors, these functions cannot be ruled out as it relates to SSV10.

3.4 Mutations of SSV10 ORF B205

Due its similarity with Cas4, mutants in SSV10 ORF B205 were tested (Table 4). Two Tn5 insertion mutants at or near-codons for
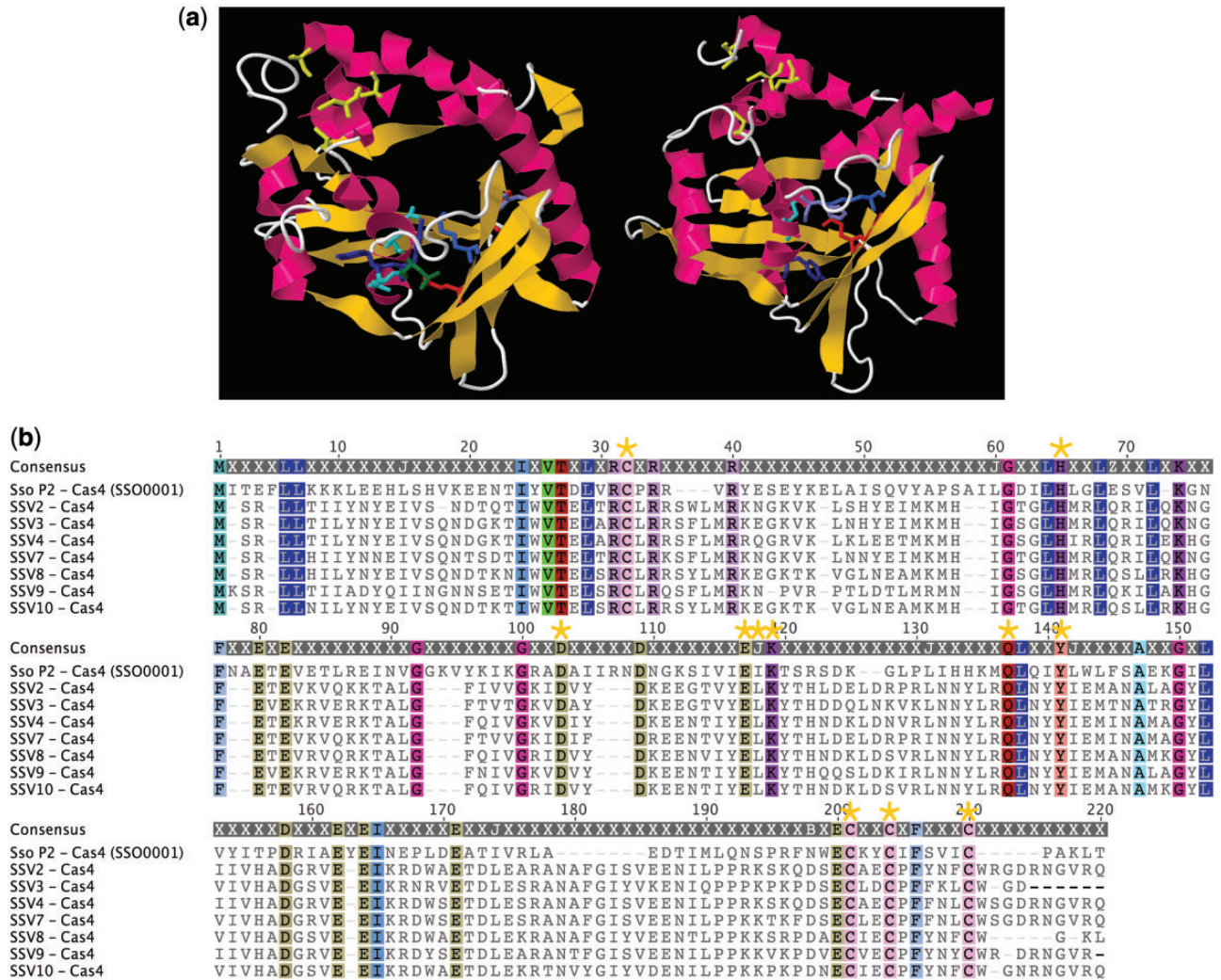
Figure 2. (a) Comparison of Cas4 in SSV10 and *S.solfataricus*. The predicted structure of the Cas4-like gene product from SSV10 (left) compared to a SSO0001 monomer from *S. solfataricus* (Lemak et al. 2013; Zhang et al. 2012) (right). The quad-cysteine [4Fe-4S] binding site residues are in yellow, while the active site residues are colored in the center of the structure. SSO0001 is rotated ~45° on the vertical access to better show the active site. (b) Alignment of Fusellovirus Cas4 homologues and *S. solfataricus* Cas4. The putative Cas4-like genes of the Fuselloviridae aligned to SSO0001 of *S. solfataricus*. Identical amino acid residues are highlighted. Active site residues and [4Fe-4S] cysteines are marked with gold stars.

active site residues within B205 (Fig. 2b) were examined for functionality. A third Tn5 mutant, located at the C-terminal amino acid, residue 203, was chosen for comparison. This C-terminal mutation also interrupts the N-terminal portion of the conserved downstream SSV10 ORF C127. A deletion of SSV10 ORF B205 (DAG821) (Table 3) was generated using LIPCR removing amino acid residues 2–196, leaving the start and stop codons intact along with a short stretch of C-terminal amino acids which overlap with the annotated start of the downstream ORF C127.

Sulfolobus cultures transformed with SSV10 containing Tn5 insertions in ORF B205 (DAG594, DAG685, and DAG788) did not generate halos after eleven separate transformations when spotted on lawns of uninfected *Sulfolobus* and were thus characterized as non-functional (Table 3). Conversely, cultures transformed with the deletion mutant DAG821 (ΔB205) generated halos in three out of eight trials. *Sulfolobus* infected with DAG821 exhibit a phenotype consistent with an active host CRISPR/Cas response to infection (Erdmann et al. 2014; Manica et al. 2011, 2013). Spot-on-lawn assays performed at 72 and 96 h post transformation generated halos, indicating production of

infectious particles in the transformed culture. Halos produced 96 h post transformation were significantly diminished compared to halos at 72 h post transformation. At 120 h post transformation, no observable halos were produced (Fig. 3). Nonetheless, PCR amplification of DNA purified from transformed cells showed that viral DNA lacking ORF B205 was still present in the culture after 1 week of growth. When this transformed strain was plated as a lawn, no spontaneous plaques formed. Moreover, cultures transformed with both SSV1 and SSV10::Tn5 mutants formed halos of growth inhibition on lawns of the transformed strain containing DAG821.

If ORF B205 and its homologues are indeed active Cas4 proteins, they may represent an archaeal virally encoded anti-CRISPR/Cas, similar to mechanisms which aid bacterial viruses in evading the host CRISPR/Cas system (Bondy-Denomy et al. 2012, 2015; Pawluk et al. 2014; Hooton and Connerton 2015; Chowdhury et al. 2017; Seed et al. 2013). This, along with the fact that insertion mutations lead to loss of function, point to ORF B205 playing a role in the evasion of host defenses, however, these insertions may have led to deleterious polar effects.
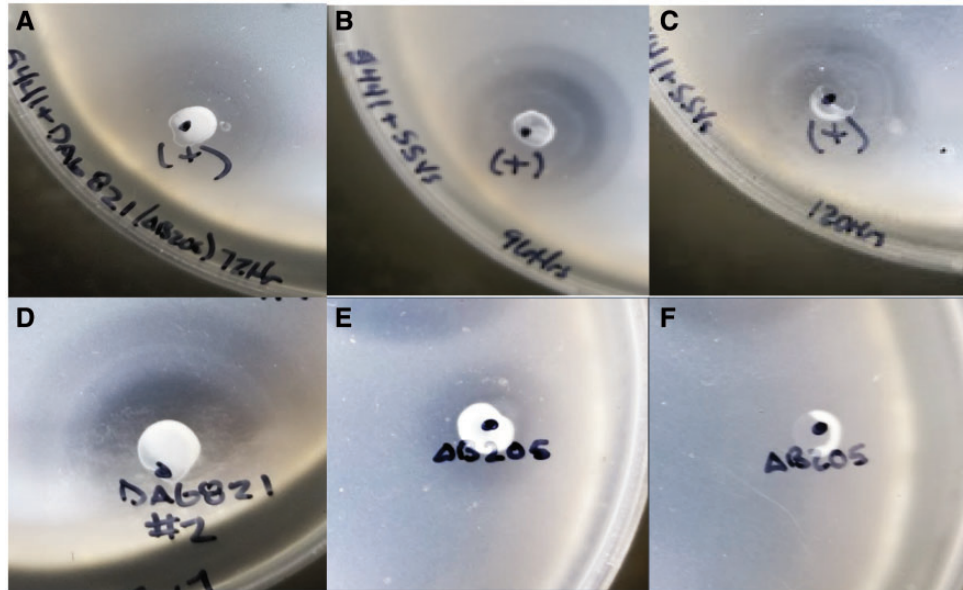
Figure 3. The SSV10 ΔB205 mutant loses the ability to slow *Sulfolobus* growth. Spot-on-lawn assay of SSV10 B205 deletion mutants. Top, The halo morphology and infectivity of the DAG593 (+) control is consistent from 72 to 120 h post transformation (h.p.t.). (A) DAG593 (+) control 72 (h.p.t.); (B) DAG593 (+) control 96 (h.p.t.); (C) DAG593 (+) control 120 (h.p.t.). Bottom, Halos of SSV10 lacking ORF B205 are consistent with the (+) control at 72 h.p.t. (D). However, infectivity is greatly reduced at 96 h.p.t. (E), and completely lost by 120 h.p.t. (F).
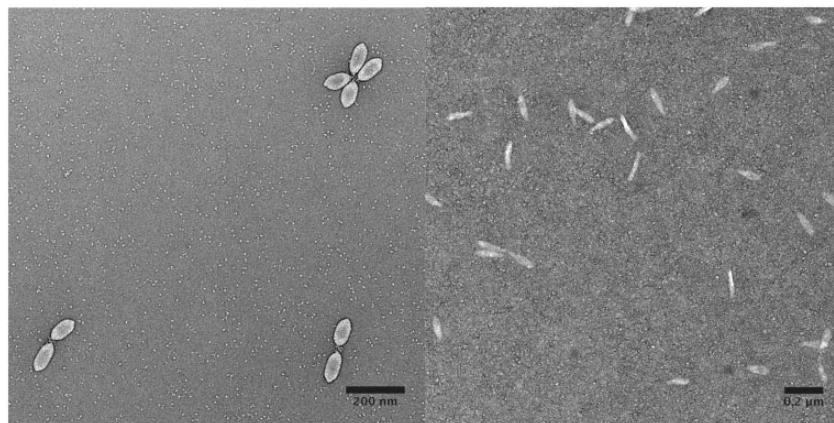
Figure 4. Electron micrographs of SSV10 and SSV10_ΔVP3. Left, Transmission electron micrograph of typical negatively stained SSV10 virus particles. Right, Transmission electron micrograph of SSV10 virus particles lacking the minor capsid gene VP3 (DAG593__ΔVP3). They are morphologically distinct from wild-type particles and have similar morphology to ΔVP3 mutants in SSV1 (Iverson et al. 2017a).

By contrast, a deletion of ORF B205 (ΔB205), generates a functional mutant SSV10. ORF B205 is thus not strictly required for SSV10 infection. In two separate trials, cultures actively infected by DAG821 stopped producing halos on lawns of uninfected cells after about 5 days post transformation, indicating that previously infected cells are eventually capable of controlling an infection or lowering the copy number of ΔB205. Yet these seemingly 'cured' cultures still contain virus sequences and were susceptible to a secondary infection by both SSV1 and SSV10. The gradual loss of function phenotype exhibited by the deletion mutant has not been observed in other SSV10 mutants and is likely caused by the absence of ORF B205.

The laboratory host *S. solfataricus* S441 used in this study is highly similar to *S. solfataricus* P2 (GenBank ID: AE006641), which has two significant matches to SSV10 in CRISPR array #11. The first of the two spacers matching SSV10 is a perfect 23 base pair match [5′-AGGTGAGCTGAATGGCTAAGAAG-3′] to the C-terminal/N-Terminal crossover region of SSV10 ORFs A89 and B150, respectively. The second spacer is 31 base pairs with two mismatches [5′-TATGGGGGtCAGTgACACTTATTCCACCGTT-3′] matching the middle of SSV10 ORF A166. Despite the presence of these spacers in *S. solfataricus* P2, SSV10 is still able to sustain an infection after transformation. Unfortunately, it is not known whether or not these CRISPR spacers are present in host strain S441 used in these studies. Regardless, it has been reported that the presence of a matching spacer does not necessarily dictate the clearing of an infecting agent in *Sulfolobus*, nor does an SSV infection automatically prompt the acquisition of new spacers (Redder et al. 2009; Garrett et al. 2011, 2015; Erdmann et al. 2014; Fusco et al. 2015a). Characterizing the infection cycle and host range of SSV10 ΔB205 will further illuminate the nature of the genetic arms race between the *Fuselloviridae* and its hosts and may lead to the unraveling of a novel viral anti-CRISPR defense mechanism in the archaea (Fig. 4).
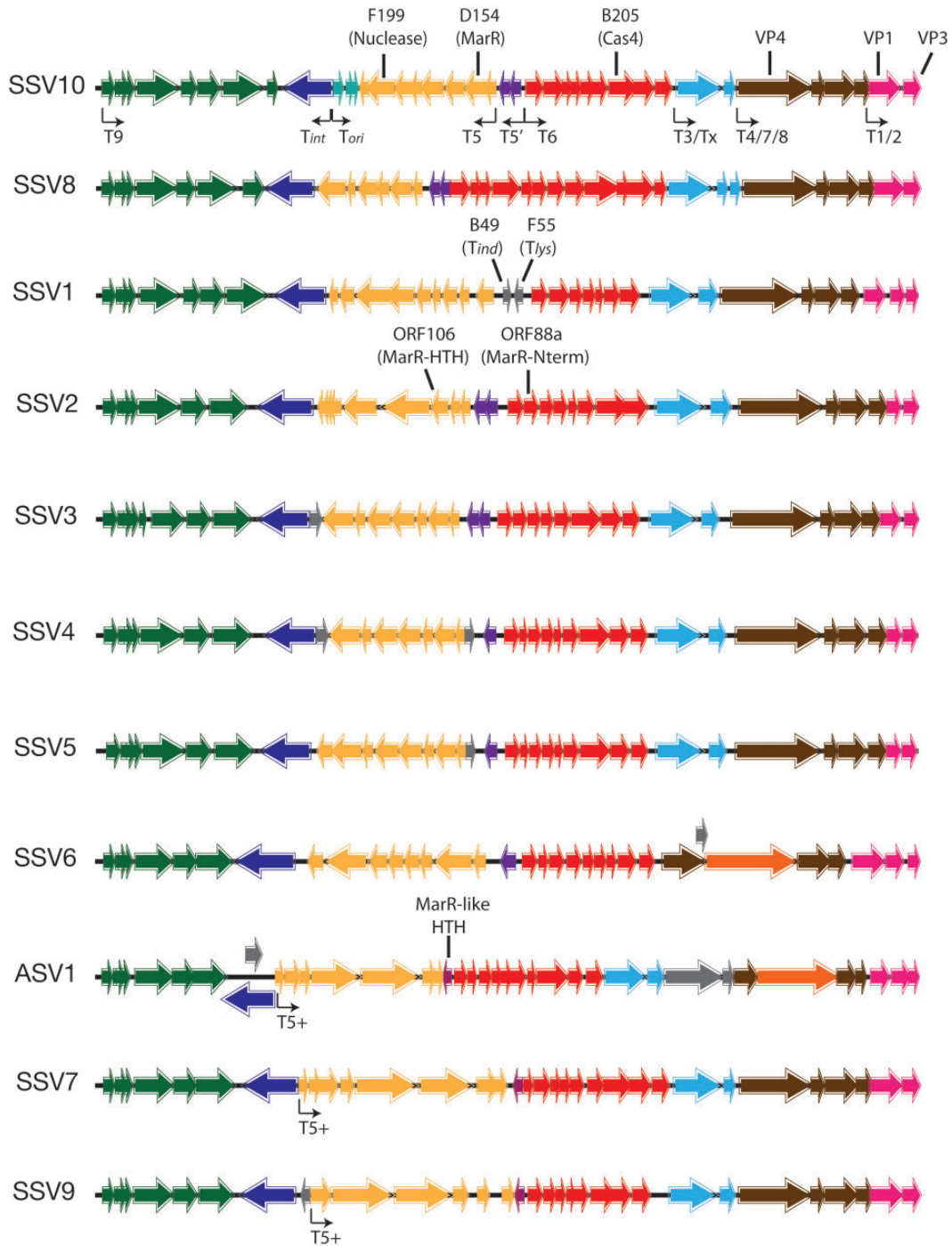
Figure 5. Comparison of Fusellovirus genomes used in this study. Genomes are grouped by similarity of genomic structure and linearized starting at the first nucleotide following VP3, or VP2 in the case of SSV1, SSV6, and ASV1. ORFs in each genome are shown as outlined arrows and are colorcoded by transcript. Promoters are indicated as bent arrows below each genome and labeled as indicated in the SSV10 genome. Structural proteins and coding regions discussed in this study are labeled above each corresponding genome. Genomes encoding a positive-strand T5 transcript (T5+), ASV1, SSV7, and SSV9 are labeled. ORFs in orange (ASV1, SSV6) indicate variant putative terminal fiber modules. ORFs in gray are those outside of the labeled transcripts.

3.5 Mutations of the minor capsid gene VP3

Two different SSV10::Tn5 mutants located in the minor capsid protein VP3 gene (DAG681 and DAG787; Table 3) were tested for functionality due to their near-identical location to Tn5 insertions in SSV1 (Iverson et al. 2017a). Both mutants were found to be nonfunctional in 6 independent transformations (Fig. 1; Table 3). Since a deletion of the SSV1 VP3 gene was active, a similar deletion was made in SSV10. In three of four independent trials, SSV10 lacking VP3 are infectious, similar to SSV1 VP3 deletion mutants. SSV10 lacking VP3 has elongated virions on TEM analysis (Fig. 5), similar to deletions of VP3 in SSV1. Insertion mutants in the SSV1 VP3 gene are excised

together with the gene in vivo, presumably by homologous recombination via direct repeats located in the C-terminal ends of VP1 and VP3 (Iverson et al. 2017a). However, Tn5 insertions in SSV10 VP3 are maintained, probably due to a lack of direct repeats in SSV10, and lead to a complete loss of infectivity (Fig. 1). Why an insertion in VP3 is not tolerated while a deletion is tolerated is not clear. There are no obvious polar effects on other ORFs. Probably a partial, defective VP3 protein is formed that disrupts assembly of mature virions, which could explain evolutionary pressure to delete insertions in SSV1 VP3 (Iverson et al. 2017a).

3.6 Synteny of partially conserved ORFs

The presence, absence, and location of not just single ORFs, but of suites of ORFs helps to decipher the complicated genetic history of fuselloviruses. For example, SSV10 ORFs F73 and D60, encoded on the putative T5′ transcript (Fig. 1), are accompanied by a promoter upstream from the F73 start codon. A homologous coding region along with highly conserved promoter elements is present in seven of eleven *Fuselloviridae* (Fig. 5). Microarray analysis of the SSV2 infection cycle supports this annotation, showing upregulation of these homologues in tandem ~8–9 h post infection (Ren et al. 2013). There are no homologues of these ORFs in SSV1, and if this transcript or any part of it is maintained in some way upstream of the T5+ transcript of ASV1, SSV7, and SSV9, the corresponding coding region is not obvious.

The T5/T5' or T5+ portion of an ancestral fusellovirus genome is predicted to have undergone an inversion or deletion event, made possible either via tandem integration or intragenome recombination (Redder et al. 2009) via repeat regions located on either end of the T5+ transcript. The T5+ transcript, only present in ASV1, SSV7, and SSV9, seems to lack ORFs predicted to be mediators of the infection process associated with the 5′-end of the T5 transcripts in other SSVs. SSV9 is known to have a very wide host range (Ceballos et al. 2012) and SSV9 appears to induce a dormant or apoptotic state in infected host cells (Bautista et al. 2015) rather than establishing a stable carrier state similar to other SSVs. ASV and SSV7 have not been studied in depth, but ASV infects *Acidianus* not *Sulfolobus*, a further indication of host range changes due to rearrangement.

3.7 MarR-like ORFs in fusellovirus genomes

Multiple antibiotic resistance regulators, or MarR proteins, are a family of transcription factors found in bacteria and archaea involved in coordinating cellular responses to external biotic and abiotic stresses (Zhu et al. 2017). In *Escherichia coli*, this is accomplished by MarR binding to the *marO* promoter, negatively regulating the *marRAB* operon (Alekshun and Levy 1997). Alternatively, the crystal structure and DNA binding mechanisms of a *S. solfataricus* MarR homologue *BldR* (Fiorentino et al. 2007; Di Fiore et al. 2009) have been demonstrated experimentally, implicating *BldR* as one of the few members of the MarR family proteins to act as a transcriptional activator rather than a repressor. A second homologue from *S. solfataricus, BldR2* (Fiorentino et al. 2011), has also been identified and shown to bind specifically to its own promoter, implicating it as a repressor more akin to *marRAB* repression in *E. coli*. SSV10 ORF D154, an MarR-like transcriptional regulator encoded in the T5 transcript, is a cryptically conserved ORF originally thought to be encoded by only a few *Fuselloviridae*. Extensive sequence analyses of SSV10 ORF D154, however, provide another indication

that ASV1, SSV7, and SSV9 are related and different from other SSVs and reveal that only SSV6 lacks a MarR-like ORF. SSV6 ORF F90 does encode a winged-helix that may be distantly derived from MarR-like regulators similar to SSV1 ORF F93 (Kraft et al. 2004b; Kraft et al. 2005). The MarR-like ORFs encoded by ASV1, SSV7, and SSV9 are also truncated, missing between 50 and 70 amino acids at the N-terminal end relative to other SSVs. These truncated ORFs are present at a genetically distinct location between the T5+ and T6 transcripts, similar to the position of the non-conserved T_{lys} transcript in SSV1 (Fig. 5). Strangely, while the promoters of the MarR-like ORF are unique and conserved between ASV1 and SSV9, the promoter upstream of the SSV7 putative MarR-like ORF is nearly identical to the T_{lys} promoter of SSV1. SSV2 encodes a seemingly bipartite version of the full-length MarR-like ORF made up of ORF 88a and ORF 106. SSV2 ORF 88a contains a conserved motif associated with the N-terminal end of the MarR-like ORF. SSV2 ORF 106, on the other hand, is homologous to the C-terminal end of the MarR-like ORF. Moreover, both ORFs contain two 8-base pair direct repeats which may be indicative of a split or would allow for them to become concatenated should a junction form. Transcript analysis of SSV2 shows that ORFs 106 and 88a, separated by ~1,200 base pairs are transcribed together between 2 and 3 h post infection (Ren et al. 2013). Curiously, a homologue of SSV2 ORF88a is found in SSV3, which itself already encodes a full length MarR-like gene. SSV10 ORF D154 and its homologues are further evidence allowing the ancestral fusellovirus lineage to be traced.

3.8 Sequence analysis reveals conserved promoter elements in the *Fuselloviridae*

SSV10 contains 10 putative promoters and corresponding transcripts (Table 4). For all SSVs, there are 'core' transcripts corresponding to the SSV1 and SSV10 transcripts T1/2, T3, T4/7/8, and T9 (Fig. 5). The promoters of these 'core' transcripts are extremely well conserved, sharing a consensus TATA box sequence of TTT[WW]AAA, with the only deviation in the SSV6 T1/2 promoter having a thymine instead of adenine at the seventh position. Conversely, the TFB-recognition elements of the T1/2 transcripts seem distinct from that of the other 'core' transcripts. The BREs of all four core transcripts share a consensus sequence of DRGSSS. However, the conservation is much greater when just the T3, T4/7/8, and T9 BREs are compared, sharing a consensus sequence of a AGGCCC, while eight out of eleven of the T1/2 BREs share a consensus sequence of SRGGGG (Fig. 6a).

SSV10 also appears to contain alternative transcripts relative to SSV1, including a unique positive strand transcript T*ori*, and transcript T5′ which is shared by six other SSVs (Fig. 5). Analysis of potential transcription start sites and the divergent transcript in SSV10 (Figs. 1, 5, and 6) indicates that all SSVs except SSV1 (Reiter et al. 1987; Fröls et al. 2007a) encode their viral integrases on a separate transcript. The promoter of this T*int* transcript is not as well conserved as the core promoters but does also exhibit a eukaryal-like TATA box with a consensus sequence of TWTTTAAC (Fig. 6b). In the case of SSV2 where the temporal regulation of transcripts differs greatly in respect to SSV1, the integrase gene was expressed late (~6 h.p.i) and likely independently of the rest of the T5 transcript (Ren et al. 2013).

The transcript Tx (Fröls et al. 2007a) (Figs. 1, 5, and 6) that was not originally annotated in SSV1 (Reiter et al. 1987) appears to be conserved in all but one SSV. The promoter elements associated with the Tx transcript are highly conserved in every fusellovirus, sharing a consensus sequence of **AAAATT**TTTTAAAC, with the
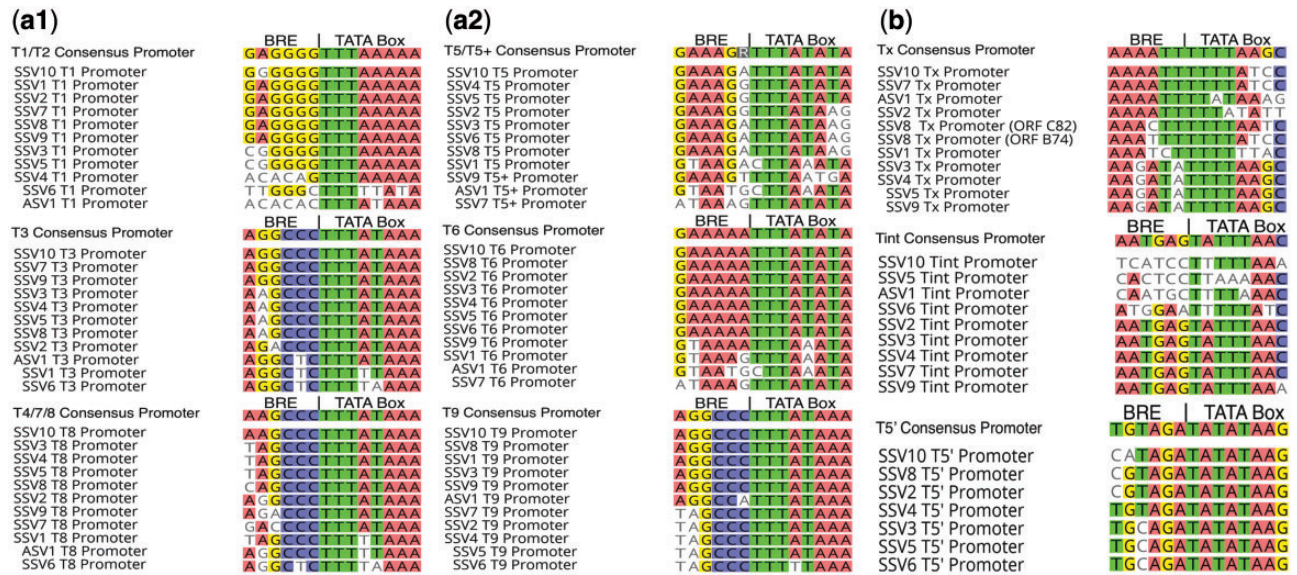
Figure 6. (a-1) and (a-2) Conservation among canonical *Fuselloviridae* promoters. Alignment of putative Fusellovirus promoter elements derived from known SSV1 transcripts and found in all other SSVs; the conserved 'core' transcript promoters are typified by their high GC-content BREs, while the T5 and T6 transcript BREs utilize a non-canonical poly-A motif. (b) Conservation among non-canonical *Fuselloviridae* promoters. Alignment of non-canonical promoters found in some, but not all, Fusellovirus genomes; these are typified by their irregular TFB recognition elements and divergence from the TTTAWAWA consensus TATA-box of the core transcripts.

exception of SSV6 (Fig. 6b). Given this promoter conservation it is likely that the non-conserved ORF directly upstream of the putative tail fiber gene, VP4, is encoded on a monocistronic transcript similar to Tx of SSV1 for all members of the *Fuselloviridae* and probably regulated similarly (Fig. 5). This region in SSV8 appears to have undergone a gene duplication event. SSV8 ORFs B74 and C82, located between the putative T3 and T4/7/8 transcripts, are preceded by highly similar promoter elements. Moreover, the ORFs themselves both encode conserved motifs shared by other C124 homologues.

SSV10 also encodes two ORFs—F73 and D60—on a short transcript upstream of the T5 transcript, herein referred to as T5′. The T5′ transcript is found in six other SSVs, the products of which were shown to be the last to be upregulated in SSV2 9 h post infection of *S. solfataricus* P2 (Ren et al. 2013) (Fig. 2). The promoters of the T5′ transcripts share a highly conserved TATA box, with a consensus sequence of TATATAAG, akin to eukaryal TATA boxes identified in *Saccharomyces cerevisiae* (Basehoar et al. 2004) (Fig. 6b).

The T_{lys} transcript that has been reported to be important for maintenance of the carrier state of SSV1 (Fusco et al. 2013, 2015b) is not conserved. The F55 product of the T_{lys} transcript and its promoter sequence are unique to SSV1. SSV10 lacks a similar transcript, instead encoding ORF E56—a conserved, homologous coding region present but unannotated in 6 other SSVs—which shares no similarity in sequence, promoter elements, or structure with the predicted copG-like transcriptional regulator F55 of SSV1 (Fusco et al. 2013). In ASV1, SSV7, and SSV9, this coding region instead corresponds to a MarR-like transcriptional regulator similar to those encoded in the T5 transcript of all other SSVs.

Promoters for the T5 and T6 'early' transcripts (Fröls et al. 2007a) are very well conserved, particularly in their TATA boxes which are identical to TATA boxes in 'core' promoters, indicating that they are probably regulated similarly in all SSVs (Fig. 6). Variations between promoters in different SSVs seem to be between putative BREs. Previous studies (Qureshi and Jackson

1998; Peng et al. 2012) have shown that TFB-recognition elements can dictate the efficacy of a given promoter and, in addition to transcriptional regulators, this may be a mechanism by which SSVs regulate their gene expression. Unfortunately, no transcript mapping has been done other than for SSV1 (Reiter et al. 1987; Fröls et al. 2007a; Fusco et al. 2013) and SSV2 (Ren et al. 2013) so all other promoter annotations are speculative. Three of the *Fuselloviridae*—ASV1, SSV7, and SSV9—lack a polycistronic T5-like negative strand transcript, instead have a putative positive strand transcript, referred to herein as T5+, with large ORFs in the corresponding part of the genome (Fig. 5).

The Tori transcript in SSV10 does not have a conserved BRE or TATA box, but ORF A49 contains a predicted large stem-loop secondary structure which could be the site for replication initiation. Both functional and nonfunctional Tn5 mutants interrupting ORF A49 were generated (Fig. 1), and the location of the nonfunctional mutant may provide evidence of a specific origin site. This differs from the predicted origin of replication in SSV1 (Cannio et al. 1998; Fröls et al. 2007a; Iverson et al. 2017a). However, no SSV origins have been functionally characterized to date.

3.9 Predicted structures of SSV10 proteins

Archaeal viruses are well-known to contain ORFs that have no sequence similarity to known proteins making functional predictions challenging (Krupovic et al. 2018). However, protein structure, particularly of viruses, is often conserved even in the absence of obvious sequence similarity (Sinclair et al. 2017). In order to predict function, all SSV10 ORFs were translated and structural predictions were generated using the Phyre2 web portal for protein modeling (Kelley et al. 2015). The 40 ORF structural predictions were partitioned into three groups—low ($\leq$44), middle (45–84), and high-confidence ($\geq$85)—based on the confidence rating generated by Phyre2 for a significant portion (30%) of the query (Table 5). Twelve of the forty annotated ORFs in SSV10 returned no significant structural matches of any kind.

Of the matches, twelve were in the low confidence bracket, four were in the middle bracket, and twelve were in the high-confidence bracket.

All of the high-confidence matches are to well-described structures, both from and beyond the *Fuselloviridae*. Homologues of both SSV10 ORF C250, a predicted DnaA-like AAA+ ATPase, and the integrase gene have been characterized previously experimentally (Koonin 1992; Zhan et al. 2015; Clore and Stedman 2007; Iverson et al. 2017a). SSV10 ORFs F199, F64, E62, F73, B104, and C127 are all predicted matches to high-resolution structures of homologous proteins from SSV1 (D63 (Kraft et al. 2004a), A100, and B129 (Lawrence et al. 2009)) or SSV8 (D212 (Menon et al. 2010) and E73 (Schlenker et al. 2012)). These ORFs are all found in the 'early' (Fröls et al. 2007a) T5, T6, or T5′ transcripts and, with the exception of the viral nuclease, are predicted to play a role in regulating transcription and translation of viral proteins (Fig. 1). Additionally, uncharacterized SSV10 ORFs B65 and A81 encode a predicted copG-like and C2H2 zinc finger-like transcriptional regulators, respectively, and are also encoded on the T6 transcript. SSV10 ORF D154 in the T5 transcript and C112 in the T8 transcript encode MarR-like and AsrR-like transcriptional regulators respectively, similar to those encoded by host cells and are likely responsible for mediating the host response to infection (Di Fiore et al. 2009; Fiorentino et al. 2011; Zhu et al. 2017). SSV10 ORF B205 encodes a Cas4-like protein product that has perfect confidence rating match over 95 per cent of the sequence to a monomer of the *S. solfataricus* Cas4 gene SSO0001. This viral Cas4 is implicated in host interference as a possible impediment to the host CRISPR/Cas response, and potentially involved in stress response and DNA repair (Hudaiberdiev et al. 2017).

Finally, SSV10 encodes eleven ORFs which contain predicted transmembrane helices. Four of these ORFs are encoded on the T3, Tx, and T8 'late' transcripts, while seven other transmembrane helix-containing ORFs are encoded on putative 'early' transcripts commonly associated with genome replication (Fig. 1). Unfortunately, there are no identifiable homologues associated with these ORFs. However, since it would seem that Sulfolobus viruses are forming virus factories at the cellular membrane (Quemin and Quax 2015; Quemin et al. 2016; Martínez-Alvarez et al. 2017), it is likely that these ORFs are involved in the rearrangement of and anchoring to said membrane in order to facilitate the production and egress of mature virions.

4 Conclusions

This work characterizes SSV10 as a novel member of the *Fuselloviridae* family, and provides a point of comparison, as its genome shares characteristics of and homology with members of the *Fuselloviridae* that are absent in SSV1. SSV10 is genetically tractable, withstanding both insertion and deletion mutations in a fashion consistent with SSV1 (Iverson et al. 2017a). SSV10 encodes a Cas4-like gene—ORF B205—deletions of which lead to a loss of infectivity over time. These experiments are the first to show a direct correlation between the Cas4-like gene present in the *Fuselloviridae* and the ability of the virus to maintain a persistent infection. SSV10 also tolerates deletion mutations of the minor capsid gene VP3 leading to aberrant, elongated morphology of the virions produced, similar to SSV1. Recently proposed structural models of hyperthermophilic archaeal viruses *Aeropyrum pernix* bacilliform virus 1 (APBV1) (Ptchelkine et al. 2017) and *Acidianus* tailed spindle virus (Hochstein et al. 2018) may provide insight into the morphological perturbations of VP3 deletions in SSV1 and SSV10. Overall, the genetic

tractability of SSV10 as well as its differential relationships to members of the *Fuselloviridae* compared to SSV1 make it an attractive model for understanding the diverse life cycles and virus–host interactions of these viruses.

Data availability

The sequence of SSV10 is available in Genbank as accession number KY563228.

Supplementary data

Supplementary data are available at *Virus Evolution* online.

Conflict of interest: None declared.

References

Alekshun, M. N., and Levy, S. B. (1997) 'Regulation of Chromosomally Mediated Multiple Antibiotic Resistance: The *Mar* Regulon', *Antimicrobial Agents and Chemotherapy*, 41: 2067–75.

Altschul, S. F. et al. (1997) 'Gapped BLAST and PSI-BLAST: A New Generation of Protein Database Search Programs', *Nucleic Acids Research*, 25: 3389–402.

Barry, E. R., and Bell, S. D. (2006) 'DNA Replication in the Archaea', *Microbiology and Molecular Biology Reviews*, 70: 876–87.

Basehoar, A. D., Zanton, S. J., and Pugh, B. F. (2004) 'Identification and Distinct Regulation of Yeast TATA Box-Containing Genes', *Cell*, 116: 699–709.

Bautista, M. A., Zhang, C., and Whitaker, R. J. (2015) 'Virus-Induced Dormancy in the Archaeon *Sulfolobus islandicus*', *mBio*, 6: e02565-14-8.

Bize, A. et al. (2009) 'A Unique Virus Release Mechanism in the Archaea', *Proceedings of the National Academy of Sciences USA*, 106: 11306–11.

Bondy-Denomy, J. et al. (2012) 'Bacteriophage Genes That Inactivate the CRISPR/Cas Bacterial Immune System', *Nature*, 493: 429–32.

—— et al. (2015) 'Multiple Mechanisms for CRISPR–Cas Inhibition by anti–CRISPR Proteins', *Nature*, 526: 136–9.

Cannio, R. et al. (1998) 'An Autonomously Replicating Transformation Vector for *Sulfolobus solfataricus*', *Journal of Bacteriology*, 180: 3237–40.

Ceballos, R. M. et al. (2012) 'Differential Virus Host-Ranges of the Fuselloviridae of Hyperthermophilic Archaea: Implications for Evolution in Extreme Environments', *Frontiers in Microbiology*, 3: 1–10.

Chan, P. P. et al. (2012) 'The UCSC Archaeal Genome Browser: 2012 Update', *Nucleic Acids Research*, 40: D646–52.

Chen, L. et al. (2005) 'The Genome of *Sulfolobus acidocaldarius*, a Model Organism of the Crenarchaeota the Genome of *Sulfolobus acidocaldarius*, a Model Organism of the Crenarchaeota', *Journal of Bacteriology*, 187: 4992–9.

Chowdhury, S. et al. (2017) 'Structure Reveals Mechanisms of Viral Suppressors That Intercept a CRISPR RNA-Guided Surveillance Complex', *Cell*, 169: 47–57.e11.

Clore, A. J. (2008) 'The Family Fuselloviridae: Diversity and Replication of a Hyperthermic Virus Infecting the Archaeon Genus Sulfolobus (Doctoral Dissertation)', *Portland State University*

——, and Stedman, K. M. (2007) 'The SSV1 Viral Integrase Is Not Essential', *Virology*, 361: 103–11.

Duggin, I. G., and Bell, S. D. (2006) 'The Chromosome Replication Machinery of the Archaeon *Sulfolobus solfataricus*', *Journal of Biological Chemistry*, 281: 15029–32.

Eilers, B. J., Young, M. J., and Lawrence, C. M. (2012) 'The Structure of an Archaeal Viral Integrase Reveals an Evolutionarily Conserved Catalytic Core yet Supports a Mechanism of DNA Cleavage in Trans', *Journal of Virology*, 86: 8309–13.

Erdmann, S., Le Moine Bauer, S., and Garrett, R. A. (2014) 'Inter-Viral Conflicts That Exploit Host CRISPR Immune Systems of Sulfolobus', *Molecular Microbiology*, 91: 900–17.

Di Fiore, A. et al. (2009) 'Structural Analysis of BldR from *Sulfolobus solfataricus* Provides Insights into the Molecular Basis of Transcriptional Activation in Archaea by MarR Family Proteins', *Journal of Molecular Biology*, 388: 559–69.

Fiorentino, G. et al. (2011) 'Identification and Physicochemical Characterization of BldR2 from *Sulfolobus solfataricus*, a Novel Archaeal Member of the MarR Transcription Factor Family', *Biochemistry*, 50: 6607–21.

—— et al. (2007) 'MarR-like Transcriptional Regulator Involved in Detoxification of Aromatic Compounds in *Sulfolobus solfataricus*', *Journal of Bacteriology*, 189: 7351–60.

Fröls, S. et al. (2007a) 'Response of the Hyperthermophilic Archaeon *Sulfolobus solfataricus* to UV Damage', *Journal of Bacteriology*, 189: 8708–18.

—— et al. (2007b) 'Elucidating the Transcription Cycle of the UV-Inducible Hyperthermophilic Archaeal Virus SSV1 by DNA Microarrays', *Virology*, 365: 48–59.

Fusco, S. et al. (2013) 'T(Lys), a Newly Identified Sulfolobus Spindle-Shaped Virus 1 Transcript Expressed in the Lysogenic State, Encodes a DNA-Binding Protein Interacting at the Promoters of the Early Genes', *Journal of Virology*, 87: 5926–36.

—— et al. (2015a) 'Transcriptome Analysis of *Sulfolobus solfataricus* Infected with Two Related Fuselloviruses Reveals Novel Insights into the Regulation of CRISPR-Cas System', *Biochimie*, 118: 322–32.

—— et al. (2015b) 'Unravelling the Role of the F55 Regulator in the Transition from Lysogeny to UV Induction of Sulfolobus Spindle-Shaped Virus 1', *Journal of Virology*, 89: 6453–61.

Garrett, R. A. et al. (2015) 'CRISPR-Cas Adaptive Immune Systems of the Sulfolobales: Unravelling Their Complexity and Diversity', *Life*, 5: 783–817.

——, Vestergaard, G., and Shah, S. A. (2011) 'Archaeal CRISPR-Based Immune Systems: Exchangeable Functional Modules', *Trends in Microbiology*, 19: 549–56. Elsevier Ltd.

Green, M. R., and Sambrook, J. (2012). *Molecular Cloning: A Laboratory Manual*, 4th edn. New York: Cold Spring Harbor Laboratory Press (June 15, 2012). <http://www.amazon.com/Molecular-Cloning-Laboratory-Edition-Three/dp/1936113422>.

Gudbergsdottir, S. et al. (2011) 'Dynamic Properties of the Sulfolobus CRISPR/Cas and CRISPR/Cmr Systems When Challenged with Vector-Borne Viral and Plasmid Genes and Protospacers', *Molecular Microbiology*, 79: 35–49.

He, F. et al. (2016) 'CRISPR-Cas Type I-a Cascade Complex Couples Viral Infection Surveillance to Host Transcriptional Regulation in the Dependence of Csa3b', *Nucleic Acids Research*, 45: 1902–13.

Hochstein, R. et al. (2018) 'Structural Studies of Acidianus Tailed Spindle Virus Reveal a Structural Paradigm Used in the Assembly of Spindle-Shaped Viruses', *Proceedings of the National Academy of Sciences*, 115: 2120–5.

Hooton, S. P. T., and Connerton, I. F. (2015) '*Campylobacter jejuni* Acquire New Host-Derived CRISPR Spacers When in Association with Bacteriophages Harboring a CRISPR-like Cas4 Protein', *Frontiers in Microbiology*, 5: 1–9.

Hudaiberdiev, S. et al. (2017) 'Phylogenomics of Cas4 Family Nucleases', *BMC Evolutionary Biology*, 17:

Iverson, E. A. et al. (2017a) 'Extreme Mutation Tolerance: Nearly Half of the Archaeal Fusellovirus Sulfolobus Spindle-Shaped Virus 1 Genes Are Not Required for Virus Function, Including the Minor Capsid Protein Gene vp3', *Journal of Virology*, 91: e02406-16.

—— et al. (2017b) 'Genetic Analysis of the Major Capsid Protein of the Archaeal Fusellovirus SSV1: Mutational Flexibility and Conformational Change', *Genes*, 8: 373.

——, and Stedman, K. M. (2012) 'A Genetic Study of SSV1, the Prototypical Fusellovirus', *Frontiers in Microbiology*, 3: 200.

Kelley, L. A. et al. (2015) 'The Phyre2 Web Portal for Protein Modelling, Prediction and Analysis', *Nature Protocols*, 10: 845–58.

Kent, W. J. (2002) 'BLAT—the BLAST -like Alignment Tool', *Genome Research*, 12: 656–64.

Koonin, E. V. (1992) 'Archaebacterial Virus SSV1 Encodes a Putative DNA A-like Protein', *Nucleic Acids Research*, 20: 1143.

Kosa, P. F. et al. (1997) 'Notes: The 2. 1-Å Crystal Structure of an Archaeal Preinitiation Complex: TATA-Box-Binding Protein-Transcription Factor (II) B', *Biochemistry*, 94: 6042–7.

Kraft, P. et al. (2005). 'Structural Studies of Crenarchaeal Viral Proteins: Structure Suggests Function', *Geothermal Biology and Geochemistry in Yellowstone National Park*, 305–16.

—— et al. (2004) 'Structure of D-63 from Sulfolobus Spindle-Shaped Virus 1: Surface Properties of the Dimeric Four-Helix Bundle Suggest an Adaptor Protein Function', *Journal of Virology*, 78: 7438–42.

—— et al. (2004) 'Crystal Structure of F-93 from Sulfolobus Spindle-Shaped Virus 1, a Winged-Helix DNA Binding Protein', *Journal of Virology*, 78: 11544–50.

Krupovic, M. et al. (2015) 'Evolution of an Archaeal Virus Nucleocapsid Protein from the CRISPR-Associated Cas4 Nuclease', *Biology Direct*, 10: 65.Biology Direct.

—— et al. (2018) 'Viruses of Archaea: Structural, Functional, Environmental and Evolutionary Genomics', *Virus Research*, 244: 181–93.

Lawrence, C. M. et al. (2009) 'Structural and Functional Studies of Archaeal Viruses', *Journal of Biological Chemistry*, 284: 12599–603.

Lemak, S. et al. (2013) 'Toroidal Structure and DNA Cleavage by the CRISPR-Associated [4Fe-4S] Cluster Containing Cas4 Nuclease SSO0001 from *Sulfolobus solfataricus*', *Journal of the American Chemical Society*, 135: 17476–87.

Liu, G., She, Q., and Garrett, R. A. (2016) 'Diverse CRISPR-Cas Responses and Dramatic Cellular DNA Changes and Cell Death in pKEF9-Conjugated Sulfolobus Species', *Nucleic Acids Research*, 44: 4233–42.

Liu, T. et al. (2017) 'Coupling Transcriptional Activation of CRISPR–Cas System and DNA Repair Genes by Csa3a in *Sulfolobus islandicus*', *Nucleic Acids Research*, 45: 8978–92. Oxford University Press.

Manica, A. et al. (2013) 'Unexpectedly Broad Target Recognition of the CRISPR-Mediated Virus Defence System in the Archaeon *Sulfolobus solfataricus*', *Nucleic Acids Research*, 41: 10509–17.

—— et al. (2011) 'In Vivo Activity of CRISPR-Mediated Virus Defence in a Hyperthermophilic Archaeon', *Molecular Microbiology*, 80: 481–91.

Martin, A. et al. (1984) 'SAV 1, a Temperate U.V.-Inducible DNA Virus-like Particle from the Archaebacterium *Sulfolobus acidocaldarius* Isolate B12', *The EMBO Journal*, 3: 2165–8.

Martínez-Alvarez, L., Deng, L., and Peng, X. (2017) 'Formation of a Viral Replication Focus in Sulfolobus Cells Infected by the Rudivirus *Sulfolobus islandicus* Rod-Shaped Virus 2', *Journal of Virology*, 91: e00486-17–12.

Menon, S. K. et al. (2010) 'The Crystal Structure of D212 from Sulfolobus Spindle-Shaped Virus Ragged Hills Reveals a New Member of the PD-(D/E)XK Nuclease Superfamily', *Journal of Virology*, 84: 5890–7.

Muskhelishvili, G., Palm, P., and Zillig, W. (1993) 'SSV1-Encoded Site-Specific Recombination System in *Sulfolobus shibatae*', *Molecular & General Genetics*, 237: 334–42.

Nakagawa, S., Niimura, Y., and Gojobori, T. (2017) 'Comparative Genomic Analysis of Translation Initiation Mechanisms for Genes Lacking the Shine–Dalgarno Sequence in Prokaryotes', *Nucleic Acids Research*, 45: 3922–31.

Palm, P. et al. (1991) 'Complete Nucleotide Sequence of the Virus SSV1 of the Archaebacterium *Sulfolobus shibatae*', *Virology*, 185: 242–50.

Pawluk, A. et al. (2014) 'A New Group of Phage anti-CRISPR Genes Inhibits the Type I-E CRISPR-Cas System of *Pseudomonas aeruginosa*', *mBio*, 5: e00896-14–7.

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*, 154: 383–91.

Peng, N. et al. (2012) 'A Synthetic Arabinose-Inducible Promoter Confers High Levels of Recombinant Protein Expression in Hyperthermophilic Archaeon *Sulfolobus islandicus*', *Applied and Environmental Microbiology*, 78: 5630–7.

Plagens, A. et al. (2012) 'Characterization of the CRISPR/Cas Subtype I-a System of the Hyperthermophilic Crenarchaeon *Thermoproteus tenax*', *Journal of Bacteriology*, 194: 2491–500.

Prangishvili, D. (2013) 'The Wonderful World of Archaeal Viruses', *Annual Review of Microbiology*, 67: 565–85.

——, Forterre, P., and Garrett, R. A. (2006) 'Viruses of the Archaea: A Unifying View', *Nature Reviews Microbiology*, 4: 837–48.

Ptchelkine, D. et al. (2017) 'Unique Architecture of Thermophilic Archaeal Virus APBV1 and Its Genome Packaging', *Nature Communications*, 8: 7–12.Springer US.

Quemin, E. R. J. et al. (2016) 'Eukaryotic-like Virus Budding in Archaea', *mBio*, 7: e01439-16–7.

—— et al. (2015) 'Sulfolobus Spindle-Shaped Virus 1 Contains Glycosylated Capsid Proteins, a Cellular Chromatin Protein, and Host-Derived Lipids', *Journal of Virology*, 89: 11681–91.

——, and Quax, T. E. F. (2015) 'Archaeal Viruses at the Cell Envelope: Entry and Egress', *Frontiers in Microbiology*, 6: 552.

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', *Molecular Cell*, 1: 389–400.

Redder, P. et al. (2009) 'Four Newly Isolated Fuselloviruses from Extreme Geothermal Environments Reveal Unusual Morphologies and a Possible Interviral Recombination Mechanism', *Environmental Microbiology*, 11: 2849–62.

Reiter, W.-D., Palm, P., and Yeats, S. (1989) 'Transfer RNA Genes Frequently Serve as Integration Sites for Prokaryotic Genetic Elements', *Nucleic Acids Research*, 17: 1907–14.

—— et al. (1987) 'Gene Expression in Archaebacteria: Physical Mapping of Constitutive and UV-Inducible Transcripts from the Sulfolobus Virus-like Particle SSV1', *Molecular & General Genetics*, 209: 270–5.

——, Palm, P., and Zillig, W. (1988) 'Analysis of Transcription in the Archaebacterium Sulfolobus Indicates That Archaebacterial Promoters Are Homologous to Eukaryotic Pol II Promoters', *Nucleic Acids Research*, 16: 1.

Ren, Y., She, Q., and Huang, L. (2013) 'Transcriptomic Analysis of the SSV2 Infection of *Sulfolobus solfataricus* with and without the Integrative Plasmid pSSVi', *Virology*, 441: 126–34. Elsevier.

Rensen, E. I. et al. (2016) 'A Virus of Hyperthermophilic Archaea with a Unique Architecture among DNA Viruses', *Proceedings of the National Academy of Sciences*, 113: 2478–83.

Romero, A., and García, P. (1991) 'Initiation of Translation at AUC, AUA and AUU Codons in *Escherichia coli*', *FEMS Microbiology Letters*, 68: 325–30.

Samson, R. Y. et al. (2013) 'Specificity and Function of Archaeal DNA Replication Initiator Proteins', *Cell Reports*, 3: 485–96. The Authors.

Schlenker, C. et al. (2012) 'Structural Studies of E73 from a Hyperthermophilic Archaeal Virus Identify the "RH3" Domain, an Elaborated Ribbon-Helix-Helix Motif Involved in DNA Recognition', *Biochemistry*, 51: 2899–910.

Schleper, C., Kubo, K., and Zillig, W. (1992) 'The Particle SSV1 from the Extremely Thermophilic Archaeon Sulfolobus Is a Virus: Demonstration of Infectivity and of Transfection with Viral DNA', *Proceedings of the National Academy of Sciences*, 89: 7645–9.

Seed, Kimberly D.. et al. (2013) 'A Bacteriophage Encodes Its Own CRISPR/Cas Adaptive Repsonse to Evade Host Innate Immunity', *Nature*, 494: 489–91.

She, Q. et al. (2001) 'The Complete Genome of the Crenarchaeon *Sulfolobus solfataricus* P2', *Proceedings of the National Academy of Sciences*, 98: 7835–40.

Sinclair, R. M., Ravantti, J. J., and Bamford, D. H. (2017) 'Nucleic and Amino Acid Sequences Support Structure-Based Viral Classification', *Journal of Virology*, 91: e02275-16–13.

Stedman, K. M., Clore, A., and Combet-Blanc, Y. (2006) 'Biogeographical Diversity of Archaeal Viruses', *Prokaryotic Diversity: Mechanisms and Significance: Published for the Society for General Microbiology*, Vol. 9780521869355: p 131–44.

—— et al. (1999) 'Genetic Requirements for the Function of the Archaeal Virus SSV1 in *Sulfolobus solfataricus*: Construction and Testing of Viral Shuttle Vectors', *Genetics*, 152: 1397–405.

—— et al. (2003) 'Relationships between Fuselloviruses Infecting the Extremely Thermophilic Archaeon Sulfolobus: SSV1 and SSV2', *Research in Microbiology*, 154: 295–302.

Stetter, K. O. (2006) 'History of Discovery of the First Hyperthermophiles', *Extremophiles*, 10: 357–62.

Torarinsson, E., Klenk, H. P., and Garrett, R. A. (2005) 'Divergent Transcriptional and Translational Signals in Archaea', *Environmental Microbiology*, 7: 47–54.

Wiedenheft, B. et al. (2004) 'Comparative Genomic Analysis of Hyperthermophilic Archaeal Fuselloviridae Viruses', *Journal of Virology*, 78: 1954–61.

Wu, Z. et al. (2014) 'DNA Replication Origins in Archaea', *Frontiers in Microbiology*, 5: 1–7.

Zhai, B. et al. (2017) 'Structure and Function of a Novel ATPase That Interacts with Holliday Junction Resolvase Hjc and Promotes Branch Migration', *Journal of Molecular Biology*, 429: 1009–29. Elsevier Ltd.

Zhan, Z., Zhou, J., and Huang, L. (2015) 'Site-Specific Recombination by SSV2 Integrase: Substrate Requirement and Domain Functions', *Journal of Virology*, 89: 10934–44.

Zhang, J., Kasciukovic, T., and White, M. F. (2012) 'The CRISPR Associated Protein Cas4 Is a 5′ to 3′ DNA Exonuclease with an Iron-Sulfur Cluster', *PLoS ONE*, 7: e47232.

Zhu, R. et al. (2017) 'Structural Characterization of the DNA-Binding Mechanism Underlying the Copper(II)-Sensing MarR Transcriptional Regulator', *Journal of Biological Inorganic Chemistry*, 22: 685–93.

Zillig, W. et al. (1993) 'Screening for Sulfolobales, Their Plasmids and Their Viruses in Icelandic Solfataras', *Systematic and Applied Microbiology*, 16: 609–28.

—— et al. (1980) 'The Sulfolobus-"Caldariella" Group: Taxonomy on the Basis of the Structure of DNA-Dependent RNA Polymerases', *Archives of Microbiology*, 125: 259–69.